THE TURKISH JOURNAL ON ADDICTIONS
www.addicta.com.tr

ORIGINAL ARTICLE

The Importance of Price, Income, and Affordability in the Demand for Cigarettes in Spain

Pedro Cadahia[1], Antonio A. Golpe[1], Juan M. Martín-Álvarez[2], Eva Asensio[2]

[1]Department of Economics. University of Huelva, Huelva, Spain
[2]Faculty of Economics and Business, Universidad Internacional de La Rioja, Logrono, Spain

ORCID iDs of the authors: P.C. 0000-0002-2474-2071, A.A.G. 0000-0003-0360-1541, J.M.M.-Á. 0000-0002-7315-9827, E.A. 0000-0001-8365-7574.

Main Points

- The importance of affordability to control tobacco consumption in Spain has grown over time.
- The separate estimates of price and income elasticity that have been carried out in Spain so far must be interpreted, considering that as of 2010, price is more important than income in explaining the demand for cigarettes.
- From a certain level of affordability (when the relative income price is around 1%), the price becomes the magnitude that directs the effectiveness of affordability as a control tool.

Abstract

In the literature it is commonly accepted that the best mechanism to control smoking is by increasing tobacco prices via taxes. However, there are some studies that indicate that the decrease in tobacco consumption when prices rise is because consumer income is not capable of counteracting said rise.
The empirical analysis was developed using a panel of data from the Spanish provinces covering 2002 to 2018. By using Machine Learning assembly models, the importance of price, GDP and affordability as a mechanism for controlling the demand for cigarettes is estimated.
The importance of affordability to control tobacco consumption in Spain has grown over time. Furthermore, until 2010, income has generally better explained the demand for cigarettes in the Spanish provinces. However, as of 2010, price is the explanatory variable of the demand function that best explains the behavior of the demand for cigarettes. In these circumstances, the separate estimates of price and income elasticity that have been carried out in Spain so far must be interpreted considering that as of 2010, price is more important than income in explaining the demand for cigarettes. Furthermore, when the relative income price reaches 1%, there seems to be a positive relationship between this variable and the importance of price in explaining the demand for tobacco.
Although the demand functions estimated so far are useful to make predictions about the behavior of cigarette demand, the government must consider that price is a good tool to control tobacco consumption from a certain point of affordability. Specifically, it appears that when the relative income price exceeds 1%, the importance of price as a tool to control legal cigarette sales grows each time the relative income price rises.
Keywords: Economics, price, public policy

Corresponding Author: Juan Manuel Martín-Álvarez
E-mail: juanmanuel.martin@unir.net

Received: June 15, 2022
Accepted: September 25, 2022
Publication Date: December 19, 2022

Introduction

The main determinants in the demand for cigarettes are the price and the income of the consumers (Chaloupka et al., 2012). Along these lines, some studies include price and income as explanatory variables in the cigarette demand function; however, the current trend is a combination of both as an explanatory variable and as a measure of the affordability of cigarettes. This measure of affordability is based on the idea that the purchasing power of consumers to buy cigarettes depends on

Cite this article as: Cadahia, P., Golpe, A. A., Martín-Álvarez, J. M., & Asensio, E. (2022). The importance of price, income, and affordability in the demand for cigarettes in Spain. *Addicta: The Turkish Journal on Addictions*, 9(3), 241-251.

the price of tobacco products in relation to consumer income (Nargis et al. 2021). In the literature, it is commonly accepted that the best mechanism to control smoking is by increasing tobacco prices via taxes. However, there are some studies that indicate that the decrease in tobacco consumption when prices rise is because consumer income is not capable of counteracting the said rise. In other words, they associate the decrease in tobacco consumption with a lower affordability of the products and not with the simple fact that prices rise. A recent study indicates that in Spain, income is a determining factor that sometimes nullifies the effect of policies when governments use tax hikes to increase the price of cigarettes (Martín-Álvarez et al., 2021).

The classical economic model of tobacco demand estimates price and income elasticity separately to measure the effects of price and income changes on tobacco demand. Price elasticity measures the sensitivity of tobacco demand to changes in the prices of tobacco products after adjusting for inflation (real prices), holding real income and other factors constant. Similarly, income elasticity measures the sensitivity of demand for tobacco products to changes in income, holding the real prices of tobacco products and other factors constant. A negative price elasticity of demand for tobacco products indicates that an increase in price causes a reduction in tobacco use, keeping everything else constant, including income. If income increases the demand for tobacco products, tobacco use is not guaranteed to decline after a price increase because the net effect of simultaneous changes in prices and income on the demand for tobacco products will depend on the relative strength of these two effects. In Spain, there is provincial heterogeneity in terms of price elasticity, in some cases exceeding unity in absolute terms in the long term (Almeida et al., 2021). This is a relevant question because previous scientific evidence indicates that the regional price elasticity in Spain goes from 0.4 to more than 1. Thus, it seems relevant that the analysis of the importance of price, gross domestic product (GDP), and affordability must also be carried out at the regional level. On the other hand, the income elasticity of cigarettes in Spain shows a marked asymmetry, while the 1% increase in income generates a 0.40% increase in the demand for cigarettes and an economic recession of 1% causes a fall in the demand for cigarettes of 3.60%, ceteris paribus (Martín Álvarez et al., 2020).

The affordability elasticity measures the sensitivity of demand for tobacco products to changes in the price of tobacco products and income growth. Therefore, a negative affordability elasticity would imply that a price increase that exceeds the effect of income growth will lead to a reduction in tobacco use. In this sense, the extremely high values of the price and income elasticities found in Spain may be caused precisely by the lower affordability of tobacco products.

There are many studies in which the price and income elasticities of the demand for tobacco are estimated. In addition, given the irruption in the market for new-generation products, there are also studies that measure the price and income elasticities of products such as electronic cigarettes or smokeless tobacco (Chaloupka et al., 2011; Shang et al., 2020; USNCI & WHO, 2016). Although price elasticity and affordability elasticity may seem to be similar concepts, it is important to make a thorough analysis of both, because sometimes until a certain level of the relative income price (RIP hereafter) is reached, price policies are not effective.

Estimating the elasticity of the affordability of tobacco is currently a hot topic. However, the birth of this trend is motivated by a primitive study of men's demand for tobacco in Great Britain from 1946 to 1971. In this work, the elasticity of demand for cigarettes with respect to price was estimated as a percentage of annual per capita disposable personal income, known as the "price – income ratio," ranging from −0.44 to −0.58 (Russell, 1973). Although this early study used the "price – income relationship," currently the RIP is used as a measure of affordability (Blecher & Van Walbeek, 2004). The higher the RIP value, the lower the affordability and vice versa. Blecher and Van Walbeek defined RIP as the percentage of income from the purchase of 100 packs of cigarettes (Blecher & Van Walbeek, 2004). They investigated the relationship between affordability and cigarette smoking by estimating the elasticity of demand for affordability, concluding that there is a negative relationship between RIP and demand for tobacco.

Affordability analysis has become very important today to assess the effectiveness of tobacco tax increases as a measure to control tobacco consumption in accordance with the guidelines for the implementation of Article 6 on price and tax measures to reduce the demand for tobacco under the Framework Convention of the World Health Organization on tobacco control (WHO FCTC) (WHO FCTC, 2021).

In this article, we analyze the importance of the explanatory variables (price and income) of the basic cigarette demand function of all the Spanish provinces during the period 2002 – 2018. Next, the RIP is analyzed to find out from what levels of affordability the price policies are effective in Spain. To the best of our knowledge, it is the first time that machine learning techniques have been used to analyze the importance of price, income, and affordability in the demand for cigarettes.

Methods

Data

Our empirical analysis was developed using a panel of data from the Spanish provinces covering 2002 to 2018—the year in which the latest data on provincial GDP was published. For cigarette consumption, we used the official annual tobacco sales and the average price of a pack of 20 cigarettes in euros, as published by the Commission for the Trade of Tobacco. The real GDP is available from the National Institute of Statistics in Spain. All series employed here are per capita (18 years or older) and expressed in real terms using the consumer price index (CPI base 2016). Table 1 shows descriptive statistics of the data used.

Empirical Methodology

Based on the theory of demand, cigarette consumption is a function of the real price and per capita of the real income.

$$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 Y_t + \varepsilon_t \tag{1}$$

where Q_t is cigarette consumption, P_t is the real average price, and Y_t is the real GDP per capita. The objective of this work was

Table 1.
Descriptive Statistics of the Data Used

		Per Capita Cigarette Sales*					Price*					Per Capita GDP*				
				Quartile					Quartile					Quartile		
Province	Years	Mean	SD	Q1	Q2	Q3	Mean	SD	Q1	Q2	Q3	Mean	SD	Q1	Q2	Q3
Albacete	16	93.45	25.84	64.12	101.18	115.86	3.07	1.19	1.93	2.93	4.39	19.72	3.34	18.01	21.27	21.47
Alicante	16	135.80	55.05	78.74	127.64	184.19	3.10	1.15	2.02	2.93	4.36	19.75	2.21	19.37	20.19	21.01
Almería	16	114.20	38.21	72.19	118.10	146.96	3.14	1.17	2.03	2.98	4.41	21.59	2.59	20.89	21.92	23.40
Álava	16	81.18	22.34	56.98	84.07	101.80	3.06	1.21	1.92	2.91	4.37	36.26	6.96	32.92	38.29	40.87
Asturias	16	88.74	21.88	64.93	95.44	107.41	3.06	1.19	1.91	2.90	4.31	20.90	3.56	19.37	22.41	23.12
Ávila	16	93.47	25.53	65.35	99.97	115.41	3.08	1.20	1.94	2.92	4.37	18.93	3.27	16.96	20.30	21.09
Badajoz	16	98.68	28.85	66.72	110.11	123.38	3.04	1.20	1.87	2.91	4.29	17.53	2.96	16.05	18.74	19.34
Islas Baleares	16	168.33	72.05	95.93	150.77	228.62	3.13	1.15	2.05	2.93	4.37	27.29	3.66	26.06	28.11	29.15
Barcelona	16	87.59	27.11	58.16	89.84	109.03	3.04	1.20	1.89	2.86	4.34	28.75	5.25	26.19	30.21	31.28
Vizcaya	16	77.93	18.21	58.77	81.43	92.03	3.06	1.21	1.92	2.91	4.37	28.92	5.62	25.51	30.90	31.83
Burgos	16	87.44	23.79	61.89	91.81	109.79	3.05	1.21	1.89	2.88	4.33	26.61	4.62	24.13	28.33	29.22
Cáceres	16	99.28	26.80	69.04	109.19	121.22	3.04	1.19	1.89	2.90	4.34	17.25	3.14	15.54	18.33	18.86
Cádiz	16	75.60	33.69	37.81	82.92	107.68	3.04	1.19	1.89	2.90	4.32	18.82	2.40	18.64	19.65	20.27
Cantabria	16	93.99	30.05	65.44	102.02	118.87	3.08	1.20	1.93	2.91	4.41	22.51	3.54	20.86	23.99	24.55
Castellón	16	102.53	33.67	66.20	103.11	133.72	3.07	1.17	1.94	2.92	4.36	25.61	3.68	24.76	26.19	27.02
Ciudad Real	16	96.62	26.99	66.04	105.98	120.09	3.06	1.20	1.91	2.91	4.37	20.66	3.31	19.17	21.80	22.65
Córdoba	16	88.33	32.20	50.46	98.48	115.96	3.03	1.21	1.88	2.88	4.35	17.91	2.97	16.83	18.99	19.45
La Coruña	16	81.50	20.58	58.94	87.07	98.53	3.05	1.20	1.89	2.89	4.35	21.68	4.30	19.17	23.58	23.98
Cuenca	16	98.42	25.96	68.50	106.47	121.25	3.08	1.20	1.94	2.93	4.42	20.53	3.97	18.54	21.77	22.65
Guipúzcoa	16	146.87	51.79	95.44	144.96	193.39	3.03	1.21	1.85	2.86	4.32	31.07	5.47	28.36	33.18	33.92
Gerona	16	267.40	97.35	169.15	261.90	358.11	3.05	1.20	1.87	2.86	4.36	29.22	4.39	27.97	30.58	31.27
Granada	16	99.06	31.07	63.37	105.28	126.55	3.08	1.19	1.94	2.93	4.38	18.28	2.97	17.18	19.32	20.06
Guadalajara	16	93.49	28.64	61.79	96.66	116.66	3.08	1.18	1.97	2.92	4.38	20.93	2.69	20.53	21.98	22.47
Huelva	16	113.75	41.13	65.68	125.89	150.66	3.04	1.19	1.88	2.89	4.31	19.40	2.70	18.84	19.91	21.07
Huesca	16	116.51	33.40	79.16	124.19	147.17	3.07	1.20	1.92	2.90	4.35	26.90	5.23	23.35	28.87	30.21
Jaén	16	95.73	27.76	64.02	106.05	118.55	3.05	1.19	1.91	2.93	4.31	17.70	2.77	16.33	18.62	19.38

(*Continued*)

Table 1.
Descriptive Statistics of the Data Used (Continued)

		Per Capita Cigarette Sales*					Price*					Per Capita GDP*				
				Quartile					Quartile					Quartile		
Province	**Years**	**Mean**	**SD**	**Q1**	**Q2**	**Q3**	**Mean**	**SD**	**Q1**	**Q2**	**Q3**	**Mean**	**SD**	**Q1**	**Q2**	**Q3**
León	16	84.99	21.14	61.87	92.17	103.14	3.08	1.21	1.92	2.91	4.35	19.93	3.27	18.45	21.61	21.93
Lleida	16	140.99	52.46	85.33	144.39	188.71	3.02	1.20	1.84	2.84	4.34	29.59	4.90	26.65	31.15	33.19
Lugo	16	73.08	15.41	56.07	78.52	87.09	3.06	1.20	1.92	2.89	4.37	20.07	4.33	17.87	20.89	22.89
Madrid	16	88.05	27.11	59.71	90.11	108.47	3.07	1.19	1.93	2.91	4.35	33.79	5.86	30.85	35.65	36.87
Málaga	16	113.73	50.43	60.59	114.07	160.94	3.10	1.17	2.01	2.92	4.34	19.02	2.68	18.81	20.02	20.43
Murcia	16	107.57	32.88	71.37	111.70	136.50	3.09	1.17	1.99	2.94	4.35	21.52	3.36	20.33	22.57	23.18
Navarra	16	139.86	40.06	97.19	148.17	174.95	3.05	1.19	1.89	2.88	4.29	30.88	4.81	28.90	32.48	33.60
Orense	16	73.21	14.59	57.04	80.84	86.36	3.07	1.20	1.92	2.90	4.38	18.59	3.54	16.47	19.68	20.82
Palencia	16	89.41	23.05	64.70	96.01	108.03	3.07	1.20	1.92	2.90	4.35	24.18	4.14	21.85	25.57	26.43
Pontevedra	16	78.35	21.73	53.76	84.15	99.32	3.05	1.19	1.89	2.89	4.36	20.62	3.67	19.11	21.86	22.74
La Rioja	16	87.69	22.32	63.63	90.91	106.82	3.06	1.19	1.93	2.90	4.37	26.56	4.36	24.49	28.18	28.94
Salamanca	16	84.75	24.26	57.80	94.72	106.90	3.08	1.20	1.92	2.92	4.29	19.64	2.85	18.45	20.62	20.98
Segovia	16	86.75	25.42	58.20	91.45	109.43	3.08	1.20	1.93	2.91	4.36	22.86	3.11	22.15	23.92	24.44
Sevilla	16	86.20	38.24	42.25	95.66	120.93	3.05	1.19	1.90	2.90	4.28	20.76	3.22	19.61	22.07	22.61
Soria	16	83.82	20.09	61.58	89.94	99.13	3.11	1.20	1.96	2.96	4.39	23.91	3.88	21.42	25.48	26.35
Tarragona	16	115.17	41.05	71.29	112.86	153.18	3.11	1.16	2.01	2.94	4.40	30.05	4.15	27.91	31.13	31.66
Teruel	16	89.73	21.70	65.42	95.51	108.95	3.09	1.21	1.95	2.94	4.40	25.17	3.99	23.31	26.90	27.96
Toledo	16	97.11	30.58	62.88	103.27	124.87	3.07	1.18	1.94	2.90	4.35	19.55	2.65	19.44	20.30	20.93
Valencia	16	98.39	31.02	64.17	102.72	123.52	3.01	1.18	1.88	2.86	4.30	23.34	3.54	21.68	24.57	25.53
Valladolid	16	85.01	24.65	58.01	89.36	105.15	3.05	1.20	1.90	2.88	4.36	24.50	4.12	22.49	26.03	26.62
Zamora	16	79.91	19.41	58.84	87.73	95.88	3.06	1.20	1.91	2.89	4.33	18.27	3.48	16.13	19.31	20.69
Zaragoza	16	94.87	26.92	64.64	99.35	118.11	3.05	1.18	1.91	2.90	4.35	26.60	4.46	24.66	28.25	29.05
Spain	62	109.90	23.47	89.69	117.86	129.85	1.82	1.08	1.05	1.20	2.42	19.39	7.01	14.21	19.72	26.45

*Per capita sales are measured in packs of 20 cigarettes per year. The price is measured in real euros of 2016. GDP per capita is expressed in thousands of real euros of 2016.
GDP = gross domestic product; SD = standard deviation.

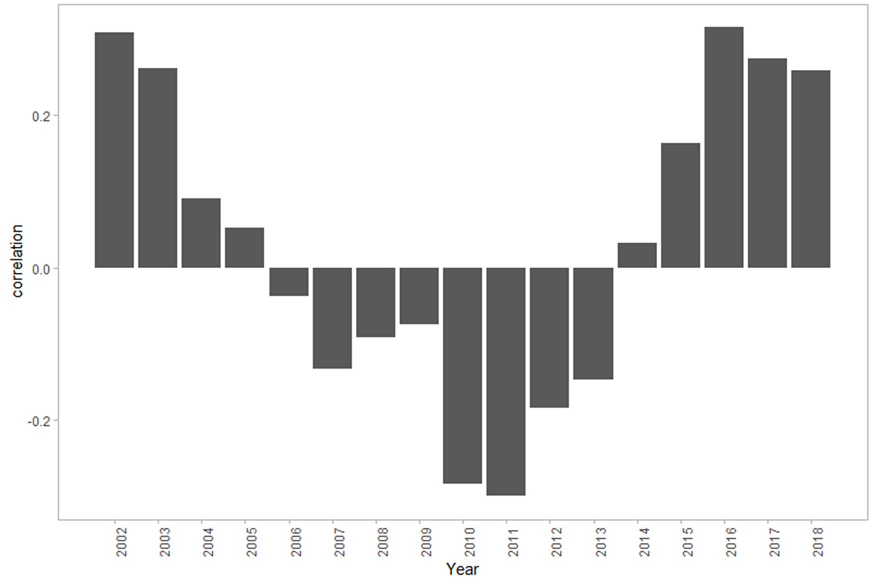

Figure 1. Pearson Correlation Between Price and Gross Domestic Product at Spanish Province Level by Year.

to identify key factors in this demand function, which is widely used in the previous literature, as well as to detect the relationship between these key factors and the affordability of cigarettes. Measuring the significance of variables is an important task, and several approaches have been proposed in the literature for addressing this question (Vladislavleva, 2013; Wei & Song, 2015; Yun et al., 2016; Yang et al., 2015). To achieve this objective, the measurement of the importance of explanatory variables in a demand function will shed light on how the importance of GDP and price has evolved as explanatory variables of per capita cigarette consumption over time. The quotient between the importance of price and the importance of GDP will allow us to observe the evolution of the importance that affordability has had over time as a tobacco control tool.

To measure the importance of price and GDP as tobacco control mechanisms, several tree-based ensemble methods have been fitted for finding the best model to interpret results. This model aims not only to predict tobacco sales by means of considering price and GDP as independent variables but also to study variable importance estimation through time in the Spanish territory. For this purpose, two datasets have been analyzed, one for a country-level analysis and the other for a provincial level analysis.[1]

In the provincial level dataset, the data are split into a training set and a test set. The training set consists of all the provinces available with the exception of the province whose tobacco sales are being predicted—namely the test set. That is, when possible, all provinces are used to predict sales in a selected province without including that chosen province. After model assessment, the variable importance metrics (VIM) are computed and studied over the time and at a province level to obtain insights.

The dataset at the country level is divided into three periods following a recent investigation that indicates that there are two structural breaks in the per capita demand for cigarettes in Spain (1969 and 2006) (Martín-Álvarez et al., 2021). The variable importance for these three periods is analyzed by fitting the winning model for each period.

To analyze the importance of variables, it must be taken into account the fact that the explanatory variables cannot be highly correlated (Dohoo et al., 1997). The collinearity issue has been proposed by means of a correlation coefficient and variance inflation factor (VIF) (Dohoo et al., 1997), and collinearity is present at the 0.9 level of a correlation coefficient or higher. Multicollinearity can be assessed by computing a score called the VIF, which measures the inflation of the variance of a regression coefficient due to multicollinearity. Furthermore, VIF values should be higher than 5 or 10 (James et al., 2014; Lin, 2008).

A correlation analysis by year between price and GDP has been performed. Figure 1 shows the results of correlation coefficients.

To complement this, the VIF test results show low values (1.41), so, with VIF and correlation analysis, we assume that there is no correlation and collinearity between variables. This is relevant to the present study, which shows the methodology and confirms several requirements for correctly interpreting the variable importance.

Once the correlation between explanatory variables has been analyzed, the next methodological step is to present some joint tree-based methods to model the relationship between a dependent variable and the characteristic vector x. An estimation of models in which the quantiles of the response are modeled to depend on the features is presented by Koenker & Bassett (1978), which is the quantile regression (QR). This method is based primarily on choosing a model for the conditional quantile; in contrast, the minimal squares estimate the conditional mean. Based on the imposed assumptions, the choice of parametric or

1 For data time issues, two datasets have been employed, as the provincial level dataset has information available from 2002 to 2018 and the country level has a wider time range.

nonparametric is available (Engle, 2004; Zhao, 2008). For the conditional α-quantile q of a scalar variable Y, $P(Y \leq q/I) = \alpha$, where the probability $0 < \alpha < 1$ is given and I denotes an information set generated by independent variables X.[2]

To measure the importance of explanatory variables of the cigarette demand function, in this work, two main methods based on trees are proposed. On the one hand, a hybrid of random forest (RF) and QR has been presented by Meinshausen (2006); this is the quantile regression forest (QRF) approach. It should be remarked that a key difference between RF and QRF is that QRF for each node of each tree maintains the values of all observations of the node, but RF only keeps the mean of the observations found in the node (Meinshausen, 2006). The quantreg Forest and ranger library, an implementation of RF (Breiman, 2001) in R software was used to fit a QRF and Ranger model respectively with the default settings (Wright & Ziegler, 2015). On the other hand, an implementation of gradient boosting machine (GBM) for QR has been selected (Friedman, 2001) due to its flexibility and efficiency in performing regression tasks (Freund et al., 1999). A quantile version of the GBM has been selected, which is the gradient boosted quantile regression. This method has been applied in several fields (Sun & Pfahringer, 2011). The accurateness of GBM predictions comes from increasingly refined approximations, which are carried out by adding tree-based models together. For a better understanding of this method, in terms of its theory and formulation of GBM, the works of Friedman (2001) and Ferreira and Figueiredo (2012) should be consulted. The R library GBM was used to fit the quantile version of GBM with the default settings.

To evaluate the model performance, a data partition was performed, and the predictive accuracy of the models was measured by splitting the data into training and test sets. In the training set, the out-of-bag (OOB) estimation error was performed in order to get an unbiased estimate of the test set error in tree-based ensemble methods (Breiman, 2001).

Every tree is made through bootstrap samples from data, in a one-third proportion, using cases out of the bootstrap sample for the *k*th tree construction.

In about one-third of the trees, a test set classification is obtained for each case. At the end of the run, take *j* to be the class that got most of the votes every time case *n* was OOB. The proportion of times that *j* is not equal to the true class of *n* averaged over all cases is the OOB error estimate. This has proven to be unbiased in many tests.

Given the 0.5-quantile predicted responses ($\hat{y}_i$) and the actual values (y_i) for the training and test sets, the error is computed as $e_i = y_i - \hat{y}_i$, using the metrics from Table 2.

Tree-based ensemble methods were selected to quantify the importance of explanatory variables in predicting cigarette consumption. These methods were chosen due to their evident suitability for the said objective. Specifically, these methods estimate the importance of an independent variable x_j due to the decrease in predictive precision when the values of x_j are randomly permuted.

Table 2.
Predictive Performance Metrics

Averaged Error metric	**Calculation Way**
Prediction error	$e_i = y_i - \hat{y}_i$
Mean squared error	$MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2$
Mean absolute error	$MAE = \frac{1}{n}\sum_{i=1}^{n} \lvert e_i \rvert$

Note: For further understanding of these formulas and its statistical properties, see Hyndman and Koehler (2006).

Several VIM based on tree-based ensembles can be found in the literature: the more traditional variance decrease VIM (VDVIM) and Gini VIM (GVIM) (Hyndman & Koehler, 2006) and other metrics such as permutation VIM (PVIM) (Wei & Song, 2015) and conditional permutation VIM (Strobl et al., 2007). Furthermore, GVIM, VDVIM, and PVIM metrics have been benchmarked in several studies not only in simulated but also in real data (Strobl et al., 2007; Wei & Song, 2015). Specifically, when continuous explicative variables are mutually uncorrelated and collinearity is not present among them, GVIM/VDVIM is expected to produce better results than PVIM (Strobl et al., 2007; Wei & Song, 2015). As a result of the variable analysis, VDVIM may lead to better results over PVIM due to the lack of correlation and collinearity. For either classification or regression models, VDVIM and GVIM are popular VIM methods; for further details of the GVIM method, see the previous literature (Hyndman & Koehler, 2006).

The VDVIM computation at tree level is that at each node N, the choice of splitting variable X from a set of split-variable candidates, as well as the splitting criteria, is based on maximizing the decrease of the metric of this node N:

$$VDVIM(N) = \frac{1}{|S|^2}\sum_{i\in S}\sum_{j\in S}\frac{1}{2}\left(x_i - x_j\right)^2 - \left(\frac{1}{|S_t|^2}\sum_{i\in S_t}\sum_{j\in S_t}\frac{1}{2}\left(x_i - x_j\right)^2 + \frac{1}{|S_f|^2}\sum_{i\in S_f}\sum_{j\in S_f}\frac{1}{2}\left(x_i - x_j\right)^2\right)$$

S is the set of pre-split sample indices, S_t is the set of sample indices for which the split test is true, and S_f is the set of sample indices for which the split test is false. Each of the above summands is indeed variance estimate, though it is written in a form that does not directly refer to the mean (Wright & Ziegler, 2015).

For each estimate of the GBM model, the Mean Squared Error (MSE) is calculated at each split of the trees. Then, it averages the improvement made by each variable x_i across all the trees that the variable x_i is used. The variables with the largest average decrease in MSE are considered most important (Ferreira & Figueiredo, 2012; Friedman, 2001).

2 For clarification of this method, see Koenker & Bassett (1978).

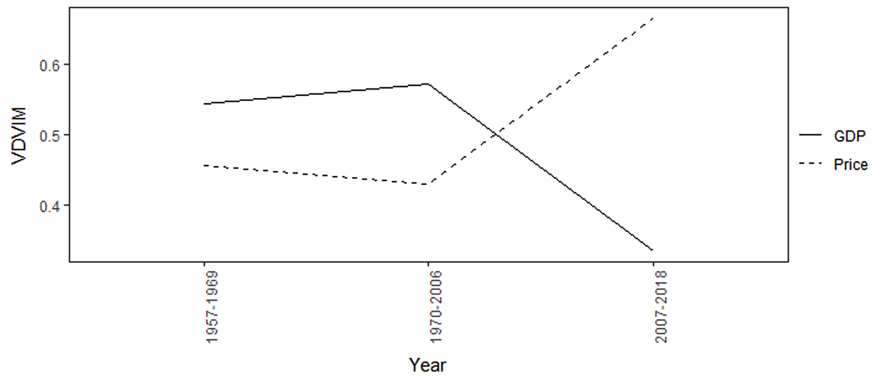

Figure 2. Importance of Price and GDP in the Three Subperiods. GDP = Gross Domestic Product; VDVIM = Variance Decrease Variable Importance Metrics.

Results

The results are presented in three parts. First, the results of the importance of variables in the cigarette demand function in Spain from 1957 to 2018 are shown. Next, the results of the provincial analysis of the importance of variables in the cigarette demand function are shown. Finally, the relationship between affordability, measured by the RIP, and the importance of the explanatory variables of the demand function for cigarettes in Spain is shown.

For measuring, at national level, the changing importance of variables for the periods, the winning model is applied in the three different periods for measuring variable importance, according to Martín-Álvarez et al. (2021). The results are shown in Figure 2 and Table 3. Variable importance of three fitted models for the established periods is plotted in Figure 2. Results show how variable importance changes over time. As can be seen, until 2006, the main driver of the demand for tobacco in Spain was income. However, during the period 2007 – 2018, price is the main explanatory variable of the demand function for cigarettes in Spain. These results are similar to what a recent study established; it is suggested that income in Spain canceled out the effect of increases in the price of cigarettes until 2006 (Martín-Álvarez et al., 2021). Therefore, it seems that these results provide the same evidence as Martín-Álvarez et al. (2021), since only from 2007, price is a more relevant explanatory variable than GDP to explain the demand for cigarettes.

A province granularity-level analysis is performed, and year-averaged importance of price and GDP metrics was computed for every model. We can see some patterns over time in Table 4 and the Density Plots for VIM Metrics for QRF Model in Figure 6. Tables 5 and 6 show other measures of goodness of fit for the models used. Results show a price relevance increase from 2007 over the importance of the GDP variable, with 2010 being the major switch over year. As can be seen, until 2010 (except in 2004), GDP is the variable that most determines cigarette consumption. However, as of 2010 (except in 2013 and 2014), it is price that is the most relevant variable in determining cigarette consumption.

To have a more precise vision of the evolution of the importance of the main explanatory variables of the cigarette demand function, the relationship between price and GDP has been included in Table 4. This price importance/GDP importance ratio (affordability importance) can be calculated to simplify the results, and the results are more interpretable:

- Affordability importance < 1: more relevance on GDP
- Affordability importance = 1: equally relevant
- Affordability importance > 1: more relevance on price

The affordability importance in the upper part of Figure 3 shows the provincial granularity level. In this sense, the patterns of all the provinces are aligned with a growing trend since 2007 and with a ratio greater than 1 as of 2010. On the other hand, the lower part of Figure 3 shows the aggregate mean of the affordability importance, which also shows an increasing trend at the national level of the ratio. These results seem to suggest that until 2010 price policies have not had the desired effect on public health. However, as of 2010, price is more important than GDP in explaining the demand for cigarettes. Figure 4 shows how from 2010 there is a notable growth of the RIP, which may be related to the indicated behavior. In other words, it seems that from a certain threshold of RIP, the price policies are more relevant to the aim of minimizing cigarette consumption.

Finally, Figure 5 shows the last relevant finding of this paper. Although it has been indicated that as of 2010 (when the RIP

Table 3.
Importance of Price and GDP in the Three Subperiods

Period	Model	GDP Importance	Price Importance	MAE (OOB)	MSE (OOB)
1957 – 1969	QRF	54.43%	45.57%	0.31	0.14
1970 – 2006	QRF	57.14%	42.86%	0.64	0.49
2007 – 2018	QRF	33.50%	66.50%	0.51	0.22

GDP = gross domestic product; OOB = out-of-bag; QRF = quantile regression forest.

Table 4.
Importance of GDP, Price, and Affordability Using Provincial Data

Year	Model	GDP Importance (%)	Price Importance	Affordability Importance	MAE (OOB)	MSE (OOB)
2002	QRF	62.88	37.12	0.59	0.1185	0.5135
2003	QRF	56.06	43.94	0.78	0.1547	0.6898
2004	QRF	44.05	55.95	1.27	0.1459	0.6685
2005	QRF	61.08	38.92	0.64	0.1515	0.5008
2006	QRF	60.04	39.96	0.67	0.1496	0.5879
2007	QRF	52.57	47.43	0.90	0.1742	0.8271
2008	QRF	55.42	44.58	0.80	0.1817	1.163
2009	QRF	56.08	43.92	0.78	0.1604	1.0105
2010	QRF	53.07	46.93	0.88	0.1453	0.9682
2011	QRF	46.56	53.44	1.15	0.1542	0.9481
2012	QRF	41.66	58.34	1.40	0.1576	1.1917
2013	QRF	50.50	49.50	0.98	0.1617	0.9383
2014	QRF	52.18	47.82	0.92	0.1391	0.8922
2015	QRF	44.27	55.73	1.26	0.1518	0.9684
2016	QRF	43.97	56.03	1.27	0.1478	0.9961
2017	QRF	44.29	55.71	1.26	0.154	1.0343
2018	QRF	46.36	53.64	1.16	0.1655	1.1451

GDP = gross domestic product; OOB = out-of-bag; QRF = quantile regression forest.

Table 5.
Results of Training and Test Sets for Error Assessment

Set	Model	MAE	MSE
Training	Ranger	0.1537	0.8849
Training	QRF	0.1561	0.0509
Training	GBM	0.1515	0.0581
Test	Ranger	0.1539	0.0495
Test	QRF	0.1527	0.0526
Test	GBM	0.1553	0.0594

GBM = gradient boosting machine; QRF = quantile regression forest.

is around 1%) price is the explanatory variable that dominates the demand function for cigarettes in Spain, it seems vreasonable to ask: Is the RIP related to the importance of price in the demand function for cigarettes? In Figure 5, it is observed how until 2010 the relationship between RIP and VIM price does not present a trend. However, as of 2010, a slight positive trend appears to be observed in the relationship between the RIP and the VIM. It seems that from a certain RIP threshold there is a slight positive relationship between affordability and the importance of price as an explanatory variable of demand for cigarettes.

Discussion

In recent years, the concern about knowing which factors drive the demand for tobacco has grown due to the deleterious effect that tobacco has on public health. Based on data from a provincial panel on cigarette sales in Spain, the results of this article are adjusted to the existing evidence that price and income are variables capable of explaining the demand for cigarettes. The results

Table 6.
VIM Statistics Comparison for Every Model

Model	VIM	Feature	Mean	Median	SD	MAE
Ranger	Relative VDVIM	GDP	0.5124	0.5168	0.0666	0.0883
Ranger	Relative VDVIM	Price	0.4876	0.4832	0.0666	0.0883
QRF	Relative VDVIM	GDP	0.5201	0.5160	0.0791	0.0889
QRF	Relative VDVIM	Price	0.4799	0.4840	0.0791	0.0889
GBM	Relative VDVIM	GDP	0.5783	0.6069	0.2460	0.2712
GBM	Relative VDVIM	Price	0.4217	0.3931	0.2460	0.2712

GBM = gradient boosting machine; GDP = gross domestic product; QRF = quantile regression forest; VDVIM, variance decrease variable importance metrics; VIM = variable importance metrics.

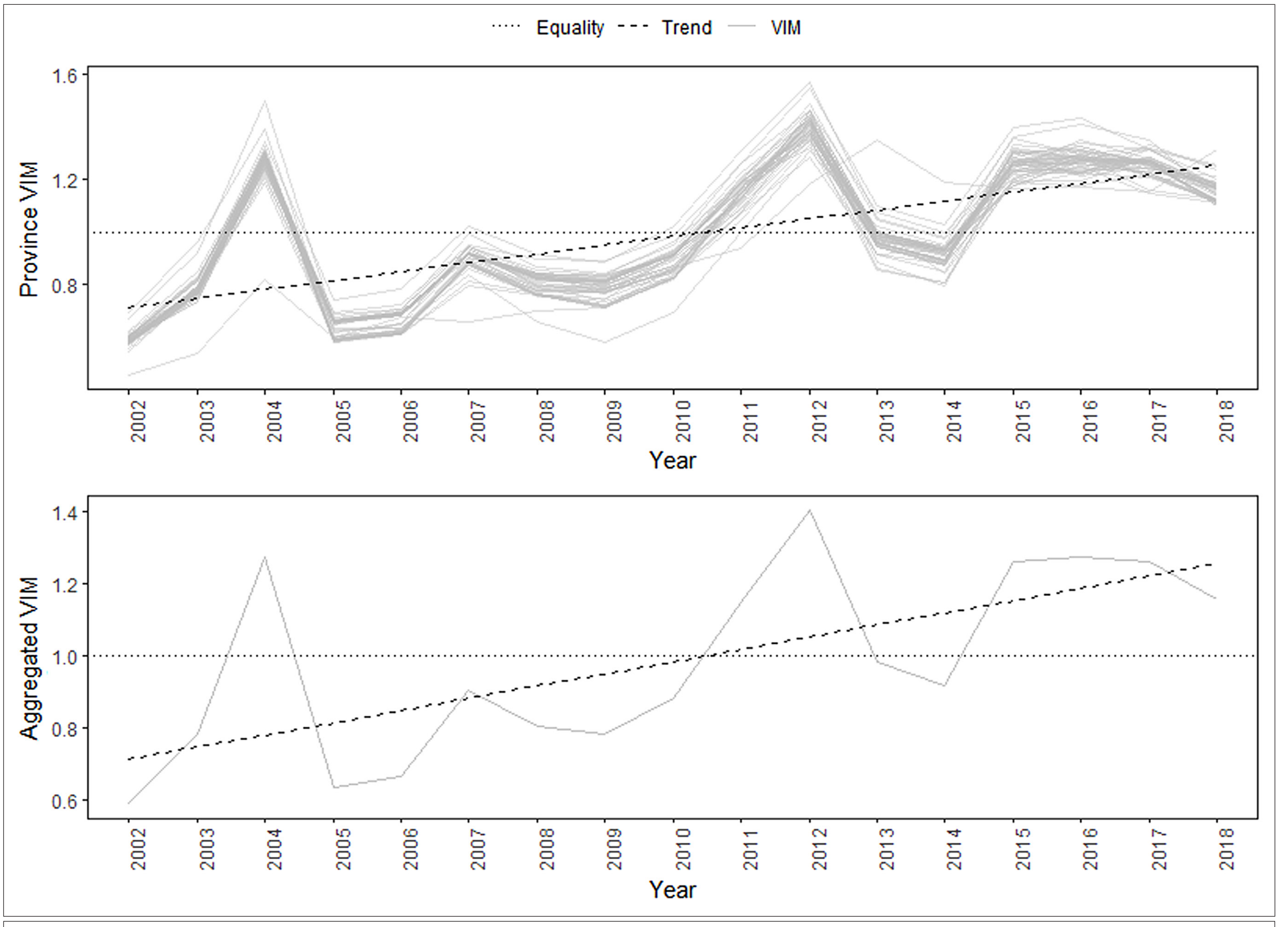

Figure 3. Evolution of Affordability Importance from 2002 to 2018 at Province Level and Aggregated Territory. VIM = Variable Importance Metrics.

also confirm that when measuring the effectiveness of tobacco price increases in reducing demand, it is important to consider the effect of income growth that can offset the effect of cigarette price increases. In other words, the findings of this work suggest that in times of economic growth, the price increases required to effectively reduce tobacco consumption in the population would be greater than the increases required under conditions of slow or no economic growth.

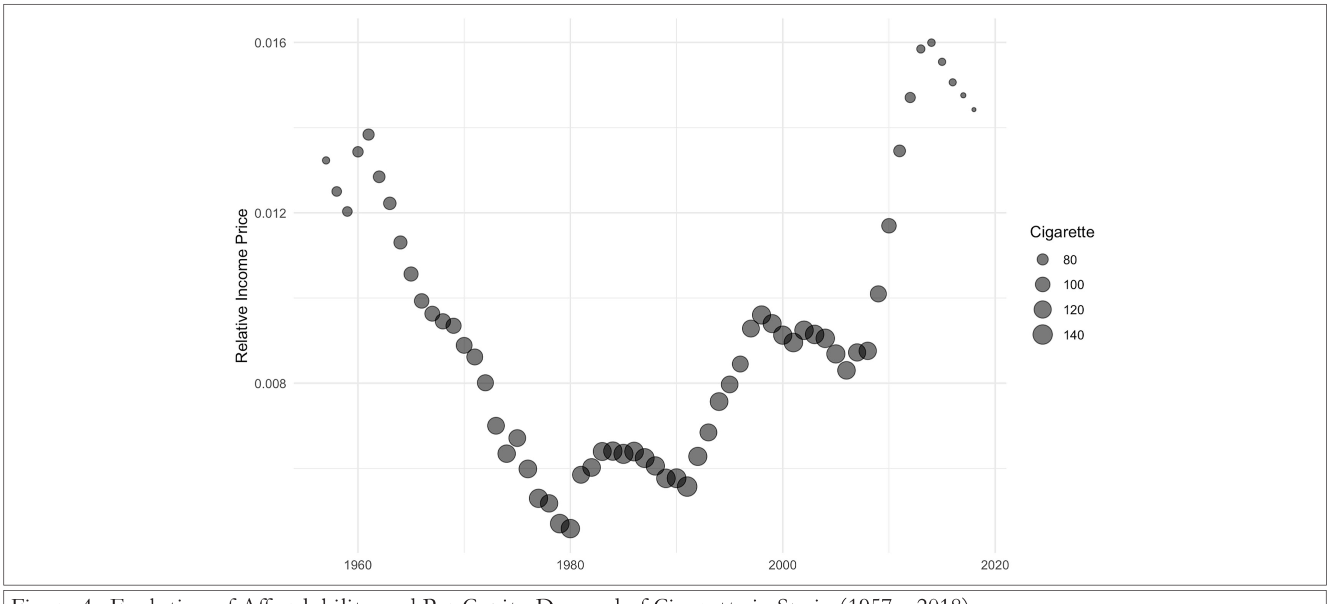

Figure 4. Evolution of Affordability and Per Capita Demand of Cigarette in Spain (1957 – 2018).

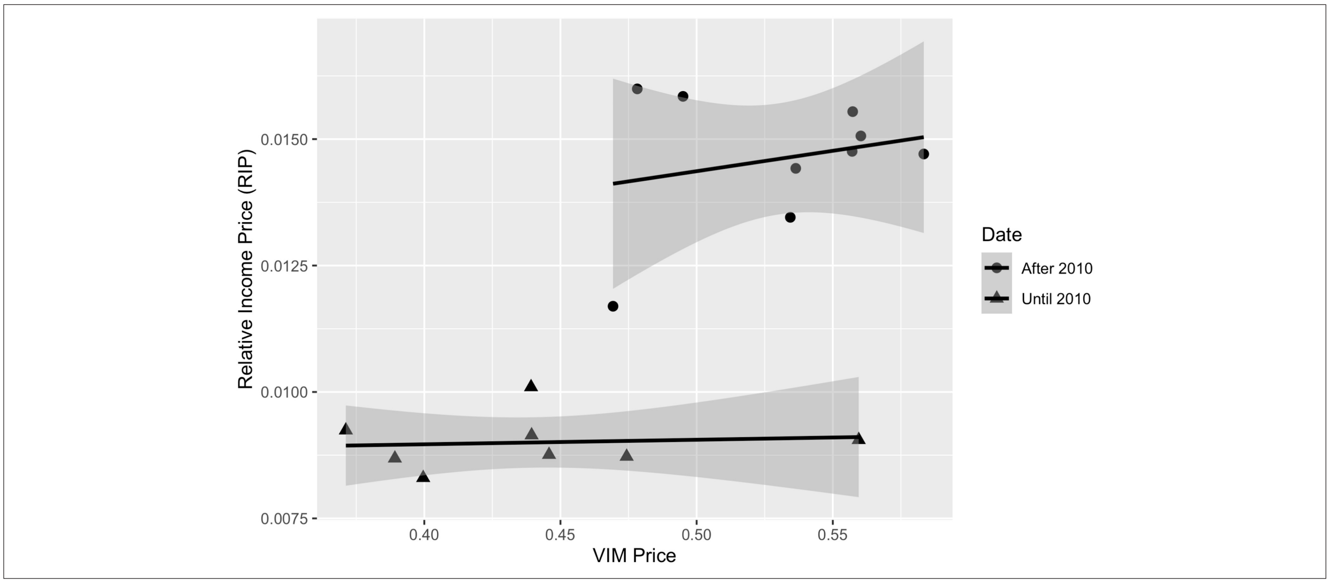

Figure 5. Scatterplot Between RIP and VIM Price. RIP = Relative Income Price; VIM = Variable Importance Metrics.

The main contribution of this article lies in indicating that the importance of affordability to control tobacco consumption in Spain has grown over time. Furthermore, until 2010, income has generally better explained the demand for cigarettes in the Spanish provinces. However, as of 2010, price is the explanatory variable of the demand function that best explains the behavior of the demand for cigarettes. In these circumstances, the separate estimates of price and income elasticity that have been carried out in Spain so far must be interpreted, considering that as of 2010, price is more important than income in explaining the demand for cigarettes. This means that, although the demand functions estimated so far are useful to make predictions about the behavior of cigarette demand, the government must consider that price is a good tool to control tobacco consumption from a certain point of affordability. In other words, for the Spanish government, the price is a more powerful way to control tobacco consumption from 2010 onward. To our knowledge, this is the first attempt to obtain estimates of the explanatory power of the main elements of the tobacco demand function. This is a relevant finding, given that it seems that when the RIP exceeds the value of 1% (as of 2010), the greater this magnitude, the more important the price becomes as an explanatory variable of the demand for cigarettes. The finding found in this paper from 2010 is the first evidence that the 1% RIP may mark a "saturation threshold" that causes consumers to react to changes in price more sharply than before that tipping point.

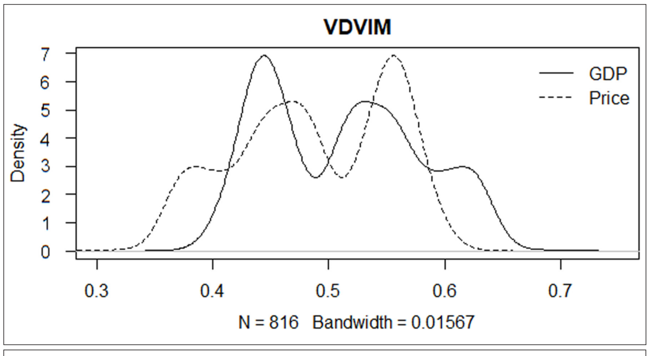

Figure 6. Density Plots for VIM Metrics for QRF Model. QRF = Quantile Regression Forest; VDVIM = Variance Decrease Variable Importance Metrics; VIM = Variable Importance Metrics.

The recommendation for researchers that emerges from this analysis is that price and income elasticity must be contextualized to make effective decisions. In many cases, the conclusions are based on whether the equality of the price and income elasticity parameters is statistically rejected. However, the findings of this paper suggest that up to certain levels of affordability, price is not a fully effective tool for tobacco control.

Affordability, measured by the RIP, is a highly relevant concept in tobacco control. This work shows how price policies are more effective beyond a certain level of affordability. However, although the level of affordability does not allow price to have a significant effect on tobacco use, this concept is an important part in the formulation of tobacco control policies. The concept of affordability involves explaining the combined effects of simultaneous changes in prices and income to policy makers and the relevance of adjusting prices according to income growth and inflation. Price elasticity as a tobacco control tool is a very widespread concept; however, although the ceteris paribus clause is an appropriate starting point, the simultaneous effect of economic growth on tobacco consumption must be considered. The concept of affordability reinforces why policy makers have difficulty understanding why in some cases the effect of price on demand is more pronounced than in others.

Limitations and Directions for Future Research

The results shown in this article are not without limitations. On the one hand, cigarette sales have been used. Although cigarettes account for almost 90% of total tobacco consumption in Spain, part of the elasticity detected may be motivated by the consumption of substitute products. In addition, the use of aggregated data prevents knowing individual behaviors. By using individual data, it could be known whether individual sociodemographic characteristics are influencing the abnormalities detected.

Ethics Committee Approval: Ethical committee approval was received from Universidad Internacional de La Rioja Ethics Committee with the letter dated December 01, 2021, and numbered PI:017/2021.

Informed Consent: N/A.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – P.C., A.A.G., J.M.M – Á., E.A.; Design – P.C., A.A.G., J.M.M – Á., E.A.; Supervision – P.C., A.A.G., J.M.M – Á., E.A.; Materials – P.C., A.A.G., J.M.M – Á., E.A.; Data Collection and/or Processing – P.C., A.A.G., J.M.M – Á., E.A.; Analysis and/or Interpretation – P.C., A.A.G., J.M.M – Á., E.A.; Literature Review – P.C., A.A.G., J.M.M – Á., E.A.; Writing – P.C., A.A.G., J.M.M – Á., E.A.; Critical Review – P.C., A.A.G., J.M.M – Á., E.A.

Declaration of Interests: The authors have no conflicts of interest to declare.

Funding: The authors declared that this study has received no financial support.

References

Almeida, A., Golpe, A. A., Iglesias, J., & Martín Álvarez, J. M. (2021). The price elasticity of cigarettes: New evidence from Spanish regions, 2002 – 2016. *Nicotine and Tobacco Research*, *23*(1), 48 – 56. [CrossRef]

Blecher, E. H., & Van Walbeek, C. P. (2004). An international analysis of cigarette affordability. *Tobacco Control*, *13*(4), 339 – 346. [CrossRef]

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5 – 32. [CrossRef]

Chaloupka, F. J., Straif, K., Leon, M. E., & Working Group, International Agency for Research on Cancer (2011). Effectiveness of tax and price policies in tobacco control. *Tobacco Control*, *20*(3), 235 – 238. [CrossRef]

Chaloupka, F. J., Yurekli, A., & Fong, G. T. (2012). Tobacco taxes as a tobacco control strategy. *Tobacco Control*, *21*(2), 172 – 180. [CrossRef]

Dohoo, I. R., Ducrot, C., Fourichon, C., Donald, A., & Hurnik, D. (1997). An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. *Preventive Veterinary Medicine*, *29*(3), 221 – 239. [CrossRef]

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, *22*(4), 367 – 381. [CrossRef]

Ferreira, A. J., & Figueiredo, M. A. (2012). Boosting algorithms: A review of methods, theory, and applications. In *Ensemble machine learning* (pp. 35 – 85). Boston, MA: Springer.

Freund, Y., Schapire, R., & Abe, N. (1999). A short introduction to boosting. *Journal of-Japanese Society for Artificial Intelligence*, *14*(771 – 780), 1612.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189 – 1232. [CrossRef]

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679 – 688. [CrossRef]

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). *An introduction to statistical learning: with applications in R*. Springer Publishing Company, Incorporated. [CrossRef]

Koenker, R., & Bassett Jr., G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33 – 50. [CrossRef]

Lin, F. J. (2008). Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. *Quality and Quantity*, *42*(3), 417 – 426. [CrossRef]

Martín Álvarez, J. M., Almeida, A., Golpe, A. A., & Vides, J. C. (2021). The influence of cigarette price on the cigarette consumption in Spain: A Logarithmic Mean Divisia Index analysis from 1957 to 2018. *Revista Espanola de Salud Publica*, *95*, 1 – 7.

Martín Álvarez, J. M., Golpe, A. A., Iglesias, J., & Ingelmo, R. (2020). Price and income elasticities of demand for cigarette consumption: What is the association of price and economic activity with cigarette consumption in Spain from 1957 to 2016? *Public Health*, *185*, 275 – 282. [CrossRef]

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, *7*(June), 983 – 999.

Nargis, N., Stoklosa, M., Shang, C., & Drope, J. (2021). Price, income, and affordability as the determinants of tobacco consumption: A practitioner's guide to tobacco taxation. *Nicotine and Tobacco Research*, *23*(1), 40 – 47. [CrossRef]

Russell, M. A. (1973). Changes in cigarette price and consumption by men in Britain, 1946 – 71: A preliminary analysis. *British Journal of Preventive and Social Medicine*, *27*(1), 1 – 7. [CrossRef]

Shang, C., Dauchy, E. P., & Feldman, N. (2020). The price elasticity of demand for heated tobacco products. *Tobacco Prevention and Cessation*, *6*(Suppl.). [CrossRef]

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*(1), 25. [CrossRef]

Sun, Q., & Pfahringer, B. (2011). Bagging ensemble selection. In Australasian Joint Conference on Artificial Intelligence (251 – 260). Berlin, Heidelberg: Springer. [CrossRef]

USNCI, & WHO (2016). *The Economics of Tobacco and Tobacco Control. National Cancer Institute Tobacco Control Monograph 21.* NIH Publication No. 16-CA-8029A. Bethesda, MD and Geneva, Switzerland: U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute and World Health Organization. Retrieved from https://cancercontrol.cancer.gov/brp/tcrb/monographs/21/index.html

Vladislavleva, E., Friedrich, T., Neumann, F., & Wagner, M. (2013). Predicting the energy output of wind farms based on weather data: Important variables and their correlation. *Renewable Energy*, *50*, 236 – 243. [CrossRef]

Wei, P., Lu, Z., & Song, J. (2015). Variable importance analysis: A comprehensive review. *Reliability Engineering and System Safety*, *142*, 399 – 432. [CrossRef]

World Health Organization, & FCTC (2021). Guidelines for implementation of article 6 of the WHO FCTC: Price and tax measures to reduce the demand for tobacco. World Health Organization Framework Convention on Tobacco Control. Retrieved from http:// www.who.int/fctc/guidelines/adopted/Guidelines_article_6.pdf.

Wright, M. N., & Ziegler, A. (2015). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *arXiv preprint arXiv:1508.04409*.

Yang, S., Tian, W., Heo, Y., Meng, Q., & Wei, L. (2015). Variable importance analysis for urban building energy assessment in the presence of correlated factors. *Procedia Engineering*, *121*, 277 – 284. [CrossRef]

Yun, Y. H., Deng, B. C., Cao, D. S., Wang, W. T., & Liang, Y. Z. (2016). Variable importance analysis based on rank aggregation with applications in metabolomics for biomarker discovery. *Analytica chimica acta*, *911*, 27 – 34. [CrossRef]

Zhao, Z. (2008). Parametric and nonparametric models and methods in financial econometrics. *Statistics Surveys*, *2*, 1 – 42. [CrossRef]