

# QAIDS Model Based on Russian Pseudo-Panel Data: Impact of 1998 and 2008 Crises <sup>1,2</sup>

Maria D. Ermolova<sup>a</sup>, Henry I. Penikas<sup>b</sup>

<sup>a</sup> International Laboratory of Decision Choice and Analysis, National Research University  
Higher School of Economics, Moscow, Russia.  
mermolova@hse.ru

<sup>b</sup> Department of Applied Economics, National Research University Higher School of  
Economics, Moscow, Russia.  
penikas@hse.ru

**Abstract:** The aim of this work is to compare shifts in the consumer behaviour of Russian households since the mid-nineties till nowadays. The research considers the consumer behaviour of the Russians over almost the maximum possible available data RLMS period, focusing on the crisis years. Special attention is paid to analysis of the effects of crises in 1998 and 2008. To reveal effects as shifts in consumer behaviour in the aftermath of two crises panel data analysis is used to estimate QAIDS model. Due to the complete sample attrition observed in RLMS dataset since 1994, pseudo-panel approach is used.

**Keywords:** QAIDS, RLMS, pseudo-panel, consumer behaviour, crisis

**JEL codes:** D12, E21

## 1 Introduction

Economic recessions change consumer behaviour through consumers' expectations that can be also formed by economic policy, economy structure or distribution of households. Structural or temporal shifts determine the subsequent economic policy, whose efficiency, in turn, also evaluated by the change in the welfare of different households. Therefore, analysis of shifts in consumer behaviour needs to be determined accurately. However, any research about life quality is highly dependent on the data used. Data may not always be suitable for study for the following reasons: selection bias (no poorest or richest people), distrust of statistical authorities (respondents often refuse to answer questions or deliberately distort the data), the lack

---

<sup>1</sup> The study was implemented in the framework of the Basic Research Program at the National Research University Higher School of Economics.

<sup>2</sup> The authors are grateful to Rustam Zakirov for conducting initial calculations and to Sergey Vinjkov for research assistance.

of representativeness relative to the general population, and a depletion of the sample over a long period of time.

The aim of this work is to compare shifts in the consumer behaviour of Russian households since the mid-nineties till 2011. In this paper, the data from the survey "Russian Longitudinal Monitoring Survey HSE" (hereafter RLMS) is used.<sup>3</sup> The paper shows that descriptive statistics or model using panel data do not provide enough information about whether 1998 or 2008 crises leads to structural or temporal effect on consumer behaviour. Using pseudo-panels it was found that the effect of 1998 crisis was stronger in consumer behaviour than the crisis of 2008.

The paper is organized in the following way. Section 1 is introduction. Section 2 discusses the works based on QAIDS model briefly. In Section 3 the theoretical demand model is described. Section 4 explains data processing. Model estimation is given in section 5. Section 6 provides the conclusion about the effects of economic shocks and the evolution of consumer behaviour in Russia.

## 2 Literature review

As far as the authors know, there are no works devoted to the study of consumer behaviour of Russian households for such a long time interval (due to the problem of sample depletion).<sup>4</sup>

There are articles covering a relatively short period of time 2000 – 2005 [Penikas, 2008] or focusing on specific aspects of consumer behaviour (differentiation of real incomes of the population on the basis of consumer choice) [Matytsin et al., 2012]. In foreign literature the number of publications on consumer behaviour is much higher, because it is closely associated with the doctrine of welfare of the population.

The article [Deaton et al., 1980] firstly provides a theoretical description of the Almost Ideal Demand Model (AIDS). AIDS has proven its viability and vitality using the British data from 1954 to 1974. [Gardes et al., 2005] concludes using AIDS model that the estimates obtained using the pseudo-panel approach is less biased compared to the cross-section data the usage of cross-section data.

[Tovar et al., 2012] pseudo-panel estimation takes into account the time dependence of the different cohorts, because the same households may be in different households over time.

## 3 Quadratic almost ideal demand model (theoretical model)

Dynamics of consumer behaviour by consumption group and by various consumption directions is considered from the perspective of analysis of coefficients of income

---

<sup>3</sup> Source: "Russia Longitudinal Monitoring survey, RLMS-HSE", conducted by the National Research University Higher School of Economics and ZAO "Demoscope" together with Carolina Population Center, University of North Carolina at Chapel Hill and the Institute of Sociology RAS. URL: <http://www.cpc.unc.edu/projects/rlms-hse>

<sup>4</sup> A gradual decrease in the number of observations.

elasticity derived from QAIDS (Quadratic Almost Ideal Demand System) (Banks et al., 1997):

$$w_{iht} = \alpha_i + \sum_{j=1}^J \gamma_{ij} \ln p_j + \beta_i \ln(x_{ht}/P_t) + c_i (\ln(x_{ht}/P_t))^2 / b(p) + Z_{ht} d_i + u_{iht} \quad (1)$$

Where  $w_{iht}$  – share of household's expenses  $h$  for sets of goods  $i = 1,2,3$  in the moment  $t$ ,  $P_t$  – Stone Price Index ( $\ln P = \sum w_k \ln p_k$ ),  $x_{ht}$  – household's income (the costs are usually used as an equivalent because respondents in surveys tend to understate their own revenues),  $Z_{ht}$  – the matrix of socio-economic characteristics,  $b(p) = \prod_k p_k^{\beta_k}$  price index, ensuring the integrability of the entire system,  $u_{iht}$  includes both individual effect and random error. To estimate the elasticities it is necessary to take derivatives of the above equation  $\ln x$  and  $\ln p_j$ :

$$u_i = \partial w_i / \partial \ln x = \beta_i + 2c_i [\ln(x_{ht}/P_t) - \ln(b(p))] \quad (2)$$

$$u_{ij} = \partial w_i / \partial \ln p_j = \gamma_{ij} - u_i \left( \alpha_j + \sum_{k=1}^K \gamma_{jk} \ln p_k \right) - c_i \beta_j (\ln(x_{ht}/P_t))^2 / b(p) \quad (3)$$

The income elasticity for each household will be defined as  $e_i = \frac{u_i}{w_i} + 1$  (4), and compensated price elasticity for good  $j$  as  $e_{ij}^c = e_{ij}^u - w_j$  (5), where  $e_{ij}^u = \frac{u_{ij}}{w_i} - \delta_{ij}$  is uncompensated elasticity ( $\delta_{ij}$  is the Kronecker symbol, that is equal to 1 when  $i = j$  and 0 in all other cases).

## 4 Data

This work is based on the second phase RLMS data, covering the period of 1994–2011. The RLMS surveys constitute an unbalanced panel, i.e. a household can vary from year to year in the survey (sample attrition). Only 35 % of the household (1 366 / 3 975 of observations) that took participation in the survey of 1994 remain in the polls by 2011.<sup>5</sup> 77% of households in the survey of 2008 are presented in the sample of 2011. The observations are placed in the same income group after data processing. It is the prevalent challenge in studying the effects of crises on different groups of households. In connection with these problems, the paper proposes to use pseudo-panels, namely to generate quasi-households on the basis of real data. Further, data processing will be described.

---

<sup>5</sup> Only 7.5% of households remain after data processing.

## 4.1 Outliers

If the difference between a one-year distribution of costs and incomes is more than 45 percentiles for a household, then the observation is recognized as atypical and removed. The threshold of 45 percentiles has been chosen in such a way to eliminate the problem of underestimating revenues, but at the same time to keep most of the sample. For example, 30 percentiles are not applicable because the sufficient sample part (16.3% of outliers) is removed in comparison with 45 percentiles (6.1 % of outliers).

## 4.2 OECD equivalence scale

The welfare of individuals of the household can be measured either "per capita" or "per consumption unit". The first approach is not applicable due to economies of scale. Two people do not consume two times more goods, because they have both public (car, refrigerator) and private goods (food) within a family. Therefore, it is necessary to implement the concept of "per consumption unit" that will depend on public-to-private goods ratio in the household.

The public-to-private goods ratio varies depending on time and country. Time is introduced through the function that depends on the age of household members. The function is a linear combination of the number of family members belonging to different groups. In our research the Oxford modified equivalence scale [Lubrano, 2010] will be used, as it is the most popular in research on consumer behaviour (see [Banks et al., 1997] and [Penikas, 2008]).

**Table 1.** Modified equivalence scale

<b>Family member</b>	<b>Coefficient</b>
The head of the household	1.0
All others, age > 14	0.5
All others, age < 14	0.3

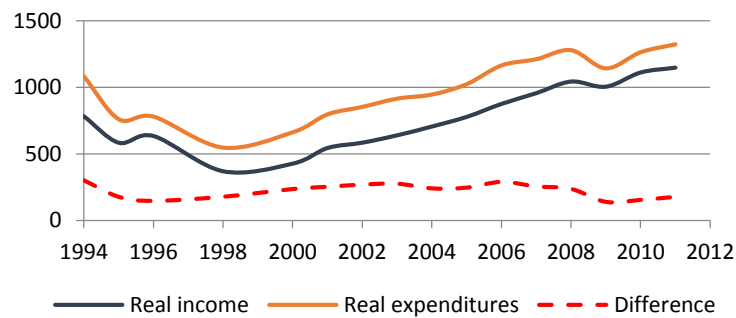
Different measurement scales can lead to different estimates of elasticities by income.

**Table 2.** The effects of equivalence scales

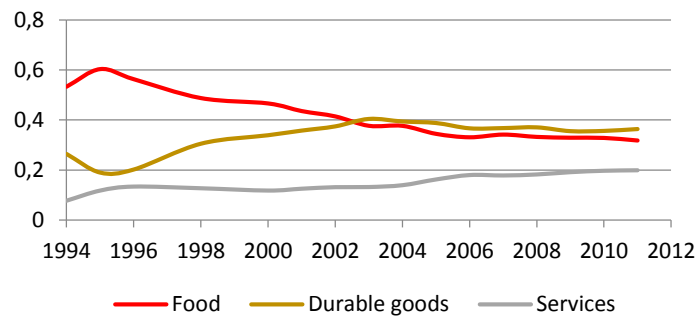
<b>Composition of household</b>	<b>Equivalence scale</b>			<b>Household income</b>
	<b>per capita</b>	<b>Oxford scale</b>	<b>Modified Oxford scale</b>	
1 adult	1	1.0	1.0	1
2 adults	2	1.7	1.5	1
2 adults, 1 child	3	2.2	1.8	1
2 adults, 2 children	4	2.7	2.1	1
2 adults, 3 children	5	3.2	2.4	1
Elasticity	1	0.73	0.53	0

### 4.3 Welfare

**Fig. 1** makes clearly visible the recession of 1998 and 2008 (effect is delayed by a year in 2009) in terms of real income and expenditures. The dynamics of real income for the RLMS sample is in line with trends in real income, represented by Federal State Statistics Service, which is in favour of the representativeness of the study sample.



**Fig. 1.** Real income and expenditures (in 1994 prices)



**Fig. 2.** Weights of expenditures by the groups of goods

The effects of the crises do not appear explicitly if weights of expenditures by good classes are examined. **Fig. 2** shows that the share of expenditures on food decreased over time, which is consistent with the growth in real incomes because the proportion of expenditure on food is a common first approximation of living standards. Over the beginning of the two thousandth's the share of durable goods was actively growing. In recent years an increase in the relative costs of services exceeded all other.

#### 4.4 Homogeneous groups of income using cohort identification

Both multi-criteria index of poor-rich (IMPR) and cluster analysis (k-means) are used to identify homogeneous groups by material welfare. The hypothesis of statistical independence of the IMPR and k-means approach is rejected at 1 % significance level, because sample correlation coefficient of quadratic conjugacy [Ayvazian et al., 1983] is  $X^2 = 29\,841$  ( $p$  – value = 0.0000). Further, it was decided to abandon the use of k-means. Using k-means there were two cases: (1) insignificant coefficients or (2) their sign does not coincide with the sign of model coefficients based on IMPR and with the sign of the correlation.

#### Multi-criteria index of poor-rich.

This method was proposed in [Gardes et al., 1999]. It was used in [Penikas, 2008]. Unlike simpler methods, it includes three main factors, each of which is assigned a score from 1 to 3 depending on the poverty group (1 – poor, 2 – average, 3 – rich):

**Table 3.** IMPR factors

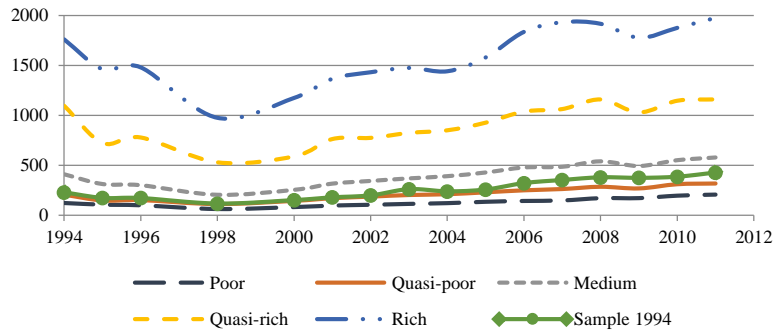
Factors	Population		
	Poor (score = 1)	Rich (score = 3)	Medium (score = 2)
Non-satiated preference relation	Food costs > 4/3 average	Food costs < 2/3 average	Otherwise
Marginalization	Total costs < 2/3 average	Total costs > 4/3 average	Otherwise
Insufficiency of financial resources	Below 25 percentile of cost distribution	Above 75 percentile of cost distribution	Otherwise

*Note:* average = group averages for each year; distribution of costs is adjusted by modified equivalence scale.

*Example:* the total score of observation with marked characteristics (gray cells) equals 6 (=1+3+2).

Scores for each factor were assigned within one year. Based on the obtained ratings for each criterion, IMPR takes 5 different values: 3 if poor, 4 if quasi-poor, 5–8 if middle class, 8 if quasi-rich, and 9 if rich. Using IMPR scoring the following distribution of households by its types was obtained. Each class has a fairly constant weight (12%, 19%, 47%, 8% and 14% respectively for poor, quasi-poor, average, quasi-rich and rich). Also, clear differences in consumption are visible for the classes. The expenditure on food exceeds all other expenses for the two poorest groups over whole time period, while the main item of expenditure is durable goods for the richer households. The highest share of expenditure on services is observed for the two poorest groups. Probably, it is the cost of housing services. All households pay for the housing services, but the impact of these services is stronger for the poor population.

Splitting the sample by income groups allows to identify the impact of the crises on income per consumption unit for the richest three groups (**Fig. 3**). The constant sample (presented over whole period since 1994) is added to show that the constant sample differs insignificantly from quasi-poor households, but absolutely not identified two richest groups for which changes in consumer behavior are obvious particularly. It suggests that the original observations cannot provide sufficient variability to examine differences in consumption behavior deeply. It stresses the relevance of using pseudo-panels.<sup>6</sup>



**Fig. 3.** Income per unit of consumption in 1994 prices

### 4.5 Quasi-households developing

The basic idea of pseudo-panel [Deaton, 1985] is the formation of cohorts that meet certain constant characteristics, such as belonging to a certain income group. Each cohort represents a quasi-household with the average values for cohort:

$$\bar{y}_{ct} = \alpha_c + \bar{x}'_{ct}\beta + \bar{u}_{ct}, \quad c = 1, \dots, C; t = 1, \dots, T, \tag{4}$$

Where  $\bar{y}_{ct}$  is the average of dependent variable in cohort  $c$  in time  $t$ ,  $\bar{x}'_{ct}$  is the average of explanatory variables,  $\alpha_c$  is fixed effect for each cohort,  $c$  is cohort's number, and  $t$  is time. For QAIDS model  $y_{ct}$  means  $w_{ict}$  (the average weight of good  $i$  for cohort  $c$  at time  $t$ ). Average revenue and descriptive statistics of the cohort are included in the vector of explanatory variables.

In the current paper the type of settlement and the average age of the household are used to identify cohorts. The optimal number of groups is formed in such way that the number of households in each group must be positive and the variation should not exceed a reasonable limit.

RLMS surveys indicate 4 main types of settlements: regional center, town, urban-type settlement, and village (44%, 27.7%, 5.7%, and 22.5% of observations, respectively). Two categories are combined into one to obtain approximately constant weights over time for each quasi-households. Urban-type settlement (town) and vil-

<sup>6</sup> The budget coefficients also differ for the three groups of goods insignificantly.

lage were joined, because there is no fundamental difference to interpret consumer behaviour (in both types there is a possibility of employment in agriculture). The weights of each settlement type are relatively stable over time (in average 44%, 28%, and 28% for the regional center, cities, and towns/villages, respectively). Expected differences concerning consumer behaviour are:

- the average income level for each settlement type is different. The larger the settlement, the greater the expected income that affects welfare. Then people who live in cities are rich people, and they need to be differentiated;
- the food expenditures are less in the rural population due to agriculture. The majority of expenditures are on services for the urban dwellers.

Three age groups are formed (**Table 4**). One can observe relatively stable weight for each group over time but still with a slight tendency to increase the percentage of senior households.

**Table 4.** Age groups

Households			
	Young	Older	Mature
%	28 % households	33 % households	39 % households
Age	<28 years	28 – 45 years	>45 years

Two criteria (settlement type and age) create 45 quasi-households ( $45 = 5$  income groups \* 3 settlement types \* 3 age groups) in each wave. Totally, there are 720 observations for the entire period ( $720 = 45$  quasi-households in a year \* 16 years). The average composition of quasi-households is 87 real households.

## 5 Model estimation

Quadratic Almost Ideal Demand Model is estimated for quasi-households both based on income groups (5 quasi-households) and joint groups taking into account income, settlement type, and age of households (45 quasi-households). The model presented in Section 3 also includes the number of consumption units for each household and dummy variables for each year to account for time effect (1994 as a base).

### 5.1 5 quasi-households

The model with the fixed effect is the most preferred model, since every quasi-household is unique and cannot be regarded as the result of a random selection from the general population. Although, consumer behavior is influenced by psychological factors, then the random effect model may be more preferred.

According to the results of F-test the model with a fixed effect is more preferred than pool model for all goods at 1% significance level. Lagrange multiplier test con-



firms the model with a random effect is chosen for food and durable goods, while a final choice for services is the model with the fixed effect. Hausman test identifies that the model with a random effect is the most preferred specification for food and durable goods consumption. The results are in **Table 5**.

**Table 5.** Model specification choice (5 quasi-households)

5 quasi-households	Fixed vs. Pooled		Random vs. Pooled		Fixed vs. Random	
	F-statistics	P-value	$\chi^2 - stat$	P-value	$\chi^2 - stat$	P-value
Foods	57.87	0.00	187.16	0.00	1.51	1.00
Durable goods	25.82	0.00	115.39	0.00	0.98	1.00
Services	66.86	0.00	0.27	0.30	-	-

Hausman test used to check the null hypothesis about whether the variables are exogenous shows that the instrumentation of variables is not necessary (no evidence to reject the null hypothesis).<sup>7</sup>

**Table 6.** Hausman's statistics for testing endogeneity (5 quasi-households)

Model	$\chi^2 - stat$	P-value
Foods	6.46	0.97
Durable goods	14.66	0.48
Services	3.32	0.99

## 5.2 Analysis of structural changes

The homogeneity of three time periods (before the first crisis, between crises and after the second crisis) was studied using correlation analysis and Chow test. The dynamic of correlations of the basic model factors shows that correlation has changed over time. There is the probability to identify a structural change. Chow test rejects the null hypothesis, i.e. there is heterogeneity, and there are two structural breaks that confirms the potential impact of crises of 1998 and 2008. The result is consistent for both 5 and 45 quasi-households.

**Table 7.** Chow's test for 3 subsamples

Model	$\chi^2 - stat$	P-value
Foods	119.9	0.00

<sup>7</sup> It should be noted that covariance matrix for every type of goods was not positive definite. It makes difficult to make a strong conclusion.

Model	$\chi^2 - stat$	P-value
Durable goods	92.88	0.00
Services	60.91	0.00

The change in the model coefficients is observed for each group of goods in the dynamics. The coefficients of demand model for services have the greatest variation.

**Table 8.** The dynamics of estimates for 5 quasi-households.

Period	Total sample	Before 1998	Over 1998–2008	After 2008
<b>Foods</b>				
lnexp	-0.39	-0.09	-0.50	-0.37
s.e.	0.07	0.23	0.06	0.13
t	-5.32	-0.40	-8.13	-2.89
lnexp2	0.01	-0.01	0.02	0.01
s.e.	0.01	0.02	0.01	0.01
t	1.95	-0.61	4.80	1.25
<b>Durable goods</b>				
lnexp	0.30	-0.43	0.35	0.00
s.e.	0.07	0.18	0.07	0.16
t	4.38	-2.34	5.36	-0.02
lnexp2	-0.01	0.04	-0.01	0.01
s.e.	0.01	0.01	0.01	0.01
t	-1.87	3.25	-2.66	1.09
<b>Services</b>				
lnexp	0.19	0.06	0.25	0.42
s.e.	0.05	0.05	0.03	0.11
t	4.24	1.07	9.82	3.82
lnexp2	-0.01	-0.01	-0.02	-0.03
s.e.	0.00	0.01	0.00	0.02
t	-4.95	-1.92	-7.79	-1.74

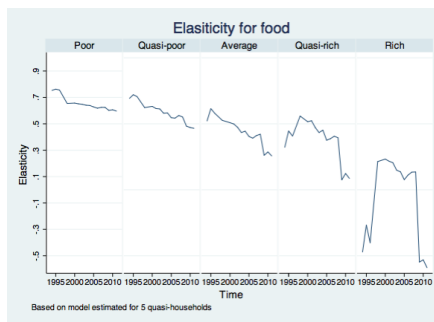
### 5.3 Elasticity analysis

Below the results of income elasticities analysis are presented using the model outputs for 5 quasi-households. The income elasticity for 45 quasi-households repeats the case with 5 quasi-households for the respective income groups.

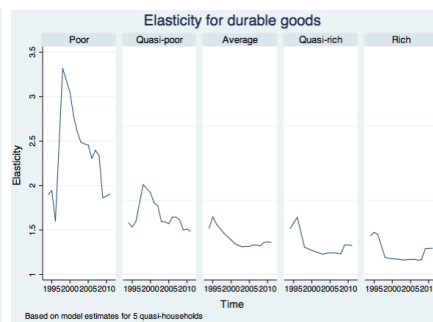
All income groups perceive food as basic necessity goods. However, the richer the group, the smaller the elasticity of food by income (that is, the less necessary the goods become). For the income group "Rich" there is a negative elasticity for the

period up to 1998. However, the estimated coefficients are statistically insignificant. Then we can argue of perfectly inelastic demand on food before 1998, i.e. a change in income has no effect on the food bought (“sticky good”). However, after 2008 a negative elasticity (calculated using the statistically significant coefficients) confirms the conclusion made previously that food products are inferior goods for rich groups.

The income elasticity of demand for durable goods has increased strongly in 1998 and then gradually decreased until 2011 mostly for poor and quasi-poor households. Over the most period durable goods are luxury goods for all income groups. The model estimates for durable goods after 2008 are statistically insignificant, then the durable goods can be recognized as “sticky good“ after 2008 until 2011.

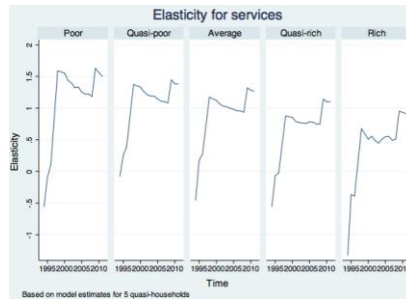


**Fig. 4.** Elasticity for food (5 quasi-households)



**Fig. 5.** Elasticity for durable goods (5 quasi-households)

The calculation with structural shifts for services showed that up to 2000 the demand for services was inelastic (statistically insignificant model estimates for the period before 1998). The calculations taking into account the structural changes also show that the services were luxury goods for the three poorest groups and normal goods for the others. Changes of the elasticities become visible at the moment of the 1998 crisis, namely the increase of elasticities in a time of crisis. The households spent their additional income on services less.



**Fig. 6.** Elasticity for services (5 quasi-households)

## 6 Conclusion

[Varian, 2014] draws attention to the need to explore new methods of data analysis for economics. Such necessity is explained by the fact that many modern solutions, including economic policies, require more complex data analysis tools than using only ordinary linear regressions. Our article provides an example of real data analysis problems motivated by the problem of the variability of consumption.

The work is aimed to study the effects of the crises of 1998 and 2008 on the consumer behaviour of Russian households. The research is based on pseudo-panels, which allowed to get rid of the sample attrition effect (a gradual decrease in the number of observations). Pseudo-panels have allowed us to examine the evolution of consumer behaviour for different groups of households according to two classifications: only by income group; and by income group, type of settlement and age of household members.

Descriptive statistics does not provide any evidence of significant impact of crises 1998 and 2008 on Russian consumption (costs weights have not changed significantly), although there was a decline in real income. The elasticity analysis and structural breaks identification shows that some effects are observed for the 1998 crisis, and there were no significant influence by the crisis of 2008. The estimation of the coefficients of dummy variables demonstrates that the effect of 1998 is the highest compared with all other years: a negative value for food products suggests that these goods became more and more necessary for the Russians while the remaining goods become relatively more luxurious. 1998 was preceded by unfavourable years after the collapse of the Soviet Union, when the population practically had no savings. Therefore the crisis affected consumer behaviour. In 2008 and 2009, the Russians have sufficient savings after favourable period for the economy during the period of 2000–2008.

## References

1. Banks, J., Blundell, R., Lewbel, A. (1997). Quadratic Engel curves and consumer demand. *Review of Economics and statistics*, 79(4), 527-539.
2. Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of econometrics*, 30(1), 109-126.
3. Deaton, A., Muellbauer, J. (1980). *Economics and consumer behavior*. Cambridge university press.
4. Deaton, A., Muellbauer, J. (1980). An almost ideal demand system. *The American economic review*, 70(3), 312-326.
5. Gardes, F., Duncan, G. J., Gaubert, P., Gurgand, M., Starzec, C. (2005). Panel and pseudo-panel estimation of cross-sectional and time series elasticities of food consumption: The case of us and polish data. *Journal of Business & Economic Statistics*, 23(2), 242-253.
6. Gardes, F., Gaubert, P., Langlois, S. (2000). Pauvrete et convergence des consommations au Canada. *CRSA/RCSA*, 36(3), 1-27.
7. Lubrano M. The econometrics of inequality and poverty. Lecture 7: Equivalence scales. 2010. Available online: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.169.3739&rep=rep1&type=pdf>
8. Tovar, A. O., Zulaica, I. G., Núñez-Antón, V. (2012). Analysis of pseudo-panel data with dependent samples. *Journal of Applied Statistics*, 39(9), 1921-1937.
9. Aivazyan S.A., Mkhitaryan V.S. (1983). *Practical Statistics*. Finances and Statistics.
10. Matytsin M.S., Yershov E.B. (2012). Research of Real Income Differentiations of Russians. *Economics Journal of the Higher School of Economics*, 16(3), 318-340
11. Penikas H.I. Analysis of Consumer Behaviour Evolution in Russia throughout 2000-2005. (2008). *Economics Journal of the Higher School of Economics*, 12(4), 512-542.
12. Varian H.R. Big Data: New Tricks for Econometrics. (2014). *Journal of Economic Perspectives*, 28(2), 3-28.