SMOKING CZECHS: MODELLING TOBACCO CONSUMPTION AND TAXATION

Karel Janda, Martin Strobl[*]

Abstract

We model the future tobacco consumption, size of smoking population and governmental tax revenues in the Czech Republic. The main model assumption states that smokers determine their future tobacco consumption behaviour as adolescents. Further assumptions make the model applicable to the data from the Czech National Monitoring Centre for Drugs and Drug Addiction. Future teenage smoking rates and average consumption are the inputs to the model; consumption growth coefficients for each age category are estimated using zero-inflated negative binomial regression. Several scenarios are built to model possible developments, including extreme cases. All our scenarios show that all model outcomes are going to grow until 2028 in a very similar pattern. In particular, the projected number of smokers in 2028 is by 4–8% higher than in 2013, the total daily tobacco consumption and tax revenue by 7–26%. This increase is induced by aging of large birth cohorts.

Keywords: smoking, tobacco, cigarettes, consumption, taxation, forecasting
JEL Classification: D12, I12

1. Introduction

Though the harmful effects of smoking on human's body are nowadays commonly known, smoking is still popular in many societies and there has not been a significant decline in smoking rate over the past few years. According to Sovinova *et al*. (2016) the long-term proportion of smokers in Czech adult populations fluctuates between 28% and 32%. According to the OECD (2014), the Czech Republic is its only member, whose percentage change in smoking rate over the period of 2000–2011 was positive (4.7%), while the OECD average was negative (−20.7%). The Czech government introduced a smoking-ban, which

should regulate the sale of tobacco, and a smoking ban in restaurants and bars. Such a measure is common in many EU countries. Just like in case of other risky behaviour (obesity, alcohol, *etc*.) this ban intends to discourage desired choices of consumers, which may (or may not) be based on imperfect information or myopic behaviour and consequently lead to their addiction.

This paper explores possible trajectories of tobacco consumption in the Czech Republic. Our theoretical model applied on empirical data and various development scenarios provide predictions for smoker population size, total daily tobacco consumption and governmental tobacco tax revenues up to 2078. The resulting outlook based on demographic projections can be used to understand the mechanisms of possible transformation in smoking behaviour as well as to see the prospects of tobacco industry and governmental tax revenues in the Czech Republic. The main assumption of the model is that smokers determine their future smoking status and consumption levels as adolescents and are unable to quit later on throughout their life. That implies that the model scenarios and outcomes are always based on future smoking rates and average consumption among teenagers.

Our model further assumes that tobacco consumption proportionally changes with age in a way that is determined as a product of individual's teenage tobacco consumption and a coefficient for his current age category. These parameters are estimated using empirical econometric analysis, namely Zero-Inflated Negative Binomial Model (ZINBM). The analysed data set comes from a survey (from Q4/2012) by the Czech National Monitoring Centre for Drugs and Drug Addiction. The data structure allows to focus on age categories of size of 5 years from the age of 15 up to 65. As the coefficients estimation identifies the determinants of consumption and smoking status, its results might be interesting from a health economist's point of view as well.

While smoking is a common research topic in the Czech Republic, majority of Czech tobacco research papers are very medically oriented (see for example Perinova *et al*. (2016)). Thus, there are only a few relevant studies like Movsisyan *et al*. (2016), which actually use quantitative models to predict causalities, future development, or the structure of the smoking population. Among those Sovinová *et al*. (2008) used the method of smoking-attributable fractions to estimate (*ex-post*) that six years earlier, in 2002, smoking could induce 19% (20,550 deaths) of overall Czech mortality. Interestingly, women represented only around 30% out of these deaths. Levy *et al*. (1997) also studied the Czech smoking related mortality. Their study models future mortality rates for several tobacco policy scenarios using SimSmoke simulations. Spilková *et al*. (2011) analysed smoking data from 2003 using multilevel modelling, which showed that education is negatively related to smoking, while unemployment and divorced status are associated with higher smoking. Moreover, men proved to be more likely to become smokers.

Because the majority of tobacco consumers initiate smoking as teenagers, many research papers focus particularly on this age group as Ross *et al*. (2003) and Powell *et al*. (2005). A good example is the work of Pertold (2009) who compared the past (primary school) and current smoking status of secondary school freshmen in order to estimate the peer effects of their current classmates. The results show that boys are liable to peer pressure but girls are not. However, the peer group in this case is limited to the classroom; in reality it could

differ or be much larger. Tesar (2011) also found the peer effects to be an important factor for youngsters, particularly, the influence of friends and family. According to a survey by Králíková *et al.* (2013) many adolescent initiators are unable to quit their tobacco addiction. Later on, they form the non-teenager percentage of the smoking population as they age.

Kvaček (2011) applied the theory of rational addiction by Becker *et al.* (1988) to study the demand for tobacco and estimated its price elasticity as −0.2. This indicates that Czech tobacco consumers are quite inelastic.

Mendez *et al.* (1998) use similar approach to our paper to predict prospective tobacco prevalence. Their model is more detailed as they account for smoking cessation and differ death rates for smokers and non-smokers. Their approach is specific to tobacco as its addiction is an important factor to reflect in the model.

When it comes to non-addictive goods, many researchers use age-(period)cohort analysis to predict future consumption. This method estimates the effects of age, birth cohort and time period, however, as have Mason *et al.* (1973) pointed out, in this case the confounding effect of these variables can cause problems in the regression analysis. This issue has been addressed by Rentz *et al.* (1991) and further by Carstensen (2007) who proposed incorporating spline functions to estimate age, period and cohort as continuous variables. Some examples of age-cohort model applications are by Mori *et al.* (2004), who used the Bayesian model modification to predict future food consumption in Japan, and by Kerr *et al.* (2004), who studied the age and cohort effects on alcohol consumption in the United States.

2. Model

2.1 Model Description and Assumptions

Our model is based on an assumption that smokers initiate smoking as adolescents and remain smokers throughout their whole life, *i.e.* the smoking prevalence and some base level of consumption (adolescent/initial consumption) is cohort specific, while coefficients determining the current consumption are age specific. Therefore, our model predictions are scenario-based – depending on the characteristics of future adolescents. Our model is given by the following equations:

$$s_{i,t} = P_{i,t} \cdot r_{t-i+1} \,, \tag{1}$$

$$S_t = s_{1,t} + s_{2,t} + \ldots + s_{n,t} = \sum_{i=1}^{n} s_{i,t} \,, \tag{2}$$

$$c_{i,t} = b_i \cdot \alpha_{t-i+1} \cdot s_{i,t} \,, \tag{3}$$

$$C_t = c_{1,t} + c_{2,t} + \ldots + c_{n,t} = \sum c_{i,t} \,, \tag{4}$$

where:

t = time; $t \in Z$; $t = 0$ is the initial period

i = age category; $i \in [\, 1, \ldots, n]$ with $i = 1$ as the youngest

$P_{i,t}$ = size of population of ith age category at time t

r_t = percentage of smokers within the youngest age category at time t

$s_{i,t}$ = population of smokers within age category i at time t

S_t = total population of smokers at time t

α_t = average tobacco consumption of the youngest cohort ($i = 1$) at time t

b_i = age category specific tobacco consumption level (factor); $b_1 = 1$; $b_i > 0$

$c_{i,t}$ = tobacco consumption of i at time t in terms of cigarettes daily

C_t = total tobacco consumption at time t in terms of cigarettes daily

Our model assumes that:

- $i = 1$ is an age category covering teenage years up to 20 years of age.
- Each age category covers the same number of consequent years of age. Each year of age belongs to only one age category. In other words, the period between i and $i + 1$ in years equals the size of one age category, such that between t and $t + 1$ the whole (surviving) population of $P_{i,t}$ moves to $P_{i+1, t+1}$. This dynamics is assumed, however, not modelled, as $P_{i+1, t+1}$ is taken as an exogenous variable (external population projection model outcomes are plugged in).
- Mortality and migration are accounted for through $P_{i,t}$.
- The probability of death is the same for smokers as for non-smokers.
- All smokers develop(ed) their smoking habit in their teenage years ($i = 1$) and are unable to quit later on.
- There is no smoking cessation and initiation after 20 years of age. Equivalently, we can assume that the rates of cessation and initiation are equal instead. As a result of that, aging after 20 years of age influences tobacco consumption volume rather than smoking status (smoker/non-smoker).
- Smokers' consumption in i is determined as the product of a corresponding prede-termined coefficient b_i and their past consumption in $i = 1$ (α_t).

The model can be modified to distinguish gender groups within age categories, which is a desirable property as smoking prevalence and consumption differs significantly between these two subgroups. The upper M or F index denotes the male or female subgroup. Then, the model has the following form (assumptions equivalent to those above apply):

$$s_{i,t}^M = P_{i,t}^M \cdot r_{t-i+1}^M \tag{5}$$

$$s_{i,t}^F = P_{i,t}^F \cdot r_{t-i+1}^F \tag{6}$$

$$s_{i,t} = s_{i,t}^M + s_{i,t}^F \tag{7}$$

$$S_t = s_{1,t} + s_{2,t} + \ldots + s_{n,t} = \sum_{i=1}^{n} s_{i,t} \tag{8}$$

$$c_{i,t}^M = b_i \cdot \alpha_{t-i+1}^M \cdot s_{i,t}^M \tag{9}$$

$$c_{i,t}^F = b_i \cdot \alpha_{t-i+1}^F \cdot s_{i,t}^F \tag{10}$$

$$c_{i,t} = c_{i,t}^M + c_{i,t}^F \tag{11}$$

$$C_t = c_{1,t} + c_{2,t} + ... + c_{n,t} = \sum_{i=1}^{n} c_{i,t} \tag{12}$$

When it comes to modelling the population size of smokers, our model follows the basic principles of a model introduced by Mendez *et al*. (1998) with some simplifications and extensions. As opposed to Mendez *et al*. (1998) we do not account for smoking cessation and do not differentiate between smokers' and non-smokers' death rates because we do not have available sufficiently reliable data specifying these death rates of smokers and non-smokers in the Czech Republic. However, our model takes sex differences and migration into account. Our assumption that future smoking status of an individual is shaped before the age of twenty is rather strong, but backed by empirical evidence of Kralikova *et al*. (2013). We assume the initial level of tobacco consumption (before twenty years of age) to be the main determinant of smoker's future level of tobacco consumption. These are therefore given as factors of the initial consumption of each age group. We estimate these factors empirically using count variable regression model on Czech survey data.

The mechanics and dynamics of the model works as follows: The model takes population projection $P_{i,t}$ for each age category; for every birth cohort, currently in this age category, it finds its original adolescent smoking rate r_{t-i+1} and average tobacco consumption α_{t-i+1} at the period of $t - i + 1$ (when this birth cohort was located in adolescent age category). Then, it multiplies the current age category size $P_{i,t}$ by the smoking rate r_{t-i+1} (smoking rate is assumed to be cohort specific and constant over time) to get the number of smokers $s_{i,t}$ in the age category and the sum for whole population S_t consequently.

Our model assumes that average consumption of a birth cohort changes throughout life, however, only by a factor of the average adolescent tobacco consumption, *i.e.* it equals $\alpha_{t-i+1}b_i$, where α_{t-i+1} is the average adolescent tobacco consumption and b_i is the corresponding coefficient specifying the change in consumption with aging. These are specific to every age category (not cohort), determined by our empirical analysis, and constant over time. Hence, for each age category i the model multiplies the number of smokers in it, $s_{i,t}$, by α_{t-i+1} and b_i in order to obtain the tobacco consumption $c_{i,t}$. And the total tobacco consumption C_t as the sum over all age categories.

With the change to the next period $t + 1$, each birth cohort moves one age category up (to $i + 1$). Then, its size changes to $P_{i+1,t+1}$, and the respective coefficient is now $b_{i+1,}$ the rest of cohort specific variables remain unchanged. This means that the changes of consumption of a cohort are triggered solely by the change of respective coefficients (from to b_i to b_{i+1}) capturing the effect of aging. In this new period, the model again needs to be supplied with the data for the adolescent population (the youngest age category) and its smoking habits as the model is unable to retrieve this information from the previous period. The population

size $P_{1,t+1}$ is provided by the external population projection. The corresponding smoking rate among teenagers and their average consumption for next periods are unknown. In our empirical analysis, we vary these variables according to our scenarios in order to see how the model results change. The above described mechanism of the model works successively for each period, such that the outcomes of the model are only dependent on the characteristics of smoking behaviour of future teenagers and their cohort size.

2.2 Estimation of the Parameters

Using a count variable in a standard OLS linear regression is not appropriate because assumptions like homoscedasticity and normality of errors might be violated. Possible alternatives suitable for this kind of data are the Poisson and negative binomial regression models. A common issue when using the Poisson regression is that equidispersion, the equality of conditional mean and variance $Var(y_i \mid \mathbf{x_i}) = E(y_i \mid \mathbf{x_i})$, is assumed in Poisson distribution and the data may (and very often they do) actually violate this rather strong assumption and exhibit overdispersion: $Var(y_i \mid \mathbf{x_i}) > E(y_i \mid \mathbf{x_i})$. This can underestimate the standard errors. Overdispersion can result from some omitted heterogeneity or due to state dependence. State dependence occurs when the observed events are not independent (*e.g.* smoking the second cigarette is not an independent event because the individual had to smoke the first before, thus the likelihood of smoking the first cigarette and the likelihood of smoking two cigarettes are not independent). Though there exists a Poisson model modification for overdispersed data, using negative binomial model which accounts for overdispersion is a more common practice. Running the likelihood-ratio test compares both models and checks for overdispersion. Hence, one can easily decide which model suits the data better.

As mentioned, the negative binomial model, unlike the Poisson model, accounts for overdispersion. This feature originates from the form of the variance of the NB2 model (notation from Cameron *et al.* (1986)), which is the standard and the most common form of the negative binomial model. Its variance function has the following form:

$$Var(y_i \mid \mathbf{x_i}) = \mu + \alpha\mu^2, \tag{13}$$

where α is the overdispersion (or heterogeneity) parameter. For example, the other (less used) version of the model, denoted NB1, has variance of $Var(y_i \mid \mathbf{x_i}) = \mu + \alpha\mu$. Notice, that if the dispersion parameter α in any of the models is zero, the model variance has properties of the Poisson model ($Var(y_i \mid \mathbf{x_i}) = E(y_1 \mid \mathbf{x_i})$). NB2 is adopted (as default option) by all major statistical programmes.

Having a count data in a regression can cause several problems, where many of them are caused by zero counts. The one that is relevant to our analysis is the inflation of zeros, *i.e.* the fact that zeros can be generated by a different process (decision) than the positive values. As in our case, where positive counts are reported only by smokers whilst non-smokers report zero. The key is to distinguish between the structural zeros and zeros of those participating in the process that generates the discrete values of the count variable. For this purpose we use the zero-inflated negative binomial model (ZINBM). Though it is

possible to use probit with ZINB, we choose to use logit as it is a more common practice. Estimating ZINB is a two step procedure consisting of estimating logit model in the first stage and estimating NB2 version of the negative binomial model in the second stage. Binary regression (logit) is used to model the decision of the individual to become a smoker and NB2 count regression is used for estimation of the smoking intensity.

Notice, that the log-likelihood function must be then different for cases when $y = 0$ and $y > 0$. The log-likelihood functions for ZINB with logit are given by Hilbe (2011) as:

$$\text{if } y = 0 : \ell(\alpha, \beta) = \Sigma_{i=1}^{n} \{\ln(1 / (1 + \exp(-x_i \beta_1)))$$

$$+ 1 / [1 + \exp(x_i \beta_1)](1 / [1 + \alpha \exp(x_i \beta)])^{\frac{1}{\alpha}}\}; \tag{14}$$

$$\text{if } y > 0 : \ell(\alpha, \beta) = \Sigma_{i=1}^{n} \{\ln(1 / [1 + \exp(-x_i \beta_1)]) + \ln\Gamma(\frac{1}{\alpha} + y_i) - \ln\Gamma(y_i + 1)$$

$$- \ln\Gamma(1\alpha) + (1\alpha)\ln(1 / [1 + \alpha \exp(x_i \beta)])$$

$$+ y_i \ln(1 - 1 / [1 + \alpha \exp(x_i \beta)])\} \tag{15}$$

Where β_1 stands for the regression coefficients for the binary part, whereas β for the count part of the model.

For the interpretation of ZINB model, the second stage NB2 part is of primary interest. Because the ZINB model has a log-form, the beta-parameter estimates represent the differences in expected counts. For a one-unit change in p^{th} predictor x_p (where $1 \leq p \leq n$) this is the difference in predicted counts:

$$\log(\hat{y} \mid \mathbf{x'}) - \log(\hat{y} \mid \mathbf{x}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ... + \hat{\beta}_p(x_p + 1) + ... + \hat{\beta}_n x_n$$

$$- (\beta_0 + \beta_1 x_1 + ... + \beta_p x_p + ... + \hat{\beta}_n x_n) \tag{16}$$

$$= \beta_p \tag{17}$$

Where $\mathbf{x}$ is the set of explanatory variables, $\mathbf{x'}$ is the set of explanatory variables after the change in p^{th} variable, $\hat{y} \mid \mathbf{x}$ is the predicted value of $\hat{y}$ from the regression of y on $\mathbf{x}$, and $\hat{y} \mid \mathbf{x'}$ is the predicted value of $\hat{y}$ from the regression of y on $\mathbf{x'}$. After exponentiation we get an Incidence Rate Ratio (IRR):

$$\text{IRR}_p = (y \mid \mathbf{x'}) / (y \mid \mathbf{x}) = e^{\hat{\beta}_p} \tag{18}$$

IRR denotes the ratio of incident rates for different values of a given predictor.

2.3 Forecasting Taxation Revenues

Since our model predicts tobacco consumption and we assume that taxation revenue changes proportionally, then the model outcomes can be expressed in terms of future taxation revenues. However, predicting future revenues from the tax on tobacco is a troublesome task as the level of taxation might change in the future and the model is unable to capture

these changes effectively. It models all predictions assuming the present price level and taxation policy. Hence, the outcome should be considered only as a qualified approximation of the future state. Here is the key equation showing very simple relation of tax revenues to tobacco consumption:

$$T_t = T_{t_0} \cdot C_t / C_{t_0} \tag{19}$$

where T_{t_0} is the present tobacco tax revenue. In our empirical analysis, the input for T_{t_0} is the 2013 tax revenue of CZK 46.82 bn. More detailed description of Czech tobacco tax policies provide Shirane *et al.* (2012).

3. Data

In our analysis we use EUROPOP2013 (European Population Projections, base year 2013) population projections from 2013 by EUROSTAT. It provides detailed projections specifically for the Czech Republic, for both genders separately and for every single year of age. Its horizon is the year of 2080.

We use a 2012 (Q4) survey collected by SCC (Czech market research company) for the Czech National Monitoring Centre for Drugs and Drug Addiction using the method of stratified sampling in all parts of the Czech Republic. This survey is run periodically every 4 years. Unfortunately, the questions differ substantially from those in previous version from 2008 and the sample is not the same as well. For example, the amount of cigarettes consumed daily were not reported prior to 2012 survey, only the smoking status (smoker/non-smoker). Also, the definition of categorical variables was slightly different (inconsistent categorization) and the survey did not provide as much detailed background information about the respondent. Thus, we could not use data from past years and form a time series. Hence, we use only the data from 2012 for cross-sectional analysis.

The purpose of this survey is to analyse the consumption of drugs (not only tobacco and alcohol but illegal ones as well) and its motivation on an individual level. All of the data is self-reported. While the raw dataset consists of 2,135 observations, we use only 1,512 of them remaining after our careful data-cleaning procedure. The reason is that many respondents left some parts of the questionnaire blank or their responses simply did not make any sense (mistakes, nonsense). In the regression model we use 38 variables described in Table 1.

Most of the listed variables are socio-demographic - describing the position of the individual in the society, his social-economic status (through education, income, *etc.*), or his personal characteristics (age, gender, *etc.*). Besides that, there is one variable, which allows to observe respondent's health behaviour other than smoking - *alcabuse.* It is a dummy that takes the value of one if the respondent is an alcoholic. We generated *Alcabuse* variable from the original statistics (which specified the frequency of drinking) using the following rule: Abusive drinkers (alcoholics) are determined by consuming five or more drinks during one occasion in a single month (or more frequently). The definition might seem broad, but the responses are possibly underrated and that should offset it.

Table 1 | Variables

Variable	Type	Description
cigsdaily	count	number of cigarettes smoked daily
smokes	dummy	= 1 if respondent smokes
alcabuse	dummy	= 1 if respondent is an alcoholic
male	dummy	= 1 if male
age	count	the age of respondent
b_1–b_6	dummy	Age categories – see Table 2
unemp	dummy	= 1 if unemployed
student	dummy	= 1 if student
maternity	dummy	= 1 if on maternity leave
retired	dummy	= 1 if retired
disabled	dummy	= 1 if disabled
$msize_1$–$msize_6$	dummy	Municipality size
edu_1–edu_5	dummy	Highest earned education
inc_1–inc_7	dummy	Income group

Source: Authors

Table 2 | Age categories

	Frequency	Per cent	Age
b_1	96	6.35	15–19
b_2	195	12.90	20–24
b_3	181	11.97	25–29
b_4	237	15.67	30–34
b_5	167	11.04	35–39
b_6	125	8.27	40–44
b_7	107	7.08	45–49
b_8	111	7.34	50–54
b_9	126	8.33	55–59
b_{10}	167	11.04	60–64

Source: Authors

Variables b_1–b_{10} are dummies corresponding to age categories from the theoretical model. Their estimated incidence rate ratios will play the role of model coefficients.

Dummy variable *smokes* is the binary dependent variable for the binary part of the Zero-Inflated Negative Binomial Model (ZINBM), while *cigsdaily* is the dependent variable for the negative binomial part of the ZINB model. Their types correspond to the approach to analysing the tobacco consumption – the first one as a dummy represents the extensive margin (smoke or not to smoke), while the other one the intensive margin (cigarettes *per* day conditional on being a smoker). The histogram below illustrates the distribution of *cigsdaily* among smokers (zero values omitted) in the sample. Apparently, the most common answers were ten and twenty cigarettes daily, most likely because of rounding.

Figure 1 | Frequency of Cigarettes/Day in the Subsample of Smokers
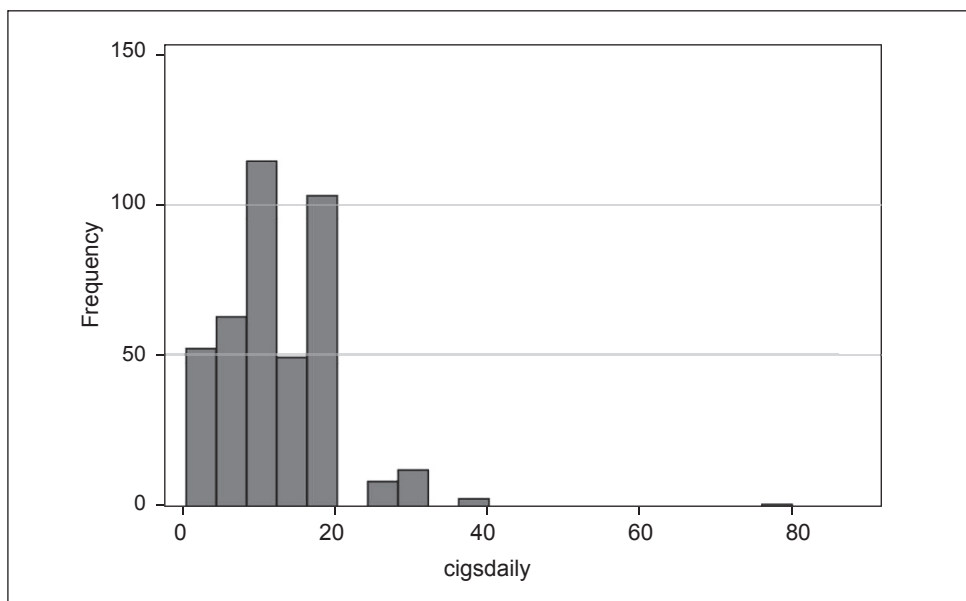
Source: Authors

4. Results

4.1 Consumption Coefficients Estimation

The ZINB regression model was estimated using MLE in STATA with *cigsdaily* as a dependent variable. The estimates of the logit part of the model have only auxiliary function to NB2 estimation, are of no interpretative value and actually vary from a standard logit results as currently a modified log-likelihood function is incorporated. The insignificant variables are omitted from the logit part including the constant term. Logit was estimated with variable *age* instead of b_1–b_{10} dummies for simplicity, *age* is significant at 5% level.

The model estimation with fully specified NB2 part (Table 11 in Appendix) shows that income, municipality size, retirement, unemployment, maternity, and disability are not significant predictors of tobacco consumption at the 5% level. Resulting model (Table 3) confirms significance for all the other variables (jointly for categorical) as well as for the intercept at the 5% significance level and *student* at 10%. The corresponding calculated IRRs supplement the estimated coefficients in Table 3.

The positive coefficient of α (the heterogeneity parameter) of 0.2429 proves that some over dispersion is really present. Though, careful testing is still necessary. The likelihood ratio test between this ZINB model and its Poisson alternative, ZIP (zero-inflated Poisson), has the following form:

$$LR = -2\ell_z IP + 2\ell_z INB \tag{20}$$

where $\ell_z IP$ is the log-likelihood value of the model preferred under null hypothesis (ZIP) and $\ell_z INB$ is of the other one (ZINB). *LR* follows chi-squared distribution. In our case, the test results in a p-value of zero, which implies that ZINB suits the data better.

The Vuong test Vuong (1989) is a test comparing the ZINB model to the standard NB2 model. It checks for significant difference between the fitted values of models. The Vuong test forms a z-statistic

$$Z = \sqrt{n}\mu / \sigma_i \tag{21}$$

where μ is the mean of u_i, σ_i is the standard deviation of u_i and

$$u_i = \ln[(\Sigma_i P_{NB2}(y_i \mid x_i)) / (\Sigma_i P_{ZINB}(y_i \mid x_i))]. \tag{22}$$

The test compares the probabilities $P(y_i \mid x_i)$ of outcome y_i given x_i. Under the null hypothesis, ZINB is the better model. The test follows the normal distribution. With a z-statistic of 12.67 and corresponding p-value of zero ZINB is again the better of the two.

These tests confirm that choosing ZINB model among other count models was the right choice. This is further confirmed by our ZINB regression where $log(\alpha)$, the logarithm of the dispersion parameter, is statistically different from minus infinity at the 99% significance level. If it were not the case, Poisson model would be more appropriate.

Let's focus on the IRRs and interpret our estimation results as this might be interesting for health economists. Among those who smoke, being an abusive drinker increases tobacco consumption by 21.75% (by a factor of 1.2175), holding all other factors constant. Similarly, being a male by a factor of 1.3749 (37.49%). The effects of indicators of education are very similar. Student's volume consumed is smaller by a factor of 0.7347 (−26.53%) than non-student's. The respective factors for *educ*2–5 are 0.6985 (−30.15%), 0.6613 (−33.87%), 0.6423 (−35.77%) and 0.5886 (−41.14%) compared to *educ*1 level.

Table 3 | Estimation Results: ZINB

Variable	IRR	Coefficient	(Std. Err.)
alcabuse	1.2175	0.197 **	(0.059)
male	1.3749	0.318 **	(0.062)
b_2	1.1300	0.122	(0.175)
b_3	1.3381	0.291	(0.195)
b_4	1.2966	0.260	(0.197)
b_5	1.2454	0.219	(0.200)
b_6	1.2124	0.193	(0.207)
b_7	1.5759	0.455 *	(0.208)
b_8	1.6906	0.525 **	(0.203)
b_9	1.6262	0.486 *	(0.204)
b_{10}	1.4531	0.374 †	(0.205)
student	0.7347	−0.308 †	(0.158)
edu_2	0.6985	−0.359 **	(0.099)
edu_3	0.6613	−0.414 **	(0.102)
edu_4	0.6423	−0.443 *	(0.191)
edu_5	0.5886	−0.530 **	(0.135)
Intercept		2.318 **	(0.174)
Inflate (logit): Table 12 in Appendix			
log (a)		−1.415	(0.100)
a		0.2429	(0.024)
N (zero/ nonzero)	1,512 (1,085/427)		
Log-likelihood	−2,209.77		
$\chi^2_{(8)}$	98.11		

Note: Significance levels: † : 10%   * : 5%   ** : 1%
Source: Authors

The IRRs of b_1–b_{10} show that the age influence is not linear, these are the coefficients sought for $b_1, \ldots, b_{10}$. Though some coefficients are insignificant, b_1–b_{10} are jointly significant at the 10% level and thus applicable in the theoretical model. These IRRs along with corresponding age categories are once again for clarity reported in the following table:

Table 4 | Age Categories and Corresponding Coefficients

Age category (i)	Age	b_i
1	15–19	1
2	20–24	1.1300
3	25–29	1.3381
4	30–34	1.2966
5	35–39	1.2454
6	40–44	1.2124
7	45–49	1.5759
8	50–54	1.6906
9	55–59	1.6262
10	60–64	1.4531
11	65+	1.4531

Source: Authors

Our data do not allow us to predict b_{11} for elderly aged 65+. Therefore, we assume that, on average, people turning 65 do not change their consumption behaviour and sustain this consumption level until their death, *i.e.*, $b_{10} = b_{11}$.

Because we work only with present data, the values of r_{t-i+1} and α_{t-i+1} for age cohorts $i = \{2, ..., 10\}$ are not defined and have to be computed or approximated. For r_{t-i+1} the situation is quite clear as the model assumes the age cohort's percentage of smokers stays constant and thus r_{t-i+1} equals the current percentage at time $t = 0$ in given age category. In case of α_{-i+1} ($t = 0$ because this is an initial model setting) we use the assumption that consumption follows the pattern given by b_1, …, b_{10}. Then, the following formula applies:

$$\alpha_{-i+1} = \overline{c}_i / b_i \qquad (23)$$

where $\overline{c}_i$ is average consumption within i at time t_0. The derived input data are summarized in the following Table 5.

Again, the absence of respondents aged 65+ (in $i = 11$) in the dataset has to be fixed. This time we plug in the results from Sovinová *et al.* (2014) survey, where 21.10% of respondents aged 65+ were smokers. The value of corresponding α_{-10} is based on an additional assumption that

$$\alpha_{-10} = \alpha_{-9}. \qquad (24)$$

The actual numbers with M/F indexes for $i = 10$ differ as α_{-10} ignores gender differences in $i = 11$ and counts the population average. This number is consistent with the findings of Sovinová *et al.* (2014).

This problem was caused by the structure of the dataset, but the missing coefficients and descriptive statistics are only the consequences of a much larger problem – the disproportional size of age group $i = 11$, which violates the assumption of having age categories of the same size. The model is able to model behaviour from 15 to 64, however, later on it lacks the ability to capture any changes and it has to rely on approximation of some constant rates. This artificial limitation and assumption violation can cause some bias in predictive power of the model. Unfortunately, the data does not allow any other solution than the one introduced above. A more thorough and detailed survey data would be required for any future improvements.

Table 5 | Model Inputs at Time t_0

i	r_{-i+1}^M	r_{-i+1}^F	α_{-i+1}^M	α_{-i+1}^F
1	0.3243	0.2373	9.08	9.07
2	0.3100	0.2421	10.31	8.93
3	0.4444	0.2900	10.59	7.32
4	0.3254	0.2432	11.12	7.31
5	0.3377	0.1778	13.09	5.87
6	0.2778	0.2830	11.59	7.09
7	0.3000	0.2807	11.00	7.42
8	0.3750	0.2727	11.18	7.18
9	0.3065	0.2031	11.94	7.14
10	0.3117	0.1556	9.98	7.96
11	0.2110	0.2110	9.24	9.24

Source: Authors

4.2 Model Scenarios

The outcomes of our model are dependent on future teenage smoking characteristics. This means that several models with different inputs must be built to capture all possible future developments. Firstly, we analyse the smoking population and daily consumption in 2013 (t_0). Secondly, we focus on possible developments of teenage smoking rates and consumption. In this respect, we consider 4 possible period-to-period changes by ±5% and ±10%. The time interval between two periods is 5 years, starting with 2013. These scenarios have this property: The positive development (+5% and +10%) scenarios' parameters grow with time at an increasing pace, while negative development

(−5% and −10%) scenarios' parameters decline with time at a decreasing pace. This model property is intentional and based on the studies of peer influence among adolescents (*e.g.* Pertold (2009)). Those proved that peer influence is an important determinant of smoking status, that implies the higher the prevalence the higher peer pressure to smoke, the lower prevalence the lower peer pressure but some compulsive smokers still remain present. This makes perfect sense for modelling smoking rates, however, not for average consumption as intensive smoking does not make existing smokers to smoke more. Therefore, the average teenage consumption will be studied through different scenarios based on data from Table 5.

4.2.1 The Initial State in 2013

The model estimates the total population of smokers in the Czech Republic in 2013 as 2.41 million. Table 6 provides more detail on the distribution within age categories. The estimated daily consumption of cigarettes is 31.72 million. See Table 7 for more detail. The tax revenue according to the Czech Ministry of Finance was CZK 46.82 billion.

Table 6 | Smokers in 2013

i	Age	Smokers	Men	Women
1	15–19	143,844	84,828	59,016
2	20–24	182,697	104,500	78,198
3	25–29	262,915	162,827	100,089
4	30–34	232,176	136,222	95,955
5	35–39	243,497	162,512	80,985
6	40–44	209,065	106,441	102,624
7	45–49	206,383	108,865	97,517
8	50–54	205,687	119,863	85,825
9	55–59	185,554	110,018	75,536
10	60–64	169,100	109,286	59,815
11	65+	372,967	–	–
Total		2,413,887		

Source: Authors

Table 7 | Daily Cigarettes Consumption in 2013

| i	Age	Daily Consumption	Men	Women
1	15–19	1,305,885	770,524	535,361
2	20–24	2,005,692	1,216,916	788,776
3	25–29	3,286,889	2,306,711	980,178
4	30–34	2,873,377	1,963,583	909,794
5	35–39	3,242,397	2,650,198	592,199
6	40–44	2,378,063	1,495,494	882,569
7	45–49	3,026,732	1,886,996	1,139,736
8	50–54	3,307,314	2,265,977	1,041,338
9	55–59	3,014,048	2,136,670	877,378
10	60–64	2,276,783	1,584,643	692,140
11	65+	5,005,614	–	–
Total		31,722,794		

Source: Authors

4.2.2 Status Quo

Suppose the current teenage smoking rate (32.43% for men and 23.73% for women) and average consumption (9.1 cigarettes) are steady-states. Then, there will be no changes in these parameters in future. The projected development of smokers in the Czech population is depicted by Figure 2, the projected consumption in Figure 3 and taxation revenues in Figure 4. For numerical results, see Table 8.
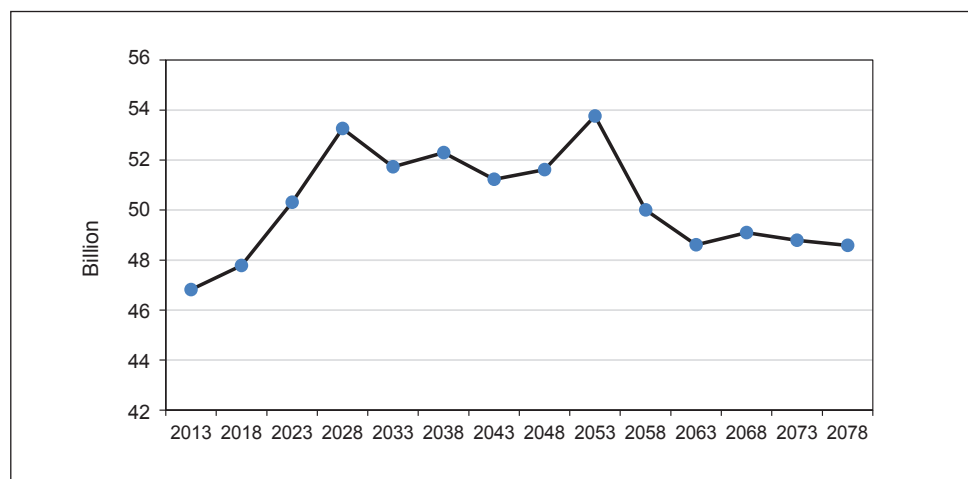
Figure 2 | Status quo Scenario: Smokers
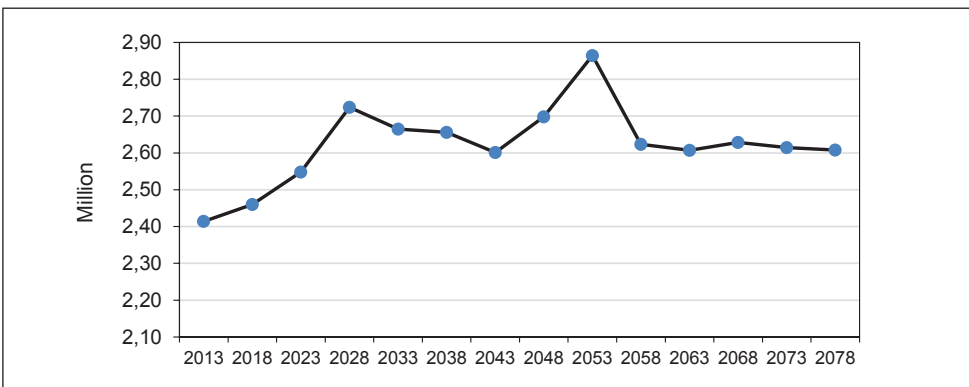
Source: Authors
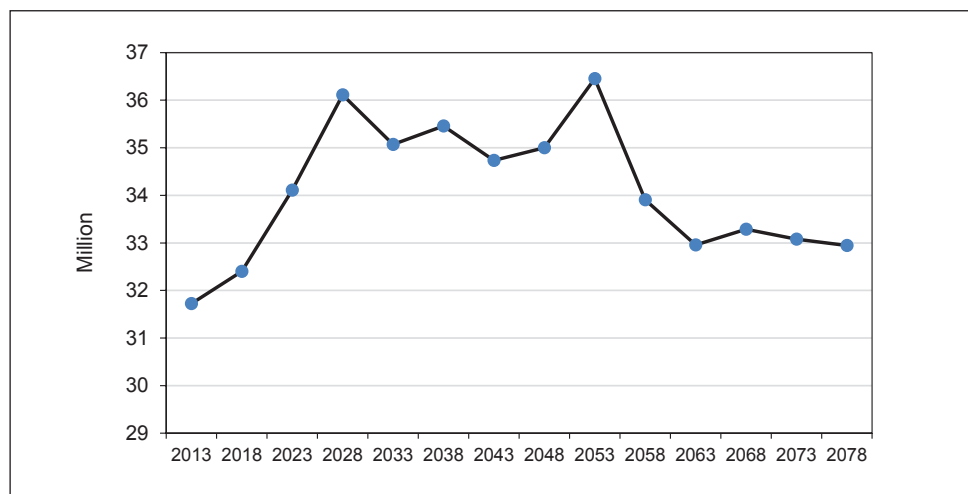
Source: Authors

Source: Authors

All of the charts follow a similar pattern – a boost until 2028, a steady decline, another surge peaking in 2053 followed by a plunge. The relation of consumption to tax revenue is proportional, hence, the figures have the same shape. On the other hand, relation of consumption to number of smokers is not so trivial and the plots actually differ. Let's focus on the growth in 2013–2023. The overall daily consumption of cigarettes grew by 7.44%, while the number of smokers only by 5.55%,

Table 8 | Status quo Scenario: Summary Data

Year	Smokers	Consumption	Tax revenue (CZK)
2013	2,413,887	31,746,646	46,820,000,000
2018	2,460,308	32,400,922	47,784,928,117
2023	2,547,749	34,109,318	50,304,471,881
2028	2,723,743	36,111,068	53,256,655,761
2033	2,664,696	35,071,554	51,723,579,196
2038	2,656,084	35,456,392	52,291,139,060
2043	2,601,042	34,737,000	51,230,178,440
2048	2,697,990	34,999,844	51,617,820,895
2053	2,864,609	36,453,267	53,761,331,904
2058	2,623,438	33,905,238	50,003,495,530
2063	2,607,316	32,959,754	48,609,093,293
2068	2,628,833	33,289,395	49,095,248,728
2073	2,614,761	33,081,122	48,788,087,615
2078	2,607,917	32,945,002	48,587,337,699

Source: Authors

4.2.3 Teenage Smoking Rate Scenarios

These scenarios assume that the percentage of smokers in the group of teenagers varies, while average teenage tobacco consumption sustains its initial level from 2013. The model predicts ±5% and ±10% period-to-period changes. These mechanisms can be expressed as:

$$r_{t+1} = r_t \cdot (1 \pm 0.05) \ \text{ or } \ _1r_{t+} = r_t \cdot (1 \pm 0.10), \ \forall t \in Z \} \ .$$

Figure 5 | Teenage Rate Scenarios: Smokers
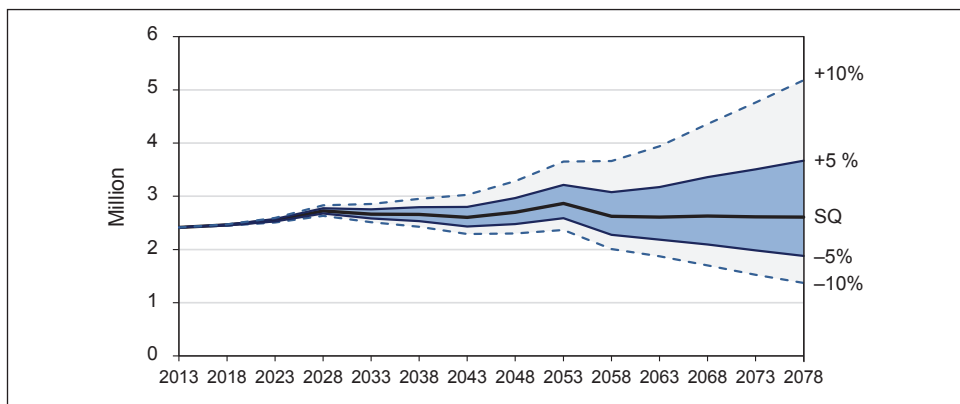
Source: Authors

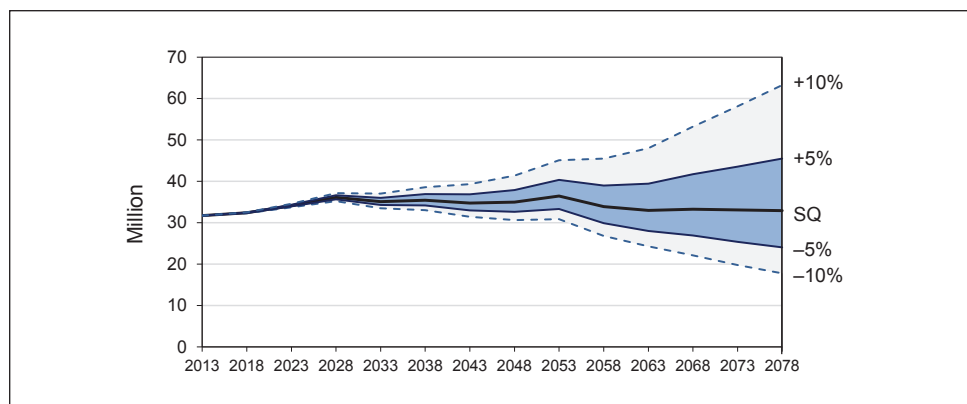Figure 5 shows the predicted smoker population size with the status quo scenario (as SQ) and period-to-period change ±5% and ±10% scenarios. Similarly, the predicted consumption is described in Figure 6 and tax revenues in Figure 7.

Figure 6 | Teenage Smoking Rate Scenarios: Total Daily Consumption

Source: Authors

Figure 7 | Teenage Smoking Rate Scenarios: Tax Revenue (in CZK)
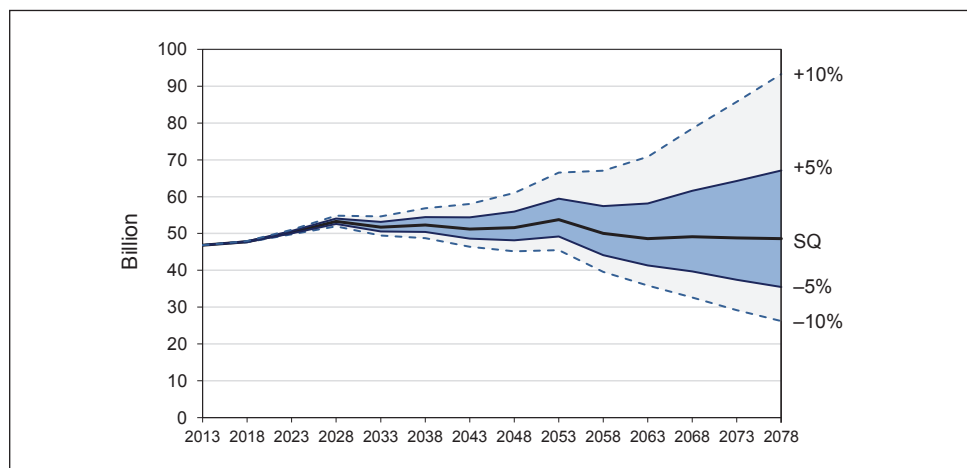
Source: Authors

Complete model predictions data are summarized in Table 9. The numbers indicate that number of smokers, daily consumption as well as tax revenues are going to increase within the next decade in any of these scenarios. These variables will ever reach their initial value from 2013 and further fall only in scenarios −5% and −10%. The drop below the initial state is predicted in period 2038–2043 at the earliest.

Table 9 | Teenage Smoking Rate Scenarios: Model Predictions

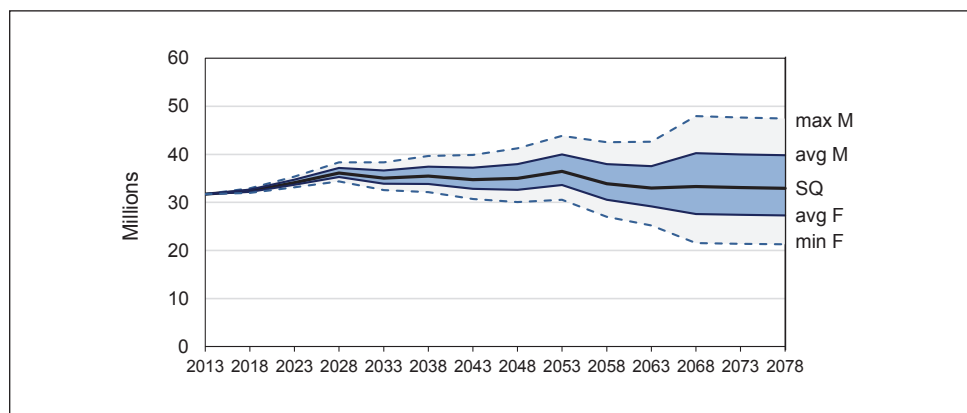|		2013	2018	2023	2028	2033	2038	2043	2048	2053	2058	2063	2068	2073	2078
Number of smokers	SQ	2 413 887	2 460 308	2 547 749	2 723 743	2 664 696	2 656 084	2 601 042	2 697 990	2 864 609	2 623 438	2 607 316	2 628 833	2 614 761	2 607 917
	−10%	2 413 887	2 447 212	2 504 882	2 631 581	2 510 865	2 428 932	2 291 753	2 298 745	2 363 974	2 008 649	1 871 699	1 702 366	1 524 379	1 370 939
	−5%	2 413 887	2 453 760	2 525 934	2 676 028	2 583 822	2 534 888	2 433 681	2 478 901	2 585 828	2 275 931	2 185 470	2 095 782	1 980 647	1 878 411
	5%	2 413 887	2 466 856	2 570 327	2 774 855	2 754 105	2 794 320	2 797 873	2 963 861	3 214 480	3 075 185	3 175 113	3 357 591	3 505 873	3 668 302
	10%	2 413 887	2 473 404	2 593 669	2 829 494	2 852 693	2 951 542	3 028 705	3 285 684	3 652 692	3 661 652	3 938 912	4 359 103	4 761 723	5 180 432
Daily consumption	SQ	31 746 646	32 400 922	34 109 318	36 111 068	35 071 554	35 456 392	34 737 000	34 999 844	36 453 267	33 905 238	32 959 754	33 289 395	33 081 122	32 945 002
	−10%	31 746 646	32 286 978	33 739 098	35 217 200	33 514 990	33 037 905	31 438 979	30 637 175	30 843 668	26 822 082	24 314 396	22 136 730	19 798 206	17 775 855
	−5%	31 746 646	32 346 424	33 938 403	35 660 909	34 274 018	34 169 756	32 975 876	32 626 007	33 350 889	29 919 937	28 028 410	26 909 260	25 401 131	24 049 711
	5%	31 746 646	32 465 316	34 357 796	36 641 347	36 030 495	36 915 026	36 883 609	37 916 186	40 331 996	38 966 348	39 423 316	41 750 417	43 556 336	45 504 318
	10%	31 746 646	32 524 762	34 577 884	37 180 430	37 039 733	38 564 568	39 339 478	41 389 835	45 127 200	45 477 389	48 023 923	53 233 196	58 130 567	63 254 834
Tax revenue	SQ	46 820	47 785	50 304	53 257	51 724	52 291	51 230	51 618	53 761	50 003	48 609	49 095	48 788	48 587
	−10%	46 820	47 617	49 758	51 938	49 428	48 724	46 366	45 184	45 488	39 557	35 859	32 647	29 198	26 216
	−5%	46 820	47 705	50 052	52 593	50 547	50 394	48 633	48 117	49 186	44 126	41 336	39 686	37 462	35 469
	5%	46 820	47 880	50 671	54 039	53 138	54 442	54 396	55 919	59 482	57 468	58 142	61 574	64 237	67 110
	10%	46 820	47 968	50 996	54 834	54 626	56 875	58 018	61 042	66 554	67 070	70 826	78 508	85 731	93 288

Note: * Consumption is expressed in number of cigarettes, tax revenues in millions of CZK
Source: Authors
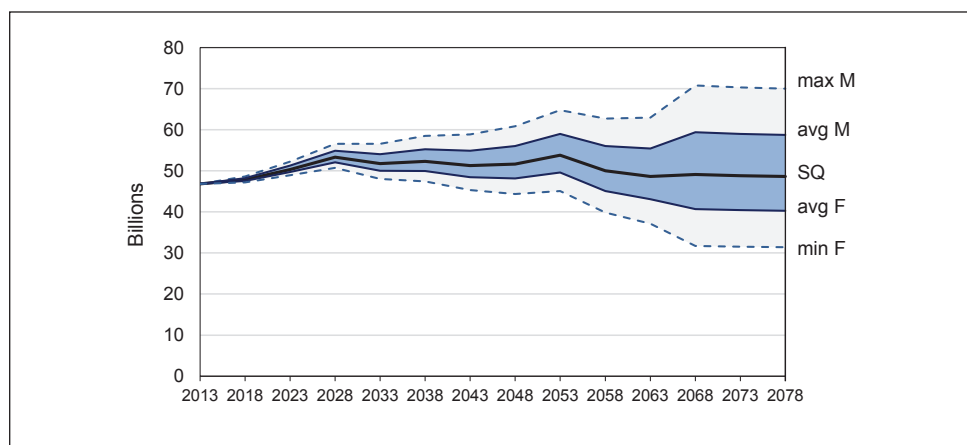
4.2.4 Teenage Tobacco Consumption Scenarios

Now, let's assume that the smoking rate among adolescents stays constant at its current rate and only consumption level changes. Interestingly, if its development mechanism were analogical to those in the previous section (5% / 10% period-to-period growth/decline), the resulting predictions and charts would be identical to those in the previous section. If current teen smokers increase the tobacco consumption by 30%, it is the same as if there were 30% more teen smokers, while they all keep their initial consumption level. Nevertheless, as have been discussed before, this approach is not realistic and we prefer to base the scenarios on empirical observations from Table 5.

Figure 8 | Teenage Consumption Scenarios: Total Daily Consumption

Source: Authors

Figure 9 | Teenage Tobacco Consumption Scenarios: Tax Revenue (in CZK)

Source: Authors

Table 10 | Teenage Consumption Scenarios: Model Predictions

		2013	2018	2023	2028	2033	2038	2043	2048	2053	2058	2063	2068	2073	2078
Daily consumption	SQ	31 722 794	32 400 922	34 109 318	36 111 068	35 071 554	35 456 392	34 737 000	34 999 844	36 453 267	33 905 238	32 959 754	33 289 395	33 081 122	32 945 002
	max M	31 722 794	32 931 790	35 386 973	38 351 816	38 314 951	39 632 162	39 874 276	41 230 184	43 868 957	42 506 746	42 647 887	47 957 533	47 650 321	47 448 162
	avg M	31 722 794	32 655 925	34 735 323	37 189 140	36 634 152	37 441 795	37 200 776	37 975 766	39 985 448	37 984 418	37 551 651	40 242 652	39 984 862	39 815 223
	avg F	31 722 794	32 202 995	33 665 408	35 280 194	33 874 524	33 845 529	32 811 274	32 632 482	33 609 287	30 559 408	29 184 367	27 575 933	27 399 284	27 283 041
	min F	31 722 794	31 985 898	33 152 578	34 365 202	32 551 784	32 121 774	30 707 306	30 071 349	30 553 078	27 000 467	25 173 776	21 504 545	21 366 789	21 276 139
Tax revenue	SQ	46 820	47 821	50 342	53 297	51 762	52 330	51 269	51 657	53 802	50 041	48 646	49 132	48 825	48 624
	max M	46 820	48 604	52 228	56 604	56 549	58 494	58 851	60 852	64 747	62 736	62 944	70 781	70 328	70 029
	avg M	46 820	48 197	51 266	54 888	54 069	55 261	54 905	56 049	59 015	56 062	55 423	59 395	59 014	58 764
	avg F	46 820	47 529	49 687	52 070	49 996	49 953	48 426	48 163	49 604	45 103	43 074	40 700	40 439	40 267
	min F	46 820	47 208	48 930	50 720	48 044	47 409	45 321	44 383	45 094	39 850	37 154	31 739	31 535	31 402

Note: * Consumption is expressed in number of cigarettes, tax revenues in millions of CZK

Source: Authors

The approach is such that we take the average of all α_{-i+1}^{M} and α_{-i+1}^{F} (separately) from the table. The resulting $\alpha^{\mathbf{M}} = 10.99$ and $\alpha^{\mathbf{F}} = 7.53$ denote the average adolescent tobacco consumption for each gender during past ten years. We plug these numbers into the model as if they were for whole adolescent population. As it is unlikely that the total teenage consumption average would fall bellow α^{F} or above α^{M}, the predicted outcomes should very likely be located inside the interval created by these scenarios.

Furthermore, we find the minimum of α_{-i+1}^{F} (5.87) and maximum of α_{-i+1}^{M} (13.09) and use them as scenarios for extreme cases. In every scenario, we assume the change of consumption takes place in the subsequent period and there are no changes further on.

As the portion of smokers among adolescents is assumed constant at its current rate, there is no point in plotting a chart for total population of smokers because that would equal to Figure 2. Predicted tobacco consumption is shown in Figure 8 and predicted tax revenues in Figure 9. The predictions of the model in a numerical form are located in Table 10.

The charts show that in the short run, the change in tobacco consumption of adolescents has only minor effect as they account only for a small part of population. However, smokers affected by these changes age and new ones come, so this transforms the population such that the portion of smokers with the new level of consumption grows steadily with time. Then, in the long run, the scenarios split and their differences grow with time.

5. Conclusions

This paper built a theoretical framework to model the future tobacco consumption, size of smoking population and governmental tax revenues in the Czech Republic. The constituted model had to be adjusted by certain limitations and assumptions (mainly due to the lack of time series data) in order to be applicable to the data by the Czech National Monitoring Centre for Drugs and Drug Addiction. While these restrictions simplified the model, the model projections should still be able to capture the future trends induced by upcoming demographic changes to the Czech population and provide approximate forecasts. On the contrary, not all assumptions are artificially imposed because of the lack of data, some are based on empirically justifiable reasoning – a good example is the main assumption of the model, stating that smokers form their tobacco consumption behaviour as adolescents.

The model predictions are dependent on the future characteristics of the adolescents – their smoking rate and average daily cigarette consumption. The data from Czech National Monitoring Centre for Drugs and Drug Addiction were used to identify the dependence of tobacco consumption (of each age category) on the prior teenage consumption and provide the resulting coefficients to the model. The same data set was used to supply input for the initial period (2013) of the model. The model inputs (adolescent tobacco consumption characteristics) for the upcoming years have been varied using several scenarios. The resulting forecasts have shown that no matter how large the change in these parameters is, the effect is little in the short run. The reason is that teenagers form only a small portion of the whole population. Nevertheless, in the long run, the effect increases as the portion of population affected by these changes grows and ages.

All of the scenarios predict a growth in the tobacco industry within the next 13 years (until 2028). In particular, the projected number of smokers in 2028 is by 4–8% higher than in 2013, the total daily tobacco consumption and tax revenue by 7–26%. This increase is induced by aging of large birth cohorts.

Interpreting the predictions in the long run is troublesome as the extreme scenarios differ substantially. For example, the total daily consumption can either double or halve in about 50 years. Focussing on the status quo scenario is the best way to understand the future development. Within the next 50 years, the perception, policies and behavioural patterns of smoking are very likely to change and debase any predictions. Nevertheless, it is interesting to see the long run outlook if the current state were to continue unchanged. The status quo scenario predicts an increase in number of smokers from 2.4 in 2013 to 2.7 million in 2028, later on fluctuating around 2.6 million with small deviation in 2048–2053 with a peak of 2.8 million in 2053. Similar pattern is observed for daily tobacco consumption and tax revenues. A boost up to 36.1 million of cigarettes in 2028 from current 31.8 million and fluctuation around 33–34 million with a deviation to 36.5 million in 2053. Tax revenues are expected to grow towards 53.3 billion of CZK in 2028 and further move around 50–52 billion compared to 46.8 in 2013.

Clearly, the question emerging from these outcomes is: Does this prospect of an increase in tobacco prevalence and consumption within the next decade call for a policy response or any taxation change? The Czech government currently introduced a new anti-tobacco policy. This law bans all smoking in restaurants and bars. This may influence the forthcoming development and the effect of this policy could partially offset the upcoming boom. However, possible trade-off is caused by the decrease in future tax revenues.

Possible model improvements and further research are strongly depending on an availability of time series data with consistent structure and necessary detail. An ideal approach would be to combine the methods of Mendez *et al.* (1998) and the age-cohort model. Such analysis would be able to estimate the cohort effects, improve the robustness of model coefficients and model the smoking behaviour much better (with cessation, higher probability of death for smokers, *etc.*). Altogether, it would result in a model with higher predictive power.

Appendix

Table 11 | Estimation Results: ZINB - All Variables

Variable	Coefficient	(Std. Err.)
alcabuse	0.197**	(0.059)
male	0.351**	(0.065)
unemp	0.118	(0.121)
student	−0.151	(0.181)
maternity	0.326*	(0.149)
retired	0.237	(0.163)
disabled	0.141	(0.177)
inc_2	0.058	(0.165)
inc_3	0.058	(0.149)
inc_4	0.118	(0.154)
inc_5	0.076	(0.155)
inc_6	0.066	(0.168)
inc_7	0.439*	(0.198)
$msize_2$	−0.066	(0.076)
$msize_3$	−0.145	(0.104)
$msize_5$	0.190	(0.088)
$msize_6$	0.003	(0.090)
edu_2	−0.289**	(0.099)
edu_3	−0.362**	(0.104)
edu_4	−0.426*	(0.198)
edu_5	−0.536**	(0.142)
b_2	0.082	(0.178)
b_3	0.297	(0.199)
b_4	0.231	(0.201)
b_5	0.218	(0.204)
b_6	0.240	(0.210)
b_7	0.446*	(0.212)
b_8	0.576**	(0.206)
b_9	0.485*	(0.210)
b_{10}	0.261	(0.241)
Intercept	2.104**	(0.215)
Inflate (logit): Table 12 in Appendices		
ln *alpha*	−1.494**	(0.102)

Note: Significance levels : † : 10% * : 5% ** : 1%

Source: Authors

Table 12 | Inflate (logit)

Variable	Coefficient	(Std. Err.)
alcabuse	−0.897**	(0.124)
age	0.017**	(0.004)
unemp	−0.641*	(0.261)
student	1.183**	(0.211)
msize$_2$	−0.222	(0.170)
msize$_3$	−0.279	(0.219)
msize$_4$	−0.564**	(0.213)
msize$_5$	−0.585**	(0.204)
msize$_6$	−0.557**	(0.197)
educ$_2$	0.346†	(0.191)
educ$_3$	0.894**	(0.180)
educ$_4$	1.352**	(0.353)
educ$_5$	1.807**	(0.243)

Note: Significance levels : † : 10% * : 5% ** : 1%

Source: Authors

References

Becker, G. S., Murphy, K. M. (1988). A Theory of Rational Addiction. *The Journal of Political Economy,* 96(4), 675–700, https://doi.org/10.1086/261558

Cameron, A. C., Trivedi, P. K. (1986). Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometrics,* 1(1), 29–53, https://doi.org/10.1002/jae.3950010104

Carstensen, B. (2007). Age–Period–Cohort Models for the Lexis Diagram. *Statistics in Medicine,* 26(15), 3018–3045, https://doi.org/10.1002/sim.2764

Hilbe, J. M. (2011). *Negative Binomial Regression.* Cambridge University Press. ISBN 9780511973420, https://doi.org/10.1017/CBO9780511973420

Kerr, W. C., Greenfield, T. K., Bond, J., Ye, Y., Rehm, J. (2004). Age, Period and Cohort Influences on Beer, Wine and Spirits Consumption Trends in the US National Alcohol Survey. *Addiction,* 99(9), 1111–1120, https://doi.org/10.1111/j.1360-0443.2004.00820.x

Kralikova, E., Kmetova, A., Zvolska, K., Blaha, M., Bortlicek, Z. (2013). Czech Adolescent Smokers: Unhappy to Smoke but Unable to Quit. *The International Journal of Tuberculosis and Lung Disease,* 17(6), 842–846.

Kvaček, J. (2011). Poptávka po cigaretách v České republice a výnosnost spotřební daně z cigaret. Master's thesis, IES FSV UK.

Levy, D. T., Ross, H., Kmetova, A., Kralikova, E., Stoklosa, M., Blackman, K., Diab, O. A., Abdelrahim, E. M., Esmail, M., Golding, J. F. et al. (1997). The Czech Republic SimSmoke: The Effect of Tobacco Control Policies on Smoking Prevalence and Smoking Attributable Deaths in the Czech Republic. *J Prev Med Public Health,* (4), 697–707.

Mason, K. O., Mason, W. M., Winsborough, H. H., Poole, W. K. (1973). Some Methodological Issues in Cohort Analysis of Archival Data. *American Sociological Review,* 38(2), 242–258, https://doi.org/10.2307/2094398

Mendez, D., Warner, K. E., Courant, P. N. (1998). Has Smoking Cessation Ceased? Expected Trends in the Prevalence of Smoking in the United States. *American Journal of Epidemiology,* 148(3), 249–258, https://doi.org/10.1093/oxfordjournals.aje.a009632

Mori, H., Clason, D. L. et al. (2004). A Cohort Approach for Predicting Future Eating Habits: The Case of At-Home Consumption of Fresh Fish and Meat in an Aging Japanese Society. *International Food and Agribusiness Management Review,* (1), 22–41.

Movsisyan, N. K., Sochor, O., Kralikova, E., Cifkova, R., Ross, H., Lopez-Jimenez, F. (2016, July). Current and Past Smoking Patterns in a Central European Urban Population: a Cross-sectional Study in a High-burden Country. *BMC Public Health,* 16(1), https://doi.org/10.1186/s12889-016-3216-5

OECD (2014). *OECD Factbook 2014.* http://dx.doi.org/10.1787/factbook-2014-en

Peřinová, P., Feketeová, E., Kemplink, D., Kovalská, P., Chlebušová, K., Nepožitek, J., Ibarburu, V., Králíková, E., Nevšímalová, S., Šonka, K. (2016). Smoking Prevalence and its Clinical Correlations in Patients with Narcolepsy-Cataplexy. *Prague Medical Report,* 117(2–3), 81–89, https://doi.org/10.14712/23362936.2016.8

Pertold, F. (2009). *Sorting into Secondary Education and Peer Effects in Youth Smoking.* CERGE-EI. Working Paper Series No. 399, https://doi.org/10.2139/ssrn.1517222

Powell, L., Tauras, J., Ross, H. (2005, September). The Importance of Peer Effects, Cigarette Prices and Tobacco Control Policies for Youth Smoking Behavior. *Journal of Health Economics,* 24(5), 950–968, https://doi.org/10.1016/j.jhealeco.2005.02.002

Rentz, J. O., Reynolds, F. D. (1991). Forecasting the Effects of an Aging Population on Product Consumption: An Age-Period-Cohort Framework. *Journal of Marketing Research,* 28(3), 355–360, https://doi.org/10.2307/3172871

Ross, H., Chaloupka, F. J. (2003). The Effect of Cigarette Prices on Youth Smoking. *Health economics,* 12(3), 217–230, https://doi.org/10.1002/hec.709

Shirane, R., Smith, K., Ross, H., Silver, K. E., Williams, S., Gilmore, A. (2012). Tobacco Industry Influence in the Czech Republic in Times of Major Political Change: Manipulating Tobacco Excise Tax and Tobacco Advertising Policies. *PLoS Med,* (6), –.

Sovinova, H., Csemy, L. (2016). Užívání tabáku v České republice 2015. Výzkumná zpráva, Státní zdravotní ústav.

Sovinová, H., Csémy, L., Procházka, B., Kottnauerová, S. (2008). Smoking-attributable Mortality in the Czech Republic. *Journal of Public Health,* 16(1), 37–42, https://doi.org/10.1007/s10389-007-0116-2

Sovinová, H., Csémy, L., Sadílek, P. (2014). Užívání tabáku v České republice 2013. *Státní zdravotní ústav.*

Spilková, J., Dzúrová, D., Pikhart, H. (2011). Inequalities in Smoking in the Czech Republic: Societal or Individual Effects? *Health & Place,* 17(1), 215–221, https://doi.org/10.1016/j.healthplace.2010.10.003

Tesar, T. (2011). Comparative Analysis of Factors Influencing Childrens Smoking. Bachelor's thesis, IES FSV UK.

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica: Journal of the Econometric Society,* 307–333, https://doi.org/10.2307/1912557