Subjective Financial Well-Being An Explorative Analysis by Algorithmic Modeling Techniques

Thomas Christiaans, Karsten Lübke and Julia Richenhagen

Abstract Taking non-rational behavior and incomplete information into account, actual income distributions and financial well-being probably deviate from their perceptions. Based on a questionnaire sample of more than 45,000 Germans we investigate which socio-economic variables and combinations of these may help to model subjective financial well-being. Additionally to age, gender, and regional information, (perceived) household income and local and country-wide rankings of income distribution are used as possible modeling variables. The link is investigated by means of ordinary least squares regression as well as tree based methods. It turns out that on our sample, additionally to subjective income class and reported actual income, age, region and, though to a lower extend, gender help to model subjective financial well-being.

Thomas Christiaans
FOM University of Applied Sciences, 57078 Siegen, Germany
✉ thomas.christiaans@fom.de

Karsten Lübke
FOM University of Applied Sciences, 44139 Dortmund, Germany
✉ karsten.luebke@fom.de

Julia Richenhagen
FOM University of Applied Sciences, 45127 Essen, Germany
✉ julia.richenhagen@fom.de

Archives of Data Science, Series A
(Online First) DOI 10.5445/KSP/1000085951/13
KIT Scientific Publishing ISSN 2363-9881
Vol. 4, No. 1, 2018

1 Introduction

The distribution of income has been a main topic of interest in economics since the times of Adam Smith and David Ricardo up to the present, as witnessed by recent contributions of e.g. Kanbur and Stiglitz (2015) and Piketty (2014). Whether the distribution of income is fair, is an important question that even influences cohesion of modern societies. Related to the question of fairness is subjective financial well-being which depends on the actual as well as on the perceived position of individuals in the social hierarchy (Adam, 2014; Bach et al, 2014; Hans Böckler Stiftung, 2013).

One of the core assumptions of mainstream economic theory has since long been the concept of unbounded rationality, meaning that economic behavior is explained by assuming that agents under full information always make an optimum choice from an available set of alternatives (cf., e.g., Kirchgässner, 2008, pp. 12–13). Because of full information, the perceived position in the social hierarchy would not deviate from the actual position in that case. However, unbounded rationality is now increasingly called into question by behavioral economists and psychologists. Taking non-rational behavior and incomplete information into account, actual income distributions and financial well-being will probably deviate from their perceptions (Cruces et al, 2013; Engelhardt and Wagener, 2016; Karadja et al, 2017).

There are various possible definitions of financial well-being. E.g., the US Consumer Financial Protection Bureau (2015) disaggregates it using four dimensions, namely having control over day-to-day and month-to-month finances, having the capacity to absorb a financial shock, being on track meeting financial goals, and having the financial freedom to make choices allowing to enjoy life. A broader discussion of possible dimensions to be included in definitions of (not just financial) well-being and associated problems can be found in Decancq et al (2015), who also critically discuss equivalence scales used to calculate equivalised income. We will take a pragmatic point of view and use the OECD modified equivalence scale to measure individual incomes (cf. Section 2).

Several studies (see, e.g., Engelhardt and Wagener, 2014; Niehues, 2016) found that perceived well-being and inequality rather than actual inequality influence how critically people assess income differences. Similarly, redistributive preferences are less influenced by the actual distribution than by perceived inequality, that is, even political preferences and choices might depend more on perceptions than on factual data. An outstanding example is the median-voter model (Meltzer and Richard, 1981), where in an election people vote for an income tax (and do not care about anything else). In a two-party system, the party being nearest to the median voter will win the election because it collects more than 50 % of votes. This theoretical result predicts much more redistribution of income than what is actually observed. Among various possible reasons is that median voters may have a biased perception of their own income compared to the mean income, which also is related to the subjective feeling of economic or financial well-being. The present paper adds to the understanding of perceived or subjective financial well-being by exploring for the first time possible additional influencing factors using algorithmic modeling techniques, thereby assessing the relative importance of perceived compared to actual income.

The interest in factors of subjective well-being in general is reflected by a large number of empirical publications (for reviews cf., e.g., Diener, 1984; Diener et al, 1999; Dolan et al, 2008). Important factors influencing this broader concept of (not just financial) subjective well-being range from income, relative income, and income inequality to non-economic factors such as physical exercises and religious activities. As a general conclusion, Dolan et al (2008) emphasize that much of the evidence is still contradictory, suffers from potentially unobserved variables and a lack of certainty about the direction of causality. A major part of research in subjective well-being focuses on economic variables, especially on income (Diener and Biswas-Diener, 2002). Diener et al (2018) argue that income is a main factor because it fulfills basic as well as psychological needs. Several articles (e.g., Diener, 1984; Diener et al, 1993) describe a positive correlation between subjective income and subjective well-being.

While these and most other authors explore well-being in general, our focus is on financial well-being in particular. As we are concerned with the subjective aspects of this more narrow concept, we used simple direct questions in the questionnaire (cf. Fowler, 1992, 1995). We also extended the analysis by explicitly asking for subjective income classes additionally to actual incomes. Based on a questionnaire sample of more than 45,000 Germans we investigate which socio-economic variables and combinations of these maybe drivers for subjective financial well-being. Additionally to (perceived) household incomes, age, gender, employment and regional information are used as possible modeling variables of subjective financial well-being. In fact, our data indicates that perceived household income as measured by the subjective income class maybe more important for modeling subjective financial well-being than the actual income.

Financial well-being is measured by the subjective assessment of respondents. Using this direct approach has the advantage of a prospectively higher data quality by not overcharging the respondents and at the same time it may unravel additional factors important for subjective financial well-being. The clearly and coherently formulated questions as well as the used common scales can be expected to also increase this aspect of data quality (cf. Fowler, 1992, 1995).

As summarized by Dolan et al (2008), most studies use simple regression methods. In contrast, we apply algorithmic modeling techniques for the analysis (cf., e.g., Breiman, 2001). Tree-based methods are used to model the link between subjective financial well-being and possibly influencing factors. To be more precise, regression trees and random forests are applied. These methods allow to construct rules, that is, combinations of variable values which may be associated to subjective financial well-being, are easier to interpret than structural equation models, and can assess variable importance. Moreover, the metric or even normal distribution assumption is not needed. These methods can help to get (new) insights into econometric problems (Varian, 2014). For the data quality reasons already mentioned and because our focus is on subjective financial well-being, we concentrate, albeit not exclusively, on economic factors.

2 Survey and Data

In a study from September 15, 2016 to October 31, 2016, supervised by Oliver Gansser (Gansser, 2016), 5,558 students conducted face-to-face interviews with a standardized questionnaire in Germany for a student research project. A quota-sampling scheme according to age and gender was used. Despite the quota, regional and socio-economic distributions may be biased and therefore are not representative for Germany. Individual cheating may have occured, too. However, the established sample size of $n = 49{,}087$ is quite large. The following variables were collected:

- Subjective financial well-being: "*How do you feel economically*?" The scale ranges from poor (1) to rich (7).
- Subjective income class (subinclass): "*Which income class do you think you fit into*?" Low (1), below average (2), above average (3), high (4)?

- Demographic variables: age, gender, postal code (PLZ1 denotes the postal code truncated to the first digit, while the overall PLZ0 refers to entire Germany).
- Reported monthly net household income, ranging on a scale from *under 1,000* (1) in 500 Euros steps to *above 5,000 Euros* (10). These were also transformed to 10 percent quantiles, i.e. deciles, based on entire Germany (income0) and on the first and second digit of postal code, respectively (income1, income2). For example "income1 = 3" reads that the reported monthly net household income of the respondent is in the third decile of all respondents within the first digit of her postal code.

In order to improve data quality by keeping questions as simple as possible, just four subjective income classes have been used. Also, an *average income class* has been left out in order to enforce respondents to make a decision whether their perceived income exceeds or falls short of average income. We sticked to a more refined scale for actual income classes that allows calculating deciles, however. Nevertheless, these self-reported subjective data may be affected by some kind of response bias.

Reported monthly net household income was then used to calculate the approximate equivalised disposable incomes according to the OECD modified equivalence scale used by the Statistical Office of the European Union (cf., e.g., Hagenaars et al, 1994; Statistisches Bundesamt, 2017). This scale assigns a value of 1 to the household head, of 0.5 to each additional adult member and of 0.3 to each child below the age of fourteen. The equivalence number is the sum of the weights of all household members. Net household income (which officially should be gross income including transfers net of taxes on income and compulsory contributions to social insurance) is then divided by the equivalence number to obtain equivalised disposable income of each of the household's members.

3 Applied Methods

We conduct an explanatory analysis by algorithmic modeling in order to reveal the factors influencing subjective financial well-being. Starting from some explorative descriptive statistics, we proceed to classification and regression trees. In order to compare results, we also briefly discuss a linear regression model.

While the classic linear regression model assumes a linear relationship between dependent (y) and independent (x) variables (cf., e.g., Friedman et al, 2001, p. 41), this assumption is not needed in tree-based methods. Here a recursive binary splitting of the space of independent variables is used. In the derived set of rectangles for all x in a rectangle the same predicition $\hat{y}$ is assigned. To build up a tree, the x-variables are split up at suitable points from the root to the final nodes. Splitting up is done according to a split criterion, that is an impurity measure of the nodes, while controlling for overfitting by some parameters. The result is a tree revealing a hierarchical structure of x-variables and corresponding values modeling the y-variable. Further advantages of tree-based methods are their fairly easy interpretation as well as the absence of any distributional assumptions. Disadvantages are a high variance, i.e. small changes in the data can lead to a different tree, and quite often a relatively bad performance compared to methods like linear regression or random forests (cf., e.g., Friedman et al, 2001, pp. 266 ff.).

In random forests, many trees are aggregated. By using bootstrapped samples, a tree is built on a random sub-sample of the independent variables (cf., e.g., James et al, 2013, pp. 320 ff.). In order to assess variable importance by random forests, one of the variables of x is permuted (and the other variables in x not) and the decrease in performance is measured (cf., e.g., Strobl et al, 2007). The result is an ordering of x-variables according to their marginal importance for predicting the y-variable. For model comparison, the root mean squared error (cf., e.g., James et al, 2013, p. 29 f.), defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}, \tag{1}$$

with n being the number of observations, is applied. It measures how close predictions for out of sample observations are compared to the actual values y. We used the R-packages `rpart` (Therneau et al, 2017) and `randomForest` (Liaw and Wiener, 2002) to conduct the analysis.

4 Exploratory Data Analysis

Figure 1 shows the sample distribution of subjective income classes according to the respondents' own assessment. The distribution is not uniform and more than half of the respondents estimate their own income as belonging to the two lower classes. It thus is slightly right skewed according to the median-mean comparison (2.00 and 2.39), although the moment coefficient of skewness (−0.09) indicates a slight skewness to the left. Thus, the median-mean comparison of subjective assessments is as it should be considering the skewness of actual income distributions (cf., e.g., Boockmann et al, 2015). Our data also show the often observed tendency to seeing oneself in the middle of the social hierarchy (e.g., Evans and Kelley, 2004; Engelhardt and Wagener, 2016).

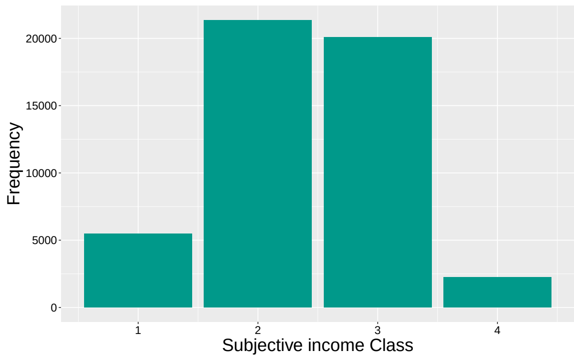

Figure 1: Distribution of subjective income class.

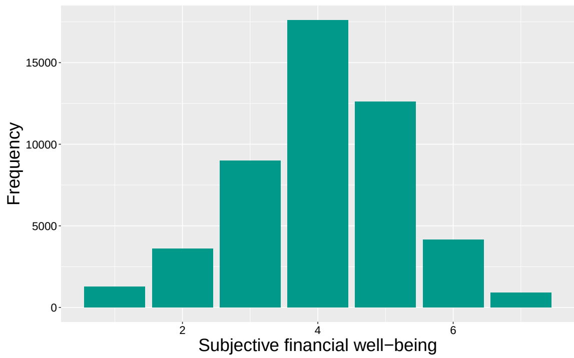

Figure 2: Distribution of subjective financial well-being.

The distribution of subjective financial well-being is shown in Figure 2. Again, it is slightly right skewed according to the median-mean comparison (median = 4.00 and mean = 4.07) but slightly left skewed according to the moment coefficient (−0.21). It also shows the tendency to the middle mentioned before.

Figure 3 shows the distribution of equivalised disposable incomes as calculated from the reported net household incomes. It is right skewed according to the median-mean comparison (median = 1,500.00 and mean = 1,669.81) as well as according to the moment coefficient (0.73). As before, this corresponds to our general knowledge about the skewness of income distributions (cf., e.g., Boockmann et al, 2015).

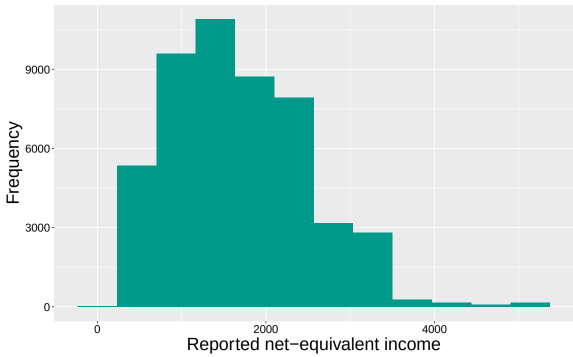

Figure 3: Distribution of reported equivalised disposable income.

Figure 4 shows that those assigning themselves to a higher subjective income class tend to report a higher subjective financial well-being (left side). Those in the lowest subjective income class (1) have the highest proportion of low subjective financial well-being (1,2) and for the highest subjective income class (4) have the highest proportion of high subjective financial well-being (6,7). Moreover, those who are in higher quantiles based on their reported income (nationwide) report a higher subjective financial well-being (right side): the proportion of those who are assigned to a higher income group have a higher proportion of high subjective financial well-being and vice versa. This is of course in line with expectations and will be further evaluated when discussing the modeling results.

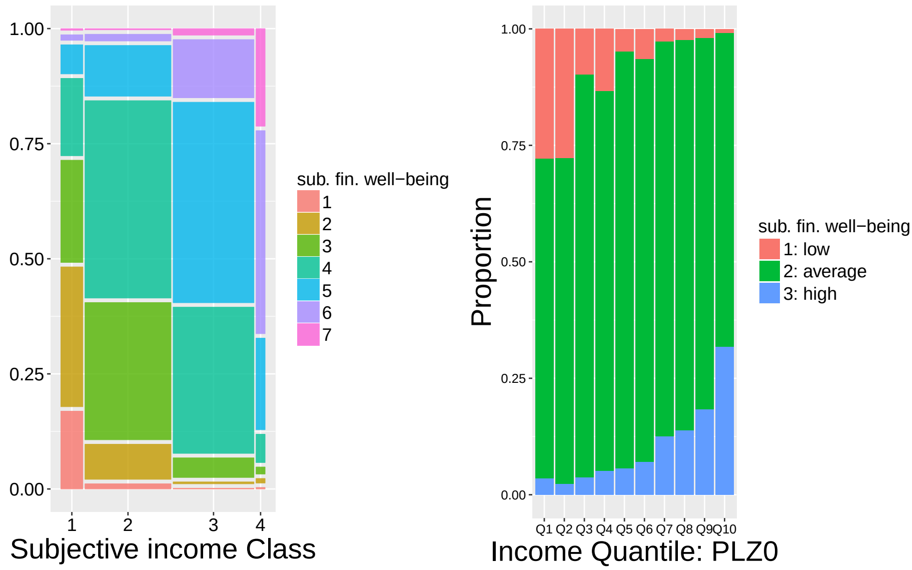

Figure 4: Distribution of subjective financial well-being by subjective income class and one-digit postal code.

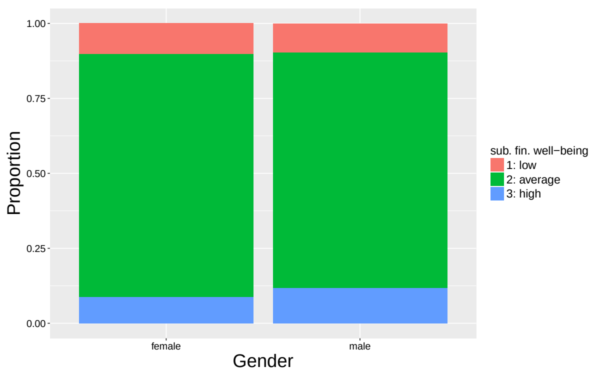

Figure 5: Impact of gender on subjective financial well-being.

Concerning demographic factors, Figure 5 shows a weak impact of gender. Slightly more men than women report high subjective financial well-being. The impact of age on subjective financial well-being is shown in Figure 6. Well-being is highest for the age group from over 44 to 54 years.

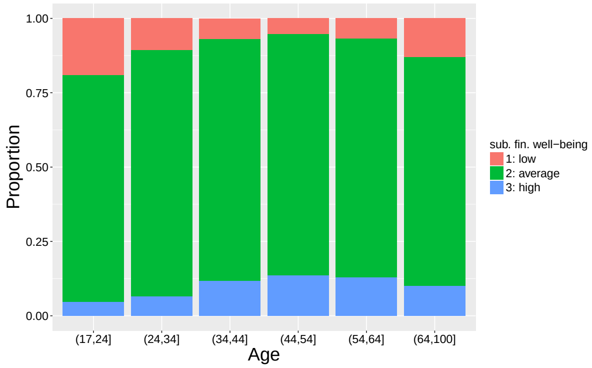

Figure 6: Impact of age on subjective financial well-being.

The impact of the region of living on subjective financial well-being is shown in Figure 7, where we have used one-digit German postal codes to classify regions. Although there is no clear-cut north-south or east-west gradient, there is a tendency for higher financial well-being in the south (postal codes 7 and 8) and also in the west (postal codes 4, 5, and 6).

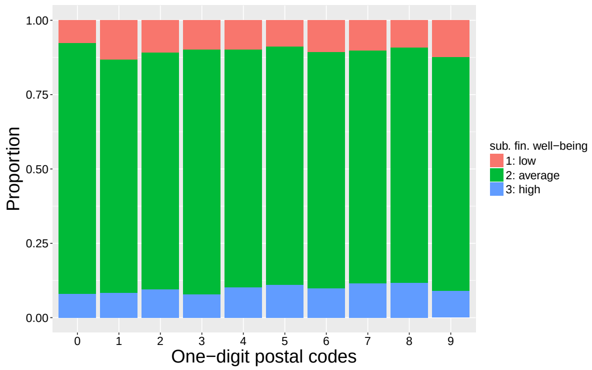

Figure 7: Impact of one-digit postal codes on subjective financial well-being.

Figure 8 shows the regression tree. The root of the tree for modeling subjective financial well-being is the subjective income class ("subinclass"), not the actual income quantile ("income"), but of course both variables are not independent. Additionally to actual income and the demographic variables, age and gender turn out to be important variables to model subjective financial well-being in the sample. Those where subjective and actual income is high rate their subjective financial well-being highest (6.0 on average – most right leaf of the tree), relatively young (or old) males with low subjective and actual incomes rate their subjective financial well-being lowest (2.1 – most left leaf of the tree).

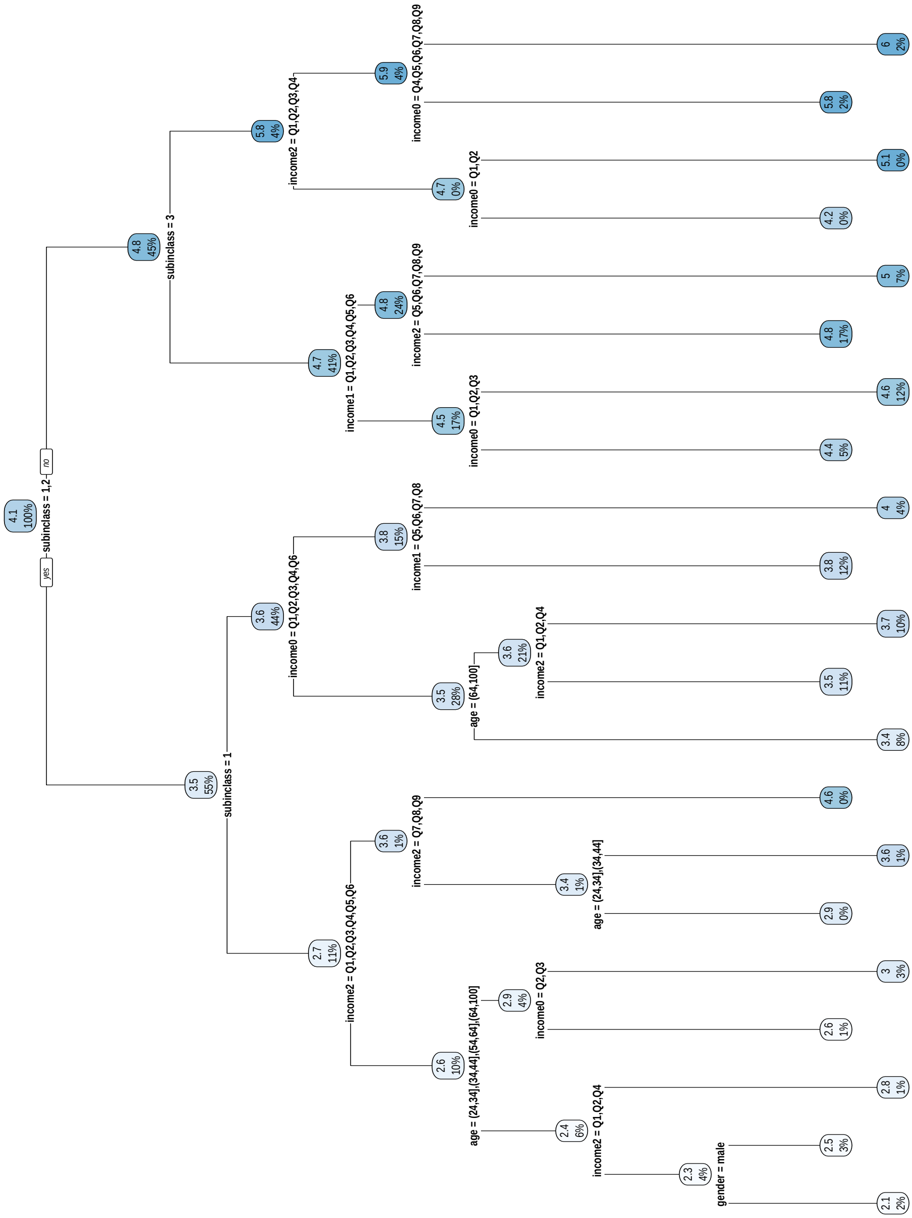

Figure 8: Regression tree for subjective financial well-being.

Variable importance in a random forest is shown in Fig. 9. One can see once again that the subjective income class (“subinclass”), conditionally on other variables, decreases a lot, both for MSE and node purity. Additionally, demographic factors help to model subjective financial well-being.

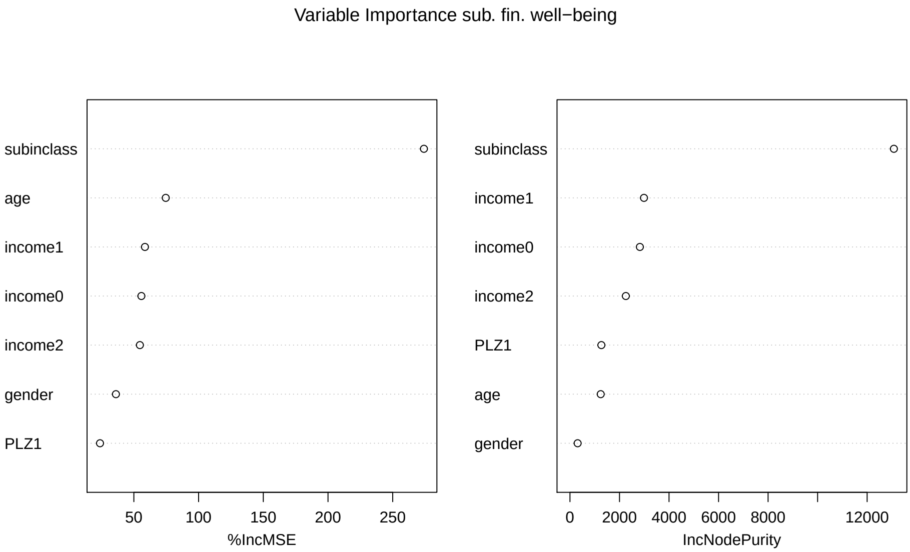

Figure 9: Importance of variables for subjective financial well-being.

The result of a linear model is given in Table 1, whereas the corresponding ANOVA table of a stepwise forward search based on AIC (cf., e.g., James et al, 2013, pp. 207-208) is given in Table 2. Since only classified (subjective) income data were collected, the income variables have been included using dummies. For example, “subinclass2 = 1” means that the respondent’s perceived income class is below average in class 2. One can see once again the high marginal value for the highest perceived income (“subinclass2 = 4”) with an estimated regression coefficient of 2.6681 as compared to the reference, the lowest subjective income class (1). Regarding the t-values, the negative estimates of age for rather young (age $(24, 34]$), old ($(64, 100]$) and male can be seen.

Table 1: Linear Regression for subjective financial well-being.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.8600	0.0423	67.66	0.0000
gendermale	−0.0493	0.0101	−4.87	0.0000
age(24,34]	−0.1696	0.0202	−8.38	0.0000
age(34,44]	−0.1153	0.0216	−5.33	0.0000
age(44,54]	−0.0681	0.0202	−3.38	0.0007
age(54,64]	−0.1327	0.0211	−6.28	0.0000
age(64,100]	−0.2009	0.0194	−10.35	0.0000
PLZ11	−0.1628	0.0438	−3.72	0.0002
PLZ12	−0.1031	0.0403	−2.56	0.0106
PLZ13	−0.0342	0.0433	−0.79	0.4295
PLZ14	−0.0758	0.0378	−2.01	0.0448
PLZ15	−0.0389	0.0383	−1.02	0.3098
PLZ16	−0.1150	0.0411	−2.80	0.0051
PLZ17	−0.0621	0.0424	−1.46	0.1432
PLZ18	−0.0366	0.0399	−0.92	0.3599
PLZ19	−0.1648	0.0444	−3.71	0.0002
income0Q2	−0.2766	0.0577	−4.79	0.0000
income0Q3	−0.0745	0.0828	−0.90	0.3682
income0Q4	0.0324	0.1239	0.26	0.7938
income0Q5	0.0981	0.1416	0.69	0.4884
income0Q6	0.0375	0.1524	0.25	0.8055
income0Q7	0.1162	0.1606	0.72	0.4692
income0Q8	0.1260	0.1666	0.76	0.4493
income0Q9	−0.0294	0.1734	−0.17	0.8655
income0Q10	−0.0318	0.1842	−0.17	0.8628
income1Q2	0.0647	0.0633	1.02	0.3065
income1Q3	0.0468	0.0887	0.53	0.5974
income1Q4	−0.0222	0.1299	−0.17	0.8640
income1Q5	0.0135	0.1475	0.09	0.9270
income1Q6	0.0545	0.1600	0.34	0.7332
income1Q7	0.0908	0.1673	0.54	0.5872
income1Q8	0.0800	0.1729	0.46	0.6439
income1Q9	0.4142	0.1802	2.30	0.0215
income1Q10	0.5432	0.1913	2.84	0.0045
income2Q2	0.0604	0.0521	1.16	0.2463
income2Q3	0.1412	0.0681	2.07	0.0382
income2Q4	0.0438	0.0787	0.56	0.5778
income2Q5	0.1609	0.0851	1.89	0.0588
income2Q6	0.1259	0.0911	1.38	0.1671
income2Q7	0.1700	0.0965	1.76	0.0781
income2Q8	0.1959	0.1014	1.93	0.0532
income2Q9	0.0689	0.1083	0.64	0.5246
income2Q10	0.1590	0.1167	1.36	0.1730
subinclass2	0.8437	0.0180	46.92	0.0000
subinclass3	1.7392	0.0194	89.56	0.0000
subinclass4	2.6681	0.0308	86.70	0.0000

Table 2: Anova Table for subjective financial well-being.

	Df	**Sum Sq.**	**Mean Sq.**	**F value**	**Pr(>F)**
subinclass	3	19232.66	6410.89	7808.53	0.0000
income0	9	1192.30	132.48	161.36	0.0000
age	5	135.44	27.09	32.99	0.0000
PLZ1	9	49.34	5.48	6.68	0.0000
income1	9	37.86	4.21	5.12	0.0000
gender	1	19.36	19.36	23.57	0.0000
income2	9	29.41	3.27	3.98	0.0000
Residuals	32695	26842.93	0.82		

Comparing Figure 8 (tree) and Figure 9 (random forest) with Table 1 (linear regression) and Table 2 (anova), it turns out that the subjective income class ("subinclass") has the strongest association with subjective financial well-being, followed by calculated reported income class and age. It should be noted that the subjective income class as well as the income quantiles on the three different aggregation levels are (highly) correlated, so that differences in importance may be spurious. Nevertheless, all variables help to improve modeling, as well as age, region and gender.

In order to investigate predictive modeling performance by means of $RMSE$ the data was split into 66.67 % training and 33.33 % test data, still leaving more than $n_{train} = 32{,}741$ observations for estimation and $n_{test} = 16{,}356$ observations for model evaluation. The root mean squared error of the test data is as follows:

- Base/Null Model: 1.22
- Tree: 0.94
- Random Forest: 0.91
- Linear Model: 0.91

Thus, approximately 25 % [$\approx (1.22 - 0.91)/1.22$] of (test data) variation of subjective financial well-being can be explained by the used variables. Despite the discrete response this shows that other variables than the ones used for modeling may also help to determine subjective financial well-being.

5 Conclusions

Our preliminary explorative analysis of subjective financial well-being shows that, despite the fact that reported income of our German sample is right skewed, subjective financial well-being is slightly left skewed. Our data provide corroborative evidence that subjective financial well-being is affected by the subjective income class additionally to actual income. Also, age, area, and gender turned out to be important variables for modeling subjective financial well-being. This insight was gained by algorithmic modeling techniques, specifically by using regression trees and random forests. To the best of our knowledge this is the first application of these techniques to this economics question.

In a next step we will try to disentangle direct and indirect effects of income, subjective income and subjective financial well-being by a causal mediation analysis. Despite the large sample size we do not claim that the sample is representative for Germany and different sources of bias may have occurred.

In future research we intend to improve the analysis by additionally checking for data quality with respect to cheating, representativeness etc. Also, more (derived) variables could be included in the analysis.

Acknowledgements We are grateful to Oliver Gansser, FOM München, for conducting the survey and sharing the data with us. The constructive comments of two anonymous referees led to significant improvement.

References

Adam H (2014) Die politische Ökonomie wachsender Ungleichverteilung der Einkommen. Wirtschaftsdienst 94(2):104–111, Springer, Berlin, Heidelberg. DOI: 10.1007/s10273-014-1633-6.

Bach S, Adam H, Niehues J, Schröder C, Frey C, Schaltegger CA, Berthold N, Gründler K (2014) Einkommens- und Vermögensverteilung – zu ungleich? Wirtschaftsdienst 94(10):691–712, Springer, Berlin, Heidelberg. DOI: 10.1007/s10273-014-1737-z.

Boockmann B, Kleimann R, Meythaler N, Nasgowitz A, Nielen S, Späth J, Bessomo AA, Lim K, Röhrs J, Stemmler H, Wegendt S, Wentz J, Pestel N, Sommer E, Biewen M, Blömer M, Dolls M, Löffler M, Markefke T, Peichl A, Stichnoth H, Ungerer M, Arndt C (2015) Forschungsprojekt: Analyse der Verteilung von Einkommen und Vermögen in Deutschland. Projektbericht an das Bundesministerium für Arbeit und Soziales. ZEW-Gutachten und Forschungsberichte, Bundesministerium für Arbeit und Soziales, Bonn (Germany). URL: http://hdl.handle.net/10419/155367.

Breiman L (2001) Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). Statistical Science 16(3):199–231, The Institute of Mathematical Statistics, Bethesda (USA). DOI: 10.1214/ss/1009213726.

Consumer Financial Protection Bureau (2015) Financial Well-Being: The Goal of Financial Education. Tech. Rep., Consumer Financial Protection Bureau (CFPB). URL: https://www.consumerfinance.gov/data-research/research-reports/financial-well-being/.

Cruces G, Perez-Truglia R, Tetaz M (2013) Biased Perceptions of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment. Journal of Public Economics 98:100–112. DOI: 10.1016/j.jpubeco.2012.10.009.

Decancq K, Fleurbaey M, Schokkaert E (2015) Inequality, Income, and Well-Being. In: Handbook of Income Distribution, Atkinson AB, Bourguignon F (eds), vol. 2A, North-Holland, Amsterdam, pp. 67–140.

Diener E (1984) Subjective Well-Being. Psychological Bulletin 95(3):542–575, American Psychological Association.

Diener E, Biswas-Diener R (2002) Will Money Increase Subjective Well-Being? Social Indicators Research 57(2):119–169, Springer Netherlands. DOI: 10.1023/A:1014411319119.

Diener E, Sandvik E, Seidlitz L, Diener M (1993) The Relationship between Income and Subjective Well-Being: Relative or Absolute? Social Indicators Research 28(3):195–223, Kluwer Academic Publishers. DOI: 10.1007/BF01079018.

Diener E, Suh EM, Lucas RE, Smith HL (1999) Subjective Well-Being: Three Decades of Progress. Psychological Bulletin 125(2):276–302, American Psychological Association.

Diener E, Oishi S, Tay L (2018) Advances in Subjective Well-Being Research. Nature Human Behaviour 2:253–260. DOI: 10.1038/s41562-018-0307-6.

Dolan P, Peasgood T, White M (2008) Do we really know what makes us happy? A Review of the Economic Literature on the Factors associated with Subjective Well-Being. Journal of Economic Psychology 29:94–122. DOI: 10.1016/j.joep.2007.09.001.

Engelhardt C, Wagener A (2014) Biased Perceptions of Income Inequality and Redistribution. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2463129. CESIFO Working Paper No. 4838.

Engelhardt C, Wagener A (2016) What do Germans Think and Know about Income Inequality? A Survey Experiment. URL: https://www.wipol.uni-hannover.de/fileadmin/wipol/publications/Wagener_What_do_Germans_think_and_know_about_income_inequality.pdf. Discussion Paper, Leibniz University of Hannover, ECINEQ Working Paper No. 389.

Evans MDR, Kelley J (2004) Subjective Social Location: Data from 21 Nations. International Journal of Public Opinion Research 16(1):3–38, World Association for Public Opinion Research (WAPOR). DOI: 10.1093/ijpor/16.1.3

Fowler FJ (1992) How Unclear Terms Affect Survey Data. Public Opinion Quarterly 56(2):218–231, American Association for Public Opinion Research (AAPOR). DOI: 10.1086/269312.

Fowler FJ (1995) Improving Survey Questions: Design and Evaluation. Applied Social Research Methods, SAGE Publications, Thousand Oaks (USA). ISBN: 978-0-803945-83-8.

Friedman J, Hastie T, Tibshirani R (2001) The Elements of Statistical Learning. Data Mining, Inference, and Prediction, 1st edn. Springer Series in Statistics, Springer, New York (USA). ISBN: 978-0-387216-06-5, DOI: 10.1007/978-0-387-21606-5.

Gansser O (2016) Weihnachtsumfrage 2016 Einkaufsverhalten der Konsumenten in Deutschland in Bezug auf Weihnachtsgeschenke. FOM Hochschule für Ökonomie & Management, Institut für Empirie & Statistik, Munich (Germany). URL: `https://www.fom.de/uploads/forschungsprojekte/downloads/161108_FOM_Weihnachtsumfrage_2016.pdf`.

Hagenaars A, de Vos K, Zaidi MA (1994) Poverty Statistics in the Late 1980s: Research Based on Micro-data. Tech. Rep., Office for Official Publications of the European Communities, Luxembourg.

Hans Böckler Stiftung (2013) Verteilung – Die stärksten Schultern werden geschont. Böckler Impuls 15:1. URL: `https://www.boeckler.de/44225_44246.htm`.

James G, Witten D, Hastie T, Tibshirani R (2013) An Introduction to Statistical Learning with Applications in R, 1st edn. Springer Texts in Statistics, Springer, New York (USA). DOI: 10.1007/978-1-4614-7138-7.

Kanbur R, Stiglitz JE (2015) Wealth and income distribution: New theories needed for a new era. VOX CEPR's Policy Portal. URL: `http://voxeu.org/article/wealth-and-income-distribution-new-theories-needed-new-era`.

Karadja M, Mollerstrom J, Seim D (2017) Richer (and Holier) Than Thou? The Effect of Relative Income Improvements on Semand for Redistribution. Review of Economics and Statistics 99(2):201–212, The MIT Press. DOI: 10.1162/REST_a_00623.

Kirchgässner G (2008) Homo Oeconomicus – The Economic Model of Behaviour and its Applications in Economics and Other Social Sciences. Springer, New York (USA). DOI: 10.1007/978-0-387-72797-4.

Liaw A, Wiener M (2002) Classification and Regression by randomForest. R News 2(3):18–22. URL: `http://CRAN.R-project.org/doc/Rnews/`.

Meltzer AH, Richard SF (1981) A Rational Theory of the Size of Government. Journal of Political Economy 89(5):914–927, The University of Chicago Press.

Niehues J (2016) Ungleichheit: Wahrnehmung und Wirklichkeit – ein internationaler Vergleich. Wirtschaftsdienst 96:13–18, Springer / Leibniz-Informationszentrum Wirtschaft (ZBW), Berlin, Heidelberg / Hamburg, Kiel (Germany). DOI: 10.1007/s10273-016-1945-9. Supplement 1.

Piketty T (2014) Capital in the Twenty-First Century. The Belknap Press of Harvard University Press, Cambridge (USA). ISBN: 978-0-674430-00-6.

Statistisches Bundesamt (2017) Leben in Europa (EU-SILC): Einkommen und Lebensbedingungen in der Europäischen Union. Wiesbaden (Germany). URL: `https://www.destatis.de/GPStatistik/receive/DESerie_serie_00000501`.

Strobl C, Boulesteix AL, Zeileis A, Hothorn T (2007) Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. Springer Nature, BMC Bioinformatics 8(1):25, BioMed Central Ltd. DOI: 10.1186/1471-2105-8-25.

Therneau T, Atkinson B, Ripley B (2017) rpart: Recursive Partitioning and Regression Trees. URL: `https://CRAN.R-project.org/package=rpart`. R package version 4.1-11.

Varian HR (2014) Big Data: New Tricks for Econometrics. Journal of Economic Perspectives 28(2):3–28. DOI: 10.1257/jep.28.2.3