the review of income and wealth

Review of Income and Wealth
Series 68, Number S1, April 2022
DOI: 10.1111/roiw.12540

KEEPING UP WITH THE NOVAKS? INCOME DISTRIBUTION AS A DETERMINANT OF HOUSEHOLD DEBT IN CESEE

by Mariya Hake

International Finance Corporation (IFC), Oesterreichische Nationalbank (OeNB)
AND

Philipp Poyntner*

Vienna University of Economics and Business

This paper constitutes an initial attempt to shed light on the role of income distribution for household debt in Central, Eastern, and Southeastern Europe (CESEE). Using household-level data from the OeNB's Euro Survey for the period 2008–2018, we address the question whether interpersonal comparisons ("Keeping up with the Novaks") are associated with the probability of having a loan and planning to take out a loan. Applying multilevel probit modeling to consider the hierarchical structure of the data, our results support the notion that higher income inequality is negatively correlated with the probability of having a loan at the bottom of the distribution, and positively at the top. We show this impact for almost all components of household debt, but evidence is strongest for mortgage and foreign currency loans. Loan plans are associated with income inequality at the very top of the income distribution.

JEL Codes: D1, D3, G5

Keywords: CESEE, household loans, income distribution, multilevel models, relative income

1. Introduction

The global financial crisis that started in 2008 has increasingly drawn attention to the importance of, and the threats arising from, household sector debt for macroeconomic stability and GDP growth. Therefore, policymakers and researchers alike have turned their attention to the factors driving household indebtedness. Our analysis constitutes a comprehensive endeavor to relate debt and income

Note: The authors would like to thank Christian A. Belabed, Susanne Höfler, Stefan Humer, Martin Šuster, Anna K. Raggl, Aleksandra Riedl, Kilian Rieder, Julia Wörz, Zuzana Fungácová, seminar participants at the Bank of Albania, BOFIT, the Vienna University of Economics and Business, NOeG, the Biannual Conference of the Economics Section of the German Association for East European Studies, the IARIW-HSE Special conference "Experiences and Challenges in Measuring Income and Wealth in Eastern Europe and CIS Countries," and two anonymous referees for helpful comments and valuable suggestions. The views expressed are strictly those of the authors and do in no way commit the Oesterreichische Nationalbank (OeNB), the International Finance Corporation (IFC), the World Bank Group or the Eurosystem.

*Correspondence to: Philipp Poyntner, Vienna University of Economics and Business, Welthandelsplatz 1, A-1020 Vienna, Austria (philipp.poyntner@wu.ac.at).

distribution for households in the countries of Central, Eastern, and Southeastern Europe (CESEE) for the period 2008–2018.

What is the relation between household debt and income inequality? Two hypotheses have been forwarded in the theoretical literature. First, positional income concerns have been recognized as a factor influencing consumption since Veblen (1899) and Duesenberry (1949), a phenomenon referred to as "Keeping up with the Joneses." As debt can be used to finance consumption, higher income inequality could lead to an increase in consumption-related debt for households who see their reference households' income increase. Therefore, we investigate if a "Keeping up with the Novaks" effect - as we fittingly rename the "Keeping up with the Joneses" effect - can be observed in CESEE. Second, income inequality could be a "signaling" factor indicating the creditworthiness of borrowers to lenders (i.e., banks). Accordingly, *ceteris paribus*, with a more unequal income distribution, lenders would become more risk-averse and tend to lend to more affluent households rather than households on the lower end of the distribution (Coibion *et al*., 2014).

A limited number of papers have addressed the empirical link between household debt and income distribution, focusing primarily on advanced economies (e.g., Iacoviello, 2008; Coibion *et al*., 2014; Kumhof *et al*., 2015 for the US, Loschiavo, 2016 for Italy, Brown *et al*., 2016 for the United Kingdom). Two very recent papers are closely related to ours. Jestl (2019) suggests a positive impact of income inequality on consumption-related household indebtedness in a sample of EU countries, but not for the CESEE countries as included in our paper. In addition, Hake and Poyntner (2019) constitutes an initial analysis of the correlation between the probability of being indebted and income inequality according to the income position.

Notably, the analysis of the household indebtedness-income inequality nexus for European Emerging Economies is an interesting case as this region has shown diverging developments as compared to advanced economies. On one hand, the levels of household debt in CESEE have remained below the levels experienced in other parts of the world (e.g., the US, OECD, Euro area countries). On the other hand, the increase in household indebtedness has been coupled with a relatively high level of income inequality, with the Gini coefficient close to 0.5 in some countries.

Against that background, our paper is one of the first to explore the link between indebtedness and income inequality in CESEE. We use data from a household survey performed in ten CESEE countries in the period 2008-2018.[1] The data offer important advantages as compared to other data sets and the studies mentioned so far. First, they are comparable across a relatively large set of countries and thus alleviate concerns about biases coming from different sampling methods. Second, our data set encompasses 11 years in a repeated cross-section (and not a

[1]The CESEE country aggregate in this paper includes EU member states (Bulgaria, Croatia, the Czech Republic, Hungary, Poland, and Romania) and (potential) EU candidate countries (Albania, Bosnia and Herzegovina, North Macedonia, and Serbia).

panel)[2] at the individual level including both the pre-crisis and post-crisis period making our results more robust to time variation concerns. Additionally, our focus on loan intentions allows us to disentangle demand and supply factors.

In our preferred specification, we apply an upward-looking measure of income inequality, the relative reference income, which gives the average income of richer households as compared to the household's own income (in line with Drechsel-Grau *et al.*, 2014). We apply multilevel methodology to account for the correlation of responses from individuals from the same region. Finally, we provide a concise test to assess which are the most suitable reference groups also including a spatial aspect.

To foreground our findings, a higher relative reference income[3] is negatively correlated with households' likelihood of having a loan for households at the bottom of the income distribution. This is consistent with the signaling channel prompting banks to redirect their loans to richer households. For households at the top of the distribution, the relation between the relative reference income and debt is positive, consistent with both the signaling channel and the "Keeping up with the Novaks" channel. We show this impact for almost all components of household debt, but evidence is strongest for mortgage and foreign currency loans. The marginal effect of an increase in the relative reference income by one unit[4] is about 2 percentage points for the top of the decile. The sign and magnitude of this effect are robust to different definitions of comparison groups. The results for the bottom of the distribution are driven by low inequality regions, whereas the results for the top of the distribution are driven by high inequality regions, hinting at a threshold effect. These results are robust to various alternative income inequality measures such as the Gini coefficient, income percentile ratios, or income shares. In addition, the analysis explores loan *intentions* and thus tries to disentangle, to the extent possible, demand and supply factors for household indebtedness. We find support for a positive correlation between loan intentions and income inequality at the top of the income distribution, supporting evidence for the "Keeping up with the Novaks" effect.

The rest of the paper is structured as follows. In Section 2, we start with a brief discussion on the theoretical and empirical literature dealing with the influence of the income distribution on household debt. The description of the data and the methodological setup follow in Sections 3 and 4, respectively. In Section 5, we turn our attention to an analysis with respect to loan purpose and currency denomination. Section 6 zooms in on the impact on households' loan intentions. Sections 7 and 8 focus on alternative measures of income inequality and test the suitability of reference groups on the regional level, respectively. The last section concludes.

[2]The data could only be used as panel data if data are aggregated at regional or country level. Because this would result in a significant loss of information, we chose to work with the microdata at individual level and pool the data.

[3]The relative reference income is defined as the average income of richer households relative to the household's own income. A high relative reference income therefore indicates that the household's reference group income is large relative to that household's own income. Detailed information on this variable can be found in Section 3.

[4]The median of the relative reference income is 2.6 in our sample.

2. Theoretical and Empirical Background

This study builds on literature that has theoretically and empirically analyzed the economic consequences of income positional concerns. Looking at the demand-side aspect, the relative income theory of consumption (Veblen, 1899, Duesenberry, 1949) states that an individual's utility function depends on the consumption of other persons (their reference group).[5] Furthermore, assuming that the individual's utility depends on interpersonal comparisons, it follows that changes in the consumption of an individual's reference group influence that person's own consumption. Drechsel-Grau *et al*. (2014) conceptualize these interpersonal comparisons in the following Euler Equation[6]:

$$\Delta ln(c_{it}) = \mu + \alpha \Delta ln(c_{ref,t}) + \Delta X_{it}\delta + \epsilon_{it}, \tag{1}$$

where c_{it} is the consumption of individual i at point t, $c_{ref,t}$ is the consumption of the reference group, μ and δ are constants with information about the interest rate, the discount factor, and parameters of the utility function, and X_i a vector of other variables that influence marginal utility. Therefore, this argument stresses the social nature of consumption decisions. If incomes at the top of the income distribution grow, given a positive α, individuals at the bottom increase their consumption, and therefore dissave/borrow.

Closely related to the above is the habit formation theory, which describes a situation where in case of economic distress, an increase in borrowing would follow when individuals try to avoid cutting down on the level of consumption already attained (e.g., Iacoviello, 2008; Fasianos *et al*., 2017). As over-borrowing has been attributed to be one of the main causes of the financial crisis 2007–2008, this effect has received particular attention in the US (Van Treeck, 2014; Mian and Sufi, 2015; Bertrand and Morse, 2016).

An extensive body of literature explores the hypothesis that individuals derive utility from status, which in turn depends on what others believe about people's income (Ireland, 1994; Charles *et al*., 2009). While income is not observable, consumption is very frequently visible. Therefore, the level of individuals' conspicuous consumption (i.e., consumption that displays social status) can be expected to depend on the income distribution of the entire sample of individuals under observation. For example, Danzer *et al*. (2014) show that internal migrants in Kazakhstan use status consumption to impress their new neighbors. Against this background, some consumption and loans can be driven by conspicuous motives. Therefore, the impact of the income distribution on the likelihood of having a loan and on the loan's purpose (e.g., consumption, car, or mortgage) will enable us to make inferences about the existence and magnitude of such motives. The expenditure cascade approach by Frank *et al*. (2014) argues in a similar vein.

[5]A notable study on the effect of income comparisons on life satisfaction, especially relating to our geographical context, is Senik (2004). Using household survey data from Russia, the paper shows that the income of the reference group favorably affects expectations about own future income, the so-called "tunnel effect" (Hirschman and Rothschild, 1973).

[6]See Drechsel-Grau *et al*. (2014) for the derivation of this result from a standard optimization problem.

The theoretical predictions of these concepts have been widely empirically assessed. For instance, Kumhof *et al*. (2015) show that for the US a surge in the income share of the top deciles could largely explain the buildup of leverage among households at the bottom of the income distribution. Building on the central assumption that income shocks are of permanent nature, the authors present a model that shows that higher leverage and financial crises are the endogenous result of a growing income share of high-income households. For Italy, Loschiavo (2016) shows that richer households living in regions with higher income inequality are more likely to be indebted than similarly rich households residing in regions with low-income inequality (and vice versa for poorer households). Carr and Jayadev (2015) is a study closely related to ours, as it relates relative income of households to debt using US household data. Analyzing the period 1999–2009, they find a confirmation for the "Keeping up with the Joneses" effect as the growth of household debt increases with the share of families with higher income than the family in question. In addition, their findings show that low-income households are more likely to leverage up than higher income households during the increase in household leverage in the early 2000s in the US. For a sample of EU countries, Jestl (2019) shows that a positive effect prevails in the third quartile and thus upper middle class in most of the countries.

The relative income theory addresses the nexus between the consumption/income distribution and spending from a demand-side angle. Considering in addition the supply-side perspective, banks might use income distribution information next to information on a household's income to assess borrowers' creditworthiness, especially in countries with low credit register coverage (as is the case in some of the CESEE countries in our sample). Coibion *et al*. (2014) refer to this as the "signaling channel" and show that banks cannot observe borrowers' ability to meet debt obligations to a sufficient extent, so they consider the observed respondents' income together with income inequality in the region or the country. Accordingly, banks tend to restrict lending funds to poorer households when income inequality increases. This reasoning is used to rationalize the finding that in the period from 2001 to 2012, low-income households in high-inequality regions in the US accumulated less debt relative to income than their counterparts in regions with lower income inequality. In addition, higher income inequality might lead to higher debt as people on the higher end of the income distribution generally have a higher propensity to save (Kumhof *et al*., 2015).

The demand and supply channels are usually activated simultaneously (Bazillier and Hericourt, 2017), so the net effects are hard to disentangle. Therefore, one possible reason why some of the studies only find an effect at the top of the distribution could be the diametrical effect of the supply side-channel. In the empirical part, we first report the net effect of the income distribution on debt which provides information on whether the demand or supply effect dominates. In a next step, we distinguish between existing loans and planned loans to be able to discern demand and supply in the analysis.

The empirical literature on the relative income hypothesis often imposes a certain reference group leaving out a discussion on the suitability of a criterion for choosing a suitable reference group. However, are households perceptive of the income of their direct neighbors, households in the same city, region, or country?

Notable exceptions include Clark and Senik (2010), who, using European survey data, find that income comparisons are directed at work colleagues. Senik (2009) finds that households surveyed in the former communist countries compare their living standard mainly with that of former schoolmates and colleagues.[7]

Often, choices defining the "Novaks" are practically bound by data constraints. This is also the case in our study, as data limitations prevent us from performing an analysis at a level more disaggregated than the regional level. However, it is still possible to test differences in outcomes for more coarse neighborhood definitions. In particular, we investigate if neighboring regions affect household debt. To do so, we separately estimate the effects of relative reference income on debt for the home region and neighboring regions. We also investigate if defining the whole country as a reference group instead of the respective region changes the results.

The paper most intimately related to this study and covering the same set of CESEE countries is Hake and Poyntner (2019). This paper features an initial explorative analysis of the correlation between the probability of having a loan and income inequality for the period 2009–2017. However, in contrast to the present study, the authors neither explore in detail the impact on different types of loan, various income inequality measures, loan intentions nor look into the spatial aspect of the definition of the reference groups.

3. Data and Descriptive Statistics

3.1. *Data*

The empirical analysis is based on data from the OeNB Euro Survey—a household survey performed in ten CESEE countries and commissioned by the Austrian Central Bank (OeNB). The survey was performed twice a year between 2007 and 2014 and has been performed annually since 2015. It includes six EU member countries (Bulgaria-BG, Croatia-HR, Czech Republic-CZ, Hungary-HU, Poland-PL, and Romania-RO) and four (potential) EU candidates (Albania-AL, Bosnia and Herzegovina-BA, Serbia-RS, and North Macedonia-MK). In each country and per wave, the target population comprises around 1000 interviewees representative of a country's population, aged 14 years or older, selected via a multi-stage stratified random sampling procedure. The data are a repeated cross section at the individual/household level. For the period of analysis of this paper (2008–2018[8]), this corresponds to a total number of observations of about 134,000.[9] The survey delivers information at the regional level, whereas the allocation roughly corresponds to NUTS 2.[10]

[7]According to Knight *et al*. (2009), Chinese households report that their reference groups are individuals in the same village.

[8]Although the survey was initiated in 2007 for the first time, we left out the first year (i.e., 2007) as the question on household debt was only introduced in the 2008 survey.

[9]The number of observations corresponds to roughly 2000 per country and year in the period 2008–2014 and roughly 1000 observations per country and year in the period 2015–2018. Data for the Czech Republic for 2008 and 2010 were incomplete and are therefore not included.

[10]For more information on the survey, see https://www.oenb.at/en/Monetary-Policy/Surveys/OeNB-Euro-Survey.html.

The survey includes questions on the use of the euro in a household's portfolio but also covers information on both the assets and liabilities of households. In particular, it informs about loans a respondent might have, either alone or with their partner, the purpose of the loans, and their currency denomination. The data also include information on whether respondents plan to take out a loan in the near future and again about its purpose and currency denomination. Apart from assessing the income of respondents in those ten CESEE countries, the survey also collects information on sociodemographic characteristics, including age, education, and employment status. Finally, the survey also includes various questions on sentiments about future developments and on past experiences.

3.2. *The Relative Reference Income*

Our income information is based on the OeNB Euro Survey question: "What is the total monthly income of the household after taxes?".[11] Between 2008 and 2016, survey respondents were asked to place their income in one of 20 income categories.[12] Before the survey data were passed on to external use, the ranges of categories were harmonized across the different countries and over the years by the survey team. In addition, they transformed amounts transformed into euro and into purchasing power units (to capture exchange rate and inflation differences) to ensure cross-country comparability.[13] We then took the average of each income category to compute the equalized household income. Only starting in the 2017 and 2018 survey waves, respondents were asked to report the exact amount of their household income.

This paper follows an approach first applied in Belabed and Hake (2018) to make adjustments to the income data coming from the survey. First, these adjustments include treating item non-response to address missing income data from some respondents. On average, 20 percent of respondents did not respond to the income question (highest share of non-response in Bosnia and Herzegovina, Bulgaria, and Romania with up to 29 percent; the Czech Republic has the lowest share with 3 percent). As we cannot assume that income information is missing completely at random, we opted against dropping these observations to avoid the issue of selection bias (Groves and Peytcheva, 2008; ECB, 2016).[14] Following Harrell (2001) and Gulyás (2009), we applied a hotdeck imputation method. Thereby, it is assumed that the item non-responses would be randomly distributed,

[11]A cross-check with the median net income reported by EU-SILC respondents for eight countries shows a close correspondence.

[12]The income ranges have been set up such that they are representative of the country's income distribution. For example, Serbian respondents were asked to place their income in one of the following brackets: 0–5000 dinars, 5001–10,000 dinars, etc.

[13]We want to stress that the data transformations described above have been performed on the survey raw data and by the survey team. The authors have been allowed to use the data only after these transformations have been done.

[14]We tested the hypothesis that the missing income data might be assigned to the more affluent part of the respondents. A probit estimation on the likelihood of not disclosing income information showed that it is rather young male respondents with university education without children who refuse to reply to this question. At the same time, students and unemployed in the data sample are also more likely to retain information on their income. Thus, these results are not clearly affirmative of the assumption that the more affluent part of the sample refuses to answer.

that is, missing at random within a group. With respect to our income variable this means that the probability of refusal is related to some socioeconomic factors, but within each socioeconomic group, the probability of missing values is independent of income. In particular, the hotdeck imputation replaces missing values in the income data with complete lines, that is, observations that are similar from the same stratum, which in this case is the observations in the same region–year pair.

Second, we applied Pareto correction to handle possible misreporting of the upper tail of the income distribution in our data set (Atkinson, 2017). Thus, we estimate a Pareto-shaped distribution for the top quintile of the distribution which should yield more realistic values of income and its distribution (in line with, for instance, Piketty and Saez, 2003; Blanchet *et al*., 2017 or Eckerstorfer *et al*., 2016). To perform a Pareto distribution for top incomes, we need two parameters: (i) the size parameter m and (ii) the shape parameter α (or Pareto's alpha). The size parameter m determines the threshold of income above which the functional form of the distribution follows a power-law. The shape parameter α determines the shape or slope of the distribution function beyond the threshold and may be considered a measure of inequality itself—a lower α indicates a higher level of inequality and vice versa.

To determine the size parameter m, we follow Eckerstorfer *et al*. (2016), who model a Pareto-shaped distribution for the top quintile of the distribution. Thereby, we focus on the country-level income distribution. Certainly, the choice of m is crucial for the resulting estimates of income concentration as it determines the part of the distribution which follows a Pareto-type distribution. Choosing a size parameter below the (unobserved) true value includes observations from the non-Pareto part of the distribution, whereas the choice of a higher value may exclude observations from the Pareto part of the distribution. In contrast to other surveys, a key advantage of our data is that we are able to compute α coefficients for every country and year in our sample (for more details, see Belabed and Hake, 2018).

After the above-mentioned income data adjustments were performed, in the following we will introduce our main income inequality variable. The main income inequality variable we apply in the present analysis is the *relative reference income* (hereinafter referred to as *relinc*). From a demand-side perspective, the literature emphasizes that interpersonal comparisons are upward-looking: households compare their consumption to richer households and adjust their consumption preferences accordingly (Ferrer-i-Carbonell, 2005; Carr and Jayadev, 2015). We therefore follow Drechsel-Grau *et al*. (2014) and define the households' reference income to account for upward-looking comparisons.[15] The reference income of a household is the average income of all households in the income deciles above the respective household's income, in the given year. The relative reference income *relinc* is defined as the ratio of the reference income to the household's own income.

The relative reference income for a given household i in a given year is therefore defined as follows:

[15]In particular, Drechsel-Grau *et al*. (2014) focus on consumption and define reference consumption as the consumption level of all households that are perceived to be richer than the respective household.

$$relinc_i = \frac{1}{K} \sum_{\substack{j=1 \\ D_j > D_i}}^{K} (Y_j) \frac{1}{Y_i}, \tag{2}$$

where $i = 1, \ldots, N$ and $j = 1, \ldots, K$ are households, D_i is the regional income decile of household i, and Y_i is the income of household i. As we have a repeated cross-section, each household appears once in the survey; therefore, the reference income is calculated for the year of the respective survey. *Relinc* is bounded by 1 from below per definition, whereas the highest value is 92 in the sample (see descriptive statistics in the Appendix). A higher *relinc* therefore indicates higher income inequality in the sense that the income of a household's reference group is large relative to the household's own income.

We decided to define the reference groups as households in the same region. We acknowledge that the "Keeping up with the Novaks" hypothesis conveys the idea that individuals are influenced by people they frequently interact with, such as neighbors, family, or colleagues. However, the regional level is the most granular level of data that allows us to have a sufficient number of observations as the number of observations at the PSU (primary sampling unit) level is too small.[16] We consider this income inequality measure to be most suitable for our analysis due to two reasons. First, it is an upward-looking measure per definition, which is suitable for our interpersonal comparison setting. Second, as per definition, this measure incorporates a household's own income, allowing for a relative reference income measure for each household.[17]

Figure 1 shows the distribution of households' relative reference income per country, while the income deciles and the respective comparisons are constructed on the regional level. Respondents in the lowest-income deciles (1st and 2nd) have the highest *relinc* of up to 92 (in Serbia), implying that the income of a household in the first decile could be up to 92 times lower than the average income of all deciles above the 1st decile. For the whole sample, the median of the relative reference income for all deciles is 2.6, while 99 percent of all observations are below 14. Judging by the descriptive evidence, there is some country-level variation as well. Accordingly, the households at the lower end of the income distribution in Bosnia and Herzegovina, North Macedonia, Romania, and Serbia tend to be relatively more disadvantaged (i.e., have a higher *relinc*) than households in the same deciles in other countries.

Important to note is that for some observations, it is possible that households in higher deciles have a lower *relinc* than households in a lower decile. This is due to the fact that we define the reference income as the mean income of households in higher deciles. Therefore, households with very similar incomes that are grouped into different deciles have different reference incomes. Take for instance an example from our data: A certain household in the North Albania region in 2009 has an

[16]In Section 5.2, we show that defining the reference group as richer households in the same country (as opposed to region) does not alter the results.

[17]In Sections 5 and 8, we show the robustness of our results using modified versions of this measure as well as other inequality measures.

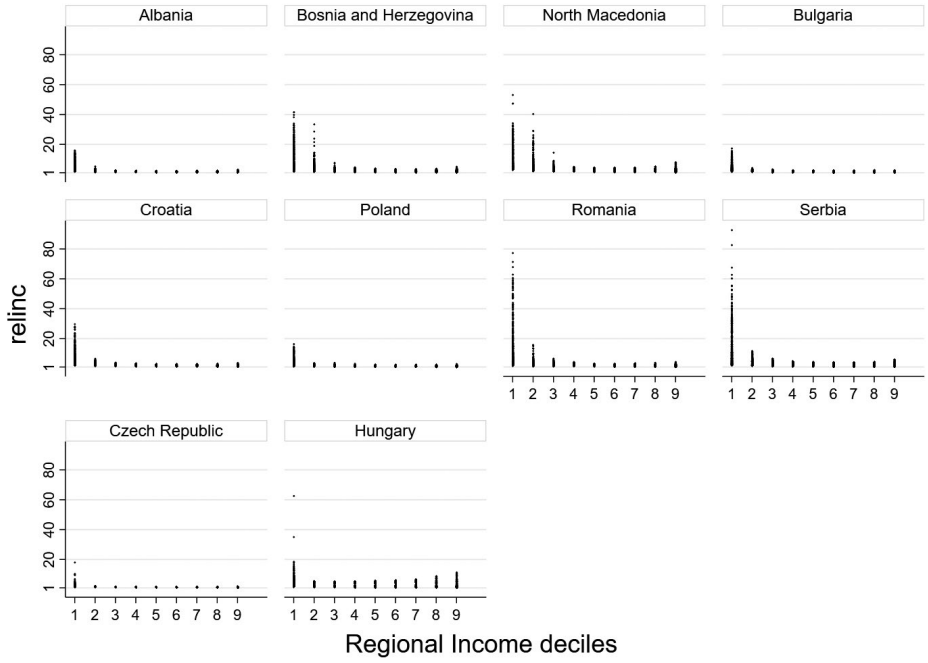

Figure 1. Relative Reference Income *relinc* by Decile and Country

Notes: The relative reference income *relinc* is the mean income of households in higher deciles divided by the households' own income. Income deciles are formed at regional level.

Source: OeNB Euro Survey, own calculations.

income of 115 euros.[18] which places it in income decile 2, whereas the mean income of households in deciles 3–10 is 303 euros. Said household's *relinc* is 2.6 (303/115). The next household with an income of 120 euros is placed in income decile 3. The mean income of households in deciles 4–10 is 329 euros. This household's *relinc* is 2.7. Therefore, it is possible that households in higher deciles have a higher *relinc*. However, poorer households in the same decile of course always have a higher *relinc*. In Section 5.2, we show that our results are not sensitive to defining the reference income as the average income of *all* richer households, including those in the same decile.[19]

3.3. *Evidence on Existing and Planned Household Loans*

The OeNB Euro Survey provides information on whether households in the CESEE sample have or intend to take out a loan. Moreover, it also contains information on the loan's purpose (i.e., consumer loan, mortgage loan, car loan, or loan

[18]As stated in the previous subsection, the data have been converted into euro and PPP.

[19]See Appendix Figure A1 for an example of the relationship between the components of the relative reference income variable and income deciles.

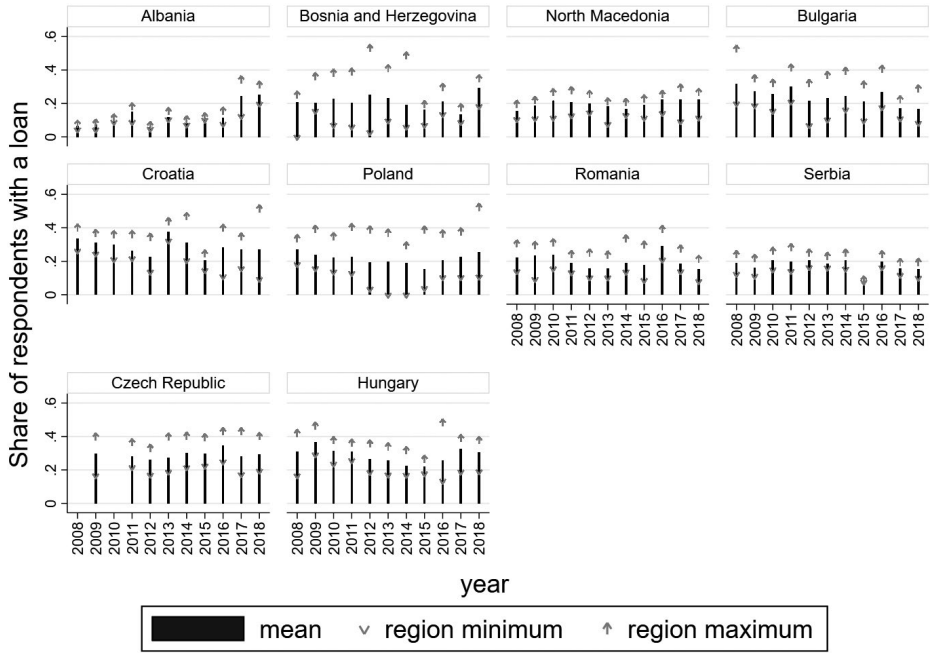

Figure 2. Current Loans

Notes: The black bars denote the mean share of households with a loan of regions in a country, the upper arrow shows the maximum share, and the *v* the minimum share.

Source: OeNB Euro Survey, own calculations.

for other purposes) and its currency denomination.[20] Therefore, our analysis focuses on the likelihood of having/planning to take out a loan and not on the respective amount of the loan.

Figure 2 shows the development of the share of households with loans over time. It also shows the minimum and maximum shares per region in the respective country. Accordingly, the variance of the regional share of respondents with loans is on average the highest in Bosnia and Herzegovina, Romania, and the Czech Republic. Overall, the average shares of indebted households are the highest in Czech Republic, Hungary, and Croatia pertaining to about one-third of the total respondents. Figure 3 shows the share of households planning to take out a loan. On average, 8.4 percent of all respondents plan a loan. The share is the highest in the Czech Republic, Hungary, and Serbia (see also Table A3 in the Appendix).

3.4. *Income Distribution and Household Debt: A First Glance at Correlations*

Figure 4 plots the region/year means of *relinc* and the mean share of households with a loan (in contrast to Figure 1, we show the decile/year means for this

[20]The loan question is asked in the following way: *"Do you, either personally or together with your partner, currently have any loans that you are still paying off?"* and *"Do you plan to take out a loan within the next year and if so, in what currency?"*, respectively. If respondents have or plan to take out a loan, they are asked to specify the purpose of the loan: *"to finance a house or apartment," "for consumption goods (furniture, travelling, household appliances, etc.)," "to finance a car," or "for other purposes."*

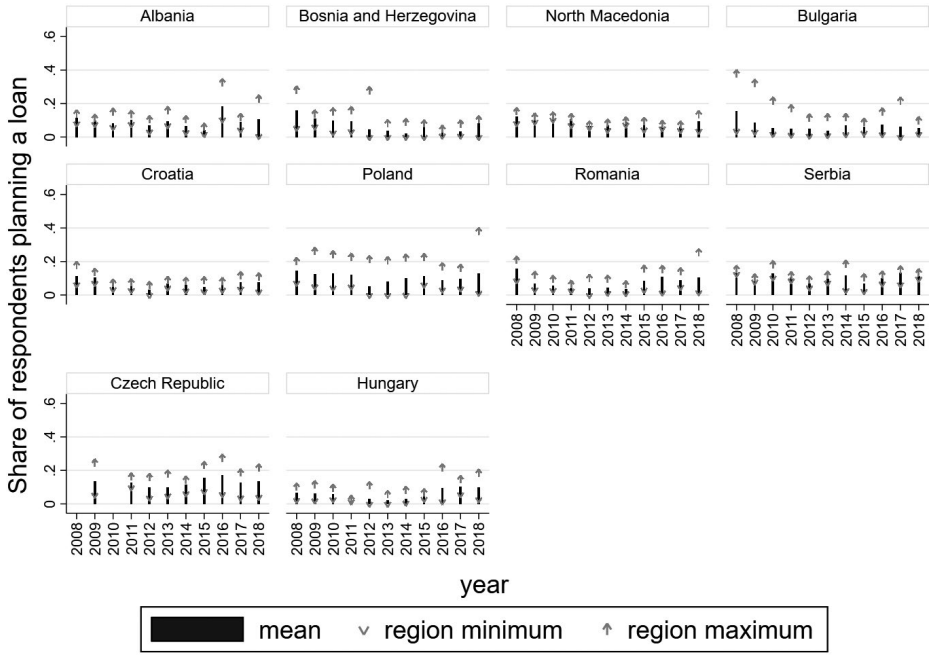

Figure 3. Planned Loans

Notes: The black bars denote the mean share of households planning to take out a loan of regions in a country, the upper arrow shows the maximum share, and the *v* the minimum share.

Source: OeNB Euro Survey, own calculations.

measure). We observe a slight negative correlation between the relative reference income and loan shares, as depicted by the fitted values indicated by the red line. In Figures 5–7, we plot loan shares and three income inequality measures as computed using the income data from the OeNB Euro Survey: the Gini coefficient, the share of the top 10 percent, and the top 1 percent households in the regional income distribution, respectively. Again, the slightly negative correlation remains irrespective of the income inequality measure used.

4. Methodology

Our chosen methodology accounts for the fact that both regional and individual factors might correlate with a household's likelihood to have or to intend to take out a loan. In particular, we first follow the assumption that the "Keeping up with the Novaks" effect is most likely located at the regional level as households that are spatially close to each other tend to have the most frequent interactions. Accordingly, the correlation between income inequality and household debt is likely to differ between households in different regions and countries, as households' characteristics interact with institutional characteristics. Therefore, the assumption that observations are conditionally independent given the covariates is likely to be violated because of interdependence of households in a region.

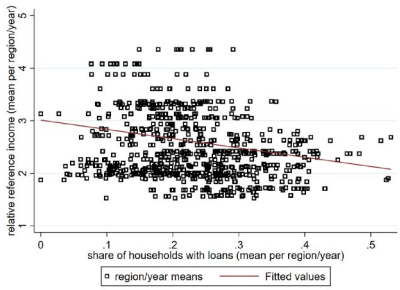

Figure 4. Correlation Between relinc and Current Loan Share

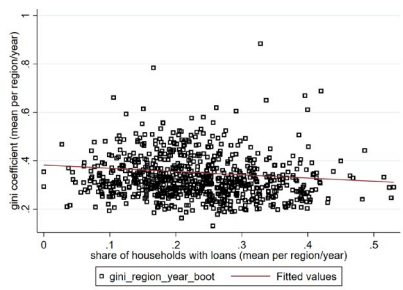

Figure 5. Correlation Between Gini Coefficient and Current Loan Share

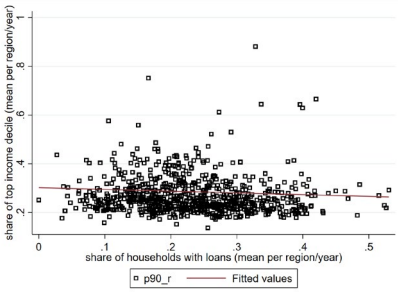

Figure 6. Correlation Between Top 10 Percent Share and Current Loan Share

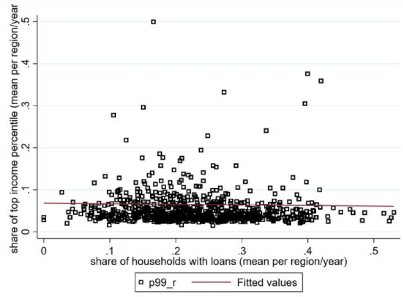

Figure 7. Correlation Between Top 1 Percent Share and Current Loan Share

Notes: The dots are region-year means for the share of households with loans and relinc, the Gini coefficient, the regional share of the top 10 percent households, and the regional share of the top 1 percent households.

Disregarding this interdependence can lead to spurious "significant" coefficients of the included variables. Against that background and to account for the nested structure of the data (i.e., "clusters"), we apply multilevel models (Rabe-Hesketh and Skrondal, 2012), widely used, for instance, in health (e.g., patients nested in hospitals) and social sciences (e.g., students nested in schools). As our outcome of interest for this paper is a binary variable (a dummy variable indicating that a household has or plans to take out a loan), we apply a multilevel probit model. It contains random effects to account for the interdependence of observations at the cluster levels (regions).

We consider the following two-level models:

$$Pr(loan_{irt} = 1 | X_{irt}, U_r) = H(X_{irt}\beta + Z_{ir}U_r), \tag{3}$$

$$Pr(planloan_{irt} = 1 | X_{irt}, U_r) = H(X_{irt}\beta + Z_{ir}U_r), \tag{4}$$

TABLE 1
ICCs

# of Levels	Level	ICC	Std. Err.	95% Conf. Int.	
1	country	0.035	0.015	0.015	0.082
1	region	0.038	0.006	0.028	0.052
2	country	0.031	0.015	0.013	0.076
2	region	0.047	0.014	0.026	0.083

Notes: ICCs calculated after estimation of multilevel probit models without covariates, dependent variable: current loan.

with $r = 1, \ldots 76$ clusters, in our case regions, consisting of $i = 1, \ldots, N$ households, observed in $t = 1, \ldots, T$ years. X_{irt} is a $1 \times p$ vector of covariates. In the baseline regression the following variables are included: income, sex, age, age squared, children, education, dummies for unemployed, self-employed, student, and retired.[21] β is a vector of regression coefficients. The $1 \times p$ vector Z_{ir} denotes the random effects in both intercepts and coefficients. U_r denotes the random effects. $H(.)$ is the standard normal cumulative distribution function. The same approach has been applied for the different types of loans as well.

The choice of the level of the analysis (region, country) is key for the model as it determines the level of the random effects. The dependent variable has to show some unexplained variance related to the cluster. Therefore, if there was no variance left after including all the variables on the individual/household level, the inclusion of the regional level would be superfluous.

The intraclass correlation coefficient (ICC) gives an indication for the correct choice of level. In the two-level specification (individual and region effects), the overall error term w_r is decomposed into e_{ir} and u_{ir}, where e_r is the random error term for the $i - th$ respondent within the $r - th$ region and is assumed to have zero mean and constant variance σ_e^2. The regional effects are estimated through u_r which is assumed random with zero mean and a constant variance σ_u^2. The partitioning of the variance in this manner defines a measure to test the suitability of the multilevel modeling, ICC. It measures the strength of "nesting" within the data hierarchy and is given as follows:

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}. \tag{5}$$

Therefore, the ICC is the proportion of variance in the dependent variable accounted for by the group level.

Table 1 suggests that introducing both regions and countries as levels into our multilevel model improves our estimations.[22] However, a small number of clusters at the second level (in our case, countries) can lead to a severe bias. While the literature cannot offer a clear-cut indication on how many level-two clusters are sufficient (e.g., Maas and Hox, 2005, Schunck, 2016), 10 countries is very low by any measure and the probability of downward bias would be present if we included

[21]Variable definitions can be found in the Appendix.
[22]For the sake of comparison, Table A5 in the Appendix shows the results of a non-multilevel probit estimation with region-, country-, and time-fixed effects. As presented, the main results hold qualitatively and quantitatively.

both regions and countries as levels. Against the background of cluster numbers and lower ICCs at country level, we opted for including only the regional level in the main estimations.[23] In addition, the country variation has been captured by country dummies, thus allowing for different intercepts but not allowing for different coefficient estimates at country level.

In the following sections, the coefficients reported should be understood as correlations and not as causal effects despite some attempts to alleviate endogeneity issues throughout the paper.

5. Household Debt at Equilibrium

5.1. *Baseline Estimations*

This section focuses on current household loans and the correlation with income inequality. We explore various features of existing loans such as the purpose of the loans (Subsection 5.3) and currency denomination (Subsection 5.4). In addition, robustness of the results along different reference groups is investigated.

Table 2 presents the results of the baseline estimations for current household loans. This realized debt is the "equilibrium" outcome, as opposed to the demand for loans we explore in Section 6. We control for major sociodemographic variables such as age, education, and employment status in all estimations.[24] In column (1), we include the constant only to show the explanatory power of the regional-level variables. Accordingly, nearly 4 percent of the variation could be explained by the regional-level variables. When including the income inequality variable as a level effect (i.e., without differentiating between the income deciles), a higher relative income correlates negatively with the likelihood to have a loan (column (2)). It is to be noted that a higher relative income indicates that the household has a lower income compared to its richer reference group. A higher relative income therefore indicates higher (upward-looking) income inequality.

Columns (3)–(7) show the results for the different income deciles. There is a positive correlation between the relative reference income above the fifth income decile of the regional distribution and the likelihood to currently have a loan. This finding is consistent with the demand-side "Keeping up with the Novaks" effect as higher reference income increases the demand for debt. At the same time, it is also consistent with the signaling channel, the supply-side channel (see Coibion *et al*., 2014), where increasing income inequality increases credit supply for the top of the distribution. For households at the bottom 20 percent of the income distribution, a higher reference income is correlated with a lower probability of having a loan. Therefore, for the bottom of the distribution, the supply-side channel that reduces loan supply for poorer households seems to dominate the demand channel.

Splitting the regions along the level of the average relative reference income of 2.6 does not change the results qualitatively. The results in columns (4) and (5) show, however, that the negative effects at the lower end of the regional income

[23]However, in Table A4 in the Appendix, we show that including countries as levels does not alter our main findings for the baseline estimates for existing loans.

[24]The characteristics refer to the survey respondent.

TABLE 2
HOUSEHOLD LOANS IN CESEE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Const	No Interaction	Baseline	Low Ineq Reg	High Ineq Reg	2011–2018	Wealth
Income distribution							
1st decile#relinc			−0.016***	−0.038***	−0.0074**	−0.011**	−0.014***
			(0.0038)	(0.0098)	(0.0029)	(0.0049)	(0.0047)
2nd decile#relinc			−0.026***	−0.060***	−0.0046	−0.019*	−0.024**
			(0.0077)	(0.015)	(0.0078)	(0.010)	(0.010)
3rd decile#relinc			−0.013	−0.034*	0.0034	−0.0056	−0.0100
			(0.0100)	(0.017)	(0.012)	(0.013)	(0.012)
4th decile#relinc			0.011	−0.018	0.034***	0.027	0.017
			(0.011)	(0.017)	(0.013)	(0.018)	(0.019)
5th decile#relinc			0.021	−0.0011	0.033**	0.050**	0.033
			(0.013)	(0.023)	(0.016)	(0.020)	(0.024)
6th decile#relinc			0.039***	0.024	0.040**	0.056***	0.040*
			(0.013)	(0.021)	(0.016)	(0.020)	(0.021)
7th decile#relinc			0.056***	0.036*	0.058***	0.076***	0.065**
			(0.014)	(0.021)	(0.017)	(0.023)	(0.028)
8th decile#relinc			0.068***	0.041*	0.072***	0.10***	0.097***
			(0.015)	(0.022)	(0.017)	(0.025)	(0.026)
9th decile#relinc			0.069***	0.058**	0.045***	0.11***	0.095***
			(0.016)	(0.027)	(0.015)	(0.031)	(0.033)
Relinc		−0.016**					
		(0.003)					
Income		0.000***	−0.000**	−0.000***	0.000**	−0.000*	−0.000**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Wealth proxies							
Savings							0.010
							(0.016)
House							0.040***
							(0.028)
Car							0.150***
							(0.022)

(Continues)

TABLE 2 (CONTINUED)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Const	No Interaction	Baseline	Low Ineq Reg	High Ineq Reg	2011–2018	Wealth
Sociodemographics							
Female		0.029***	0.028***	0.021	0.044***	0.029**	0.047***
		(0.0092)	(0.0092)	(0.013)	(0.012)	(0.014)	(0.014)
Age		0.10***	0.10***	0.10***	0.11***	0.11***	0.10***
		(0.0039)	(0.0039)	(0.0051)	(0.0054)	(0.0045)	(0.0046)
Age squared		−0.0012***	−0.0012***	−0.0012***	−0.0012***	−0.0012***	−0.0012***
		(0.000043)	(0.000043)	(0.000056)	(0.000063)	(0.000049)	(0.000052)
Children		0.13***	0.13***	0.15***	0.11***	0.14***	0.14***
		(0.013)	(0.012)	(0.014)	(0.019)	(0.014)	(0.014)
Education		0.15***	0.14***	0.10***	0.19***	0.13***	0.11***
		(0.016)	(0.016)	(0.022)	(0.023)	(0.023)	(0.022)
Unemployed		−0.31***	−0.30***	−0.19***	−0.40***	−0.35***	−0.34***
		(0.028)	(0.028)	(0.039)	(0.026)	(0.035)	(0.036)
Self-employed		−0.0057	−0.0062	0.039	−0.095	−0.022	−0.048
		(0.031)	(0.031)	(0.034)	(0.060)	(0.041)	(0.046)
Student		−0.68***	−0.67***	−0.75***	−0.55***	−0.67***	−0.67***
		(0.058)	(0.058)	(0.096)	(0.059)	(0.070)	(0.079)
Retired		−0.13***	−0.13***	−0.16***	−0.11***	−0.13***	−0.11**
		(0.023)	(0.023)	(0.032)	(0.034)	(0.040)	(0.042)
Constant	−3.53***	−3.58***	−3.31***	−3.53***	−3.60***	−3.61***	
	(0.086)	(0.092)	(0.12)	(0.11)	(0.12)	(0.12)	
ICC (regional)	0.042	0.014	0.015	0.014	0.012	0.019	0.020
N	134,092	116,666	116,666	68,855	47,811	58,135	49,164

Dependent variable: binary response indicating if respondents have a loan. Estimation method: multilevel modeling. Country and time fixed effects are included in all estimations. The intraclass correlation coefficient (ICC) denotes the explained portion of the variance by inclusion of the regional (second) level covariates. Robust standard errors in parentheses. Variables are defined in the Appendix.

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

distribution are driven by low income inequality regions. The positive correlation for the top of the distribution is mainly driven by high-inequality regions. This is a policy-relevant result which shows that financial inclusion (the number of people reached by financial services) might be deterred by higher income inequality.

In a next step, we add proxies for wealth (i.e., whether households hold savings, cars, or property), in column (7). The latter components were only included from 2011 on; therefore column (5) repeats the baseline estimation in column (3) but for the period 2011–2018. The results suggest that our findings are not driven by confounding wealth effects affecting debt intake. However, the coefficients for the wealth controls should be taken with a grain of salt due to endogeneity issues. The included sociodemographics largely show the expected signs. Older respondents are more likely to have a loan, but this effect is nonlinear starting to decline above the age of 53, and this is also in line with the negative sign of the "retired" dummy. Respondents with a university degree are also more likely to have a loan. As for female respondents and respondents with children, it is to be noted that these coefficients are to be understood as correlations for households with a female partner and a household with children, thus probably not directly comparable with other studies due to this specific feature of the loan-questions. Accordingly, households with a female partner and households with children are more likely to take out a loan.

As the coefficients so far show the qualitative stance, we compute the average marginal effects to get a better idea of the economic size of the effect (Figure 8). For instance, the average marginal effect of respondents' relative reference income in the seventh income decile is 0.02, which means that an increase in the relative reference income in this decile by 1 unit (for instance, from 2 to 3) would increase the likelihood of a household in this decile to have a loan by 2 percentage points. Considering that the share of indebted households in the upper deciles is higher than in the rest of the income distribution, the overall effect of the relative reference income on household indebtedness is non-negligible.

5.2. *Household Indebtedness in CESEE: Different Reference Groups*

In this section, our aim is to explore if the correlation found so far is driven by our definition of a reference group. In particular, we try to answer the question: Is the upward-looking component primarily driven by households that have the same age, education, or only by those close to the households' own income rank?

The estimation results are presented in Table 3. First, for the sake of easier comparison, in column (1) we include the baseline estimation from Table 2, where the reference group are all households in higher regional income deciles.

In column (2), the reference group is defined as only those households that are in the next higher decile at regional level. The reason is that households might focus above all on households that are richer but still close to their own income rank (Drechsel-Grau *et al.*, 2014). Interestingly, our expectations for larger coefficients in this case are confirmed (for the marginal effects see Figure 9). This could indicate that households compare the most with their immediate "income neighbors."[25]

[25]It is, however, also possible that this result is driven by the fact that when only using the neighboring decile, *relinc* is per definition smaller than it is when using all richer deciles.

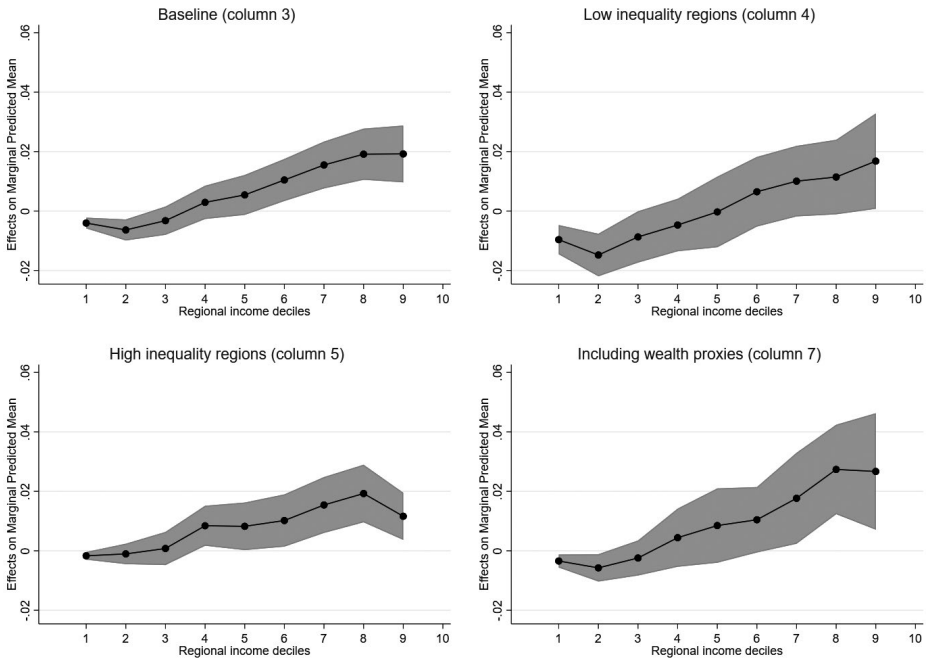

Figure 8. Marginal Effects of Relative Reference Income at Income Deciles for the Estimations in Table 2, Columns 3, 4, 5, and 7

Notes: Average marginal effects of relative reference income for selected specifications of estimations presented in Table 2. The grey area depicts the 95 percent confidence interval.

Source: OeNB Euro Survey, own calculations.

In columns (3) and (4) we define the reference groups to be at the same educational level[26] and in the same age group[27] at the regional level, respectively. The results from the baseline estimations remain unchanged. A caveat in this analysis is the at times relatively low number of observations that are used to calculate the reference income for each education/age group-year cell. While empty cells do not occur frequently, the reference income is calculated as the average income of a low number of observations. Therefore, while it is reassuring that the signs of our coefficients are mostly unchanged, effects sizes should be interpreted with a grain of salt.

Finally, the reference group in column (5) includes all households which are more affluent (including the ones in the household's own regional income decile). An advantage of this approach is that we also obtain an estimate of the top income decile—where no significant effect can be found. Our main takeaway from the estimations so far is that there is a robust pattern not entirely driven by specific subgroups (see Figure 10 and 11).

[26]Education has been defined in categories low, medium, and high.
[27]14–24 years; 25–54 years; 55 years and older.

TABLE 3
HOUSEHOLD LOANS IN CESEE—ALTERNATIVE RELATIVE REFERENCE INCOME CONCEPTS

Relative Reference Income Group	(1) Baseline	(2) Next Decile	(3) Same Education	(4) Same Age Group	(5) All Richer
Income distribution					
1st decile#relinc	−0.016***	−0.053***	−0.013***	−0.014***	−0.026***
	(0.004)	(0.010)	(0.004)	(0.003)	(0.005)
2nd decile#relinc	−0.026***	−0.066***	−0.025***	−0.026***	−0.048***
	(0.008)	(0.023)	(0.008)	(0.008)	(0.008)
3rd decile#relinc	−0.013	−0.021	−0.0011	−0.017**	−0.038***
	(0.010)	(0.022)	(0.008)	(0.0082)	(0.008)
4th decile#relinc	0.011	0.020	0.012	−0.001	−0.029***
	(0.01)	(0.022)	(0.010)	(0.001)	(0.010)
5th decile#relinc	0.021	0.036*	0.019*	0.005	−0.030***
	(0.013)	(0.020)	(0.011)	(0.008)	(0.011)
6th decile#relinc	0.039***	0.069***	0.036***	0.016	−0.004
	(0.013)	(0.021)	(0.011)	(0.010)	(0.010)
7th decile#relinc	0.056***	0.090***	0.053***	0.024***	0.000
	(0.014)	(0.022)	(0.009)	(0.011)	(0.009)
8th decile#relinc	0.068***	0.110***	0.060***	0.031***	0.016*
	(0.015)	(0.022)	(0.012)	(0.010)	(0.009)
9th decile#relinc	0.069***	0.072***	0.058***	0.019*	0.012**
	(0.016)	(0.016)	(0.013)	(0.010)	(0.006)
10th decile#relinc					0.011
					(0.010)
Income	−0.000**	−0.000***	−0.000*	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sociodemographics					
Female	0.028***	0.029***	0.029***	0.029***	0.030***
	(0.0092)	(0.0091)	(0.0091)	(0.0093)	(0.010)
Age	0.10***	0.10***	0.10***	0.10***	0.10***
	(0.0039)	(0.0039)	(0.0040)	(0.0040)	(0.0042)

(Continues)

TABLE 3 (CONTINUED)

Relative Reference Income Group	(1) Baseline	(2) Next Decile	(3) Same Education	(4) Same Age Group	(5) All Richer
Age squared	−0.0012***	−0.0012***	−0.0012***	−0.0012***	−0.0012***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Children	0.13***	0.13***	0.13***	0.13***	0.140***
	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)
Education	0.14***	0.14***	0.14***	0.15***	0.14***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)
Unemployed	−0.30***	−0.30***	−0.30***	−0.30***	−0.300***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.031)
Self-employed	−0.0062	−0.0041	−0.0060	−0.0021	−0.014
	(0.031)	(0.031)	(0.032)	(0.032)	(0.028)
Student	−0.67***	−0.67***	−0.67***	−0.67***	−0.69***
	(0.058)	(0.058)	(0.058)	(0.059)	(0.053)
Retired	−0.130***	−0.130***	−0.120***	−0.130***	−0.130***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.024)
cons	−3.58***	−3.56***	−3.59***	−3.56***	−3.39***
	(0.092)	(0.091)	(0.087)	(0.091)	(0.092)
ICC (regional)	0.015	0.015	0.015	0.015	0.014
N	116,666	116,488	116,041	111,957	98,141

Dependent variable: binary response indicating if respondents have a loan. Estimation method: multilevel modeling. Country and time fixed effects included in all estimations. Intraclass correlation coefficient denotes the explained portion of the variance by inclusion of the regional (second) level covariates. Robust standard errors in parentheses. Variables are defined in the Appendix.

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

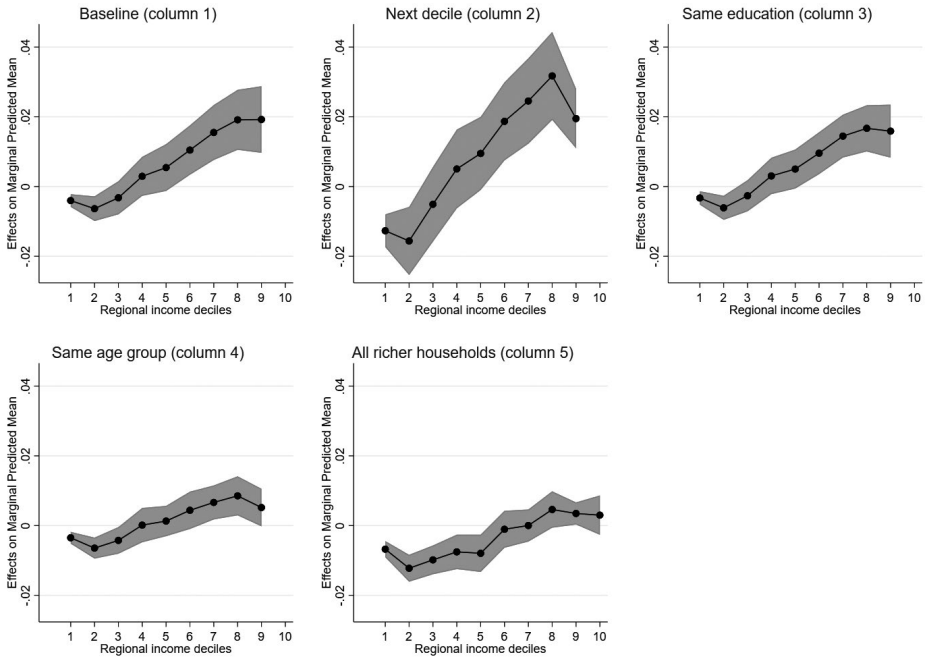

Figure 9. Marginal Effects—Various Reference Groups as Included in the Estimations in Table 3

Notes: Average marginal effects of relative reference income for selected specifications of estimations presented in Table 3. The grey area depicts the 95 percent confidence interval.

Source: OeNB Euro Survey, own calculations.

5.3. *Exploring Heterogeneity: Household Loans by Purpose*

In a next step, we investigate if the nexus between the reference income and existing household debt differs depending on the purpose of the "equilibrium" loans. We use information gathered by the OeNB Euro Survey from 2010 to 2014, where respondents were asked about the loan purpose according to four categories: mortgage, car, consumption, and other-purpose loans.

Traditionally, households in the CESEE countries show a high preference for owning their residences (Del Pero *et al.*, 2016). Apart from a high preference for buying a home, favorable credit supply conditions and increased availability of housing have led to a higher demand in mortgage-financed housing (see Rosan and Zauder, 2019). Accordingly, column (1) of Table 4 shows a high positive correlation between reference income and mortgage loans. Interestingly, the coefficients' signs are positive even in the third and fourth decile. This might suggest a stronger demand-driven link. For consumption and other-purpose loans, we see a pattern similar to our baseline specification. For loans financing cars, only the second decile shows a significant (negative) coefficient. Overall, the results in columns (1), (3), and (4) support the "Keeping up with the Novaks" hypothesis.

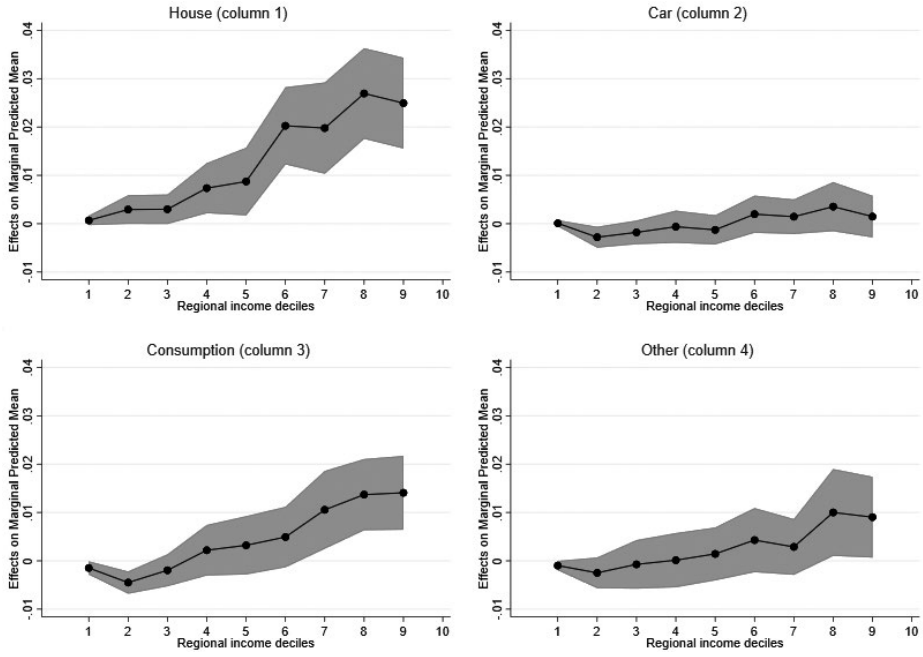

Figure 10. Marginal Effects of Relative Reference Income on Loans by Purpose at Income Deciles for the Estimations in Table 4

Notes: Average marginal effects of relative reference income for loan purposes as presented in Table 4. The grey area depicts the 95 percent confidence interval.

Source: OeNB Euro Survey, own calculations.

5.4. *The Currency Composition of Household Loans*

A potential concern for any causal interpretation of our results is omitted variable bias. In particular, if a factor influences both the income distribution and the probability of having a loan at the same time, this could lead to a bias in the estimated coefficients. Some endogeneity concerns could be alleviated if we take a look at the currency denomination of the loans, as described in detail below.

In the Euro Survey, respondents are asked to state the currency denomination of their loans and place their replies into one of the following five categories: (1) the whole amount is in foreign currency, (2) the whole amount is in local currency, (3) the loan is predominantly denominated in foreign or (4) in local currency, and (5) the loan is equally split. Based on the information whether a household has a loan that corresponds to one of these five categories, we perform several estimations. The estimation results are presented in Table 5.

Our baseline results show that the strongest link between reference income and household indebtedness can be found for loans that are fully denominated in foreign currency (this applies to nearly 20 percent of households with a loan) or

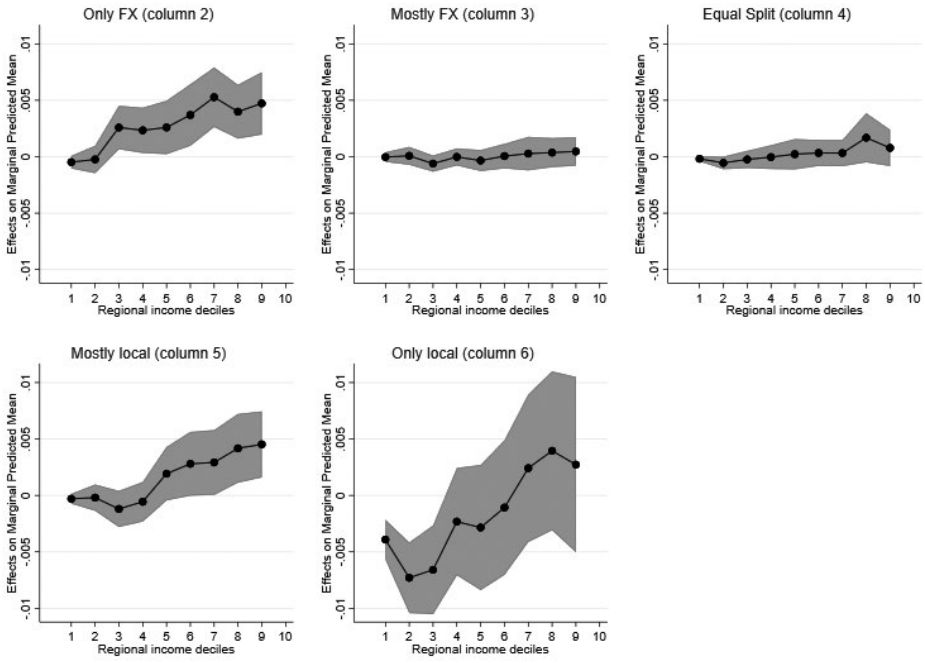

Figure 11. Marginal Effects at Income Deciles for the Estimations in Table 5

Notes: Average marginal effects of relative reference income for loans by currency composition presented in Table 5. The grey area depicts the 95 percent confidence interval.

Source: OeNB Euro Survey, own calculations.

for loans fully denominated in local currency (about 58 percent households with a loan). There is also some correlation for predominantly local currency loans. For the remaining categories (mostly foreign or equally split), no significant correlation can be found (however, the sample size of these categories is also significantly lower).

These results alleviate some omitted variable bias/endogeneity concerns, given that shocks that affect income inequality and loans at the same time are more likely to affect loans in local currencies but not loans denominated in foreign currency. An example of this are the favorable economic conditions leading to a higher level of local currency credit supply and higher reference income at the same time. As the survey data also show, most of the households with loans fully denominated in foreign currency have those loans in a foreign-owned bank (56 percent). Accordingly, if a shock causes the bias, the effect on foreign currency loans should be smaller or nil, which is, overall, not the case as our results show. The only notable differences can be observed in the lower part of the distribution. This might be related to a potentially stronger signaling effect for domestically owned banks, as domestically owned banks potentially acquire more information about the local income distribution.

6. Zooming-in on the Demand of Household Loans in CESEE

Looking at the literature, it becomes apparent that disentangling the effects of demand and supply factors on loans is a daunting task. As shown in Section 5, existing loans in our survey data constitute equilibrium transactions without any information on whether the result was supply- and/or demand-driven. In this section, we attempt to take a closer look at the demand side of the correlation between household loans and income inequality using a different question from our data set concerning the intentions of respondents to take up a loan. In particular, we ground our analysis in this section on a question where respondents have to indicate whether they plan to approach a bank to demand a loan in the following 12 months.[28] In a subsequent step, information is provided on the purpose of the intended loan (i.e., house, car, consumption, or other purposes).[29] For this part of the analysis, we use data from 2011 to 2018 due to data availability considerations.

The results are presented in Table 6. Overall, we find some indication for a positive correlation between a larger relative reference income and loan plans above the seventh decile, although this correlation tends to be weaker than in the case of current "equilibrium" loans. Interestingly, when testing whether this result is valid for all planned loan categories, it remains valid only for loans for car purchases. This could indicate conspicuous motives for loan demand as cars have been shown to be a highly visible (consumption) good (e.g., Heffetz, 2011). However, loans aimed at acquiring consumption goods do not seem to support the conspicuous motive. Overall, our data do not allow much interpretation in this direction. The results in columns (7) and (8) hint at threshold effects because the strongest correlation can be found for high inequality regions.

Loan intentions could also be influenced by households' sentiments about past developments with respect to the country or related to past experiences of the respondents themselves. Against that background, in Table 7 we include information from the OeNB Euro Survey on (1) developments of the financial situation of the household in the past 12 months, and (2) the current economic situation of the country.[30] Our results show that a good financial situation of the household and a good economic situation of the country increase the likelihood to plan to take out a loan. Similar to Table 3, we could not detect any significant effect for household savings. Including those variables does not alter our main results. Going forward, we investigated whether over the long period of time we cover in our paper there were any changes with respect to the correlation between the likelihood to plan a loan and income inequality. Therefore, we split the period before and after 2012 that marks the Euro area sovereign debt crisis that also impacted the CESEE countries. The results show that the correlation seems to have strengthened slightly post 2012 and only for the more affluent deciles pointing toward a higher risk aversion

[28]The question in the questionnaire is: "Do you plan to take out a loan within the next year and if so, in what currency?"

[29]"What is the purpose of your loan or your loans? To finance a house or apartment, for consumption goods (furniture, travelling, household appliances, etc.), for a car or for other purposes." This question has been included in the questionnaire since 2011.

[30]The questions in the survey are worded as follows: "Over the last 12 months, the financial situation of my household has got better" and "Currently, the economic situation of my country is very good." The replies of the respondents range between "strongly agree" and "strongly disagree" in five categories.

TABLE 4
Household Loans in CESEE by Purpose

	(1) House	(2) Car	(3) Consumption	(4) Other
Income distribution				
1st decile#relinc	0.008	0.000	−0.012*	−0.011**
	(0.006)	(0.005)	(0.006)	(0.005)
2nd decile#relinc	0.032**	−0.042**	−0.036***	−0.029
	(0.014)	(0.018)	(0.011)	(0.021)
3rd decile#relinc	0.033**	−0.021	−0.013	−0.004
	(0.014)	(0.020)	(0.015)	(0.026)
4th decile#relinc	0.057***	−0.013	0.013	0.006
	(0.020)	(0.024)	(0.019)	(0.030)
5th decile#relinc	0.053**	−0.030	0.014	0.012
	(0.023)	(0.022)	(0.019)	(0.027)
6th decile#relinc	0.106***	0.006	0.019	0.032
	(0.022)	(0.023)	(0.020)	(0.028)
7th decile#relinc	0.095***	−0.001	0.052**	0.023
	(0.025)	(0.023)	(0.022)	(0.028)
8th decile#relinc	0.114***	0.013	0.063***	0.064**
	(0.023)	(0.029)	(0.020)	(0.032)
9th decile#relinc	0.107***	0.005	0.079***	0.070**
	(0.024)	(0.027)	(0.021)	(0.030)
Income	0.000***	0.000**	−0.000***	−0.000***
		(0.000)	(0.000)	(0.000)
Cons	−3.694***	−3.723***	−3.748***	−3.416***
	(0.222)	(0.167)	(0.147)	(0.195)
ICC (regional)	0.014	0.011	0.022	0.017
N	56,841	56,841	56,841	56,841

Dependent variable: binary response indicating if respondents have a loan, by loan purpose. Estimation method: multilevel modeling. Sample where loan purpose question was included: 2010–2014. Country and time fixed effects included in all estimations. Intraclass correlation coefficient denotes the explained portion of the variance by inclusion of the regional (second) level covariates. Robust standard errors in parentheses. Coefficient estimates for sociodemographics variables excluded in the table. Variables are defined in appendix.

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

on the part of the banks but also to a lower demand. Finally, we tested whether not having already a loan would increase the likelihood of planning one and if so, for which households. The estimations in column (5) support our baseline results for the respondent without loans and in that case even less affluent households (i.e., in the fourth decile) increase the likelihood to plan to take out a loan.

7. Alternative Measures of Income Inequality

In a next step of our analysis, we also test the robustness of our estimations to alternative income inequality measures. For instance, we apply the Gini coefficient (ranging between 0 and 1) that is widely used in the literature instead of the relative reference income. The Gini coefficient incorporates movements in the whole income distribution, meaning that it is "noisier" than the stricter upward-looking comparison measure which we use in our baseline analysis. Therefore, we expect the effect of the Gini coefficient on household indebtedness to be smaller than the effect of the relative reference income. For instance, if a household is in the

middle of the distribution, Gini changes attributed to income losses at the bottom of the distribution might not affect the household's behavior because the household is comparing itself in an upward-looking rather than downward-looking way. Indeed, as column (1) of Table 8 shows, the overall pattern is similar to our baseline inequality measure, but the correlation shown by the coefficient is weaker.

Another alternative income inequality measure that we use is the ratio between income in the 90th and 10th percentile. As depicted in column 2, this estimation yields results that are qualitatively similar to the results using the relative reference income. In addition, we also use the ratio between the 75th and the 25th percentile (column 3), and the results show that the bottom and top deciles have (weakly) significant coefficients as compared to the baseline estimations.

Going forward, columns (4)–(7) show the results for the share of the top 1 percent, top 5 percent, and top 10 percent on total income, respectively, yielding very similar results to the baseline estimation with relative reference income. For the shares of the bottom 10 percent and bottom 20 percent, correlations are only present at the very bottom of the distribution. The last column finally calculates the difference between the log income of percentiles 90 and 10, as used by Coibion *et al*. (2014).

Overall, all the estimations above suggest that: (i) Our main result, that is, that higher income inequality is correlated with a lower probability of having a loan for the bottom third of the distribution and a higher probability of having a loan for the top of the distribution, is confirmed by various inequality measures other than the relative reference income used above. (ii) The strongest link between income inequality and debt can be captured by the P90/P10 percentile ratio and the share of the top 1 percent on total income. Generally, movements at the top of the income distribution seem to play a key role for both the signaling channel and the "Novaks" (i.e., demand) channel, while movements at the bottom of the regional income distribution tend to be less central.

8. Where Are "The Novaks"? Keeping Up with Other Regions

Until now, we have constructed the relative reference income using the income of more affluent households in the same region. In this section, we will focus on relaxing the assumption that respondents compare themselves only with "Novaks" from the same region. For that purpose, we also include households in neighboring regions and test whether this adds to the explanatory value. The reason for that is a high probability that households compare themselves not only to their "neighbors" (households in the same region) but also to other households in their relative vicinity because of various possible social ties. In addition, from a supply-side perspective, banks that operate trans-regionally could have clients in other regions and thus include information from all regions of operation in their considerations.

Against that background, we use an explanatory spatial analysis to test if including the relative reference income for neighboring regions affects our results.[31]

[31]Various other "neighbor" concepts are possible, for example, households closer than 100 km, 200 km, etc. The analysis performed here is only exploratory, and a full spatial analysis is beyond the scope of this paper.

TABLE 5
HOUSEHOLD LOANS IN CESEE BY CURRENCY

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	Only fx	Mostly fx	Equal	Mostly Local	Only Local
Income Distribution						
1st decile#relinc	−0.016***	−0.007*	−0.001	−0.010	−0.005	−0.021***
	(0.004)	(0.004)	(0.010)	(0.007)	(0.004)	(0.005)
2nd decile#relinc	−0.026***	−0.004	0.004	−0.036*	−0.003	−0.041***
	(0.008)	(0.009)	(0.018)	(0.021)	(0.010)	(0.010)
3rd decile#relinc	−0.013	0.033***	−0.033	−0.014	−0.022	−0.037***
	(0.010)	(0.011)	(0.022)	(0.024)	(0.016)	(0.012)
4th decile#relinc	0.011	0.030**	−0.001	−0.002	−0.010	−0.012
	(0.011)	(0.012)	(0.018)	(0.030)	(0.016)	(0.013)
5th decile#relinc	0.021	0.033**	−0.017	0.011	0.028*	−0.015
	(0.013)	(0.014)	(0.025)	(0.033)	(0.016)	(0.015)
6th decile#relinc	0.039***	0.045***	0.003	0.016	0.039**	−0.006
	(0.013)	(0.015)	(0.025)	(0.027)	(0.018)	(0.016)
7th decile#relinc	0.056***	0.060***	0.012	0.016	0.041**	0.012
	(0.014)	(0.013)	(0.031)	(0.027)	(0.019)	(0.016)
8th decile#relinc	0.068***	0.048***	0.016	0.061*	0.055***	0.019
	(0.015)	(0.013)	(0.026)	(0.032)	(0.018)	(0.017)
9th decile#relinc	0.069***	0.055***	0.019	0.034	0.058***	0.014
	(0.016)	(0.015)	(0.025)	(0.031)	(0.017)	(0.019)
Income	−0.000	0.000	0.000	−0.000	−0.000	−0.000**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cons	−3.575***	−4.134***	−4.581***	−4.156***	−3.758***	−3.348***
	(0.092)	(0.176)	(0.196)	(0.278)	(0.145)	(0.105)
ICC (regional)	0.015	0.022	0.010	0.021	0.018	0.015
N	116,666	116,666	116,666	116,666	116,666	116,666
Share of total loans		20.11%	4.42%	3.63%	13.61%	58.23%

Dependent variable: binary response indicating if respondents have a loan. Estimation method: multilevel modeling. Country and time fixed effects included in all estimations. Intraclass correlation coefficient denotes the explained portion of the variance by inclusion of the regional (second) level covariates. Robust standard errors in parentheses. Coefficient estimates for sociodemographics variables excluded in the table. Variables are defined in the Appendix.

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

TABLE 6
LOAN INTENTIONS OF HOUSEHOLDS IN CESEE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Const	All	House	Car	Consumption	Other	Low Ineq Regions	High Ineq Regions
Income Distribution								
1st decile#relinc		0.004	0.006	0.004	0.001	−0.005	0.032	0.005
		(0.006)	(0.004)	(0.006)	(0.009)	(0.008)	(0.043)	(0.004)
2nd decile#relinc		−0.028*	−0.012	−0.028*	−0.001	0.013	−0.009	0.015
		(0.017)	(0.015)	(0.017)	(0.026)	(0.014)	(0.040)	(0.010)
3rd decile#relinc		0.040*	−0.007	0.040*	0.022	−0.028	−0.005	0.013
		(0.023)	(0.026)	(0.023)	(0.029)	(0.026)	(0.048)	(0.015)
4th decile#relinc		−0.008	−0.063**	−0.008	0.034	0.009	−0.011	0.039*
		(0.030)	(0.029)	(0.030)	(0.036)	(0.031)	(0.050)	(0.021)
5th decile#relinc		0.029	−0.004	0.029	0.028	−0.019	0.003	0.079***
		(0.032)	(0.031)	(0.032)	(0.045)	(0.031)	(0.050)	(0.027)
6th decile#relinc		0.015	0.010	0.015	0.038	0.034	−0.022	0.027
		(0.025)	(0.025)	(0.030)	(0.041)	(0.037)	(0.055)	(0.030)
7th decile#relinc		0.044*	0.022	0.044*	0.034	0.014	0.008	0.046*
		(0.038)	(0.038)	(0.025)	(0.041)	(0.034)	(0.057)	(0.025)
8th decile#relinc		0.026*	0.007	0.026	0.023	−0.004	−0.008	0.065**
		(0.025)	(0.025)	(0.025)	(0.045)	(0.035)	(0.055)	(0.032)
9th decile#relinc		0.064**	0.005	0.064**	0.022	−0.006	0.014	0.045**
		(0.027)	(0.33)	(0.027)	(0.038)	(0.035)	(0.054)	(0.022)
Sociodemographics								
Female		−0.145***	0.007	−0.146***	0.024	−0.069***	−0.191	−0.007
		(0.029)	(0.022)	(0.029)	(0.019)	(0.023)	(0.014)	(0.032)
Age		0.021***	0.021**	0.021***	0.028***	0.037***	0.029***	0.038***
		(0.007)	(0.006)	(0.007)	(0.005)	(0.007)	(0.004)	(0.006)
Age squared		−0.000***	−0.000***	0.000***	0.000***	0.000***	−0.000***	−0.00***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Children		0.004	0.029*	−0.004	0.023	0.016	0.010	0.019
		(0.020)	(0.015)	(0.020)	(0.017)	(0.017)	(0.013)	(0.013)
Education		0.056**	0.035	0.056**	−0.027	−0.001	0.061***	0.057**
		(0.027)	(0.028)	(0.027)	(0.029)	(0.030)	(0.019)	(0.027)

(Continues)

TABLE 6 (CONTINUED)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Const	All	House	Car	Consumption	Other	Low Ineq Regions	High Ineq Regions
Unemployed		−0.183***	−0.189***	−0.183***	−0.181***	−0.060	−0.176***	−0.211***
		(0.045)	(0.026)	(0.045)	(0.047)	(0.041)	(0.019)	(0.033)
Self-employed		−0.015	0.146***	0.015	−0.043***	0.200	0.099***	0.096
		(0.063)	(0.052)	(0.063)	(0.050)	(0.042)	(0.039)	(0.079)
Student		−0.580***	−0.742***	−0.580***	−0.616***	−0.101	−0.564***	−0.455***
		(0.086)	(0.73)	(0.086)	(0.084)	(0.080)	(0.063)	(0.084)
Cons	−1.423***	−1.556***	−1.556***	−2.135***	−0.743***	−0.105	−1.591***	−1.94***
	(0.025)	(0.132)	(0.132)	(0.175)	(0.132)	(0.133)	(0.166)	(0.166)
ICC (regional)	0.039	0.013	0.013	0.013	0.017	0.014	0.015	0.012
N	113,722	71,322	71,322	71,322	71,322	76,297	76,297	23,390

Dependent variable: the share of respondents, who plan to take out a loan in the following 12 months (dummy variable). Estimation method: multilevel modeling. Country and time fixed effects included in all estimations. Intraclass correlation coefficient denotes the explained portion of the variance by inclusion of the regional (second) level covariates. Robust standard errors in parentheses. Variables are defined in the Appendix.

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

TABLE 7
LOAN INTENTIONS OF HOUSEHOLDS IN CESEE

	(1)	(2)	(3)	(4)	(5)
	Sentiments	Savings	Pre-2012	Post-2012	Without Current Loan
Income Distribution					
1st decile#relinc	−0.001	−0.001	−0.001	0.005	0.005
	(0.005)	(0.004)	(0.006)	(0.005)	(0.005)
2nd decile#relinc	−0.008	−0.001	−0.002	0.015	0.009
	(0.009)	(0.009)	(0.017)	(0.012)	(0.010)
3rd decile#relinc	0.000	0.000	−0.000	0.021	0.014
	(0.010)	(0.015)	(0.016)	(0.023)	(0.018)
4th decile#relinc	0.011	0.011	0.011	0.034	0.037*
	(0.017)	(0.017)	(0.021)	(0.026)	(0.021)
5th decile#relinc	0.046***	0.030*	0.047**	0.037	0.051**
	(0.017)	(0.018)	(0.024)	(0.026)	(0.023)
6th decile#relinc	0.029	0.002	−0.011	0.046*	0.091
	(0.020)	(0.021)	(0.028)	(0.028)	(0.023)
7th decile#relinc	0.044*	0.026	0.031	0.059**	0.045*
	(0.038)	(0.018)	(0.025)	(0.026)	(0.026)
8th decile#relinc	0.068***	0.022	0.034	0.039	0.032
	(0.022)	(0.022)	(0.027)	(0.026)	(0.022)
9th decile#relinc	0.072***	0.033**	0.048**	0.047**	0.041**
	(0.019)	(0.017)	(0.021)	(0.026)	(0.022)
Sentiments					
Past sit HH	0.049**				
	(0.024)				
Econ sit. country	0.119***				
	(0.030)				
Savings		0.026			
		(0.020)			
Sociodemographics					
Female	−0.008	−0.011	0.006	−0.030*	0.011
	(0.018)	(0.014)	(0.021)	(0.017)	(0.016)

(Continues)

TABLE 7 (CONTINUED)

	(1) Sentiments	(2) Savings	(3) Pre-2012	(4) Post-2012	(5) Without Current Loan
Age	0.030***	0.030***	0.031**	0.031***	0.027***
	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)
Age squared	−0.000***	−0.000***	−0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Children	0.001	0.011	0.030	0.032**	0.017
	(0.013)	(0.010)	(0.014)	(0.013)	(0.011)
Education	0.058***	0.067**	0.069***	0.058**	0.064***
	(0.019)	(0.020)	(0.020)	(0.029)	(0.020)
Unemployed	−0.184***	−0.186***	−0.185***	−0.177***	−0.186***
	(0.023)	(0.045)	(0.028)	(0.034)	(0.023)
Self-employed	0.090*	0.104***	0.060	0.110***	0.094***
	(0.049)	(0.034)	(0.045)	(0.037)	(0.036)
Student	−0.536***	−0.508***	−0.505***	−0.603***	−0.578***
	(0.063)	(0.056)	(0.081)	(0.061)	(0.064)
Cons	−1.677***	−1.556***	−1.679***	−2.135***	−1.631***
	(0.153)	(0.132)	(0.143)	(0.175)	(0.116)
ICC (regional)	0.018	0.013	0.021	0.013	0.019
N	59,018	84,995	51,775	71,322	76,769

Dependent variable: the share of respondents, who plan to take out a loan in the following 12 months (dummy variable). Estimation method: multilevel modeling. Country and time fixed effects included in all estimations. Intraclass correlation coefficient denotes the explained portion of the variance by inclusion of the regional (second) level covariates. Robust standard errors in parentheses. Variables are defined in the Appendix.

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

TABLE 8
ALTERNATIVE INEQUALITY MEASURES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Gini	P90/P10	P75/P10	Top1%	Top5%	Top10%	Bottom10%	Bottom20%	logP90–logP10
Income Distribution									
1st decile#ineq	−0.332*	−0.025***	−0.077***	−1.422***	−0.696***	−0.407**	−3.608*	−1.857**	−0.091***
	(0.186)	(0.005)	(0.029)	(0.442)	(0.225)	(0.204)	(1.988)	(0.906)	(0.033)
2nd decile#ineq	−0.147	−0.014***	−0.047	−0.737*	−0.377*	−0.189	−3.076	−1.502*	−0.052
	(0.170)	(0.004)	(0.029)	(0.428)	(0.193)	(0.179)	(1.981)	(0.909)	(0.033)
3rd decile#ineq	−0.021	−0.008**	−0.023	−0.605**	−0.209	−0.050	−1.309	−0.804	−0.018
	(0.177)	(0.004)	(0.030)	(0.297)	(0.193)	(0.187)	(1.850)	(0.860)	(0.033)
4th decile#ineq	0.112	0.000	−0.003	−0.089	0.037	0.112	−1.041	−0.627	0.010
	(0.173)	(0.004)	(0.030)	(0.258)	(0.182)	(0.181)	(1.888)	(0.860)	(0.033)
5th decile#ineq	0.146	0.002	0.001	−0.060	0.101	0.158	−0.443	−0.422	0.018
	(0.192)	(0.004)	(0.031)	(0.296)	(0.221)	(0.207)	(2.083)	(0.910)	(0.034)
6th decile#ineq	0.231	0.007*	0.017	0.178	0.239	0.257	0.449	−0.024	0.036
	(0.186)	(0.004)	(0.031)	(0.270)	(0.205)	(0.197)	(1.773)	(0.815)	(0.034)
7th decile#ineq	0.302	0.009**	0.026	0.494**	0.379*	0.345*	0.420	−0.011	0.050
	(0.188)	(0.004)	(0.031)	(0.239)	(0.206)	(0.200)	(1.896)	(0.844)	(0.034)
8th decile#ineq	0.385**	0.014***	0.040	0.666***	0.492**	0.435**	0.615	0.204	0.072**
	(0.185)	(0.004)	(0.032)	(0.237)	(0.195)	(0.193)	(1.844)	(0.823)	(0.035)
9th decile#ineq	0.346*	0.012***	0.032	0.675**	0.454**	0.396*	0.095	−0.042	0.060*
	(0.191)	(0.003)	(0.030)	(0.287)	(0.218)	(0.206)	(1.907)	(0.844)	(0.033)
10th decile#ineq	0.457**	0.017***	0.050*	0.864**	0.636***	0.532***	1.332	0.483	0.085***
	(0.183)	(0.003)	(0.028)	(0.350)	(0.216)	(0.199)	(2.105)	(0.938)	(0.030)
Cons	−3.543***	−3.494***	−3.482***	−3.487***	−3.510***	−3.538***	−3.465***	−3.451***	−3.506***
	(0.114)	(0.089)	(0.109)	(0.085)	(0.097)	(0.109)	(0.096)	(0.099)	(0.104)
ICC (regional)	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
N	131,923	131,570	131,923	131,923	131,923	131,923	131,923	131,923	131,923

Dependent variable: binary response indicating if respondents have a loan. Estimation method: multilevel modeling. Country and time fixed effects included in all estimations. Intraclass correlation coefficient denotes the explained portion of the variance by inclusion of the regional (second) level covariates. Sociodemographic variables included in all estimations, omitted in the table. Robust standard errors in parentheses. Variables are defined in the Appendix.

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

TABLE 9
Spatial Dimension: Neighboring Regions

	(1)	(2)	(3)
	Baseline	*Relinc Neighbor*	*Relinc Country*
Income Distribution			
1st decile#relinc	−0.016***	−0.013	−0.015***
	(0.004)	(0.010)	(0.004)
2nd decile#relinc	−0.026***	−0.041*	−0.022***
	(0.008)	(0.022)	(0.008)
3rd decile#relinc	−0.013	−0.011	−0.009
	(0.010)	(0.025)	(0.011)
4th decile#relinc	0.011	−0.006	0.015
	(0.011)	(0.032)	(0.012)
5th decile#relinc	0.021	0.015	0.021
	(0.013)	(0.036)	(0.014)
6th decile#relinc	0.050***	0.080*	0.039***
	(0.014)	(0.041)	(0.013)
7th decile#relinc	0.056***	0.070**	0.055***
	(0.014)	(0.031)	(0.014)
8th decile#relinc	0.068***	0.078***	0.068***
	(0.015)	(0.026)	(0.016)
9th decile#relinc	0.069***	0.048*	0.078***
	(0.016)	(0.029)	(0.018)
1st decile#relinc neighbor		−0.003	
		(0.009)	
2nd decile#relinc neighbor		0.015	
		(0.021)	
3rd decile#relinc neighbor		−0.002	
		(0.024)	
4th decile#relinc neighbor		0.017	
		(0.030)	
5th decile#relinc neighbor		0.006	
		(0.034)	
6th decile#relinc neighbor		0.005	
		(0.029)	
7th decile#relinc neighbor		−0.014	
		(0.028)	
8th decile#relinc neighbor		−0.010	
		(0.025)	
9th decile#relinc neighbor		0.019	
		(0.020)	
Cons	−3.575***	−3.575***	−3.577 ***
	(0.092)	(0.093)	(0.091)
ICC (regional)	0.015	0.015	0.014
N	116,666	116,666	116,462

Dependent variable: binary response indicating if respondents have a loan. Estimation method: multilevel modeling. *relinc neighbor* denotes the relative reference income of households in higher income deciles in neighboring regions of the same country. Country and time fixed effects included in all estimations. Intraclass correlation coefficient denotes the explained portion of the variance by inclusion of the regional (second) level covariates. Coefficient estimates for sociodemographics variables omitted in the table. Robust standard errors in parentheses. Variables are defined in the Appendix.

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

To disentangle "home region" and "neighbor region" effects, we construct the neighbors' reference income separately. *Relinc neighbor* denotes the relative reference income of households in higher income deciles in neighboring regions of the same country. Table 9 depicts the baseline regression in column (1), while column

(2) contains the added neighbor region income effect. This result suggests that the bulk of effects is for households in the same region, not neighboring regions. However, as some of the results for the influence of "home region"-relative income appear to lose their significance, possible co-movements between relative reference income at home and in neighboring regions could be at work.

Finally, we test a different definition for the reference group, namely that the reference group is defined as households in higher income deciles in the same country (as opposed to the same region). The results presented in column (3) show that our main results are robust, with negative effects of higher relative reference income at the bottom of the distribution and positive effects on loans at the top. This result suggests that the "Novaks" seem to be present at both country and regional levels, with little changes in quantitative terms of our estimates.

9. Conclusions

In this paper, we analyze the nexus between household debt and income inequality in CESEE, using data from a household survey for the period 2008–2018. We compute and apply various income inequality measures on the regional level—a primer for many countries and years. Our empirical analysis aims at pinpointing the association between income inequality and the probability of currently being indebted or planning to take out a loan. As households are nested into regions, we use multilevel models that account for intra-regional correlation.

Our main findings are twofold. First, there is a negative correlation between the relative mean income of richer households and the probability of being indebted for the very bottom of the distribution. This result suggests a strong signaling effect of income inequality, meaning that presumably, banks react to rising income inequality by targeting richer households with more, and cheaper, credit while restraining credit supply to poorer households. Second, there is a positive correlation between relative reference income and the probability of having a loan for the upper half of the regional income distribution. This result is consistent with the aforementioned supply-side channel, but also with a demand-side channel ("Keeping up with the Novaks"). This result is robust against variations in the definitions of relative reference income and holds also when using traditional inequality measures such as the Gini coefficient, 90/10 percentile ratios, and the shares of the top 1 percent and top 5 percent of the income earners.

We also take a very disaggregated view of the nexus between the demand for loans and income inequality by looking at loans with different purposes. Although the coefficients in this case are quantitatively weaker, we can confirm a positive correlation at the very top of the distribution, supporting the "Keeping up with the Novaks" effect.

References

Atkinson, A. B., "Pareto and the Upper Tail of the Income Distribution in the UK: 1799 to the Present," *Economica*, 84, 129–56, 2017.

Bazillier, R., and J. Hericourt, "The Circular Relationship Between Inequality, Leverage, and Financial Crises," *Journal of Economic Surveys*, 31, 463–96, 2017.
Belabed, C. A., and M. Hake, "Income Inequality and Trust in National Governments in Central, Eastern and Southeastern Europe," Working Paper 222, Oesterreichische Nationalbank, 2018.
Bertrand, M., and A. Morse, "Trickle-Down Consumption," *Review of Economics and Statistics*, 98, 863–79, 2016.
Blanchet, T., J. Fournier, and T. Piketty, "Generalized Pareto Curves: Theory and Applications," Wid working paper 2017/3, World Inequality Database. 2017.
Brown, S., P. Ghosh, and K. Taylor, "Household Finances and Social Interaction: Bayesian Analysis of Household Panel Data," *Review of Income and Wealth*, 62, 467–88, 2016.
Carr, M. D., and A. Jayadev, "Relative Income and Indebtedness: Evidence From Panel Data," *Review of Income and Wealth*, 61, 759–72, 2015.
Charles, K. K., E. Hurst, and N. Roussanov, "Conspicuous Consumption and Race," *Quarterly Journal of Economics*, 124, 425–67, 2009.
Clark, A. E., and C. Senik, "Who Compares To Whom? The Anatomy of Income Comparisons in Europe" *Economic Journal*, 120, 573–94, 2010.
Coibion, O., Y. Gorodnichenko, M. Kudlyak, and J. Mondragon, "Does Greater Inequality Lead to More Household Borrowing? New Evidence From Household Data," National Bureau of Economic Research Working Paper, 2014.
Danzer, A. M., B. Dietz, K. Gatskova, and A. Schmillen, "Showing Off to the New Neighbors? Income, Socioeconomic Status and Consumption Patterns of Internal Migrants" *Journal of Comparative Economics*, 42, 230–45, 2014.
Del Pero, A . S., W. Adema, V. Ferraro, and V. Frey, "Policies to Promote Access to Good-Quality Affordable Housing in OECD Countries," OECD Social, Employment and Migration Working Papers, 2016.
Drechsel-Grau, M., and K. D. Schmid, "Consumption-Savings Decisions Under Upward-Looking Comparisons," *Journal of Economic Behavior & Organization*, 106, 254–68, 2014.
Duesenberry, J. S., *Income, Saving, and the Theory of Consumer Behavior*. Harvard University Press,Cambridge, 1949.
ECB, "The Eurosystem Household Finance and Consumption Survey: Methodological Report for the Second Wave," Technical Report, ECB Statistics Paper, 2016.
Eckerstorfer, P., J. Halak, J. Kapeller, B. Schütz, F. Springholz, and R. Wildauer, "Correcting for the Missing Rich: An Application to Wealth Survey Data," *Review of Income and Wealth*, 62, 605–27, 2016.
Fasianos, A., H. Raza, and S. Kinsella, "Exploring the Link Between Household Debt and Income Inequality: An Asymmetric Approach," *Applied Economics Letters*, 24, 404–09, 2017.
Ferrer-i-Carbonell, A., "Income and Well-Being: An Empirical Analysis of the Comparison Income Effect," *Journal of Public Economics*, 89, 997–1019, 2005.
Frank, R., A. Levine, and O. Dijk, "Expenditure Cascades, Review of Behavioral," *Economics*, 1, 55–73, 2014.
Groves, R. M., and E. Peytcheva, "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis," *Public Opinion Quarterly*, 72, 167–89, 2008.
Gulyás, A., *Item Nonresponse in the OeNB Euro Survey* (Ph.D. thesis). University of Vienna. 2009.
Hake, M., and P. Poyntner, "Household Loans in CESEE From A New Perspective: The Role of Income Distribution," *Focus on European Economic Integration*, 75–93, 2019.
Harrell, F. E., *Regression Modeling Strategies, with Applications to Linear Models, Survival Analysis and Logistic Regression*. Springer-Verlag, New York, 2001.
Heffetz, O., "A Test of Conspicuous Consumption: Visibility and Income Elasticities," *Review of Economics and Statistics*, 93, 1101–17, 2011.
Hirschman, A. O., and M. Rothschild, "The Changing Tolerance for Income Inequality in the Course of Economic Development: With A Mathematical Appendix," *Quarterly Journal of Economics*, 87, 544–66, 1973.
Iacoviello, M., "Household Debt and Income Inequality, 1963–2003," *Journal of Money, Credit and Banking*, 40, 929–65, 2008.
Ireland, N. J., "On limiting the Market for Status Signals," *Journal of Public Economics*, 53, 91–110, 1994.
Jestl, S., *The Impact of Income Inequality on Household Indebtedness in Euro Area Countries*. Vienna Institute for International Economic Studies, 2019.
Knight, J., S. Lina, and R. Gunatilaka, "Subjective Well-Being and its Determinants in Rural China," *China Economic Review*, 20, 635–49, 2009.
Kumhof, M., R. Rancière, and P. Winant, "Inequality, Leverage, and Crises," *American Economic Review*, 105, 1217–45, 2015.

Loschiavo, D., "Household Debt and Income Inequality: Evidence From Italian Survey Data," Bank of Italy Temi di Discussione Working Paper No 1095, 2016.
Maas, C. J., and J. J. Hox, "Sufficient Sample Sizes for Multilevel Modeling," *Methodology*, 1, 86–92, 2005.
Mian, A., and A. Sufi, *House of Debt: How They (and You) Caused The Great Recession, and How We Can Prevent it From Happening Again*. University of Chicago Press, 2015.
Piketty, T., and E. Saez, "Income Inequality in the United States, 1913–1998," *Quarterly Journal of Economics*, 118, 1–41, 2003.
Rabe-Hesketh, S., and A. Skrondal, "Multilevel and Longitudinal Modeling Using Stata," Technical Report, StataCorp LP, 2012.
Rosan, M., and K. Zauder, "Risky Business or Basic Needs Fulfilment? An Analysis of Croatian Households' Debt," *Presentation at the Twenty-Fifth Dubrovnik Economic Conference*. 2019.
Schunck, R., "Cluster Size and Aggregated Level 2 Variables in Multilevel Models," *A Cautionary Note, Methods, Data, Analyses*, 10, 120, 2016.
Senik, C., "When Information Dominates Comparison: Learning From Russian Subjective Panel Data," *Journal of Public Economics*, 88, 2099–123, 2004.
_____, "Direct Evidence On Income Comparisons and Their Welfare Effects," *Journal of Economic Behavior & Organization*, 72, 408–24, 2009.
Van Treeck, T., "Did Inequality Cause the US Financial Crisis?," *Journal of Economic Surveys*, 28, 421–48, 2014.
Veblen, T., *The Theory of the Leisure Class. An Economic Study in the Evolution of Institutions*. The Macmillan Company, New York/London, 1899.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site:

APPENDIX

Figure A1: An Example for the Relationship Between Income, Reference Income and Relative Reference Income *Note*: This figure shows the relationship between income (triangles, left axis), mean reference income (rectangles, left axis) and relative reference income (dots, right axis) for the North Albania region 2015. The relative reference income relinc is the mean income of households in higher deciles divided by the households' own income. The vertical lines depict the cut-off points of regional income deciles, denoted d1-d9. Source: OeNB Euro Survey, Own Calculations.

Table A1: Variables description

Table A2: Summary Statistics of Main Variables

Table A3: Descriptive Statistics (Average Over 2009 to 2018, Weighted)

Table A4: Household Loans in CESEE—Country and Region Levels

Table A5: Household Loans in CESEE—Probit Estimation