Heterogeneous preferences for nutritional and health attributes of Frozen Fried Potatoes: A developing country perspective

Julia González [1] *and M. Victoria Lacaze* [2]

[1]*University of Illinois, USA* [2]*Universidad Nacional de Mar del Plata, Argentina*

This article intends to explore consumers' heterogeneous preferences for health and nutritional attributes of frozen fried potatoes (FFP) by analyzing an important urban area of a Latin American country, Argentina, where FFP remain to be the most preferred side dish when people choose to eat fast food, among the available options of a growing market. Results shed light on what drive differences in consumers' valuation and how demographic characteristics affect price sensitivity and substitution patterns, which are the main magnitudes considered when introducing any market regulation or public policy. By accounting for the heterogeneity-in-preferences estimation also provides insightful information for suppliers for optimally launch price differentiation strategies.

IAAE

1. Introduction

Frozen fried potatoes (FFP) is an extensively consumed food in developed countries. Although the market has reached maturity in the United States and Western Europe, a swiftly growth in developing countries is related to pronounced modifications in working patterns and social habits that push an increasing frequency of meals away from home. Particularly in Argentina, where potato is the most consumed vegetable by all income levels, the processed potato industry is rapidly developing and, as a consequence, FFP are starting to play a more relevant role in consumers' diet. The local processed potato industry is characterized by high concentration, and also a high degree of horizontal and vertical differentiation is verified. Straight-cut fries -named *"bastón"* in some Latin American countries- are the main product of this industry, even though slices, noisettes, and croquettes are also offered.

Although restricted because of the high prices when compared with fresh potatoes, Argentines households' direct demand for FFP is primarily supplied by super and hypermarkets where McCain and Alimentos Modernos are the dominating manufacturing companies. The same multinational firms rule a maturing fast food market: while McCain supplies McDonald's, Alimentos Modernos provides Burger King through two own brands. The consumption boom not only encouraged other firms to enter the market, as Wendy's, Subway, and Kentucky Fried Chicken recently did, but also created profitable opportunities for local brands, such as the pioneer Mostaza, Nac & Pop, and Betos. The market expansion includes some more-healthy options, e.g. rolls, soups and cakes that are advertised as 'reduced calories', 'diabetes friendly','lower in sodium', or 'gluten-free' choices. But FFP keeps its place as the most preferred side dish when people choose to eat fast food.

There are virtually no research on the characteristics, evolution, and development of the domestic market of FFP in Argentina. Few exceptions are past studies committed to analyze contractual relationships and integration schemes between potato producers and agro-industry stakeholders, but there are not investigations concerned with understanding and identifying consumers' preferences for these products. Uncovering how households'

characteristics have an impact on FFP choices is a crucial input for decision making of the supply chain stakeholders.

The purpose of this article is threefold: to explore consumers' heterogeneous preferences for health and nutritional attributes of FFP in the Argentine domestic market; to identify different patterns of consumption, and different effects of changes in market structure, based on the individuals' demographic characteristics; and to compare the impact of the estimation of heterogeneous vs. assumed homogeneous demographic characteristics on suppliers' decisions. In order to achieve the proposed goals, the study uses an estimation of a discrete choice model of households' direct demand which results conclude that both income and age affect consumers' appraisal of FFP attributes (González & Lacaze, 2012).

2. Theoretical framework

2.1. Product differentiation and discrete-choice logit models

Product differentiation as a research topic of agricultural economics dates back to the decade of the 1920s, when Waugh (1928) published his seminal work devoted to analyze the relationship between price and characteristics of vegetables in the United States. Later, Houthakker (1951-52) incorporated the product characteristics in their utility maximization models, while Lancaster (1966) postulated that it is the properties or characteristics of the good from which utility is derived. The Simple Logit Model (henceforth, SLM) makes use of this conceptual framework and solves some challenges that arise when estimating demand functions for differentiated products. Specifically, it overcomes the dimensionality problem by projecting the products onto a characteristics space (McFadden, 4. However, all individuals are assumed to be identical except for the error term, which entails strong restrictions on elasticity coefficients. On the one hand, the own-price elasticities are almost perfectly proportional to prices when the market share of the outside good is close to one (McFadden, 1981). On the other hand, the cross-sensitivity of demand is the same regardless the good whose price changes, and therefore consumers substitute towards other products in proportion to market shares. More flexible substitution patterns are achieved with the Nested Logit Model, whose estimation requires a priori clustering of products. Although the cross-price elasticity coefficients are different between groups, they remain equal within them. Finally, the Random Coefficients Discrete Choice Model (henceforth,

RCDCM) allows for heterogeneous own- and cross-price elasticity patterns as explained below.

2.2.   *Estimating a demand system in differentiated-product markets*

The estimation of demand has been a key part of Industrial Organization applied studies that examine the functioning of differentiated-products markets. While Lancaster (1971) and McFadden (1974) seminal work initiated the theoretical and econometric groundwork for characteristic-based demand systems, applications increased significantly after the literature succeeded to circumvent the restriction on elasticities calculation and to explicitly account for the unobservable characteristics. In this sense, Pakes (1986) introduced simulation estimators to enable the researcher to use a micro behavioral model with heterogeneous agents to structure the empirical analysis of aggregate data. Therefore, from the observed distribution of consumer characteristics and an appropriate functional form, an aggregation process can be made. Then, a demand system conditional on consumers' characteristics and a vector of parameters which determines the relationship between those characteristics and preferences over products (or over product characteristics) can be formulated and estimated from market level data.

One step further, Berry (1994) and Berry, Levinsohn, and Pakes (henceforth BLP, 1995) developed a method for estimating RCDCM of demand, allowing for flexible own-price elasticities driven by the different price sensitivity of diverse consumers who purchase various products, but not by functional-form assumptions about how price enters the indirect utility function as the SLM does. The RCDCM also allows for cross-price substitution patterns driven by product characteristics, but not constrained by any arbitrary segmentation of the market as the Nested Logit Model does, yet at the same time taking advantage of these segmentation procedures.

To be estimated, the RCDCM requires market-level price and quantity data for each product in a series of markets. Socioeconomic information regarding the distribution of demographics might be available, but a key assumption is that individual decisions are not observed. A product is defined by a set of characteristics, being some of them observable by the researcher and others are unobservable but also influence demand and are explicitly introduced in the model. BLP (1995) use a contraction mapping which transforms the

demand system into a system of equations that is linear in these unobservable factors. To the extent that producers know those unknown characteristics when set the prices, they are correlated with the disturbance term and instruments can be used to overcome the arisen endogeneity problem once the system becomes linear in the error term.

Nevo (2001), whose approach is applied in González & Lacaze (2012), follows BLP's algorithm but extends their work in some prominent ways. First, he assumes observed product characteristics as chosen by firms that account for consumer preferences. Since they are not treated as exogenous or at least predetermined, the assumption to identify demand parameters differs with BLP's. Second, he identifies the demand side without needing to rely on the functional form of the supply side. Third, he adds product-specific dummy variables as product characteristics, capturing those that do not vary by market. Therefore, the correlation between prices and the part of the unobserved quality that is market-invariant is fully accounted for and does not require instruments. Forth, he models heterogeneity as a function of the nonparametric distribution of demographics, thereby partially relaxing the parametric assumptions previously used.

2.3. RCDCM specification

2.3.1. Demand

Suppose there are t = 1,…, T markets, each with i = 1,…, I consumers and for each market aggregate quantities, average prices, and product characteristics for J (j = 1,…, J) products are observed. The conditional indirect utility of consumer i from product j at market t is

(1) $$U_i(x_{jt}, \xi_{jt}, p_{jt}; \theta)$$

where, for product j in market t, x_{jt} are observed product characteristics, ξ_{jt} are the unobserved product characteristics that producers not only observe but also take into account when setting the prices, and p_{jt} is the price, while τ_i are individual characteristics and θ are unknown parameters. In the particular specification of demand, utility function is expressed as

(2) $$u_{ijt} = x_j \beta_i^* - \alpha_i^* p_{jt} + \xi_j + \Delta\xi_{jt} + \varepsilon_{ijt}$$

where $(\alpha_i^* \beta_i^*)$ are K+1 individual-specific coefficients, x_j is a K-dimensional (row) vector of observable product characteristics, $\Delta\xi_{jt}$ is a market specific deviation from the mean

valuation of the unobserved characteristics, ξ_j, and ε_{ijt} is a mean-zero stochastic term. This indirect utility can be derived from a quasi-linear utility function, which is free of wealth effects, a reasonable assumption for food products like FFP.[1] Consequently this alters the way that price and income enter equation (2).

The next component of the model describes how consumer preferences vary as a function of the individual characteristics. They consist of demographics that are observed, D_i, and additional characteristics that are unobserved, v_i. The distribution of consumer taste parameters is modeled as a multi-variate normal conditional on demographics. Let $\gamma_i^* = (\alpha_i^*, \beta_{il}^*, \dots, \beta_{iK}^*)$ and $\gamma = (\alpha, \beta_l, \dots, \beta_K)$ where K is the dimension of the observed characteristics vector, therefore

(3)
$$\gamma_i^* = \gamma + \Pi D_i + \Sigma v_i, \quad v_i \sim N(0, I_{K+1})$$

where Π is a $(K + 1) \times d$ matrix of coefficients that measures how the taste coefficients vary with demographics, and Σ is a scaling matrix. The specification of equation (3) implicitly makes assumptions about both functional form and distributions.

The specification of the demand system is completed with the introduction of an outside good, because consumers may decide not to purchase any of the available products. Without this allowance, a homogeneous price increase of all products does not change quantities purchased. The indirect utility from this outside option is

$$u_{i0t} = \xi_0 + \pi_0 D_i + \sigma_0 v_{i0} + \varepsilon_{i0t}$$

The mean utility of the outside good, ξ_0, is not identified, so it is normalized to zero.

Let $\theta = (\theta_1, \theta_2)$ be the vector containing all parameters of the model. While the vector $\theta_1 = (\alpha, \beta)$ contains the linear parameters, $\theta_2 = (\Pi, \Sigma)$ comprehends the nonlinear parameters. Combining equations (2) and (3):

(4)
$$u_{ijt} = \delta_{jt}\left(x_j, p_{jt}, \xi_j, \Delta\xi_{jt}; \theta_1\right) + \mu_{ijt}\left(x_j, p_{jt}, v_i, D_i; \theta_2\right) + \varepsilon_{ijt}$$
$$\delta_{jt} = x_j\beta - \alpha p_{jt} + \xi_j + \Delta\xi_{jt}; \ \mu_{ijt} = \left[p_{jt}, x_{jt}\right]' * (\Pi D_i + \Sigma v_i)$$

Where $\left[p_{jt}, x_{jt}\right]$ is a $(K + 1) \times 1$ vector, δ_{jt} represents the mean utility, and $\mu_{ijt} + \varepsilon_{ijt}$ is a mean-zero heteroskedastic deviation from that mean that captures the effects of the random coefficients. The estimation exploits this separation to reduce the number of parameters that enter in a non-linear fashion and to generate linear moment conditions.

---

[1] For other goods this is an unreasonable assumption, as in BLP (1995), who must include wealth effects.

It is assumed that consumers purchase one unit of the good that gives the highest utility, which is a reasonable assumption since most people consume only one kind of FFP at a time. This implicitly defines the set of individual-specific variables that lead to the choice of good j:

$$A_{jt}(x, p_{.t}, \delta_{.t}; \theta_2) = \{(D_i, v_i, \varepsilon_{it}) \mid u_{ijt} \geq u_{ilt} \ \forall l = 0, 1, \dots, J\}$$

Assuming ties occur with zero probability, the market share of the j-th product as a function of the mean utility levels of all the $J + 1$ goods, given the parameters, is

(5) $$s_{jt}(x, p_{.t}, \delta_{.t}; \theta_2) = \int_{A_{jt}} dP^*(D, v, \varepsilon) = \int_{A_{jt}} dP_D^*(D) dP_v^*(v) dP_\varepsilon^*(\varepsilon)$$

where $P^*(\cdot)$ denotes population distribution functions. A straightforward estimation strategy is to choose parameters that minimize the distance between the market shares predicted by equation (5) and the observed shares. Possibly the simplest assumption that can be made in order to solve the integral in equation (5) is that consumer heterogeneity enters the model only through the separable additive random shock, ε_{ijt}. This implies $\theta_2 = 0$ and therefore, $\beta_{ij}^* = \beta_j$, $\alpha_l^* = \alpha$ for all i, and equation (2) becomes

(6) $$u_{ijt} = x_j \beta - \alpha p_{jt} + \xi_j + \Delta \xi_{jt} + \varepsilon_{ijt} \quad i = 1, \dots, I_t \ \ j = 1, \dots, J \ \ t = 1, \dots, T$$

If ε_{ijt} is distributed i.i.d. with a Type I extreme value distribution, the Multinomial Logit Model arises. The market shares relative to the total market, including the outside good, are

(7) $$s_{jt} = \frac{\exp(x_{jt} \beta - \alpha p_{jt} + \xi_j + \Delta \xi_{jt})}{1 + \sum_{k=1}^{J} \exp(x_k \beta - \alpha p_{kt} + \xi_k + \Delta \xi_{kt})}$$

Although this model and the extreme value distribution assumption are appealing due to its tractability, substitution patterns are restricted to depend only on the market shares. Since in most cases the market shares are small, the term $\alpha(1 - s_{jt})$ is nearly constant and then the own-price elasticities are proportional to own price. Therefore, a standard pricing model predicts a higher markup for the lower-priced brands. Regarding the cross-price elasticities, if two products have similar market shares, even if they possess different characteristics, then the substitution pattern from a third product whose price changes toward either of them will be the same.

In the Nested Logit Model, which implies a closed form expression for the integral in equation (5), the i.i.d. extreme value assumption is replaced by a variance components structure. All products are grouped into exhaustive and mutually exclusive sets. If the price

of a product changes, consumers are more likely to substitute to other products within the group. Consequently, the cross-price elasticities are different between groups but remain equal within the groups. But in some cases, the division of segments is not clear or does not fully account for the substitution patterns. Besides, the segmentation of the market would be multi-layered in some industries, implying that the order of the nests matters.

The RCDCM assumes the i.i.d. extreme value distribution assumption but own-price elasticities are not necessary driven by the functional form because they depend on prices and available demographic information. The partial derivative of the market shares will no longer be determined by a single parameter α. Instead, each individual will have a different price sensitivity, which will be averaged to a mean price sensitivity using the individual specific probabilities of purchase as weights. The composite random shock is no longer independent of the product and individual characteristics. Thus, if the price of a product rises, consumers are more likely to purchase products with similar characteristics, rather than the product with the biggest market share. Also, households with similar characteristics will tend to have similar purchasing patterns.

The price elasticities of the market shares implied by equation (7) are

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}}\frac{p_{kt}}{s_{jt}} = \begin{cases} \alpha p_{jt}(1 - s_{jt}), & if\ j = k \\ -\alpha p_{kt}s_{kt}, & if\ j \neq k \end{cases}$$

For the Nested Logit Model, they are

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}}\frac{p_{kt}}{s_{jt}} = \begin{cases} -\alpha p_j\left[1/1 - \sigma\left(1 - s_{j|A}\right) + s_{j|A}(1 - s_A)\right], & if\ j = k \\ \alpha p_l s_{j|A} s_A[1/1 - \sigma - (1 - s_A)], & if\ j \neq k\ within\ a\ group \\ \alpha p_k s_k, & if\ j \neq k\ among\ groups \end{cases}$$

Finally, the price elasticities of the market shares defined by equation (5) for the RCDCM are

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}}\frac{p_{kt}}{s_{jt}} = \begin{cases} p_{kt}/s_{jt}\displaystyle\int \alpha_{it}s_{ijt}(1 - s_{ijt})dP_D^*(D)dP_v^*(v), & if\ j = k \\ -p_{kt}/s_{jt}\displaystyle\int \alpha_{it}s_{ijt}s_{ikt}\,dP_D^*(D)dP_v^*(v), & if\ j \neq k \end{cases}$$

Equation (5) has no an analytical closed form for RCDCM. Consequently, simulation techniques have to be applied to find a solution. Since the main data source includes aggregate sales data, heterogeneity can be modeled either by assuming a parametric distribution of $P^*(\cdot)$ or as a function of the empirical nonparametric distribution of demographics. Both approaches allow obtaining flexible product substitution patterns, but

the latter also provides more information regarding how demographics affect consumers' heterogeneous preferences.

2.3.2. Consumer welfare

The measure we use to evaluate changes in consumer welfare is the compensating variation. In the SLM, all individuals are equal except for the error term. Consequently, the welfare change provoked by a counterfactual change is the same for every individual and expressed as the difference between consumer surplus with the starting prices and the consumer surplus with the new prices p^*:

$$(log \sum_{j \in H} e^{u_{j(p*)}} - log \sum_{j \in H} e^{u_{j(p)}})/\alpha$$

The compensating variation does not have an analytical solution for the RCDCM because α_i^* in equation (2) is a function of income. In this case, the individual compensating variation CV_i, has to be computed iteratively, and is equal to $-\Delta y_i$, where Δy_i solves

$$u_i(y_i, p) = u_i(y_i + \Delta y_i, p^*)$$

where y_i is the income of individual i and p is the vector of prices in the initial situation. The mean compensating variation in the population is given by

(8) $$CV = N \int CV_i \, dP_D^*(D) dP_v^*(v)$$

where N is the total number of consumers. The two needed assumptions for computing these changes in consumer surplus are 1) as with the observed characteristics, there is no change in the unobserved components ξ_{jt}; 2) there are no changes in the utility from the outside good.

3. Households demand estimation

3.1. Dataset construction

We used the discrete choice approach to analyze the demand for FFP. First, three different specifications of a SLM were estimated. The first one only included observed product characteristics, the second added product dummy variables, and a Two-Stages Least

Squares (TSLS) regression was finally estimated including instrumental variables and product fixed effects.[2]

Estimation results revealed that the effects of including product-specific dummy variables and of using instrumental variables are significant both in statistical and economical terms. But in order to overcome the restrictive and unrealistic substitution patterns yielded by SLG, a RCDCM of demand was estimated. The data required to consistently estimate this model consist of market shares and prices in each market, product attributes, and socioeconomic characteristics of individuals. In order to identify the variable part of the coefficients in the utility function in equation (2), and since information about individual purchases is not available; scanner data made available by a traditional supermarket chain in Mar del Plata was matched with a database which provided the distribution of demographics across population in each market.

Mar del Plata is the second largest city of Buenos Aires Province, the seventh of the country, and is the main urban center of the major potato production area of Argentina, located in the southeast Province of Buenos Aires.

The scanner database was provided by Supermercados Toledo S. A. and consists of the value of monthly sales and the quantity sold for each product and each of the 23 branches of the supermarket, from July 2005 to December 2009. The sales data cover 18 FFP products supplied by three firms (McCain, Alimentos Modernos, and Granja del Sol) through four brands (McCain, FarmFrites, Granja del Sol, and RapiPap), and were classified in six segments or varieties (*bastón*, golden longs, noisette, rondelles, smiles, and croquettes) offered in several container sizes. Nutritional information on calories, saturated fat, fiber, and sodium was collected by visual inspection of the products' nutrition facts labels. Unit value per serving was calculated as a proxy for price, by dividing the value of sales by the quantity of servings sold, which was computed as the package size divided by the serving size and multiplied by the quantity of units sold.

Information on the distribution of demographics was obtained by sampling individuals from the *Encuesta Permanente de Hogares* (EPH), which is carried out by the National Statistics Bureau *(Instituto Nacional de Estadísticas y Censos)* in the most important cities of the country. The socioeconomic variables of interest are per capita income and average

---

[2] The TSLS specification is the one referred to henceforth.

age of the household members, which is related with both household size and presence of children.

In order to match both data sets it is necessary to define the criterion for aggregating sales data and sampling simulated individuals, i.e. to define a market. Since the EPH does not provide the geographical location of surveyed households, it is not possible to delimit a market as a combination of a geographical area and a unit of time, which would be a branch-month combination. Therefore, a market was defined as an income-month combination following three steps. First, per capita average income of each Mar del Plata census tract was calculated using data from a previous household survey on potato consumption patterns. Second, the potential customers of each supermarket branch were identified according to the population of the census tract in which the branch is located. Finally, the branches were classified by the income level of their potential buyers (high, middle-high, middle, middle-low, and low), and sales data of branches with the same income level were aggregated by month and product. Thus, the data were structured in 270 markets (5 income levels and 54 months) and 2,145 observations (considering different products sold in each market). The demographic characterization of each market was accomplished by randomly drawing simulated individuals from the corresponding period and quintile of the EPH.

Lastly, to calculate the market shares it was necessary to assess the market size, i.e. the total potential demand for FFP of the supermarket chain. This was obtained as the 35%[3] of the total potential demand of the city, which in turn was calculated by imputing the FFP consumption frequency of "real consumers"[4] to the entire city population. This was done for each of the branches regarding their potential customers, and then the market size for each income-month combination was calculated. The market share for each product in each market was determined by dividing the quantity of servings sold by the market size.

---

[3] This is the Supermercados Toledo share of total supermarket sales in Mar del Plata, according to the opinion of key actors in the supermarket industry.

[4] This refers to the FFP consumption frequency of those polled in the potato consumption survey who declared they consume FFP.

3.2.   Analyzing the demand for FFP

A preliminary descriptive analysis shows that *bastón* is the most popular variety followed by noisette, despite its relatively high price. On the other hand, croquettes and rondelles are the segments with the least market shares. Customers can take advantage of economies of scale in these products, since price per serving decrease as container size increases, at equal value of the other characteristics. For all varieties and brands, consumers of high income-level face higher prices than consumers of low income-level for any product variety, which suggests the presence of a price discrimination strategy implemented by suppliers.

When estimating the RCDCM of demand, we found that, on average, consumers' utility increases as the FFP content of fiber and calories increase, and as the content of fat decreases. McCain products were revealed as the least valued FFP. If compared with the base group (golden longs, rondelles, and croquettes), the valuation of *bastón* is negative, and the valuation of noisette is positive. Both income and age reinforced the negative effect of price on utility. Households are more sensitive to the negative effect of fat and sodium the wealthier they are, and are also more sensitive to the positive effect of fiber.

The effect of random shocks to tastes on price and fat coefficients was not significant, suggesting that the heterogeneity in the coefficients is mostly explained by the included demographics. On the contrary, calories, fiber, sodium, and smiles presented statistically significant coefficients, implying that at least part of the parameter variability was captured by unobserved individual characteristics. This is especially interesting for sodium and smiles, since the average effect of these variables on utility was not statistically different from zero, but even so these results indicate there is heterogeneity in preferences for these attributes.

Based on the results from the RCDCM, all own-price elasticities were found negative and greater than one in absolute value. As for cross-price elasticities they were all positive as expected since the products are substitute goods.

Lastly, hypothetical changes in the FFP industry structure were simulated and their effect on prices, market shares, and consumer surplus was evaluated. On the one hand, a merger between the two smaller firms would enable an increase of firms' market power which explain the higher resulting prices and the decreasing consumers' welfare. This would cause an increase in the market share of the outside good, which is more pronounced for the

non-merged firm's FFPs. Moreover, the sales of all firms would decrease. This result is consistent with the relatively elastic demand for FFP. On the other hand, if the market turned into a single-product firm industry, the prices would drop and hence the consumer surplus would increase. Indeed, the reduction in prices encourages some households to enter the FFP industry and hence there would be an increase in the market shares of all products and sales. The lower prices in this scenario have to do with the lack of a portfolio effect.

Regarding the relationship between individual compensating variation and demographic variables, the influence of the counterfactual changes in market structure would be greater the higher the income and age. These results might be driven by the heterogeneity in price sensitivity.

4. Results and discussion

4.1. *Demographic variables distribution*

An expected lognormal distribution is verified when analyzing the empirical non-parametric distribution of income (Figure 1, panel a). When considering the distribution of income by income level it can be appreciated that higher groups present more value dispersion (Figure 1, panel b).

The empirical nonparametric distribution of household average age follows a bimodal distribution (Figure 1, panel c), i.e. there are two types of representative households in terms of this demographic, where the lower mode has more density than the upper one, especially in the case of the low-income level, which is consistent with lower-income level households composed by children and parents that are younger than those in higher-income households. When analyzing age by income level, the higher the income the less scattered the age distribution. Lower modes move right as income grows, and upper modes are more heterogeneous among income levels. Particularly, two nearly overlapping patterns can be shown: one between the distributions of age for middle-high and high income, and the other between the distributions for middle-low and middle income (Figure 1, panel d).

4.2.   Utility function coefficients distribution

An analysis of the distribution of the estimated individual parameters for the RCDCM and the comparison against the coefficients estimated from the SLM has been performed. According to the purposes of the RCDCM, it is expected that the distribution of each analyzed coefficients would be affected by the distribution of the demographics that are interacted with. Consequently, the coefficient distributions could probably replicate the shape of the demographics allocation. This effect is related with the magnitude of each coefficient, which can be compared since all variables were re-scaled to be of the same order of magnitude.

Regarding price coefficient distribution, which was supposed to depend on both household income and average age, it can be verified that age distribution has a more important effect than income distribution. The location of the SLM coefficient near zero verifies the underestimation of the price effect on consumer utility (Figure 2, panel a).When analyzing the coefficient of calories content, the interaction with income was found non-significant, but this nutritional variable is significantly affected by other demographic variables not explicitly accounted for, but included as random shocks to tastes. In the SLM, this parameter has no virtually impact on utility (Figure 2, panel b). As previously discussed, the marginal valuation of fat, fiber, and sodium is accentuated by increasing income, and for the last two parameters, also has a significant impact from unobserved characteristics. When consumers' heterogeneity is accounted for, the coefficients for fat and fiber changes their signs. Besides, the unobserved characteristics distribution has a deeper impact on them than income distribution does. Even though sodium has not a significant mean parameter (which happens in the SLM as well) the full model confirms the statistically significant presence of heterogeneity in sodium valuation. This variability in preferences may be due to the fact that healthier FFP (i.e., with lower sodium content) are usually at the same time less tasty. The major part of sodium-coefficient's variability is captured by unobserved individual characteristics and a less important portion is got by income. Finally, although the expected interaction between age and smiles (a kid-oriented segment) has no significant impact on utility at a 10% level,[5] its valuation is affected by heterogeneous characteristics

---

[5]This result could be due to not interaction of this variable with income as well.

and therefore its distribution is quite scattered. The mean parameter of this variable is non-significant for both SLM and the full model.

4.3.   Elasticities

The relationship among each demographic and the own- and cross-price elasticities of the inside goods -and also for the outside good- are analyzed both in general terms and also for each segment.[6] First, the correlation coefficients of each demographic and elasticity are calculated. Then, differences of elasticities in general terms and within each income level and age category are analyzed.

4.3.1.   Own-price elasticities

While the relationship between price sensitivity and demographics is inverse -i.e., the older and the wealthier the individual, the lesser the sensitivity- in the case of the own-price elasticity this relationship is more complex. Beginning at a low income level, if income grows the elasticity downs in absolute value, but as income continues rising elasticity becomes to increase, as results from the sign of the coefficients of income and squared income in the first column of Table 1. This behavior is explained because, as previously found, individuals with different income levels face different prices and consequently buy different quantities, resulting in price-quantity ratios that increase with income. Pricing strategies then reverse the behavioral pattern of sensitivity (i. e. utility parameters) and transpose the observed pattern in terms of percentage sensitivity (elasticity). The result is that, on average, low-income consumers have the higher own-price elasticity and middle-low income ones are the more inelastic, but then elasticity starts to increase until it reaches high-income, the second larger own-price elasticity (Table 2). Within each income level, an income-elasticity direct relationship is verified (Table 1 and Figure 3 -panel a and b-).
Own-price elasticity increases with age but at a decreasing rate, i.e., the relationship is attenuated for the oldest individuals (Table 1, column 1). As a result, if the average own-price elasticity by ranges of age is analyzed, low-age households present the less elastic

---

[6] A similar analysis has been performed in terms of brands but has not been included in the article since average brand elasticities are very influenced by the segments that each firm offers, which are not uniformly distributed.

demand and high-age consumers the more elastic ones (Table 3). This implies that although the increase is at a decreasing rate, it does not reach a maximum considering the range of age of the studied households. The same pattern is repeated when considering each age segment (Table 1 and Figure 3 -panels c and d-). Since age affects negatively the price sensitivity in the utility function, it has to be the case that the older the consumer the less the quantity of FFP purchased in order to obtain own-price elasticities that are increasing in absolute value with age. This result reinforces the expectations before performing RCDCM, i.e. that older consumers buy less FFP than younger ones probably because of health concerns. At the same time, it forces to get an explanation for the negative effect of age on price sensitivity in the utility function, different from younger households being more concerned about health issues because of the presence of children. Actually, what could be driving this inelasticity of older consumers is that they reasonably might have a more structured diet, and then changes in prices would not cause great product substitutions. Younger households could probably have a less structured diet; they are more likely to buy much more FFPs if prices go down but also much less if prices go up.

To visualize the relationship between own-price elasticity and demographic variables within income levels, first it can be seen that, within income levels, lower income households present lower average age (Figure 3), which is in line with the results in a previous subsection that show the same pattern but between income levels (Figure 1, panel d). This relationship however is non-significant within middle income level. Adding this result with the fact that own-price elasticity is higher as income and age rise, it explains the higher size of the bubbles in the chart as points are in an upper position and to the right side of the figure.

The analysis of own-price elasticity by product segment shows that croquettes and noisette are the more elastic varieties, followed by *bastón* and smiles; golden longs and rondelles are then the type of FFP whose demand is less reactive to changes in price (Table 4). However, middle-low and middle-high neighborhoods present some differences in these patterns, since croquettes is the more inelastic segment for that kind of households. On the other hand, rondelles is one of the more elastic products in middle-high neighborhoods, while *bastón* presents the higher elasticity within middle-low consumers (Figure 4, panel a). At the same time, although croquettes is the more elastic segment in general,

considering ranges of ages it only happens in older households, while lower-age consumers have a relatively less elastic demand for noisette and smiles (Figure 4, panel b).

As previously pointed out, SLM estimations yields an inaccurate coefficient of price sensitivity, and this makes own-price elasticities from the SLM much lower than those obtained from the full model. On the other hand, since in that model differences in elasticities between income levels are due only to differences in the ratio price-quantity, own-price elasticity is higher in absolute value for higher income neighborhoods (except for middle-low ones), because of the aforementioned pricing strategies of the firms. In the RCDCM this effect is counteracted because of the greater impact of price on utility of lower income consumers (Figure 5, panel a). Another difference between own-price elasticities from those models is the greater dispersion of coefficients between product segments in the case of SLM (Figure 5, panel b).

4.3.2. Cross-price elasticities

The same pattern that in the case of own-price elasticity occurs when analyzing the relationship among substitution patterns of FFPs regarding demographic variables. The sensitivity of consumers to change which product to purchase when faced to a change in prices turns from decreasing to increasing as income rises (Table 5). However, in this case the change in the direction of the relationship is so strong that high-income households are those who present the higher average cross-price elasticity (Table2). Within each income level, more elasticity is in general verified for wealthier consumers (Table 5). Cross-price elasticity also increases with age at a decreasing rate (Table 5), although in this case it does reach a maximum at the third range of age considered (Table 3). At the inside of each of those ranges, the relationship is also direct but not at a decreasing rate in all of them (Table 5).

Bastón and smiles, although not the product segments with higher own-price elasticity, are those that consumers substitute the more for other FFP when prices change. Golden longs and rondelles present the lower cross-price elasticity, which in addition with their low own-price elasticity make them the products with a more stable demand (Table 4), except in the youngest households at which they have the higher cross-price elasticity after *bastón.* On the other hand, smiles FPP is the segment with the higher cross-price elasticity in the older

households, which makes sense since this is a kid-oriented product (Figure 6, panel b). Finally, croquettes are again much lesser substitutable for middle-income consumers (Figure 6, panel a).

Due to the flexibility of substitution patterns achieved with the RCDCM, it is possible to compare how much the demand of a product is affected by changes in prices of other products (sensitivity) and how changes in its price affect other products' demand (influence). Although sensitivity and influence are of the same order of magnitude for each segment -a similar result was found when comparing products in the first stage of this research- *bastón* is the only kind of FFP whose price affects other product's demand more than it is affected by others (Figure 7).

On the other hand, results show that consumers choose to substitute more between products of the same variety, but when decide to change the segment they also change the brand. Additionally, they tend to interchange products that are identical except for their package size (Table 6).

These results confirm one of the advantages of RCDCM when compared with SLM: substitution patterns differ regarding the characteristics of products, but SLM estimates one cross-price elasticity for each product which implies that a change in that product's price affect in the same way the demand for all the rest. Another disadvantage is that, as in the case of the own-price elasticity, the underestimation of the price coefficient generates an underestimation of the substitution magnitude. Lastly, the results suggest that the SLM overestimates the relative cross-price elasticity of golden longs and noisette (Figure 8).

One interesting kind of cross-price elasticity, which was not possible to estimate by SLM, is the elasticity of the outside good: the sensitivity of consumers to stop buying any FFP when the price of the one that they were consuming rises, or to start purchasing FFPs when the price of one of them goes down. On average, this measure is lesser than the cross-price elasticity between inside goods, which means that consumers tend to stay rather than to go out of the market when the price of the product they purchase changes. Golden longs have the second larger coefficient (Table 4). Taking into account previous results regarding the price elasticities of this segment, it might be the case that consumers have a stable demand of it and, when its price increases, they prefer relatively more than other consumers going out of the market than replacing it for another kind of FFP.

In terms of demographic characteristics of households, in the lowest age levels the outside alternative sensitivity of smiles is lower, which means that they prefer relatively more than older households to purchase less -or substitute for other segments- rather than stop purchasing smiles when its price goes up. On the other hand, changes in *bastón*'s prices are significantly less able to generate consumers drop FFP market in older neighborhoods (Figure 9).

4.4.   Compensating variation

The inclusion of consumer heterogeneity in RCDCM estimation not only achieves a distribution of welfare impacts across households, which implies that all of them are affected in different ways for what happens in FFP industry, but also obtains an accurate measure of the price coefficient that at the same time yields reliable figures for the compensating variation. In contrast, the aforementioned underestimation of the price coefficient in the utility function from the SLM causes an overestimation of the compensating variation, which is the same for all consumers (Figure 10). This is especially true in the case of the simulated merge between two of the firms, where the compensating variation generated for the SLM is extremely out of the range of the distribution of welfare changes yielded by the full model.

5. Concluding comments

The most recent literature in empirical Industrial Organization provides demand estimation strategies that account for the differences in product valuation among consumers, and therefore in demand decision making, driven by the heterogeneity in individual demographic characteristics. On the one hand, it allows obtaining consistent and flexible estimators for preference parameters, which in turn results in accurate measures of market characteristics in terms of aggregate demand behavior and market structure. On the other hand, it models heterogeneity in such a way that it is possible to analyze the distribution of the preferences for different attributes in terms of demographics, and the consequences of the shape of these distributions on individual-specific elasticities and on how different consumer profiles are affected by changes in market characteristics. Those advantages are of great importance since they generate correct elasticity estimations, which are the main

magnitude considered in any market regulation or public policy. We exploited the first mentioned advantage in a preceding work, to then dig into the second advantage for the case of the FFP industry in an important urban area of Argentina, a developing country where fresh potato is the most consumed vegetable for all income levels, and also where the demand for potato-processed products is swiftly growing.

Our results shed light on what drive differences in consumers' valuation of nutritional and health attributes of FFP and how demographic characteristics affect price sensitivity and substitution patterns, providing insightful information for the decision making of supply chain stakeholders. Knowing how individuals differ in their valuation of healthier processed food and how they react when faced to changes in prices is important for the agents in food industries to optimally launch price differentiation strategies or assessing the impact of the introduction of new goods in the market. Therefore, our work contributes to improve the results of the food supply chain operations in order to increase consumers' satisfaction, while at the same time highlights the importance of flexible demand estimations to infer about consumer preferences and market structure.

6. References

Ackerberg, D., Benkard, L., Berry, S.,&Pakes, A. (2007).Econometric tools for analyzing market outcomes. In J.Heckman and E. Leamer (Eds.), The Handbook of Econometrics, Amsterdam: North-Holland.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. Rand Journal of Economics, 25(2), 242-262.

Berry, S., Levinsohn, J.,&Pakes, A. (1995). Automobile prices in market equilibrium. Econometrica 63(4): 841-890.

Chamberlain, G. (1982). Multivariate regression models for panel data. Journal of Eonometrics 18(1), 5-46.

Chan, T. (2006). Estimating a continuous hedonic-choice model with an application to demand for soft drinks. Rand Journal of Economics 37(2): 1-17.

Erdem, T., Imai, S.,&Keane, M. (2003). Brand and quantity choice dynamics under orice uncertainty. Quantitative Marketing and Economics 1(1): 5-64.

Gentzkow, M. (2006). Valuing new goods in a model with complementarities: online newspapers. NBER working paper # 12562. Retrieved from http://www.nber.org/papers/w12562.pdf

González, J., & Lacaze, M.V. (2012). Preferences, market structure, and welfare evaluations in the Argentinean FFP industry: a case in Buenos Aires Province. Agribusiness, 28(3), 341-360.

Houthakker, H. (1951-52). Compensated changes in qualities and quantities consumed. Review of Economic Studies, 19, 155-164.

Hendel, I., & Nevo, A. (2002). Sales and consumer inventory. NBER working paper # 9048. Retrieved from http://www.nber.org/papers/w9048

Lancaster, K. (1971). Consumer demand: a new approach. Columbia University Press, New York.

Lancaster, K. (1966). A new approach to consumer theory. Journal of Political Economy, 74(2), 132-157.

McFadden, D. (1981). Econometric models of probabilistic choice. In C. Manski and D. McFadden (Eds.), Structural Analysis of Discrete Data, Cambridge: MIT Press.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), Frontiers of Econometrics, New York: Academic Press.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. Econometrica69(2), 307-342.

Pakes, A. (1986). Patents as options: some estimates of the value of holding European patent stocks. Econometrica 54(4): 755-784.

Waugh, F. (1928). Quality factors influencing vegetable prices. Journal of Farm Economics, 10(2), 185-196.

**Tables and Figures**

Figure 1. Distribution of demographic variables

a) Income	b) Income by income level

| c) Age	d) Age by income level
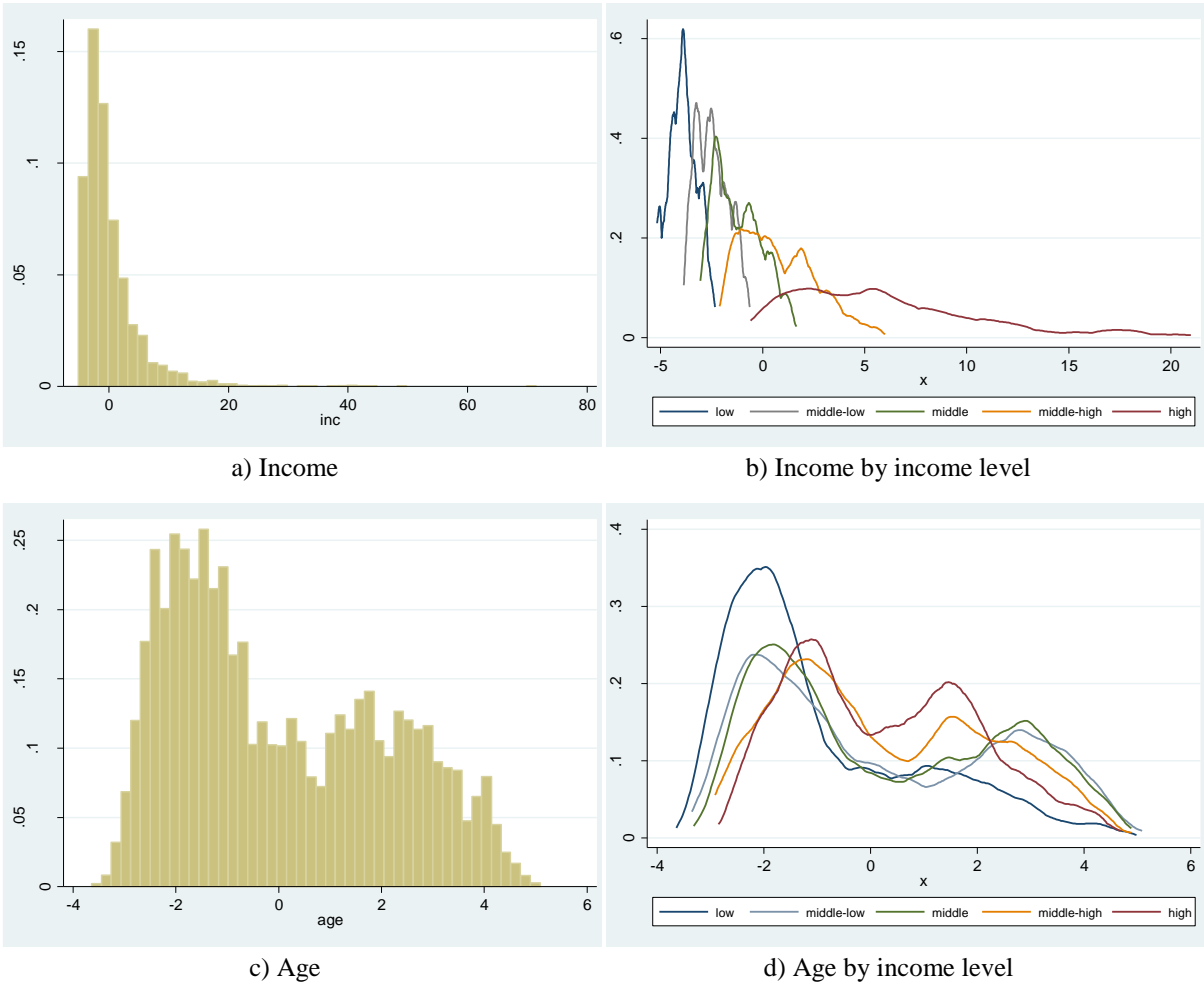
Note: *Income* is hourly per capita income, in Argentine pesos. *Age* is household average age, in years. Both demographic variables are expressed as deviation from the mean.

Figure 2. Distribution of individual coefficients for each random parameter

a) Price

b) Calories

c) Fat

d) Fiber

e) Sodium

f) Smiles
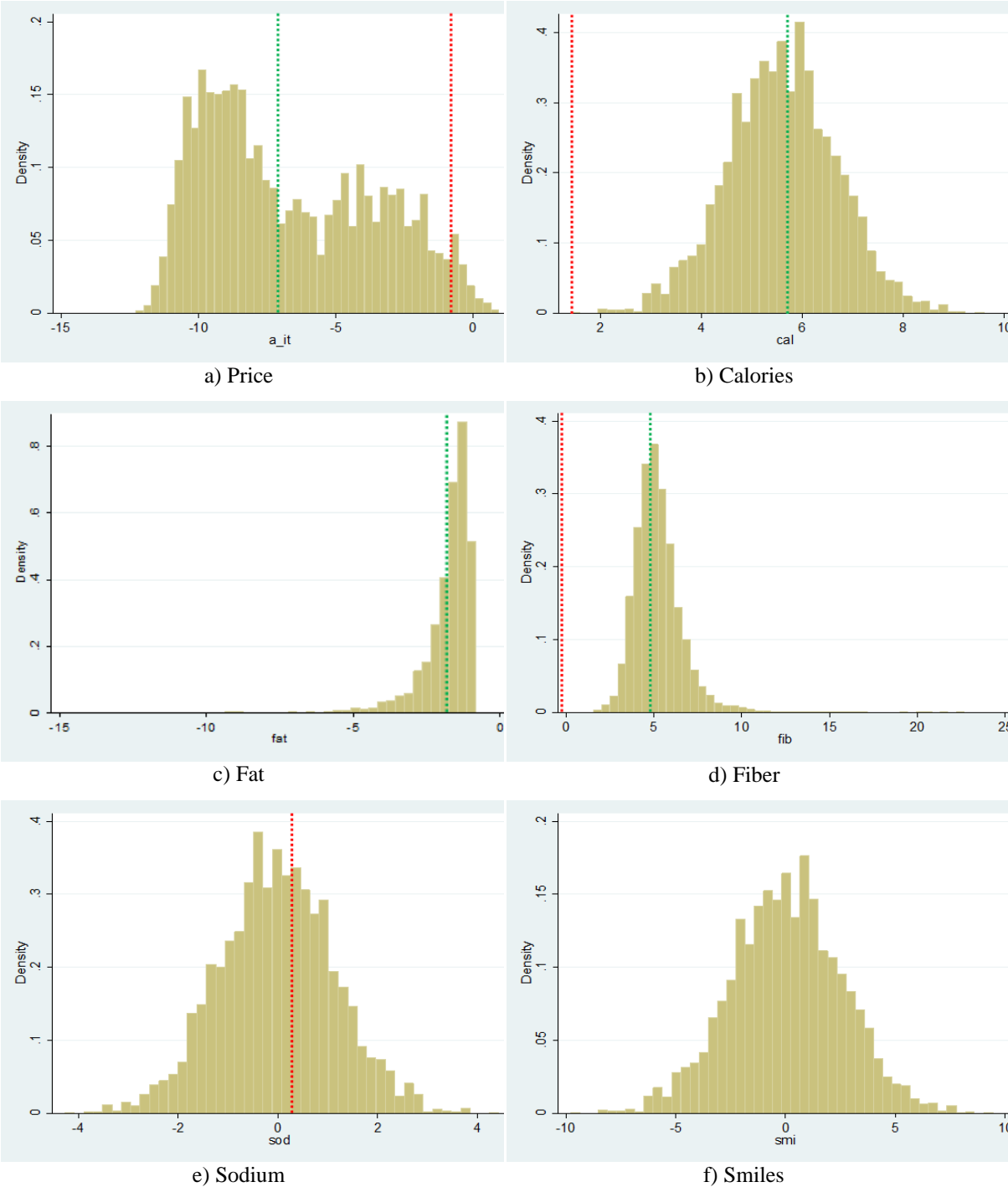
Note: A dotted green line represents the value of the RCDCM's mean parameter, and a dotted red line shows the value of SLM coefficient. If lines are missing the coefficients are non-significant.

Figure 3. Relationship between own-price elasticity and demographics, by income level[1]

Low income level   *High income level*

a) Income

b) Income

c) Age

d) Age

e) Income and age

f) Income and age
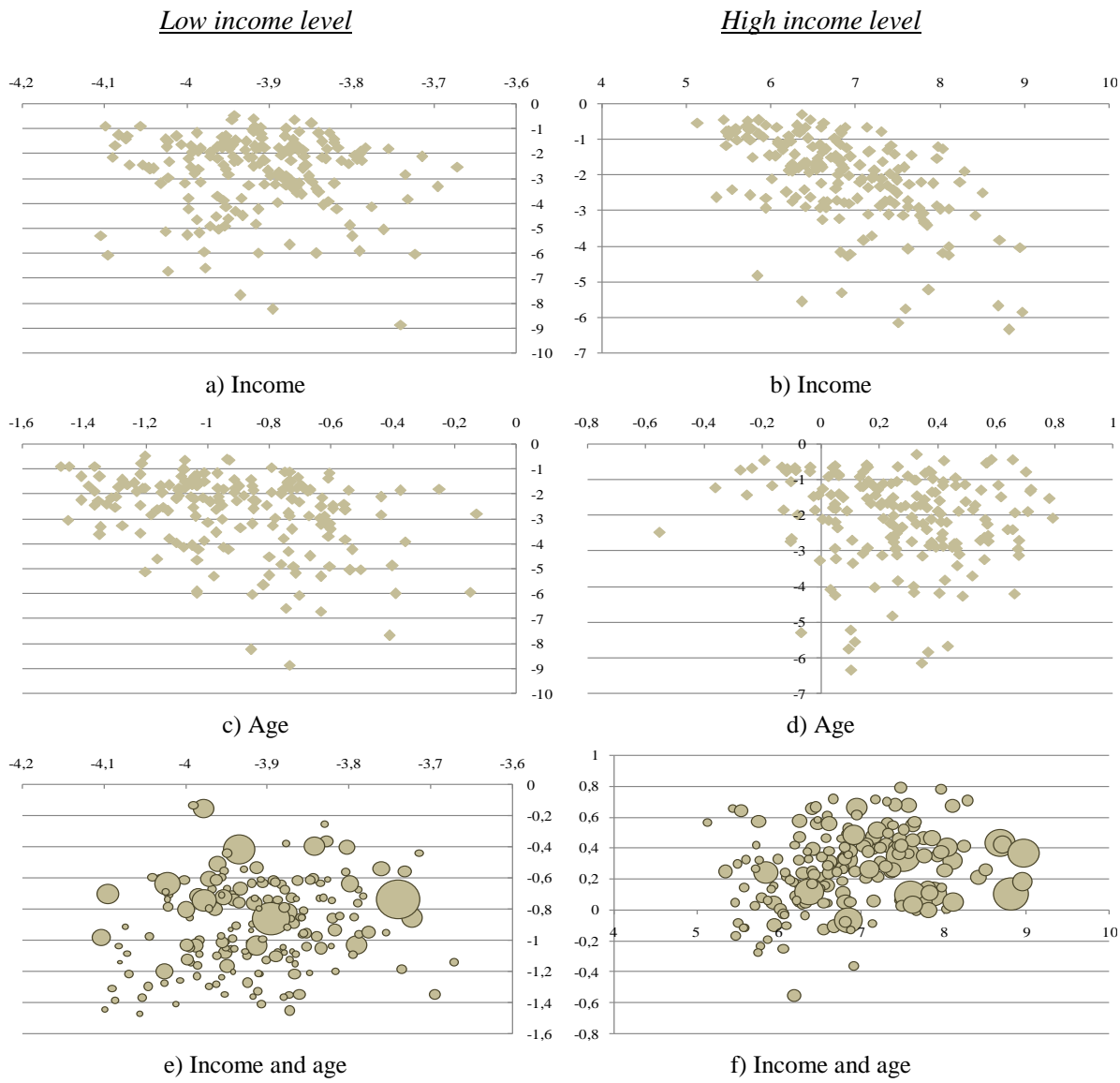
Notes: the vertical axe of figures in panels a, b, c, and d represents cross price elasticity. In bubble charts (panels e and f), the horizontal axe represents income, the vertical axe represents age, and the size of the bubble displays the own price elasticity. [1] This figure only reports low- and high-income levels for brevity.

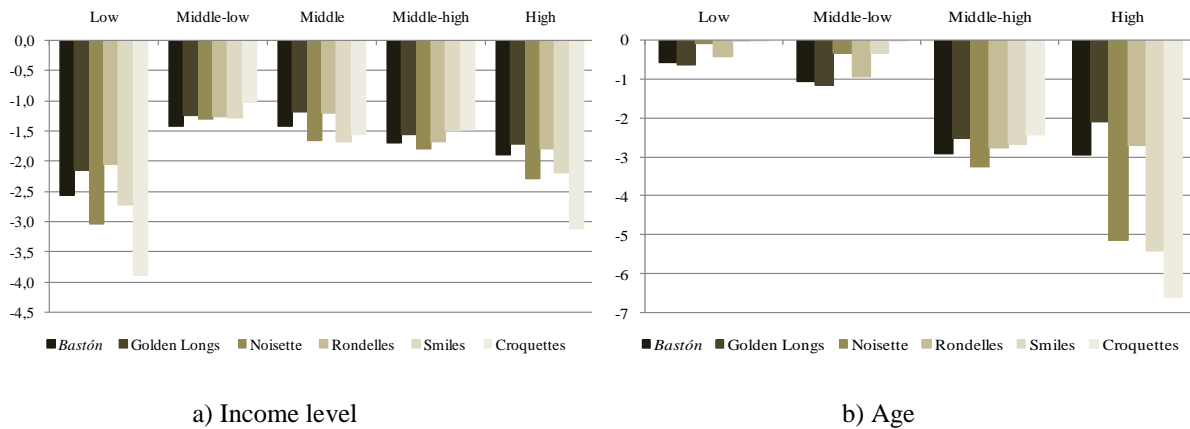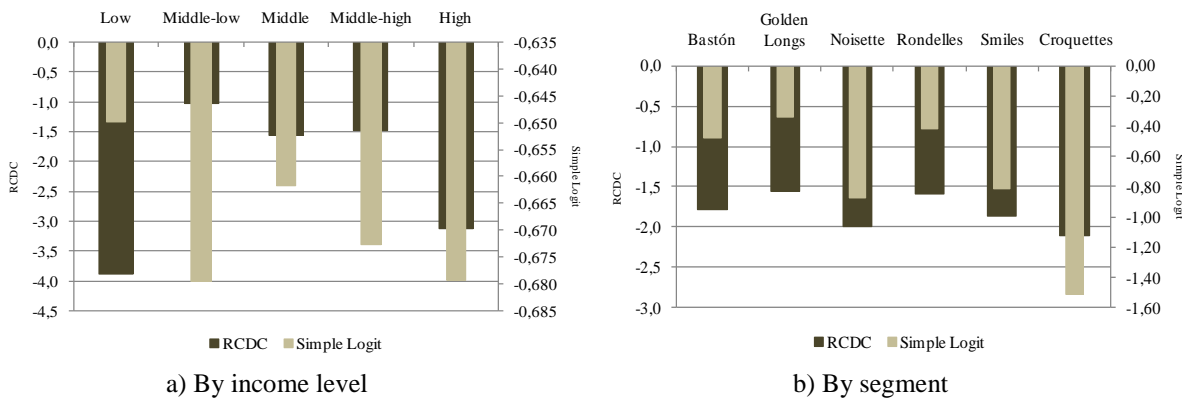Figure 4. Own-price elasticity for each FFP segment, by demographic variables

a) Income level

b) Age

Figure 5. Own-price elasticity for each FFP segment, retrieved from SL and RCDC models, by demographic variables

a) By income level

b) By segment

Note: Results retrieved from SLM and RCDCM are respectively expressed in right and left vertical axes because of their great differences in magnitude.

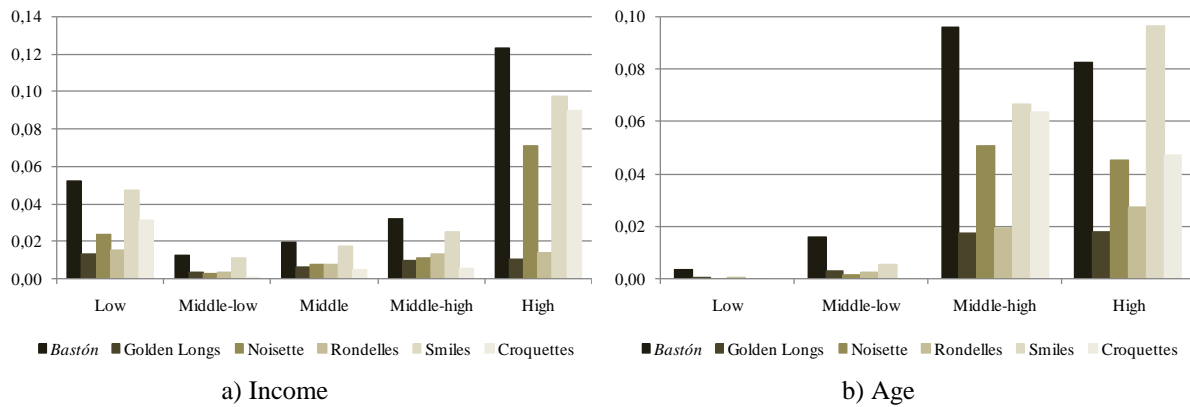Figure 6. Cross-price elasticity for each FFP segment, by demographic variables

a) Income
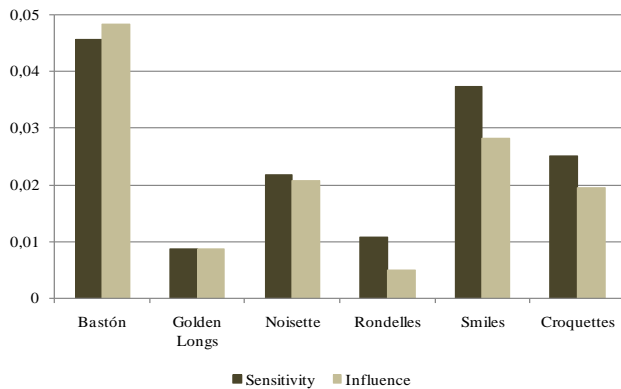
b) Age

Figure 7. Influence and sensitivity, by FFP segment

Figure 8. Cross-price elasticity from SL and RCDC models
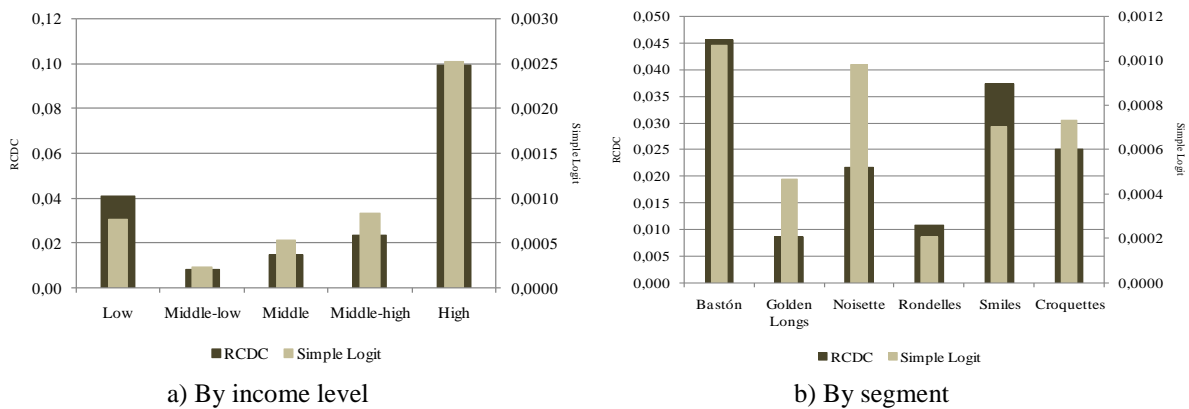
a) By income level

b) By segment

Note: Results retrieved from SLM and RCDCM are respectively expressed in right and left vertical axes because of their great differences in magnitude.

Figure 9. Cross-price elasticity of the outside good, by FFP segments and demographic variables
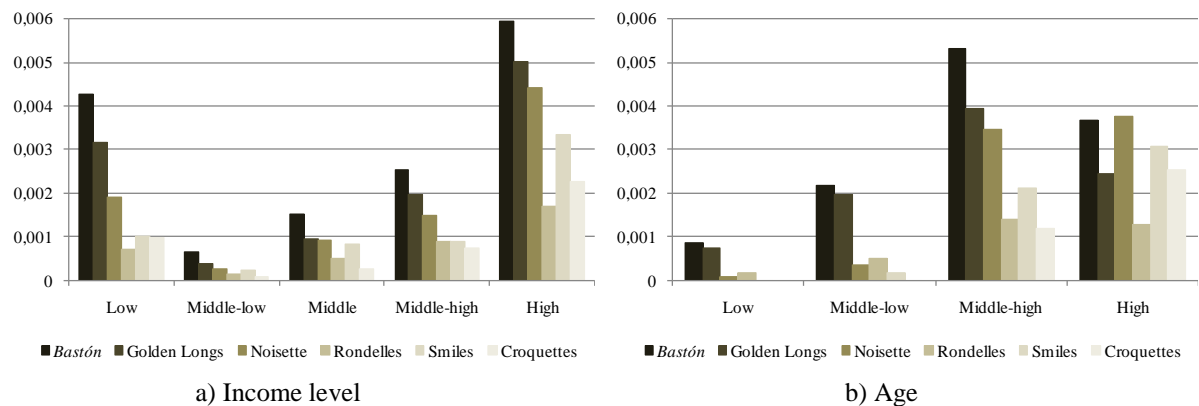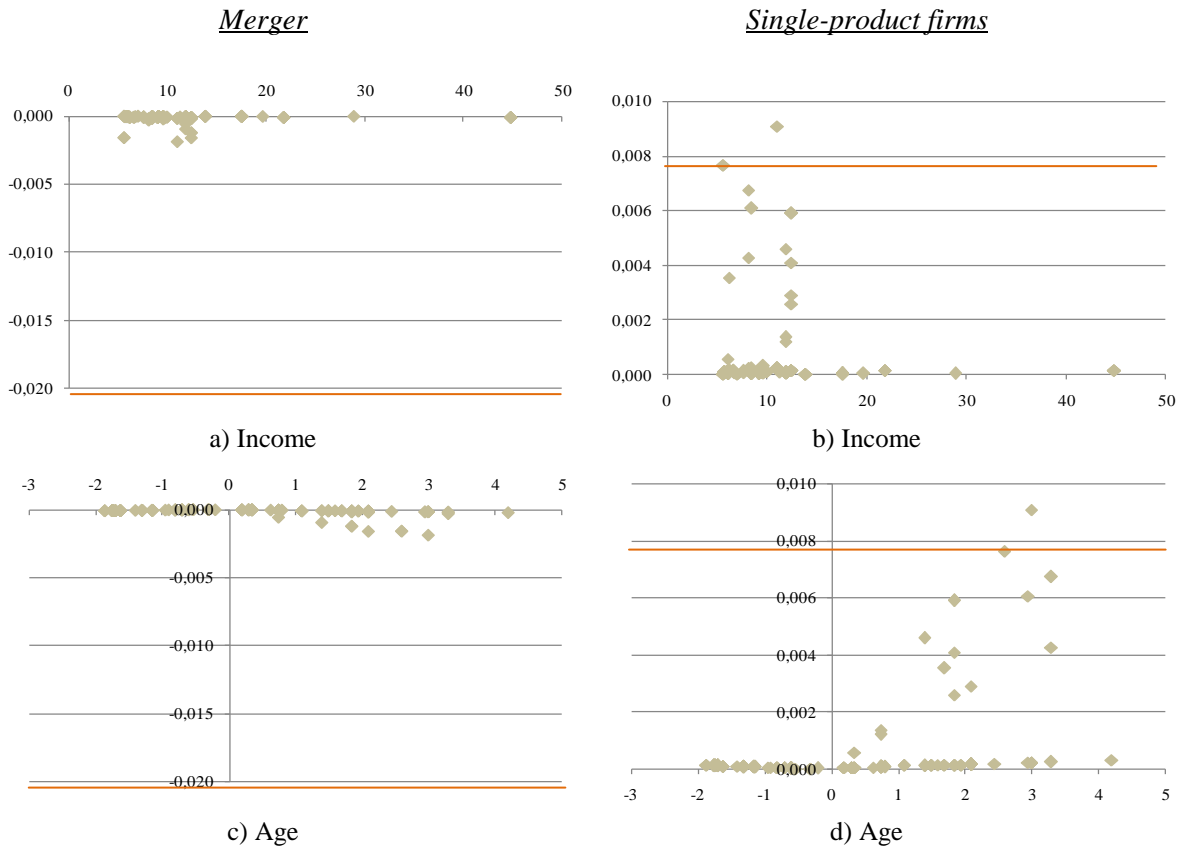

a) Income level

b) Age

Figure 10. Compensating variation for hypothetical changes from SLM and RCDCM

Merger   *Single-product firms*

a) Income   b) Income

c) Age   d) Age

Note: the vertical axe represents compensating variation. Compensating variation of SLM is indicated with a red line.

Table 1. Relationship between own-price elasticity and demographic variables

Demographic variable	Own-price elasticity	Own-price elasticity by income level				
		Low	Middle-low	Middle	Middle-high	High
Income	0.3027	-6.1709	-0.0853[a]	-0.2426	-0.0345[a]	0.4608
Income^2	-0.0338	-0.6491	-0.0365[a]	-0.0073[a]	0.0102[a]	-0.0386
Age	-0.7657	-1.5837	-0.5306	-0.6330	-0.7176	-0.4882
Age^2	0.0726	-0.1288	0.0694	0.0023[a]	0.0214	0.0765

Note: Reported figures are the estimated coefficients of a linear regression in which the dependent variable is the own-price elasticity. [a] indicates the coefficient is non-significant at a 10% level.

Table 2. Average elasticities, by income level

Income level	Own-price elasticity		Cross-price elasticity		
	SLM	RCDCM	SLM	RCDCM	Outside good
Low	-0.64989	-2.74182	0.00077	0.04099	0.00310
Middle-low	-0.67952	-1.33985	0.00023	0.00830	0.00046
Middle	-0.66160	-1.49509	0.00053	0.01463	0.00119
Middle-high	-0.67276	-1.67813	0.00084	0.02367	0.00198
High	-0.67918	-2.09734	0.00252	0.09900	0.00496

Table 3. Average elasticities, by age

Age	Own-price elasticity RCDCM	Cross-price elasticity	
		RCDCM	Outside good
Low	-0.38128	0.00195	0.00053
Middle-low	-0.75159	0.00978	0.00137
Middle-high	-2.93178	0.07607	0.00419
High	-3.94067	0.06795	0.00347

Table 4. Average elasticities, by FFP segment

Segment	Own-price elasticity		Cross-price elasticity			
	SLM	RCDCM	SLM	Sensitivity	Influence	Outside good
Bastón	-0.47667	-1.78458	0.00107	0.04566	0.04838	0.00294
Golden Longs	-0.34007	-1.55277	0.00047	0.00856	0.00856	0.00227
Noisette	-0.87119	-2.00184	0.00098	0.02170	0.02068	0.00173
Rondelles	-0.41391	-1.59142	0.00021	0.01075	0.00487	0.00079
Smiles	-0.81513	-1.87251	0.00070	0.03741	0.02810	0.00120
Croquettes	-1.51174	-2.10267	0.00073	0.02503	0.01950	0.00087

Table 5. Relationship between cross-price elasticity and demographic variables

Demographic variable	Cross-price elasticity	Cross-price elasticity by income level				
		Low	Middle-low	Middle	Middle-high	High
Income	-0.0175	0.1610	0.0137	0.0032	0.0028	-0.0591
Income^2	0.0026	0.0168	0.0025	-0.0001[a]	0.0003	0.0037
Age	0.0154	0.0322	0.0046	0.0067	0.01257	0.0492
Age^2	-0.0019	0.0055	-0.0010	0.0012	0.0001[a]	-0.0048

Note: Reported figures are the estimated coefficients of a linear regression in which the dependent variable is the cross-price elasticity. [a] indicates the coefficient is non-significant at a 10% level.

Table 6. Relationship between cross-price elasticity and brand, segment, and content

Variable	Cross price elasticity
Same segment	0.0371
Same brand	-0.0070
Same segment and brand (dif. package size)	0.0155

Note: Reported figures are the coefficients of a linear regression in which the dependent variable is the cross price elasticity.