

2019

IARIW-World Bank

Special IARIW-World Bank Conference “New Approaches to Defining and Measuring Poverty
in a Growing World” Washington, DC, November 7-8, 2019

Measuring the Middle Class in Kazakhstan: A Subjective Approach

Maria Grazia Pittau

Roberto Zelli

Paper Prepared for the IARIW-World Bank Conference

Washington, DC, November 7-8, 2019

Measuring the Middle Class in Kazakhstan

A Subjective Approach

M. Grazia Pittau

Roberto Zelli



WORLD BANK GROUP

Poverty and Equity Global Practice

August 2018

Abstract

This paper proposes a model-based approach to estimate income boundaries for identifying the middle class in Kazakhstan over 2003–15. The approach exploits the subjective evaluation of Kazakhstan households about their social status, relating self-declared social class membership to income. Income data come from the Kazakhstan Household Budget Survey, which also includes a specific module on quality of life and perceived social status. As social status is intrinsically an ordinal response, the paper

estimates a proportional odds model with income as the key explanatory variable. Although other factors influence the self-perception of being in the middle class, income is by far the most important determinant. Benchmarking on 2013, the estimated middle class lower bound is \$14 at 2011 purchasing power parity and the upper bound is \$52. The Kazakhstan middle class has increased massively in size and income concentration. The increase is essentially due to a growth effect rather than a redistributive cause.

This paper is a product of the Poverty and Equity Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/research>. The authors may be contacted at grazia.pittau@uniroma1.it and roberto.zelli@uniroma1.it.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Measuring the Middle Class in Kazakhstan: A Subjective Approach

M. Grazia Pittau and Roberto Zelli^{1 2}

JEL: D31; C25; I30.

Keywords: Middle class; Income distribution; Kazakhstan.

¹ Affiliated with Sapienza University and the World Bank Poverty group.

² This paper draws on the work of the Kazakhstan middle class analysis project led by Sarosh Sattar, task team leader. We are grateful to Nicholas T. Longford for invaluable help, especially in data processing. We would like to thank Francisco Ferreira and Maria Ana Lugo for their useful comments and suggestions. We also thank Sarosh Sattar, Luis Felipe López-Calva, William Hutchins Seitz and participants at the ECAPOV HBSS seminar in April 2017 at the World Bank, Washington DC. A special thank to Aibek Baibagysh Uulu for providing assistance and clarification when needed.

1. Introduction

Kazakhstan has achieved rapid poverty reduction since independence. Based on the international poverty line of \$1.90 a day at 2011 international prices, headcount poverty fell from 10.5 percent in 2001 to 0.04 percent in 2013.³ However, the rapidity of the change has generated questions about what “poverty” and “middle class” mean.

An emerging middle class is critical because of its potential as an engine of growth (Easterly 2001). Historically, those in the ‘middle’ have vigorously accumulated both physical and human capital (OECD 2011a, 2011b). By driving consumption and domestic demand, consolidating Kazakhstan’s incipient middle-income group into a stable middle class would contribute to a solid foundation for economic progress. The expansion of the middle classes has also contributed to democratic movements and progressive but moderate political reforms, especially those that promote inclusive growth.

While most economists agree that middle class status is characterized by a relatively high income, there is no consensus on where to draw the line, since living above the poverty line does not necessarily ensure middle-class status.

The very concept of middle class has been extensively debated in the economic and sociological literature (for a review, see Atkinson and Brandolini 2013). Sociologists tend to identify the middle class in terms of its functional position in the society. Economists, however, tend to identify it in terms of thresholds in the income or consumption distribution. These thresholds can be either relative (e.g., percentiles or percentage of median income) or absolute (a certain amount of income per day or month).

In advanced economies, where the middle class usually comprises households in the middle of the income distribution, economic thresholds are identified using percentiles or 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 because median incomes do not necessarily identify middle class

³ World Development Indicators, World Bank (<https://datacatalog.worldbank.org/dataset/world-development-indicators>).

households in terms of living standards.

Recently, a “vulnerable” class, located between the poor and the middle class, has been defined (López-Calva and Ortiz-Juarez 2014). Vulnerability relates to the risk of falling into poverty over time due to adverse events. For instance, a vulnerable household may not have sufficient resources to continue to acquire all the necessities to maintain an adequate living standard over a reasonably long period. Even in countries like Kazakhstan where extreme poverty is very low, the presence of a large vulnerable group may prevent households with median income—those in the middle of the income distribution—from becoming middle class. Assessing vulnerability, Ferreira et al. (2013) used per capita income of US\$10 to US\$50 a day in 2005 PPP terms to define the middle class in Latin America and the Caribbean.

Can the US\$10–US\$50 range be exported and applied directly to Kazakhstan? While the range is solidly grounded for Latin America, it may not be appropriate elsewhere. For instance, the same method applied to Nigeria estimates the lower threshold at US\$3 per capita a day (Corral, Molini, and Siwatu 2015).

For this reason, we have devised an approach to estimate country-specific thresholds for identifying the Kazakhstan middle class. Our method, which is based on an absolute income approach, estimates lower and upper bounds based on how Kazakhstan citizens perceive their own status relative to their reported income. One module of the 2013 Kazakhstan Household Survey (see Section 2) dealt directly with self-declared social class. Our analysis focuses on 2013 because the module appeared only in that round of the survey, but we have extended the analysis to 2014 and 2015 and have also looked back to 2003, 2006, and 2010 to better describe the profile of the middle class and analyze possible changes in the Kazakhstan middle class over a longer period.

In what follows, Section 2 discusses the Household Budget Survey of Kazakhstan. Section 3 details the econometrics method for estimating the lower and upper income bounds that define the middle class. Section 4 highlights the characteristics of the middle class, and Section 5 explores factors beyond income that could influence the self-perception of being middle class. After controlling for these factors, income remains the most important determinant of middle-class

status, giving support to the income-based approach. Section 6 reports the evolution in size and in income shares of the middle class between 2003 and 2015. Section 7 sets out the conclusions.

2. Data Issues

This analysis is based on data from the Household Budget Survey of Kazakhstan (HBS) conducted periodically between 2003 and 2015 by the Statistics Agency of the Republic of Kazakhstan (www.stat.gov.kz) and harmonized by the World Bank Poverty and Equity Group for Europe and Central Asia (ECA). Survey samples were generally 12,000 households, except in 2010, when the sample was doubled to 24,000. Households are currently interviewed quarterly about sources of income and amounts received, but the evolution of the sampling design introduced inconsistencies into our analysis. For example, in earlier surveys a household provided data for only one quarter, so annual income is extrapolated by multiplying household income by four. Later, each household was contacted twice, in either quarters 1 and 2, or 3 and 4.

The households in the samples are associated with sampling weights. Sampling designs are summarized in Table 1. The surveys are stratified using a geo-administrative division of the country into 16 regions further divided into their urban and rural areas (see Table 2). The rural-urban subsample sizes, which were similar in the surveys for 2013 through 2015, roughly reflect the distribution of population in the region. Because the cities of Astana and Almaty are entirely urban areas, there are 30 rather than 32 strata. Within a stratum there are at most two distinct values of the sampling weight.

Table 1: KHBS Survey Sampling Designs

Year	Households	Strata (Regions)	Quarters per household
2003	12,000		4
2006	12,000	30(16)	1
2010	24,000	30(16)	2
2013	12,000	30(16)	4
2014	12,000	30(16)	4

2015	12,000	30(16)	4
------	--------	--------	---

A small number of data entries are missing. Because we are concerned mainly with household income, Table 3 summarizes the income data gaps related to households that are represented in only some quarters. Of course, a household may move out of the domain or the country or be dissolved by death, so some of the missing data cannot be regarded as nonresponse. In any case, because there are so few gaps, we make no provisions to address this issue. Similarly, we cannot establish whether some changes in the composition of the households are a result of a genuine change or nonresponse.

Table 2: The Regions

Region	Sample Sizes (2013 – 15)		Region	Sample Sizes	
	Rural	Urban		Rural	Urban
Akmola	480	360	Karaganda	360	600
Aktobe	480	360	Kostanay	450	360
Almaty	480	240	Kyzylorda	360	240
Almaty City	—	898	Mangystau	270	330
Astana City	—	660	North Kazakhstan	390	270
Atyrau	240	300	Pavlodar	480	360
East Kazakhstan	480	420	South Kazakhstan	480	300
Jambyl	420	270	West Kazakhstan	420	240

Table 3: Missing Household Income Data

	Households with data for				All
	1 quarter	2 quarters	3 quarters	4 quarters	
2003	90	160	96	11 645	11 991
2006	12 000	—	—	—	12 000
2010	183	23 816	—	—	23 999
2013	45	107	49	11 799	12 000
2014	54	81	54	11 809	11 998
2015	123	93	56	11 713	11 985

Note: Gap counts are in italics.

The income of a household consists of the total of the income of each

member from all sources. Sources are listed in the survey questionnaire. In all analyses, household incomes are annualized by aggregating the quarterly data and are in per capita terms. Incomes have been regionally deflated on the basis of the unit values of food reported in the survey.

3. Identifying the Middle Class in Kazakhstan

3.1 The Absolute Income-based Approach

The absolute approach identifies the middle class as those households with income or consumption within a specific range, fixing a lower threshold of particular income (e.g., \$10 per household member) and a corresponding upper threshold (e.g., \$50). Thresholds are usually set in purchasing power parity (PPP) terms for the sake of international comparison.

Since the welfare aggregate for official poverty in Kazakhstan is income (World Bank, Country Poverty Brief, October 2017), we focus on income aggregate.⁴ The fundamental question is naturally how to set the thresholds; to some extent, the choices for the middle-class range are arbitrary. Banerjee and Duflo (2008) acknowledged use of an ad hoc range of values considering two groups of households: the lower middle class whose daily per capita expenditures valued at PPP are between \$2 and \$4, and the upper middle class between \$6 and \$10. Ravallion (2010) sets the lower bound of a “developing world middle class” at the global poverty line (\$2 per day) and the higher bound (\$13) at the U.S. poverty line. Milanovic and Yitzhaki (2002), in defining the global middle class, set the lower threshold equal to the mean earnings in Brazil (\$12 per day) and choose the upper bound of \$50 per day—the average income in Italy, the least wealthy G7 member.

Given the rate of absolute poverty in Kazakhstan, it is not appropriate to categorize anyone who is not poor as middle class, as Banerjee and Duflo (2008) and Ravallion (2010) did for a group of developing countries. It is more appropriate to follow the vulnerability approach, which defines the middle class as those who are not vulnerable to falling into poverty in a few years’ time, as López-Calva and

⁴ On average, however, income and consumption are in balance. In 2013 the log-balances, defined as the logarithms of the ratios of consumption and income, have a mean of -0.0113 , corresponding to consumption that is 1.13% lower than income, and standard deviation of 0.335. Details on the income and consumption distributions are available upon request.

Ortiz-Juarez (2014) and Ferreira et al. (2013) have done for Latin America and the Caribbean. López-Calva and Ortiz-Juarez (2014) set the lower bound of middle-class income in Latin America as the maximum income that ensures economic stability. Households were defined as economically stable and invulnerable only if the probability of their falling below the national poverty line within five years was 10 percent or less. The income level associated to the 10 percent probability was defined as the lower bound of the middle class.

Using longitudinal data in their study of Latin America and the Caribbean, Ferreira et al. (2013) established as the lower threshold of the middle class the value of US\$10 a day in 2005 PPP terms. This estimation was corroborated by a subjective approach in which respondents were asked to report their social class; the majority of respondents with that income identified themselves as middle class rather than poor. The upper bound was established as \$50 (2005 PPP international dollars) as a reasonable level that excludes the richest 2% of Latin American households.⁵ Though the US\$50 threshold is fairly arbitrary, it has since gained ‘authority’ by being widely adopted.

Although these fixed thresholds are in PPP terms, only to a limited extent can they be used to define the middle class in Kazakhstan because they have been attributed to countries at different stages of social and economic development. Country-specific thresholds are therefore preferable. However, in Kazakhstan the method proposed by López-Calva and Ortiz-Juarez (2014) cannot be applied directly for two reasons: (1) it would yield an extremely low probability of falling into poverty because of the very low poverty rate in Kazakhstan, and (2) household longitudinal data for at least four years are not available. In the next section, therefore, we explore a new methodology for estimating country-specific thresholds for identifying the middle class in Kazakhstan.

3.2 Estimation of the thresholds

Our approach to estimating absolute thresholds for identifying the Kazakhstan middle class exploits individual perception of social status and relates self-declared social class to income. We follow a vulnerability-poverty approach by identifying a

⁵ The upper threshold lies in the 92nd percentile for Chile, the 97th for Mexico and the 98th for Peru.

vulnerable group between the poor and the middle class. Income thresholds are estimated using the *Quality of life of the population* module in the 2013 survey in which respondents were asked to report their social class. Once the thresholds were estimated for 2013, we calibrated the corresponding lower and upper bounds for the years for which this information was not available.

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 Rich⁶ (see Table 4).

Table 4: Responses to the Question on Self-Perceived Social Status, 2013

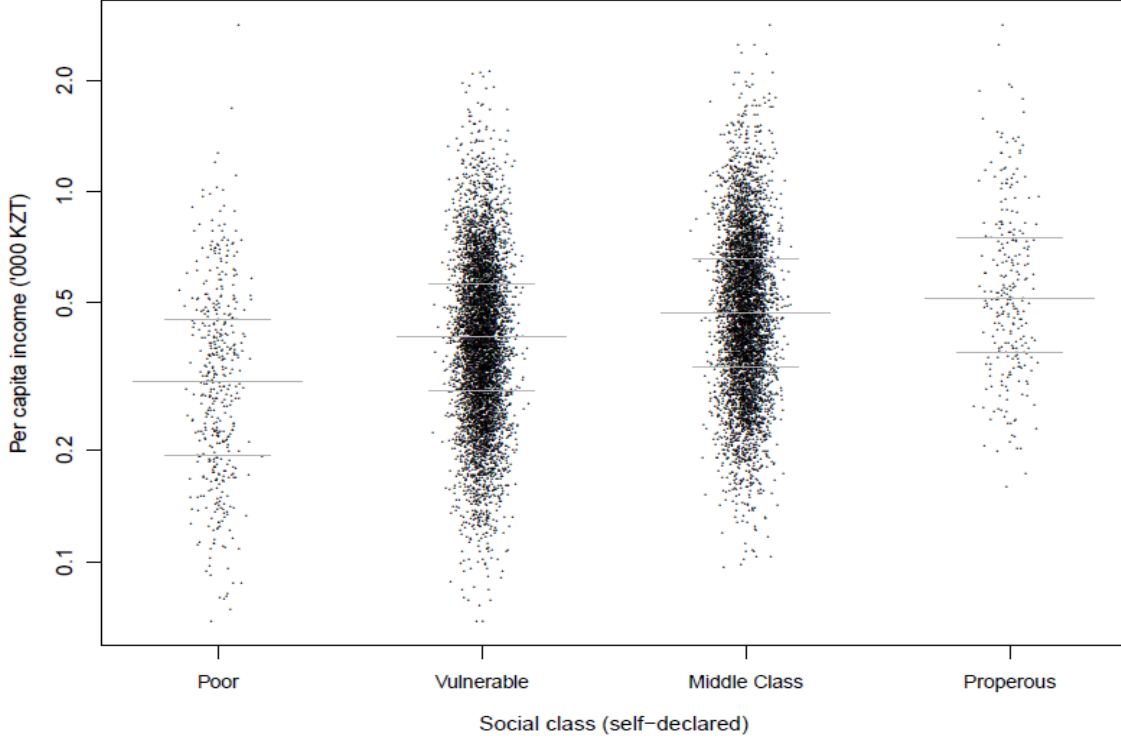
What social group would you refer the household you head to?	Responses (%)
Low-income (do not have enough funds for food, clothes and footwear)	3.7
Not poor, but not middle-class (enough funds to buy food-products, clothes and footwear, pay for housing and utility services, however we encounter difficulties with purchase of durable goods)	49.8
Middle class (tier, level): we do not encounter any difficulties with purchase of food-products, main non-food products and services, but have insufficient funds to purchase additional dwelling (apartments, houses, summer house), an expensive car and etc.	44.3
Top-middle class (tier, level): we consume high-quality products, live in comfortable conditions, have work, our own income-generating business and/or property, but we do not have enough free time for recreation	2.2
Prosperous (rich): have enough resources (knowledge, health, finance, property, time) for comfortable life	

Figure 1 shows correspondences between self-declared social class and per capita income as recorded in the survey Household Income and Expenditure Module. Not surprisingly, relatively few respondents (5.9%) placed their households in the extreme classes 1 (Poor, 3.7%) and 4/5 (Prosperous, 2.2%). The majority declared their households “Not poor but not middle class”—i.e., vulnerable—(49.8%), followed by those who assigned themselves to the middle group (44.3%). Although

⁶ Since few identified their households as “Top middle class” or “Rich,” ECA-POV merged them into a single “Prosperous” category.

there is substantial overlap of the income distribution by social classes, the median pattern shows a clear upward trend, ensuring confidence in the validity of using income in identifying social classes.

Figure 1: Income and Self-reported Social Status



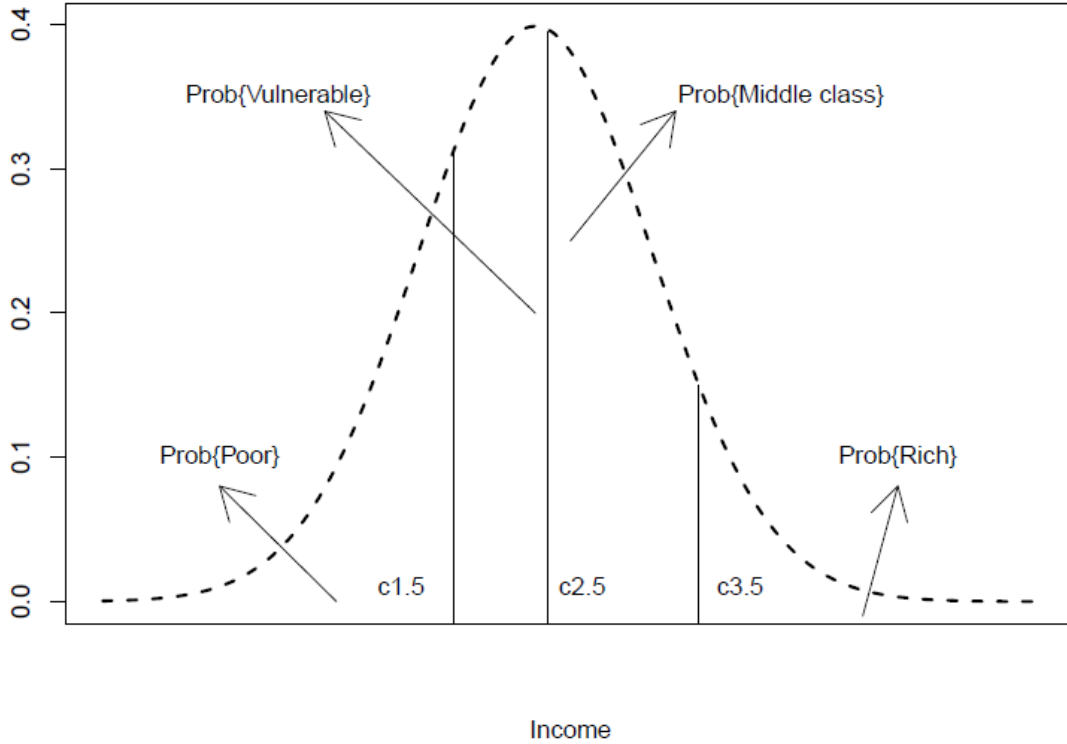
Note: Each household is represented by a dot. The medians are marked by longer horizontal bars, and the quartiles by shorter bars.

Since self-declared status can be considered an ordered variable, both lower and upper bounds are estimated by fitting an ordered logistic regression model. We estimate the model considering as logistic case the *proportional odds logistic regression* (Agresti 2013). The ordinal logistic model can be briefly illustrated as follows: Perceived social status is a categorical outcome (y) that can take the values 1, 2, 3, or 4, corresponding to Poor, Vulnerable, Middle, and Prosperous. Each category of y is associated with a continuous unobserved (latent) outcome, z , generally defined as a linear combination of different predictors with independent errors e_i that follow the logistic or Gaussian distribution. Consequently:

$$y_i = \begin{cases} \text{poor,} & \text{if } z_i < c_{1.5} \\ \text{vulnerable,} & \text{if } z_i \in (c_{1.5}, c_{2.5}) \\ \text{middle,} & \text{if } z_i \in (c_{2.5}, c_{3.5}) \\ \text{prosperous,} & \text{if } z_i \geq c_{3.5} \end{cases} \quad (1)$$

In our analysis we opted for the logistic distribution of the error term, and therefore $z_i = x_i + e_i \sim \text{logistic}(x_i, \sigma)$, where x is per capita income, e represents measurement errors, and σ can be interpreted as a fuzziness parameter.⁷ Figure 2 illustrates the ordered categorical model and shows how the distance between any two adjacent cutpoints $c_{1.5}, \dots, c_{3.5}$ affects the probability that $y = 1, 2, 3$, or 4.

Figure 2: Cutpoints in an Ordered Categorical Logistic Model



Note: In this example, $K = 4$ categories and the cutpoints are $c_{1.5}$, $c_{2.5}$ and $c_{3.5}$. The figure illustrates the distribution of the latent outcome, z , corresponding to a given value of the linear predictor, $X\beta$. For each, the cutpoints show where the outcome y will equal Poor, Vulnerable, Middle class, and Prosperous.

⁷ The logistic distribution looks very much like a normal distribution. We could always approximate a logistic distribution with unit variance with the Gaussian density with variance $\sigma^2 = \pi^2/3$. Therefore, the choice between logistic or Gaussian distribution for the error terms affects the estimates only by a constant equal to $\pi/\sqrt{3}$.

Due to identification issues, the most widespread statistical software programs estimate a “traditional” parametrization of model (1) in which $z_i = \beta x_i + e_i$ and $e_i \sim \text{logistic}(x_i, 1)$. However, estimating all the thresholds in the same scale of income requires a slightly different parameterization (Gelman and Hill 2007). Using this parameterization, we can directly interpret the cutpoints $c_{1.5}$, $c_{2.5}$, and $c_{3.5}$ of model (1) as thresholds on the same scale of income and the standard deviation of the error terms σ as the gradual transition from one class to the next.

The estimated cut-points (intercepts) make it possible to estimate the probability of being Poor, Vulnerable, Middle Class, and Prosperous. For example:

$$\text{Prob}\{y_i = \text{poor} | x_i\} = \text{Prob}\{x_i + e_i \leq c_{1.5}\} = \Phi(c_{1.5} - x_i),$$

where $\Phi(w) = \frac{e^w}{1+e^w}$ is the logistic distribution function.⁸

To reduce uncertainty in estimating the thresholds we incorporate into the second class the 441 respondents (3.7%) who identify themselves as poor, ending up with three social classes—Poor/Vulnerable, Middle Class, and Prosperous—and two thresholds, those for the lower and upper bounds of the middle class, to be estimated. Estimation of model (1) in 2013 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, and a size of 43.5 percent.⁹ These values correspond to US\$14.00 and US\$52.20 per day in 2011 international dollars. 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

⁸ Depending on the assumption on the distribution of the error terms, logistic or Gaussian, we estimate a logistic or a probit regression model. Logit and probit differ in how the function that links the categorical outcome and the linear predictor is defined. 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.

⁹ We fit the ordered logistic model using the `bayespolr` (bayesian “proportional odds logistic regression”) function, which is part of the `ARM` package in R (Gelman et al., 2016). For robust estimation the model was fitted with trimming: we trimmed the top and the bottom 2% for each social class. Estimation of the probit model leads instead to a lower threshold equal to 474,900 tenge and an upper threshold of 1,639,100 tenge corresponding to a size of the middle class equal to 43.2 percent.

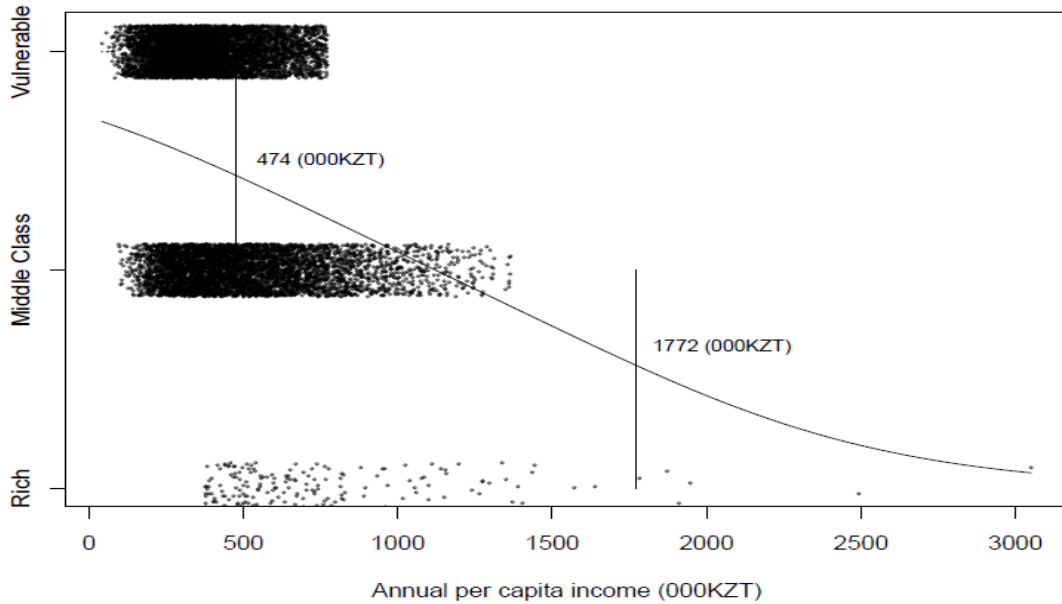
365 to get per day values.

Figure 3 shows the estimated lower and upper thresholds and the expected social status as a function of income (x), defined as:

$$\begin{aligned}
 E(y|x) &= 1 \cdot \Pr(y = 1|x) + 2 \cdot \Pr(y = 2|x) + 3 \cdot \Pr(y = 3|x) = \\
 &= 1 \cdot \left(1 - \text{logit}^{-1}\left(\frac{x - c_{1.5}}{\sigma}\right)\right) + 2 \cdot \left(\text{logit}^{-1}\left(\frac{x - c_{1.5}}{\sigma}\right) - \text{logit}^{-1}\left(\frac{x - c_{2.5}}{\sigma}\right)\right) + 3 \\
 &\quad \cdot \text{logit}^{-1}\left(\frac{x - c_{2.5}}{\sigma}\right)
 \end{aligned}$$

where $\text{logit}^{-1}(x)$ is the logistic distribution function $\Phi(\cdot)$.

Figure 3: Per Capita Income and Class Status, “Quality of Life” Respondents

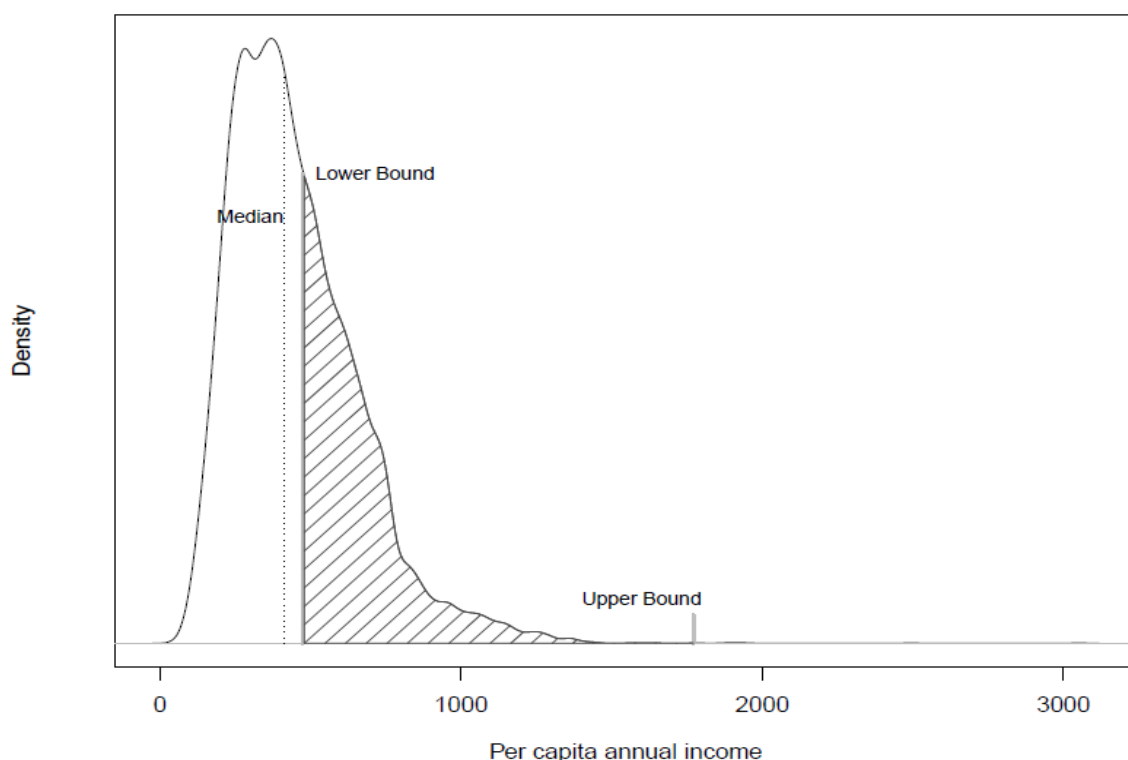


Note: Vertical lines show estimated thresholds and curve shows expected responses as estimated by a multinomial ordered logistic model. The dots are for incomes of respondents in each declared social class.

4. Characteristics of the Middle Class

In 2013, 43.5 percent of households fell into the income range that defines the middle class. Figure 4 shows the income distribution of Kazakhstan households. A kernel density estimator has been used to estimate its shape, using the Sheather-Jones criterion to select the bandwidth.

Figure 4: Middle Class Kazakh Households, 2013, Percent



Based on the data recorded about household members, their ages and employment status, the head of the household, and location, we define the following household types:¹⁰

- Urb Household in an urban area
- Chd1 One child
- Chd2 At least two children
- ChdM More children than adults
- Fem Female head
- Empo No household income from employment or self-employment
- Ret1 One retired person
- Ret2 At least two retired persons

¹⁰ Note that these types are not exclusive. For example, type Chd2 is subsumed in Chd1. The same kind of analysis can be done for other possible household types.

Table 5 shows the incidence in 2013 of different household types in the middle class and in the whole population. Whenever the value is greater than one, the household type is over-represented in the middle class; when less than one, it is under-represented. In 2013 households with retired members and those headed by a female were over-represented. Families with at least one child were underrepresented, and the presence of children in the family has a severely negative effect on the possibility of the household being in the middle class. For example, the percentage of households with more children than adults is 16 percent in the whole population but only 5.7 percent of middle-class families.

**Table 5: Distribution of Household Types,
Middle Class and Total Population**

Type	Description	Middle class	Total pop.	Ratio
		(1)	(2)	(1)/(2)
Ret1	At least one retired person in the HH	31.2	27.9	1.12
Ret2	At least two retired persons in the HH	7.3	6.3	1.16
Chd1	At least one child in the HH	51.9	70.0	0.74
Chd2	At least two children in the HH	18.0	37.5	0.48
ChdM	More children than adults in the HH	5.7	16.0	0.36
Fem	Female head of HH	56.0	50.7	1.11
Empo	No income from employment in the HH	20.3	21.3	0.95

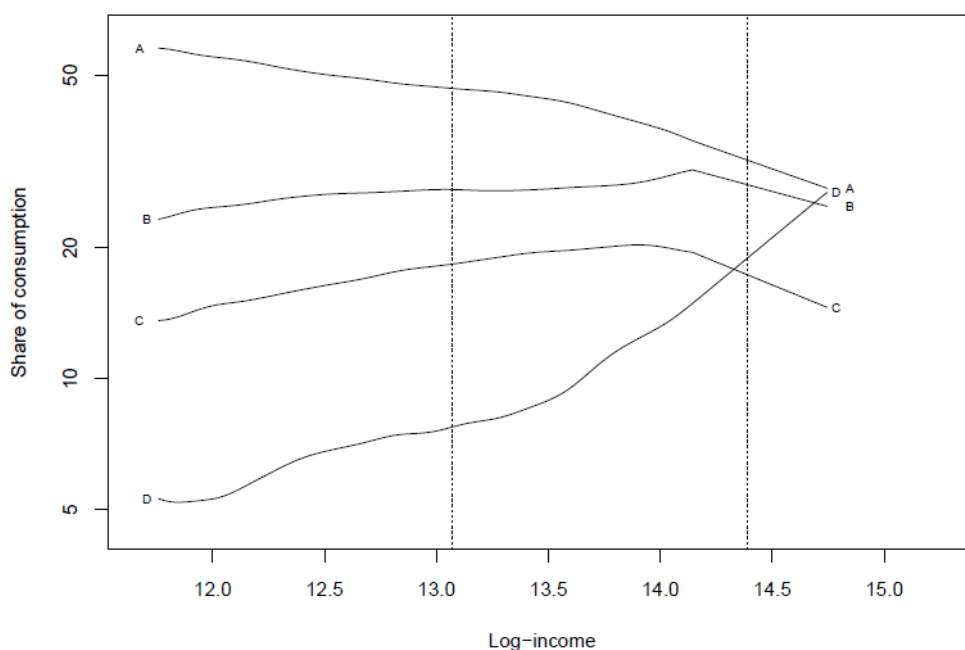
Here we examine patterns of consumption of Kazakh households at different levels of income, to evaluate whether the poor and vulnerable, the middle class, and the rich have different consumption patterns. Consumption is recorded in monetary amounts for 14 non-overlapping categories of goods and services. For convenience, we classify them into four groups (see Table 6): Food, Essentials, Optional items, and Luxuries. The classification is somewhat arbitrary, also because it is constrained by the data recorded. For example, the category ‘Travel’ includes everyday travel to work (essential), occasional shopping and social meeting in a nearby town (optional), and long-distance travel for holidays (luxury).

Table 6: Consumption Groups and Categories

Group		Category	
A	Food	a	Food
B	Essentials	c	clothing
		d	housing/utilities
		f	transport
		m	rent
C	Optional Items	b	alcohol and tobacco
		e	furnishing/hh equipment
		g	communication
		i	education
		k	miscellaneous
		l	health
D	Luxuries	h	recreation
		j	hotels/restaurants
		n	durables

Figure 5 relates the components of consumption to income. The smoothed fractions of the consumption (within total household consumption) are plotted against household per capita log-income. The vertical lines correspond to the lower and upper income thresholds estimated for the middle class. The plot shows that throughout the income range, food (A) is the dominant component, although the fraction spent on food decreases with prosperity (from 57 to 27 percent), consistently with the Engel's law. Within the middle-class thresholds, the food share goes from 46.6 to 35.3 percent. Throughout the range of income, the shares of necessities (B) and optional items (C) also increase, but more slowly. Within the middle-class range, the shares of B and C are almost stable, at around 27 and 20 percent, respectively. The share of luxuries (D) is very low until around 270,000 tenge and then rapidly increases. Within the middle-class income range, the share of luxuries varies from 7.7 to 14.9 percent.

Figure 5: Shares of consumption groups, per cent.



Note: The groups are: A. Food; B. Essentials; C. Optionals; D. Luxuries. The vertical dashes mark the (log) thresholds of the middle class.

5. Beyond Income: Other Factors?

From the economic standpoint, income is the best indicator of living standards, but there are determinants beyond income that can help to identify whether people belong in the middle class. Among these are ownership of, e.g., a home or car; access to amenities like energy supplies or the Internet; access to credit and to such services as schools, universities, and health care; type of employment; and other demographic characteristics. If the role of these can be determined, it is possible to evaluate what government policies help the middle class grow and keep it stable and sizable.

In this section, we explore the effect of both income and other determinants on the probability of belonging to the middle class, as defined by the self-perception of Kazakhstan citizens. The main goal is to evaluate whether, after controlling for other determinants, income is still essential. The dependent variable takes the value of 1 if respondents identify themselves as middle class, 0 otherwise. Explanatory variables include household per capita income, access to facilities, health status,

education, occupation, and some other demographic characteristics. These variables combine information from a thoughtful merger of responses to different sections of the 2013 HBS questionnaires.

We estimate different logistic regression models for three sets of predictors: income and access to facilities, other demographic characteristics, and region of residence. Table 7 shows estimated coefficients and corresponding standard errors for the three models. 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 (Gelman 2008) and the relative importance of each predictor can be evaluated.

Table 7: Estimated Likelihood of Being in the Middle Class, 2013, Logistic Regression Results

	Model A			Model B			Model C		
	Coeff.	s.e.		Coeff.	s.e.		Coeff.	s.e.	
Income	0,926	0,067	***	1,731	0,090	***	2,297	0,103	***
Safe district	0,850	0,068	***	0,765	0,071	***	0,501	0,075	***
Chronic disease	-0,215	0,041	***	-0,230	0,043	***	-0,054	0,048	
Access to gas	0,385	0,039	***	0,400	0,043	***	0,125	0,061	**
Land line or cell phone	0,311	0,125	**	0,244	0,129	*	0,451	0,138	***
Access to PC at home	0,345	0,039	***	0,144	0,044	***	0,247	0,048	***
Age: 20-35				-0,136	0,054	**	-0,078	0,059	
Age: 50-60				0,170	0,052	***	0,144	0,057	**
Age: 60-80				-0,054	0,080		-0,106	0,087	
Education: primary				0,115	0,272		0,063	0,325	
Education: tertiary				0,254	0,047	***	0,288	0,052	***
Urban				-0,208	0,046	***	-0,060	0,053	
Female				-0,275	0,042	***	-0,219	0,048	***
Ret1				0,168	0,060	***	0,155	0,065	**
Ret2				0,244	0,091	***	0,284	0,099	***
Empo				-0,182	0,060	***	-0,264	0,065	***
HH size: 1 component				-0,835	0,087	***	-0,980	0,095	***
HH size: 3 components				0,340	0,062	***	0,354	0,068	***
HH size: 4 components				0,518	0,066	***	0,524	0,073	***
HH size: 5 components				0,699	0,076	***	0,647	0,085	***
HH size: 6 or more components				0,734	0,078	***	0,600	0,089	***
Aktobe							2,710	0,140	***
Almaty							-0,494	0,132	***
Almaty city							0,154	0,126	
Astana city							0,051	0,129	
Atyrau							1,129	0,139	***
East Kaz							0,922	0,113	***
Jambyl							1,539	0,125	***
Karaganda							-0,860	0,125	***
Kostanay							-1,256	0,148	***
Kyzlorda							1,196	0,126	***
Mangystau							0,412	0,139	***
North Kaz							0,918	0,121	***
Pavlodar							1,624	0,114	***
South Kaz							1,282	0,123	***
West Kaz							0,739	0,130	***
(Intercept)	-1,474	0,141	***	-1,389	0,156	***	-1,093	0,114	***

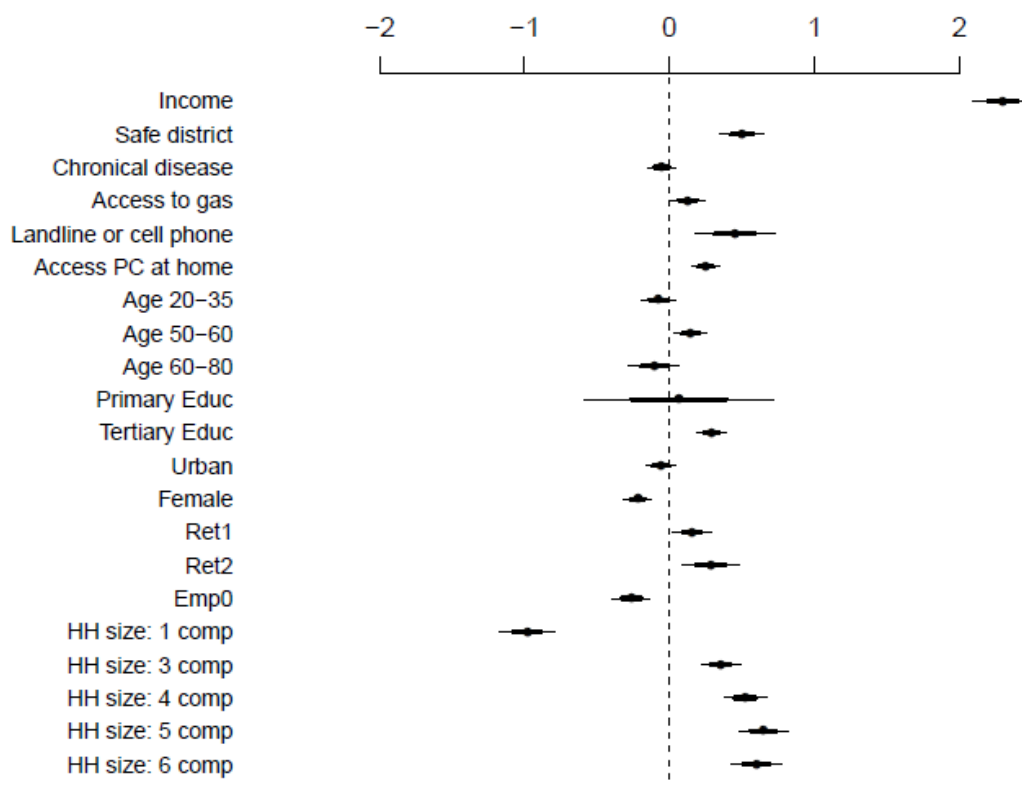
*** Significant at 0.1%; **significant at 1%; *significant at 5%.

In Model A, the probability of belonging to the middle class is estimated as a function of: household per capita income; safety of the district where respondents live (0 unsafe, 1 safe); presence of a person in the household with a chronic disease (proxy for access to health care); access to gas (1 yes, 0 no); access to communication (having a landline or at least one cell phone); and access to a PC at home (proxy for access to the Internet). Income is the most significant predictor in terms of size. Living in a safe area, having access to gas, communication and a PC increase the probability that the household is middle class, though the probability decreases when at least one member has chronic health problems.

As evident from Model B (Table 7), sociodemographic characteristics like a man being the household head, middle-aged, or having reached a tertiary education have a significant positive impact. Families receiving one or two pensions (Ret1 and Ret2) also increase the probability to belong to the middle class. Instead, living in an urban area, not having income from employment (Empo), and the household head being a woman reduce the probabilities. After controlling for these characteristics, income is by far the strongest predictor.

In Model C we finally control also for the region of residence, since the distribution of opportunities can be geographically unequal. As expected, there is a strong regional effect: living in Almaty, for example, reduces the probability of being in the middle class and Akmola (the baseline) raises it; residents of Jambyl or Pavlodar are more likely to consider themselves middle class than residents of Akmola. After controlling for place of residence, income is still clearly the factor that does most to shape the middle class in Kazakhstan. Figure 6 shows the estimated coefficients of Model C and their standard errors.

Figure 6: Estimated Coefficients (± 2 Standard Errors) of Kazakh Household Characteristics, 2013, Model C.



Note: Dependent variable: Self-declared status. Regional coefficients have been omitted for space considerations.

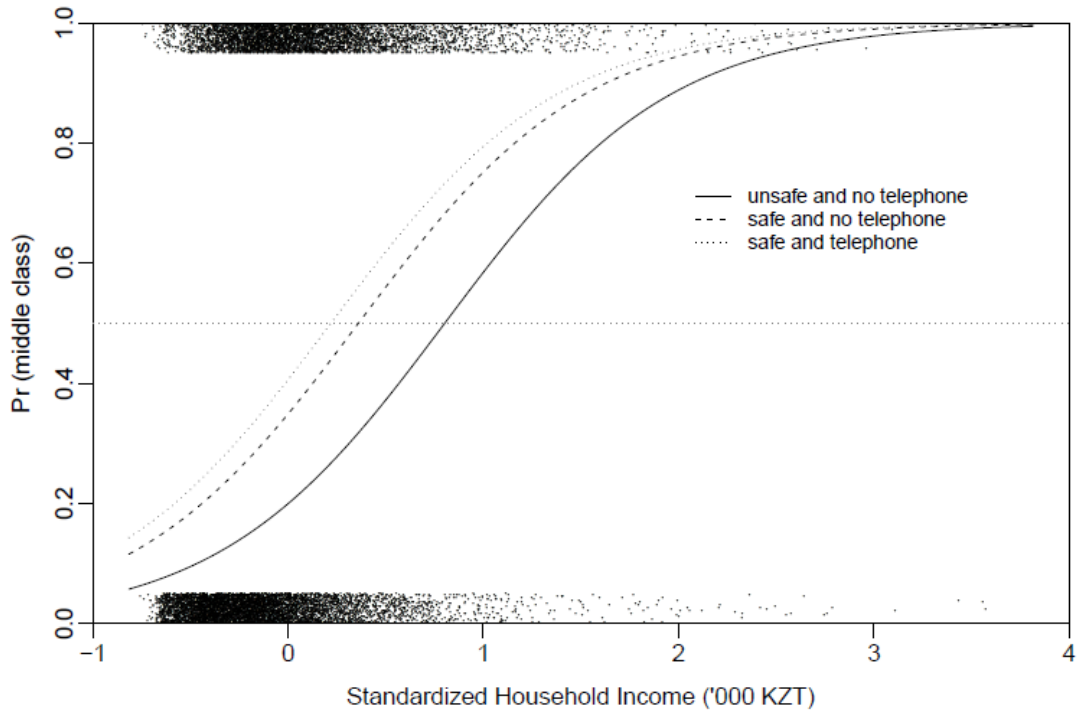
Household size plays an important role in shaping self-assessment of middle class. Other things being equal, as the number of components increases, the likelihood of the household to identify itself as part of the middle class also increases. This is due to the presence of economies of scale in consumption that large families benefit from.

Based on the coefficient estimates, we can evaluate how much change in *per capita* income is needed by households of various sizes to have the same probability to be in the middle class. For example, a household composed by a single person needs 220,000 tenge more than the per capita income of a two-component household with otherwise similar characteristics. Instead, a three-component household needs 80,000 tenge per capita less than the per capita income of a household with two components to have the same probability to belong to the middle class. This evidence suggests that other possible schemes of adjusting for household composition could take into account economies of scale and eventually

distinguish between children and adult members of the household.

Figure 7 gives examples of the predicted probability of being in the middle class as a function of income. The curves represent, *ceteris paribus*, the probabilities to belong to middle class for households that live in a safe or unsafe district; and have or have not access to communication (landline or cell phone). Assuming perfect compensability, the income compensation required to keep the same probability of belonging to the middle class for a family living in an unsafe area is 113,000 tenge with respect to a family living in a safe district.

Figure 7: Household Income and Probability of Being in the Middle Class



Note: Jittered data are overlain. Curves refer to whether the district of residence is safe or unsafe and whether the household has a fixed line or at least one cell phone. For each curve, other variable inputs are held constant at their baseline values.

6. Evolution of the Middle Class, 2003–15

Based on the estimated thresholds from Model (1) in 2013 (see Table 7 and section 3), we estimated upper and lower boundaries for the middle class for the other years surveyed. The boundaries of each year t have been calculated by multiplying the 2013 boundaries by the CPI of year t fixing 2013 as base. The resulting boundaries

in 2011 international dollars are constructed to be all the same.

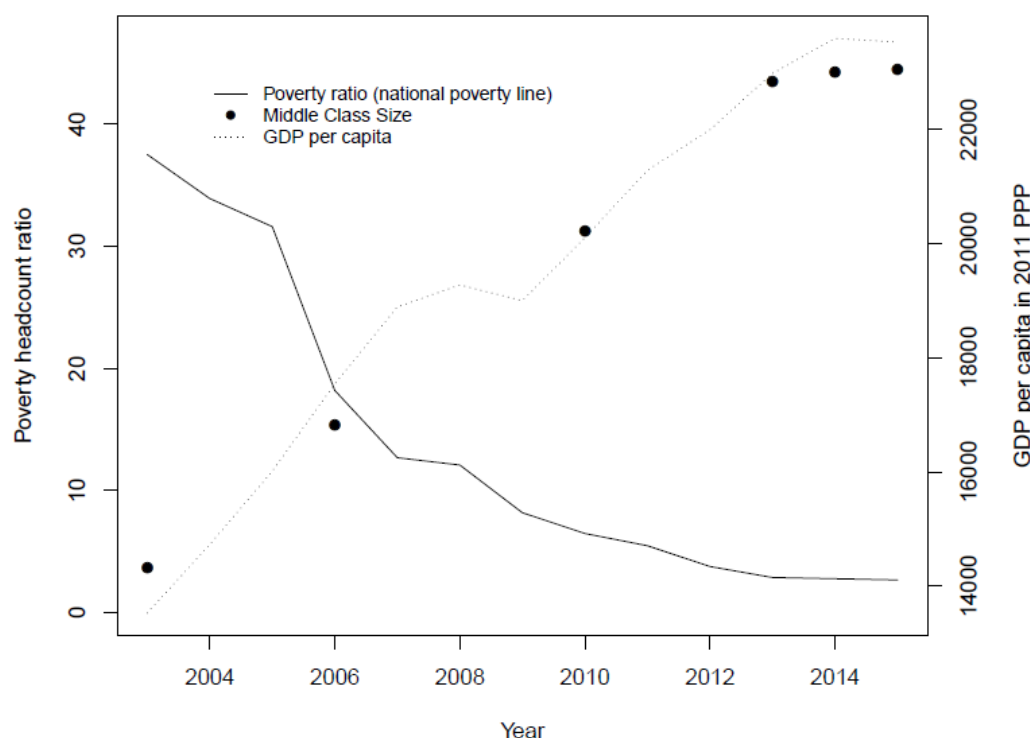
Table 8: Estimated Middle Class Thresholds, 2003–15

Year	Lower Bound			Upper Bound	
	Annual Per Capita Income, '000 KZT	Income Per Day, 2011 US\$ PPP		Annual Per Capita Income '000 KZT	Income Per Day, 2011 US\$ PPP
2003	211.1	14.0		789.4	52.2
2006	263.6	14.0		985.6	52.2
2010	393.2	14.0		1,470.1	52.2
2013	474.0	14.0		1,772.0	52.2
2014	505.8	14.0		1,891.0	52.2
2015	539.5	14.0		2,016.7	52.2

Based on the estimated thresholds, Table 9 reports how many households have qualified within the intervals since 2003, along with the share of overall income the middle class received. The results confirm that since 2003 the middle class of Kazakhstan has significantly expanded—from 3.7 percent of households in 2003 to 44.4 percent in 2015. The trend appears to be almost consistent throughout the period: between 2003 and 2006 there was considerable growth in the percentage of Kazakhstan households that moved from poor/vulnerable to the middle class; between 2006 and 2010 the middle class doubled; from 2010 to 2013 this group continued to grow substantially, continuing to grow slightly into 2015. In terms of income, the middle class had only 12.1 percent of total disposable income in 2003 but in 2006 it shot up to almost one-third, reached half in 2010, and accounted for almost two-thirds in the last few years.

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 8).

Figure 8: Official National Poverty Ratio, GDP per capita and middle class size between 2003 and 2015.



Source: World Bank, World Development Indicators.

Was economic growth the only source of the middle-class expansion, as it seems from the graph? To answer this question, we decomposed the changes in the size of the middle class into a “location” effect and a “shape” effect, using a nonparametric version¹¹ of the methodology based on Datt and Ravallion’s (1992) decomposition. 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 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

¹¹ The method is presented in Massari et al. (2009).

can be attributed to redistribution without growth. The last three columns of Table 9 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 9: Evolution of the Kazakh Middle Class: Size, Shares of Total Income, 2003–15, Percent

Year	Middle class (share %)	Share of Total Income	Δ share middle class	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

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 of the middle group in the last years (2014-2015) that would otherwise have reduced due to the growth effect.

7. Conclusions

What constitutes the middle class is hotly debated; different concepts result in different conclusions about the size of the middle class and its evolution. Taking an income-based approach, in this paper we propose a way to estimate absolute thresholds for identifying the middle class in Kazakhstan. Instead of taking boundaries from recent studies for other countries, we formulated a model-based approach to estimating country-specific boundaries. The approach relies on how citizens identified their own status in response to a module of the 2013 HBS questionnaire.

Our main findings are these:

- Taking an absolute approach, for 2013 we estimated absolute boundaries of

about \$14 and \$52 at 2011 PPP international prices.

- Some groups of households, such as families with one or more retired members are over-represented in the middle class, and others, such as families with more children than adults or with no income from employment, are under-represented.
- Consumption aggregates and income aggregates do not differ significantly. Shares of consumption groups (food, essentials, optional items, and luxuries) show different patterns between poor/vulnerable, middle and prosperous classes.
- The Kazakh middle class has increased massively in size and in income concentration. 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 stayed relatively stable into 2015. The increase in the size of the middle class is essentially due to a growth effect.
- Among factors beyond income we found to influence the self-perception of being middle class were access to a household gas supply, to communication, and to the Internet, living in safe areas, and being educated. Household size and regional residence also play a significant role. Nevertheless, income was by far the most important determinant of middle-class status, which supports the income-based approach.

References

- Agresti, A. 2013. *Categorical Data Analysis*, 3rd Edition. New York: John Wiley & Sons.
- Atkinson, A. B., and A. Brandolini. 2013. "On the Identification of the Middle Class." In *Income Inequality*, edited by J. C. Gornick and M. Jantti, 77–100. Stanford, CA: Stanford University Press.
- Banerjee, A. V., and E. Duflo. 2008. "What Is Middle Class about the Middle Classes around the World?". *Journal of Economic Perspectives* 22(2): 3–28.
- Corral, P., V. Molini, and O. G. Siwatu. 2015. "No Condition Is Permanent: Middle Class in Nigeria in the Last Decade." Policy Research Working Paper 7214, World Bank Group, Washington, DC.
- Datt, G., and M. Ravallion (1992). "Growth and Redistribution Components of Changes in Poverty Measures". *Journal of Development Economics*, 8(2): 275-295.
- Easterly, W. (2001). "The Middle Class Consensus and Economic Development". *Journal of Economic Growth* 6(4): 317–35,
- Ferreira, F.H.G., Messina J., Rigolini J., López-Calva L.F., Lugo M.A., Vakis R. (2013). *Economic Mobility and the Rise of the Latin American Middle Class*. Washington, DC: World Bank.
- Gelman, A. 2008. "Scaling Regression Inputs by Dividing by Two Standard Deviations." *Statistics in Medicine*, 27: 2865–73.
- Gelman, A., and J. Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press.
- Gelman, A., Y.-S. Su, M. Yajima, J. Hill, M. G. Pittau, J. Kerman, et al. 2016. arm: data analysis using regression and multilevel/hierarchical models. R package, version 1.9-3. Available at: <http://CRAN.R-project.org/package=arm>.
- López-Calva, L. F., and E. Ortiz-Juarez . 2014. "A Vulnerability Approach to the Definition of the Middle Class." *Journal of Economic Inequality* 12: 23–47.

- Massari, R., Pittau, M.G., and R. Zelli (2009). “A dwindling middle class? Italian evidence in the 2000’s”. *Journal of Economic Inequality*, 7, 333-350.
- Milanovic, B., and S. Yitzhaki. 2002. “Decomposing World Income Distribution: Does the World Have a Middle Class?”. *Review of Income and Wealth* 48(2): 155-78.
- OECD. 2011a. *Latin American Economic Outlook 2011: How Middle Class is Latin America?* Paris: OECD Publishing.
- . 2011b. *Perspectives on Global Development: Social Cohesion in a Shifting World*. Paris: OECD Publishing.
- Pew Research Center. 2016, May. “America’s Shrinking Middle Class: A Close Look at Changes Within Metropolitan Areas.” Washington, DC: PRC.
- R Development Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Ravallion, M. 2010. “The Developing World’s Bulging (but Vulnerable) Middle Class.” *World Development* 38(4): 445–54.