PhD-FDEF-2023-006
The Faculty of Law, Economics and Finance

DISSERTATION

Defence held on 09/03/2023 in Luxembourg
to obtain the degree of

Docteur de l'université du Luxembourg
En Sciences Économiques

by

Alemayehu Demissew TAYE
Born on 29 May 1989 in Hirna (Ethiopia)

Essays on the Prediction and Measurement of Individual Well-being

Dissertation Defence Committee

Dr Conchita D'AMBROSIO, dissertation supervisor
Professor, Université du Luxembourg

Dr Chiara PERONI,
Head, STATEC Research

Dr Chiara BINELLI,
Associate Professor, University of Bologna

Dr Eugenio PELUSO, Chair
Head of Department, Living Conditions, Luxembourg Institute of Socio-Economic Research

Dr Michela BIA, Vice-chair
Research Scientist, Luxembourg Institute of Socio-Economic Research

Acknowledgements

I would like to thank all of the people without whom I would not have been able to write this dissertation.

First, my sincere appreciation goes to my PhD supervisor, Conchita D'Ambrosio. Her continuous guidance, insightful suggestions, and refinements immensely helped the successful completion of this dissertation. She has been patient and kind with her time throughout my PhD journey. Most of all, I appreciate her close supervision and the freedom she provided for me to explore more, allowing me to gain knowledge and skills from other fields. I genuinely feel privileged to have the opportunity to be trained and supervised by her.

I would also like to thank Chiara Peroni and Alexandre Tkatchenko for supporting my journey as a member of my thesis supervision committee. They have always been there whenever I needed support and guidance. I have learned a whole lot from their comments and advice. I am grateful to the staff of STATEC research for their warm welcome and interesting discussions during my research visits. I am also incredibly grateful to Eugenio Peluso, Michela Bia, and Chiara Binelli, who kindly agreed to be part of my PhD jury.

I would like to extend my sincere gratitude to all PhD and post-doc colleagues at the University of Luxembourg, particularly those in the *Economic and Social Well-being* research group (Liyousew, Anthony, Rémi, Vincent, Niccolò, Giorgia, Sonia, and Gemma). Without their encouragement, none of this would have been possible. A special "thank you" goes to my best friend and colleague, Liyousew G. Borga (and his wife, Rokia), whom I cherish for their encouragement, support, and friendship all along.

Financial support of the PRIDE17/12252781 through the Data-driven computational modelling and application (DRIVEN) doctoral training unit granted by the National Research Fund of Luxembourg (FNR) is gratefully acknowledged.

Finally, my most sincere appreciation goes to my wife, Salem. Her continuous encouragement, unfailing love, patience, and prayer gave me mental stamina. A special mention goes to my son Kenna who has been a great source of joy and inspiration that sustained me this far. I am also forever thankful to my brother Solomon for his constant support and encouragement in all my endeavours.

All errors remaining in this text are entirely mine.

Luxembourg
April 7, 2023 Alemayehu Demissew Taye

Dedications

To my late father Demissew Taye Bedhane, and
To my late mother Gennet Wodajo Desta

Contents

List of Figures

List of Tables

Abstract

In the first chapter, building on the definition of vulnerability as expected poverty, we train supervised machine-learning algorithms and a baseline Ordinary least squares regression (OLS) using the German Socio-Economic Panel (version 37) data for years 1984-2020 under two scenarios: 1) considering only cross-sectional data; 2) using over time information on the relative position of the household in the income distribution. Random forests (RF), Gradient boosted trees (GBT), and Neural networks (NN) predict the vulnerable group on average by 23%, 18%, and 16% more than the OLS in the first scenario. The hit rate and the overall accuracy of all vulnerability estimates increase in the second scenario, but the sensitivity gains shrink to 19.3%, 12%, and 5.5%, respectively. With Shapely values from the RF model, we explain the sources of vulnerability and their evolution. We find that weak ties to the labour market, single-person households, the number of dependants in the family, living in East Germany, and the sociodemographic characteristics of household head are associated with vulnerability to poverty.

In the second chapter, using the European Union Statistics on Income and Living Conditions (EU-SILC) microdata and applying machine learning (ML) algorithms, I explore the questions: 1) How accurately can one classify unseen individuals' deprivations status given their observable personal, household, and country-specific factors? 2) What is the performance of targeting subsets of features, such as sociodemographic, socioeconomic, health, and location, to identify the deprived? 3) What are the key predictors and their partial effects? Key results of the empirical analysis demonstrate that the relative accuracy gained by using the sophisticated tree-based ML algorithm is positive and significant compared to that of the standard Generalised linear model (7.3% relative gain with the Extreme gradient boosted trees and 5.9% with the Random forests). Socioeconomic factors yield a classification accuracy as close as when the whole set of features is considered. Feature importance and partial effect analysis identified with Shapley's value reveal insightful relationships consistent with theoretical and empirical evidence.

In the third chapter, using a unique harmonised real-time data set from the COME-HERE longitudinal survey that covers five European countries (France, Germany, Italy, Spain, and Sweden) and applying a non-parametric machine learning model, we identify the main individual and macro-level predictors of self-protecting behaviours against the coronavirus disease 2019 (COVID-19) during the first wave of the pandemic. Exploiting the interpretability of a Random forest algorithm via Shapely values, we find that a higher regional incidence of COVID-19 triggers higher levels of self-protective

behaviour, as does a stricter government policy response. The level of individual knowledge about the pandemic, confidence in institutions, and population density also ranks high among the factors that predict self-protecting behaviours. We also identify a steep socioeconomic gradient with lower levels of self-protecting behaviours being associated with lower income and poor housing conditions. Among socio-demographic factors, gender, marital status, age, and region of residence are the main determinants of self-protective measures.

Introduction

This dissertation consists of three essays examining the measurement and the prediction of individual well-being.

Previous studies show the effective use of prediction methods for policy analysis, for instance, in education, labour market policy, public health, criminal justice, finance, and social policy (Kleinberg et al., 2015; Rockoff et al., 2011; Kang et al., 2013; Chandler et al., 2011; Kleinberg et al., 2018). These illustrative studies demonstrate that some relevant policy problems are inherently prediction problems. However, despite the widespread application of prediction for policy problems, the limitations of the traditional empirical approach bound our ability to predict with accuracy on real-world data. In response, a flexible data-driven machine-learning method has gained increasing traction in many fields of study over the past few years. Big data, cheap computational power, and algorithmic advance fuel the recent resurgence of interest in applied machine learning (ML).[1] More recently, economists have started experimenting with machine learning algorithms as feasible alternatives to address various prediction problems (Athey, 2018; Mullainathan and Spiess, 2017; Varian, 2014).

Using machine learning techniques and three different large micro datasets, this dissertation answers research questions in three vital areas of individual well-being: measurement of vulnerability to poverty, classification of material and social deprivations, and identifying factors associated with individuals self-protective behaviour during a public health crisis.

The first chapter examines vulnerability to poverty as an important measure of economic well-being to proactively target those families who will experience poverty in the future. In a joint work with Conchita D'Ambrosio, we use the German Socio-Economic Panel (SOEP) data covering 1984-2020 to build a data-driven interpretable model for predicting vulnerability to poverty and understanding its direct and indirect

[1]Machine learning is a sub-field of artificial intelligence. Artificial intelligence (AI) refers to systems or machines that emulate the problem-solving and decision-making capabilities of the human intelligence and can iteratively improve their abilities based on the data (Russell, 2010).

determinants. The most pursued definition in the literature characterises vulnerability as the probability of becoming poor in the future. However, the existing empirical implementation of this concept imposes strong assumptions about the distribution of income and its mapping with observable household characteristics that could not necessarily be warranted in real-world data, which could, in turn, negatively impact the accuracy of vulnerability estimates.

This study contributes to the literature by providing the first application of an explainable machine learning approach (ExpML) to predict vulnerability to poverty. We use Random forests, Gradient boosted trees, Neural networks, and penalised regressions as examples of ML algorithms. As the vulnerability issue is even more prevalent in developing nations where panel data is commonly scarce, the study builds and examines the performance of the ExpML model under two scenarios—only using cross-sectional information and then incorporating the past relative position of the households in the income distribution. Finally, we document the drivers of household income, the profiles of the vulnerable group and their evolution over the past three decades in Germany.

The results from this study demonstrate a substantial advantage of the non-parametric Random forests algorithm in accurately targeting the vulnerable group. These results are robust to varying the cross-sectional year, extending the prediction time horizon, and using alternative vulnerability cutoffs. Additionally, by applying the Shapley values on the RF algorithm, we identify that the past year's income decile, household's ties to the labour market, human capital, household's composition (by age and activity status), and region of residence are the potent signal to the sources of vulnerability.

The second chapter is concerned with material deprivations, which is an absolute, non-monetary, outcome-based ('standards of living') measure of individual well-being. Using supervised machine learning algorithms, such as extreme gradient boosted trees (Xgboost) and Random forests (RF), this chapter analyses the prediction of unseen individuals' deprivation status and identifies the main predictors and their partial effects. It investigates the performance of targeting subsets of indicators, such as sociodemographic, socioeconomic, health, and location, to identify the deprived. I carry out the empirical analysis using the European Union Statistics on Income and Living Conditions (EU-SILC) microdata, a nationally representative sample of individuals aged 16 or above in 28 European countries, covering a range of material and social deprivation indicators, sociodemographic, socioeconomic, and health variables. This study makes several contributions to the related literature. First, as far as the author is concerned, this study is the first to evaluate the usefulness of the nonparametric extreme gradient

boosted trees classifier algorithm augmented with Shapley values to model the complex nature of deprivations by comparing the ML algorithms with the traditional logistic regression and documenting their heterogeneity by country. Second, using the newly revised 13-item deprivation indicators proposed by Guio et al. (2016), this study analyses deprivations in material and social aspects instead of those based on the old 9-item indicators, which mainly capture household-level hardship in durable goods. Finally, while remaining highly relevant to the previous studies, this study reveals non-linear relationships between key variables and the likelihood of poor living conditions.

The results indicate that the relative accuracy gained by using the more sophisticated ML algorithm is positive and significant compared to logistic regression. The socioeconomic and location features have the best classification power compared to sociodemographic and health characteristics. The relative feature importance identified with Shapley's values shows that the individual's relative economic position to others, general health status, level of education, age, and housing wealth is the most prominent predictor of deprivation. These results are robust to using different cross-sectional years.

The third chapter answers the question: what predict self-protecting behaviours during the earliest wave of the pandemic? During the first wave of the pandemic, without a vaccine or therapeutic measures, governments had to rely on behavioural interventions to slow the spread of the virus and reduce the number of infections. Authorities launched confinement policies - such as lockdowns, travel restrictions, and social distancing requirements - and preventive sanitary measures - such as mask-wearing and frequent handwashing. Even though it is apparent that human behaviour largely influences the spread of the virus, it is unclear which factors are most strongly associated with protective behaviours. Identifying these factors is of paramount importance for devising effective policies to manage the current pandemic as well as to be better prepared for future ones. In a joint work with Liyousew G. Borga, Samuel Greiff, Claus Vögele, and Conchita D'Ambrosio, we use a unique harmonized real-time data set, the COME-HERE (COVID-19, MEn- tal HEalth, Resilience, and Self-regulation) longitudinal survey, collected by the University of Luxembourg in five European countries (France, Germany, Italy, Spain, and Sweden) covering a range of individual-level information such as sociodemographic variables, income and wealth, health and behavioural risk factors, awareness about the pandemic, and trust in major public institutions. Additionally, we consult country-level information describing the evolution of the pandemic itself, as well as the policy responses to COVID-19, from the Blavatnik School of Government of the University of Oxford COVID-19 government response tracker.

Using a machine learning approach, we examine how individual characteristics and government policy responses predict self-protecting behaviours during the earliest wave of the pandemic. This study makes several contributions to the related literature. First, we use a combination of complex non-parametric machine learning model and state-of-the-art model explanation method to explain factors impacting the adoption of self-protecting behaviours during the COVID-19 pandemic. To the best of our knowledge, this is the first attempt in the literature. Second, we demonstrate a tree-based algorithm's advantages and relative gains over linear regression. Third, we train a highly predictive model with original data with a universe of items specifically constructed to measure the behavioural change in response to COVID-19. This allows us to minimise the bias from measurement error, which is a common limitation in related studies that only focus on a few measures. Moreover, we use causal theory to justify the reason for including or omitting variables, and our data offer a large number of features. This enables us to minimise bias in the partial effects of features that may occur when an important variable is omitted in a causal model. Fourth, our approach allows for the presence of interaction effects among key features. And finally, we identify key policy relevant individual and social predictors of self-protecting behaviours and document their heterogeneity by country.

The results suggest that a higher regional incidence of COVID-19 triggers higher levels of self-protective behaviour, as does a stricter government policy response. The level of individual knowledge about the pandemic, confidence in institutions, and population density also ranks high among the factors that predict self-protecting behaviours. We also identify a steep socioeconomic gradient with lower levels of self-protecting behaviours being associated with lower income and poor housing conditions. Among sociodemographic factors, gender, marital status, age, and region of residence are the main determinants of self-protective measures. This work is forthcoming in Scientific Reports.

References

Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda*, pages 507–547. University of Chicago Press.

Chandler, D., Levitt, S. D., and List, J. A. (2011). Predicting and preventing shootings among at-risk youth. *American Economic Review*, 101(3):288–92.

Guio, A.-C., Marlier, E., Gordon, D., Fahmy, E., Nandy, S., and Pomati, M. (2016). Improving the measurement of material deprivation at the european union level. *Journal of European Social Policy*, 26(3):219–333.

Kang, J. S., Kuznetsova, P., Luca, M., and Choi, Y. (2013). Where not to eat? improving public policy by predicting hygiene inspections using online reviews. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1443–1448.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. (2018). Human decisions and machine predictions. *Quarterly Journal of Economics*, 133(1):237–293.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review*, 105(5):491–95.

Mullainathan, S. and Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.

Rockoff, J. E., Jacob, B. A., Kane, T. J., and Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education finance and Policy*, 6(1):43–74.

Russell, S. J. (2010). *Artificial intelligence a modern approach*. Pearson Education, Inc.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–28.

Chapter 1

Vulnerability to Poverty: an Explainable Machine Learning Approach *

with Conchita D'AMBROSIO (University of Luxembourg)

*The data used in this study come from the German Socio-Economic Panel Study (SOEP) at the German Institute for Economic Research (DIW), Berlin.

1.1 Introduction

Following the World Development Report, 2000/01 entitled 'Attacking Poverty,' the analysis of households' well-being has evolved to incorporate a measure that allows predicting individuals and families that might suffer well-being loss due to a high chance of facing poverty in the future. Proactive social policy recognises the paramount importance of also targeting vulnerability to poverty (VP) instead of solely relying on the current poverty status. Result in increased attention among researchers, several approaches have been devised to conceptualise and estimate vulnerability to poverty. The favoured approach in the literature defines "Vulnerability as Expected Poverty" (VEP), that is, the household's probability of becoming poor in the future (Chaudhuri et al., 2002; Chaudhuri, 2003; Christiaensen and Subbarao, 2005; Christiaensen and Boisvert, 2000). This approach is preferred for its interpretability since the household's welfare losses are expressed in terms of consumption expenditures or disposable income. VEP can also be applied using panel and cross-sectional data. However, the practical implementation of VEP imposes strong assumptions regarding the distribution of the well-being indicator and its mapping with observable household characteristics that could not necessarily be warranted in real-world data and could, in turn, negatively impact vulnerability estimates' accuracy (see, e.g., Klasen and Povel, 2013; Dutta et al., 2011). Similarly, (Bérgolo et al., 2012; Celidoni, 2013; Haughton and Khandker, 2009; Hohberg et al., 2018) point out the need for improvement as the VP estimate cannot accurately target the vulnerable group.

To this aim, we propose a data-driven technique using a state-of-the-art an explainable machine learning (ExpML) approach to predict VP. Besides the advantage of panel data in allowing the dynamic analysis of vulnerability, evaluating the accuracy of VP estimates requires knowledge of realised poverty status of the predicted vulnerable households, which is only obtained from panel data. Thus, this study uses a high-quality longitudinal panel data of households: the German Socio-Economic Panel (SOEP). Our results show that the empirically optimised ML algorithms: Random forests (RF), Gradient boosted trees (GBT), and Neural networks (NN) improve the prediction of the vulnerable group on average by 23%, 18%, and 16% compared to the baseline Ordinary least squares model (OLS) when cross-sectional data is used. When we consider past information on the relative position of the household in the income distribution, the hit rate and the overall accuracy of all vulnerability estimates increase. However, this impact is more pronounced in the linear models than in the more sophisticated machine learning algorithms. From Shapley values and profile analysis of vulnerable

households, we find factors such as households tie to the labour market, single-person households, the number of dependants in the family, region of residence (East Germany vs. West), and the sociodemographic characteristics of household head to be associated with vulnerability to poverty in Germany.

This study makes several contributions to the related literature. First, to the best of our knowledge, this study is the first to apply ExpML to predict VP. Second, since VP is very relevant also in the developing world where panel data is commonly scarce, we estimate our models under two scenarios. 1) when only cross-sectional data is available vs. 2) when historical information on the relative position of the households in the income distribution is available. Finally, we document the drivers of household income, the determinant of VP and their evolution over the past three decades in Germany.

The rest of the chaper is structured as follows. In Section 1.2 we define the concept of vulnerability to poverty and provide an overview of the existing empirical approaches to measuring it. The data used and the measurement of the outcome variable and set of features is described in Section 1.3. In Section 1.4 we present the empirical strategy to predict VP with ML, set out prediction metrics, and offer a detailed treatment of the considered ML algorithms. In Section 1.5 are contained the results and discussion. Section 1.6 concludes.

1.2 Review of Related Literature

In the context of poverty analysis, *vulnerability* can be theoretically defined as the likelihood of an individual to suffer a substantial shock that reduces its welfare below a socially accepted level (Kühl, 2003). This definition characterises vulnerability to poverty as a forward-looking, ex-ante evaluation of welfare losses as opposed to the assessment of poverty, which is an ex-post concept. Therefore, assessing households' vulnerability status instead of the current poverty level could allow social protection policies to proactively target those families with a higher likelihood of experiencing it (Celidoni, 2013, 2015). The welfare losses are usually assessed in terms of household disposable income or consumption expenditures.

Studies that aim to quantify vulnerability to poverty empirically can be classified into the following four approaches: (i) "vulnerability as uninsured exposure to risk" (VER), (ii) "vulnerability as a low expected utility" (VEU), (iii) "vulnerability by mean risk" (VMR), and (iv) "vulnerability as expected poverty" (VEP) (Gallardo, 2018).[1] The first

[1]For an extensive critical review of each approach, see Gallardo (2018) and Klasen and Povel (2013).

approach retrospectively assesses whether an observed income shock translates into consumption changes. Vulnerability, as uninsured exposure to risk (VER), is identified via regression analysis by estimating the impact of idiosyncratic and covariate income shocks on consumption expenditure (e.g., Cafiero and Vakis, 2006; Dutta et al., 2011). Although VER has advanced the field of consumption smoothing and insurance, it doesn't directly match the concept of vulnerability to poverty since the impact of shocks is assessed ex-post. This approach is built on the notion that it's essential to focus on the change/growth rate of consumption per capita with respect to the previous period instead of the current consumption per capita falling below certain level of consumption deemed adequate. Moreover, the concept of VER does not directly capture the likelihood of adverse shocks (idiosyncratic and covariate shocks), and the individual risk preference is also not accounted for in this approach (Grimm et al., 2016).

The second approach, VEU, as formalised by Ligon and Schechter (2003) measures vulnerability in terms of utility gaps. The utility gaps consider the discrepancy between utility derived from certainty-equivalent consumption and expected utility obtained from actual consumption (Calvo and Dercon, 2005; Günther and Maier, 2014). Although vulnerability as a low expected utility has the advantage of capturing the risk preference of the household, it has limitations stemming from non-interpretability and its dependency on the form of an assumed utility function and its risk aversion parameter (Celidoni, 2013; Christiaensen and Subbarao, 2005; Gaiha and Imai, 2008).

The third approach characterises vulnerability to poverty by mean risk, and a household's vulnerability is assessed based on a mean-risk dominance criterion. The VP estimate under this approach evaluates the household's preference ordering between the contingent welfare outcomes, which are estimated using the expected value of an individual's consumption and an aggregate risk parameter measured with the standard deviation of consumption (Chiwaula et al., 2011; Gallardo, 2018). Individuals are then declared vulnerable if their expected consumption level plus the estimated aggregate risk falls below a poverty threshold. This strand of literature was formerly treated under the VEP category of Hoddinott and Quisumbing's (2003) classification of VP approaches but was later presented as a stand-alone approach in Gallardo (2018) as it does not require a probability distribution function.

Finally, vulnerability as expected poverty measures households' vulnerability as the likelihood of falling into poverty in the period to come. This approach is a widely applied method in the literature because of its advantage of suitability to be applied using either cross-sectional or panel data, and it is also easily interpretable as household disposable income expressed in monetary terms (Celidoni, 2013; Gallardo, 2018; Hohberg et al.,

2018).[2] Formally specified by Chaudhuri et al. (2002) as:

$$V_{ht} = Pr(y_{h,t+1} \leq z), \tag{1.1}$$

where $V_{h,t}$ is vulnerability to poverty of the household h in period t, $y_{h,t+1}$ is the household disposable income in period $t+1$. The poverty line is z. A standard application to empirically estimate this probability with cross-sectional data identifies the income-generating process as follow:

$$ln(y)_{ht} = \boldsymbol{X}_h\beta + \epsilon_h, \tag{1.2}$$

where y_h is disposable income, $\mathbf{X}_h$ represents a bundle of observable household characteristics, but also national or climate variables or past years consumption indicators (e.g., Christiaensen and Subbarao, 2005). ϵ_h is a random error term that captures any idiosyncratic shocks when shock are identically and independently distributed over time. Then the expected disposable income, $E(.)$ and its variance, $Var(.)$ (shown in the following equations) are estimated with ordinary least square regression via three-step feasible generalised least squares (FGLS). For a step-by-step strategy to estimating vulnerability via the FGLS (see, Chaudhuri et al., 2002).

$$\hat{E}[ln(y_h)|\boldsymbol{X}_h] = \boldsymbol{X}_h\hat{\beta}, \tag{1.3}$$

$$\widehat{Var}[ln(y_h)|\boldsymbol{X}_h] = \widehat{\sigma_{e,h}} = \sqrt{\boldsymbol{X}_h\hat{\theta}}, \text{ and} \tag{1.4}$$

$$\widehat{V}_h = \widehat{Pr}(ln(y_h) < ln(z)|\boldsymbol{X}_h) = \boldsymbol{\Phi}\left(\frac{ln(z) - \boldsymbol{X}_h\hat{\beta}}{\sqrt{\boldsymbol{X}_h\hat{\theta}}}\right). \tag{1.5}$$

Under a log normal distributional assumption of household disposable income, the probability of vulnerability will be drawn parametrically from the standard normal distribution $\Phi(.)$. Commonly in most applications (for instance, see, Chaudhuri et al., 2002; Chaudhuri, 2003; Christiaensen and Subbarao, 2005; Hohberg et al., 2018; Kamanou and Morduch, 2002; Landau et al., 2012; Pritchett et al., 2000; Zereyesus et al., 2017), households with a probability larger than or equal to 0.5 are identified as vulnerable.

The measurement of VP is inherently very challenging since it depends on the adequate prediction of the future (Haughton and Khandker, 2009). This study builds

[2]However, panel data is required to estimate vulnerability to poverty under the VEU and VER approaches (Azeem et al., 2018).

on vulnerability as expected poverty. This is our preferred approach since it is forward-looking, interpretable, and popular in the extant literature. We then propose entirely data-driven machine learning approaches that target the vulnerable group with better accuracy while still utilising the VEP's original definition.

More specifically, we investigate the possible technical refinements in the VEP approach with respect to the following limitations:

(i) The vulnerability estimate via the FGLS method has multiple sources of potential noises, which are coming from the first-stage estimation of the expected value of household income, its variance, and the probability distribution in the final stage. However, employing complex models or improving the model's fitting parameters can minimise the reducible error component (model bias or approximation error) in the first-stage regression (Hastie et al., 2009).

(ii) The log-normality assumption of the household income distribution could not necessarily be warranted in real-world data. Several empirical applications posit departure from log normality (Battistin et al., 2009; Celidoni, 2013; Hohberg et al., 2018; McDonald and Ransom, 2008; Sohn et al., 2015).

(iii) Several authors challenge an assumption that there is a parametric linear relationship between household income y_h and each observable household characteristic $\mathbf{X}_h$ in the first-stage regression. For instance, Kamanou and Morduch (2004) estimate vulnerability to poverty using the observed distribution of the errors of the first-stage equation, using bootstrapping.

(iv) Several studies point out this approach's limitation regarding an arbitrary probability cutoff below which a household is regarded as vulnerable. Hohberg et al. (2018) argue that the vulnerability threshold at 50 percent disregard the variability a household faces. Similarly, other studies posit that when the expected value of household income coincide with the poverty line (on the log scale), the probability is 50 percent independent of the standard deviation (Gallardo, 2018; McCarthy et al., 2016).

In Section 1.4 we discuss our proposed method in detail: And how we aim to tackle the above drawbacks. We explain how the machine learning technique works in this context. We also outline the steps we take, such as model training, cross-validation, hyperparameter optimisation, model testing, and model interpretation.

1.3 Data and Summary Statistics

To build the ML-based vulnerability estimates and assess their predictive power, we use the high-quality longitudinal panel data of private households, the German Socio-Economic Panel (SOEP) (Goebel et al., 2019). We employ SOEP-core data for years 1984-2020 (v37). Consistent with the extant literature, we keep the unit of analysis at the household level, assuming the possibility of resource sharing within the households. We restrict our analyses to households with equivalent disposable income below 100,000 Euros because the wealthy households above this limit will be less susceptible to a short period ahead of poverty. Our main empirical exercises require that a given household's income should be available in three consecutive waves. For instance, in a one-year ahead VP prediction, we train the algorithms using available information in the first two waves and the third wave is used to validate the performance of our prediction. To check the robustness of our prediction to different years (the obtained prediction accuracy should be survey-year independent) and to see how the income-generating function evolves in Germany, we perform the empirical exercise retrospectively at three different years i.e., 1997, 2007, and 2017 with sample size of 5827, 9433, and 11026 households, respectively. In order to do so, we create a three-year balanced panel: 1996-1998, 2006-2008, and 2016-2018. Last, we expand our analysis to a three-year ahead vulnerability to poverty. We again use the first two waves of each sample to train ML models and then conduct performance evaluation with the fifth wave. Put formally, for instance, using the information of year t and the previous year $(t-1)$, we predict if the household would become poor in $t+3$. We then evaluate the accuracy using the realised poverty status in $t+3$.

1.3.1 Measures

The outcome variable in the income-generating function is the equivalized household disposable income. To adjust for price variation across time, the annual disposable income expressed in PPP (at constant 2010 price) is used. To control for differences in the household size and the resulting economies of scale, we assign the square root of household size to compute the equivalent income.

The features set employed to train the income-generating function are variables at the household level and the household's head characteristics. In particular, features on the household's head consist of socio-demographic variables, such as gender, the achieved highest level of education degree (with International Standard Classification of Educa-

tion (ISCED) three levels aggregation: low, middle, and high), marital status (single, married, widowed, divorced, and separated). In addition, we consider variables that measure the labour market participation of the household's head, including the number of hours worked, occupation class (self-employed, blue-collar, and white-collar), and employment industry (based on one-digit industry aggregation). Features relating to family head's health conditions have also been considered. These measures are objective health indicators, such as disability status and hospitalisation i.e., a dichotomous variable asking whether a person stayed overnight in a hospital in the previous year.

Household characteristics include the previous year household's relative position in the distribution of equivalent disposable income (by decile), household size, the federal state of residence (sixteen states), household's composition by age group (i.e., the number of individuals below 18, [18, 34), [34, 59), and above 59), and home ownership status. We measure a household's ties to the labour market by the number of individuals in the family belonging to either of three categories of activity status (i.e., the number of individuals employed as full-time, part-time, and those unemployed). We also consider nationality and region of residence (West Germany vs. East Germany).

1.3.2 Summary Statistics

Table 1.1 presents descriptive statistics of selected features over the three cross-sectional years. All samples mainly consist of household head of working age, male, married, with vocational education degrees. With respect to the form of dwelling ownership, nearly 39% of the households owned their residence in the 1997 sample, but this proportion increased to 51% in 2007 and 47% in 2017. On average, approximately 40% of the sample has an inactive or unemployed household head, which could be because of a higher percentage of retirees, as the average age of the household head was 52 years of age.

Table. 1.1 *Summary statistics of selected variables*

I. Household characteristics	**1997**	**2007**	**2017**
Household size	2.65 (1.31)	2.34 (1.18)	2.50 (1.36)
HH's composition by age:			
Under 18	0.62 (0.95)	0.42 (0.81)	0.62 (1.03)
18-34 years old	0.62 (0.78)	0.39 (0.67)	0.36 (0.66)
34-59 years old	0.96 (0.87)	0.91 (0.86)	0.91 (0.85)
$\geq$ 60 years old	0.45 (0.72)	0.62 (0.81)	0.59 (0.79)
HH's ties to labour market:			
№ of full-time worker	0.87 (0.78)	0.70 (0.74)	0.65 (0.70)
№ of non-working	0.83 (0.81)	0.85 (0.82)	0.79 (0.83)
№ of part-time worker	0.41 (0.59)	0.41 (0.60)	0.50 (0.63)
Home owner	38.74	51.28	46.74
Region of residence:			
West Germany	71.58	74.72	77.34
II. Information on household head			
Age	48.67 (15.55)	53.47 (15.62)	54.04 (15.60)
Gender (female = 1)	37.84	40.28	48.75
Occupation class:			
Blue-collar	37.03	25.85	23.69
White-collar	45.2	51.14	56.82
Self-employed	5.19	4.55	3.84
Marital Status:			
Divorced	8.62	11.78	15.46
Married	63.94	57.98	54.25
Separated	2.99	3.12	3.71
Single	13.55	15.84	17.61
Widowed	10.18	10.74	8.68
Education:			
High	24.54	35.15	36.06
Middle	57.69	55.94	54.62
Low	17.77	8.90	9.32
Health status:			
Hospitalized	12	12.84	14.71
Disability	12.78	13.85	13.29
N	5827	9433	11026

Notes: This table presents summary statistics of key features used in the final prediction. The numbers in the tables show the percentages. The other numbers, followed by parenthesis, are the average and standard deviation, respectively.

1.4 Method

We estimate vulnerability to poverty in two steps. First, we apply Supervised ML models to learn the household income-generating function from the training data. We make abstract from a large body of theoretical and empirical literature about the household-

level income-generating function, we mainly refer to (Miles, 1997; Chaudhuri et al., 2002; Christiaensen and Subbarao, 2005). We build the models under two scenarios of the household income-generating process: In the first scenario, we consider only cross-sectional data, i.e., without $y_{h,t-1}$ in the following specification and this follows Chaudhuri et al.'s (2002) formulation of the income-generating process (described in Section 1.2). In the second scenario, we introduce $y_{h,t-1}$, the past information on the relative position of the household in the income distribution as an additional covariate. We thus have

$$y_{\text{ht}} = f\left(y_{h,t-1}, \mathbf{X}_h\right) + \epsilon_{\text{ht}} \tag{1.6}$$

where y_{ht} is the equivalent disposable household income, f is some fixed but unknown function of input features, $y_{h,t-1}$ is the lag of equivalent disposable household income, $\mathbf{X}_h$ represents a vector of observable household characteristics (described in Section 1.3), ϵ_{ht} is a random error term that captures any idiosyncratic shocks that contribute to differential household disposable income. The accuracy of $\widehat{y}_{\text{ht}}$ as a predictor of y_{ht} depends on the component of the error term often called reducible error: It is reducible because one can potentially minimize it with the use of suitable ML algorithm to estimate $f(.)$.

$$E\left(y_{\text{ht}} - \widehat{y}_{\text{ht}}\right)^2 = E\left[f(.) + \epsilon - \widehat{f}(.)\right]^2 = \underbrace{\left[f(.) - \widehat{f}(.)\right]^2}_{\text{Reducible}} + \underbrace{Var\left(\epsilon\right)}_{\text{Irreducible}} \tag{1.7}$$

where $E\left(y_{\text{ht}} - \widehat{y}_{\text{ht}}\right)^2$ is the expected value of the squared difference between the predicted and actual value of y_{ht}, and Var (ϵ) represents the variance associated with the error term.

The supervised ML technique has specific desirable properties relevant to our goal of predicting $f(.)$ with better accuracy by extracting the most *generalizable pattern* on the unseen individuals. To control *over-fitting* in the learning process, particular importance is given to model hyperparameters (details in the subsequent sections). The ML algorithms can handle high *dimensionality*, and more importantly, the flexibility of ML allows the model to fit complex relationships in the data that were not specified in advance.[3] Hence, the FGLS's assumptions about the distribution of the dependent variable and the parametric mapping of input features onto the outcome could no longer be restrictive.

[3]For a dedicated discussion of ML technique vis-a-vis the traditional econometric approach, see (Athey and Imbens, 2019; Mullainathan and Spiess, 2017).

In the second stage, we identify vulnerability to poverty at time t as being in poverty in the next period. Unlike the FGLS based estimates, in this step, we perform binary categorisation of household vulnerability status without relying on the lognormality of the income distribution. Formally,

$$\widehat{v}_{ht} = \begin{cases} 1, & \text{if } \widetilde{y}_{h,t+1} = \widehat{f}(y_{ht}, \mathbf{X}_h) \leq z \\ 0, & \text{if } \widetilde{y}_{h,t+1} = \widehat{f}(y_{ht}, \mathbf{X}_h) > z \end{cases} \tag{1.8}$$

where $\widetilde{y}_{h,t+1}$ is the forecast of household equivalized income in period $t+1$ that is inferred from the first-stage regression. To identify households' vulnerability to poverty, one must specify the appropriate poverty line z. We experiment with the commonly used poverty thresholds: a) the absolute poverty line computed as the mean relative poverty lines over the years, approximately 11000 Euros. b) the relative poverty lines set at 60% of the median the observed equivalized income at time t, and c) the relative poverty lines set at 60% of the median of forecasted income for period $t+1$.

Finally, to see of how well these models perform in predicting VP vis-a-vis the baseline OLS, we set evaluation metrics that capture local and global accuracy. We construct *sensitivity*, the so-called hit rate, the probability of detection, or the true positive rate (TPR), i.e., a percentage of households predicted vulnerable in period t ($\widehat{v}_{ht} = 1$) and actually became poor in period $t+1$ ($p_{h,t+1} = 1$). In addition, we look at overall *accuracy*, which is a proportion of households in the sample N that were correctly classified as vulnerable ($\widehat{v}_h = 1$) and non-vulnerable ($\widehat{v}_h = 0$).

$$sensitivity = \frac{\sum_{h=1}^{n} \mathbf{I}\left(p_{h,t+1} = 1 | \widehat{v}_{\text{ht}} = 1\right)}{\sum_{h=1}^{n} \mathbf{I}\left(p_{h,t+1} = 1 | \widehat{v}_{\text{ht}} = 1\right) + \sum_{h=1}^{n} \mathbf{I}\left(p_{h,t+1} = 1 | \widehat{v}_{\text{ht}} = 0\right)} \tag{1.9}$$

$$accuracy = \frac{\sum_{h=1}^{n} \mathbf{I}\left(p_{h,t+1} = 1 | \widehat{v}_{\text{ht}} = 1\right) + \sum_{h=1}^{n} \mathbf{I}\left(p_{h,t+1} = 0 | \widehat{v}_{\text{ht}} = 0\right)}{N}. \tag{1.10}$$

Hence, in addition to the predicted VP status of households, we utilise the observed household poverty status in the last year of each panel when constructing the sensitivity and accuracy of each model. The idea is that, for example, say we were in 1997 and observe a household's information during this year and the previous year. Using this information, we predict if a given household will be poor in 1998. Then once the household's actual poverty status was revealed in 1998, we retrospectively validate the accuracy of our prediction using the realised poverty status of 1998. In a similar vein, for a three-year ahead VP prediction, we employ information of 1997 and the previous

year to make the prediction if a given household will be poor in 2000 and then validate the accuracy of our estimate using the realised poverty status in 2000.

The remaining part of this section describes the selected ML algorithms.

1.4.1 Which ML Algorithms?

Supervised ML consists of a myriad of algorithms, and hence the question of which algorithms work best seeks case-specific treatment.[4] To suggest a robust algorithm for predicting vulnerability to poverty, we rely on an extensive experimental approach. We employ five ML models from different function classes, namely the Least absolute shrinkage and selection operator (LASSO), Ridge regression (RIDGE), Random forests, Gradient boosted trees, and deep learning Neural networks model.

LASSO and Ridge regression are widely used to predict continuous outcomes variables. These models are analogous to OLS with a crucial exception that these models penalise the model's complexity by adding regularisation parameters in the least square loss function and shrinking all non-zero coefficients.[5] Both models are generally efficient when applied to data sets with many features. However, the critical difference between LASSO and Ridge is that the former automatically discard variables insignificant to the model's predictive power using a replicable selection mechanism—by pushing some of the regression coefficients precisely equal to zero. Thus, resulting in a parsimonious solution and helps avoid multicollinearity and overfitting problems that often afflict OLS models, especially when the number of explanatory factors is large (Tibshirani, 1996). On the other hand, ridge regression keeps the estimated coefficients different from zero (Hoerl and Kennard, 1970). Generally, it suits prediction problems where sparsity is not the priority issue. In this study, these models are utilised to examine if we can improve VP prediction without diverging from the linearity assumption of the standard approach in the extant literature.

Random forests (RF) are an ensemble of many *de-correlated* trees. These trees are grown from subsamples of the training set using many random subsets of the features

[4] A well-known theorem in machine learning literature states that there is 'no free lunch' in ML– meaning that no one algorithm is best for all problems (Wolpert, 2002; Wolpert and Macready, 1997). So, determining a suitable algorithm often remains at the researcher's discretion.

[5]

$$\widehat{\beta} = argmin_{\beta} \sum_{i=1}^{N} \left(Y_i - \beta' X_i\right)^2 + \lambda \left(||\beta||_q\right)^{1/q}; \widehat{\beta} = \widehat{\beta}_{\text{ols}} \text{ if } \lambda = 0; \widehat{\beta} = \widehat{\beta}_{\text{lasso}} \text{ if } q = 1; \widehat{\beta} = \widehat{\beta}_{\text{ridge}} \text{ if } q = 2,$$

were $||\beta||_q = \sum_{k=1}^{K} |\beta_k|^q$ and $\lambda \geq 0$ is a complexity parameter that controls the amount of shrinkage: the bigger the value of λ, the greater the amount of shrinkage.

(Breiman, 2001). In a regression task, the final prediction is then rendered as the average prediction across all trees, while in a classification task forest's best prediction is chosen on a majority vote. Hence, random forests overcome the overfitting problem that is often afflicting the decision tree regressor by averaging many noisy but unbiased trees. This ensemble of trees increases model stability by inducing smoother estimation of the function underlying the data-generation process, which results in a robust and accurate prediction. For a rigorous mathematical characterisation of reduction in overfitting, while retaining the predictive accuracy of the trees, see (Breiman, 2001; Hastie et al., 2015).

Gradient boosted trees (GBT) is an ensemble of trees in which the model is built sequentially with base learners fitting simple models to current pseudo-residuals by least-square at each iteration (Friedman, 2002). In ML, boosting is a technique of combining several simple supervised ML models (known as weak learners) into a powerful composite model (Schapire, 2003; Freund et al., 1999; Drucker, 1997; Mason et al., 1999). The term '*gradient*' in GBT refers to how pseudo-residuals are obtained with the gradient descent of the loss function being minimised with respect to the model's values at every training data point. Unlike the RF, an ensemble of trees in the GBT is not built parallelly and independently. Instead, the weak learners are first trained on a random subsample of the full training data, and then in each boosting iteration, a tree is grown sequentially on a modified data (the residuals as an outcome variable) by correcting past mistakes with sample weighting—i.e., more weights on the wrong predicted samples (Friedman, 2001, 2002).

Neural networks (NN) is a flexible supervised ML algorithm that can generally learn nonlinear function $f(\mathbf{x};\theta)$ (governed by unknown parameter θ) by mapping an input vector $\mathbf{x}$ onto an output vector y. This mapping consists of one or more nonlinear layers referred to as *hidden layer* and NN with many of these layers called '*deep*' neural networks. Each node in the hidden layer (often called '*neuron*') transforms the value from the previous layer with a weighted sum. Then a nonlinear activation function is applied to yield the final result (Goodfellow et al., 2016). Hence, the NN's complexity is controlled by the number of hidden layers, the number of neurons per hidden layer, the connectivity of each layer (the architecture of the networks), and the type of activation function used. Unlike the classical linear models, the coefficients of NN (often called weights) are trained flexibly by first assigning some arbitrary values and then updated by evaluating the loss function, which is specified as mean squared error for regression or cross-entropy for classification. The updating task is conducted via Stochastic Gradient Descent (SGD) of the loss function being minimised with respect to a parameter to be

updated (Goodfellow et al., 2016; Bottou, 2010; LeCun et al., 2012).

1.5 Results and Discussion

1.5.1 Prediction Process with Different ML Algorithms

This study trains the above algorithms as implemented in the scikit-learn package (version 0.22.2), the most prominent Python library for ML (Pedregosa et al., 2011). The objective is to find a generalizable estimate of the income-generating function outlined in Equation 6 under the two scenarios: 1) when only cross-sectional data is available and 2) when historical information on the relative position of the households in the income distribution is available. For the sake of systematic presentation, henceforth, specifications 6 (a) and 6 (b) denote scenario one and two, respectively. Before feeding the dataset into the models, it is always essential to perform the necessary feature engineering/pre-processing (Zheng and Casari, 2018; Kotsiantis et al., 2006). As the SOEP dataset is complete and clean, we pre-process the non-numerical categorical non-ordinal variables (e.g., federal state of residence, marital status, occupation industry) and perform feature scaling. Assigning an integer value, say from 1 to k, to each of k possible categories of non-ordinal variables makes the model assume a natural ordering between categories, resulting in poor performances or unexpected results. Therefore, we apply one-hot encoding.[6] Although the tree-based algorithms are insensitive to the monotonic transformation of features, heterogeneous scales may generate incomparable or convoluted importance in models that are smooth functions of the input or anything that involves a matrix. In general, quick convergence in the Neural networks also requires the predictors to be on the same scale. Hence, we train the linear models and NN using standardised predictors.

To avoid data leakage, we use 80% random sample of the dataset as training set and the remaining as test set. Then the models are trained on the training set model's generalisability to the unseen households is tested using the test set. The hyperparameters are empirically tuned via grid search with five-fold cross-validation (CV). The optimal hyperparameters are obtained based on high mean test prediction accuracy measured by R-squared in the process of five-fold CV. We tune the regularised linear models by the complexity parameter. In the RF algorithm, we discover that the maximum depths of trees (*'max_depth'*) and the number of features used in each split (*'max_feature'*), are

[6]One-hot encoding is a technique to create a group of dummy variables—each variable represents a possible category if the variable does not belong to multiple categories at once.

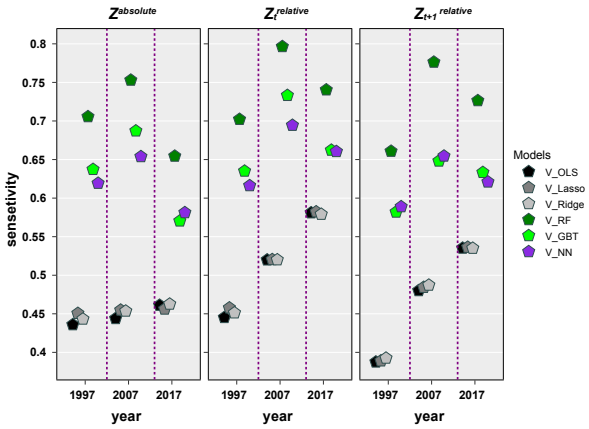
consequential hyperparameters to the predictive accuracy. We control the learning process of GBT using the number of boosting stages (*'n_estimator'*), the maximum depth of the tree (*'max_depth'*) and learning rate ($0 <$ *'learning_rate'* ≤ 1). Last, considering the size of our datasets, a NN with two hidden layers is built. Many hyperparameters could be tuned for Neural Nets. In this work, hyperparameters related to neural networks architectures (i.e., the number of neurons per hidden layer) and the learning rate are tuned jointly; we run the above procedure over a grid of three-dimensional values (See Appendix 1.B for a comprehensive training workflow, model inspection, and the chosen hyperparameters of all ML algorithms under both specifications on the three panels).

1.5.2 Performance Evaluations

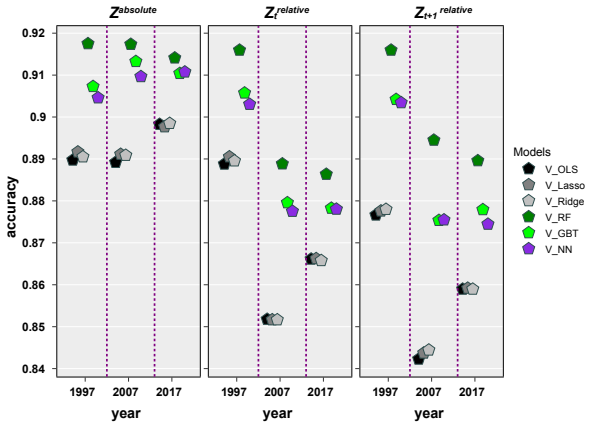
We are now able to predict a year ahead of VP in 1997, 2007, and 2017. Besides parsimony, the ten years gap between the years will allow us to capture heterogeneity in VP prediction over time, which could be explained by a mixture of the German business cycle over the three decades and the variation that is sample-specific.

First, we analyse the performance of the modes' vulnerability estimate with specification 6 (a) under the three alternative poverty cutoffs: $\mathbf{z}^{absolute}$, $\mathbf{z}_t^{relative}$, and $\mathbf{z}_{t+1}^{relative}$ (as described in Section 1.4). Starting from the 1997 and focusing on the estimated vulnerability by the linear machine learning algorithms vis-a-vis the baseline OLS with the relative vulnerability cutoff seen in the second column of Figure 1.1 (a), it is possible to notice that the true positive rate or the hit rate of the baseline model is small (approximately 44.5%). We find negligible relative sensitivity and accuracy gains from LASSO and Ridge regressions. These observations hold for the vulnerability estimates of 2007 and 2017 regardless of the type cutoffs used. Hence, we can conclude that regularising the linear model could not improve vulnerability estimates. These algorithms generally work better when the OLS model is overfitting. They penalise model's complexity by adding a penalty term on the least square loss function. For instance, LASSO generally performs better than OLS when a slim fraction of the included features significantly impacts the response. In our VP setup, however, we train the algorithms on large sample sizes and sets of clean and non-redundant predictors, which are theoretically expected to impact the response. In both LASSO and Ridge regression the tuned complexity parameters are close to zero (See Figure 1.7 and Figure 1.8 in the appendix), meaning that the coefficient estimates from the regularised models are very close to the ones with OLS.

Figure. 1.1 *Performance of models' VP estimate under specification 6 (a)*



(a) *Sensitivity*



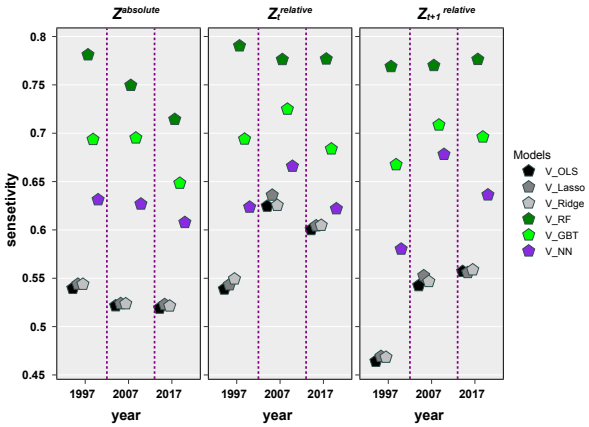
(b) *Accuracy*

In contrast, we obtain remarkable prediction accuracy by relaxing the over-simplifying linearity assumption of FGLS. For instance, predictions with the RF algorithm target future poverty with much better sensitivity of 26%, 28%, and 16% than the baseline estimate in 1997, 2007, and 2017. Similarly, with the Gradient boosted trees, the gain is 23%, 22%, and 9% more per each survey year analyzed, followed by the Neural Network's prediction, which resulted in 18%, 24%, and 6% higher sensitivity vis-a-vis the OLS estimates. Furthermore, results show that the patterns of sensitivity and accuracy gains with the non-linear ML models are robust to the poverty cutoffs.

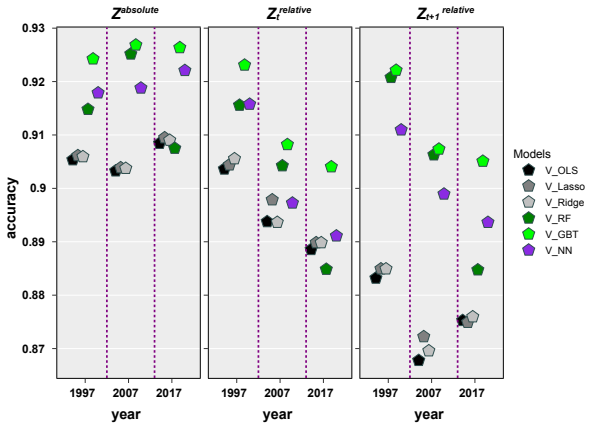
These results suggest the presence of interactions and non-linearities in the household income-generating process. In other words, if the feature's contribution is independent of one another and there was linear mapping between these input features and the target variable, then algorithms from the linear function class would perform well. This empirical exercise suggests that the Explainable ML approach to VP with empir-

ically optimised tree-based algorithms (and Neural networks) shows high potential in targeting future poverty with better accuracy, as these models are specifically designed to learn non-linearity and feature interactions.

Figure. 1.2 *Performance of models' VP estimate under specification 6 (b)*



(a) *Sensitivity*



(b) *Accuracy*

Next, we investigate if there is accuracy gain by considering the household's relative position in the national income distribution in the past year. This is performed by re-training the models to learn specification 6 (b) , which includes the deciles of the past year's household disposable equivalent income. We decided to consider only the preceding one year's income information because including many lagged information did not improve model accuracy significantly but resulted in unstable model interpretability due to the high collinearity of these features. Hence, as can be seen in the second column of Figure 1.2 (a), the hit rate of the VP estimate based on the re-trained model increased to 54%, 62%, and 60% over the three years in the baseline OLS model. In

other words, averaged over the years, the linear models predict vulnerability by approximately 7% more hit rate than they do when past household income is not in the model. Nonetheless, the average prediction gains with the introduction of the income feature are 3.5, 1.2., and -3.3 percent in RF, GBT, and NN, respectively. These results suggest that historical information regarding the relative position in the income distribution has a more prominent role in increasing the prediction performances of linear models. Last, from this experiment, we can also reiterate the conclusions that the gain from LASSO and ridge regression is marginal compared to the baseline estimate vs. the gain from Random forests, which is 25, 15, and 18 percent more than the baseline estimate.

1.5.3 Drivers of Household Income and their Evolution: ML Interpretability

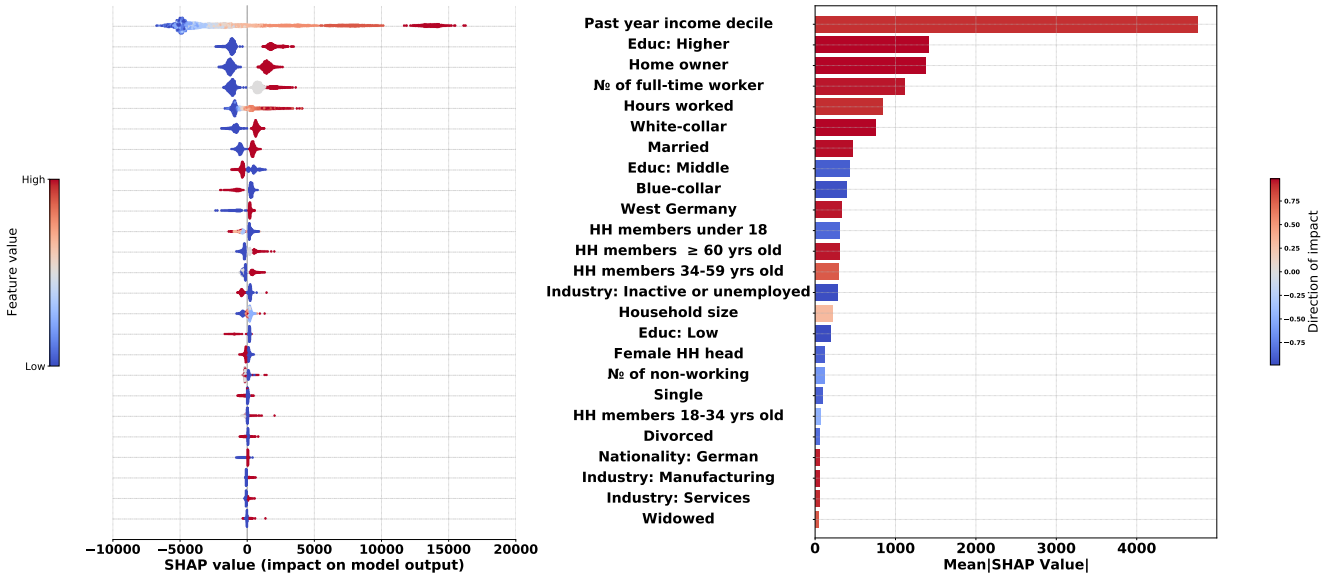
Understanding the prediction of a household income-generating process is equally essential to obtain actionable insights that can ultimately contribute to social policies that aim at poverty eradication. Many state-of-the-art algorithms such as RF and GBT often produce better accuracy. However, unlike parametric linear models, these ML algorithms lack model-specific or intrinsic interpretability. Nevertheless, these algorithms report feature importance, i.e., increase in the mean squared errors (MSE) of prediction when a feature is randomly permuted from the model. However, the feature ranking generated by these models does not provide the direction and magnitude of the association between the predictors and the target variable, posing a challenge to the domain expert to have a gist of ML-produced results. To overcome this limitation, we use the visualisation tool SHapley Additive exPlanations (SHAP) proposed by Lundberg and Lee (2017), to reveal insights into a complex outcome. SHAP is based on a solution concept in a cooperative game setup that aims to 'fairly' allocate the gains among players as suggested in the seminal work of Shapley (1953). SHAP has the advantage of consistency and provides both local and global interpretability (see Guidotti et al., 2018; Pedregosa et al., 2011; Molnar, 2020, for a comprehensive review of black-box ML model interpretation techniques).

In what follows, we will look at Shapley's value analysis to identify the key predictors and disentangle their partial effects. First, we analyse Shapley values of features capturing households' characteristics without past year income decile, i.e., model 6 (a); these results are in Appendix 1.D. We then introduce the past income decile to see how the features' impact change (model 6 (b)), as shown in Figure 1.3. We report the

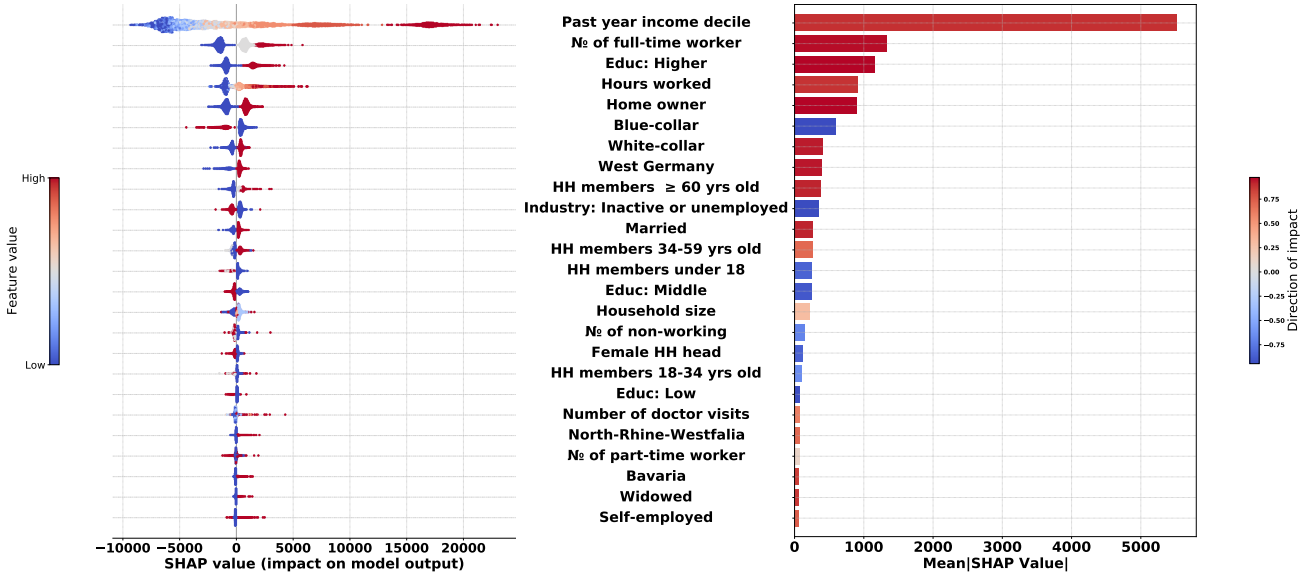
top 25 significant predictors of household income during the three cross-sectional years in the text for illustration reasons—full results showing all covariates used to estimate the income-generating function are reported in Appendix 1.C. The top 25 predictors are placed in descending order of significance to the prediction, and the direction of impacts of these variables are consistent with the economic intuition (well summarised in the bar graph). Model 6 (b) has a higher mean test R-squared (shown in Figure 1.9 in the appendix) and higher average hit rate (and overall accuracy) of VP predictions than model 6 (a) for all the years analysed. This fact clearly shows that the households' relative position in the income distribution of a given year is an essential positive predictor of their income in the following year. Our model captures an interesting positive non-linear association for all the years investigated, as plotted in panel (a) of Figure 1.3. Controlling for other household characteristics, on average, the wealthiest families (in the 10th decile of the previous year's income distribution) will have 13,000 Euros more in the next year than the average income. In comparison, the poorest households (in the first decile) will likely have approximately 5000 Euros less than the average income. We observe a consistent gentle slope until the fifth decile, which gets steeper above the seventh decile in all the years analysed, suggesting that the between-decile variation gets bigger only on the right side of the distribution.

Note that the beeswarm plots on the left side of each panel in the following figure are SHAP summary plots of the best-performing Random forests model during the years. This plot displays the top 25 features of prediction (the top on the y-axis is the most important) and the distribution of the impacts of each predictor on the model prediction, which includes a set of distributions where each dot corresponds to a household. When multiple dots arrive at the same coordinate in the plot, they pile up to show density. The colours correspond to the feature values: red for larger values and blue for smaller ones. A negative SHAP value (extending to the left) shows a decrease in the outcome variable, while a positive (extending to the right) shows an increased value. The long-right tails in the summary plot indicate that the variables are highly predictive for some respondents but not others, i.e., predictors with minor global importance can still be very important for specific respondents. The bar plot on the right side of each panel displays three-fold information. (i) The direction of association captured by the correlation between the feature and SHAP values (red for positive and blue for negative). (ii) The darkness of each colour gradient shows the strength of the direction of the association. (iii) The feature's marginal impact magnitude is measured as the absolute SHAP values average and indicated by the horizontal length of the bars.

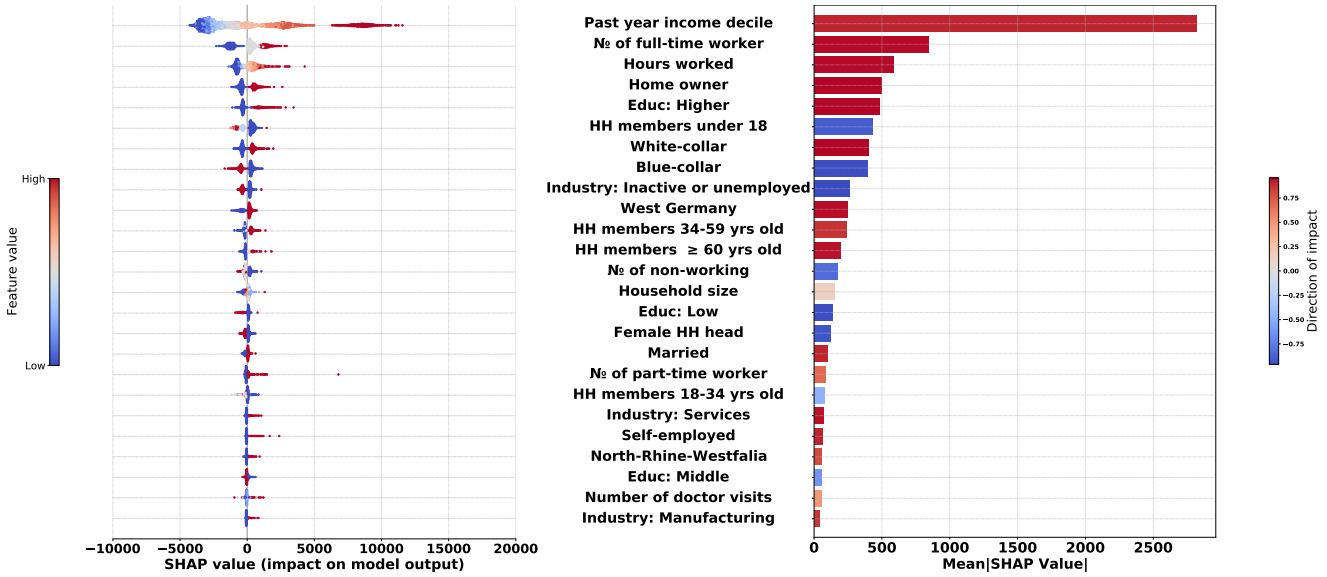
Figure. 1.3 *Key predictors of household income over the years*



(a) *2017*



(b) *2007*



(c) *1997*

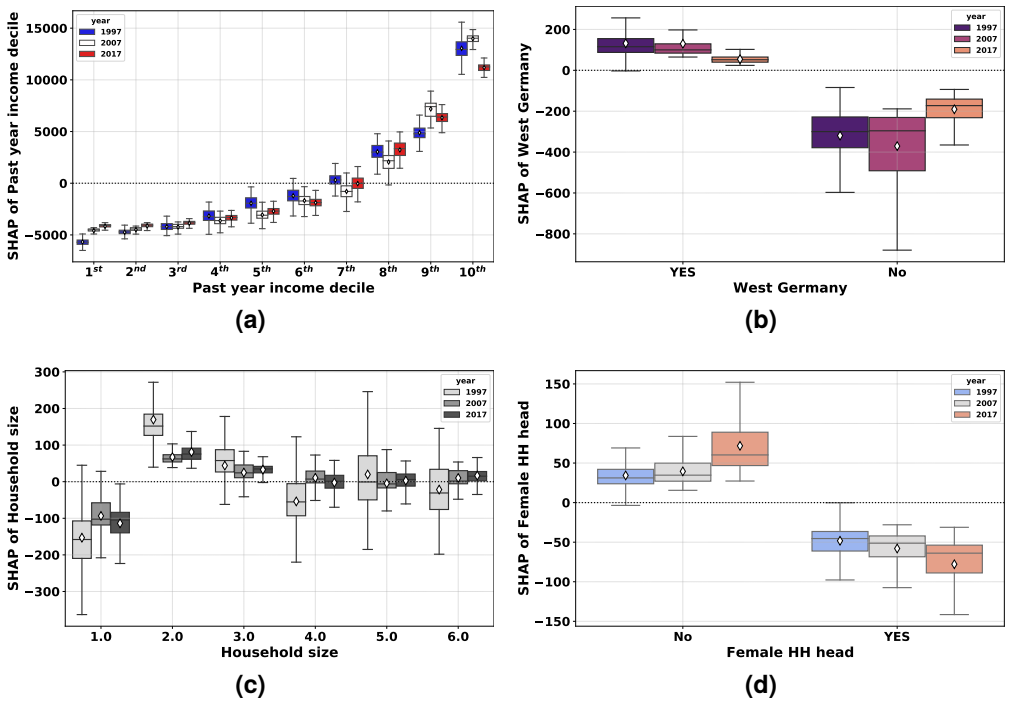
Feature showing the households' strong ties to the labour market is the second most important predictor in 2007 and 1997 and holds the fourth tier in 2017. The number of full-time workers in the family is positively associated with households' disposable income. The magnitude of impact (mean absolute SHAP value) of this variable increased when we drop the past year's household income decile from the model. The number of individuals in part-time employment is also a positive predictor but has a smaller impact than the full-time workers. In contrast, the number of non-working individuals in the households is inversely associated with the target variable with a minor effect on the model. We find a similar positive pattern between hours

worked by the head of the family and household income. We see a sizeable earnings difference between occupational groups: households whose heads of the family work in a blue-collar job are associated with lower income than those whose heads are self-employed or working in white-collar jobs. The relative importance of these features has been relatively consistent across the years.

The highest achieved level of education is the next critical predictor. As can be seen in Figure 1.3 families headed by highly educated individuals tend to have a higher income than those with a middle and lower level of education. This relationship is consistent with the human capital literature (Grossman, 2017; Heckman and Carneiro, 2003; Jolliffe, 2002; Miles, 1997). In terms of feature importance over time, the high education variable has ascended to being the second and the third most important feature in 2017 and 2007, respectively, from the fifth position in 1997. This pattern can be linked to the observed increase in the proportion of degree holders over time in the dataset. The magnitude of the impact of these features diminishes when we control for the household's position in the past year income distribution.

Homeownership status is the other influential predictors of household disposable income and thereby vulnerability to poverty—appear in third, fifth, and fourth positions over the years. Consistent with the positive association between housing wealth and disposable income (and consumption) documented in the literature, (e.g., Aladangady, 2017; Browning et al., 2013). Households living in owner-occupied homes are associated with higher household disposable income than those living in rented housing. In the next tier of feature importance, we see a household's location as a key predictor of a household's disposable income in Germany. Looking at the average income by region of residence depicted in Figure 1.3 (b), we find a persistent difference by region of residence, i.e., households in West Germany are associated with higher household incomes than in East Germany (the former German Democratic Republic). This attributed to the wage differentials between the regions. Previous studies attribute the source of the wage gap to the characteristics of establishments than observable differences in employees' demographics (Gollin et al., 2014; Heise and Porzio, 2019; Moretti, 2011; Smolny and Kirbach, 2011), suggesting that establishments in the West pay a significantly higher real wage than in the East, and labour market frictions prevent workers' reallocation across firms and across spaces. We observe a narrowing of this gap in 2017 compared to the previous decades, stipulating a diminishing impact of regional differences over the three decades.

Figure. 1.3 *Partial effect of past year's income, household size, gender of head of the household & region of residence*



Notes: This figures display SHAP dependence boxplots of the past year income decile and a regional dummy variable in the three years. The diamond symbol in the boxes denotes the average of SHAP value distribution per each category. In panel (iii) for the sake of presentation, we exclude households with a size of seven or bigger which represent only 1% of the sample.

The household's size and family's composition by age are consistently in the top 25 predictors in all the years we analyse. The number of individuals living together affects the income distribution because of income sharing within households. However, we observe a non-monotonic association between the number of individuals in the family and their disposable income, as shown in Figure 1.3 (c). A single-person household is associated with a lower equivalent income; in contrast, two-person families (followed by a family with three members) are associated with the higher one. As the household size increases further, the positive impact vanishes; and the household composition underpins this effect. Consistent with the prediction from a life-cycle model (e.g., Gourinchas and Parker, 2002), a family is less likely to be poor as it comprises more working-age adults (above 34 years of age) and the elderly (60 years of age and above) due to accumulated wealth effect, and fewer children (below 18 years of age). We do not find a similar positive impact of having more young adults between 18 and 34 in the family. One reason for this result is that this age group coincides with the life stage of dependent college students or family formation and childbearing stage, which implies

low or no income. In Figure 1.3 (d), we see a slight persistent disparity in equivalent disposable income according to the gender of the family head: a female-headed household predicts lower income than a family with a non-female head. With respect to family status: households headed by a single person (never married), separated, or divorced is associated with lower equivalent income than households with married or even widowed family heads. Last, disability status negatively predicts the household equivalent income in all the years examined and appears among the top 25 predictors of 2007.

1.5.4 Profiles of Vulnerable Households

Once we can accurately estimate those households that will become poor in the next period and determine the factors associated with household income, the next crucial step is understanding the direct determinants of vulnerability. To do so, we study the predicted vulnerable households for the years with respect to several indicators of household profiles: family status, occupation class, and region of residence. This analysis is based on the Random forests' VP estimate under specification 6 (b) with a poverty cutoff of 60% of the median of the observed equivalized income at time t (shown in Figure 1.2). See Appendix 1.E for the profile analysis that is based on specification 6 (a). For each profile element, the percentage of the predicted vulnerable households is reported in Table 1.2 below.

Table. 1.2 *Sociodemographic profiles of vulnerable households*

Sociodemographic variables	1997	2007	2017
Female HH head	60.88%	60.74%	63.89%
Marital Status:			
Divorced	20.78%	26.70%	26.83%
Married	26.65%	20.96%	23.57%
Separated	8.07%	6.70%	6.16%
Single	20.54%	31.70%	34.66%
Widowed	21.52%	13.19%	8.34%
Occupation class:			
Blue-collar	38.63%	34.47%	34.37%
White-collar	23.72%	27.23%	28.64%
Self-employed	3.18%	2.98%	2.76%
Region: West Germany	7.41%	8.86%	11.22%

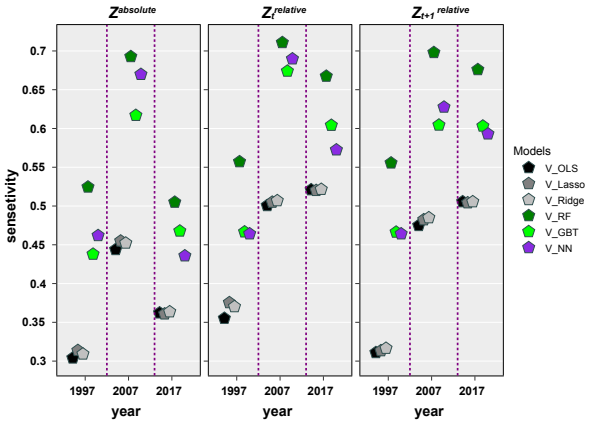
From the profile analysis, in 1997 and 2007, we observe that approximately 61% of vulnerable households have a female household head, and this figure rose to 64% in 2017. Looking at the relationship status of heads of the vulnerable families, we

find that single-person households represent the largest percentage, followed by those families whose leaders are divorced, married, widowed, and separated. These rankings are consistent for the years 2007 and 2017. However, in 1997, most of the vulnerable were families with married heads. This can be because households with such a profile represented the vast majority of the sample in 1997 compared to the other years (see, Table 1.1). Similarly, the percentage of vulnerable single-headed households is lower than in the other years by 10% at least. However, families with widowed heads that were vulnerable were more present in 1997 as opposed to the other years. Next, we assess the vulnerable families with respect to employment status and occupation class of the head of the household. Consistent over the years, families with blue-collar heads represent the largest percentage of the vulnerable, followed by those with unemployed, white-collar, and self-employed profiles. We observe an increasing pattern in the proportion of vulnerable households with white-collar heads over the years, while the opposite is true for blue-collar household heads. Note that this pattern is reflected in the composition of the SOEP samples over the years, as described in Section 2. Moreover, the more significant proportion of households identified as vulnerable was from East Germany. But this percentage slightly dropped (by 4%) from 1997 to 2017, which was also reflected in the partial effect of the region of residence.

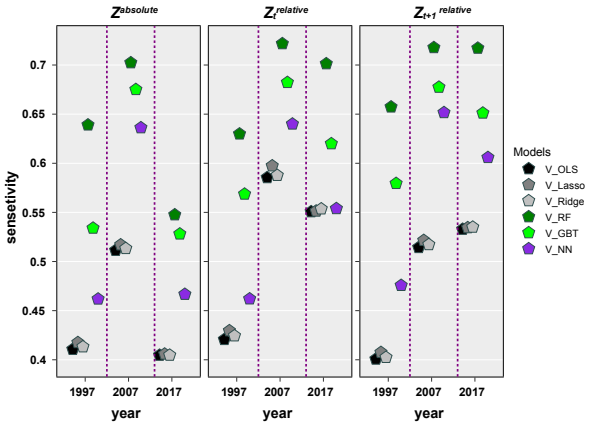
1.5.5 Extending the Time Horizon

Thus far, we have analysed one year ahead vulnerability. We also considered n years ahead vulnerability, i.e., whether the household will fall into poverty in the n period. To be comparable with previous research, for instance, (Hohberg et al., 2018), we expand the time horizon of our prediction to three years ahead of vulnerability to poverty. As the future poverty line is unknown, we use the absolute and relative poverty lines (based on the income distribution in year t and the forecasted income in time $t + 1$). Figure 1.4 displays the results of the ML algorithms and the baseline OLS's sensitivity score under the two specifications.

Figure. 1.4 *Performance of models' in predicting three years ahead VP*



(a) *Specification 6 (a)*



(b) *Specification 6 (b)*

The non-linear ML algorithms have shown powerful predictive performance compared to the standard linear model in identifying if households will become poor in period $t + 3$. However, the level of sensitivity degenerates for all models (looking at the second column of each panel) compared to the one year ahead vulnerability prediction. For instance, the prediction of vulnerability with linear models falls on average by 6.5% with both specifications. Similarly, the VP prediction with Random forests drops by 10% with both specifications.

1.6 Conclusions

This study examines households' vulnerability to poverty as an important measurement that can be used for proactive social protection policies to target those families

who will experience poverty in the future. The most pursued approach in the literature defines "vulnerability as expected poverty", that is, the household's probability of becoming poor in the future. This study proposes a data-driven approach to better predict VP with explainable ML algorithms. Several empirical experiments were conducted using the German Socio-Economic Panel (SOEP), a high-quality longitudinal panel data of private households in Germany. First, we analyse if better accuracy can be obtained without departing from the FGLS's linearity assumption by employing regularised linear ML algorithms. Second, we introduce flexibility by embedding the household income-generating function into non-parametric complex ML algorithms, such as Neural networks, Random forests, and Gradient boosted trees. We ran these experiments under two scenarios. 1) when only cross-sectional data is available vs. 2) when historical information on the relative position of the households in the income distribution is available. Finally, we address the FGLS's arbitrary VP cutoff by considering absolute and relative poverty lines that are standard in poverty analysis. To check whether the results are dependent on the specific cross-sectional year from which VP is predicted, we conducted the analysis retrospectively for multiple years (1997, 2007, and 2017). From the results of this study, the following conclusions can be obtained:

The regularised linear ML algorithms do not enhance the level of predictive sensitivity compared with the baseline OLS. With empirically optimised, flexible non-parametric ML algorithms, such as Random forests, Gradient boosted trees, and Neural networks, one can target VP with better sensitivity. Notably, a solid predictive performance can be obtained with RF. This predictive power is robust to varying the cross-sectional year, extending the time horizon of prediction, and using different vulnerability cutoffs.

A higher sensitivity of VP prediction can be obtained by utilising the previous year's income information. A more significant impact is reflected in the linear models than in the nonlinear ML models. Including multiple past years' household income does not further enhance the predictive power of models on our data. However, uncaptured signals could be fetched by utilising multi-period past information in an environment where household income is highly volatile from year-to-year.

The highly predictive RF model can be effectively interpreted by SHAP values locally and globally. Although feature importance inferred from the coefficients of the OLS estimates reported in the related literature and Shapley values from RF produce results differently, the two models arrive at the same conclusion about the critical predictor of VP. Looking at the Shapley values ranking of the most influential predictors, the past year income decile, variables relating to household's ties to the labour market, human

capital, household's composition (by age and activity status), and region of residence are the potent signal to the sources of VP.

Future studies can complement the highly predictive data-driven approach by non-parametrically estimating the degree to which a given household will face the risk of future poverty. Effective social policy can be devised with an accurate targeting of the disadvantaged group; hence it is essential if researchers assess and report the sensitivity of VP prediction when using panel data. We find a comparable pattern in terms of the predictive power of models for a given cutoff employing absolute and relative poverty concepts. Farther sensitivity can also be achieved by adjusting the cutoff.

References

Aladangady, A. (2017). Housing wealth and consumption: Evidence from geographically-linked microdata. *American Economic Review*, 107(11):3415–46.

Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11(1):685–725.

Azeem, M. M., Mugera, A. W., and Schilizzi, S. (2018). Vulnerability to multi-dimensional poverty: An empirical comparison of alternative measurement approaches. *Journal of Development Studies*, 54(9):1612–1636.

Battistin, E., Blundell, R., and Lewbel, A. (2009). Why is consumption more log normal than income? gibrat's law revisited. *Journal of Political Economy*, 117(6):1140–1154.

Bérgolo, M., Cruces, G., Ham, A., et al. (2012). Assessing the predictive power of vulnerability measures: evidence from panel data for argentina and chile. *Journal of Income Distribution*, 21(1):28–64.

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Browning, M., Gørtz, M., and Leth-Petersen, S. (2013). Housing wealth and consumption: a micro panel study. *Economic Journal*, 123(568):401–428.

Cafiero, C. and Vakis, R. N. (2006). *Risk and vulnerability considerations in poverty analysis: Recent advances and future directions*. World Bank, Social Protection Washington.

Calvo, C. and Dercon, S. (2005). Measuring individual vulnerability.

Celidoni, M. (2013). Vulnerability to poverty: An empirical comparison of alternative measures. *Applied Economics*, 45(12):1493–1506.

Celidoni, M. (2015). Decomposing vulnerability to poverty. *Review of Income and Wealth*, 61(1):59–74.

Chaudhuri, S. (2003). Assessing vulnerability to poverty: concepts, empirical methods and illustrative examples. *Department of Economics, Columbia University, New York*.

Chaudhuri, S., Jalan, J., and Suryahadi, A. (2002). Assessing household vulnerability to poverty from cross-sectional data: A methodology and estimates from indonesia. *Department of Economics, Columbia University, New York*.

Chiwaula, L. S., Witt, R., and Waibel, H. (2011). An asset-based approach to vulnerability: The case of small-scale fishing areas in cameroon and nigeria. *Journal of Development Studies*, 47(2):338–353.

Christiaensen, L. J. and Boisvert, R. N. (2000). On measuring household food vulnerability: Case evidence from northern mali.

Christiaensen, L. J. and Subbarao, K. (2005). Towards an understanding of household vulnerability in rural kenya. *Journal of African Economies*, 14(4):520–558.

Drucker, H. (1997). Improving regressors using boosting techniques. In *ICML*, volume 97, pages 107–115. Citeseer.

Dutta, I., Foster, J., and Mishra, A. (2011). On measuring vulnerability to poverty. *Social Choice and Welfare*, 37(4):743.

Freund, Y., Schapire, R., and Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612.

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

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378.

Gaiha, R. and Imai, K. (2008). *Measuring vulnerability and poverty estimates for rural India*. Number 2008/40. WIDER Research Paper.

Gallardo, M. (2018). Identifying vulnerability to poverty: A critical survey. *Journal of Economic Surveys*, 32(4):1074–1105.

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The german socio-economic panel (soep). *Jahrbücher für Nationalökonomie und Statistik*, 239(2):345–360.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). The agricultural productivity gap. *Quarterly Journal of Economics*, 129(2):939–993.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.

Gourinchas, P.-O. and Parker, J. A. (2002). Consumption over the life cycle. *Econometrica*, 70(1):47–89.

Grimm, M., Waibel, H., and Klasen, S. (2016). *Vulnerability to poverty: Theory, measurement and determinants, with case studies from Thailand and Vietnam*. Springer.

Grossman, M. (2017). 2. The Human Capital Model. In *Determinants of Health*, pages 42–110. Columbia University Press.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):1–42.

Günther, I. and Maier, J. K. (2014). Poverty, vulnerability, and reference-dependent utility. *Review of Income and Wealth*, 60(1):155–181.

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

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC press.

Haughton, J. and Khandker, S. R. (2009). *Handbook on poverty+ inequality*. World Bank Publications.

Heckman, J. J. and Carneiro, P. (2003). *Human capital policy*. National Bureau of Economic Research Cambridge, Mass., USA.

Heise, S. and Porzio, T. (2019). Spatial wage gaps and frictional labor markets. *FRB of New York Staff Report*, (898).

Hoddinott, J. and Quisumbing, M. A. R. (2003). Methods for microeconometric risk and vulnerability assessments. *World Bank Social Protection Discussion Paper Series*, 324.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Hohberg, M., Landau, K., Kneib, T., Klasen, S., and Zucchini, W. (2018). Vulnerability to poverty revisited: Flexible modeling and better predictive performance. *Journal of Economic Inequality*, 16(3):439–454.

Jolliffe, D. (2002). Whose education matters in the determination of household income? evidence from a developing country. *Economic Development and Cultural Change*, 50(2):287–312.

Kamanou, G. and Morduch, J. (2002). *Measuring vulnerability to poverty*. Number 2002/58. WIDER Discussion Paper.

Kamanou, G. and Morduch, J. (2004). Measuring vulnerability to poverty, chapter in dercon, s. *Insurance against Poverty*.

Klasen, S. and Povel, F. (2013). Defining and measuring vulnerability: State of the art and new proposals. In *Vulnerability to poverty*, pages 17–49. Springer.

Kotsiantis, S., Kanellopoulos, D., Pintelas, P., et al. (2006). Handling imbalanced datasets: A review. *GESTS international transactions on computer science and engineering*, 30(1):25–36.

Kühl, J. J. (2003). Disaggregating household vulnerability–analyzing fluctuations in consumption using a simulation approach. *Manuscript, Institute of Economics, University of Copenhagen, Denmark*.

Landau, K., Klasen, S., and Zucchini, W. (2012). Measuring vulnerability to poverty using long-term panel data. Technical report, SOEP papers on Multidisciplinary Panel Data Research.

LeCun, Y. A., Bottou, L., Orr, G. B., and Müller, K.-R. (2012). Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–48. Springer.

Ligon, E. and Schechter, L. (2003). Measuring vulnerability. *Economic Journal*, 113(486):C95–C102.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Mason, L., Baxter, J., Bartlett, P., and Frean, M. (1999). Boosting algorithms as gradient descent. *Advances in neural information processing systems*, 12.

McCarthy, N., Brubaker, J., and De La Fuente, A. (2016). Vulnerability to poverty in rural malawi. *World Bank Policy Research Working Paper*, (7769).

McDonald, J. B. and Ransom, M. (2008). The generalized beta distribution as a model for the distribution of income: estimation of related measures of inequality. In *Modeling income distributions and Lorenz curves*, pages 147–166. Springer.

Miles, D. (1997). A household level study of the determinants of incomes and consumption. *Economic Journal*, 107(440):1–25.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Moretti, E. (2011). Local Labor Markets. volume 4B, pages 1237–1313. Elsevier.

Mullainathan, S. and Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830.

Pritchett, L., Suryahadi, A., and Sumarto, S. (2000). *Quantifying vulnerability to poverty: A proposed measure, applied to Indonesia*. Number 2437. World Bank Publications.

Schapire, R. E. (2003). The boosting approach to machine learning: An overview. *Nonlinear estimation and classification*, pages 149–171.

Shapley, L. S. (1953). A value for n-person games. In Kuhn, H. W. and Tucker, A. W., editors, *Contributions to the Theory of Games (AM-28), Volume II*, pages 307–318. Princeton University Press.

Smolny, W. and Kirbach, M. (2011). Wage differentials between East and West Germany: are they related to the location or to the people? *Applied Economics Letters*, 18(9):873–879.

Sohn, A., Klein, N., Kneib, T., et al. (2015). A semiparametric analysis of conditional income distributions. *Schmollers Jahrbuch*, 135(1):13–22.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Wolpert, D. H. (2002). The supervised learning no-free-lunch theorems. In *Soft computing and industry*, pages 25–42. Springer.

Wolpert, D. H. and Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82.

Zereyesus, Y. A., Embaye, W. T., Tsiboe, F., and Amanor-Boadu, V. (2017). Implications of non-farm work to vulnerability to food poverty-recent evidence from northern ghana. *World Development*, 91(0305-750X):113–124.

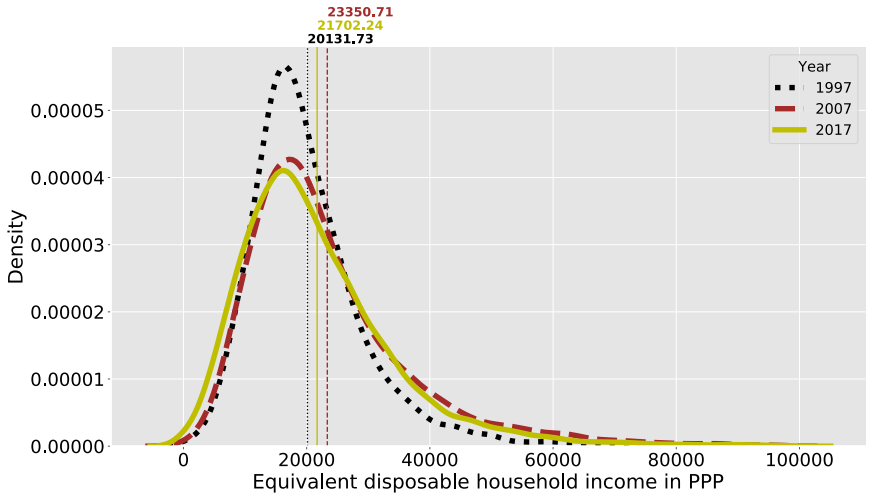
Zheng, A. and Casari, A. (2018). *Feature engineering for machine learning: principles and techniques for data scientists*. " O'Reilly Media, Inc.".

Appendices

1.A Distribution of HH Equivalent Income and its Evolution

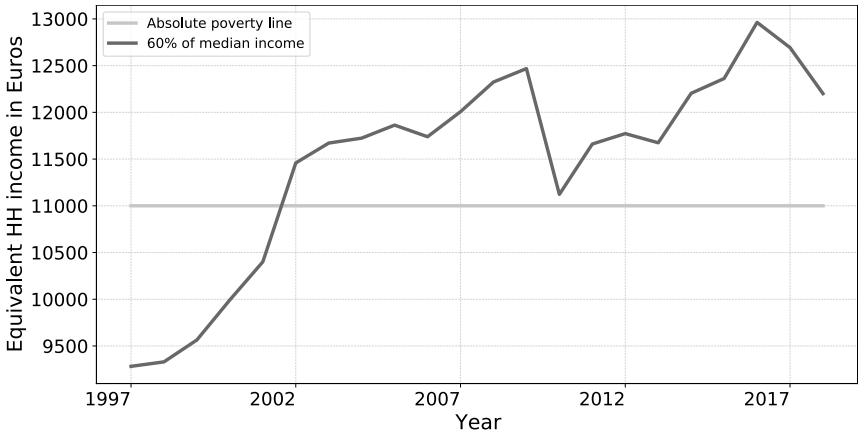
Figure 1.5 shows the distribution of the dependent variable in the three years. The distribution's shape has changed over time, with a dramatic shift of the density from the middle towards the right and left tail of the distribution, showing increased average equivalent disposable household income from 1997 to 2007. Likewise, in 2017, the distribution spread with a more pronounced shift towards the right tail, resulting in a decreased mean distribution from 2007; this reveals the high mobility of households from the middle class to the upper and lower class during these years.

Figure. 1.5 *Evolution of household income distribution in 1997, 2007, and 2017*



Notes: This figure shows the evolution of the empirical densities of equivalent household disposable income. The annotated figures correspond to the mean value of each distribution.

Figure. 1.6 *Absolute and relative poverty lines in Germany 1997 to 2017*



1.B How we Train the Models

In this section we describe the prediction process for the supervised ML models more specifically on sampling of the data, tuning of the parameters and how to evaluate them. In all prediction exercises we used six models as implemented in the Scikit-learn.

Data Pre-processing and Overall Workflow

For the VP prediction exercise described in the main part of the paper, we want to train a generalizable estimate of income-generating function (as outlined in equation 6a and 6b) that can be used to predict VP status of households. In training these models (including the baseline OLS model), we followed the following workflow.

Before training the models, we perform the necessary feature engineering/pre-processing (Zheng & Casari, 2018). Our covariates come in different forms: (a) numerical discrete, e.g., household size and the number of hours worked, (b) dichotomous variables, such as gender and disability status, (c) categorical ordinal, namely last year's income decile, and (d) categorical non-ordinal, e.g., region of residence, marital status, occupation class. Assigning an integer value, say from 1 to k, to each of k possible categories of non-ordinal variables makes the model assume a natural ordering between categories, resulting in poor performance or unexpected results. Therefore, we apply one-hot encoding, a technique to create a group of dummy variables—each variable represents a possible category if the variable does not belong to multiple categories at once. Moreover, heterogeneous scales of features may generate incomparable or convoluted importance in models that are smooth functions of the input or anything that

involves a matrix. On the other hand, the tree-based algorithms are insensitive to the monotonic transformation of features. Quick convergence in the Neural networksalso requires the predictors to be on the same scale. Hence, we train the linear models and NN using standardised predictors.

After random splitting of the datasets. On the training sample, we run each of the algorithms: For models that do not require any tuning, e.g., OLS, we simply fit the model on the full training sample and store estimated function. On the other hand, for algorithms that involve a regularisation parameter, we empirically select the optimum hyperparameters on the training sample through grid search with cross-validation. Once we have chosen an optimal hyperparameter, we re-train the models with tuned hyperparameter on the full training sample.

We next assess the performance of each algorithm in predicting household income on the 20% unseen households. We employ the commonly used evaluation metrics in regression settings, namely Mean Absolute Errors (MAE) and the (R^2), to evaluate how intelligently the models predict for out-of-sample households.

$$\text{MAE}^s = \frac{1}{n}\sum_{i=1}^{n} |y_i^s - \widehat{y}_i^s \mid \text{ and } R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i^s - \widehat{y}_i^s\right)^2}{\sum_{i=1}^{n}\left(y_i^s - \overline{y}^s\right)^2} \tag{1.11}$$

where $s \in \{\text{Trainset}, \text{Testset}\}$. The best model is the one that has a smaller value in both MAE (ideally close to zero) and a higher R^2, meaning that it produces predictions that are very close to the true responses. Moreover, the prediction accuracy on the train and test set should be comparable enough to establish the model's generalisability.

Tuning Hyperparameters and Inspecting CV Results

For each ML algorithms, we select complexity parameters using Grid search with five-fold cross-validation is applied to identify the optimal hyperparameters in each model that control over-fitting and performed in three steps:

i) We randomly partition the training set into five subsamples

ii) For every possible combination of hyperparameters, we train the models on the four subsamples and generate predictions for the fifth, and

iii) Repeat this process five times so that each subsample is used only once to generate the prediction. The optimal hyperparameters are updated based on the average of these ten prediction results. The tuning step ends when we find an optimal

hyperparameter that produces minimum average prediction errors (or high R-squared).

Inspecting the results of empirical tuning

The hyperparameter spaces we search over are a crucial input to the empirical tuning exercise, and the tuning process will be poor if the optimal hyperparameter is outside these spaces. To understand whether this could be the case, we inspect the cross-validated performance by hyperparameter values and check whether the chosen hyperparameter value is inside the range provided.

The figures below show the hyperparameter tunning of specification 6 (a) and 6 (b) on the three panels. For the sake of parsimony, we discuss in text the chosen hyperparameters of models that are trained on the 2017 sample.

In the regularised linear models, the selection of the complexity parameter is crucial, and the optimal level is often determined via grid search with cross-validation (see Hastie et al., 2017; Varian, 2014, for a detailed explanation). From Figure 1.7 (c) and Figure 1.8 (c), it is noted that the regularisation degree of 0.084 in LASSO and 0.001 in the Ridge regression produce better performance with an average test R-squared of 0.44 in both models. However, when a feature measuring the past year's household income decile is included, the model performance vastly increases to 0.665 in LASSO and 0.664 in RIDGE with a regularisation parameter of 0.263, 0.009, respectively seen in Figure 1.7 (f) and Figure 1.8 (f).

Figure. 1.7 *LASSO*

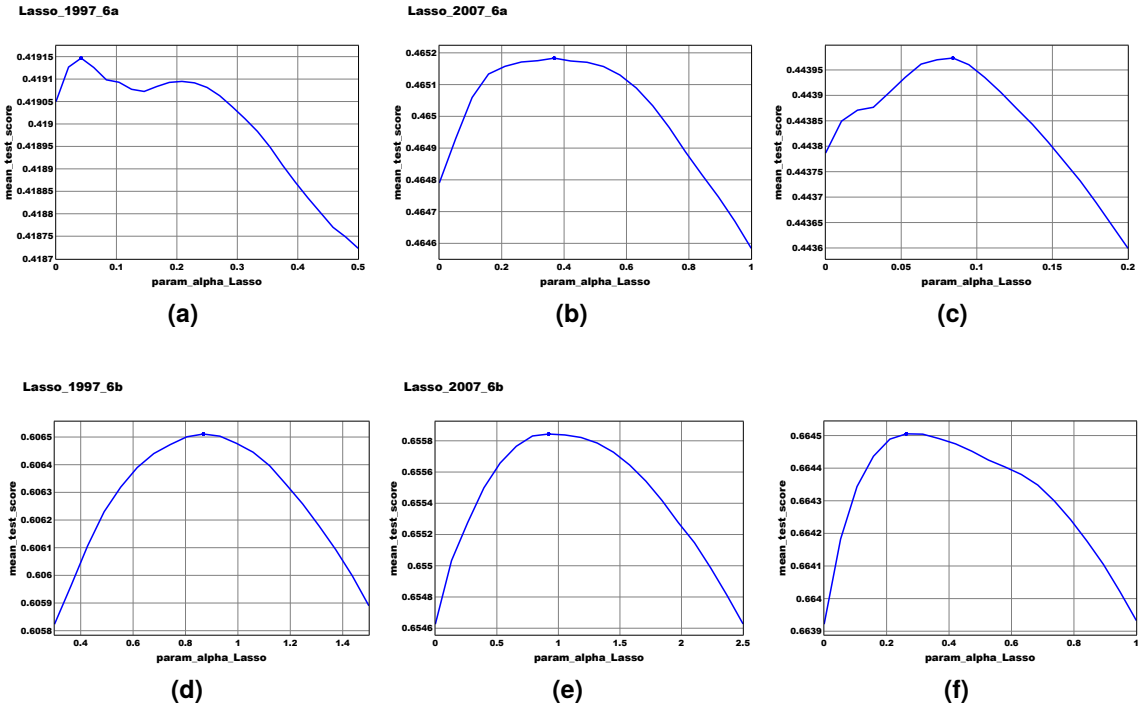
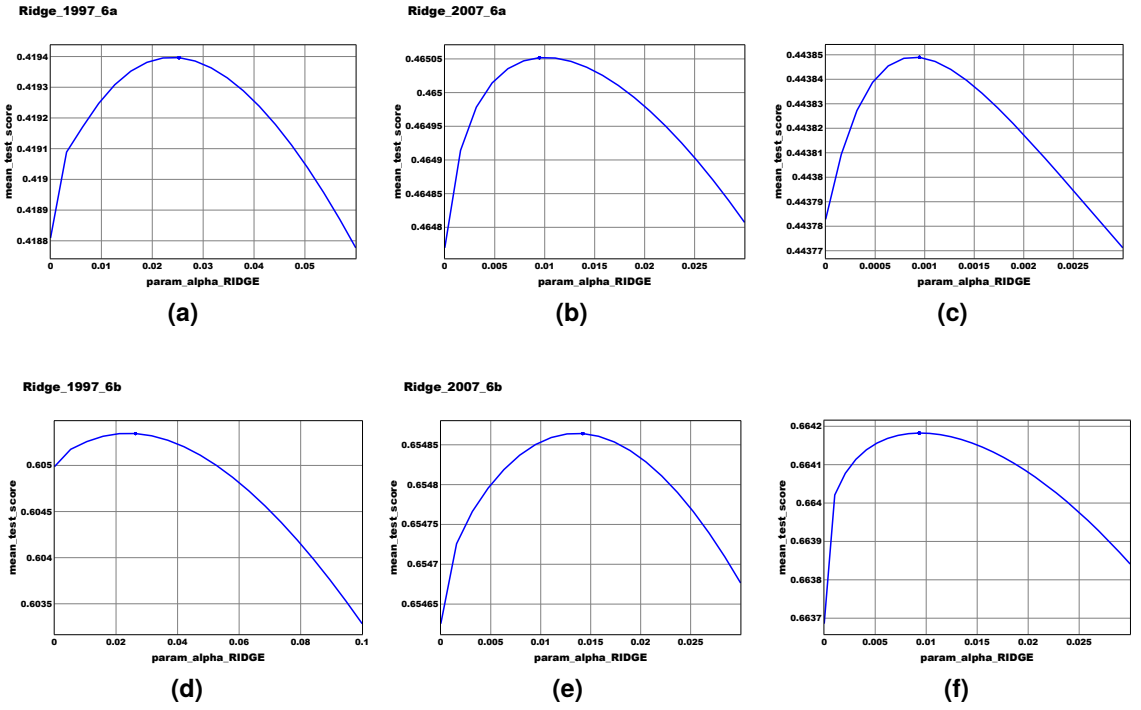
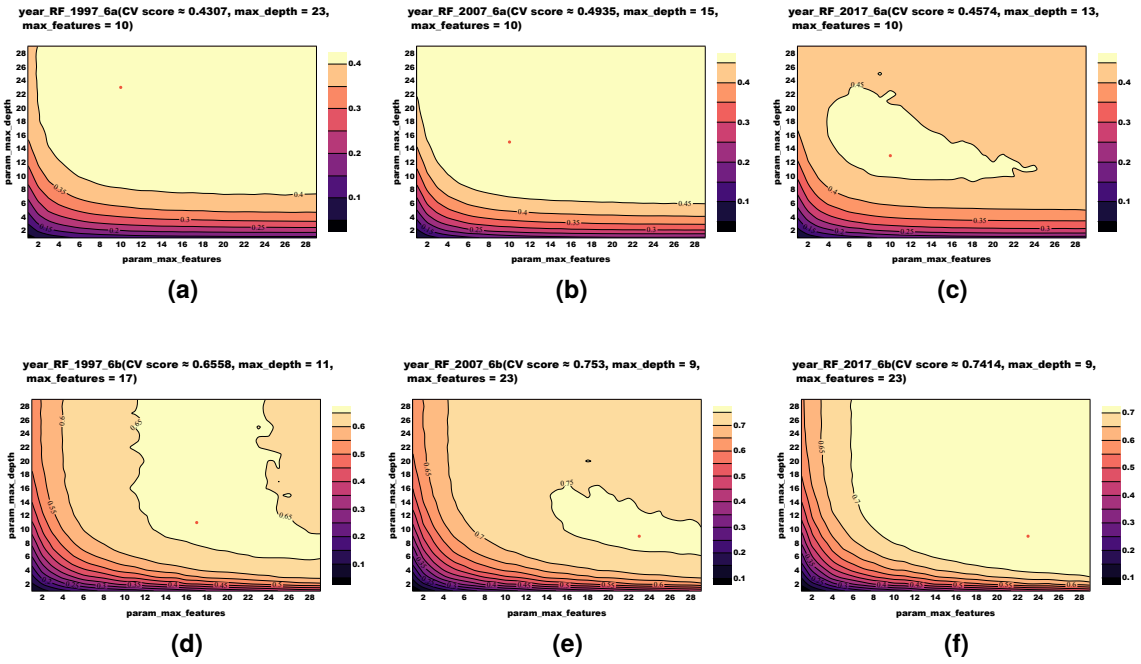


Figure. 1.8 *Ridge regression*



In this study, we find that tuning hyperparameters, such as the maximum depths of trees (*'max_depth'*) and the number of features used in each split (*'max_feature'*), are consequential to the predictive accuracy. The max_depth controls the model's complexity—i.e., the shallower the trees, the more observations end up in each terminal node resulting in poor performance on the training set. At the same time, the max_feature dictates the size of random subsets of features from which the best split is chosen as we split nodes while growing the trees. Setting max features very low could lower the model's variance but worsen the model's performance on the training set. Therefore, selecting a proper value of max_depth and max_feature is paramount (Probst et al., 2019). From Figure 1.9 (c) and (f), it can be concluded that when the max_depth equals 13 and the max_feature equals 10, the RF performance is the greatest, with the average R-squared of 0.457. Once we include the past year's income information, the optimal max_depth is 9 and max_feature is 23 that result in the best performance of the model with an R-squared of 0.745.

Figure. 1.9 *Random forests*



Although the RF performs better than the regularised linear models in both specifications, we check if we can further improve the prediction accuracy with GBT. The Gradient boosted trees are usually more sensitive to hyperparameters settings than RF but can provide better accuracy if the hyperparameters are set correctly (Müller &

Guido, 2016). In this study, the learning process is controlled by optimising selected hyperparameters: such as the number of boosting stages (*'n_estimator'*), the maximum depth of the tree (*'max_depth'*) and learning rate $(0 <$ 'learning_rate' $\leq 1)$. The reasons for choosing these hyperparameters are increasing n_estimator improves the accuracy of the training set but setting it too high may lead to overfitting. Analogous to RF, the role of max_depth here too is to control the model's complexity. The learning_rate regulates the contribution of each boosting iteration: setting the learning rate lower requires more boosting iterations. From the obtained result seen in Figure 1.10 (c) and (f), when n_estimators is 275, the learning rate equals 0.057, and max_depth equals 3, the model performance is the greatest with 0.478 mean R-squared on the test set. The R_squared of the GBT model increased to 0.74 when we ran the grid search by incorporating the past year's household income. The selected values of n_estimators, learning rate, and max_depth equal 175, 0.036, and 3, respectively.

Figure. 1.10 *Gradient boosted trees*

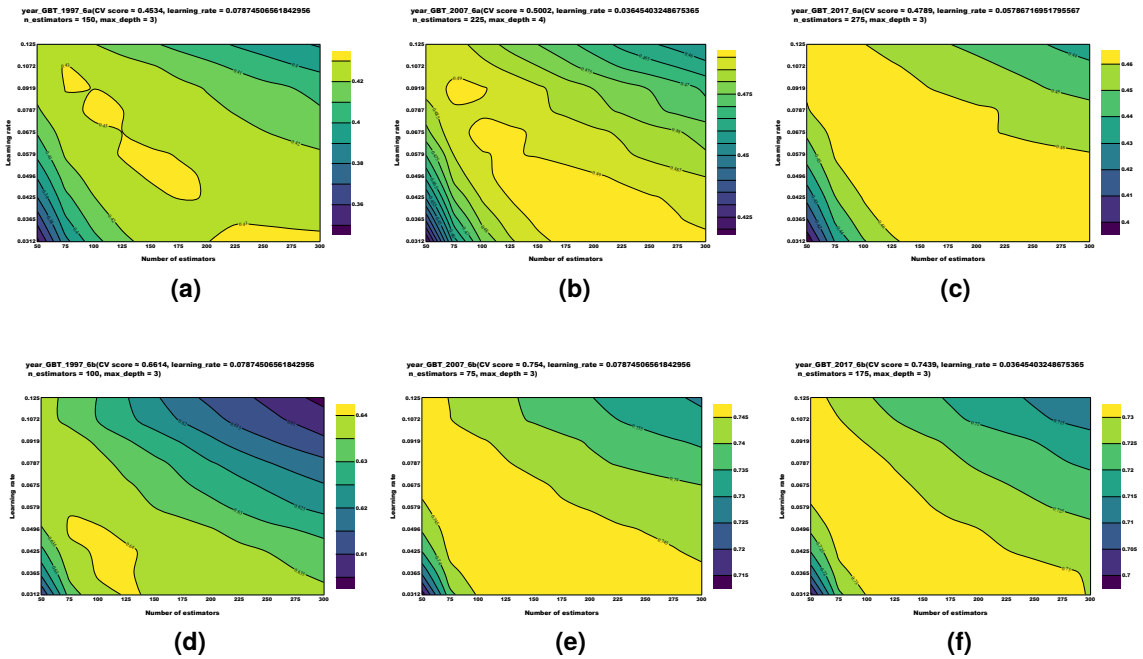
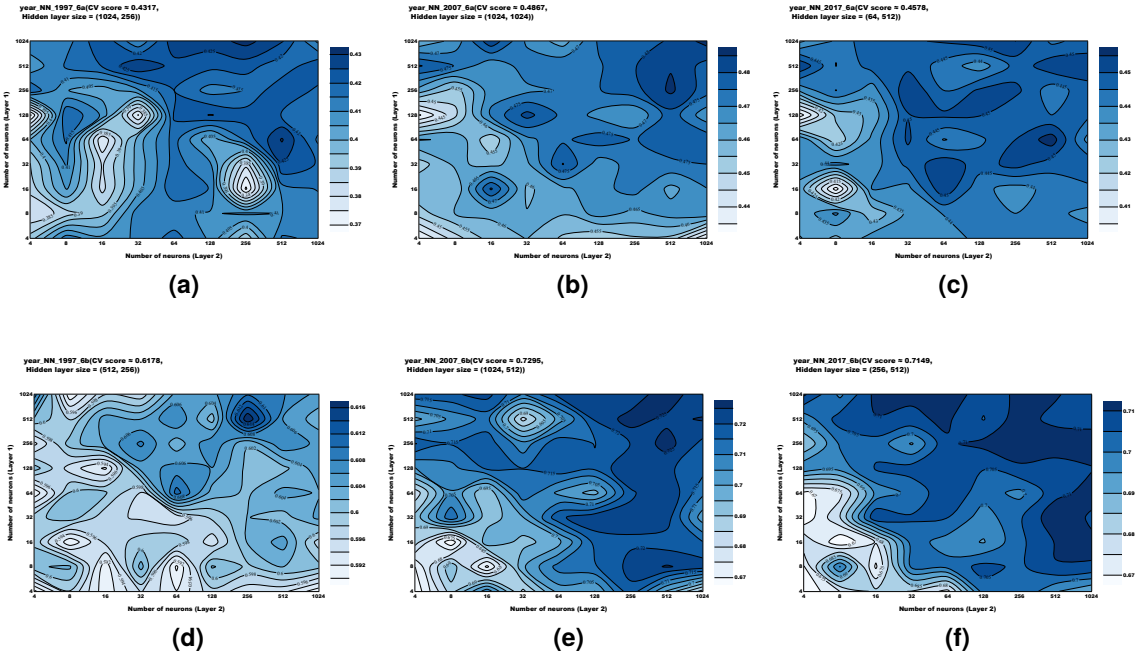
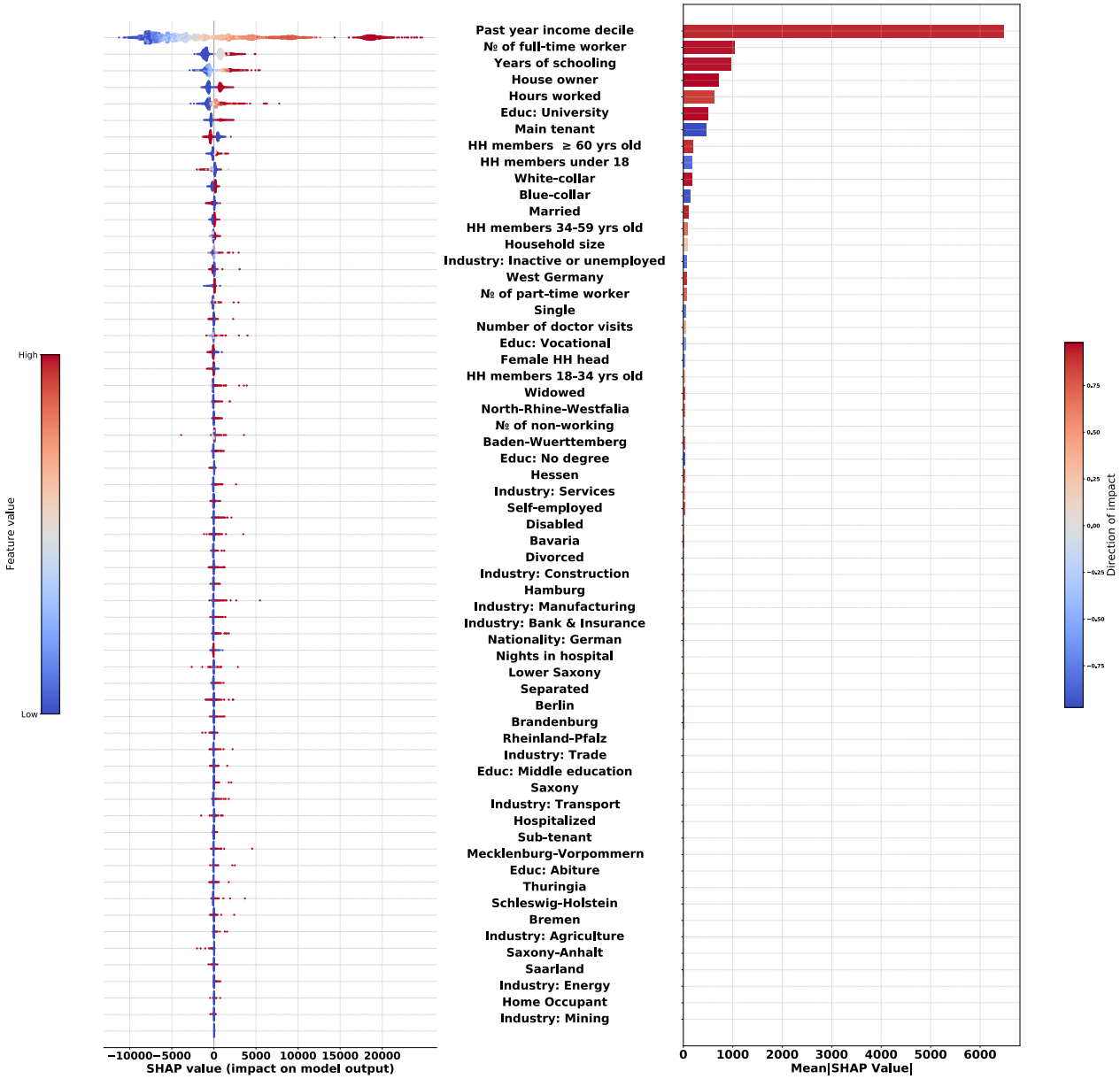


Figure. 1.11 *Neural networks*

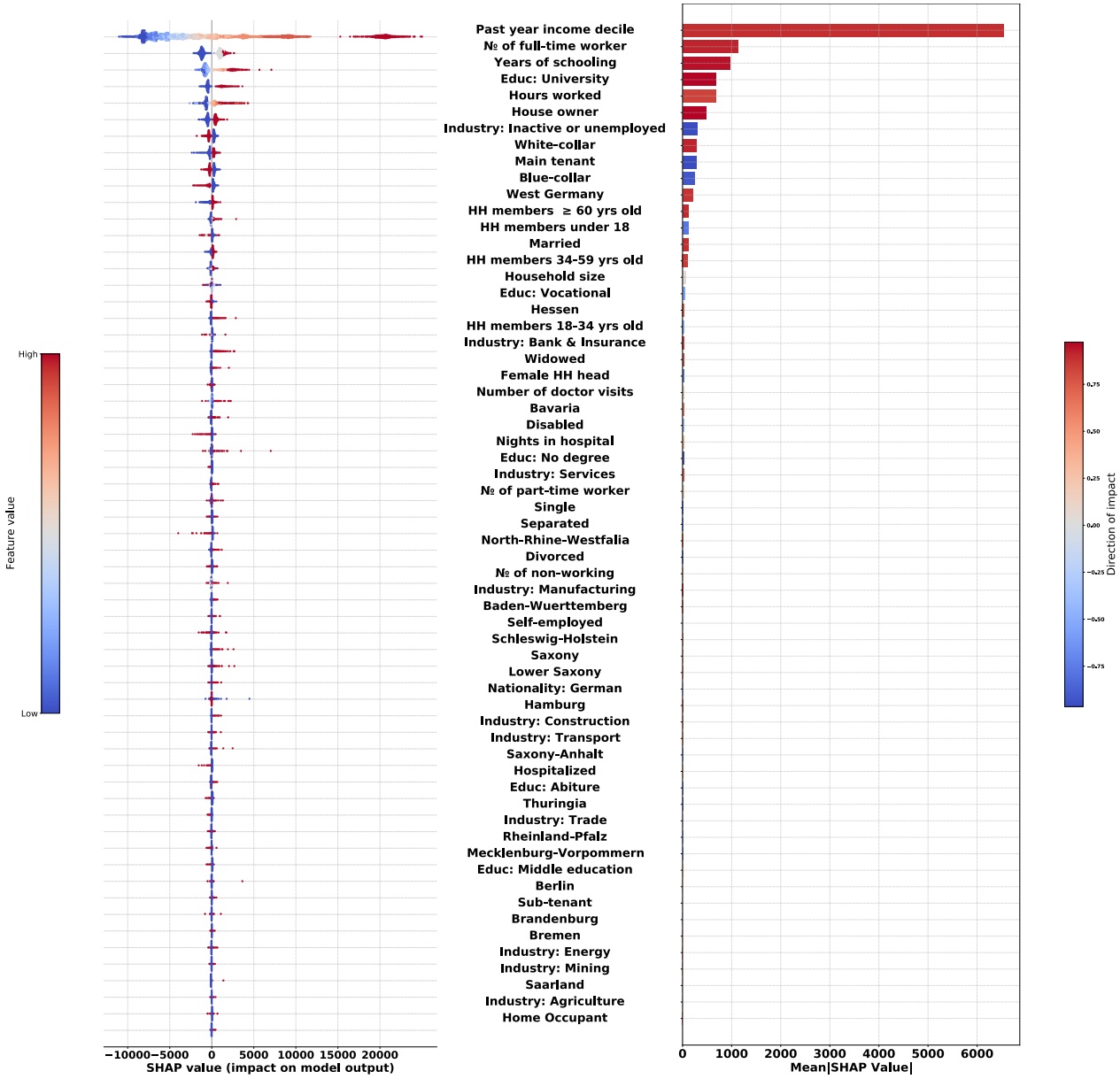


1.C Shapley Values of all Features for all Years (equation 6b)

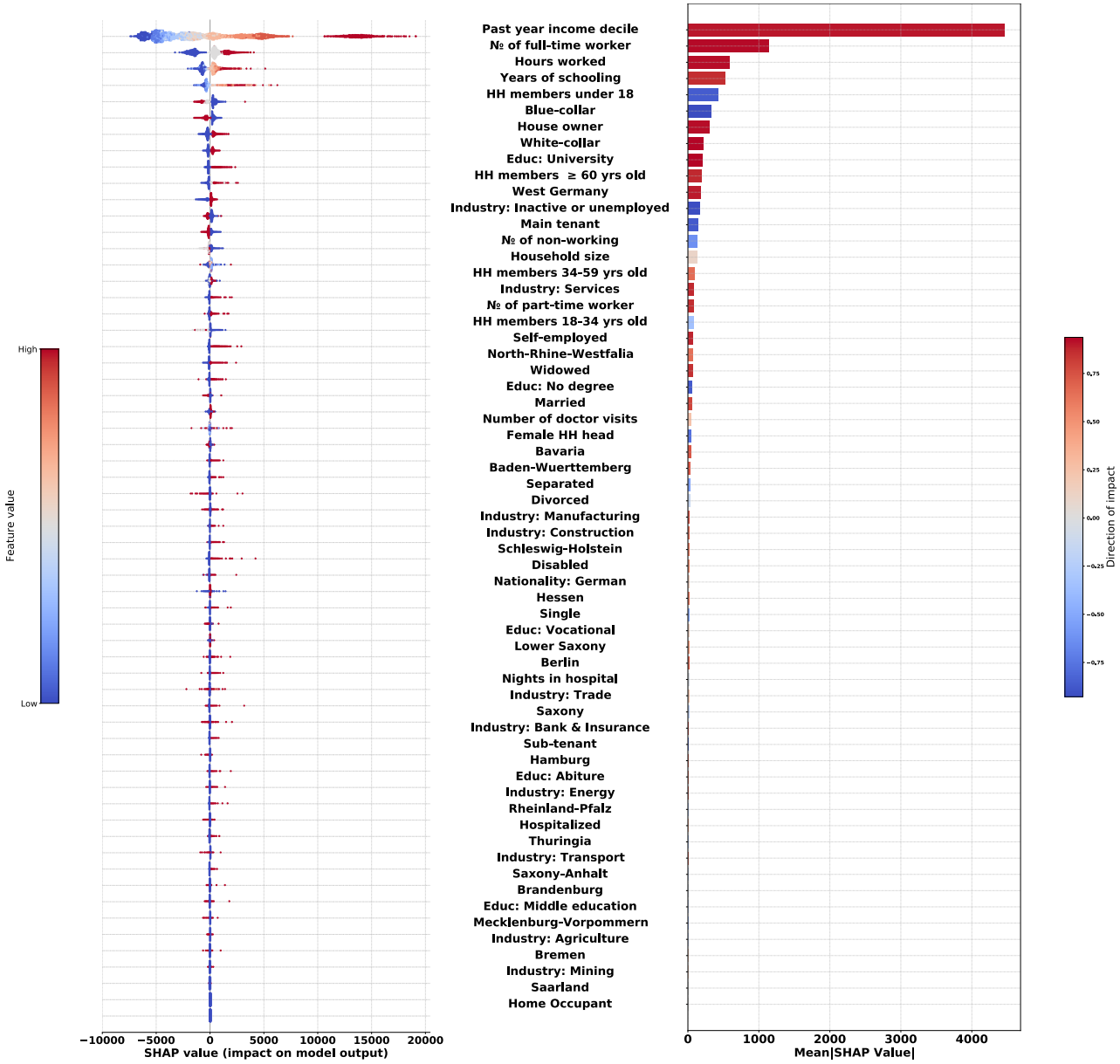
Figure. 1.12 *SHAP summary plot of all features*



(a) *2017*


Past year income decile
№ of full-time worker
Years of schooling
Educ: University
Hours worked
House owner
Industry: Inactive or unemployed
White-collar
Main tenant
Blue-collar
West Germany
HH members ≥ 60 yrs old
HH members under 18
Married
HH members 34-59 yrs old
Household size
Educ: Vocational
Hessen
HH members 18-34 yrs old
Industry: Bank & Insurance
Widowed
Female HH head
Number of doctor visits
Bavaria
Disabled
Nights in hospital
Educ: No degree
Industry: Services
№ of part-time worker
Single
Separated
North-Rhine-Westfalia
Divorced
№ of non-working
Industry: Manufacturing
Baden-Wuerttemberg
Self-employed
Schleswig-Holstein
Saxony
Lower Saxony
Nationality: German
Hamburg
Industry: Construction
Industry: Transport
Saxony-Anhalt
Hospitalized
Educ: Abiture
Thuringia
Industry: Trade
Rheinland-Pfalz
Mecklenburg-Vorpommern
Educ: Middle education
Berlin
Sub-tenant
Brandenburg
Bremen
Industry: Energy
Industry: Mining
Saarland
Industry: Agriculture
Home Occupant
−10000 −5000 0 5000 10000 15000 20000
SHAP value (impact on model output)
0 1000 2000 3000 4000 5000 6000
Mean|SHAP Value|
High
Low
Feature value
Direction of impact

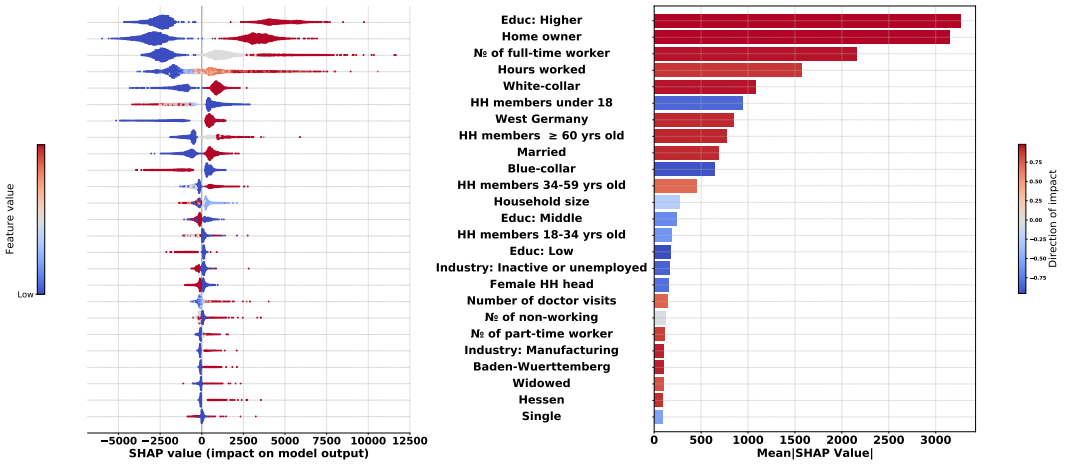
(a) *2007*


Past year income decile
№ of full-time worker
Hours worked
Years of schooling
HH members under 18
Blue-collar
House owner
White-collar
Educ: University
HH members ≥ 60 yrs old
West Germany
Industry: Inactive or unemployed
Main tenant
№ of non-working
Household size
HH members 34-59 yrs old
Industry: Services
№ of part-time worker
HH members 18-34 yrs old
Self-employed
North-Rhine-Westfalia
Widowed
Educ: No degree
Married
Number of doctor visits
Female HH head
Bavaria
Baden-Wuerttemberg
Separated
Divorced
Industry: Manufacturing
Industry: Construction
Schleswig-Holstein
Disabled
Nationality: German
Hessen
Single
Educ: Vocational
Lower Saxony
Berlin
Nights in hospital
Industry: Trade
Saxony
Industry: Bank & Insurance
Sub-tenant
Hamburg
Educ: Abiture
Industry: Energy
Rheinland-Pfalz
Hospitalized
Thuringia
Industry: Transport
Saxony-Anhalt
Brandenburg
Educ: Middle education
Mecklenburg-Vorpommern
Industry: Agriculture
Bremen
Industry: Mining
Saarland
Home Occupant
−10000 −5000 0 5000 10000 15000 20000
SHAP value (impact on model output)
0 1000 2000 3000 4000
Mean|SHAP Value|
High
Low
Feature value
Direction of impact

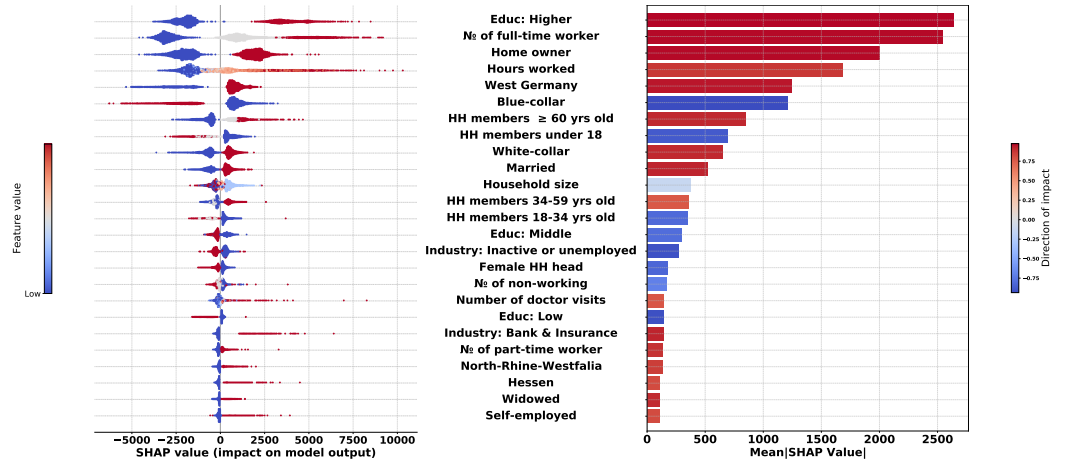
(a) *2007*

1.D Key Predictors without the Past Income Decile in the Model

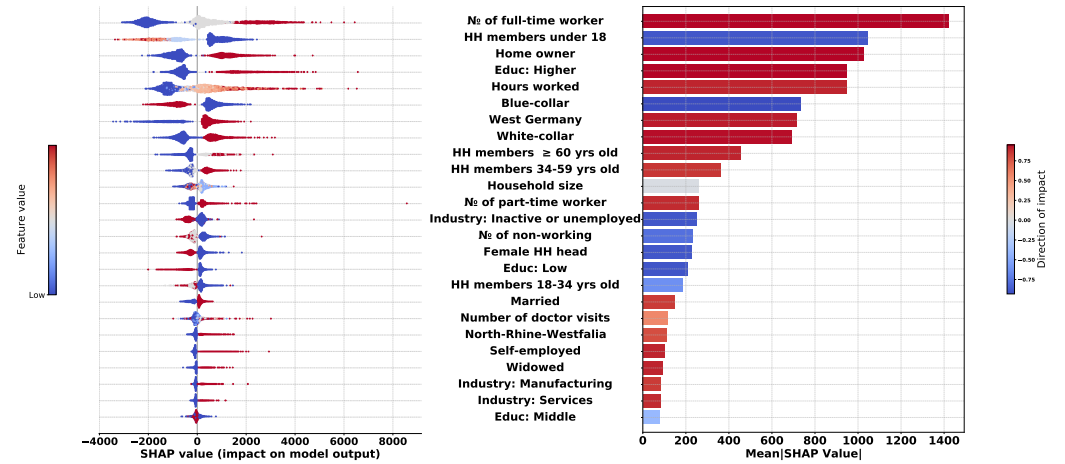
Figure. 1.13 *Top 25 predictors when historical income is not accounted for*



(a) *2017*



(b) *2007*



(c) *1997*

1.E Profiles Analysis of Vulnerable Households Under Specification 6 (a)

Table. 1.3 *Sociodemographic profiles of vulnerable households* 6 (*a*)

Sociodemographic variables	1997	2007	2017
Female HH head	62.55	61.64	66.13
Marital status			
Divorced	27.64	34.18	28.64
Married	22.91	8.55	15.82
Separated	9.45	7.64	7.35
Single	20.36	45.82	45.08
Widowed	18.18	3.09	2.99
Occupation class:			
Blue-collar	44	44	38.98
Self-employed	2.55	2.36	1.87
White-collar	21.09	25.09	25.65
Region: West Germany	4.94	4.73	6.44

Chapter 2

Predicting Material and Social Deprivations with ML*

*The data used in this study come from the European Union Statistics on Income and Living Conditions (EU-SILC).

2.1 Introduction

The debate on who to consider as poor was largely influenced by cultural, political and country-specific factors (Rowntree, 1901; Minter, 1992; Saunders, 2004; Townsend, 1970; Sen, 1982). Two competing approaches are prevalent. Poverty as well-being losses that is 'indirectly' measured in monetary terms; and poverty as 'directly' assessed by looking at the outcome or the final conditions of individuals, which are generally perceived as well-being or standards of living (Townsend, 1979; Sen, 1985, 1992, 2009). There is a consensus in the literature that these two approaches are complementary instead of substitutes as they do not perfectly overlap—an income poor individual may not be deprived, while a non-poor individual might suffer from deprivations (Perry, 2002; Nolan et al., 1996). Several countries and multi-national organisations have already employed material deprivation as one of crucial measures for policy analysis.[1] At the same time, there is a flurry of empirical studies predominantly examining the role of key policy variables in affecting material deprivations (for instance, Nelson, 2012; Saltkjel and Malmberg-Heimonen, 2017; Julkunen, 2002). However, despite the widespread use of this measure, the question of how accurately one can identify individuals based on their risk of deprivation remains unexplored. Being able to identify those individuals with accuracy and in a cost-effective manner is imperative for policy design to direct social programs to the deprived and facilitate the optimal allocation of limited resources.

Using state-of-the-art supervised machine learning (ML) algorithms, such as Extreme gradient boosted trees (Chen and Guestrin, 2016) and Random forests (Breiman, 2001) grounded in economic theory and data, this study answers the questions: 1) How accurately can one classify unseen/out-of-sample individuals' deprivation status given their observable personal, household, and country-specific factors? 2) What is the performance of targeting subsets of indicators, such as sociodemographic, socioeconomic, health, and location, to identify the deprived? 3) What are the key predictors and their partial effects? I use the EU-SILC (the European Union Statistics on Income and Living Conditions) microdata. The empirical analysis is based on the 2014 and 2018 cross-sectional EU-SILC User Data Base (EUSILC UDB), a nationally representative sample of 457,475 and 416,973 individuals aged 16 or above in European countries covering a range of material and social deprivation indicators, sociodemographic, socioeconomic, and health variables.

[1]For instance, the European Union (EU) utilise at-risk of poverty or social exclusion (AROPE) to assess poverty in the member countries. The AROPE indicator is based on three sub-dimensions: income poverty or material and social deprivation or low work intensity (Eurostat, 2022b; Guio et al., 2016).

The results indicate that the ML models, specifically the Extreme gradient boosted trees (Xgboost) algorithm excels at accurately classifying unseen individuals: Xgboost and Random forests achieved an accuracy of 88% and 86% based on classification accuracy evaluated by the area under the ROC curve (Wodon, 1997; Hanley and McNeil, 1982). Put differently, the relative accuracy gained by using the more sophisticated algorithm is positive and significant as compared to that of logistic regression (7.3% relative gain with Xgboost and 5.9% with the random forest). The socioeconomic and location features have the best classification power compared to sociodemographic and health characteristics. The relative feature importance identified with Shapley's values shows that the individual's relative economic position to others, general health status, level of education, age, and housing wealth is the most prominent predictor of deprivation. These results are robust to using different cross-sectional years.

This study makes several contributions to the related literature. First, as far as the author is concerned, this study is the first to evaluate the usefulness of the non-parametric extreme gradient boosted trees classifier algorithm augmented with Shapley values to model the complex nature of deprivations. By comparing the ML algorithms with that of the Generalised linear model (i.e., logistic regression) and documenting their heterogeneity by country. Second, using the newly revised 13-item deprivation indicators proposed by Guio et al. (2016), this study analyses deprivations in material and social aspects instead of those based on the old 9-item indicators, which mainly capture household-level hardship in durable goods. Finally, while remaining highly relevant to the previous studies, this study reveals non-linear relationships between key variables and the likelihood of poor living conditions.

The remainder of the chapter is structured as follows. I describe the dataset, define the outcome variable and provide summary statistics in Section 2.2. Section 2.3 presents the method of the analysis. In Section 2.4 are contained the results and discussion, and Section 2.5 presents concluding remarks and future direction.

2.2 Data and Summary Statistics

2.2.1 Data

This study relies on the European Union Statistics on Income and Living Conditions (EU-SILC) as a data source. The EU-SILC provides the reference source of statistics on income, poverty, social exclusion, and living conditions in the EU member countries (Eurostat, 2022a). The survey also covers other domains of living conditions and

their determinants, which enables the analysis of the multidimensional phenomena of poverty and social exclusion and the joint analysis of its different dimensions. The main analysis are based on the 2018 cross-sectional EU-SILC User Data Base (EUSILC UDB 2018), which contains a nationally representative sample of 457,475 individuals aged 16 or above in 28 European countries. The 2014 cross-sectional EU-SILC User Data Base (EUSILC UDB 2014), which contains a nationally representative sample of 416,973 individuals is used to check the robustness of the prediction to different survey year. In the dataset individuals provide rich micro-level information on material and social deprivation indicators, demographic information, socioeconomic characteristics, education status, labour market information, and health status. In the subsequent paragraphs, I explain how I define the outcome variable and measurement of features used in the prediction exercises.

2.2.2 Outcome Measurement

To measure multidimensional aspects of individual living standards, I employ the Guio et al.'s (2016) newly proposed thirteen-item scale material and social deprivation (henceforth MSD) indicators. The MSD items are based on responses to questions that inquire if respondents experience forced inability in the past twelve months to: 1) "face unexpected financial expenses," 2) "afford a one-week annual holiday away from home," 3) "avoid arrears on mortgage or rent, utility bills or hire purchase instalments in the last 12 months," 4) "afford a meal with meat, chicken, fish (or vegetarian equivalent) every second day," 5) "afford to keep their home adequately warm," 6) "replace worn-out furniture," 7) "have access to a car/van for personal use," 8) "replace worn-out clothes with some new ones," 9) "have two pairs of properly fitting shoes", 10) "spend a small amount of money each week on him/herself (pocket money)," 11) "have regular leisure activities," 12) "get together with friends/family for a drink/meal at least once a month," 13) "have an internet connection". The MDS indicators reflects deprivations at both household-level (item 1-7) and at the individual level (item 8-13). Under the assumption of resource sharing within the same household, individuals in the same families are assigned identical deprivation status in items collected at the household level. Unlike the old 9-item material deprivation indicators, the revised items account for economic strain or financial stress (item 1-6), enforced lack of durable (item 7 & 13), basic needs (item 8 & 9), social inclusion or activity (item 10-12). To correctly label individuals based on their deprivations status, I first re-coded the items in the

same direction as binary variables, i.e., deprivation indicator takes 1, if the person does not have the item because she cannot afford it; otherwise it takes value 0. Regarding household's arrears (embedded in item 3), the household is said to be non-deprived in this dimension if it has never been in any type of arrears or if the questions does not apply to the household, that is the case when the respondent either owns a house or is living rent free, pay no utility bills or no loan payment. Last, respondents with missing information in any of the deprivation items are excluded from the analysis. Then individuals are identified as materially and social deprived if they lack at least 5 items of the above 13 (Guio et al., 2016).

2.2.3 Feature Set and Summary Statistics

I consider a host of micro-level features related to demographic, health status, socioeconomic, and household characteristics. These variables include age, gender, marital status, education, employment status, occupation class, household income, home ownership status, and health status. The comprehensive list of features and how each predictor was constructed is discussed in Appendix 2.A.

Table 2.1 presents the summary statistics of key characteristics of individuals in the pooled and country-level samples. In the pooled sample, 457475 respondents were included, mainly consisting of working-age individuals (mean age of 52); with upper secondary level of education; with the self-assessed general health status of "good"; and 53% of the sample is composed of female respondents. The average household comprises nearly three members and has approximately 17111 Euros equivalent disposable income. The equivalent disposable income is computed using purchasing power parities (PPPs) adjusted household annual disposable income divided by the square root of family size. Most of the per-country samples have a comparable composition of individuals with respect to age, gender, household size, education and general health. However, unsurprisingly, there is considerable heterogeneity among countries with respect to the percentage of individuals who are at risk of income poverty (AROP).

Table. 2.1 *Description of selected set of features in* 2018

Country	N	Gender	Age	HH size	AROP	Education	Gen. health
Austria	10546	0.53	50.9 (17.9)	2.5 (1.3)	0.15	4.45 (1.1)	2.05 (1.0)
Belgium	10566	0.51	49.4 (18.3)	2.8 (1.4)	0.15	4.30 (1.5)	2.06 (0.9)
Bulgaria	14929	0.53	54.5 (18.0)	3.0 (1.7)	0.20	4.04 (1.2)	2.41 (1.0
Croatia	18186	0.52	52.9 (18.5)	3.1 (1.6)	0.21	3.94 (1.1)	2.47 (1.2)
Cyprus	9317	0.53	49.9 (18.9)	3.1 (1.4)	0.16	4.03 (1.6)	1.93 (1.0)
Czechia	16054	0.53	53.0 (18.3)	2.6 (1.2)	0.08	4.23 (0.9)	2.41 (0.9)
Denmark	5466	0.53	55.9 (17.5)	2.1 (1.1)	0.09	4.43 (1.3)	2.20 (0.9)
Estonia	12036	0.54	50.4 (19.0)	3.0 (1.5)	0.19	4.44 (1.2)	2.60 (0.9)
Finland	9390	0.48	51.0 (17.7)	2.4 (1.3)	0.14	4.62 (1.2)	2.15 (0.8)
France	19127	0.53	51.3 (18.3)	2.7 (1.3)	0.13	4.13 (1.4)	2.23 (0.9)
Germany	21202	0.52	53.5 (16.7)	2.3 (1.1)	0.14	4.67 (1.1)	2.30 (0.9)
Greece	48903	0.52	54.4 (18.5)	2.8 (1.3)	0.17	3.65 (1.6)	2.01 (1.1)
Hungary	14357	0.56	53.9 (18.2)	2.8 (1.5)	0.12	4.08 (1.1)	2.55 (1.0)
Italy	39969	0.53	53.6 (18.5)	2.7 (1.3)	0.19	3.64 (1.3)	2.22 (0.8
Latvia	10383	0.58	53.6 (18.5)	2.7 (1.5)	0.23	4.41 (1.1)	2.77 (0.9)
Lithuania	9521	0.56	53.5 (18.1)	2.7 (1.4)	0.21	4.60 (1.2)	2.78 (0.9)
Luxembourg	8169	0.51	45.9 (17.4)	3.2 (1.4)	0.17	4.15 (1.4)	2.16 (0.9)
Malta	8487	0.51	49.1 (18.5)	3.0 (1.3)	0.15	3.55 (1.3)	2.16 (0.8)
Netherlands	11468	0.54	54.5 (17.2)	2.2 (1.3)	0.11	4.41 (1.3)	2.12 (0.8)
Norway	5699	0.48	49.5 (17.8)	2.4 (1.3)	0.13	4.68 (1.2)	2.00 (0.9)
Poland	28205	0.56	51.9 (18.1)	3.1 (1.6)	0.15	4.42 (1.8)	2.49 (1.0)
Portugal	29292	0.53	51.9 (18.3)	2.9 (1.3)	0.18	3.12 (1.5)	2.66 (0.9)
Romania	15537	0.52	52.5 (18.1)	3.0 (1.5)	0.22	3.87 (1.1)	2.21 (0.9)
Serbia	13777	0.51	50.4 (18.4)	4.0 (2.1)	0.24	3.89 (1.2)	2.46 (1.1)
Slovenia	21924	0.51	48.7 (18.3)	3.5 (1.5)	0.11	4.26 (1.1)	2.34 (1.0)
Spain	28153	0.52	51.5 (18.1)	3.0 (1.3)	0.21	3.75 (1.6)	2.17 (0.9)
Sweden	5429	0.50	51.9 (18.4)	2.5 (1.3)	0.15	4.54 (1.3)	1.99 (0.9)
Switzerland	10204	0.53	51.3 (17.5)	2.6 (1.3)	0.15	4.70 (1.1)	1.87 (0.8)
Pooled	457475	0.53	52.1 (18.3)	2.8 (1.4)	0.17	4.04 (1.4)	2.28 (0.96)

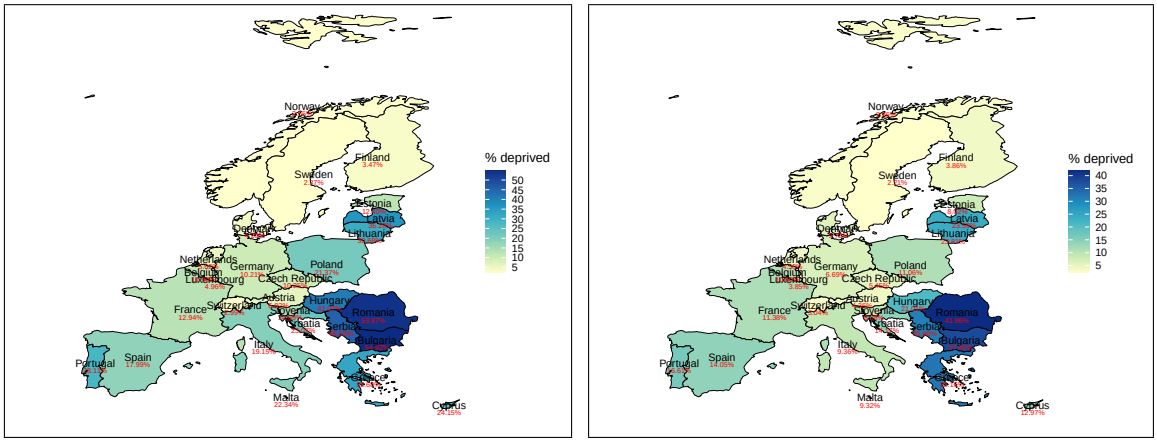
Notes: This table presents summary statistics of selected features in the 2018 sample. The AROP is defined as those below 60% of the national median household equivalent disposable income. The figures in parentheses are standard deviations.

2.2.4 Distribution of Material and Social Deprivations

Figure 2.1 plots the proportion of deprived individuals in the EU countries in 2014 and 2018; and some notable patterns are evident in the plot. The cross-country range in terms of percentage of deprived individual is non-negligible: from 2.37% in Sweden, $\approx$ 2.6% in Norway and 3% Switzerland upto $\approx$ 55% in Bulgaria and 54% in Romania for the year 2014. The percentage of deprivation in the pooled sample for the same year was 21.2%. By and large, the pattern repeats for the year 2018 but the percentage diminishes in the highly deprived countries, for instance to 42% and 38% in Bulgaria

and Romania, respectively. The percentage of deprivation in the pooled sample for the same year declined to 15.6%. This clearly indicate that there is wide divergence in average living standards across the countries. This difference is much wider than that in at-risk of poverty rates (as can be seen in the sixth column of Table 2.1).

Figure. 2.1 *Percentage of Material and Social Deprivations in the EU countries in 2014 and 2018*



(a) *2014* **(b)** *2018*

Source: author's computation using EU-SILC cross-sectional data.
Notes: The figures show the percentage of materially and socially deprived individuals in the 28 European countries and refer to 2014 and 2018 values. Slovakia, Ireland, and Iceland are excluded from the analysis due to missing information for 2018.

The distribution of the intensity of material and social deprivations shows a right-skewed distribution. For instance, in the pooled sample: The proportion of individuals who are non-deprived in all 13 dimension of MSD items represent the largest percentage, 39.2% and 45% in 2014 and 2018, respectively. In contrast, the percentage of individuals who are deprived in all dimensions represent the smallest group 0.23% and 0.12% (see Appendix 2.A). This pattern repeats in the per-country samples. However, there is extreme heterogeneity between the countries in terms of the percentage of population who possess all 13 items: 81% in Sweden and 78% in Switzerland, while 23% in Bulgaria 17.6% in Romania (in 2018).

If we now look more specifically at deprivations per each single item, we observe the largest percentage of deprivation: to "afford one-week annual holiday away from home" (Romania $\approx$ 60% followed by Croatia $\approx$ 56%) and to "face unexpected financial expense" (Latvia $\approx$ 60% followed by Croatia $\approx$ 54%). In contrast, deprivations in the internet connectivity and get together with friends/family are experienced with small fraction of the population, e.g., (Romania $\approx$ 21% followed by Serbia $\approx$ 11.3%) and (Hungary $\approx$ 22.6% followed by Romania $\approx$ 22.5%), respectively (figures are reported in

Appendix 2.A).

2.3 Method

Many features contribute to the individuals' risk of material and social deprivations. The primary goal of this study is to predict micro-level material and social deprivation status with better accuracy, to identify the key determinants and examine their patterns of association with the likelihood of deprivation. The second objective is to analyse the predictive performances of targeting subsets of features, such as sociodemographic, socioeconomic, health, and location, in identifying the deprived.

When it comes to classifying models, there is a myriad of supervised machine learning algorithms. I favour tree-based algorithms, such as Extreme gradient boosted trees (Xgboost) and Random forests (RF). Because, these algorithms are well-established in the ML literature and are shown to be highly effective in tabular datasets. I prefer Xgboost over the other boosting trees, such as Gradient boosted trees (GBT), because the Xgboost algorithm has several desirable properties. These include: 1) The algorithm's core is parallelisable and does parallelisation within a single tree resulting in high computation speed. 2) Xgboost is highly scalable. Unlike bagging techniques, where trees are grown to their maximum level, boosting grows trees with fewer splits. 3) Xgboost is a more regularised (L1 and L2) form of GBT, which can result in better generalisability of the model. 4) Extreme gradient boosted trees algorithm supports sparsity-aware split finding by default, which helps to handle sparse data (See, Chen and Guestrin, 2016, for a detailed analysis).[2]

Xgboost is an extension of Gradient boosted trees (GBT), an ensemble of trees like the RF, but trees are grown neither randomly nor independently (Chen and Guestrin, 2016). Instead, the weak learners (weak classifiers) are fit first, and then each tree is fit sequentially on a full dataset by correcting the prediction error of their predecessor. In order to correct the past mistakes, observations are weighted by the error rates of the previous trees—i.e., more weights on the wrongly predicted samples Friedman (2001, 2002). Although GBT performs better in various classification and regression tasks, they are susceptible to overfitting and have many tuning parameters. Hence, Xgboost offers subtle modelling details to GBT: it introduces more regularisation to control overfitting; performs parallelisation within a single tree that result in high computation speed

[2]Sparsity in the input feature can happen due to either missing values or frequent zero entries in the data, which is common in socio-economic data.

compared to the standard GBT. Moreover, model-agnostic interpretation tools, e.g., Shapley values, are fast to compute because trees are grown with few splits, compared to RF, which grows trees to maximum depth.

Random forests are an ensemble of many de-correlated trees. These trees are grown from random subsamples of the training set and random subsets of features (Breiman, 2001). Each tree is a sequence of rules that splits the sample into subsets called leaves. The prediction for each leaf is the average outcome of observations on that leaf, and trees are fit to minimise the mean squared error. In a regression task, the final prediction is then rendered as the mean prediction across all trees, and in a classification task forest's best prediction is chosen on a majority vote. As a result, RF can overcome the overfitting problem that is often afflicting the decision tree regressor by averaging many noisy but unbiased trees.[3] This ensemble of trees increases model stability by inducing smoother estimation of the function underlying the data-generation process, which results in a robust and accurate prediction. Random forests are able to model complex interactions between features. Appendix 2.B reports in more details about hyper-parameter choices and the mechanics of these algorithms.

To assess the performance of the models, I use a metrics that is based on the Receiver Operating Curve (ROC), which shows the trade-off between true and false positives for a given model (Hanley and McNeil, 1982). Mainly, I focus on the area under the ROC, often referred to as the area under the curve (AUC). The AUC measures the likelihood that a randomly selected pair of observation is correctly ordered in terms of predicted outcomes. A model that is no better than chance would have an AUC of 0.5, an AUC of 0.7 - 0.8 indicates an acceptable classification power, and a perfect model would have an AUC of 1. As the output of different classifying models that use binary response variables produces continuous probability showing the degree to which each individual is a member of the two labels, a probability threshold is required to predict membership to each class. The advantage of the AUC is that it does not require a probability threshold above which an individual will be declared to be materially and socially deprived. Choosing a specific threshold involves a trade-off between true positive rate (sensitivity) and true negative rate (specificity). The choice of this threshold depends on one's relative patience for false positives vs. false negatives, which could often be specific to domain of application. For instance, a lenient policymaker would favour a smaller probability threshold that leads to higher sensitivity at the cost of specificity. On the other hand, a strict policymaker (with scarce resources) might favour

[3]For a rigorous mathematical characterisation of overfitting reduction while retaining its predictive accuracy, see (Hastie et al., 2017; Breiman, 2001).

a higher probability threshold, boosting specificity at the cost of sensitivity. This study is after overall performance at any given probability threshold, which regards the concern of both types of policymakers; hence, this study rely on AUC. Alternative evaluation metrics, such as Accuracy, Precision, and F1-score are reported in Appendix 2.C.3.

Finally, I use the visualisation tool SHapley Additive exPlanations (SHAP) proposed by Lundberg and Lee (2017a) to explain the contribution of each feature to the prediction of material and social deprivation using Shapley values. SHAP is based on a solution concept in a cooperative game setup that aims to 'fairly' allocate the gains among players as suggested in the seminal work of Shapley (1953). SHAP has the advantage of consistency and provides both local and global interpretability (see Guidotti et al., 2018; Molnar, 2020, for a comprehensive review of black-box ML model interpretation techniques).

2.4 Results and Discussion

2.4.1 Is MSD Status Predictable?

I train the models using a random sample constituting 80% of individuals (in the pooled as well as the per-country datasets). The model's generalisability to the unseen individuals is tested using the remaining 20% as a test set. Figure 2.2 plots the in-sample and out-of-sample performance of the models using 2014 and 2018 pooled datasets. Some notable patterns are evident in the plot. First, the two tree-based algorithms yield a more substantial classification power. In general, a model with an AUC of 0.8 and above is regarded as very good, and one with an AUC of 0.9 and above is considered an excellent classifier. The Xgboost, RF, and logit achieved an AUC of 0.88, 0.86, and 0.82 on the test set. Put differently, the relative classification accuracy gained by using the more sophisticated algorithm is 7.3% with Xgoost and 5.9% with the random forest. Exhibiting that the extreme gradient boosted trees algorithm excels at accurately classifying material and social deprivation status on unseen individuals.

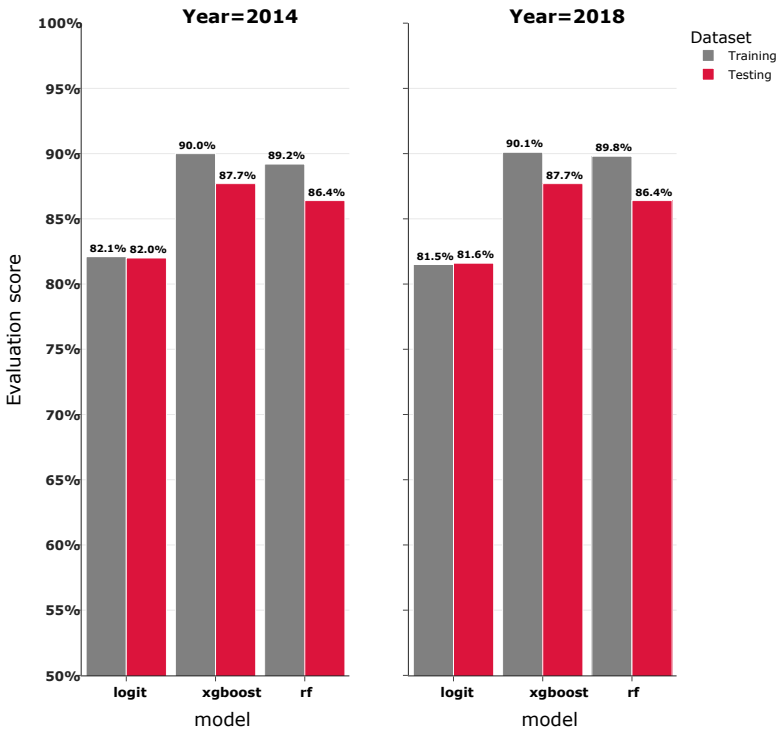
The obtained classification accuracy gains with the flexible non-parametric tree-based algorithms can be explained by the potential interaction effects between features, which logistic regression can not capture inherently.

Second, approximately, equal classification accuracy is obtained on in-sample and out-of-sample individuals with the logit. This classifier does not tend to over-fit as opposed to the ML algorithms. Specifically, the RF with a default hyperparameter configuration notoriously overfits the data. To this end, I applied ten-fold cross-validation

for model selection. That is, the results are obtained after carefully tuning the Random forests using the depth of the trees and the number of features sampled when growing the trees. Similarly, the reported figures of the Xgboost algorithm are attained after tuning the model hyperparameters, such as "the maximum depth of a tree", "the minimum loss reduction required to make a further split" (gamma), "the learning rate", "the number of trees" (number of estimators).

Finally, the results above are significantly stable during the two cross-sectional years. Moreover, the superior performance of the tree-based ML algorithm has also been replicated in the per-country predictions exercises (results are reported in Appendix 2.E.1).

Figure. 2.2 *In-sample and Out-of-sample performance of prediction models, area under the ROC curve*



Notes: Each model is trained on 80% of pooled and per-country datasets and tested on the remaining 20% of the datasets. The AUC compute the trade-off between the True Positive rate and the False Positive rate at different probability thresholds.

However, it is essential to keep in mind some caveats when interpreting the findings, which are based on a single "*hold-out*" technique, i.e., the simple random data splitting into two mutually exclusive subsets referred to as training and test sets. The concern with this approach is that the obtained predictive accuracy can be biased when the hold-out group is either over-represented or under-represented by the deprived individuals.

In the applied machine learning literature, the common panacea to overcome this issue is accuracy estimation based on cross-validation (Kohavi et al., 1995). I use k-fold stratified cross-validation so that the same proportion of class labels are reflected in each fold as the entire dataset. I repeated the sampling procedure 100 times in order to reduce the variance of the estimates.

Table 2.2 reports the mean test scores (in terms of the AUC) with 95% confidence intervals in parentheses. In the last row of the table, it is noted that the mean test AUC is 0.88 for Xgboost, $\approx$ 0.87 for RF, and 0.81 for the logit. The figures are comparable to the ones obtained with the single hold-out approach, which can be due to a sufficiently large number of individuals represented in both the training and test sets.

Table. 2.2 *AUC of pooled model with sequentially inserted sets of features, year* 2018

Subset of predictors	Xgboost (1)	Random Forests (2)	Logit (3)
Sociodemographic	0.690 [0.689, 0.692]	0.691 [0.690, 0.693]	0.674 [0.673, 0.676]
Socioeconomic	0.840 [0.839, 0.841]	0.837 [0.836, 0.838]	0.808 [0.807, 0.808]
Health	0.642 [0.641, 0.642]	0.641 [0.640, 0.642]	0.638 [0.637, 0.639]
Country dummies	0.715 [0.714, 0.718]	0.715 [0.714, 0.717]	0.715 [0.714, 0.718]
All predictors	0.880 [0.878, 0.881]	0.867 [0.876, 0.879]	0.812 [0.810, 0.813]

Notes: The AUC compute the trade-off between the True Positive rate and the False Positive rate at different probability thresholds. The figures in square parentheses are 95% confidence interval of AUC computed via ten-fold stratified sampling repeated 100 times.

Rows (1-4) compare models composed entirely of one or the other feature subset to disentangle the predictive power of subsets of features. I estimate the prediction metrics by training the models by taking one set of predictors at a time. The performance of models described in the previous paragraphs can be used as a benchmark to assess the predictive power of subsets of indicators. First, let us examine the predictive power of demographic characteristics alone. This subset includes variables such as age, gender, marital status (with four categories), level of education, household size, and household typology indicators (i.e., couple with child(ren), couple with no child, one-person household, and single parent household). The results show that demographic factors alone fall short of producing an acceptable predictive power, although by just a little.

Row 2 indicates that individuals' socioeconomic factors are the best predictor. This set of features includes equivalent income deciles, home ownership, employment status, occupation types (ISCO code 08, ten categories classification), and contract type. The predictive accuracy is almost as good as when all variable is used, especially in the logit model. By contrast, the area under the ROC curve is the lowest for the health indicators

(general health status, chronic illness, and limited activity), suggesting that predicting the deprivation status of individuals using only health features is narrowly better than random guesses (an AUC of 0.5 - 0.7 is regarded as poor classification power). Last, I assess the predictive performance of models trained using country dummies to separate the role of unobserved country-specific factors in the pooled dataset. The result shows that unobserved country-specific factors alone predict material and social deprivation status with a good classification power.

In sum, the socioeconomic factors and country-specific elements contain more signals in correctly classifying people based on their risk of material and social deprivation in the pooled prediction. These observations remained unaltered when the 2014 data was used (results are in Appendix 2.C.2). Of course, the above feature grouping includes a different number of multiple predictors; hence, one variable could contribute much more than others. In the subsequent Section, I examine in a more detailed manner the relative importance of the most prominent features and their pattern of association with the likelihood of deprivation.

2.4.2 What Predicts it? Explainability with Shapley Values (SHAP)

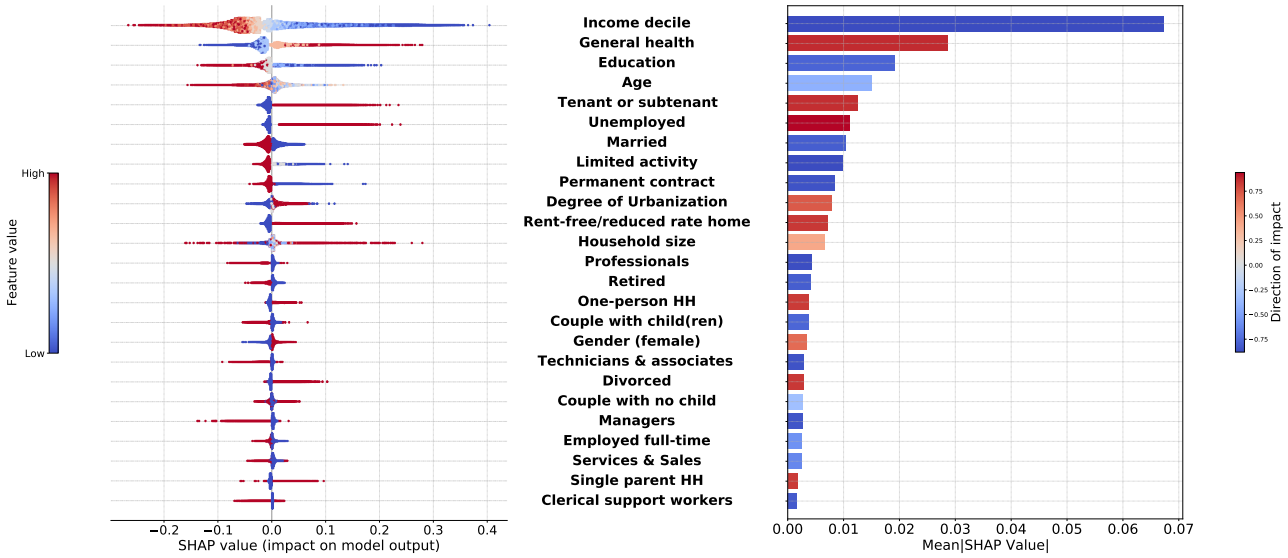
After training the highly performing Xgboost classifier (including all features along with country-specific effects in the pooled dataset and with regional fixed effects in the per-country datasets), I proceed with an explainability analysis of the model using SHAP. Figure 2.3 summarise the key variables predictive of individual-level material and social deprivation status in the 2018 dataset. See Appendix 2.D for detailed results showing all features used in the classification exercise. In order to check the stability of the results in Figure 2.4, I plot the feature importance using the 2014 dataset

In the figures below, panel a) visualise the key features that contribute significantly to classifying material and social deprivation status. On the y-axis, the top 25 predictors are placed in descending order of importance by global contribution to the classification (measured in mean absolute SHAP value). Numbers in the x-axis are the SHAP values of each observation. Each dot represents an individual respondent; hence, the number of dots against each feature reflects the sample size of the training set. The dot's position along the x-axis is the feature's impact on the probability of deprivation for that respondent. When multiple dots arrive at the same coordinate in the plot, they pile up to show the density of effect sizes. The colours correspond to the feature values: red for larger values and blue for smaller ones. A negative SHAP value (extending to the left) shows a reduced likelihood of deprivation, while a positive (extending to the right)

shows an increased probability of deprivation. Asymmetric distribution with respect to the zero SHAP line in the summary plot suggests that the feature's impact is non-linear. The long-right tails in the summary plot indicate that the variables are highly predictive for some respondents but not others, i.e., predictors with minor global importance can still be very important for specific respondents. In panel b) I summarise the three-fold information with a bar plot to aid quick reading. The direction of association between a feature and the likelihood of deprivation is colour-coded (red for positive and blue for negative). The darkness of each colour gradient shows the strength of the direction of the association. The horizontal length of the bars shows the magnitude of the feature's global impact measured as the average of absolute SHAP values.

The results show that features related to the relative position in the income distribution and home ownership status are negatively associated with the likelihood of being materially and socially deprived. These variables are measured at the household level. Under the assumption of resource sharing within the same household, every individual in the same family is assigned identical equivalent incomes and assumed to equally benefit from the housing wealth. Similarly, individual-specific features such as good health conditions, higher level of education, being married (including couples with children), and not having a condition that limits activity (for at least the past six months) are associated with a lower likelihood of deprivations. Variables related to individuals' ties to the labour market, such as being employed with a permanent contract, being retired, and engaging in occupations like professionals, technical, managers, service, and clerical jobs, show a negative correlation with the probability of being deprived. While living in a densely populated locality, living in a rent-free/reduced-rent home, being from a household characterised as a single-person household or a single-parent family, and being divorced are positively associated with the likelihood of being materially and socially deprived. As can be seen in the figures below the relative importance of these features and their corresponding sings are stable and consistent across the survey years.

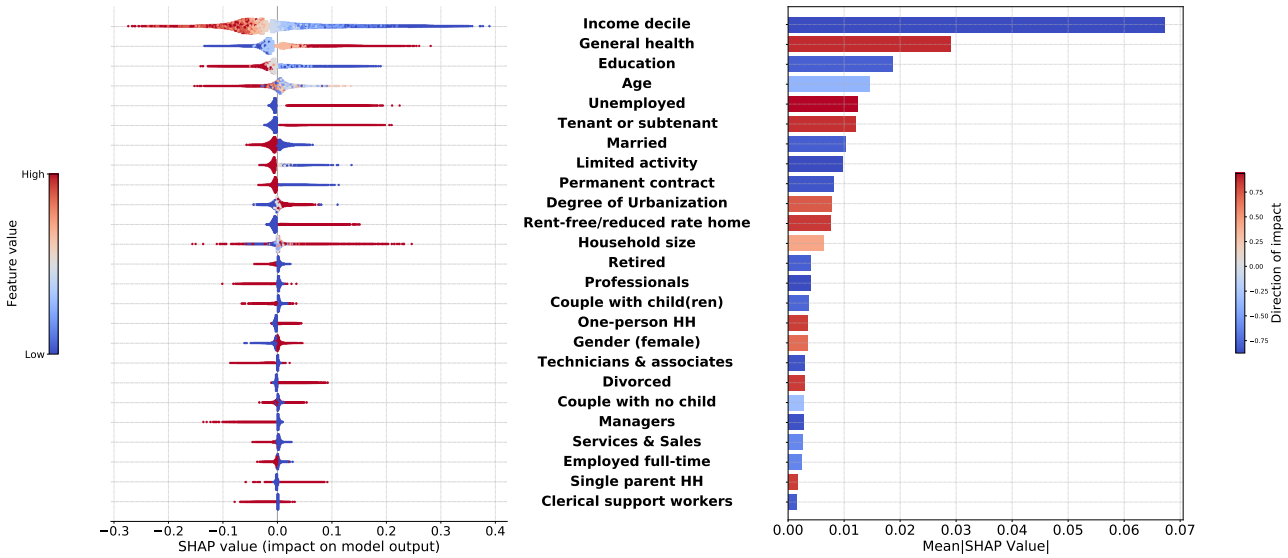
Figure. 2.3 *Key predictors of material and social deprivations, year 2018*



(a) *Summary plot* **(b)** *Global impacts (magnitude and direction) of top features*

Notes: This figure shows the SHAP feature importance from the Xgboost algorithm trained on the pooled data set of 28 European countries to classify individual-level material and social deprivations. In this classification task country-level fixed effects are accounted for. General health status takes value one if respondents assess their health status as "very good"..., and value five if the health status is "very bad."

Figure. 2.4 *Key predictors of material and social deprivations, year 2014*



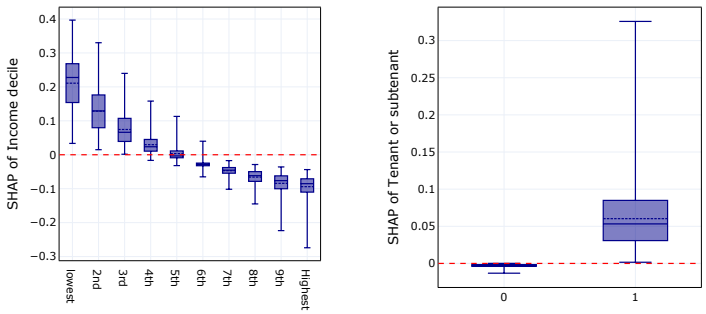
(a) *Summary plot* **(b)** *Global impacts (magnitude and direction) of top features*

Notes: This figure shows the SHAP feature importance from the Xgboost algorithm trained on the pooled data set of 28 European countries to classify individual-level material and social deprivations. In this classification task country-level fixed effects are accounted for. General health status takes value one if respondents assess their health status as "very good"..., and value five if the health status is "very bad."

2.4.3 Partial Effects of Selected Variables: SHAP from the Xgboost Model

I now turn to a detailed analysis of crucial predictors' SHAP values to disentangle the different factors' separate effects and their patterns of association with the probability of being deprived materially and socially. SHAP by Lundberg and Lee (2017b) provides partial dependence plot to shows the marginal effect of one or two features on the predicted outcome of a machine learning algorithm. Since most of the features in socioeconomic studies are measured as a categorical variable, I create box plots to show the difference in the average marginal contribution of each category in a more robust way. It displays whether the relationship between the target outcome and a feature is linear, monotonic, or more complex. The y-axis of the box plots indicates the SHAP value of the variable in terms of probability (SHAP = 0 is the baseline —i.e., the average predicted probability), and on the x-axis are the variable's values. The dotted horizontal line in the box-plots shows the distribution's mean value in each group. Since individuals' living standards are likely to be determined by household characteristics, individual-level demographics, and socioeconomic variables, I structure the results according to factors related to income and housing wealth, ties to the labour market, household characteristics, and sociodemographic features.

Figure. 2.5 *Partial effects of income and housing wealth*



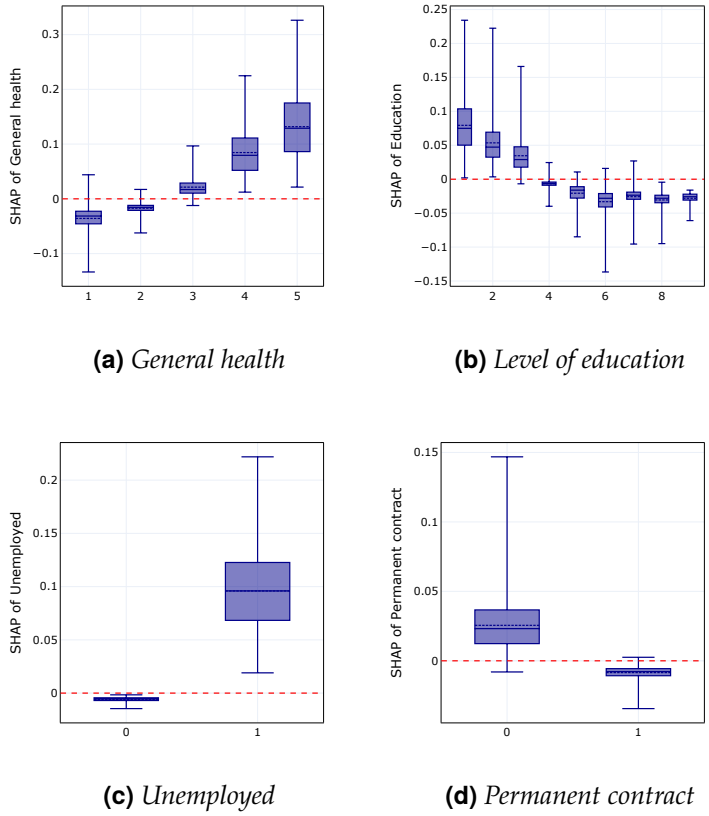
(a) *Income distribution (decile)* **(b)** *Home ownership*

Notes: This figure shows the partial effect of income distribution (in decile) and home ownership status (1 = tenant or subtenant, 0 = otherwise) in terms of Shapley values. The dotted line in the box-plot shows the mean value of the distribution in each group.

Figure 2.5 plots the partial effects of the equivalent disposable income and housing wealth on the probability of material and social deprivation. Under the assumption of resource sharing within the same household, individuals in the same families are assigned identical equivalent incomes. In a similar vein, the impact of home-ownership

status is assumed to be the same for individuals in a given household. The results show that these factors are highly predictive of multidimensional deprivation and ranked at first and fourth position consistently during the two periods studied. As seen in panel (a), the impact of income is steeper on the left-hand side of the income distribution than on the right-hand side. That is, individuals in the lowest equivalent income decile have approximately a 21% chance of being materially and socially deprived compared to those with the median income. However, the wealthiest individuals (in the highest income decile) are 10% less likely to be deprived. Similarly, in panel (b), it is evident that those without housing wealth are 5.5% more likely to be deprived than the tenant occupants. The nexus of income poverty and living standards is then consistent with those discussed in previous work, i.e., although highly correlated, income poverty and individual's living standard do not perfectly overlap (Fusco et al., 2011; Nolan et al., 1996; Gordon et al., 2000; Perry, 2002). 1) available resources and disposable income are not identical. Thus, the current consumption can be affected by accumulated savings, debt repayment, past investments, in-kind transfers, and access to free/subsidised goods and services. 2) Household heterogeneity, personal costs, such as expenses related to health, housing, mobility, and education, are different from household to household as these are dependent on individual member circumstances. What is shown in panel (b) also reflects that home-ownership status can reinforce the individual's command over the available resources. Thus, re-distributive housing policies and housing allowances can attenuate the risk of living conditions-deprivation (Dewilde, 2022).

Figure. 2.6 *Partial effects of individuals' features related to health, education and employment*



(a) *General health* **(b)** *Level of education*

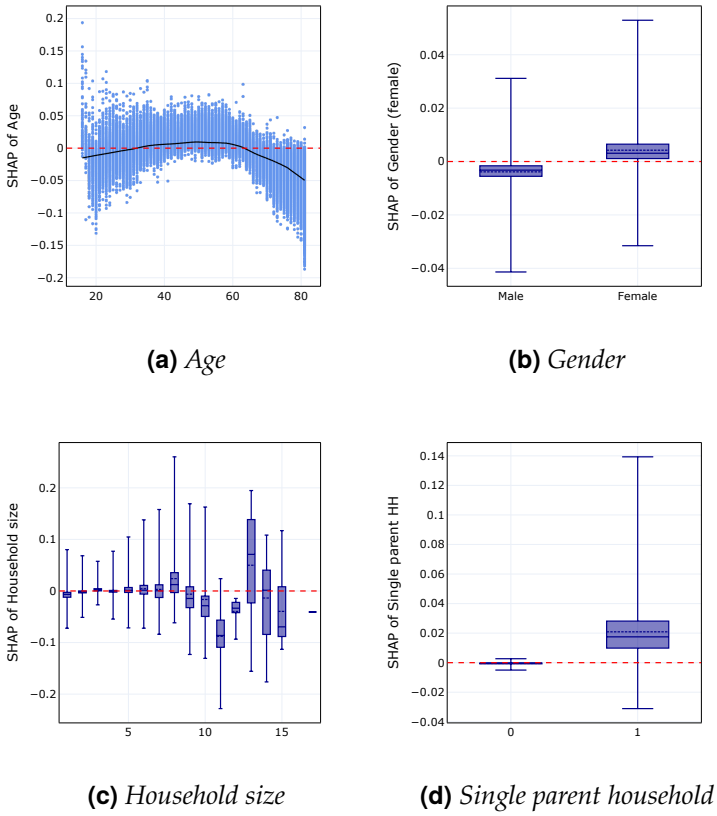
(c) *Unemployed* **(d)** *Permanent contract*

Notes: This figure shows the partial effect of individuals health status (1 = very good ,..., and 5 = very bad), employment status (unemployed =1), contract type (1 = permanent contract), and education status (1 = Less than primary; 2 = primary; 3 = lower secondary education; 4 = upper secondary education; 5 = Post-secondary non-tertiary; 6 = Short cycle tertiary; 7 = Bachelor or equivalent; 8= Master or equivalent; and 9 = Doctorate or equivalent). The dotted line in the box-plot shows the mean value of the distribution in each group.

Figure 2.6 displays the partial effects of disadvantages, such as health problems, low level of education, unemployment and contract type. Notable pattern is evident in the figure; panel (a) shows that individuals with a "very bad" health condition are 17.3% more likely to be deprived compared to healthy individuals with a "very good" health condition. This strong positive association between poor health conditions and material and social deprivation can reflect what (Sen, 1999) coined as the "coupling" between disadvantages. Poor health conditions could result in reduced earnings. Poor health might also inhibit the ability to convert the reduced income into adequate material and social necessities due to extra costs related to medical treatment for inadequate health conditions. Panel (b) shows that individuals with a higher achieved level of education are less likely to be materially and socially deprived. For instance, an individual with less than primary education is 7.5% more likely to be materially and socially constrained

than an individual with upper secondary education. This can reflect the well-known evidence that human capital positively affects one's quality of living through returns to education in the labour market. However, the result of this study reveals a non-linear association between schooling and deprivation. That is, there is no statistically significant difference in the likelihood of deprivation among individuals with high levels of education (i.e., between short-cycle tertiary education and above) compared to lower and middle education levels. As seen in panel (c), unemployed individuals are approximately 10% more likely to be materially and socially deprived than employed. The last panel in the figure above plots the effect of employment contracts: individuals with a permanent contract are 2.6% less likely to be deprived than those with temporary contracts. This evidence is consistent with those found in past research (Fusco et al., 2011; Halleröd and Larsson, 2008; Saltkjel and Malmberg-Heimonen, 2017; Testi and Ivaldi, 2009), where material deprivation was shown to be strongly associated with individual disadvantages.

Figure. 2.7 *Partial effects of individuals' age, gender and their family structure*



(a) *Age* **(b)** *Gender*

(c) *Household size* **(d)** *Single parent household*

Notes: The dark hump-shaped line in panel (a) is a Locally Weighted Scatter plot Smoothing (LOESS). Single parent household (1 = if the person's family is described as single parent household). The dotted line in the box-plot shows the mean value of the distribution in each group. In panel (c), note that a large majority of the sample has a family size of seven or less.

Figure 2.7 (a-d) examine the role of key sociodemographic factors. The impact of age is non-monotonic; instead, it shows a hump-shaped association with the risk of deprivation. Middle-aged (40-60 years of age) people were more prone to suffer from a higher risk of material and social deprivation. Younger and older people are associated with a lower likelihood of deprivation, with a relatively steeper slope for the older age group. Regarding gender, women are more likely than men to suffer the risk of deprivation. However, the magnitude of the difference is relatively slim. Concerning the difference with respect to family structure, one average, one-person household and single parents households were more likely to suffer from a higher risk of deprivation. However, I observe no meaningful pattern between family size and risk of deprivation.

2.4.4 Heterogeneity Analysis by Country

Detailed results of feature importance and direction of association in each of the 28 countries separately are presented in Appendix 2.E. The family's relative position in the income distribution and home ownership status are critical factors in accurately classifying individuals based on their risk of material and social deprivation. Consistent with the results from the pooled model, income is at the top in terms of relative importance in the per-country predictions, while the latter appears in the top five in a great majority of the nations. Individual-specific features, e.g., good general health conditions (also in terms of a long-term illness that limits activity), higher level of education, and family status (being married), are associated with a lower likelihood of deprivations. And these features consistently appear among the top 25 in all countries. Consistent with the results from the pooled sample, variables related to individual employment status, contract type, occupation class, gender, and family characteristics also appear among the top 25 predictors in all countries.

2.5 Conclusion

Previous studies examine material deprivation and the role of specific policy variables. Still, the question of how accurately one can classify unseen/out-of-sample individuals' deprivation status given their observable personal, household, and country-specific factors remain unexplored. Using the EU-SILC microdata, this study shows a reliable and computation-efficient approach to predicting material and social deprivation (MSD) status with better accuracy. Key results of these analyses demonstrated that the non-parametric extreme gradient boosted trees (Xgboost) algorithm followed by the Random forests provides a better classification accuracy than the Generalised linear model on the pooled analysis, and the same holds in the per-country prediction exercises. The socioeconomic features alone yield a classification accuracy as close as when the whole set of features is used in the pooled analysis, followed by the unobserved country-specific factors.

Although these complex tree-based algorithms are highly predictive, they lack model-specific interpretability. Hence, augmenting the highly predictive Xgboost algorithm with Shapley's values helps to effectively explain the complex relationship between MSD and sets of predictors. In addition, overfitting poses a challenge in the current study. However, a transparent model selection process via k-fold cross-validation reduces the potential impact of overfitting. Hence, the relative feature importance and

partial effects identified with Shapley's values reveal insightful relationships between key policy variables and the likelihood of poor living conditions. The individual's relative economic position to others, general health status, level of education, age, and housing wealth is the most prominent predictor of material and social deprivation class. This study reveals a slightly nonlinear association between income and deprivation status, i.e., a substantial effect of income reflected on the left side of the income distribution. At the same time, it identifies a hump-shaped association between age and deprivation status. These results are generalisable to different cross-sectional years.

Future analyses can complement this study and multidimensional poverty analysis, in general, by exploiting more advanced machine learning algorithms. For instance, this study follows the EU's definition of deprivation: an enforced lack of five items (out of the thirteen) deemed necessary and desirable to lead an adequate life. In this setup, an individual's risk of deprivation is shown to be effectively modelled using a binary classifier ML algorithm. In contrast, in a more strict sense, deprivations can also be seen as the failure to access one or more dimensions of material and social items (known as the "union" identification criterion). In this setting, the classification task can be approached with *multilabel* classification technique – not to be confused with multi-level classification. Multilabel classification algorithms can model multiple binary vectors of deprivation items on input features. The current development in ML, *algorithm adaptation method*, for instance, Random Forests - Predictive Clustering Trees (RF-PCT) are shown to effectively model multidimensional outcomes directly without reducing rich dimensions of outcomes to a one-dimensional problem. Although RF-PCT is a powerful tool in this setting, it still lacks model interpretability.

References

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Dewilde, C. (2022). How housing affects the association between low income and living conditions-deprivation across europe. *Socio-Economic Review*, 20(1):373–400.

Eurostat (2022a). Eu-silc microdata. https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions.

Eurostat (2022b). Sustainable development in the european union. https://ec.europa.eu/eurostat/documents/3217494/14665254/KS-09-22-019-EN-N.pdf/2edccd6a-c90d-e2ed-ccda-7e3419c7c271?t=1654253664613.

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

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4):367–378.

Fusco, A., Guio, A.-C., and Marlier, E. (2011). Income poverty and material deprivation in european countries.

Gordon, D., Adelman, L., Ashworth, K., Bradshaw, J., Levitas, R., Middleton, S., Pantazis, C., Patsios, D., Payne, S., Townsend, P., et al. (2000). Poverty and social exclusion in britain.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):1–42.

Guio, A.-C., Marlier, E., Gordon, D., Fahmy, E., Nandy, S., and Pomati, M. (2016). Improving the measurement of material deprivation at the european union level. *Journal of European Social Policy*, 26(3):219–333.

Halleröd, B. and Larsson, D. (2008). Poverty, welfare problems and social exclusion. *International Journal of Social Welfare*, 17(1):15–25.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36.

Hastie, T., Tibshirani, R., and Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction*. springer open.

Julkunen, I. (2002). Social and material deprivation among unemployed youth in northern europe. *Social Policy & Administration*, 36(3):235–253.

Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada.

Lundberg, S. M. and Lee, S.-I. (2017a). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*, pages 4768–4777.

Lundberg, S. M. and Lee, S.-I. (2017b). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Minter, W. (1992). Africa's problems... african solutions. retrieved june 9, 2006.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Nelson, K. (2012). Counteracting material deprivation: The role of social assistance in europe. *Journal of European Social Policy*, 22(2):148–163.

Nolan, B., Whelan, C. T., et al. (1996). Resources, deprivation, and poverty. *OUP Catalogue*.

Perry, B. (2002). The mismatch between income measures and direct outcome measures of poverty. *Social Policy Journal of New Zealand*, pages 101–127.

Rowntree, B. (1901). Poverty-a study of town life (ed., 1922), macmillan, london.

Saltkjel, T. and Malmberg-Heimonen, I. (2017). Welfare generosity in europe: A multilevel study of material deprivation and income poverty among disadvantaged groups. *Social Policy & Administration*, 51(7):1287–1310.

Saunders, P. (2004). Towards a credible poverty framework: from income poverty to deprivation.

Sen, A. (1982). *Poverty and famines: an essay on entitlement and deprivation*. Oxford university press.

Sen, A. (1985). Well-being, agency and freedom: The dewey lectures 1984. *Journal of Philosophy*, 82(4):169–221.

Sen, A. (1992). *The political economy of targeting*. World Bank Washington, DC.

Sen, A. (1999). *Development as Freedom*. Oxford: Oxford University Press.

Sen, A. (2009). The idea of justice. london: Allen lane. *Link: https://goo. gl/tBfKgq*.

Shapley, L. S. (1953). A value for n-person games. In Kuhn, H. W. and Tucker, A. W., editors, *Contributions to the Theory of Games (AM-28), Volume II*, pages 307–318. Princeton University Press.

Testi, A. and Ivaldi, E. (2009). Material versus social deprivation and health: a case study of an urban area. *European Journal of Health Economics*, 10(3):323–328.

Townsend, P. (1970). Measures and explanations of poverty in high income and low income countries: The problems of operationalizing the concepts of development, class and poverty. *The concept of poverty*, pages 1–45.

Townsend, P. (1979). *Poverty in the United Kingdom: a survey of household resources and standards of living*. Univ of California Press.

Wodon, Q. T. (1997). Targeting the poor using roc curves. *World Development*, 25(12):2083–2092.

Appendices

2.A Data and Some Descriptive Analysis

Feature Set

Sociodemographic features: I consider age, gender (female = 1), marital status, household size, household typology, and the highest level of education attained by the individual. Marital status is recoded as four binary variables: Married, separated, widowed, and divorced (reference category: never married). The household typology indicator is recoded into four: dichotomous features: one-person household, couple with no dependent child, couple with dependent child(ren), and single-parent household (reference category: other households. The highest level of educational attainment was measured in nine ordinal categories (Less than primary, primary, lower secondary education, upper secondary education, post-secondary non-tertiary, short cycle tertiary, bachelor or equivalent, master or equivalent, and doctorate or equivalent).

Socioeconomic features: This subset of measures includes the distribution of income (decile). The income decile is computed using purchasing power parities (PPPs) adjusted household annual disposable income equivalised by the square root of family size. The housing wealth of individuals is measured with two dummy variables: 1) tenant and 2) free/reduced rate accommodation, considering homeowners as a reference group. Industry of occupation variables is based on the ten ISCO code 08 classification categories (reference group: Elementary occupations). I also consider indicators of the individual's economic status, nine binary variables: employed (full-time or part-time), self-employed, unemployed, disabled, inactive, retired, student or in the military, and domestic task or care responsibility. In addition, I use a dummy variable measuring the type of individual's employment contract (1 = permanent, 0 = temporary).

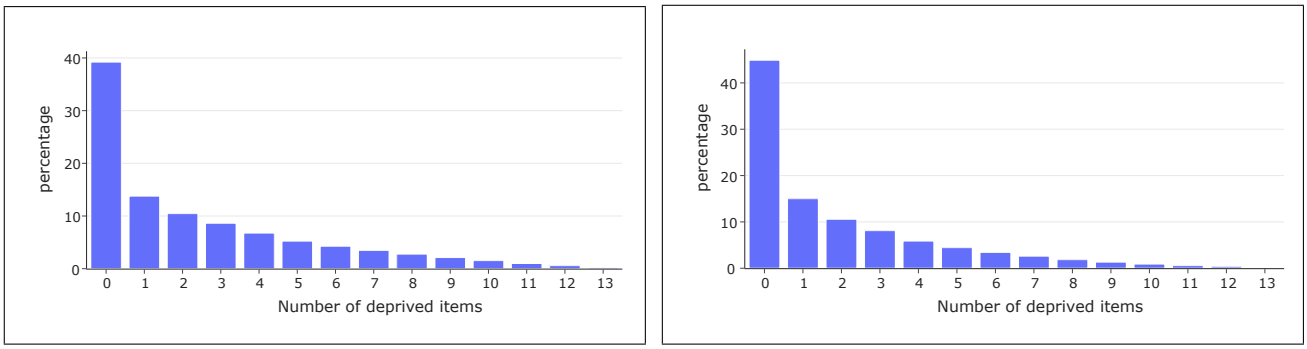
Health features: These include an indicator for chronic illness of individuals—this feature takes value one if the individual was chronically ill during the past six months or more. The "general health" condition of individuals is measured in the EU-SILC

data with a five-level Likert scale (1 = very good,..., and 5 = very bad). Last but not least, I also examine the health feature of individuals relating to "limited activity": This dummy variable captures the limitation in the activities of individuals for at least the past six months because of their health problems (1 = if there is a limitation in activity, 0 = otherwise).

Location features: These include dummy variables of country of residence (or regional dummies in the country-level analysis). Moreover, I also capture the degree of urbanisation with three-level ordinal categories; which takes value 1 if thinly-populated (outside urban cluster); 2 if intermediately-populated (at least 300 inhabitants per square km and minimum of 5000 population); 3 if densely-populated (at least 1500 inhabitants per square km and minimum of 50000 population).

Distribution of Material and Social Deprivation

Figure. 2.9 *Proportion of individuals by the reported number of items deprived 2018*



(a) *2014* **(b)** *2018*

Figure. 2.10 *Percentage of material and social deprivation rate in the 28 EU countries by gender in 2018*

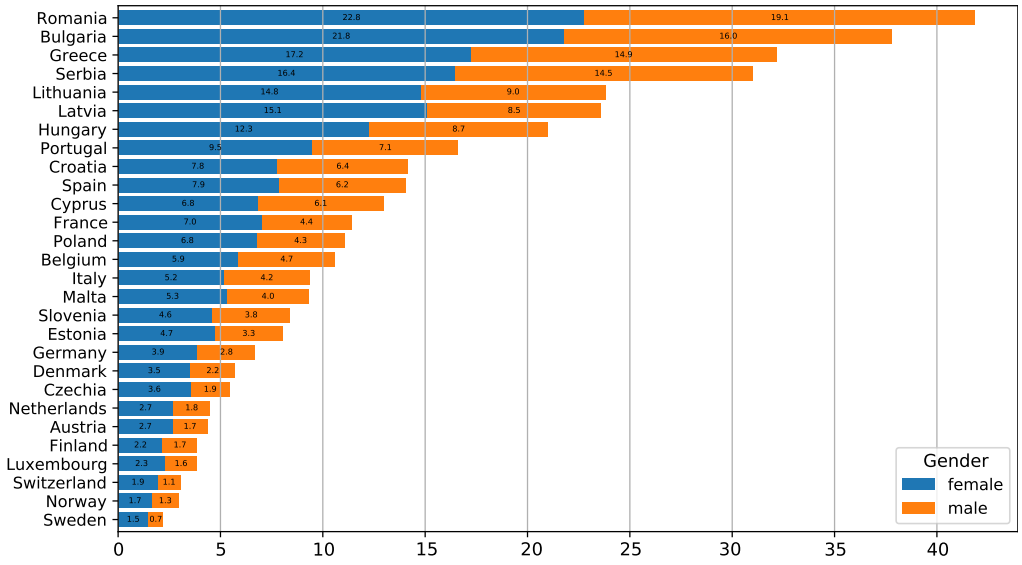
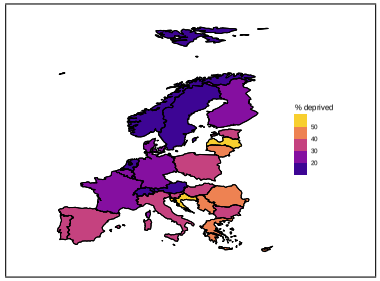
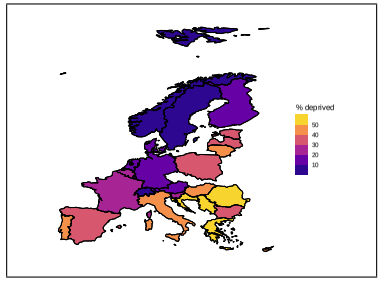


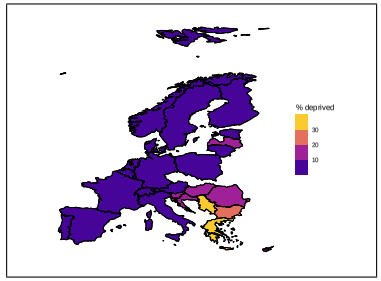
Figure. 2.8 *Patterns of deprivations in each dimension in the EU countries in 2018*



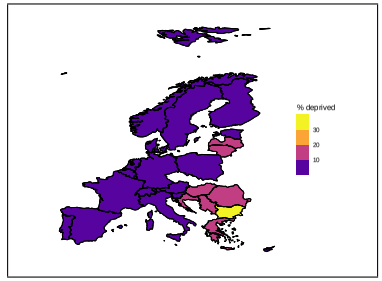
(a) *Financial expenses*



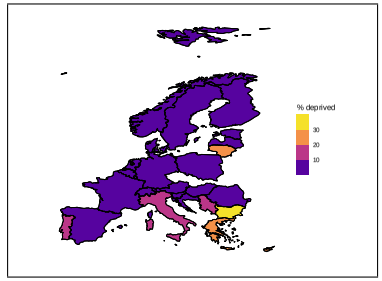
(b) *One-week holidy*



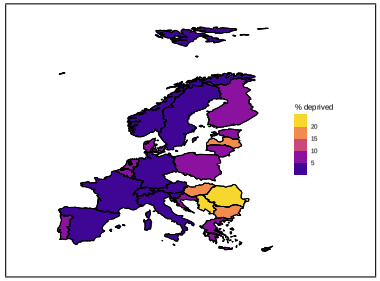
(c) *Avoid arrears*



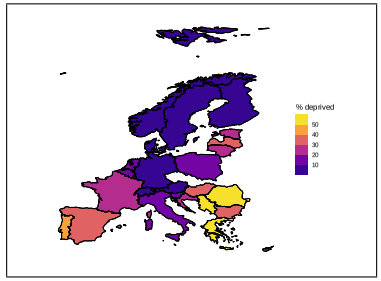
(d) *Afford a balanced meal*



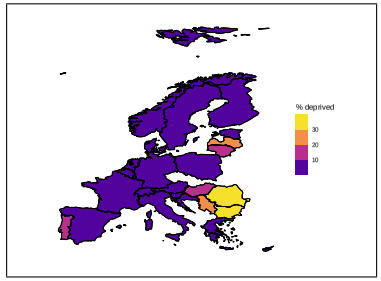
(e) *Keep home adequately warm*



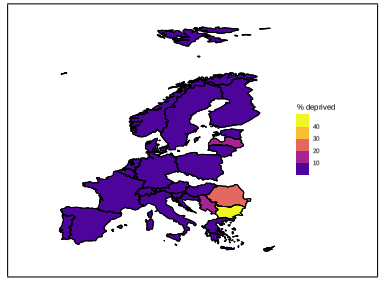
(f) *Access to a car/van for personal use*



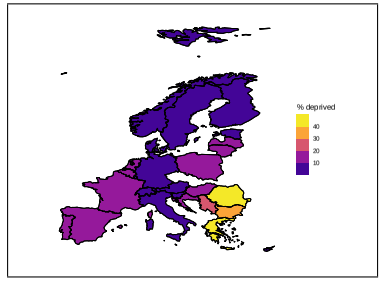
(g) *Replace worn-out furniture*



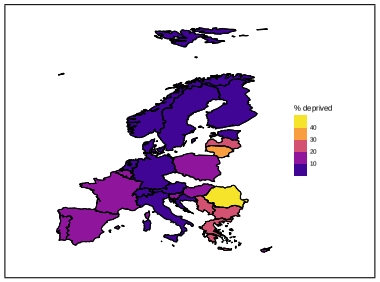
(h) *Replace worn-out clothes*



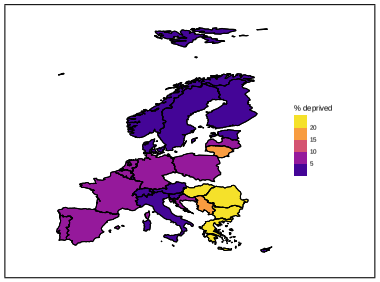
(i) *Two pairs of shoes*



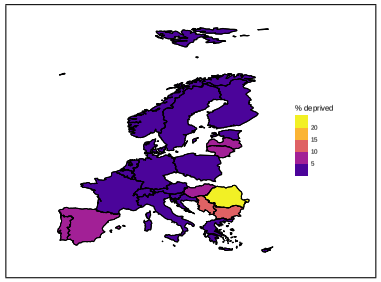
(j) *Pocket money for oneself*



(k) *Regular leisure activities*



(l) *Get together (friends/family)*



(m) *Internet connection*

Source: author's computation using EU-SILC cross-sectional data.
Notes: This figure show the cross-country pattern of material and social deprivation in European countries. Slovakia, Ireland, and Iceland are excluded from the analysis due to missing information for 2018.

2.B Methodological Details

2.B.1 The Mechanics of Xgboost

Suppose there are m observable individual characteristics of size n, which can be used to classify individuals' observed $\mathbf{y}$ deprivation status. A tree ensemble model learns K additive function to classify deprivation status of individual i.

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^{K} f_k(\mathbf{x}_i), \tag{2.1}$$

where f_k represents an independent tree with structure q and leaf weight w. In order to learn the set of functions used in the model we minimise the following regularised loss function.

$$\mathcal{L}(\mathbf{x}_i) = \sum_i \ell(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \tag{2.2}$$

where $\Omega(f) = \gamma T + \frac{1}{2}\lambda||w||^2$, T is the number of leaves in the tree, and w is the weight of the leaf ℓ is a differentiable convex loss function that measures the prediction error, Ω penalises the model's complexity (Chen and Guestrin, 2016).

2.B.2 Hyperparameter space

Table. 2.3 *Hyperparameter grid of Xgboost*

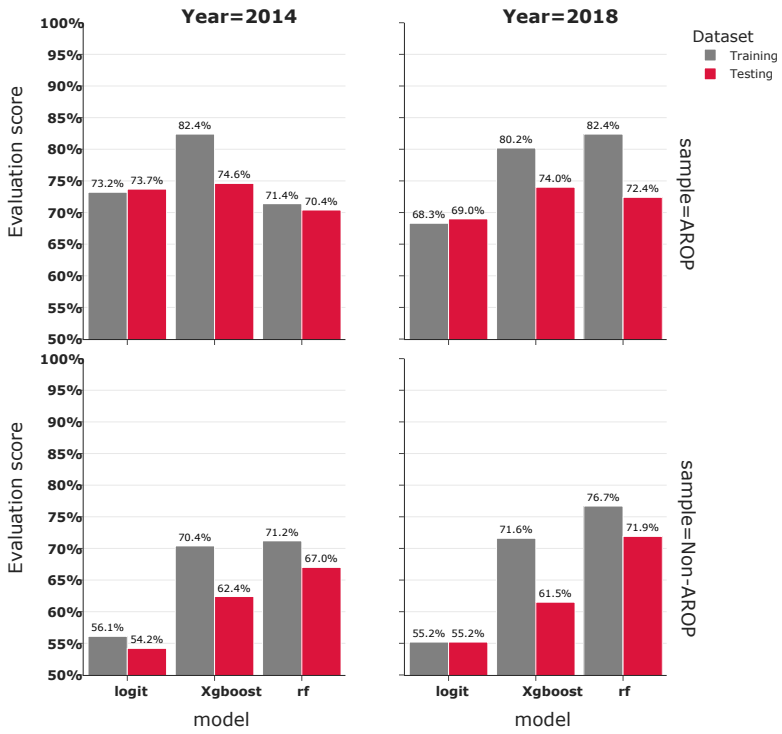
Hyperparameters	Description	Search space
learning rate	Step size shrinkage used in model update.	[0, ..., 0.5]
number of estimators	Number of trees in base learner.	$\{100, ..., 150\}$
maximum depth	Maximum depth of a tree.	$\{5, ..., 30\}$
gamma	Minimum loss reduction required for further partition.	[0, ..., 0.3]
sub-sampling of columns	Sub-sample ratio of features/columns used for fitting the individual tree.	[0.8, ..., 1]

Notes: This table shows combination of hyperparameter space over which the optimal Xgboost model is selected with a ten-fold cross-validation.

2.C Robustness Checks

2.C.1 Subsample Performance Evaluation by At-risk-of-poverty Rate

Figure. 2.11 *In-sample and Out-of-sample performance of prediction models by AROP subsamples, area under the ROC curve*



Notes: Each model is trained on 80% of pooled dataset and tested on the remaining 20% of the dataset. The ROC_AUC is the area under the ROC curve, a measure of trade-off between the true positive rate and the false positive rate at different probability thresholds.

2.C.2 AUC by Different Sets of Features

Table. 2.4 *AUC of pooled model with sequentially inserted set of features, year* 2014

Subset of predictors	Xgboost (1)	Random Forests (2)	Logit (3)
Sociodemographic	0.695 [0.689, 0.702]	0.693 [0.690, 0.695]	0.674 [0.673, 0.676]
Socioeconomic	0.812 [0.810, 0.813]	0.813 [0.812, 0.815]	0.782 [0.781, 0.783]
Health	0.644 [0.643, 0.645]	0.644 [0.643, 0.645]	0.641 [0.639, 0.642]
Location FE	0.726 [0.725, 0.728]	0.726 [0.725, 0.728]	0.726 [0.725, 0.728]
All variables	0.881 [0.880, 0.883]	0.870 [0.868, 0.872]	0.821 [0.820, 0.822]

Notes: The ROC_AUC is the area under the ROC curve, a measure of trade-off between the true positive rate and the false positive rate at different probability thresholds. The figures in square parentheses are 95% confidence interval of AUC computed via ten-fold stratified sampling repeated 100 times.

2.C.3 Alternative Prediction Metrics

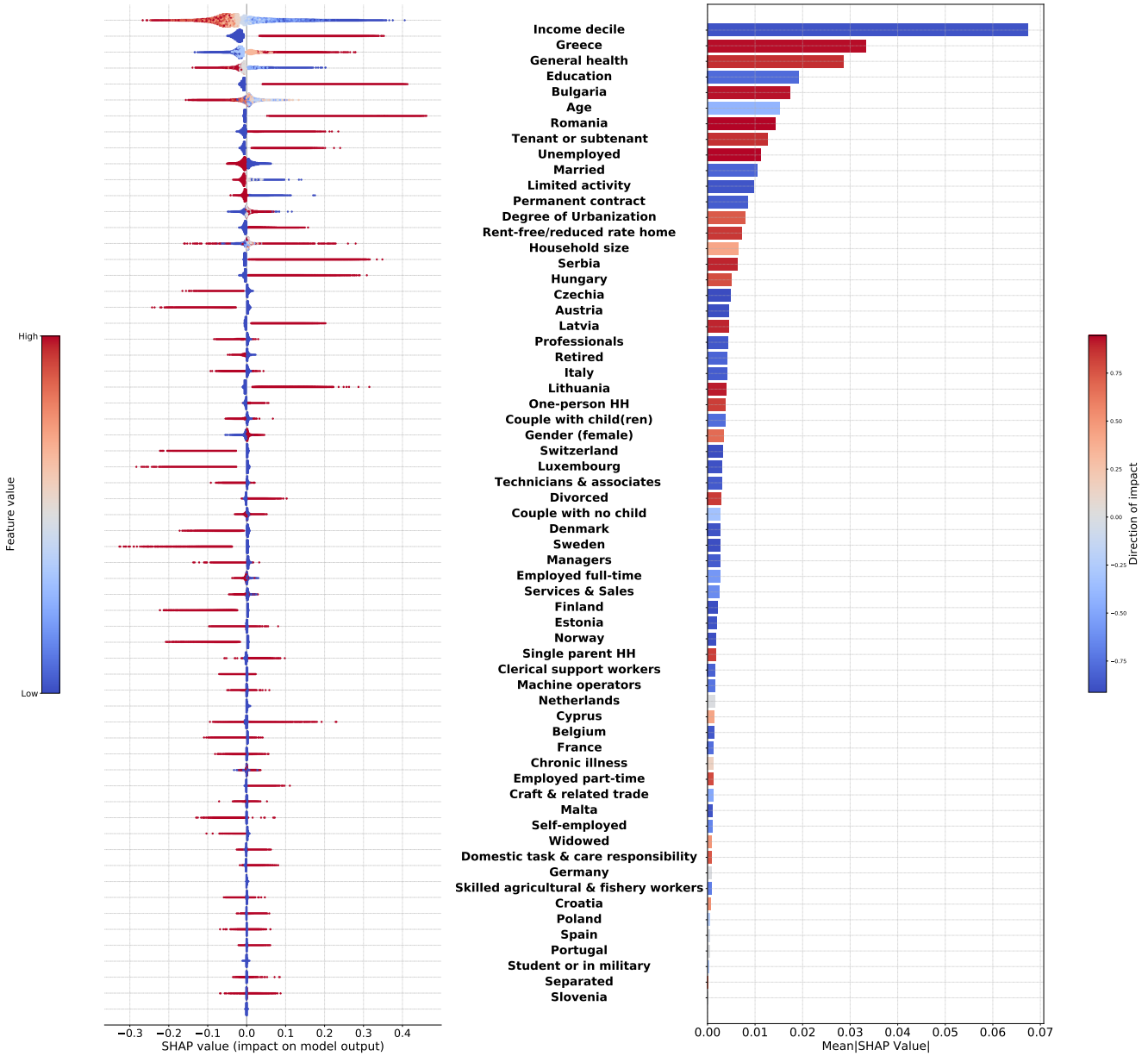
Table. 2.5 *Performance of Xgboost with model specific hyperparameters configuration compared with baseline Logit model*

	Accuracy		Precision		F1-score	
	Xgboost	Logit	Xgboost	Logit	Xgboost	Logit
Austria	0.9922 (0.95)	0.9557 (0.94)	0.9923[c](0.94[c])	0.9135 (0.89)	0.9919[c](0.95)	0.9341 (0.92)
Belgium	0.9786[c](0.89)	0.8929 (0.89)	0.9786[a](0.88[b])	0.7972 (0.79)	0.9777[b](0.88[c])	0.8424 (0.84)
Bulgaria	0.853[a](0.73[b])	0.6208 (0.61)	0.8521[a](0.73[a])	0.3855 (0.38)	0.8519[a](0.73[a])	0.4756 (0.47)
Croatia	0.9146[c](0.85)	0.8596 (0.85)	0.9126[a](0.82[b])	0.7389 (0.73)	0.9021[b](0.82)	0.7947 (0.79)
Cyprus	0.9517[c](0.86)	0.8736 (0.86)	0.9532[a](0.84[b])	0.7633 (0.75)	0.947[b](0.84)	0.8147 (0.8)
Czechia	0.9769 (0.94)	0.9454 (0.94)	0.9772[c](0.92)	0.8938 (0.89)	0.9741[c](0.93)	0.9189 (0.92)
Denmark	0.9931[c](0.92)	0.9434 (0.94)	0.9932[b](0.9)	0.8909 (0.89)	0.9929[c](0.91)	0.9164 (0.91)
Estonia	0.9611 (0.92)	0.918 (0.92)	0.9618[b](0.9[c])	0.8428 (0.85)	0.9558[c](0.91)	0.8788 (0.88)
Finland	0.9956 (0.96)	0.9604 (0.96)	0.9956[c](0.95)	0.9225 (0.92)	0.9955[c](0.95)	0.9411 (0.94)
France	0.9516[c](0.89)	0.8859 (0.88)	0.9519[a](0.87[b])	0.7867 (0.78)	0.9465[b](0.87[c])	0.8334 (0.82)
Germany	0.9619 (0.93)	0.9333 (0.93)	0.9602[b](0.92[c])	0.8712 (0.86)	0.9564[c](0.92)	0.9012 (0.89)
Greece	0.7774[b](0.72[c])	0.6788 (0.67)	0.777[a](0.71[a])	0.4608 (0.46)	0.7563[a](0.69[a])	0.5489 (0.55)
Hungary	0.9084[b](0.83[c])	0.7884 (0.79)	0.9095[a](0.81[a])	0.6215 (0.62)	0.9013[a](0.81[b])	0.6951 (0.69)
Italy	0.9397 (0.91)	0.9058 (0.9)	0.9378[b](0.89[c])	0.8205 (0.81)	0.9289[c](0.89)	0.861 (0.85)
Latvia	0.8842[b](0.77)	0.7658 (0.75)	0.8844[a](0.74[a])	0.5865 (0.57)	0.875[a](0.74 [b])	0.6642 (0.65)
Lithuania	0.8874[b](0.78)	0.7609 (0.75)	0.8882[c](0.75)	0.8181 (0.81)	0.879[a](0.75[b])	0.6579 (0.64)
Luxembourg	0.9914 (0.95)	0.9604 (0.96)	0.9914[c](0.94)	0.9224 (0.94)	0.9908[c](0.95)	0.941 (0.95)
Netherlands	0.9848 (0.94)	0.9557 (0.96)	0.9847[c](0.93)	0.9133 (0.92)	0.9835[c](0.93)	0.934 (0.94)
Norway	0.9989 (0.96)	0.9674 (0.96)	0.9989[c](0.95[c])	0.9429 (0.94)	0.9989 (0.96)	0.9541 (0.95)
Poland	0.9334[c](0.88)	0.8868 (0.89)	0.9343[b](0.85)	0.7979 (0.8)	0.9221[b](0.86)	0.8351 (0.84)
Portugal	0.8936[c](0.84)	0.8345 (0.82)	0.8905[a](0.81[b])	0.6964 (0.68)	0.8783[b](0.81[c])	0.7592 (0.75)
Romania	0.813[a](0.7[a])	0.5788 (0.58)	0.8124[b](0.7 b)	0.7342 (0.61)	0.8119[a](0.7[a])	0.426 (0.42)
Serbia	0.8654[a](0.75[c])	0.6899 (0.69)	0.8656[a](0.73[a])	0.4759 (0.47)	0.8599[a](0.73[a])	0.5633 (0.56)
Slovenia	0.9382 (0.91)	0.9145 (0.92)	0.9384[b](0.87)	0.8363 (0.84)	0.923[c](0.89)	0.8737 (0.88)
Spain	0.9228[c](0.87)	0.8591 (0.85)	0.9211 (0.84[b])	0.8789 (0.73)	0.9135[b](0.85[c])	0.794 (0.79)
Sweden	0.9995 (0.97)	0.9768 (0.97)	0.9995 (0.96)	0.9543 (0.95)	0.9995 (0.97)	0.9654 (0.96)
Switzerland	0.9914 (0.97)	0.9684 (0.96)	0.9914[c](0.96)	0.9378 (0.94)	0.9907 (0.96)	0.9529 (0.95)

[a] At least 20% relative gain.
[b] Between 10% and 20% relative gain.
[c] At most 10% relative gain.

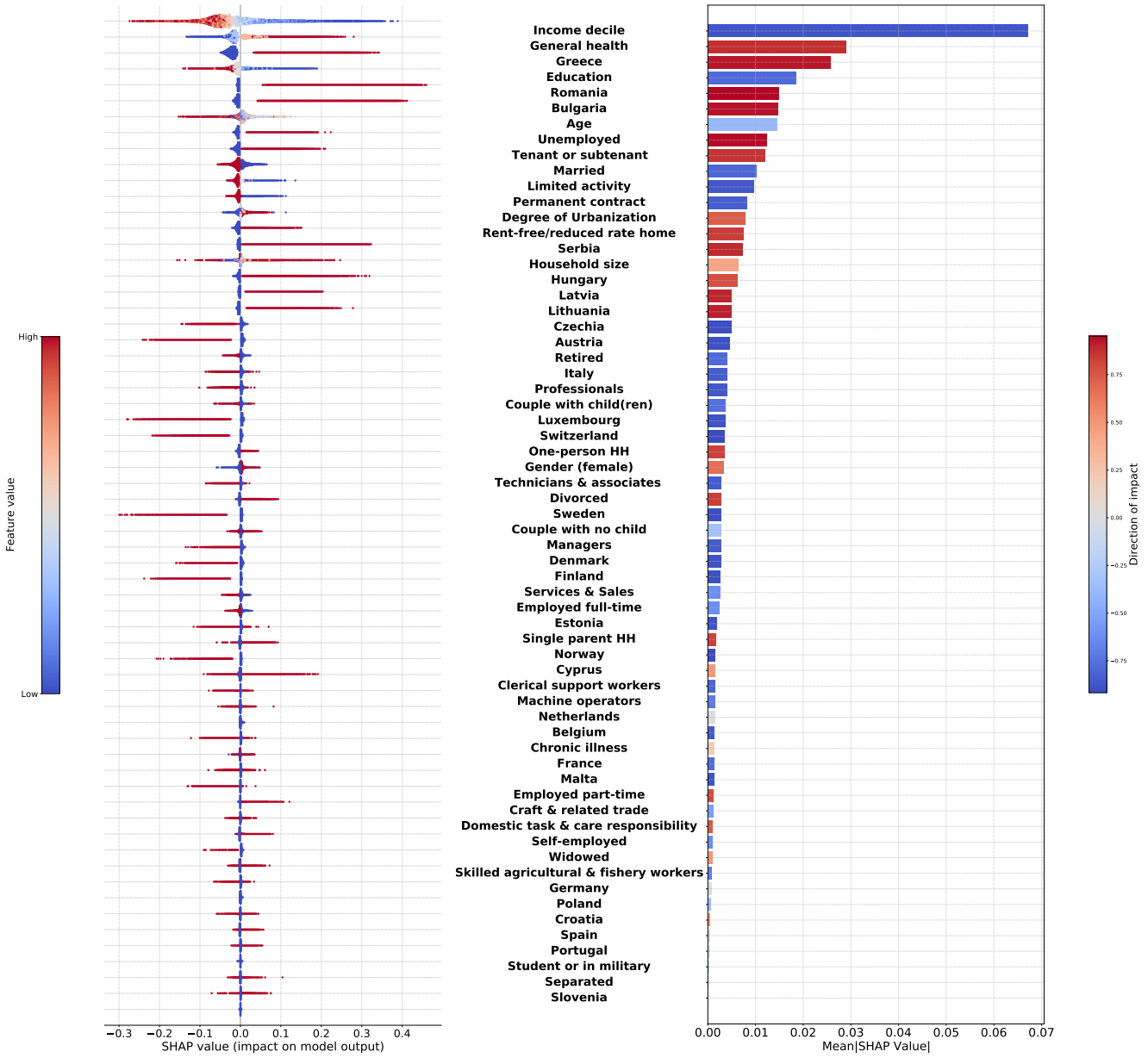
2.D Shapley Values of all Feature

Figure. 2.12 *SHAP summary plot of all features, year 2018*



(a) *a* **(b)** *b*

Figure. 2.13 *SHAP summary plot of all features, year 2014*

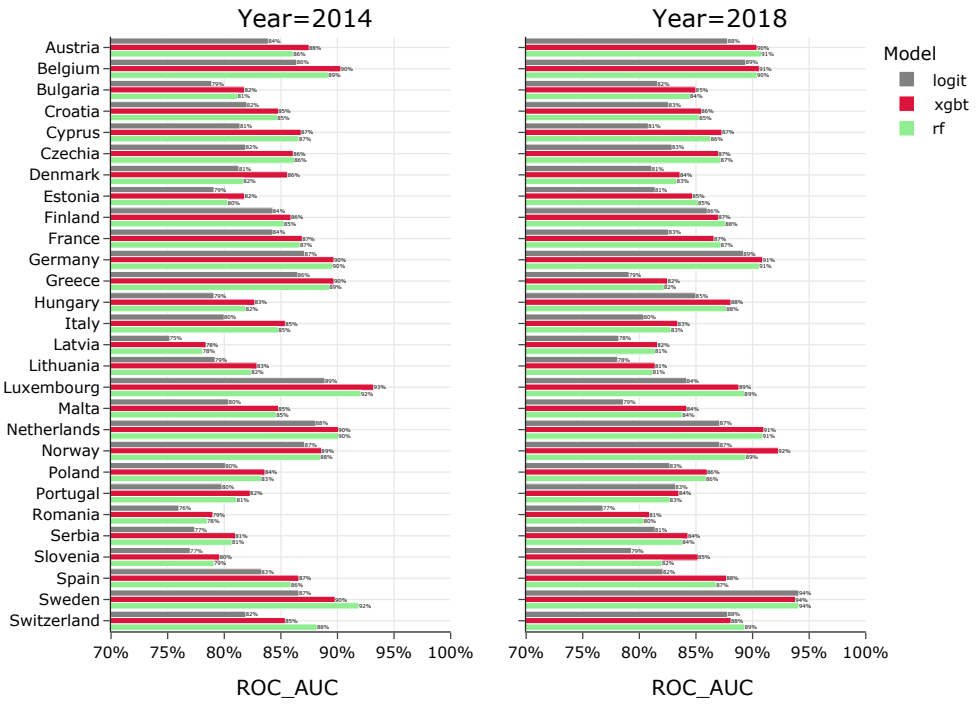


(a) *a* **(b)** *b*

2.E Heterogeneity Analyses by Country

2.E.1 AUC in the Per-country Analysis

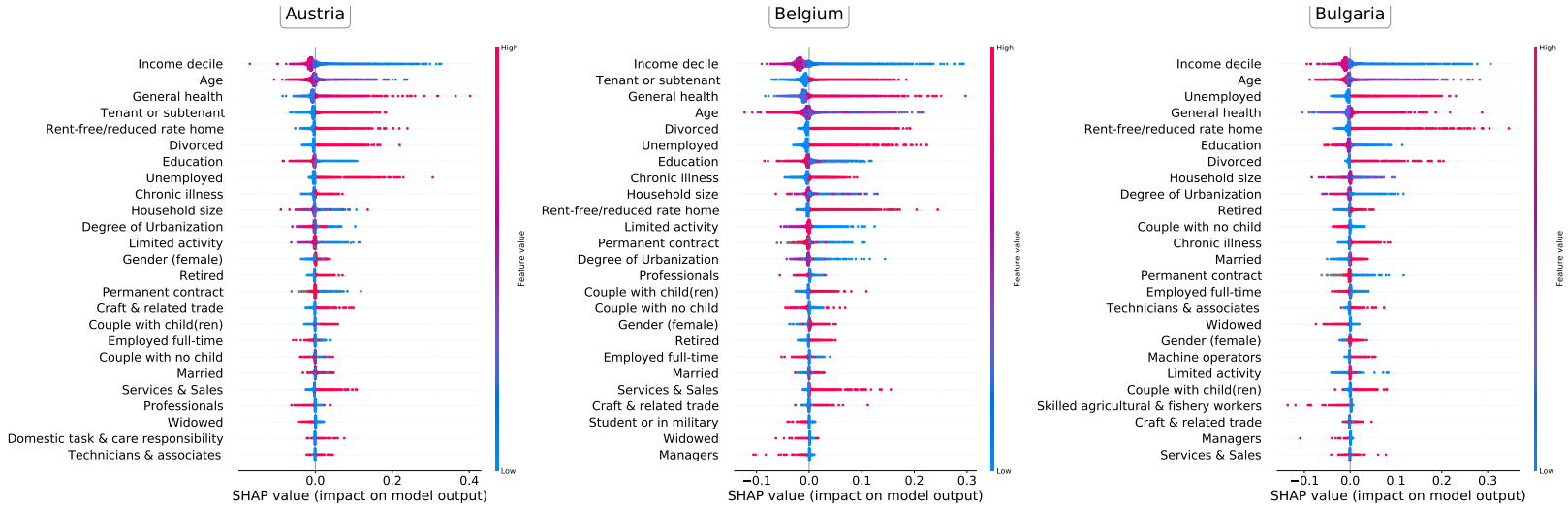
Figure. 2.14 *Out-of-sample performance of prediction models per each country, area under the ROC curve*



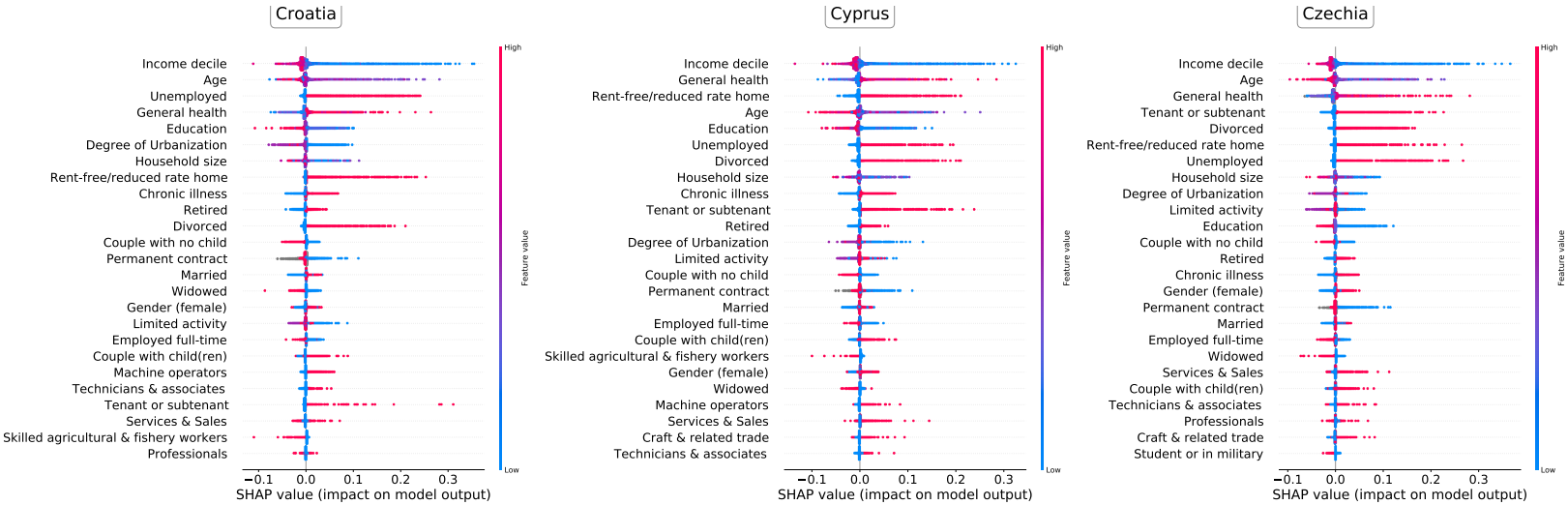
Notes: Each model is trained on 80% of pooled and per-country datasets and tested on the remaining 20% of the datasets . The ROC_AUC is the area under the ROC curve, a measure of trade-off between the true positive rate and the false positive rate at different probability thresholds. The region of residence information is missing in Germany, Netherlands and Portugal. Degree of urbanisation variable has all missing values for Germany, Netherlands, and Slovenia.

2.E.2 SHAP in the Per-country Analysis

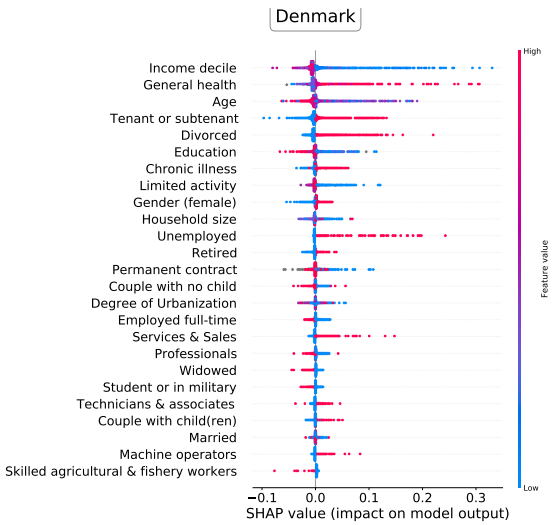
Figure. 2.15 *SHAP summary plots for each country*



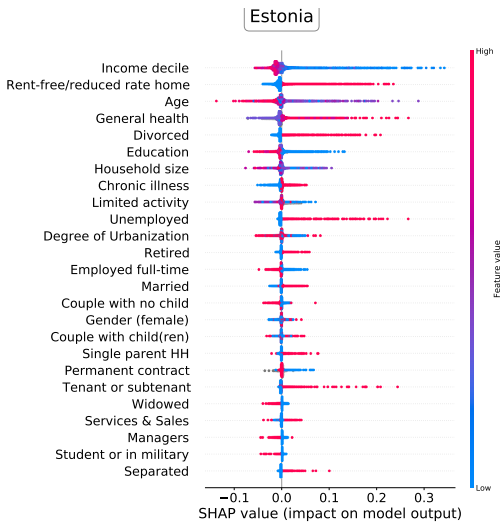
(1) *Austria* **(2)** *Belgium* **(3)** *Bulgaria*



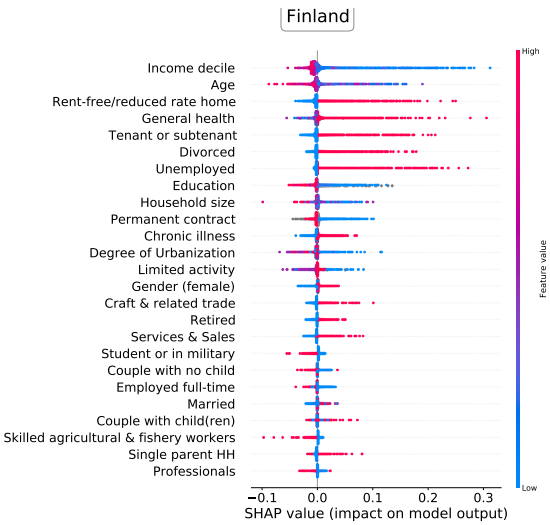
(4) *Croatia* **(5)** *Cyprus* **(6)** *Czechia*



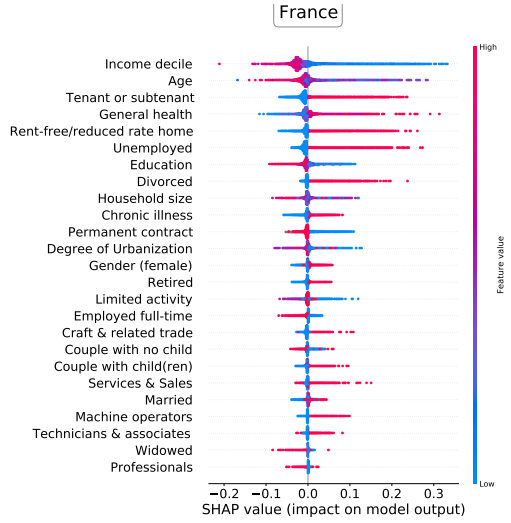
(7) *Denmark*



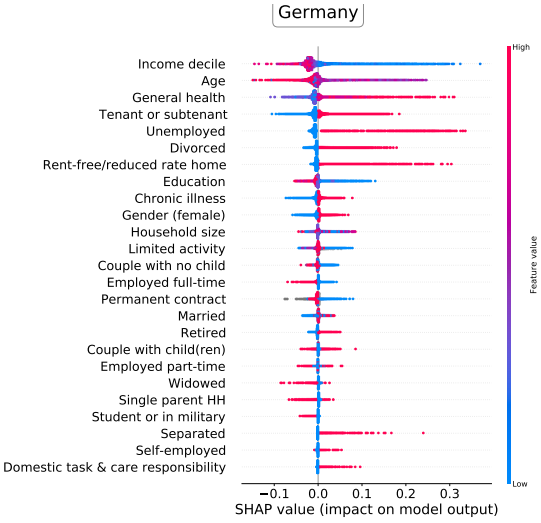
(8) *Estonia*



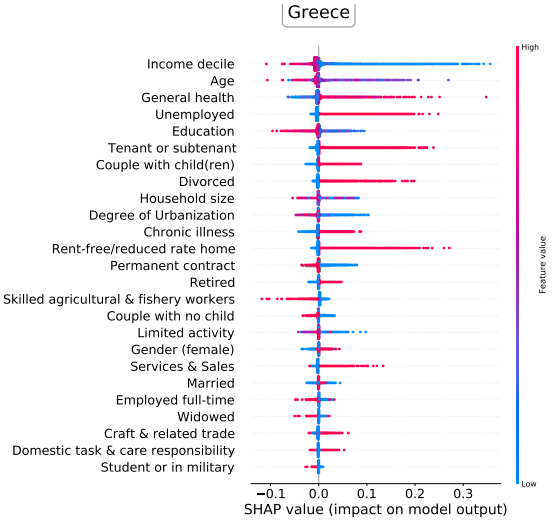
(9) *Finland*



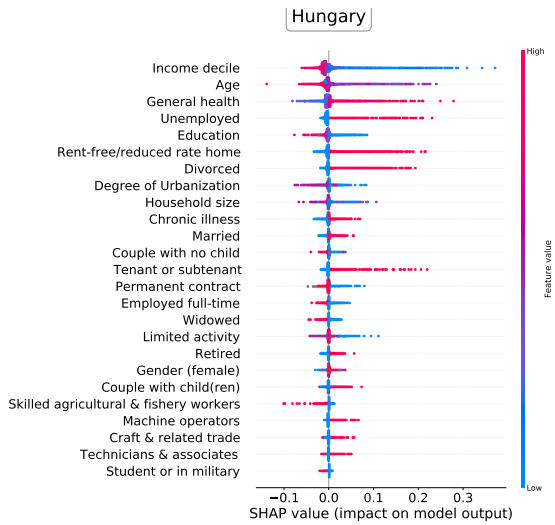
(10) *France*



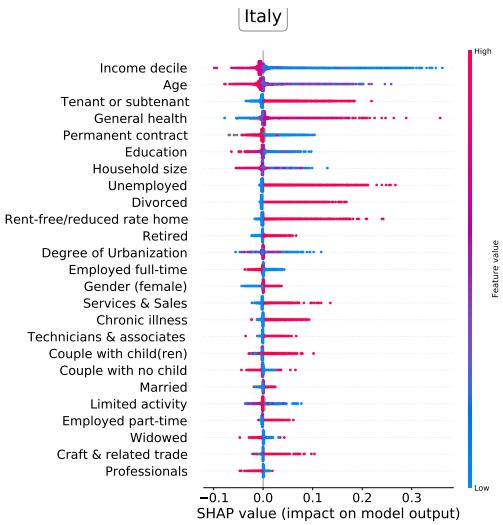
(11) *Germany*



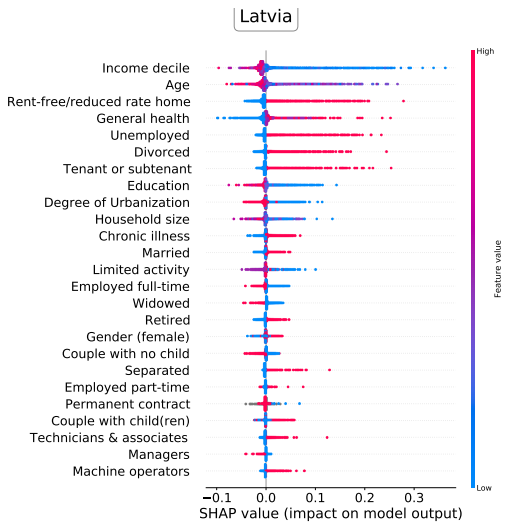
(12) *Greece*



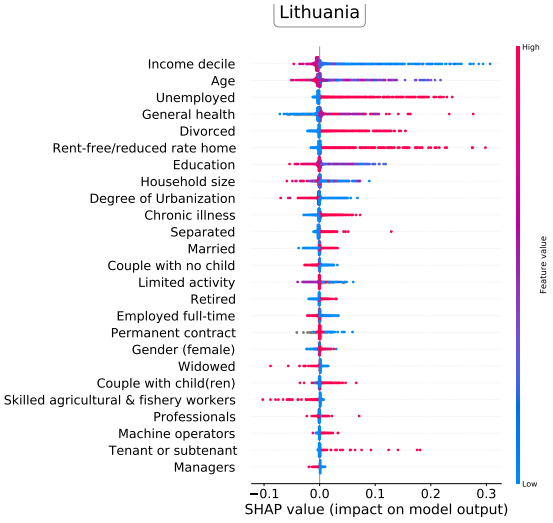
(13) *Hungary*



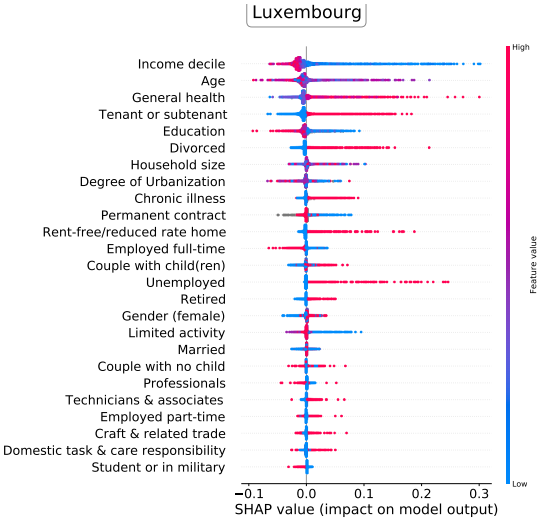
(14) *Italy*



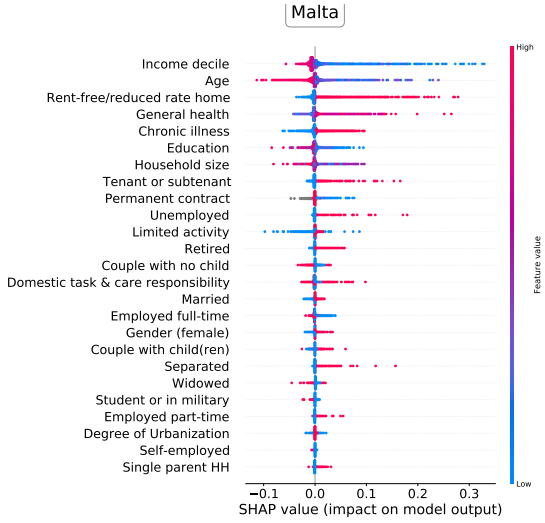
(15) *Latvia*



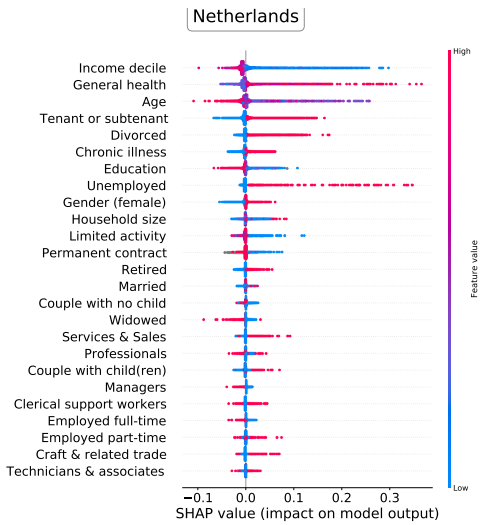
(16) *Lithuania*



(17) *Luxembourg*

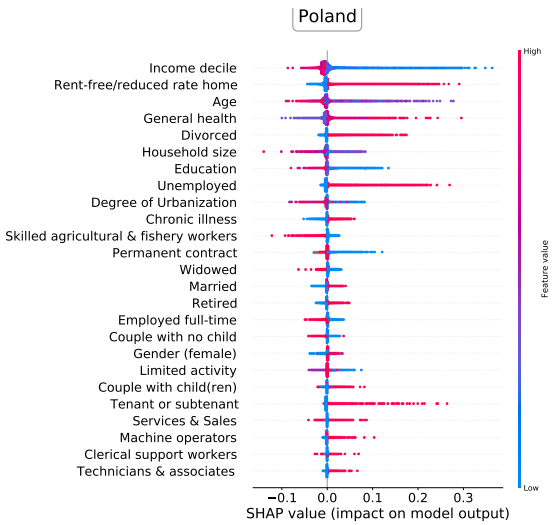


(18) *Malta*

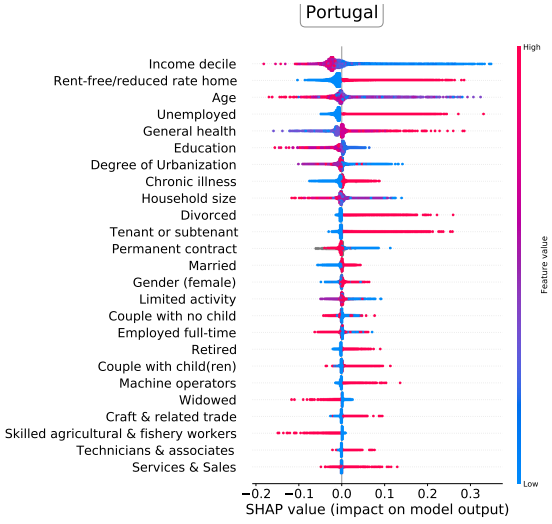


(19) *Netherlands*

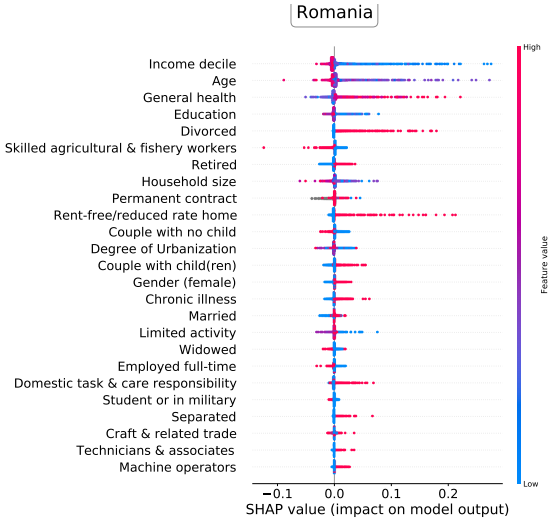
(20) *Norway*



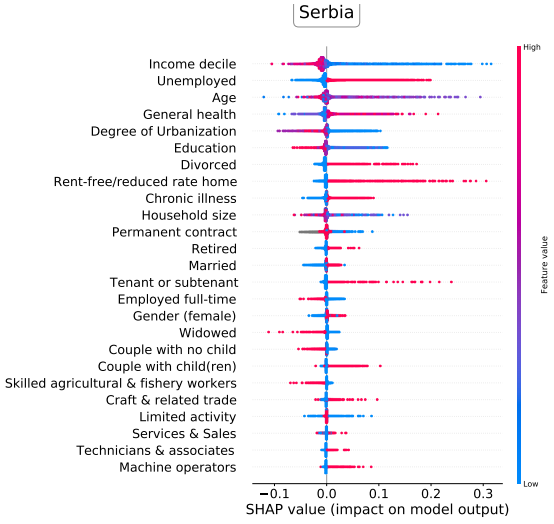
(21) *Poland*



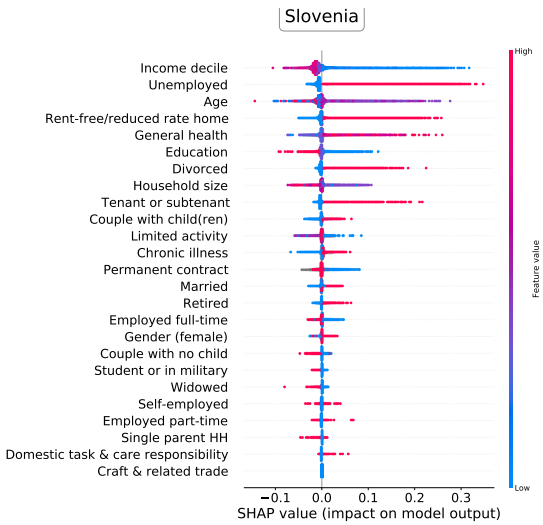
(22) *Portugal*



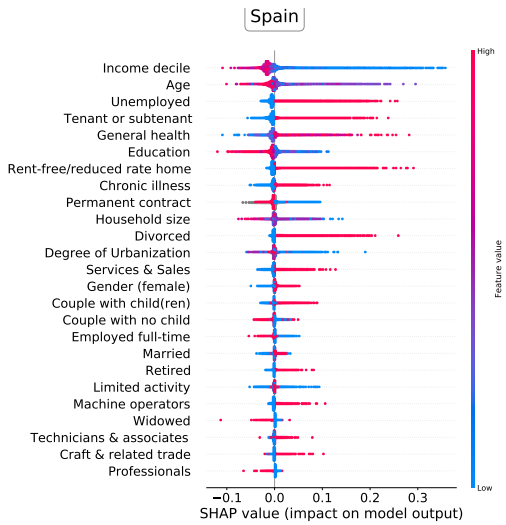
(23) *Romania*



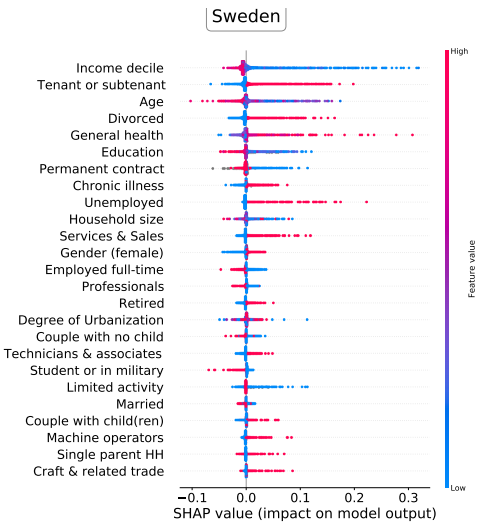
(24) *Serbia*



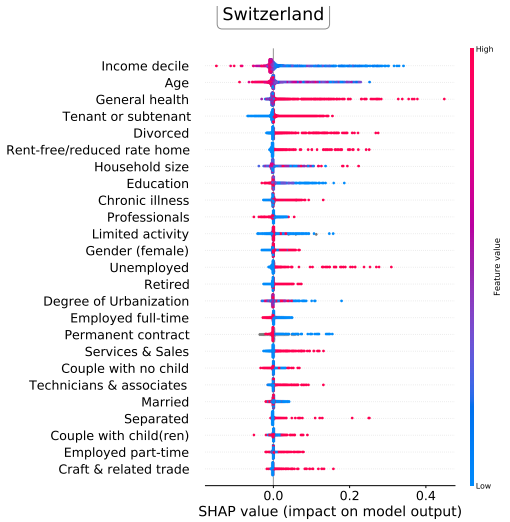
(25) *Slovenia*



(26) *Spain*

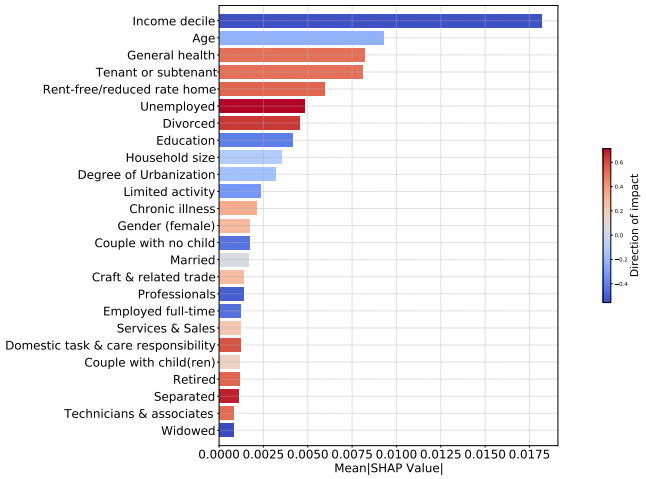


(27) *Sweden*

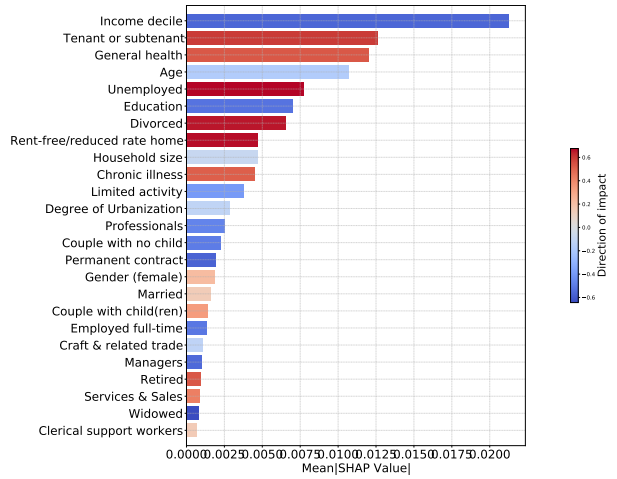


(28) *Switzerland*

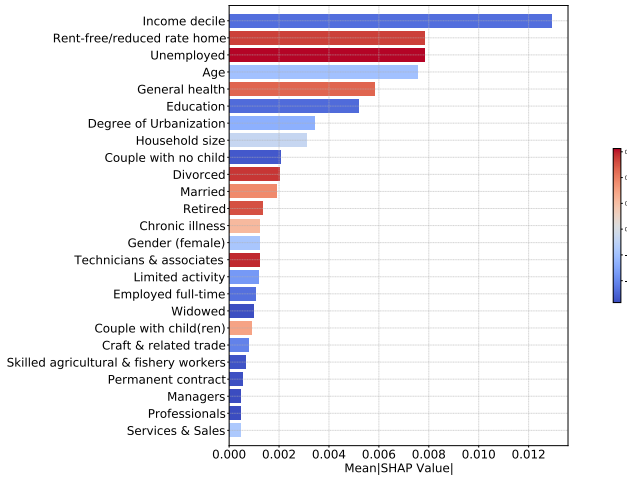
Figure. 2.12 *Global impact of the key features for each country*



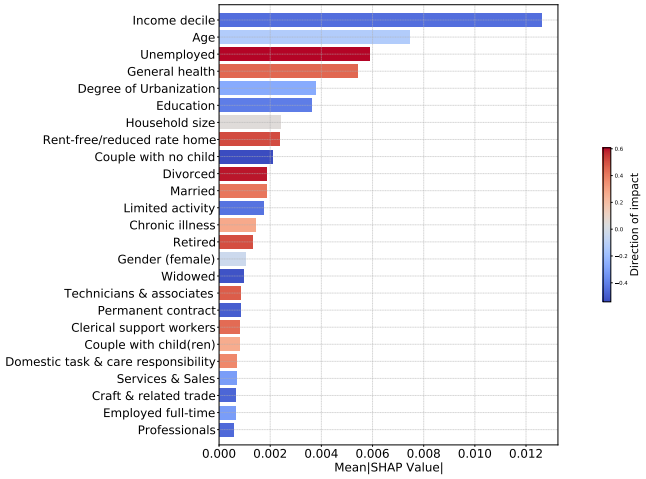
(1) *Austria*



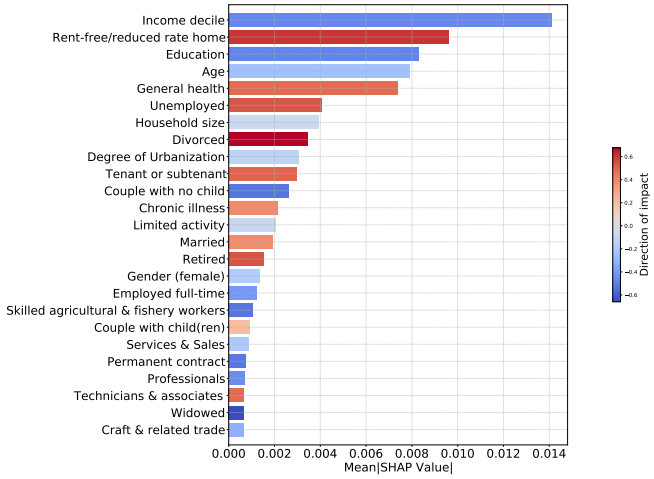
(2) *Belgium*



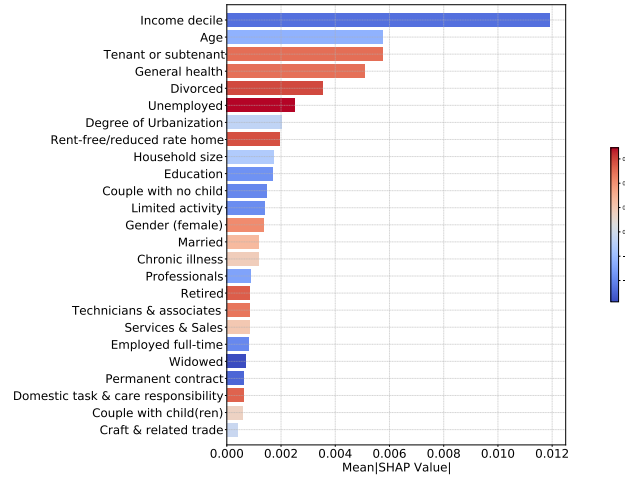
(3) *Bulgaria*



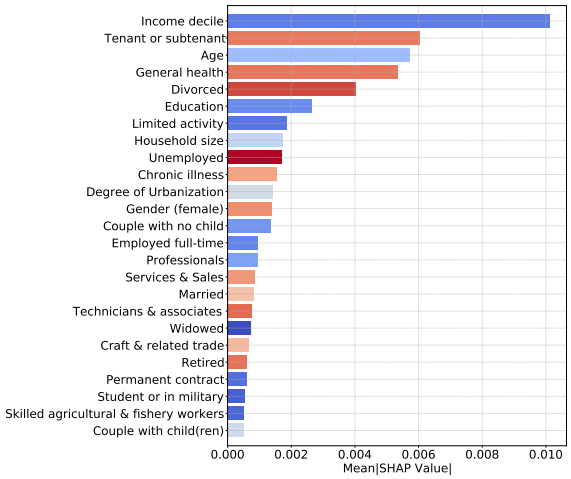
(4) *Croatia*



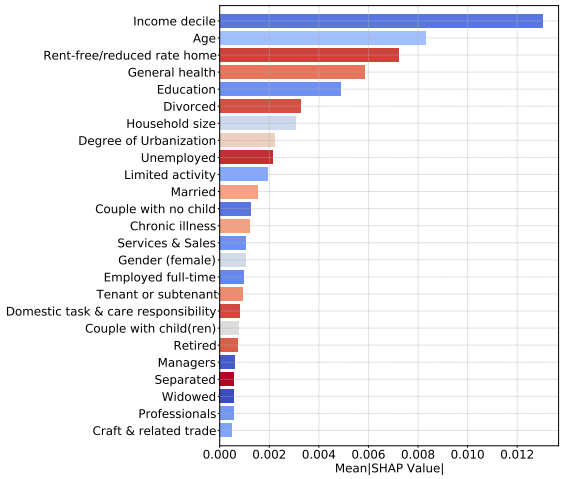
(5) *Cyprus*



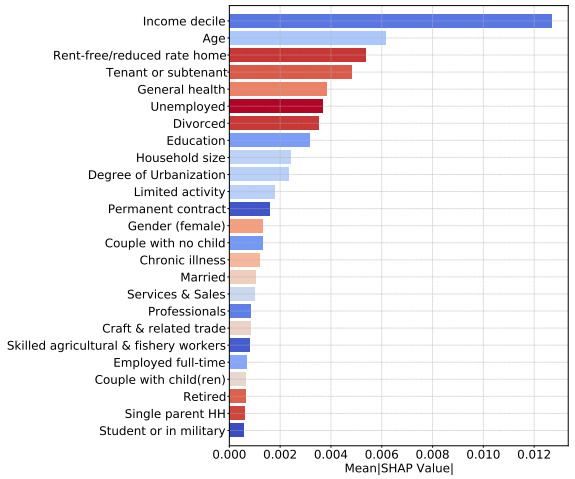
(6) *Czechia*



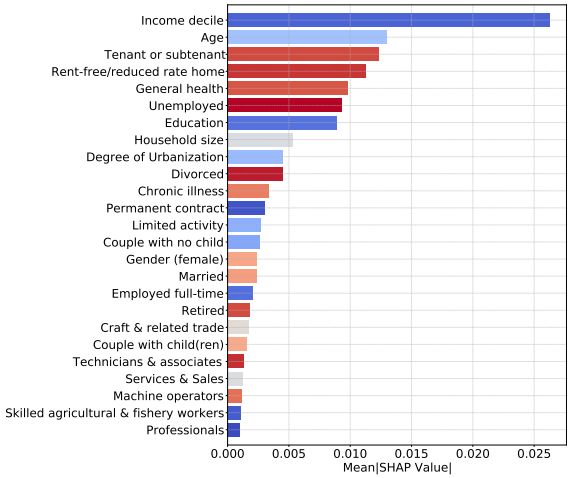
(7) *Denmark*



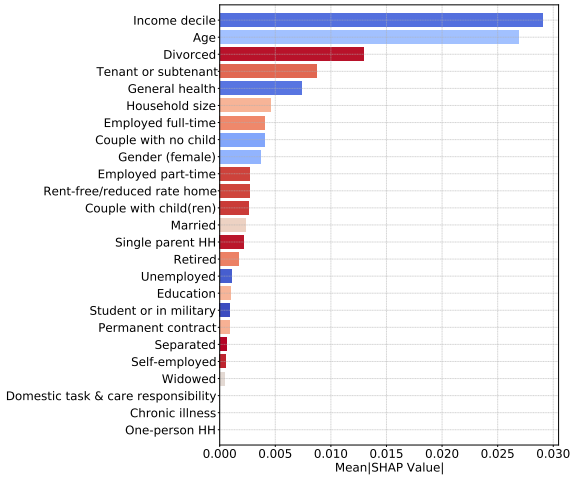
(8) *Estonia*



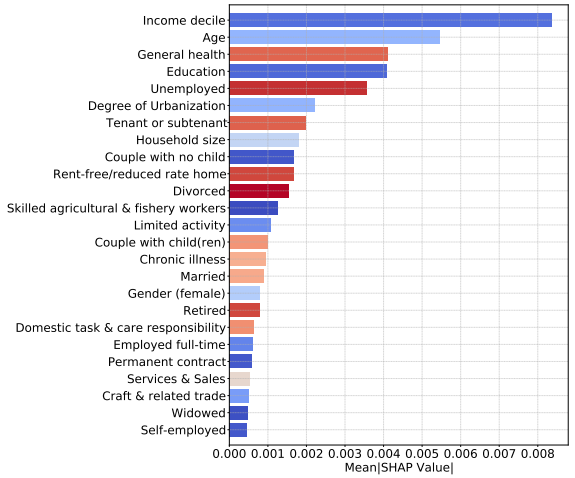
(9) *Finland*



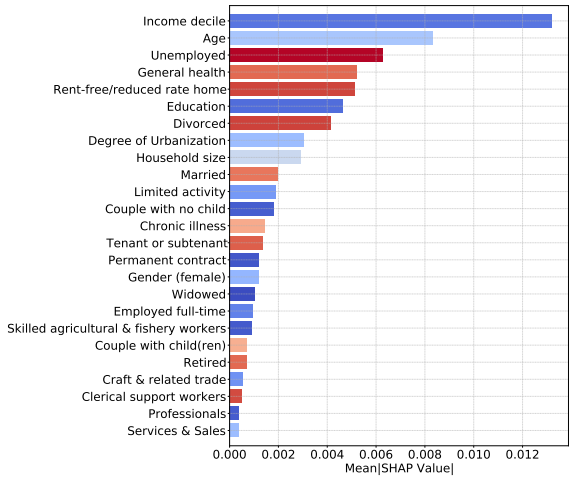
(10) *France*



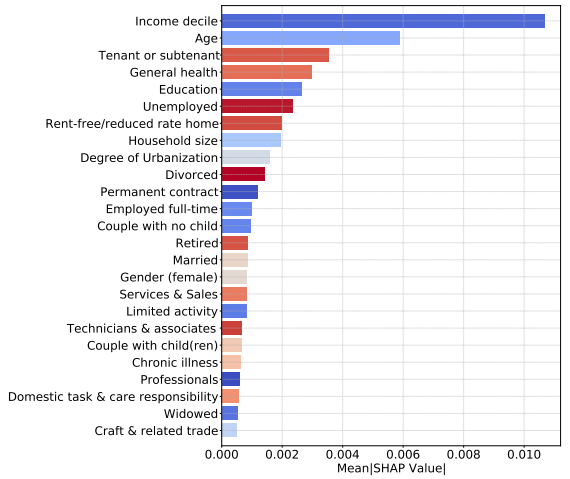
(11) *Germany*



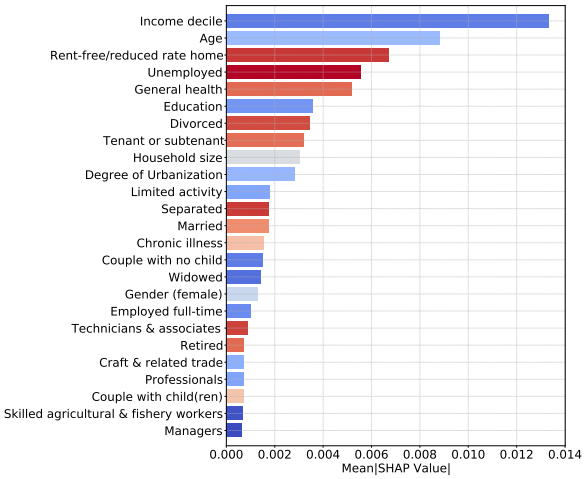
(12) *Greece*



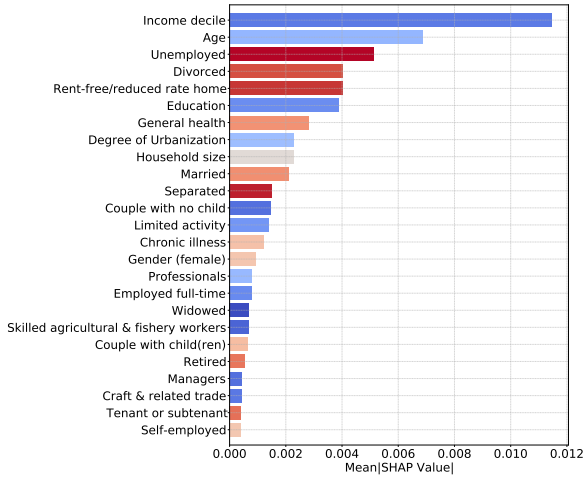
(13) *Hungary*



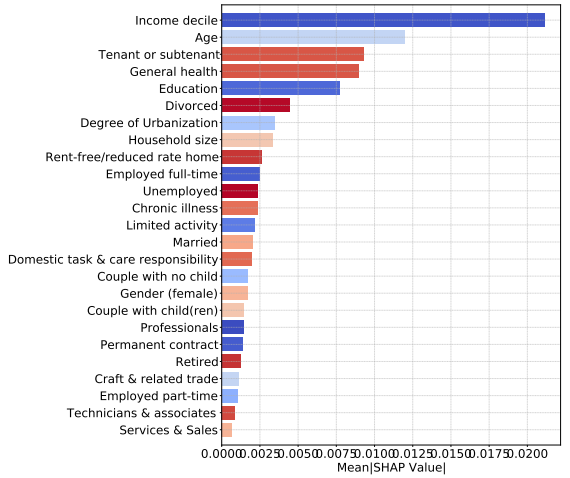
(14) *Italy*



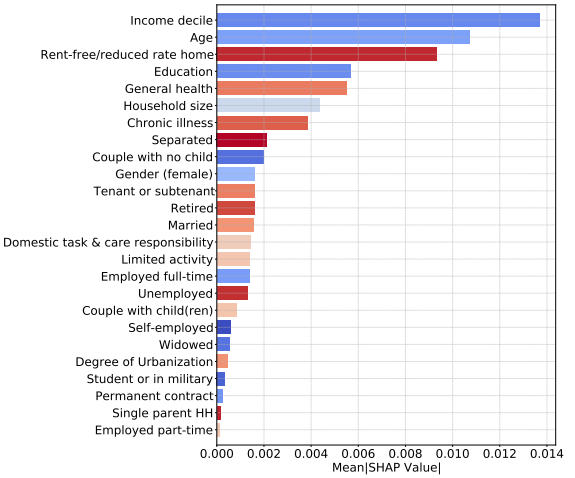
(15) *Latvia*



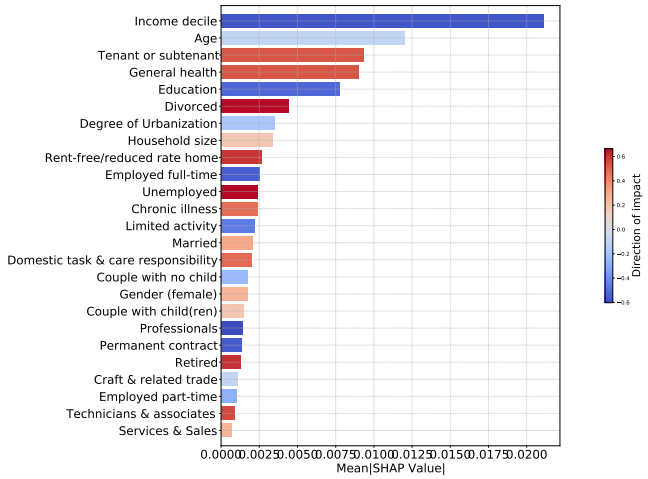
(16) *Lithuania*



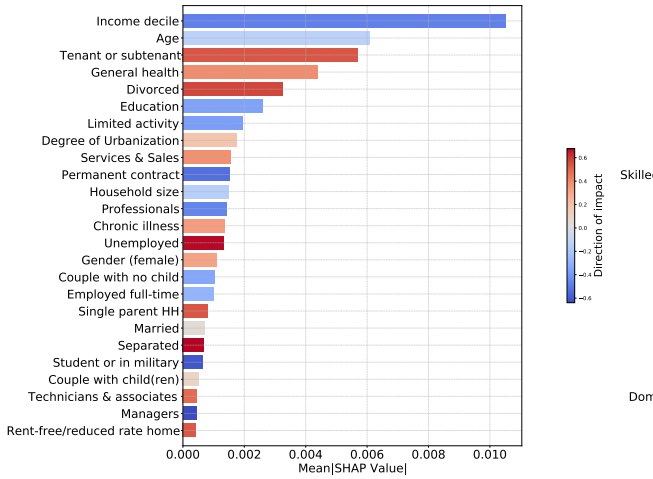
(17) *Luxembourg*



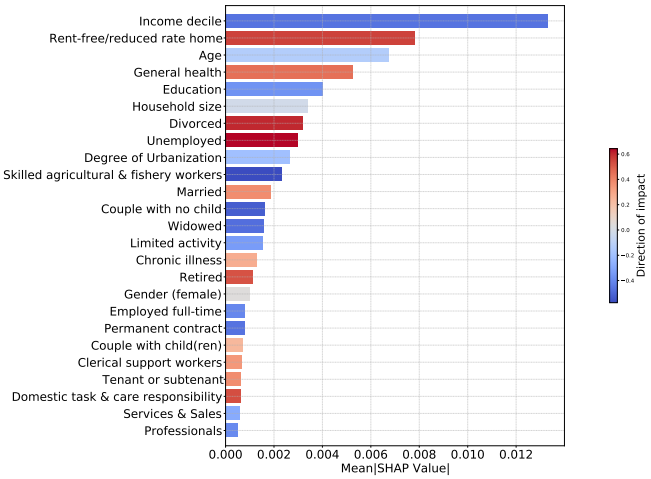
(18) *Malta*



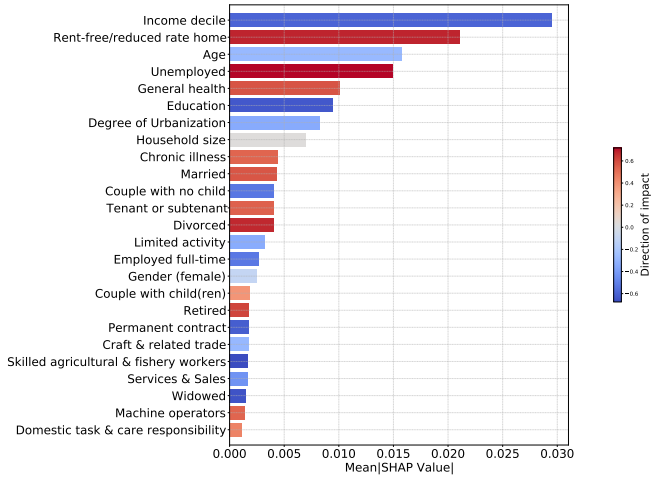
(19) *Luxembourg*



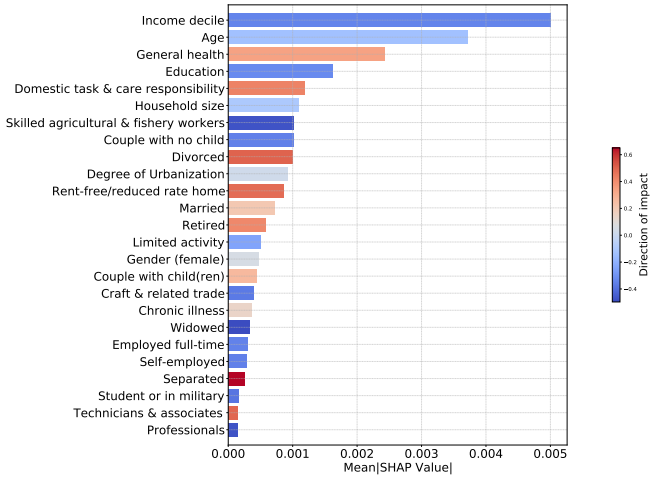
(20) *Norway*



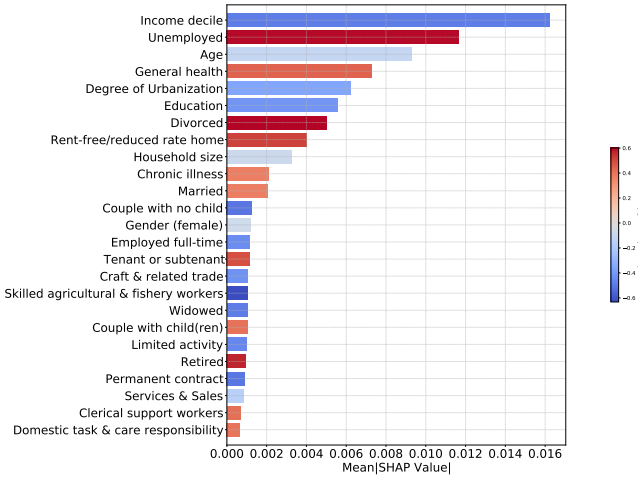
(21) *Poland*



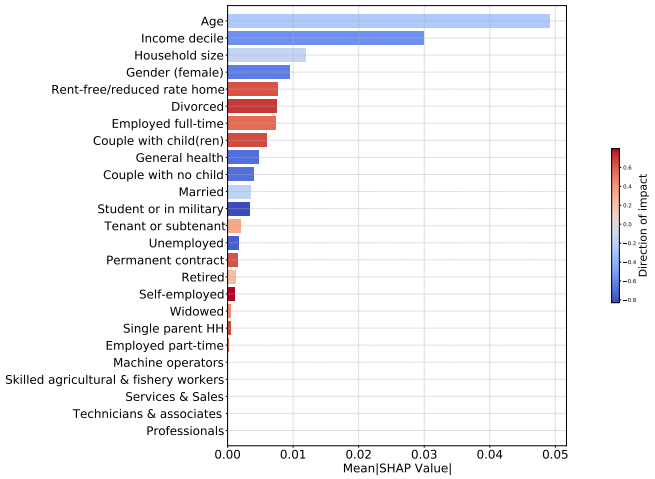
(22) *Portugal*



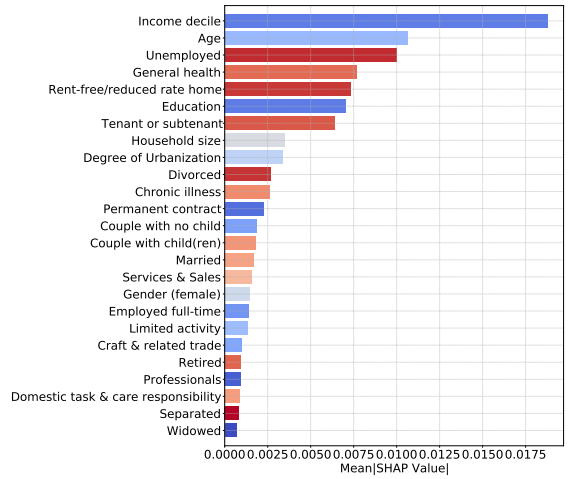
(23) *Romania*



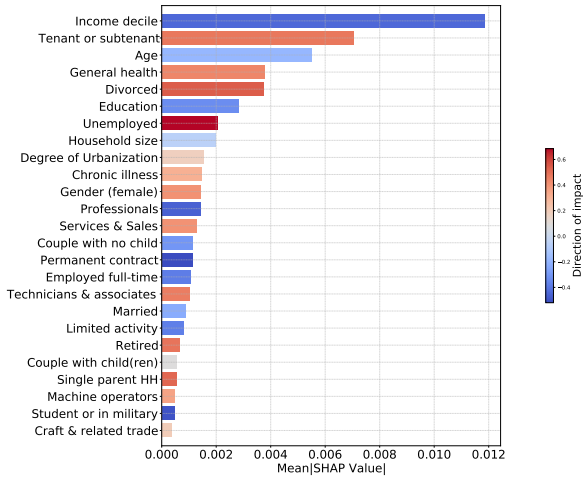
(24) *Serbia*



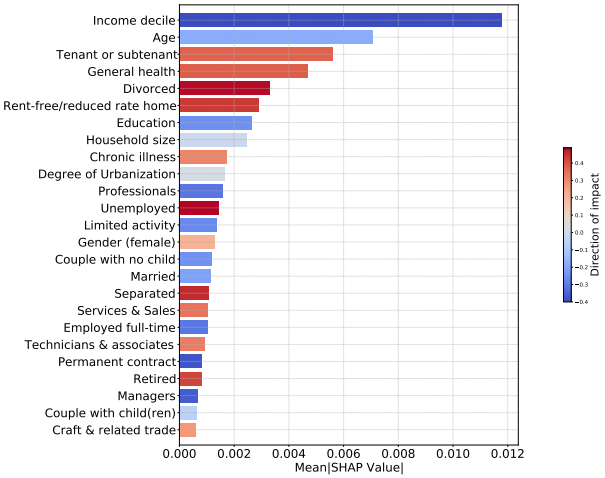
(25) *Slovenia*



(26) *Spain*



(27) *Sweden*



(28) *Switzerland*

Chapter 3

Predictors of Self-protecting Behaviours During the Early Wave of the COVID-19 Pandemic: A Machine Learning Approach *

With Liyousew G. BORGA (Luxembourg Institute of Health),
Claus VÖGELE (University of Luxembourg),
Samuel GREIFF (University of Luxembourg),
and Conchita D'AMBROSIO (University of Luxembourg)

*The data used in this study mainly come from COME-HERE (COVID-19, MEn-tal HEalth, Resilience, and Self-regulation) longitudinal survey collected by the University of Luxembourg in five European countries (France, Germany, Italy, Spain, and Sweden). COME-HERE is sponsored by the André Losch Fondation, Art2Cure, Cargolux, CINVEN Foundation and COVID-19 Foundation, under the aegis of the Fondation de Luxembourg, and Fonds National de la Recherche Luxembourg (14840950 – COME-HERE). The views expressed here are those of the authors and not necessarily those of the abovementioned sponsors.

3.1 Introduction

On 11 March 2020 the World Health Organisation (WHO) declared COVID-19, the disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a pandemic. During the first wave of the pandemic, without a vaccine or therapeutic measures, governments had to rely on behavioural interventions to slow the spread of the virus and reduce the number of infections. Authorities launched confinement policies - such as lockdowns, travel restrictions, and social distancing requirements - and preventive sanitary measures - such as mask wearing and frequent handwashing. Even though it is apparent that human behaviour largely influences the spread of the virus, it is unclear which factors are most strongly associated with protective behaviours. Identifying these factors is of paramount importance for devising effective policies to manage the current pandemic as well as to be better prepared for future ones.

Using a machine learning approach, we examine how individual characteristics and government policy responses predict self-protecting behaviours during the earliest wave of the pandemic. We use a unique harmonised real-time data set, the COME-HERE (COVID-19, MEntal HEalth, Resilience, and Self-regulation) longitudinal survey, collected by the University of Luxembourg in five European countries (France, Germany, Italy, Spain, and Sweden) covering a range of individual-level information such as sociodemographic variables, income and wealth, health and behavioural risk factors, awareness about the pandemic, and trust in major public institutions. In addition, we consult country level information describing the evolution of the pandemic itself, as well as the policy responses to COVID-19, from the Blavatnik School of Government of the University of Oxford COVID-19 government response tracker (Hale et al., 2021).

Our work relates to two strands of the literature. First, we complement studies that try to understand self-protecting behaviours during the COVID-19 pandemic. Based on health behaviour models, Aubert and Augeraud-Véron (2021) frame self-protection efforts as the outcome of a Nash equilibrium in which individuals choose the best response to their economic, psychological, and epidemiological environment. Similarly, Brown and Ravallion (2020) posit that people trade-off the perceived benefit of social distancing against the cost, which includes costs in adjusting to the pandemic from a pre-pandemic level and the expected future cost of being infected.

In line with these model-based predictions, there is evidence suggesting that the most important determinants of self-protecting behaviours are socioeconomic factors. Papageorge et al. (2021) find that in the US, the costs of self-protective behaviours are unevenly distributed across socio-demographic groups such that economically disad-

vantaged individuals, who presume low risk of adverse health effects, are less likely to engage in protective behaviours. Ashraf (2020) shows a strong negative association between COVID-19 cases and socioeconomic conditions in a panel of 77 countries. Previous results also show that socioeconomic disparities in health behaviours create an imbalance between individual responses and socially optimal outcomes (Adda, 2016; Cutler and Lleras-Muney, 2010; Pampel et al., 2010).

Other factors influencing safe behaviour that have been identified in the related literature include average neighbourhood income (Wright et al., 2020), internet access (Chiou and Tucker, 2020), and political affiliation (Barrios and Hochberg, 2020; Painter and Qiu, 2020; Simonov et al., 2020). Other studies emphasise the role of perceived severity of COVID-19 and perceived susceptibility to it (Abel et al., 2021; Jadil and Ouzir, 2021; Vacondio et al., 2021), self-efficacy (Jørgensen et al., 2021), demographics (Dohle et al., 2020), and information source and credibility (Lep et al., 2020).

The second strand of the literature our paper relates to is the growing number of studies that apply a machine learning approach in social science investigations. Machine learning is a class of flexible algorithmic and statistical techniques that focus on prediction (regression) and classification, as well as clustering (grouping) (Athey, 2018). Recent studies in the social sciences have used machine learning methods to estimate how the effect of a particular intervention differs across characteristics of individuals (Athey and Imbens, 2016; Grimmer et al., 2017; Imai and Ratkovic, 2013; Künzel et al., 2019; Wager and Athey, 2018). With regards to the COVID-19 pandemic, a few studies employed machine learning methods to isolate and predict key indicators of compliance with social distancing rules (Van Lissa et al., 2020), fear and perceived health (Eder et al., 2021) and psychological distress (Prout et al., 2020).

Our results are consistent with the predictions of epidemiological and health belief models that posit that health behaviour is activated based on locally relevant perceptions of threats, susceptibility, and benefits in engaging in protective behaviour (Karl et al., 2022); and with models of social epidemiology that established a socioeconomic gradient in health behaviour (Pampel et al., 2010). We find that a higher regional incidence of COVID-19 triggers higher levels of self-protective behaviour, as does a stricter government policy response. The level of individual knowledge about the pandemic, confidence in institutions, and population density also rank high among the factors that predict self-protecting behaviours. Among socio-demographic factors, gender, marital status, age, and region of residence are the main determinants of self-protective measures. Consistent with the evidence in the extant literature that documents a socioeconomic gradient in health, we show that housing features and household income

are among the strongest positive predictors of self-protecting behaviours.

Our study makes several contributions to the related literature. First, we use a combination of complex non-parametric machine learning model and state-of-the-art model explanation method to explain factors impacting the adoption of self-protecting behaviours during the COVID-19 pandemic. To the best of our knowledge this is the first attempt in the literature. Second, we demonstrate the advantages and relative gains of a tree-based algorithm over linear regression. Third, we train a highly predictive model with original data with a large number of features and a universe of items specifically constructed to measure behavioural change in response to COVID-19. This allows us to minimise the bias from both measurement error and omitted variables, which is a common limitation in related studies that only focus on few measures and a limited set of features. Fourth, our approach allows for the presence of interaction effects among key features. And finally, we identify key policy relevant individual and social predictors of self-protecting behaviours and document their heterogeneity by country.

The remainder of the chapter is structured as follows. In Section 3.2 we describe the data, define the variables of interest and present summary statistics. The method employed in the study is in Section 3.3. We present our results and discuss their implications in Section 3.4. Section 3.5 concludes.

3.2 Data

Our study primarily relies on the COME-HERE (COVID-19, MEntal HEalth, REsilience, and Selfregulation) survey as a data source. COME-HERE is a longitudinal survey conducted by the University of Luxembourg to investigate the psychological and socioeconomic effects of the COVID-19 pandemic and related social distancing measures across Europe. The survey is administered by Qualtrics on nationally representative samples of individuals in France, Germany, Italy, Spain, and Sweden. This study was approved by the Ethics Review Panel of the University of Luxembourg. All research was carried out in accordance with relevant guidelines, and informed consent was obtained from all participants. Respondents were asked to complete an online questionnaire that takes approximately 20 minutes. The survey collects information at the individual and household levels. The first wave of the survey took place in late April and early May 2020, and over 8,000 people participated. The same pool of individuals was re-contacted six additional times in June, August, and December 2020 and March, June, and October 2021. For more information see Vögele et al. (2020).

The present results are based on the first wave of the survey with over 8,000 participants who answered a range of questions on their sociodemographic characteristics, socioeconomic variables, pre-existing health conditions and risk behaviours, mental health, resilience, social support, and trust in major institutions. This dataset is representative of the population countries in terms of age, gender and region of residence. In the next paragraphs, we outline in greater detail how we define the variables and features used in the prediction exercises.

Our outcome variable is a composite measure of self-protecting behaviours. The measure is constructed from a range of questions from the specially designed Coronavirus Behaviour Scale (*CBS*). The *CBS* is a 14-item self-report measure of behaviour change due to the Coronavirus pandemic. It contains two subscales, with nine items assessing reasonable behaviours (e.g., shaking hands less) and five items assessing unreasonable behaviours (e.g., buying more toilet paper than usual). Responses are given on a 5-point Likert scale ranging from 1 strongly disagree to 5 strongly agree. We focus our analysis on the reasonable behaviour subscale. This subscale is based on responses to the following questions: "Because of coronavirus, I am planning to or have already 1) cleaned and disinfected surfaces in my home more often, 2) bought medical masks, 3) stayed home when I feel ill, 4) started avoiding crowded spaces, 5) bought disinfectant, 6) started washing my hands more, and 7) shaken hands with people less". We compute a composite index from the responses to the seven items of the reasonable CBS subscale to define self-protecting behaviours of individual i ($CBS_i = \frac{1}{K}\sum_{k=1}^{K} \text{item}_k$).

We consider a host of sociodemographic, economic, and social contextual variables that predict self-protecting behaviours. These predictors include age, gender, marital status, education, employment status, household income and housing conditions, health status, behavioural risk factors (such as alcohol and cigarette consumption), confidence in institutions, and incidences of COVID-19 and responsive policy measures. In Appendix 3.A, we discuss in detail how each predictor variable is constructed and the necessary features pre-processing.

Table 3.1 presents the summary statistics of key characteristics of individuals in the pooled and country level samples. In the pooled sample, 8063 respondents were included, and the sample mainly consisted of working-age individuals (mean age of 47.5). Approximately 55.4% of respondents reported ties to the labour market (44% employed full-time and 11.5% part-time). Twenty percent of the sample are unemployed, and about 23.7% of the participants are retired. Nearly 58% are married and cohabiting, 8.7% are married but living apart, and 33.4% of the respondents are single. Close to 9% of the participants reported that they live in a house without any open-air access

(such as garden, terrace, or balcony). Nearly 52% of respondents report at least one pre-existing health condition. In terms of household income, 21% of the respondents are in relative poverty.

Table. 3.1 *Descriptive statistics of selected variables*

	Pooled	France	Germany	Italy	Spain	Sweden
Gender (Female = 1)	51.7	52.5	52	52.3	51.1	50.2
Age category:						
$[18, 20)$	2.5	3.6	2.6	1.9	1.3	3.3
$[20, 30)$	16.3	14.9	13.7	23.2	14.2	15.3
$[30, 40)$	17.4	17.8	16.6	18.2	17.9	16.3
$[40, 50)$	17.9	17.8	16.2	18.2	21.4	15.5
$[50, 60)$	16	15.5	17.8	13.6	15.8	17.5
$[60, 70)$	19.3	20.5	22.4	16.5	20.8	14.9
70 and above	10.6	9.8	10.8	8.3	8.6	17.3
Family status:						
Married & cohabiting	57.9	61	57.2	54.4	61.7	54
Married but living apart	8.7	3.9	9.6	13.5	8.9	7.6
Single	33.4	35.2	33.3	32.1	29.5	38.4
Employment status:						
Full-time	43.9	47.7	44.8	40.1	46.9	38.7
Part-time	11.5	9.3	13.8	13.4	9.6	11.3
Unemployed	20.9	17	16.6	28.4	20.8	21.7
Retired	23.7	26.1	24.7	18.1	22.6	28.3
Education:[a]						
Low	7.8	6.7	4	9.3	10.1	9.4
Middle	37.5	38.6	28.3	51.6	22.7	49.9
Higher	55	54.7	67.7	39.1	67.2	40.7
Relative poverty	21.3	18.6	18.6	26.9	26.8	12.9
Housing feature:						
Balcony	53.9	28.7	56.5	74.7	54.2	55.8
Park	4.4	4.9	3.9	4.4	3.7	5.6
Garden	47.1	61.4	54	45.8	24.3	50.9
Terrace	44.3	49.3	45.3	38.1	53	31.9
No home feature	8.8	11.4	7.9	3.7	13.3	7.2
No pre-existing illness	48.2	54.9	44.7	50.3	46.5	43.3
Confidence in						
Government	4.65 (1.82)	3.94 (1.79)	5.09 (1.49)	4.92 (1.74)	4.05 (2.00)	4.90 (1.88)
Health services	5.39 (1.41)	5.14 (1.43)	5.50 (1.29)	5.52 (1.36)	5.45 (1.43)	5.32 (1.57)
Essential services	5.66 (1.30)	5.35 (1.28)	5.78 (1.25)	5.81 (1.30)	5.88 (1.23)	5.44 (1.40)
Knowledge about COVID-19	5.24 (1.22)	4.90 (1.27)	5.25 (1.15)	5.58 (1.10)	5.38 (1.19)	5.04 (1.27)
CBS	3.86 (0.80)	4.01 (0.69)	3.54 (0.79)	4.19 (0.68)	4.04 (0.76)	3.40 (0.77)
N	8063	1706	1720	1710	1711	1216

Notes: This table presents summary statistics of respondents' key characteristics and target variable in each sample used in the final prediction. The first column shows the summary statistics of selected variables of the pooled sample. The other columns list summary statistics of selected variables in each country. In parenthesis are the standard deviations. The numbers in the tables show the percentages. The figures followed by parenthesis are the mean and standard deviation of the variables, respectively.

[a] For the sake of presentation, the eight levels of education are recoded into three-level ISCED aggregation.

3.3 Method

Many features contribute to self-protecting behaviours against COVID-19. Our objective is to investigate this complex relationship with better accuracy. We consider

vectors of features capturing socioeconomic status (such as income and home features), pre-existing health conditions, health and behavioural risk factors, socio-demographic factors, severity of the disease and government policy responses, trust in major institutions, and knowledge about the pandemic.

Modelling this relationship with a parametric approach requires the researcher to make certain assumptions about the data-generating process, imposing structure on the data, which may eventually lead to functional misspecification. With supervised ML algorithm we do not need to assume any structure about the data-generating process but directly learn a function that maps the input features to the target variable. The learning process is controlled with hyperparameters that regulate some aspect of the learning algorithm. These hyperparameters are commonly optimised via a grid search over the hyperparameter space.

Although highly predictive, most complex ML algorithms often lack model-specific interpretability unlike parametric linear models. To circumvent this limitation, we propose an *explainable* ML (ExpML) that provides an opportunity to improve prediction accuracy over standard OLS using a 'black-box' supervised ML algorithm while still revealing interesting insights into a complex outcome. To do so, we build a Random Forests (RF) model and complement it with model-agnostic interpretative tools.

We favour Random Forests (RF) over other ML algorithms because they are easily scalable to accommodate a large dataset with higher dimensionality without losing statistical efficiency. They can naturally capture complex interactions between features and their predictions are robust to outliers due to repeated sampling. RF is a collection of many *de-correlated* trees. Trees are grown on a bootstrapped sample with a random subset of feature vectors. Then the final predictions are produced as a mean value of predictions from each tree (in a classification problem, the final predictions are yielded based on a majority vote).[1] In this study, all predictions are done with a popular library for the Python programming language called scikit-learn (version 0.22.2).[2]

We build a RF model consisting of 500 trees. To control over-fitting, we apply a grid search with five-fold cross-validation and identify the optimal hyperparameters. We optimise two important hyperparameters that will ensure randomness in RF: the

[1]The steps in a typical RF algorithm are as follows: (i) Draw a bootstrap sample from the training data and randomly select k variables from p variables, where $k \ll p$. (ii) Select the best split among the k variables. The maximum number of k is a hyperparameter. (iii) Split the node into two daughter nodes. (iv) Repeat step i to iii until the terminal nodes have been reached— until no farther splits are possible. (v) Repeat the above steps for T number of times to grow a forest of T trees.

[2]Scikit-learn (Pedregosa et al., 2011) is an open-source project containing several state-of-the-art ML algorithms and their extensive documentation.

maximum depths of trees ('max depth') and maximum number of features used in each split ('max features'). The deeper the tree, the more splits it has, and it captures more information about the data. Similarly, by choosing a reduced number of features we can increase the stability of the tree and reduce variance and over-fitting.

The dataset is randomly divided into training (80%) and testing (20%) sets. For each combination of hyperparameters, we undertake the following three steps: (i) randomly partition the training set into five subsamples, (ii) train the models on the four subsamples and generate predictions for the fifth, and (iii) repeat this process five times so that each subsample is used only once to generate the prediction. The optimal hyperparameters are updated based on the average of these five prediction results. The tuning step ends when we find an optimal hyperparameter that produces minimum average prediction errors.[3]

We assess the predictive accuracy of each Random Forests and a baseline OLS model because having a reliable estimate of predictive performance (e.g., a significant relative performance gain from using RF vis-a-vis the baseline model) is an essential requirement for the socioeconomic interpretation of the results we are after in the study. To quantify the extent to which the predicted value for a given respondent is close to the actual value of that individual, we use the most common metrics in regression settings: Mean Absolute Errors (MAE) and Root Mean Squared Errors (RMSE). We use MAE to show how the model fares when prediction errors are linearly weighted. The best model will be the one that has a smaller value in both MAE and RMSE (ideally close to zero), meaning that it produces predictions that are very close to the true responses. In our data, both metrics produce similar conclusions.

We find that the RF model dominates OLS in all prediction tasks in both the pooled and per-country datasets, can be seen Table 3.2. The superior prediction performance of RF shows that we could capture signals relevant to self-protecting behaviour, which the linear model failed to capture (as measured by the percentage reduction of the prediction error).

Finally, we use the visualisation tool SHapley Additive exPlanations (SHAP) proposed by Lundberg and Lee (2017) to explain the contribution of each feature to the prediction of self-protecting behaviours using Shapley values. SHAP is based on a solution concept in a cooperative game setup that aims to 'fairly' allocate the gains among players as suggested in the seminal work of Shapley (1953). SHAP has the advantage of consistency and provides both local and global interpretability (see Guidotti et al.,

[3]The explored hyperparameters space, the cross-validation results, and the final hyperparameter combination used in the final model are available on request from the authors.

2018; Molnar, 2020, for a comprehensive review of black-box ML model interpretation techniques).

Table. 3.2 *Accuracy of RF (with 500 trees) with model specific hyperparameters configuration compared with baseline OLS model*

Prediction performance (MAE and RMSE) on training and testing sample [1]								
Model	Metrics	Sample	Pooled and per country datasets					
			Pooled	France	Germany	Italy	Spain	Sweden
RF[2]	MAE [a]	Trainset	0.454	0.473	0.469	0.412	0.468	0.468
		Testset	0.461	0.505	0.608	0.494	0.581	0.567
	RMSE[b]	Trainset	0.588	0.617	0.603	0.538	0.614	0.603
		Testset	0.584	0.652	0.785	0.621	0.774	0.720
OLS	MAE	Trainset	0.547	0.487	0.571	0.475	0.538	0.534
		Testset	0.562	0.542	0.634	0.497	0.604	0.594
	RMSE	Trainset	0.711	0.627	0.728	0.618	0.696	0.697
		Testset	0.709	0.687	0.834	0.642	0.794	0.775
Relative improvement over ordinary least squares								
	MAE	Trainset	-17.00%	-2.90%	-17.86%	-13.26%	-13.01%	-12.36%
		Testset	-17.97%	-6.83%	-4.10%	-0.60%	-3.81%	-4.55%
	RMSE	Trainset	-17.30%	-1.60%	-17.17%	-12.94%	-11.78%	-13.49%
		Testset	-17.63%	-5.09%	-5.88%	-3.27%	-2.52%	-7.10%

[a] Mean Absolute Error
[b] Root Mean Squared Error
The target variable is reasonable COVID-19 behavioural Scale.
[1] The data has been sliced into training (80%) and test set (20%)
[2] RF is optimized using the *maximum depth* of the trees and *the number of features* sampled to grow a tree. These hyperparameters have been obtained with a 5-fold cross-validated grid search.

3.4 Results

3.4.1 Identifying the Top 30 Predictors

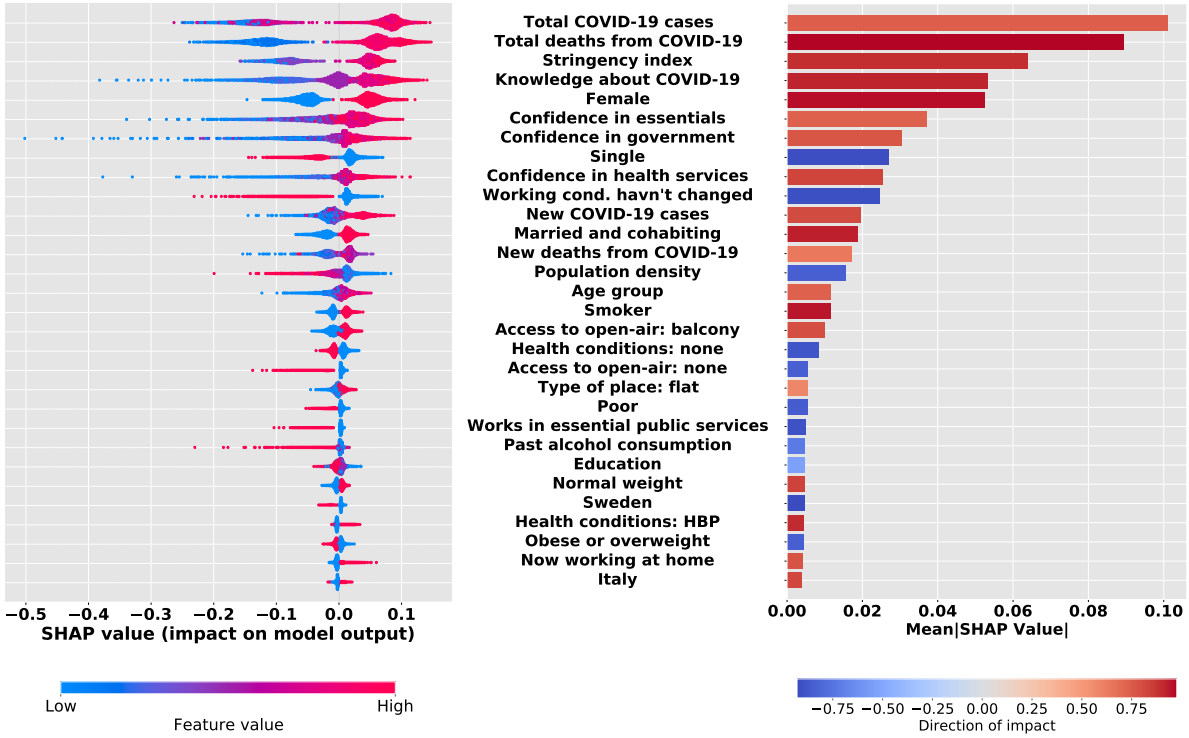
We identify the top 30 features in predicting self-protecting behaviours. Figure 3.1 (a) presents a SHAP summary plot that succinctly displays the importance of these features, the magnitude of impact, and direction of a feature association with self-protecting behaviours (see Appendix 3.B, for detailed results showing all features used in the prediction exercise)[4]. On the y-axis, the top 30 predictors of self-protecting

[4]We also perform a robustness check of feature ranking using an alternative method to SHAP feature importance. See Figure 3.7 in Appendix 3.D for feature ordering using permutation feature importance. There is a big difference between these two alternative methods: Permutation feature importance is based on the decrease in model performance whereas SHAP is based on the magnitude of feature attributions.

behaviours are positioned in descending order of average contribution to the prediction (i.e., the global contribution measured in mean absolute SHAP value). On the x-axis the SHAP values for each observation are presented – negative SHAP values are interpreted as reduced self-protecting behaviour, while positive SHAP values are interpreted as increased level of self-protecting behaviours.

Each dot represents an individual respondent; hence, the number of dots against each feature reflects the sample size of the training set. The dot's position along the x-axis is the feature's impact on the model's prediction for that respondent. When multiple dots arrive at the same coordinate in the plot, they pile up to show the density of effect sizes. The long-left tails in the summary plot indicate that the predictors are highly predictive for some respondents but not others, i.e., predictors with minor global importance can still be very important for specific respondents.

Figure. 3.1 *Top 30 predictors of self-protecting behaviours*



(a) *Summary plot* **(b)** *Global impacts (magnitude and direction) of top features*

Notes: Panel (a) is the SHAP summary plot for the Random Forests trained on the pooled data set of five European countries to predict self-protecting behaviours responses against COVID-19. The plot displays the top 30 features on prediction (the top on the y-axis is the most important) and the distribution of the impacts of each predictor on the model prediction, which includes a set of distributions where each dot corresponds to an individual. When multiple dots arrive at the same coordinate in the plot, they pile up to show density. The colours correspond to the feature values: red for larger values and blue for smaller ones. A negative SHAP value (extending to the left) shows reduced self-protecting behaviour, while a positive (extending to the right) shows an increased self-protecting behaviour. Panel (b) displays three-fold information: (i) The direction of association captured by the correlation between the feature and SHAP values (red for positive and blue for negative); (ii) strength of the direction of association shown by the darkness of each colour gradient; (iii) The magnitude of feature's marginal impact measured as the average of absolute SHAP values.

In Figure 3.1 (b), we summarise our key findings for an easier global explanation of the impact of the features on the model and their association with self-protecting behaviours. The horizontal length of each bar shows the magnitude of impact on the model. The correlations of each feature with self-protecting behaviours are shown by the colour code of each bar. We can see that features such as total number of deaths from COVID-19, stringency index, knowledge about COVID-19, being female, being married and co-habiting, having a pre-existing high blood pressure (HBP) and being smoker (including those who currently smoke or ex-smokers) have strong positive associations (correlation coefficient ≥ 0.9) with self-protecting behaviours, whereas being single, not experiencing COVID-19 induced work-related changes, not having pre-existing health conditions, not smoking, residing in Sweden, and working in essential public services all

show strong negative associations (correlation coefficient ≤ -0.9) with self-protecting behaviours.

In the next subsections, we examine how each of the top features contributes to the model's output. Since self-protecting behaviours are likely to depend on locally relevant perceptions of threats, susceptibility, and benefits in engaging in protective behaviour, we structure our results according to factors related to epidemiological, socioeconomic, and specific health conditions. As most of the predictors of self-protecting behaviours are categorical, we create box plots to show the difference in the average marginal contribution of each category in a more robust way. The y-axis of the box plots shows the SHAP value of the variable, and on the x-axis are the values that the variable takes.

3.4.2 Epidemiological Factors

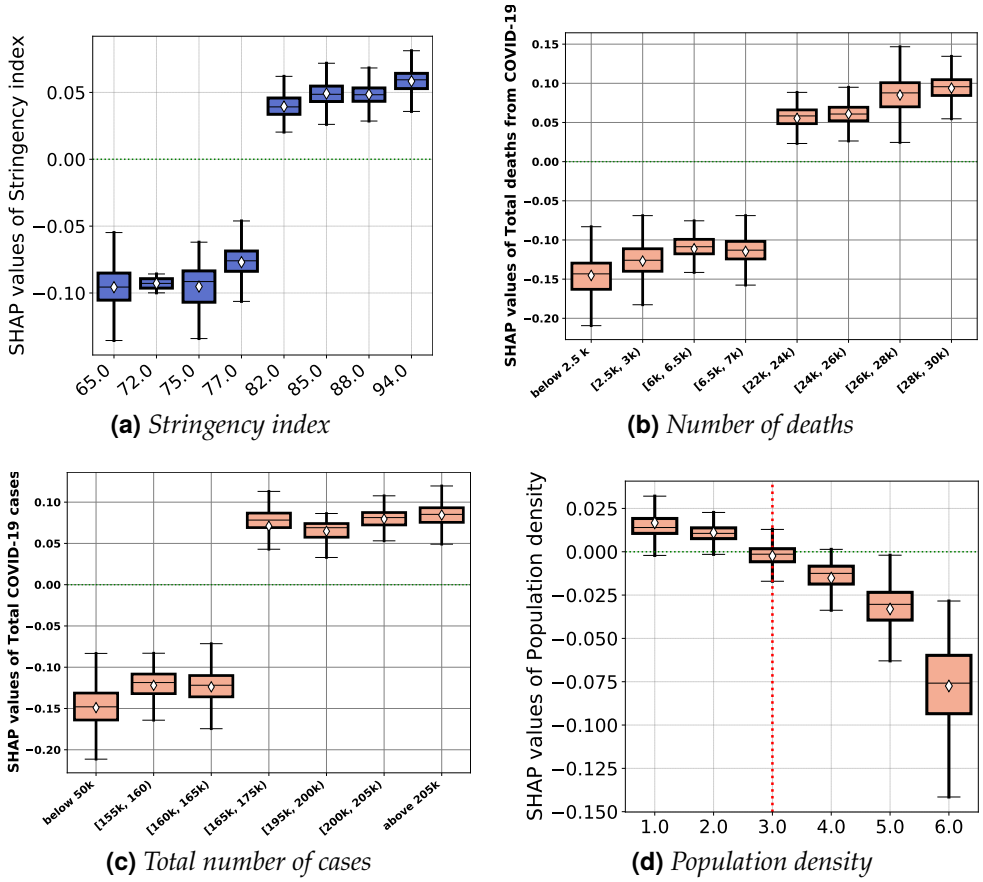
Epidemiological models of the spread of infections postulate that people start reacting against contracting a disease with self-protective measures whenever they are informed about the disease and when the burden of the disease is in a recognisable stage (Kassa and Ouhinou, 2011). The variables typically used in epidemiological models include population density, stringency of lockdown measures, number of new infections and fatalities, confidence in institutions, and knowledge about COVID-19. Our results confirm that these indicators are all important predictors of self-protecting behaviour.

Figure 3.2 depicts the effects of stringency policy response, local infection rate, confidence in institutions, and level of knowledge about COVID-19 on self-protecting behaviours. Our results show that total number of cases and deaths linked to COVID-19, and the OxCGRT stringency index are ranked first, second and third most important predictors of self-protecting behaviours respectively. A striking positive non-linear pattern exists between the stringency index and self-protecting behaviours. From panel (a) of Figure 2, we see that a stringency index of 80 (with the index ranging between 0 and 100, 100 = strictest response) is an important threshold over which any increase in strictness of policy response triggers a level of self-protecting behaviours higher than the mean prediction (positive SHAP values).

Panels (c) and (d) of Figure 3.2 contain the partial effects of infection rate on individuals' behaviour. Both measures of infection rate show a step like association with self-protecting behaviours. Behavioural responses remain quite unresponsive with respect to national infection rates for a while and jump higher following the total number of deaths from COVID-19 spiking up to above 20,000 from somewhere below 7,000 deaths. From panel (b), we observe a somewhat non-linear negative association be-

tween the degree of population density and the level of self-protecting behaviours. Overall, living in a densely populated area/city/town/village is associated with lower self-protecting behaviours.

Figure. 3.2 *Partial effects of stringency policy response and local infection rate*



(a) *Stringency index* **(b)** *Number of deaths*

(c) *Total number of cases* **(d)** *Population density*

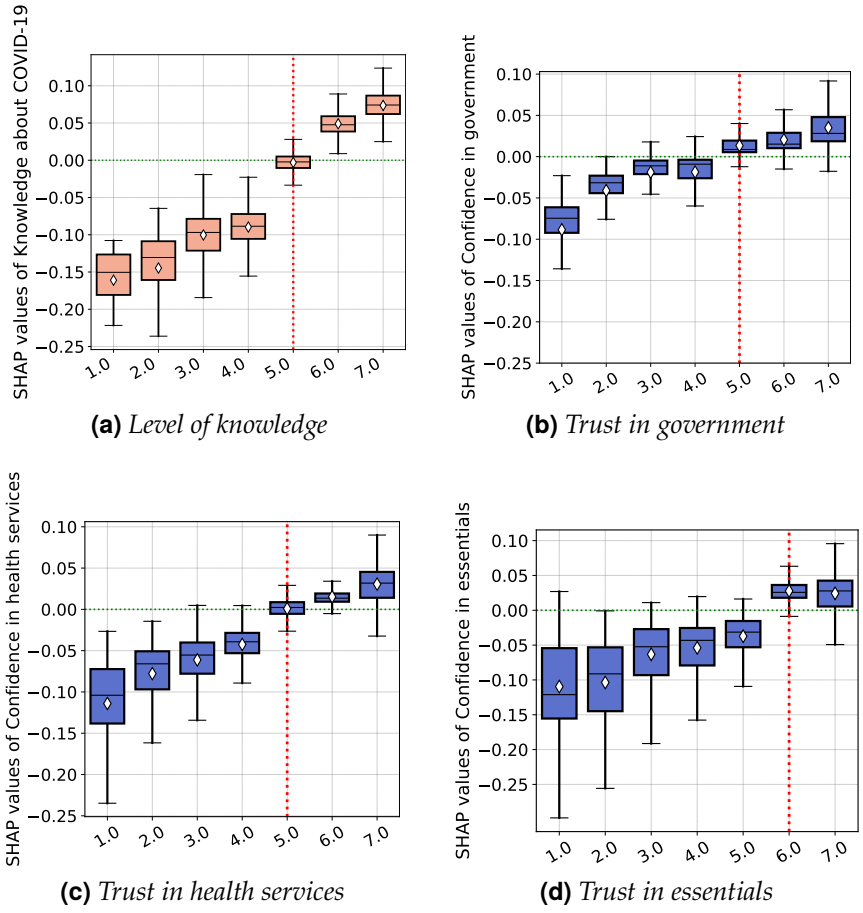
Notes: This figure displays SHAP dependence boxplots of the stringency index. The diamond symbol in the boxes denotes the average of the SHAP value distribution per each value of the stringency index during the first wave of the pandemic. In panel (d), the labels in the x-axis correspond to the number of people in the respondent's residential area, 1 "Isolated dwelling", 2 "less than 2,000", 3 "between 10,000 and 2,000", 4 "between 50,000 and 10,000", 5 "between 100,000 and 50,000", and 6 "more than 100,000". The values in the x-axis of panels (a) and (c) are the number of cases and deaths summarised in a few bins.

Confidence in the essentials, government, and health services are the top 6th, 7th and 9th most important features in our model, respectively. We find a consistent positive association between confidence in the essentials, government, health services, and the SHAP of each feature with correlation coefficients of 0.74, 0.77, and 0.84, respectively (Figure 3.3, panels (b) - (d)). We also identify an important level of confidence threshold (shown by the red dotted vertical line in the panels) over which the three features about confidence in major institutions are associated with positive SHAP values. We find higher threshold for confidence in essentials (degree of confidence = 6), compared to

the other two features (degree of confidence = 5).

Panel (a) of Figure 3.3, reveals a strong positive relationship between self-assessed level of respondent's knowledge about COVID-19 and level of self-protecting behaviours. It is the 4th most important predictor of self-protecting behaviours. We discover an important threshold (level of knowledge = 5) over which the level of knowledge is associated with positive SHAP values.

Figure. 3.3 *Partial effect of confidence in institutions and level of knowledge about COVID-19*



(a) *Level of knowledge* **(b)** *Trust in government*

(c) *Trust in health services* **(d)** *Trust in essentials*

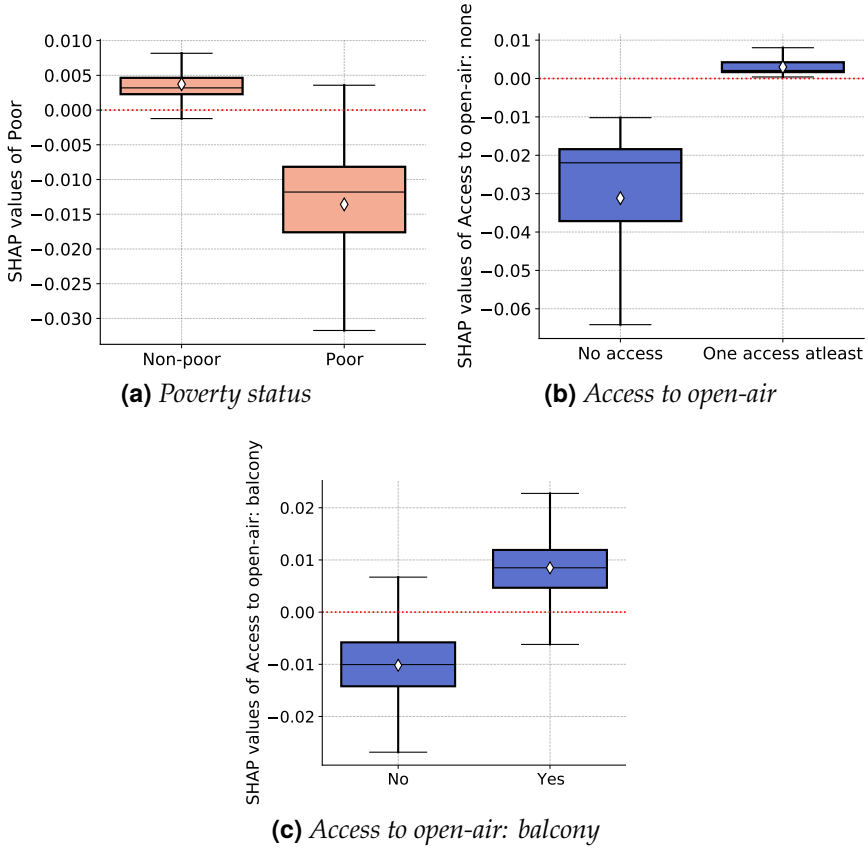
Notes: Each panel (a) - (d) displays SHAP dependence boxplots of features related to trust in institutions and knowledge about COVID-19. The diamond symbol in the boxes denotes the average of SHAP value distribution per each category.

Interestingly, all features relating to confidence in authorities and knowledge about COVID-19 have long left tails shown in the summary plot, suggesting that the lower degree of trust and awareness about the pandemic is exceptionally detrimental to self-protecting behaviours than a higher level of confidence is in enhancing protective behaviour.

3.4.3 Socioeconomic Factors

Our results identify a strong socioeconomic gradient in self-protecting behaviours. Panels (a) - (c) of Figure 3.4 present partial effects of key socioeconomic features on self-protecting behaviours. Results depicted in panel (a), shows a strong negative association between the poverty status and the level of self-protecting behaviours. On average, the non-poor have a self-protecting behaviours level that is higher than the mean prediction of the model $(E(\hat{f}(x)) = 3.866)$ while the poor exhibit a level of self-protecting behaviours smaller than the average prediction of the model. Hence, the economically better-off individuals (those non in poverty, that is with an equivalent household income above 60 percent of the country's median) tend to exhibit a higher level of engagement in protective behaviours against COVID-19. This suggests that adherence to some sanitary recommendations (e.g., physical distancing) is a costly option to relatively poor households. Hence, authorities should devise a buffering mechanism that could enable the financially vulnerable individuals to cope with the economic burden of pandemic health behaviours.

Figure. 3.4 *Partial effects of income and housing features*



(a) *Poverty status*

(b) *Access to open-air*

(c) *Access to open-air: balcony*

Notes: Each panel (a) - (c) display SHAP dependence boxplots of income and housing features. The diamond symbol in the boxes denotes the average of SHAP value distribution per each category.

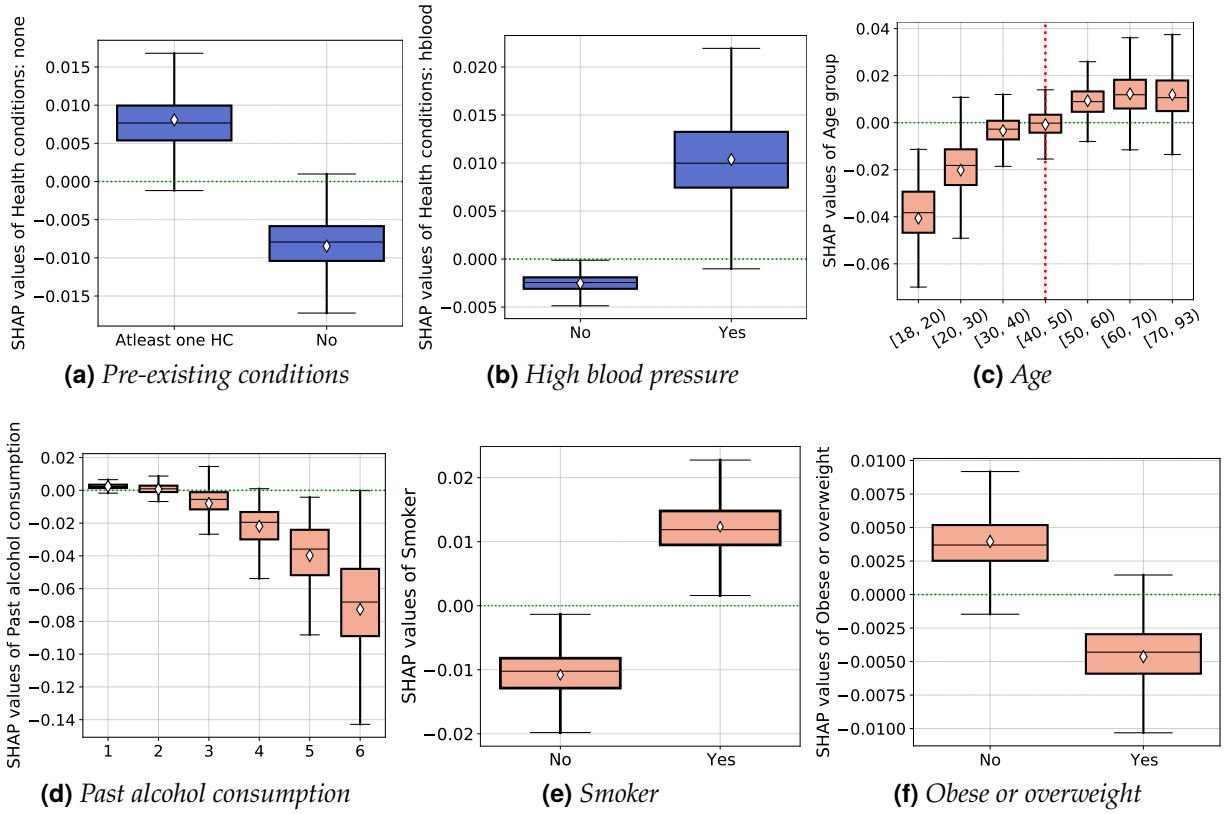
We also find that access to open-air at home are among the strong predictor of self-protecting behaviours and they positively impact respondents' protective behaviours. We assume that better home features are indicators of higher wealth. RF prediction shows that respondents that live in homes without access to open-air have, on average, self-protecting behaviours that are well below the mean prediction. Having access to open-air with at least one home feature guarantees the level of self-protecting behaviours that is above the mean prediction. Of the different home features considered in our model, having a balcony appears the most important in terms of relative contribution to the model while access to garden, terrace, and park do not feature in the top 30. The reason that balcony is among the main factor determining people's behaviours is that during COVID-19, it is through their balconies that people enjoy a sense of being outdoors while confined at home. Its functionality as a result has been diversified to be as a work corner, a place to exercise, a place of contemplation, and fresh air without the anxiety of breaking the lockdown rules.

Housing features, such as access to open-air and living in accommodation that are not overcrowded, enable individuals to comply with protective measures better. In addition, lack of access to outdoor spaces affects dwellers' mental health during the pandemic influencing decisions to engage in protective behaviours (Yang and Xiang, 2021). Hence, precautions may seem more relevant in dense localities and poor living arrangements. Local authorities should prioritise devising temporary open spaces accessible to residents during pandemics. The long-run housing policy should also prioritise open-air access as an integral part of housing design.

3.4.4 Specific Health Conditions

The role of specific health conditions has been emphasised regarding the severity of COVID-19 once infected (Brown and Ravallion, 2020). We also find that having at least one pre-existing health condition is strongly associated with higher self-protecting behaviour, as shown in panel (a) of Figure 3.5. Moreover, we assess how each pre-existing condition fares in terms of contributing to self-protecting behaviours prediction. As evident from panel (b), respondents with the pre-existing condition of high blood pressure tend to engage in a higher level of self-protecting behaviours than the others.

Figure. 3.5 *Partial effects of pre-existing health conditions and behavioural risk factors*



(a) *Pre-existing conditions* **(b)** *High blood pressure* **(c)** *Age*

(d) *Past alcohol consumption* **(e)** *Smoker* **(f)** *Obese or overweight*

Notes: Notes: Each panel (a) - (f) display SHAP dependence boxplots of pre-existing health conditions and behavioural risk factors. The diamond symbol in the boxes denotes the average of SHAP value distribution per each category. In figure (d), the labels in the x-axis correspond to the number alcohol consumption in number of glasses in an average week: 1" <5", 2"[5,10)", 3"[10,15)", 4"[15,20)", 5"[20,25)", 6" ≥ 25".

Panels (c)-(e) of Figure 3.5 display the impact of health and behavioural risk factors (described in Section 2) on the overall model prediction. We find evidence that past alcohol consumption and obesity are negatively related to self-protecting behaviours. There is strong evidence that alcohol consumption is negatively associated with self-protecting behaviours. Obesity is associated with increased risk of severe illness from COVID-19 (Dietz and Santos-Burgoa, 2020; Mendoza-Jiménez et al., 2021). At the same time, obesity is strongly related to low socioeconomic status (McLaren, 2007), that may explain the negative relationship between obesity and self-protecting behaviours.

Although smoking is an adverse health behaviour, interestingly, our results show that smokers (including those who currently smoke or ex-smokers) tend to manifest a higher level of self-protecting behaviours than non-smokers. Hence, this increased risk of COVID-19 for smokers might motivate them to exhibit higher compliance. Consistent

with Dai et al. (2020), which find smokers tend to suffer severe outcome of COVID-19 while alcohol consumption was not linked to severe complication from the virus, we find that alcohol consumption does not motivate greater self protection. Similarly, Reddy et al. (2021) argue that smokers tend to have more severe symptoms of COVID-19 compared to non-smokers.

At the beginning of the pandemic older people have been found to be disproportionately more likely to have severe symptoms from COVID-19 leading to hospitalisation and death. We find that the nature of the relationship between age and self-protecting behaviours is somewhat non-linear, i.e., for age groups older than 50 years of age, we do not find a significant difference in the average level of self-protecting behaviours. Moreover, 40 years of age appears to be an essential threshold over which the number of years is associated with positive SHAP values.

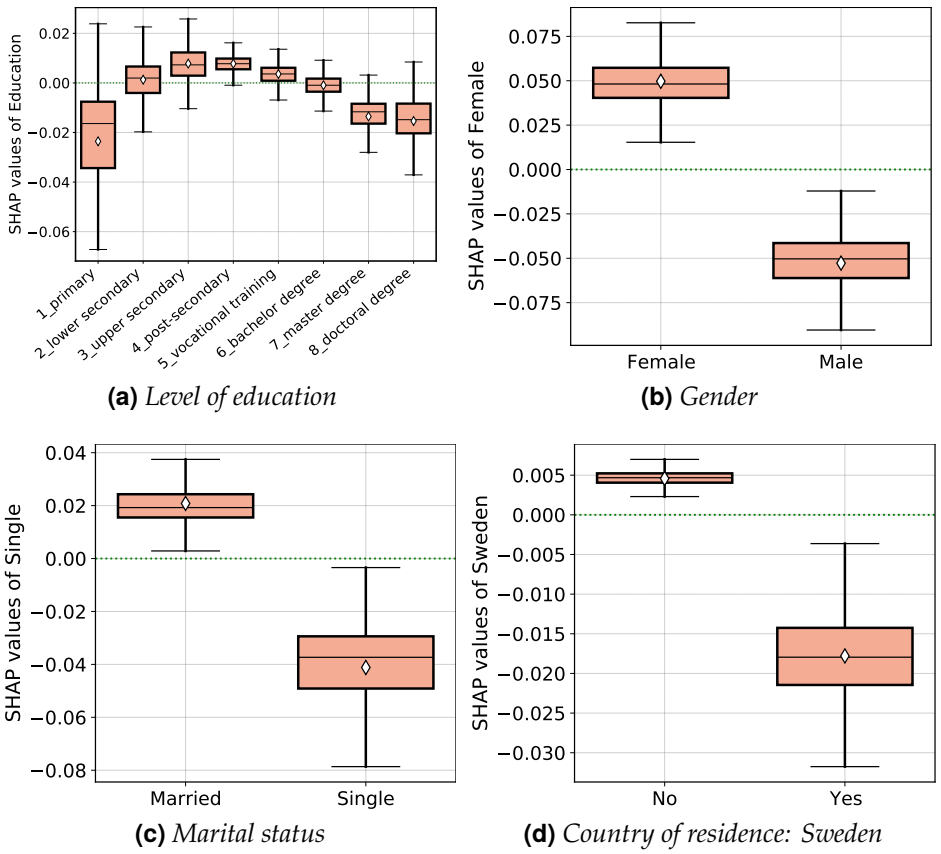
3.4.5 Sociodemographic Factors

The next set of features we consider are the main socio-demographic indicators, such as gender, family status, country of residence and level of education of respondents. In panel (a) of Figure 3.6, we examine the partial effect of individuals' level of education and their corresponding self-protecting behaviours. We do not find a monotonically increasing protective behaviour with educational attainment. Our results show a non-linear association between education and self-protecting behaviours - the lowest level of self-protection is associated with the lowest education group. It peaked with the middle education group and then slightly decreased for the highest education group. The increasing part (in the partial effect of education) re-affirms the strong association between socioeconomic status and self-protective behaviours. The slight drop in self-protection from middle to the highest education level could be explained by differential exposure to the pandemic for the high socioeconomic group. One such important difference is remote work. For instance, in our data, 36% of individuals with higher education levels responded that they were working at home during the first wave of the pandemic. This remote working arrangement is disproportionately higher for the more educated groups than the middle (16.5%) and lower (6%) education groups. Thus, it is plausible to argue that individuals with higher educational attainments have a low risk of getting infected with COVID-19 in the first place than those with low and middle educational attainment (Niedzwiedz et al., 2020).

Our model identifies gender as the 5th top predictor of self-protecting behaviours. As evident from panel (b) of Figure 3.6, there is substantial gender differences in health-

protective behaviours. Females tend to exhibit a higher level of self-protective behaviour. Gender differences in fear and risk perception (Alsharawy et al., 2021) can potentially explain this pattern. The intra-household bargaining theory can also explain the residual effect – where women engage more in household caretaking roles while men continue to work in person (Alon et al., 2020; Biroli et al., 2021; Collins et al., 2021; Xue and McMunn, 2021)

Figure. 3.6 *Partial effects of socio-demographic factors*



(a) *Level of education* **(b)** *Gender*

(c) *Marital status* **(d)** *Country of residence: Sweden*

Notes: Each panel (a) - (d) display SHAP dependence boxplots of socio-demographic features. The diamond symbol in the boxes denotes the average of SHAP value distribution per each category.

From panel (c) of Figure 3.6, we see the partial effects of family status of individuals in determining the level of self-protecting behaviours in our model. Single respondents (ranked 8th) react less to the pandemic. This study also isolates a positive, partial impact of being married and cohabiting vs. married but living apart. Stickley et al. (2021) documented that loneliness is linked with lower engagement in COVID-19 protective behaviours, suggesting that individuals who live alone could find it more challenging to restrain themselves from meeting others as they lack companionship at home.

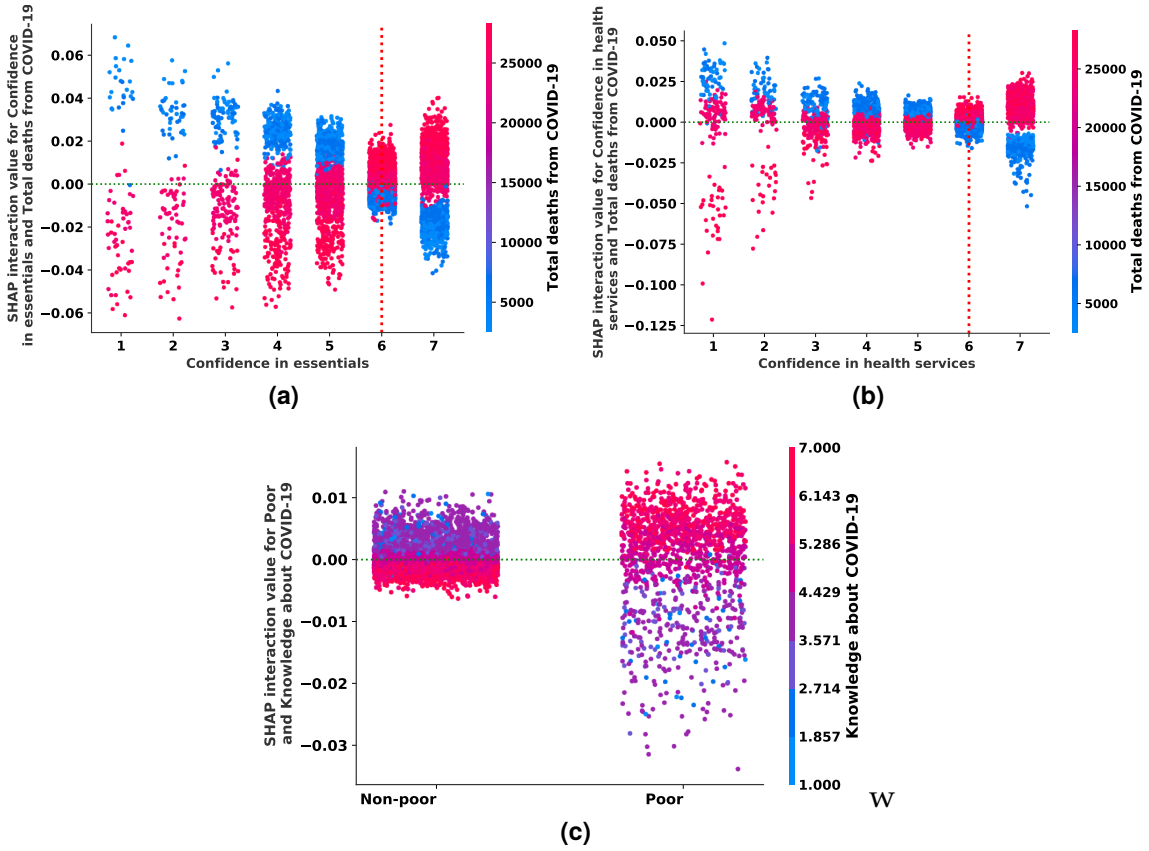
In panel (d) of Figure 3.6, we present the partial effect of country of residence on the level of self-protecting behaviours. The results show that compared to respondents from Italy, Spain, France, and Germany, respondents from Sweden tend to exhibit lower level of self-protecting behaviours during the first wave of the pandemic. Sweden was the only country never in lockdown.

3.4.6 Measuring Interaction Effects

In the preceding sections we highlighted the top-ranking features that are predictive of self-protecting behaviours. In some cases, it is plausible to assume that the relationship between the target variable and a feature depends on the value of another feature. Such contextual dependence between features that jointly impact predictions are known in the ML literature as feature interactions. Our RF model automatically accounts for these interactions, and we now systematically report interaction effects that have some policy relevance.

The interaction between confidence in essentials with the number of deaths from the COVID-19 shows that the relationship between self-protecting behaviours and confidence in essentials alerted by the national incidence rate of the pandemic, observed in Figure 3.7 (a). Those individuals who are very confident (level of confidence on the Likert scale six and above) that essential services will be maintained during the pandemic exhibit a higher level of self-protecting behaviours proportional to the number of deaths. On the other hand, those respondents with lower confidence in essentials respond asymmetrically to the pandemic's national incidence rate. We find a similar impact on confidence in health services as seen in Figure 3.7 (b).

Figure. 3.7 *Feature interaction effects*



Notes: Each panel (a) - (c) displays SHAP feature dependence plots of the RF model with the largest interaction effect. Artificial jitter (0.5) was added along the x-axis to better show the overlapping distribution of the points.

Next, we investigate the interaction effect between knowledge about COVID-19 and the relative poverty status of individuals: in Figure 3.7 (c), we see a more pronounced positive effect of knowledge about the pandemic and self-protective behaviours among the poor group than the non-poor, meaning that among the poor those who show a higher level of self-protection are the ones who have better awareness about the pandemic. However, in the non-poor groups, the interaction effect is slim, and self-protection behaviours are not necessarily tied to the level of knowledge. This result is in line with the partial effect of education examined above. It is safe to assume that a high level of education and better knowledge about COVID-19 are highly correlated.

These interaction effects imply that enhancing an individual's trust in institutions could enable people to show behavioural responses with respect to the national incidence rate. Moreover, awareness creation about the pandemic that targets the low socioeconomic group could result in a better behavioural response.

3.4.7 Heterogeneity by Country

Detailed result of feature importance and direction of association in each of the five countries separately is presented in Appendix 3.C. Confidence in public institutions, self-assessed knowledge about COVID-19, gender, age, and marital status are the feature among the top ranked predictors of self-protecting behaviours in each country. Consistent with the results from the pooled sample, COVID-19 infection and death counts also appear among the top 30 predictors in all countries.

3.5 Conclusion

In response to the COVID-19 pandemic, several national governments have applied lockdown restrictions and requested their citizens to undertake some behavioural changes to reduce the infection rate. Even though it is apparent that human behaviour largely influences the spread of illness, it is unclear what factors most strongly correlate with protective behaviour. We examine the most relevant predictors of individual self-protective behavioural responses during the earliest wave of the COVID-19 pandemic with a data-driven approach of ExpML on real-time survey data from five European countries. Key findings of this paper shed light on essential policy variables that help to slow the current pandemic and guide the design of future policies.

We show that COVID-19 disease progression and the responsive government policies induce an increase in self-protecting behaviours provided that both the number of COVID-19 infection/death rate and the strictness of policies are sufficiently high. The government policy setting and the local infection rates play an important role in individuals' responses to the health crisis. These factors guide how individuals adopt the protective guidelines. Higher local infection rates increase people's perceived risk of getting infected, inducing responses to protective behaviours. This was further reflected in the macro features relating to the country of residence. Compared to respondents in Sweden, respondents from Italy, Spain, France, and Germany tend to exhibit a higher level of protecting behaviour during the first wave of the pandemic. This can be explained by the stringent lockdown measures in these countries in April and early May 2020. The Swedish government promoted individual responsibility rather than mandatory restrictions during this period.

Our findings shed light on some social and economic features that policymakers could modify to bridge the socioeconomic gap we identified in self-protecting behaviours. Consistent with economic models that predict suboptimal individual be-

haviour in the presence of strong externalities where the costs of protective behaviour are unevenly distributed across socio-demographic groups, we find that people with lower socioeconomic status are less likely to adhere to protective behaviours. First, we show that increased knowledge about COVID-19 and trust in government, health care, and essential services is crucial to adopting protective behaviour. Second, we find that a lower level of self-protecting behaviours is associated with lower income. Third, we observe that suitable living arrangements and localities are positively related to health-protective behaviours.

Effective communication by authorities that restores trust, providing basic support to the financially disadvantaged, increasing access to open air for those residing in houses that lack basic outdoor features are some of the policy levers that government could pull to incentivise changes in protective behaviour.

References

Abel, M., Byker, T., and Carpenter, J. (2021). Socially optimal mistakes? Debiasing COVID-19 mortality risk perceptions and prosocial behavior. *Journal of Economic Behavior & Organization*, 183:456–480.

Adda, J. (2016). Economic activity and the spread of viral diseases: Evidence from high frequency data. *Quarterly Journal of Economics*, 131(2):891–941.

Alon, T., Doepke, M., Olmstead-Rumsey, J., and Tertilt, M. (2020). The impact of COVID-19 on gender equality. Technical report, National Bureau of economic research. WP 26947.

Alsharawy, A., Spoon, R., Smith, A., and Ball, S. (2021). Gender differences in fear and risk perception during the COVID-19 pandemic. *Frontiers in Psychology*, page 3104.

Ashraf, B. N. (2020). Economic impact of government interventions during the COVID-19 pandemic: International evidence from financial markets. *Journal of behavioral and experimental finance*, 27:100371.

Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda*, pages 507–547. University of Chicago Press.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.

Aubert, C. and Augeraud-Véron, E. (2021). The relative power of individual distancing efforts and public policies to curb the COVID-19 epidemics. *PLoS One*, 16(5):e0250764.

Barrios, J. M. and Hochberg, Y. (2020). Risk perception through the lens of politics in the time of the covid-19 pandemic. Technical report, National Bureau of Economic Research. WP 27008.

Biroli, P., Bosworth, S., Della Giusta, M., Di Girolamo, A., Jaworska, S., and Vollen, J. (2021). Family life in lockdown. *Frontiers in psychology*, 12.

Brown, C. S. and Ravallion, M. (2020). Inequality and the coronavirus: Socioeconomic covariates of behavioral responses and viral outcomes across US counties. Technical report, National Bureau of Economic Research. WP 27549.

Chiou, L. and Tucker, C. (2020). Social distancing, internet access and inequality. Technical report, National Bureau of Economic Research. WP 26982.

Clark, A., d'Ambrosio, C., and Lepinteur, A. (2021). The Fall in Income Inequality during COVID-19 in Five European Countries.

Collins, C., Landivar, L. C., Ruppanner, L., and Scarborough, W. J. (2021). COVID-19 and the gender gap in work hours. *Gender, Work & Organization*, 28:101–112.

Cutler, D. M. and Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics*, 29(1):1–28.

Dai, M., Tao, L., Chen, Z., Tian, Z., Guo, X., Allen-Gipson, D. S., Tan, R., Li, R., Chai, L., and Ai, F. (2020). Influence of cigarettes and alcohol on the severity and death of COVID-19: a multicenter retrospective study in Wuhan, China. *Frontiers in physiology*, 11:588553.

Dietz, W. and Santos-Burgoa, C. (2020). Obesity and its implications for COVID-19 mortality. *Obesity*, 28(6):1005.

Dohle, S., Wingen, T., and Schreiber, M. (2020). Acceptance and adoption of protective measures during the COVID-19 pandemic: The role of trust in politics and trust in science. *Social Psychological Bulletin*, 15(4):1–23.

Eder, S. J., Steyrl, D., Stefanczyk, M. M., Pieniak, M., Martínez Molina, J., Pešout, O., Binter, J., Smela, P., Scharnowski, F., and Nicholson, A. A. (2021). Predicting fear and perceived health during the COVID-19 pandemic using machine learning: A cross-national longitudinal study. *Plos one*, 16(3):e0247997.

Grimmer, J., Messing, S., and Westwood, S. J. (2017). Estimating heterogeneous treatment effects and the effects of heterogeneous treatments with ensemble methods. *Political Analysis*, 25(4):413–434.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):1–42.

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., and others (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, 5(4):529–538.

Imai, K. and Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470.

Jadil, Y. and Ouzir, M. (2021). Exploring the predictors of health-protective behavior during the COVID-19 pandemic: A multi-country comparison. *Environmental Research*, page 111376.

Jørgensen, F., Bor, A., and Petersen, M. B. (2021). Compliance without fear: Individual-level protective behaviour during the first wave of the COVID-19 pandemic. *British Journal of Health Psychology*, 26(2):679–696.

Karl, J. A., Fischer, R., Druică, E., Musso, F., and Stan, A. (2022). Testing the effectiveness of the health belief model in predicting preventive behavior during the covid-19 pandemic: The case of romania and italy. *Frontiers in psychology*, 12:6454.

Kassa, S. M. and Ouhinou, A. (2011). Epidemiological models with prevalence dependent endogenous self-protection measure. *Mathematical biosciences*, 229(1):41–49.

Künzel, S. R., Sekhon, J. S., Bickel, P. J., and Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116(10):4156–4165.

Lep, Z., Babnik, K., and Beyazoglu, K. H. (2020). The Role of Information Credibility in Emotional Responses and Engagement in Self-Protective Behaviour within Days of the COVID-19 Outbreak: A Cross-Sectional Study. *Frontiers in Psychology, https://doi.org/10.3389/fpsyg*.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*, pages 4768–4777.

McLaren, L. (2007). Socioeconomic status and obesity. *Epidemiologic reviews*, 29(1):29–48.

Mendoza-Jiménez, M.-J., Hannemann, T.-V., and Atzendorf, J. (2021). Behavioral risk factors and adherence to preventive measures: Evidence from the early stages of the COVID-19 pandemic. *Frontiers in public health*, 9.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Niedzwiedz, C. L., O'Donnell, C. A., Jani, B. D., Demou, E., Ho, F. K., Celis-Morales, C., Nicholl, B. I., Mair, F. S., Welsh, P., and Sattar, N. (2020). Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank. *BMC medicine*, 18:1–14.

Painter, M. and Qiu, T. (2020). Political beliefs affect compliance with covid-19 social distancing orders. *Covid Economics*, 4:103–123.

Pampel, F. C., Krueger, P. M., and Denney, J. T. (2010). Socioeconomic disparities in health behaviors. *Annual review of sociology*, 36:349–370.

Papageorge, N. W., Zahn, M. V., Belot, M., Van den Broek-Altenburg, E., Choi, S., Jamison, J. C., and Tripodi, E. (2021). Socio-demographic factors associated with self-protecting behavior during the Covid-19 pandemic. *Journal of Population Economics*, 34(2):691–738.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Prout, T. A., Zilcha-Mano, S., Aafjes-van Doorn, K., Békés, V., Christman-Cohen, I., Whistler, K., Kui, T., and Di Giuseppe, M. (2020). Identifying predictors of psychological distress during COVID-19: a machine learning approach. *Frontiers in Psychology*, 11:3063.

Reddy, R. K., Charles, W. N., Sklavounos, A., Dutt, A., Seed, P. T., and Khajuria, A. (2021). The effect of smoking on COVID-19 severity: A systematic review and meta-analysis. *Journal of Medical Virology*, 93(2):1045–1056.

Shapley, L. S. (1953). A value for n-person games. In Kuhn, H. W. and Tucker, A. W., editors, *Contributions to the Theory of Games (AM-28), Volume II*, pages 307–318. Princeton University Press.

Simonov, A., Sacher, S. K., Dubé, J.-P. H., and Biswas, S. (2020). The persuasive effect of fox news: non-compliance with social distancing during the covid-19 pandemic. Technical report, National Bureau of Economic Research. WP 27237.

Stickley, A., Matsubayashi, T., and Ueda, M. (2021). Loneliness and COVID-19 preventive behaviours among Japanese adults. *Journal of Public Health*, 43(1):53–60.

Vacondio, M., Priolo, G., Dickert, S., and Bonini, N. (2021). Worry, Perceived Threat and Media Communication as Predictors of Self-Protective Behaviors During the COVID-19 Outbreak in Europe. *Frontiers in psychology*, 12:231.

Van Lissa, C. J., Stroebe, W., vanDellen, M., Leander, P., Agostini, M., Gutzkow, B., Kreienkamp, J., Belanger, J., Draws, T., Grygoryshyn, A., et al. (2020). Early indicators of covid-19 infection prevention behaviors: Machine learning identifies personal and country-level factors. *PsyArXiv*.

Vögele, C., Lutz, A., Yin, R., and D'Ambrosio, C. (2020). How do different confinement measures affect people in Luxembourg, France, Germany, Italy, Spain and Sweden. *COME-HERE: First Report. Luxembourg. Available online at https://wwwen. uni. lu/research/highlights/how_do_diff erent_confinement_measures_affect_people_first_re sults, updated on*, 7(7):2020.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Wright, A. L., Sonin, K., Driscoll, J., and Wilson, J. (2020). Poverty and economic dislocation reduce compliance with COVID-19 shelter-in-place protocols. *Journal of Economic Behavior & Organization*, 180:544–554.

Xue, B. and McMunn, A. (2021). Gender differences in unpaid care work and psychological distress in the UK Covid-19 lockdown. *PloS one*, 16(3):e0247959.

Yang, Y. and Xiang, X. (2021). Examine the associations between perceived neighborhood conditions, physical activity, and mental health during the COVID-19 pandemic. *Health & Place*, 67:102505.

Appendices

3.A Data: Predictors of Self-protecting Behaviours

Sociodemographic factors: We consider gender, age, marital status, education, employment status, and residence. Age is recoded into seven-level ordinal categories, such as individuals below 20 years of age, five categories between 20 to 70 years of age, each spanning a range of 10 years, and individuals older than 70. Marital status is recoded into three level non-ordinal variables to allow us to understand the effect of cohabitation: single (includes individuals who never married, divorced, or widowed), married and cohabiting, and married but living apart. The highest level of educational attainment was measured with eight levels (primary, lower secondary, upper secondary, post-secondary, vocational training, bachelor's degree, master's degree, and doctoral degree).

The respondents' ties to the labour market were measured with four binary variables, employed full-time, employed part-time, retired, and unemployed. Furthermore, respondents were asked if they have been working in any 'key sectors' defined by governments. We recoded these measures into nine binary features: essential public services; energy and oil; health; food sector; water distribution; waste removal; finance and insurance; and transport. Moreover, respondents were also asked to report changes in employment status, work arrangements, and job losses since the implementation of social distancing. We measure the population density of the respondents' living area with a six-level ordinal variable according to the inhabitants in the living area: 1 "Isolated dwelling", 2 "less than 2,000", 3 "between 2,000 and 10,000", 4 "between 10,000 and 50,000", 5 "between 50,000 and 100,000", and 6 "more than 100,000".

Income and housing: Respondents were asked to record their net monthly disposable household income in January 2020 with the following income bands (all in Euros): 0-1250; 1250-2000; 2000-4000; 4000-6000; 6000-8000; 8000-12500; and >12500. A "prefer not to say" option was available as an alternative answer. Following Clark et al. (2021)

we take the mid-point of each band and adjusted it for purchasing power using 2019 Euros for household final consumption expenditures as the reference. We assign 15000 Euros to the open-ended top income band. To account for economies of scale across members of the same household, each income figure is equivalized using the square root of family size, and the resulting value is attributed to each household member. We then compute the relative poverty status of individuals using 60% of the national median of the equivalent income distribution.

We capture features relating to dwelling characteristics, such as housing type (with six categories), and the existence of an outdoor space in the housing (housing features, e.g., balcony, terrace, garden). We also compute a binary variable to capture the household-level overcrowding (1 = if the number of required rooms is bigger than the number of current rooms in the household), following the Eurostat definition (see OECD, 2021, https://data.oecd.org/inequality/housing-overcrowding.htm)

Health conditions: Participants were asked to report if they had pre-existing chronic diseases such as diabetes, heart disease, high blood pressure, pollen allergy, lung disease, asthma, cancer, and disability. Respondents were also asked if they had ever been clinically diagnosed with any mental disorders in the past and during the last two weeks.

Health and behavioural risk factors: We consider indicators for smoking, drinking, and BMI measures. Smoking was assessed by a binary variable, which takes value one if the respondents are currently smoking or were ex-smokers, and value zero if the respondents are non-smokers. Alcohol consumption is measured in the number of glasses in an average week and during the last week. BMI was calculated based on the weight and height of the respondents and categorised into three categories: "overweight or obese" if BMI greater than or equal to 25; "normal" if BMI is between 18.5 and 25; and "underweight" if BMI is below 18.5. In addition, we use a dummy variable measuring if respondents take any medication in a higher dose or frequency than prescribed.

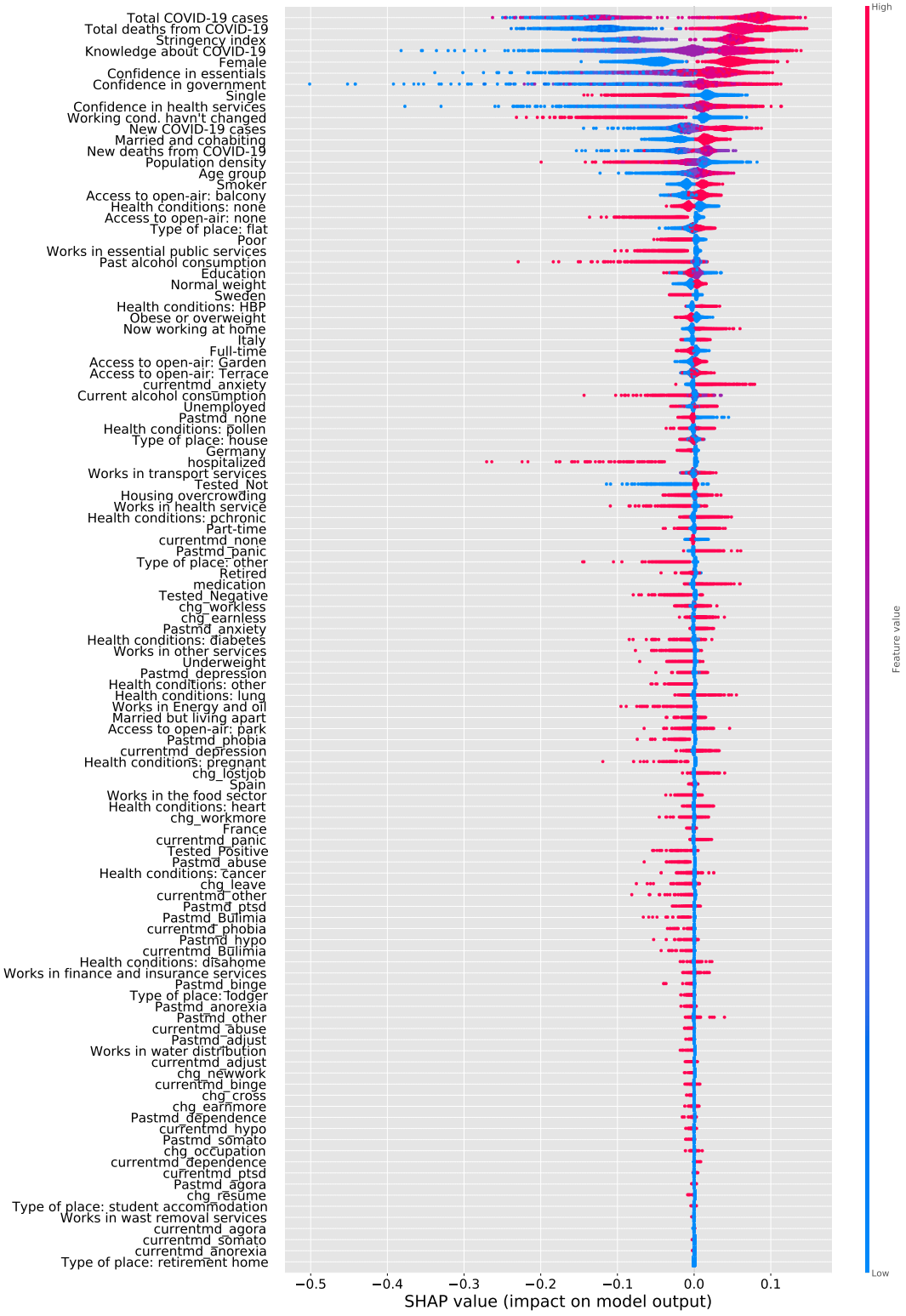
Confidence in institutions and knowledge about COVID-19: Participants rated using a 7-point Likert scale (ranging from 1 = "not at all confident" to 7 = "full confidence") their degree of confidence that 1) the government can handle COVID-19 well, 2) the health services can cope during COVID-19, and 3) essentials will be maintained during COVID-19. Individuals were also asked to rate their knowledge about COVID-19 using a 7-point Likert scale, with higher scores indicating better knowledge.

Incidences of COVID-19 and policy responses: We consult the Oxford COVID-19 Government Response Tracker (OxGRT) to access country-level information on the government policy response to COVID-19 and variables measuring the incidence and evo-

lution of the pandemic (Hale et al., 2021). We use the number of total and new COVID-19 cases and deaths to measure the evolution of the pandemic. The stringency index is a composite index computed using nine response indicators that gauge the "strictness" of the confinement measures of governments, such as: “school closures”; “workplace closures”; “cancellation of public events”; “restrictions on public gatherings”; “closures of public transport”; “stay-at-home requirements”; “public information campaigns”; “restrictions on internal movements”; and “international travel controls”. We merge this information with the COME-HERE data by date of interview to maintain variation at the individual level.

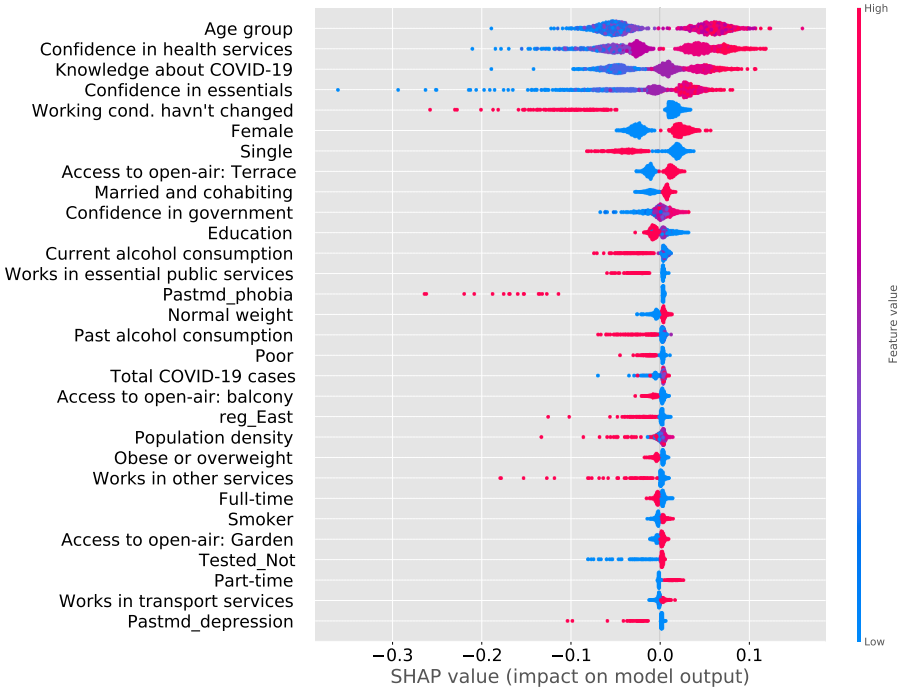
3.B Shapley Values of all CBS's Predictors

Figure. 3.8 *SHAP summary plot of all features*

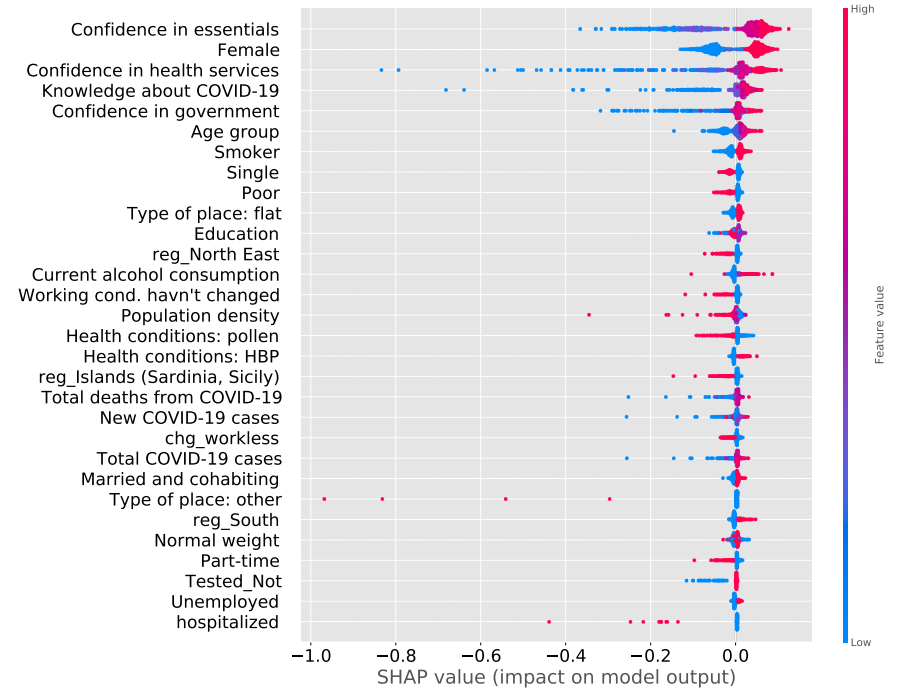


3.C Heterogeneity by Country

Figure. 3.9 *SHAP summary plots for each country*

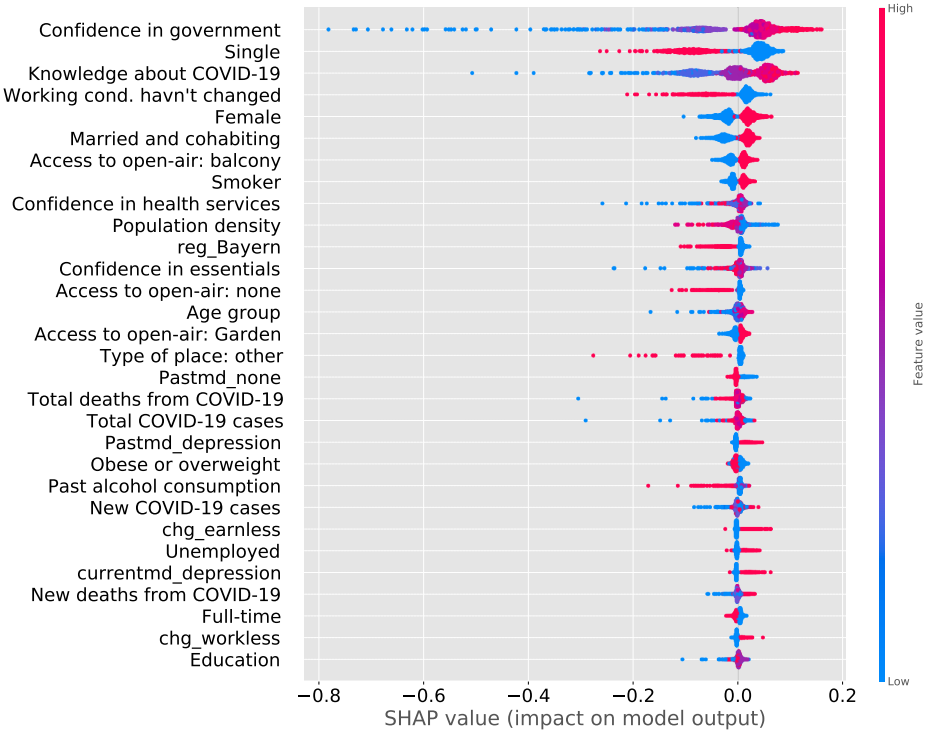


(a) *France*

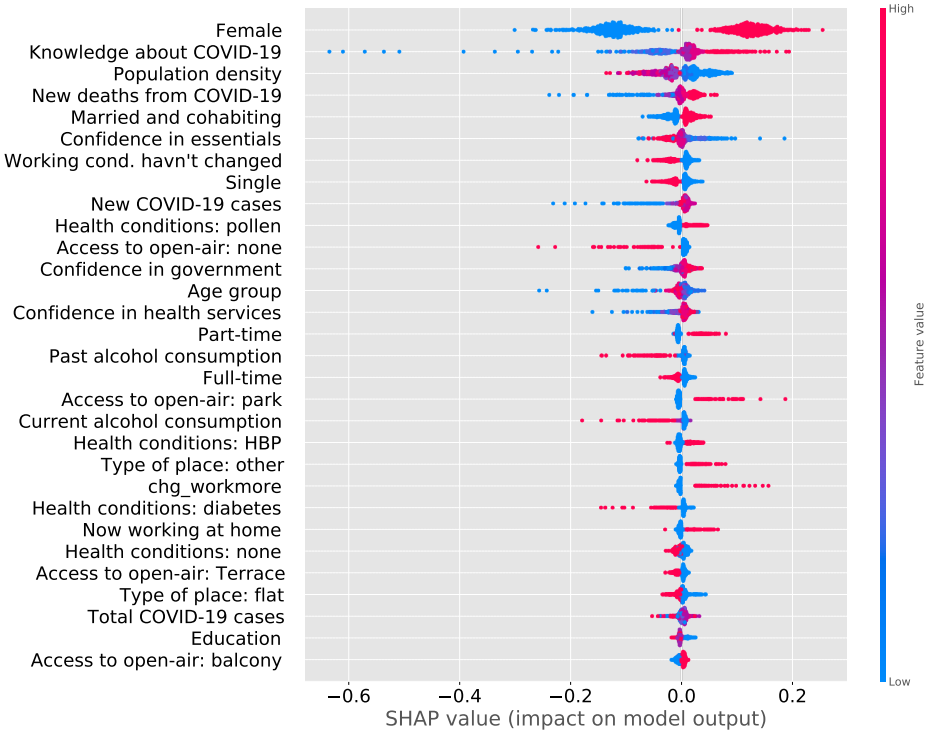


(b) *Italy*

(c) *Spain*

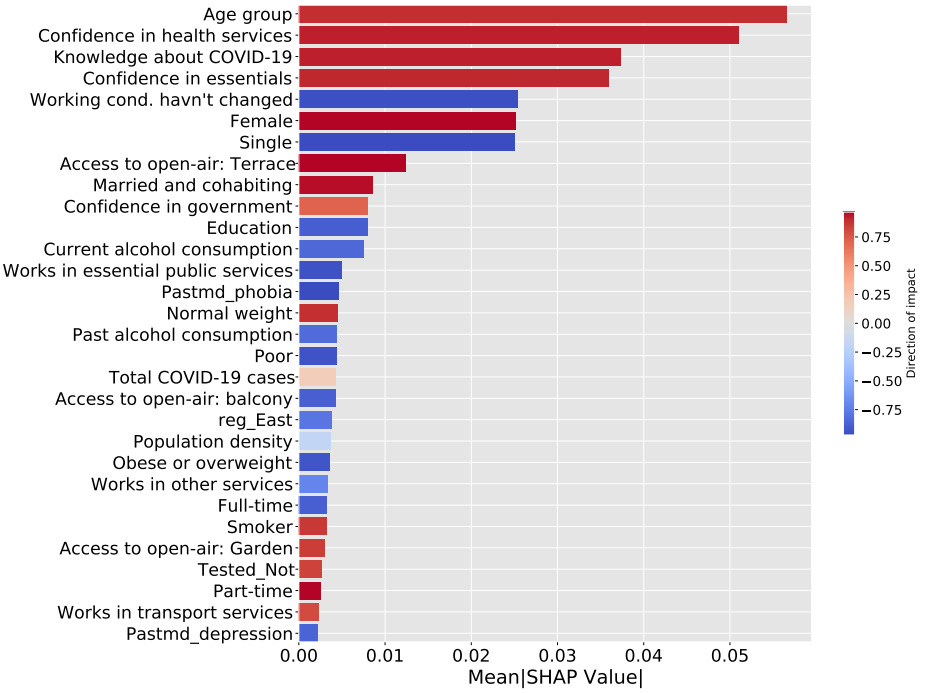


(d) *Germany*

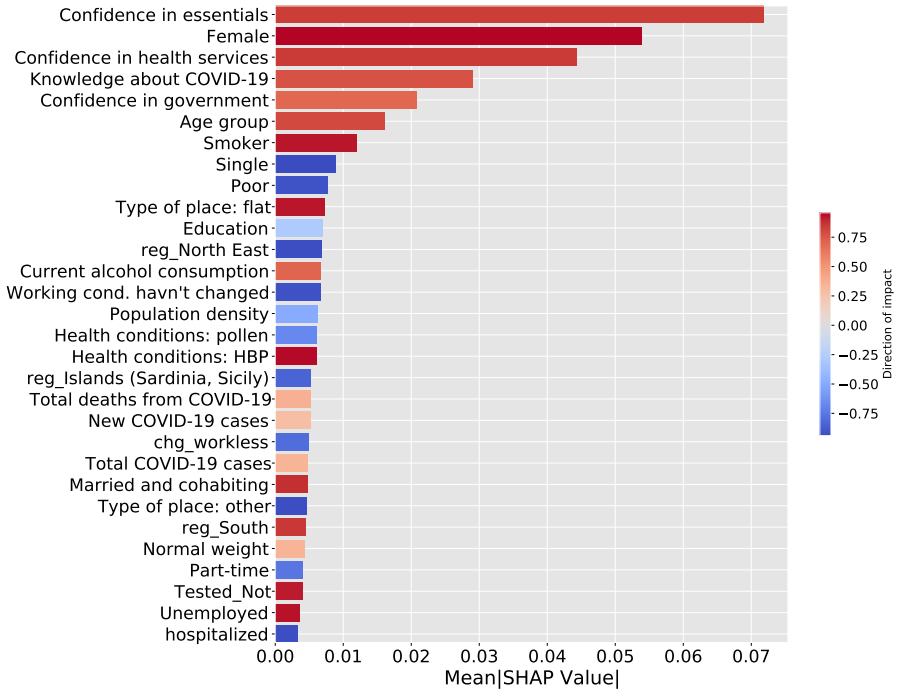

Female
Knowledge about COVID-19
Population density
New deaths from COVID-19
Married and cohabiting
Confidence in essentials
Working cond. havn't changed
Single
New COVID-19 cases
Health conditions: pollen
Access to open-air: none
Confidence in government
Age group
Confidence in health services
Part-time
Past alcohol consumption
Full-time
Access to open-air: park
Current alcohol consumption
Health conditions: HBP
Type of place: other
chg_workmore
Health conditions: diabetes
Now working at home
Health conditions: none
Access to open-air: Terrace
Type of place: flat
Total COVID-19 cases
Education
Access to open-air: balcony
−0.6
−0.4
−0.2
0.0
0.2
SHAP value (impact on model output)
High
Feature value
Low

(e) *Sweden*

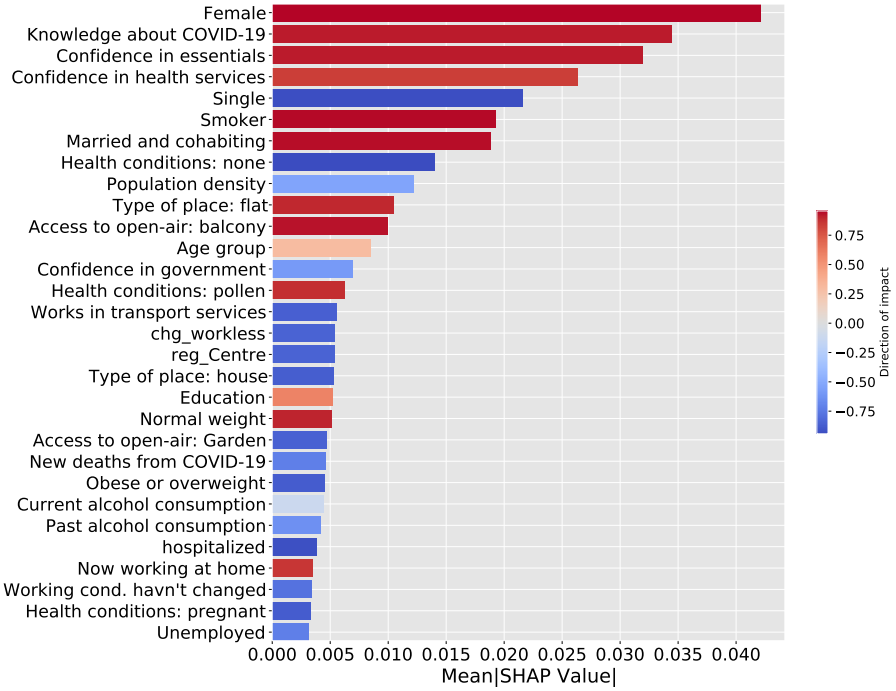
Figure. 3.8 *Global impact of the top 30 features for each country*



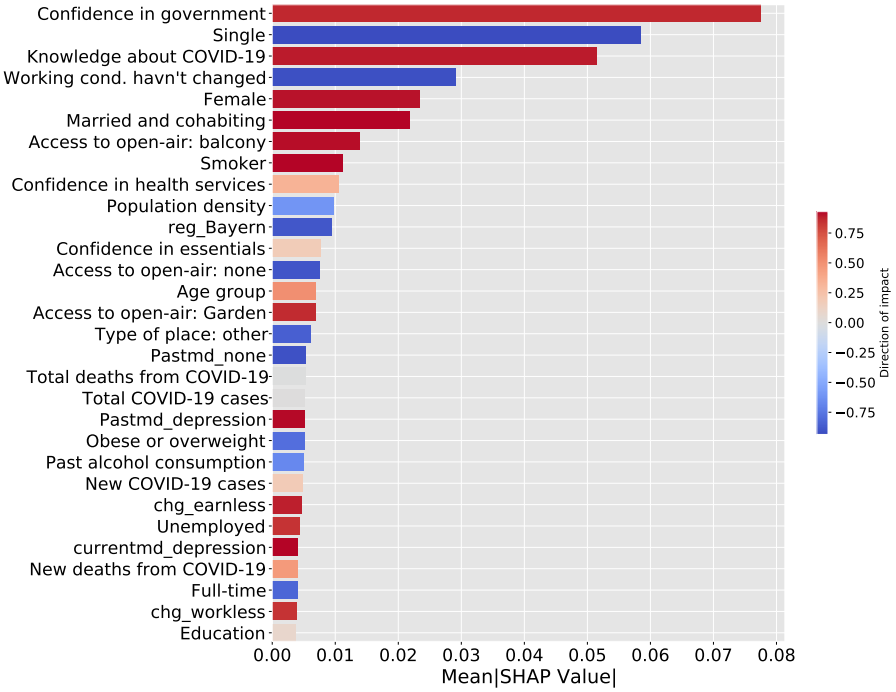
(a) *France*



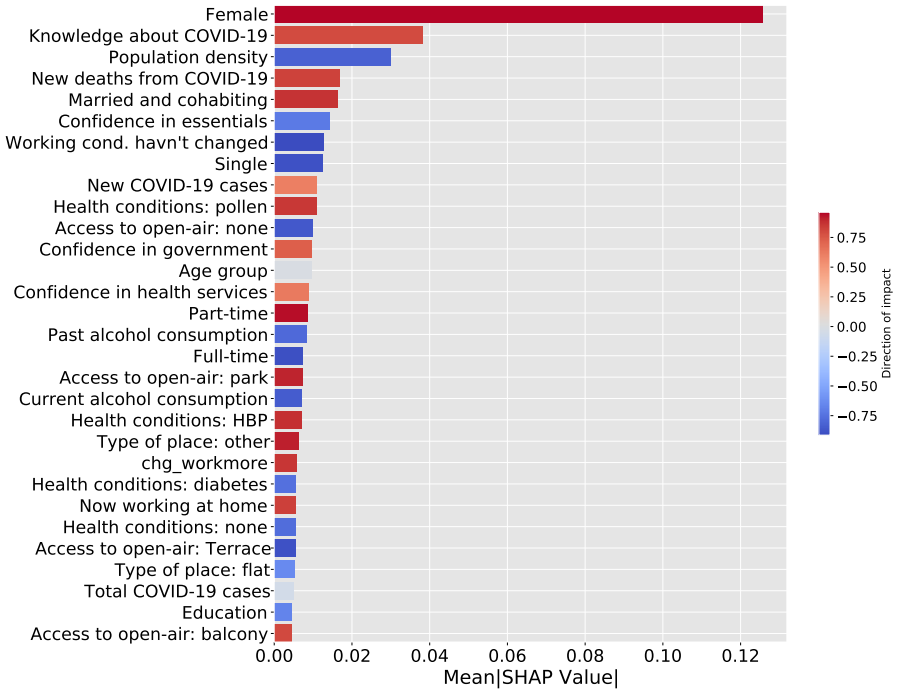
(b) *Italy*



(c) *Spain*



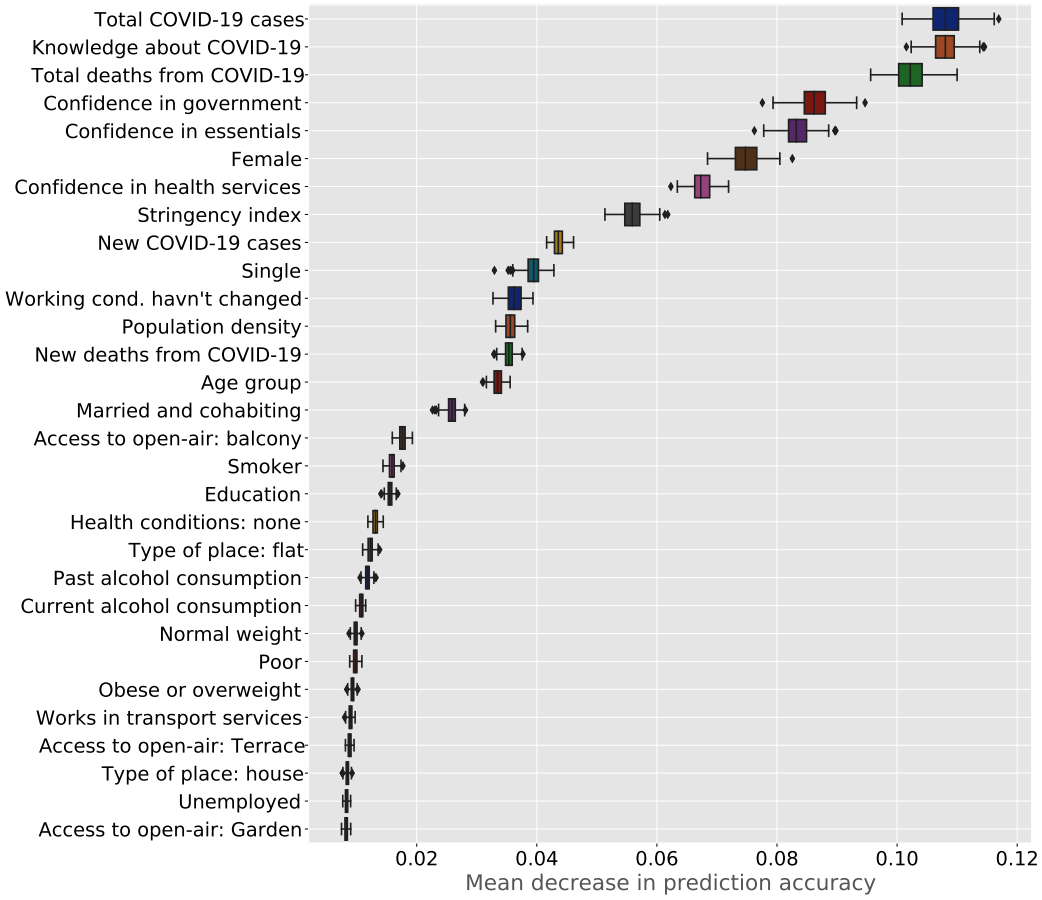
(d) *Germany*


Female
Knowledge about COVID-19
Population density
New deaths from COVID-19
Married and cohabiting
Confidence in essentials
Working cond. havn't changed
Single
New COVID-19 cases
Health conditions: pollen
Access to open-air: none
Confidence in government
Age group
Confidence in health services
Part-time
Past alcohol consumption
Full-time
Access to open-air: park
Current alcohol consumption
Health conditions: HBP
Type of place: other
chg_workmore
Health conditions: diabetes
Now working at home
Health conditions: none
Access to open-air: Terrace
Type of place: flat
Total COVID-19 cases
Education
Access to open-air: balcony
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Mean|SHAP Value|
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
Direction of impact

(e) *Sweden*

3.D Robustness Checks

Figure. 3.7 *Permutation feature importance of the top 30 feature*



Overall Conclusion

Conclusion

In this dissertation, we examine the prediction and measurement of individual well-being. We address research questions in three vital areas of individual well-being: measurement of vulnerability to poverty, classification of material and social deprivations, and identifying factors associated with individuals self-protective behaviour during a public health crisis in ways that are different from the traditional approach using explainable machine learning techniques and three large micro datasets. These studies reveal interesting and interrelated results with important social policy relevance.

The first chapter examines vulnerability to poverty as an essential measure of economic well-being to proactively target those families who will experience poverty in the future. After conducting several empirical experiments and robustness checks using the German Socio-Economic Panel (SOEP) data covering 1984-2020, this study proposes a data-driven explainable machine learning method to predict vulnerability to poverty better. Main results from feature importance with Shapley values show that the past year's income decile, household's ties to the labour market, human capital, household's composition (by age and activity status), and region of residence are the potent signal of the sources of vulnerability.

Chapter two analyses the prediction of unseen individuals' deprivation status and identifies the main predictors and their partial effects using the European Union Statistics on Income and Living Conditions (EU-SILC) microdata and applying machine learning (ML) algorithms. The key results indicate that the relative accuracy gained by using the more sophisticated ML algorithm is positive and significant compared to logistic regression. The socioeconomic and location features have the best classification power compared to sociodemographic and health characteristics. The relative feature importance and partial effects identified with Shapley's values reveal that individual's relative economic position to others, general health status, level of education, age, and

housing wealth is the most prominent predictor of material and social deprivation. In addition, this study reveals a slightly nonlinear association between income and the likelihood of deprivation and a hump-shaped association between age and deprivation status.

The third chapter investigates the most relevant predictors of individual self-protective behavioural responses during the first wave of the COVID-19 pandemic employing a data-driven approach of ExpML on real-time survey data from five European countries. Key findings of this paper shed light on essential policy variables that help to slow the current pandemic and guide the design of future policies with important implications for researchers and policy makers. Effective communication by authorities that restores trust, providing basic support to the financially disadvantaged, increasing access to open air for those residing in houses that lack basic outdoor features are some of the policy levers that government could pull to incentivise changes in protective behavior.

In conclusion, it is essential to consider some limitations when interpreting the findings of these studies. First of all, many complex ML algorithms might, on the one hand, be highly predictive but might, on the other hand, lack intrinsic model interpretability. To circumvent this issue, we employed explainable ML that has the advantage of providing an opportunity to improve prediction accuracy over the traditional approach using a 'black-box' supervised ML algorithm while still revealing interesting insights into a complex outcome. More specifically, we built ML models and complemented them with model-agnostic interpretative tools. Second, although our explainable ML approach is grounded in economic theory and data, we set out by making the distinction that this dissertation is concerned with prediction policy problems preventing causal conclusions. Third, in chapter two, we only report a limited number of interactions between features (out of all possible binary combinations) and limited our analyses to those we assumed to have the highest levels of policy relevance. Finally, overfitting poses a problem in these studies. To this end, we applied k-fold cross-validation for model selection to reduce the potential impact of overfitting.

Future Work

One potential direction for future work is to explore the estimation of the extent to which a given household is vulnerable to future income poverty. We only evaluate the performance of algorithms to a dichotomous outcome using poverty cutoff, but it may

be valuable to extend the analysis to include the extent of vulnerability. Additionally, we only focused on the German data, but comparing the method using data from other countries with different business cycles may be beneficial.

Another area for future research is to examine the effectiveness of different machine learning algorithms in predicting individuals' material deprivation. While I used a tree-based classifier and logistic regression in the study, numerous other algorithms, such as neural networks or causal forests, could be explored. Comparing the performance of different algorithms could provide valuable insights into the most effective approaches for accurately targeting disadvantaged groups. Additionally, I only focused on a binary indicator of deprivation, but evaluating other definitions, such as the extent and intensity of deprivation, may be beneficial.

Finally, exploring the over-time dynamics of self-protective behaviours against COVID-19 and their predictors may be worthwhile. For example, people may be more likely to wear masks in the early stages of the pandemic when there is less information about the other therapeutic measures but may become less vigilant as more information becomes available. In addition, it is crucial to identify the causal impact of socioeconomic variables on protective behaviours. Investigating these types of questions could help to inform public health interventions better.