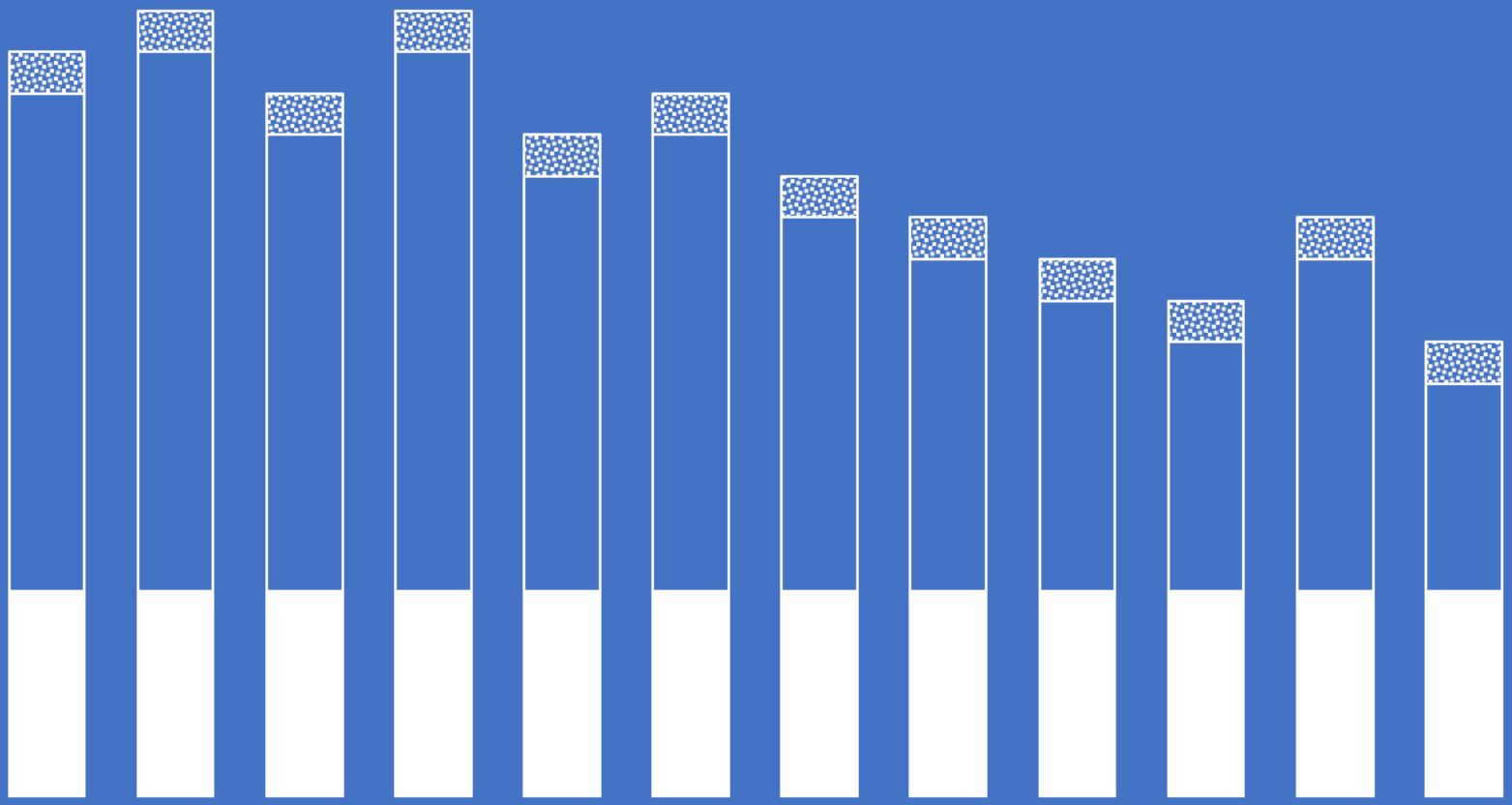
tobaccotaxation
Economic Research Informing
Tobacco Taxation Policy

INSTITUTUL ASPEN ROMÂNIA

Country Report
Romania



June 2024

Bucharest

Impacts of Tobacco Excise Increases on Cigarette Consumption and Government Revenues in Romania

Authors

Vlad Neră u, PhD

George Ștefan, PhD

Marius Geantă, PhD

Rodica Milena Zaharia, PhD

Disclaimer

Aspen Institute Romania is working in cooperation with the Institute of Economic Sciences from Belgrade, which is coordinating a regional network of researchers in Southeastern Europe on tobacco taxation. The project was funded by the Tobacconomics team, now at Johns Hopkins University (JHU) (previously housed at the University of Illinois Chicago, or UIC), to conduct economic research on tobacco taxation in Romania. JHU is a partner of the Bloomberg Initiative to Reduce Tobacco Use. The views expressed in this document cannot be attributed to, nor do they represent, the views of the Institute of Economic Sciences, JHU, UIC, or Bloomberg Philanthropies.

Acknowledgments

This present work benefited from substantial assistance through a two-day econometric model implementation seminar held in March by the Institute of Economic Sciences, Belgrade, for which the authors are grateful. The research also benefited from the logistical support of the Aspen Institute Romania team and the various national and international stakeholders that participated in consultative events and policy discussions over the course of 2023.

Contents

Executive Summary

Consumption of tobacco products in Romania is higher than the EU average and imposes a significant financial burden on smoking households as well as negative externalities for society as a whole. This report analyzes the increase in cigarette prices due to excise tax increases and the associated impacts on cigarette consumption and household expenditure in different income groups, as well as the impact on budgetary revenues.

The data used for estimating the price and income elasticity of demand are microdata gathered through the Household Budget Surveys (HBS) received from the National Institute of Statistics for the period between 2015 and 2021. The present study estimates price and income elasticity for both prevalence and intensity of smoking for the entire population and by income group.

The main finding of the study is that price increases due to an increase in cigarette excise tax would lead to lower consumption, higher budgetary revenues, and positive redistribution effects.

- **Recent economic growth in Romania resulted in increased income, making cigarettes relatively more affordable and likely contributing to increases in consumption.** In the last decade, Romania experienced higher than EU average economic growth rates and managed to achieve the status of high-income country,[1] which led to a consequent increase in population income. The average annual income increase amounted to approximately 10 percent, which is much higher than the increase in cigarette prices over the same time, thus making cigarettes more affordable in relative terms. This is very likely to have contributed to an increase in smoking prevalence from less than 16 percent in 2015 to more than 19 percent in 2021.

- **Significant cigarette excise tax increases would reduce consumption.** An increase in excise duties that translates into a sufficient increase in the price per pack would reduce

[1] According World Bank definitions, high-income economies are those with a Gross National Income per capita of $13,846 or more.

cigarette consumption. The reduction in consumption differs by income group. Households in the low-income category are much more receptive to a price increase, reducing their consumption the most, which is partly due to the fact that some people will stop smoking altogether. It is important that the excise tax increase is sizeable as the results suggest that any price increase below 11.5 percent nationwide would actually lead to an increase in consumption assuming income increases at the historical pace.

- **Increasing excise taxes on tobacco in Romania can reduce the budget deficit.** Irrespective of how cigarette manufacturers choose to respond to excise tax increases (full pass through of a tax increase onto the consumer; under-shifting, meaning sacrificing some profit by not passing the entire tax increase onto the consumer; or over-shifting, meaning raising prices by more than the tax increase, thereby likely increasing profits) an excise tax increase on tobacco would have a positive impact on budget revenues.

- **Increasing tobacco excise tax is both progressive and has a redistributive effect.** The greatest share of the financial burden of an excise increase would be borne by high-income households, because most of the smokers (both in number of smokers and intensity of smoking) are in this category. In the case of low-income households, the decrease in consumption would be caused by some people quitting smoking, which would free up part of the family budget to be allocated towards more productive expenditures.

1. Introduction

In light of growing concerns within the EU regarding life expectancy, and especially the extension of years of life lived in healthy conditions, greater attention is being directed towards the reduction of avoidable causes of premature death.

> "The tobacco epidemic is one of the biggest public health threats the world has ever faced, killing over 8 million people a year around the world. More than 7 million of those deaths are the result of direct tobacco use while around 1.3 million are the result of non-smokers being exposed to second-hand smoke. Around 80 percent of the 1.3 billion tobacco users worldwide live in low- and middle-income countries, where the burden of tobacco-related illness and death is heaviest. In Romania there are 4.1 million smokers, and the annual number of deaths attributable to tobacco smoking is over 36 thousands.[2] Tobacco use contributes to poverty by diverting household spending from basic needs such as food and shelter to tobacco. This spending behaviour can be difficult to curb because tobacco is so addictive."[3]

Last year, Romania adopted the National Plan to Combat and Control Cancer,[4] and the regulations for implementing this plan are currently being developed. This plan notes that, among the risk factors, smoking occupies the first position as the main cause for the onset of lung cancer and is a major risk factor for a number of other cancers, including breast and colon. Smoking is also responsible for heart disease, stroke, lung diseases, diabetes, and chronic obstructive pulmonary disease (COPD), which includes emphysema and chronic bronchitis. Smoking also increases the risk for tuberculosis, certain eye diseases, and problems of the immune system, including rheumatoid arthritis.[5]

Also in this context, a Horizon Project funded by the European Commission, Personalized CANcer Primary Prevention Research through Citizen Participation and digitally enabled social innovation,[6] was launched this year, in which the effects of smoking on the probability of developing any form of cancer will be

[2] WHO. (2021). WHO global report on trends in prevalence of tobacco use 2000-2025, fourth edition (4th ed). World Health Organization. https://apps.who.int/iris/handle/10665/348537
[3] World Health Organization. (2023). WHO report on the global tobacco epidemic, 2023: protect people from tobacco smoke. World Health Organization. https://apps.who.int/iris/handle/10665/372043.
[4] https://ms.ro/media/documents/Planul_Na%C8%9Bional_de_Combatere_%C8%99i_Control_al_Cancerului_RIQiTXG.pdf
[5] Center for Disease Control and Prevention https://www.cdc.gov/tobacco/basic_information/health_effects/index.htm
[6] https://rethink-health.eu/projects/4p-can-project-launch-personalised-cancer-primary-prevention/

analyzed. Reducing tobacco consumption or even preventing the start of consumption will be key elements in reducing cancer risk factors.

Compared to much of the EU, SEE countries are characterized by high levels of tobacco consumption and low prices of cigarettes (Zubović & Vladisavljević, 2020). High tobacco consumption imposes a significant economic burden on households in the region, while at the same time, the negative effects of tobacco consumption have long-lasting effects on health and well-being in general. Numerous studies such as "Tobacco taxes as a tobacco control strategy"[7] (2012) by Chaloupka et al., the regional study titled "Impacts of tobacco excise increases on cigarette consumption and government revenues in Southeastern European Countries"[8] (2019), and the WHO *Technical manual on tobacco tax policy and administration*[9] (2021), emphasize that tobacco taxation is one of the most important policies to reduce tobacco consumption.

The main objective of this research is to estimate the price elasticity of demand for cigarettes, because it plays a crucial role in providing policy makers with essential insights, particularly enabling them to evaluate and model the potential impact of adjusting cigarette taxes on tobacco consumption. Additionally, accurately estimating price elasticity of demand plays a vital role in better forecasting how alterations in cigarette taxes will influence government revenues.

The econometric model for estimating price and income elasticities is based on the theoretical framework of the two-part model developed by Manning and Mullahy (2001). This model estimates the overall demand elasticity as a (corrected) sum of two elasticities: prevalence elasticity and conditional demand (in other words, intensity) elasticity. The prevalence elasticity is estimated via a logit model and generalized linear model (GLM) is used for the estimation of conditional demand (intensity).

In order to estimate the price and income elasticity of demand, Household Budget Survey (HBS) data were used. The data were received from the Romanian National Institute of Statistics for the period of 2015 to 2021. First, we estimated the cigarette price and income elasticity of demand on the extensive (prevalence elasticity) and the intensive (conditional demand, or intensity elasticity) margins. After

[7] https://tobaccocontrol.bmj.com/content/tobaccocontrol/21/2/172.full.pdf
[8] https://tobacconomics.org/research/impacts-of-tobacco-excise-increases-on-cigarette-consumption-and-government-revenues-in-southeastern-european-countries/
[9] https://www.who.int/publications/i/item/9789240019188

estimating the demand elasticity for the entire sample, we divided the sample into three income groups (low income, middle income, and high income), in order to evaluate if the change in the price of cigarettes affects smokers differently depending on their income. And lastly, we conducted a simulation of the impacts of an increase in tobacco excise and price on consumption and government revenue.

2. Data Description and Pre-processing

This chapter describes the data and methodology used in the report. It details the approach for calculating the price participation and intensity (conditional) elasticity of cigarettes in further detail. The mechanism for estimating price elasticity at various income levels is also covered in this chapter. The effects of a price increase on consumption and tax income are then predicted using the estimates. The same two-part econometric models and simulation techniques were used in the regional study from 2019 on "Impacts of tobacco excise increases on cigarette consumption and government revenues in Southeastern European Countries" (Zubović & Vladisavljević, 2019). However, there are minor variations in the model definition and years of available data due to slight discrepancies in the data that are available and country-specific factors.

To calculate the price and income elasticities of cigarette consumption, microdata from household budget survey (HBS) data are used in all analyses. The HBS provides the necessary information for the assessment of income, expenses, and consumption of the population.

Households from all socioeconomic categories are included in the research: employees, employers, self-employed workers in agriculture (farmers) or members of agricultural associations (farmers from commercial agricultural companies), self-employed workers in non-agricultural activities (e.g., tradesmen, traders, and freelancers), members of non-agricultural cooperatives (craft, consumer, and credit cooperatives), unemployed, pensioners, and other categories.

The assignment of households to one of these socioeconomic categories is established based on the declared main occupational status of the head of household.

The HBS is organized as a continuous quarterly survey conducted over a period of 3 consecutive months. The 2021 survey included a sample of 9,504 permanent dwellings, divided into independent monthly subsamples of 3,168 permanent dwellings. The response rate was 80.5 percent (76.5 percent in urban areas and 85.8 percent in rural areas).

In order to extract the sample, a two-stage survey design was used:

- In the first step, 792 research centers (primary sampling units) were selected from the Population and Housing Census of October 2011 (RPL'2011) using the stratified and balanced extraction method of Primary Units (PU) within each stratum, constituting the multifunctional sample of territorial zones (EMZOT'2011 "master" sample) as a survey basis for selective surveys in households in the intercensal period. The stratification criteria were county and type of residence environment: urban or rural. Intersecting these criteria resulted in 88 strata (in the Municipality of Bucharest, the selection was made separately for each of the six administrative sectors). EMZOT is a sample of 792 research centers distributed in all counties of the country and in the sectors of the Municipality of Bucharest (450 in the urban environment and 342 in the rural environment).
- In the second step, 9,504 permanent homes were selected per quarter, in three monthly waves of 3,168 according to a systematic selection algorithm. The homes extracted in the second step are assimilated to the secondary sampling units. From each research center, 12 homes were included in the sample quarterly, respectively four homes monthly.

The sample size was calculated to ensure national and regional representativeness for the main survey variables. The survey sample is extracted from the EMZOT-2011 master sample, based on the data recorded in the 2011 Population and Housing Census. EMZOT-2011 is a database composed of approximately 1,500,000 households, selected according to probabilistic criteria, with the aim of serving as a sampling base for all household survey research for the period 2015–2021.

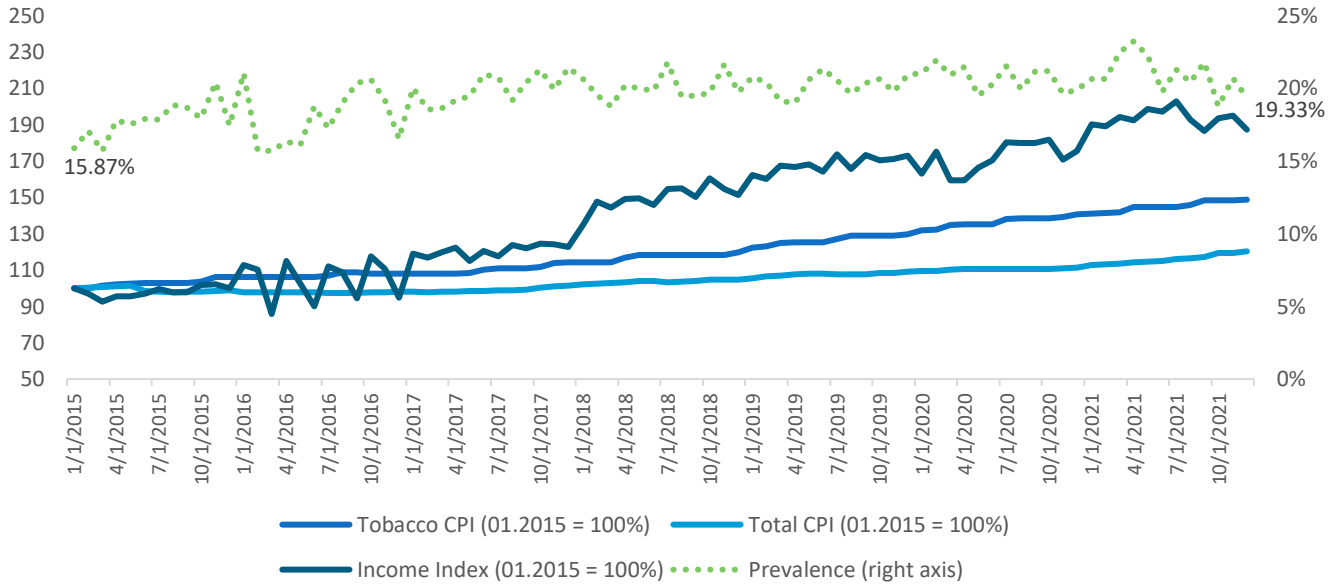
A list of the relevant survey questions on smoking and for the econometric model can be found in Appendix 7.

3. Descriptive Statistics and Variable Definitions

Before moving to the econometric model and price and income elasticity estimations, a quick birds-eye view of the HBS data could provide some insights into the Romanian economic context and tobacco consumption. A critical aspect regarding the consumption of cigarettes is the correlation between the increase in income and the increase in cigarette prices.

As can be seen in Figure 1, even though in the analyzed period 2015–2021 the tobacco consumer price index (CPI) was higher than the general CPI, the increase in the average income of households was considerably higher. The cumulative increase in tobacco CPI was somewhere around 50 percent, while the increase in average income was approximately 90 percent. Therefore, despite the price increases of tobacco products, cigarettes have actually become more affordable. The result of this "cheapening" in relative terms of cigarettes likely contributed to the increase in prevalence rates, though we do not test this claim empirically in this study. The observed prevalence rate increased in the analyzed period from a little below 16 percent to more than 19 percent.

Figure 1. Income, inflation, and tobacco consumption



Source: National Institute of Statistics and HBS data

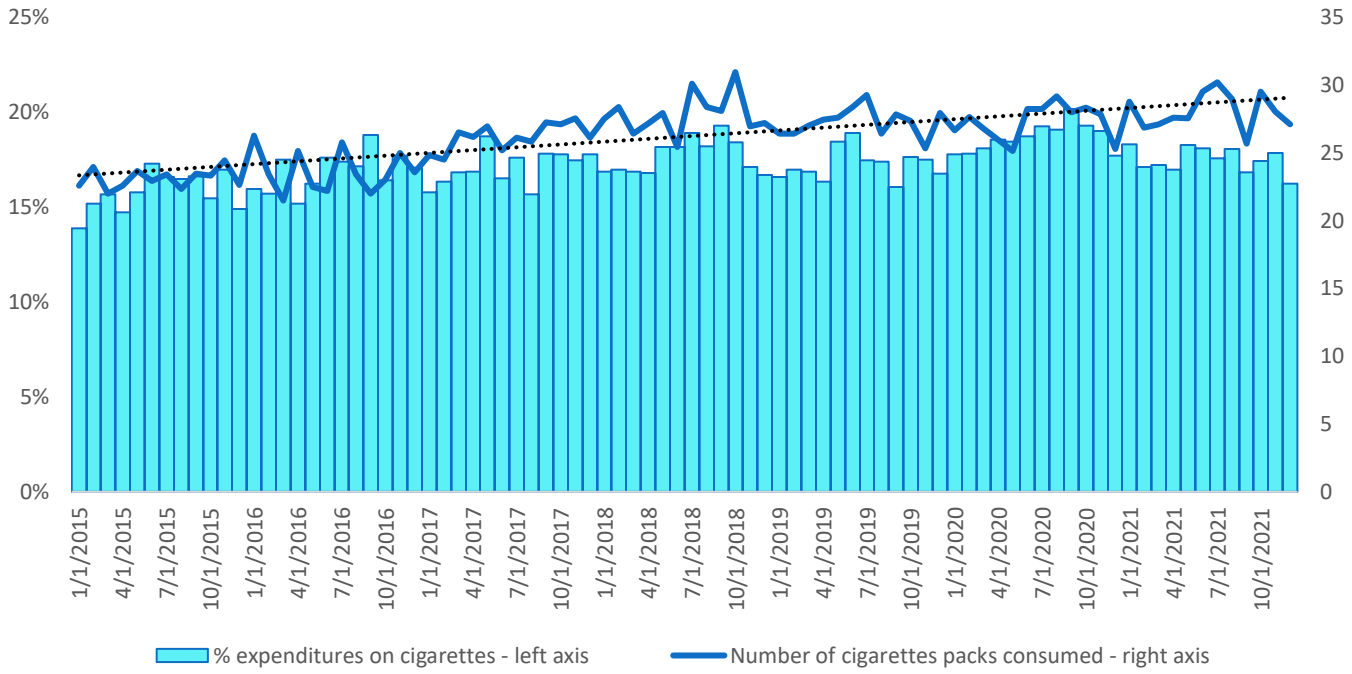
* Note: For Tobacco CPI and Total CPI publicly available data from the National Institute of Statistics from Romania (NIS) were used. For income index and prevalence, data provided by NIS from HBS data were used. The graph displays the information for income after the winsorizing[10] step that was performed in the data cleaning process (that is the same data used in the model).

During this period, the consumption of electronic cigarettes, vaping, and heated tobacco products also increased. This would have been expected to result in reduced prevalence rates for cigarette consumption, due to a switching effect; however, this did not happen. The reasons why this did not happen can be multiple but likely include the fact that the broader increase in income was considerably higher than the increase in the price of cigarettes over the same time period. Another likely factor is that—due to the much looser regulations regarding heated tobacco products—smokers who use heated tobacco products in places where cigarettes are prohibited may be the same people who smoke manufactured cigarettes in places where smoking is allowed. This dual use of heated tobacco products and traditional cigarettes shows they can be complementary products not necessary substitutes.

HBS data also provide insights regarding the share of smoking expenses in the total expenses of a household (see Figure 2 below). Total household spending on consumption is the sum of food expenses, non-food expenses, and services, all found in the HBS survey. On average over the seven-year period analyzed, cigarette expenses account for 17 percent of the total average monthly expenses of a household, which represent an enormous financial burden for the smoking households. Though there may be some over-reporting of expenditure, this average almost certainly represents one of the highest shares of spending on tobacco globally.

[10] Winsorizing represents a method of averaging that initially replaces the smallest and largest values with the observations closest to them. This is done to limit the effect of outliers or abnormal extreme values, or outliers, on the calculation.

Figure 2. Expenditure on cigarettes as share of total household expenditure and number of cigarette packs consumed [11]



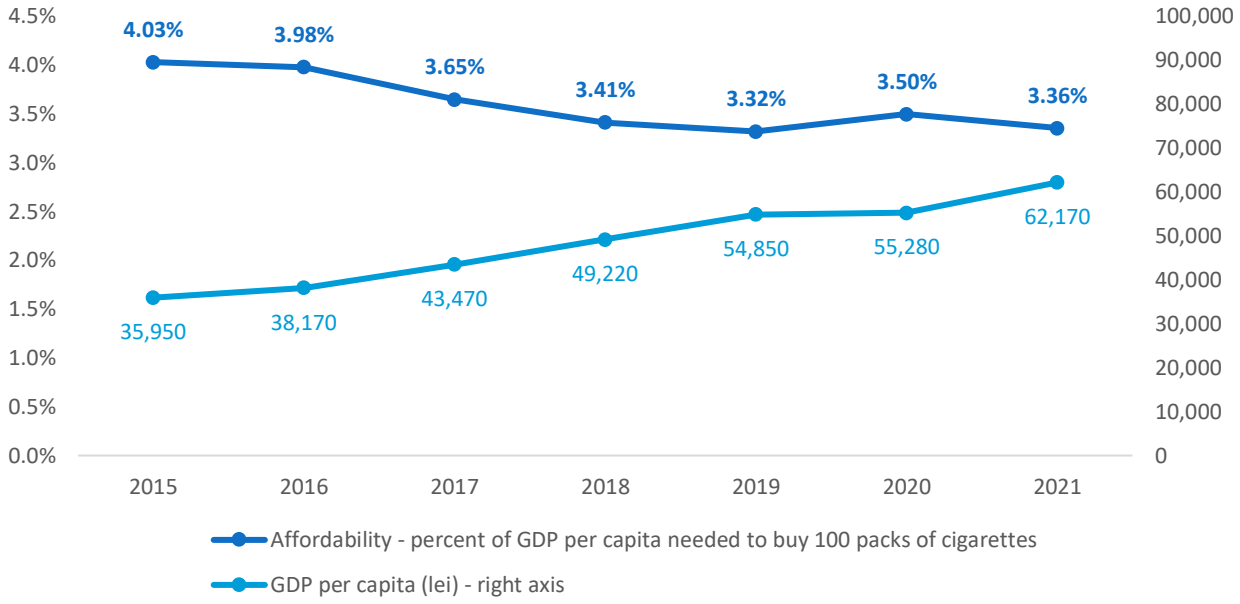
Source: Authors' calculations based on National Institute of Statistics and HBS data

The number of cigarettes consumed on average per household is slightly increasing. The number of cigarette packs (each containing 20 cigarette sticks) consumed in a respective month was determined by dividing the household expenditure on cigarettes by the weighted average price (WAP) (Tobacco Consumer Price Index adjusted) for the respective month. Because both household income and expenses increased significantly in this seven-year period, and at a much faster rate than the price of cigarettes, with the same allocation of 16 percent of total expenses a household could consume approximately 27 packs of cigarettes at the end of 2021 compared to 22 packs of cigarettes in March 2015.

Affordability determines the accessibility of cigarette consumption. Accessibility is most commonly determined by using the relative income price (RIP). In this case, the gross domestic product (GDP) per capita was used. Figure 3 shows the percentage of GDP per capita required to buy 100 packs (of 20

cigarettes each). In 2015, 4.03 percent of GDP per capita was needed to buy 100 packs of cigarettes, while in 2021, only 3.36 percent of GDP per capita was needed to buy the same quantity. This means that, in relative terms, cigarettes have become more affordable because income measured here as GDP per capita (Blecher, E. & van Walbeek, C., 2008) increased faster than the price of cigarettes. In the seven-year period analyzed (2015–2021), cigarettes have become 16.7 percent cheaper in relative terms.

Figure 3. Evolution of affordability and GDP per capita



Source: Authors' calculations based on HBS data, National Institute of Statistics

4. Price Elasticity of Cigarette Consumption in Romania

4.1. Model description

When considering tobacco consumption rates, a significant portion of the total population is composed of non-smokers. This implies that the distribution has a discrete component and a continuous component. This results in the consumption variable taking a value of zero for non-smokers, whereas the rest have strictly positive values. The distribution can be described as:

$$y_i = \begin{cases} 0, & n = 0, 1, \dots, n_i \\ y_i > 0, & n = n_{i+1}, n_{i+2}, \dots n_{i+N} \end{cases} \quad (1)$$

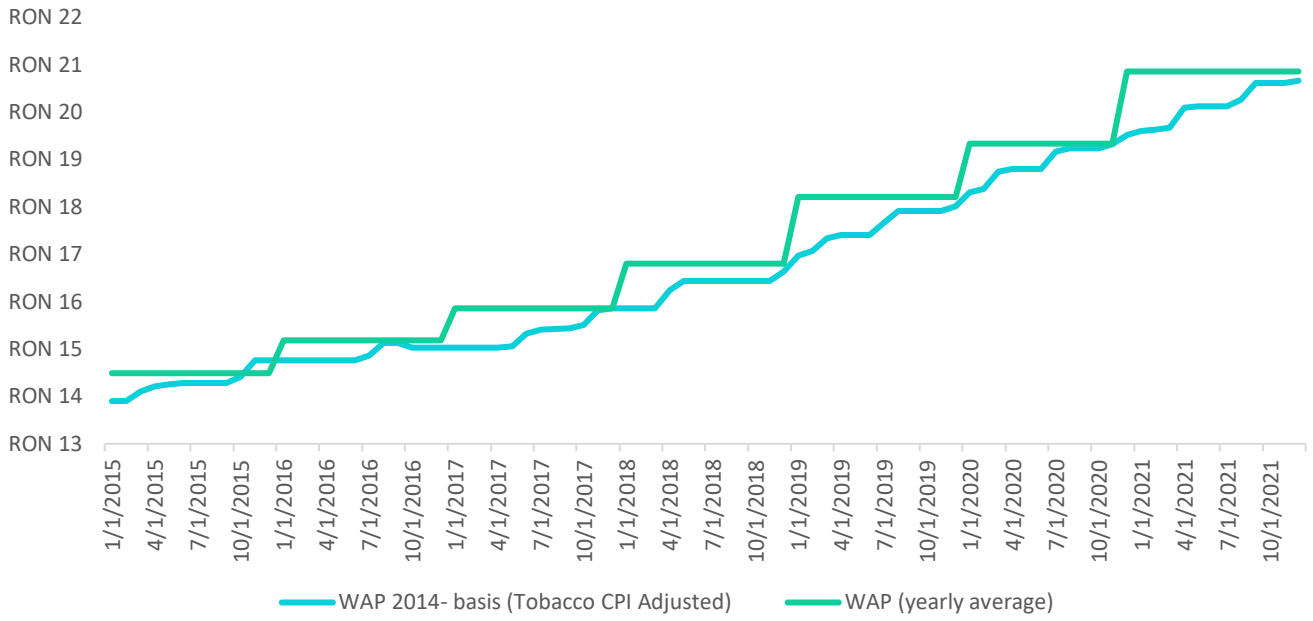
where:

- y_i represents the quantity of consumed cigarettes by a household, and
- n_i represents the household i.

The study examines the distribution of cigarettes. This distribution highlights that individuals—when considering market prices, their financial limits, and the satisfaction they get from smoking—make two primary choices: 1) whether to smoke or not and 2) if they choose to smoke, how much to consume. Existing literature posits that these two decisions should be analyzed separately in what is termed as the two-part model (Belotti et al., 2015). This approach is especially relevant when a value of y=0 is frequently observed. This is evident in cigarette consumption, as the global smoking rate stands at about 17.5 percent (WHO, 2021), while the smoking rate in this specific study is situated between 18 and 24 percent for cigarettes.

Price and income are the two primary factors that both models take into account. The computation of price elasticity, income elasticity, prevalence, and intensity of cigarette smoking is based on these two factors. WAP provided from administrative sources is reported on a yearly basis. In order to generate monthly data, the WAP for 2014 is used as a starting point to compute monthly WAP based on tobacco CPI data. By adjusting the 2014 WAP with monthly tobacco CPI, the resulting values (Figure 4) are slightly lower than the yearly WAP observed on the Ministry of Finance website.

Figure 4. Evolution of weighted average price (WAP)



Source: Authors' calculations based on NIS and Taxation and Customs Union data

As the models are estimated separately and independently, the total price and income elasticities are calculated as the corrected sum of the prevalence and the conditional demand (intensity) elasticity. Total elasticity cannot be calculated as a simple sum of the two elasticities. Instead, this sum needs to be corrected for the fact that a change in the smoking prevalence can attenuate the effect of the conditional demand (intensity) elasticity. Consequently, two models are formulated using the following equations:

$$P\ (y_i > 0) = f(\beta_1 p_{\ j}\ + \beta_2 i_{i,j}\ \ \gamma H_{i,j} + \zeta CF_j) \tag{2}$$

$$E\ (y_i \mid y_i > 0) = \beta_1 p_{\ j}\ + \beta_2 i_{i,j} +\ \ \gamma H_{i,j} + \zeta CF_j \tag{3}$$

where:

- $p_{\ j}$ denotes the price in period j,
- $i_{i,j}$ represents the income for household i in period j,
- $H_{i,j}$ is a set of household specific variables (detailed description in Table 1), and
- CF_j represents a set of control variables (detailed description in Table 1).

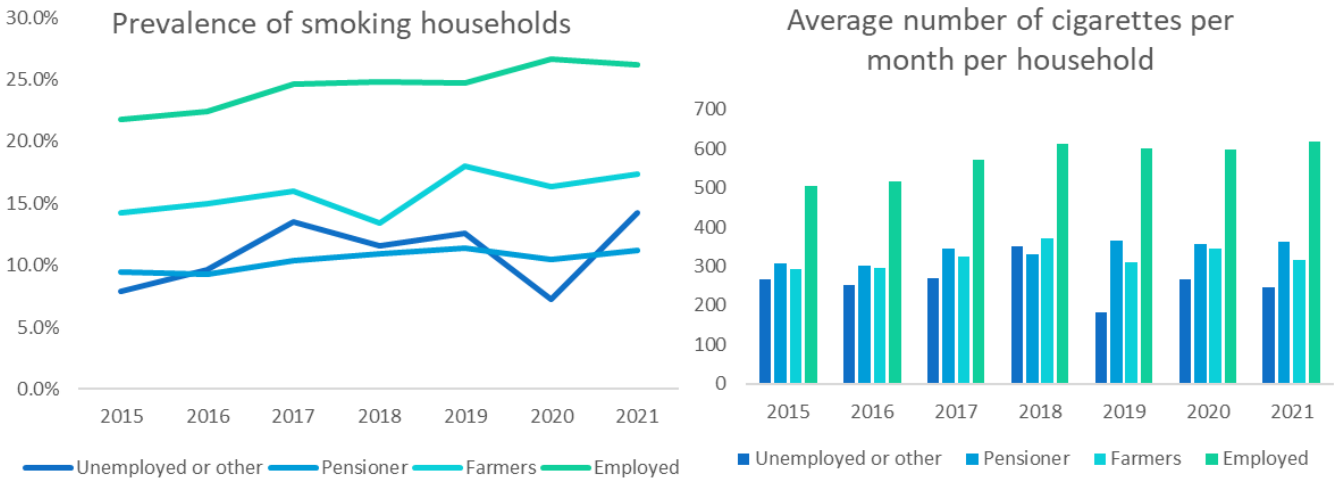
Equation (2) depicts prevalence while equation (3) focuses on intensity. Together, these models form a system of two equations detailing the demand for cigarettes.

It is crucial to consider the potential reciprocal relationship between prices and demand indicators when evaluating price elasticities. Prior research assessing the independence of tobacco prices has deduced that such prices can be considered exogenous (Karki et al., 2003; Kyaing, 2003; NCI, 2016; Kostova & Dave, 2015), even when derived from a similar aggregation level (Huang et al., 2018). Lastly, it is worth noting that prices are not solely driven by market dynamics. State-determined excise taxes also significantly shape prices. Moreover, in the SEE region, price alignment with the EU heavily sways prices, meaning a dip in demand would not necessarily modify cigarette prices.

However, to assess possible endogeneity problems, a Hausman test is performed on the model. The test suggests that there are possible endogeneity issues. To address this problem, an instrumental variable approach is used. First, the price is estimated using the same regressors as in the model and an instrumental variable. In this case, the excise tax is chosen as an instrumental variable for price as it clearly has an impact on the price, but the level of the excise should have no impact on smoking decisions. The estimated level of price is used as a regressor instead of the original price variable.

The control variables considered in the model are represented by household size, proportion of males in the household, proportion of adults in the household, legislative measures imposing a ban on smoking indoors, highest level of education in the household, and highest level of occupation in the household. In the preprocessing step, education is grouped into six levels: 1 - less than primary, 2 - primary, 3 - secondary, 4 - post secondary, 5 - tertiary BA, 6 - tertiary MA, PhD. Occupation status is grouped into four levels: unemployed or other (unemployed, family support, student, housewife, other status), pensioner, farmer (agricultural freelancer, member of an agricultural cooperative), and employed (employee, owner, freelancer, member of a non-agricultural cooperative). The lowest level of occupation status is unemployed or other, followed by pensioner, then farmer. The highest level of occupation status is employed.

Figure 5. Prevalence and intensity based on household occupational status



Source: Authors' calculations based on NIS HBS data

Given that the control factors from the model can have a significant impact on coefficients that ultimately determine the elasticities, of all the model variations, the only difference between the best candidate models (models 3 and 5) is the presence of the highest occupational status in the household. We consider these control factors to be relevant and important to be kept in the model as there is significant heterogeneity between occupational groups, as depicted in Figure 5. The observed heterogeneity manifests both in terms of prevalence, with a clear distinction between groups (household with higher occupational status have higher prevalence), as well as in intensity (household with higher occupational status consume, on average, more cigarettes than those with lower occupational status).

For the purpose of this estimation, five variations of models are tested (detailed results presented in the Appendix). The best functional form of the model is identified as model (5) in log form, which has the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) scores for the second part of the model, passes the link test, and does not suffer from multicollinearity. However, model (3) has the lowest AIC and BIC scores for the first part. And even though model (3) does not pass the link test, it is right on the limit with a score of 0.049. Model (3) also includes variables that show the occupational status within the household, while model (5) does not. For these reasons, model (3) is preferred as a functional form, even though its failure to pass the link test may lead to biased or inefficient parameters.

After the selection of model (3) as the preferred model, the goodness of fit is assessed for deciles, with results indicating that there is no evidence of systematic bias in the residuals across the deciles of predicted values, indicating a good model fit across different levels of the predictors.[12]

After identifying the best form of the two-part model and using the same log specifications, the model is re-run, conditional on each separate income group, to assess if there are significant differences of elasticity by income group. The whole sample of observations, approximatively 105,000, is divided into three equal-size observation groups representing low-income, middle-income, and high-income groups.

4.2. Estimation of prevalence elasticity

The initial segment of the model examines how tobacco prices influence a household's choice to smoke, given the set of independent variables. This decision is commonly represented using a binary choice model. The distinguishing factor between a binary choice and the standard linear regression model is the character of the dependent variable. Binary choice models, rather than modeling a continuous variable, focus on the likelihood that the dependent variable y_i equals one—indicating households that spend on or consume cigarettes—as opposed to zero, which signifies households that do not consume cigarettes. As a result, the model employs a (nonlinear) function of the linear combination of independent variables to articulate the probability of a household incurring positive tobacco expenses.

For the first part of the model that estimates prevalence, a logit model is used to estimate equation (2). County-level, month, and year cluster corrected standard errors are applied to account for any differences that may manifest locally or that have appeared over the course of the studied interval, as well as heteroscedasticity-robust standard errors to control for potential heteroscedasticity in both parts of the model. For the purpose of this report, the variables are used both in level as well as logarithm form to test the robustness of the results. Also, further tests have been performed to assess the performance of each of the estimated models in order to select the best functional form for the model. These tests include

[12] The test was performed both for the original model with F-test was F(10, 20637) = 0.94, and Prob > F = 0.4921, as well as for the model updated to accommodate endogeneity problems with a F(10, 20637) = 0.78, and Prob > F = 0.6475

Akaike and Bayesian Information Criteria, collinearity diagnostics, a link test, and goodness of fit tests (the test results for each model are presented in appendices 2 to 5).

Given the structure of HBS data, where households change each year, a typical panel regression is not applicable in its standard form, as it generally relies on observing the same units across different time periods to capture unobserved individual heterogeneity. To address this, the current model is a pooled regression, which combines all data into a single model without accounting for the individual effects specific to each unit or time interval. If cross-sectional units change each period, pooled regression can still be applied as it combines all data into a single model without accounting for the individual effects specific to each unit.

4.3. Estimation of intensity elasticity

In intensity models, such as Equation (3), the dependent variable is typically represented in log form, as it helps to stabilize non-constant error variance; however, similar to Equation (2), both level and log form are estimated and tested for performance. A standard practice in health economics, in this case, is to use the generalized linear model (GLM) with the gamma family and a log link function. This method has been proposed as a more robust alternative to a log regression specification (Manning et al., 2005). In this situation, GLM is the preferred model as the ordinary least squares (OLS) estimator requires re-transformation, which can cause a prediction bias.

In order to implement a two-part model, each component is estimated and tested separately and then aggregated into a two-part model.

4.4. Data description

Before estimating the models, some of the variables need to be transformed in order to increase the quality of the modeling process. This section provides a description of all the variables used and the transformations that are performed on each of them before the modeling process is conducted. The descriptive statistics are presented in full in Appendix 1.

Table 1. Variable names and descriptions

Variable name in the model	Description
hsize	The total household size computed for each combination of year and household (including children).
nadults	The number of adults in each household was determined by counting all persons above 14 years of age.
adultratio	The ratio between the number of adults and household size
namales	The number of males in each household
maleratio	The ratio between the number of males and household size
educc	The level of education of the household member with the highest education. In the HBS data, there are 12 categories of education. For the purpose of this study, the number of categories was reduced to six: 1 - less than primary, 2 - primary, 3 - secondary 4 years, 4 - post-secondary, 5 - tertiary BA, 6 - tertiary MA, PhD.
deduc1 to deduc6	A set of dummy variables corresponding to each level of education (some of the specifications and tests needed this type of data structure). In all the specifications of the models, one of these variables is dropped.
htype	The occupational status of the household member with the highest occupational status. In the HBS data, there are 14 categories of occupational status. For the purpose of this study, the number of categories was reduced to four: 1 - unemployed or other, 2 - pensioner, 3 - farmer, 4 - employed.
htdum1 to htdum4	A set of dummy variables corresponding to each level of occupation. In all the specifications of the models, one of these variables is dropped.
ecig	Household expenditure on cigarettes
insdban	A dummy variable that highlights when there was a change in Law that prohibited smoking in indoor places like restaurants, bars, office buildings, etc.

wap	Given that the quantities of cigarettes are not collected in HBS in Romania, to determine the price, we had to rely on administrative information regarding the weighted average price per pack (which is available at a yearly level from the Ministry of Finance, or MoF). In order to infer monthly data, we used the WAP data for 2014 as a basis and computed the monthly values for January 2015 to December 2021 using tobacco CPI data provided by NIS. This approach led to slightly different average yearly WAP rates compared to the Ministry of Finance values for WAP.
ncig	The number of cigarettes consumed by the household. Since quantity data are not readily available in HBS, this was determined by using expenditure data and WAP data. The number of cigarettes consumed was estimated as the ratio between the expenditure on cigarettes and the WAP multiplied by 20 (the number of cigarettes in a pack).
idcig	A dummy variable that highlights if a household has any expenditure on cigarettes.
wap2	The square of price (WAP)
lnp	The natural logarithm transformation of price
lnp2	The square of log transformation of price
mce30	The income at household level extracted from HBS data. The income was divided by 10,000 to reduce the possibility that the models will be affected by the difference in scales between variables. Furthermore, to mitigate the impact of outliers on the model, income above the 99th percentile was replaced by the income corresponding to the 99th percentile.
mce2	The square of income (mce30)
lny	The natural logarithm transformation of income
lny2	The square of the log transformation of income
income_grp	A new variable containing information about the income group was created. For each year the households were divided into terciles based on their income:

	lowest-income households, middle-income households, and highest-income households.

5. Results

5.1. Prevalence elasticity

The results for prevalence elasticity (extensive margin), using variables in logarithm, are presented in Table 2, indicating a price elasticity of -0.111 and an income elasticity of 0.240. The detailed results of all the model specifications are presented in Appendix 2 and Appendix 3 (the selected model, model (3), is presented in Appendix 3).

Table 2. Prevalence elasticity

	ey/dx	std. err.	z	P>|z|	95% conf. interval	
Price elasticity	-0.111***	0.0711	-1.58	0.115	-0.251	0.027
Income elasticity	0.240***	0.0219	10.94	0	0.197	0.283

Source: Authors' calculations

All of the estimated models perform modestly for the first part. However, adding more interaction terms or higher order terms is not supported in literature, and in the majority of the cases it also induces severe multicollinearity into the model. The current form that is kept is the one that 1) does not present multicollinearity problems and 2) for the second part of the model has values very close to the lowest values for the information criteria (AIC and BIC) and includes variables on occupational status within the household.

One possible cause for these results is the structure of the data: the data contain information at the household level when the decision to smoke is rather a personal one, thus making it difficult to use household data to classify actions (that is, the decision to smoke or not) that are taken by an individual from that household (for example, even if one or more individuals decides to smoke or quit smoking, if another person from that household already smokes and keeps smoking, the information about the household does not change, even though, at the individual level, changes have occurred).

5.2. Intensity elasticity

The results for intensity elasticity (intensive margin), using variables in logarithm, are presented in Table **Table 2**, indicating a price elasticity of -0.432 and an income elasticity of 0.508. The detailed results of all of the model specifications are presented in Appendix 4 and Appendix 5 (the selected model, model (3), is presented in Appendix 5).

Table 3. Intensity elasticity

	ey/dx	std. err.	z	P>|z|	95% conf. interval	
Price elasticity	-0.432***	0.0471	-9.18	0	-0.525	-0.340
Income elasticity	0.508***	0.0146	34.85	0	0.480	0.537

Source: Authors' calculations

Compared to prevalence intensity for the second part, the model performs better in the link test. Although the value (0.049) is still less, than the threshold of 0.05, it is not significantly less. It also passes the specification and the Collin test for multicollinearity. The same caveats as with the prevalence model still apply, even though they are mitigated by the fact that the decision on how many cigarettes to buy (being continuous not discrete) is evaluated at the household level, as previously argued by Deaton and Ng (1988). Given that the actual consumption is derived from average prices and the fact that the real quantity is not collected in the surveys, but computed from average prices, one of the drawbacks of this approach is that it does not capture quality switching. However, national-level data should capture the phenomenon accurately.

5.3. Two-part model

The two-part approach essentially takes both Equation (2) and (3) and estimates together. Although the literature suggests that these two decisions can be modeled independently, total elasticity cannot be calculated as the simple sum of the two elasticities. Instead, this sum needs to be corrected for the fact

that a change in the smoking prevalence can attenuate the effect of the conditional demand (intensity) elasticity. The total price and income elasticities are the sum of prevalence and intensity elasticities, **resulting in a price elasticity of -0.545 and an income elasticity of 0.749**. Both price and income elasticities of demand are within the range observed in the literature for medium-to-high- and high-income countries like Romania. The model used to estimate the elasticities together with the variables that were included and their related coefficients can be seen in Table 4.

Table 4. Two-part model estimations

	Prevalence model		Intensity model			
VARIABLES	Coef.	std. err.	Coef.	std. err.	Estimated elasticities	std. err.
Linear prediction	-0.139	(0.088)	-0.433***	(0.047)	-0.545***	(0.085)
lny	0.299***	(0.027)	0.508***	(0.015)	0.749***	(0.026)
lny2	-0.050***	(0.010)	0.036***	(0.006)		
hsize	-0.030***	(0.005)	0.016***	(0.002)		
maleratio	1.248***	(0.037)	0.307***	(0.019)		
adultratio	-0.390***	(0.078)	0.160***	(0.042)		
educc==less than primary	-0.451***	(0.076)	-0.110**	(0.051)		
educc==primary	-0.075**	(0.033)	-0.047***	(0.017)		
educc==post secondary	-0.091***	(0.027)	-0.005	(0.013)		
educc==tertiary BA	-0.099**	(0.040)	-0.043**	(0.020)		
educc==tertiary MA, PhD	-0.138***	(0.024)	-0.031***	(0.012)		
htype== 1.0000	-0.266**	(0.116)	0.010	(0.070)		
htype== 2.0000	-0.573***	(0.028)	-0.008	(0.014)		
htype== 3.0000	-0.178***	(0.041)	-0.053**	(0.025)		
Constant	-0.584**	(0.279)	7.323***	(0.152)		
Observations	104,919		104,919		104,919	

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Authors' calculations

5.4. Income group estimation

Using the same log specification as above, the model is re-run, conditional on each separate income group. The elasticities are presented in Table **5**5, while the detailed results of estimations are presented in Appendix 6.

Table 5. Elasticities estimation – national level and income group level

	Group 1 - low income		Group 2 - middle income		Group 3 - high income		Total	
VARIABLES	Prevalence	Intensity	Prevalence	Intensity	Prevalence	Intensity	Prevalence	Intensity
Price elasticity	-0.349**	-0.386***	-0.114	-0.520***	-0.107	-0.463***	-0.112	-0.433***
	(0.158)	(0.0973)	(0.127)	(0.0800)	(0.111)	(0.0738)	(0.071)	(0.047)
Income elasticity	0.566***	0.800***	0.130	0.534***	0.232***	0.469***	0.240***	0.508***
	(0.175)	(0.120)	(0.137)	(0.0846)	(0.0362)	(0.0251)	(0.022)	(0.015)
Observations	34,937	4,146	34,987	7,538	34,995	8,963	104,919	20,647

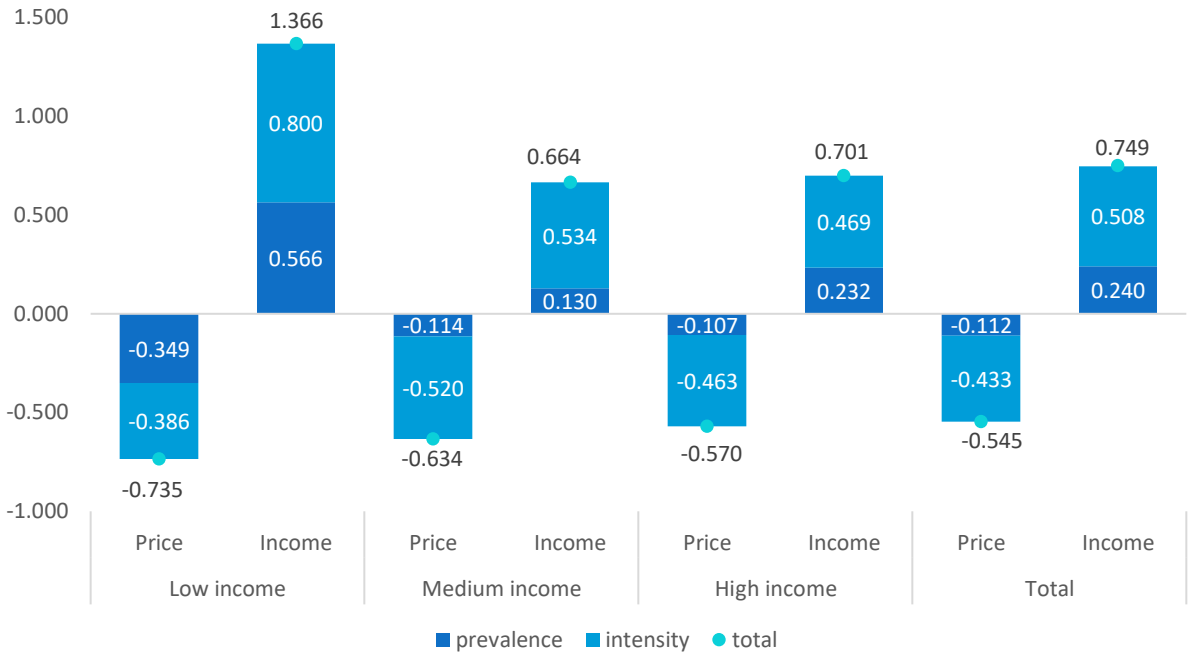
Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Authors' calculations

Based on the estimates of prevalence and conditional demand elasticities from the previous sections, total demand elasticity is calculated and presented below, by income group, in Figure 6.

Figure 6. Elasticities – national level and by income group



Source: Authors' calculations based on the estimated elasticities from Table 6

Total price elasticity is the highest for low-income households at -0.735, which means that a 10-percent price increase leads to a decrease in consumption by 7.35 percent. Price elasticity is the lowest in the high-income group at -0.570, with the effect of price on prevalence being almost insignificant, meaning that most of the increase in price will lead to a reduction in consumption rather than quitting.

It is important to note that a similar percentage increase in income as in prices will offset the increase in cigarette prices, especially for the low-income category where income elasticity is highest, at almost 1.4. This is relevant since, as explained in the descriptive statistics chapter, income increases over the last decade have been substantial.

6. Simulation of Price and Excise Increase on Consumption and Government Revenue

In this chapter, the previously estimated price and income elasticities are used to simulate the impact of price and excise tax increases on consumption and government revenue. The total price and income elasticities are calculated as a corrected sum of prevalence elasticity and intensity. The starting point of the analysis is cigarette consumption, which is obtained from the administrative data on cigarette packs for the year for which the latest HBS is available, which is 2021. In order to account for the impact of an increase in income on consumption, total income at the household level is used. In the analyzed period of seven years, the compounded annual growth rate of total household income was 10 percent.

Six scenarios are simulated, presenting the estimated impact of three alternative price increases of five, 10, and 20 percent (considering full pass through of the excise increase from the producers, meaning an excise increase of 8.8, 17.5, and 35.1 percent, respectively, would be needed for the previous price increases), while maintaining the historical average of income increase of 10 percent. Another three scenarios are simulated for the same increases on prices, but these are corroborated with an income increase of only five percent, given the current economic context.

6.1. Simulation of the impact of price on demand and expenditures of income groups

The entire sample of approximately 105,000 households is divided into three approximately equal groups of 35,000 households, each based on income criteria. As can be seen in Table 7, the number of smoking households in the high-income group is more than double the number of smoking households in the low-income group. Based on the cigarette expenses of each income group, the share of cigarette consumption from the total cigarette consumption of each group is computed. Income elasticities for each group are used to determine the impact on consumption. We assume a 10-percent income increase based on the historical Compounded Annual Growth Rate (CAGR) for 2015–2021. Total elasticity (prevalence and intensity) is used for each income category and for the total population to assess the impact of price and income increases. Other data used for this simulation include the consumption of

cigarette packs and total tax revenues from the EU Taxation and Customs Union[13] for Romania, the weighted average price per pack of cigarettes, the excise rate in the baseline year, the value-added tax (VAT) rate of 19 percent, and income CAGR.

To simulate the transmission of the excise tax increase in the final price of cigarettes, the conditions of 2021 are maintained, where the total taxes (ad valorem excise + specific excise + VAT) represent 72.88 percent of the price of a pack of cigarettes, while the producer's share is 27.22 percent. Keeping these assumptions, in order to obtain the price increases used in the simulation of five, 10, and 20 percent, excise taxes should increase by 8.8, 17.5, and 35.1 percent, respectively (considering full pass through of excise).

In table 6, three different scenarios are presented taking into account the assumption that the income increase will be at the historical CAGR of 10%. The three scenarios simulate what will happen to consumption and budget revenues if prices increases will be of 5%, 10% and 20% respectively. Table 6, below, presents both the overall impact as well as the impact on each income group (low, medium, high).

[13] EU Taxation and Customs Union - Excise Duties on Tobacco section: https://taxation-customs.ec.europa.eu/taxation-1/excise-duties/excise-duties-tobacco_en

Table 6. Tobacco consumption and tax revenue simulation with 5%,10%, and 20% price increase while maintaining the historical CAGR of income (10%)

Price increase 5%	Consumption (mil. packs)			Government revenues (mil. RON)		
Income increase 10%	Baseline	Scenario	Change	Baseline	Scenario	Change
Low - income	81	88	9.5%	1,224	1,407	15.0%
Middle - income	295	305	3.3%	4,484	4,862	8.4%
High - income	651	677	4.0%	9,898	10,804	9.2%
Total	**1,026.5**	**1,069.5**	**4.2%**	**15,605.9**	**17,073.1**	**9.4%**
Price increase 10%	Consumption (mil. packs)			Government revenues (mil. RON)		
Income increase 10%	Baseline	Scenario	Change	Baseline	Scenario	Change
Low - income	81	85	5.3%	1,224	1,418	15.8%
Middle - income	295	295	-0.1%	4,484	4,927	9.9%
High - income	651	657	0.9%	9,898	10,987	11.0%
Total	**1,026.5**	**1,036.4**	**1.0%**	**15,605.9**	**17,331.1**	**11.1%**
Price increase 20%	Consumption (mil. packs)			Government revenues (mil. RON)		
Income increase 10%	Baseline	Scenario	Change	Baseline	Scenario	Change
Low - income	81	78	-3.0%	1,224	1,424	16.3%
Middle - income	295	275	-6.9%	4,484	5,011	11.7%
High - income	651	617	-5.2%	9,898	11,261	13.8%
Total	**1,026.5**	**970.0**	**-5.5%**	**15,605.9**	**17,695.7**	**13.4%**

We can observe from the above table that even a 10% increase in price would result in an increase in consumption, that is because income elasticity is higher than price elasticity. Maintaining the historical increase in income of 10%, we simulated what the necessary increase in price should be in order to have no change in consumption. For that to happen, the increase in price should be of 11.5%.

Table 7 presents the same simulation as Table 6, with the only difference being a more conservative increase in income of 5% in comparison with the historical CAGR of 10%.

Table 7. Tobacco consumption and tax revenue simulation with 5%,10%, and 20% price increase with a conservative 5% income increase

Price increase 5%	Consumption (mil. packs)			Government revenues (mil. RON)		
Income increase 5%	Baseline	Scenario	Change	Baseline	Scenario	Change
Low - income	81	83	2.9%	1,224	1,322	8.0%
Middle - income	295	295	0.0%	4,484	4,711	5.0%
High - income	651	655	0.6%	9,898	10,450	5.6%
Total	**1,026.5**	**1,032.6**	**0.6%**	**15,605.9**	**16,483.3**	**5.6%**
Price increase 10%	Consumption (mil. packs)			Government revenues (mil. RON)		
Income increase 5%	Baseline	Scenario	Change	Baseline	Scenario	Change
Low - income	81	80	-1.0%	1,224	1,333	8.9%
Middle - income	295	285	-3.2%	4,484	4,773	6.4%
High - income	651	635	-2.4%	9,898	10,627	7.4%
Total	**1,027**	**1,001**	**-2.5%**	**15,606**	**16,733**	**7.2%**
Price increase 20%	Consumption (mil packs)			Government revenues (mil RON)		
Income increase 5%	Baseline	Scenario	Change	Baseline	Scenario	Change
Low - income	81	73	-8.9%	1,224	1,338	9.4%
Middle - income	295	266	-9.8%	4,484	4,855	8.3%
High - income	651	597	-8.3%	9,898	10,892	10.0%
Total	**1,026.5**	**936.5**	**-8.8%**	**15,605.9**	**17,085.3**	**9.5%**

Source: Authors' calculations based on previously estimated price and income elasticities

7. Conclusions

The results show that any of the modeled price increases would lead to an increase in budget revenues. At the same time—and this is probably the most important conclusion—**a price increase of less than 11.5 percent,** given the historical evolution of household incomes from 2015 up to 2021, **would still lead to an increase in cigarette consumption.**

In the analyzed period between 2015 and 2021, based on HBS data, a CAGR of household incomes of approximately 10 percent is observed. This means that despite the increase in the price of cigarettes (their price throughout this period increased faster than the consumer price index), in relative terms, cigarettes became cheaper in 2021 compared to 2015. Thus, with the median income of a person in the low-income group, 43 packs of cigarettes per month could be bought in 2015, while in 2021, the same person could buy 59 packs of cigarettes. In the case of a person in the high-income group, 79 packs of cigarettes could be purchased in 2015 compared to 122 packs in 2021. This "cheapening" of cigarettes eventually translates to an increase in prevalence, which increased from 15.9 percent in 2015 to 19.3 percent in 2021. It should be noted that this study accounts only for manufactured cigarettes, so the increase in the prevalence of smokers (manufactured cigarettes) happened in parallel with the increase in the number of smokers of novel tobacco products.

In the case of smoking households, the cost of purchasing cigarettes represent a real financial burden. In the analyzed period, these expenses represented approximately 17 percent of the total average monthly expenses of a household. Research from other countries consistently demonstrates that expenditures on smoking result in a crowding-out effect for other expenses, such as expenses for higher quality food, better education, or clothing. These expenses also have intergenerational effects: children who grow up in smoking households, in addition to the effects of second-hand smoking on health, are also put at a disadvantage because less funds are allocated to them, due to the decrease in the available family budget because of cigarette expenses.

Regarding the simulations based on income groups, it is observed that an increase in the price of cigarettes has substantially different effects depending on the income category in which each smoking

household is situated. The elasticity of demand (price and income) is much higher in the case of low-income households. Thus, a price increase would have a higher impact on low-income households than on those with high incomes. It should be noted that the income elasticity of demand is very high for the low-income group, and thus a price increase could be offset by an income increase, as can be seen in Table 6.

The different reaction to a price increase depending on the income group can be seen both in the case of people who, in the event of a price increase, would give up smoking altogether and, among people who would continue to smoke, would decrease their consumption. In the case of low-income households, the decrease in consumption would occur mostly due to the fact that some people would stop smoking altogether. For people with high incomes, the price increase would translate into a decrease in intensity (in other words, they would smoke less) while only very few would stop smoking altogether, as can be observed in Figure 5.

Finally, it is worth mentioning that, as can be seen in tables 6 and 7, most of the cigarette consumption is done by households with high incomes. Additionally, in the case of households with low incomes, where the price elasticity of demand is highest, prevalence elasticity plays a significant part. Thus, an increase in excise taxes, and implicitly in the price, has a progressive character: the highest part of the financial burden is borne by high-income households. In the case of low-income households, the decrease in consumption comes also from the fact that some people will quit smoking, which would free up part of the family budget to be allocated for more productive purposes.

8. Policy Recommendations

- Given the national and European context, where there are intense concerns regarding the health of the population (with smoking being the main factor in triggering lung cancer, as the National Plan against Cancer also shows), the main goal of public policies regarding tobacco control should be the decrease in consumption due to the resulting health and economic benefits for the general population.
- The increase in excise taxes, and implicitly in the price of cigarettes, is a win-win situation, which would result in both a decrease in consumption and an increase in budget revenues. However, given the increase in the population's income from 2015 until 2021, it can be observed that a 10-percent price increase (while maintaining the historical CAGR of income for the same period) would lead to an increase in consumption, which occurred during the analyzed period. **A minimum 11.5-percent increase in price would be needed if a decrease in consumption is desired.**
- Notably, heated tobacco products currently benefit from a much looser regulation compared to cigarettes (such as indoor smoking, publicity, images on packaging), and thus smokers who are not allowed to smoke cigarettes indoors often smoke heated tobacco products indoors, diminishing the purpose of regulations against tobacco and exposing non-smokers to second-hand smoking. To prevent these harms, heated tobacco products should be subject to the same regulations and taxation as cigarettes.

References

Belotti, F., Deb, P., Manning, W. G., & Norton, E. C. (2015). Twopm: Two-Part Models. The Stata Journal, 15(1), 3-20. https://doi.org/10.1177/1536867X1501500102

Blecher, E. & van Walbeek, C. (2008) An Analysis of Cigarette Affordability. Paris: International Union Against Tuberculosis and Lung Disease. https://assets.tobaccofreekids.org/global/pdfs/en/TAX_Cigarette_affordability_report_en.pdf

Deaton, A., & Ng, S. (1998). Parametric and nonparametric approaches to price and tax reform. *Journal of the American Statistical Association, 93*(443). Applications and Case Studies. https://www.princeton.edu/~deaton/downloads/Parametric_and_Nonparametric_Approaches_Price_Tax_Reform.pdf

Huang, J., Gwarnicki, C., Xu, X., Caraballo, R.S., Wada, R., & Chaloupka, F.J. (2018) A comprehensive examination of own- and cross-price elasticities of tobacco and nicotine replacement products in the U.S. Prev Med. Dec;117:107-114.

Kostova, D., & Dave, D. (2015). Smokeless tobacco use in India: Role of prices and advertising. Social Science & Medicine, 138, 82-90.

Manning, W.G., & Mullahy, J. (2001). Estimating log models: to transform or not to transform? Journal of Health Economics, 20(4), 461–494.

Manning, W.G, Basu, A., & Mullahy, J. (2005) Generalized modeling approaches to risk adjustment of skewed outcomes data. J Health Econ. May;24(3):465-88.

Vladisavljević, M., Zubović, J., Jovanović, O., Đukić, M., & Jolović, N. (2020). How do prices of manufactured cigarettes and roll-your-own tobacco affect demand for these products? Institute of Economic Sciences, Belgrade.

Zubović, J., Vladisavljević, M., Gjika, A., Zhllima, E., Imami, D., Gligorić, D., Mićić, L., Preradović, D., Pepić, A., Prekazi, B., Pula, E., Najdovska, N.T., Mugoša, A., Čizmović, M., Laković, T., Popović, M., Đukić, M., & Jovanović, O. (2019). Impacts of Tobacco Excise Increases on Cigarette Consumption and Government Revenues in Southeastern European Countries [Report]. IES. https://www.tobacconomics.org/files/research/561/Regional-report-2019.pdf

Zubović, J., & Vladisavljević, M. (Eds.). (2019). Impacts of tobacco excise increases on cigarette consumption and government revenues in Southeastern European Countries. http://tobaccotaxation.org/cms_upload/pages/files/Regional-report-2019.pdf

Romanian Government, Ministry of Health. (2023). National plan to combat and control cancer. https://ms.ro/media/documents/Planul_Na%C8%9Bional_de_Combatere_%C8%99i_Control_al_Cancerului_RIQiTXG.pdf

WHO. (2023). WHO report on the global tobacco epidemic, 2023: protect people from tobacco smoke. World Health Organization. https://apps.who.int/iris/handle/10665/372043

Appendices

Appendix 1. Descriptive statistics

Variable	Obs	Mean	Std. dev.	Min	Max
year	105,010	2018	2.000	2015	2021
month	105,010	6.5	3.5	1.0	12.0
Wap (weighted average price)	105,010	16.8	2.1	13.9	20.7
Nrgl (Order number of the household within the dwelling)	105,010	1.0	0.1	1.0	4.0
sex	105,010	1.5	0.5	1.0	2.0
Lunn (month)	105,010	6.4	3.3	1.0	12.0
Ann (year)	105,010	1962.0	18.0	1916.0	2006.0
Nat (nationality)	105,010	1.1	0.4	1.0	5.0
Nive (The last level of education of the highest degree graduated)	105,010	5.9	2.3	1.0	14.0
Stocup (Occupational status in the reference month)	105,010	6.4	4.0	1.0	14.0
Stocupan (Main occupational status in the last 12 months)	105,010	6.3	4.0	1.0	14.0
Idhh (household ID)	105,010	8075.0	4686.0	1.0	16449.0
reg nuts2 (region – nuts2)	105,010	4.2	2.3	1.0	8.0
Mediu (environment)	105,010	2.0	1.0	1.0	3.0
Weight	105,010	262.5	245.2	35.9	3375.0
district	105,010	21.9	13.4	1.0	52.0
age	105,010	56.1	18.0	15.0	100.0
Hsize (Household size)	105,010	4.0	2.8	1.0	27.0
Adultratio	105,010	1.0	0.1	0.1	1.0
Maleratio	105,010	0.4	0.3	0.0	1.0
Maxedu (highest level of education in household)	104,988	3.6	1.4	1.0	6.0
educ avg years (average years of education in an Household**)**	105,010	10.4	2.7	0.0	20.0
Htype (The occupational status of the household's member with the highest occupational status)	105,010	3.3	0.9	1.0	4.0
hhd avg activity	105,010	3.0	1.0	1.0	5.0
urban	105,010	0.5	0.5	0.0	1.0
tot	104,999	1175.0	866.2	0.0	12920.0
Ecig (Household expenditure on cigarettes)	20,657	447.7	332.5	4.0	3464.0
tot1	84,015	1287.0	1369.0	0.3	79698.0
tot2	104,932	1028.0	971.8	0.0	36733.0
mce30 (The income at household level)	105,010	0.6	0.5	0.0	3.4
total cons (total consumption)	105,010	3232.0	2653.0	0.0	83477.0
insd ban (indoor smoking ban)	105,010	0.8	0.4	0.0	1.0
Ncig (The number of cigarettes consumed by the household)	105,010	103.6	268.1	0.0	3442.0

ldcig (A dummy variable that highlights if a household has any expenditure on cigarettes)	105,010	0.2	0.4	0.0	1.0
Disid	105,010	1775.0	1083.0	1.0	3947.0
income grp	105,010	2.0	0.8	1.0	3.0
htdum1 (A set of dummy variables corresponding to each level of occupation)	105,010	0.0	0.1	0.0	1.0
htdum2	105,010	0.3	0.5	0.0	1.0
htdum3	105,010	0.1	0.2	0.0	1.0
htdum4	105,010	0.6	0.5	0.0	1.0
hsize2 (Household size squared)	105,010	24.0	35.5	1.0	729.0
lnp (The natural logarithm transformation of price)	105,010	2.8	0.1	2.6	3.0
lny (The natural logarithm transformation of income)	104,941	-0.9	0.9	-5.5	1.2
lnp2 (The square of log transformation of price)	105,010	7.9	0.7	6.9	9.2
lny2 (The square of the log transformation of income)	104,941	1.7	2.3	0.0	30.0
wap2 (The square of price (wap))	105,010	286.2	71.0	193.3	427.3
mce2 (The square of income (mce30))	105,010	0.6	1.2	0.0	11.5
maleratio2 (The square of male ratio)	105,010	0.3	0.2	0.0	1.0
adultratio2 (The square of adult ratio)	105,010	0.9	0.2	0.0	1.0

Appendix 2. Prevalence elasticity estimations – variables used in level

	(1)	(2)	(3)	(4)	(5)
VARIABLES	idcig	idcig	idcig	idcig	idcig
weighted average price	0.00713	0.115	-0.00690	0.196*	-0.0264***
	(0.00475)	(0.111)	(0.00548)	(0.111)	(0.00479)
weighted average price^2		-0.00347		-0.00552*	
		(0.00314)		(0.00314)	
income	0.311***	0.826***	0.833***	0.301***	1.529***
	(0.0246)	(0.0622)	(0.0621)	(0.0249)	(0.0558)
income^2		-0.190***	-0.192***		-0.389***
		(0.0209)	(0.0209)		(0.0200)
household size	-0.0147***	-0.0237***	-0.0242***	-0.0135***	-0.0296***
	(0.00431)	(0.00449)	(0.00447)	(0.00433)	(0.00447)
males in household	1.193***	1.209***	1.209***	1.193***	1.260***
	(0.0346)	(0.0351)	(0.0351)	(0.0346)	(0.0345)
adults in household	-0.307***	-0.357***	-0.358***	-0.306***	-0.701***
	(0.0779)	(0.0780)	(0.0780)	(0.0779)	(0.0781)
insd_ban		0.0242	0.0466*	0.0288	
		(0.0345)	(0.0277)	(0.0344)	
education level 1	-0.805***	-0.742***	-0.741***	-0.803***	-0.965***
	(0.0724)	(0.0727)	(0.0728)	(0.0724)	(0.0721)
education level 2	-0.253***	-0.204***	-0.203***	-0.252***	-0.295***
	(0.0318)	(0.0323)	(0.0323)	(0.0318)	(0.0318)
education level 4	-0.0430	-0.0740***	-0.0749***	-0.0407	-0.114***
	(0.0265)	(0.0267)	(0.0267)	(0.0265)	(0.0265)
education level 5	-0.0531	-0.0990**	-0.0986**	-0.0570	-0.159***
	(0.0401)	(0.0402)	(0.0402)	(0.0401)	(0.0399)
education level 6	-0.0975***	-0.134***	-0.136***	-0.0927***	-0.182***
	(0.0236)	(0.0239)	(0.0239)	(0.0237)	(0.0239)
household occupation 1	-0.757***	-0.639***	-0.638***	-0.757***	
	(0.111)	(0.111)	(0.112)	(0.111)	
household occupation 2	-0.755***	-0.672***	-0.670***	-0.757***	
	(0.0253)	(0.0273)	(0.0273)	(0.0253)	
household occupation 3	-0.398***	-0.315***	-0.314***	-0.400***	
	(0.0393)	(0.0405)	(0.0405)	(0.0393)	
Constant	-1.647***	-2.632***	-1.591***	-3.258***	-1.357***
	(0.111)	(0.947)	(0.113)	(0.946)	(0.112)
Obs.	104,988	104,988	104,988	104,988	104,988
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					
		Information criteria			
AIC	99,484	99,388	99,387	99,477	100,090
BIC	99,617	99,550	99,540	99,630	100,205

		Link test			
_hatsq z score	-13.04	-13.33	-13.33	-13.02	-14.82
_hatsq p value	0.0000	0.0000	0.0000	0.0000	0.0000
		LM test prob > chi2			
10 groups	0.0000	0.0000	0.0000	0.0000	0.0000
20 groups	0.0000	0.0000	0.0000	0.0000	0.0000
50 groups	0.0000	0.0000	0.0000	0.0000	0.0000
		Collin			
Mean VIF	1.51	91.27	3.19	95.16	3.27
Det(correlation matrix)	0.1037	0.0000	0.0068	0.0001	0.0201

Appendix 3. Prevalence elasticity estimations – variables used in log

	(1)	(2)	(3)	(4)	(5)
VARIABLES	idcig	idcig	idcig	idcig	idcig
ln(weighted average price)	-0.295***	5.438	-0.211**	10.51*	-0.540***
	(0.0943)	(5.376)	(0.0956)	(5.365)	(0.0817)
ln(weighted average price)^2		-0.989		-1.808*	
		(0.941)		(0.940)	
ln(income)	0.398***	0.297***	0.299***		0.465***
	(0.0192)	(0.0271)	(0.0270)		(0.0259)
ln(income)^2		-0.0503***	-0.0500***	-0.127***	-0.0606***
		(0.00959)	(0.00958)	(0.00731)	(0.0101)
household size	-0.0366***	-0.0293***	-0.0296***	0.00148	-0.0366***
	(0.00436)	(0.00454)	(0.00453)	(0.00356)	(0.00451)
males in household	1.235***	1.247***	1.247***	1.254***	1.303***
	(0.0361)	(0.0365)	(0.0365)	(0.0364)	(0.0365)
adults in household	-0.415***	-0.389***	-0.390***	-0.263***	-0.699***
	(0.0780)	(0.0780)	(0.0780)	(0.0775)	(0.0778)
insd_ban	0.0339	0.0116	0.0351	0.0110	
	(0.0282)	(0.0361)	(0.0282)	(0.0361)	
education level 1	-0.536***	-0.452***	-0.451***	-0.427***	-0.543***
	(0.0744)	(0.0764)	(0.0764)	(0.0763)	(0.0762)
education level 2	-0.106***	-0.0751**	-0.0748**	-0.0865***	-0.106***
	(0.0327)	(0.0332)	(0.0332)	(0.0331)	(0.0330)
education level 4	-0.0974***	-0.0904***	-0.0911***	-0.0461*	-0.128***
	(0.0267)	(0.0266)	(0.0266)	(0.0265)	(0.0264)
education level 5	-0.121***	-0.101**	-0.100**	-0.0182	-0.150***
	(0.0398)	(0.0399)	(0.0399)	(0.0391)	(0.0396)
education level 6	-0.171***	-0.137***	-0.138***	-0.0132	-0.183***
	(0.0229)	(0.0237)	(0.0237)	(0.0207)	(0.0236)
household occupation 1	-0.357***	-0.268**	-0.266**	-0.293**	
	(0.115)	(0.116)	(0.116)	(0.116)	
household occupation 2	-0.580***	-0.574***	-0.573***	-0.649***	
	(0.0276)	(0.0277)	(0.0277)	(0.0268)	
household occupation 3	-0.194***	-0.179***	-0.178***	-0.258***	
	(0.0411)	(0.0410)	(0.0410)	(0.0400)	
Constant	-0.0910	-8.450	-0.414	-16.63**	0.858***
	(0.278)	(7.651)	(0.285)	(7.627)	(0.258)
Obs.	104,919	104,919	104,919	104,919	104,919
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					
		Information criteria			
AIC	99,124	99,100	99,099	99,216	99,597

BIC	99,267	99,262	99,252	99,369	99,712
		Link test			
_hatsq z score	-13.1	-11.94	-11.94	-11.18	-9.42
_hatsq p value	0.0000	0.0000	0.0000	0.0000	0.0000
		LM test prob > chi2			
10 groups	0.0000	0.0000	0.0000	0.0000	0.0000
20 groups	0.0000	0.0000	0.0000	0.0000	0.0000
50 groups	0.0000	0.0000	0.0000	0.0000	0.0000
		Collin			
Mean VIF	1.66	724.68	2.42	767.58	2.56
Det(correlation matrix)	0.0506	0.0000	0	0	0.029

Appendix 4. Intensity elasticity estimations – variables used in level

	(1)	(2)	(3)	(4)	(5)
VARIABLES	ncig	ncig	ncig	ncig	ncig
weighted average price	-0.0122***	-0.0667	-0.0255***	0.0253	-0.0259***
	(0.00262)	(0.0586)	(0.00288)	(0.0585)	(0.00259)
weighted average price^2		0.00117		-0.00118	
		(0.00166)		(0.00165)	
income	0.465***	1.044***	1.041***	0.460***	1.127***
	(0.0135)	(0.0306)	(0.0304)	(0.0136)	(0.0273)
income^2		-0.208***	-0.207***		-0.229***
		(0.00969)	(0.00965)		(0.00903)
household size	0.0291***	0.0182***	0.0184***	0.0296***	0.0159***
	(0.00222)	(0.00226)	(0.00225)	(0.00224)	(0.00221)
males in household	0.240***	0.267***	0.267***	0.240***	0.262***
	(0.0198)	(0.0194)	(0.0194)	(0.0198)	(0.0193)
adults in household	0.237***	0.180***	0.181***	0.236***	0.158***
	(0.0436)	(0.0430)	(0.0430)	(0.0437)	(0.0421)
insd_ban		0.0387**	0.0311**	0.0417**	
		(0.0185)	(0.0152)	(0.0187)	
education level 1	-0.289***	-0.222***	-0.222***	-0.287***	-0.256***
	(0.0561)	(0.0538)	(0.0539)	(0.0565)	(0.0518)
education level 2	-0.168***	-0.111***	-0.111***	-0.167***	-0.129***
	(0.0168)	(0.0165)	(0.0165)	(0.0168)	(0.0164)
education level 4	0.0308**	0.00221	0.00247	0.0316**	-0.00281
	(0.0130)	(0.0131)	(0.0131)	(0.0131)	(0.0130)
education level 5	-0.00275	-0.0434**	-0.0435**	-0.00521	-0.0508**
	(0.0199)	(0.0200)	(0.0200)	(0.0199)	(0.0199)
education level 6	0.0106	-0.0282**	-0.0276**	0.0133	-0.0353***
	(0.0115)	(0.0117)	(0.0117)	(0.0116)	(0.0116)
household occupation 1	-0.262***	-0.146*	-0.147*	-0.264***	
	(0.0814)	(0.0766)	(0.0766)	(0.0812)	
household occupation 2	-0.150***	-0.0600***	-0.0606***	-0.151***	
	(0.0127)	(0.0133)	(0.0133)	(0.0126)	
household occupation 3	-0.220***	-0.131***	-0.131***	-0.222***	
	(0.0244)	(0.0244)	(0.0243)	(0.0245)	
Constant	5.630***	6.009***	5.658***	5.305***	5.670***
	(0.0636)	(0.503)	(0.0628)	(0.503)	(0.0621)
Obs.	20,655	20,655	20,655	20,655	20,655
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					
		Information criteria			
AIC	297,363	297,212	297,210	297,362	297,221
BIC	297,475	297,347	297,337	297,489	297,221

		Link test			
_hatsq z score	-1.62	-0.69	-0.69	-1.38	-2.5
_hatsq p value	0.1050	0.4920	0.4910	0.1660	0.0120
		Collin			
Mean VIF	1.51	91.27	3.19	95.16	3.27
Det(correlation matrix)	0.1037	0.0000	0.0068	0.0001	0.0201

Appendix 5. Intensity elasticity estimations – variables used in log

	(1)	(2)	(3)	(4)	(5)
VARIABLES	ncig	ncig	ncig	ncig	ncig
ln(weighted average price)	-0.432***	-1.040	-0.507***	8.930***	-0.471***
	(0.0506)	(2.822)	(0.0498)	(2.928)	(0.0440)
ln(weighted average price)^2		0.0931		-1.516***	
		(0.494)		(0.512)	
ln(income)	0.436***	0.512***	0.512***		0.519***
	(0.0125)	(0.0145)	(0.0144)		(0.0141)
ln(income)^2		0.0375***	0.0374***	-0.0832***	0.0366***
		(0.00647)	(0.00647)	(0.00711)	(0.00632)
household size	0.0217***	0.0154***	0.0154***	0.0735***	0.0145***
	(0.00230)	(0.00226)	(0.00225)	(0.00179)	(0.00221)
males in household	0.323***	0.307***	0.307***	0.313***	0.305***
	(0.0191)	(0.0193)	(0.0193)	(0.0206)	(0.0193)
adults in household	0.182***	0.158***	0.158***	0.406***	0.156***
	(0.0429)	(0.0425)	(0.0425)	(0.0448)	(0.0418)
insd_ban	0.0294*	0.0326*	0.0304**	0.0171	
	(0.0152)	(0.0190)	(0.0151)	(0.0199)	
education level 1	-0.0732	-0.110**	-0.110**	-0.126**	-0.111**
	(0.0478)	(0.0503)	(0.0503)	(0.0539)	(0.0496)
education level 2	-0.0295*	-0.0466***	-0.0466***	-0.0923***	-0.0512***
	(0.0169)	(0.0170)	(0.0170)	(0.0199)	(0.0171)
education level 4	7.37e-05	-0.00550	-0.00545	0.0730***	-0.00659
	(0.0131)	(0.0130)	(0.0130)	(0.0135)	(0.0130)
education level 5	-0.0299	-0.0452**	-0.0452**	0.0978***	-0.0451**
	(0.0198)	(0.0199)	(0.0199)	(0.0199)	(0.0199)
education level 6	-0.00806	-0.0319***	-0.0318***	0.177***	-0.0340***
	(0.0116)	(0.0117)	(0.0116)	(0.0107)	(0.0116)
household occupation 1	0.0765	0.0101	0.00974	-0.0628	
	(0.0694)	(0.0699)	(0.0699)	(0.0798)	
household occupation 2	-0.00581	-0.00702	-0.00715	-0.165***	
	(0.0141)	(0.0137)	(0.0137)	(0.0149)	
household occupation 3	-0.0368	-0.0524**	-0.0525**	-0.203***	
	(0.0248)	(0.0246)	(0.0246)	(0.0254)	
Constant	7.225***	8.270**	7.512***	-7.708*	7.444***
	(0.155)	(4.022)	(0.152)	(4.168)	(0.142)
Observations	20,647	20,647	20,647	20,647	20,647
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					
		Information criteria			
AIC	297,026	297,010	297,008	297,478	297,003
BIC	297,145	297,145	297,135	297,605	297,099

Link test					
_hatsq z score	5.61	1.98	1.97	5.93	1.88
_hatsq p value	0.0000	0.0480	0.0490	0.0000	0.0610
Collin					
Mean VIF	1.66	724.68	2.42	767.58	2.56
Det(correlation matrix)	0.0506	0.0000	0.0087	0.0000	0.029

Appendix 6. Detailed model estimation of the two-part model at income group level

	Group 1		Group 2		Group 3	
VARIABLES	idcig	ncig	idcig	ncig	idcig	ncig
lnp	-0.942***	-0.387***	-0.697***	-0.613***	-0.123	-0.551***
	(0.168)	(0.0886)	(0.155)	(0.0769)	(0.146)	(0.0735)
lny	1.228***	0.796***	0.478***	0.556***	0.327***	0.503***
	(0.183)	(0.120)	(0.172)	(0.0834)	(0.0488)	(0.0257)
lny2	0.131***	0.0993***	-0.114	0.00232	0.0503	0.0499*
	(0.0420)	(0.0297)	(0.101)	(0.0499)	(0.0586)	(0.0282)
hsize	0.0358*	-0.0172	-0.104***	-0.0133***	-0.0150***	0.0229***
	(0.0187)	(0.0106)	(0.00966)	(0.00480)	(0.00535)	(0.00264)
maleratio	1.641***	0.238***	0.960***	0.304***	0.909***	0.434***
	(0.0525)	(0.0292)	(0.0655)	(0.0310)	(0.0980)	(0.0494)
adultratio	-1.021***	-0.0497	-0.804***	-0.0288	-0.185	0.236***
	(0.202)	(0.112)	(0.123)	(0.0647)	(0.123)	(0.0630)
deduc1	-0.505***	-0.133***	0.161	0.350		
	(0.0800)	(0.0434)	(0.305)	(0.268)		
deduc2	-0.103**	-0.0402*	-0.0569	-0.0274	0.318*	-0.0459
	(0.0414)	(0.0214)	(0.0588)	(0.0287)	(0.164)	(0.0930)
deduc4	-0.0295	0.0582	-0.147***	-0.0150	-0.131***	-0.0263
	(0.0808)	(0.0404)	(0.0407)	(0.0203)	(0.0387)	(0.0192)
deduc5	0.0266	0.147**	-0.307***	-0.00730	-0.0897*	-0.0880***
	(0.145)	(0.0746)	(0.0747)	(0.0368)	(0.0508)	(0.0254)
deduc6	-0.0540	0.168***	-0.207***	-0.0595***	-0.152***	-0.0483***
	(0.119)	(0.0544)	(0.0424)	(0.0205)	(0.0313)	(0.0154)
o.deduc1					-	-
Constant	2.643***	7.787***	1.891***	8.167***	-0.800*	7.494***
	(0.644)	(0.349)	(0.503)	(0.250)	(0.449)	(0.230)
Observations	34,937	4,146	34,987	7,538	34,995	8,963

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix 7. Relevant questions in HBS on smoking and for the econometric model

1. The person's gender
2. Month and year of birth
3. The last level of education of the highest degree graduated
4. Main occupational status in the last 12 months
5. The revenues obtained
6. The expenses for the purchase of non-food goods (even if for each non-food product there are 3 columns - the unit of measure, the quantity and the amount spent, in the case of tobacco products data is collected **only** on the amount spent for their purchase)
7. Cigarette expenses
8. Cigar expenses
9. Spending on other tobacco products