tobacconomics

Economic Research Informing
Tobacco Control Policy

The Impact of *Cigarette Price Increases on Daily Smoking Prevalence and Initiation in Argentina*

INSTITUTE FOR
HEALTH RESEARCH
AND POLICY

UIC

Citation

Franco-Churruarin, F. & Gonzalez-Rozada, M. (2022). *The impact of cigarette price increases on daily smoking prevalence and initiation in Argentina*. A Tobacconomics Research Report. Chicago, IL: Tobacconomics, Health Policy Center, Institute for Health Research and Policy, University of Illinois Chicago. *www.tobacconomics.org*

Authors

This Research Report was written by Martin Gonzalez-Rozada, PhD, Professor of Econometrics, Department of Economics, Universidad Torcuato Di Tella, and Fiona Franco-Churruarin, Researcher, Department of Economics, Universidad Torcuato Di Tella. It was peer-reviewed by Corné van Walbeek, PhD, Professor of Economics, University of Cape Town, Cape Town, South Africa; and Guillermo Paraje, PhD, Professor, Business School, Universidad Adolfo Ibáñez, Santiago, Chile.

This Research Report is funded by Bloomberg Philanthropies.

About Tobacconomics

Tobacconomics is a collaboration of leading researchers who have been studying the economics of tobacco control policy for nearly 30 years. The team is dedicated to helping researchers, advocates, and policy makers access the latest and best research about what's working—or not working—to curb tobacco consumption and its economic impacts. As a program of the University of Illinois Chicago, Tobacconomics is not affiliated with any tobacco manufacturer. Visit *www.tobacconomics.org* or follow us on Twitter *www.twitter.com/tobacconomics*.

Key Messages

Increasing cigarette prices through excise tax increases would reduce the initiation of daily smoking in Argentina.

Increasing prices would reduce daily smoking prevalence, regardless of the population group considered.

Daily smoking prevalence is lower among those with higher education, those in higher wealth quartiles, and women.

Price increases delay the age of daily smoking initiation. An increase of 10 percent in cigarette prices would delay smoking initiation by around four months.

Executive Summary

There is abundant evidence documenting the negative consequences of smoking. Over the last few decades, Argentina has implemented effective policies to reduce tobacco consumption, and smoking prevalence has been decreasing for the past 15 years. However, prevalence is still high compared to other countries in the region.

In Argentina, smoking usually starts in adolescence. On average, people start smoking daily at 17 years old, with many starting as young as the age of ten. Early smoking initiation predicts long-term nicotine dependence, affecting smoking behavior for the duration of a person's life. In practice, it is not possible to predictively identify potential smokers. Therefore, there is a compelling need to address the issue of onset and prevalence of smoking in young individuals.

Evidence shows that price increases reduce smoking prevalence and delay smoking onset. This research report contributes to the understanding of the determinants of smoking onset and prevalence by estimating daily prevalence price and smoking onset elasticities.

This study uses data from the 2018 National Risk Factors Survey (Encuesta Nacional de Factores de Riesgo, ENFR). Data from this survey show that 16.8 percent of Argentinians smoke every day. Prevalence of daily smoking is lower for women than for men. Survey data also show that smoking prevalence decreases as wealth increases. It is much more common for less wealthy individuals to be daily smokers: prevalence among the less wealthy is almost 21 percent, while it is only 13.8 percent for the wealthiest.

This report finds that daily prevalence price elasticity is -0.108, meaning that an increase of 10 percent in cigarette prices would reduce prevalence by 1.1 percent. This study finds that the prevalence price elasticity does not differ significantly between genders and wealth quartiles.

Another finding of this study is that while daily smoking prevalence is affected by an increase in prices, there is no evidence price affects overall prevalence in the group of both daily and less-than-daily

smokers. This suggests the effect of an increase in prices is more likely to discourage smokers from smoking every day, but not from smoking altogether; that is, some daily smokers may reduce to less-than-daily smoking rather than quit.

The data show the most common age at which Argentinians start smoking daily is 17 years old, with men starting, on average, earlier than women. The youngest age at which children start smoking is between the ages of eight and ten. The price elasticity of smoking onset is around 0.43. This means that an increase of 10 percent in prices would delay smoking initiation by 4.3 percent; this implies that, at mean smoking starting age, smoking would be delayed by around four months. This result is robust to different specifications.

The evidence presented in this research report suggests that increasing excise taxes on cigarettes, which leads to higher retail prices, would reduce daily smoking prevalence and induce a delay in smoking initiation.

1. Introduction

Smoking starts with the first few puffs, usually during childhood or early adolescence (see Wellman et al., 2016). Symptoms of nicotine dependence can manifest soon after onset in some adolescents, often well before daily or even regular smoking begins (see DiFranza et al., 2000, 2007; O'Loughlin et al., 2003; Gervais et al., 2006; O'Loughlin et al., 2009). Moreover, early onset predicts long-term adult smoking (see Chassin et al., 1990).

Nicotine addiction is the fundamental reason that individuals continue using tobacco products, and this persistent tobacco use contributes to many diseases (see USDHHS, 2010). Gonzalez-Rozada and Montamat (2019) have shown that, in Argentina, increases in cigarette prices reduce consumption more among people with a shorter history of addiction than for those that have consumed tobacco products for longer. This evidence suggests the importance of addressing the tobacco epidemic through control policies at early ages. It also highlights what is arguably the greatest challenge in tobacco policy: to have an effect on people who are addicted to tobacco.

Delaying the age at which individuals start daily smoking is associated with fewer health harms (Institute of Medicine, 2015). Since it is not possible to identify those individuals who, after their first use of tobacco, will adopt the habit of sustained smoking, the need to prevent the first puff is compelling (see Klein, 2006 and Gervais et al., 2006). There is substantial evidence that, of those individuals who have ever tried smoking, about one-third become daily smokers (USDHHS 1994, p. 67). Of those smokers who try to quit, less than five percent are successful at any one time (CDC, 2002, 2004). Therefore, any efforts to reduce tobacco initiation must consider the addiction potential of cigarettes.

This research report analyzes the determinants of daily smoking initiation, and, in particular, the impact of increasing cigarette prices via increasing cigarette excise taxes on prevalence and the onset of cigarette use. For smoking prevalence, the authors first estimate a probit model and then integrate this into a split-population model to quantify the impact of changes in cigarette prices on smoking onset.

This research report is organized as follows. Section 2 describes the data used in the estimations and presents a summary of descriptive statistics. Section 3 discusses the methodology used to estimate the probability of smoking and the smoking onset price elasticity. Results are presented in Section 4. Finally, Section 5 discusses these results and draws conclusions. An appendix provides further details on the estimation procedures and other analyses.

2. Data

2.1. Survey data

This research report uses data from the 2018 edition of the National Risk Factors Survey (Encuesta Nacional de Factores de Riesgo, or ENFR) carried out by the National Statistics and Census Institute (INDEC). This survey is part of the Surveillance System for Non-Communicable Diseases and the Integrated System of Household Surveys and provides information on risk factors such as tobacco consumption as well as nutrition, physical activity, alcohol consumption, medical care, and non-communicable diseases. The ENFR features 29,224 observations from individuals over the age of 18. Table 1 presents descriptive statistics for the

Table 1

Descriptive statistics of the ENFR survey

Variables	Total	Men	Women
Daily smoker	16.78% (0.40%)	19.83% (0.64%)	14.01% (0.49%)
Daily and less-than-daily smoker	23.46% (0.45%)	27.39% (0.71%)	19.90% (0.56%)
Age of daily smoking initiation	17.03 (0.57)	16.56 (0.78)	17.62 (0.82)
Price per pack (20 cigarettes) (AR$)	74.18 (0.32)	74.21 (0.46)	74.13 (0.41)
Highest level of education attained			
No formal education	7.32% (0.26%)	7.17% (0.38%)	7.45% (0.35%)
Primary	35.44% (0.50%)	38.45% (0.77%)	32.71% (0.66%)
Secondary	39.29% (0.51%)	39.09% (0.78%)	39.47% (0.68%)
Tertiary and university	17.95% (0.37%)	15.29% (0.54%)	20.37% (0.51%)
Employment status			
Employed	61.71% (0.50%)	72.58% (0.70%)	51.82% (0.69%)
Unemployed	6.17% (0.28%)	5.67% (0.39%)	6.63% (0.38%)
Out of labor force	32.11% (0.48%)	21.75% (0.63%)	41.54% (0.68%)

Table 1

Descriptive statistics of the ENFR survey (cont'd)

Variables	Total	Men	Women
Wealth index*			
1st quartile (poorest)	65.95 (0.31)	66.25 (0.45)	65.68 (0.42)
2nd quartile	85.10 (0.05)	85.10 (0.08)	85.10 (0.07)
3rd quartile	92.54 (0.04)	92.57 (0.05)	92.52 (0.05)
4th quartile (richest)	97.89 (0.03)	97.88 (0.05)	97.91 (0.04)
Age at survey**	43.95 (0.18)	43.77 (0.27)	44.11 (0.24)
18 to 24	15.37% (0.42%)	15.79% (0.63%)	15.00% (0.55%)
25 to 44	39.98% (0.51%)	39.51% (0.77%)	40.41% (0.68%)
45 to 64	28.65% (0.46%)	29.34% (0.71%)	28.02% (0.60%)
65 and older	15.99% (0.34%)	15.35% (0.51%)	16.57% (0.45%)

Note: Standard errors in parentheses.
**: Average value of the wealth index in each quartile of the distribution*
***: Average age in the first row. The following rows have the percentage of observations who are in each age group, conditional to being in each column: (1) aggregate or unconditional, (2) men, (3) women*
Source: Authors' calculations based on ENFR 2018

variables of interest in the survey, and the following paragraphs describe them.

The definition of a daily smoker adopted in this study includes individuals who self-report as smokers that smoke at least one cigarette every day. The overall prevalence of daily smoking is 16.78 percent. Prevalence among men is higher, with 19.83 percent of men smoking daily, while only 14.01 percent of women smoke every day. A looser definition considers anyone who reports that they smoke at the time of the survey, regardless of the frequency. Using this measure (daily and less-than-daily smokers) the smoking prevalence is about 23.5 percent and is larger for men (27.4 percent) than for women (20.0 percent).

The average age at which people take up smoking is 17. Men start smoking, on average, around a year earlier than women, with 16 and a half years being the mean starting age for men and almost 18 for women.

The average price paid for a pack of 20 cigarettes at the time of the survey was AR$ 74.18 (US$ 1.93 at the average exchange rate of September–December 2018[1] of 38.46 pesos per dollar).

The survey also provides information about education levels, with a focus on levels of education completed. In the survey, slightly more than seven percent of the population have not completed any formal level of education, with women having a somewhat higher value in this category. Those whose highest completed level of education is primary school (usually six years) comprise 35.44 percent of the population. There is a higher proportion of men in this category, with 38.4 percent of men and 32.71 percent of women having finished only primary school. Almost 40 percent of the population have finished secondary school (12 years), with very small differences between men and women. Finally, almost 18 percent of Argentinians finished tertiary or university studies. Postsecondary education is more prevalent for women, with 20.37 percent of women having postsecondary education compared to 15.29 percent of men.

Regarding employment status, 61.71 percent of the population is employed, and there is a gender difference of more than 20 percentage points. Table 1 reports that 72.58 percent of men are employed while only slightly more than half of women are. In terms of unemployed individuals, 6.17 percent were unemployed at the time of the survey (5.67 percent of men and 6.63 percent of women). The remaining 32.11 percent of the population is out of the labor force, with more women being out of the labor force than men.

Due to the high amount of missing data in reported monthly income, the authors compute a wealth index using Principal Components Analysis (PCA). Weights for this index are defined with the first principal component. The variables included in the PCA are binary and reflect socioeconomic characteristics of the person surveyed, such as education above secondary school and household possessions. The index ranges from 0 to 1 and is higher for individuals with more characteristics. This report defines quartiles for this index and categorizes individuals in them according to the value of the wealth index.

The average age in the survey is around 44 years. Authors define age groups that will be relevant covariates in the estimation, using age groups from 18 to 24 (15.37 percent), 25 to 44 (39.98 percent), 45 to 64 (28.65 percent), and those 65 and older (15.99 percent). Ages are relevant because, even though the estimation of prevalence price elasticity will use individuals of all ages, the estimation of onset price elasticity restricts the sample to those individuals between 18 and 32, representing 17 percent of the population.

Table 2 shows prevalence of smoking by age. The eldest individuals, aged 65 and older, are those who smoke the least, with a prevalence of 8.71 percent. They are followed by the youngest group, aged 18 to 24, with smoking prevalence of 14.01 percent. The groups in between these two have a smoking prevalence of around 19 percent, and it is slightly higher for the group aged 45–64 than for the 25–44 group.

[1] In the ENFR 2018, households were surveyed from September to December.

Table 2

Daily smoking prevalence by age groups

Age groups	Prevalence	Men	Women
18-24	14.01% (1.09%)	18.72% (1.82%)	9.51% (1.20%)
25-44	19.07% (0.67%)	22.63% (1.07%)	15.90% (0.81%)
45-64	19.58% (0.76%)	22.03% (1.16%)	17.20% (0.99%)
65 and older	8.71% (0.66%)	9.47% (0.98%)	8.07% (0.88%)

Note: Standard errors in parentheses
Source: Authors' calculations based on ENFR 2018

Table 3

Daily smoking prevalence by wealth quartiles

Wealth quartiles	Aggregate	Men	Women
1st quartile (poorest)	20.94% (0.97%)	25.64% (1.52%)	16.54% (1.19%)
2nd quartile	16.43% (0.77%)	20.56% (1.26%)	12.67% (0.91%)
3rd quartile	15.88% (0.73%)	17.80% (1.14%)	13.99% (0.92%)
4th quartile (richest)	13.84% (0.66%)	15.00% (1.05%)	12.84% (0.85%)

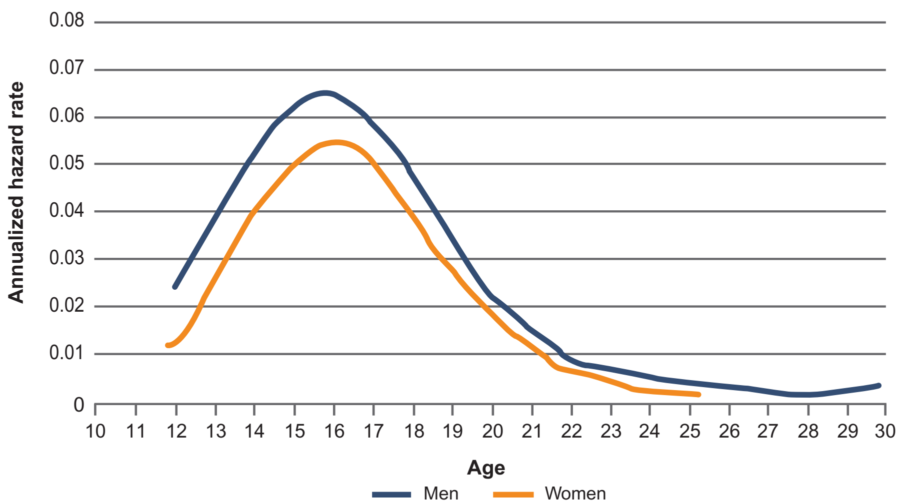
Note: Standard errors in parentheses
Source: Authors' calculations based on ENFR 2018

Table 3 shows smoking prevalence disaggregated by wealth quartiles. Smoking is the most prevalent among the less wealthy, with 20 percent of people in the poorest wealth quartile being daily smokers. In Argentina, daily smoking prevalence decreases as wealth increases, so that in the richest quartile, smoking prevalence is only around 14 percent.

The authors of this report assess the impact of increasing cigarette prices on the onset of daily smoking. For a first look at this issue, Figure 1 shows the smoothed hazard function of smoking initiation. As shown in the figure, people have a positive risk of initiating smoking daily from around age 12. Accordingly, in the modeling below, an individual is

Figure 1

Smoothed hazard function



Source: Authors' elaboration based on ENFR 2018

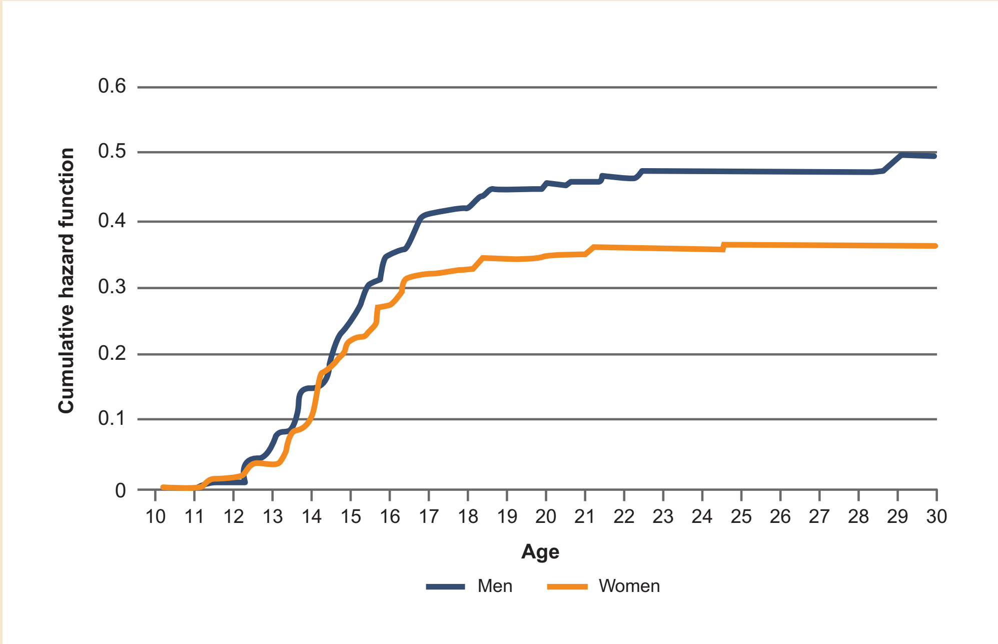
considered to be at risk of starting to smoke daily at the age of 10. Teenage men around the age of 16 have the highest risk of starting to smoke daily, while, for women, the highest risk is around 17 years old. The figure also shows that some people start smoking before their teenage years and might be at considerable risk up to their early twenties. At all ages, men are at higher risk of smoking initiation than women.

Figure 2 shows the cumulative hazard function of daily smoking initiation by gender. The cumulative hazard function describes the total amount of risk of initiating smoking (from this point on "initiating smoking" means initiating daily smoking) that has been accumulated up to each age in the x-axis. The cumulative hazard of starting smoking begins to increase around 13 years old for both men and women.

Around 21 years of age, the figure shows that the cumulative hazard of starting smoking among men is about 30% higher than that of women, and this relationship holds at older ages. Moreover, the slope of both curves is different, suggesting that between 13 and 17 years old the risk of initiating smoking for men increases at a faster rate (steeper slope) than for women. For women, the acceleration in the risk of starting to smoke is slower than for men, but after the age of 21, the slope of both cumulative hazard curves stabilizes.

Figure 2

Nelson-Aalen cumulative hazard estimates



Source: Authors' elaboration based on ENFR 2018

2.2. The price variable

The ENFR 2018 survey has questions regarding the last purchase of cigarettes for personal consumption. This research report computes the self-reported implicit paid price per cigarette. Following the recommendations in the *Economics of Tobacco Toolkit* (2010), the authors first check for endogeneity of the self-reported price using the Rivers-Vuong (1988) test statistic. The Rivers-Vuong procedure is similar to the Hausman (1978) test for endogeneity in the linear model but applied to a prevalence estimation. It consists of two steps.

In the first stage, the authors estimate, by least squares, the reduced form of the potentially endogenous self-reported price variable on the instrument and all exogenous variables of the model and generate the residuals of this estimation. In the second step, the authors estimate the prevalence equation using a probit model with the residuals of the first stage as an explanatory variable. In this second step, the Rivers-Vuong test for endogeneity consists of testing if the coefficient accompanying the residuals variable is equal to zero. Rejection of this hypothesis indicates that the self-reported price is endogenous.

Since non-smokers did not provide cigarette prices in the survey, before applying the test the authors must impute a price for them as if they had been smokers. This is done by using a random regression imputation. The procedure is as follows. First, a regression is estimated for the smokers in the survey, using as a dependent variable, the self-reported price paid at the last purchase and, as explanatory variables: gender (woman=1), age, labor and education categories, wealth quartiles, and region fixed effects. This is done after having dropped all observations of smokers who did not report price. Then, the authors input prices for non-smokers using the predicted price from this regression plus a random draw from a normal distribution with mean and standard deviation equal to the mean and standard deviation of the residuals (Table A1 in the Appendix shows the complete procedure). The average of this random imputation price per pack of 20 cigarettes is US$ 1.91 (AR$ 73.55).

For the instrumental variable used in the Rivers-Vuong test procedure, the report needs to assign both smokers and non-smokers the average price by primary sampling unit (PSU) as explained in the *Economics of Tobacco Toolkit* (2010). Unfortunately, ENFR 2018 only reports the results by province, so using average cigarette prices by province will generate only 24 different prices. Given that, by survey design, households in each province-PSU should have the same probability of being sampled—proportional to the households in that PSU—the authors use this information to generate 1,495 pseudo-PSUs. Pseudo-PSU classifications are created by grouping within each province all households with the same design weights (i.e. the same probability of being sampled). Authors cannot be sure that these are real PSUs since households could have the same design weight in two or more PSUs, but is the best that can be done to have enough variability in the instrument variable. The instrumental price variable is constructed by regressing the self-reported price per pack on binary variables for the pseudo-

Table 4
Average price by deciles

Deciles	Average random imputation price		Average price by PSU	
	Logged price	Actual price	Logged price	Actual price
1	4.23	68.89	4.11	60.90
2	4.25	70.30	4.20	66.72
3	4.26	70.91	4.24	69.63
4	4.27	71.41	4.25	70.45
5	4.28	71.90	4.27	71.29
6	4.28	72.46	4.29	72.68
7	4.29	72.97	4.31	74.57
8	4.30	73.47	4.33	76.29
9	4.30	74.04	4.37	78.85
10	4.33	75.83	4.45	85.87

Note: All prices are in logs and measured in AR$.
Source: Authors' elaboration based on ENFR 2018

PSU. The predicted price from this regression assigns the average price by pseudo-PSU to each individual in the sample, whether they are smokers or non-smokers. The average of this average price per pseudo-PSU is US$ 1.90 (AR$ 72.97).

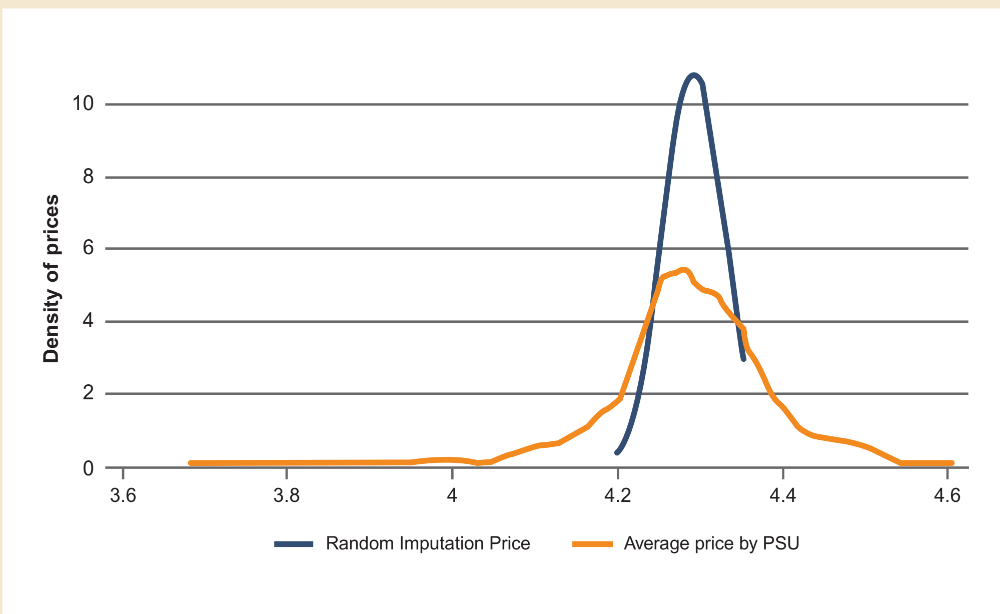
Table 4 summarizes the distribution of these measures of prices by showing the average price by deciles for each variable. The first two columns refer to the random imputation price, whereas the last two columns refer to the average price by PSU. In both cases, the first column shows the price variable expressed in logs, and the second column shows the actual price (in this case, in AR$). From the comparison of both price variables, it seems that the stochastic random regression imputation price estimates larger prices at the lower deciles but smaller prices at the higher deciles. The table suggests that the average price by PSU has a larger variability than the random regression imputation price.

Figure 3 plots the kernel density estimates of the two price variables (measured in logs of AR$). It is clear from the figure what Table 4 suggests: the variability of the random regression imputation price is smaller than the average price by PSU.

Table 5 shows the results of the Rivers-Vuong second step probit estimation. The first column of the table shows the exogenous explanatory variables, the natural logarithm of the self-reported price, *log(price)*, and the residuals from the first stage estimation. As the table shows, the coefficient accompanying the residuals is not statistically different from zero at usual significance levels, given that its t-statistic has a p-value of 0.627. This result indicates there is not evidence of the self-reported price being an endogenous variable, which suggests the smoking prevalence should be estimated using a regular probit model. Results of the first step of the Rivers-Vuong test procedure

Figure 3

Kernel density estimates of prices



Source: Authors' elaboration based on ENFR 2018 and the partial correlation between the self-reported price and the instrument can be found in the Appendix tables A2 and A3, respectively.

For the estimation of the impact of cigarette prices on smoking onset, the data first need to be transformed into a pseudo-panel in order to assign to each smoker the cigarette price at the smoking initiation date. For this calculation, this study uses the average weighted price of a 20-cigarettes pack constructed by the Ministry of Agriculture, Livestock and Fisheries of the Nation.[2] The authors use the Consumer Price Index (CPI) to convert this price into real terms. The index is set to 100 in November 2003. Figure 4 shows the evolution of the real price of cigarettes from January 1995.

Table 5
Rivers-Vuong endogeneity test

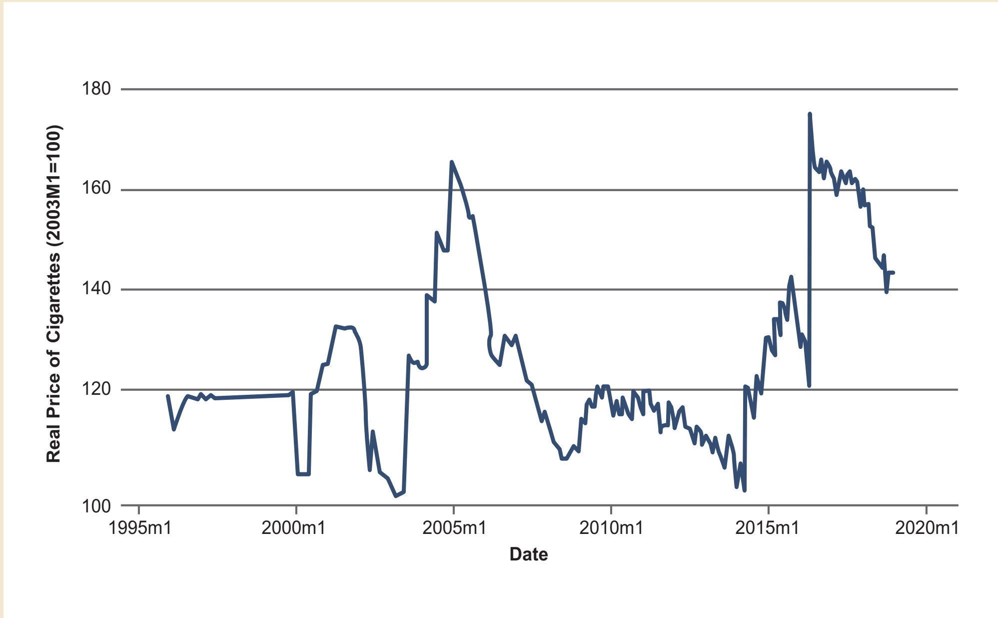
Dependent variable: smokedaily	Estimated coefficient	Linearized std. err.	t-statistic	p-value
log(price)	-0.7586	1.4839	-0.5100	0.6090
Woman=1	-0.2134	0.0320	-6.6700	0.0000
Age categories				
25 to 44	0.2517	0.0549	4.5800	0.0000
45 to 64	0.2259	0.0668	3.3800	0.0010
65 and older	-0.3123	0.0804	-3.8800	0.0000
Education				
Primary	0.0933	0.0723	1.2900	0.1970
Secondary	-0.0596	0.0806	-0.7400	0.4590
Tertiary and university	-0.2901	0.0804	-3.6100	0.0000
Region				
Pampas	0.0245	0.0738	0.3300	0.7390
North-west	-0.2639	0.0611	-4.3200	0.0000
North-east	-0.2711	0.0901	-3.0100	0.0030
Cuyo	-0.0014	0.1122	-0.0100	0.9900
Patagonia	-0.0567	0.0780	-0.7300	0.4670
Residuals	**0.7161**	**1.4743**	**0.4900**	**0.6270**
Constant	2.3403	6.3833	0.3700	0.7140
Number of observations = 29,224				

Source: Authors' estimation based on ENFR 2018

[2] This average weighted price can be found here: https://www.magyp.gob.ar/sitio/areas/tabaco/estadisticas/_archivos/000001-Volumen%20de%20Paquetes%20de%20Cigarrillos%20Vendidos%20por%20Rango%20de%20Precio%20(2008-2019).pdf

Figure 4

Evolution of the real price of cigarettes



Source: Authors' elaboration based on Ministry of Agriculture, Livestock and Fisheries of the Nation and Consumer Price Index (CPI) from INDEC

3. Methodology

This study uses survival analysis estimation focusing not only on smoking probability, but also on the onset of cigarette use. For smoking prevalence, the authors estimate a probit model, and, for smoking, onset a split-population model is used (Schmidt & Witte, 1989).

Since the ENFR survey has a single record per individual for their starting age of smoking, the authors construct a pseudo panel. Based on the reported age of initiation, authors create for each individual a duration spell. Duration refers to the time that elapses between the risk age of smoking onset and the age of starting. Therefore, a spell begins at the risk age (which this report assumes to be 10 years old), and either ends on the period the individual reported to have started smoking or at the survey date if they never started.

The main idea behind the use of a split-population model is to account for the fact that not all individuals who have an incomplete spell will eventually start smoking, as opposed to the traditional assumption of standard duration models that they all will. The duration process applies then only to those individuals who are predicted to eventually fail, which in this context means that they start smoking. The likelihood of each observation is weighted with the probability that the individual will ever start smoking. Formally expressed, the log-likelihood function to be maximized is:

$$ln\,(L) = \sum w_i \, \{c_i \, ln\,[\phi(\alpha' z_i)\, f(t|s_i = 1, x_i(t)] + (1 - c_i)\;\; ln\,[1 - \phi(\alpha' z_i) + \phi(\alpha' z_i)\, S(t|s_i = 1, x_i(t)]\} \qquad (1)$$

where c_i is a dummy variable equal to 1 if individual i ever smoked and 0 otherwise, s_i is another dummy equal to 1 if the individual will eventually start smoking and 0 if they never do. Φ is the standard normal cumulative function, and z_i is a vector of time-invariant covariates. f refers to the chosen conditional density function to model duration, S is the respective survival function, and w is a survey weight. $x_i(t)$ is a vector of time-varying covariates, including the price of cigarettes.

The contribution to the log likelihood (1) for individual i, who is an observed smoker in the sample ($c_i = 1$, uncensored observations) is simply the natural logarithm of the probability of daily smoking, $\Phi(\alpha' z_i)$, multiplied by the probability density function of starting smoking at the observed starting age $f(t/s_i = 1, x_i(t))$. For those individuals i observed as not starting smoking ($c_i = 0$, censored observations) the contribution is the natural logarithm of the probability of no daily smoking, $1 - \Phi(\alpha' z_i)$, plus the probability of starting after the age observed in the survey, $\Phi(\alpha' z_i)S(t/s_i=1, x_i(t))$ (Forster & Jones, 2001).

Notice that in the traditional split-population model the probability to start smoking is constant for all individuals, $\Phi(\alpha' zi) = k$, while here with a more general setup not all individuals have the same probability of starting to smoke. Smoking prevalence depends on the socioeconomic characteristics of the individuals. That is,

$$Pr(y_i = 1 \mid z_i) = \Phi(\alpha' z_i) \qquad (2)$$

where $y_i=1$ indicates that individual i smokes and z_i is a vector of explanatory variables including the log of the imputed self-reported cigarette price, the wealth index, a dummy for women, education and labor categories, and dummy variables for province of residence. In using these province-fixed effects, it is assumed there is no movement of individuals between provinces.

Using (2) as part of the log likelihood (1) means that, instead of estimating a single coefficient k for smoking prevalence, as in the traditional split-population model, the authors need to estimate the coefficients of a nonlinear function. This makes the log likelihood (1) to be maximized highly nonlinear and difficult to fit because the convergence to a maximum is more likely to fail (Jenkins, 2001). To avoid this problem, the strategy adopted here is first to use a probit model to estimate equation (2), $\widehat{\Phi}(\widehat{\alpha' z_i})$ and then introduce this estimation into equation (1) to estimate the duration coefficients. This procedure has the advantage of allowing the authors to compute the prevalence elasticity directly from equation (2) using,

$$\epsilon_i = \frac{\partial \Phi(\alpha' z_i)}{\partial \ln(cp_i)} \times \frac{1}{\Phi(\alpha' z_i)} \qquad (3)$$

where ln (cp_i) is the log of the imputed self-reported cigarette price. Equation (3) is a function that gives a different elasticity for each i. Therefore, when reporting the estimated elasticity, the average prevalence price elasticity is presented over the relevant group of people.

This study follows Forster and Jones (2001), who also use a split-population model to study the effect of tobacco taxes on smoking initiation, choosing the distribution of duration time to be log-logistic. This means that the density function in (1) is:

$$f\left(t \middle| s = 1, x_i(t)\right) = \frac{1}{\gamma} \frac{\psi^{1/\gamma} t^{1/\gamma - 1}}{[1 + (\psi t)^{(1/\gamma)}]^2} \qquad (4)$$

where $\psi = \exp\,(-\beta' x_i(t))$. The authors refer to γ as the shape parameter because it governs the shape of the density and the hazard. The hazard function of the log-logistic model is:

$$\lambda\left(t \middle| s = 1, x_i(t)\right) = \frac{1}{\gamma} \frac{\psi^{1/\gamma} t^{1/\gamma - 1}}{[1 + (\psi t)^{(1/\gamma)}]} \qquad (5)$$

The log-logistic model belongs to the continuous time accelerated failure time (AFT) class of models. Since this study uses monthly data, and the event of interest happens years after starting to be at risk, the assumption of continuous time is a reasonable one. The AFT class of models leads to an intuitive interpretation of coefficients because they are interpreted as the proportional change in survival time for a unit change in the regressor (Jenkins, 2005). In the case of regressors measured in logarithms, the coefficient accompanying it is an elasticity. The authors seek to estimate the price elasticity of daily smoking onset η_p, which is:

$$\eta_p = \frac{\partial \ln (T)}{\partial \ln (p)} = \beta_1 \qquad (6)$$

and so this study's results can be interpreted as a one-percent increase in prices (in real terms) leads to a β_1%-increase in daily smoking onset. As mentioned previously, an increase in smoking onset suggests a delay in the age at which individuals start smoking. The delay is calculated in months after the risk age of 10, which is the (dependent) time variable in the model. Thus, the delay in months at a given age *a* and risk age *r* (both in years) after a given price change of Δ_p is:

$$D(\beta_1,\Delta_p,a,r) = \beta_1 \cdot \Delta_p \cdot 12(a - r) \qquad (7)$$

where $\Delta_p = (p_1 - p_0)/p_0$. After calculating this, it is easy to recover the delay in years. The delay cannot be compared to the results of studies in which the individuals are assumed to be at risk at other starting ages (Guindon, 2014).

In order to estimate the split-population model with time-varying covariates, the authors expand the survey data from the risk age of smoking onset up to the date of the survey. This procedure is constrained by the availability of cigarette price data. This report assumes the starting risk age is 10 years old, and since the series of real cigarette prices started in January 1996, this data set keeps all individuals who, at the time of the survey (2018), were between 18 and 32 years old. There are 7,927 individuals in the survey between those ages, which accounts for 27.12 percent of the sample size. The consequence of this is that the results apply mostly to young individuals, but this is not a limitation given that one of the main problems of smoking onset is that it occurs at young ages. The authors expand this data set replicating each observation to create a pseudo-panel of monthly frequency with only time-invariant covariates and combine it with time-varying covariates such as the price of cigarettes.

Assignment of months is randomized due to the fact that people report the age in years at which they began smoking daily. Since the survey asks only about the age at which individuals started daily smoking, the authors input the price of a month of the year at which they started smoking at random. This cannot be done for those individuals who had not started smoking at the time of the survey. The solution to this problem adopted here is to attribute these individuals the cigarette price at the time of the survey. This procedure is constrained by the availability of cigarette price data. Prices are available from January 1996. For those observations whose age of starting smoking corresponds to a calendar month year before January 1996, the authors do not have any price to assign. If they were to be included, these observations would not be seen until the beginning of the time at risk.

The other explanatory variables in the duration part of the model are time invariant. The authors assign the value of the covariate at the date of the survey for each individual *i* in the new database. Thus, covariates vary between individuals but are fixed in time. This study shows estimates of two specifications of the model. The first specification to be estimated uses baseline covariates and province-fixed effects, while the second specification does not include the province-fixed effects. An assumption the authors implicitly make is that there are no movements between provinces over time.

4. Results

4.1. Smoking prevalence

Table 6 shows four alternative estimations. The first two columns of the table show estimations for daily smokers, whereas the last two columns show the estimations for daily and less-than-daily smokers. Column (1) presents a probit estimation using the random imputation based on the self-reported price described in Section 2.2. Column (2) shows the estimation of a probit model using the average self-reported price by pseudo-PSU. Because the empirical evidence presented in Section 2.2 suggests the random regression imputation price is an exogenous variable, the authors have greater confidence in the calculations shown in column (1) of the table. Columns (3) and (4) show the estimation using the same models as in column (1) and (2), respectively, but for daily and less-than-daily smokers.

In all specifications authors control for the following covariates: gender (woman=1), wealth index, age categories (18–24 as the base group, 25–44, 45–64, and 65 and older), education categories (no formal education, completed primary or less, more than primary up to complete secondary, and more than secondary education), and employment status (employed as the base group, unemployed, and out of the labor force). All specifications also use regional indicator variables.

The estimated coefficient on the price variable (in logs) is negative as expected, but is not statistically significant. Nonetheless, since this is a non-linear model, the coefficients do not have the usual marginal effect interpretation; the marginal effect has to be explicitly calculated. The coefficients only indicate the sign of the marginal effect. The relevant quantitative measure is the marginal effect, or elasticity in this case. This calculation is the standard derivative of the logged prevalence against the logged price. As a consequence, the standard errors of the estimated coefficients and the marginal effects (or elasticity) are fundamentally different. The estimated coefficient standard error only depends on the uncertainty of that coefficient estimation while the standard error of the estimated marginal effect (or elasticity) not only depends on the estimated coefficient uncertainty but also on the uncertainty of all other estimated coefficients in the model through the estimated density function. The prevalence price elasticity for daily smokers is negative and statistically significant in both specifications. This evidence suggests that increasing cigarette prices would reduce the daily smoking prevalence in Argentina. Both specifications show similar prevalence price elasticities. In particular, in column (1) the estimated prevalence price elasticity is -0.108, implying that a 10-percent increase in cigarette prices would be associated with a reduction of 1.1 percent in smoking prevalence.

When considering not only daily but also less-than-daily smokers as a measure of smoking prevalence, prevalence price elasticity is not statistically significant—meaning that, in this sample, the methods used did not find a statistically significant relationship between prices and overall smoking prevalence, as opposed to daily smoking prevalence. This could suggest that an increase in prices might induce frequent smokers to smoke less and

Table 6

Prevalence price elasticity estimation

Deciles	Daily smoking		Daily and less-than-daily smoking	
	Imputation price	Average price by PSU	Imputation price	Average price by PSU
Price of cigarettes (in logs)	-0.0605 (0.144)	-0.1110 (0.179)	0.0767 (0.135)	0.1652 (0.163)
Gender (woman=1)	-0.1571 (0.039)***	-0.1761 (0.033)***	-0.1748 (0.035)***	-0.1907 (0.03)
Wealth index	-0.0012 (0.175)	-0.4637 (0.140)***	0.1838 (0.152)	-0.5687 (0.125)
Age category				
25-44 years old	0.2170 (0.065)***	0.2225 (0.056)***	0.1556 (0.058)***	0.1676 (0.049)
45-64 years old	0.1654 (0.069)***	0.2511 (0.058)***	0.0297 (0.061)	0.0477 (0.051)
65 years and older	-0.3346 (0.085)***	-0.1498 (0.070)***	-0.5504 (0.074)***	-0.3925 (0.061)
Education categories				
Primary education	0.0513 (0.090)	0.0881 (0.066)	0.0483 (0.08)	0.0971 (0.06)
Secondary education	-0.0113 (0.096)	-0.0294 (0.071)	-0.0319 (0.085)	-0.0371 (0.065)
Tertiary/university education	-0.2801 (0.103)***	-0.2557 (0.078)***	-0.2887 (0.091)***	-0.2451 (0.071)
Labor category				
Unemployed	0.2492 (0.136)*	0.0946 (0.070)	0.0472 (0.078)	0.0647 (0.066)
Out of labor force	-0.1142 (0.087)	-0.2150 (0.043)***	-0.2817 (0.045)***	-0.1953 (0.039)
Region fixed effects	YES	YES	YES	YES
Intercept	-0.8670 (0.625)	-0.0715 (0.071)	-0.9538 (0.605)	-0.8225 (0.707)
Prevalence price elasticity	-0.1079 (0.017)***	-0.1702 (0.026)***	0.1225 (0.216)	0.2221 (0.219)

** 10%, ** 5%, and *** 1%*
Note: Standard errors in parentheses
Source: Authors' elaboration based on ENFR 2018

become less-than-daily smokers rather than to quit smoking altogether. These estimations suggest that the magnitude of the effect is different depending on the prevalence measure adopted. For the daily prevalence of smoking, the price elasticity is statistically significant, while for daily and less-than-daily prevalence it is not.

For daily smokers, the coefficient on gender in column (1) is negative and statistically significant, indicating that, on average, the smoking prevalence is lower for women than for men. Column (1) also suggests a very weak association between daily prevalence and wealth. Age is an important determinant of smoking prevalence. Results in column (1) suggest that prevalence is higher for people aged 25-64 and lower for individuals aged 65 and older, compared to the base category (age 18-24). The results in the table also suggest that those with higher education have a lower smoking prevalence.

Table 7 shows that prevalence price elasticity is slightly larger, in absolute value, for women than for men. However, this difference is not statistically significant. An increase of 10 percent in price induces a decline in the smoking probability of 1.1 percent for women and 1.0 percent for men.

Increases in prices affect mostly older people in Argentina. A 10-percent increase in cigarette prices will reduce the probability of smoking by 1.3 percent among those 65 years and older and 1.1 percent for those between 18 and 24 years old.

The prevalence price elasticity is similar for all quartiles of wealth. This result indicates that an increase in prices would induce a reduction in prevalence across all wealth

Table 7

Prevalence price elasticity by categories

Categories	Prevalence price elasticity	Standard error
Gender		
Men	-0.1017	0.0158***
Women	-0.1135	0.0165***
Age		
18-24 years old	-0.1113	0.0124***
25-44 years old	-0.0995	0.0117***
45-64 years old	-0.1026	0.0123***
65 years and older	-0.1346	0.0118***
Education categories		
No formal education	-0.1177	0.0200***
Complete primary	-0.1053	0.0178***
Complete secondary	-0.1047	0.0150***
Tertiary/university	-0.1160	0.0155***
Wealth quartiles		
Q1 (poorest)	-0.1049	0.0162***
Q2	-0.1078	0.0179***
Q3	-0.1097	0.0179***
Q4 (richest)	-0.1091	0.0165***

** 10%, ** 5%, and *** 1%*
Source: Authors' elaboration based on ENFR 2018

groups. The same thing seems to be true when considering education levels: those with higher levels of education have a prevalence price elasticity similar to those with less education. All these figures are statistically significant at the usual levels.

4.2. Smoking onset

Table 8 presents the duration estimation results that show estimates for the time that elapses between the risk age of daily smoking onset (assumed to be age 10) and the age of starting. The split-population model, equation (1), uses the daily prevalence equation presented in column (1) of Table 6. The duration component of the model is presented in accelerated failure time format, and, therefore, the estimated coefficients can be interpreted as regression coefficients for the logarithm of time until failure. For an explanatory variable expressed in natural logarithm its coefficient can be interpreted as an elasticity (see Forster & Jones, 2001). A positive coefficient indicates that higher values of the explanatory variable delay the initiation in smoking.

Column (1) is the baseline model, using as explanatory variables: gender (woman=1), wealth quartiles, age, education, labor categories, and province-fixed effects. Column (2) is the baseline specification without province-fixed effects. In both

Table 8

Split-population estimates using real cigarette prices

	[1]	[2]
Real price of cigarettes (in logs)	0.4334 [0.252]*	0.4973 [0.250]**
Gender (woman=1)	0.0183 [0.067]	0.0224 [0.062]
Wealth quartiles		
Q3	-0.0695 [0.101]	-0.0603 [0.097]
Q2	0.0076 [0.118]	0.0504 [0.115]
Q1 (poorest)	-0.0694 [0.145]	-0.0234 [0.12]
Age categories	YES	YES
Labor categories	YES	YES
Education categories	YES	YES
Fixed effects by provinces	YES	NO
Intercept	2.1131 [1.266]*	1.6651 [1.281]
Shape	0.2291 [0.013]***	0.2385 [0.014]***

** 10%, ** 5%, and *** 1%*
Note: Bootstrapped standard errors in brackets
Source: Authors' elaboration based on ENFR 2018

specifications the price elasticity of delay of smoking initiation is between 0.43 and 0.50. This elasticity is statistically significant at the usual levels, and it suggests that an increase in cigarette prices of 10 percent would delay the age of smoking initiation between 4.3 and 5.0 percent. This result implies that, at a mean starting age of 18 years,[3] an increase of 10 percent in prices would delay daily smoking initiation by around four months. This is calculated as 4.3 percent of the 12 x 8 months after the person turned 10, at which age it is assumed they are at risk of starting to smoke. Using the specification in column (2), this delay would be five months. The coefficient on the gender variable is not statistically significant, indicating that women initiate smoking at a similar age than men. The estimate of the shape parameter of the hazard rate is positive and statistically less than one, implying that the smoking hazard rate first rises with time and then falls monotonically as suggested in Figure 1 above.

Table 9

Split-population estimates using price-gender interactions

	[1]	[2]
Price-man interaction (in logs)	0.4313 [0.252]*	0.4950 [0.250]**
Price-woman interaction (in logs)	0.4347 [0.252]*	0.4993 [0.251]**
Wealth quartiles		
Q3	-0.0696 [0.101]	-0.0605 [0.097]
Q2	0.0074 [0.118]	0.0502 [0.115]
Q1 (poorest)	-0.0697 [0.145]	-0.0237 [0.119]
Age categories	YES	YES
Labor categories	YES	YES
Education categories	YES	YES
Fixed effects by provinces	YES	NO
Intercept	2.1235 [1.268]*	1.6764 [1.281]
Shape	0.2291 [0.013]***	0.2385 [0.014]***
Number of observations = 7,450		

** 10%, ** 5%, and *** 1%*
Note: Bootstrapped standard errors in brackets
Source: Authors' elaboration based on ENFR 2018

[3] In the cross-section sample, the average starting age is 17 years old, but once the pseudo panel is constructed, the mean starting age of smoking changes to 18 years old.

Table 9 inquires if changes in cigarette prices have differential effects on smoking initiation of women versus men. The structure of the table is similar to Table 8; the only difference is instead of using the logarithm of the cigarette price, it shows the interaction of this price variable with gender. As can be seen from the table, there are no statistical differences between the estimated coefficients on the price-gender interaction variable in both specifications. In the authors' preferred specification (column 1, including province fixed effects), an increase of 10 percent in cigarette prices would delay smoking initiation by 4.3 percent for both men and women.

Table 10 shows the results of the split-population model, where the logarithm of the cigarette price is interacted with a dummy variable adopting the value of 1 for those individuals in the lowest quartile of wealth (price-poor interaction). As in the case of gender, the table shows no significant difference between the impacts of an increase in prices on smoking initiation of those with low wealth compared to those with greater wealth. This means an increase in prices affects poor and non-poor people in the same way.

Table 10

Split-population estimates using price-wealth interactions

	[1]	[2]
Price-poor interaction (in logs)	0.4439 [0.247]*	0.5059 [0.244]**
Price-not poor interaction (in logs)	0.4545 [0.251]*	0.5103 [0.246]**
Gender (woman=1)	0.0184 [0.069]	0.0191 [0.062]
Age categories	YES	YES
Labor categories	YES	YES
Education categories	YES	YES
Fixed effects by provinces	YES	NO
Intercept	1.9777 [1.258]	1.5817 [1.221]
Shape	0.2302 [0.012]***	0.2401 [0.014]***
Number of observations = 7,450		

** 10%, ** 5%, and *** 1%*
Note: Bootstrapped standard errors in brackets
Source: Authors' elaboration based on ENFR 2018

5. Discussion

This study shows that increasing retail cigarette prices would reduce the probability of smoking and delay the age at which people start daily smoking. Therefore, a public policy of increasing excise taxes on cigarette consumption that leads to cigarette price increases is a relevant strategy to reduce the proportion of people who smoke daily and also to increase the average age at which people start smoking. The authors find that increasing cigarette prices by 10 percent would induce a reduction in the probability of smoking by 1.1 percent. This estimation is lower, in absolute magnitude, than the one found by Vellios and van Walbeek for South Africa and Gonzalez-Rozada and Franco-Churruarin for Mexico and Brazil.

When considering the price elasticity of smoking onset, a 10-percent increase in cigarette prices would delay the age of smoking initiation by around four months (at the mean starting age of 18). This evidence shows that young smokers are sensitive to increases in cigarette prices, suggesting that increasing excise taxes on cigarette consumption could be an important public policy to delay smoking initiation in Argentina.

Argentina has a very complex tax structure on cigarette consumption. There are four federal taxes affecting cigarette consumption: the additional emergency tax (IAE), the value added tax (VAT), the special tobacco fund (FET), and the internal tax (II). The tax base of each one is different but almost all are ad valorem types of taxes. The cigarette tax reform applied in 2016 increased the ad valorem tax rate of the II from 60 to 75 percent, inducing an average increase in cigarette retail prices of more than 40 percent (see Gonzalez-Rozada, 2020, for a detailed explanation of the tax reform). This is important because it highlights the fact that in Argentina it is relatively easy to increase cigarette retail prices through increases in excise tax.

6. Conclusion

In this policy report, the authors estimate the impact of increasing cigarette prices on daily smoking prevalence and on the age of smoking initiation in Argentina. The empirical evidence presented suggests that prices, gender, age, and wealth are important determinants of daily smoking prevalence in Argentina. On average, daily smoking prevalence is lower for women than for men, and it is negatively associated with wealth.

The addictive nature of tobacco products is at the center of many health problems, and adolescence is a key phase in which addiction might develop. The evidence presented in this report suggests that increases in cigarette prices are, on average, linked to less frequent smoking and a delay in the development of the habit of daily smoking. Delaying or reducing smoking at young ages is expected to improve health outcomes over the life course. Hence, a policy of increasing excise taxes with the objective of increasing cigarette prices could be very effective to reduce daily smoking and delay smoking initiation.

References

Centers for Disease Control and Prevention (CDC). (2002). Cigarette smoking among adults in the US, 2000. *Morbidity and Mortality Weekly Report, 51*(29), 642–5. [PubMed]

Centers for Disease Control and Prevention (CDC). (2004). Cigarette smoking among adults in the US, 2002. *Morbidity and Mortality Weekly Report, 53*(20), 427–31. [PubMed]

DiFranza, J. R., Rigotti, N. A., McNeill, A. D., &. (2000). Initial symptoms of nicotine dependence in adolescents. *Tobacco Control, 9*(3), 313–319. http://dx.doi.org/10.1136/tc.9.3.313

DiFranza, J. R., Savageau, J. A., Fletcher, K. E., &. (2007). Symptoms of tobacco dependence after brief intermittent use: The Development and Assessment of Nicotine Dependence in Youth study II. *Archives of Pediatrics & Adolescent Medicine, 161*(7), 704–710. http://dx.doi.org/10.1001/archpedi.161.7.704

Chassin, L., Presson, C. C., Sherman, S. J., & Edwards, D. A. (1990). The natural history of cigarette smoking: Predicting young-adult smoking outcomes from adolescent smoking patterns. *Health Psychology, 9*(6), 701–716. http://dx.doi.org/10.1037/0278-6133.9.6.701

Forster, M., & Jones, A. (2001). The role of tobacco taxes in starting and quitting smoking: Duration analysis of British data. *Journal of the Royal Statistical Society Series A, 164*(3), 517-547.

Gervais, A., O'Loughlin, J., Meshefedjian, G., Bancej, C., & Tremblay, M. (2006). Milestones in the natural course of onset of cigarette use among adolescents. *Canadian Medical Association Journal, 175*(3), 255–261. http://dx.doi.org/10.1503/cmaj.051235

Gonzalez-Rozada, M. (2020). Impact of a recent tobacco tax reform in Argentina. *Tobacco Control, 29*, s300–s303. doi:10.1136/tobaccocontrol-2019-055238

Gonzalez-Rozada, M., & Montamat, G. (2019). How increasing tobacco prices affects the decision to start and quit smoking: Evidence from Argentina. *International Journal of Environmental Research and Public Health, 16*(19), 3622. https://doi.org/10.3390/ijerph16193622

Gonzalez-Rozada, M., & Franco-Churruarin, F. (2020). *Increasing cigarette excise tax would delay smoking initiation in Mexico*. A Tobacconomics Policy Brief, University of Illinois at Chicago.

Gonzalez-Rozada, M., & Franco-Churruarin, F. (2020). *Effects of increasing cigarettes prices on smoking onset and prevalence in Brazil*. Mimeo.

Guindon, G. E. (2014). The impact of tobacco prices on smoking onset in Vietnam: Duration analysis of retrospective data. *European Journal of Health Economics*, 15, 19–39.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica, 46*(6), 1251-1271. doi:10.2307/1913827

Klein, J. (2006). Adolescents and smoking: The first puff may be the worst. *Canadian Medical Association Journal, 175*(3), 262–263.

O’Loughlin, J., DiFranza, J. R., Tyndale, R. F., &. (2003). Nicotine-dependence symptoms are associated with smoking frequency in adolescents. *American Journal of Preventive Medicine, 25*(3), 219–225. http://dx.doi.org/10.1016/S0749-3797(03)00198-3

O’Loughlin, J., Gervais, A., Dugas, E., & Meshefedjian, G. (2009). Milestones in the process of cessation among novice adolescent smokers. *American Journal of Public Health, 99*(3), 499–504. http://dx.doi.org/10.2105/AJPH.2007.128629

Rivers, D., & Quang, H. V. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics, 39*(3), 347-366. https://econpapers.repec.org/article/eeeeconom/

Schmidt, P., & Witte, A. D. (1989). Predicting criminal recidivism using ‘split population’ survival time models, *Journal of Econometrics, 40*(1), 141-159.

US Department of Health and Human Services (USDHHS). (1994). *Preventing tobacco use among young people: A report of the Surgeon General*. Atlanta: US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

US Department of Health and Human Services (USDHHS). (2010). *How tobacco smoke causes disease: The biology and behavioral basis for smoking-attributable disease: A report of the Surgeon General*. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Vellios, N., & van Walbeek, C. (2016). Determinants of regular smoking onset in South Africa using duration analysis. *BMJ Open, 6*, e011076. doi:10.1136/bmjopen-2016-011076

Wellman, R., Dugas, E. N., Dutczak, H., O’Loughlin, E. K., Datta, G. D., Lauzon, B., & O’Loughlin, J. (2016). Predictors of the onset of cigarette smoking. *American Journal of Preventive Medicine, 51*(5), 767–778.

World Health Organization. (2010). Economics of tobacco toolkit: Economic analysis of demand using data from the Global Adult Tobacco Survey (GATS). ISBN 978 92 4 150016 6

Appendix

Table A1
First step regression of price imputation

Dependent variable: log price	Estimated coefficient	Linearized std. err.	t-statistic	p-value
Women	0.0008	0.0081	0.1000	0.9210
Age categories				
25 to 44	-0.0069	0.0133	-0.5200	0.6050
45 to 64	-0.0295	0.0140	-2.1100	0.0350
65 and older	-0.0346	0.0191	-1.8200	0.0700
Education				
Primary	0.0267	0.0308	0.8700	0.3860
Secondary	0.0385	0.0315	1.2200	0.2220
Tertiary and university	0.0299	0.0322	0.9300	0.3540
Wealth quartiles				
3rd quartile	0.0159	0.0107	1.4800	0.1380
2nd quartile	0.0038	0.0120	0.3200	0.7520
1st quartile (poorest)	0.0066	0.0139	0.4700	0.6370
Region				
Pampas	-0.0396	0.0100	-3.9700	0.0000
North-west	-0.0286	0.0112	-2.5500	0.0110
North-east	-0.0507	0.0160	-3.1700	0.0020
Cuyo	-0.0665	0.0147	-4.5400	0.0000
Patagonia	-0.0445	0.0111	-3.9900	0.0000
Constant	4.2978	0.0365	117.8200	0.0000

Table A2

First stage of the Rivers-Vuong test

Dependent variable: log price	Estimated coefficient	Linearized std. err.	t-statistic	p-value
Instrument for price	0.1205	0.0175	6.8800	0.0000
Women	0.0000	0.0029	-0.0100	0.9890
Age categories				
25 to 44	-0.0002	0.0049	-0.0400	0.9700
45 to 64	-0.0255	0.0051	-4.9900	0.0000
65 and older	-0.0318	0.0055	-5.7500	0.0000
Education				
Primary	0.0234	0.0065	3.5700	0.0000
Secondary	0.0317	0.0067	4.7400	0.0000
Tertiary and university	0.0264	0.0070	3.7400	0.0000
Region				
Pampas	-0.0389	0.0037	-10.5400	0.0000
North-west	-0.0273	0.0040	-6.8900	0.0000
North-east	-0.0475	0.0044	-10.8700	0.0000
Cuyo	-0.0611	0.0050	-12.3100	0.0000
Patagonia	-0.0417	0.0042	-9.9600	0.0000
Constant	3.7855	0.0753	50.2400	0.0000
Number of observations = 29,224				

Table A3

Partial correlation between self-reported price and instrument

Dependent variable: self-reported price (in logs)	Estimated coefficient	Linearized std. err.	t-statistic	p-value
Average price by PSU (in logs)	**0.1289**	**0.0185**	**6.9700**	**0.0000**
Women	-0.0002	0.0029	-0.0700	0.9460
Wealth index	0.0476	0.0222	2.1400	0.0320
Age categories				
25 to 44	-0.0001	0.0049	-0.0200	0.9860
45 to 64	-0.0254	0.0050	-5.0400	0.0000
65 and older	-0.0336	0.0056	-5.9800	0.0000
Education				
Primary	0.0214	0.0065	3.2700	0.0010
Secondary	0.0326	0.0070	4.6400	0.0000
Tertiary and university	0.0282	0.0076	3.7100	0.0000
Wealth quartiles				
3rd quartile	0.0208	0.0044	4.7100	0.0000
2nd quartile	0.0140	0.0052	2.6700	0.0080
1st quartile (poorest)	0.0273	0.0079	3.4400	0.0010
Fixed effects by province			YES	
Intercept	3.6992	0.0814	45.4500	0.0000
Number of observations = 29,224				

www.tobacconomics.org

INSTITUTE FOR
HEALTH RESEARCH
AND POLICY

UIC