the review of income and wealth

Review of Income and Wealth
Series 69, Number 1, March 2023
DOI: 10.1111/roiw.12553

ANCHORING MEASUREMENT OF THE MIDDLE-INCOME CLASS TO SUBJECTIVE EVALUATION

BY M. GRAZIA PITTAU and ROBERTO ZELLI*

Sapienza University of Rome

What constitutes the middle class is hotly debated. Following an income-based approach, a main issue concerns how to fix the income boundaries that define the middle-income tier. This paper offers a novel model-based approach to the use of self-reported class evaluation for identifying those boundaries. The self-declared status responses are modeled using a non-conventional parametrization of an ordered logistic model. In this parametrization, the cut-points of the model are directly interpretable as income boundaries, and the variance of the errors captures the idiosyncratic heterogeneity of the outcome variable. The use of subjective data is exemplified in the estimation of the middle class in Kazakhstan over the period 2003–2015.

JEL Codes: C25, D31, I30

Keywords: middle class, income boundaries, self-perception, ordered logistic model, Kazakhstan

1. Introduction

"A strong and prosperous middle class is important for the economy and society as a whole, and notably to sustain consumption and investment in education, health and housing" (OECD, 2019, p. 32). The expansion of the middle class has also contributed to democratic movements and progressive but moderate political reforms, especially those that promote inclusive growth.

There is unanimous consensus that middle class refers to a group of people who share the same socioeconomic status. However, there is no agreement on what middle status represents and its definition varies according to different academic traditions and perspectives (Atkinson and Brandolini, 2013). Sociologists tend to identify the middle class in terms of its functional position in the society, typically occupation and educational level. Economists instead tend to characterize the

Notes: We are grateful to Nicholas Longford, Francisco Ferreira, Maria Ana Lugo, and two anonymous reviewers for their useful comments and suggestions on previous versions of the paper. We also thank Luis Felipe López-Calva, William Hutchins Seitz, Aibek Baibagysh Uulu, and participants at the ECAPOV HBSS seminar at the World Bank and at the IARIW-World Bank Conference on "New Approaches to Defining and Measuring Poverty in a Growing World." We are indebted to Sarosh Sattar for having involved us in the issue of shaping middle class in Kazakhstan.

Funding information: Open Access Funding provided by Universita degli Studi di Roma La Sapienza within the CRUI-CARE Agreement. [Correction added on 21st May 2022, after first online publication: CRUI funding statement has been added.]

*Correspondence to: Roberto Zelli, Department of Statistical Sciences, Sapienza University of Rome, Piazzale Aldo Moro, 5, 00185 Roma RM, Italy (roberto.zelli@uniroma1.it).

middle class status using income as benchmark and the term middle-income class rather than middle class. The general practice is to fix a lower and an upper threshold in the income distribution, and measuring the size of the middle-income class as the share of the population with income between those bounds. The thresholds can be either relative (percentiles or percentage of median income) or absolute (a certain amount of income per day or month), in accordance with the methodology to identify poverty lines.[1] The use of income as an indicator of the class status has considerable advantages because it is well grounded on welfare economics and is comparable across countries and time. It also has several cons, one of which is that fixing the income levels for the middle is a debatable task and the procedure somehow arbitrary.

Middle class has also been defined using subjective measures. Subjective social class is based on people's perceptions, asking people to make a "subjective" judgment about their social class. Classes people are usually asked to choose from are: lower, middle, upper-middle, and rich, which are explicitly hierarchically ordered. These self-perceptions are generally related to "objective" data, but suffer from heterogeneity due to factors that are not considered relevant for identification such as idiosyncratic-mood effects, personality traits, or simply errors (Ravallion, 2014): After all people are free to believe, they are in any social class they choose. This issue has largely prevented the use of self-declared social status in measuring the middle class. At the same time, it cannot be ignored that those who refer to a social class should have a basic idea of the income level needed to fall into that class. Whatever else they have in common, those who feel belonging to the middle class have similar sentiments regarding their position in the income distribution. Thus, subjective questions provide some basis for identifying the relevant middle income bounds, and so objective measures may benefit from people's perception of belonging to the middle class.

What the paper aims at providing is a rigorous model-based way to empirically identify income boundaries of the middle class exploiting the information coming from subjective evaluation of own-welfare. We recognize income as the class identification metric, but interaction between objective and subjective social class is used to anchor the monetary metric to subjective evaluations. The self-declared status responses are conveniently modeled using an ordered logistic model with a parametrization different from the conventional one. In this parametrization, the cut-points or thresholds are on the income scale and directly interpretable as income boundaries, and the variance of the errors captures the idiosyncratic heterogeneity of the outcome variable.

The use of subjective data to calibrate the middle-income class is exemplified in the estimation of the middle-class size in Kazakhstan, whose official household budget survey (KHBS) was planned to include subjective measure of social classes. Rapid growth of emerging countries over the past decades has led to increasing

[1]In advanced economies, where the middle class is perceived as the group in the middle of the income distribution, economic thresholds are identified using values around the median income, such as 0.75 and 1.25 times the median (Pew Research Center, 2016). In developing economies, scholars opt for absolute thresholds since median incomes do not necessarily ensure a standard of living compatible with a middle class status. A rather different approach for categorizing individuals without resort to such boundaries is suggested by Anderson *et al.* (2016) using mixture models. In that case, the classification is partial in the sense that only the probability of class membership can be determined for each individual.

interest in the development of world's new middle classes (e.g., Easterly, 2001; Milanovic and Yitzhaki, 2002; Banerjee and Duflo, 2008; Ravallion, 2010). On the importance of a consolidation of a middle class in Kazakhstan, see, among others, Verme (2000), Daly (March 2008), and OECD (2016).

In what follows, Section 2 reviews the previous attempts at combining objective and subjective measures of middle class and details our method to estimate the boundaries that define the middle-income class. Section 3 exemplifies our approach for estimating the middle income boundaries in Kazakhstan. It evaluates how people's perceptions are related to their income, it derives the middle class income thresholds, and it shows that income is the most important determinant in shaping self-perceived middle class. It also reports the evolution in size and income concentration of the middle class in the country. Section 4 sets out some concluding remarks.

2. Calibrating Middle-Income Class Boundaries on People's Perceptions

2.1. *What Do People's Perceptions Tell Us About Middle Class?*

There is a long tradition in public opinion surveys of accepting the idea that people can be asked direct questions as a way of reporting middle class placement (Bird and Newport, 2017), even though people may misperceive their objective position. Biased perception of people's position in the income ladder has been widely documented (Cruces *et al*., 2013), but also class self-perception may suffer from potential bias, and it is not clear in which direction the bias may occur. An over-perception of being middle class has been documented in Evans and Kelley (2004), using a large data set including 21 countries. This "middle class bias" has been partially explained by the way people form their perceptions: they look at their reference group and not at the whole population, and most people are objectively toward the middle of their reference groups. This evidence has been confirmed, on average, by OECD (2019) using more recent data and a different set of countries. Interestingly, some Eastern European countries have shown instead an under-perception that has been related to a "lingering effect" (Curtis, 2013). Compared to people in established democracies, people in former socialist societies were found much more likely to identify themselves as lower class and less likely to identify themselves as higher class, regardless of the level of economic development (and income inequality) reached by the country.

The identification of an objective social class entails a direct determination of a person's social class based on socioeconomic variables, as income, wealth, education, and occupation. This approach usually considers income as the most powerful determinant, and household budget surveys constitute the main data source for fixing the income boundaries and for estimating the size of the middle class. Survey answers on middle class self-placement are often correlated with socioeconomic and demographic variables like family heritage and background, prestige of the area where people live but also strongly with income. The relationship has been documented to be in the expected direction, with respondents living in higher-income households ranking themselves higher in terms of class (Ricci, 2016). Consequently, information coming from subjective approaches can be considered a useful instrument to shed light in the estimation of income boundaries.

The two approaches, objective and subjective, diverge in measurement methodology and remain far apart (Ravallion and Lokshin, 2002). Several studies document the mismatch of the results coming from these two perspectives (e.g., Lora and Fajardo, 2013; Sosnaud *et al.*, 2013), but less attention is given to possible concordant schema between objective and subjective measures of middle class. Self-perceived class questions are rarely included in households budget surveys but, when included, lower- and upper-income thresholds could be identified combining the information coming from the two data sources,[2] mixing objective and subjective methodologies.

Without abandoning the use of income estimates that are proper standard objective measures, is there any chance to combine quantitative data with subjective assessments? We propose a model-based method that identifies the middle-income boundaries, linking the self-declared status to income using a latent variable formulation. Our general idea grounds on the Leyden school of subjective poverty measurement (Van Praag, 1971) and looks for the lowest-income level around which more people regard themselves as middle class than as poor or lower class and, analogously, for the highest-income level that best separates the self-declared middle class from the upper class. Data-driven approaches have been proposed for this purpose. A naïve method is to map directly the share of population self-placed in each class into the income distribution (Cashell, 2007). For example, the dividing line between lower and middle class is estimated as the percentile in the income distribution corresponding to the cumulated share of population that consider themselves to be poor or lower class. An "inspection" method (Ferreira *et al.*, 2013) is based on the graphical representation of two income density functions, one representing those who declare to be lower class, and the other one those who consider themselves as middle. The lower bound of the middle class is defined as the income value at which the two functions cross. Similarly, the upper bound is defined as the income level at which the income density of the self-declared middle intersects the density of the self-declared upper class. In the following, we detail our methodology along with an exemplifying empirical analysis.

2.2. *Model Specification*

An ordered logistic (probit) model is the natural framework to relate an ordered categorical outcome to a set of predictors. Consider a categorical outcome y, observed on n individuals, $i = 1, \cdots, n$, with K natural ordered alternatives assuming values $1, \ldots, k, \ldots, K$. There is no assumption of equidistance between alternatives or categories. What is only required is that the response categories have a natural ordering, but no cardinal comparison is required. To address the ordering, we focus on the cumulative logits:

$$\log\left(\frac{\text{Prob}(y \leq k)}{\text{Prob}(y > k)}\right) = \log\left(\frac{\pi_1 + \cdots + \pi_k}{\pi_{k+1} + \cdots + \pi_K},\right) \tag{1}$$

[2]Statistical matching could be helpful to match the surveys whenever the information is not available in budget surveys data.

where $\pi_1, \cdots, \pi_K$ denote the response probabilities, satisfying $\sum_{k=1}^{K} \pi_k = 1$. All K categories have nonzero probability of occurring. The proportional odds model assumes that each predictor exerts the same effect on each cumulative logit regardless of k:

$$\log\left(\frac{\text{Prob}(y \le k)}{\text{Prob}(y > k)}\right) = c_k + \mathbf{X}'\boldsymbol{\beta}. \tag{2}$$

Each cumulative logit has its own intercept, but each predictor only has a single coefficient β (see Agresti, 2013). The intercepts c_k represent the cut-off points that need to be estimated.

Estimation of Prob $(y = k|\mathbf{X}'\boldsymbol{\beta})$ for the first category ($k = 1$) is:

$$\text{Prob}\left(y=1|\mathbf{X}'\boldsymbol{\beta}\right) = \text{Prob}\left(y \le 1|\mathbf{X}'\boldsymbol{\beta}\right) = \frac{\exp(c_k + \mathbf{X}'\boldsymbol{\beta})}{1+\exp(c_k + \mathbf{X}'\boldsymbol{\beta})}. \tag{3}$$

Estimating the probabilities for other categories requires a bit more work:

$$\text{Prob}\left(y=k|\mathbf{X}'\boldsymbol{\beta}\right) = \text{Prob}\left(y \le k|\mathbf{X}'\boldsymbol{\beta}\right) - \text{Prob}\left(y < k-1|\mathbf{X}'\boldsymbol{\beta}\right). \tag{4}$$

The corresponding latent formulation of the proportional odds model lies in explaining an underlying unobserved continuous random variable z while we observe only y, which takes value k ($k = 1, \cdots, K$) whether or not z is between two thresholds: $y_i = k$ if $c_{k-1} < z_i \le c_k$. As z crosses a series of increasing unknown thresholds, we move up the ordering of categories coming up to an ordered logistic model. The unrestricted formulation of the latent model assumes the form:

$$y_i = \begin{cases} 1 & \text{if } \; z_i \le c_1 \\ 2 & \text{if } \; z_i \in (c_1, c_2] \\ \cdots & \\ k & \text{if } \; z_i \in (c_{k-1}, c_k] \\ \cdots & \\ K-1 & \text{if } \; z_i \in (c_{K-2}, c_{K-1}] \\ K & \text{if } \; z_i > c_{K-1}, \end{cases} \tag{5}$$

where

$$z_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i, \;\; \epsilon_i \sim \text{N}(0, \sigma^2).$$

The latent variable z depends on a regressor $n \times M$ matrix $\mathbf{X}$ and an M parameter vector $\boldsymbol{\beta}$ plus the error term.

Based on the model (5) and the following equation (4):

$$\begin{aligned}\text{Prob}\{y_i = k|\mathbf{x}_i'\} &= \text{Prob}\{c_{k-1} < \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i \leq c_k\}\\ &= \text{Prob}\{c_{k-1} - \mathbf{x}_i'\boldsymbol{\beta} < \epsilon_i \leq c_k - \mathbf{x}_i'\boldsymbol{\beta}\}\\ &= \Phi(c_k - \mathbf{x}_i'\boldsymbol{\beta}) - \Phi(c_{k-1} - \mathbf{x}_i'\boldsymbol{\beta}),\end{aligned}$$

where Φ is the cumulative density function of the error term.[3]

This K-choice ordered model has $M + K - 1$ parameters, where M denotes the number of regressors excluding the intercept, plus the scale parameter of the error term, all of them to be estimated. In the model with a complete set of parameter estimates, the issue of unidentifiable parameters is well known. Different combinations of the parameters have in fact the same implications on the observed data y. Thus, model (5) has an essential indeterminacy, and it is standard to impose a restriction on the variance of the error term so that the parameters can be uniquely identified. This restriction imposes the variance equal to a fixed value, usually one (see, among others, Cameron and Trivedi, 2005).

The variance restriction (and the exclusion of the intercept) yields the standard specification, as researchers are mainly interested in the marginal effects in the probabilities due to the regressors. In fact the sign of the parameters $\boldsymbol{\beta}$ can be easily interpreted as determining whether or not the latent variable increases with the regressors. However, this is not the only possible specification, because there are at least two other specifications of the same model, related to each other by simple linear transformations (Gelman and Hill, 2007). Here we are interested in estimating thresholds on the income scale, which are interpretable as income boundaries, and in estimating the variance of the errors, as a measure of the idiosyncratic heterogeneity of the self-declared status. This requires a slightly different parametrization of model (5):

$$y_i = \begin{cases} 1 & \text{if} \quad z_i \leq c_{1.5} \\ 2 & \text{if} \quad z_i \in (c_{1.5}, c_{2.5}] \\ 3 & \text{if} \quad z_i \in (c_{2.5}, c_{3.5}] \\ 4 & \text{if} \quad z_i > (c_{3.5}), \end{cases} \tag{6}$$

where

$$z_i = x_i + \epsilon_i, \;\; \epsilon_i \sim \text{N}(0, \sigma^2).$$

The values 1, 2, 3, and 4 correspond to the alternatives of the self-declared status, say Poor, Vulnerable, Middle class, and Rich. In this parametrization, $c_{1.5}$, $c_{2.5}$, and $c_{3.5}$ can be directly interpreted as thresholds on the scale of the input x, here income, and the standard deviation of the error terms σ as the gradualness of the transition from Poor to Vulnerable, from Vulnerable to Middle class, and from Middle to Rich, as illustrated in Figure 1. Estimation of model (6) is presented in Section 3.1.

[3]Different assumptions on the distribution of the error term (logistic or Gaussian) lead to different outcome models (logistic or probit). The logistic model uses the cumulative distribution function of the logistic distribution, while the probit model uses the cumulative distribution function of the standard normal distribution. Both methods yield similar, although not identical, inferences. Our model is estimated under the logistic specification. See also Section 3.1.

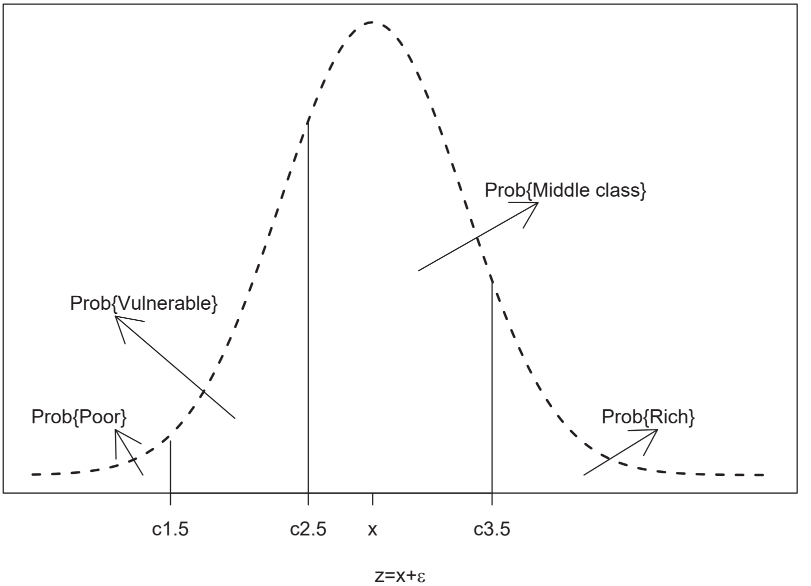

Figure 1. Illustration of cut-off points in Model (6), with $K = 4$
Notes: The graph shows the distribution of the latent outcome z corresponding to a given value of income x. The shape of the distribution depends on the estimated σ. The cutpoints $c_{1.5}$, $c_{2.5}$, and $c_{3.5}$ are the boundaries of the categories and are on the same scale of x.

3. Income Boundaries Estimation

3.1. *Kazakhstan Data and Estimation*

The Household Budget Survey of Kazakhstan (KHBS), conducted periodically by the Statistics Agency of the Republic of Kazakhstan, represents a valuable source of information on household incomes and expenditures. The survey covers about 12,000 households, which are interviewed on a quarterly basis. The sample is designed to be representative at the regional level. Households in the survey are associated with sampling population weights. We consider per capita household income because it is the welfare aggregate for official poverty in Kazakhstan. Household income is the total of the income of each member from all sources. Household incomes are annualized and regionally deflated on the basis of the unit values of food reported in the survey to capture geographical cost-of-living differences.

In 2013, a *Quality of life of the population* module was introduced in the survey, and respondents were asked to self-declare their social status. The subjective module of the questionnaire asked heads of household to place themselves in ordered classes: poor; not poor, but not middle class; middle class; top middle class, and rich.[4] The questionnaire explicitly refers to a class between the poor and the middle (not poor not middle), which we renamed vulnerable. This term has been

[4]As very few households identified themselves as "rich," the two categories were merged in a single category labeled "prosperous."

TABLE 1
Responses to the Question on Self-Perceived Social Status, 2013

What Social Group Would You Refer the Household You Head to?	Responses (%)
Poor	3.7
Not poor, but not middle class (vulnerable)	49.8
Middle class	44.3
Prosperous: Top-middle class and rich	2.2

TABLE 2
Estimated Parameters of the Proportional Odds Logistic Model in Its Standard Parametrization

Parameters	Estimates	SE	*t*-Value
β_{income}	0.00330	0.00010	33.181
Intercepts			
$c_{1\|2}$	1.56550	0.04803	32.592
$c_{2\|3}$	5.85225	0.09737	60.102

used to identify people who are not poor but have a relatively high probability of becoming poor in the future (for a comprehensive survey, see Ceriani, 2018).

Table 1 shows that relatively few respondents placed themselves into the poor class (3.7 percent). The majority declared to belong to the vulnerable group (49.8 percent), followed by those who assigned themselves to the middle group (44.3 percent). Very few self-rated themselves as prosperous (2.2 percent). It is interesting to note the share of those who self-allocated as poor in 2013 is between the poverty head count ratio according to the national official poverty line (2.9 percent), and the poverty head count ratio at $5.50 per day in 2011 purchasing power parity (PPP) (4.4 percent).[5]

To reduce uncertainty in estimating the thresholds, we incorporate into the second-class respondents who identify themselves as poor, ending up with three social classes—poor/vulnerable, middle, and prosperous—and two thresholds, those for the lower and upper bounds of the middle class. In our analysis the ordered logistic model, under the proportional odds assumption, is estimated using the function `stan_polr` (`polr` stands for proportional odds logistic regression) in the library `rstanarm` in R (Carpenter *et al.*, 2017; Stan Development Team, 2020). Estimation is performed in a Bayesian framework via Monte Carlo Markov Chains (MCMC) rather than maximum likelihood estimation. Uncertainty is summarized using simulations from the entire posterior distribution. Estimation of the standard parametrization of the proportional odds logistic model is presented in Table 2.

From the table, it is clear that the fitted model has the traditional parametrization in which the income coefficient β and two cutpoints are estimated, and the independent errors are set as $\epsilon_i \sim$ logistic $(x_i, 1)$.[6] To be able to transform the estimated thresholds

[5]The Statistical Agency of Kazakhstan uses an absolute poverty line to arrive at the official estimates of income poverty in the country. The international poverty lines of US$5.50 per day is the World Bank poverty line for upper middle-income countries such as Kazakhstan. A more recent approach introduced by the World Bank tends to capture also the relative aspects of poverty, introducing a societal poverty line (SPL). The SPL can be considered an hybrid line because it combines the US$1.90 a day absolute line with a relative component (Jolliffe and Prydz, 2021).

[6]The unit logistic distribution is essentially equivalent to a Gaussian distribution with standard deviation equal to 1.6.

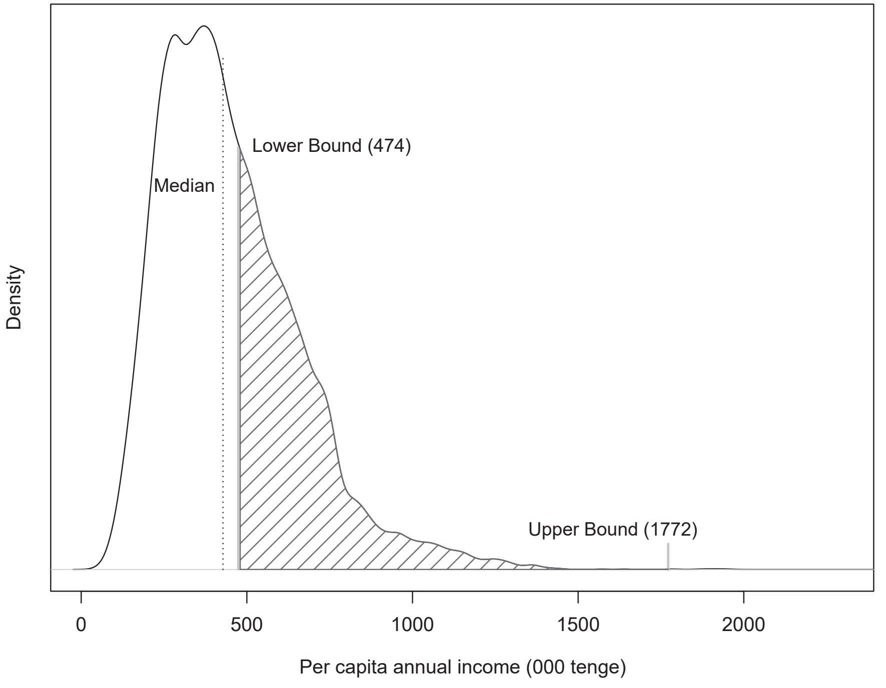

Figure 2. Estimated Income Distribution in Kazakhstan and the Size of the Middle Class (Dashed Area), Year 2013
Notes: Lower and upper bounds are the estimated cutpoints of model (6).

in the same scale of income and to estimate the fuzziness parameter σ, we need to re-parametrize the estimated intercepts shown in Table 2 in terms of model (6), as follows:

$$\hat{c}_{1.5} = \frac{\hat{c}_{1|2}}{\left|\hat{\beta}\right|};\hat{c}_{2.5} = \frac{\hat{c}_{2|3}}{\left|\hat{\beta}\right|};\hat{\sigma} = \frac{1}{\left|\hat{\beta},\right|}$$

being $\hat{\beta}$, $\hat{c}_{1|2}$, and $\hat{c}_{2|3}$ the estimates of the standard parametrization of Table 2.

As evident in Table 2, the uncertainty of the estimated intercept $c_{2|3}$ is larger than the estimated $c_{1|2}$ leading to a less accurate estimation of the upper threshold.[7] Availability of subjective data in different years would help in assessing the robustness of the estimated thresholds, especially the upper bound.

This re-parametrization yields a lower income threshold of 474,000 tenge and an upper of 1,772,000 tenge, which correspond to the 56th and the 99th percentiles of the weighted income distribution.[8] The corresponding estimated size of the mid-

[7]See also the discussion in Ferreira *et al*. (2013) about the emphasis on the upper income threshold. As a matter of fact, using a probit specification instead of a logistic specification, the estimated lower bound does not significantly change, while the point estimate of the upper threshold decreases by 129,000 tenge. The size of the middle class, however, remains almost the same.

[8]In the presence of middle class bias, our estimated thresholds can be considered as lower and upper bounds of the "true" boundaries. If instead the bias is in the opposite direction ("lingering"), our estimated thresholds can be considered as more conservative than the "true" boundaries.

dle class is equal to 43.5 percent, which is slightly different from what people declared (see Table 1). Figure 2 shows the income distribution of Kazakhstan households, where a kernel density estimator has been used to estimate its shape, using the Sheather-Jones criterion to select the bandwidth. As evident from the figure, the lower bound of the middle-income class is greater than the median income. The presence of a large vulnerable group prevents households with median income—technically those in the middle of the income distribution—from being middle class.

Although country-specific and expressed in local currency, our estimated values can be compared with other thresholds after converting them in PPP. The values of 474,000 and 1,772,000 tenge correspond to US$14.0 and US$52.2 per capita per day at 2011 international dollars, respectively.[9]

Milanovic and Yitzhaki (2002), with the aim of defining a global middle class, set the lower threshold equal to the mean earnings in Brazil ($12 per capita per day) and choose the upper bound at $50 per day, which is equal to the average income in Italy, the least wealthy country among the G7 members. Within the "vulnerability-to-poverty approach," using longitudinal data in their study on Latin American and the Caribbean (LAC) region, Ferreira *et al*. (2013) and López-Calva and Ortiz-Juarez (2014) established a range of per capita income of US$10–50 a day at 2005 PPP terms to define the middle class in those countries.[10] The same methodology applied to Nigeria has provided an estimate of the lower threshold at US$3.9 in 2011 PPP-adjusted USD (Corral Rodas *et al*., 2019). Updating measures from 2005 PPPs to 2011 PPPs using the domestic inflation rate over the period 2005–2011 (Ferreira *et al*., 2016), it is interesting to note that the threshold that divides the vulnerable from the middle group in Kazakhstan is similar to the corresponding threshold established for LAC countries (roughly US$13.0 in 2011 PPP terms) and much higher than the lower bound estimated for Nigeria. It is also worth mentioning that applying the inspection approach of Ferreira *et al*. (2013) to our data, that is, crossing the income distributions of the self-perceived vulnerable and the self-perceived middle class estimated by kernel densities, yields a lower bound of 442,000 tenge, corresponding to US$13.0 per capita per day in 2011 terms.

Figure 3 shows the estimated lower and upper thresholds and the expected social status as a function of income $\mathbb{E}(y|x)$, along with the incomes of the respondents by self-reported social class. Although overlapping of the income distributions by self-declared classes is present, there is clear evidence of a decreasing pattern of the expected status whenever income increases.

There is discrepancy between subjective class identity and objective (income-based) class position. This discordance can be explained by other factors beyond income that might influence those who identify themselves as middle class (Ravallion and Lokshin, 2002). These factors, as sharing values regarding their

[9]Using the consumption price index for 2013 with the base in 2011 as deflator, annual values in tenge at current prices have been converted to 2011 international dollars using the 2011 PPP conversion factor for private consumption (World Bank, International Comparison Program database), and then divided by 365 to get per day values.

[10]Households were defined economically stable or not vulnerable if their probability of falling below their national poverty line on a 5-year horizon was less than 10 percent. If this probability was above 10 percent, households were considered vulnerable to poverty. The predicted income associated with the 10 percent probability was defined as the lower bound of the middle class.

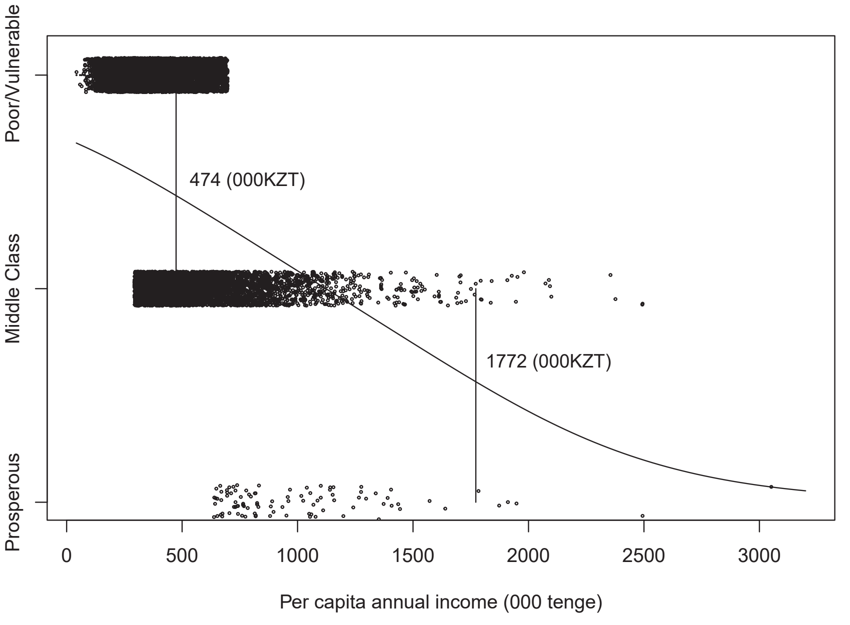

Figure 3. Expected Social Status as a Function of Income

Notes: Vertical lines show estimated thresholds, and curve shows expected responses as estimated by the ordered logistic model (6). The *dots* show the incomes of respondents in each declared social class, jittered for a clearer representation.

own personal futures, access to education, common living standard in general, attitudes, and behavior, can be either observed or unobserved. In the next section, we empirically show that after controlling for potential observable determinants, we can still rely on income as a proxy for class.

3.2. *Other Significant Factors Beyond Income?*

In an attempt to understand the main sources of the discordance between self-reported social status and income, in this section we empirically evaluate whether—after controlling for the main observable determinants widely accepted in the literature—income plays a dominant role. The overview of the potential drivers is not exhaustive, as there might be contributing factors outside the KHBS available data.[11]

Figure 4 summarizes the estimations of a logistic model that explains the subjective membership to the middle class in terms of per capita household income and other observable characteristics. These predictors are used as control variables, and they include neighborhood characteristics, housing, health conditions, access to services and ownership of durable goods, education, type of household, and demographic characteristics of the household head. We also control for region of residence.[12]

[11]The survey does not include any monetary measure of financial and real wealth.

[12]To better separate the characteristics of the middle class, we excluded from the model those who declared themselves prosperous. Regional coefficients are omitted in the figure for sake of space. The estimated geographical effects can be attributed to perceptions of relative welfare within the local community.

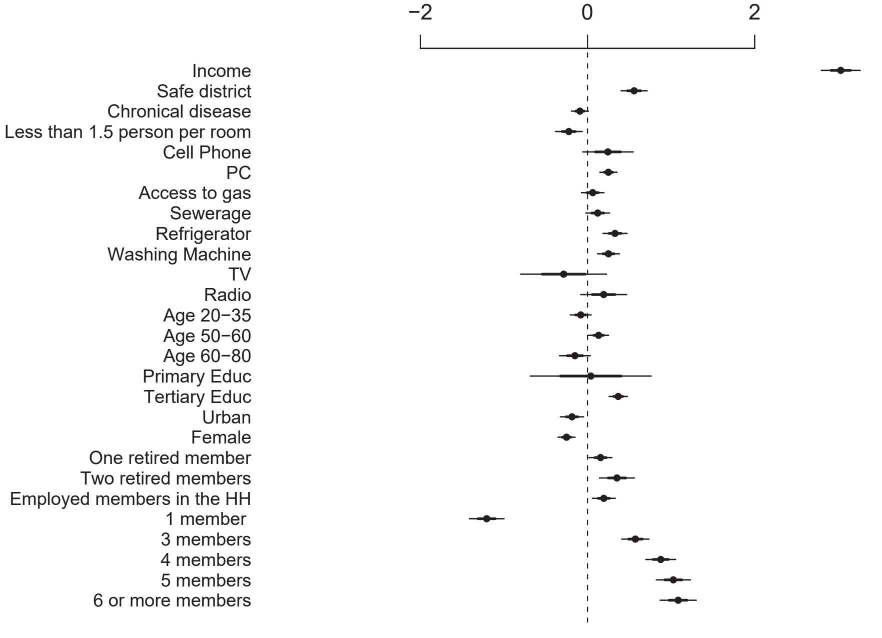

Figure 4. Estimated Coefficients ± 2 Standard Errors of Determinants of Self-Perceived Middle Class

Income is mean-centered and scaled by two times its standard deviation, so that the resulting coefficient can be interpreted like those of binary predictors, and therefore the relative importance of each predictor can be precisely evaluated (Gelman, 2008). As shown in the figure, although the relevance of other determinants is not negligible, per capita income clearly stands out among the predictors, with a positive effect on the perception of being middle class. If per capita income has been the only objective measure to reflect peoples' perceptions, the coefficients on the control variables in the logistic model would have been not statistically significant. As expected, living in a safe area and in a comfortable house (less than 1.5 person per room), having access to a connected PC, and having a refrigerator or a washing machine at home increase the probability that the respondents perceive themselves as middle class. The differences by age groups are minimal. For education, respondents with the highest education feel somewhat more middle class than what their income would suggest. The presence of chronic health problems in the family does not seem to influence the perceived status, *ceteris paribus*. Respondents living in households receiving the bulk of their income from pensions (one or two pensions received by members of the family) tend to feel better off than their income implies. Being a woman reduces the probabilities to self-declaring to be middle class.

The most substantial influence, other than income, on the self-assessment of middle class is due to household size. Other things being equal, as the number of components increases, the likelihood of the respondent to identify herself as part of the middle class also increases. This is probably due to the presence of economies of scale that large families benefit from, which is not captured when income is measured in per capita term. As a robustness check, we estimated the model with equivalent income using the square root of the household size as equivalence scale. The coefficients associated with household sizes are much lower, but still significant. Therefore, the presence of economies of scale only partially explains the empirical evidence. Actually, the difference between two-person households and more-than-two person households is not very relevant (roughly, at the most 5 percent differences in probability of feeling middle class). It is living alone that reduces the perception of belonging to the middle class. A one-person household has 11 percent less probability than a two-person household, other things being equal. At any rate, equivalent income is still the most important predictor in shaping middle class self-perception. These findings are similar to the analysis of Carletto and Zezza (2006) conducted to evaluate the gap between subjective poverty and objective welfare measures in Albania.

In nuce, the existence of "confounding" non-income factors in survey-based measures of subjective middle class is a concern (Ravallion, 2014). These variables certainly enrich the understanding of the discrepancy between subjective and objective measures. However, built on our findings, income is the variable that matters the most, and if one variable is to be picked, for sake of comparisons and implementation, that should be income.

3.3. *On the Evolution of the Middle-Income Class in Kazakhstan, 2003–2015*

Once estimated the income boundaries of the middle class for the year in which the subjective module was available, they are updated backward and forward using the consumer price index. Consistent with these estimated boundaries, it was possible to trace the evolution of the middle class in Kazakhstan between 2003 and 2015.

The consistent growth of the middle class parallels that of the per capita GDP and the rapid reduction in poverty experienced by the country after the period of economic turmoil immediately following independence (see Figure 5).

Was economic growth the only source of the middle-class expansion, as it seems from the graph? To answer this question, we decompose the changes in the size of the middle class into a "location" effect and a "shape" effect, using a nonparametric version of the methodology based on Datt and Ravaillon's (1992) decomposition. The method is presented in Massari *et al*. (2009).

The location effect measures the change in the middle-class share that can be attributed to balanced growth, corresponding to an equal relative increase of each household income. This growth, equal to an increase in the location parameter of the distribution (the mean or the median income), generates a shift in the income density in a distributionally neutral fashion. The shape effect refers to the change in the middle class size attributable to changes in the income curve holding the mean/median constant. In other words, this is the change that would have occurred if only the observed change in the shape of the income distribution had occurred without any shift in the mean/median of the curve. Therefore, this effect can be attributed to redistribution without growth. The last three columns of Table 3 show the decomposition of the change of the middle-class share (Δ share) into the growth effect and the redistribution effect over different sub-periods, using the initial year of each sub-period as the reference year.

TABLE 3
Evolution of the Kazakh Middle Class: Size, Shares of Total Income, 2003–2015, Percent

Year	Middle Class Size	Share of Income	Δ Share Size	Growth Effect	Redistribution Effect
2003	3.7	12.1	—	—	—
2006	15.4	30.9	11.7	12.8	−1.1
2010	31.2	50.3	15.8	14.0	1.8
2013	43.5	61.7	12.3	12.0	0.3
2014	44.2	62.1	0.7	1.1	−0.4
2015	44.4	62.7	0.2	−0.2	0.4

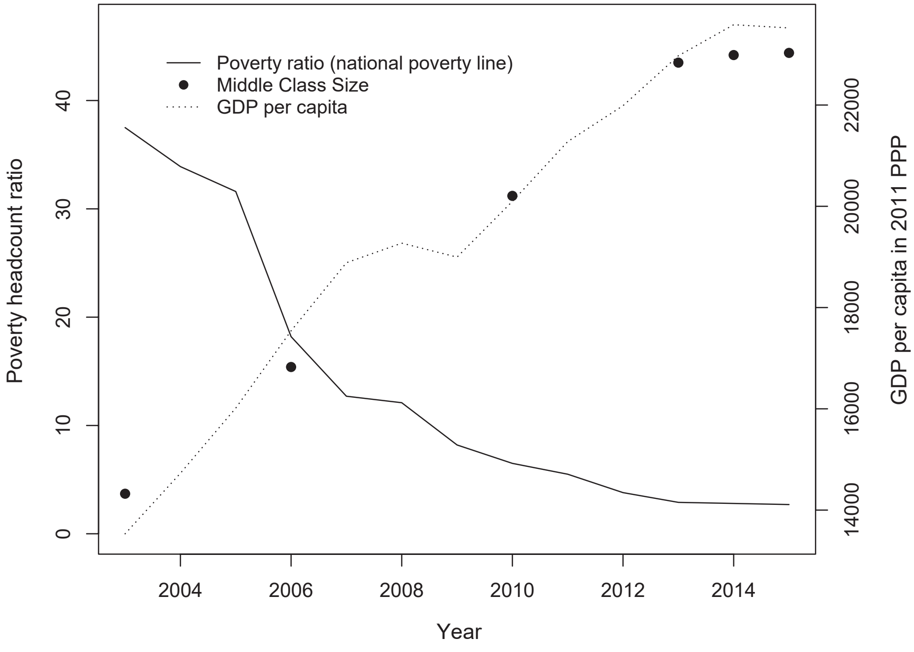

Figure 5. Official National Poverty Ratio, GDP Per Capita, and Middle Class Size Between 2003 and 2015

Overall, the growth component played the prominent role in increasing the share of the middle class, especially in the first decade of the 2000s. The redistribution component marginally contributed to increase the middle class between 2006 and 2013, but it had a negative effect between 2003 and 2006 and in 2013–2014. Redistribution was responsible for the slight increase in the middle group in the years 2014–2015 that would otherwise have reduced due to the growth effect.

4. Conclusions

What constitutes the middle class is hotly debated. Different concepts result in non-concordant measurements. Even adopting an income-based approach, the middle class varies according to different definitions of "statistical middle." We proposed a novel schema to estimate the income boundaries of the middle class driven by subjective data corresponding to self-declared class position. The boundaries are estimated by re-parameterizing the multinomial ordered logistic model in which the response variable is people's subjective class identity. This approach is grounded on the evidence of a strong association—although not perfect—between class perceptions and objective class positions driven by income. The discrepancy is due to the association between middle class perception and many factors, observable or unobservable, in line with previous research. Our results highlighted the specially important role of income after controlling for the most significant observable factors, leaving to individual unobservable traits the main cause of the divergence between subjective class identity and objective class position. The method has been exemplified in an application to Kazakhstan in which we estimated middle class boundaries using the 2013 household budget survey, in which self-perception questions were introduced, and so both subjective and objective data in the same survey were available. The estimated lower and upper boundaries, converted in 2011 PPP international prices, are about \$14 and \$52, respectively, corresponding to the 56th and 99th percentile of the per capita income distribution. Interestingly, the lower bound of the middle-income class is greater than the median income. The presence of a large vulnerable group prevents individuals with median income—technically those in the middle of the income distribution—from being middle class. The Kazakh middle class has increased massively in size and in income concentration. Using the deflated values of the income boundaries for all other years, between 2003 and 2010 the middle class expanded from 3.7 to 31.2 percent of the population. Thereafter it continued to grow substantially, reaching 43.5 percent by 2013 and being relatively stable in the subsequent 2 years. The increase in the size of the middle class is essentially due to a growth effect.

To give solidity to our results, it would have been useful to replicate the model on different years, but data availability was limited to 1 year. Comparisons over time and across countries with similar data are left for future research.

References

Agresti, A., *Categorical Data Analysis*. 3rd ed. John Wiley & Sons, New York, 2013.

Anderson, G. J., A. Farcomeni, M. G. Pittau, and R. Zelli, "A New Approach to Measuring and Studying the Characteristics of Class Membership: The Progress of Poverty, Inequality and Polarization of Income Classes in Urban China," *Journal of Econometrics*, 191, 348–59, 2016.

Atkinson, A. B., and A. Brandolini, "On the Identification of the Middle Class." in J. C. Gornick and M. Jantti (eds), *Income Inequality*. Stanford University Press, 77–100, 2013.

Banerjee, A. V., and E. Duflo, "What is Middle Class About the Middle Classes Around the World?," *Journal of Economic Perspectives*, 22, 3–28, 2008.

Bird, R., and F. Newport, "What Determines How Americans Perceive Their Social Class?" 2017. https://news.gallup.com/opinion/polling-matters/204497/determines-americans-perceive-social-class.aspx

Cameron, A. C., and P. K. Trivedi, "Microeconometrics," *Theory and Applications*. Cambridge University Press, Cambridge, 2005.

Carletto, G., and A. Zezza, "Being Poor, Feeling Poorer: Combining Objective and Subjective Measures of Welfare in Albania," *Journal of Development Studies*, 42, 739–60, 2006.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, "Stan: A Probabilistic Programming Language," *Journal of Statistical Software*, 76, 1–32, 2017.

Cashell, B. W., "Who are the 'Middle Class?'" (RS22627). Congressional Research Service, Washington, DC, 2007.
Ceriani, L., "Vulnerability To Poverty: Empirical Findings," in C. D'Ambrosio (ed), *Handbook of Research on Economic and Social Well-being*. Edward Elgar Publishing, 284–99, 2018.
Corral Rodas, P. A., V. Molini, and G. Oseni, "No Condition is Permanent: Middle Class in Nigeria in the Last Decade," *Journal of Development Studies*, 55, 294–310, 2019.
Cruces, G., R. Perez-Truglia, and M. Tetaz, "Biased Perception of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment," *Journal of Public Economics*, 98, 100–12, 2013.
Curtis, J., "Middle Class Identity in the Modern World: How Politics and Economics Matter," *Canadian Review of Sociology/Revue canadienne de sociologie*, 50, 203–26, 2013.
Daly, J. C. K., "Kazakhstan's Emerging Middle Class," *Silk Road Paper*, Central Asia-Caucasus Institute, March 2008.
Datt, G., and M. Ravaillon, "Growth and Redistribution Components of Changes in Poverty Measures," *Journal of Development Economics*, 8, 275–95, 1992.
Easterly, W., "The Middle Class Consensus and Economic Development," *Journal of Economic Growth*, 6, 317–35, 2001.
Evans, M. D. R., and J. Kelley, "Subjective Social Location: Data from 21 Nations," *International Journal of Public Opinion Research*, 16, 3–38, 2004.
Ferreira, F. H. G., J. Messina, J. Rigolini, L. F. López-Calva, M. A. Lugo, and R. Vakis, *Economic Mobility and the Rise of the Latin American Middle Class*. World Bank, Washington, DC, 2013.
Ferreira, F. H. G., S. Chen, A. Dabalen, Y. Dikhanov, N. Hamadeh, D. Jolliffe, A. Narayan, E. B. Prydz, A. Revenga, P. Sangraula, U. Serajuddin, and N. Yoshida, "A Global Count of the Extreme Poor in 2012: Data Issues, Methodology and Initial Results," *Journal of Economic Inequality*, 14, 141–72, 2016.
Gelman, A., and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, 2007.
Gelman, A., "Scaling Regression Inputs by Dividing by Two Standard Deviations," *Statistics in Medicine*, 27, 2865–73, 2008.
Jolliffe, D., and E. B. Prydz, "Societal Poverty: A Relative and Relevant Measure," *World Bank Economic Review*, 35, 180–206, 2021.
López-Calva, L. F., and E. Ortiz-Juarez, "A Vulnerability Approach to the Definition of the Middle Class," *Journal of Economic Inequality*, 12, 23–47, 2014.
Lora, E., and J. Fajardo, "Latin American Middle Classes: The Distance Between Perception and Reality," *Economia - Journal of the Latin American and Caribbean Economic Association*, 14, 35–54, 2013.
Massari, R., M. G. Pittau, and R. Zelli, "A Dwindling Middle Class? Italian Evidence in the 2000's," *Journal of Economic Inequality*, 7, 333–50, 2009.
Milanovic, B., and S. Yitzhaki, "Decomposing, World Income Distribution: Does the World have a Middle Class?" *Review of Income and Wealth*, 48, 155–78, 2002.
OECD, "Multi-Dimensional Review of Kazakhstan Volume 1," *Initial Assessment*. OECD Development Pathways, OECD Publishing, Paris, 2016.
______, *Under Pressure: The Squeezed Middle Class*. OECD Publishing, Paris, 2019.
Pew Research Center, "America's Shrinking Middle Class: A Close Look at Changes Within Metropolitan Areas," *PRC*, May 2016.
Ravallion, M., and M. Lokshin, "Self-Rated Economic Welfare in Russia," *European Economic Review*, 46, 1453–73, 2002.
Ravallion, M., "The Developing World's Bulging (but Vulnerable) Middle Class," *World Development*, 38, 445–54, 2010.
______, "Poor, or Just Feeling Poor? On Using Subjective Data in Measuring Poverty." in A. E. Clark and C. Senik (eds), *Happiness and Economic Growth: Lessons from Developing Countries*. 2014. Oxford University Press, coll. "Studies of Policy Reform", Oxford.
Ricci, C. A., "Perceived Social Position and Objective Inequality: Do they Move Together? Evidence from Europe and the United States," *Italian Economic Journal*, 2, 281–303, 2016.
Sosnaud, B., D. Brady, and S. M. Frenk, "Class in Name Only: Subjective Class Identity, Objective Class Position, and Vote Choice in American Presidential Elections," *Social Problems*, 60, 81–99, 2013.
Stan Development Team, *RStan: The R Interface to Stan*, R package version 2.21.2 2020, http://mc-stan.org/
van Praag, B. M. S., "The Welfare Function of Income in Belgium: An Empirical Investigation," *European Economic Review*, 2, 337–69, 1971.
Verme, P., "The Choice of the Working Sector in Transition: Income and Non-Income Determinants of Sector Participation in Kazakhstan," *Economics of Transition*, 8, 691–731, 2000.