European Commission

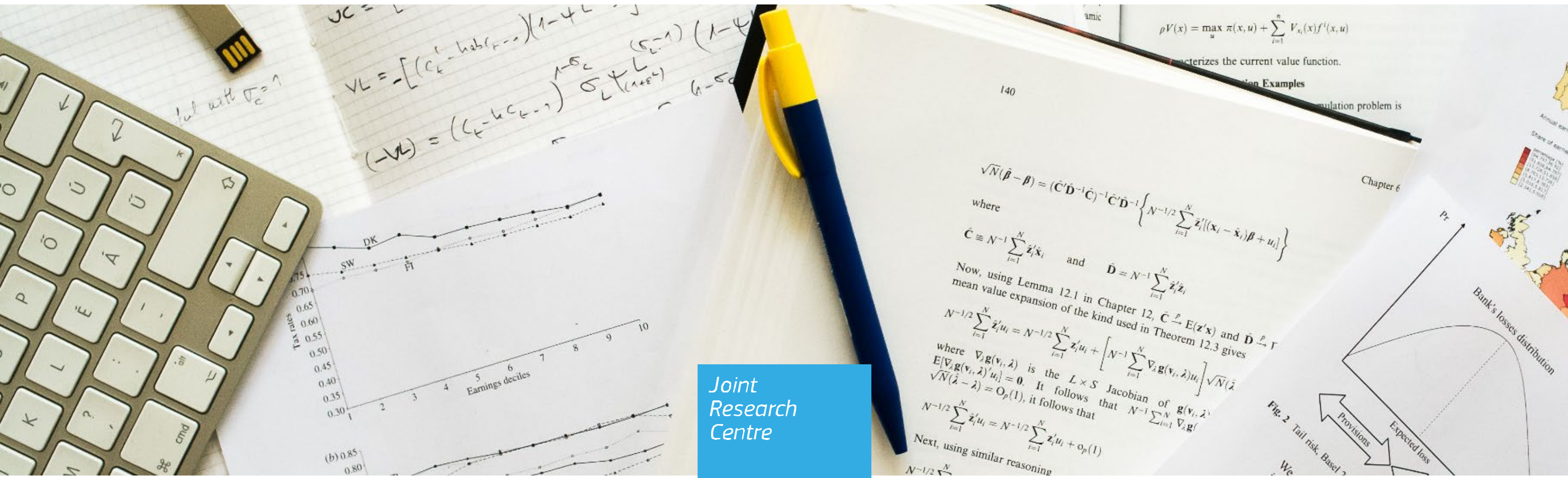
# The vulnerability aspect of happiness: The European middle class perspective


Benczur, P.

Kvedaras, V.


*JRC Working Papers in Economics and Finance, 2023/5*

2023



Joint Research Centre

**Contact information**
Name: Peter Benczur
Email: peter.benczur@ec.europa.eu

**EU Science Hub**
https://joint-research-centre.ec.europa.eu

JRC132978

Ispra: European Commission, 2023



How to cite this report: Benczur, P., and Kvedaras, V., *The vulnerability aspect of happiness: The European middle class perspective, JRC Working Papers in Economics and Finance, 2023/5,* European Commission, Ispra, Italy, 2023, JRC132978.

## Executive summary

This study evaluates the importance of middle class vulnerability to poverty for subjective well-being (happiness) in European countries. A solid middle class is necessary for political stability, an active society, and successful economic development. The size of the middle class needs to be both large and sufficiently stable. A sizeable but vulnerable middle class is a sign of weakness and may pose a risk to society, potentially leading to economic and political sustainability issues and social problems after even only mild shocks.

Most of the recent studies on the middle class in European countries have concentrated on its size. The vulnerability of the middle class has been evaluated only indirectly, through the identification of worrying structural changes, but without an explicit measure of vulnerability. In contrast, this paper uses an explicit measure of vulnerability to poverty in order to gauge the income vulnerability and its importance for the well-being of middle-class members. In particular, we look at the probability of becoming poor, or more generally, falling under certain critical income thresholds. The methodology of Chaudhuri et al. (2002) then serves as a basis for developing an extension that ensures a better distributional fit and assessment of vulnerability.

Using the special module of the European Union Statistics on Income and Living Conditions micro-data on well-being and national aggregates from the World Happiness Report, the vulnerability indicator used is shown to be statistically significant for happiness perceptions at both the individual and aggregate levels. At the micro level, the economic importance of income vulnerability for the middle class is on a par with (or even greater than) economic factors such as severe material deprivation, difficulty of finding a job, or the acute financial burden of housing or debt repayments. The aggregate income vulnerability of the middle class is also one of the most significant factors affecting national happiness, along with gross domestic product per capita and perceptions of corruption.

We also find that the more relevant critical income threshold for the middle class studied in this paper is the lower income bound of the middle class (75% of the national median income) and not the poverty line (60% of the median). It looks therefore that it is the prospect of the loss of income and potentially social status – rather than the fear of deprivation – that matters for middle class members. A general income insecurity concept connected with class-specific income thresholds might therefore be needed for the analysis of middle (and upper) class well-being.

Our results underscore the importance of income vulnerability and insecurity, and indicate that there are large costs of anxiety related to the (severe) downside risk of income. They therefore support policies to reduce income uncertainty, especially for the lower-middle class. Finally, the results derived using only the aggregate data – and therefore to be taken with caution – underline the importance of corruption perceptions for national happiness, and might thus serve to remind politicians and policy makers that reducing corruption and setting common rules is important for the subjective perception of wellbeing and the happiness of nations.

# The vulnerability aspect of happiness: The European middle class perspective

Péter Benczúr and Virmantas Kvedaras*

European Commission, Joint Research Centre, Ispra (VA), Italy

17th May 2023

**Abstract**

This paper evaluates the importance of middle class vulnerability to poverty for subjective well-being (happiness) in European countries. Using the special module of the European Union Statistics on Income and Living Conditions micro-data on well-being and national aggregates from the World Happiness Report, the vulnerability indicator used is shown to be statistically significant for happiness perceptions at both the individual and aggregate levels. At the micro level, the economic importance of income vulnerability for the middle class is on a par with (or even greater than) economic factors such as severe material deprivation, difficulty of finding a job, or the acute financial burden of housing or debt repayments. We find that the lower income bound of the middle class (75% of the national median income) and not the poverty line (60% of the median) is more relevant for individual hapiness of middle class members. The aggregate income vulnerability of the middle class is also one of the most significant factors affecting national happiness, along with gross domestic product per capita and perceptions of corruption. The results indicate that anxiety related to the (severe) downside risk of income has large costs in terms of happiness. They therefore support the development of policies to reduce income uncertainty.

Keywords: happiness, income insecurity, middle class, poverty, subjective well-being, vulnerability.

*The opinions expressed are those of the authors only and do not necessarily represent the European Commission's official position.

# 1 Introduction

A solid middle class is necessary for political stability, an active society, and successful economic development (Alesina and Perotti, 1996; Easterly, 2001; Stiglitz, 2012; OECD, 2019). The size of the middle class needs to be both large and sufficiently stable. A sizeable but vulnerable middle class is a sign of weakness and may pose risks to the society, potentially leading to economic and political sustainability issues and social problems after even only mild shocks. Most of the recent studies on the middle class in European countries have concentrated on its size, with emerging concerns about the potentially decreasing (ILO, 2016, OECD, 2019, and Dendorfer and Kranzinger, 2021), volatile (Eurofound, 2019) or rather constant (Salido and Carabaña, 2020) middle class. Conclusions partially depend on the period analyzed and the definition of the middle class that is chosen (see, e.g., Atkinson and Brandolini, 2013, on this issue).

The vulnerability of the middle class has been evaluated only indirectly in these studies, through the identification of worrying structural changes, but without an explicit measure of vulnerability. However, the literature identifies several alarming global trends that have recently weakened the middle class. These include, among others, automation and robotisation, digitalisation, the rise of artificial intelligence, and increasing housing costs (see, e.g., OECD, 2019, and Korinek et al., 2021).

There is scant literature using explicit indicators of middle class vulnerability to analyse the experience of the European middle class.[1] This is especially true for studies examining the impact of middle class vulnerability on subjective well-being indicators such as happiness. This is partially because emerging broader economic insecurity concerns have shifted attention towards alternative aspects and multidimensional problems (see, e.g., Osberg and Sharpe, 2014, Hacker, 2018, Richiardini and He, 2020, Romaguera-de-la-Cruz, 2020, Ranci et al., 2021, as well as Osberg, 2021, for a recent general overview). New views have also emerged that define the middle class using income vulnerability itself, instead of measuring the vulnerability of a middle class defined in more traditional ways (see, e.g., López-Calva and Ortiz-Juarez, 2014). This approach was recently used by Simona-Moussa (2020) to identify vulnerable and less vulnerable population groups in Switzerland. The study showed that happiness is sig-

[1] An important exception is Bussolo et al. (2018), who found that the size of the middle class has remained similar in the EU countries, but its vulnerability has increased over time. Somewhat more distantly but equally importantly, Ranci et al. (2021) used a multidimensional approach within a more general view of economic insecurity, and found that economic insecurity is already expanding to the EU middle class.

nificantly lower only for the most vulnerable parts of the population. However, the results hold barely at the 10% level of statistical significance and there is no explicit evaluation of absolute vulnerability of the middle class.

In contrast, this paper uses a more traditional, vulnerability-to-poverty-measuring approach to gauge the vulnerability of a given middle class defined in the standard way, relying on household equivalised income[2] relative to the median income (see, e.g., Atkinson and Brandolini, 2013, on the middle class definition with a number of potential alternatives). The methodology of Chaudhuri et al. (2002) and Chaudhuri (2003) is used as a basis for the measurement of individual income vulnerability, with a proposed additional transformation of errors that ensures a better distributional fit and assessment of vulnerability.[3] It was also extended to an aggregate vulnerability measure of the middle class as a whole. Following Eurofound (2019) and OECD (2019), the middle class is defined in this paper as the population from households whose equivalised net income is between three-quarters–to–twice the median of household equivalised net income in their country. This approach has recently been applied using a similar methodology to evaluate how the risk of poverty affects happiness by Caria and Falco (2018), albeit not for the middle class, without covering the aggregate level, nor for a broad spectrum of European countries.

We also considered the approach proposed by López-Calva and Ortiz-Juarez (2014), but this requires longitudinal/panel data. Good quality panels are available only for few European countries. The quality of the EU SILC longitudinal data lags far behind the quality of its cross-sectional data because of much smaller samples and calibration issues (see, e.g., Krell et al., 2017). A more traditional approach that required only cross-sectional data was therefore preferred. Despite the 'traditional approach' used, the vulnerability to poverty indicator was highly statistically significant at both the individual and aggregate levels. The inclusion of vulnerability also sizably reduced the relevance of income for happiness. Additional augmentation with financial cushion indicators rendered the income—but not the vulnerability—indicator insignificant in most cases.

[2]The *European Union Statistics on Income and Living Conditions* (EU SILC) equivalised income indicator is used, which draws on the OECD equivalence scale (see `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Equivalised_income`).

[3]The proposed approach might significantly affect the results of known (and recent) extensions of vulnerability measurement, including, e.g., Hohberg et al. (2018) and Rohde et al. (2020). It can be expected to yield a more precise valuation of vulnerability and insecurity when the distribution even of the logarithm of income has fat tails. However, it is not granted that a straightforward functional transformation used in this paper fits every empirical case. More general extensions, including the use of nonparametric estimators, will therefore be discussed in Section 3.1.1.

Considered at the micro-level, the conditional probability of middle class individuals becoming poor was highly statistically significant in explaining their subjective evaluation of happiness in the largest EU countries and the UK. The economic importance of the income vulnerability indicator was on a par with (or even greater than) economic factors such as severe material deprivation, inability to find a job, or having a heavy financial burden, for example, from housing or debt repayments. Along with gross domestic product per capita and perceptions of corruption, the vulnerability to poverty of the middle class was a significant factor for the aggregate national happiness scores, too. Many other control variables were insignificant.

Looking from a broader perspective, the paper goes beyond contributing solely to the literature on the middle class and its vulnerability in European countries. It also adds to the literature in other interconnected areas. First, it is essential to underscore that vulnerability to poverty is linked to expectations about potential negative outcomes. This relates the findings to the more general literature on economic insecurity, which stresses the importance of expected income and anxiety connected to potential downside risks (see, e.g., Osberg, 2021). The proposed methodological extension is also equally relevant for any downside risk measurement that relies on a conditional income model (see, e.g., Rohde et al., 2020). Second, the empirical importance of income vulnerability for happiness at the individual and aggregate levels also contributes to a broader subjective well-being literature that has flourished recently in connection with the need to understand the drivers of individual and aggregate well-being by going beyond the usual income or gross domestic product per capita indicators (see, e.g., Fleurbaey, 2009, Frey and Stutzer, 2010, 2019, Stiglitz et al., 2018, and Van den Bergh, 2022).

The paper is structured as follows. Section 2 describes the employed data and variables. Section 3 defines the methodology, including the measurement of individual and aggregate vulnerability of the middle class and the econometric models used. Section 4 presents the main empirical results at both individual and aggregate levels. Section 5 adds a few important extensions, and Section 6 concludes.

# 2 Data and variables

This section outlines the data and variables used in the study. Subsection 2.1 describes the micro data and corresponding variables, and Subsection 2.2 covers the country-level data and variables.

## 2.1 Micro level data and variables

To investigate individual vulnerability at the micro level, and to derive the corresponding aggregate measure of middle class vulnerability, we used the special module of the European Union Statistics on Income and Living Conditions (EU-SILC) cross-sectional survey: 2018 – Material deprivation, well-being and housing difficulties (see Eurostat, 2020, for an assessment of the implementation with the characterization of the available variables). This covers questions on subjective well-being, including happiness. The module provides information on well-being for each current household member, or, if applicable, for selected respondents aged 16 and over. Here age refers to the person's age at the end of the income reference period. Data were collected in a personal interview with all current household members aged 16 and over or, if applicable, with each selected respondent (see *ibidem*).

The number of individual observations on the happiness indicator ('Being happy', code PW090T in EU SILC) ranges between approximately ten and thirty thousand in various countries, with the number available for the middle class members being lower (see Table A3 in Appendix A.1, which shows the number of observations for the six countries under consideration). Following the approach of Baldazzi et al. (2019), the original scale (1–5) was transformed to a binary dependent variable that takes a value of 1 for the two most intense/prolonged happiness levels ('All of the time' and 'Most of the time') and 0 for the rest ('Some of the time', 'A little of the time', and 'None of the time').[4]

The income and other explanatory variables at household and individual levels were taken from the EU-SILC dataset, specifically from the Household Data File (H-File) and Personal Data File (P-File) (see Eurostat, 2021). Tables A3 and A4 in Appendix A.1 show the summary statistics of the data and their further characterization. The poverty line, used to delineate those 'falling into poverty'—or being at risk of poverty (AROP)—was fixed at 60%

[4] A preliminary analysis of the full scale using ordered logistic regression instead of a binary dependent variable gives qualitatively analogous (and more significant) results. However, it is much more demanding in terms of the time needed for the estimation of mixed-effects models with tens of thousands of observations such as in this study.

of the median household equivalised net income (see Menyhért et al., 2021, for a general evaluation of the poverty measurements and possible alternatives in the EU context). The middle class includes the population from households with net equivalised income ranging from three quarters to two medians of household equivalised net income in countries, following Eurofound (2019) and OECD (2019).

The set of variables used to evaluate the expected conditional income by individuals was initially drawn from Simona-Moussa (2020), and extended by adding other variables from Chang (2013), Caria and Falco (2018), and Rohde et al. (2020). Individual and locational characteristics were controlled for, including age, education, gender, health, years of work experience, type of occupation, sector of activity, region, and urbanization level (Table A1 in AppendixA.1 sets out the variables and links to previous research). However, the share of the total income variation explained by these variables alone was quite modest. This is because the income equivalisation process levels income within a household and mixes the various characteristics of household members. This reduces their individual influence on income even though the contribution can still differ whenever the individual weights, representing the underlying population structure, vary for different members of the household. This probably was one reason why Simona-Moussa (2020) also included household-level characteristics such as the household size and the number of household members working.[5]

In this paper, we exploited the relatively large data samples in the countries under study to develop a more extensive and nuanced set of additional household-level characteristics. This substantially increased the explanatory power of income prediction models. The first addition was the proportions of household members with specific features such as their economic activity, education, gender, occupation within a given household (see Table A4 in Appendix A.1). Second, we introduced a synthetic household composition indicator using the structure outlined in Table A5 of Appendix A.1. This aimed to jointly capture the composition of households in terms of members' activity, age, and education levels. In some cases, its introduction almost doubled the precision of the income prediction models.[6] This composite variable was used in addition to (linear) household shares, because they alone cannot capture various nonlinearities pertinent to household composition, and using all interactions would have lead to a heavy over-parametrization of the model. Even within this constrained

[5]Simona-Moussa (2020) used the changes in the variables because she relied on a longitudinal framework.

[6]We also considered gender separation and several other variables, but they turned out to be less informative empirically and were therefore not used. The number of dimensions was kept low using heavy aggregation to avoid having too many states.

structure, the number of effective states of the synthetic composition variable varied between approximately 300 and 700 in the countries studied. Given this large size, it was modelled using random effects, which will be outlined in more detail in Section 3.

## 2.2 National level data and variables

We used the individual data to derive the aggregate middle class vulnerability to poverty (the methodology is explained in Section 3.1.2). To assess the importance of vulnerability at the aggregate (country) level, we used the latest available national data on happiness from the World Happiness Report (WHR), based on aggregated answers to the Cantril (1965) ladder question from the Gallup World Poll. These are freely available.[7] The data accompanying the Helliwell et al. (2022) WHR were used, which cover both the national happiness scores, and the set of explanatory variables.[8]

The set of explanatory variables includes the logarithm of GDP per capita, social support, healthy life expectancy at birth, freedom to make life choice, generosity, and perceptions of corruption.[9] (Table A6 of Appendix A.1 provides of the aggregate data). The vulnerability indicator used 2018 data, so the value of these additional explanatory variables was also fixed at the level of 2018. The dependent variable—the national happiness scores—provided in the WHR represents the 'evaluations (answers to the Cantril ladder question) for each country, averaged over 2019-2021' (see page 16 in Helliwell et al., 2022). This means that we explore the predictive power of these variables to explain the average national happiness scores over three years ahead.[10] This forward-looking approach has the benefit of avoiding simultaneity-induced endogeneity. However, we also used an extension to show similarity of the results at the aggregate level using a dynamic panel approach with the Generalized Method of Moments (GMM) estimation based on particular instruments.

[7]We did not have access to the micro data from the Gallup survey. This data set could also be used to evaluate the significance of vulnerability at a disaggregated level in an analogous way, using the EU SILC survey data in this paper.

[8]The data from the files 'DataForTable2.1.xlsx" and 'Appendix_2_Data_for_Figure_2.1.xlsx" were used which are available at `https://worldhappiness.report/ed/2022/#appendices-and-data` (accessed on September 12, 2022). A few additional macro variables, such as income inequality and (un)employment were introduced as further controls, using data downloaded from the Eurostat online database available at `https://ec.europa.eu/eurostat/web/lfs/data/database` (see Table A6 in Appendix A.1 for the complete list of variables and their sources).

[9]As well as positive and negative affects that seem likely to be highly endogenous with happiness.

[10]The results also remained analogous using data from separate years, i.e., the 'Life Ladder" indicator available from the dataset accompanying the WHR.

# 3 Methodology

This section presents the methodology used in this study. Subsection 3.1 defines how individual and aggregate middle class vulnerability is measured and Subsection 3.2 describes the econometric models used.

## 3.1 Vulnerability measurement

There are various approaches to defining and measuring vulnerability, as well as a broader concept of economic insecurity (see Ligon and Schechter, 2003, Celidoni, 2013, Fujii, 2016, and Gallarado (2018), on the former, and Bossert and D'Ambrosio, 2013, Osberg, 2015, Richiardini and He, 2020, and Rohde et al., 2020, on the latter). In this section, we describe our setup to measure the vulnerability at both the individual and aggregate levels.

The measurement of individual income vulnerability using cross-sectional data followed mainly the expected income approach of Chaudhuri et al. (2002) and Chaudhuri (2003). It was extended with additional transformations of errors, because the standard correction for heteroscedasticity—see Chaudhuri, 2003, and the more recent application by, e.g., Caria and Falco, 2018—was clearly not sufficient to achieve a Gaussian distribution (see Figure A1 in Appendix A.4). For more elaborate approaches with axiomatic underpinning and expected utility and prospect theory-based treatment see, e.g., Dutta et al. (2011) and Rohde et al. (2020).

After defining the measurement of individual vulnerability in Subsection 3.1.1, Subsection 3.1.2 then describes the derivation of aggregate middle class vulnerability from the individual level. Essentially, it is defined as the share of people with individual vulnerability greater than a chosen significance level. To simplify the presentation, only generic notation is used in the next subsection omitting any indexing by individuals. This indexing will be introduced in the aggregate evaluation section, which requires operations over individuals.

### 3.1.1 Measurement of (conditional) individual vulnerability

Vulnerability is represented by the probability that income will be below some threshold $\bar{y} \in \mathbb{R}_+$:

$$p = \mathbb{P}(y \leq \bar{y}), \tag{1}$$

where $\mathbb{P}$ is the probability measure and $y$ denotes income. In the empirical application, we used the (logarithm of the) household equivalised net income (as $y$) with the (logarithm of the) AROP line fixed as the threshold $\bar{y}$. $\bar{y}$ is a 'poverty threshold', and the probability in eq. (1) represents therefore the (unconditional) vulnerability to poverty.

Let an appropriate income model be represented by

$$y = f(\boldsymbol{x}) + \varepsilon, \quad \varepsilon|\boldsymbol{X} \sim F_{\varepsilon|\boldsymbol{X}}, \tag{2}$$

where $\boldsymbol{X} \in \mathbb{R}^k$ is a vector of real-valued explanatory random variables with realizations $\boldsymbol{x}$; $f(\boldsymbol{x})$ is the modelled/predicted part of income with some function $f : \mathbb{R}^k \to \mathbb{R}$; $F_{\varepsilon|\boldsymbol{X}}$ is the (conditional) cumulative distribution function of zero mean errors $\varepsilon$. Placing eq. 2 into 1, it follows that the (conditional) probability for income to be below the threshold $\bar{y}$ is

$$p(\boldsymbol{x}) = \mathbb{P}\big(\varepsilon \leq \bar{y} - f(\boldsymbol{x})\big) = F_{\varepsilon|\boldsymbol{X}}\big(\bar{y} - f(\boldsymbol{x})\big). \tag{3}$$

Eq. (3) shows the conditional vulnerability to poverty at a given value of explanatory variables defined by $\boldsymbol{x}$.

Since in practice $f(\boldsymbol{x})$ is directly unobserved, it is replaced by its consistent estimator $\widehat{f}(\boldsymbol{x})$, yielding the respective estimated probability

$$\widetilde{p}(\boldsymbol{x}) = F_{\varepsilon|\boldsymbol{X}}\big(\bar{y} - \widehat{f}(\boldsymbol{x})\big), \tag{4}$$

provided $F_{\varepsilon|\boldsymbol{X}}$ is known. Otherwise, using a consistent estimator $\widehat{F}_{\varepsilon|\boldsymbol{X}}$ of $F_{\varepsilon|\boldsymbol{X}}$, gives a consistent estimator of the probability of interest

$$\widehat{p}(\boldsymbol{x}) = \widehat{F}_{\varepsilon|\boldsymbol{X}}\big(\bar{y} - \widehat{f}(\boldsymbol{x})\big). \tag{5}$$

In the empirical literature on vulnerability to poverty, it is not uncommon to assume that $F_{\varepsilon|\boldsymbol{X}}$ is a Gaussian distribution in eq. (4). If this were true, it would be possible to simply use the standard Gaussian distribution function $\Phi$

$$\widetilde{p}_{\Phi}(\boldsymbol{x}) = \Phi\left(\frac{\bar{y} - \widehat{f}(\boldsymbol{x})}{\widehat{\sigma}_{\varepsilon|\boldsymbol{X}}}\right), \tag{6}$$

after the standardization with the consistent estimate of standard deviation of errors $\widehat{\sigma}_{\varepsilon|\boldsymbol{X}}$, potentially also allowing for heteroscedasticity.[11]

In our empirical application, the straightforward Gaussianity of errors did not seem to hold even after taking the potential heteroscedasticity of errors into account.[12] This result is often found in similar studies on income vulnerability and insecurity (see, e.g., Rohde et al., 2020), although it is most often erroneously skipped. Consequently, the estimate given by eq. (6) might be severely biased and inconsistent (see Figure A1 in Appendix A.4, which evaluates the potential severity of this issue for the empirical case in this study and shows the value of the residual transformation approach[13] that will be described shortly). To avoid these potential problems, at least two alternatives are possible. First, a nonparametric estimator $\widehat{F}_{\varepsilon|\boldsymbol{X}}$ could be used in eq. (5) without assuming a specific functional form of the error distribution. Second, provided that a functional transformation $g$ of errors exists[14] such that the transformed errors are Gaussian, i.e.,

$$g(\varepsilon) \sim \mathcal{N}\left(0, \sigma^2_{g(\varepsilon)}\right), \tag{7}$$

it holds[15]

$$p(\boldsymbol{x}) = \mathbb{P}\big(\varepsilon \leq \bar{y} - f(\boldsymbol{x})\big) = F_{\varepsilon|\boldsymbol{X}}\big(\bar{y} - f(\boldsymbol{x})\big) = \Phi\left(\frac{g\big(\bar{y} - f(\boldsymbol{x})\big)}{\sigma_{g(\varepsilon)}}\right),$$

and the probability could be estimated by

$$\widetilde{p}^{(g)}_{\Phi}(\boldsymbol{x}) = \Phi\left(\frac{g\big(\bar{y} - \widehat{f}(\boldsymbol{x})\big)}{\widehat{\sigma}_{g(\varepsilon)}}\right). \tag{8}$$

[11] The set of explanatory variables affecting the regression function and the variance of errors might not coincide, but we can generically think of $\boldsymbol{X}$ as covering both sets. The function $f(\boldsymbol{x})$ can trivially be invariant (constant with respect) to a particular subset of the joint set of all variables.

[12] As judged even by the simple quantile-quantile plots. The formal statistical tests also rejected Gaussianity, but they are known to be powerful to detect tiny deviations in large samples with little actual real value (see, e.g., Kim and Ji, 2015 and Abadir, 2020, as well as a broader review by Kim and Robinson, 2019).

[13] An advantage of the residual transformation approach is that it retains 'parsimoniousness' of the logarithmic transformation of income underscored by Rohde et al. (2020). It is also possible to search for other transformations of the dependent variable instead of applying the procedure proposed here. However, the empirical applications on vulnerability measurement seem almost exclusively to use the logarithm of income (see Rohde et al., 2020, for relevant references in their Footnote 6).

[14] Here $g$ is assumed to be a real-valued, continuous, and strictly increasing function. Additional conditions of integrability for the existence of moments and some order of differentiation for the application of certain estimators might be further required in general or implied by eq. (7).

[15] See Appendix A.2 for the mathematical argument. The simulations showing close similarity of empirical results whenever the IHST transform is applied to Gaussian errors are available upon request.

In the empirical application, a simple inverse hyperbolic sine transform (IHST)[16] produced a good approximation to requirement (7), at least judging by the quantile-quantile plots of the transformed errors against the quantiles of a Gaussian distribution (see Figure A1 in Appendix A.4, which will be also discussed in Section 4). More general families of transformations that allow for additional parameters could be used for even a better fit (see, e.g., Tsai et al., 2017, for a recent suggestion and a review of alternatives) but they are outside of the scope of this paper. The transformation approach with a given function is very simple and fast, and was therefore used for the estimations in this paper. The results using the nonparametric approach were similar, but the nonparametric estimator is much more demanding in terms of time whenever the cross-validation-based bandwidth selection is used, and was therefore applied only for robustness checks of the main results.[17]

### 3.1.2 From individual to aggregate vulnerability of the middle class

To aggregate individual vulnerabilities, the respective conditional probability defined in eq. (3) was augmented by an index: $p_i := p(\boldsymbol{x}_i)$, $i \in \mathbb{N}$, where $i$ explicitly indexed vulnerability of different individuals having a vector of characteristics $\boldsymbol{x}_i$.

To get an aggregate characteristic of vulnerability for the whole population or some subpopulation of interest, we set an exogenous significance level (of vulnerability) $\alpha \in [0, 1]$, at which a person is evaluated to be (or not to be) vulnerable. People with $p_i$ greater than the chosen significance level were considered to be vulnerable to poverty, and vice versa.[18] The proportion of the middle class that is at risk of poverty at an $\alpha$-significance level of vulnerability is given by:

$$S_{MC}(\alpha) = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \mathbb{1}\{p_i \geq \alpha\} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \mathbb{1}\{f(\boldsymbol{x}_i) + q_{\mathcal{E}}(\alpha) \leq \bar{y}\}, \tag{9}$$

where the number of individuals in the middle class is denoted by $N_{MC} \in \mathbb{N}$; $\mathbb{1}\{\cdot\}$ is an indicator function taking a value of one, whenever the condition defined in the brackets holds, and zero otherwise; and $q_{\mathcal{E}}(\alpha)$ is the $\alpha$-quantile of the error distribution.

[16] In particular, for a $z \in \mathbb{R}$, the transformation $\log(z + \sqrt{1 + z^2})$ is applied (Johnson, 1949). Its further generalizations introducing additional parameterization can be considered, too (see, e.g., Burbidge et al., 1988, Carroll et al., 2003, and Bellemare and Wichman, 2020).

[17] Results are available upon request.

[18] Here vulnerable means people with a larger (not smaller) probability than a given significance level—it is important not to confuse this significance level of vulnerability with the one used for statistical inferences.

The expression in the middle of eq. (9) shows that the aggregate vulnerability here is defined as the proportion of individually vulnerable middle class members.[19] The expression on the right side shows that this vulnerability can be understood as a perturbation of the expected income by the $\alpha$-level quantile income shock: vulnerable people are those whose income drops below the threshold $\bar{y}$ after experiencing this type of shock. The *proportion at risk indicator* in eq. (9) is based on the expected income $f(\boldsymbol{x}_i)$ $\big(= \mathbb{E}(y|\boldsymbol{X} = \boldsymbol{x}_i)\big)$ rather than the actual income, which makes it well-aligned with the *income insecurity* approach that stresses the relevance of expected and not actual income (see, e.g., Osberg, 2021). The empirical counterpart of eq. (9) was obtained using the estimates of $p_i$ (or $f(\boldsymbol{x}_i)$ and $q_\varepsilon(\alpha)$).

Besides fixing a particular $\alpha$-level, it is possible to evaluate the *trace of vulnerability*, given by the function of $S_{MC}(\alpha)$ in terms of $\alpha$, i.e., the set of pairs $\{S_{MC}(\alpha), \alpha\}$ for all $\alpha$ values, represented, e.g., by the respective plot of $S_{MC}(\alpha)$ against $\alpha \in [0, 1]$. This might be especially relevant for more specialized comparative studies, e.g., when a couple of countries are being compared. It would show the behavior of conditional vulnerability at various $\alpha$ levels that might be distinct, i.e., some countries may be relatively more or less vulnerable at different levels of $\alpha$. However, this was left outside of the scope of the empirical part of the paper.

## 3.2 Econometric models

Three econometric models were used to develop the results in the empirical work in Section 4, and are described in this section. Subsection 3.2.1 defines the empirical counterpart of eq. (2), i.e., the income prediction model. Subsection 3.2.2 explains the econometric equation used to evaluate the significance of the individual vulnerability indicator for happiness of middle class people at the micro level. Section 3.2.3 specifies the econometric model of national happiness scores used to gauge the importance of the middle class vulnerability to poverty at the country level using the proportion of the middle class that is vulnerable.

The evaluation of significance of the vulnerability to poverty at the micro and aggregate levels go in parallel, and are not interdependent in terms of modelling. However, they both rely on the income prediction model, which serves as a basis to derive the individual and aggregate middle class vulnerability to poverty.

[19] It is also possible to consider some distributional characteristics of $\{\widehat{p}_i\}_{i=1}^{N_{MC}}$, e.g., the mean or median. However, these characteristics were (much) less significant in the empirical application in the next section, and are therefore not discussed further.

There is a lengthy list of explanatory variables, and the models are therefore described conceptually with the control variables of less interest covered by a single vector of variables. For a full list of explanatory variables in each model see the data section (Section 2), and the lists of explanatory variables in Tables A1, A2, and A6 in Appendix A.1.

### 3.2.1 Individual-income prediction model

As hinted in Section 2.1, the income model covers both the fixed and random effects, and the following mixed effects model was used:

$$y_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1'\mathbf{P}_i + \boldsymbol{\beta}_2'\mathbf{H}_{h[i]} + \kappa_{c[i]} + \boldsymbol{\mu}_{\eta[i]} + \varepsilon_i. \tag{10}$$

$y_i$ is the (logarithm of) household equivalised net income of an individual indexed by $i$. $\mathbf{P}_i$ is a vector of personal (individual) characteristics with the parameter vector $\boldsymbol{\beta}_1$. $\mathbf{H}_{h[i]}$ is a vector of household shares—i.e., the proportion of household members with a particular feature—with the parameter vector $\boldsymbol{\beta}_2$. $h[i]$ implicitly indexes a household of a person with individual index $i$. $\kappa_{c[i]}$ is the household-composition random effects—see Table A5 of Appendix A.1 for its structure— that are assumed to be normally distributed.[20] The implicit index $c[i]$ represents a particular composition structure of a household to which a person indexed by $i$ belongs. $\boldsymbol{\mu}_{\eta[i]}$ is the vector of various additional fixed effects at various levels (e.g., age, sector, region). Finally, $\varepsilon_i$ denotes the remaining zero mean error term assumed to satisfy certain common conditions.

In relation to eq. 2, the modeled part $f(\boldsymbol{x})$, connected to the expected income, corresponds to

$$f(\boldsymbol{x}_i) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1'\mathbf{P}_i + \boldsymbol{\beta}_2'\mathbf{H}_{h[i]} + \kappa_{c[i]} + \boldsymbol{\mu}_{\eta[i]}, \tag{11}$$

where $\boldsymbol{x}_i$ covers all the explanatory variables, including the groups of fixed effects and random effects.

The restricted maximum likelihood estimator (see Bates et al., 2015) was used to estimate model (10) weighted with individual sampling weights.[21] The predicted income $\widehat{y}_i = \widehat{f}(\boldsymbol{x}_i)$ and residuals $\widehat{\varepsilon}_i = y_i - \widehat{f}(\boldsymbol{x}_i)$ were obtained by replacing unobserved terms in eq. (11) with

[20] Note that the estimates in models like this are robust to the violation of this assumption (see, e.g., Bell et al., 2019, Schielzeth et al., 2020).

[21] We prefer the Restricted Maximum Likelihood Estimator here, because the standard Maximum Likelihood estimator of variance and covariance parameters in Mixed Effects models is biased. For the empirical implementation we used `R` package `lme4` (Bates et al., 2015).

the relevant estimates, i.e, $\widehat{\boldsymbol{\beta}}_0, \ldots, \widehat{\boldsymbol{\beta}}_2$, $\widehat{\kappa}_{c[i]}$, and $\widehat{\boldsymbol{\mu}}_{\eta[i]}$ from the estimated model (10). The predicted income and residuals were then used to derive the individual and overall middle class vulnerability to poverty as described in Section 3.1. The IHST-transform-based individual vulnerability to poverty measurement was performed using eq. (8).

### 3.2.2 The model of happiness at the micro level

To evaluate if the vulnerability to poverty is significantly related to the subjective well-being indicator of happiness for middle class individuals, we used a logistic mixed effects model. This model allowed for the potential household random effects while controlling for a number of fixed effects. The household-level random effect term was included to account for the potential correlation of individual happiness within households (see Baldazzi et al., 2019, for a similar motivation).

The dependent variable of individual happiness intensity was denoted by $h_i$ for a subject indexed by $i$. It takes a value of one for a sufficiently intense/prolonged happiness and zero otherwise. Considering the population in the middle class, logistic regression was then used to model the probability of intense happiness by:

$$\mathbb{P}(h_i = 1) = \operatorname{logit}^{-1}\left(\theta p_i + \boldsymbol{\theta}_Z{}' \boldsymbol{Z}_i + \lambda_{h[i]}\right), \ \lambda_h \sim \mathcal{N}(0, \sigma_\lambda^2), \tag{12}$$

where $\operatorname{logit}^{-1}(x) = \frac{\exp(x)}{1+\exp(x)}$, $x \in \mathbb{R}$. $p_i$ is the individual vulnerability to poverty with the parameter of interest $\theta$ expected to be negative. The remaining control variables were covered by vector $\boldsymbol{Z}_i$, including a constant and various fixed effects from eq. (10), and $\boldsymbol{\theta}_Z$ is the corresponding vector of parameters. $\lambda_{h[i]}$ is the random household effects term, where $h := h[i]$ shows an implicit household index.

The parameter estimates were underpinned by the Maximum Likelihood estimation (weighted with individual sampling weights) (see Bates, 2022).

### 3.2.3 The model of national happiness scores

The previous equations were considered at the individual level to show the significance of individual vulnerability. The aggregate vulnerability of the middle class, derived as the proportion of members who are vulnerable, as defined by eq. (9) and plotted in Figure A2 of Appendix A.6, was then used to check its relevance for the national happiness scores in the following model

$$H_c = \phi V_c + \boldsymbol{\phi}_W{}' \boldsymbol{W}_c + \nu_c. \tag{13}$$

For each country indexed by $c$, $H_c$ denotes the (logarithm of the) national happiness score. The parameter of interest $\phi$ is expected to be negative, and the related (log) vulnerability of the middle class, $V_c := \log(\widehat{S}_{MC,c}(0.1))$, is measured at the $\alpha = 0.1$ level in the baseline estimation. $\boldsymbol{W}_c$ is a vector of various additional control variables, including a constant term. Finally, $\nu_c$ is the remaining zero mean error term.

The Ordinary Least Squares estimator was used to estimate eq. (13) with heteroscedasticity-consistent standard errors.[22]

# 4 Main empirical results

The results of the estimation of the auxiliary income prediction model are shown in Appendix A.3. This section tackles the substantive question of the significance of the vulnerability to poverty indicator for happiness. Subsection 4.1 reports the empirical results on the individual vulnerability to poverty and happiness of middle class members. Subsection 4.2 describes the estimation results with the aggregate middle class vulnerability at national level, and its relationship with national happiness scores. All equations were estimated using cross-sectional data (just at different aggregation levels).

## 4.1 Individual vulnerability and happiness of middle class members

The estimated analogue of eq. (12), considering members of the middle class identified using the EU SILC micro-data,[23] is shown in Table 1. which presents the results for the five largest EU countries by population[24] and the United Kingdom (UK). The five EU countries are Germany (DE), France (FR), Italy (IT), Spain (ES), and Poland (PL).

Table A8 in Appendix A.5 contains the complete estimation results, and Table 1 shows the most relevant variables (those that were either consistently significant across different coun-

[22]The estimator proposed by MacKinnon and White (1985) was used for the baseline, and the results with other alternatives are available upon request.

[23]People were considered to be middle class if they were members of a household with equivalised household net income in the range $[0.75M, 2M]$, where $M$ is the median household equivalised net income in that country.

[24]They also had relatively large sample sizes in their surveys.

tries or are of direct economic interest). The set of variables considered by Simona-Moussa (2020) served as an initial basis,[25] with a few additional variables from Kahneman and Deaton (2010), Chang (2013), Caria and Falco (2018), and Baldazzi et al. (2019) (see Table A2 in Appendix A.1 for details). We also added some further economic determinants—the presence of a heavy financial burden connected either with housing or debt repayment—that seemed relevant from an economic perspective. Finally, besides the urbanization level, we added additional variables (crime and environmental concerns) to represent the living environment. This aspect is underscored by, e.g., Welsch (2006), Ferrer-i-Carbonell and Gowdy (2007), and Brenig and Proeger (2018).

An extensive set of explanatory variables is more likely to experience multicollinearity issues, with consequent implications for a larger variance of estimators and, possibly, the need to drop some dimensions. However, our aim was to show the significance of a negative coefficient of the vulnerability indicator with a relatively large number of controls. It is worth pointing out that the significance of the vulnerability variable becomes even stronger under the reduced set of control variables, and its consistency in terms of the sign of the impact was more stable relative to many 'standard determinants' of happiness.

The vulnerability to poverty indicator is highly significant in Table 1 along with other economic dimensions such as material deprivation and financial burden. The (log) income variable is somewhat less significant, but became highly significant (with a larger coefficient) if the vulnerability indicator was removed from the equation. This could suggest that studies that considered income without vulnerability, might have overstated at least part of the significance of income.

It is interesting to compare the coefficient of the vulnerability indicator with those of the other economic determinants (apart from income) that are also in the $[0, 1]$ range. Severe vulnerability—i.e., whenever the probability of being below the AROP line is close to one—has similar or even larger effect on happiness (in absolute terms) than 'real consequences' such as being severely materially deprived, facing serious financial troubles, or being unable

[25] The variable 'To feel down' used in Simona-Moussa (2020) represents a very similar state of well-being as happiness and was therefore omitted from the list of explanatory variables. The perception of health state is also known to be highly endogenous with happiness (see, e.g., Sabatini, 2014, or Pierewan and Tampubolon, 2015). In the equation of happiness, a more exogenously determined health limitation variable was therefore used as in Baldazzi et al. (2019). This relies on a clearer health limitation criterion, which is more exogenouse to perceptions. Trust and social connectivity/loneliness variables were also initially excluded from the baseline specification because of their high likelihood of endogeneity with happiness (see e.g., Guven, 2011, and Kaliterna-Lipovčhan and Prizmić-Larsen, 2016), and were considered separately (see Section 5).

**Table 1: Logistic mixed-effects regression of individual happiness.**

| | *Dependent variable:* Individual Happiness Intensity | | | | | |
|---|---|---|---|---|---|---|
| | DE | ES | FR | IT | PL | UK |
| Income (log) | 0.260** | 0.363*** | 0.298** | 0.114 | 0.298*** | 0.118 |
| | (0.101) | (0.115) | (0.129) | (0.089) | (0.113) | (0.125) |
| Vulnerability (individual) | −0.808*** | −0.659*** | −0.878*** | −0.813*** | −0.485** | −0.735*** |
| | (0.249) | (0.216) | (0.330) | (0.185) | (0.223) | (0.264) |
| Severe material deprivation | −0.498** | −0.468** | −1.041*** | −0.992*** | −1.010*** | −0.347’ |
| | (0.252) | (0.185) | (0.213) | (0.122) | (0.173) | (0.216) |
| Heavy burden of housing cost | −0.322*** | −0.281*** | −0.444*** | −0.233*** | −0.319*** | −0.346*** |
| | (0.081) | (0.058) | (0.071) | (0.045) | (0.054) | (0.099) |
| Heavy burden of debt repayment | −0.623*** | −0.333*** | −0.435*** | −0.564*** | −0.362*** | −0.515*** |
| | (0.147) | (0.091) | (0.107) | (0.142) | (0.105) | (0.118) |
| Cannot find job | 0.001 | −0.188 | −0.593** | −0.385** | 0.117 | 0.240 |
| | (0.296) | (0.185) | (0.251) | (0.167) | (0.576) | (0.287) |
| Strong health limit. [ref. Unlimit.] | −1.287*** | −1.738*** | −0.792*** | −1.523*** | −0.964*** | −1.171*** |
| | (0.101) | (0.120) | (0.096) | (0.120) | (0.097) | (0.096) |
| Health limit. [ref. Unlimit.] | −0.616*** | −0.911*** | −0.599*** | −0.602*** | −0.505*** | −0.543*** |
| | (0.064) | (0.068) | (0.073) | (0.057) | (0.067) | (0.075) |
| Environmental problems | −0.019 | −0.438*** | −0.314*** | −0.234*** | 0.049 | −0.250*** |
| | (0.059) | (0.092) | (0.079) | (0.071) | (0.075) | (0.085) |
| Crime problems | −0.364*** | −0.140’ | −0.183** | −0.103 | −0.081 | −0.299*** |
| | (0.075) | (0.090) | (0.081) | (0.071) | (0.121) | (0.072) |
| Married [ref. Never married] | 0.487*** | 0.669*** | 0.214** | −0.049 | 0.577*** | 0.547*** |
| | (0.083) | (0.085) | (0.087) | (0.068) | (0.097) | (0.090) |
| With children | −0.003 | 0.020 | 0.275** | 0.152** | 0.156* | 0.046 |
| | (0.089) | (0.086) | (0.120) | (0.072) | (0.083) | (0.107) |
| Retired [ref. Active] | 0.428*** | 0.316** | 0.267 | 0.650*** | 0.297* | 1.014*** |
| | (0.164) | (0.135) | (0.198) | (0.137) | (0.155) | (0.260) |
| Age | −0.070*** | −0.095*** | −0.071*** | −0.040*** | −0.077*** | −0.081*** |
| | (0.013) | (0.011) | (0.012) | (0.009) | (0.012) | (0.013) |
| $(\text{Age}/100)^2$ | 0.044*** | 0.065*** | 0.050*** | 0.025*** | 0.045*** | 0.076*** |
| | (0.015) | (0.012) | (0.012) | (0.010) | (0.012) | (0.014) |
| Old age (> 60) | 0.275** | 0.346*** | 0.236’ | 0.039 | 0.095 | 0.094 |
| | (0.115) | (0.123) | (0.152) | (0.096) | (0.112) | (0.130) |
| Female | 0.080* | −0.134** | −0.217*** | −0.025 | −0.003 | −0.119** |
| | (0.047) | (0.056) | (0.059) | (0.045) | (0.054) | (0.059) |
| | | | | ⋮ | | |
| (Intercept) | 0.541 | −0.191 | 5.537 | 1.049 | −0.052 | 1.365 |
| | (1.078) | (2.129) | (5.163) | (1.690) | (1.984) | (1.331) |
| FE: Household size | + | + | + | + | + | + |
| FE: Sector of activity | + | + | + | + | + | + |
| FE: Urbanization | - | + | + | + | + | + |
| FE: Region | - | + | + | + | + | + |
| RE: SD of household RE | 0.769 | 0.800 | 0.473 | 0.709 | 0.697 | 0.801 |
| SD of Residual | 0.789 | 0.679 | 0.800 | 0.721 | 0.756 | 0.755 |
| Degrees of freedom | 13,824 | 15,812 | 9,899 | 17,760 | 12,513 | 10,068 |
| R2 (conditional) | 0.231 | 0.331 | 0.197 | 0.242 | 0.253 | 0.266 |
| R2 (marginal) | 0.092 | 0.200 | 0.142 | 0.126 | 0.143 | 0.123 |
| RMSE | 0.40 | 0.38 | 0.42 | 0.43 | 0.40 | 0.38 |
| Log Likelihood | -6,295 | -4,299 | -3,272 | -6,311 | -4,407 | -3,908 |
| AIC | 12,679 | 8,734 | 6,687 | 12,729 | 8,933 | 7,941 |
| BIC | 13,018 | 9,256 | 7,206 | 13,150 | 9,380 | 8,396 |

*Notes:* *p<0.1; **p<0.05; ***p<0.01. The model estimates the probability of being happy most or all of the time. The household random effects term accounts for the potential correlation of happiness within households. Individual survey weights were used for the model-consistent standard errors that are shown in parentheses. For factors with more than two variables, the reference level is shown in square brackets. The regional and urbanization dimensions were unavailable for Germany.

to find a job. For instance, in a hypothetical situation where a person has just recently entered the middle class and therefore supposedly climbed out of a state of severe material deprivation, the increase in the perception of happiness could be substantially higher if also the vulnerability to poverty were insignificant, i.e., the risk of falling back to poverty was negligible. Economic policies designed to support a sufficiently high level of unconditional basic income would therefore help to improve both of these simultaneously, i.e., by reducing the material deprivation and the downside risk of income at the same time.

Summarizing the results at the micro level, there seems to be strong evidence of the significance and importance of middle class vulnerability to poverty from both a statistical and economic point of view.

## 4.2 Aggregate middle class vulnerability and national happiness

The results for eq. (13) are shown in Table 2 on the following page, with different iterations showing variations in the set of control variables included. Apart from a single case with all the explanatory variables reported in Column (5), the table includes only very simple specifications with just a few parameters because of the small number of observations. However, to increase confidence in the importance of vulnerability, a large number of potential individual controls was further explored, and the results are included in Appendix A.7. The size of the middle class is also included among the explanatory variables in Table 2, to check that the middle class vulnerability indicator is not simply capturing the 'omitted middle class size effect'.

Besides the (logarithm of the) gross domestic product (GDP) per capita, the proportion of middle class members who are vulnerable to poverty is highly significant in all the cases reported in Table 2. However, only a few other variables become significant. First, the North-West Europe (NWE) dummy is significant, hinting that countries in this region may differ in terms of happiness from the Southern European (SoE) and Central and Eastern European (CEE) countries (see Appendix A.7 for the (in)significance of other geographic region indicators). However, even the North West Europe dummy became insignificant after the inclusion of the *perceptions of corruption* variable, suggesting that the geographic difference could be driven by perceptions of corruption, potentially reflecting different traditions, maturity of the society, and the higher quality of governance in countries in North West Europe. This result is consistent with the previous literature establishing the importance of corruption for

**Table 2: Cross-sectional regression of national happiness scores.**

| | *Dependent variable:* National Happiness Scores (log) | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| North-West Europe | 0.074** | 0.074** | 0.016 | | |
| | (0.032) | (0.030) | (0.033) | | |
| GDP per capita (log) | 0.113** | 0.113** | 0.087* | 0.094** | 0.100** |
| | (0.052) | (0.051) | (0.045) | (0.040) | (0.047) |
| Middle class vulnerability (log) | −0.053** | −0.050** | −0.045** | −0.045*** | −0.041* |
| | (0.025) | (0.019) | (0.017) | (0.016) | (0.021) |
| Middle class size (log) | −0.018 | | | | |
| | (0.108) | | | | |
| Perceptions of corruption | | | −0.165** | −0.186*** | −0.193*** |
| | | | (0.061) | (0.046) | (0.054) |
| Social support | | | | | 0.035 |
| | | | | | (0.375) |
| Healthy life expectancy at birth | | | | | 0.0001 |
| | | | | | (0.008) |
| Freedom to make life choices | | | | | 0.016 |
| | | | | | (0.115) |
| Generosity | | | | | −0.049 |
| | | | | | (0.082) |
| Intercept | 0.580 | 0.596 | 1.005* | 0.959** | 0.847 |
| | (0.543) | (0.541) | (0.492) | (0.452) | (0.758) |
| Observations | 28 | 28 | 28 | 28 | 28 |
| $R^2$ | 0.780 | 0.780 | 0.829 | 0.828 | 0.831 |
| Adjusted $R^2$ | 0.742 | 0.752 | 0.800 | 0.806 | 0.772 |
| F Statistic | 20.380*** (df = 4; 23) | 28.330*** (df = 3; 24) | 27.947*** (df = 4; 23) | 38.405*** (df = 3; 24) | 14.061*** (df = 7; 20) |

*Notes*: *p<0.1; **p<0.05; ***p<0.01. The table shows the ordinary least squares estimates of eq. (13). The robust standard errors are shown in brackets. The middle class size and vulnerability indicators are based on the EU SILC data, and the data for national happiness scores and other socioeconomic variables were taken from the World Happiness Report 2022.

subjective well-being at various aggregation levels (see, e.g., Welsch, 2008, Tay et al., 2014, Li and An, 2020, and Ma et al., 2022).

After controlling for the perceptions of corruption, all the other variables remained insignificant; Column (5) shows the main set of variables from the WHR and Appendix A.7 shows individual checks augmented with further variables from Eurostat. The middle class size variable could not compete in terms of significance with the vulnerability indicator and the GDP per capita variable (see Column (1)). The middle class size is not reported in other specifications because of its insignificance. Middle class vulnerability to poverty therefore seems to be an important constituent of the national happiness scores reported by the WHR.

After establishing the importance of the vulnerability to poverty at an individual level in the previous section, and given the substantial share of the middle class in the population—ranging from around 50% to 80% in the EU countries—the significance of the aggregate middle class vulnerability indicator for national happiness is not an unexpected finding. However, the aggregate analysis at the country level highlights its outstanding importance among a number of other socioeconomic determinants.

# 5 Further extensions

This section discusses some important extensions to the analysis. Subsection 5.1 includes a few additional determinants of individual happiness, and Subsection 5.2 provides the estimation results using aggregate data in the dynamic panel framework.

## 5.1 Additional factors of individual vulnerability

In Subsection 4.1, a couple of variables were intentionally omitted for different reasons. This subsection describes some other changes to the base specification. First, individuals might worry less about their vulnerability if they have sufficient means to withstand a temporary negative shock. Two additional variables were therefore added to capture this aspect: a direct forward-looking measure of capacity to face unexpected financial expenses, and the capacity to afford one-week annual holiday away from home. These were intended to measure the potential presence of a 'savings buffer'. Second, trust and loneliness were excluded from the base estimation because of the risk of endogeneity. Happiness might be affected by them and may also influence their perception, because happy people tend to trust more and evaluate things more positively (see, e.g., Guven, 2011 and Kaliterna-Lipovčhan and Prizmić-Larsen, 2016).[26] They are included in this section as potentially important additional controls that could have created an omitted variable bias, if excluded. They are of little direct interest in this paper, and therefore no additional instrumenting was performed, hoping that the bias contribution to the mean squared error is relatively small compared with the variance increase because of a potential use of the instrumental variable estimator.

The next three tables show the estimation results when each two pairs of variables augment the base specification used in Subsection 4.1 either separately (Tables 3 and 4) or jointly (Table 5). To save space, only the economic variables of interest are reported with these additional controls.[27]

The results of these estimations were marginally weaker for the vulnerability to poverty. However, the significance of income and severe material deprivation was much more affected, especially when the financial safety variables were included in Table 3 or all additional variables were added in Table 5. The income variable remained significant in just one case, and sever material deprivation was insignificant in half of the cases. This reinforces the previous conclusion that the direct income variable might be much less important after taking the financial capacity and safety into account. Whereas, the vulnerability to poverty indicator became (marginally) insignificant in just one case.

Finally, the last extension in this subsection hints at the potential relevance of a more general income insecurity than only just vulnerability to poverty. It suggests that middle class members might be worrying more about belonging to the middle class than becoming poor. In particular, taking the specification used for Table 5 with a single change in the vulnerability threshold, Table 6 shows the results whenever the middle class income lower boundary (three-quarters of the median income) was used as $\bar{y}$ in eq. (8) instead of the AROP line, defined as three-fiths of the median income.

When the vulnerability indicator was connected to the middle class lower income boundary and not the AROP line, it was more significant in half of the countries (DE, PL, and UK) and the values of information criteria were lower in all but one case. This therefore suggests

[26]Kaliterna-Lipovčhan and Prizmić-Larsen (2016) stated, "Examining the causal relationship between self-reported happiness and measures of social capital Guven (2011) found that happiness induces a higher level of trust to others. < ... > His study also showed that happier people perform more volunteer work, are more attached to their neighborhoods, and participate more in community events, social gatherings, cultural events, local politics, and religious events.", and are therefore less lonely.

[27]The complete estimation results are available upon request.

**Table 3: Logistic mixed-effects regression of individual happiness (with capacity to face unexpected financial expenses and have holidays away from home).**

| | *Dependent variable:* Individual Happiness Intensity | | | | | |
|---|---|---|---|---|---|---|
| | DE | ES | FR | IT | PL | UK |
| Income (log) | 0.142 | 0.240** | 0.144 | 0.073 | 0.147 | 0.054 |
| | (0.103) | (0.117) | (0.132) | (0.091) | (0.116) | (0.126) |
| Vulnerability (individual) | −0.604** | −0.505** | −0.679** | −0.755*** | −0.338 | −0.586** |
| | (0.253) | (0.218) | (0.334) | (0.186) | (0.227) | (0.267) |
| Severe material deprivation | 0.040 | −0.212 | −0.685*** | −0.942*** | −0.727*** | −0.173 |
| | (0.264) | (0.188) | (0.220) | (0.127) | (0.178) | (0.220) |
| Heavy burden of housing cost | −0.271*** | −0.152** | −0.357*** | −0.194*** | −0.230*** | −0.259** |
| | (0.082) | (0.061) | (0.073) | (0.047) | (0.056) | (0.102) |
| Heavy burden of debt repayment | −0.487*** | −0.230** | −0.328*** | −0.541*** | −0.312*** | −0.428*** |
| | (0.150) | (0.092) | (0.109) | (0.143) | (0.107) | (0.121) |
| Cannot find job | 0.035 | −0.171 | −0.589** | −0.386** | 0.103 | 0.258 |
| | (0.302) | (0.185) | (0.252) | (0.168) | (0.583) | (0.287) |
| Capacity to face financial shocks | 0.217*** | 0.188** | 0.211*** | −0.234*** | 0.034 | 0.098 |
| | (0.070) | (0.075) | (0.079) | (0.059) | (0.067) | (0.082) |
| Capacity to have holidays | 0.490*** | 0.368*** | 0.385*** | 0.396*** | 0.539*** | 0.299*** |
| | (0.109) | (0.075) | (0.086) | (0.055) | (0.066) | (0.096) |
| Degrees of freedom | 13,758 | 15,810 | 9,858 | 17,758 | 12,408 | 10,066 |
| R2 (conditional) | 0.233 | 0.335 | 0.204 | 0.249 | 0.266 | 0.266 |
| R2 (marginal) | 0.098 | 0.207 | 0.149 | 0.131 | 0.154 | 0.124 |
| RMSE | 0.40 | 0.38 | 0.41 | 0.43 | 0.40 | 0.38 |
| Log Likelihood | -6,243 | -4,278 | -3,243 | -6,292 | -4,341 | -3,900 |
| AIC | 12,580 | 8,696 | 6,633 | 12,697 | 8,807 | 7,929 |
| BIC | 12,934 | 9,233 | 7,166 | 13,133 | 9,268 | 8,399 |

*Notes:* *p<0.1; **p<0.05; ***p<0.01. The model estimates the probability of being happy most or all of the time. The household random effects term accounts for the potential correlation of happiness within households. Individual survey weights were used for the model-consistent standard errors shown in parentheses.

that this specification should be preferred (compare the results in Tables 5 and 6).[28] In analogy to the question considered by Rohde et al. (2017) in terms of impact on health, it can be concluded that a more general economic insecurity concept of staying in the middle class might be more relevant for the happiness of middle class members from the countries in this study than 'solely' the more narrow indicators of vulnerability to poverty and deprivation.[29]

[28] In a specification with both measures included in the equation, the AROP-linked vulnerability was always insignificant and had the wrong sign in all countries but Spain. Despite the multicollinearity of the two vulnerability measures, the lower-bound-based vulnerability was significant for Germany and Poland, with just marginal insignificance for Italy and the UK: here the p-values were below 0.15.

[29] A number of additional robustness checks were performed at the micro level and are available upon request. These included: nonparametric estimation of the vulnerability term; additional age restrictions; different metrics of vulnerability; smaller number of determinants; squared income as well as the income prediction error term; different impact intensity for upper and lower middle class members; income classes instead of income (log-) level (which made income insignificant although the vulnerability term remained significant); and without the vulnerability term, which made income much more significant.

**Table 4: Logistic mixed-effects regression of individual happiness (with trust and loneliness variables).**

| | *Dependent variable:* Individual Happiness Intensity | | | | | |
|---|---|---|---|---|---|---|
| | DE | ES | FR | IT | PL | UK |
| Income (log) | 0.184* | 0.345*** | 0.250* | 0.120 | 0.280** | 0.117 |
| | (0.103) | (0.116) | (0.130) | (0.090) | (0.114) | (0.126) |
| Vulnerability (individual) | −0.749*** | −0.583*** | −0.777** | −0.813*** | −0.429* | −0.653** |
| | (0.252) | (0.218) | (0.333) | (0.186) | (0.225) | (0.267) |
| Severe material deprivation | −0.505** | −0.391** | −0.937*** | −0.932*** | −0.944*** | −0.378* |
| | (0.255) | (0.187) | (0.217) | (0.122) | (0.176) | (0.218) |
| Heavy burden of housing cost | −0.271*** | −0.276*** | −0.430*** | −0.201*** | −0.305*** | −0.313*** |
| | (0.082) | (0.058) | (0.071) | (0.045) | (0.054) | (0.101) |
| Heavy burden of debt repayment | −0.592*** | −0.349*** | −0.454*** | −0.541*** | −0.304*** | −0.491*** |
| | (0.149) | (0.091) | (0.107) | (0.143) | (0.106) | (0.119) |
| Cannot find job | 0.036 | −0.138 | −0.579** | −0.410** | 0.064 | 0.245 |
| | (0.299) | (0.187) | (0.251) | (0.169) | (0.580) | (0.289) |
| Trust | 0.805*** | 0.649*** | 0.535*** | 0.541*** | 0.573*** | 0.509*** |
| | (0.058) | (0.055) | (0.074) | (0.049) | (0.057) | (0.066) |
| Loneliness | −1.823*** | −1.477*** | −1.930*** | −1.224*** | −1.217*** | −1.746*** |
| | (0.260) | (0.250) | (0.264) | (0.237) | (0.254) | (0.306) |
| Degrees of freedom | 13,822 | 15,810 | 9,897 | 17,758 | 12,511 | 10,066 |
| R2 (conditional) | 0.260 | 0.352 | 0.220 | 0.258 | 0.272 | 0.283 |
| R2 (marginal) | 0.129 | 0.227 | 0.169 | 0.146 | 0.163 | 0.143 |
| RMSE | 0.39 | 0.37 | 0.41 | 0.43 | 0.40 | 0.38 |
| Log Likelihood | -6,152 | -4,216 | -3,220 | -6,241 | -4,342 | -3,857 |
| AIC | 12,398 | 8,571 | 6,587 | 12,594 | 8,809 | 7,845 |
| BIC | 12,752 | 9,108 | 7,120 | 13,030 | 9,270 | 8,314 |

*Notes:* *p<0.1; **p<0.05; ***p<0.01. The model estimates the probability of being happy most or all of the time. The household random effects term accounts for the potential correlation of happiness within households. Individual survey weights were used for the model-consistent standard errors shown in parentheses.

## 5.2 Aggregate vulnerability in the dynamic panel framework

Up to this point, this study has focused on a cross-sectional framework, because the micro-data-based evidence on happiness relied on the special 2018 module of the EU SILC survey. The evaluation of national happiness in Section 4.2 also used the 2018-based aggregate middle class vulnerability indicator at national level. By taking the (logarithm of the) aggregate Cantril ladder—the *Life Ladder*[30]—at national level from the WHR to represent national happiness, this subsection shifts to a (dynamic) panel framework, because panel data are available for the Life Ladder indicator. In this case, the predictive interpretation of the relationship does not apply, but the lags of all variables were used as instruments in the GMM to overcome the potential contemporaneous endogeneity in the relationship.

The income prediction model underlying aggregate vulnerability measurement can also be estimated from the usual EU SILC data without the need to rely on its special module on well-

[30]See page 15 of the WHR for more details.

**Table 5: Logistic mixed-effects regression of individual happiness (with all additional variables: trust, loneliness, capacity to face unexpected financial expenses and have holidays away from home).**

| | *Dependent variable:* Individual Happiness Intensity | | | | | |
|---|---|---|---|---|---|---|
| | DE | ES | FR | IT | PL | UK |
| Income (log) | 0.077 | 0.226* | 0.108 | 0.079 | 0.135 | 0.059 |
| | (0.105) | (0.118) | (0.133) | (0.091) | (0.117) | (0.127) |
| Vulnerability (individual) | −0.568** | −0.438** | −0.588* | −0.755*** | −0.288 | −0.515* |
| | (0.255) | (0.220) | (0.338) | (0.187) | (0.229) | (0.270) |
| Severe material deprivation | −0.021 | −0.153 | −0.607*** | −0.883*** | −0.672*** | −0.217 |
| | (0.267) | (0.190) | (0.223) | (0.127) | (0.181) | (0.221) |
| Heavy burden of housing cost | −0.226*** | −0.155** | −0.348*** | −0.162*** | −0.220*** | −0.234** |
| | (0.083) | (0.062) | (0.073) | (0.047) | (0.056) | (0.103) |
| Heavy burden of debt repayment | num-0.466*** | −0.251*** | −0.351*** | −0.520*** | −0.258** | −0.411*** |
| | (0.152) | (0.093) | (0.110) | (0.144) | (0.108) | (0.122) |
| Cannot find job | 0.042 | −0.124 | −0.573** | −0.410** | 0.062 | 0.261 |
| | (0.302) | (0.187) | (0.252) | (0.169) | (0.587) | (0.289) |
| Capacity to face financial shocks | 0.205*** | 0.176** | 0.204** | −0.232*** | 0.019 | 0.090 |
| | (0.071) | (0.076) | (0.080) | (0.059) | (0.068) | (0.082) |
| Capacity to have holidays | 0.435*** | 0.348*** | 0.355*** | 0.391*** | 0.529*** | 0.276*** |
| | (0.110) | (0.076) | (0.087) | (0.055) | (0.067) | (0.097) |
| Trust | 0.793*** | 0.641*** | 0.523*** | 0.539*** | 0.561*** | 0.505*** |
| | (0.058) | (0.055) | (0.074) | (0.050) | (0.057) | (0.066) |
| Loneliness | −1.772*** | −1.455*** | −1.910*** | −1.229*** | −1.226*** | −1.724*** |
| | (0.261) | (0.250) | (0.265) | (0.238) | (0.262) | (0.307) |
| Degrees of freedom | 13,756 | 15,808 | 9,856 | 17,756 | 12,406 | 10,064 |
| R2 (conditional) | 0.262 | 0.356 | 0.227 | 0.265 | 0.283 | 0.283 |
| R2 (marginal) | 0.133 | 0.232 | 0.175 | 0.151 | 0.173 | 0.144 |
| RMSE | 0.39 | 0.37 | 0.41 | 0.43 | 0.40 | 0.38 |
| Log Likelihood | -6,107 | -4,197 | -3,192 | -6,223 | -4,281 | -3,851 |
| AIC | 12,312 | 8,538 | 6,537 | 12,563 | 8,689 | 7,835 |
| BIC | 12,681 | 9,091 | 7,084 | 13,015 | 9,165 | 8,319 |

*Notes:* *p<0.1; **p<0.05; ***p<0.01. The model estimates the probability of being happy most or all of the time. The household random effects term accounts for the potential correlation of happiness within households. Individual survey weights were used for the model-consistent standard errors shown in parentheses.

being. This section therefore provides the estimation results using a more general dynamic panel approach, compared with the static regressions in Section 4.2. It therefore allowed more observations by using additional years, and accounted for the potential importance of the dynamic (lagged) terms of happiness.

Table 7 shows the estimation results[31] using the 2014-2020 data. This period was con-

[31] The specification was:

$$L_{i,t} = \rho L_{i,t-1} + \boldsymbol{\theta}' \boldsymbol{W}_{i,t} + \eta_i + \delta_t + \nu_{it},$$

where $\eta_i$ and $\delta_t$ are the country and period fixed effects, $L_{i,t}$ denotes the (logarithm of the) Life Ladder in a country indexed by $i$ and period $t$, and the rest of the notation is as used in eq. (13). At the estimation stage, all explanatory variables were instrumented by their lagged value, and the dependent variable was instrumented by its second lag. The efficient 'system GMM' estimator, exploiting the levels and changes of instruments, was applied with two-way adjustment (for time and cross-sectional dimensions) and two-step robust estimation.

**Table 6: Logistic mixed-effects regression of individual happiness with the income vulnerability evaluated at the lower middle class income boundary, i.e., at 0.75 of the median income.**

| | *Dependent variable:* Individual Happiness Intensity | | | | | |
|---|---|---|---|---|---|---|
| | DE | ES | FR | IT | PL | UK |
| Income (log) | 0.050 | 0.213* | 0.085 | 0.055 | 0.114 | 0.038 |
| | (0.106) | (0.119) | (0.135) | (0.092) | (0.118) | (0.128) |
| Vulnerability (individual) | −0.511*** | −0.371** | −0.456* | −0.670*** | −0.306* | −0.471** |
| | (0.186) | (0.183) | (0.236) | (0.156) | (0.181) | (0.213) |
| Severe material deprivation | −0.029 | −0.153 | −0.613*** | −0.888*** | −0.670*** | −0.216 |
| | (0.267) | (0.190) | (0.223) | (0.127) | (0.181) | (0.221) |
| Heavy burden of housing cost | −0.228*** | −0.154** | −0.348*** | −0.162*** | −0.220*** | −0.234** |
| | (0.083) | (0.062) | (0.073) | (0.047) | (0.056) | (0.103) |
| Heavy burden of debt repayment | −0.465*** | −0.251*** | −0.349*** | −0.520*** | −0.259** | −0.410*** |
| | (0.152) | (0.093) | (0.110) | (0.144) | (0.108) | (0.121) |
| Cannot find job | 0.053 | −0.119 | −0.568** | −0.398** | 0.062 | 0.266 |
| | (0.302) | (0.188) | (0.252) | (0.169) | (0.588) | (0.289) |
| Capacity to face financial shocks | 0.201*** | 0.176** | 0.202** | −0.234*** | 0.018 | 0.089 |
| | (0.071) | (0.076) | (0.080) | (0.059) | (0.068) | (0.082) |
| Capacity to have holidays | 0.434*** | 0.349*** | 0.353*** | 0.389*** | 0.526*** | 0.274*** |
| | (0.110) | (0.076) | (0.087) | (0.055) | (0.067) | (0.097) |
| Trust | 0.793*** | 0.641*** | 0.523*** | 0.539*** | 0.561*** | 0.504*** |
| | (0.058) | (0.055) | (0.074) | (0.050) | (0.057) | (0.066) |
| Loneliness | −1.768*** | −1.453*** | −1.908*** | −1.231*** | −1.222*** | −1.721*** |
| | (0.261) | (0.250) | (0.265) | (0.238) | (0.262) | (0.307) |
| Degrees of freedom | 13,756 | 15,808 | 9,856 | 17,756 | 12,406 | 10,064 |
| R2 (conditional) | 0.262 | 0.356 | 0.227 | 0.265 | 0.284 | 0.283 |
| R2 (marginal) | 0.133 | 0.232 | 0.175 | 0.152 | 0.173 | 0.144 |
| RMSE | 0.40 | 0.38 | 0.43 | 0.44 | 0.40 | 0.38 |
| Log Likelihood | -6,106 | -4,197 | -3,192 | -6,223 | -4,280 | -3,850 |
| AIC | 12,310 | 8,538 | 6,536 | 12,561 | 8,688 | 7,833 |
| BIC | 12,679 | 9,091 | 7,084 | 13,013 | 9,164 | 8,317 |

*Notes:* *p<0.1; **p<0.05; ***p<0.01. The model estimates the probability of being happy most or all of the time. The household random effects term accounts for the potential correlation of happiness within households. Individual survey weights were used for the model-consistent standard errors shown in parentheses.

strained by two facts. First, its end is limited by availability of data on income from the EU SILC.[32] Second, the consistency of the GMM estimator of dynamic panels relies on the increasing number of cross-sectional dimension relative to time, and the period length was therefore intentionally limited to (the most recent) 7 years, or a quarter of the total number of countries.[33]

The structure of Table 7 is similar to that of Table 2, apart from the fact that there was no need to separately include any regional effect because these were accounted for by the individual country effects through the applied two-ways transformation. A slight difference

[32]The last available year varies for different countries, but for most of them the latest observation is from 2020.

[33]Similar results were obtained shifting the seven-year window backwards.

**Table 7: Dynamic panel estimation with Life Ladder dependent variable.**

| | *Dependent variable:* Life Ladder (log) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| lag of the dependent | 0.336* | 0.313* | 0.221 | 0.154 |
| | (0.201) | (0.167) | (0.185) | (0.132) |
| GDP per capita (log) | 0.188*** | 0.158*** | 0.093* | 0.090* |
| | (0.060) | (0.056) | (0.051) | (0.047) |
| Middle class vulnerability (log) | −0.105 | −0.053** | −0.037** | −0.040** |
| | (0.083) | (0.025) | (0.017) | (0.020) |
| Middle class size (log) | −0.379 | | | |
| | (0.310) | | | |
| Perceptions of corruption | | | −0.167*** | −0.149** |
| | | | (0.047) | (0.073) |
| Social support | | | | −0.023 |
| | | | | (0.244) |
| Healthy life expectancy at birth | | | | −0.002 |
| | | | | (0.004) |
| Freedom to make life choices | | | | 0.076 |
| | | | | (0.073) |
| Generosity | | | | 0.090 |
| | | | | (0.066) |
| Observations | 189 | 189 | 189 | 189 |
| Countries | 28 | 28 | 28 | 28 |
| p-value(Hansen test) | 0.596 | 0.464 | 0.560 | 0.998 |
| d.f.(Hansen test) | 10 | 11 | 12 | 16 |
| p-value(AR(2)) | 0.318 | 0.434 | 0.484 | 0.619 |
| p-value(Wald test for coefficients) | $< 0.001$ | $< 0.001$ | $< 0.001$ | $< 0.001$ |

*Notes:* *p<0.1; **p<0.05; ***p<0.01. The table shows the two-step 'system GMM' estimates with two-way adjustment for country and year effects. The standard errors are shown in brackets. The middle class size and vulnerability indicators were based on EU SILC data, and national Life Ladder valuations and other socioeconomic variables were taken from the World Happiness Report 2022.

in comparison with the static regression result is that, besides the significance of the lagged dependent variable in some specifications, the inclusion of the middle class size in Column (1) made both the middle class vulnerability and size insignificant while preserving the expected sign only for the former. In the remaining specifications the qualitative results are retained.

These results therefore confirmed the previous results despite the use of the (dynamic) panel framework with a contemporaneous Life Ladder indicator instead of the average national scores (over the three years ahead) as the dependent variable. This therefore strongly suggests that middle class vulnerability to poverty is a significant factor explaining the variation in the national Life Ladder scores.

# 6 Conclusion

Looking at the middle class in a group of European countries, the analyses in this study at both individual and aggregate levels confirm the statistical significance and economic importance of vulnerability to poverty for individual happiness perceptions and national happiness scores. In the European countries analysed in this study, at least half of the population were in the middle class when defined as within the range of 0.75–2 times the country's median income. The results at the individual and aggregate levels are therefore naturally mutually consistent. However, the different levels of analysis show some particular complementary aspects.

From the micro perspective, the impact of vulnerability to poverty on the individual happiness of middle class members in the countries examined is very tangible and of similar order to 'real' troubles (severe material deprivation, inability to find a job, or a heavy financial burden of housing or debt repayments). The negative influence remains there even after controlling for the presence of a financial cushion, which renders income but not vulnerability insignificant. The weight of the importance of vulnerability therefore needs to be aligned adequately in both scientific studies and empirical policies. This means that vulnerability should be used as a regular control variable in happiness research and a constantly tracked indicator for economic policy purposes.

At the national level, middle class vulnerability turned out to be among the very few variables that were persistently relevant for the national happiness score across different specifications. Its significance was analogous to variables such as GDP per capita and perceptions of corruption, leaving many other macroeconomic and socioeconomic policy variables insignificant. The size of the middle class itself turned out to be less important.[34] The 'disappearing middle class' narrative should therefore be reoriented towards one of a potentially 'weakening middle class', consistent with Bussolo et al. (2018).

For more vulnerable population segments considered by e.g., Simona-Moussa (2020), the 'at risk of poverty' line is the most relevant threshold for vulnerability valuations. However, for the middle class studied in this paper, the lower income bound of the middle class is an even more relevant threshold. In line with Chang (2013), Rohde et al. (2017), and Ranci et al. (2021), it looks that the prospect of loss of income and potential social status—rather than the fear of deprivation—matters for middle class members. A general income insecurity concept connected with class-specific income thresholds might be needed for the analysis of middle (and upper) class well-being, and should be explored in more depth in the future.

These results underscore the importance of income vulnerability and insecurity, and indicate that there are large costs of anxiety related to the (severe) downside risk of income. They therefore support policies that reduce income uncertainty.

The proposed methodological extension, exploiting a simple transformation of residuals, should also be considered as a potential solution in empirical applications whenever the assumption about the Gaussianity of errors is substantially violated. This approach can be expected to deliver a better distributional fit and a more precise valuation of income vulnerability and insecurity. However, it is not granted that such a straightforward functional transformation like the one used in this paper fits every empirical case. More general extensions, including the use of other flexible parametric functions and nonparametric estimators, might therefore be needed.

An even broader avenue for future research would be to use longitudinal or repeated cross-sectional data. Such data would allow not only tracking income risk and vulnerability over time, but also getting an understanding about their different sources relevant for socioeconomic policies. For example, one could look at the contribution of various transformations and structural changes to income vulnerability, and the relative importance of common and idiosyncratic shocks with their potentially unequal influence on individual happiness.

Finally, a side result derived using only the aggregate data—and therefore to be taken with caution—might serve to remind politicians and policy makers that reducing corruption

[34]Partially because of its correlation with income per capita (positive) and vulnerability (negative).

and setting common rules is important for the happiness of nations. Fundamental equality in terms of the principles and application of law and order matters for the subjective perception of welfare. This is in line with similar results established at various aggregation levels by, e.g., Welsch (2008), Li and An (2020), and Ma et al. (2022).

# References

Abadir, A. (2020). Statistical Nonsignificance in Empirical Economics. *AER: Insights*, 2(2), 193–208.

Alesina, A., and Perotti, R. (1996). Income Distribution, Political Instability, and Investment. *European Economic Review*, 40(6), 1203–1228.

Atkinson, A., and Brandolini, A. (2013). On the Identification of the Middle Class. As Chapter Two in *Income Inequality Economic Disparities and the Middle Class in Affluent Countries* (edited by: Janet C. Gornick and Markus Jäntti), Standford University Press, 2013.

Baldazzi, B., De Carli, R., Lo Castro, D., Savioli, M., Siciliani, I., and Tinto, A. (2019). Analysis of Determinants of Life Satisfaction. In *BES Report 2019: Equitable and Sustainable Wwll-Being in Italy*, ISTAT, `https://www.istat.it/it/files//2019/12/BES-2019-en.pdf`.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using `lme4`. *Journal of Statistical Software*, 67(1), 1—48.

Bates, D. (2022). Computational methods for mixed model. `R` Vignette: Computational Methods, `https://cran.r-project.org/web/packages/lme4/vignettes/Theory.pdf`

Bell, A., Fairbrother, M., and Jones, K. (2019). Fixed and random effects models: making an informed choice. *Quality & Quantity*, 53, 1051—1074.

Bossert, W., and D'Ambrosio, C. (2013). Measuring Economic Insecurity. *International Economic Review*, 54(3), 1017–1030.

Bellemare, M. F. and C. J. Wichman (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, 82, 50—61.

Brenig, M., and Proeger, T. (2018). Putting a price tag on security: Subjective well-being and willingness-to-pay for crime reduction in Europe. *Journal of Happiness Studies*, 19(1), 145—166.

Burbidge, J. B., L. Magee and A. L. Robb (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, 83, 123-–7.

Bussolo, M., Karver, J., and López-Calva, L.F. (2018). Is there a middle-class crisis in Europe? Brookings Institution blog on Future Developments, `https://www.brookings.edu/blog/future-development/2018/03/22/is-there-a-middle-class-crisis-in-europe/`.

Cantril, H. (1965). *The Pattern of Human Concerns*. New Brunswick, NJ: Rutgers University Press.

Carroll, C. D., K. E. Dynan and S. D. Krane (2003). Unemployment risk and precautionary wealth: Evidence from households' balance sheets. *Review of Economics and Statistics*, 85, 586—604.

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

Caria, S.A., and Falco, P. (2018) Does the Risk of Poverty Reduce Happiness? *Economic Development and Cultural Change*, 67(1), 1–28.

Chang, W.-C. (2013). Climbing up the Social Ladders: Identity, Relative Income, and Subjective Well-being. *Social Indicators Research*, 113(1), 513–535.

Chaudhuri, S. (2003). Assessing vulnerability to poverty: concepts empirical methods and illustrative examples. Columbia University, Mimeo (2003).

Chaudhuri, S., Jalan, J., and Suryahadi, A. (2002). Assessing household vulnerability to poverty from cross-sectional data: A methodology and estimates from Indonesia. Discussion Paper Series 0102-52 Department of Economics. Columbia University, New York.

Dendorfer, J. and Kranzinger, S. (2021). The Decline of the Middle Class: New Evidence for Europe. *Journal of Economic Issues*, 55, 914–938.

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

Easterly, W. (2001). The Middle Class Consensus and Economic Development. *Journal of Economic Growth*, 6, 317–335.

Eurofound (2019). Recent developments in the state of the middle classes, Publications Office of the European Union, Luxembourg, `https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef19053en.pdf`.

Eurostat (2020). 2018 – Material deprivation, well-being and housing difficulties. Assessment of the implementation, March 2020. European Commission, Eurostat, Directorate F: Social Statistics, Unit F-4: Income and living conditions; Quality of life, `https://ec.europa.eu/eurostat/documents/1012329/8706724/2018+EU-SILC+module_assessment.pdf`.

Eurostat (2021). Methodological Guidlines and Description of EU-SILC Target Variables. 2021 operation (Version 4_09/12/2020), `https://circabc.europa.eu/sd/a/f8853fb3-58b3-43ce-b4c6-a81fe68f2e50/Methodological%20guidelines%202021%20operation%20v4%2009.12.2020.pdf`.

Ferrer-i-Carbonell, A., and Gowdy, J.M. (2007). Environmental degradation and happiness. *Ecological Economics*, 60, 509–516.

Fleurbaey, M. (2009). Beyond GDP: The Quest for a Measure of Social Welfare. *Journal of Economic Literature*, 47(4), 1029–1075.

Frey, B.S., and Stutzer, A. (2010). Happiness: A New Approach in Economics. *CESifo DICE Report*, 8(4), 3–7.

Frey, B.S., and Stutzer, A. (2019). Public Choice and Happiness. Chapter 39 in *The Oxford Handbook of Public Choice* (eds. Roger D. Congleton, Bernard Grofman and Stefan Voigt), Volume 1, Oxford: Oxford University Press, 779–795.

Fujii, Tomoki. (2016). Concepts and Measurement of Vulnerability to Poverty and Other Issues: A Review of Literature. SSRN Electronic Journal. 10.2139/ssrn.2893054.

Gallarado, M. (2018). Identifying Vulnerability to Poverty: A Critical Survey. *Journal of Economic Surveys*, 32(4), 1074–1105.

Guven, C. (2011). Are happier people better citizens? *Kyklos*, 64, 178–192.

Hacker, J. S., Economic security. In Stiglitz, J., J. Fitoussi and M. Durand (eds.) (2018). For Good Measure: Advancing Research on Well-being Metrics Beyond GDP, OECD Publishing, Paris, https://doi.org/10.1787/9789264307278-en.

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., and Wang, S. (Eds.). (2022). World Happiness Report 2022. New York: Sustainable Development Solutions Network, https://worldhappiness.report/ed/2022/.

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

ILO (2016). Europ's Dissapearing Middle Class? Evidence from the World of Work (ed. D. Vaughan-Whitehead), Edward Elgar Publishing, 2016 (DOI 10.4337/9781786430601).

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36, 149-–76.

Kahneman, D., and Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *PNAS*, 107(38), 16489–16493.

Kaliterna-Lipovčhan, L., and Prizmić, Z., (2016). What differs between happy and unhappy people? *SpringerPlus*, 5, 225.

Kim, J.H., and Ji, P.I. (2015). Significance testing in empirical finance: A critical review and assessment. *Journal of Empirical Finance*, 34, 1–14.

Kim, J.H., and Robinson, A.P. (2019). Interval-Based Hypothesis Testing and Its Applications to Economics and Finance. *Econometrics* 2019, 7(2), 21.

Korinek, A., Schindler, M., and Stiglitz, J. (2021). Technological Progress, Artificial Intelligence, and Inclusive Growth. IMF Working Paper WP/21/166, https://www.elibrary.imf.org/downloadpdf/journals/001/2021/166/article-A001-en.xml.

Krell, K., Frick, J.R., and Grabka, M.M. (2017). Measuring the Consistency of Cross-Sectional and Longitudinal Income Information in EU-SILC. *Review of Income and Wealth*, 63(1), 30–52.

Li, Q., and An, L. (2020). Corruption Takes Away Happiness: Evidence from a Cross-National Study. *Journal of Happiness Studies*, 21, 485–504.

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

López-Calva, L.F., Ortiz-Juarez, E. (2014). A vulnerability approach to the definition of the middle class. *Journal of Economic Inequality*, 12, 23–47.

Ma, J., Guo, B., and Yu, Y. (2022). Perception of Official Corruption, Satisfaction With Government Performance, and Subjective Wellbeing—An Empirical Study From China. *Frontiers in Psychology*, 13.

MacKinnon, J.G., and White, H. (1985). Some heteroscedasticity consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29, 53–57.

Menyhért, B., Cseres-Gergely, Z., Kvedaras, V., Benedetta, M., Filippo, P., and Slavica, Z. (2021). Measuring and monitoring absolute poverty (ABSPO)—Final report. European Commission, Joint Research Centre, `https://data.europa.eu/doi/10.2760/787821`.

OECD (2019). Under Pressure: The Squeezed Middle Class, OECD Publishing, Paris, `https://doi.org/10.1787/689afed1-en`.

Osberg, L. (2015). How Should One Measure Economic Insecurity? OECD Statistics Working Papers, No. 2015/01, OECD Publishing, Paris, `https://doi.org/10.1787/5js4t78q9lq7-en`.

Osberg, L., and Sharpe, A. (2014). Measuring Economic Insecurity in Rich and Poor Nations. *Review of Income and Wealth*, 60(1), S53–S76.

Osberg, L. (2021). Economic insecurity and well-being, DESA Working Paper No. 173, `https://www.un.org/en/file/119928/download?token=EWdmhpy4`.

Pierewan, A.C., and and Tampubolon, G. (2015). Happiness and Health in Europe: A Multivariate Multilevel Model. *Applied Research Quality Life*, 10, 237–252.

Ranci, C., Beckfield, J., Bernardi, L., and Parma, A. (2021). New Measures of Economic Insecurity Reveal its Expansion Into EU Middle Classes and Welfare States. *Social Indicators Research*, 158, 539–562.

Richiardini, M.G. and He, Z. (2020). Measuring economic insecurity: A review of the literature. CeMPA Working Paper Series 1/20, `https://www.iser.essex.ac.uk/wp-content/uploads/files/working-papers/cempa/cempa1-20.pdf`.

Rohde, N., Tang, K.K., Osberg, L., and Rao, D.S.P. (2017). Is it vulnerability or economic insecurity that matters for health? *Journal of Economic Behavior & Organization*, 134, 307–319.

Rohde, N., Tang, K.K., D'Ambrosio, C., Osberg, L., and Rao, D.S.P. (2020). Welfare-based income insecurity in the us and germany: evidence from harmonized panel data. *Journal of Economic Behavior & Organization*, 176, 226–243.

Romaguera-de-la-Cruz, M. (2020). Measuring Economic Insecurity Using a Counting Approach. An Application to Three EU Countries. *Review of Income and Wealth*, 66(3), 558–583.

Sabatini, F. (2014). The relationship between happiness and health: evidence from Italy. *Social Science & Medicine*, 114, 178–87.

Salido, O. and Carabaña, J. (2020). On the Squeezing of the Middle Class: Overview and Prospects for the EU-15. *European Review*, 28(2), 325–342.

Schielzeth, H., Dingemanse, N.J., Nakagawa, S., Westneat, D., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N.A., Garamszegi, L.Z., Araya-Ajoy, Y.G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, 11, 1141–1152.

Simona-Moussa, J. (2020). The Subjective Well-Being of Those Vulnerable to Poverty in Switzerland. *Journal of Happiness Studies*, 21, 1561–1580.

Stiglitz, J. (2012). The Price of Inequality: How Today's Divided Society Endangers Our Future. W. W. Norton & Company.

Stiglitz, J., J. Fitoussi and M. Durand (eds.) (2018). For Good Measure: Advancing Research on Well-being Metrics Beyond GDP, OECD Publishing, Paris, `https://doi.org/10.1787/9789264307278-en`.

Tay, L., Herian, M.N, and Diener, E. (2014). Detrimental Effects of Corruption and Subjective Well-Being. *Social Psychological and Personality Science*, 5, 751–759.

Tsai, A.C., Liou, M., Simak, M., and Cheng, P.E. (2017). On hyperbolic transformations to normality. *Computational Statistics* & *Data Analysis*, 115, 250–266.

Van den Bergh, J.C.J.M. (2022). A procedure for globally institutionalizing a 'beyond-GDP' metric. *Ecological Economics*, 192.

Welsch, H. (2006). Environment and happiness: Valuation of air pollution using life satisfaction data. *Ecological Economics*, 58, 801–813.

Welsch, H. (2008). The welfare costs of corruption. *Applied Economics*, 40, 1839–1849.

# A Appendices

List of appendices:

- Appendix A: Data

  Table A1: Explanatory variables of income

  Table A2: Explanatory variables of individual happiness

  Table A3: Number of observations by country at micro level

  Table A4: Summary statistics of the household share indicators

  Table A5: Summary statistics of the synthetic household compostion indicator

  Table A6: Summary statistics of the aggregate (national) data

- Appendix B: The error transformation approach
- Appendix C: Income prediction models
- Appendix D: Quantile-Quantile plots of residuals
- Appendix E: Estimated model of individual happines
- Appendix F: Aggregate middle class vulnerability at national level
- Appendix G: Potential impact of additional variables at the aggregate level

## A.1 Appendix A: Data

**Table A1: Explanatory variables of income.**

| Variable | Description. [Underlying SILC code] | Considered also by (or motivation to include additionally) |
|---|---|---|
| Age | Fixed effects of age by year. [PX020] | Chang (2013),Caria and Falco (2018), Rohde et al. (2020),Simona-Moussa (2020). |
| Gender | A dummy variable: 1 - female, 0 - otherwise. [RB090] | Chang (2013), Caria and Falco (2018), Rohde et al. (2020), Simona-Moussa (2020). |
| Education | A dummy variable: 1 - tertiary, 0 - otherwise. Tertiary is defined by ISCED levels 5-8. [PE040] | Chang (2013), Caria and Falco (2018), Rohde et al. (2020), Simona-Moussa (2020). |
| Civil status | Fixed effects: never married (ref. level), married, divorced or separated, and widowed. [PB190] | Caria and Falco (2018), Rohde et al. (2020), Simona-Moussa (2020). |
| Children | A dummy variable: 1 - household with children, 0 - otherwise. [HX060] | Rohde et al. (2020). |
| Experience | Logarithm of number of years spent in paid work [PL200] | Chang (2013), Caria and Falco (2018). |
| Contract type | A dummy variable: 1 - permanent contract, 0 - otherwise. [PL140] | Added: Conditions for temporary contracts may differ. |
| High-skill occup. | Occupations with ISCO code starting with 1-3 and 6-7. [PL051] | As an informative aggregate of occupation in Simona-Moussa (2020). |
| White-collar occup. | Occupations with ISCO code starting with 1-5. [PL051] | As an informative aggregatet of occupation in Simona-Moussa (2020). |
| Hours worked | Logarithm of sum of hours worked in main and other jobs. [PL060, PL100] | Caria and Falco (2018), Rohde et al. (2020). |
| Part-time job | A dummy variable: 1 - part-time, 0 - otherwise. [PL031] | Rohde et al. (2020). |
| Self-employed | A dummy variable: 1 - self-employed with or without employees, 0 - otherwise. [PL040] | Chang (2013). |
| Basic activity status | Fixed effects: at work (ref. level), unemployed, retired, inactive. [RB210] | *Not working* in Rohde et al. (2020), *Change in number of members working* in Simona-Moussa (2020). |
| Health | A dummy variable: 1 - good and very good health, 0 - otherwise. [PH010] | Simona-Moussa (2020). |
| HH size | Logarithm of number of household members. [HX040] | Rohde et al. (2020), Simona-Moussa (2020). |
| Sector | Fixed effects of aggregate NACE sector. [PL111] | Simona-Moussa (2020). |

| | | |
|---|---|---|
| Region | Fixed effects of regions at NUTS 1 or 2 level depending on availability. [DB040] | Simona-Moussa (2020). |
| Urbanization | Fixed effects of degree of urbanization: densely (ref. level), intermediate, and thinly populated. [DB100] | Added: a highly significant determinant of income modtivated by its use in Baldazzi et al. (2019) for happiness. |
| HH share: In labor force | Share of HH member in labor force. [RB210] | Simona-Moussa (2020). |
| HHA share: Female | Share of women at work. [RB090, RB210] | Added: to better regard household-equivalisation of income. |
| HHA share: High-skill occupation | Share having high-skill occupation. [PL051, RB210] | *Idem.* |
| HHA share: White-collar occupation | Share having white-collar occupation. [PL051, RB210] | *Idem.* |
| HHA share: Contract type | Share of working under permanent contract. [PL140] | *Idem.* |
| HHA share: Self-employed | Share of self-employed. [PL040] | *Idem.* |
| HHA share: Unemployed. | Share of unemployed. [RB210] | *Idem.* |
| HHA share: Tertiary education | Share of working with tertiary education. [PE040] | *Idem.* |
| HHA share: Good health | Share with good and very good health. [PH010] | *Idem.* |
| HHA share: Sector activity | Separate shares of working in NACE sectors: A, B–E, J–K, L–N, R–U. [PL111, RB210] | *Idem.* |
| HH composition | A synthetic indicator constructed according to a structure presented in Tabl A5. | Added: to better account for a composition structure of a household. |

*Notes:* EU SILC is the source of initial data (codes of variables are shown in square brackets). Abbreviations: EU SILC – European Union Statistics on Income and Living Conditions, HH – household, HHA – household members active in labor market, ISCED – International Standard Classification of Education, NACE – Statistical Classification of Economic Activities in European Union, NUTS – Nomenclature of Territorial Units for Statistics.

**Table A2: Explanatory variables of individual happiness.**

| Variable | Description. [Underlying SILC code where applicable] | Considered also by (or motivation to include additionally) |
|---|---|---|
| Vulnerability | Individual vulnerability of middle class members as defined in eq. (8). | Similar considerations by Caria and Falco (2018) and Simona-Moussa (2020),and implicitly by Rohde et al. (2020).[35] |
| Income | Logarithm of equivalised disposable net income. [HX090] | Kahneman and Deaton (2010), Chang (2013), Caria and Falco (2018), Baldazzi et al. (2019). |
| Age | Age, its square, and an old age dummy (age $\geq$ 60; as in Kahneman and Deaton, 2010). [PX020] | Kahneman and Deaton (2010), Chang (2013), Caria and Falco (2018), Baldazzi et al. (2019), Simona-Moussa (2020) |
| Gender | A dummy variable: 1 - female, 0 - otherwise. [RB090] | Kahneman and Deaton (2010), Chang (2013), Simona-Moussa (2020). |
| Education | Fixed effects: primary, secondary (ref. level), and tertiary education levels (ISCED 0–2, 3–4, and 5–8, respectively). [PE040] | Kahneman and Deaton (2010), Chang (2013), Baldazzi et al. (2019), Simona-Moussa (2020). |
| Civil status | Fixed effects: never married (ref. level), married, divorced or separated, and widowed. [PB190] | Kahneman and Deaton (2010), Chang (2013), Caria and Falco (2018), Simona-Moussa (2020). |
| Children | A dummy variable: 1 - household with children, 0 - otherwise. [HX060] | Kahneman and Deaton (2010), Chang (2013). |
| Basic activity | Fixed effects: in labor force (ref. level), retired, other inactive. [RB210] | Baldazzi et al. (2019), Simona-Moussa (2020). |
| Cannot find a job | A dummy variable: 1 - low work intensity because one cannot find a job, 0 - otherwise. [PL120] | Chang (2013), Baldazzi et al. (2019), Simona-Moussa (2020). |
| Contract type | A dummy variable: 1 - permanent contract, 0 - otherwise. [PL140] | Added: Temporary contracts provide less security and more anxiety about the future. |
| Health limitation | Fixed effects: strongly limited, limited, and no lmitation (ref. level). [PH030] | Kahneman and Deaton (2010), Baldazzi et al. (2019), Simona-Moussa (2020). |

[35] Through the income variance-induced ‘income unpredictability’ influence on expected utility-driven welfare indicators.

| | | |
|---|---|---|
| Environmental problems | A dummy variable: 1 - present, 0 - absent. [HS180] | Added: motivated by, e.g., Welsch (2006) and Ferrer-i-Carbonell and Gowdy (2007). |
| Crime problems | A dummy variable: 1 - present, 0 - absent. [HS190] | Baldazzi et al. (2019). |
| Severe material deprivation | A dummy variable: 1 - yes, 0 - no [RX060] | Baldazzi et al. (2019), Simona-Moussa (2020). |
| Heavy burden of housing cost | A dummy variable: 1 - yes, 0 - no. [HS140] | Added: economic well-being aspect. |
| Heavy burden of debt repayment | A dummy variable: 1 - yes, 0 - no. [HS150] | Added: economic well-being aspect. |
| Capacity to face unexpected financial expenses | A dummy variable: 1 - yes, 0 - no. [HS060] | Added: economic well-being aspect. |
| Capacity to have holidays | A dummy variable: 1 - yes, 0 - no, depending on the capacity to afford paying for one week annual holiday away from home [HS040] | Simona-Moussa (2020). |
| Trust | A dummy variable: 1 - satisfied to completely satisfied with trust in others (levels 8–10 out of 0–10), 0 - otherwise. [PW190T] | Chang (2013), Simona-Moussa (2020). |
| Feeling lonely | A dummy variable: 1 - feeling lonely most or all of the time, 0 - otherwise. [PW230T] | To capture the social (non)engagement considered by Simona-Moussa (2020). |
| Sector | Fixed effects of aggregate NACE sector. [PL111] | Baldazzi et al. (2019). |
| Region | Fixed effects of regions at NUTS 1 or 2 level depending on availability. [DB040] | Baldazzi et al. (2019), Simona-Moussa (2020). |
| Urbanization | Fixed effects of degree of urbanization. [DB100] | Baldazzi et al. (2019). |

*Notes:* EU SILC is the source of initial data (the codes of variables are shown in square brackets). Abbreviations: EU SILC – European Union Statistics on Income and Living Conditions, ISCED – International Standard Classification of Education, ISCO – International Standard Classification of Occupations, NACE – Statistical Classification of Economic Activities in European Union, NUTS – Nomenclature of Territorial Units for Statistics.

**Table A3: Number of observations by country at micro level (for middle class population).**

| Variable | DE | ES | FR | IT | PL | UK |
|---|---|---|---|---|---|---|
| Age | 16609 | 19667 | 16716 | 29046 | 25160 | 22677 |
| Basic activity: At work | 8485 (51.1) | 8353 (42.5) | 7114 (42.6) | 12648 (43.5) | 10774 (42.8) | 11261 (49.7) |
| Basic activity: Unemployed | 176 (1.1) | 1188 (6) | 502 (3) | 993 (3.4) | 594 (2.4) | 227 (1) |
| Basic activity: Retired | 4679 (28.2) | 3938 (20) | 4743 (28.4) | 7653 (26.3) | 6468 (25.7) | 5583 (24.6) |
| Basic activity: Other inactive | 3269 (19.7) | 6187 (31.5) | 4357 (26.1) | 7609 (26.2) | 7324 (29.1) | 5606 (24.7) |
| Being happy most or all of the time: 1 | 9716 (67.4) | 11568 (71.8) | 7136 (51.8) | 9181 (36.2) | 8764 (41.8) | 7458 (39.5) |
| Being happy most or all of the time: 0 | 4297 (29.8) | 4312 (26.8) | 2893 (21) | 8633 (34.1) | 3839 (18.3) | 2680 (14.2) |
| Being happy most or all of the time: Unavail. | 403 (2.8) | 232 (1.4) | 3743 (27.2) | 7514 (29.7) | 8362 (39.9) | 8745 (46.3) |
| Cannot find a job: 1 | 82 (0.5) | 292 (1.5) | 134 (0.8) | 455 (1.6) | 43 (0.2) | 119 (0.5) |
| Cannot find a job: 0 | 16527 (99.5) | 19375 (98.5) | 16582 (99.2) | 28591 (98.4) | 25117 (99.8) | 22558 (99.5) |
| Capacity to face unexpected expenditure: 1 | 13710 (82.5) | 14786 (75.2) | 13006 (77.8) | 21714 (74.8) | 18291 (72.7) | 16859 (74.3) |
| Capacity to face unexpected expenditure: 0 | 2833 (17.1) | 4880 (24.8) | 3668 (21.9) | 7332 (25.2) | 6697 (26.6) | 5816 (25.6) |
| Capacity to have holidays: 1 | 15573 (93.8) | 14677 (74.6) | 14121 (84.5) | 19157 (66) | 17803 (70.8) | 19235 (84.8) |
| Capacity to have holidays: 0 | 999 (6) | 4989 (25.4) | 2559 (15.3) | 9889 (34) | 7274 (28.9) | 3440 (15.2) |
| Civil status: Never Married | 3528 (21.2) | 4767 (24.2) | 4754 (28.4) | 8158 (28.1) | 3344 (13.3) | 5210 (23) |
| Civil status: Married | 8662 (52.2) | 9546 (48.5) | 7031 (42.1) | 13645 (47) | 11998 (47.7) | 10698 (47.2) |
| Civil status: Separated or divorced | 1296 (7.8) | 917 (4.7) | 1151 (6.9) | 1267 (4.4) | 806 (3.2) | 1919 (8.5) |
| Civil status: Widowed | 934 (5.6) | 1403 (7.1) | 894 (5.3) | 2552 (8.8) | 1816 (7.2) | 1107 (4.9) |
| Civil status: Not applicable or unavailabe | 2189 (13.2) | 3034 (15.4) | 2886 (17.3) | 3424 (11.8) | 7196 (28.6) | 3743 (16.5) |
| Contract type: Permanent | 11974 (72.1) | 9158 (46.6) | 9282 (55.5) | 15854 (54.6) | 11588 (46.1) | 9874 (43.5) |
| Contract type: Temporary | 1215 (7.3) | 3066 (15.6) | 1212 (7.3) | 2785 (9.6) | 2607 (10.4) | 365 (1.6) |
| Contract type: Not applicable or unavailable | 3420 (20.6) | 7443 (37.8) | 6222 (37.2) | 10407 (35.8) | 10965 (43.6) | 12438 (54.8) |
| Crime problems: 1 | 1830 (11) | 1961 (10) | 2235 (13.4) | 2923 (10.1) | 1136 (4.5) | 3245 (14.3) |
| Crime problems: 0 | 14779 (89) | 17706 (90) | 14433 (86.3) | 26123 (89.9) | 24024 (95.5) | 11787 (52) |
| Education: Primary | 1324 (8) | 7894 (40.1) | 3602 (21.5) | 11169 (38.5) | 2716 (10.8) | 3072 (13.5) |
| Education: Secondary | 7859 (47.3) | 3854 (19.6) | 5900 (35.3) | 10930 (37.6) | 11231 (44.6) | 3943 (17.4) |
| Education: Tertiary | 5237 (31.5) | 4885 (24.8) | 4008 (24) | 3523 (12.1) | 3997 (15.9) | 5425 (23.9) |
| Education: Not applicable or unavailable | 2189 (13.2) | 3034 (15.4) | 3206 (19.2) | 3424 (11.8) | 7216 (28.7) | 10237 (45.1) |
| Environmental problems: 1 | 3799 (22.9) | 1862 (9.5) | 2212 (13.2) | 2959 (10.2) | 3557 (14.1) | 1885 (8.3) |
| Environmental problems: 0 | 12810 (77.1) | 17805 (90.5) | 14470 (86.6) | 26087 (89.8) | 21603 (85.9) | 13148 (58) |
| Experience[36] | 16609 | 19667 | 16716 | 29046 | 25160 | 22677 |
| Feeling lonely: 1 | 123 (0.7) | 166 (0.8) | 155 (0.9) | 284 (1) | 129 (0.5) | 106 (0.5) |
| Feeling lonely: 0 | 16486 (99.3) | 19501 (99.2) | 16561 (99.1) | 28762 (99) | 25031 (99.5) | 22571 (99.5) |
| Gender: Female | 8559 (51.5) | 10036 (51) | 8543 (51.1) | 14943 (51.4) | 12984 (51.6) | 11501 (50.7) |
| Gender: Male | 8050 (48.5) | 9631 (49) | 8173 (48.9) | 14103 (48.6) | 12176 (48.4) | 11176 (49.3) |
| Health limitation: Strongly limited | 846 (5.1) | 725 (3.7) | 1236 (7.4) | 1468 (5.1) | 1355 (5.4) | 2083 (9.2) |
| Health limitation: Limited | 2159 (13) | 2721 (13.8) | 2145 (12.8) | 4769 (16.4) | 3031 (12) | 3015 (13.3) |
| Health limitation: Not limited | 11257 (67.8) | 13187 (67.1) | 10120 (60.5) | 18990 (65.4) | 13487 (53.6) | 13639 (60.1) |
| Health perception: Good-to-very good | 9660 (58.2) | 12074 (61.4) | 9085 (54.3) | 18290 (63) | 10320 (41) | 13784 (60.8) |
| Health perception: Fair-to-bad | 4729 (28.5) | 4559 (23.2) | 4398 (26.3) | 7018 (24.2) | 7637 (30.4) | 4959 (21.9) |
| Heavy burden of debt repayment: 1 | 420 (2.5) | 2060 (10.5) | 1382 (8.3) | 746 (2.6) | 1850 (7.4) | 1031 (4.5) |
| Heavy burden of debt repayment: 0 | 16118 (97) | 17607 (89.5) | 15318 (91.6) | 28300 (97.4) | 23310 (92.6) | 13995 (61.7) |
| Heavy burden of housing cost: 1 | 1603 (9.7) | 8975 (45.6) | 3372 (20.2) | 11398 (39.2) | 13757 (54.7) | 1664 (7.3) |

[36] Set to zero in cases where it is not applicable or unavailable.

| | | | | | | |
|---|---|---|---|---|---|---|
| Heavy burden of housing cost: 0 | 14903 (89.7) | 10692 (54.4) | 13311 (79.6) | 17648 (60.8) | 11327 (45) | 13369 (59) |
| High-skill occupation: 1 | 8688 (52.3) | 6600 (33.6) | 6979 (41.8) | 11817 (40.7) | 9706 (38.6) | 9663 (42.6) |
| High-skill occupation: 0 | 7921 (47.7) | 13067 (66.4) | 9737 (58.2) | 17229 (59.3) | 15454 (61.4) | 13014 (57.4) |
| Hours worked | 14416 | 16117 | 13772 | 25622 | 20965 | 18883 |
| Household equivalised income | 16609 | 19667 | 16716 | 29046 | 25160 | 22677 |
| Household size | 16609 | 19667 | 16716 | 29046 | 25160 | 22677 |
| Self-employed: 1 | 596 (3.6) | 2151 (10.9) | 907 (5.4) | 3798 (13.1) | 2169 (8.6) | 1726 (7.6) |
| Self-employed: 0 | 16013 (96.4) | 17516 (89.1) | 15809 (94.6) | 25248 (86.9) | 22991 (91.4) | 20951 (92.4) |
| Severely materially deprived: 1 | 141 (0.8) | 397 (2) | 302 (1.8) | 1079 (3.7) | 578 (2.3) | 385 (1.7) |
| Severely materially deprived: 0 | 16468 (99.2) | 19270 (98) | 16414 (98.2) | 27967 (96.3) | 24582 (97.7) | 22292 (98.3) |
| Trust: 1 | 3767 (22.7) | 7154 (36.4) | 2228 (13.3) | 4490 (15.5) | 4184 (16.6) | 3375 (14.9) |
| Trust: 0 | 12842 (77.3) | 12513 (63.6) | 14488 (86.7) | 24556 (84.5) | 20976 (83.4) | 19302 (85.1) |
| Urbanization: Densely populated | – | 9629 (49) | 6674 (39.9) | 8387 (28.9) | 7943 (31.6) | 12041 (53.1) |
| Urbanization: Intermediate | – | 4571 (23.2) | 3101 (18.6) | 12017 (41.4) | 6362 (25.3) | 7277 (32.1) |
| Urbanization: Thinly populated | – | 5467 (27.8) | 6941 (41.5) | 8642 (29.8) | 10855 (43.1) | 3355 (14.8) |
| White-collar occupation: 1 | 11160 (67.2) | 8589 (43.7) | 8543 (51.1) | 14551 (50.1) | 9384 (37.3) | 13081 (57.7) |
| White-collar occupation: 0 | 5449 (32.8) | 11078 (56.3) | 8173 (48.9) | 14495 (49.9) | 15776 (62.7) | 9596 (42.3) |
| With children: 1 | 6099 (36.7) | 9480 (48.2) | 7930 (47.4) | 11395 (39.2) | 12420 (49.4) | 9898 (43.6) |
| With children: 0 | 10510 (63.3) | 10187 (51.8) | 8786 (52.6) | 17651 (60.8) | 12740 (50.6) | 12779 (56.4) |

*Notes:* The proportion of observations is indicated in parentheses for qualitative factors with several states. EU SILC is the source of initial data. Abbreviations: DE – Germany, ES – Spain, FR – France, IT – Italy, PL – Poland, UK – United Kingdom; EU SILC – European Union Statistics on Income and Living Conditions.

**Table A4: Summary statistics of household share indicators (in 2018).**

| Variable | DE | | ES | | FR | | IT | | PL | | UK | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | St.Dev. | Mean | St.Dev. | Mean | St.Dev. | Mean | St.Dev. | Mean | St.Dev. | Mean | St.Dev. |
| HHA share: Female | 0.36 | 0.35 | 0.33 | 0.35 | 0.34 | 0.35 | 0.28 | 0.35 | 0.34 | 0.32 | 0.36 | 0.34 |
| HHA share: Good health | 0.58 | 0.46 | 0.65 | 0.45 | 0.54 | 0.47 | 0.65 | 0.46 | 0.52 | 0.44 | 0.63 | 0.45 |
| HHA share: White-collar occupation | 0.69 | 0.6 | 0.67 | 0.69 | 0.59 | 0.6 | 0.68 | 0.71 | 0.53 | 0.59 | 0.71 | 0.59 |
| HHA share: High-skill occupation | 0.55 | 0.55 | 0.49 | 0.59 | 0.48 | 0.55 | 0.57 | 0.65 | 0.61 | 0.66 | 0.51 | 0.53 |
| HHA share: Tertiary education | 0.34 | 0.49 | 0.41 | 0.57 | 0.34 | 0.51 | 0.22 | 0.47 | 0.26 | 0.44 | 0.31 | 0.51 |
| HH share: In labor force | 0.53 | 0.38 | 0.51 | 0.32 | 0.46 | 0.34 | 0.46 | 0.33 | 0.49 | 0.3 | 0.51 | 0.35 |
| HHA share: Permanent contract | 0.7 | 0.6 | 0.6 | 0.62 | 0.59 | 0.56 | 0.65 | 0.68 | 0.6 | 0.67 | 0.51 | 0.62 |
| HHA share: Self-employed | 0.03 | 0.12 | 0.08 | 0.18 | 0.04 | 0.13 | 0.11 | 0.21 | 0.09 | 0.19 | 0.08 | 0.19 |
| HHA share: Unemployed | 0.03 | 0.14 | 0.09 | 0.19 | 0.05 | 0.15 | 0.06 | 0.16 | 0.04 | 0.12 | 0.02 | 0.1 |
| HHA share: Works in sector A | 0.01 | 0.07 | 0.04 | 0.18 | 0.02 | 0.11 | 0.03 | 0.17 | 0.08 | 0.24 | 0.01 | 0.08 |
| HHA share: Works in sectors B-E | 0.14 | 0.29 | 0.11 | 0.27 | 0.11 | 0.27 | 0.16 | 0.33 | 0.15 | 0.3 | 0.08 | 0.22 |
| HHA share: Works in sectors J-K | 0.06 | 0.2 | 0.04 | 0.17 | 0.04 | 0.18 | 0.04 | 0.16 | 0.03 | 0.14 | 0.06 | 0.2 |
| HHA share: Works in sectors L-N | 0.06 | 0.2 | 0.07 | 0.22 | 0.06 | 0.21 | 0.08 | 0.24 | 0.04 | 0.17 | 0.1 | 0.25 |
| HHA share: Works in sectors R-U | 0.03 | 0.13 | 0.05 | 0.18 | 0.03 | 0.14 | 0.05 | 0.19 | 0.02 | 0.11 | 0.05 | 0.17 |

*Notes:* EU SILC is the source of initial data. Abbreviations: DE – Germany, ES – Spain, FR – France, IT – Italy, PL – Poland, UK – United Kingdom; EU SILC – European Union Statistics on Income and Living Conditions, HH – household, HHA – household members active in labor market,

**Table A5: The structure of the synthetic household composition indicator and summary statistics of its components (in 2018).**

| | | Active | | | | | | Inactive | |
|---|---|---|---|---|---|---|---|---|---|
| | | With tertiary education | | | Without tertiary education | | | Retired | Other |
| Country | Statistics | Age: $< 30$ | $30 - 50$ | $> 50$ | $< 30$ | $30 - 50$ | $> 50$ | | |
| DE | Mean | 0.033 | 0.102 | 0.066 | 0.085 | 0.145 | 0.103 | 0.228 | 0.238 |
| | St.Dev. | 0.147 | 0.222 | 0.199 | 0.199 | 0.250 | 0.247 | 0.399 | 0.285 |
| ES | Mean | 0.033 | 0.127 | 0.042 | 0.051 | 0.160 | 0.093 | 0.162 | 0.332 |
| | St.Dev. | 0.123 | 0.232 | 0.146 | 0.145 | 0.237 | 0.205 | 0.303 | 0.274 |
| FR | Mean | 0.041 | 0.102 | 0.033 | 0.064 | 0.134 | 0.085 | 0.241 | 0.300 |
| | St.Dev. | 0.162 | 0.210 | 0.141 | 0.177 | 0.222 | 0.220 | 0.400 | 0.292 |
| IT | Mean | 0.014 | 0.057 | 0.029 | 0.066 | 0.177 | 0.115 | 0.189 | 0.350 |
| | St.Dev. | 0.078 | 0.169 | 0.124 | 0.162 | 0.254 | 0.226 | 0.333 | 0.293 |
| PL | Mean | 0.026 | 0.080 | 0.019 | 0.083 | 0.184 | 0.093 | 0.203 | 0.312 |
| | St.Dev. | 0.113 | 0.183 | 0.099 | 0.164 | 0.221 | 0.200 | 0.320 | 0.273 |
| UK | Mean | 0.035 | 0.089 | 0.042 | 0.089 | 0.149 | 0.105 | 0.179 | 0.311 |
| | St.Dev. | 0.144 | 0.204 | 0.159 | 0.198 | 0.240 | 0.246 | 0.359 | 0.314 |

*Notes:* EU SILC is the source of initial data. Abbreviations: EU SILC – European Union Statistics on Income and Living Conditions; DE – Germany, ES – Spain, FR – France, IT – Italy, PL – Poland, UK – United Kingdom.

Table A5 shows the structure and average proportion (with standard deviation) of each component of the synthetic indicator. The indicator is constructed by stacking the number of members with each characteristic. For example, let us consider a household with five members. Three of them are in labor force (one is aged 27 and has tertiary education, two members have secondary education and are aged 45 and 49), one is retired, and there is one child. This household had the following value of the synthetic indicator:

$$1\_0\_0\_0\_2\_0\_1\_1.$$

**Table A6: Summary statistics of national data.**

| Variable | Obs. | Mean | Median | St.dev. | Source | Period |
|---|---|---|---|---|---|---|
| | | | **Cross-sectional data (2018)** | | | |
| Freedom to make life choices | 28 | 0.81 | 0.84 | 0.11 | WHR | 2018 |
| Generosity | 28 | -0.09 | -0.13 | 0.15 | WHR | 2018 |
| Gross domestic product per capita (log) | 28 | 10.64 | 10.6 | 0.35 | WHR | 2018 |
| Happiness score | 28 | 6.64 | 6.48 | 0.58 | WHR | 2018 |
| Healthy life expectancy at birth | 28 | 69.87 | 70.74 | 1.96 | WHR | 2018 |
| Perceptions of corruption | 28 | 0.66 | 0.75 | 0.25 | WHR | 2018 |
| Social support | 28 | 0.91 | 0.92 | 0.04 | WHR | 2018 |
| Middle class size | 28 | 0.64 | 0.64 | 0.06 | EU SILC-based | 2018 |
| Middle class vulnerability | 28 | 0.35 | 0.34 | 0.12 | EU SILC-based | 2018 |
| Gini coefficient of equival. disp. income | 28 | 49.22 | 48.75 | 4.91 | Eurostat (online) | 2018 |
| Income quintile share ratio | 28 | 4.88 | 4.32 | 1.24 | Eurostat (online) | 2018 |
| Inflation (consumer) | 28 | 0.02 | 0.02 | 0.01 | Eurostat (online) | 2018 |
| HICP | 28 | 103.86 | 103.53 | 1.86 | Eurostat (online) | 2018 |
| Unemployment rate | 28 | 6.58 | 5.7 | 3.62 | Eurostat (online) | 2018 |
| | | | **Panel data (2014–2020)** | | | |
| Freedom to make life choices | 187 | 0.81 | 0.84 | 0.12 | WHR | 2014 - 2020 |
| Generosity | 187 | -0.04 | -0.07 | 0.16 | WHR | 2014 - 2020 |
| Gross domestic product per capita (log) | 187 | 10.59 | 10.56 | 0.35 | WHR | 2014 - 2020 |
| Healthy life expectancy at birth | 187 | 69.74 | 70.53 | 1.99 | WHR | 2014 - 2020 |
| Life Ladder | 187 | 6.45 | 6.45 | 0.72 | WHR | 2014 - 2020 |
| Perceptions of corruption | 187 | 0.67 | 0.78 | 0.25 | WHR | 2014 - 2020 |
| Social support | 187 | 0.91 | 0.92 | 0.05 | WHR | 2014 - 2020 |
| Middle class size | 189 | 0.63 | 0.64 | 0.06 | EU SILC-based | 2014 - 2020 |
| Middle class vulnerability | 189 | 0.35 | 0.34 | 0.11 | EU SILC-based | 2014 - 2020 |

*Notes:* Data sources: main indicators are from WHR (see Helliwell et al., 2022), and additional robustness controls are from Eurostat online database (`https://ec.europa.eu/eurostat/web/lfs/data/database`). Abbreviations: EU SILC – European Union Statistics on Income and Liging Conditions; HICP – harmonized index of consumer prices; WHR – World Happiness Report.

## A.2 Appendix B: The error transformation approach

Let us consider a real-valued random variable $\varepsilon$ with the cumulative distribution function $F_\mathcal{E}(a) = \mathbb{P}\left(\varepsilon \le a\right)$. Suppose its transformation $V = g(\varepsilon) \sim \mathcal{N}(0, \sigma_V^2), \sigma_V \in \mathbb{R}_+$, for a given $g : \mathbb{R} \to \mathbb{R}$ that is a continuous and strictly increasing function (defined on the whole range of $\varepsilon$). Then it holds

$$\mathbb{P}\left(\varepsilon \le a\right) = \mathbb{P}\left(g(\varepsilon) \le g(a)\right) = \mathbb{P}\left(V \le g(a)\right) = \mathbb{P}\left(\frac{V}{\sigma_V} \le \frac{g(a)}{\sigma_V}\right) = \Phi\left(\frac{g(a)}{\sigma_V}\right) \tag{14}$$

because of the following properties granting each consequent equality:

i) $g : \mathbb{R} \to \mathbb{R}$ is a continuous and strictly increasing function;

ii) $V = g(\varepsilon)$;

iii) $\sigma_V > 0$;

iv) $V \sim \mathcal{N}(0, \sigma_V^2) \Rightarrow \frac{V}{\sigma_V} \sim \mathcal{N}(0, 1)$ with $\mathbb{P}\left(\frac{V}{\sigma_V} \le z\right) = \Phi\left(z\right)$, $z \in \mathbb{R}$, where $\Phi$ is the cumulative distribution function of a standard normal random variable.

Letting $a := \bar{y} - f(\boldsymbol{x})$, $\sigma_V := \sigma_{g(\mathcal{E})}$, and using a conditional distribution of errors instead of the unconditional one applied in the generic result above gives

$$p(\boldsymbol{x}) = \mathbb{P}\big(\varepsilon \le \bar{y} - f(\boldsymbol{x})\big) = \Phi\left(\frac{g\big(\bar{y} - f(\boldsymbol{x})\big)}{\sigma_{g(\mathcal{E})}}\right). \tag{15}$$

## A.3 Appendix C: Income prediction models

Table A7 shows the empirical counterpart of eq. (10) which was estimated with the data of whole population, i.e., including also those who are below an income threshold. The employed set of basic features included age, gender, education, job/occupation type, health, employment sector, as well as geographic characteristics (urbanization level and region indicator). The individual- and household-level dimensions are present in Table A7. The latter included the proportions of household (HH) members with a particular characteristic and the synthetic household composition indicator defined according to the structure shown in Table A5. Neither regional nor urbanization indicators were available for Germany in the EU SILC. They are therefore missing in column DE in Table A7.

**Table A7: Income prediction models.**

| | *Dependent variable:* Household Equivalised Net Income (log) | | | | | |
|---|---|---|---|---|---|---|
| | DE | ES | FR | IT | PL | UK |
| HHA share: Female | −0.098*** | −0.069*** | −0.011 | −0.083*** | −0.006 | −0.055*** |
| | (0.013) | (0.012) | (0.011) | (0.012) | (0.010) | (0.012) |
| HHA share: Good health | 0.014 | 0.047*** | 0.029*** | 0.009 | 0.012 | 0.145*** |
| | (0.013) | (0.014) | (0.011) | (0.014) | (0.009) | (0.012) |
| HHA share: White collar occupation | 0.121*** | 0.069*** | 0.087*** | 0.092*** | 0.050*** | 0.113*** |
| | (0.012) | (0.009) | (0.009) | (0.007) | (0.007) | (0.010) |
| HHA share: High skill occupation | 0.100*** | 0.077*** | 0.108*** | 0.060*** | 0.045*** | 0.203*** |
| | (0.011) | (0.009) | (0.009) | (0.008) | (0.006) | (0.010) |
| HHA share: Tertiary education | −0.022 | 0.001 | 0.032** | 0.079*** | 0.068*** | −0.075*** |
| | (0.020) | (0.013) | (0.016) | (0.014) | (0.014) | (0.017) |
| HH share: In labor force | 0.091*** | 0.177*** | 0.024** | 0.096*** | 0.034*** | 0.007 |
| | (0.013) | (0.009) | (0.010) | (0.009) | (0.007) | (0.009) |
| HHA share: Permanent contract | 0.091*** | 0.177*** | 0.024** | 0.096*** | 0.034*** | 0.007 |
| | (0.013) | (0.009) | (0.010) | (0.009) | (0.007) | (0.009) |
| HHA share: Self-employed | 0.082** | −0.030 | −0.122*** | 0.016 | −0.199*** | −0.200*** |
| | (0.035) | (0.026) | (0.028) | (0.022) | (0.020) | (0.022) |
| HHA share: Unemployed | −0.605*** | −0.740*** | −0.290*** | −0.736*** | −0.670*** | −0.902*** |
| | (0.083) | (0.058) | (0.062) | (0.069) | (0.057) | (0.068) |
| HHA share: Works in sector A | −0.137* | −0.015 | −0.014 | −0.107*** | −0.480*** | 0.010 |
| | (0.071) | (0.026) | (0.038) | (0.027) | (0.017) | (0.058) |
| HHA share: Works in sectors B–E | 0.152*** | 0.131*** | 0.139*** | 0.073*** | 0.009 | 0.133*** |
| | (0.018) | (0.017) | (0.016) | (0.015) | (0.012) | (0.020) |
| HHA share: Works in sectors J–K | 0.171*** | 0.174*** | 0.117*** | 0.131*** | 0.101*** | 0.318*** |
| | (0.025) | (0.028) | (0.024) | (0.028) | (0.025) | (0.023) |
| HHA share: Works in sectors L–N | −0.004 | 0.017 | 0.079*** | −0.073*** | −0.022 | 0.116*** |
| | (0.025) | (0.020) | (0.021) | (0.019) | (0.020) | (0.018) |
| HHA share: Works in sectors R–U | −0.148*** | −0.235*** | −0.079** | −0.286*** | 0.022 | −0.050** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | (0.034) | (0.024) | (0.031) | (0.025) | (0.030) | (0.024) |
| Female | 0.027*** | −0.009 | −0.001 | −0.002 | −0.013** | 0.009 |
| | (0.008) | (0.008) | (0.006) | (0.008) | (0.005) | (0.007) |
| Good health | 0.080*** | 0.049*** | 0.052*** | 0.043*** | 0.048*** | −0.004 |
| | (0.009) | (0.009) | (0.007) | (0.009) | (0.007) | (0.008) |
| White-collar occupation | 0.094*** | 0.114*** | 0.114*** | 0.115*** | 0.084*** | 0.058*** |
| | (0.010) | (0.009) | (0.008) | (0.009) | (0.008) | (0.009) |
| High-skill occupation | 0.054*** | 0.045*** | 0.068*** | 0.026*** | 0.009 | 0.053*** |
| | (0.009) | (0.009) | (0.008) | (0.008) | (0.007) | (0.009) |
| Tertiary education | 0.105*** | 0.132*** | 0.124*** | 0.142*** | 0.078*** | 0.088*** |
| | (0.010) | (0.011) | (0.010) | (0.013) | (0.010) | (0.010) |
| Hours worked (IHST) | 0.154*** | 0.167*** | 0.102*** | 0.232*** | −0.007 | 0.158*** |
| | (0.011) | (0.016) | (0.010) | (0.020) | (0.006) | (0.008) |
| Work experience (IHST) | 0.033*** | 0.019*** | 0.040*** | 0.029*** | 0.017*** | −0.029*** |
| | (0.005) | (0.004) | (0.004) | (0.004) | (0.004) | (0.003) |
| Permanent contract | 0.104*** | 0.092*** | −0.022** | 0.031*** | 0.049*** | 0.011 |
| | (0.012) | (0.010) | (0.009) | (0.011) | (0.008) | (0.012) |
| Self-employed | 0.007 | −0.016 | −0.061*** | −0.060*** | −0.012 | −0.033** |
| | (0.021) | (0.015) | (0.014) | (0.013) | (0.011) | (0.015) |
| Unemployed | 0.207*** | 0.188*** | 0.092*** | 0.370*** | −0.024 | 0.524*** |
| | (0.038) | (0.029) | (0.035) | (0.062) | (0.018) | (0.056) |
| Retired [ref. Active] | 0.106*** | 0.115*** | −0.012 | 0.208*** | −0.127*** | 0.406*** |
| | (0.032) | (0.032) | (0.036) | (0.062) | (0.017) | (0.051) |
| Inactive [ref. Active] | 0.215*** | 0.262*** | 0.175*** | 0.420*** | −0.040*** | 0.533*** |
| | (0.023) | (0.029) | (0.033) | (0.061) | (0.014) | (0.049) |
| Married [ref. Never married] | 0.043*** | 0.008 | 0.008 | −0.023** | 0.011 | 0.042*** |
| | (0.012) | (0.010) | (0.009) | (0.011) | (0.009) | (0.010) |
| Divorced or separ. [ref. Never marr.] | −0.036*** | −0.052*** | 0.001 | −0.034** | −0.041*** | 0.007 |
| | (0.013) | (0.016) | (0.012) | (0.016) | (0.014) | (0.012) |
| Widowed [ref. Never married] | 0.166*** | 0.155*** | 0.121*** | 0.129*** | 0.016 | 0.001 |
| | (0.017) | (0.018) | (0.016) | (0.017) | (0.013) | (0.017) |
| With children | 0.003 | −0.095*** | 0.005 | −0.096*** | 0.056*** | −0.038*** |
| | (0.016) | (0.014) | (0.014) | (0.013) | (0.009) | (0.015) |
| Household size (log) | 0.127*** | 0.038 | −0.005 | −0.052 | −0.060** | 0.034 |
| | (0.041) | (0.032) | (0.033) | (0.039) | (0.031) | (0.031) |
| Intercept | 8.778*** | 8.877*** | 9.337*** | 9.213*** | 8.733*** | 8.682*** |
| | (0.073) | (0.063) | (0.092) | (0.089) | (0.059) | (0.075) |
| FE: Age | + | + | + | + | + | + |
| FE: Sector of activity | + | + | + | + | + | + |
| FE: Urbanization | - | + | + | + | + | + |
| FE: Region | - | + | + | + | + | + |
| RE: SD of houshehold composition | 0.280 | 0.294 | 0.250 | 0.333 | 0.324 | 0.225 |
| SD of Residual | 0.420 | 0.432 | 0.327 | 0.521 | 0.374 | 0.459 |
| Degrees of freedom | 21,716 | 27,034 | 19,690 | 39,210 | 33,185 | 31,152 |
| R2 (conditional) | 0.591 | 0.546 | 0.576 | 0.432 | 0.548 | 0.578 |
| R2 (marginal) | 0.410 | 0.337 | 0.328 | 0.201 | 0.208 | 0.477 |
| RMSE | 0.44 | 0.53 | 0.38 | 0.59 | 0.43 | 0.50 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Log Likelihood | -15,846.620 | -26,171.440 | -11,382.420 | -41,609.550 | -23,096.900 | -26,309.490 |

*Notes:* *p<0.1; **p<0.05; ***p<0.01. EU SILC is the source of initial data. Abbreviations: HH – household, HHA – household members active in labor market, IHST – inverse hyperbolic sine transform, FE – fixed effects, RE – random effects, SD – standard deviation, SILC – European Union Statistics on Income and Living Conditions; DE – Germany, ES – Spain, FR – France, IT – Italy, PL – Poland, UK – United Kingdom. The economic activity codes (A–U) are from the Statistical Classification of Economic Activities in the European Community.

The explanatory power of the models in terms of the (conditional) $R^2$ is good. Notice that the number of observations is larger than in typical similar cross-sectional studies.[37] The contribution of the household proportions and, especially, of the household composition indicator was predominant here.[38]

The results are mostly similar across the countries. Part of the variation was induced by the multicollinearity of household and individual characteristics. For instance, the gender variable was often insignificant at individual level (even positive in column DE) while it tended to be significantly negative for the household proportion of women. In general, when household proportions were insignificant, the corresponding individual coefficients typically were, and vice versa. The signs of significant coefficients are mostly in line with the expected impact. An exception is the proportion of tertiary educated in the UK that has a significant negative coefficient. It is nevertheless smaller than the postive individual-level coefficient. The predicted total impact therefore would still be positive, e.g., for households with one person.

When the interest lies in identifying the impact of variables, the described lack of identification due to multicollinearity might be important and require additional actions. In this paper, the income equations were needed only for predictions. The presented equations therefore were used to get $\widehat{f}$ and $\widehat{\varepsilon}$ without any further engagement. These predictions took into account not only the coefficients shown in the table, but also all the fixed and random effects.

[37] Compare, e.g., with the results in Chang (2013).

[38] The results that include only the individual-level characteristics are available upon request.

## A.4 Appendix D: Quantile-Quantile plots of residuals

Figure A1 on page 53 here

## A.5 Appendix E: Estimated model of individual happines

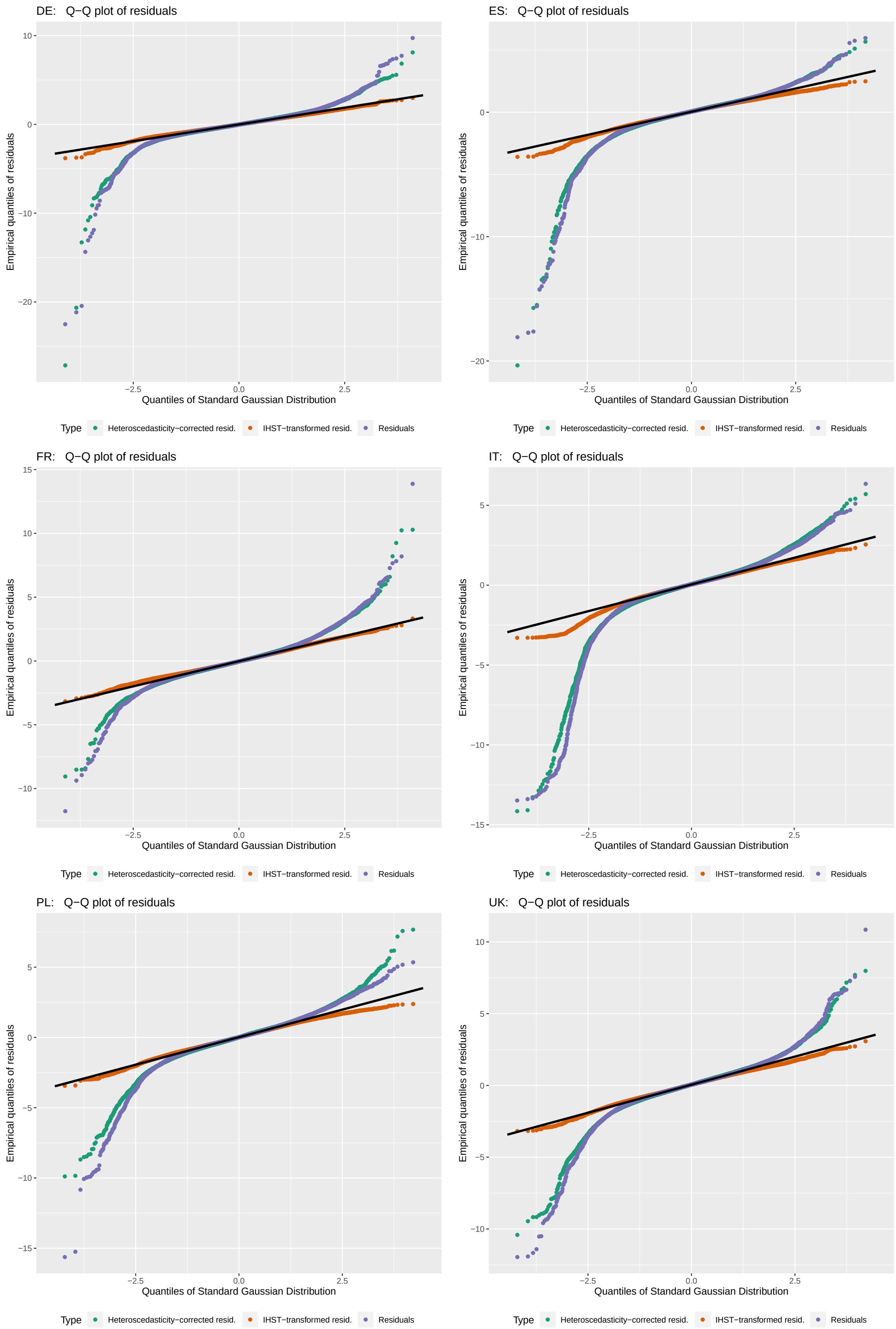
**Table A8: Logistic mixed-effects regression of individual happiness (with a complete list of explanatory variables).**

| | *Dependent variable:* Individual Happiness Intensity | | | | | |
|---|---|---|---|---|---|---|
| | DE | ES | FR | IT | PL | UK |
| Income (log) | 0.260** | 0.363*** | 0.298** | 0.114 | 0.298*** | 0.118 |
| | (0.101) | (0.115) | (0.129) | (0.089) | (0.113) | (0.125) |
| Vulnerability (individual) | −0.808*** | −0.659*** | −0.878*** | −0.813*** | −0.485** | −0.735*** |
| | (0.249) | (0.216) | (0.330) | (0.185) | (0.223) | (0.264) |
| Severe material deprivation | −0.498** | −0.468** | −1.041*** | −0.992*** | −1.010*** | −0.347’ |
| | (0.252) | (0.185) | (0.213) | (0.122) | (0.173) | (0.216) |
| Heavy burden of housing cost | −0.322*** | −0.281*** | −0.444*** | −0.233*** | −0.319*** | −0.346*** |
| | (0.081) | (0.058) | (0.071) | (0.045) | (0.054) | (0.099) |
| Heavy burden of debt repayment | −0.623*** | −0.333*** | −0.435*** | −0.564*** | −0.362*** | −0.515*** |
| | (0.147) | (0.091) | (0.107) | (0.142) | (0.105) | (0.118) |
| Cannot find job | 0.001 | −0.188 | −0.593** | −0.385** | 0.117 | 0.240 |
| | (0.296) | (0.185) | (0.251) | (0.167) | (0.576) | (0.287) |
| Strong health limit. [ref. Unlimit.] | −1.287*** | −1.738*** | −0.792*** | −1.523*** | −0.964*** | −1.171*** |
| | (0.101) | (0.120) | (0.096) | (0.120) | (0.097) | (0.096) |
| Health limit. [ref. Unlimit.] | −0.616*** | −0.911*** | −0.599*** | −0.602*** | −0.505*** | −0.543*** |
| | (0.064) | (0.068) | (0.073) | (0.057) | (0.067) | (0.075) |
| Environmental problems | −0.019 | −0.438*** | −0.314*** | −0.234*** | 0.049 | −0.250*** |
| | (0.059) | (0.092) | (0.079) | (0.071) | (0.075) | (0.085) |
| Crime problems | −0.364*** | −0.140’ | −0.183** | −0.103 | −0.081 | −0.299*** |
| | (0.075) | (0.090) | (0.081) | (0.071) | (0.121) | (0.072) |
| Married [ref. Never married] | 0.487*** | 0.669*** | 0.214** | −0.049 | 0.577*** | 0.547*** |
| | (0.083) | (0.085) | (0.087) | (0.068) | (0.097) | (0.090) |
| With children | −0.003 | 0.020 | 0.275** | 0.152** | 0.156* | 0.046 |
| | (0.089) | (0.086) | (0.120) | (0.072) | (0.083) | (0.107) |
| Retired [ref. Active] | 0.428*** | 0.316** | 0.267 | 0.650*** | 0.297* | 1.014*** |
| | (0.164) | (0.135) | (0.198) | (0.137) | (0.155) | (0.260) |
| Age | −0.070*** | −0.095*** | −0.071*** | −0.040*** | −0.077*** | −0.081*** |
| | (0.013) | (0.011) | (0.012) | (0.009) | (0.012) | (0.013) |
| $(\text{Age}/100)^2$ | 0.044*** | 0.065*** | 0.050*** | 0.025*** | 0.045*** | 0.076*** |
| | (0.015) | (0.012) | (0.012) | (0.010) | (0.012) | (0.014) |
| Old age (> 60) | 0.275** | 0.346*** | 0.236’ | 0.039 | 0.095 | 0.094 |
| | (0.115) | (0.123) | (0.152) | (0.096) | (0.112) | (0.130) |
| Female | 0.080* | −0.134** | −0.217*** | −0.025 | −0.003 | −0.119** |
| | (0.047) | (0.056) | (0.059) | (0.045) | (0.054) | (0.059) |
| Temporary contract [ref. Permanent] | −0.264*** | −0.038 | −0.123 | −0.148** | −0.099 | 0.047 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | (0.085) | (0.076) | (0.106) | (0.073) | (0.078) | (0.168) |
| No contract/Other [ref. Permanent] | −0.233** | 0.011 | −0.104 | −0.060 | 0.013 | 0.019 |
| | (0.092) | (0.073) | (0.078) | (0.055) | (0.078) | (0.080) |
| Oth.inactive [ref. Employed] | −0.244* | 0.226** | 0.101 | 0.547*** | 0.111 | 0.487* |
| | (0.144) | (0.114) | (0.184) | (0.123) | (0.142) | (0.251) |
| Primary educ. [ref. Second. educ.] | −0.118 | 0.021 | −0.051 | −0.136*** | −0.214** | −0.013 |
| | (0.086) | (0.071) | (0.072) | (0.050) | (0.084) | (0.082) |
| Tertiary educ. [ref. Second. educ.] | −0.180*** | 0.110 | −0.067 | 0.020 | 0.007 | −0.094 |
| | (0.053) | (0.077) | (0.076) | (0.065) | (0.068) | (0.070) |
| Separated or divorced [ref. Never married] | 0.058 | 0.032 | 0.210** | −0.107 | −0.036 | 0.224** |
| | (0.093) | (0.121) | (0.107) | (0.098) | (0.131) | (0.107) |
| Widowed [ref. Never married] | 0.043 | 0.113 | −0.061 | −0.155' | 0.203' | 0.085 |
| | (0.115) | (0.124) | (0.129) | (0.095) | (0.124) | (0.145) |
| Intercept | 0.541 | −0.191 | 5.537 | 1.049 | −0.052 | 1.365 |
| | (1.078) | (2.129) | (5.163) | (1.690) | (1.984) | (1.331) |
| FE: Household size | + | + | + | + | + | + |
| FE: Sector of activity | + | + | + | + | + | + |
| FE: Urbanization | - | + | + | + | + | + |
| FE: Region | - | + | + | + | + | + |
| RE: SD of household RE | 0.769 | 0.800 | 0.473 | 0.709 | 0.697 | 0.801 |
| SD of Residual | 0.789 | 0.679 | 0.800 | 0.721 | 0.756 | 0.755 |
| Degrees of freedom | 13,824 | 15,812 | 9,899 | 17,760 | 12,513 | 10,068 |
| R2 (conditional) | 0.231 | 0.331 | 0.197 | 0.242 | 0.253 | 0.266 |
| R2 (marginal) | 0.092 | 0.200 | 0.142 | 0.126 | 0.143 | 0.123 |
| RMSE | 0.40 | 0.38 | 0.42 | 0.43 | 0.40 | 0.38 |
| Log Likelihood | -6,295 | -4,299 | -3,272 | -6,311 | -4,407 | -3,908 |
| AIC | 12,679 | 8,734 | 6,687 | 12,729 | 8,933 | 7,941 |
| BIC | 13,018 | 9,256 | 7,206 | 13,150 | 9,380 | 8,396 |

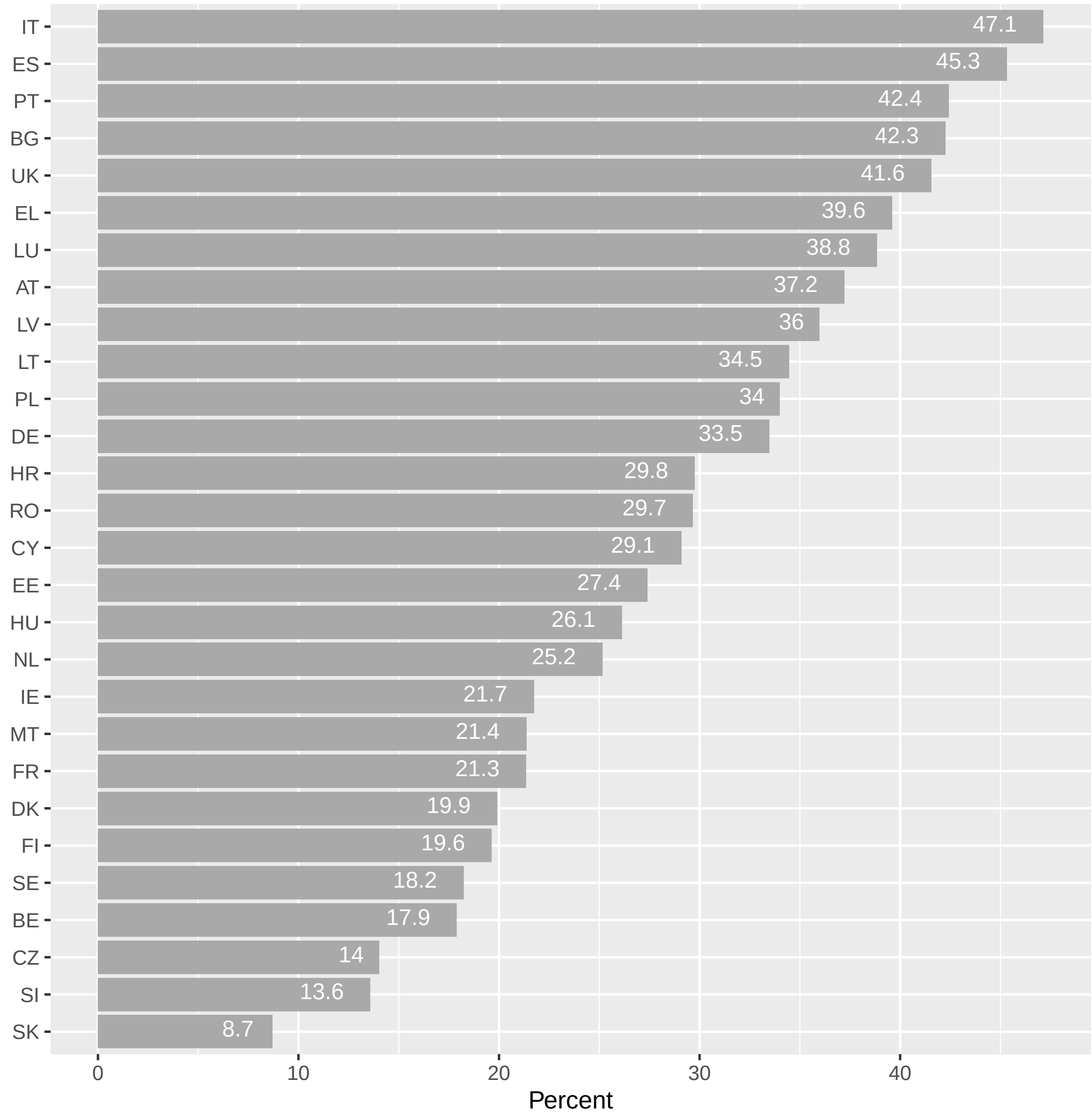
*Notes:* *p<0.1; **p<0.05; ***p<0.01. The model estimates the probability of being happy most or all of the time. The household random effects term accounts for the potential correlation of happiness within households. Individual survey weights were used for the model-consistent standard errors that are shown in parentheses. For factors with more than two variables, the respective reference level is shown in square brackets. The regional and urbanization dimensions were unavailable for Germany.

**Figure A1: Quantile – Quantile (Q–Q) plots of residuals.** Color code: blue – original residuals, green – heteroscedasticity-corrected distribution of residuals, brawn – IHST-transformed residuals.



## A.6 Appendix F: Aggregate middle class vulnerability at national level

**Figure A2: Proportion of vulnerable middle class in 2018 (at $\alpha = 0.1$).** The ISO 3166-1 alpha-2 country codes are shown on the left side.




## A.7 Appendix G: Potential impact of additional variables at the aggregate level

Table A9 on the following page shows some additional results. Here Column (2) of Table 2 is agumented with additional control variables individually. Only *perceptions of corruption* is significant that is shown in the last column.

**Table A9: Cross-sectional regressions of national happiness scores (with individual control variables).**

| | *Dependent variable:* | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) |
| NWE | $0.075^{**}$ | $0.072^{*}$ | $0.069^{**}$ | $0.073^{**}$ | $0.079^{**}$ | $0.074^{**}$ | $0.073^{**}$ | $0.072^{**}$ | $0.070^{**}$ | $0.074^{**}$ | $0.067^{**}$ | $0.073^{**}$ | $0.074^{**}$ | 0.016 |
| | (0.030) | (0.035) | (0.032) | (0.031) | (0.032) | (0.031) | (0.031) | (0.031) | (0.028) | (0.031) | (0.029) | (0.033) | (0.030) | (0.033) |
| GDP per capita (log) | $0.114^{**}$ | $0.114^{**}$ | $0.126^{**}$ | $0.113^{**}$ | $0.107^{**}$ | $0.113^{**}$ | $0.114^{**}$ | $0.115^{**}$ | $0.112^{**}$ | $0.110^{*}$ | $0.103^{*}$ | $0.111^{**}$ | $0.112^{*}$ | $0.087^{*}$ |
| | (0.055) | (0.055) | (0.056) | (0.053) | (0.051) | (0.051) | (0.053) | (0.052) | (0.053) | (0.056) | (0.053) | (0.053) | (0.060) | (0.045) |
| Middle class vulnerability (log) | $-0.049^{**}$ | $-0.049^{**}$ | $-0.051^{***}$ | $-0.049^{**}$ | $-0.043^{*}$ | $-0.051^{*}$ | $-0.050^{**}$ | $-0.050^{**}$ | $-0.046^{*}$ | $-0.050^{**}$ | $-0.042^{*}$ | $-0.050^{**}$ | $-0.050^{**}$ | $-0.045^{**}$ |
| | (0.022) | (0.022) | (0.018) | (0.022) | (0.022) | (0.027) | (0.020) | (0.019) | (0.023) | (0.019) | (0.021) | (0.019) | (0.020) | (0.017) |
| CEE | 0.003 | | | | | | | | | | | | | |
| | (0.020) | | | | | | | | | | | | | |
| SOE | | −0.003 | | | | | | | | | | | | |
| | | (0.020) | | | | | | | | | | | | |
| EMU | | | −0.017 | | | | | | | | | | | |
| | | | (0.016) | | | | | | | | | | | |
| Unemployment rate | | | | −0.0003 | | | | | | | | | | |
| | | | | (0.002) | | | | | | | | | | |
| Income inequality (Gini) | | | | | −0.001 | | | | | | | | | |
| | | | | | (0.002) | | | | | | | | | |
| Income inequality (S8020) | | | | | | 0.0003 | | | | | | | | |
| | | | | | | (0.013) | | | | | | | | |
| Inflation (consumer) | | | | | | | 0.105 | | | | | | | |
| | | | | | | | (1.162) | | | | | | | |
| HICP (log) | | | | | | | | 0.303 | | | | | | |
| | | | | | | | | (0.473) | | | | | | |
| Social support | | | | | | | | | 0.133 | | | | | |
| | | | | | | | | | (0.261) | | | | | |
| Life expectancy[a] | | | | | | | | | | 0.001 | | | | |
| | | | | | | | | | | (0.005) | | | | |
| Freedom of choice[b] | | | | | | | | | | | 0.111 | | | |
| | | | | | | | | | | | (0.081) | | | |
| Generosity | | | | | | | | | | | | 0.007 | | |
| | | | | | | | | | | | | (0.065) | | |
| Confidence in government | | | | | | | | | | | | | 0.002 | |
| | | | | | | | | | | | | | (0.056) | |
| Perceptions of corruption | | | | | | | | | | | | | | $-0.165^{**}$ |
| | | | | | | | | | | | | | | (0.061) |
| Constant | 0.581 | 0.584 | 0.465 | 0.601 | 0.711 | 0.591 | 0.583 | −0.836 | 0.491 | 0.576 | 0.625 | 0.611 | 0.603 | $1.005^{*}$ |
| | (0.578) | (0.574) | (0.587) | (0.561) | (0.561) | (0.540) | (0.557) | (2.366) | (0.539) | (0.573) | (0.539) | (0.565) | (0.626) | (0.492) |

*Notes:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$. The table shows the ordinary least squares estimates of eq. (13). The robust standard errors are shown in brackets. The vulnerability indicator was based on EU SILC data. Other socioeconomic variables were taken either from the World Happiness Report 2022 or . the Eurostat online database. Abbreviations: CEE - Central and Eastern Europe, EMU - Economic and Monetary Union, GDP - Gross Domestic Product , NWE - North-West Europe, HICP - Harmonized Index of Consumer Prices, SoE - Southern Europe.

[a] Healthy life expectancy at birth.

[b] Freedom to make life choices.

## GETTING IN TOUCH WITH THE EU

### In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

### On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

## FINDING INFORMATION ABOUT THE EU

### Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

### EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

### EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

### Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

## The European Commission's science and knowledge service

Joint Research Centre

### JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

**EU Science Hub**
joint-research-centre.ec.europa.eu

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

EU Science