U.PORTO
FEP FACULDADE DE ECONOMIA
UNIVERSIDADE DO PORTO

MARKET SIZING FOR "REDUCED-RISK PRODUCTS" IN THE TOBACCO INDUSTRY

Liliana Cristina Correia da Silva

Internship report
Master in Management

Supervised by
Paulo João Figueiredo Cabral Teles

2018/2019

Acknowledgments

First of all, I would like to thank my supervisor Paulo Teles for all the help and patience throughout the development of this dissertation. I couldn't be any more grateful for his active and careful review of all my work, his constant words of encouragement, and for the trust he placed in me.

Thanks to all my colleagues at LTPlabs, especially Paulo Sousa, Paulo Pereira and Teresa Bianchi Aguiar, for being willing to help me whenever it was needed and for the valuable knowledge shared.

I would also like to thank all of my friends who direct or indirectly contributed to the achievement of another step in my academic journey. Thank you for being there not only in moments of joy, but also in moments of sorrow, you are in my heart.

Finally, to my pillars, my family! A big thank you to my parents and my brother for supporting me all the time, for letting me follow my dreams and for providing me with everything needed to successfully reach my goals. You are the best family I could have.

Abstract

Companies that launch innovative products in the marketplace face the challenge of predicting how market size will evolve. This study was motivated by a company of the tobacco industry, following a strategy of future competitiveness in a novel "Reduced-Risk Products" (RRP) category, considered as lower risk to human health than traditional tobacco.

Our purpose is to analyse the dissemination of RRP with the aim of predicting its evolution until 2025 for 40 countries. Despite the existence of several forecast studies, none of them have been focused on the products under analysis. We propose a market size forecast methodology based on diffusion models and complemented by statistical methods and optimization. The data used was collected through public and external sources as well as internal information regarding the market provided by the company.

By applying this methodology, improvements are expected in the decision-making of tobacco companies regarding countries where the vaping segment is already present and in the assessment of the investment involved in entering into a given market. Companies will be able to define the strategic priorities in relation to RRP in order to meet customers' expectations and achieve success in several countries around the world.

Key-words: Bass Diffusion Model, diffusion models, diffusion of innovations, forecasting, innovation, market sizing, Reduced-Risk Products, strategy, tobacco, vaping categories

JEL-Codes: C13, C82, O31, O33

Resumo

As empresas que lançam produtos inovadores enfrentam o desafio de prever a evolução do seu tamanho de mercado. Este estudo foi motivado por uma empresa da indústria do tabaco, que segue uma estratégia de competitividade futura numa nova categoria de produtos "Produtos de Risco Reduzido" (RRP), considerados como de menor risco para a saúde humana do que o tabaco tradicional.

O estudo pretende analisar a disseminação dos RRP, com o objetivo de prever a sua evolução até 2025 para 40 países. Apesar da existência de vários estudos de previsão, nenhum deles é focado nos produtos em análise. Nós propomos uma metodologia para previsão do tamanho do mercado baseada em modelos de difusão e complementada com métodos estatísticos e otimização. Os dados utilizados foram recolhidos através de fontes públicas e externas, bem como informação interna sobre o mercado fornecida pela empresa.

Através da aplicação desta metodologia, esperam-se melhorias na tomada de decisão das empresas da indústria do tabaco no que respeita aos países onde o segmento de vaping já está presente, bem como na avaliação do investimento envolvido na entrada num dado mercado. As empresas serão capazes de definir prioridades estratégicas no que respeita aos RRP, de forma a satisfazer as expectativas dos clientes e a alcançar o sucesso em vários países do mundo.

Palavras-chave: Modelo de Difusão de Bass, modelos de difusão, difusão de inovações, previsão, inovação, dimensão de mercado, Produtos de Risco Reduzido, estratégia, tabaco, categorias de vaping

JEL-Codes: C13, C82, O31, O33

Contents

Acronyms and Symbols

RRP	Reduced-Risk Products
GII	Global Innovation Index
GDP	Gross Domestic Product
GDP*pc*	Gross Domestic Product *per capita*
RMC	Ready Made Cigarettes
Nb	Number
UK	United Kingdom
USA	United States of America

List of Figures

List of Tables

Chapter 1

Introduction

Over the years, it has become increasingly important for companies to be competitive in the market. Competition is rising and customers are becoming more demanding and with ever-increasing expectations. In order to ensure the fulfillment of every customer's needs and deliver the right product at the right time, companies should adopt the most suitable strategic approach and innovation could be a key point. With the enhancement of new technologies, several companies are able to innovate constantly in a number of ways, such as products/services, processes, new markets, new forms of commercial, business or financial organization.

Due to the development of countries, a more informed population and greater health concerns, there is a need for the emergence of new products that best fit these new standards. Facing some of these changes, especially the trend towards higher health concerns, the tobacco industry is, for obvious reasons, one of the most affected. Thus, it is crucial for companies in this sector to find new alternatives to grow and adapt to the new market demands. A few years ago, vaping products or 'Reduced-Risk Products' (RRP), considered as alternative tobacco products that are of lower-risk to human health than traditional tobacco, began to appear. Two main categories of vaping products have emerged: E-vapor, also known as electronic cigarettes, and T-vapor, also referred as heated tobacco products. In this report, the focus will be on the this industry and the analysis will be based on a consulting project we have integrated from October 2018 to March 2019. This type of products might 'revolutionize' the tobacco industry and the purpose of this report is to provide a quantitative analysis in order to understand the market size evolution from 2019 to 2025 in 40 countries which will be fundamental for companies in this sector as a key tool for decision-making.

Despite the existing literature on diffusion models, commonly used to model the evolution of innovative products, the change on customers' behavior, the globalization and the growing competition have unleashed an immensity of new issues. Many studies have been done on the prediction of market size for new or emerging products but there is no evidence of existing studies focused on the kind of products under analysis in this report. Thus, this gap in the current literature will be bridged.

In order to reach the goal of this report, the research was based on the following research questions:

1. What are the main drivers that influence the adoption of this type of products?

2. What is the forecast market size for tobacco products called 'Reduced-Risk Products' between 2019 and 2025?

With this study, companies will be able to decide in which countries the launch of this type of product would be a priority and how to gauge the investment involved in entering a new market. In addition, taking into consideration the expected demand of RRP, they could plan their decisions in detail such as regarding production and stocks management, resource planning, development of new products, marketing strategies and supply chain management, in order to meet consumers' expectations and achieve success in several countries.

For this research, the data used was collected from public and external sources as well as internal market information provided by the tobacco company that requested this study. In order to reach the expected market size, quantitative methods were used, focusing on the Bass Diffusion Model and its refinements, which are described in detail in the next chapter. Additionally, since we faced very limited data for some countries under analysis, we resort to statistical methods to complement the diffusion models approach. First, a range of drivers were selected and their relevance in estimating historical data from 2015 to 2018 was analysed. With this, it was possible to select the most relevant variables to be considered in the forecasting work.

This dissertation is structured in 5 distinct chapters. Chapter 2 covers an overview of the current state of the art on Diffusion Models, including the most relevant aspects of this topic. Chapter 3 presents the problem under analysis in this report. Also, a brief summary of the approach taken to market size estimation for the historical period is discussed. Chapter 4 describes the detailed methodology applied to achieve the final purpose of the study. Chapter 5 depicts the results of the implementation of the proposed methodology. Finally, Chapter 6 summarises the main conclusions drawn and provides suggestions for further work.

Chapter 2

Literature Review

The goal of this chapter is to introduce the reader into the current state of the art in market size forecasting for new products, by covering diffusion models and its applications in the past. Section 2.1 presents the main definitions regarding this theme according to the literature. Then, section 2.2 introduces the Fourt and Woodlock Model and the Bass Diffusion Model, considered as the first diffusion model used in marketing (Radas, 2005). Furthermore, its refinements and extensions provided by several authors will be presented in the subsection 2.2.3 as well as some aspects regarding the international diffusion process in the subsection 2.2.4. Finally, section 2.3 addresses a critical analysis of the literature reviewed.

2.1 Relevant definitions according to the literature

As mentioned before, this research will focus on forecasting the market sizes of 'Reduced-Risk Products' in the tobacco industry and on studying the diffusion process of innovations. We will start by defining some relevant concepts according to the literature, namely, innovation, product innovation, diffusion of innovation, diffusion models and forecasting.

Given competition at a global level, several companies need to distinguish from competitors by launching innovative products or providing innovative services. Over the years, several authors have suggested different definitions for the term 'innovation' depending on the perspective they intended to be focused on. Rogers (1962) define innovation as an "idea, practice, or object that is perceived as new by an individual or other unit of adoption" (p. 11). On the other hand, authors such as OECD and Eurostat (2018), have proposed definitions that require the introduction of a new product/service on the marketplace and are significantly different from the range of products/services previously offered by the firm.

The concept of 'diffusion of innovation' also has various definitions in the literature. Rogers (1962) describes the diffusion as "The process by which an innovation is communicated through certain channels over time among the members of a social system" (p. 5). However, recent authors associate this term to the adoption rate of an innovation (Smith, 2010) and to the market penetration of new products/services (Peres, Muller, & Mahajan, 2010). Furthermore, it is important to understand how innovations are disseminated throughout their life cycle, leading to the consideration of Diffusion Models (Peres et al., 2010).

For the concept of 'Forecasting', the definitions found do not diverge significantly, being perceived as a projection of future events or conditions (Golden, Milewicz, & Herbig, 1994; Mentzer & Moon, 2005). Usually, the projection of a future situation involves the analysis of historical data.

A synthesis of the definition of these relevant concepts is presented in Table 2.1.

2.2 Diffusion Models

Over the years, several authors have sought to understand the spread of innovations in the marketplace throughout their life cycle. The diffusion theory proposes that there is a time gap in the adoption of new products by different members of a society (Schmittlein & Mahajan, 1982).

The forecast of the diffusion of innovations has been a subject studied extensively over several decades, with pioneering studies emerging in the 1960s, such as the Fourt and Woodlock Model. The developments that followed this decade consisted, mainly, of modifying the existing models, specially, adding more flexibility (Meade & Islam, 2006). The purpose of a diffusion model is to present how the dissemination of an innovation will occur over time among a given set of potential adopters (Mahajan, Muller, & Bass, 1990), being the Bass Diffusion Model considered as "the most widely used diffusion model" by Shen (2015, p. 24).

In this study, we will only analyse the diffusion process on the demand side, since the purpose is to understand consumer adoption patterns and not aspects related to the supply behaviour, as addressed by Mansfield (1961).

2.2.1 The Fourt and Woodlock Model

Fourt and Woodlock (1960) sought to predict the success or failure of new grocery products by taking into account consumer panel statistics. By computing sales volume through

Table 2.1: Examples of definitions of relevant concepts

Concept	Definition of the concept	Author(s) (date)
Innovation	"An innovation is a new or improved product or process (or combination thereof) that differs significantly from the unit's previous products or processes and that has been made available to potential users (product) or brought into use by the unit (process)"	OECD and Eurostat (2018, p. 20)
	An innovation could be any "idea, practice, or object that is perceived as new by an individual or other unit of adoption"	Rogers (1962, p. 11)
Product innovation	"A product innovation is a new or improved good or service that differs significantly from the firm's previous good or services and that has been introduced on the market."	OECD and Eurostat (2018, p. 21)
Diffusion of innovation	"Diffusion describes the rate at which consumers adopt the innovation."	Smith (2010, p. 9)
	It is defined as "the process of the market penetration of new products and services, which is driven by social influences"	Peres et al. (2010, p. 91)
	"The process by which an innovation is communicated through certain channels over time among the members of a social system."	Rogers (1962, p. 5)
Diffusion models	Diffusion models "seek to understand the spread of innovations throughout their life cycle"	Peres et al. (2010, p. 91)
Forecasting	"Forecasting is predicting, projecting, or estimating some future event or condition which is outside an organization's control and provides a basis for managerial planning"	Golden et al. (1994, p. 33)
	"Forecasting is a projection into the future of expected demand, given a stated set of environmental conditions"	Mentzer and Moon (2005, p. 41)

number of buyers, frequency of purchase and quantity of product purchased, and by analysing consumer behavior regarding the type of purchase, a separation among initial and repeated acquisitions was established.

The penetration curves model is based on two variables, a ceiling value (the fraction of households expected to be reached when the market is saturated) and a constant of proportionality which, multiplied by the percentage of missing households to achieve the potential market, provides the percentage of adoptions in each period. Once penetration curves have shown a decline in successive increments of new consumers and cumulative penetration is usually much less than 100% of households, high repeat ratios of purchase were required for the success of a product (Fourt & Woodlock, 1960). By assuming similar conditions to those of the first period and by adjusting the repeat ratios of purchase to the type of product and the periodicity considered plausible for a repurchase, the repeat ratios of purchase were computed and, consequently, it was possible to predict the sales volume for the various periods concerned (Fourt & Woodlock, 1960).

2.2.2 The Bass Diffusion Model

Bass (1969) developed a "growth model for the timing of initial purchase of new products" (p. 215). The model was tested for eleven consumer durables and incorporated the following assumptions: the time of the first purchase is linearly related to the number of previous consumers; unit sales of a product coincide with the number of buyers and replacement sales are excluded; the social network is homogeneous and fully connected; two types of influences affect the decision of adoption, namely external influences, such as advertising, and internal influences which are interactions between current and potential adopters in the society; the potential market of a new product is determined at the moment of its introduction and remains constant over time; the diffusion of an innovation is independent from all other innovations that already exist.

Bass (1969) adopted the classification of members of a social system based on the propensity to innovate provided by Rogers (1962, p. 22), which includes the following five classes of innovation adopters: Innovators; Early adopters; Early majority; Late majority; and Laggards. The main difference between the classes is related to the time of innovation adoption. The 'Innovators' are the first adopters of an innovation, whose decision is not affected by the other members of the social system, while the other classes are influenced by the number of previous buyers, being classified as imitators.

First, the probability of purchase at time T, considering that no purchase had been made previously, can be determined by a linear function of the number of previous consumers (Bass, 1969, p. 216),

$$P(T) = p + \frac{q}{m}Y(T) \tag{2.1}$$

where $Y(T)$ is the total number purchasing in the $(0,T)$ interval, being computed by $Y(T) = m\int_0^T f(t)\,dt$, where $f(T)$ is the likelihood of purchase at T. Since $Y(0) = 0$, p is the probability of an initial purchase in the beginning, i.e., the coefficient of innovation or coefficient of external influence; q is the coefficient of imitation or coefficient of internal influence; m is the number of potential adopters of the product. As the number of previous consumers increases, the $\frac{Y(T)}{m}$ ratio rises, reflecting the pressure exercised on imitators. With this, we have (Bass, 1969, p. 217)

$$f(T) = \frac{(p+q)^2}{p}\frac{e^{-(p+q)T}}{\left(\frac{qe^{-(p+q)T}}{p}+1\right)^2}, \tag{2.2}$$

$$F(T) = \int_0^T f(T)dt = \frac{1-e^{-(p+q)T}}{\frac{qe^{-(p+q)T}}{p}+1}, \text{ where } F(0) = 0 \tag{2.3}$$

and

$$P(T) = \frac{f(T)}{1-F(T)} = p + \frac{q}{m}Y(T) = p + qF(T) \tag{2.4}$$

where $F(T)$ is the percentage of potential adopters who have already adopted the product at time T.

Generally, a diffusion curve can have two basic shapes, in part due to the relation between the parameters p and q. A typical diffusion of innovation presents a S-shaped curve as shown in the first graph of Figure 2.1. In this case, the coefficient of innovation (p) is lower than the coefficient of imitation (q), where the internal influence is very relevant (Shen, 2015). On the other hand, when the coefficient of innovation (p) is higher than the coefficient of imitation (q), meaning that the effect of internal influence was smaller than the external one, the cumulative number of adopters increases over time at a decreasing rate (Mahajan &

Peterson, 1985). In this case, the diffusion curve assumes a modified exponential shape, as shown in the second graph of Figure 2.1.

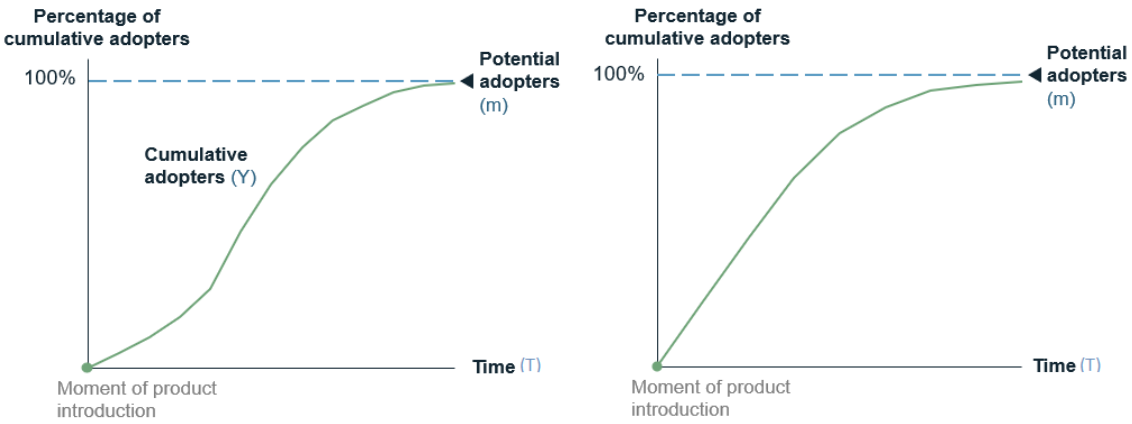

Figure 2.1: Diffusion curves

In order to calculate the growth rate of sales at T, the $S(T)$ function was introduced, where $S(T)$ denotes the sales rate of a given product at time T (Bass, 1969, p. 218),

$$\begin{aligned} S(T) &= mf(T) = P(T)[m - Y(T)] = pm + (q-p)Y(T) - \frac{q}{m[Y(T)]^2} \\ &= \frac{m(p+q)^2}{p\dfrac{e^{-(p+q)T}}{\left(\dfrac{qe^{-(p+q)T}}{p}+1\right)^2}}. \end{aligned} \tag{2.5}$$

The derivative of $S(T)$, denoted by $S'(T)$, is

$$S'(T) = \frac{\dfrac{m(p+q)^3 e^{-(p+q)T}}{p}\left(\dfrac{qe^{-(p+q)T}}{p} - 1\right)}{\left(\dfrac{qe^{-(p+q)T}}{p}+1\right)^3} \tag{2.6}$$

and enables the computation of the time T^* where sales growth is maximum, i.e., the time of the sales growth peak, T^*=$\dfrac{-1}{(p+q)\log\left(\dfrac{p}{q}\right)}$ (Bass, 1969, p. 218). Furthermore (p. 218),

$$S(T^*) = \frac{m(p+q)^2}{4q}, \tag{2.7}$$

$$Y(T^*) = \int_{T^*}^{0} S(t)dt = \frac{m(p+q)}{2q}. \tag{2.8}$$

Additionally, the expected time to purchase is $E(T) = \frac{1}{q}\log\left(\frac{p+q}{p}\right)$ (Bass, 1969, p. 219).

Parameter estimation is another crucial aspect in Bass model. So, in order to estimate p, q and m, from discrete time series data, Bass (1969, p. 219) used the following analogue:

$$S_T = a + bY_{T-1} + cY_{T-1}^2; \text{ where } T = 2, 3, ... \tag{2.9}$$

where S_T denotes sales at time T and $Y_{T-1} = \sum_{t=1}^{T-1} S_{T-1}$, i.e., the cumulative sales from period 1 to $T-1$. In this equation, a estimates pm, b estimates $q-p$ and c estimates $-\frac{q}{m}$. Thus, $q-p = -mc - \frac{a}{m} = b$ and $cm^2 + bm + a = 0$ (Bass, 1969, p. 219). Solving this quadratic equation yields $m = \frac{-b \pm \sqrt{b^2-4ca}}{2c}$. Then, differentiating S_T (Y_{T-1}) with respect to Y_{T-1}, we get $\frac{dS_T}{dY_{T-1}} = b + 2cY_{T-1}$ (Bass, 1969, p. 219). Setting this derivative equal to zero, the solution is (Bass, 1969, p. 219)

$$Y_{T-1}^* = \frac{-b}{2c} = \frac{m(q-p)}{2q} = Y(T^*) \tag{2.10}$$

and consequently

$$S_T(Y_{T-1}^*) = a - \frac{b^2}{2c} + \frac{b^2}{4c} = m\frac{(p+q)^2}{4q} = S(T^*). \tag{2.11}$$

Therefore, Bass (1969) concluded that the time of sales rate $(S(T))$ and of cumulative sales (S_T) maxima is the same.

The model parameters were estimated from the regression in expression (2.9). According to the data and the regression R^2 values for the eleven products analysed by Bass, the model seems to describe the growth rate behaviour very well. The actual sales values were very close to the values predicted by the regression and the estimates of the parameter c are negative as required and those of the parameter m are quite plausible. Despite the observation of some deviations from the trend, the largest were mainly due to short-term income variations.

Concerning forecasting, there are two possible cases with respect to the data available – no-data or limited-data. Bass (1969) argues that the estimation of the parameter m can be achieved through the analysis of the potential market size and, on the other hand, the estimates of p and q can be obtained by analysing the purchase motives. Some parameters can be very sensitive to small variations, mainly when there are few observations, so the plausibility of the estimates should be closely examined. Taking this into account, Bass (1969, p. 223) adopted the continuous model $\int_0^x S(t)dt$ instead of the discrete analogue $\sum_{t=0}^{x=1} S_t$ and, in order to mitigate the possible bias due to this adaptation, the following function of the discrete model was used:

$$\text{if } S_T = S(T), \text{ thus } S_T = a + bk(T)Y_{T-1} + ck^2(T)Y_{T-1}^2 \tag{2.12}$$

where $k(T) = \dfrac{Y(T)}{Y_{T-1}}$.

Bass (1969, p. 223) noted that, for any probability distribution such that $f(x) = \frac{1}{k}[F(x+1) - F(x)]$ and $F(0) = 0$, then $\sum_{t=0}^{x-1} f(t) = \frac{1}{k}F(x)$. Furthermore, the exponential distribution possesses these two properties. Consequently, $\sum_{t=0}^{x-1} \dfrac{F(x)}{f(t)} = k$ for this distribution. In the Bass growth model (Bass, 1969, p. 223), the function $f(T)$ is approximately exponential in character when p and T are small. Thus, $f_{apx}(T) = \dfrac{1}{k[F_{apx}(T+1) - F_{apx}(T)]}$ and $\dfrac{1}{k} = \dfrac{p+q}{e^{(p+q)} - 1}$ (Bass, 1969, p. 223). Furthermore, when T is small, S_T can be written as $S_T = a + b'Y_{T-1} + c'Y_{T-1}^2$, where $b' = kb$ and $c' = k^2c$. Then, $m = km'$, $q = \dfrac{1}{kq'}$ and $p = \dfrac{1}{kp'}$ (Bass, 1969, p. 223).

As a conclusion, parameter estimation derived from regression analysis provided good information regarding sales growth. It provided satisfactory predictions of the time and magnitude of the sales peak for the eleven products studied, as well as exhibited the slow down of growth rates near the peak. Thus, the Bass model is considered very useful for understanding the adoption and diffusion of new products.

2.2.3 Extensions of the main diffusion models

In the fundamental diffusion models of a new product, the fact that the dissemination depends essentially on the variable 'time' and the assumption that marketing decision variables remain constant throughout the life of the product, has led to the need of their extension and refinement (Mahajan & Muller, 1979).

Several modifications proposed are mainly focused on the incorporation of external and marketing mix variables into the estimation of the three main parameters considered in diffusion models (potential market, coefficient of external influence and coefficient of internal influence) or alternative approaches of estimating them. Furthermore, various studies have also emerged attempting to extend diffusion models to successive generations of technologies and to a micro-modeling approach which takes into consideration the heterogeneity of consumers.

2.2.3.1 Extensions concerning the main parameters

The number of potential adopters of a new product (parameter m in the Bass Model, Bass (1969)) is an important parameter for understanding the saturation level that may be reached at the end of a product life. In addition to the characteristics of the market, intentions data of eventual buyers collected through surveys must also be considered, since it could be a complementary source to perceive customers' behaviour (Bass, 2004).

Instead of being determined at the time of product introduction and not varying over time, some authors argue that the potential market (m) can be considered as a function of relevant endogenous or exogenous variables, such as population growth (Sharif & Ramanathan, 1981), increase of the number of households (Peterson & Mahajan, 1978), GDP-related variables (Islam & Meade, 1996), increase of the number of retailers which commercialize the product (Jones & Ritz, 1991), technological changes or government measures (Mahajan & Muller, 1979).

On the other hand, many authors have defended that some marketing relevant variables may be determinant in the estimation of potential market over the years, such as the price of the product (Horsky, 1990; Kalish, 1985; Peterson & Mahajan, 1978) and advertising (Dodson & Muller, 1978).

Horsky (1990) claimed that there is a negative relation between the saturation level and two other variables: the average price of the product and the dispersion of the distribution of income. The proportion of the population that will buy the new durable product is represented by $P(p_i < w)$, where p_i represents the effective price of the product and w expresses the wage rate (Horsky, 1990, p. 346). Thus, a reduction in the price of the product can make it accessible to a group of individuals with narrower budgetary constraints, leading to an increase of the number of potential buyers. In its turn, Kalish (1985) argued that potential adopters were the individuals whose reservation level (level to which the purchase of the product was acceptable) was higher than the risk adjusted price of buying the product.

Several authors have argued that also the probability of adoption, which includes the coefficients of external and internal influence (parameters p and q in the Bass Model, Bass (1969)), should be estimated considering marketing mix variables, such as advertising and price.

Authors like Thompson and Teng (1984) and Simon and Sebastian (1987) stated that the coefficient of imitation should be linked to advertising actions, both past and current advertising expenditures, once the effect may extend over time. In addition, Horsky and Simon (1978) suggested that the coefficient of external influence should reflect advertising expenses, since it may have an impact on the decision of adoption of innovators, i.e.,

$$p(T) = p_1 + a_2 A(T) \tag{2.13}$$

where $p(T)$ expresses the coefficient of external influence, $A(T)$ represents the level of advertising expenditures at time T, p_1 and a_2 are constants.

Additionally, also the price of the product is a fundamental variable for the decision-making of its purchase. The introduction of a price index into the Bass hazard function, in order to investigate pricing strategies have been done in various studies (Dolan & Jeuland, 1981; Kalish & Lilien, 1983; Robinson & Lakhani, 1975). Robinson and Lakhani (1975) stated that an optimal constant price or an optimal pricing strategy based on a low starting price that increases to a peak and decreases again were the most appropriate. Furthermore, Robinson and Lakhani (1975) related the coefficient of internal influence with the price of the product, i.e.,

$$q(T) = q_1 \exp(-EP(T)) \tag{2.14}$$

where $q(T)$ is the coefficient of internal influence, q_1 is a constant, E is the price elasticity, and $P(T)$ is the price at time T.

Being aware of the emergence of several extensions to the initial diffusion models, Bass, along with two other authors, also developed a generalized version of his initial diffusion model of 1969. Bass, Krishnan, and Jain (1994) asserted that the coefficient of innovation and the coefficient of imitation may change with time, adding the variable $x(T)$ to the equation

(2.4), which represents the "current marketing effort" (p. 204), i.e.,

$$P(T) = \frac{f(T)}{1-F(T)} = [p+qF(T)]x(T). \tag{2.15}$$

Considering two decision variables, the price of the product at time T denoted by $P_r(T)$ and advertising at time T denoted by $\text{Adv}(T)$, the variable $x(T)$ can be written as (Bass et al., 1994, p.207):

$$x(T) = 1 + \left[\frac{P_r(T) - P_r(T-1)}{P_r(T-1)}\right]\beta_1 + \left[\frac{\text{Adv}(T) - \text{Adv}(T-1)}{\text{Adv}(T-1)}\right]\beta_2 \tag{2.16}$$

being expected that parameters β_1 and β_2 have negative and positive signs, respectively.

Another aspect highlighted by authors such as Peterson and Mahajan (1978) is the relation between an innovation and others innovations already present in the marketplace. This is more critical when the products analysed are complementary (e.g. washers and dryers) or substitutes, once future sales of one of them have influence on the sales of the other. Taking as an example two complementary products, the coefficient of internal influence of product 1 should be computed considering also some elements related to product 2, i.e.,

$$q(T) = q_1 + q_2\frac{N_2(T)}{N_1(T)}, \tag{2.17}$$

where q_1 expresses the coefficient of internal influence among the potential buyers and current adopters at time T of product 1, q_2 is the positive influence of the buyers of product 2 on the potential adopters of product 1, $N_1(T)$ and $N_2(T)$ represent the number of buyers of product 1 and product 2 at time T, respectively.

2.2.3.2 Alternative approaches to the estimation of the main parameters

An alternative approach for the estimation of the coefficients of innovation and imitation for a given product was adopted for several authors, such as Tansurat and Gerdsri (2015) and Abu and Ismail (2013).

Being aware that technological changes have a huge impact on almost every company, Tansurat and Gerdsri (2015) attempted to predict the diffusion pattern of the OLED (Organic Light Emitting Diodes) technology in portable devices (smartphones and tablets). The analysis was based on the Bass Diffusion Model and the estimation of the parameters p and

q derived from a type of product considered to be analogous, in this case, the cellular telephone. Tansurat and Gerdsri (2015) concluded that the current cumulative sales of tablets and smartphones were in line with the cumulative sales obtained with the Bass model.

Additionally, Abu and Ismail (2013) carried out a study to predict demand for a new car denominated 'Inspira' launched by a Malaysian car manufacturing company. Since the available data on the demand for 'Inspira' were limited, Abu and Ismail (2013) chose to carry out the forecast by analogy. First, it was identified an existing car with similar characteristics to the new one regarding structural features, level of need, and pattern of sales and growth. After that, the estimation of the diffusion parameters (p and q) could be obtained using the diffusion history of the analoge product. Abu and Ismail (2013) concluded that the estimation of p and q and the results obtained by the Bass model provided a robust and efficient demand forecast. Furthermore, it was found that the use of p and q values of similar products with identical characteristics can be useful for prediction.

2.2.3.3 Successive generations of technologies

The time between the emergence of successive generations of several products, especially, high-technology products, has decreased. This has led to a growing need to understand the impact of a recent generation on previous ones, so the substitution effects must be taken into consideration (Norton & Bass, 1987). Successive generations of this type of products will lead to the reduction of the potential market of the earlier generations, once some potential customers will prefer the most recent version and some current consumers may switch for this one (Norton & Bass, 1987).

In the equation (2.18), we can observe the sales forecast of a product before the emergence of a recent generation, where $S_i(t)$ represents the sales of generation i in period T, m_i refers to the initial potential market for product i, and $F_i(T)$ is the percentage of potential adopters of product i who have already adopted the product at time T (Norton & Bass, 1987, p. 1074).

$$S_i(T) = m_i F_i(T). \tag{2.18}$$

Typically, the incorporation of a new technology into a given product happens before the previous generation reaches the saturation level, causing an impact on the potential market of both generations. In the expression (2.19), where T_2 expresses the time of introduction of the second generation, we can notice the decrease on the potential market of generation 1

(Norton & Bass, 1987, p. 1074):

$$S_1(T) = F_1(T)m_1 - F_2(T - T_2)F_1(T)m_1, \text{ for } T > 0. \tag{2.19}$$

On the other hand, the potential market of generation 2 is also influenced, increasing with respect to the initial value, as present in the expression (2.20) (Norton & Bass, 1987, p. 1074).

$$S_2(T) = F_2(T - T_2)[m_2 + F_1(T)m_1], \text{ for } T > T_2. \tag{2.20}$$

This extension of a basic diffusion model can be applied to several generations simultaneously.

2.2.3.4 A micro-modeling approach

Chatterjee and Eliashberg (1990), being aware that members of the population are heterogeneous with respect to individual preferences, initial perceptions about the performance of an innovation, and reliance on perceived information about a product, developed a diffusion model which incorporates a micromodeling approach. Chatterjee and Eliashberg (1990) stated that innovation's performance, degree of risk averse, price of the innovation, price sensitivity and responsiveness to product information were the main determinants considered for the decision of adoption.

Despite of the availability of product's price information a priori, there is uncertainty regarding the product's performance. The authors defended that as the potential adopters have access to more information about the product's performance, their perceptions may change, affecting the expected utility and, consequently, the time of adoption. Chatterjee and Eliashberg (1990) assumed that there were three types of consumers taking into account the distance to the adoption before the product launch and the consumer's reservation price. Diversified individual characteristics explains the variety of diffusion patterns and, in turn, the diffusion curve is reached by aggregating the behavior in terms of adoption time at individual level among the potential adopters.

2.2.4 International diffusion of innovations

The process of diffusion varies greatly among countries, even when dealing with the same products or countries that belong to the same continent (Ganesh, 1998; Helsen, Jedidi, & DeSarbo, 1993; Mahajan & Muller, 1994). Talukdar, Sudhir, and Ainslie (2002) argued that when companies intend to adopt market expansion strategies, they must assess the attractiveness of a market which is related to the potential market and the speed of diffusion of the product in the market. Thus, in the study of international diffusion of innovation, the analysis of the acceptance of multinational products is important.

Concerning the influence of the moment of innovation introduction on the speed of diffusion, several authors such as Dekimpe, Parker, and Sarvary (2000); Ganesh, Kumar, and Subramaniam (1997); Takada and Jain (1991); Tellis, Stremersch, and Yin (2003) claim that, in countries where innovation was introduced later, the diffusion process is faster and, according to Van Everdingen, Fok, and Stremersch (2009), the time to takeoff is shorter. This means that the entry time lag may have a positive impact on the process of diffusion. Furthermore, customers from a country considered the innovation level of acceptance in other countries as a signal which leads to the reduction of their perceptions of risk and increases the legitimacy of the adoption of the product in question (Dekimpe et al., 2000; Takada & Jain, 1991).

Concerning developed and developing countries, some differences were noted about the adoption and diffusion of new products (Talukdar et al., 2002). These authors used the discrete time version of Bass Diffusion Model, and through the Hierarchical Bayes estimation methodology to estimate the model, studied the diffusion of 6 product categories during the first 9 years, in 31 countries, including developed and developing ones. Despite the positive impact of delayed product introduction on the speed of product adoption in developing countries, the adoption rate is slower than in developed countries. In developing countries, the achievement of peak sales is on average 17.9% longer and the average penetration potential is approximately one-third of that for developed countries. Moreover, positive changes in urbanization level or international trade can lead to the increase of the potential penetration (Talukdar et al., 2002).

Country's cultural characteristics can also affect the diffusion of a product (Takada & Jain, 1991), as well as macroeconomic variables. A country's wealth, usually measured by GDP per capita, positively influences the diffusion of a product (Dekimpe et al., 2000; Desiraju, Nair, & Chintagunta, 2004; Helsen et al., 1993; Putsis Jr, Balasubramanian, Kaplan, & Sen, 1997; Talukdar et al., 2002).

In addition, the 'neighborhood effect' across countries can also be incorporated into diffusion models (Brown, 1981; Gore & Lavaraj, 1987). Peterson and Mahajan (1978) developed a study in the United States of America with the aim to understand how geographical boundaries affect the introduction of an innovation in the surrounding markets. The authors concluded that the relative number of total adoptions is higher in markets closer to the originating market of innovation, i.e. the neighborhood effect decreases with increasing distance to this market, thus reducing the size of the potential market of more distant markets.

Furthermore, Talukdar et al. (2002) have found that previous processes of adoption and diffusion of other products in one country (country effect) may be useful to explain the penetration level, whereas the processes of adoption and diffusion of a specific product in several countries (product effect) where it has been previously introduced may be useful to explain the coefficients of external and internal influence. With that, Talukdar et al. (2002) concluded that by combining data on patterns of past diffusions across countries and products, predictive power can be improved.

2.3 Critical analysis of the literature reviewed

Diffusion models, particularly the Bass Diffusion Model, have been significantly used in understanding the dissemination of innovations through their life cycle. The results of the studies that used the Bass model were very satisfactory. However, there are some aspects that should be mentioned.

First, the Bass model was developed under a series of assumptions which could be a limitation in its application. The need to relax some of them has been felt by several authors over the years, who focused their work on refining this methodology in order to improve the accuracy of the forecast. Furthermore, several extensions have been considered, such as the use of diffusion models for successive generations of technologies.

Secondly, the Bass model was mostly based and applied on durable consumer products. Despite the differences between the type of products under analysis and those for which Bass model was mainly applied, some relevant conclusions concerning both the discrepancies among different countries and the relation between some variables and the parameter estimation were drawn. Such works have provided good contributions to our study since we will deal with 40 countries around the globe. While the devices of vaping products could be considered as a durable product, this is not the case of the refills (see Chapter 3). However, the number of purchases of devices may be a good proxy for the number of individual users of RRP. Lastly, the lack of literature on similar products such as traditional cigarettes is a limitation for related studies.

In this study, the Bass Diffusion Model will be the base model and some additional variables will be considered in order to refine its application and adapt it to the products in question.

Chapter 3

The Problem

In the first section of this chapter, a brief introduction to RRP will be presented. Subsequently, the purpose of this study will be depicted in detail, including the several dimensions and granularity levels under analysis.

Section 3.2 is focused on data collection and is divided into two subsections on distinct themes. In subsection 3.2.1, the process of estimating historical data will be briefly addressed, as well as the process that provided the computation of the market size through three dimensions. On the other hand, in subsection 3.2.2, several related market variables will be described.

3.1 Problem description

Some years ago, vaping products, considered as alternative tobacco products that are of lower-risk to human health than traditional tobacco, began to appear. There are two main categories of vaping products considered in the market: E-vapor, also known as electronic cigarettes, and T-vapor, also referred as heated tobacco products. An electronic cigarette is a device that does not contain tobacco and "creates an inhalable vapor by electronically heating a liquid" (Japan Tobacco International – A global tobacco company, 2019). On the other hand, heated tobacco products contain tobacco that is heated (not combusted) to create an inhalable vapor (Japan Tobacco International – A global tobacco company, 2019). Each category brings its own distinct characteristics, appealing to different consumers with diverse needs.

This segment of products might have a significant impact on the tobacco evolution, so the purpose of this study is a quantitative analysis of its market size from 2019 to 2025, on an annual basis, in 40 countries from five continents. The market size will be computed in different dimensions and granularity levels in order to achieve a more detailed approach, as explained below.

The three major dimensions under analysis are value, volume and number of consumers. The 'value' dimension consists in the prediction of the market value in $\$US$. The same currency is considered for all countries by the respective exchange rates, in order to allow a direct comparison of the values among the markets. Concerning the 'volume' dimension, the metric of 'sticks equivalents' is used in order to convert the different units of each product category into a comparable metric between the several vaping products and also, between this segment and traditional tobacco. Finally, the dimension of 'number of consumers' consists of the determination of the number of individuals who consume regularly some vaping product.

Within the main dimensions, a more specific granularity was considered taking into account three kinds of market splits: product category, item type, and consumer type. In what concerns product category, the products under analysis were separated into secondary categories: 'E-vapor' was divided into Cigalike, Open tank, Pen and Liquid pod and 'T-vapor' was divided into Tobacco stick and Infused tobacco. A brief description of each product category is presented in Table 3.1.

Concerning the 'consumer type', the division was made between two different types, dualists and solus. Dualists are the consumers who currently smoke both RRP and traditional tobacco, and, on the other hand, solus are the adopters who came from traditional tobacco or who did not smoke before and exclusively consume vaping products. Moreover, in the 'item type' split, a distinction between devices and refills was considered only for the 'value' dimension. This is due to the fact that the 'volume' dimension corresponds almost exclusively to refills and it does not make sense to divide the number of consumers through this metric.

3.2 Data collection

In order to estimate the market size in the three main dimensions mentioned in the previous section, we started by collecting the data for the past period. First, we obtained the estimation of historical market sizes for the period between 2015 and 2018 for the different countries under analysis through several available data sources. Subsequently, we collected data of several variables that could be useful for the estimation of the main parameters considered in the Bass Diffusion Model (Bass, 1969), which was the main model considered in the methodology chosen for this study, and for the statistical methods used.

Table 3.1: RRP product type definition

Product category	Definition
E-vapor	An electronic cigarette is a device that does not contain tobacco and creates an inhalable vapor by electronically heating a liquid
Cigalike	Products with pre-filled liquid capsules similars to a cigarette concerning the shape and size, with single use or with interchangeable capsules
Pen	Products with a format similar to a pen with pre-filled liquid interchangeable capsules
Liquid Pod	Products with a pod format with pre-filled liquid interchangeable capsules
Open tank	Products that can be refilled with liquid manually from liquid bottles
T-vapor	Products that contain tobacco which is heated (not combusted) to create an inhalable vapor
Tobacco stick	Device and consumables of directly heated tobacco, in the form of a cigarette
Infused tobacco	Device and consumables of indirectly heated tobacco

3.2.1 Historical data

First of all, the collection of historical data on key dimensions and splits was needed as a starting point for predicting values for the near future.

Since in the tobacco industry, companies were unaware of the exact values of market sizes of vaping category from previous years, data was collected from a number of sources, including external sources and internal information on the tobacco market provided by the company of the tobacco industry that requested this study. Available data sources by type of information are presented in Table 3.2[1]. However, it is important to clarify that some data sources did not provide information for all the countries under analysis, but only cover a part of them.

[1]Due to confidentiality issues, the names of the external data sources considered could not be shown.

Table 3.2: Data sources — Historical period

Source type	Data source	Content (granularity)
External research or report	External source 1	Sales value/volume (country/product)
	External source 2	Sales value (country)
	External source 3	Sales value (country/product type), price
	Players' reports	Sales value/volume (country)
Internal data or study	Internal surveys	Incidence, consumption ($US, sticks)
	Non-RRP market model[a]	Sales value/volume (country)
	Country-specific report[b]	Consumption ($US, sticks)

[a]Traditional tobacco [b]Available data source for a few countries

For the historical period that, in this study, corresponds to the time interval between 2015 and 2018, and taking into consideration the limited available data, estimates were obtained by two different ways in order to achieve a larger number of estimates for the same dimension. One of them consisted of extracting the final values provided by each data source while the second one was an indirect way. It consisted of combining information on different variables from several data sources, resulting in a total of 7 estimates. This approach was adopted since a higher number of estimates may lead to a better assessment of the credibility of each data source by understanding the discrepancy among the values found.

As shown in Table 3.2, the dimension 'value' was the one with the most available data sources, so it will be the starting point for calculating the remaining dimensions. In order to obtain the final range of estimates for the 'value' dimension, four approaches were considered, namely, traditional retail sales, non-RRP market, specific category's sales, and direct estimate, as detailed next.

Retail is the most traditional channel to sell vaping products and many estimates are only carried for this channel. In the 'traditional retail sales' approach, the market value of vaping products of country $\boldsymbol{A}$ at time $\boldsymbol{T}$ in \$*US*, represented by $\boldsymbol{MV_{TA}}$, is obtained through the following expression,

$$MV_{TA} = \frac{\text{Ret.sales}_{TA}}{\text{Ret.share}_{TA}} \tag{3.1}$$

where the 'Ret.sales' variable represents the total of RRP sales in the traditional retail and the variable 'Ret.share' denotes the share of market of traditional retail in $\boldsymbol{T}$, i.e., the percentage of vaping products sales from traditional retail relatively to the total sales of this type of product in T. With this approach, estimates provided by External source 1, which includes only traditional retail data, can be considered for countries with online commerce.

The non-RRP market value has decreased in several countries due to the emergence of the alternative tobacco products under analysis. Taking this into consideration, the 'non-RRP market' approach consists in comparing the actual non-RRP market value with the expected market value, assuming that it would follow the 'natural evolution' observed in the historical period. The 'natural evolution' was estimated based on the non-RRP value growth trend over the past 10 years. This approach allows the understanding of the fall caused in the non-RRP market due to the entrance of RRP in the market, and by applying a conversion rate, the market value of these innovative products is found,

$$MV_{TA} = (NE_{TA} - AM_{TA}) \times CR_{TA} \tag{3.2}$$

where NE represents the value of non-RRP market considering the 'natural evolution', AM denotes the actual non-RRP market value, and CR denotes the conversion rate from non-RRP value to RRP value.

The 'specific category's sales' approach is exclusively directed to data sources that provide only information on a specific category of vaping products, as is the case of External source 3 which covers only the E-vapor category. Thus, in order to allow the computation of the total market value through this data source, we consider the market share of E-vapor category, as represented in the expression (3.3).

$$MV_{TA} = \frac{\text{Cat.sales}_{TA}}{\text{Cat.share}_{TA}} \tag{3.3}$$

where the 'Cat.sales' variable denotes the RRP sales of a specific category of products and the variable 'Cat.share' represents the share of market of the same category in the total RRP market at time T for a given country.

In the last approach considered, the 'direct estimate', we collected estimates directly from a specific data source, such as External source 2 and country-specific reports. In these sources, the market value of RRP is available for several countries.

After compiling all the estimates provided by the several approaches, we noticed that each estimate of market value was yielding different results, which led to the need of quantifying the reliability of each data source and assign them different weights. The goal was to obtain a final market estimate for each country being weighted by a combination of distinct estimates. In Figure 3.1, the determination of the 'best guess' market value is represented with the flag for three of the countries under study. Therefore, a mathematical optimization problem was framed in which the decision variables to be optimized were the reliability weights of all estimates and the objective was to minimize the distance between each estimate and the 'best guess' market value. The weights were optimized for each country based on

$$\min \sum_i \sum_j [MV_{Tij} - \sum_k (w_{Tik} \times MV_{Tik})]^2, \tag{3.4}$$

$$dist_{Tij} = [MV_{Tij} - \sum_k (w_{Tik} \times MV_{Tik})]^2 \tag{3.5}$$

and

$$w_{Tij} = \frac{\sum_j dist_{Tij}}{dist_{Tij}} \tag{3.6}$$

where T is the year T, i represents a given country, j and k represent a given data source, MV_{Tij} is the market value provided by a specific data source for a given country, $\sum_k (w_{Tik} \times MV_{Tik})$ is the 'best guess' market value for the country i, $dist_{Tij}$ expresses the distance between the market value provided by the data source j and the 'best guess' market value for the country i at time T. The w_{Tij} represents the weight assigned to each estimate to be optimized for a given country and intends to prioritize estimates closer to the 'best guess' market value. In the end of the optimization process, the w_{Tij} were normalized in order to yield values between 0 and 1, leading to the final reliability weight assigned to each estimate for a given country.

To solve this problem we have used a heuristic where, first, for all countries, 100% of reliability was equally distributed for each available estimate provided by the data sources. Then, by multiplying the respective confidence-based weight by the value provided by each estimate and adding the weighted values, a preliminary 'best guess' for the final result was found for each of the 40 countries.

Subsequently, the optimization process began trying to minimize the quadratic distance of each estimate to the preliminary 'best guess' market value. At each interaction a new 'best

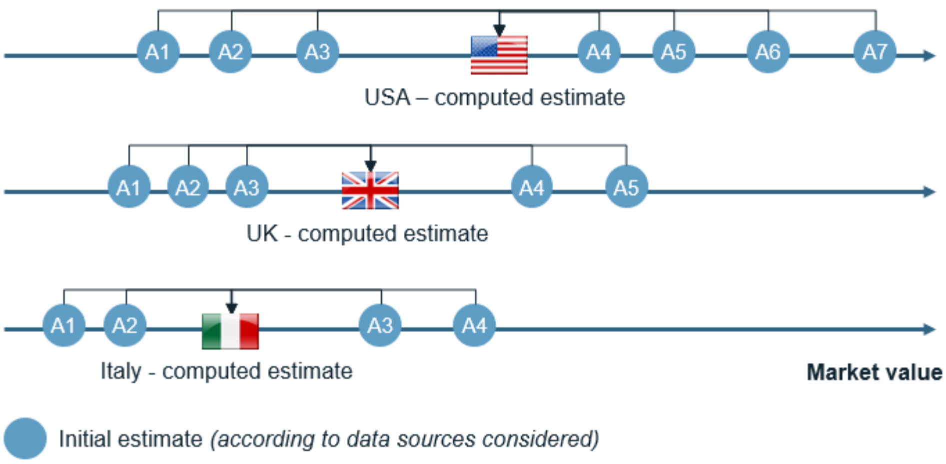

Figure 3.1: Determination of the 'best guess'

guess' was determined and a greater reliability weight was given to data sources whose estimates were closer to the 'best guess' (expression 3.6). Convergence was achieved after 200 iterations and the optimization process ended. The final 'best guess' market value and reliability weights were collected and assigned to each estimate. In order to improve the reliability assigned to each data source, each country was assigned to a particular cluster through the k-means clustering method (Marôco, 2018, Chapter 11) taking into consideration the range of available data sources for each country and the quadratic distance of each estimate to the 'best guess' market value. This was considered since some data sources were more reliable for some specific group of countries than others.

The results of the dimension 'value' were estimated and with the aim of computing the dimension 'volume', the variable 'average price' for each product category was applied to the market value (expression (3.7)) and, on the other hand, for the dimension 'number of consumers', the 'average consumption per adopter' was used (expression (3.8)).

$$Mvol_T = \frac{MV_T}{Price_T}, \tag{3.7}$$

$$Cons_T = \frac{Mvol_T}{CPA_T} \tag{3.8}$$

where the $Mvol$ represents the RRP market volume, $Price$ is the RRP average price, $Cons$ denotes the number of consumers and CPA denotes consumption of sticks equivalents per adopter in T. The procedure that links the three main dimensions under analysis is displayed in Figure 3.2.

Figure 3.2: Procedure linking the main dimensions — Historical period

For the three categories of splits mentioned above, the available data were more scarce, once the detail needed was very specific. Surveys were the main source for the computation of the 'consumer type' division. This data source provided information on incidence and consumption separately for solos and dualists, allowing the breakdown of the market value by type of consumer. On the other hand, for product-related splits (category and item type), the main sources used were the external reports. These data sources provided countries' sales value, broken down by product reference, allowing the sales value to be obtained by product type and item type.

3.2.2 Variables

The dependent variables of this study were the RRP's market size in the three dimensions mentioned above. Several variables have already been identified as potential forecast drivers for the estimation of these market sizes. The potential explanatory variables were divided into four classes: macroeconomic variables, market environment, non-RRP data and geographical context (Table 3.3).

Concerning macroeconomic variables, two indicators were selected: Gross Domestic Product *per capita* (GDP*pc*) in $*US* and Global Innovation Index. A positive impact on the size of RRP's market is expected for both these variables. Authors like Talukdar et al. (2002) and Desiraju et al. (2004) state that (GDP*pc*) positively influences the diffusion of a product. "The Global Innovation Index (GII), by creating metrics through which innovation can be measured across the globe, helps identify ways that innovation can better serve society and the challenges we face" (Cornell University, INSEAD, & WIPO, 2018, p. 9). We assume that the Global Innovation Index is a proxy for the opening of countries to innovations, so it is expected that countries with a higher GII will be more receptive to innovative products.

The second class corresponds to the variables of market environment composed by four indicators: time since RRP introduction (years), number of products offered, RRP price (in $*US*) and number of RRP players present in the considered country. The variable of time since RRP introduction in a specific market is considered as a relevant potential variable because the time factor is expected to have a positive influence on market size, i.e., it is expected that in the first years after the introduction of a product, the market size continuously increases. On the other hand, in our view, the number of products offered and the number of RRP players present are related to the access to the product. Thus, in countries where the access to the product is facilitated (e.g. higher number of players with introduced products), the adoption of a product is positively affected (Talukdar et al., 2002). Moreover, the price of a product is a factor that influences the purchase decision and should be included as an indicator (Horsky, 1990; Kalish, 1985).

Concerning the class of the 'non-RRP data', we selected three indicators: smoking incidence (%), share of premium RMC[2] (%), smoking bans (existence/nonexistence) and non-RRP price (in $*US*). In countries with higher smoking incidence and smoking bans of traditional tobacco – for example, indoor smoking – the likelihood of existing more smokers of vaping products is higher and, consequently, the size of the RRP market is positively affected. The share of premium RMC, which represents the proportion of traditional premium tobacco brands in the total RMC market value, could be an indicator of the living standards of a country's population. Furthermore, in countries where RRP price is significant lower than non-RRP price, it is expected that more smokers will be more willing to adopt vaping products.

The geographical context includes several indicators, such as the GDP*pc*, time since RRP introduction and number of vaping products available in neighboring countries. Considering the 'neighborhood effect', geographical boundaries affect the introduction and diffusion of an innovation in the surrounding markets (Peterson & Mahajan, 1978), so a positive relationship between these variables and the size of vaping products market is expected.

[2] Ready Made Cigarettes

Table 3.3: Chosen variables for the forecast model

Class	**Variables**
Macroeconomic variables	GDP*pc* Global Innovation Index
Market environment	Time since RRP introduction Number of products offered RRP price Number of RRP players
Non-RRP data	Smoking incidence Share of premium RMC Smoking bans of traditional tobacco Non-RRP price
Geographical context	GDP*pc* in neighboring countries Number of products offered in neighboring countries Time since RRP introduction in neighboring countries

Chapter 4

Forecasting Methodology

This chapter addresses the methodology developed to predict the size of RRP market in the tobacco industry for 40 countries. The Bass Diffusion Model (Bass, 1969) corresponds to the chosen base model, so in section 4.1, the estimation of its main parameters will be explained. Furthermore, the alternative approaches considered will also be detailed. In section 4.2, the procedure adopted to reach the forecast of the three main dimensions considered in this study ('value', 'volume' and 'number of consumers') for the period from 2019 to 2025 will be addressed.

4.1 Parameter estimation

The Bass Diffusion Model was used for the forecast of RRP market size in the three dimensions addressed in chapter 3. The equation (2.3) of Bass model regarding the percentage of potential adopters who have already adopted the product at time T $(F(T))$, based on three parameters m, p and q, was the main expression considered in this methodology. In this section, we will explain how these three parameters were estimated.

As explained in chapter 2, one of the assumptions of Bass model (Bass, 1969) is the equality between the number of consumers and the number of purchases. However, in our study, although the item type 'device' could be considered a durable product, we are also dealing with non-durable products when considering the item type 'refill'. Thus, we had to relax this assumption in order to compute the 'value' and the 'volume' dimensions, taking into account the annual average consumption in sticks equivalents and the average price of vaping products.

4.1.1 The potential adopters

The first stage consisted of estimating the potential adopters for each country, where two approaches were considered depending on the available data for a particular country. The potential market was estimated separately for the two main vaping categories (E-vapor and T-vapor).

On one hand, some countries had available data provided from internal surveys which allowed the estimation of acceptance rate of vaping products among people who have experienced RRP. Additionally, it was also possible to estimate the proportion of individuals who have never tried any vaping product. As we referred in subsection 2.2.3.1, surveys can be an useful source for collecting intention purchase data (Bass, 2004).

For modeling purposes, we limited the scope to estimate the potential number of RRP users to the number of current smokers. First, by deducting the incidence of dualists from the smoking incidence, we obtain the percentage of individuals who consume exclusively traditional tobacco. Thus, by multiplying this rate by the adult population, the number of smokers is computed for a given country. Subsequently, it is important to understand how vaping products are received among smokers, i.e, the acceptance rate of this type of products, denoted by 'acc', which is achieved by obtaining the percentage of individuals who have experienced a vaping product and consume it regularly,

$$acc = \frac{RC}{TR} \tag{4.1}$$

where RC is the number of respondents who experienced RRP and consume it regularly and TR denotes the total number of respondents who tried a vaping category. Furthermore, we assume that an individual who tried a vaping category and was not satisfied, will not try it again in the future, so we restrict the potential adopters to the smokers who have never tried the particular RRP category until 2018,

$$nt = \frac{NTR}{SR} \tag{4.2}$$

where nt denotes the never-tried rate, NTR corresponds to the number of individuals inquired who have never tried a particular RRP category, and SR represents the total number of respondents of a survey to a specific country.

Following this approach, the computation of the number of potential adopters was determined as follows:

$$m = pop \times (smk - dual) \times acc \times nt + cons \tag{4.3}$$

where,

m = number of potential adopters for E-vapor or T-vapor category, respectively

pop = adult population of 2017

smk = smoking incidence, i.e., the percentage of adult population who smoked traditional tobacco, in 2018

$dual$ = incidence of dualists of E-vapor or T-vapor category, i.e., the proportion of the adult population who regularly consumed traditional tobacco and some RRP, in 2018

acc = acceptance rate, i.e., the percentage of people who had experienced some vaping product and consumed it regularly, in 2018

nt = percentage of the people inquired who have never tried any RRP from a given vaping category, in 2018

$cons$ = the number of consumers of E-vapor or T-vapor category, in 2018

The adult population for each country was collected in the public source 'World Bank', where 2017 was the last year with available data at the time of development of this study. The incidence of dualists and the number of consumers were found through the methodology explained for the historical period. In addition, smoking incidence was provided by internal data regarding the non-RRP market and finally the acceptance rate and never-tried rate were provided by internal surveys.

For countries for which the previous approach could not be used because of the inexistence of internal surveys, we resort to statistical methods to estimate the number of potential consumers. This alternative approach consisted of three main stages: data collection on potential explanatory variables (see Chapter 3), understanding the relationship between these variables and the number of potential adopters for countries with internal surveys, and applying the most appropriate model to the remaining countries.

The data collection on several potential explanatory variables mentioned in the section 3.2.2 was made for the 40 countries under analysis from 2015 to 2018. Considering the countries with available internal surveys, the main purpose was to understand the relationship between these potential drivers (Table 3.3) and the difference among the potential consumers for E-vapor or T-vapor category and the current number of adopters of a given country and year, corresponding to $(m - cons)$ for each vaping category, as shown in expression (4.3). In all the countries under analysis, the trading of E-vapor products started until the end of 2018. However, in some of them, T-vapor had not been introduced, so for these particular cases, the output variable corresponds to the potential T-vapor market once the actual number of consumers is zero.

In order to avoid the use of unnecessary variables, their relevance was assessed by statistical methods. A correlation analysis was performed which allowed the understanding of the degree of association among multiple variables, in order to avert the introduction of very closely related variables (multicollinearity). Subsequently, after excluding variables with strong correlation between them, a regression model was fitted in order to assess the relevance of the explanatory variables.

Regarding the training methods, the holdout method and the cross-validation are two of the most common (Kohavi, 1995). The holdout method or test sample estimation is a simpler method, which divides the data into two mutually exclusive subsets. These two data sets with distinct functions throughout the process of choosing the best model correspond to a training data set and a test data set. The first one could include a sample with some of the observations from countries with internal surveys used to train the model, i.e., the observations used to understand the relationship between the chosen variables and the output variable. On the other hand, the test data set might be composed by the remaining observations from countries with internal surveys that would be used to test the previously trained models with the aim to provide an unbiased assessment of each one. However, one of the main disadvantages of this method is that only a fraction of the available information is shown to the algorithm to be trained.

Trying to overcome this limitation, cross-validation, often referred as k-fold cross-validation, is an alternative technique for estimating the test error of a predictive model. The training data set is randomly divided into k mutually exclusive subsets of observations with similar size. Then, k models are produced and each is trained on different k-1 folds and tested on the remaining one. The process is repeated k times and the result corresponds to the averaged one. This method allows the consideration of all data, increasing the confidence over the previous one, being the method used in this study.

A range of methods[1] were considered and tested. The Generalized Linear Model (GLM) was the best model for predicting the number of potential adopters for countries with data from internal survey, i.e., its prediction error[2] outperformed the remaining methods, so they were abandoned. After being defined that a GLM-based approach would be the chosen for this case, using the data set with all the observations from countries with internal surveys, a new model was trained without leaving any observations out. This model was then used to estimate the potential market for the countries that lacked surveys in the year of 2018.

4.1.2 Coefficients of innovation and imitation

After estimating the parameter m for both E-vapor and T-vapor categories and the number of consumers for the historical period, we were able to determine the cumulative percentage of adopters ($F(T)$) for those years for each country,

$$F(T) = \frac{cons(T)}{m} \tag{4.4}$$

where $cons(T)$ represents the cumulative number of consumers of period T for a given vaping category.

Moreover, in order to collect data on the entry time of RRP in a given market, we considered two sources of information. On one hand, an internal report that contained the date of the launch of the products from the four major players in this segment. On the other hand, a manual collection through the Google Trends website (Google Trends, 2019) was conducted. The search term 'vaping' was used as a proxy for the E-vapor category since most of the countries were supplied by small brands not covered by the internal report mentioned before. On the Google Trends website, the countries were selected one by one and the information for all available years was considered (from 2004 to the present). The entry year was defined based on the first year since search on the term 'vaping' began to increase consistently.

[1] Such as Generalized Linear Model (GLM), Gradient Boosting Machine (GBM) and Random Forest (RF).

[2] The Mean Absolute Percentage Error (MAPE) was the accuracy measure used to assess the prediction error.

The estimation of the remaining two parameters of Bass Diffusion Model (Bass, 1969), the coefficient of innovation (p) and the coefficient of imitation (q), was obtained through two different approaches depending on the availability of historical data for each vaping category (E-vapor/T-vapor) for a particular country. For the countries with historical data for a given vaping category, the estimation of the two parameters (p and q) was obtained on the basis of a non-linear optimization approach as presented in Figure 4.1 and described next. After setting the parameters to be optimized (p and q), it was important to identify the constraints of the problem, i.e., what limits our decisions. In this study, we consider only one constraint, the value of each of these parameters must be between 0 and 1. Subsequently, parameters p and q were estimated through the method of least squares once it was not possible the estimation by the Bass estimation method presented in equation (2.9) because it requires a high number of observations which we do not have. The objective was to estimate these parameters by minimizing the deviation between the historical values for the cumulative percentage of potential consumers and those estimated through the Bass curve as a function of parameters p and q (expression (2.3)) for every year, as follows

$$\min \sum_{T} [\hat{F} - F(T)]^2 \text{ subject to } 0 \leqslant q \leqslant 1 \text{ and } 0 \leqslant p \leqslant 1 \tag{4.5}$$

where $\hat{F}$ corresponds to the prediction resulting from expression (2.3) for the year T and $F(T)$ is the actual cumulative percentage of potential adopters for the period T for a given vaping category and country. By optimizing the values of p and q, the sum of the squared deviations among both values was minimized (see the procedure in Figure 4.1).

On the other hand, a second approach was considered for the countries where a specific vaping category had not existed until 2018 or in which only existed a year ago. In order to estimate the coefficients of innovation and imitation for this range of countries, we followed the next steps. First, each country was assigned to a particular cluster defined based on variables, such as smoking incidence, GII and GDP*pc* through the k-means clustering method (Marôco, 2018, Chapter 11). Subsequently, for the countries of a given cluster where there was a certain category of vaping products in 2017, the average values of the coefficients p and q were found. Then, those average values were assigned to the countries of a given cluster where that specific vaping category had not been introduced yet or in which only existed a year ago.

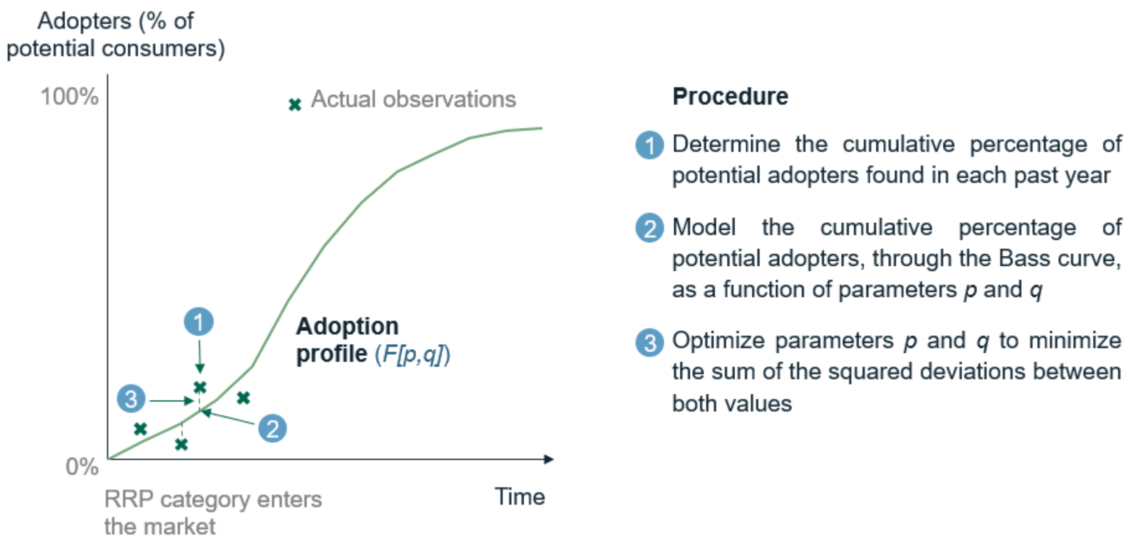

Figure 4.1: Optimization process for the estimation of parameters p and q

4.2 Estimation of dimensions

After estimating the parameters p and q of Bass model (Bass, 1969), we were able to estimate the percentage of cumulative adopters for the forecast years for the two main vaping categories and for each country. Applying these percentages to the number of potential adopters found through the approaches explained in subsection 4.1.1, we determined the first main dimension under analysis, the 'number of consumers' from 2019 to 2025 for each vaping category:

$$cons(T) = F(T) \times m \tag{4.6}$$

where T corresponds to the years between 2019 and 2025.

As previously mentioned, we are dealing with durable (devices) and non-durable (refills) products so the number of consumers did not correspond to the number of purchases. Thus, in order to address the 'volume' dimension, we consider the average consumption in sticks equivalents (metric that allows the comparison of the several products from the both main vaping categories) of each product category, item type and consumer type of 2018 (expression (4.7)). In addition, the average price per stick of each product category and item type of 2018 was used for the estimation of the market 'value' dimension (expression (4.8)).

$$Mvol(T) = cons(T) \times CPA(T) \tag{4.7}$$

and

$$Mval(T) = MVol(T) \times Price(T), \tag{4.8}$$

where $\boldsymbol{Mvol}$ represents the RRP market volume, $\boldsymbol{Price}$ is the RRP average price, $\boldsymbol{cons}$ denotes the number of consumers, $\boldsymbol{Mval}$ represents the RRP market value and $\boldsymbol{CPA}$ denotes consumption of sticks equivalents per adopter in $\boldsymbol{T}$ for each vaping category. The procedure is displayed in Figure 4.2.

Figure 4.2: Procedure linking the main dimensions — Forecast period

In the historical period, the values of the main dimensions were divided into product-related and consumer-related splits as referred in chapter 3. The dimension 'value' was the starting point as explained in section 3.2.1 and its results were divided into seven product categories (Table 3.1), two item types (device and refill) and two consumer types (solus and dualists). Each product category present in a given country represented a fraction of the market value and all the existing categories must add up to 100%. Through the consideration of the average consumption in sticks equivalents per product category existing in a given country and the average price, the results for the 'number of consumers' and 'volume' dimensions were estimated as well as the results for the several splits.

For the forecast period, two approaches were considered. One for the countries where both vaping categories (E-vapor and T-vapor) have already existed in 2018, where we assumed that the results of the splits under analysis, the average consumption in sticks equivalents and the average price will remain constant over time, being equal to the ones of 2018.

In the second approach, for the countries in which the T-vapor category has not emerged until 2018, we assumed that the entry year of the product category 'Tobacco stick' is 2019. For the T-vapor category, we considered only 'Tobacco stick', as it seems the most promising product category among the two existing ones (the product category 'Infused tobacco' exists in very few countries to the current date). In order to estimate the average consumption in sticks equivalents and the average price of this product category for countries with historical data exclusively for the E-vapor category, we considered the clusters referred in section 4.1.

For the countries of a given cluster where 'Tobacco stick' existed in 2018, the average consumption in sticks equivalents and the average price of the 'Tobacco stick' product category for both item types were found. Then, those average values were assigned to the countries of a given cluster where the T-vapor category had not been introduced yet. After this step, the estimation of the results of 'value' and 'volume' dimensions were determined according to the procedure displayed in Figure 4.2.

Chapter 5

Results

This chapter shows the results obtained through the application of the methodology based on diffusion models and complemented by optimization and statistical methods described in chapter 4. Section 5.1 describes the chosen variables that were considered in the model, including the results of the correlation analysis and the assessment of their relevance. Section 5.2 presents results of the estimation of the three main Bass parameters for some countries under analysis. Section 5.3 describes the results for the historical and forecast periods of the main dimensions considered.

5.1 Variable selection

The chosen variables presented in Table 3.3 were analysed with the purpose of being included in statistical models for the estimation of the parameter m of the Bass model.

In an initial assessment, a correlation study was conducted in order to discard the variables with high correlation among them. In order to ensure that the differences in size between the variables selected would not affect the results found, their values were normalized before performing the correlation analysis. This analysis is crucial to guarantee the model integrity, once a high correlation among the variables leads to model redundancy (Guyon & Elisseeff, 2003). A graphical display of the correlation matrix is presented in Figure 5.1, where the ball sizes are proportional to the correlation absolute values, and the correlation coefficients are in Figure B.1 (Appendix B).

After analysing the correlation among the initial variables selected, we decided to discard variables with a correlation higher than 0.7 in absolute value. In Table 5.1, the pairs of variables which verified this condition are presented.

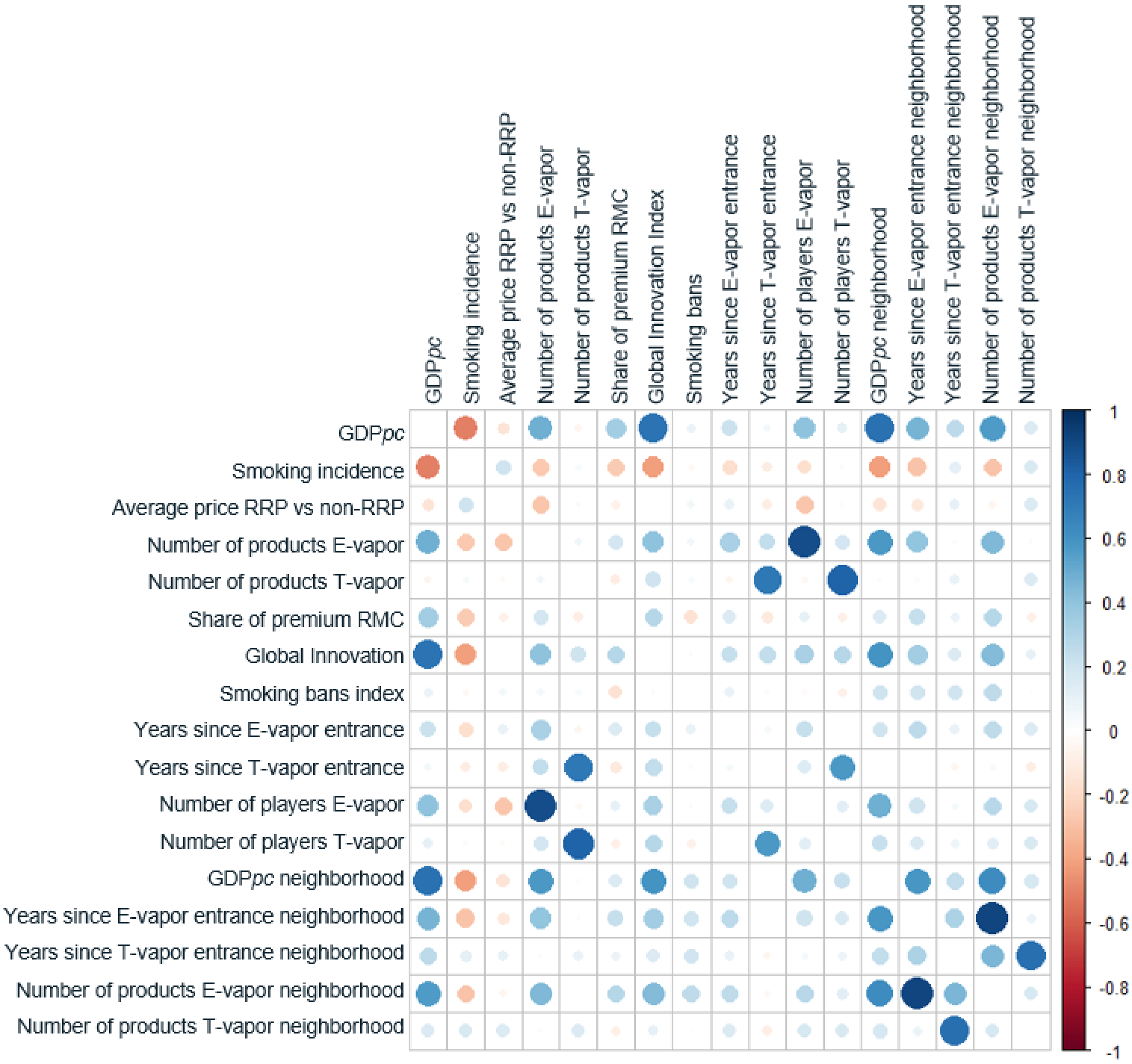

Figure 5.1: Correlation Matrix between the explanatory variables

As a result, for the set of variables from Table 3.3 whose relevance would be assessed, we decided to exclude the following variables: *GDPpc of neighboring countries, Global Innovation Index, Number of players of both vaping categories, Years since T-vapor entrance and Years since both vaping categories entrance of neighboring countries* in order to avoid multicollinearity issues. Thus, the remaining variables (Table 5.2) were initially introduced in the GLM applied to the training data set.

The model dependent variable was the difference between the number of potential consumers and the number of RRP adopters of 2018 in relation to the adult population of 2017 for each country. Therefore, since this variable is a proportion, the "logit" link function was considered which means that this is a logistic regression model. The *p-values* and the estimate of the parameters associated to each variable are presented in Table 5.3 for both vaping categories.

Table 5.1: Pairs of variables with correlation higher than 0.7

Variable A	**Variable B**	**Correlation coefficient**
GDP*pc*	Global Innovation Index	0.73
GDP*pc*	GDP*pc* neighborhood	0.74
Nb players E-vapor	Nb products E-vapor	0.88
Nb players T-vapor	Nb products T-vapor	0.80
Nb products T-vapor	Years since T-vapor entrance	0.72
Nb products E-vapor neighborhood	Years since E-vapor entrance neighborhood	0.91
Nb products T-vapor neighborhood	Years since T-vapor entrance neighborhood	0.76

Table 5.2: Remaining explanatory variables after correlation analysis

GDP *per capita*	Smoking incidence	Share of premium RMC
Smoking bans	Nb of E-vapor products	Nb of T-vapor products
Time since E-vapor entrance neighborhood	Nb of E-vapor products neighborhood	Time since E-vapor entrance
RRP/non-RRP price		

Only the explanatory variables whose estimated parameter was statistically significant (*p-value* lower than 0.05) were considered in the GLM used in the test data set (the *p-value* of the estimate associated with the variable 'smoking bans of traditional tobacco' is 5% and therefore we decided to keep it in the model). Thus, the final model estimation results are presented in Table 5.4 for both vaping categories.

The residual deviance chi-square statistics (lack-of-fit test) are 1.25 and 0.682 for the E-vapor and the T-vapor categories respectively, with p-values > 0.999 in both cases, clearly non-significant. The model chi-square statistics (regression overall significance test) are 270.266 and 90.052 respectively with p-values < 0.0001 in both cases, clearly significant. Furthermore, the pseudo-R^2 are 0.995 and 0.992 respectively, both very high. Thus, the fitted models are adequate for both vaping categories.

Table 5.3: Estimation results of the GLM

Variable	E-vapor		T-vapor	
	Estimate	p-value	Estimate	p-value
Share of premium RMC	-0.577	0.013	2.677	0.206
Average price RRP vs non-RRP	0.142	0.520	1.230	0.175
GDP *pc*	0.000	0.048	0.000	0.367
Smoking incidence	-0.034	0.101	0.196	0.043
Years since E-vapor introduction	0.058	0.673	0.078	0.877
Number of products offered (E-vapor)	0.238	0.161	0.828	0.170
Number of products offered (T-vapor)	-0.477	0.044	0.082	0.918
Smoking bans of traditional tobacco	0.129	0.392	1.251	0.042
Number of products offered in neighboring countries (E-vapor)	0.006	0.969	0.076	0.961
Number of products offered in neighboring countries (T-vapor)	0.978	0.003	2.606	0.033

Table 5.4: Final model estimation results

Variable	E-vapor		T-vapor	
	Estimate	p-value	Estimate	p-value
Share of premium RMC	-0.416	0.042	-	-
GDP *pc*	0.000	0.001	-	-
Smoking incidence	-	-	0.025	0.049
Number of products offered (T-vapor)	-0.530	0.013	-	-
Smoking bans of traditional tobacco	-	-	0.395	0.050
Number of products offered in neighboring countries (T-vapor)	0.894	0.001	1.511	0.007

5.2 Bass parameters

As previously mentioned, the main Bass parameters under analysis were the potential market (m) and the coefficients of innovation and imitation (p and q).

Throughout the analysis of potential market results, we will always express the results in relative terms, i.e., the ratio between the number of potential RRP adopters and the number of adults in 2017 for a given region[1]. We took into account the number of adult individuals of 2017 once it was the most recent year with available data for the adult population in the World Bank, at the time of the study.

[1]Due to confidentiality issues, country-level results could not be displayed.

As can be seen in Table 5.5[2], in general, for the several countries under analysis, the T-vapor category presents a higher potential market than E-vapor one. These results were found by considering the sum of the number of potential consumers and the sum of the adult population in the 40 countries.

Table 5.5: Potential market (% of adult population in 2017)

Scope	E-vapor	T-vapor	Both vaping categories
40 countries	2.61	4.72	7.33

The prediction of greater success of T-vapor products might exist due to the higher similarity between heated tobacco and traditional tobacco than between electronic cigarettes and traditional tobacco. Additionally, the detailed results of potential incidence for vaping products for each region on the scope are presented separately in Table A.2. It is important to note that all results were computed at the country level, nevertheless, in this report, the results are presented aggregated by region due to confidentiality requirements.

In some countries, vaping products are expected to be particularly successful, with a potential vaping incidence higher than 10%, such as Japan, the United Kingdom and South Korea. On the other hand, Sweden and Colombia will be some of the countries where a lower percentage of the adult population smokes RRP, which might be justified by the lower smoking incidence in these countries. Sweden is considered the European country with a smaller rate of smokers (5% in 2017) and in Colombia around 11% of adults smoked in 2017.

According to our results, E-vapor category will have more success, especially, in the United Kingdom, Italy and Poland, where it is expected that more than 4% of the adult population will consume this type of products. On the other hand, it is expected that T-vapor products will be particularly successful in Japan and South Korea, with a potential vaping incidence rate higher than 9%. Currently, in these countries, the T-vapor category is largely spread already (Table A.3), having reached more than 60% of the maximum potential estimated. This can be due to the fact that these countries have a high rate of smokers and, for instance, Japanese consumers are considered to be very receptive to the adoption of new technologies. In the remaining countries, T-vapor products still have a lot of potential.

The large discrepancy in the results of Table A.3 among the two vaping categories could be explained by the later introduction of T-vapor products and the inexistence of them in

[2]The values are presented as a rate due to confidentiality issues of the consulting project.

many of the countries under analysis. Moreover, E-vapor market is almost already saturated in several countries, such as the Sweden, Austria and Denmark, representing more than 70% of the expected number of consumers.

Concerning the two remaining parameters (p and q), the results found for a country in each region are displayed in Table A.4. According to the results reached, the mean value of p for the E-vapor category for the 40 countries was higher than the average value of p for T-vapor one. Furthermore, the opposite relationship is observed for the mean value of the coefficient of imitation for all the countries.

Both internal and external influences seem important for both vaping categories and for the majority of the countries. E-vapor products were the first to be introduced in the market, so they have existed for more years and in more countries than the T-vapor category. Since for this category, we had a greater number of observations available in the historical period, this facilitated the understanding of the possible behavior of the diffusion curves. For several countries, a significant portion of the estimated potential market for E-vapor category has already been achieved in 2018, which shows the great importance of external influence, especially for countries where this type of products showed significant growth in the early years.

5.3 Main dimensions

We will describe the results through the market share of vaping products in the total of tobacco market for the 'value' and 'volume' dimensions. To this purpose, internal data regarding non-RRP sales in value ($\$US$) and volume (sticks equivalents) for the several countries under analysis was considered:

$$Market\,share\,vol(T) = \frac{RRP\,vol(T)}{RRP\,vol(T) + non\text{-}RRP\,vol(T)} \tag{5.1}$$

and

$$Market\,share\,val(T) = \frac{RRP\,val(T)}{RRP\,val(T) + non\text{-}RRP\,val(T)} \tag{5.2}$$

where $RRP\,vol$ and $non\text{-}RRP\,vol$ represent the quantity of sticks equivalents for RRP and non-RRP, respectively; and $RRP\,val$ and $non\text{-}RRP\,val$ represent the RRP and non-RRP market in $\$US$, respectively for both vaping categories in 2018.

Furthermore, the incidence found through expression (5.3) was the metric used to express the results for the 'number of consumers' dimension:

$$Inc(T) = \frac{cons(T)}{pop(2017)} \tag{5.3}$$

where Inc represents the incidence for period T, $cons$ is the number of consumers in T and pop expresses the number of adults in 2017 of a given country for both dimensions.

This section is divided into two subsections. In subsection 5.3.1, the results found for the historical period considered (from 2015 to 2018) will be expressed for the three dimensions under analysis. Subsequently, in subsection 5.3.2, the forecast results for the period from 2019 to 2025 will be presented for the 40 countries of the study. The definition of the forecast period took into consideration a period of time that seems reasonable for the understanding of future market behaviour and provides a view that allows companies in the tobacco industry to make investment decisions and planning in the short term. Moreover, the countries of the scope (Table A.1) are countries where at least one of the major vaping categories currently exists and in which the company that requested this study had strategic interest.

5.3.1 Historical period

In the period between 2015 and 2018, the E-vapor category attracted more adopters, with more than twice the number of T-vapor consumers (Table 5.6). The growth was slowing down in both categories, but in 2018 the rate was lower in E-vapor one, which might be justified by the difference in entry years. The E-vapor products began to appear around 2009 in the United Kingdom and in the USA. On the other hand, T-vapor category was only introduced in the market in 2014 in Italy and Japan.

Until 2017 T-vapor products existed in very few countries and were not relevant in many regions (A.5). However, two main Asian markets should be highlighted in this vaping category, Japan and South Korea, where heated tobacco presents a significant growth and is specially successful. For the remaining countries, almost all of them present very low incidence rates.

For the E-vapor category, Italy, the USA and Denmark are among the countries with the highest percentage of adults who consume vaping products. In addition, E-vapor is not so popular among the individuals of several countries, such as South Korea, Colombia and Japan.

Several discrepancies in the consumption of sticks equivalents for each product category and type of consumer provided by the available internal report were taken into consideration. The consumption per adopter in sticks equivalents for a given product category was considered equal for all countries, once the internal report to which we had access only provided information on a very limited number of countries. Therefore, the average of the consumption per adopter for all countries with available data was assumed for the remaining ones.

Table 5.6: Global incidence (% of adult population of 2017) — Historical Period

	Year			
Vaping category	**2015**	**2016**	**2017**	**2018**
E-vapor	0.90	1.14	1.39	1.64
T-vapor	0.01	0.15	0.56	1.06
Both	0.91	1.29	1.95	2.70

This has generally led to the fact that the metric considered for the 'volume' dimension, i.e., the share of market of RRP measured in sticks equivalents, presents a similar behaviour to the 'number of consumers' dimension (Table 5.7), being higher in the E-vapor category. However, since the consumption measured in sticks equivalents for the E-vapor products was greater than for the T-vapor one, the differences between the two vaping categories were more significant. The detailed data on all the regions under analysis is presented in Table A.6.

Table 5.7: Market share in % (volume) — Historical Period

	Year			
Vaping category	**2015**	**2016**	**2017**	**2018**
E-vapor	3.10	4.22	5.04	6.13
T-vapor	0.02	0.34	1.31	2.54
Both	3.12	4.56	6.34	8.68

Regarding the 'value' dimension, the scenario we faced was quite different from the previous ones (Table 5.8). Globally, in 2015, the discrepancy among E-vapor and T-vapor results was very significant, since T-vapor category represented only 0.03% of the total tobacco market while the E-vapor one 1.58% due to the reduced presence of heated tobacco products.

However, the difference in market shares between the two vaping categories was declining, especially in the last two years of the historical period, which might be explained by a very strong growth of the T-vapor category considerably above the growth of the E-vapor category and price differences across vaping categories, since one stick equivalent of E-vapor products tend to be cheaper.

Table 5.8: Market share in % (value) — Historical Period

	Year			
Vaping category	**2015**	**2016**	**2017**	**2018**
E-vapor	1.58	2.06	2.44	2.91
T-vapor	0.03	0.35	1.36	2.72
Both	1.60	2.41	3.80	5.63

This led to the existence of a lower discrepancy between E-vapor and T-vapor results when compared with the results of the remaining dimensions. Additionally, the results found for both vaping categories were very close in 2018. The detailed data on all the regions under analysis is presented in Table A.7.

5.3.2 Forecast period

From 2018 to 2025, we predict that the E-vapor category will attract an additional 0.8 p.p. of the adult population of the 40 countries under analysis, while T-vapor products will be consumed by more 1.7 p.p. of the adults (Table 5.9). A higher growth of the T-vapor category is expected over the forecast period as compared to the alternative vaping category, which may be justified by the later introduction of heated tobacco products on the marketplace and the assumption that T-vapor will be introduced in 2019 in all the countries where it did not exist until then.

In 2025, it is expected that the E-vapor market will be very close to the saturation point, while T-vapor category may grow a bit more. Globally, for the group of countries under analysis, around 5% of its adult population will adopt any vaping product in this year. Although some countries are very close to the saturation level, overall, the incidence of the vaping adopters is expected to grow steadily by 2025. The detailed results for each region are presented in Tables A.8 and A.9, for E-vapor and T-vapor categories, respectively.

It is expected that in the United Kingdom, Italy and Ireland more than 4% of the adult population will consume electronic cigarettes, i.e., E-vapor products, being the countries that should be highlighted through this metric in 2025. On the other hand, in Japan, Colombia and Portugal, a low incidence is expected, so the adoption of this type of products by the adult population will be slight.

Heated tobacco products will be more popular among adults in Japan and South Korea, as in the current situation. Moreover, Canada, the Netherlands and Poland are among the countries with the highest growth from 2018 to 2025 in this vaping category. On the other hand, in Germany, the United Kingdom and the USA, consumers will not reach 0.1% of the adult population. It is notorious that in countries where a vaping category has already a strong position, the other category will have a difficult path[3].

Table 5.9: Global incidence (% of adult population of 2017) — Forecast Period

	Year						
Vaping category	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**2025**
E-vapor	1.91	2.08	2.22	2.31	2.38	2.43	2.47
T-vapor	1.44	1.73	1.97	2.19	2.39	2.57	2.72
Both	3.35	3.82	4.18	4.50	4.77	5.00	5.19

For the estimation of the market share of vaping products in relation to non-RRP market in sticks equivalents and in $*US*, we considered product type divisions (i.e. considering the product categories mentioned in Chapter 3) and consumer type splits similar to those of 2018. In the forecast period, also the average consumption per adopter in sticks equivalents for each product category and type of consumer was assumed equal for all countries, since we only had access to this type of information for a limited number of countries.

This methodology showed that, although both vaping categories show continuous growth until 2025, it is expected a more pronounced growth in the T-vapor one. According to our results, the vaping segment will have a significant weight in the total tobacco market volume, in 2025, representing about 16% of market share (Table 5.10). Despite of the significant growth observed in T-vapor category, electronic cigarettes category measured in sticks equivalents continue to represent a higher market share than the alternative vaping category due to the higher consumption previously mentioned. Tables A.10 and A.11 present the detailed expected results for each region, for E-vapor and T-vapor category, respectively.

[3]Due to confidentiality issues, the exact results for each country could not be displayed.

Table 5.10: Market share in % (volume) — Forecast Period

Vaping category	**2019**	**2020**	**2021**	**Year** **2022**	**2023**	**2024**	**2025**
E-vapor	7.35	7.92	8.36	8.70	9.03	9.27	9.45
T-vapor	3.45	4.15	4.75	5.30	5.78	6.22	6.58
Both	10.80	12.07	13.10	14.00	14.81	15.49	16.04

Regarding the market share in value, computed considering the sales of RRP and non-RRP em $*US*, the scenario faced was quite different. Since 2019, the market share of T-vapor category is expected to exceed the E-vapor one. This may be justified because a higher growth of the sales of heated tobacco products is expected and it has been assumed that in 2019 there will be T-vapor products in all the 40 countries. Additionally, the differences between the prices of vaping products detected in 2018 will remain constant over the years under study, with T-vapor prices being higher than those for E-vapor, leading to an increase in the discrepancy in market shares of value dimension, as shown in Table 5.11. Overall, the vaping segment is expected to represent around 12.31% of the total tobacco sales in $*US*, in 2025. Furthermore, the E-vapor and T-vapor categories are expected to show continuous growth until 2025. The detailed results for each region are presented in Tables A.12 and A.13, for E-vapor and T-vapor category, respectively.

Table 5.11: Market share in % (value) — Forecast Period

Vaping category	**2019**	**2020**	**2021**	**Year** **2022**	**2023**	**2024**	**2025**
E-vapor	3.26	3.52	3.74	3.93	4.14	4.28	4.40
T-vapor	3.74	4.57	5.33	6.07	6.74	7.37	7.91
Both	7.00	8.09	9.08	10.01	10.88	11.66	12.31

Once the values of market shares are computed based on the size of traditional tobacco market, it is important to understand in which countries is expected a higher growth in the three dimensions under analysis. Due to confidentiality issues, the values of each dimension cannot be displayed in this internship report. However, in Figure 5.2, the bars height measures the expected growth in value ($*US*) for the RRP category from 2018 to 2025, i.e., corresponds to the difference between the forecast sales in 2025 and the sales for 2018 for the

40 countries. The value displayed in this Figure corresponds to the sum of the value growth of E-vapor and T-vapor categories.

The countries with an associated blue bar are countries where the value growth ($*US*) of T-vapor category is expected to be higher than the E-vapor one. On the other hand, when the growth of value of E-vapor category is expected to be greater than the T-vapor one, the countries have a green bar. According to our results, T-vapor category is expected to grow more strongly in about 3/4 of the countries, which can be explained by the later introduction of T-vapor products in the marketplace and by the assumption that in 2019 in the 40 countries of the scope will exist heated tobacco products.

Furthermore, Japan, Italy and Canada are among the countries where higher growth in sales value ($*US*) is expected. In these countries, the increase in T-vapor category sales from 2018 to 2025 is expected to be more significant than E-vapor one. The USA is the country with the highest growth in sales value in the set of countries where growth of E-vapor category is dominant.

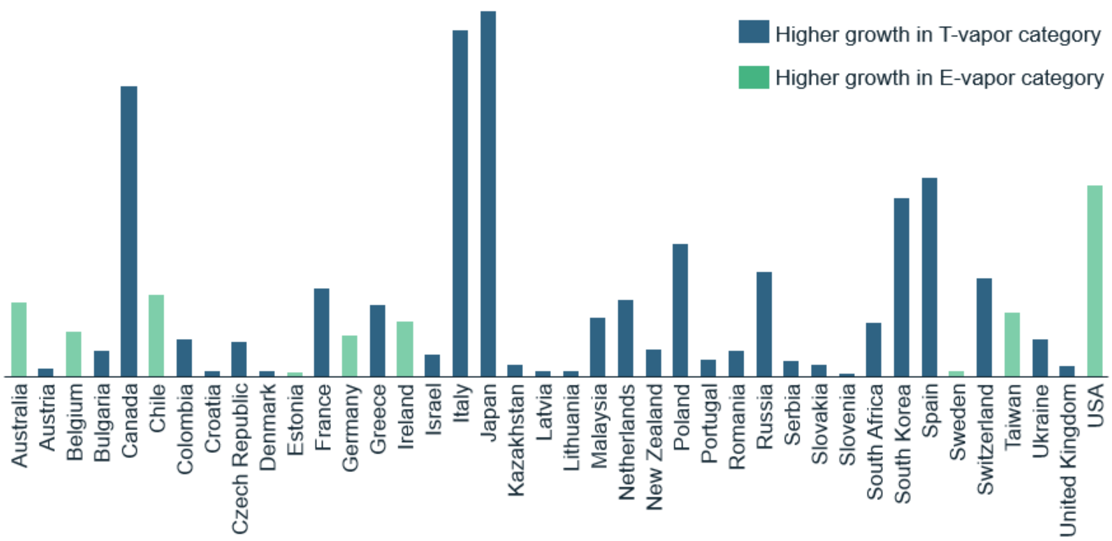

Figure 5.2: Growth of the RRP category sales from 2018 to 2025 in $*US*

On the other hand, Estonia, Slovenia and Lithuania are among the countries where we expect a low sales growth in $*US*.

Finally, the model results found with this consulting project were made available through an user-friendly online tool (Appendix C). The most important purposes of this online tool, which will allow users to interact with the model results, are presented in Table C.1.

Chapter 6

Conclusion

Companies that launch innovative products in the marketplace face the challenge of predicting how market size will evolve. The developed work was based on a consulting project for the prediction of vaping products market size for 40 countries. The main goal was to provide its forecasts in three main dimensions (sales in value, sales in volume and number of consumers) until 2025 which could be useful for the business decisions of companies of the tobacco industry.

Despite of the existing literature on forecast for new products, there is no evidence of similar studies for the type of products under analysis. Thus, this dissertation aimed to contribute to the scientific research, focusing on the forecast work for alternative tobacco products, using diffusion models and statistical methods.

This work began with the reconstruction of the historical period market sizes considering the available data sources we had access. After that, data from several potential explanatory variables was collected with the aim of using statistical methods. Subsequently, a methodology based on the Bass Diffusion Model, which is considered as the most widely diffusion model currently used, and complemented by statistical methods and optimization was applied in order to anticipate the future dissemination of vaping products worldwide.

For the initial range of variables selected, after running a correlation analysis, the range of variables considered in the GLM were: *GDPpc, smoking incidence, price RRP, price non-RRP, number of vaping products of both categories, share of premium RMC, smoking bans, time since both vaping categories entrance, number of players of both vaping categories, and time since T-vapor entrance and number of products from neighboring countries*. Moreover, after running the model, the variables that were relevant for the E-vapor category were *share of premium RMC, GDPpc, num-*

ber of T-vapor products offered in a given country and in neighboring countries, and, on the other hand, for the T-vapor category, we remain with *smoking incidence, smoking bans of traditional tobacco and number of T-vapor products offered in neighboring countries*.

Concerning the number of adopters in 2025, there is no significant difference between E-vapor and T-vapor categories. However, on the one hand, in the 'volume' dimension, tobacco companies are expected to sell a higher quantity of E-vapor products than T-vapor ones. On the other hand, as T-vapor products average price per stick equivalent is expected to remain higher than E-vapor one, we faced a different scenario in the 'value' dimension, with T-vapor products representing a larger share of the tobacco market.

The results will enable the tobacco company that requested this study to understand in which countries a higher adoption of RRP and a greater profitability are expected in order to guide investment decisions and the launch of a given vaping product. Although we will only have the actual values of the vaping market in the coming years and only then we can assess the quality of these forecasts, diffusion models were used in the past to forecast sales for other innovative products and, according to the literature, the results were satisfactory.

Nonetheless, the use of diffusion models to predict sales of non-durable products with limited data, such as the dissemination of vaping products, is still a challenge and up for enhancement. Moreover, when we are dealing with human behavior, there is an inherent unpredictability and uncertainty associated.

For future research, we suggest the replication of the present methodology in a few years, when more data will be available. Furthermore, with the existence of more historical data, the estimation of the parameters p and q can be done through the methodology suggested by Bass (1969) (Section 2.2.2). This will also allow the comparison between these estimates with those obtained by the optimization approach considered in this study. Additionally, the scope can also be extended to more countries and finer granularity as well as scenario analysis to understand the impact of the variation of exogenous and endogenous factors.

References

Abu, N., & Ismail, Z. (2013). New car demand modeling and forecasting using Bass diffusion model. *American Journal of Applied Sciences*, *10*(6), 536–541. doi: 10.3844/ajassp.2013.536.541

Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science*, *15*(5), 215–227. doi: 10.1287/mnsc.15.5.215

Bass, F. M. (2004). Comments on "A new product growth for model consumer durables the Bass model". *Management Science*, *50*(12_supplement), 1833–1840. doi: 10.1287/mnsc.1040.0300

Bass, F. M., Krishnan, T. V., & Jain, D. C. (1994). Why the Bass model fits without decision variables. *Marketing Science*, *13*(3), 203–223.

Brown, L. A. (1981). *Innovation diffusion: A new perspective*. London: Methuen.

Chatterjee, R. A., & Eliashberg, J. (1990). The innovation diffusion process in a heterogeneous population: A micromodeling approach. *Management Science*, *36*(9), 1057–1079. doi: 10.1287/mnsc.36.9.1057

Cornell University, INSEAD, & WIPO. (2018). *The Global Innovation Index 2018: Energizing the world with innovation* (11th ed.). Ithaca, Fontainebleau, and Geneva.

Dekimpe, M. G., Parker, P. M., & Sarvary, M. (2000). Global diffusion of technological innovations: A coupled-hazard approach. *Journal of Marketing Research*, *37*(1), 47–59. doi: 10.1509/jmkr.37.1.47.18722

Desiraju, R., Nair, H., & Chintagunta, P. (2004). Diffusion of new pharmaceutical drugs in developing and developed nations. *International Journal of Research in Marketing*, *21*(4), 341–357. doi: 10.1016/j.ijresmar.2004.05.001

Dodson, J. A., & Muller, E. (1978). Models of new product diffusion through advertising and word-of-mouth. *Management Science*, *24*(15), 1568–1578. doi: 10.1287/mnsc.24.15.1568

Dolan, R. J., & Jeuland, A. P. (1981). Experience curves and dynamic demand models: Implications for optimal pricing strategies. *Journal of Marketing*, *45*(1), 52–62. doi: 10.1177/002224298104500106

Fourt, L. A., & Woodlock, J. W. (1960). Early prediction of market success for new grocery products. *Journal of Marketing*, *25*(2), 31–38. doi: 10.1177/002224296002500206

Ganesh, J. (1998). Converging trends within the European Union: Insights from an analysis of diffusion patterns. *Journal of International Marketing*, *6*(4), 32–48. doi: 10.1177/1069031x9800600403

Ganesh, J., Kumar, V., & Subramaniam, V. (1997). Learning effect in multinational diffusion of consumer durables: An exploratory investigation. *Journal of the Academy of Marketing Science*, *25*(3), 214–228. doi: 10.1177/0092070397253003

Golden, J., Milewicz, J., & Herbig, P. (1994). Forecasting: Trials and tribulations. *Management Decision*, *32*(1), 33–36. doi: 10.1108/00251749410050642

Google Trends. (2019). *Google trends.* Retrieved 2019-04-01, from `https://trends.google.pt/trends/explore?date=all&geo=PT&q=vaping`

Gore, A., & Lavaraj, U. (1987). Innovation diffusion in a heterogeneous population. *Technological Forecasting and Social Change*, *32*(2), 163–167. doi: 10.1016/0040-1625(87)90037-0

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(7), 1157–1182.

Helsen, K., Jedidi, K., & DeSarbo, W. S. (1993). A new approach to country segmentation utilizing multinational diffusion patterns. *The Journal of Marketing*, *57*(4), 60–71. doi: 10.2307/1252219

Horsky, D. (1990). A diffusion model incorporating product benefits, price, income and information. *Marketing Science*, *9*(4), 342–365. doi: 10.1287/mksc.9.4.342

Horsky, D., & Simon, L. S. (1978, May). Advertising in a model of new product diffusion. Presented at the TIMS/ORSA National Meeting in New York City.

Islam, T., & Meade, N. (1996). Forecasting the development of the market for business telephones in the UK. *Journal of the Operational Research Society*, *47*(7), 906–918. doi: 10.1057/jors.1996.116

Japan Tobacco International – A global tobacco company. (2019). *Reduced-risk products – our vaping products.* Retrieved 2019-02-15, from `https://www.jti.com/about-us/what-we-do/our-reduced-risk-products`

Jones, J. M., & Ritz, C. J. (1991). Incorporating distribution into new product diffusion models. *International Journal of Research in Marketing*, *8*(2), 91–112. doi: 10.1016/0167-8116(91)90018-3

Kalish, S. (1985). A new product adoption model with price, advertising, and uncertainty. *Management Science*, *31*(12), 1569–1585. doi: 10.1287/mnsc.31.12.1569

Kalish, S., & Lilien, G. L. (1983). Optimal price subsidy policy for accelerating the diffusion of innovation. *Marketing Science*, *2*(4), 407–420. doi: 10.1287/mksc.2.4.407

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai* (Vol. 14, pp. 1137–1145).

Mahajan, V., & Muller, E. (1979). Innovation diffusion and new product growth models in marketing. *Journal of Marketing*, *43*(4), 55–68.

Mahajan, V., & Muller, E. (1994). Innovation diffusion in a borderless global market: Will the 1992 unification of the European Community accelerate diffusion of new ideas, products, and technologies?. *Technological Forecasting and Social Change*, *45*(3), 221–235. doi: 10.1016/0040-1625(94)90047-7

Mahajan, V., Muller, E., & Bass, F. M. (1990). New product diffusion models in marketing: A review and directions for research. *Journal of Marketing*, *54*(1), 1–26. doi: 10.1177/002224299005400101

Mahajan, V., & Peterson, R. A. (1985). *Models for Innovation Diffusion.* California: Sage.

Mansfield, E. (1961). Technical change and the rate of imitation. *Econometrica*, *29*(4), 741–766. doi: 10.2307/1911817

Marôco, J. (2018). *Análise estatística com o SPSS Statistics* (7th ed.). Pêro Pinheiro: Report-Number.

Meade, N., & Islam, T. (2006). Modelling and forecasting the diffusion of innovation – A 25-year review. *International Journal of Forecasting*, *22*(3), 519–545. doi: 10.1016/j.ijforecast.2006.01.005

Mentzer, J., & Moon, M. (2005). *Sales forecasting management: A demand management approach*

(2nd ed.). Thousand Oaks, CA: SAGE Publications.
Norton, J. A., & Bass, F. M. (1987). A diffusion theory model of adoption and substitution for successive generations of high-technology products. *Management Science*, *33*(9), 1069–1086. doi: 10.1287/mnsc.33.9.1069
OECD, & Eurostat. (2018). *Oslo Manual 2018: Guidelines for collecting, reporting and using data on innovation* (4th ed.). Paris, France: OECD Publishing. Retrieved from `https://www.oecd-ilibrary.org/content/publication/9789264304604-en`
Peres, R., Muller, E., & Mahajan, V. (2010). Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing*, *27*(2), 91–106. doi: 10.1016/j.ijresmar.2009.12.012
Peterson, R. A., & Mahajan, V. (1978). Multi-product growth models. *Research in Marketing*, *1*(20), 1–23.
Putsis Jr, W. P., Balasubramanian, S., Kaplan, E. H., & Sen, S. K. (1997). Mixing behavior in cross-country diffusion. *Marketing Science*, *16*(4), 354–369. doi: 10.1287/mksc.16.4.354
Radas, S. (2005). Diffusion models in marketing: How to incorporate the effect of external influence?. *Economic Trends and Economic Policy*, *15*(105), 30–51.
Robinson, B., & Lakhani, C. (1975). Dynamic price models for new-product planning. *Management Science*, *21*(10), 1113–1122. doi: 10.1287/mnsc.21.10.1113
Rogers, E. M. (1962). *Diffusion of innovations*. New York: The Free Press.
Schmittlein, D., & Mahajan, V. (1982). Maximum likelihood estimation for an innovation diffusion model of new product acceptance. *Marketing Science*, *1*(1), 57–78. doi: 10.1287/mksc.1.1.57
Sharif, M. N., & Ramanathan, K. (1981). Binomial innovation diffusion models with dynamic potential adopter population. *Technological Forecasting and Social Change*, *20*(1), 63–87. doi: 10.1016/0040-1625(81)90041-X
Shen, Y. (2015). *Forecasting Twitter topic popularity using Bass Diffusion Model and machine learning* (Unpublished master's thesis). Massachusetts Institute of Technology, Cambridge.
Simon, H., & Sebastian, K.-H. (1987). Diffusion and advertising: The German telephone campaign. *Management Science*, *33*(4), 451–466. doi: 10.1287/mnsc.33.4.451
Smith, D. (2010). *Exploring innovation* (2nd ed.). London: McGraw-Hill Higher Education.
Takada, H., & Jain, D. (1991). Cross-national analysis of diffusion of consumer durable goods in Pacific Rim countries. *The Journal of Marketing*, *55*(2), 48–54. doi: 10.2307/1252237
Talukdar, D., Sudhir, K., & Ainslie, A. (2002). Investigating new product diffusion across products and countries. *Marketing Science*, *21*(1), 97–114. doi: 10.1287/mksc.21.1.97.161
Tansurat, P., & Gerdsri, N. (2015). Forecasting of diffusion pattern: A case example of OLED technology. *IEEE International Conference on Industrial Engineering and Engineering Management*, 275-279. doi: 10.1109/IEEM.2014.7058643
Tellis, G. J., Stremersch, S., & Yin, E. (2003). The international takeoff of new products: The role of economics, culture, and country innovativeness. *Marketing Science*, *22*(2), 188–208. doi: 10.1287/mksc.22.2.188.16041
Thompson, G. L., & Teng, J. T. (1984). Optimal pricing and advertising policies for new product oligopoly models. *Marketing Science*, *3*(2), 148–168. doi: 10.1287/mksc.3.2.148
Van Everdingen, Y., Fok, D., & Stremersch, S. (2009). Modeling global spillover of new product takeoff. *Journal of Marketing Research*, *46*(5), 637–652. doi: 10.1509/jmkr.46.5.637

Appendix A

Results Tables

Table A.1: Countries in the study

Region				
Western Europe	**Eastern Europe**	**Asia Pacific**	**Americas**	**MENEA[a]**
Austria	Bulgaria	Australia	Canada	Israel
Belgium	Croatia	Japan	Chile	South Africa
Denmark	Czech Republic	Malaysia	Colombia	
Estonia	Kazakhstan	New Zealand	USA	
France	Poland	South Korea		
Germany	Romania	Taiwan		
Greece	Russia			
Ireland	Serbia			
Italy	Slovakia			
Latvia	Slovenia			
Lithuania	Ukraine			
Netherlands				
Portugal				
Spain				
Sweden				
Switzerland				
United Kingdom				

[a] Middle East, Near East, Africa

Table A.2: Potential market (% of adult population of 2017)

Region	E-vapor	T-vapor	Both
Americas	3.16	4.24	7.40
Asia Pacific	0.95	8.73	9.68
Eastern Europe	2.40	2.55	4.95
MENEA	2.22	1.67	3.88
Western Europe	3.31	4.61	7.93

Table A.3: Number of RRP consumers in 2018 (% of estimated market potential)

Region	E-vapor	T-vapor	Both
Americas	70.97	0.52	30.62
Asia Pacific	37.23	59.64	57.44
Eastern Europe	50.95	10.93	30.37
MENEA	56.31	6.69	35.03
Western Europe	67.11	2.82	29.69

Table A.4: Estimated coefficients of innovation and imitation

		E-vapor		T-vapor	
Region	**Country**	**p**	**q**	**p**	**q**
Americas	USA	0.032	0.440	0.002	0.101
Asia Pacific	Japan	0.017	0.445	0.012	0.990
Eastern Europe	Poland	0.076	0.291	0.003	0.700
MENEA	Israel	0.025	0.620	0.004	0.660
Western Europe	Denmark	0.022	0.620	0.021	0.420

Table A.5: Incidence (% of adult population of 2017) — Historical Period

	E-vapor				T-vapor			
Region	**2015**	**2016**	**2017**	**2018**	**2015**	**2016**	**2017**	**2018**
Americas	1.20	1.58	1.82	2.24	0	0	0.01	0.02
Asia Pacific	0.45	0.26	0.33	0.35	0.03	0.75	2.83	5.21
Eastern Europe	0.83	0.98	1.07	1.22	0.01	0.03	0.10	0.28
MENEA	0.25	0.42	0.82	1.25	0	0.03	0.11	0.11
Western Europe	1.03	1.46	1.93	2.22	0.01	0.03	0.06	0.13

Table A.6: Market share in % (volume dimension) — Historical Period

Region	E-vapor 2015	2016	2017	2018	T-vapor 2015	2016	2017	2018
Americas	5.42	7.22	7.64	8.59	0	0	0.03	0.09
Asia Pacific	0.85	0.91	1.10	1.49	0.07	1.62	6.24	11.49
Eastern Europe	1.84	2.20	2.44	3.40	0.02	0.08	0.24	0.60
MENEA	2.60	4.51	8.65	17.01	0	0.17	0.92	1.02
Western Europe	4.14	6.08	7.94	9.38	0.01	0.06	0.16	0.34

Table A.7: Market share in % (value dimension) — Historical Period

Region	E-vapor 2015	2016	2017	2018	T-vapor 2015	2016	2017	2018
Americas	2.47	3.17	3.61	4.42	0	0	0.04	0.11
Asia Pacific	0.27	0.33	0.41	0.48	0.06	1.59	6.28	11.93
Eastern Europe	1.32	1.57	1.70	1.94	0.02	0.12	0.34	0.84
MENEA	0.79	1.28	2.47	3.69	0	0.19	0.76	0.78
Western Europe	1.61	2.22	2.85	3.38	0.03	0.11	0.27	0.54

Table A.8: Incidence of E-vapor category (% of adult population of 2017) — Forecast Period

	Year						
Region	2019	2020	2021	2022	2023	2024	2025
Americas	2.52	2.72	2.87	2.98	3.07	3.15	3.21
Asia Pacific	0.49	0.63	0.85	1.10	1.47	1.70	1.87
Eastern Europe	1.51	1.67	1.82	1.93	2.03	2.10	2.15
MENEA	1.68	1.93	2.08	2.15	2.18	2.20	2.21
Western Europe	2.54	2.74	2.90	3.04	3.18	3.30	3.40

Table A.9: Incidence of T-vapor category (% of adult population of 2017) — Forecast Period

	Year						
Region	2019	2020	2021	2022	2023	2024	2025
Americas	0.05	0.09	0.14	0.24	0.36	0.52	0.66
Asia Pacific	6.78	7.64	8.04	8.22	8.26	8.31	8.36
Eastern Europe	0.47	0.68	0.93	1.21	1.51	1.81	2.05
MENEA	0.22	0.36	0.54	0.76	0.98	1.19	1.35
Western Europe	0.29	0.53	0.84	1.16	1.42	1.61	1.75

Table A.10: Market share of E-vapor category in % (volume dimension) — Forecast Period

	Year						
Region	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**2025**
Americas	10.09	10.91	11.48	11.88	12.17	12.40	12.59
Asia Pacific	1.98	2.32	2.83	3.40	4.23	4.79	5.21
Eastern Europe	4.81	5.30	5.69	5.99	6.22	6.37	6.48
MENEA	21.37	23.67	24.69	24.96	24.90	24.71	24.52
Western Europe	10.56	11.10	11.39	11.57	11.72	11.85	11.97

Table A.11: Market share of T-vapor category in % (volume dimension) — Forecast Period

	Year						
Region	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**2025**
Americas	0.09	0.19	0.37	0.67	1.12	1.67	2.18
Asia Pacific	14.52	16.11	16.79	17.05	16.98	17.00	17.06
Eastern Europe	1.02	1.49	2.02	2.61	3.24	3.84	4.34
MENEA	1.62	2.53	3.71	5.09	6.49	7.73	8.69
Western Europe	0.78	1.41	2.22	3.04	3.69	4.15	4.45

Table A.12: Market share of E-vapor category in % (value dimension) — Forecast Period

	Year						
Region	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**2025**
Americas	4.92	5.28	5.54	5.72	5.85	5.95	6.02
Asia Pacific	0.61	0.79	1.06	1.37	1.82	2.12	2.35
Eastern Europe	2.34	2.58	2.77	2.92	3.05	3.13	3.18
MENEA	4.77	5.42	5.73	5.83	5.83	5.79	5.74
Western Europe	3.70	3.95	4.13	4.28	4.43	4.56	4.68

Table A.13: Market share of T-vapor category in % (value dimension) — Forecast Period

	Year						
Region	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**2025**
Americas	0.23	0.40	0.67	1.10	1.72	2.47	3.17
Asia Pacific	15.01	16.64	17.39	17.75	17.80	17.93	18.06
Eastern Europe	1.42	2.03	2.72	3.51	4.37	5.23	5.96
MENEA	1.53	2.48	3.77	5.31	6.89	8.29	9.37
Western Europe	1.18	2.12	3.31	4.51	5.49	6.22	6.75

Appendix B

Correlation Matrix

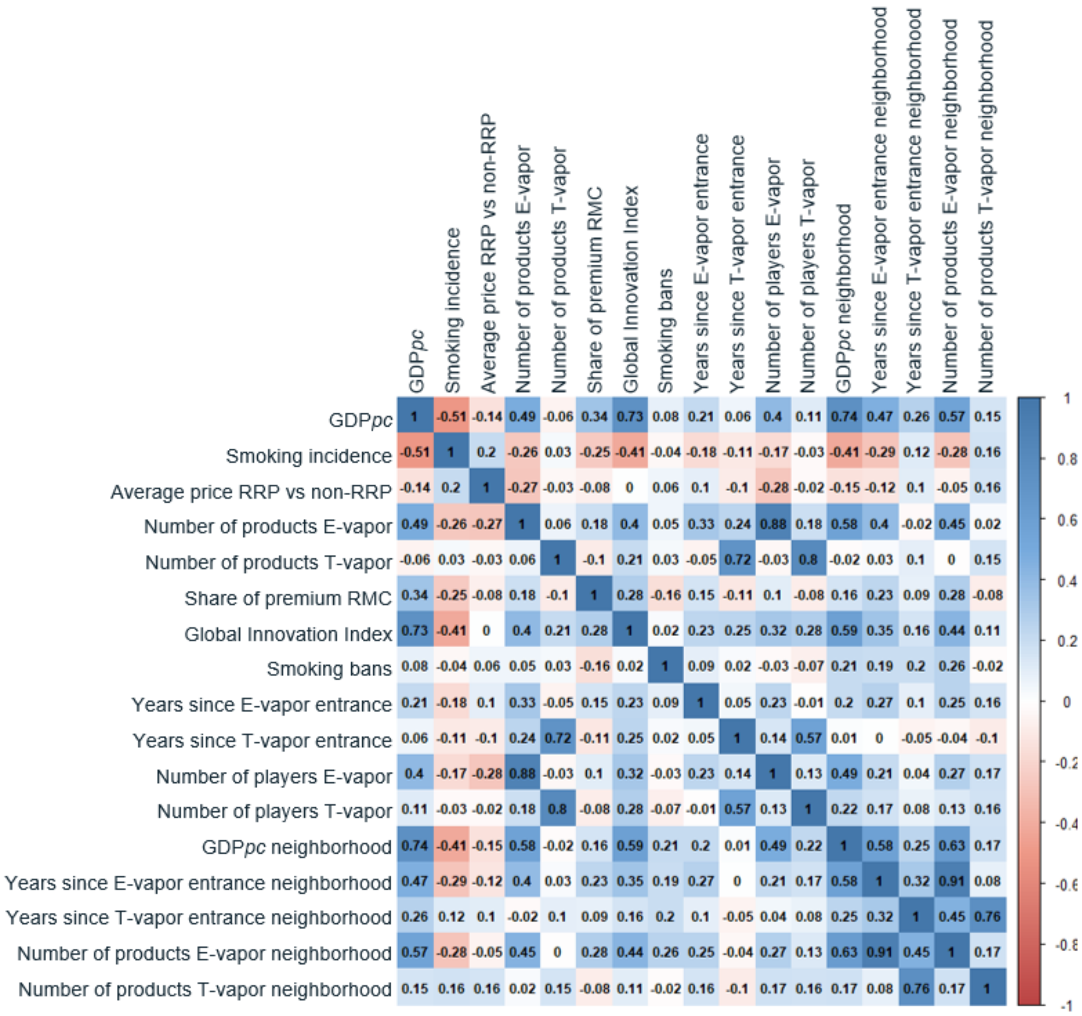

Figure B.1: Correlation Matrix between the explanatory variables with the correlation coefficients

Appendix C

Online tool

The objective of this consulting project was to build a market model to estimate historical and future market sizes for the "Reduced-Risk Products" segment from 2015 to 2025. The market model results were made available through an user-friendly online tool (Figure C.1[1]).

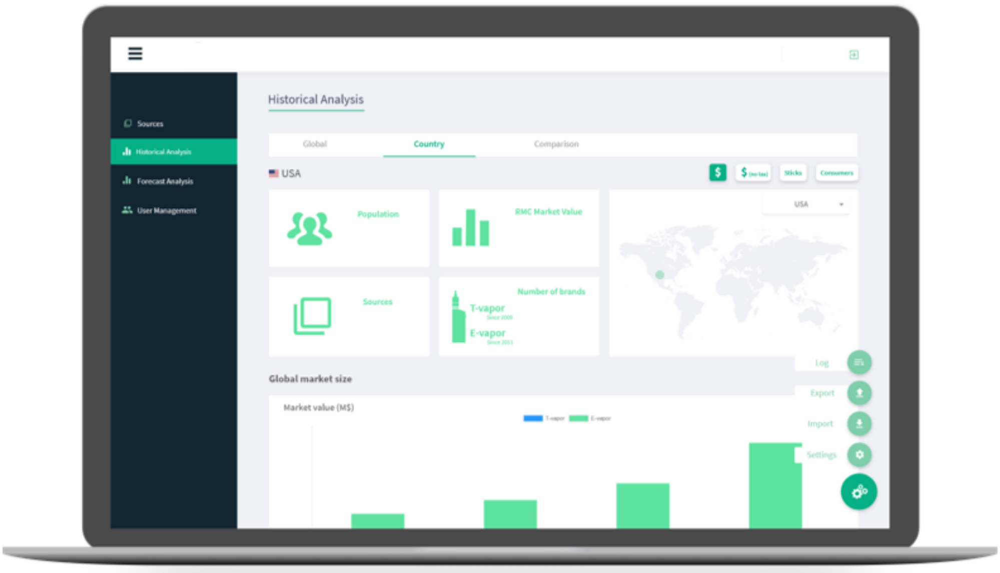

Figure C.1: Online tool

The description of each section that composed the online app is presented in Table C.2.

[1]The numbers available in the online tool were omitted because of confidentiality issues.

Table C.1: Main purposes of the online tool

Objectives
Ongoing update of the models' inputs by uploading the most recent information
Analyse available data sources by country and their reliability level
Permanent online access to latest results
Interpretation of results by analysing different granularity levels and checking model's parameters
Adjust final estimates taken into consideration additional information available or stakeholders' business sense
Monitor users' adjustments

Table C.2: Main sections of the online tool

Section	**Description**
Sources	Source management section, allowing users to update the model's inputs and check the reliability of each source
Historical Analysis	Historical model's most recent estimates, including multiple splits and enabling users to edit parameters and values
Forecast Analysis	Forecast model's most recent estimates, allowing users to change some key market drivers in order to adapt the results to the market evolution
User management	User management system, allowing users to add new users and edit their permissions