Social Indicators Research (2021) 153:561–584
https://doi.org/10.1007/s11205-020-02505-6

ORIGINAL RESEARCH

Over-indebted Households in Poland: Classification Tree Analysis

Grzegorz Wałęga[1] · **Agnieszka Wałęga**[2]

Accepted: 21 September 2020 / Published online: 3 October 2020

Abstract
Increasing a personal debt burden implies greater financial vulnerability and threats for macroeconomic stability. It also generates a risk of the households over-indebtedness. The assessment of over-indebtedness is conducted with the use of various objective and subjective measures based on the micro-level data. The aim of the study is to investigate over-indebted households in Poland using a unique dataset obtained from the CATI survey. We discuss and compare the usefulness of various over-indebtedness measures across different socio-economic characteristics. Due to the differences in over-indebtedness across single measures, we perform a more complex assessment using a mix of indicators. As an alternative to other commonly criticised over-indebtedness measures, we apply the "below the poverty line" (BPL) measure. In order to obtain the profile of over-indebted households, we use classification and regression tree analysis as an alternative to logit or probit models. We find that DSTI ("debt service to income") ratio underestimates the extent of over-indebtedness in vulnerable groups of households in comparison with the BPL. We highlight the necessity to use different measures depending on the adopted definition of over-indebtedness. A psychological burden of debts is particularly strong among older and poorly educated respondents. We also find that the age structure of over-indebted households in Poland differs from this structure in countries with a broader access to consumer credits. Our results can be used to enrich the methods of assessing the household over-indebtedness.

Keywords Over-indebtedness · Classification tree analysis · Poverty line · Survey data

1 Introduction

The last decades were a period of a substantial increase in household debt worldwide (Barba and Pivetti 2009; Karwowski et al. 2019). Household indebtedness ratios have been trending up since 2000 in nearly all OECD countries. In the years 2000–2017, the average

✉ Grzegorz Wałęga
grzegorz.walega@uek.krakow.pl

[1] Department of Microeconomics, Cracow University of Economics, Rakowicka 27, 31-510 Kraków, Poland

[2] Department of Statistics, Cracow University of Economics, Rakowicka 27, 31-510 Kraków, Poland

debt to income ratio increased almost twice (OECD 2020). Household debts that remained after the recent crisis are increased by new ones resulting from easier access to credit and growing house prices, as well as the improvement of the financial situation and consumers' sentiments (Turinetti and Zhuang 2011; Zabai 2017). The substantial increase in loans is also observed in Poland. The amount of outstanding debt in Poland exceeds 676 billion PLN, which is equivalent to around 62% of net disposable household income in 2017 compared to 22% in 2004 (OECD 2020).

With an increasing debt burden attention should be paid to over-indebted households, because growing indebtedness in nominal and relative terms exposes debtors to greater financial vulnerability, especially to external shocks. The latest economic slowdown has demonstrated the significant role of debt in the financial strain of households and their financial fragility (Bańkowska et al. 2015; Hiilamo 2018). Over-indebtedness is recognized not only as a root of the unbalanced households budgets and consumption deprivation but also is seen as a threat to subjective well-being (Keese and Schmitz 2014; Tay 2017). A debt strain is commonly regarded as stress and depression factor (Gathergood 2012). Furthermore, a growing body of literature reports that the micro-level indebtedness can have serious implications for the macroeconomic financial stability (Mian et al. 2017; Coletta et al. 2019; Ramsay and Williams 2020). For this reason, we believe that monitoring over-indebtedness is a necessity. When the number of households with financial difficulties is growing, over-indebtedness becomes both an economic and social issue. Particularly the relatively small, but heavily indebted household fraction exerts a considerable impact on welfare costs (Campbell 2006; Girouard et al. 2006; HFCN 2009). The effective policy, therefore, requires an appropriate and comprehensive measurement of this phenomenon.

It is relatively easy to quantify over-indebtedness of households when aggregate measures are used. However, even though this approach identifies the average household position on the debt market, it does not reveal the situation of a particular household (Anderson et al. 2016; Ferretti and Vandone 2019). The aggregate view masks for instance information about the ways in which individual households perceive their debt. Therefore, the micro-level approach is a better source of data on the financial position of households.

Our study is a contribution to the growing literature that investigates the over-indebtedness of households. Although the indebtedness in Poland is significantly lower than in highly developed countries, the increase of a debt burden leads to a situation in which households are likely to experience difficulties in managing their debts and to become more prone to over-indebtedness. This raises many questions regarding whether the recent acceleration of credit growth in Poland affects over-indebtedness and what is the financial condition of different types of households. However, to the best of our knowledge, few articles addressing over-indebted households in Poland have been published so far.

The aim of the article is to investigate the scale of over-indebted households in Poland employing different measures of over-indebtedness. In order to better capture the differences between various indicators of over-indebtedness we also discuss their matrix. We provide an in-depth analysis of the socio-economic characteristics of household based on a unique dataset obtained from the CATI survey of indebted Polish households conducted in 2018. In our opinion, this study, although it refers primarily to Polish experiences, is a valuable voice in the discussion on the measures of households over-indebtedness and the methods of its assessment, whose number is still inadequate.

Our contribution to subject literature is twofold. The study is the very first use of "below the poverty line" (BPL) measure to assess over-indebtedness in Poland. This approach allows adopting a different perspective and sheds light on the social dimension of over-indebtedness, especially in a group of vulnerable households. In our opinion, commonly

used objective measures of over-indebtedness, which are based on arbitrarily set thresholds, can yield a distorted profile of over-indebted households.

Secondly, we apply Classification and Regression Tree method to obtain the profile of over-indebted households. We believe that the CART analysis, so far not used in studying of over-indebted households, will allow identifying the importance of these characteristics of households which affect the risk of becoming over-indebted. This method is an alternative to previously applied ones based mainly on logit models.

This paper is organized as follows. The Sect. 2 offers a brief overview of the definition of over-indebtedness and the way of its measuring. The third section outlines the research methods, while the next one discusses the dataset and the structure of the sample. The objective and subjective over-indebtedness indicators and the CART results are presented in Sect. 5. The final sections contain discussion and conclusions.

2 Definition and Measurement of Over-indebtedness

There are many approaches to defining and measuring over-indebtedness. This diversity results from different socio-economic and legislative backgrounds of over-indebtedness in the international contexts. Most authors focus on the identification of over-indebtedness and its causes and consequences in the financial systems (Ntsalaze and Ikhide 2016; Hyytinen and Putkuri 2018).

Oxera (2004) defines over-indebtedness as a situation in which a household is not only in arrears on a structural basis but also if it is at significant risk of getting into arrears on this basis. Over-indebtedness can also be defined in the context of the household's ability to meet its financial obligations. Haas (2006) defines over-indebtedness as a situation in which household's income "in spite of a reduction of the living standard, is insufficient to discharge all payment obligations over a long period of time". According to Angel and Heitzmann (2015), over-indebtedness usually results from the household's illiquidity.

Anderloni et al. (2012) propose a definition of "financially vulnerable" households which have problems with arrears and default in loan commitments. What causes households' financial fragility is not only their over-commitment resulting from excess indebtedness but also other conditions of financial instability (e.g. unbalanced budgets, adverse shocks). While describing fragile households, Brunetti et al. (2016) pay attention to non-optimal portfolio allocation rather than the absolute debt level. Over-indebtedness can also be considered as the negative financial margin of a household and its ability to cope by liquidating financial assets (Ampudia et al. 2016; Bettocchi et al. 2018).

Another approach to defining over-indebtedness is proposed by the European Commission (2008, 2010). A common definition used across the EU has indicated a set of criteria to be applied to identify over-indebtedness, which include, among others, a social context. It is emphasised that the household experiencing over-indebtedness is unable to meet its recurrent or unexpected expenses and that the relatively high commitment payments push it below the poverty line. Thus, this debt strain substantially reduces its ability to meet its needs and adversely affects its well-being.

To sum up, it is possible to indicate several common features of the definition of over-indebtedness, such as:

- the economic dimension (the ability to repay the debt),
- the temporal dimension (the problem is not incidental, but has at least medium or long term time horizon),
- the social dimension (the necessity to substantially reduce the expenses that have to be met before the repaying the debt),
- the psychological dimension (stress caused by over-indebtedness).

As there is no consensus regarding the definition of over-indebtedness in the literature (Kempson 2002; Bridges and Disney 2004; Kempson et al. 2004; Bańkowska et al. 2015; D'Alessio and Iezzi 2016; Bourova et al. 2019), likewise there is no consensus on how to measure it (Table 1). This is in good agreement with Betti et al. (2007) who claim that the measurement of over-indebtedness to a large extent is based on a wide range of ad hoc statistical indicators calculated using public and private data sets.

Objective over-indebtedness indicators based on the quantitative ratios, such as debt to income (DTI), debt to wealth or assets (DTW/DTA), and debt-service to income (DSTI) are employed by many authors (Brown and Taylor 2008; Faruqui 2008; Keese 2009; Magri and Pico 2012; Jappelli et al. 2013). Some of them use these indicators to distinguish between secured and unsecured debt (del Rio and Young 2008; French and Vigne 2018). It is assumed that unsecured debt is characterized by a relatively high rate of interest, and it tends to be relatively the most expensive way of borrowing (Brown and Taylor 2008), so the higher risk of over-indebtedness can be expected.

Among the measures listed above, debt-service to income ratio (DSTI) seems the most relevant indicator. Most studies put the limit at 30% or 40% of debt-service to income (Tiongson et al. 2009; Michelangeli and Pietrunti 2014; Sánchez-Martínez et al. 2016; D'Alessio and Iezzi 2016; Terraneo 2018). In countries with the well-developed credit markets (e.g. Great Britain, USA), 50% cost of debt to income ratio is identified as a threshold beyond which households are deemed a significant burden (Oxera 2004). In a modified approach interest payments and minimum repayments as a proportion of disposable income are taken into account as an approximation of what households are normally required to repay (Oxera 2004; Faruqui 2008). The key is to find an adequate threshold for the identification of over-indebted households (Ampudia et al. 2016).

One of the solutions for overcoming limitations of the DSTI ratio is to adopt a financial (economic) margin measure. This indicator—apart from debt repayments—takes into account necessary living expenses (D'Alessio and Iezzi 2016) or expected expenditure (Bettocchi et al. 2018). Thus, the assessment of over-indebtedness is made in the context of the lifecycle of consumption and borrowing of households. When the negative financial margin emerges the household can be treated as over-indebted because it finds it hard to meet its financial obligations without deteriorating its living standard and might, therefore, default on its debts.

An interesting approach to identifying over-indebted households is the application of the poverty line as a threshold. The measures that deal with the poverty line are rather intuitive and refer to a commonly accepted benchmark: if deducting its debt payment from the household income puts it below the poverty line, over-indebtedness occurs. To the best of our knowledge, D'Alessio and Iezzi (2013) and Ntsalaze and Ikhide (2016) seem to be the only authors who refer to this indicator in assessing over-indebtedness on the basis of the micro-level data.

Other indicators which are used in the assessment of over-indebtedness are a number of loans (NL) and a number of arrears (NA). Kempson (2002) identifies a relationship between reporting being in arrears and having four or more credit commitments. A large number of loans may indicate difficulties in self-control and budget management, however,

Table 1 Overview of over-indebtedness measures. *Source*: Own study

Ratio (measure) name	Description	Drawbacks
Debt to income (DTI)	The ratio of outstanding debt to household annual income	Comparing a stock measure (household debt) with a flow measure (income)
		Limited data on outstanding debt
Debt to wealth (DTW) or debt to assets (DTA)	The ratio of outstanding debt to household wealth (assets)	Wealth (assets) can be difficult to assess
		Different types of wealth (assets) which can be adopted
Debt-service to income (DSTI)	The ratio of total (or minimum) borrowing repayments to household monthly income	Depending on the total income the same level of DSTI results in different risk of the financial strain
		Lump-sum repayments may disrupt the assessment
		Detailed information on minimum payments at micro-level are usually unavailable
Financial margin (FM)	The level of household income which is left after servicing debts and paying necessary living expenses	Data on detailed categories of expenses is limited
Below the poverty line (BPL)	Spending on total borrowing repayments take household below the poverty line	Even small debt repayments may result in crossing the poverty line
Number of loans (NL)	Having four or more credit commitments	In itself does not indicate over-indebtedness
Number of arrears (NA) and their duration	Being in arrears	May result from other than structural or long-term causes
Subjective burden of debt (SB)	Self-assessment of being over-indebted	Strong subjectivity bias
		The way of wording the question (definition of over-indebtedness) may significantly affect the results

a large number of outstanding small debts does not necessarily imply financial difficulties. Because of a high probability of obtaining "false-positive" results given by these indicators, they are used as a supplementary.

A subjective burden of debts indicator (SB) is used as a direct measure of the probability of falling into arrears based on how a household views of itself (Oxera 2004; Betti et al. 2007). This approach takes into consideration, the psychological load of having a debt and repaying it. Thus, over-indebted consumers are defined as those who consider themselves to be over-indebted. A subjective assessment of over-indebtedness is based on opinions and preferences of household members, but also—which is not taken into consideration while using objective indicators—their expectations regarding their future financial situation (Białowolski and Węziak-Białowolska 2014). Generally, over-indebted households are identified as those which express difficulty or serious difficulty in making debt payments. Other proposals of subjective measures of over-indebtedness can be found in Kempson (2002). Christelis et al. (2009) and McCarthy (2011) refer to the question of how to "make ends meet". Confidence in the ability to cope with unexpected expenses is another factor which can be analysed in the subjective assessment of over-indebtedness (Lusardi et al. 2011).

Not many studies of the over-indebtedness of households in Poland based on micro-level data have been conducted so far. The existing ones are mostly based on household budget surveys conducted by the government executive agency Statistics Poland (Główny Urząd Statystyczny—GUS) or the Polish representative household panel 'Social Diagnosis' (Białowolski and Węziak-Białowolska 2017; Białowolski 2019). Zajączkowski and Żochowski (2007) and Anioła-Mikołajczak (2017) use the DSTI ratio and the negative financial margin calculated on the basis of household budget surveys. Similarly, Wałęga and Wałęga (2018) apply a logit model to prove the relationship between households' socio-economic characteristics and the probability of their excessive debt. Świecka (2009) investigates delays in repayments using data obtained during 581 individual interviews. Data on household indebtedness in Poland can also be extracted from the EU-SILC database, but they are limited only to the DTI ratio of vulnerable households. The Household Finance and Consumption Survey conducted by the National Bank of Poland (NBP, i.e. Narodowy Bank Polski) and the EBC (European Central Bank) verify household indebtedness using the objective ratios DSTI, DTI and DTA (NBP 2017). Despite this interest, so far no one has conducted an in-depth analysis of over-indebtedness in Poland based on a complex set of measures.

3 Research Methods

Over-indebted households are selected from the sample with the use of the indicators proposed in literature (see Betti et al. 2007; D'Alessio and Iezzi 2016). Since each indicator reveals only one aspect of indebtedness using a combination of indicators might yield more accurate profile of households coping with a heavy debt burden (Brunetti et al. 2016; Ampudia et al. 2016). In accordance with Brunetti (2016), we use qualitative and quantitative indicators of financial malaise to assess indebtedness and perceived hardship.

Due to the flaws of the DTI ratio and a lack of data on the level of outstanding debt, we decide to use the DSTI ratio and the BPL measure. We classify households as over-indebted if spending on total borrowing repayments takes them below the poverty line (BPL), which equals to 60% of the median income using the modified OECD scale of

equivalence or if their debt-service to income ratio exceeds 30% (DSTI30). The adopted threshold for DSTI follows its value used by the NBP. We are unable to use the DTW (DTA) ratio and the financial margin because of a lack of the detailed data. In addition, the number of credit agreements (4 or more—NL4) and being more than 3 instalments in arrears (A3) are taken into account. A subjective burden of debt (SB) (the respondents' answers to the question whether they consider themselves to be over-indebted) is an additional indicator. We believe that such an approach allows obtaining a comprehensive profile of over-indebted households.

The segmentation of over-indebted households is conducted by using the classification and regression tree (CART) algorithm. This method is an alternative to many statistical techniques, such as multiple regression, logistic regression, or analysis of variance, used for exploring patterns in complicated datasets uncovered by linear models (De'ah and Fabricius 2000; Frisman et al. 2008). Tree-based methods are particularly popular in statistical data classification (Loh 2014) and are applied not only to economics (see, e.g. Williams et al. 1987; Keely and Tan 2008; Manasse and Roubini 2009; Galletta 2016; Bilton et al. 2017), insurance, and consumer credits, but also to the areas such as engineering, medicine, biology, and marketing (e.g. De'ah and Fabricius 2000; Dacko et al. 2016). We believe that the CART analysis has not been used to investigate the determinants of over-indebtedness so far.

The decision to employ the CART method is determined by the qualitative nature of the data obtained in the survey and its adequacy to the research problem. The CART analysis is highly effective in qualitative prediction in which the applicability of many other methods is limited.

The CART is a data mining technique which, by identifying patterns in data, selects the variables yielding the best prediction of individuals' types from among a set of explanatory variables (Galletta 2016). The first comprehensive study devoted to the classification tree algorithms was presented by Breiman et al. (1984), who introduced the CART algorithm. As a non-parametric approach without distributional assumptions, the CART can handle datasets containing variables of categorical, scale, and ordinal measurement types. Decision trees can perform well even if assumptions are somewhat violated by the dataset and they can also handle outliers, missing and unbalanced values in both response and explanatory variables (Low and Lai 2016). In comparison to linear and logit models, tree-based models can be visualised, more easily understood and interpreted when the predictors are a mix of numeric variables and factors. They require little data preparation whereas other techniques often require performing some operations (e.g. data normalization, creation of dummy variables, or the removal of their blank values). The CART results are invariant to monotone transformations of its independent variables. This algorithm performs an automatic variable selection and can establish interactions among variables (Sharma and Kumar 2016). One of CART method disadvantages is the fact that it splits only by one variable and decision trees may be unstable[1].

The proposed method allows generating decision trees. The dependent variable in the classification trees is measured on a weak scale (nominal or ordinal) and in case of regression trees—on a strong scale (at least an interval scale). The algorithm recursively partitions the data into nodes by iterated binary splits[2]. The root node (i.e. the whole sample) is,

[1] For a more detailed description of the CART method, its pros and cons and a comparison with other methods, see Feldman and Gross (2005) or Hastie et al. (2009).

[2] For a detailed technical explanation, see Breiman et al. (1984).

therefore, divided into other nodes (i.e. subsamples) by following a set of rules[3] which—from among all predictors—find the ones that allow for the most discriminative split (Poterie et al. 2019). The following classification function is used (Gatnar 2012):

$$y = \sum_{k=1}^{K} \alpha_k I\left(\mathrm{x}_i \in R_k\right) \tag{1}$$

where x_i is a multidimensional observation, R_k ($k = 1, \ldots, K$) are the subspaces (segments) of space X^m (m-dimensional variable space), α_k are the parameters of the model, while I is an index function that takes the value 1 (when $x_i \in R_k$) or 0 (when $x_i \notin R_k$).

If the dependent variable is a nominal variable, model (1) is called discriminatory and is represented by a classification tree. In this case, the parameters are determined as (Gatnar 2012):

$$\alpha_k = arg \max_s p\left(P_s|\mathrm{x}_i \in R_k\right) \tag{2}$$

where P_s ($s = 1, \ldots, u$) denotes the class to which the observation x_i belongs.

The classification is accomplished by testing the level of impurity of all possible splits. This procedure continues by creating branches and other nodes until certain conditions are met. A subset which does not split further is known as a terminal node or leaf. The terminal nodes define the predicted type for each individual whose characteristics match the traced path. The classification error, a Gini index, or entropy measure are most commonly used to assess the homogeneity of the subspace R_k in a classification tree. An undesirable phenomenon accompanying the construction of classification trees is the excessive complexity of the model, which is associated with an increase in the error value for the test set. When a full decision tree is built, it is usually necessary to prune some of its branches, which makes the results both easier to understand and more precise in classifying alternative data-sets (Han et al. 2011). V-fold cross validation is used to select the best-pruned tree, i.e. the one that is the least complex and whose cost of cross-validation is as close to minimal as possible (Breiman et al. 1984; Wu and Kumar 2009).

The interpretation of the CART results follows the classification rules created for each tree path linking the start and end nodes. Its purpose is to identify the combinations of predictive factors that determine the existence of a specific value of a dependent variable. The CART method made it possible to systematize the predictors (variables) in terms of their impact on the dependent variable. The importance of an attribute is based on the sum of the improvements in all nodes in which the attribute appears as a splitter. Predictor importance can be computed by summing—over all nodes in the tree(s)—the drop (delta) in node impurity (delta(I) for classification) and expressing these sums relative to the largest sum found over all predictors. The results are presented on a scale of 0–1 (Wu and Kumar 2009).

[3] The rules for the CART method concern (Manasse and Roubini 2009): (i) splitting each node into two child nodes; (ii) deciding when to stop growing the tree; and (iii) assigning each terminal node to a class outcome (e.g., over-indebted vs. not over-indebted).

4 Dataset and the Structure of the Sample

This study is based on the dataset obtained from the CATI survey conducted among Polish households in the second quarter of 2018. All the respondents were adults aged 18 years or over with at least one loan commitment (secured or unsecured). CEM (Market and Public Opinion Research Institute), a professional market and opinion research agency, partnered in the data collection phase of the survey. The initial sample consisted of almost 35,500 cell phone numbers (in 2018 96.7% of households in Poland possessed a cell phone) selected using random digit dialling. Finally 1107 individuals from all over Poland were interviewed, so the response rate for this study was 3.2% (calculated as the ratio of the number of completed telephone interviews to the number of all telephone contacts). It can be assumed that the data are representative of indebted households in Poland due to random sampling and the sample size. However, our dataset includes only indebted households so it cannot be compared with the EU-SILC or the nationwide Household Budget Survey.

The analysis addresses the situation of Polish indebted household. The dataset provides information on their indebtedness and attitudes towards debt as well as selected demographic characteristics. It includes the variables describing the respondents and their households: gender, age group, the level of education, the number of household members and the size of the place of residence. In order to evaluate the economic condition of a household, the following set of indicators is selected: the main source of income, a monthly net household income (in PLN), the level of debt repayment (in PLN). The number of loans, a subjective burden of debt, and the ability to repay debts by current income or savings are additional variables included in the dataset. The socio-economic characteristics of households are comprised in Table 2.

5 Results

5.1 Over-indebtedness in Poland

Table 3 demonstrates the percentage of over-indebted households in various socio-economic dimensions. On the basis of the NL4 and A3 measures it can be concluded that the problem of over-indebtedness among Polish households is not serious, as the percentage of households in arrears more than 3 instalments does not exceed several per cent. A small percentage of households with 4 or more loans confirms that they are coping with debt management adequately. It stays in line with previous studies on the Polish households in which the percentage of households with 2–3 instalments in arrears is estimated at 4% and the percentage of households against whom debt recovery proceedings have been commenced is 2% (Świecka 2009). The DSTI30 and BPL ratio show a more detailed profile of the households whose debt burden is considerable, but causing no serious financial difficulties and arrears.

Households in which the respondents are aged over 65 have the highest level of DSTI30 ratio (25.7%). The BPL ratio indicates a heavy debt burden also among the respondents aged 18–24. The older the respondents, the higher the percentage of those who report that their debts are a source of serious worries (SB).

The percentages of over-indebted households across the level of education are relatively constant but definitely higher among the respondents with the lowest education

Table 2 Socio-economic characteristic of the indebted households in the sample (%). *Source*: own calculations based on the CATI survey data

Specification	Structure
Gender	
Male	59.17
Female	40.83
Age of the respondent	
18–24	6.14
25–34	23.49
35–44	23.67
45–54	21.23
55–64	15.99
65 +	9.49
Level of education of the respondent	
Lower secondary or lower	4.07
Vocational	14.72
Secondary vocational	25.11
General secondary	16.71
Higher	39.39
Number of household members	
1	10.84
2	25.29
3	27.55
4	24.66
5	7.41
6 +	4.25
Place of residence	
Rural area	25.66
Town up to 50,000 inhabitants	28.65
Town with 50,000–100,000 inhabitants	15.96
City with 100,000–500,000 inhabitants	16.14
City over 500,000 inhabitants	13.60
Main source of income	
Self-employed	15.00
Blue-collar worker	33.06
White-collar worker	32.43
Working in agriculture	3.25
Pensioner	12.83
Disability pensioner	1.63
Household net monthly income*	
Up to 1500 PLN	6.34
1501–2000 PLN	9.12
2001–3000 PLN	15.66
3001–4000 PLN	17.34
4001–6000 PLN	28.05
6001–8000 PLN	10.51
Above 8000 PLN	12.98
Source of debt	
Banks	68.20

Table 2 (continued)

Specification	Structure
Instalments	34.87
Workplace	10.93
Family/friends	5.60
Other financial institutions and payday	3.70
Monthly debt repayments	
Up to 200 PLN	20.86
(200 PLN—500 PLN >	28.05
(500 PLN—1000 PLN >	24.23
(1000 PLN—2000 PLN >	20.04
Over 2000 PLN	6.83

*The average household net monthly income in the sample is 5499 PLN and the median—4500 PLN. The average equivalised monthly income is 2958 PLN and the median—2381 PLN

level. Those with higher education are characterised by a relatively low risk of over-indebtedness (DSTI30—16%; BPL—13%), and a similar pattern emerges for the SB indicator.

Generally, according to the DSTI30 measure over-indebtedness decreases when the size of the household increased. However, the opposite pattern is observed when the BPL indicator is taken into consideration—the level of over-indebtedness between a household with one and five or more persons is 1.4–2.6 times higher than in case of other households.

The results reveal a strong impact of income on over-indebtedness risk. The percentage of households with the highest income classified as over-indebted is remarkably lower both in case of the objective indicators (DSTI30 and BPL) and the subjective measure (SB).

As far as the main source of income is concerned, three groups of households in Poland—those of pensioners, disability pensioners and people living off agriculture—are at higher risk of over-indebtedness than other groups.

As a rule, the level of over-indebtedness increases with the level of debt repayment, but surprisingly, according to the BPL measure, in our study the tendency is opposite—over-indebtedness decreases with the level of debt repayment—below 200 PLN: 34.1% of over-indebted households, followed by 30.6% with debt repayment between 500 and 1000 PLN being over-indebted, while 20.3% per cent are over-indebted households among those with debt repayment over 2000 PLN. It can be speculated that low-income households that incur loans for small amounts also carry a low burden in nominal terms. In this case, however even small nominal repayments significantly increase the risk of falling below the poverty line.

Using a cross-sectional analysis a general profile of over-indebted households in Poland can be built. Over-indebtedness affects primarily such households in which the age of the reference person is between 45 and 54 or 65 +, and this person is poorly educated. Among the most vulnerable groups of households are those that make a living off agriculture or are pensioners. Households with more than four members from the rural areas are over-indebted more frequently than other types of households. Whether a household is over-indebted to a great extent depends on its economic situation—the most at risk are the ones with a monthly net household income up to 3000 PLN spending more than 2000 PLN per month on debt repayments.

Table 3 Socio-economic characteristics of the over-indebted households (%), 2018. *Source*: own calculations based on the CATI survey data

Specification	NL4	A3	DSTI30	BPL	SB
Age of the respondent					
18–24	2.94	1.47	10.29	35.48	7.35
25–34	4.62	1.54	12.69	25.21	14.23
35–44	6.11	2.29	16.41	21.99	17.56
45–54	5.53	1.28	20.00	37.26	18.72
55–64	3.39	3.95	15.82	33.33	16.38
65 +	1.90	0.00	25.71	45.74	26.67
Level of education of the respondent					
Lower secondary or lower	6.67	4.44	35.56	78.05	31.11
Vocational	3.07	3.68	16.56	46.45	20.25
Secondary vocational	2.52	2.16	15.47	35.71	15.83
General secondary	4.86	2.70	16.76	38.37	16.76
Higher	6.19	0.46	15.60	12.92	15.37
Number of household members					
1	0.00	3.33	30.00	38.05	21.67
2	3.57	1.07	15.36	26.27	15.00
3	5.25	1.97	17.70	28.21	19.67
4	6.96	1.47	13.92	25.31	15.38
5	6.10	1.22	13.41	44.74	19.51
6 +	2.13	6.38	6.38	65.79	6.38
Place of residence					
Rural area	4.24	2.47	20.14	40.77	18.37
Town up to 50,000 inhabitants	3.80	0.95	12.97	29.18	13.29
Town with 50,000–100,000 inhabitants	3.98	2.27	17.05	30.00	20.45
City with 100,000–500,000 inhabitants	7.30	1.69	16.29	32.30	15.17
City over 500,000 inhabitants	4.67	2.00	18.67	14.18	20.67
Monthly net household income					
Up to 1500 PLN	3.17	6.35	23.81	98.43	30.16
1501–2000 PLN	3.26	2.17	32.61	91.21	20.65
2001–3000 PLN	3.16	2.53	20.25	62.66	20.89
3001–4000 PLN	3.43	0.57	13.71	28.57	16.00
4001–6000 PLN	4.24	2.12	18.02	4.96	13.78
6001–8000 PLN	6.60	1.89	8.49	0.00	18.87
Above 8000 PLN	9.92	0.00	7.63	0.76	12.98
Main source of income					
Self-employed	3.61	1.81	15.66	21.77	13.86
Blue-collar worker	4.64	2.19	15.30	38.90	18.31
White-collar worker	5.85	0.84	14.48	14.06	13.09
Working in agriculture	8.33	2.78	27.78	58.06	22.22
Pensioner	1.41	1.41	23.24	44.53	23.24
Disability pensioner	11.11	5.56	22.22	66.67	27.78
Monthly debt repayments					
Up to 200 PLN	0.44	1.31	0.44	34.10	3.93
(200 PLN—500 PLN >	1.62	1.95	5.19	34.63	14.94

Table 3 (continued)

Specification	NL4	A3	DSTI30	BPL	SB
(500 PLN—1000 PLN>	4.51	1.88	16.17	30.61	19.92
(1000 PLN—2000 PLN>	7.73	2.27	38.18	25.38	25.45
Over 2000 PLN	21.33	2.67	54.67	20.31	32.00

Table 4 Matrix of over-indebtedness indicators in Poland in 2018 (%). *Source*: own calculations based on the CATI survey data

Specification	NL4	A3	DSTI30	BPL	SB
NL4	4.61	0.63	1.69	1.49	2.26
A3	0.63	1.9	0.50	0.79	1.36
DSTI30	1.69	0.50	16.71	11.72	7.25
BPL	1.49	0.79	11.72	30.78	8.34
SB	2.26	1.36	7.25	8.34	17.07

The marginal frequencies are along the diagonal

Due to the differences in over-indebtedness across single measures, we perform a more complex assessment using a mix of indicators. Table 4 presents the matrix of over-indebtedness indicators calculated for the Polish households. The analysis reveals that 43.5% of the households which took part in the survey are over-indebted according to at least one of five indicators, 17.4% according to at least two indicators simultaneously, and 6.3% according to at least three indicators. Only 0.9% of households turn out over-indebted according to four or five indicators.

The BPL indicator reveals that 30.8% of households in debt are classified as over-indebted, while their respective percentage identified by the DSTI30 is only 16.7% and drops to 11.7% when the DSTI30 and BPL ratios are combined.

Adding one more dimension—the subjective perception of a debt burden (SB)—to the assessment of over-indebtedness reduces the percentage of over-indebtedness by up to three times. 8.3% of households are classified as over-indebted by the BPL and SB indicators and 7.3% by the SB and DSTI30. In other compositions of two indicators applied simultaneously, the percentage of over-indebted households in the sample does not exceed 2.3%.

5.2 CART Analysis

We use the CART algorithm to determine the impact of socio-economic characteristics of households on the probability of classifying a given household as over-indebted. The CART model is built using STATISTICA ver. 13.1 software. We create separate models for each over-indebtedness measure. The V-fold cross-validation at $v = 10$ (typical value) proves that the best models (and therefore used in analysis) are based on the DSTI30 ratio and the BPL indicator. Its value for the first model is 0.1704, and for the second one 0.1811.

The number of categories for each independent variable is limited in order to preserve the compact size of the tree. The following socio-economic characteristics are treated as independent variables in the analysis: A monthly net income of the household (2000 PLN

Table 5 The importance of the CART predictors. Source: own calculations based on the CATI survey data.

Specification	DSTI30 (model 1)	BPL (model 2)
Age of respondent	1.00	0.43
Place of residence	0.71	0.30
Gender	0.53	0.05
Monthly net income	0.23	1.00
Number of household members	0.21	0.23
Level of education	0.20	0.66

or less, 2000–6000 PLN and above 6000 PLN), the place of residence: A rural area, a town (up to 100,000 residents), a city (over 100,000 residents), the number of household members (up to 3 persons and over 3 persons), the age of the respondents (three age groups: up to 34, 35–54 and 55 or more), the level of education (three levels: vocational at most, secondary, and higher), and gender.

We assess the validity of the predictors in model 1 and model 2. The ranking of the importance of variables is presented in Table 5. The ranking obtained in two models differ: the age of the respondent and the place of residence exert the greatest impact on assigning a household to the group of over-indebted households in model 1 while a monthly net income and the level of education have the greatest influence in model 2. The number of household members has limited importance in both models.

Model 1 is based on the CART with the DSTI30 dependent variable. The first split depends on the household income and suggests that households with a low income (below 2000 PLN) are classified as over-indebted (37.8%) with higher probability than households with a high income (14.8%).

The next splitting variable on the left branch is the respondents' age. Young people (up to 34 years old) are grouped against those between 35–54 and 55 or older. The households in the age groups over 34 are subsequently divided depending on the place of residence. The probability of being over-indebted is higher among households from rural areas than among households from cities.

The right branch of the tree demonstrates that households belonging to the income group 2000–6000 PLN is characterised by 17% probability of being over-indebted, while those with income higher than 6000 PLN—by only 8% probability. The respondents' education plays a significant role in determining over-indebtedness in middle-income groups.

Homogeneous subsets resulting from the CART splitting may be treated as subsets of households classified as over-indebted depending on their socio-economic characteristics. The results of the CART analysis can be interpreted for selected segments of households designated by the end nodes of the tree (Fig. 1):

- the highest probability (0.59) that the household is over-indebted is found among the households located in a rural area with income up to 2000 PLN with the reference person at the age over 34 (node ID 6),
- the relatively high probability (0.42) of over-indebtedness is also found among middle-income and well-educated households headed by men, located in rural areas or in a city and consisting of up to three members (node ID 16),
- if the household income is above 6000 PLN (nodes ID 28 and 29), the probability of being over-indebted is extremely low but slightly higher for households located in rural areas,

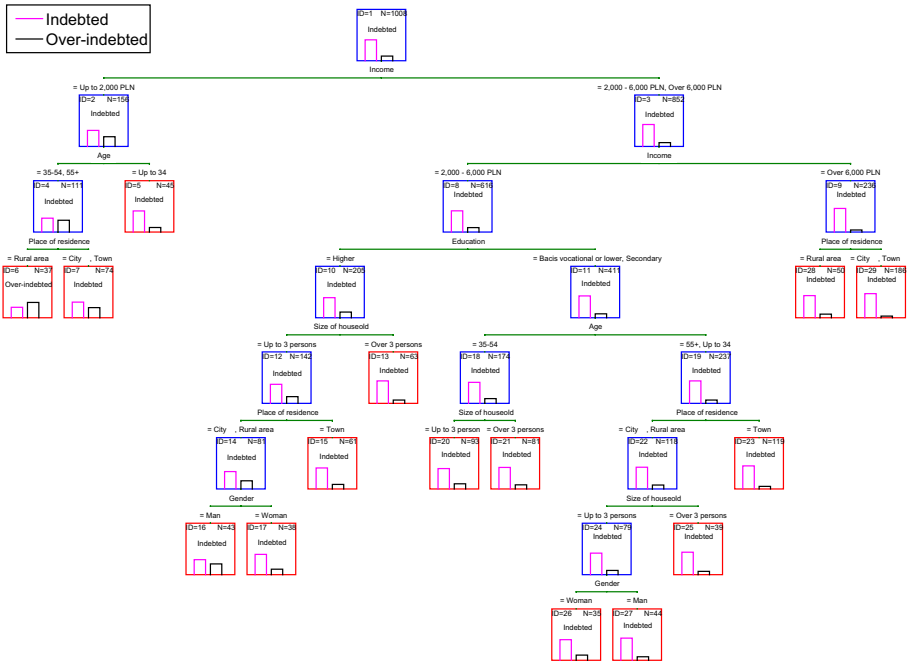

Fig. 1 The CART with the DSTI30 dependent variable (model 1). *Source*: Own elaboration based on the CATI survey data

- the lowest probability (0.10) of being classified as over-indebted is found among middle-income households and those located in a town headed by either a young (up to 34) or 55 + person with secondary or lower education (node ID 23).

The tree in model 2 is generated for the BPL dependent variable and—after pruning—consists of 5 divided nodes and 6 end nodes (Fig. 2). The same set of independent variables as in model 1 is taken into account, however, ultimately the classification tree in this model is slimmer. Only four variables (an income, the number of household members, education and age) turn out important for the segmentation of over-indebted households.

Some interpretations of the household segments designated by the end nodes of the CART algorithm (model 2) are as follows:

- if the household has a monthly income of up to 2000 PLN (node ID 2), the probability of being classified as over-indebted is 0.94; while among the high-income households (a monthly income above 6000 PLN), the probability of being over-indebted is almost 0 (node ID 5);
- the substantial probability that the household will be classified as over-indebted is found in case of poorly-educated households headed by a person older than 34 and belonging to the middle-income group, with more than three household members (node ID 10).

In both trees the first split depends on the household income. In the BPL model (in the group of households with an income of up to 2000 PLN and above 6000 PLN) the

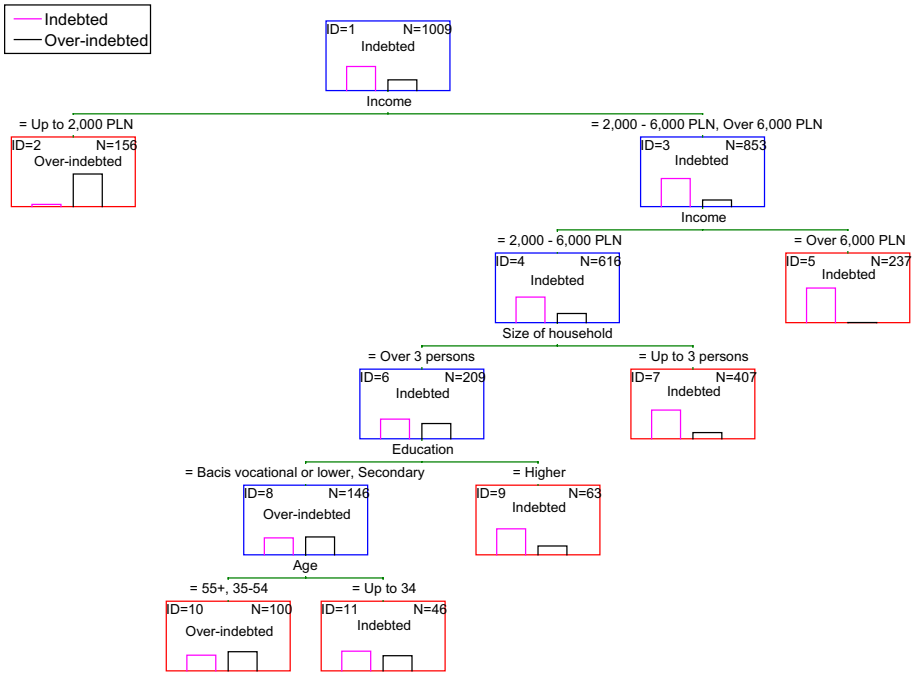

Fig. 2 The CART with the BPL dependent variable (model 2). *Source*: Own elaboration based on the CATI survey data

classification of households as over-indebted is not affected by other variables. In both models the household size, the education level, and the age of the reference person have a significant impact on the classification of households as over-indebted in the group of households with a middle-income (2000 PLN–6000 PLN).

6 Discussion

Although the individual level of a debt burden of Polish households increased in the last decade, the results of our study indicate that the risk of over-indebtedness is—on average—still relatively low. A complex assessment based on a matrix of indicators reveals that taking into account more than one indicator significantly lowers the number of households classified as over-indebted. It can be argued that specific over-indebtedness measures do not capture the same set of households and refer to different dimensions of over-indebtedness. Thus, the proper selection of the indicator or indicators depending on the adopted definition of over-indebtedness becomes an issue.

Apart from several aggregative analyses of over-indebtedness in Poland, there are no studies examining a debt burden by employing different measures in different types of households. Unfortunately, a lack of comparative data for other countries, especially the analyses of their over-indebtedness based on the micro-level data, makes comparisons difficult.

Only a minority of households could be considered over-indebted when arrears are taken into account. This is in line with the results from previous, fragmentary studies of Polish households (Świecka 2009; Anioła-Mikołajczak 2017). The percentage of over-indebted households in relation to the DSTI30 obtained in our study is similar to the results presented by Świecka (2016) and by the NBP (2017). However, these results are not fully comparable due to differences in survey methodology and because our sample includes only indebted households.

Using the BPL indicator in the assessment of over-indebtedness allows obtaining an insight into Polish households. The debt of households should be considered not only in economic or psychological terms, but also in social terms. Firstly, the share of households classified as over-indebted by the BPL indicator is 2–3 times higher than those classified as such by the DSTI30 ratio. The results are consistent with these regarding the Italian (D'Alessio and Iezzi 2016) and South African (Ntsalaze and Ikhide 2016) households. The discrepancies stem from the fact that each indicator refers to a different aspect of over-indebtedness, which on the whole, confirms that it is a complex phenomenon (Coin et al. 2013; D'Alessio and Iezzi 2013; Disney et al. 2008). Secondly, the most remarkable result of our analysis is that the DSTI30 ratio strongly underestimates the extent of over-indebtedness especially in vulnerable groups of households in comparison with the BPL. This is noticeable in the case of the youngest and the oldest age cohorts, the poorly-educated persons, and either the lowest income groups or debt repayments. The same situation is observed in households with 5 or more members. The problem of over-indebtedness which leads to poverty is noticed among disability pensioners and people living off agriculture. Thus, even relatively low debt repayments can have an adverse impact on the financial situation of their household and can reduce its ability to meet basic needs. This finding may suggest that over-indebtedness and poverty affect the same group of households. Nevertheless, it needs to be interpreted with caution due to potential reverse causality. In our opinion, however, this indicates the need for a further in-depth analysis of the social dimension of over-indebtedness, because typical income-based measures (DTI, DSTI) are not able to capture this problem.

We are also aware that assessing of over-indebtedness by the DSTI30 ratio or by the BPL indicator require caution. Heavily indebted households with high repayments may experience a much greater subjective debt burden (and therefore the consequences associated with it) than indicated by the objective measures. The high level of repayment burden is primarily psychological. This confirms the results of previous studies (Hojman et al. 2016) which report that being in debt is a mental burden and is associated with stress. Thus, it is advisable to include a subjective indicator in the over-indebtedness assessment of these households. In turn, combining the BPL and A3 may be better measures of economic distress of over-indebtedness for low-income households. We highlight the need to analyse the indicators jointly, as the proper identification of the level of over-indebtedness seems to require using different indicators depending on the socio-economic group analysed.

The CART model allows us to identify the dependencies within the set of indebted households. The results, described by the dividing nodes, are visualised and can be more easily understood and interpreted than when other methods of modelling over-indebtedness are used. Although, the CART models do not provide a p value to test the significance of variables, but it is still possible to examine the importance of particular variables and the order of their interactions. We find that a monthly net income, the level of education and the age of the reference person treated as predictors in both our CART models play a crucial role in the differentiation of indebted and over-indebted households. This stays in

line with the results of other studies based on the dynamic probit model (Chichaibelu and Waibel 2018) or logit model (Anioła-Mikołajczak 2016).

Our study provides considerable insight into the age structure of over-indebted households in Poland. It differs from such structure in countries with a wider access to consumer credits (Cox et al. 2002; Betti et al. 2007). We find that young–age groups in Poland are less likely to be over-indebted as indicated by the DSTI30 ratio and the SB indicator. It can be explained by the still less developed credit market in Poland and its major imperfections. In this respect, Poland seems to be rather similar to countries with a relatively low level of household income, like Slovakia or Slovenia, where widespread borrowing is not very popular (Grejcz and Żółkiewski 2017). Similar conclusions regarding countries with the more restricted (or less developed) consumer credit markets are reported by Betti et al. (2007). On the other hand, the greatest number of over-indebted Polish households as such by the DSTI30, BPL and SB is found in old-age cohorts (55–64 and 65 +), which are characterised by a lower income and usually a higher consumption/income ratio (following the U-shape age profile of this ratio). Therefore, they will tend to increase the burden of repayments on income (higher the DSTI30 ratio and the subjective burden perception SB). These findings contradict the results reported by D'Alessio and Iezzi (2016) for Italian households with heads over 65, Faruqui (2008) for Canadian case and by Haq et al. (2018b) for Pakistan.

We demonstrate that the probability of being over-indebted is substantially higher among poorly-educated households than among the households with a well-educated reference person. This finding supports the importance of financial knowledge in the occurrence of over-indebtedness (Campbell 2006; French and McKillop 2016). This dependence can be explained by the fact that better-educated persons have a greater ability to evaluate and foresee their economic capacity to repay the debts (Disney and Gathergood 2011; Białowolski et al. 2019). However, it is fundamental to notice that a significant percentage of poorly-educated group of households matches low-income one. This is especially confirmed by the percentage of poorly-educated and over-indebted households classified by the BPL indicator (almost 80%).

As expected, high incomes limit the percentage of households that can be classified as over-indebted (based on the DSTI30), which also refers to the subjective perception of a debt burden. It stays in line with the results obtained for Italian households (D'Alessio and Iezzi 2016; Giarda 2013). Additionally, our study finds a positive relationship between a household income and the number of credit commitments. Similar results are obtained by Haq et al. (2018a).

The psychological debt burden is exceptionally strongly felt by the elderly and the poorly educated. Interestingly, in these groups of households the percentage of households that feel subjectively over-indebted is significantly higher than the percentage of households on the basis of objective indicators. The finding that the perception of debt in this group of households is particularly acute is supported by previous studies of Drentea and Lavrakas (2000) and Melzer (2011).

Our empirical study of over-indebtedness finds that objective indicators involving debt service to income align fairly well with the subjective perceptions of burden. It is in line with the results reported by e.g. Rinaldi and Sanchis-Arellano 2006; Keese 2012; D'Alessio and Iezzi 2013; Chichaibelu and Waibel 2018.

Our study is not free from certain limitations. First of all, the set of microdata on household indebtedness that we use is only available as a cross-section. To gain more insights on this topic, time series or panel data could be used. Our survey is conducted using the CATI method (cell phone numbers). The main concern in this type of surveys is usually

a high non-response rate. Nevertheless, recent studies demonstrate that this rate is not as important a measure of survey data quality as it was once thought (Keeter et al. 2006). Despite these problems, and bearing in mind that similar surveys have not been conducted in Poland so far, we believe that the sample size is large enough to justify the application of our findings to the general population of indebted households.

Secondly, over-indebtedness measures have some limitations. As noted by Betti et al. (2007), it is difficult to define an optimal level of indebtedness, as the level which leads to over-indebtedness depends on particular circumstances or a particular stage of the lifecycle. Moreover, some of the available indicators, mainly those subjective ones based on the responses to questions about economic difficulties may be affected by rather strong subjectivity bias (Brunetti et al. 2016).

Unfortunately, it is not possible to investigate and estimate the number of households that face a significant risk of becoming over-indebted in a representative manner. Using survey data at the household level allows assessing households which are currently in arrears. However, these data need to be interpreted with caution, as the extent of the problem may be underestimated. On the one hand, some households in arrears are misclassified as over-indebted because their financial problems may result from forgetfulness rather than structural problems. On the other hand, the subjective measures can lead to underestimation due to the inability of certain households to assess their financial situation correctly.

7 Conclusions

Using micro-level data in assessing over-indebtedness of households allows us to shed more light on the vulnerabilities in this sector in Poland. Our study is based on a unique dataset obtained from the CATI survey conducted among indebted Polish households in 2018. Our study is the very first which examines a debt burden of Polish households by employing different measures.

In general, our study indicates that the over-indebtedness in Poland concerns relatively low fraction of households. As stated in the Introduction, we use the BPL indicator for the first time in the assessment of over-indebtedness in Poland. We find it valuable with reference to the social dimension of over-indebtedness. In our view, the results demonstrate that using the BPL indicator helps to overcome the limitations of the DSTI30, particularly the ones which affect underestimating of over-indebtedness in vulnerable groups of households.

In this paper we concisely introduce the CART method, its main advantages and disadvantages, and guidelines for its implementation as a classification tool. We apply the classification trees constructed by combining the data on socio-economic characteristics of indebted households for the DSTI30 and BPL indicators of over-indebtedness. To the best of our knowledge, this is the first time when this tool is used to obtain the profile of over-indebted households. The use of the CART method identified the most important socio-economic household characteristics that increase the probability of being over-indebted.

We agree with Betti et al. (2007) that a widely accepted and accurate definition of consumer over-indebtedness is still to be provided, which is also true about a consensus on how to measure it or where to draw the line between ‘normal’ debt and over-indebtedness. The fact that there is no simple aggregate measure of ‘normal’ or ‘excessive’ consumer indebtedness justifies the application of a multi-indicator approach in assessing

over-indebtedness. We try to follow this need by creating a matrix of over-indebtedness indicators.

The results of our study call for systematic monitoring of over-indebtedness risks, which might appear particularly in vulnerable, low income households. The study indicates several important implications that can be of interest to policy makers. The protection of households against over-borrowing requires adequate financial education, provided especially for those most poorly educated and from the lowest income classes. Thus it is highly recommended to design effective support instruments preventing over-indebted households from falling into poverty or bankruptcy.

Acknowledgments This study was supported by funds from the National Science Centre (NCN, Poland) through grant no. 2015/19/D/HS4/02569.

References

Ampudia, M., van Vlokhoven, H., & Żochowski, D. (2016). Financial fragility of euro area households. *Journal of Financial Stability, 27*, 250–262. https://doi.org/10.1016/j.jfs.2016.02.003.

Anderloni, L., Bacchiocchi, E., & Vandone, D. (2012). Household financial vulnerability: An empirical analysis. *Research in Economics, 66*(3), 284–296. https://doi.org/10.1016/j.rie.2012.03.001.

Anderson, G., Bunn, P., Pugh, A., & Uluc, A. (2016). The Bank of England/NMG Survey of household finances. *Fiscal Studies, 37*(1), 131–152. https://doi.org/10.1111/j.1475-5890.2016.12091.

Angel, S., & Heitzmann, K. (2015). Over-indebtedness in Europe: The relevance of country-level variables for the over-indebtedness of private households. *Journal of European Social Policy, 25*(3), 331–351. https://doi.org/10.1177/0958928715588711.

Anioła-Mikołajczak, P. (2016). Over-indebtedness of households in Poland and its determinants. *Acta Scientiarum Polonorum. Oeconomia, 15*(4), 17–26.

Anioła-Mikołajczak, P. (2017). The impact of age on polish households financial behavior–indebtedness and over-indebtedness. *Optimum. Economic Studies, 85*(1), 106–116.

Bańkowska, K., Lamarche, P., Osier, G., & Pérez-Duarte, S. (2015). Measuring household debt vulnerability in the euro area. *Indicators to support monetary and financial stability analysis: data sources and statistical methodologies, 39*, 1–25.

Barba, A., & Pivetti, M. (2009). Rising household debt: Its causes and macroeconomic implications—a long-period analysis. *Cambridge Journal of Economics, 33*(1), 113–137. https://doi.org/10.1093/cje/ben030.

Betti, G., Dourmashkin, N., Rossi, M., & Ping Yin, Y. (2007). Consumer over-indebtedness in the EU: measurement and characteristics. *Journal of Economic Studies, 34*(2), 136–156. https://doi.org/10.1108/01443580710745371.

Bettocchi, A., Giarda, E., Moriconi, C., Orsini, F., & Romeo, R. (2018). Assessing and predicting financial vulnerability of Italian households: a micro-macro approach. *Empirica, 45*(3), 587–605. https://doi.org/10.1007/s10663-017-9378-2.

Białowolski, P. (2019). Patterns and evolution of consumer debt: evidence from latent transition models. *Quality and Quantity, 53*(1), 389–415. https://doi.org/10.1007/s11135-018-0759-9.

Białowolski, P., & Węziak-Białowolska, D. (2014). The index of household financial condition, combining subjective and objective indicators: An appraisal of Italian households. *Social Indicators Research, 118*(1), 365–385. https://doi.org/10.1007/s11205-013-0401-0.

Białowolski, P., & Węziak-Białowolska, D. (2017). What does a Swiss franc mortgage cost? The tale of Polish trust for foreign currency denominated mortgages: Implications for well-being and health. *Social Indicators Research, 133*(1), 285–301. https://doi.org/10.1007/s11205-016-1363-9.

Białowolski, P., Cwynar, A., Cwynar, W., & Węziak-Białowolska, D. (2019). Consumer debt attitudes: The role of gender, debt knowledge and skills. *International Journal of Consumer Studies*. https://doi.org/10.1111/ijcs.12558.

Bilton, P., Jones, G., Ganesh, S., & Haslett, S. (2017). Classification trees for poverty mapping. *Computational Statistics and Data Analysis, 115*, 53–66.

Bourova, E., Ramsay, I., & Ali, P. (2019). The experience of financial hardship in Australia: Causes, impacts and coping strategies. *Journal of Consumer Policy, 42*(2), 189–221. https://doi.org/10.1007/s10603-018-9392-1.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (Eds.). (1984). *Classification and regression trees (the wadsworth statistics/probability series)*. NY: Chapman and Hall.

Bridges, S., & Disney, R. (2004). Use of credit and arrears on debt among low-income families in the United Kingdom. *Fiscal Studies, 25*(1), 1–25. https://doi.org/10.1111/j.1475-5890.2004.tb00094.x.

Brown, S., & Taylor, K. (2008). Household debt and financial assets: evidence from Germany, Great Britain and the USA. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 171*(3), 615–643. https://doi.org/10.1111/j.1467-985X.2007.00531.x.

Brunetti, M., Giarda, E., & Torricelli, C. (2016). Is financial fragility a matter of illiquidity? An appraisal for Italian households. *Review of Income and Wealth, 62*(4), 628–649. https://doi.org/10.1111/roiw.12189.

Campbell, J. Y. (2006). Household finance. *The Journal of Finance, 61*(4), 1553–1604. https://doi.org/10.1111/j.1540-6261.2006.00883.x.

Chichaibelu, B. B., & Waibel, H. (2018). Over-indebtedness and its persistence in rural households in Thailand and Vietnam. *Journal of Asian Economics, 56*, 1–23. https://doi.org/10.1016/j.asieco.2018.04.002.

Christelis, D., Jappelli, T., Paccagnella, O., & Weber, G. (2009). Income, wealth and financial fragility in Europe. *Journal of European Social Policy, 19*(4), 359–376. https://doi.org/10.1177/1350506809341516.

Coin, D., Vacca, V., Conti, AM., Leva, L., Liberati, D., Manzoli, M., Marangoni, D., Mocetti, S., Saporito, G., Sironi, L. (2013) Households indebtedness and financial vulnerability in the Italian regions, *Questioni di Economia e Finanza* (Occasional Papers) 163, Bank of Italy, Economic Research and International Relations Area.

Coletta, M., De Bonis, R., & Piermattei, S. (2019). Household debt in OECD countries: The Role of supply-side and demand-side factors. *Social Indicators Research, 143*(3), 1185–1217. https://doi.org/10.1007/s11205-018-2024-y.

Cox, P., Whitley, J., & Brierley, P. (2002). Financial pressures in the UK household sector: Evidence from the British household panel survey. *Bank of England Quarterly Bulletin*. Available at: https://ssrn.com/abstract=708984. Accessed 25 Feb 2019.

D'Alessio, G., & Iezzi, S. (2013). Household over-indebtedness: Definition and measurement with Italian data. *Bank of Italy Occasional Paper*, (149). https://doi.org/10.2139/ssrn.2243578.

D'Alessio, G., & Iezzi, S. (2016). Over-indebtedness in Italy: How widespread and persistent is it? *Bank of Italy Occasional Paper*, (349). https://doi.org/10.2139/ssrn.2243578.

Dacko, M., Zając, T., Synowiec, A., Oleksy, A., Klimek-Kopyra, A., & Kulig, B. (2016). New approach to determine biological and environmental factors influencing mass of a single pea (*Pisum sativum* L.) seed in Silesia region in Poland using a CART model. *European Journal of Agronomy, 74*, 29–37. https://doi.org/10.1016/j.eja.2015.11.025.

De ah, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology, 81*(11), 3178–3192. https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2.

del Rio, A., & Young, G. (2008). The impact of unsecured debt on financial pressure among British households. *Applied Financial Economics, 18*(15), 1209–1220. https://doi.org/10.1080/09603100701604233.

Disney, R., & Gathergood, J. (2011). Financial literacy and indebtedness: New evidence for UK consumers. *Discussion Papers 11/05*, University of Nottingham, centre for finance, credit and macroeconomics (CFCM).

Disney, R., Bridges, S., & Gathergood, J. (2008). Drivers of over-indebtedness. Report to the UK department for business.

Drentea, P., & Lavrakas, P. J. (2000). Over the limit: The association among health, race and debt. *Social Science & Medicine, 50*(4), 517–529. https://doi.org/10.1016/S0277-9536(99)00298-1.

European Commission (2008), Towards a common operational European definition of over-indebtedness, Directorate-General for Employment Social, Affairs and Equal Opportunities.
European Commission (2010). Over-indebtedness: New Evidence from the EU-SILC Special Module. Research Note 4/2010, Brussels.
Faruqui, U. (2008). *Indebtedness and the household financial health: an examination of the Canadian debt service ratio distribution* (No. 2008, 46). Bank of Canada Working paper.
Feldman, D., & Gross, S. (2005). Mortgage default: Classification trees analysis. *The Journal of Real Estate Finance and Economics, 30*(4), 369–396. https://doi.org/10.1007/s11146-005-7013-7.
Ferretti, F., & Vandone, D. (2019). *Personal debt in Europe: The EU financial market and consumer insolvency*. Cambridge: University Press.
French, D., & McKillop, D. (2016). Financial literacy and over-indebtedness in low-income households. *International Review of Financial Analysis, 48*, 1–11. https://doi.org/10.1016/j.irfa.2016.08.004.
French, D., & Vigne, S. (2018). The causes and consequences of household financial strain: A systematic review. *International Review of Financial Analysis*. https://doi.org/10.1016/j.irfa.2018.09.008.
Frisman, L., Prendergast, M., Lin, H. J., Rodis, E., & Greenwell, L. (2008). Applying classification and regression tree analysis to identify prisoners with high HIV risk behaviors. *Journal of Psychoactive Drugs, 40*(4), 447–458. https://doi.org/10.1080/02791072.2008.10400651.
Galletta, S. (2016). On the determinants of happiness: A classification and regression tree (CART) approach. *Applied Economics Letters, 23*(2), 121–125. https://doi.org/10.1080/13504851.2015.1054066.
Gathergood, J. (2012). Debt and depression: causal links and social norm effects. *The Economic Journal, 122*(563), 1094–1114. https://doi.org/10.1111/j.1468-0297.2012.02519.x.
Gatnar, E. (2012). Drzewa klasyfikacyjne i regresyjne – podejście wielomodelowe. In M. Walesiak & E. Gatnar (Eds.), *Statystyczna analiza danych z wykorzystaniem programu R* (pp. 238–259). Warszawa: Wydawnictwo Naukowe PWN.
Giarda, E. (2013). Persistency of financial distress amongst Italian households: Evidence from dynamic models for binary panel data. *Journal of Banking and Finance, 37*(9), 3425–3434. https://doi.org/10.1016/j.jbankfin.2013.05.005.
Girouard, N., Kennedy, M., & André, C. (2006). *Has the Rise in debt made households more vulnerable? OECD economics department*. Paris: OECD Publishing.
Grejcz, K., & Żółkiewski, Z. (2017). Household wealth in Poland: The results of a new survey of household finance. *Bank i Kredyt, 48*(3), 295–326.
Haas, O. J. (2006). *Overindebtedness in Germany*. Employment sector, social finance program, working paper No. 44. Geneva: International labour office.
Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and techniques* (3rd ed.). San Francisco: Morgan Kaufmann Publishers.
Haq, W., Ismail, N. A., & Satar, N. M. (2018a). Household debt in different age cohorts: A multilevel study. *Cogent Economics and Finance, 6*(1), 1455406. https://doi.org/10.1080/23322039.2018.1455406.
Haq, W., Ismail, N. A., & Satar, N. M. (2018b). Investigation of household debt through multilevel multivariate analysis: Case of a developing country. *Journal of Reviews on Global Economics, 7*, 297–316. https://doi.org/10.6000/1929-7092.2018.07.26.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer-Verlag. https://doi.org/10.1007/978-0-387-84858-7.
Hiilamo, H. (2018). *Household debt and economic crises: Causes, consequences and remedies*. Cheltenham: Edward Elgar Publishing. https://doi.org/10.4337/9781785369872.
Hojman, D. A., Miranda, Á., & Ruiz-Tagle, J. (2016). Debt trajectories and mental health. *Social Science and Medicine, 167*, 54–62. https://doi.org/10.1016/j.socscimed.2016.08.027.
Household Finance and Consumption Network (HFCN). (2009). Survey Data on Household Finance and Consumption. Research Summary and Policy Use. ECB Occasional Paper No. 100. January 2009.
Hyytinen, A., & Putkuri, H. (2018). Household optimism and overborrowing. *Journal of Money, Credit and Banking, 50*(1), 55–76. https://doi.org/10.1111/jmcb.12453.
Jappelli, T., Pagano, M., & Di Maggio, M. (2013). Households' indebtedness and financial fragility. *Journal of Financial Management, Markets and Institutions, 1*(1), 23–46. https://doi.org/10.12831/73631.
Karwowski, E., Shabani, M., & Stockhammer, E. (2019). Dimensions and determinants of financialisation: Comparing OECD countries since 1997. *New Political Economy, 25*, 1–21. https://doi.org/10.1080/13563467.2019.1664446.
Keely, L. C., & Tan, C. M. (2008). Understanding preferences for income redistribution. *Journal of Public Economics, 92*(5–6), 944–961. https://doi.org/10.1016/j.jpubeco.2007.11.006.
Keese, M. (2009). Triggers and determinants of severe household indebtedness in Germany. *Ruhr Economic Paper*, (150). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1518373. Accessed 15 May 2019.

Keese, M. (2012). Who feels constrained by high debt burdens? Subjective vs objective measures of household debt. *Journal of Economic Psychology, 33*(1), 125–141. https://doi.org/10.1016/j.joep.2011.08.002.

Keese, M., & Schmitz, H. (2014). Broke, ill, and obese: Is there an effect of household debt on health? *Review of Income and Wealth, 60*(3), 525–541. https://doi.org/10.1111/roiw.12002.

Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *International Journal of Public Opinion Quarterly, 70*(5), 759–779. https://doi.org/10.1093/poq/nfl035.

Kempson, E. (2002). *Over-indebtedness in Britain*. London: Department of Trade and Industry.

Kempson, E., McKay, S., & Willitts, M. (2004). *Characteristics of families in debt and the nature of indebtedness (No 211)*. Leeds: Corporate Document Services.

Loh, W. Y. (2014). Fifty years of classification and regression trees. *International Statistical Review, 82*(3), 329–348. https://doi.org/10.1111/insr.12016.

Low, C. T., & Lai, P. C. (2016). Personal factors influencing the perception of quality of life in Hong Kong–a classification tree approach. *Procedia Environmental Sciences, 36*, 70–73. https://doi.org/10.1016/j.proenv.2016.09.014.

Lusardi, A., Schneider, D. J., & Tufano, P. (2011). *Financially fragile households: Evidence and implications (No w17072)*. Cambridge: National Bureau of Economic Research.

Magri, S., & Pico, R. (2012). Italian household debt after the 2008 crisis. Bank of Italy occasional papers, 134.

Manasse, P., & Roubini, N. (2009). "Rules of thumb" for sovereign debt crises. *Journal of International Economics, 78*(2), 192–205. https://doi.org/10.1016/j.jinteco.2008.12.002.

McCarthy, Y. (2011). Behavioural Characteristics and Financial Distress. ECB Working Paper, 1303. Available at: https://ssrn.com/abstract=1761570. Accessed 12 April 2019.

Melzer, B. T. (2011). The real costs of credit access: Evidence from the payday lending market. *The Quarterly Journal of Economics, 126*(1), 517–555. https://doi.org/10.1093/qje/qjq009.

Mian, A., Sufi, A., & Verner, E. (2017). Household debt and business cycles worldwide. *The Quarterly Journal of Economics, 132*(4), 1755–1817. https://doi.org/10.1093/qje/qjx017.

Michelangeli, V., & Pietrunti, M. (2014). A microsimulation model to evaluate Italian households' financial vulnerability. *Bank of Italy Occasional Paper*, (225). https://doi.org/10.2139/ssrn.2564087.

NBP – National Bank of Poland (2017). Zasobność gospodarstw domowych w Polsce. Raport z badania 2016 r., Warsaw, Poland: National Bank of Poland.

Ntsalaze, L., & Ikhide, S. (2016). Household over-indebtedness: Understanding its extent and characteristics of those affected. *Journal of Social Sciences, 48*(1–2), 79–93. https://doi.org/10.1080/09718923.2016.11893573.

OECD (2020), Household debt (indicator). doi: https://doi.org/10.1787/f03b6469-en. Accessed on 21 February 2020.

Oxera. (2004). Are UK households over-indebted? Commissioned by the Association for Payment Clearing Services (APACS), British Bankers Association (BBA), Consumer Credit Association (CCA) and the Finance and Leasing Association (FLA). Available at: https://www.oxera.com/wp-content/uploads/2018/03/Are-UK-households-overindebted-1.pdf. Accessed 28 March 2019.

Poterie, A., Dupuy, J. F., Monbet, V., & Rouvière, L. (2019). Classification tree algorithm for grouped variables. *Computational Statistics, 34*, 1–36. https://doi.org/10.1007/s00180-019-00894-y.

Ramsay, I., & Williams, T. (2020). Peering forward, 10 years after: International policy and consumer credit regulation. *Journal of Consumer Policy, 43*, 209–226. https://doi.org/10.1007/s10603-019-09436-x.

Rinaldi, L. & Sanchis-Arellano, A. (2006). Household Debt Sustainability: What Explains Household Non-Performing Loans? An Empirical Analysis. ECB Working Paper No. 570. Frankfurt am Main: European Central Bank. Available at SSRN: https://ssrn.com/abstract=872528. Accessed 20 February 2020.

Sánchez-Martínez, M. T., Sanchez-Campillo, J., & Moreno-Herrero, D. (2016). Mortgage debt and household vulnerability: Evidence from Spain before and during the global financial crisis. *International Journal of Housing Markets and Analysis, 9*(3), 400–420. https://doi.org/10.1108/IJHMA-07-2015-0038.

Sharma, H., & Kumar, S. (2016). A survey on decision tree algorithms of classification in data mining. *International Journal of Science and Research (IJSR), 5*(4), 2094–2097.

Świecka, B. (2009). *Niewypłacalność gospodarstw domowych: przyczyny, skutki, przeciwdziałanie*. Warszawa: Difin.

Świecka, B. (2016). Ocena sytuacji finansowej osób starszych w polsce ze szczególnym uwzględnieniem zasobności i zadłużenia [assessment of the financial situation of the elderly in poland, with

particular emphasis on wealth and debt]. *Problemy Zarządzania, 2*(59), 118–134. https://doi.org/10.7172/1644-9584.59.8.

Tay, L., Batz, C., Parrigon, S., & Kuykendall, L. (2017). Debt and subjective well-being: The other side of the income-happiness coin. *Journal of Happiness Studies, 18*(3), 903–937. https://doi.org/10.1007/s10902-016-9758-5.

Terraneo, M. (2018). Households' financial vulnerability in Southern Europe. *Journal of Economic Studies, 45*(3), 521–542. https://doi.org/10.1108/JES-08-2016-0162.

Tiongson, E., Gueorguieva, A. I., Levin, V., Subbarao, K., Sugawara, N., Sulla, V., et al. (2009). The crisis hits home: Stress testing households in Europe and Central Asia. *The World Bank*. https://doi.org/10.1596/978-0-8213-8222-6.

Turinetti, E., & Zhuang, H. (2011). Exploring determinants of US household debt. *Journal of Applied Business Research (JABR), 27*(6), 85–92.

Wałęga, A., & Wałęga, G. (2018). *Living on the edge: over-indebted households in Poland.* Conference Proceedings of 21st Annual International Conference Enterprise and Competitive Environment Mendel University in Brno.

Williams, M. A., Joskow, A. S., Johnson, R. L., & Hurdle, G. J. (1987). Explaining and predicting airline yields with non-parametric regression trees. *Economics Letters, 24*(1), 99–105.

Wu, X., & Kumar, V. (Eds.). (2009). *The top ten algorithms in data mining*. FL: CRC Press.

Zabai, A. (2017). Household debt: Recent developments and challenges. *BIS Quarterly Review, 4*, 39–54.

Zajączkowski, S., & Żochowski, D. (2007). *Obciążenia gospodarstw domowych spłatami długu: rozkłady i stress testy: na podstawie badań budżetów gospodarstw domowych GUS* (p. 221). Warsaw, Poland: National Bank of Poland Working Papers.