Substitution between Cigarettes and E-cigarettes in Canada: with Exercises in Price Construction

Mahnoor Babar (261032538)

McGill University

12th August 2022

Abstract

Despite the claims of e-cigarettes being a healthier alternative to cigarettes, there are fears of substitution from e-cigarettes to regular cigarettes. While other types of analyses are prevalent, there is a gap in the literature with regards to the empirical analysis of the market, especially in Canada. By analysing both individual and aggregate data, this paper considers the possible substitution between tobacco products with a primary focus on cigarettes and e-cigarettes using random utility discrete choice models. The results show that more so that the substitution from e-cigarettes to cigarettes, the real problem may be the substitution from cigarettes as well as other tobacco products towards e-cigarettes.

1 Introduction

With the decline in smoking among the younger demographic, tobacco companies introduced a new product in 2011- e-cigarettes. This "healthier", "cooler" alternative quickly proliferated and became especially popular among middle and high school students.

One the one hand, e-cigarettes are touted as a smoking cessation tool as individuals can get their nicotine fix without some of the harmful chemicals found in regular

cigarettes. On the other hand, there are fears that e-cigarette use may in-turn be leading to cigarette use once individuals are addicted and in search of a stronger product. While the former is likely the case for individuals that start off with regular cigarettes and so a "better" use for e-cigarettes, the later is likely for individuals who started off using e-cigarettes and so get into trying cigarettes for the first time meaning that newer generations being addicted to nicotine. This makes it essential to understand the direction of substitution even though, establishing causality may be difficult.

Using the CTADS data (elaborated in Section 3), I found that the percentage of respondents who had 'ever smoked' e-cigarettes rose from 2013 to 2017. As can be seen from the figure below, despite increase '30-day use' numbers there is a small decline between 2015 and 2017. This makes it crucial to understand the true trajectory of e-cigarette smoking patterns. Importantly,

Using the aforementioned dataset along with information from the census, I am able to analyse both individual and aggregate level data. While individual level variables include information regarding use of tobacco products, demographic and health variables; aggregate information includes share of employed in the population, median income etcetera as well as prices of tobacco products. I use two random coefficient discrete choice model to analyse the data and find that substitution from e-cigarettes to regular cigarettes may not be the real problem here. In fact, the data shows that the high substitution towards e-cigarettes not just from cigarettes but also other tobacco products may be the real issue if e-cigarette use is considered hazardous to health.

To give an overview of the document following this introduction, Section 2 presents the literature review; Section 3 gives an overview of the data used as well as describes an exercise in the construction of e-cigarette prices without which this analysis would not have been possible; Section 4 consists of the methodology and econometric specification used to attain the object; Section 5 contains the results obtained using a multinomial logit as well as a nested logit model; Section 6 puts forth the limitations of this study; and finally, Section 7 concludes the document.

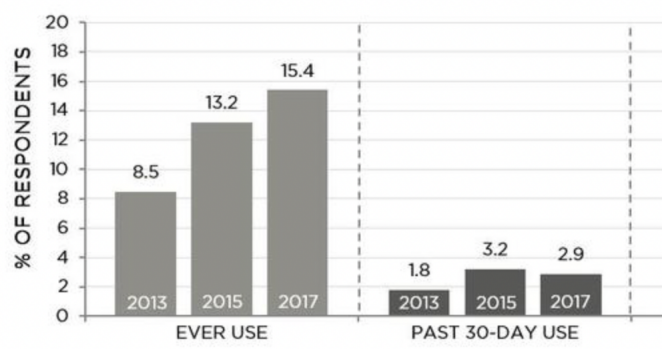

Figure 1: Canadians 'Ever Use' and '30-Day Use' of e-cigarettes

2 Literature Review

My research is related to four distinct but related area:

The first and most basic area is that of smoking determinants. Different socio-economic and health characteristics, for instance, in have a role to play in whether or not an individual takes up smoking. Van Loon et al. (2005) looks at separate determinants for smoking initiation as well as cessation. He finds physical and mental health characteristics to be associated with initiation whereas age, marital status and socio-economic factors were consistently associated with smoking cessation. This study uses the logistic regression.

Dautzenberg and Bricard (2015) use self-reported data from high schools in Montreal, Canada and find that adolescents are more prone to taking up smoking if they are younger, male, from a non-white ethnicity, their parents have lower education levels, have low socio-economic status and if they also use alcohol or other drugs such as marijuana.

This older but interesting work by Schachter (1978) considers that (back then) well known claims that smoking improves the user's mood or allows them to perform better at work. In debunking these myths, he ascribed these claims to the user need to regulate nicotine levels. In doing so, he elucidates the cyclical role of psychological factors that lead to smoking and also get exacerbated because of smoking.

For e-cigarettes in particular Vardavas et al. (2015) find the determinant to be age, sex, marital status, rural or urban residence, and economic status by using logistic regression. This is a study of 27 countries from the European Union. I took inspiration from all these studies to decide on the regressors of my models as will be evident in Section 4.

Second, a significant portion of my paper is dedicated to the construction of e-cigarette prices. Therefore, my work follows prior work that constructed prices in markets where they may not be known. One market where this is common is that of as yet illicit drugs such as marijuana. Jacobi and Sovinsky (2016) address the issue by constructing an empirical distribution of the individual level prices by banking on the prevalence of different types of marijuana used from individual use data and market level price information. Their main goal is to model and estimate the impact of legalisation on use of marijuana. The consumer behavior model includes the impact of access on and disutility of consumption in the Australian market.

The next work in this category is by (Dautzenberg and Bricard, 2015) who try to estimate the number of times an individual uses their e-cigarette per day. Usually, in surveys and studies, individuals are asked to record the number of times per day they smoke cigarettes or e-cigarettes. While this may be straight forward for cigarettes (one can count them), it becomes more complicated for e-cigarettes as one can try counting the number of puffs but otherwise there is no simple way to do so. Therefore, this study uses a type of e-cigarette that records the number of times it is used. The study finds that while there is great variation between individuals and for the same individual during different days of the week etcetera, the median number of puffs is 132. Although this study was conducted in France, I am assuming it is applicable to Canada. While this study does not directly itself construct prices, it is instrumental in helping me construct prices as will be seen in Section 3.

Further, this medical research from California (Yao et al. (2020)) constructs e-cigarette prices along with cigarette prices using 2012–2017 Nielsen Retail Scanner Data. They categorised e-cigarettes as disposable or reusable and so find different prices for the two categories. They estimate separate fixed affects models for the impact of e-cigarette and cigarette prices on per capita disposable e-cigarette, reusable e-cigarette, and cigarette sales controlling for year, quarter, market, and smoke-free air law coverage. They find the price of USD 9.80 for disposable e-cigarettes and USD 19.11 for reusable e-cigarettes.

Third, my paper follows research before me attempting to find substitution patterns between cigarettes and e-cigarettes. Snider et al. (2017) uses self-reported data through a crowd-sourcing tool to find that price and market conditions tend to impact purchasing behavior of regular cigarettes and e-cigarettes users differently which result in different frequency of use patterns. Moreover, it is identified that demand for

cigarettes was the lowest and demand for e-cigarettes was highest in those with greater frequency of e-cigarette use ; when both products were available together, daily e-cigarette users purchased more e-cigarettes, but e-cigarettes served as a substitute for cigarettes in all groups regardless of frequency of use, and lastly the demand for regular cigarette was lower in frequent e-cigarette users when e-cigarettes were also available.

Generally speaking, there is likely substitution from cigarettes to e-cigarettes and vice versa. The former is possibly a cessation strategy for smokers interested in moving towards less harmful products that still satisfy their nicotine cravings (if they choose). The latter would be when someone starts off using e-cigarettes but eventually move towards the "stronger" counterpart - cigarettes. This is also something that has been called the "gateway effect" for drugs.

Chapman et al. (2019) considers this gateway effect for e-cigarettes by analysing longitudinal studies. They find that smoking has been on a decline among the youth in the northern hemisphere even before the introduction of e-cigarettes in the market. But because big tobacco companies needed to continuously attract the younger generation in order to stay relevant, it is in their interest to convince the younger demographic of the appeal of e-cigarettes. For them the declining smoking initiation in the youth is a signal that the gateway effect is not as threatening. They further they rightly point out the difficulty of excluding other ways in which in which vaping and later cigarette smoking may be related. Similarly, Etter (2018) too find little support for this theory despite the political influence it has had over the years.

When it comes to the use of e-cigarettes as smoking cessation Caponnetto et al. (2013) carry a randomised control trial to find that for those trying to quit smoking, the use of e-cigarettes both with or without nicotine, decreased cigarette consumption and resulted in enduring tobacco abstinence. For the 12 and 52 week studies, reduction was documented in 22.3% and 10.3% while abstinence from tobacco smoking was documented in 10.7% and 8.7% respectively. On the other hand, Ghosh and Drummond (2017) finds little evidence for the the efficacy of e-cigarettes as a smoking cessation tool by evaluating data from four randomised-control trials.

This class of literature also gives me an idea of what to expect from my results.

Lastly, my research follows prior works in Industrial Organisation literature that look at substitution between various goods that use estimation techniques that I was interested in using for my analysis. Tuchman (2019) studies the effects of e-cigarette advertising in the United States. She uses individual and aggregate level data to show

that individuals reduce demand for regular cigarettes as a result because they consider the two substitutes. A structural demand model is specified which incorporates addiction and exploits heterogeneity across households in a regression discontinuity difference-in-differences approach. Using estimates from the demand model, the impact of a counter-factual ban is evaluated showing that demand for regular cigarettes may increase.

Similarly, the work of Dubois et al. (2018) has been a huge source of inspiration behind my work. They consider the effect of banning potato chips advertisements in the UK. They exploit consumer level heterogeneity in exposure to television advertising using the random coefficients discrete choice model and simulate a ban on advertising. They find any benefits of the ban might be offset by firms reducing prices and consumers moving towards nutritionally worse foods.

Moreover, Shapiro (2018) considers spillover effects of television advertising in the anti-depressant market on rivals' demand. He used a discrete choice model and banks on discontinuity in advertising created in border regions to determine the extent to which advertising affects rival demand. A supply side analysis to finds significant spillovers and that firms advertise less where positive shocks to rival advertising are likely to occur.

While all these works are looking into the effects of advertising, I have stuck to focusing on the substitution effect alone due to incomplete advertising data available to me (more in the Appendix). This set of works provides critical methodology to help identify substitution between cigarettes and e-cigarettes.

Although works on cigarette and e-cigarette substitution have been done, they have mixed or inconclusive results. My contribution, therefore, is to use Industrial Organization techniques perspective. My contribution to the literature, therefore is the use of IO techniques to extract the extent of substitution between cigarettes and e-cigarettes, which to the best of my knowledge has not been conducted in this manner.

3 Data

3.1 Smoking Information

In order to carry out my research, I have used the Canadian Tobacco, Alcohol and Drugs Survey (CTADS) for the three years it is available: 2013, 2015 and 2017. It is a general population survey on the use of tobacco, alcohol and drug such as marijuana among Canadians aged 15 years and older. It is conducted by Statistics Canada on behalf of Health Canada. The total observations for all three survey years are 44,146 but represents a weighted total of the Canadian population for each of the years. It has been conducted via telephone interviews. The data contains information on the province, whether the residential locality is urban or rural, household size, ages, information on the frequency of regular and e-cigarette use, detailed information on attitudes around smoking, quitting attempts along with reasons for doing so are also available. While the survey has remained largely consistent, some new questions are added in 2015 and 2017, especially on e-cigarette use. The dataset was accessed to McGill University's data office as it is otherwise restricted.

Further, to bolster my analysis I felt the need to use additional variables which according to the literature possibly affected the choice of consuming tobacco products. For this purpose I used the National Household Survey (NHS). The NHS complements the census data to provide information about Canadians by their demographic, social and economic characteristics. Because this data is available every five years, I used data from 2011, 2016 and 2021. The 2016 census reports 2015 data and so has been assigned directly to 2015. An average of 2011 and 2016 data has been used for 2013; and an average of 2016 and 2021 has been used for 2017. I extracted the variables median total income, employment rate, share of immigrated population in the province and rate of tertiary education completion. These I took for males and female separately in an attempt to increase the variation in my data. These were matched onto my existing CTADS dataset simply for each gender in a given province.

3.2 Use of Tobacco Products

	Survey year			
	2013	2015	2017	Total
Use(%)				
No	56.46	54.44	59.42	56.82
Cigarette	4.03	3.30	2.47	3.23
E-cigarette	2.57	3.22	6.44	4.15
Cigar	16.19	16.98	12.32	15.09
Tobacco	1.27	1.06	1.00	1.10
Cigarette and E-cigarette	1.84	1.98	1.41	1.73
Cigarette and Cigar	3.52	2.01	1.63	2.34
Cigarette and Tobacco	0.20	0.23	0.13	0.19
E-cigarette and Cigar	2.59	3.96	4.88	3.86
E-cigarette and Tobacco	0.18	0.25	0.42	0.29
Tobacco and Cigars	3.88	3.88	2.90	3.53
Cigarette, Cigar and Tobacco	0.90	0.77	0.55	0.73
Cigarette, Tobacco and E-cigarette	0.17	0.24	0.24	0.22
E-cigarette, Tobacco and Cigars	0.83	1.56	1.77	1.41
E-cigarette, Cigarette and Cigars	3.62	3.93	2.78	3.43
All four	1.76	2.20	1.66	1.87
Total	100.00	100.00	100.00	100.00
Observations	13,660	14,919	15,567	44,146

Table 1: Use of Tobacco Products by Survey Year

Table 1 above shows the percentage use of each of the categories of tobacco products as well as their use in any combination. This guides the form of the dependent variable. First, our data is equally representative of the three years as they have very similar number of observations. The total dataset for all three years comprises of 44,146 households. The outside option here is the use of no tobacco products. As can be seen from the table, more than 50% of the individuals do not smoke any tobacco products.

Among those who only consume a single product, cigars seem to be the most popular with about 15% of the individuals consuming them. The next most popular category

here is e-cigarettes attracting over 4% of the consumers each year. However unlike other products combinations, the use of e-cigarettes has more than doubled between 2015 and 2017.

Among the pairs of products commonly used together, e-cigarettes and cigars as well as tobacco and cigars are a popular choice, attracting over 3% of the consumers each. The use of e-cigarettes, cigarettes and cigars attracts an equal proportion of consumers. A small proportion of consumers - nearly 2% on average use all four tobacco products.

3.3 Demographics

Table 2 shows demographic information for each of the survey years. For each of the years, the female population is slightly over-represented as compared to the male population. Changes in marital status across the years is reflective of the changing societal norms and attitudes towards marriage and relationships. The average age for each of the years seems pretty different, especially for 2017 which is more than 10 years younger on average as compared to both prior years. This means age might be an important control variable as it may significantly affect the choice of tobacco products used. Average income seems to have successively increased over the years.

	Survey year			
	2013	2015	2017	Total
Sex				
Male	45.47%	46.08%	47.45%	46.37%
Female	54.53%	53.92%	52.55%	53.63%
Marital Status				
Living common-law/Married	39.19%	59.78%	27.10%	41.89%
Widowed/Divorced/Separated	13.32%	14.30%	6.97%	11.41%
Single/Never married	47.50%	25.91%	65.93%	46.70
Locality				
Urban	71.19%	70.92%	74.3%	72.2%
Rural	28.81%	29.08%	25.7%	27.8%
Immigration				
Share of population	0.127%	0.145%	0.006%	0.090%
Tertiary Education				
Male	41.53%	43.30%	47.03%	44.11%
Female	55.42%	57.82%	60.86%	58.12%
Employment rate				
Male	64.78%	64.86%	63.43%	64.32%
Female	57.33%	57.06%	56.59%	56.89%
Age (Mean)				
Male	38.44	40.37	30.35	36.16
Female	41.44	44.17	32.17	39.16
Income (Mean)				
Male	35508	41686	42422	40077
Female	26169	28580	32261	29093

Table 2: Demographic information by Survey Year

3.4 Prices and Taxes

In addition, I will be using information available on prices for each of the three years. For cigarettes, cigars and tobacco these were retrieved through Statistica (2021a) and contain information disaggregated at the provincial level. The unit for cigarettes is one cigarette, for cigars it is one cigar, and for tobacco it is one gram. Table 3 below shows national average price for each of the tobacco products for the survey years as well as the revenue per capita information from the same source. Revenue information is critical to calculating provincial prices as we will see later in this section. The e-cigarette prices mentioned in this table were calculated using the CPI and will be explained in the next subsection.

	Survey year		
	2013	2015	2017
Prices(CAD)			
Cigarettes	0.479	0.478	0.469
E-Cigarette	0.253	0.333	0.450
Cigar	0.638	0.716	0.740
Tobacco	0.288	0.307	0.325
Revenue per Capita(CAD)			
Cigarettes	453.4	502.4	483.1
E-Cigarette	12.5	24.1	33.1
Cigar	9.8	9.2	8.1
Tobacco	8.5	6.9	6.1

Table 3: Average Prices and Revenue per Capita

Figure 2 illustrates the changes in the average prices of tobacco products over the years. According to this figure, prices increase for all products over the years except for cigarettes which show very little change. However, this average price masks considerable variation in price of cigarettes between provinces and across years as we see next.

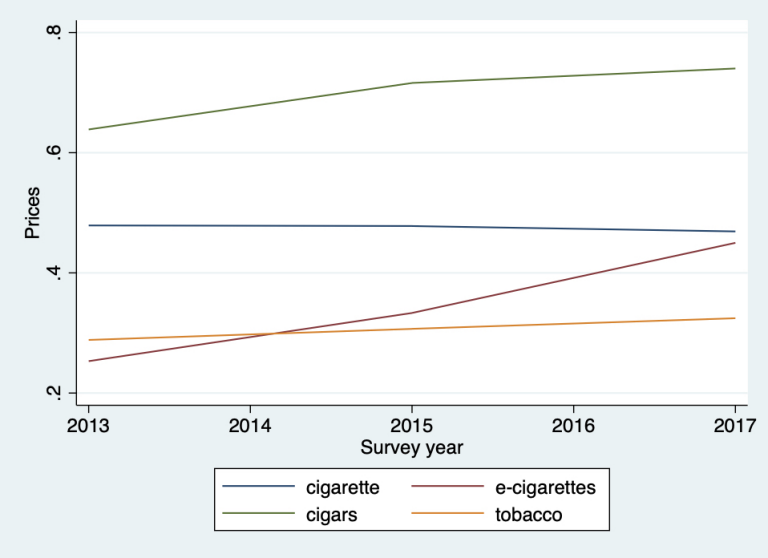

Figure 2: Average Prices by Product

	Survey year								
	2013			2015			2017		
	P	T	W	P	T	W	P	T	W
Province									
NL	0.485	0.205	0.280	0.483	0.235	0.248	0.487	0.245	0.242
PE	0.514	0.225	0.289	0.494	0.225	0.269	0.501	0.250	0.251
NS	0.537	0.235	0.302	0.507	0.255	0.251	0.515	0.275	0.239
NB	0.420	0.190	0.230	0.419	0.190	0.229	0.425	0.255	0.170
QC	0.385	0.125	0.259	0.375	0.149	0.226	0.382	0.149	0.233
ON	0.402	0.124	0.279	0.403	0.139	0.263	0.409	0.165	0.245
MB	0.565	0.290	0.270	0.588	0.295	0.293	0.601	0.295	0.311
SN	0.529	0.250	0.279	0.513	0.250	0.263	0.526	0.270	0.256
AB	0.444	0.200	0.244	0.470	0.200	0.270	0.482	0.250	0.232
BC	0.487	0.223	0.223	0.520	0.239	0.239	0.546	0.239	0.239

Table 5: Cigarette Prices, Taxes and Wholesale Prices by Province

Table 5 shows the price, taxes and wholesale prices (difference of prices and taxes) for cigarettes. Price information was retrieved from Statistica (2021b) while provincial excise tax information was first found in Worrell and Hagen (2021) but has actually been sourced from the statistics collected by the Propel Centre for Population Health Impact. Since this data is only available from 2006 to 2014, I used excel's trend

function to generate prices for the outer years (I checked these generated prices against the national price averages available and the trend seems to do a decent job. These were converted into per piece prices (divided by 200). As we can see from these numbers, prices and excise taxes on cigarettes are different across province with the highest prices in Manitoba in each of the years and the lowest prices in Quebec. Tax rates seem to correspond to prices as provinces with the highest prices also have the highest taxes.

Figure 3, 4 and 5 show cigarette prices, taxes and wholesale prices over the years for each of the ten provinces of Canada, respectively. On the whole, both cigarette prices and taxes seem to be on the rise in almost all provinces. However, wholesale prices paint a different story. It seems on the whole, whole sale prices are declining throughout the provinces with some exceptions.

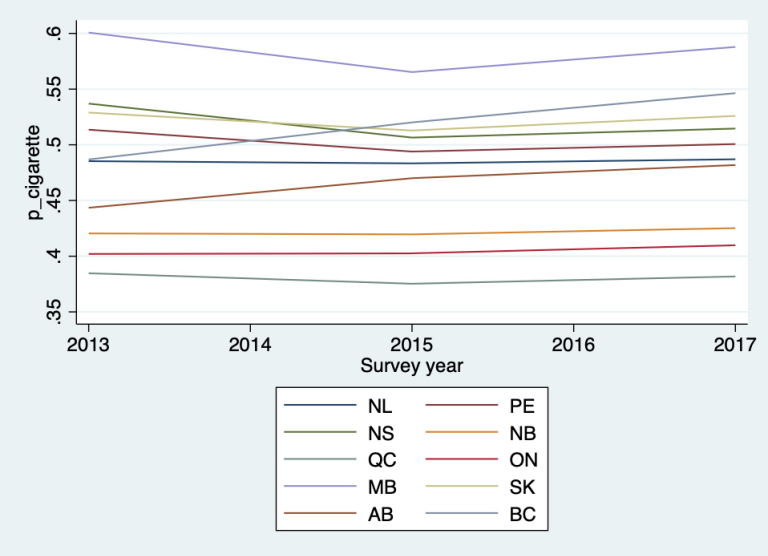

Figure 3: Cigarette Prices by Province

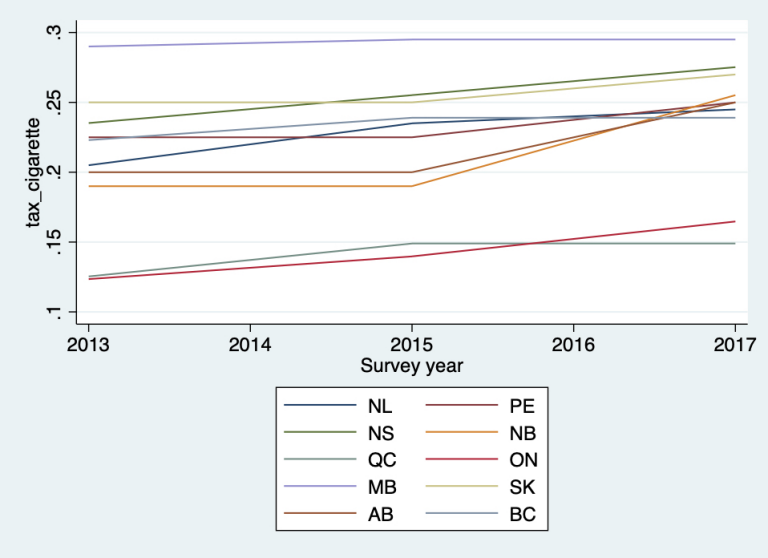

Figure 4: Cigarette Taxes by Province

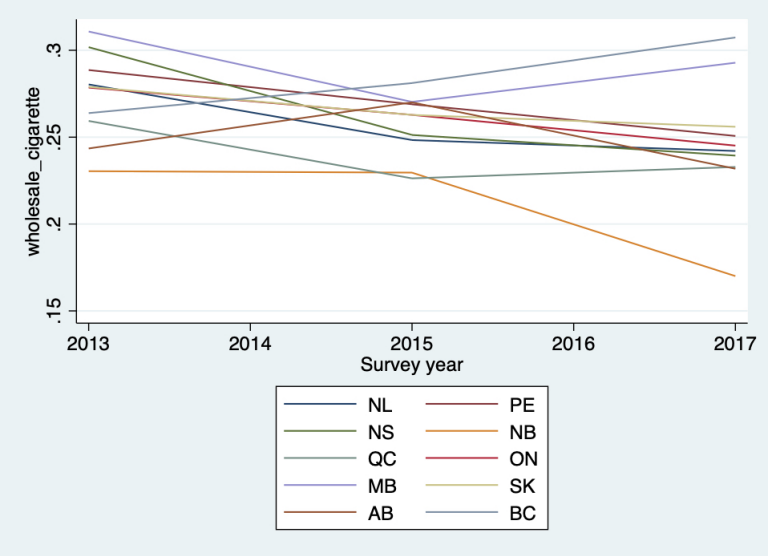

Figure 5: Cigarette Wholesale Prices by Province

Another category of taxes applicable to all tobacco products are the General Sales Tax (GST) and Provincial Sales Tax (PST). These have been given by Table 6 for each of the survey years. While these too have changed over the years, they have stayed relatively more persistent than excise taxes. Some of the provinces have eliminated PST and instead use Harmonised Sales Tax (HST), given under GST in this table. Over these years, Nova Scotia has had the highest GST while Quebec has had the highest PST.

	Survey year					
	2013		2015		2017	
	GST	PST	GST	PST	GST	PST
Province						
Newfoundland and Labrador	13%	0%	13%	0%	15%	0%
Prince Edward Island	5%	10%	14%	0%	15%	0%
Nova Scotia	15%	0%	15%	0%	15%	0%
New Brunswick	13%	0%	13%	0%	15%	0%
Quebec	5%	9.97%	5%	9.97%	5%	9.97%
Ontario	13%	0%	13%	0%	13%	0%
Manitoba	5%	7%	5%	7%	5%	8%
Saskatchewan	5%	5%	5%	5%	5%	6%
Alberta	5%	0%	5%	0%	5%	0%
British Columbia	5%	7%	5%	7%	5%	7%

Table 4: GST and PST by Province

3.4.1 Price Construction

In order to obtain an adequate demand model, I was lacking one crucial ingredient: the prices of e-cigarettes! Without these, there would be no demand. This is a relatively new product in the market and so there isn't as much information available about the category that is ready to use. It is important to note the e-cigarettes or vaping devices are differentiated products that come in a range of prices. By carrying out this exercise, I am aiming for a somewhat average price. I am also not distinguishing between a disposable and reusable device. All additional costs of e-juice, battery etcetera are considered part of this price. Since, I am also aiming for prices that are comparable with the other categories of tobacco products, I am taking a "puff measure". While different e-cigarette brand and researchers provide different answers to the equivalence of e-cigarettes and cigarettes, I am taking 5 puffs of an e-cigarette to be equal to that of a cigarette. Hence, the unit price for e-cigarettes is the price of 5 puffs.

I consider this price calculation exercise to be one of my unique contributions to this literature. Such an exercise can be particularly useful for other products where prices are difficult to observe such as illicit drugs. This is done using a combination of micro and macro data. This was achieved using a combination of micro and macro data as my main dataset is a household survey at the province level that records individual choices. This allows estimation of prices at the province-year level.

In order to calculate this price, I try several different methods. The first way to find price is using the CPI index. The CPI for tobacco is retrieved from Statistica

(2021b) which uses StatsCan data available from 2003 to 2021. It uses 2002 as the base year. Canada uses the Laspeyres index which is as follows (Price analytical series, StatsCan):

$$CPI = \frac{P_t * Q_b}{P_b * Q_b}$$

where t refers to time period of interest and b refers to base year. This means I only need relevant prices to complete my calculation although quantities are available through Drew (2021) and so are the list of products. I add product prices for years of interest (per unit), take e-cigarette prices as unknown:

$$CPI = \frac{P_{2013\,s} + P_{2013\,c} + P_{2013\,t} + X_{2013\,e}}{P_{2002\,s} + P_{2002\,c} + P_{2002\,t}}$$

For 2002 (the base year), e-cigarette prices were not required as the product had not entered the market. Prices for 2002 were once again extrapolated using excel's trend function which summed up to 1! The other products being summed include cigarettes, cigars and tobacco as previously mentioned [1]. Using this method I get the average national prices for e-cigarettes already mentioned in Table 3 above.

The next task was to find prices for each of the provinces for the survey years. As seen with the prices of cigarettes, average prices can mask considerable variation. Hence, I used revenue information as follows [2]:

$$Price = \frac{Average Revenue from-cigarettes^{3}(by province)}{total no. of units smoked(by province)}$$

To obtain the numerator, e-cigarette revenue per capita was used (as shown in Table 3) and multiplied with the provincial population. The denominator is constructed using the CTADS survey using measures for the average no. of times a smoker smokes per day, multiplied by the total no. of smokers.

Total no. of units smoked (by province) =
Average no. smoked x 360 x no. of smokers in the province

Each of these is obtained by re-weighing the dataset using the weights given in the survey and then extracting provincial averages. This method of calculating prices works very well for cigarettes. That can be said with certainty as the obtained prices corroborates with the prices already available. This serves as a check that the method being used is correct. Calculating prices for e-cigarettes posed one (little) challenge. The dataset contains no variable that notes the frequency of use like it does for cigarettes. This is likely to be a challenge as well since the no. of units of cigarettes consumed is pretty obvious. But for e-cigarettes it is not so clear. Nobody records (or even thinks about) the number of times they puff their devices. My first idea was to go on the internet and look for a measure of the no. of times people use e-cigarettes per day. This study by Dautzenberg and Bricard (2015) in France uses connected e-cigarette devices to track the no. of puffs people take during the day. While this number changes

[1] No accessory costs are included in this calculation as I did not even find the category explicitly listed in the product list

[2] Another option was to use expenditure data from the Canadian Household expenditure survey. However, there the e-cigarette expenditure category was omitted by StatsCan for certain provinces in a year where the data was not considered good enough (marked by an F). Hence, this idea was dropped.

$$Price = \frac{Average HH expenditure one-cigarettes(by province)}{total no. of units smoked(by province)}$$

significantly between individuals, the day of the week etcetera, the median no. of puffs observed is about 132. Going by the 5 puffs per cigarette measure, this is the equivalent of nearly 26 cigarettes. This number will be used in the calculation. The assumption here is that this number is equivalent to that in Canada.

For cigarettes, the average number smoked was multiplied by 360 to obtain a yearly measure. However, for e-cigarettes the data for 2017 (only) records the number of days out of 30 that an individual uses e-cigarettes. For all of Canada, the average is about 9 out of 30 days. Hence, in order to complete the denominator formula sketched above, the average number of puffs per day are used from Dautzenberg and Bricard (2015) previously mentioned, and the number of days out of 30 are used. For 2017, this corresponds to a price of 0.4498 for all of Canada. It can be seen that this price 2017 matches almost perfectly with that calculated for 2017 using the CPI method. This gives me confidence in my calculations. The number of days out of 30 is not available for 2013 and 2015 however, I have used the data available for 2017 and used the pumfid (randomly generated variable identifying individual households) and simply ascribed the 2017 values to it. This method is definitely not without its flaws. It assumes the same amount of use for all 3 years [4] and that the same households were using e-cigarettes throughout the 3 years. However, it does give me enough variation in the data to keep my analysis going. Using this variable in my formula to find e-cigarette prices by province generates remarkably good results. I claim that because the national average prices found using the CPI are 0.293, 0.333 and 0.450 for the three survey years using CPI. When I take averages of the provincial prices generated using the revenue formula, the averages are 0.330, 0.354 and 0.419 respectively. I am excited to see that the prices are quite similar even though the range of the prices is smaller using revenue. Since, I will benefit from using maximum variation in the data, I will be using the provincial prices (found using the revenue formula) for my analysis.

Lastly, all of these prices were corroborated by looking at prices on different Canadian website in 2013, 2015 and 2017 using the Wayback machine. Further, both online and in-store prices for the present were checked. According to the Statistica Tobacco Products webpage, over 99% of all e-cigarette sales come from physical stores.

Table 6 shows the prices obtained for e-cigarettes by province that resulted from the exercise described in this section. Figure 6 captures this variation graphically. In terms of averages, prices have consistently increase but within provinces the trajectory is not as clear.

[4] there seemed to be no "correct" way to generate some change over the years without making additional assumptions

	Survey year		
	2013	2015	2017
Province			
Newfoundland and Labrador	0.2804	0.3162	0.3269
Prince Edward Island	0.2786	0.4496	0.2972
Nova Scotia	0.3204	0.3352	0.2852
New Brunswick	0.2321	0.3956	0.5068
Quebec	0.4466	0.2170	0.4260
Ontario	0.3450	0.3956	0.5502
Manitoba	0.5900	0.2900	0.4197
Saskatchewan	0.2544	0.3726	0.4197
Alberta	0.2743	0.3796	0.3147
British Columbia	0.2208	0.3923	0.3510

Table 6: E-cigarette Prices by province and year

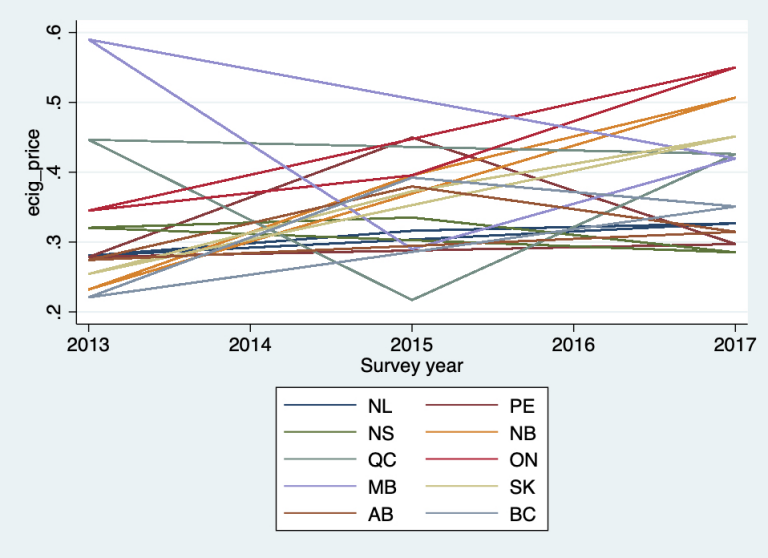

Figure 6: E-cigarette (constructed) Prices by province and year

4 Methodology

4.1 Model

This analysis concerns the substitution between cigarettes and e-cigarettes. An individual choose whether or not to consume one of several categories j of tobacco product in the given market m which is defined as one of the ten provinces of Canada in a given year. The indirect utility that a consumer i receives from using a tobacco product depends on a number of factors such as the price p_{mj} payed for the product category in a given province for each of the survey years, the demographic characteristics d_i such as sex, age, marital status, rural or urban locality, population characteristics such as immigration share of the population, tertiary education share of the population, employment rate and median income x_{mi}; as well as health variable including both physical and mental health h_i. The tax variables include cigarette specific taxes, GST and PST t_{mj}. Province fixed effects have also been included.

In order to extract the full effect of both the age and marital status variables, an interaction term for the two has been included. This helped extract choices of say older, married individuals and how they might be different. In particular, the indirect utility can be represented as such:

$U_{im} = \alpha_{0ij} + \alpha_{1ij} p_{mj} + \beta_{1i}\mathbf{d}_i + \beta_{2i}\mathbf{x}_j + \beta_{3i}\mathbf{h}_i + \xi_{1i}\mathbf{t}_{mj} + \epsilon_{imj}$

where ϵ_{ipt} is the idiosyncratic error term and α_{0j}, α_{1j}, β_1, β_2 and ξ_1, are (vectors of) parameters to be estimated.

The utility from not using tobacco products may be given by

$U_{im} = \alpha_{0ij} + \epsilon_{imj}$

An individual i chooses to smoke a tobacco product in a given market if:

$U_{imj}(\alpha_{0ij}, \alpha_{1ij}, \beta_{1i}, \beta_{2i}, \beta_{3i}, \xi_1) \geq U_{im0}(\alpha_{0i0}, \alpha_{1i0}, \beta_{1i}, \beta_{2i}, \beta_{3i}, \xi_1) > 0$

This approach inform us of the extensive margin. While this approach borrows from models that are now mainstream in Industrial Organisation literature for various applications, I am not familiar with any paper that attempts to use this approach on the topic of smoking.

4.2 Econometric Specification

The random utility discrete choice model shall be used to estimate the model sketched above. The random coefficient distribution is to be conditional on demographic groups i. Preference heterogeneity across consumers seems likely as smoking decisions may differ across income, education and other parameters. Random coefficients in discrete choice demand models are essential to estimate realistic substitution patterns. For instance, I am interested in substitution between different tobacco products.

In order to conduct this analysis, I use two different specifications to model choice of tobacco products. The first is a multinomial logit model that encapsulates the choice of using any category of tobacco product or not. The second is a nested logit

model that considers the aforementioned choice to be nested such that in the first nest an individual decides between either smoking or not. This is a degenerate nest as there are no more choices to be made once an individual decides not the smoke. For the individuals who decide to smoke, the second nest includes the choice between the different categories of tobacco products available for consumption.

The most basic discrete choice model is the simple logit which requires the dependent variable to be one of two variables. Because I am dealing with more than two choices, I opted for the multinomial logit where the dependant variable takes the value 0 if an individual uses no tobacco products, 1 if cigarettes are primarily used, 2 if e-cigarettes are primarily used, 3 if both cigarettes and e-cigarettes are primary drugs, 4 if cigars alone are used and 5 if tobacco or another combination of products is used. To elaborate category 1, for example, also includes use of cigarettes with tobacco or cigarettes with cigars. A major advantage of discrete choice over a simple regression is that it allows greater flexibility in handling non-linear relations. This model is estimated using STATA's mlogit package.

The logit model assumes Irrelevance of Alternative Assumptions (IIA). This is a result of the independent identically distributed error term assumption (iid). What this means for our analysis is that ratio of probabilities of any two alternatives is only equivalent to the characteristics of those two alternative and not any other alternatives (as the case may be).

Further, this implies proportional substitution across alternatives. This can be a helpful property for a well-specified model but it can also be restrictive if it not suitable for a situation (Train (2009)). In my particular case, more than 50% of the individuals in the sample chose not to smoke anything whereas the other half was distributed between the remaining alternatives specified previously under the dependent variable.

Hence, in an attempt to better specify the model, I have used the nested logit model using STATA's nlogit package. The nesting structure has been illustrated in Figure 7. What this allows me to do is to make nests in a way that captures choice. In the first level, an individual decides whether to smoke or not. If an individual decides to smoke, the second level considers the decision to smoked cigarettes, e-cigarettes, both of those, cigars or tobacco and other combinations of products. This nesting of alternatives allows for substitution between smoking particular tobacco products and not smoking to be different, even though proportional substitution between the different tobacco products is the same.

Therefore, the nested logit model allows IIA to hold within each nest but not between nests. In the two different nests the ratio of probability depends on attributes of other alternatives within the nest (Train (2009)).

Possible issue to identifications are as follows:

(i) There is likely a correlation between prices (especially constructed prices for e-cigarettes and the error term ϵ_{imj}. This can lead to an endogeneity problem. This can be resolved by allowing time effects to vary by product category and demographic

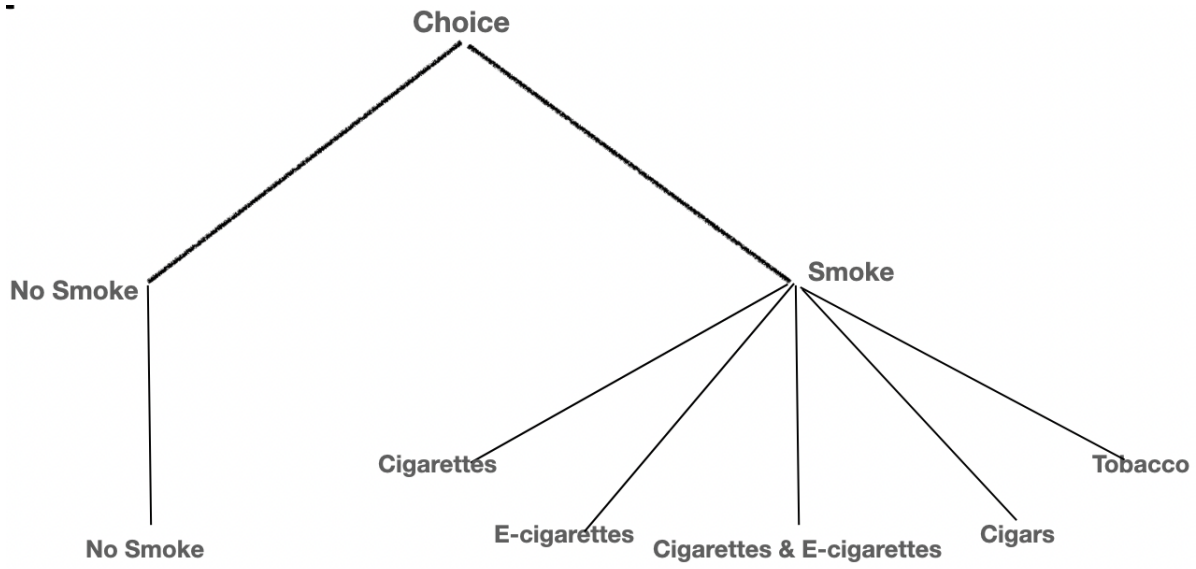

Figure 7: Two-level Nesting Structure

group. This absorbs shocks to demand to particular products.

(ii) Another identification issue is how we identify the distribution of unobserved heterogeneity that is modeled as random coefficients. This can be handled by using data that are at the micro level and that are longitudinal so that we observe each individual making repeated choices. In the dataset I use, certainly every household does not make decisions over the entire survey period but some do. Micro data has been shown to be particularly useful in identifying and estimating substitution patterns as seen in Dubois et al. (2018).

5 Results

This section looks at the results from the two models specified to understand the demand of tobacco products and the choices made by various individuals.

Before estimating the two random utility discrete choice model, I conducted two separate OLS regressions. The first where the independent variable is the number of cigarettes consumed per day and the same regressors previously mentioned. Only the coefficients for urban locality, age, marital status and physical health are significant. The second regression takes the no. of days out of 30 an individual uses e-cigarettes. Here, only the price of e-cigarettes is significant (with a negative coefficient). The Rsquare are 15.37 and 4.44 respectively. Based on the insignificance of coefficients it can be concluded that the linear model is unable to capture the variation in the dependent variable. Because I did not want this model to take away from my current

analysis, OLS results can be found in the Appendix.

5.1 Multinomial Logit Model

Table 8 shows the results of the multinomial (polytomous) logistic regression. The dependant variable is a categorical variable which considers four tobacco products: cigarettes, e-cigarettes, cigars and tobacco. The variable itself takes five different values for which product or combination of products the representative used. It is 0 when the individual[5] chooses not to smoke at all; 1 when the individual uses cigarettes either alone or in combination with cigar or tobacco; 2 when the individual uses e-cigarettes either alone or in combination with cigar or tobacco; 3 if the individual uses both cigarettes and e-cigarettes maybe even in combination with cigars and tobacco; 4 if the individual smokes cigars only. This category was kept exclusive as this is where the second largest group of consumers (15% of our sample falls), the first being non-smokers at 56%; lastly the category is 5 if the user uses a combination other than the ones already mentions, or even uses all four products. While I could have kept the dependant variable with 16 different categories - one for each combination- it was causing my analysis to suffer as some of the categories were pretty small and it would have led to three times the interpretation currently required.

The model below is specified with the non-smoking category as the base category so all interpretations will be in reference to that category. The Pseudo R-squared value is 0.0891 indicating our model shows an 8.91% improvement in fit relative to the null model. The likelihood ratio chi squared test compares the fit of our model with the complete set of predictors with the null model. In this case the P-value is zero and therefore it is significant, we can infer that at least one of the population regression slopes is significantly different from zero. Based on the chi squared test, we can say that the model containing the full set of predictors represents a significant improvement in fit over the null model.

Now, we look at the regression table. Our baseline category is that of cigarette smokers and so we see no coefficients for this category. We can now determine which of the indicators significantly explain whether an individual from a household is a non-smoker vs. a cigarette smoker. Interpretation is done variable by variable for each of the five categories. Starting from the non-smokers category and going up to the tobacco plus category for each variable. [6]

[5] Although this is a household level survey, each household representatives responds to questions regarding their use of tobacco products from their own perspective (i.e. the variables I am looking at records the individual's choice not the household's choice necessarily and so I will be using the word representative or individual hereon wards.

[6] The idea was to defined market as the province year. However, adding province-year to any of my model has resulted in the year variables dropping out due to collinearity and the prov-year interactions to be always insignificant. Therefore, the market has been limited to province only.

VARIABLES	No	Cigarette	E_cigarette	Cig and E_cig	Cigar	Tobacco+
Urban		0.178***	-0.0574	-0.0517	0.0304	0.323***
		(0.0449)	(0.0404)	(0.0507)	(0.0328)	(0.0457)
Age		0.00654***	-0.0503***	-0.0242***	0.0112***	0.000975
		(0.00130)	(0.00166)	(0.00156)	(0.000904)	(0.00135)
Sex		-0.468**	-1.092***	-1.006***	-0.838***	-2.512***
		(0.212)	(0.177)	(0.232)	(0.149)	(0.304)
Div.Marital		0.329***	0.196**	0.347***	-0.251***	-0.153**
		(0.0595)	(0.0946)	(0.0796)	(0.0472)	(0.0751)
Sin.Marital		-0.218***	-0.0566	-0.123**	-0.516***	-0.542***
		(0.0553)	(0.0494)	(0.0580)	(0.0383)	(0.0564)
Phy Health		0.282***	0.0666***	0.458***	-0.0477***	0.0995***
		(0.0235)	(0.0226)	(0.0263)	(0.0178)	(0.0256)
Men Health		0.181***	0.253***	0.269***	0.118***	0.193***
		(0.0239)	(0.0203)	(0.0249)	(0.0180)	(0.0256)
Educ		0.00457	0.00735	0.0245**	-0.00652	-0.00487
		(0.0110)	(0.00999)	(0.0124)	(0.00804)	(0.0156)
Immig		-0.255	-0.826	-0.210	0.0184	0.281
		(0.712)	(0.587)	(0.762)	(0.483)	(0.751)
Empl		-0.00254	-0.0261**	-0.0285*	-0.0338***	-0.0304*
		(0.0142)	(0.0126)	(0.0162)	(0.0106)	(0.0184)
ln Inc		1.002**	0.209	-0.287	1.162***	0.477
		(0.486)	(0.394)	(0.529)	(0.347)	(0.604)
P Cig		5.799**	0.410	-4.099	0.430	3.701
		(2.442)	(2.278)	(2.730)	(1.678)	(2.519)
P Ecig		0.440	0.192	0.101	0.309	0.271
		(0.299)	(0.268)	(0.322)	(0.207)	(0.342)
P Cigar		-0.930	14.96***	13.21***	2.442	-1.806
		(3.240)	(2.702)	(3.318)	(2.176)	(3.475)
P Tob		-19.97	-34.05***	-45.64***	-20.85**	1.280
		(13.11)	(10.37)	(13.32)	(8.839)	(13.71)
Tax Cig		1.129	-2.101	-2.302	0.603	-3.232
		(2.378)	(1.931)	(2.550)	(1.668)	(2.517)
GST		-0.0228	0.118*	-0.0944	0.0213	-0.0771
		(0.0817)	(0.0652)	(0.0845)	(0.0574)	(0.0904)
PST		-0.00576	0.122**	-0.0502	0.0126	-0.0505
		(0.0739)	(0.0598)	(0.0772)	(0.0522)	(0.0800)

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 7: Multinomial Logit results

Moving on to the interpretation of individual coefficients. For an individual in rural area, smoking cigarettes is more likely relative to a person in an urban locality compared to the baseline category of not smoking. The coefficient can be read as 0.178 unit increase in the log-odds on an individual from the rural locality smoking cigarettes relative to not smoking, as compared to their urban counterparts. This coefficient is significant at the 99% confidence level. For e-cigarettes, cigarettes and e-cigarettes, and cigar categories, the coefficient remains insignificant. Similarly the use of tobacco and other combinations is more likely to be used by rural as compared to urban dwellers, relative to the baseline category of not smoking. The coefficient can be read as 0.323 unit increase in the log-odds on an individual from the rural locality smoking tobacco relative to not smoking, as compared to their urban counterparts. This coefficient is significant at the 99% confidence level. It can be said that cigarettes and tobacco may be a more popular choice in rural areas due to the kind of stress levels associated with rural like whereas e-cigarettes and cigars are also sometime a more 'fashionable' choice and therefore relatively preferred in an urban setting.

Moving on to age. Individuals with greater age are more likely to consume cigarettes and tobacco but less likely to consume, e-cigarettes or a combination of cigarettes and e-cigarettes. This is an indicating towards the popularity of e-cigarettes among the younger demographic. Each of the coefficients imply that with an additional years increase in age, the log-odds of an individual smoking cigarettes is 0.006 units greater compared to individuals a year younger; the log-odds of an individual smoking e-cigarettes is 0.05 units lower compared to individuals a year younger; the log-odds of an individual smoking cigarettes and e-cigarettes is 0.024 units lower compared to individuals a year younger; and the log-odds of an individual smoking cigars is 0.011 units greater compared to individuals a year younger. All these coefficients are significant at the 99% confidence level. The coefficient for tobacco remains insignificant.

Females relative to males are less likely to smoke cigarettes as compared to not smoking as the regression slope is negative and significant with a coefficient if 0.46 implying that the log odds of a female smoking cigarettes is 0.46 units less than that of falling into the no smoking category. This result is statistically significant at the 95% confidence level. Similarly, females relative to males are less likely to smoke e-cigarettes; both cigarettes and e-cigarettes; cigars; and tobacco as compared to not smoking as the regression slopes are negative and significant. The coefficient can be read as follows: the log odds of a female smoking e-cigarettes is 1.09 units less than that of falling into the no smoking category; the log odds of a female smoking both cigarettes and e-cigarettes is 1.006 units less than that of falling into the no smoking category; the log odds of a female smoking cigars is 0.838 units less than that of falling into the no smoking category; and the log odds of a female smoking tobacco is 2.51 units less than that of falling into the no smoking category. Each of these results is statistically significant at the 99% confidence level. This results perhaps shows the relatively less proclivity of women to smoke any tobacco product relative to not smoking, in general.

Relative to married individuals, divorced or widowed individuals are more likely to smoke cigarettes compared to not smoking. The coefficient is implying that the log

odds of a divorced or widowed individual smoking cigarettes is 0.32 units more likely than that of falling into the no smoking category. This result is statistically significant at the 99% confidence level. Similarly, relative to married individuals, divorced or widowed individuals are more likely to smoke e-cigarettes compared to not smoking as well as both cigarettes and e-cigarettes. The coefficients are implying that the log odds of a divorced or widowed individual smoking e-cigarettes is 0.19 units more likely than that of falling into the no smoking category and the log odds of a divorced or widowed individual smoking both cigarettes and e-cigarettes is 0.35 units more likely than that of falling into the no smoking category. These result are statistically significant at the 95% and 99% confidence level respectively. On the other handd, relative to married individuals, divorced or widowed individuals are less likely to smoke both cigars or tobacco compared to not smoking. The coefficient is implying that the log odds of a divorced or widowed individual smoking cigars is 0.25 units less likely than that of falling into the no smoking category the log odds of a divorced or widowed individual smoking tobacco is 0.15 units less likely than that of falling into the no smoking category. These results is statistically significant at the 99% and 95% confidence level respectively.

Single individuals relative to married individuals, are less likely to smoke cigarettes compared to not smoking. The coefficient is implying that the log odds of a divorced or widowed individual smoking cigarettes is predicted to decrease by 0.21 units relative the no smoking category. This result is statistically significant at the 99% confidence level. Similarly, relative to married individuals, single individuals are less likely to smoke both cigarettes and e-cigarettes compared to not smoking. The coefficient is implying that the log odds of a single individual smoking both is predicted to be 0.12 units less than that of falling into the no smoking category. This result is statistically significant at the 95% confidence level. Similarly, relative to married individuals, single individuals are less likely to smoke both cigars or tobacco compared to not smoking. The coefficient is implying that the log odds of a single individual smoking cigars is predicted to decrease by 0.25 units relative no smoking category and the log odds of a single individual smoking tobacco is predicted to be 0.54 units less likely than that of falling into the no smoking category. These results is statistically significant at the 99% confidence level.

Less healthy individuals are more likely to smoke cigarettes, e-cigarettes, both and tobacco relative to not smoking, compared to healthier individuals. There is possibly a reverse causation here as individuals become less healthy the more they use tobacco products. The coefficients imply that the log odds of a less healthy individual smoking cigarettes is predicted to increase by 0.28 units relative the no smoking category; the log odds of a less healthy individual smoking e-cigarettes is predicted to increase by 0.06 units relative the no smoking category; the log odds of a less healthy individual smoking both cigarettes and e-cigarettes is predicted to increase by 0.45 units relative the no smoking category; the log odds of a less healthy individual smoking cigarettes is predicted to increase by 0.09 units relative the no smoking category. These result is statistically significant at the 99% confidence level.

Less healthy individuals are more likely to use cigars relative to not smoking, compared to healthier individuals. The log odds of a less healthy individual smoking cigarettes is predicted to decrease by 0.04 units relative the no smoking category. This result is statistically significant at the 99% confidence level.

When it comes to mental health, individuals with weaker mental health are more likely to smoke all products, compared to healthier individuals. While it is likely that individuals with weaker mental health are more likely to opt for smoking, it is also likely that individuals who take up smoking experience more mental health issues leading to a cyclical pattern that may be hard to break from. The coefficients imply that the log odds of an individual with weaker mental health smoking cigarettes is predicted to increase by 0.18 units relative the no smoking category; the log odds of an individual with weaker mental health smoking e-cigarettes is predicted to increase by 0.25 units relative the no smoking category; the log odds of an individual with weaker mental health smoking both cigarettes and e-cigarettes is predicted to increase by 0.26 units relative the no smoking category; the log odds of an individual with weaker mental health smoking cigars is predicted to increase by 0.11 units relative the no smoking category; and the log odds of an individual with weaker mental health smoking tobacco is predicted to increase by 0.19 units relative the no smoking category. These result are statistically significant at the 99% confidence level.

The coefficient for education is positive and significant for individuals with higher education more likely to smoke both cigarettes and e-cigarettes relative to not smoking. The coefficient implied the log odds of an individual with higher education smoking both cigarettes and e-cigarettes is predicted to increase by 0.02 units relative the no smoking category. This result is statistically significant at the 95% confidence level. The other results remain insignificant.

The coefficient for employment is negative and significant for employed individuals indicating they are less likely to smoke e-cigarettes relative to not smoking. The coefficient implies the log odds of an employed individual smoking e-cigarettes is predicted to decrease by 0.002 units relative the no smoking category. This result is statistically significant at the 95% confidence level. The other results remain insignificant. The regression slope for employment is negative and significant for employed individuals indicating they are less likely to smoke both cigarettes and e-cigarettes relative to not smoking. The coefficient implies the log odds of an employed individual smoking e-cigarettes is predicted to decrease by 0.02 units relative the no smoking category. This result is statistically significant at the 90% confidence level. The other results remain insignificant. Coefficients for cigar and tobacco use are very similar and also statistically significant but the one for smoking cigarettes is insignificant. This could likely point to unemployed individual being more stressed as a result of being unemployed and therefore taking up smoking.

The coefficient for log of income is positive and significant for employed individuals indicating that people with higher income are more likely to smoke cigarettes relative to not smoking. The coefficient implies the log odds of people with greater incomes is predicted to increase by 1.002 units relative the no smoking category. This result

is statistically significant at the 95% confidence level. The coefficient for cigars is also positive and significant at the 99% confidence level. The remaining coefficients are insignificant.

Price of cigarettes has a positive and statistically significant regression slope for cigarettes only. Higher prices are more likely to cause individuals to smoke more cigarettes relative to not smoking. This again could indicate cigarette prices rising in response to increased demand since it is has a relatively inelastic demand. According to the coefficient, for increased prices log odds of individuals smoking cigarettes is predicted to increase by 1.002 units relative the no smoking category. This result is statistically significant at the 95% confidence level.

E-cigarette prices remain insignificant for all categories. Cigar prices have a positive and significant affect on the probability consuming e-cigarettes category as well as the cigarettes and e-cigarettes category. With each unit increase in price of cigars, the log odds of consumption of the aforementioned categories increases by 14.9 and 13.2 units respectively. These results are statistically significant at the 99% confidence level. Tobacco prices have a negative and significant affect on the probability consuming cigarettes, e-cigarettes as well as the cigarettes and e-cigarettes category. With each unit increase in price of tobacco, the log odds of consumption of the aforementioned categories decreases by 34.05, 45.64 and 20.85 units respectively. These results are statistically significant at the 99% confidence level

no smoking and cigar smoking category relative to cigarette smoking category. The regression slope for education is positive for cigarette and e-cigarette as well as tobacco plus category relative to cigarette smoking category. All four of these results are statistically insignificant. For e-cigarette smoking category relative to cigarette smoker category, the regression slope is positive and significant at the 90% confidence interval implying the log odds of a person falling into this category as compared to the cigarette smoking category increases by 0.103 units with a unit increase in tertiary education. This could indicate an increased awareness of tobacco products that are relatively better to use than others.

Following this full model, a restricted model is run containing only price and quantity variables as regressors. This model has an Rsquared of 6%. The lrtest for these returns a p-value of 0. Since this is less than 0.05, we can reject the null at the 99% confidence interval to say the full model does sufficiently increase the explanatory power of the model over the restricted one.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	No	Cigarette	E_cigarette	Cig and E_cig	Cigar	Tobacco+
Urban	-0.0181***	.0124***	-0.0046**	-0.0042**	.0001	0.0144***
Age	0.00058	.0010***	-0.0025***	-.0009***	0.0016***	0.0001**
Sex	0.2383***	-0.0042	-0.0368***	-0.0295**	-0.0622***	-0.1053***
Marital Status	0.0342***	.0005	0.0020	0.0012	-0.0278***	-0.0101***
Physical Health	-.0313***	.0210***	.0009	0.0201***	-0.0133***	0.0024*
Mental health	-.0428***	.0103***	0.0096***	0.0099***	0.0066***	0.0061***
Education	-.00057	.0003	0.0003	0.0011**	-0.0010	-0.0002
Immigrants	0.0350	-0.0149	-0.0398	-0.0066	0.0095	0.0168
Employment	0.0056***	0.0006	-0.0008	-0.0009	-0.0033***	-0.0010
ln Income	-0.1703***	0.0686	-0.0032	-0.0264	0.1209***	0.0104
Price Cigarette	-.3692	0.4711	-0.0081	-0.2266**	-0.0224	0.1553
Price Ecigarette	-.0687*	0.0301	0.0043	-0.0005	0.0268	.0079
Price Cigar	-1.0474***	-0.1906	0.7102***	0.5913	0.1016	-0.1651
Price Tobacco	5.5326***	-1.1381	-1.3619***	-1.8850***	-1.6253*	0.4778
Tax Cigarette	.1326	0.1128	-0.1000	-0.1050	0.1101	-0.1505
GST	.0006	-0.0018	0.0062**	-0.0044	0.0030	-0.0037
PST	-0.0019	-0.0006	0.0062***	-0.0024	0.0013	-0.0026

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

Table 8: Predicted Probabilities

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	No	Cigarette	E_cigarette	Cig and E_cig	Cigar	Tobacco+
Urban	-0.3719	0.1773	-0.1131	-0.1118	0.0015	0.3732
Age	0.0351	0.4472	-1.8011	-0.7414	0.4603	0.1094
Sex	0.5870	-0.0730	-1.0813	-0.9256	-0.7024	-0.4168
Marital Status	0.1124	0.0117	0.0789	0.0050	-0.4198	-0.4169
Physical Health	-0.1094	0.5136	0.3829	0.8969	-0.2132	0.1094
Mental Health	-0.1361	1.9696	0.3663	0.4018	0.0974	0.2471
Education	-0.0478	0.0151	3.3483	1.2234	-0.0147	-0.2946
Immigrants	0.0051	-0.0152	-0.0686	-0.0122	0.0063	0.0306
Employment	0.5481	0.4352	-1.025	-1.1736	-1.4917	1.2854
ln Income	-2.8466	8.0020	-0.6363	-5.6244	59.2613	2.2776
Price Cigarette	-0.2810	2.5012	0.0740	-2.1989	-0.0078	1.4639
Price Ecigarette	0.0407	-0.1245	0.0303	0.0034	0.0726	0.0594
Price Cigar	1.1771	-1.4931	9.4984	8.4648	0.5235	2.3315
Price Tobacco	2.728	-4.1680	-7.9909	-11.8381	3.6661	-2.9573
Tax Cigarette	0.0459	0.2729	-0.4127	-0.4645	0.1749	-2.0826
GST	0.0088	0.1890	1.1078	-0.8387	0.2092	-0.6975
PST	-0.0102	0.0223	0.3963	0.1659	0.0324	-0.1744

Table 9: Elasticities of Demand

Table 8 above shows the margins or predicted probabilities for each of the outcome. Starting with the first variable urban, for an individual in rural area relative to urban area, the probability of the individual falling in the not smoking category decreases by 1.81 percentage points. This result is significant at the 99% confidence level. For the same group, the probability of falling in the cigarette category increases by 1.24 percentage points. This result is significant at the 99% confidence level. The probability of falling in the e-cigarette category decreases by 0.46 percentage points. This result is significant at the 95% confidence level. The likelihood of falling in the cigarette and e-cigarette category decreases by 0.42 percentage points. This result is significant at the 95% confidence level. Results for the cigar category are insignificant. The likelihood of falling in the tobacco+ category increases by 1.44 percentage points. This result is significant at the 99% confidence level.

The next variable is age, for every one year increase in the individual's age, the probability of the individual falling in the cigarette category increases by 0.1 percentage points. This result is significant at the 99% confidence level. Similarly, the probability of falling in the e-cigarette category decreases by 0.25 percentage points. This result is significant at the 99% confidence level and is consistent with the observation that e-cigarettes are a more popular choice among the younger demographic. The likelihood of falling in the cigarette and e-cigarette category decreases by 0.02 percentage points. This result is significant at the 99% confidence level. The probability of falling in the cigar category increases by 0.16 percentage points. This result is significant at the 99% confidence level and may point to decreased use of cigars in older ages. The probability of falling in the tobacco+ category increases by 0.01 percentage points. This result is significant at the 99% confidence level. Results for the no smoking category are insignificant.

For sex, the probability females falling in the no smoking category, relative to males is 23.8 percentage points greater. This result is significant at the 99% confidence level. Results for the cigarette category are insignificant. The likelihood females falling in the e-cigarette category, relative to males is 3.68 percentage points lower. This result is significant at the 99% confidence level. The likelihood females falling in the cigarette and e-cigarette category, relative to males is 2.95 percentage points lower. This result is significant at the 99% confidence level. The likelihood females falling in the cigar category, relative to males is 1.6 percentage points greater. This result is significant at the 99% confidence level. The likelihood females falling in the tobacco category, relative to males is 0.01 percentage points greater. This result is significant at the 99% confidence level. These results point to the lower inclination of females to smoke any cigarettes and e-cigarettes but slightly higher towards cigars and tobacco, relative to males.

For marital status, moving from married to divorced, and divorced to single, the probability of an individual falling into the no smoking category increases by 3.42 percentage points. This result is significant at the 99% confidence level. The probability of an individual falling into the cigar category decreases by 2.78 percentage points. This result is significant at the 99% confidence level. The probability of an individual

falling into the tobacco category decreases by 1.01 percentage points. This result is significant at the 99% confidence level. Results for cigarette category and e-cigarette category remain insignificant.

For physical health, for individuals with decreasing health levels, the probability of falling into the no smoking category decreases by 3.13 percentage points. This result is significant at the 99% confidence level and shows the obvious relation the smoking leads to poor health. The probability of falling into the cigarette category increases by 2.1 percentage points. This result is significant at the 99% confidence level. The probability of falling into the cigarette and e-cigarette category decreases by 2.01 percentage points. This result is significant at the 99% confidence level. The probability of falling into the cigar category decreases by 1.33 percentage points. This result is significant at the 99% confidence level. The probability of falling into the tobacco+ category increases by 0.24 percentage points. This result is significant at the 99% confidence level. Results for the e-cigarette category are insignificant. Of all these results, the cigar category is the only one indicating the use of cigars leading to better health which is counter-intuitive. However, it may point to the use of cigars by individuals who are relatively wealthy and so can afford to maintain good health despite use of cigars. For mental health, for individuals with decreasing mental health levels, the probability of falling into the no smoking category decreases by 4.25 percentage points. The probability of falling into the cigarette category increases by 1.03 percentage points. The probability of falling into the e-cigarette category decreases by 0.96 percentage points. The probability of falling into the cigarette and e-cigarette category decreases by 0.92 percentage points. The probability of falling into the cigar category decreases by 1.33 percentage points. The probability of falling into the tobacco+ category increases by 0.61 percentage points. All results are significant at the 99% confidence level. These results are very clear. Smoking any tobacco product has a negative impact on mental health. The causation can go both ways: use of tobacco products may be leading to poor mental health, and poor mental health may be leading to take-up of tobacco products.

The education variable is only significant for cigarettes and e-cigarettes category. Increased education levels lead to a 0.11 percentage point increase in the probability of falling into this category. This result is significant at the 95% confidence level. While the expectation is that increased education levels lead to more information on the adverse affects of using tobacco products, these results may be pointing to the perception that e-cigarettes are relatively better than cigarettes.

Increased employment levels make the probability of falling into the no-smoking category 5.6 percentage points higher. This result is significant at the 99% confidence level. Moreover, the probability of falling into the cigar category 0.33 percentage points lower. This result is significant at the 99% confidence level. This may point to the use of tobacco products among the employed as there might be a culture of use of tobacco products and peer pressure or the desire to fit in may lead to increase takeup of tobacco products.

Increased income levels make the probability of falling into the no-smoking category 17.03 percentage points lower. This result is significant at the 99% confidence level.

The probability of falling into the cigar smoking category 12.09 percentage points higher. This result is significant at the 99% confidence level indicating cigars to be a popular choice at higher income levels. These results corroborate the results from the employment variable.

Higher price of cigarettes make the probability of falling into the cigarette and e-cigarette category is 22.66 percentage points less likely. This result is significant at the 95% confidence level. Remaining results are insignificant.

Higher price of e-cigarettes make the probability of falling into the no smoking category 6.87 percentage points less likely. This result is significant at the 90% confidence level. Remaining results are insignificant. This result once again seems to point to rising e-cigarette demand despite increasing prices.

Higher price of cigars make the probability of falling into the no smoking category 104.7 percentage points less likely. This result is significant at the 99% confidence level. The probability of falling into the e-cigarette category 71.02 percentage points more likely. This result is significant at the 99% confidence level and indicates possible substitution towards e-cigarettes.

Higher price of tobacco make the probability of falling into the no smoking category 553.2 percentage points more likely. This result is significant at the 99% confidence level. The probability of falling into the e-cigarette category 136.19 percentage points less likely. This result is significant at the 99% confidence level. The probability of falling into the cigarette and e-cigarette category 188.5 percentage points more likely. This result is significant at the 99% confidence level. The probability of falling into the cigar category is 162.5 percentage points more likely. This result is significant at the 99% confidence level.

With higher GST, the probability of falling into e-cigarette category increased by 0.02 percentage points. This result is significant at the 95% confidence level. With higher PST, the probability of falling into e-cigarette category increased by 0.06 percentage points. This result is significant at the 99% confidence level. These results may actually be pointing out at the reverse causality. With both e-cigarette use, GST and PST rising over the years, it may be the e-cigarette consumption has increased despite increase in GST and PST.

Table 9 above shows elasticities of demand for each of the categories. These elasticities have been calculated at the mean. Only own elasticities are being reported below as cross price elasticities as cross price elasticities are constant under the IIA assumption.

Own price elasticity of demand for cigarettes is elastic[7] as indicated by the cigarettes category with a value of -0.281. Every 1% increase in price of cigarettes causes demand for cigarettes to fall by 250%. This points towards the existence of a close substitute that consumers could turn to in the face of rising prices. For e-cigarettes own price elasticity of demand is inelastic[8] with a value of -0.0303. Every 1% increase in price

[7]Elastic own price demand is when the elasticity value is greater than 1, indicating that a change in price causes more than proportional change in demand.

[8]Inelastic own price demand is when the elasticity value is less than 1, indicating that a change in

of e-cigarettes causes demand for e-cigarettes to increase by 3.03%. For cigars too, own price elasticity of demand for cigars is inelastic with a value of -0.523. Every 1% increase in price of cigars causes demand for cigars to increase by 52.35%. This could be as cigars seem to have a very specific market that attracts the wealthy and so demand does not sway as much in face of rising prices and in fact may even be more of a factor for existing or prospective consumers. Lastly for tobacco, own price elasticity of demand is elastic with a value of -2.957. Every 1% increase in price of tobacco causes demand for tobacco to decrease by 295%.

The Likelihood Ratio (LR) test is carried out to determine whether our multinomial logit model add significantly to the explanation of the dependant variable. It assesses the goodness of fit of two competing statistical models based on the ratio of their likelihoods. In the present case, the full model is the model presented above whereas the null model does not consist of any of the regressors. So the null hypothesis is that the current model does not explain significant variation in fit of the dependant variable. The results obtained show an LR chi2 value of 11064.17 and a p-value of 0. Indicating we can reject the null hypothesis that at the 99% confidence level.

5.1.1 Substitution

Next, I calculate the marginal rate of substitution between products the different product categories. This was done using ratios of coefficients of the multinomial logit regression. The coefficients are equal to marginal utilities.

VARIABLES	Cigarette	E-cigarette	Cig and E_cig	Cigar	Tobacco+
Cigarette	-	0.46	0.97	5.68	0.34
E-cigarette	13.17	-	39.58	7.90	4.72
Cig and E_cig	11.83	0.32	-	0.25	0.32
Cigar	-6.23	0.01	0.30	-	-0.71
Tobacco+	-0.29	-0.01	0.08	-0.12	-

Table 10: Marginal Rate of Substitution

The row indicates the category an individual may substitute away from while the column indicates the category they may substitute towards. To interpret the results, for a person smoking cigarettes, the marginal rate of substitution towards e-cigarettes is 13.17 meaning that in terms of value, an additional dollar spent on cigarettes is equivalent to spending $13.17 on e-cigarettes. Similarly, from the same category, an additional dollar spent on cigarettes is equivalent to spending $11.83 on both cigarettes and e-cigarettes. Both these categories show very high substitution is possible from

price causes less than proportional change in demand.

cigarettes. On the other hand, an additional dollar spent on cigarettes is equivalent to spending -$6.23 on cigars and -$0.29 on tobacco. Both these categories indicate it is less likely for substitution to take place away from cigarettes towards tobacco but more towards cigars.

Moving on to substitution from e-cigarettes, the marginal rate of substitution towards cigarettes is 0.46 meaning that in terms of value, an additional dollar spent on cigarettes is equivalent to spending $0.46 on e-cigarettes. Similarly, from the same category, an additional dollar spent on e-cigarettes is equivalent to spending $0.32 on both cigarettes and e-cigarettes. Moreover, an additional dollar spent on e-cigarettes is equivalent to spending $0.01 on cigars and -$0.01 on tobacco. All these results indicate it is much less likely for substitution to take place away from e-cigarettes towards these categories.

For individuals using both cigarettes and e-cigarettes, the marginal rate of substitution towards cigarettes is 0.97 meaning that in terms of value, an additional dollar spent on both is equivalent to spending $0.97 on e-cigarettes. Since this result shows near MRS near 1, consumers are almost indifferent between the two categories. Similarly, from the same category, an additional dollar spent on e-cigarettes is equivalent to spending $39.58 on e-cigarettes only. When compared with the results for cigarettes it can be seen it is highly likely that consumers of both products will move towards e-cigarettes rather than cigarettes. Moreover, an additional dollar spent on cigarettes and e-cigarettes is equivalent to spending $0.30 on cigars and $0.08 on tobacco. These results indicate it is much less likely for substitution to take place away from cigarettes and e-cigarettes towards cigars and tobacco.

For cigar users, the marginal rate of substitution towards cigarettes is 5.68 meaning that in terms of value, an additional dollar spent on cigars is equivalent to spending $5.68 on e-cigarettes. Similarly, from the same category, an additional dollar spent on e-cigarettes is equivalent to spending $7.9 on e-cigarettes only. These results indicate it is highly likely that consumers of cigars substitute towards cigarettes and e-cigarettes. Moreover, an additional dollar spent on cigars is equivalent to spending $0.25 on cigarettes and e-cigarettes both and -$0.12 on tobacco. These results indicate it is much less likely for substitution to take place away from cigars towards both cigarettes and e-cigarettes towards or tobacco.

Lastly, for tobacco users, the marginal rate of substitution towards cigarettes is 0.34 meaning that in terms of value, an additional dollar spent on tobacco is equivalent to spending $0.34 on e-cigarettes. Moreover, an additional dollar spent on tobacco is equivalent to spending $0.32 on cigarettes and e-cigarettes both and -$0.71 on tobacco. These results indicate it is much less likely for substitution to take place away from cigars towards both cigarettes, both cigarettes and e-cigarettes towards or cigars. On the other hand, from the same category, an additional dollar spent on e-cigarettes is equivalent to spending $4.72 on e-cigarettes only. This result indicate it is highly likely that consumers of tobacco substitute towards e-cigarettes.

To summarise, the substitution towards e-cigarettes is pretty high. This is true not just for cigarettes but also from other tobacco products. Simultaneously, substitution

towards cigarettes is not as high. Moreover, individuals using both cigarettes and e-cigarettes are nearly indifferent to using cigarettes. Based on these results, it can be inferred that substitution towards cigarettes may not be the issue. This result on the "gateway effect" is inline with that found in previous literature. However, if e-cigarette use is considered to be a problem, efforts need to be made to curb it. An unanticipated result of this analysis has been realisation of the high use of cigars among the population as well as the high degree of substitution from e-cigarettes to cigars and vice versa.

5.2 Nested Logit

Unlike the multinomial logit model, the nested logit captures the fact that the different tobacco products are correlated and they share unobserved attributes that are common to the fact that they are all tobacco products and so are distinct from "not smoking". By the IIA assumption, the probability of choosing between any of the tobacco categories is constant under the multinomial logit model. The nested structure allows for greater flexibility as the probability of not smoking is no longer equal to smoking tobacco products as it was in the multinomial logit model.

Interpreting the results of the model, the upper model reflects the marginal probability (choice of nest) whereas the lower model reflects conditional probability (choice of alternative within nest). The p-value for the likelihood ratio chi-square test is zero indicating that the model with full set of predictors represents a significant improvement in fit as compared to the null model. Hence, we can infer that at least one of the regression slopes is non-zero. The pseudo R squared value is 0.123 indicating that relative to the null model, our model gives an improvement in fit by 12.3%.

Starting at the first nest which contains the decision to smoke or not, given by the parameter "use" in the table above. The coefficient cost is the price for each of the products in the market. Based on the positive coefficient, it can be that an increase in income makes in more likely for an individual to smoke compared to the base parameter of not smoking. It can be interpreted as such: a 1% increase in price of smoking products leads to 0.84 units increase in the log-odds that a individual will take up smoking. This result is significant at the 95% confidence interval. This makes sense as a non-smoker has even less of an incentive to smoke in the face of rising prices. Next is the coefficient on tax. This constitutes of tax on cigarettes, General Sales Tax (GST) and Provincial Sales Tax (PST). The positive coefficient of 1.4 indicates an increase in tax makes it more likely that an individual will take up smoking. The coefficient implies that a 1% increase in taxes would lead to an increase in the log-odds of an individual smoking by 1.4 units. This result is significant at the 90% confidence interval however it seems counter intuitive, it makes sense if we consider that provinces would increase taxes in response to increased demand.

	(1)	(2)	(3)
VARIABLES	use	smoke	/typesmoke
ln Income		0.330***	
		(0.119)	
Urban		0.0834***	
		(0.0228)	
Age		-0.00714***	
		(0.000858)	
Div.Marital		4.150***	
		(0.132)	
Sin.Marital		3.600	
		(0)	
Marr.Marital#Age		0	
		(0)	
Div.Marital#Age		-0.0270***	
		(0.00147)	
Sin.Marital#Age		0.0258***	
		(0.00147)	
Sex		-0.624***	
		(0.0485)	
Marital Status		-0.290	
		(0.0271)	
Education		-0.0184***	
		(0.001777)	
Immigrants		-0.0255	
		(0.130)	
Employment		0.00659*	
		(0.00331)	
Price	-0.841**		
	(0.371)		
Tax	1.393*		
	(0.816)		
no_tau			1
			(0)
smoke_tau			2.924
			(2.009)

Standard errors in parentheses

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

Table 11: Nested Logit Results

Next let's look at the nest of alternatives when someone decides to smoke, i.e the choice between cigarettes, e-cigarettes, both the former, cigars or other combinations. The coefficient on lnincome is 0.33 implies an increase in income makes it more likely that that a 1% increase in income leads to a 0.0033 units increase in the log-odds of an individual taking up smoking relative to the baseline parameter of not smoking, this result is significant at the 99% confidence interval. Greater income simply reflects to having more disposable income to spend on tobacco products which probably leads to greater consumption. The coefficient on urban implies as an individual moves from urban to rural locality, it is more likely that they will smoke relative to the baseline category of not-smoking. The log odds of them taking up smoking increases by 0.08 units. This result is significant at the 99% confidence interval and may point to increased stressors in rural life that make people take up smoking. An increase in age makes it more likely for an individual to smoke relative to the baseline category of not smoking. A 1 year increase in age leads to 0.008 units decrease in the log odds of a non-smoker smoking, indicating a greater use of tobacco products among the youth.Train (2009) As compared to married individuals, divorced/widowed individuals are more likely to take up smoking. The log odds of divorced/widowed individuals taking up smoking are 4.15 units greater than married individuals. This result is significant at the 99% confidence interval. Moreover, looking at the interactions between marital status and age: widowed/divorced individuals are more likely to smoke as compared married individuals relative to the baseline category of not smoking. A 1 year increase in age of widowed/divorced individuals as compared to married individuals leads to a decrease of 0.02 units in the log odds of smoking. Whereas, for single individuals relative to married individuals, the log odds of smoking increases by the same amount. Both these results are significant at the 99% confidence interval. Between genders,females are less likely to smoke. As we move from males to females, the log-odds of smoking decreases by 0.6 units. This result is significant at the 99% confidence interval and likely points to the relatively less inclination of women to smoke (in general). Similarly, increase in education indicates the individual is less likely to smoke. A one year increase in education leads to the log odds of smoking to decrease by 0.02 units. This result is significant at the 99% confidence interval and points to increased awareness regarding the adverse effects of using tobacco products with greater education.

Finally, employed individuals are more likely to smoke as compared to unemployed individuals. The log-odds of smoking increases by 0.006 units as individuals move from unemployed to employed. This result is significant at the 99% confidence interval and may point to increase stressors of employment or even a smoking culture among employed individuals that could lead to increased take-up of smoking.

	Nest 1 Probability	Nest 2 Probability	Conditional Probability
Choice			
No	0.568	0.568	1
Cigarettes	0.432	0.065	0.150
E-Cigarettes	0.432	0.097	0.224
Cig and E-cigs	0.432	0.053	0.125
Cigar	0.432	0.150	0.350
Tobacco+	0.432	0.065	0.150

Table : Probabilities and conditional probabilities within nests

Next, we can predict the probabilities within the first and second nests for each of the observations in our dataset. The table above shows average probabilities within the two nests as well as the conditional probability for the second nest. Due to the IIA assumption holding within but not between nests, we can see that the probability of an individual not smoking is 0.57 whereas the probability of an individual smoking is 0.48. This assumes equal substitution between the five different categories within the smoking branch. Nest 2 probability then shows probability of falling in each of the categories of 'choice' and simply corresponds to the percentages of the sample falling into each of the buckets. Lastly, let's look at conditional probability. Conditional probability for not smoking is one because this is a degenerate branch. Conditional on an individual smoking, the probability of the individual smoking cigarettes primarily is 0.15; conditional on an individual smoking, the probability of the individual smoking e-cigarettes primarily is 0.22; conditional on an individual smoking, the probability of the individual smoking both cigarettes and e-cigarettes is 0.125; Conditional on an individual smoking, the probability of the individual smoking cigars primarily is 0.35; and conditional on an individual smoking, the probability of the individual smoking tobacco and/or other combinations of products is 0.15.

The coefficient for log sum constant for this regression is 0.109 indicating a 0.891 correlation between the first and the second nests. The log sum coefficient implies that substitution within nests is more than the substitution across nests which makes sense for the analysis underway. We can use the t-test for this coefficient to test the hypothesis that this model could have been estimated using a simple logit model i.e. the null hypothesis being the log sum or correlation coefficient is equal to zero. The p-value for this test is 0 indicating that we can reject the null hypothesis at 99% confidence level. Hence, the model is better suited to the current nested structure rather than a simple logit model.

The results of this estimation also presented the LR test for IIA = 1.84, with a p-value of 0.1747. Indicating that we cannot reject the null hypothesis that IIA holds between nests. This points to the nested logit model being the appropriate for this

case.

Conducting the Likelihood ratio test on our two nested set of predictors, we obtain a p-value of 0 for both the nests indicating that the predictors add significantly to our null model to explain the dependant variable. The hypothesis that the second nest does not add significantly to the first is rejected at 99% confidence. The LR is 34764.59 for nest one and 130.28 for nest 2 respectively.

5.2.1 Substitution

Using the coefficients from the second nest, the marginal rates of substitution between products are as follows:

For cigarette users, the marginal rate of substitution towards e-cigarettes is 0.66 meaning that in probability of use, an additional unit of cigarettes is equivalent to 0.67 of e-cigarettes. However, the MRS for substitution towards both products is 1.2, indicating a high substitution possibility. Moreover, an additional unit of cigarettes is equivalent to 0.43 units of cigars and 1 of tobacco. These results indicate it is much less likely for substitution to take place away from cigarettes towards cigars. Between tobacco and cigarettes, consumers are indifferent.

For e-cigarette users, the marginal rate of substitution towards cigarettes is 1.49 meaning that in terms of probability of use, an additional unit of cigarettes is equivalent to 1.49 of cigarettes. This points to e-cigarettes being a possible "gateway" drug for cigarettes. Similarly, an additional unit of e-cigarettes is equivalent to 1.79 of both cigarettes and e-cigarettes. This corroborates the previous result as consumers may now use both. Moreover, an additional unit of e-cigarettes is equivalent to 0.64 units of cigars and 1.49 of tobacco. These results indicate it is much less likely for substitution to take place away from e-cigarettes towards cigars but highly likely for substitution to take place towards tobacco.

To summarise, these results show possible substitution from cigarettes to both products. Simultaneously, e-cigarette users have a high MRS towards both products as well. While these results partly corroborate results of multinomial logit, indication of substitution towards cigarettes from e-cigarettes is different from the previous model. However, since I have a better understanding of the multinomial logit results, those have been used in the conclusion.

6 Limitations

The first limitation to this study has been that of data. Working with freely available datasets available through the university network may mean I have missed out on quality variables in my analysis. The unavailability of price of e-cigarettes meant I spent a significant portion of the time allotted for this project on finding prices.

The identification issue mentioned in Section 4.2 on the possible correlation between prices and the error term can be taken care of using instrumental variables (IVs) for price. Possible IVs include the acreage of land in province allocated to tobacco farming. This fulfils the relevance and exclusion requirements since it only affect demand of tobacco products through prices and not directly. Unfortunately, I could not find data on this variable and so it is a limitation of my work. One, I hope to work on in the future.

Further, a number of assumptions have been made in order to make sense of the data as well as the through the estimation techniques. If any of these fall short, the results of the analysis will become questionable.

7 Conclusion

The question of substitution between cigarettes and e-cigarettes is a complex topic worthy of attention. Through this analysis, I have tried to empirically find the substitution patterns between tobacco products using random coefficients discrete choice models. While it is still difficult to establish causation owing to possible unaccounted factors in this relationship, the results show likely high substitution from all tobacco products to e-cigarettes. Further, substitution towards cigarettes is not as high. In addition, individuals using both cigarettes and e-cigarettes are nearly indifferent to using cigarettes alone. According to these results, it can be said that substitution towards cigarettes may not be the issue. However, if e-cigarette use is considered to hazardous to health, efforts need to be made to control its use.

References

Caponnetto, P., Campagna, D., Cibella, F., Morjaria, J. B., Caruso, M., Russo, C., and Polosa, R. (2013). Efficiency and safety of an electronic cigarette (eclat) as tobacco cigarettes substitute: a prospective 12-month randomized control design study. *PloS one*, 8(6):e66317.

Chapman, S., Bareham, D., and Maziak, W. (2019). The gateway effect of e-cigarettes: reflections on main criticisms. *Nicotine and Tobacco Research*, 21(5):695–698.

Dautzenberg, B. and Bricard, D. (2015). Real-time characterization of e-cigarettes use: the 1 million puffs study. *J. Addict. Res. Ther*, 6(229.10):4172.

Drew, D. (2021). An analysis of the 2022 consumer price index basket update, based on 2021 expenditures. https://www150.statcan.gc.ca/n1/pub/62f0014m/62f0014m2022009-eng.htm.

Dubois, P., Griffith, R., and O'Connell, M. (2018). The effects of banning advertising in junk food markets. *The Review of Economic Studies*, 85(1):396–436.

Etter, J.-F. (2018). Gateway effects and electronic cigarettes. *Addiction*, 113(10):1776–1783.

Ghosh, S. and Drummond, M. B. (2017). Electronic cigarettes as smoking cessation tool: are we there? *Current opinion in pulmonary medicine*, 23(2):111.

Jacobi, L. and Sovinsky, M. (2016). Marijuana on main street? estimating demand in markets with limited access. *American Economic Review*, 106(8):2009–45.

Schachter, S. (1978). Pharmacological and psychological determinants of smoking. *Annals of Internal Medicine*, 88(1):104–114.

Shapiro, B. T. (2018). Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants. *Journal of political economy*, 126(1):381–437.

Snider, S. E., Cummings, K. M., and Bickel, W. K. (2017). Behavioral economic substitution between conventional cigarettes and e-cigarettes differs as a function of the frequency of e-cigarette use. *Drug and alcohol dependence*, 177:14–22.

Statistica (2021a). Average cigarette prices in canada from 2006 to 2014, by region. https://www-statista-com.proxy3.library.mcgill.ca/statistics/449082/average-cigarette-prices-by-region-canada/.

Statistica (2021b). Cigarette prices in canada by province. https://www-statista-com.proxy3.library.mcgill.ca/outlook/cmo/tobacco-products/canada.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Tuchman, A. E. (2019). Advertising and demand for addictive goods: The effects of e-cigarette advertising. *Marketing science*, 38(6):994–1022.

Van Loon, A. J. M., Tijhuis, M., Surtees, P. G., and Ormel, J. (2005). Determinants of smoking status: cross-sectional data on smoking initiation and cessation. *The European Journal of Public Health*, 15(3):256–261.

Vardavas, C. I., Filippidis, F. T., and Agaku, I. T. (2015). Determinants and prevalence of e-cigarette use throughout the european union: a secondary analysis of 26 566 youth and adults from 27 countries. *Tobacco control*, 24(5):442–448.

Worrell, M. and Hagen, L. (2021). At-a-glance-cigarette affordability in canadian provinces: a 10-year review. *Health Promotion and Chronic Disease Prevention in Canada: Research, Policy and Practice*, 41(10):315.

Yao, T., Sung, H.-Y., Huang, J., Chu, L., Helen, G. S., and Max, W. (2020). The impact of e-cigarette and cigarette prices on e-cigarette and cigarette sales in california. *Preventive Medicine Reports*, 20:101244.

8 Appendix

8.1 Ordinary Least Squares Regression

The OLS mimics a demand specification for both cigarettes and e-cigarettes. The Ordinary Least Squares (OLS) regression is used to estimate parameters of a linear model for a continuous dependent variable such as our quantity variables for cigarettes and e-cigarettes. While this is a simple yet powerful estimator, it falls short in capturing more of the variation in the dependent variable resulting from the independent variables.

Table A above shows demand equations for cigarettes and e-cigarettes respectively. The Rsquared for the first regression is 0.1537 implying the regressors capture about 15% of the variation in quantity of cigarettes demanded. For the second regression, the Rsquared is 0.0444 implying the regressors capture about 4.4% of the variation in quantity of cigarettes demanded. The p value for the full cigarette model is 0, and for the full e-cigarette model, it is 0.086 implying that for both models one can reject the null hypothesis.

For each individual living in rural relative to urban area, quantity demanded of cigarettes per day goes up by 1.4 cigarettes. This result is statistically significant at the 99% confidence level. Similarly, for a 1-year increase in age, quantity of cigarettes demanded per day increases by 0.136. This result is statistically significant at the 99% confidence level. For divorced or widowed individuals, relative married individuals, quantity or number of cigarettes demanded each day increased by 1.458 cigarettes. This result is statistically significant at the 99% confidence level. Results are insignificant for single relative to married individuals. For a one unit decrease in health, quantity or number of cigarettes demanded each day increased by 1.30 cigarettes. This result is statistically significant at the 99% confidence level. All other variables have insignificant impact on demand of cigarettes. For e-cigarettes, only the price of e-cigarettes is significant at the 99% confidence level.

In order to gauge additional statistical inferences from my OLS models, I carried out the one-sided hypothesis test. The alternate hypothesis being that coefficient on income is significantly positive for the e-cigarette demand equation while the null being that the coefficient is infact less than or equal to zero. Since the t-stat for this variable is 1.35, it is greater then the corrected critical value of 1.28 for this test. Therefore, we can reject the null hypothesis at the 90% significance level and say that despite the coefficient being insignificant, the income coefficient is significantly positive determinant of e-cigarette demand.

	(1)	(2)
VARIABLES	quantity_cig	quantity_ecig
Urban	1.416***	0.457
	(0.249)	(0.386)
Age	0.136***	0.00242
	(0.00770)	(0.0119)
Aex	-0.249	0.355
	(1.019)	(1.613)
Div.Marital	1.458***	0.216
	(0.353)	(0.770)
Single.Marital	-0.0886	-0.845*
	(0.273)	(0.476)
Physical Health	1.308***	-0.143
	(0.129)	(0.217)
Mental Health	0.191	0.166
	(0.121)	(0.194)
Education	-0.101	-0.0291
	(0.0620)	(0.0977)
Immigrants	-1.422	-6.665
	(3.909)	(5.443)
Employment	-0.0866	-0.107
	(0.0829)	(0.138)
Income	0.000105	0.000127
	(6.43e-05)	(9.38e-05)
Price cigarette	15.17	36.82
	(13.52)	(23.06)
Price e-cigarette	0.865	-8.841***
	(1.663)	(2.663)
Price cigars	9.627	23.52
	(17.17)	(25.46)
Price tobacco	-85.45	-115.7
	(68.57)	(97.59)
Tax cigarette	-10.77	7.854
	(13.24)	(19.33)
GST	0.470	0.678
	(0.441)	(0.642)
PST	0.442	0.458
	(0.400)	(0.594)
R-squared	0.154	0.044

8.2 Advertising Information

Advertising information was collected through AdSpenser which itself gathers data from the Kantar World Panel- a research organisation that collects consumer data through surveys and then analyses it on behalf of its clients into marketable insights. This is a popular source of advertising data as it has been used in numerous works including Dubois et al. (2018) Junk which has been the inspiration behind this paper. However, the aforementioned work had access to data that is a lot more detailed including information on an individual households exposure to a given brand's commercial. Personally, I have been able to access information on a particular brands expenditure on advertising in Canada for a given year. The information is disaggregated into categories such as point-of-sales advertising, magazine advertising and even online advertising for both regular and e-cigarette brands. In most cases however, it is the total advertising amount that is consistently available and so will be the main data used as part of this analysis. It must be noted, however, that most forms of advertising are banned in Canada. As of the 2020 Tobacco and Vaping Products Act, ad placement in banned where 'the youth can see them'. While the advertising data being used precedes this particular ban, it is likely that dissagregated information is not shared by companies because it was not per say 'allowed'.

Some of the earlier works on advertising uses total advertising expenditure as part of the demand model. For instance, Bishop and Yoo (1985) use total cigarette advertising expenditure in a given year as part of their demand equation. Similarly, Batalgi and Levin (1986) use a dummy variable which takes the value of the per capita index of advertising for the years advertising was not banned and zero otherwise.

For the purpose of this analysis, I first created separate variables for per capita advertising expenditure in each category (cigarettes, e-cigarettes, cigars and tobacco) based on whether a person had ever used that particular product. This assumes that the person had been persuaded by advertising. These variables are added up to form the single variable ad exposure. According to this variable, nearly half of the 46086 individuals in the sample are non-smokers and so have no exposure to advertising.

Table B summarises advertising expenditure on each of the categories:

	Survey Year		
Category	2013	2015	2017
E-Cigarettes	73971.1	31282.1	34770
Cigarettes	81457.2	37019.7	19647.9
Cigars	4639.9	638.1	147.5
Tobacco+[9]	30062.7	80093.4	65308.3
Anti-Smoking Products[10]	33939.2	37372.6	41384.3

Table B: Ad Expenditure (000) Canada

Back in 2013, the e-cigarette market had free-reign when in came to advertising its products. In fact the first regulations in this area were put in place in 2018. This can

be seen by the high expenditure. For 2013 and 2015, the regular cigarette advertising share has been greater than e-cigarette however the e-cigarette share takes over in 2017. In the data, it can be seen that some companies such as Marlboro have not provided any information. While this puts in question the validity of the data, each of the categories as well as the years still look comparable (or proportionate) to each other so I shall continue to use this data. Further, the data also reveals that a small number of large firms produce regular cigarettes while a larger number of smaller firms are producing e-cigarettes indicating that it is still a burgeoning market. Expenditure on anti-smoking products is roughly consistent. Whereas, expenditure on cigars had greatly declined while that on tobacco increased greatly in 2015. This is probably in line with the introduction of new snus products Canada.