Understanding Differences in Segment-specific Willingness-to-pay for the Fair Trade Label

By Friederike Paetz and Daniel Guhl

We conduct a discrete choice analysis in Germany to study consumer behaviour for orange juice in a Fair Trade context. In particular, we estimate Finite Mixture-Multinomial Logit models for analysing the segment-specific willingness-to-pay for the Fair Trade label. The average willingness-to-pay is 48 Eurocent (relative price premium of 24%), and we find substantial heterogeneity in values between segments (10 to 95 Eurocent). A profiling of the segments reveals these differences can be attributed to differences in segment members’ gender, consumption frequency, and consciousness-for-fair-consumption. Segments with a large proportion of women, a high level of consciousness-for-fair-consumption and frequent consumption of orange juice have a higher willingness-to-pay for the Fair Trade label.

1. Motivation and literature review

Currently, the increasing competition in saturated markets of fast-moving consumer goods drives companies to highly sophisticated marketing strategies. Increasing brand equity and reducing prices are two tools that are frequently used to increase consumer preferences and purchase intentions and, therefore, to increase companies’ sales and ultimately profits (cp. Cobb-Walgren et al. 1995). A product enhancement with a social attribute such as the Fair Trade (FT) label could be interpreted as an additional tool. If respondents gain a utility surplus from the social product attribute and honour it mon- etarily, the company’s profits may also increase. Hence, focusing on consumers’ social preferences is an interesting research field for marketing managers and academics. In particular, assessing consumers’ willingness-to-pay (WTP) for the social attribute is highly interesting: a positive WTP serves as a necessary condition for the price premium and, therefore, for increasing profits because adhering to FT standards may result in additional costs for commodities (Fair Trade Deutschland 2014, p. 4).

In marketing practice, profit-oriented companies use the prevailing trend of socially responsible consumption and add social product features to differentiate their products from their competitors. For example, Starbucks promises their customers that 100 percent of their coffee is ethically sourced (certified through external audit systems such as FT), which they achieved in 2015 (Starbucks 2015). This research has proven interesting not only from a practical point of view but also from an academic research perspective. Understanding consumer behaviour in the context of fair consumption has emerged as a vast research field in recent years, as supported by the literature reviews of Andorfer and Liebe (2012) and Tully and Winer (2014). In particular, the consumption of FT products as a component of socially responsible consumption has drawn high interest in practice and marketing theory. Within the context of FT in marketing theory, many studies have focused on consumers’ WTP for the FT label. They found that an aggregated assessment of consumers’ WTP for the FT label attribute might be overly restrictive: some consumers may have a highly positive WTP while others may yield a small or even negative WTP for the FT label attribute. Hence, an aggregated view could result in biased WTP estimates and,
finally, a loss in profits for companies that derive price premia from these biased WTPs.

The problems caused by neglecting WTP heterogeneity (as is done within aggregated approaches) have been identified in the relevant literature, and several studies have highlighted the importance of considering WTP heterogeneity in the FT context. The seminal study of Basu and Hicks (2008), for example, compared differences in the WTP for FT coffee between German and US consumers. They found that German consumers are more sensitive than US consumers, leading to a stronger decline in German consumers’ WTP for increasing relative inequality between participants and non-participants in the FT program (cf. Basu and Hicks 2008, p. 13). Rotaris and Danielis (2011) identified socio-demographic differences in price premia for the FT label in an Italian sample in the product category of coffee. Consumers who are female or young or who regularly buy FT products are more sensitive to FT issues. Yang et al. (2012) determined individual WTP within the coffee category and observed gender and consumption-related differences in a Chinese sample. Female and long-term consumers (or those who intend to increase their coffee consumption) were willing to pay high price premia for the FT label attribute. De Pelsmacker et al. (2005a) asked Belgian respondents about their preferences for FT coffee. They found that the respondents to whom the FT label attribute had the highest importance when making a purchase decision were willing to pay the highest relative price premium of 35%. In contrast, those respondents who mostly cared exclusively about brand or flavour were only willing to pay a relative price premium of 5%. On average, De Pelsmacker et al. (2005a) calculated a relative price premium of 10%. This discrepancy shows, once again, that if WTP heterogeneity is neglected, marketing managers are in danger of losing profits when deciding price premia based on aggregate WTP calculations.

So far, the relevant literature has used varying approaches to evaluate the drivers of differences in consumers’ WTP for the FT label attribute. The recent literature apparently uses non-aggregated approaches, such as individual or segment-level analyses, to incorporate consumers’ WTP heterogeneity. In particular, segment-level WTP estimates are commonly used in marketing practice because they provide a sound basis for companies’ (segment-specific) pricing differentiation strategies and, hence, contribute to companies’ revenue. Furthermore, managers prefer segment-level results because they are easy to communicate and comprehend. However, in the context of preference-based segmentation approaches for determining WTP for the FT label, the so-called one-stage segmentation approaches (simultaneous market segmentation and estimation of segment-specific preferences) are less popular in the marketing literature. These approaches, however, allow a deeper understanding and a more precise prediction of consumer behaviour compared to two-stage (i.e., estimation of individual preferences and subsequent clustering) segment approaches (Ramaswamy and Cohen 2007, p. 297) and may thus also contribute to strategic marketing decisions in practice.

In the context of ethical consumption, the study of Auger et al. (2008) is an example that actually uses a one-stage segmentation approach. The authors found three consumer segments that differed in the importance that they attached to several social product features (child labour, minimum wage, dangerous working conditions and living standards). However, they neither classified their results to the FT context nor derived segment-specific WTP for social product attributes. To contribute to the scarce academic research within this field, we conduct a discrete choice analysis for orange juices and estimate a preference-based segmentation model, i.e., Finite Mixture-Multinomial Logit (FM-MNL) model, calculating segment-specific WTP for the FT label attribute. We address the following research questions: Do differences exist in the WTP for the FT label between different preference-based consumer segments? How can these differences be explained in terms of the socio-demographic or psychographic characteristics of segment members?

Usually, the influence of demographic variables such as gender or age is examined, but the current results are conflicting: Rotaris and Danielis (2011) identified an influence of demographic variables on consumers’ WTP for an Italian sample, and Yang et al. (2012) did so for a Chinese sample, but De Pelsmacker et al. (2005b) did not observe demographic differences in consumer preferences for the FT label attribute in a Belgian sample. Because these conflicting results were found for samples in different cultures, cultural differences in the demographic drivers of consumer WTP are likely. We focus on a German sample.

In addition to these demographic variables, we use the psychographic variable “Consciousness-for-fair-consumption” (cfc) (cp. Balderjahn et al. 2013) to test its influence on consumers’ WTP. Considering psychographic variables to be potential drivers of WTP differences seems reasonable because De Pelsmacker et al. (2005a, p. 366) already stated that demographic variables might not be sufficient to fully describe the socially responsible consumer. Psychographic variables may also affect consumers’ purchase behaviours. The cfc construct serves as a psychographic variable and is tailored to the FT context. Such variables are highly underrepresented in relevant studies so far. The work of Arnot et al. (2006) is one of the few studies to consider a related variable: “familiarity with the concept of FT coffee”. The authors found that respondents who are familiar with the concept of FT coffee are more likely to choose a FT coffee. Balderjahn and Peyer (2012) examined the cfc variable in the FT context and discovered a relationship between consumers’ cfc levels and the FT label’s influence on consumers’ purchase decisions for a German sample. Hence, this further investigation of the cfc variable’s potential as a driver for WTP differences contributes to the recent literature.
Finally, we address consumers’ consumption frequency (e.g., weekly, monthly, or less often consumption) as a behavioural construct. Consumption-related variables, such as “generally buying FT products: yes/no” (cp. Rotariss and Daniellis 2011) or “long-term consumption” (i.e., more than five years) (cp. Yang et al. 2012), have been considered in the recent literature. However, these variables relate primarily to consumers’ familiarity with the (FT) product category than to the actual consumption frequency of the focal product. However, the frequency of purchase and, therefore, the amount of money spent for the focal product may influence consumers’ WTP in general.

The remainder of this paper is structured as follows: We explain data collection and analysis methods in Section 2. We summarize the empirical results of our discrete choice experiment for the estimated FM-MNL model in Section 3, including the profiling of consumer segments. We close in Section 4 by drawing conclusions and discussing limitations as well as future research issues.

2. Methods

2.1. Data collection

To obtain the data, we conducted a discrete choice experiment via an online survey. For the following analysis, we can distinguish two different types of data: choice data and respondent data. Specific details regarding the data of our empirical study are presented in the next section; here, we focus on the way the data were collected.

Choice data

The survey contained a discrete choice experiment that consisted of 16 choice sets. We used the first 14 choice sets of each respondent for model estimation and the last two choice sets for model validation. Note that this number is well below the number of 20 choice sets, which Johnson and Orme (1996) suggested as an upper bound before the data quality begins to degrade. Each choice set included three (hypothetical) orange juice alternatives as well as a ‘no purchase’ option. The latter is essential to precisely measure WTP, because a forced choice situation (i.e., a choice set without a ‘no purchase’ option) may cause an underestimation of price sensitivity (Allen et al. 2014). To describe the orange juice alternatives, we used the following four attributes (with the corresponding levels in parentheses): brand (Albi, Granini, Hohes C, Valensina), type of packaging (Tetra Pak (carton), PET bottle (plastic)), display of an FT label (yes, no), and price (per litre) (1.09 €, 1.39 €, 1.69 €, 1.99 €). We followed the guidelines of Orme (2002). For the price levels and different types of packaging, we used those levels that are prevalent in German grocery stores. For the brand levels, we chose leading national orange juice brands from the German market (Statista 2016). Note that we use these attributes and levels to create orange juice alternatives that are hypothetical but which respondents nevertheless perceive as realistic choice options. The specific variation of attributes in our experimental design enables us to measure their effects on utility, including the impact of a currently not available FT label for the top four national orange juice brands. The online survey was created and administered using Sawtooth Discover. The individual-specific fractional factorial design, without any attribute prohibitions, is near orthogonal and has high D-efficiency. In particular, our 4 x 2 x 2 x 4 design has a D-efficiency of over 99 % compared to a design created using the complete enumeration approach of Sawtooth’s SSI Web module, which has optimal efficiency for main effect designs. However, our design also contains a certain amount of overlap within and between attributes and avoids dominant alternatives (see Sawtooth Software 2014). The use of a discrete choice experiment tackles the problem of the often-discussed gap between consumer attitudes to buy FT products due to social desirability and their actual behaviour in real purchase situations. It is therefore advisable to use a realistic choice situation, in which respondents must trade-off among several product attributes (cf. DePelsmacker et al. 2005, p. 513). Discrete choice experiments closely mimic real choice situations and are therefore known to relax the overestimation of the influence of social product attributes, such as the FT label attribute (cf. Auger and Devinney 2007).

Respondent data (background variables)

As mentioned before, we aim to analyse (potential) WTP differences for the FT label across consumer segments. To this end, we employ several promising respondent-level background variables, which we derived from prior literature. In particular, we use demographics (age and gender), orange juice consumption frequency, and cfc. While questions eliciting the former variables are straightforward (e.g., “how frequently do you drink orange juice”), some clarification of the cfc scale is in order. Balderjahn et al. (2013, p. 546) presented a scale measuring the “latent disposition of consumers to prefer products that are produced and traded in compliance with fair labour and business practices”. This psychographic construct is directly related to preferences for FT products: Consumers with higher cfc should also have higher WTP for a FT label. Balderjahn et al. (2013) tested the cfc scale in three independent studies and found high validity and reliability across all samples. Hence, we are confident in the quality of this scale. The respondents’ cfc levels are measured on a 7-point rating scale, with six items each corresponding to a specific labour standard (i.e., compliance with workers’ rights; freedom from forced labour; abolition of illegal child labour; non-discrimination in the workplace; compliance with international statutory labour standards; fair wages for workers). For each item, a belief component \( B_p \) (“I only buy a product if I believe that in its production ...”) and an im-
2.2. Analysis

Our analysis consists of several parts. The main steps are depicted in Fig. 1 and briefly discussed. The appendix contains technical details regarding model, estimation, and segmentation. The data described before can be separated into 3 data sets that serve different purposes in our analysis: (1) choice data (in-sample), which we use for model estimation; (2) choice data (out-of-sample), which we use for model validation and selection; and (3) respondent data (background variables), which we use for profiling segments.

We employ the FM-MNL model to determine segment-specific preferences. This model is rooted in random utility theory and allows us to account for preference heterogeneity (see, e.g., Elshiewy et al. 2017, for an overview). The FM-MNL model, as a one-stage segmentation approach, permits a simultaneous market segmentation and estimation of segment-specific preferences (see, e.g., Vriens et al. 1996 for a comparison of one-stage and two-stage segmentation approaches). The deterministic part of the utility function is linear-additive in the attribute levels of an alternative, and parameters can vary across segments. All non-price attributes are dummy-coded; hence, the corresponding parameters measure the utility of an attribute with respect to the reference level (i.e., brand “Albi”, packaging “plastic”, FT label “no”). The (negative) effect of price on utility is assumed to be linear, i.e., a vector utility function (Orme 2007, p. 2). Although the use of partworth utility functions may account for non-linearities in price, we decided to use a linear price function for several reasons. First, non-linearities in the price function are negligible in the case of our data, so a more complicated model with additional parameters is unnecessary. Second, the use of a linear price parameter simplifies our subsequent calculation of respondents’ WTP because WTP is then price-independent (Louviere et al. 2000, p. 280).

Because the number of segments is unknown before the analysis, we estimate FM-MNL models with 2 to 5 segments to cover the typical number of segments (3–4 segments) in marketing literature and practical applications (Tuma and Decker 2013, p. 11). We also add a version with one segment only (i.e., the regular MNL model) as a benchmark. We use several starting values in the maximum likelihood estimation to avoid convergence to local optima (cf. Grün 2008).

After model estimation, we must select the final model for analysis (cf. Melnikov and Mahtia 2010, p. 88). We employ multiple measures to this end (see, e.g., Wedel and Kamakura 2000, for an overview) and also consider the interpretability of resulting segments. We evaluate model fit using log-likelihood (LL) values and first choice hit rates. For the latter, the two extra choices (out-of-sample) are used. In addition, we use the Bayesian information criterion (BIC) and Entropy. Whereas the BIC helps in model selection by penalizing model fit with model complexity (number of parameters), Entropy informs about how distinct the assignment of respondents to the segments is (for a model that contains a particular number of segments).

Once we have decided which model is best according to the different measures described above, we can obtain segment-specific estimation results. For each segment, we can obtain parameter estimates and the segment size.
Furthermore, we can derive segment-specific attribute importance (i.e., the utility of an attribute relative to the range of utility values) and WTP values for the FT label (i.e., the utility of the FT label rescaled in monetary units). These measures are common in discrete choice analysis and help us to understand and compare raw parameter estimates. The latter is based on the marginal rate of substitution between the FT label attribute and price. For our models with a dummy-coded FT label attribute and a linear price effect, we have the following for each segment (Tully and Winer 2014):

\[
WTP_{\text{FT label}} = \frac{U(x_{\text{FT label}} = 1) - U(x_{\text{FT label}} = 0)}{-\alpha} = \beta_{\text{FT label}},
\]

where \( \beta_{\text{FT label}} \) measures the utility difference of an alternative with and without an FT label and \( \alpha \) is the marginal effect of price \( p \) on utility \( U \) (see Small and Rosen 1981 for a general discussion of welfare measures in a discrete choice analysis). Note that we assume an economically meaningful negative price-effect (i.e., all else being equal, consumers prefer to pay less). Additionally, this measure is not an elasticity measure because we are not interested in the relative change in utility caused by a relative change in an independent variable, such as price. We are interested in the absolute change in utility due to the presence of the FT label, and we express this change in monetary units (using \( \alpha \)) for ease of interpretation. Finally, in contrast to direct approaches where respondents state their WTP directly, indirect methods suffer less from an overrating of respondents’ WTP (see, e.g., Breidert et al. 2006 for further details).

After selecting the final model, we calculate respondent’s probability of belonging to a particular segment based on the parameter estimates and individual choices. Then, we assign each respondent to the segment where he/she has the maximum (so-called) posterior membership probability (“discrete segmentation”) (see DeSarbo et al. 1995). Intuitively, the potential segmentation error is smaller with non-overlapping segments if the solution has high Entropy (i.e., well-separated segments).

Because our research aims not only to identify differences in segment-specific preferences but also to understand whether respondents’ characteristics can explain these differences, we “profile” these segments after assigning each respondent to a single segment. This means we use respondent-level background variables and analyse their distributions post hoc across segments (cp. Wedel and Kamakura 2000, p. 145). Such an approach is simple to perform by using standard methods of multivariate analysis, and the communication of the results is straightforward. In particular, for categorical variables, such as gender or age-class, \( \chi^2 \)-tests of the corresponding contingency tables are appropriate (e.g., \( H_2 \); the frequency of women are the same across segments). For continuous variables, such as cfc-values, F-tests based on a one-way ANOVA are adequate (e.g., \( H_3 \); mean cfc-values are the same across segments).

### 3. Empirical study

The questionnaire was distributed online at two German universities in spring 2015. Overall, 360 respondents completed the survey: 58% were female, and most respondents were 25 years old or younger (64%). Regarding consumption frequency, many respondents consumed orange juice ‘at least once a week’ (41%), followed by ‘one to three times per month’ (36%). Only 22% of the respondents stated they consume orange juice ‘less often than once per month’.

The 360 respondents made 5040 and 720 discrete choice decisions in- and out-of-sample, respectively. The in-sample choice shares are very similar for the three orange juice alternatives (each about 29%), which was an expected result because the alternatives were “unlabeled” in the experiment (i.e., all attributes varied freely over all alternatives in each choice set). The choice share for the “no purchase” option is 13%. The choice shares for the out-of-sample data are 29%, 28%, and 26% for the orange juice alternatives and 17% for the ‘no purchase’ option. Although these values are quite comparable to the corresponding in-sample values, the choice share for the ‘no purchase’ option is slightly higher.

The cfc-scale has high reliability (Cronbach’s \( \alpha = .96 \)), and the results of an exploratory factor analysis indicate the scale’s unidimensionality (e.g., all factor loadings \( \geq .85 \); proportion of variance explained = .80). The results are very similar to those of Balderjahn et al. (2013), and we conclude that the scale measures the cfc construct well.

#### 3.1. Model selection

Using in-sample choice data, we estimated several FM-MNL models with a varying number of segments. As explained earlier, we also estimate as a benchmark a simple MNL model with one segment. Tab. 1 displays the number of parameters to be estimated and the statistical model fit (i.e., LL value, BIC statistic, Entropy, and first choice hit rates).

Apparently, the LL value decreases with an increasing number of segments. This was expected because considering more segments improves model fit and therefore causes higher LL values. To select a model that both fits the data well and is parsimonious, it is advisable to inspect the BIC statistic, which accounts for the model’s LL value and the model-specific number of parameters.

<table>
<thead>
<tr>
<th># Segments</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td># Parameters</td>
<td>7</td>
<td>15</td>
<td>23</td>
<td>31</td>
<td>39</td>
</tr>
<tr>
<td>LL</td>
<td>-5323.63</td>
<td>-4936.28</td>
<td>-4656.41</td>
<td>-4472.77</td>
<td>-4332.60</td>
</tr>
<tr>
<td>BIC</td>
<td>10706.94</td>
<td>10000.64</td>
<td>9580.91</td>
<td>9209.82</td>
<td>8997.67</td>
</tr>
<tr>
<td>Entropy</td>
<td>.85</td>
<td>.88</td>
<td>.89</td>
<td>.90</td>
<td></td>
</tr>
<tr>
<td>Hit rate</td>
<td>.54</td>
<td>.58</td>
<td>.63</td>
<td>.64</td>
<td>.65</td>
</tr>
</tbody>
</table>

Note: The proportional chance criterion (Morrison 1969) for the out-of-sample data is .26.

Tab. 1: Number of parameters and performance measures

https://doi.org/10.15358/0344-1369-2017-4-37


Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

Paetz/Guhl, Understanding Differences in Segment-specific Willingness-to-pay for the Fair Trade Label

MARKETING · ZFP · Issue 4 · 4. Quarter 2017 41
Here, the BIC statistic also decreases monotonically with an increasing number of segments, but the decline slows between the four- and five-segment solutions (BIC-difference of 212.15) compared to the three- to four-segment solutions (BIC-difference of 299.09). The entropy-based measure saturates at quite high levels (88 % and 90 % for solutions with more than two segments), which indicates well-separated segments. The first choice hit rates indicate an excellent out-of-sample fit compared to the values of a trivial benchmark (proportional chance criterion, see Morrison 1969). All models increase the fit by more than 100 %, but again, we see that the model with five segments is only marginally better than the four-segment solution. Based on the measures reported in Tab. 1 one may favour the five-segment-solution. However, a closer inspection of this solution reveals that it contains several non-significant parameter estimates indicating overfitting. Furthermore, two (of the five) segments are quite similar in their estimates. This prevents a unique interpretation of segments, which is crucial from a managerial point of view. Because it is a very good trade-off between model fit, predictive validity, and interpretability of segments, we select the four-segment solution. This solution also provides a sound basis for our subsequent profiling task.

3.2. Estimation results

Tab. 2 displays the associated segment-specific parameter estimates of the four-segment solution. The segments 2, 3, and 4 most prefer the brand Hohes C, but the preference order of the other brands differs across segments (e.g., segment 1: Hohes C/Albi, Valensina, and Granini; segment 4: Hohes C, Granini, Valensina, and Albi). Additionally, all segments prefer orange juice with the FT label. Furthermore, the linear price parameter has a reasonable negative sign in all segments, implying that respondents’ preferences decrease as prices increase. Segment 4 favours the PET bottle, in contrast to segments 1 and 3 that prefer a carton (Tetra Pak). Furthermore, all segments have a meaningful size (19 % to 34 %), which indicates that the four-segment solution does not suffer from overfit.

The WTP for the FT label attribute is positive and significant in all segments and ranges between 10 Eurocent (segment 2) and 95 Eurocent (segment 3). Standard errors of the WTP estimates are computed using a parametric bootstrap with 10,000 draws (Krisny and Robb 1986). The aggregated (i.e., segment-size weighted) WTP for the FT label within our orange juice example is approximately 48 Eurocents (95 % CI [42.00, 54.93]), which translates into a relative price premium of approximately 24 %, based on an average estimated WTP of 1.96 € for orange juice alternatives in the experiment. The magnitudes of the WTP estimates are reasonable given recent meta-analysis results from Tully and Winer (2014), who evaluated 80 studies dealing with socially responsible consumption. If we subset their results (see Appendix A of Tully and Winer 2014, pp. 267–271) to food-related studies where people are the beneficiary of the FT label (as in our study), the relative price premium is approximately 29 %. Thus, our estimate is slightly lower but still comparable with the results from literature, which strengthens our confidence in the validity of our results.

An inspection of segment-specific attribute importance in Tab. 3 reveals further insights that aid in explaining these segment-specific differences in consumers’ WTP for the FT label attribute.

While segment 2 almost exclusively cares about the price of orange juice (82 %), the brand attribute is highly important (51 %) for segment 4. Segment 1 attaches the greatest importance to the price attribute (49 %), but it also (nearly equally) cares about packaging (21 %) and the FT label (22 %). Members of segment 3 care primarily about the FT label (36 %) but also consider the price attribute (34 %) in their purchase decision process. This behaviour coincides with the results of Devinney et al. (2012), who concluded that companies should (also) heavily focus on appropriate pricing for their (socially enhanced) products because even consumers who care

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Segment 3</th>
<th>Segment 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand</td>
<td>.09</td>
<td>.08</td>
<td>.08</td>
<td>.51</td>
</tr>
<tr>
<td>Price</td>
<td>.49</td>
<td>.82</td>
<td>.34</td>
<td>.27</td>
</tr>
<tr>
<td>FT label</td>
<td>.22</td>
<td>.09</td>
<td>.36</td>
<td>.08</td>
</tr>
<tr>
<td>Packaging</td>
<td>.21</td>
<td>&lt;.01</td>
<td>.21</td>
<td>.13</td>
</tr>
</tbody>
</table>

Tab. 2: Segment-specific estimates of parameters and FT label WTP

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Segment 3</th>
<th>Segment 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>No purchase</td>
<td>-3.21* (.33)</td>
<td>-11.26* (.73)</td>
<td>-3.13* (.32)</td>
<td>-2.06* (.53)</td>
</tr>
<tr>
<td>Granini</td>
<td>-40* (.17)</td>
<td>.09 (.20)</td>
<td>.06 (.11)</td>
<td>1.28* (.29)</td>
</tr>
<tr>
<td>Hohes C</td>
<td>.27 (.15)</td>
<td>.65* (.20)</td>
<td>.40* (.10)</td>
<td>2.13* (.31)</td>
</tr>
<tr>
<td>Valensina</td>
<td>-37* (.17)</td>
<td>.35* (.19)</td>
<td>.25* (.12)</td>
<td>.68* (.28)</td>
</tr>
<tr>
<td>FT label (yes)</td>
<td>1.62* (.14)</td>
<td>.74* (.23)</td>
<td>1.73* (.12)</td>
<td>.35* (.13)</td>
</tr>
<tr>
<td>packaging (carton)</td>
<td>1.59* (.20)</td>
<td>.02 (.15)</td>
<td>1.02* (.09)</td>
<td>-.53* (.12)</td>
</tr>
<tr>
<td>price</td>
<td>-40* (.24)</td>
<td>-7.29* (.60)</td>
<td>-1.84* (.14)</td>
<td>-1.26* (.18)</td>
</tr>
<tr>
<td>seg. size (%)</td>
<td>.21* (.01)</td>
<td>.27* (.01)</td>
<td>.34* (.01)</td>
<td>.19* (.01)</td>
</tr>
<tr>
<td>WTP</td>
<td>.40* (.03)</td>
<td>.10* (.03)</td>
<td>.95* (.08)</td>
<td>.27* (.12)</td>
</tr>
</tbody>
</table>

Tab. 3: Segment-specific attribute importances

Note: * indicates significance at p < .05. Standard errors in parenthesis.
about social product attributes (e. g., the FT label) also care about core product attributes (cf. Auger et al. 2008, p. 190). This calls for a holistic and integrated marketing mix strategy including investments in socially responsible production practices.

Segment-specific preference characteristics are also present in the corresponding WTP values. While the price-sensitive segment 2 yields the smallest WTP for the FT label of 10 Eurocents, social segment 3 exhibits the highest FT label WTP of 95 Eurocents. In sum, this answers our first research question on WTP differences between different preference-based segments.

3.3. Profiling segments

Based on the four-segment solution for the FM-MNL model from Section 3.2, we now assign respondents to particular segments and profile the segments using individual background variables. This reveals further valuable insights concerning our second research question of how these differences can be explained in terms of the socio-demographic or psychographic characteristics of segment members.

First, we analyse the cfc variable. The group means for segments 1 to 4 are .22, -.23, .19, and -.27, respectively. A one-way ANOVA clearly shows that the cfc-means are not the same across groups \( (F = 6.17, df = 3, p < .01) \). Segments with higher average cfc-values also show higher FT label WTP (e. g., segment 1 and 3). This replicates the results of Balderjahn and Peyer (2012), who report a positive influence of consumers’ cfc level on consumers’ WTP for the FT label.

Fig. 2, furthermore, illustrates gender-specific differences between segments. The proportion of women in segment 1 (61 %) and segment 3 (67 %) exceeds those in segment 2 (51 %) and segment 4 (51 %). Hence, segments with a higher WTP for the FT label include higher proportions of women. These gender-specific differences between segments is (weakly) significant based on a contingency analysis \( (\chi^2 = 7.39, df = 3, p = .06) \). This mirrors the findings of gender’s influence on consumers’ WTP found in Italian (Rotaris and Danielis 2011) and Chinese (Yang et al. 2012) samples.

The contingency analysis did not find significant differences in age (classes) between segments \( (\chi^2 = 1.21, df = 3, p = .75) \), which is in line with the results of Yang et al. (2012). However, a contingency analysis discovered segment-specific differences in the frequency of orange juice consumption \( (\chi^2 = 13.62, df = 6, p = .03) \), which mirrors the results of Rotaris and Danielis (2011) in the Italian sample. Segments 1 and 3 contain higher proportions of members who consume orange juice at least once a week than do segments 2 and 4, which exhibit a lower WTP for the FT label. Hence, those segments consuming orange juice more often are willing to pay higher price premia for the FT label than those segments consuming less often.

4. Conclusion

In this contribution, we focused on segment-specific differences in consumers’ WTP for the FT label as well as on determinants for these WTP differences. We conducted a discrete choice experiment for orange juice at two German universities and estimated FM-MNL models. We selected a four-segment solution and calculated segment-specific WTPs for the FT label by using the marginal rate of substitution between the FT label and price. We found substantial differences in segment-specific WTP for the FT label: between 10 and 95 Eurocents. The average value in the sample is 48 Eurocents, which corresponds to a relative price premium of 24 %. The profiling of segments with individual background variables revealed that these differences can be related to differences in segment members’ gender, consumption frequency, and cfc level. Segments with a larger proportion of women, a higher cfc level or a more frequent consumption of orange juice have, all else being equal, a higher WTP for the FT label. However, consumers’ age did not affect their WTP in our sample.
Our results also have relevant managerial implications. In contrast to a priori segmentation, which is commonly based on one individual background variable, preference-based one-stage segmentation approaches with the subsequent segment profiling with several individual background variables contribute to a deeper understanding of social consumer segments and are easy to conduct. Marketing managers can now identify that the target segment for an FT orange juice consists of women with a high cfc level who consume orange juice frequently. Marketing mix strategies, e.g., managerial decisions on product line extensions, price differentiation or tailored promotional campaigns, could be built on this information and potentially lead to increasing company profits because consumers are addressed more appropriately.

Our study is based on a convenience (student) sample, where young and well-educated respondents are oversampled. Because students are known to be more receptive to the FT context in general (cp. Yang et al. 2012, p. 24), our WTP estimates may be upwardly biased. Additionally, we used stated choices from a discrete choice experiment, rather than revealed market data, which may also contribute to inflated WTP estimates (cp. Völkner 2006). However, no top German orange juice brand has a FT label so far, which renders impossible the use of revealed choice data for WTP calculations in our specific application. Furthermore, our results in terms of WTP value seem to be conservative compared with the results of the recent meta-analysis of Tully and Winer (2014). Additionally, the corresponding general price elasticities of the FM-MNL model are elastic, reasonable in magnitude (e.g., between -6 and -1.5) and comparable to other studies using market data (cf. Weber and Steiner 2012).

Hence, we are confident that our WTP estimates for the FT label are not (heavily) overestimated. Furthermore, our main findings primarily concern the segment-specific WTP differences in consumers’ gender, cfc level and consumption frequency, and we are confident that our findings are replicable in a representative German sample.

We identified several drivers for WTP differences and since the recent (sparse) literature already observed differences in WTP between different cultures, future research should focus more heavily on cultural differences in WTP determinants. In this context, the investigation of emerging versus developed countries may constitute an especially interesting research field and furthermore relates to prospective investigations of consumer income as a WTP driver. Some researchers (e.g., Sonnier et al. 2007) also advocate specifying and estimating discrete choice models directly in WTP spaces instead of preference spaces and then, as we did, computing WTP values using equation (1). Differences between both approaches might be relevant if a continuous specification of consumer heterogeneity is used (i.e., mixed logit models, see Elshiewy et al. 2017). In our application, the differences for the FM-MNL should be negligible. However, we still think that employing a discrete choice model in WTP space is a fruitful avenue for future research in the context of analysing WTP for FT labels.

Hence, utility \( U_{jt} \) is linear additive in the effects of the covariates \( x_{jt} \) and \( p_{jt} \) as well as the error term \( \varepsilon_{jt} \). Furthermore, respondents are utility maximisers, i.e., respondent \( j \) picks an alternative \( i \) if no other alternative \( q \neq i \) in choice set \( t \) has a higher utility. Together with the particular choice for the distribution of the random components of the model \( \varepsilon_{jt} \) this leads to the well known multinomial logit (MNL) model for the choice probabilities \( Pr_{jt} \):

\[
Pr_{jt} = \frac{\exp(U_{jt})}{\sum_{s=1}^{S} \exp(U_{jt})}
\]  

(A2)

These choice probabilities are conditional on respondent \( j \) belonging to segment \( s \). The unconditional choice probability for alternative \( i \) is:

\[
Pr_{j} = \sum_{s=1}^{S} \pi_{s} Pr_{jts}.
\]  

(A3)

Here the conditional choice probabilities (A2) are multiplied by the unknown size of a segment \( \pi_{s} \), which can be interpreted as the \textit{a priori} probability of finding a respondent in segment \( s \):

\[
\pi_{s} = \frac{\exp(\gamma_{s})}{\sum_{s=1}^{S} \exp(\gamma_{s})}
\]  

(A4)

Hence, for the segment sizes, we also assume a MNL (sub)model, where \( \gamma_{s} \) are parameters to be estimated, with \( \gamma_{s} = 0 \) as additional restriction for identification. Note that \( \sum_{s=1}^{S} \pi_{s} = 1 \) and \( 0 < \pi_{s} < 1 \). Model parameters are estimated using maximum likelihood, and we employ the following log-likelihood function of the FM-MNL model:

\[
\text{log-likelihood} = \sum_{t=1}^{T} \sum_{s=1}^{S} \sum_{j=1}^{J} \left( \sum_{i=1}^{I} \pi_{s} \log Pr_{jts} \right).
\]

Appendix

In this appendix, we provide technical details regarding the methods (model, estimation, and segmentation) used in the analysis. We follow closely DeSarbo et al. (1995) and Wedel and Kamakura (2000). See also Elshiewy et al. (2017) for an overview of multinomial logit models in marketing, incl. the Finite Mixture-Multinomial Logit (FM-MNL) model used here.

We assume that respondents come from a population composed of several unobserved segments and that they have the following general (indirect) utility function:

\[
U_{jt} = \sum_{k=1}^{K} \beta_{jt} x_{jt} + \alpha_{jt} p_{jt} + \varepsilon_{jt},
\]  

(A1)

with:

- \( j = 1, \ldots, J \) respondents,
- \( t = 1, \ldots, T \) choice alternatives,
- \( \pi_{s} \) latent segments,
- \( k = 1, \ldots, K \) attributes and dummy variables excluding price,
- \( x_{jt} \) for respondent \( j \), the \( k \)th dummy variable of alternative \( i \) in choice set \( t \),
- \( p_{jt} \) for respondent \( j \), the price variable of alternative \( i \) in choice set \( t \),
- \( \beta_{jt} \) effect of the \( k \)th attribute for segment \( s \),
- \( \alpha_{jt} \) effect of the price for segment \( s \),
- \( y_{jt} \) choice variable, equals 1 if respondent \( j \) chooses alternative \( i \) in choice set \( t \) and 0 otherwise,
- \( \varepsilon_{jt} \) random components, assumed to be i.i.d. extreme value type I distributed.
\[ LL(\theta) = \sum_{j=1}^{J} \ln \left( \sum_{s=1}^{S} \sum_{\tau} \pi_s \prod_{j=1}^{K+2} P_{ji}\tau | s, y_{ij}\tau | s \right), \]  
(A5)

where the vector \( \theta = (\alpha_1, \ldots, \alpha_l, \beta_1, \ldots, \beta_K, \gamma_1, \ldots, \gamma_l) \) contains all parameters to be estimated. Note that we exploit the structure of the data in (A5), where due to the data collection design all respondents have \( T \) observations, because the whole product of conditional choice probabilities for the observed sequence of choices for each respondent is weighted by the segment sizes. We estimate all \((K + 2) \cdot S - 1\) parameters simultaneously using gradient-based methods. Specifically, we use the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (see Greene 2008, p. 1071) implemented in the gmmnl package in R (Sarrias and Daziano 2017). Because mixture models may have multiple local optima (see Wedel and Kamakura 2000), multiple starting values should be tested to find the global optimum (cf. Grün 2008, p. 235).

Once the parameters have been estimated, Bayes’ theorem can be used to compute the posterior probability for respondent \( j \) belonging to segment \( s \):

\[ \hat{\pi}_s = \frac{\hat{\pi}_s \cdot \prod_{j=1}^{K+2} \hat{p}_{ji}\tau | s}{\sum_{s} \hat{\pi}_s \cdot \prod_{j=1}^{K+2} \hat{p}_{ji}\tau | s}. \]  
(A6)

In (A6), the estimated prior probabilities are re-weighted using the estimated likelihood of each respondent \( j \) conditional on segment \( s \). These probabilities can be directly used for a “fuzzy” (i.e., probabilistic) segmentation, where respondents can be fractional members of multiple segments (cf. Wedel and DeSarbo 1994). However, we follow DeSarbo et al. (1995) and apply a non-overlapping segmentation where we form discrete segments by assigning each respondent \( j \) to the segment \( s \) where the value for \( \hat{\pi}_s \) is highest.

References


Keywords
Fair Trade Label, Discrete Choice Experiment, Finite Mixture-Multinomial Logit Model.