Understanding Heterogeneous Consumer Preferences in Chinese Milk Markets: A Latent Class Approach

Xiang Wu, Bin Hu, Jie Xiong

▶ To cite this version:

Understanding Heterogeneous Consumer Preferences in Chinese Milk Markets: A Latent Class Approach

Xiang Wu, Bin Hu and Jie Xiong[1]

Please cite as:

Abstract

We examine heterogeneous consumer preferences in Chinese milk markets. Using a discrete choice experiment, we examine how the brand, quality certification, traceability label and price influence consumers' milk choices. We identify four consumer segments using a latent class model: price conscious (9.8%), balanced thinking (19.8%), health conscious (57.5%), and environment conscious (12.9%) consumers. These four segments have distinct preferences: price conscious consumers prefer

green certification; balanced thinking consumers have the highest willingness to pay for traceability labels; health conscious consumers have strong brand awareness; and environment conscious consumers prefer organic certification and traceability labels and use price as a quality signal. Such diversity of consumer preference can be explained by four psychological factors: price consciousness, food safety concerns, health consciousness and environmental concerns.

1. Introduction

The Chinese dairy industry has experienced rapid growth in the past decade, with the production of dairy products increasing from 17.9 million tons in 2007 to 27.8 million tons in 2015. Scholars attribute this growth to industrial policies, income growth, urbanisation, and changes in consumption habits (Fuller *et al*., 2006; Wang *et al*., 2015). Nevertheless, the Chinese dairy industry needs further improvements in various aspects, such as industrial structure, resource use efficiency, product quality and brand competitiveness (Yu, 2012). Most dairy companies are small-scale farms, which largely ignore milk quality and safety, and thus earn very low profits (Yu, 2012). These domestic dairy products cannot meet consumers' emerging requirements, including high quality, traceability and environmentally-friendly production (Liu *et al*., 2013). Chinese consumers lack confidence in domestic dairy products, since they can hardly determine the quality of dairy products (Bai *et al*., 2013; Liu *et al*., 2013). Thus, the Chinese government issued the 'National dairy industry development plan (2016–2020)' to cope with these problems. This plan considers the dairy industry as a representation of food safety, national health and agricultural modernisation, putting forward several goals such as improving product quality and building strong brands. In recent years, the Chinese government has adopted stringent industry regulations and closed unqualified dairy companies. Changes in consumer demands and industry policies are fostering the development of the Chinese dairy industry.

In response to these changing demands, dairy companies can choose several strategies to consolidate consumers' confidence and influence their quality perceptions. Consumers learn about product quality from various cues such as price, brand and certifications. In the food industry, quality certification and traceability labels are considered as promising quality signals (Aung and Chang, 2014). Dairy companies can decide whether to obtain quality certifications from the existing three-level certification system in China: hazard-free, green and organic certifications (Liu *et al.*, 2013). Furthermore, recent advances in internet of things (IoT) technologies such as radio-frequency identification (RFID), near-field communication (NFC), and two-dimensional barcode (QR) technologies bring new opportunities to the traceability systems (Yoo *et al.*, 2015). Hence, dairy companies can enhance consumer perceptions of product quality through building strong brands, obtaining quality certifications, and establishing traceability systems. However, all these strategies require substantial investment. Dairy companies need to make prudent decisions based on a thorough understanding of consumer preferences and willingness to pay for these quality signals (Cheng *et al.*, 2015).

Consumer preferences for goods and services are heterogeneous (Boxall and Adamowicz, 2002). We can assume that several consumer segments exist in the market; and consumer preferences are homogeneous in the same segment. Given differing consumer preferences across segments, a company can decide which consumer segments to serve, and what customised product and marketing strategy to choose for each segment. Previous studies have demonstrated how to identify consumer segments by using choice experiment data and latent class models (Apostolidis and McLeay, 2016; Boxall and Adamowicz, 2002; Pouta *et al.*, 2010). Latent class models (LCMs) provide preference estimates and consumer profiles for each segment. Furthermore, LCM relates preference heterogeneity to consumer characteristics, such as socio-demographics (Boxall and Adamowicz, 2002). More importantly, Boxall and Adamowicz (2002) and some other following studies such as Rao (2014) highlight the important but understudied psychological factors in influencing consumers' preferences, for example attitudes and perceptions. However, consumer segmentation based on psychological factors is valuable for differentiated marketing strategies. To fill this research gap, we also examine how psychological factors form heterogeneous preferences in the LCM. Our review of established

literature on food consumption behaviour indicates four key psychological factors: price consciousness, food safety concerns, health consciousness and environmental concerns (Hughner *et al.*, 2007; Michaelidou and Hassan, 2008; Nasir and Karakaya, 2014).

Our contributions are twofold. First, we use a latent class model to identify consumer segments in Chinese milk markets as well as their heterogeneous preferences. A few studies on milk preference consider preference heterogeneity, most of which implement the random parameter model (RPL) to reveal the existence and scale of heterogeneity (Olynk *et al.*, 2012; Yamamoto *et al.*, 2008; Wolf *et al.*, 2011; Wu *et al.*, 2014). The RPL model, however, cannot explain the sources of heterogeneity (Boxall and Adamowicz, 2002). Using LCM, we identify consumer segments, estimate preferences of each segment, and explain preference heterogeneity based on consumer characteristics. In addition, we use psychological variables rather than socio-demographic characteristics to identify the source of preference heterogeneity.

2. Literature Review

Previous studies have examined consumer preferences and willingness to pay (WTP) for milk or dairy products by using contingent valuation methods (CVM), rating-based conjoint analysis (Ares *et al.*, 2010; Costanigro *et al.*, 2015; Miklavec *et al.*, 2015), choice experiments (Bai *et al.*, 2013; Olynk *et al.*, 2012; Wu *et al.*, 2014), and supermarket scanner data (Dhar and Foltz, 2005). The most related literature for our study is that which uses choice experiments (Brooks and Lusk, 2010; Wolf *et al.*, 2011; Wu *et al.*, 2014; Yamamoto *et al.*, 2008). For example, Wu *et al.* (2014) conduct a choice experiment to investigate consumers' WTP for infant milk formula with different organic certification labels. Similarly, Wolf *et al.* (2011) evaluate the value of various milk production attributes by choice experiments. Choice experiments can simulate real purchase scenarios and minimise hypothetical bias (Jedidi and Jagpal, 2009).

In choice experiments, researchers should carefully determine the product attributes and the set of product profiles to be chosen by consumers (Jedidi and Jagpal, 2009). Previous studies have discussed some key attributes, including price (Bai *et al.*, 2013; Brooks and Lusk, 2010; Wu *et al.*, 2014), nutritional information (Miklavec *et al.*,

2015), type of milk or yogurt (Ares *et al*., 2010; Bai *et al*., 2013), brand (Ares *et al*., 2010; Wu *et al*., 2014), package (Kildegaard *et al*., 2011), production process (Wolf *et al*., 2011; Yamamoto *et al*., 2008), quality certification (Bai *et al*., 2013; Wolf *et al*., 2011), and traceability label (Bai *et al*., 2013). Because consumers pay increasing attention to food safety and environmental protection, quality certification and traceability labels are of interest for researchers (Jin *et al*., 2017; Wu *et al*., 2016; Xu and Wu, 2010). Therefore, we include quality certification and traceability label in the choice experiment, in addition to the other two primary attributes, namely, price and brand.

In established studies, scholars mainly apply a multinomial logit (MNL) or RPL model to estimate consumer preferences for milk attributes (Bai *et al*., 2013; Olynk *et al*., 2012; Wu *et al*., 2014; Yamamoto *et al*., 2008). An MNL model assumes consumer homogeneity and thus benefits computational simplicity. However, such a model also implies the independence of an irrelevant alternatives (IIA) assumption (Jedidi and Jagpal, 2009). By contrast, an RPL model takes preference heterogeneity into account by allowing parameters to vary randomly over individuals and relaxes the IIA assumption inherent in the MNL model (Garrod *et al*., 2014). The RPL model, however, fails to explain the sources of heterogeneity (Boxall and Adamowicz, 2002). In reality, we often need to know the consumer segments and their profiles for implementing customised marketing strategies. Thus, we use an LCM to estimate consumer preferences, identify consumer segments, and explain preference heterogeneity based on consumer characteristics.

To explain why consumer preferences vary across segments, most studies refer to socio-demographic characteristics but ignore psychological factors (Boxall and Adamowicz, 2002). However, we pay particular attention to the role of psychological factors for two main reasons. First, consumer perceptions of the attributes or alternatives form the basis for preferences (Rao, 2014). For example, consumers' perceptions of food safety and environment issues can influence their preferences for organic certification and thus can be important for differentiating products in the market. Studies of consumer food perceptions also indicate the importance of psychological factors. For instance, Magnusson *et al*. (2003) analyse the roles of human health, environment and animal welfare motives in predicting purchase

intentions for organic foods. Similarly, Nasir and Karakaya (2014) find that health orientation significantly influences consumers' intention to purchase organic foods. By contrast, Michaelidou and Hassan (2008) report that food safety is the most important predictor of attitude. In addition, Hughner *et al*. (2007) suggest that price consciousness leads to a discrepancy between consumer attitude and actual behaviour. Therefore, we select health consciousness, food safety concerns, environmental concerns, and price consciousness to identify the source of preference heterogeneity.

3. Methods

3.1. Choice experiment design

We include four milk attributes in our choice experiment design: brand, quality certification, traceability label, and price.

- Brand: we choose the top three brands in China as levels, that is, *GuangMing*, *MengNiu* and *YiLi*.

- Quality certification: we choose the three legally defined quality certifications in China (Liu *et al*., 2013). *Hazard-free food* is of good quality, nutritious and safe governed by national standards which set limits for harmful and toxic residues. With regard to *green food*, the materials used throughout the production process have to meet the specified standards of environment protection. In the *organic food* production process, producers should not use artificially synthesised fertilisers, pesticides, growth regulators, livestock and poultry feed additives, or genetically engineered technology. Furthermore, producers should renew hazard-free and green certifications every three years, and renew organic certification annually (Liu *et al*., 2013).

- Traceability label: two representative traceability labels in the dairy industry are numeric code label and QR code label. Thus, we choose three levels, namely, no traceability label, numeric code traceability label, and QR code traceability label.

- Price: by taking the prices of milk sold in supermarkets as a reference and conducting a pilot test, we set three price levels, namely, 3.5 Chinese Yuan Renminbi (CNY), 5 CNY and 6.5 CNY for 250ml fluid milk.

-

Table 1

Attributes and associated levels

Attributes	Levels
Brand	GuangMing, MengNiu, YiLi
Quality certification	Hazard-free, Green, Organic
Traceability label	No, numeric code, QR code
Price (in CYN)	3.5, 5, 6.5

Our choice experiment design includes four attributes, each with three levels (Table 1). The full factorial design involves $3^4 = 81$ profiles, the number of which is too large for respondents to evaluate. To reduce the number of profiles, we construct an orthogonal array for the 3^4 design which consists of nine profiles by using support.CEs package in R (Aizaki, 2012). We further follow Rao (2014) and use the *shifting method* to generate choice sets from this orthogonal array. This procedure generates nine choice situations. Figure 1 provides an example choice question in our choice experiment.

	A	B	C	D
Brand	MengNiu	YiLi	GuangMing	
Quality certification	Hazard-free	Green	Organic	None of the
Traceability label	QR	No	Numeric	above
Price (in CYN)	6.5	3.5	5	

When you want to buy 250ml fluid milk, which one would you choose?
- A
- B
- C
- None of the above

Figure 1. An example choice question from the choice experiment

Our questionnaire includes three sections. The first section lists questions about the socio-demographic variables, including gender, age, education level, occupation and income. The second section presents our generic choice experiment. Each respondent

was asked to answer nine choice situations, each of which requires the respondent to choose a milk product from three alternatives and an opt-out option. The last section measures respondents' perceptions about food safety, health, environment and price by a five-point Likert scale. We adapted the measurement items of food safety concerns (FS) and health consciousness (HL) from Michaelidou and Hassan (2008), those of price consciousness (PC) from Alford and Biswas (2002), and those of environmental concerns (EC) from Fujii (2006).

The questionnaire was hosted and distributed by a well-recognised and widely used Chinese online survey service provider, Sojump (http://www.sojump.com). Sojump has 2.6 million nationwide registered members, from which Sojump randomly invites members to answer paid surveys. It has been used for consumer research in numerous studies, such as He *et al.* (2018) and Jia *et al.* (2018). Hence, we can get access to the nationwide representative samples. We conducted a pilot test with 50 respondents to ensure the comprehensiveness of the survey and determine reasonable price levels.

3.2. A latent class model

Choice experiments are rooted in the Lancaster (1966) theory of consumer choice and random utility theory (McFadden, 1974). Respondent i ($1 \leq i \leq N$) evaluates J alternatives in choice situation $k(1 \leq k \leq K)$ and selects the j-th alternative. Random utility theory (RUT) argues that individuals make their choices by maximising the latent utilities of alternatives (McFadden, 1974, 1980, 2001). That is to say, the latent utility of the j-th alternative is larger than other alternatives in choice situation k,

$$U_{i,j,k} > U_{i,l,k} \text{ for any } 1 \leq l \leq J \text{ and } l \neq j. \tag{1}$$

Furthermore, RUT assumes that the unobservable utility includes two separate components: a deterministic (observable) component $V_{i,j,k}$ and a random error (unobservable) component $\epsilon_{i,j,k}$, that is,

$$U_{i,j,k} = V_{i,j,k} + \epsilon_{i,j,k}. \tag{2}$$

The Lancaster (1966) theory of consumer choice models the observed utility $V_{i,j,k}$ as a linear function of the product attributes $X_{j,k}$,

$$V_{i,j,k} = \beta_{i,ASC} ASC_{j,k} + \beta_i X_{j,k}, \text{ for } 1 \leq j \leq J, \tag{3}$$

where $X_{j,k}$ is the product profile of alternative j in choice situation k. β_i characterises respondent i's preference vector towards the product attributes. The dummy variable $ASC_{j,k} = 0$ indicates that the j-th alternative is the opt-out option. Therefore, $ASC_{J,k} = 0$ and $ASC_{j,k} = 1$ for $1 \leq j \leq J - 1$. $\beta_{i,ASC}$ represents respondent i's preference towards the three brands included in our choice experiment compared to the opt-out option. Product profile $X_{j,k}$ has four attributes as shown in Figure 1, namely, brand, quality certification, traceability label and price. We consider price as a continuous variable[2] and use the price coefficient $\beta_{i,Price}$ to calculate respondents' willingness to pay for a specific attribute. Considering that the other three attributes are categorical variables, we take the first level of each attribute as our reference level. Furthermore, we normalise the scale of utility by setting the observed utility of the opt-out option as $V_{i,J,k} = 0$.

If $\epsilon_{i,j,k}$ obeys the Gumbel distribution (also known as the generalised extreme value distribution Type one), we have a logit choice probability (McFadden, 1974):

$$\text{prob}_{i,j,k} = \frac{\exp(\beta_{i,ASC}ASC_{j,k} + \beta_i X_{j,k})}{\sum_{l=1}^{J} \exp(\beta_{i,ASC}ASC_{j,k} + \beta_i X_{l,k})}. \tag{4}$$

We further assume the existence of S segments in a population and respondent i belongs to segment $s(s = 1, \ldots, S)$. Consumer preferences are homogeneous within a segment but vary across different segments. Thus, the preference vector is specific to a segment, that is, $\beta_i = \beta_s$. We rewrite the utility function as:

$$V_{i|s,j,k} = \beta_{s,ASC}ASC_{j,k} + \beta_s X_{j,k} + \epsilon_{i,j,k}.$$

The probability that respondent i in segment s chooses alternative j among J alternatives in choice situation k is:

$$\text{prob}_{i|s,j,k} = \frac{\exp(\beta_{s,ASC}ASC_{j,k} + \beta_s X_{j,k})}{\sum_{j=1}^{J} \exp(\beta_{s,ASC}ASC_{j,k} + \beta_s X_{j,k})}. \tag{5}$$

Consider a latent membership likelihood function $m_{i,s}$ that classifies respondent i into segment s, where $\sum_{s=1}^{S} m_{i,s} = 1$. This segment membership can be further explained

---

[2] We use (price – 3.5) instead of the original price values.

by respondent-specific variables Z_i:

$$m_{i,s} = \gamma_s Z_i + \xi_{i,s}, \tag{6}$$

where γ_s specifies how respondent-specific variables influence segment membership; and $\xi_{i,s}$ is an error term which obeys the Gumbel distribution (McFadden, 1974). Similar to equation (4), the probability that respondent i has membership in segment s is (Boxall and Adamowicz, 2002):

$$\text{prob}_{i,s} = \frac{\exp(\gamma_s Z_i)}{\sum_{s=1}^{S} \exp(\gamma_s Z_i)}. \tag{7}$$

Therefore, the probability that respondent i chooses alternative j in choice situation k is given by (McFadden, 1974, 1980):

$$\text{prob}_{i,j,k} = \sum_{s=1}^{S} \text{prob}_{i|s,j,k} \times \text{prob}_{i,s}.$$

Let $y_{i,j,k} = 1$ if respondent i chooses alternative j in choice situation k, and $y_{i,j,k} = 0$ otherwise. Thus, the log-likelihood function for the sample is:

$$\ln L = \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{J} y_{i,j,k} \cdot \ln \left(\sum_{s=1}^{S} \text{prob}_{i|s,j,k} \times \text{prob}_{i,s} \right). \tag{8}$$

Equations (5), (7) and (8) jointly define the log-likelihood function. We can estimate β_s, $\beta_{s,ASC}$ and γ_s parameter vectors via maximum likelihood methods using the *gmnl* package in R (Sarrias and Daziano, 2017; Train, 2009).

4. Results

4.1. Descriptive statistics of respondents

Our data were collected by using an online paid questionnaire in May 2017. After excluding 187 invalid responses, we obtained 691 usable responses. These samples result in 6,219 choices. Table 2 gives the descriptive statistics of these respondents.

Table 2

Descriptive statistics of respondents

	Frequency	Sample (%)	China Population (%)

Gender[1]			
Male	297	43	51.2
Female	394	57	48.8
Age (years)[1]			
18 – 35	312	45.2	26.2
36 – 50	358	51.8	34.0
> 50	21	3	39.8
Education level[2]			
Below college	30	4.3	90.5
College	608	88	9.2
Above college	53	7.7	0.3
Occupation[1]			
Student	48	6.9	28.5
Worker	643	93.1	71.5
Income (CNY/month)[3]			
< 3,000	53	7.7	53.7
3,000 – 5,000	164	23.7	21.5
5,000 – 8,000	315	45.6	12.9
> 8,000	159	23	11.9

Notes: (1) Data source: China Population and Employment Statistics Yearbook 2017.
(2) Data source: The Sixth National Population Census of the People's Republic of China conducted in 2010.
(3) Data source: the 42nd Statistical Report on Internet Development in China published in 2018.

The demographic statistics of our sample deviate from that of the national population, especially for the education level, occupational status and income level. The main reason for this inconsistency is that milk consumers are only a small part of the national population. The annual per capita consumption of milk in China is still low; although increasing disposable income, urbanisation, and health consciousness are leading to growth in milk consumption (Renub Research, 2018). Not everyone has access to fluid milk in China where the proportions of the rural population and low-income class remain high. As a consequence, milk consumers have higher education and income levels than the national population. Besides, our online sampling procedure may also lead to a potential bias. In an ideal but unlikely achievable context, we should randomly select respondents at the entrance to the carefully-chosen fluid milk outlets. We note that conducting online choice experiment surveys

is now commonplace; and an online survey platform with a balanced, nationwide respondent sample can perform well (He *et al*., 2018; Jia *et al*., 2018; Sheremet *et al*., 2017). Furthermore, the descriptive statistics in Table 2 are consistent with previous studies that developed based on the sample obtained from the Sojump survey platform, for instance, see Jia *et al*. (2018). Thus, we expect that this sampling bias is limited and will not significantly affect the results of our study. Comparing with established literature both methodologically and empirically, it is expected that our potential sampling bias is limited and will affect our results marginally.

Before estimating the econometric models, we assess our measurement model for psychological factors, such as attitudes and perceptions (Rao, 2014). We report the reliability and validity in Tables A1 and A2 (online Appendix). First, internal consistencies are adequate, with Cronbach's α and CR exceeding 0.7 and factor loadings exceeding 0.6 for all constructs (see Table A1). Second, the AVE is above 0.5 and the square root of the AVE is much larger than other cross-correlations for all constructs (see Table A2), which suggests good validity. Finally, we conduct CFA and find the measurement model provides a satisfactory fit to the dataset: χ^2/df = 1.517, comparative fit index (CFI) = 0.991, Tucker-Lewis Non-normed fit index (TLI) = 0.988, root-mean-square error of approximation (RMSEA) = 0.027, and standardised root mean square residual (SRMR) = 0.028. These tests suggest that the measurement model provides a solid foundation for further analysis.

4.2. Choosing the number of segments

We select the optimal value of the number of latent classes S by comparing different models. Following Boxall and Adamowicz (2002), we use the minimum Akaike Information Criterion (AIC) and the minimum Bayesian Information Criterion (BIC) to select the optimal number of latent classes S^*.

Table 3

A comparison of LCM with different number of segments

	P	LL	McFadden's R^2	AIC	BIC
S=1	8	–6,394	0.159	12,803	12857
S=2	21	–6,176	0.188	12,394	12,535
S=3	34	–6,078	0.200	12,224	12,453

S=4	47	–5,995	0.211	12,085	**12,401**
S=5	60	–5,939	0.219	11,999	12,403
S=6	73	–5,900	0.224	11,947	12,439
S=7	86	–5,858	0.229	11,889	12,468
S=8	99	**–5,827**	**0.233**	**11,852**	12,518
S=9	112	–5,835	0.232	11,894	12,648
S=10	125	–5,872	0.228	11,994	12,836
RPL	15	–6,371	0.162	12,771	12,872

Notes: (1) Sample size is 6,219 choices from 691 respondents.
(2) McFadden R^2 is calculated as 1–LL/LL(0).
(3) AIC is calculated as 2P–2LL. BIC is calculated as ln(N) ×P–2LL.

Table 3 presents the statistical summary for $S = 1,2,\ldots,10$ latent classes and the RPL models. The number of parameters (P) and the log-likelihood (LL) are listed in the first two columns. For the RPL model, we hold the price preference fixed, and assume other preferences to be normally distributed across the respondents. The key statistics of the RPL model suggest improvement over the MNL model. Furthermore, the log likelihood values at convergence (LL) and McFadden's R^2 increase from $S = 1$ to $S = 8$, indicating the existence of heterogeneity in consumer preferences.

The AIC suggests the optimal number of latent classes $S^* = 8$, while the BIC suggests $S^* = 4$. Moreover, the marginal change in AIC after $S = 4$ is considerably smaller than those before $S = 4$, suggesting that adding a segment to the 4-segment model does not result in much improvement. Consequently, $S^* = 4$ provides a parsimonious description of the latent class structure.

4.3. Characterising the segments

We present the segment membership parameters γ_s in Table 4. The levels of significance and effect sizes of these parameters jointly tell us how psychological factors relate to a specific latent class, or 'segment'. Particularly, we highlight the maximum effect size among all significant effects in each row. For each segment, we can identify the psychological factors which seem most important. We set the segment membership parameters for the first class (i.e. the reference class) $\gamma_1 = 0$ to identify the latent class model. Hence, the other three classes are described relative to the

reference class. The last row gives the share of each segment.

Table 4

Estimation of segment membership parameters

	Class 1	Class 2	Class 3	Class 4
intercept	0	–2.57***	–1.25*	–5.04***
		(0.60)	(0.50)	(0.80)
Food safety concerns (FS)	0	0.20*	**0.26*****	0.05
		(0.09)	(0.07)	(0.09)
Health consciousness (HL)	0	0.51***	**1.01*****	0.96***
		(0.10)	(0.09)	(0.11)
Environmental concerns (EC)	0	0.33**	-0.01	**0.97*****
		(0.10)	(0.09)	(0.13)
Price consciousness (PC)	**0**	–0.30***	–0.54***	–0.85***
		(0.08)	(0.07)	(0.08)
Share (%)	9.8	19.8	57.5	12.9

Notes: Standard errors are in parentheses. Significance levels: *: 0.05, **: 0.01, ***: 0.001.

We label each segment by comparing the segment membership parameters γ_s. The first segment represents *price conscious* consumers who show the least interest in food safety, health and environment ($\gamma_{1,FS}$, $\gamma_{1,HL}$ and $\gamma_{1,EC}$ are all negative). We label the second segment as *balanced thinking* consumers, because these consumers take all aspects into account without emphasising any specific aspect. All parameters in the second column are significant at $p = 0.05$, indicating that this consumer segment is concerned about all four aspects (food safety, health, environment and price). Furthermore, all four parameters $\gamma_{2,FS} = 0.2$, $\gamma_{2,HL} = 0.51$, $\gamma_{2,EC} = 0.33$ and $\gamma_{2,PC} = -0.3$ rank second or third in respective rows. These results suggest that the second consumer segment tries to balance all four aspects. The third segment, labeled as *health conscious* consumers, cares most about food safety and health related issues ($\gamma_{3,FS} = 0.26$ and $\gamma_{3,HL} = 1.01$ are the largest in respective rows). Furthermore, health conscious consumers do not care about price ($\gamma_{3,PC} = -0.54$ ranks third in the respective row) and show no interest in environmental issues (the coefficient $\gamma_{3,EC}$ is

not significant). The last segment, called *environment conscious* consumers, is most concerned about environmental issues ($\gamma_{4,EC} = 0.97$ is the largest in the corresponding row) but least about price ($\gamma_{4,PC} = -0.85$ is the smallest in the respective row). Environmental conscious consumers also show considerable interest in health issues ($\gamma_{4,HL} = 0.96$ ranks second in the respective row) but do not seem to care about food safety ($\gamma_{4,FS}$ is not significant at $p = 0.05$). This result appears somehow abnormal but may have some possible reasons. In a normal quality certification system, environmental protection standards are very stringent for the food production process. Hence, any food that meets the requirements of environmental protection will not face food safety problems. Therefore, environment conscious consumers may consider it as unnecessary to worry about food safety issues.

The last row in Table 4 shows the market share of each segment. The health conscious segment has the largest share (i.e. accounts for 57.5%), followed by the balanced thinking segment (19.8%) and the environment conscious segment (12.9%), while the price conscious segment has the smallest share (only 9.8%). This result provides a holistic view of consumer segmentation in the Chinese milk market.

4.4. Heterogeneous preferences

Estimated preference parameters and WTP are shown in Tables 5 and 6, respectively. We take the MNL model as the baseline. All parameters in the MNL model are significant at $p = 0.05$. Respondents are unlikely to choose the opt-out option since $\beta_{ASC} = 1.23 > 0$. Consumers prefer YiLi ($\beta_{YiLi} = 0.14$) and MengNiu ($\beta_{MengNiu} = 0.12$) to GuangMing, which is consistent with the ranking of market shares. For quality certifications, consumers prefer organic ($\beta_{Organic} = 0.63$) and green certifications ($\beta_{Green} = 0.39$) to hazard-free certification, which might reflect that standards of organic certification are more stringent than green certification. Furthermore, consumers prefer QR code traceability label ($\beta_{QR} = 1.2$), followed by numeric code traceability label ($\beta_{Numeric} = 1.13$). The price coefficient $\beta_{Price} = -0.4 < 0$, as we expected.

However, before we turn to the WTPs implied by these estimates, we note that price has a positive coefficient for the environment conscious segment, which means that

we cannot impute sensible WTPs for this segment. We might conjecture that price, for this particular segment, is being interpreted as a quality cue for environmentally friendly production, but we cannot infer this from our data and analysis.

Table 5

Preference estimates for each latent class

	MNL	Price conscious	Balanced thinking	Health conscious	Environment conscious
ASC	1.23***	4.35	-2.66*	1.33***	-0.85
	(0.09)	(2.36)	(1.20)	(0.14)	(0.50)
MengNiu	0.12**	0.37	0.40	0.28***	-0.43*
	(0.04)	(0.51)	(0.44)	(0.06)	(0.18)
YiLi	0.14***	-0.47	0.76	0.23***	-0.24
	(0.04)	(0.45)	(0.50)	(0.06)	(0.21)
Green	0.39***	1.22**	1.36**	0.38***	0.20
	(0.04)	(0.47)	(0.51)	(0.06)	(0.24)
Organic	0.63***	0.98	1.24**	0.57***	1.45***
	(0.04)	(0.55)	(0.48)	(0.06)	(0.20)
Numeric	1.13***	0.31	4.87***	0.93***	1.88***
	(0.04)	(0.42)	(0.82)	(0.07)	(0.26)
QR	1.20***	0.42	5.34***	0.99***	2.20***
	(0.04)	(0.51)	(1.03)	(0.07)	(0.25)
Price	−0.40***	−2.23***	−1.28***	−0.33***	**0.33*****
	(0.01)	(0.31)	(0.23)	(0.02)	(0.07)
Log-likelihood	−6,390		−5,995.5		
McFadden's R^2	0.16		0.21		

Notes: (1) Sample size is 6,219 choices from 691 respondents.
   (2) Standard errors are in parentheses. Significance levels: *: 0.05, **: 0.01, ***: 0.001.

Table 6 shows the consumers' WTP for each attribute based on the price coefficient for all but the environment conscious segment (12.9% of the total sample).

Table 6

WTP estimates for each latent classes

	MNL	Price conscious	Balanced thinking	Health conscious
MengNiu	0.31	0.16[a]	0.31[a]	**0.85**
YiLi	0.35	-0.21[a]	0.60[a]	**0.71**
Green	0.97	0.55	1.07	1.18
Organic	1.60	0.44[a]	0.97	1.76
Numeric	2.85	0.14[a]	**3.81**	2.84
QR	3.04	0.19[a]	**4.18**	3.05

Note: (1) [a]: insignificance at $p = 0.05$.

(2) We use CNY/250ml as price unit, that is, 2.12 \$/gal.

(3) We calculate WTP directly from Table 5 and thus omit the standard errors.

We take the preferences and WTP from the MNL model as a baseline to reveal the uniqueness of each segment.

- For price conscious consumers (9.8% of the respondents), only two coefficients are significant at $p = 0.05$: the coefficient of green certification 1.22 and the price coefficient –2.33. We may infer that price conscious consumers are mainly concerned about price, and pay limited attention to other attributes except green certification. Nevertheless, their WTP for green certification is below the average level (0.55 < 0.97, Table 6).

- Balanced thinking consumers (19.8% of the respondents) are more likely to choose the opt-out option (–2.66 for ASC), but do not prefer a specific brand. They care about quality certifications, traceability labels and price. These consumers have average WTP for quality certifications (1.07 CNY for green certification and 0.97 CNY for organic certification). Interestingly, they show very high WTP for traceability labels, which are approximately 1 CNY higher than the average. It is possible that this segment of consumers is concerned about all aspects (i.e. food safety, health, environment and price), which are satisfied

by food traceability systems and therefore show high WTP for traceability labels.

- For health conscious consumers (57.5% of the respondents), all coefficients are significant. Furthermore, most coefficients are similar to those of the MNL model, while brand preferences are strong.

- Environment conscious consumers (12.9% of the respondents) prefer GuangMing to MengNiu brand. They show special interest in organic certification, numeric and QR code traceability labels, with the coefficients being around twice of those in the MNL model. They may think that organic certification and traceability systems can provide valuable information on environmentally-friendly production methods. By contrast, green production may not meet their requirements on environment protection. Furthermore, the price coefficient 0.33 > 0 seems to contradict our traditional wisdom. A potential explanation might be from *signaling theory*. Environment conscious consumers prefer milk products with organic certification and traceability labels, which are usually associated with high costs. Therefore, price is considered as a quality signal by these consumers. Considering that we chose reasonable price levels in the choice experiment, we conjecture that this consumer segment prefers high but reasonable prices.

5. Discussions and Conclusions

We use a choice experiment and a latent class model to analyse Chinese consumer preferences and WTP for milk attributes. We identify four latent consumer segments as well as their characteristics and identify the source of heterogeneity by individual psychological factors.

We have three main findings. First, the MNL model suggests that consumers are willing to pay most for traceability labels, followed by quality certifications, and least for brands. Second, we identify four consumer segments in the milk market, that is, price conscious consumers (9.8% in market share), balanced thinking consumers (19.8%), health conscious consumers (57.5%), and environment conscious consumers (12.9%). We further identify that each consumer segment has its own psychological characteristics. Price conscious consumers care most about price and pay limited attention to food safety, health and environment. Health conscious consumers are

most concerned about food safety and health, focus less on price, and have little interest in environmental issues. Environment conscious consumers care most about environmental issues, care least about price, and have considerable interest in health issues. Balanced thinking consumers consider all aspects but do not prefer any specific aspect.

These findings provide managerial implications for dairy companies. To convey more added value to consumers, dairy companies should first focus on food traceability, followed by quality certification. Companies can improve their profits by targeting one or more consumer segments. As examples, they can build a strong milk brand to serve health conscious consumers who have the largest market share and the highest WTP for brands, or provide organic, traceable and high-priced milk to meet a presently limited environmentally conscious market segment.

A limitation of our study is the potential bias of our online sample, which is clearly not representative of the general Chinese population, but which we argue is likely to be reasonably representative of the portion of the population which currently regularly purchases milk. Our study also provides a promising direction for future research. The price coefficient of environment conscious consumers is positive, suggesting that this segment of consumers uses price as a quality signal, in which case signal consistency matters. Consequently, understanding consumer preferences for milk products from a signaling theory perspective will be worthwhile.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix 1: Measurement items for psychological factors

Table A1. Summary of reliability and validity

Table A2. Correlations of constructs

References

Aizaki, H. 'Basic functions for supporting an implementation of choice experiments in R', *Journal of Statistical Software*, Vol. 50(1), (2012) pp. 1–24.

Alford, B. L. and Biswas, A. 'The effects of discount level, price consciousness and sale proneness on consumers' price perception and behavioral intention', *Journal of Business Research*, Vol. 55(9), (2002) pp. 775–783.

Apostolidis, C. and McLeay, F. 'Should we stop meating like this? Reducing meat consumption through substitution', *Food Policy*, Vol. 65, (2016) pp. 74–89.

Ares, G., Giménez, A. and Deliza, R. 'Influence of three non-sensory factors on consumer choice of functional yogurts over regular ones', *Food Quality and Preference*, Vol. 21(4), (2010) pp. 361–367.

Aung, M. M. and Chang, Y. S. 'Traceability in a food supply chain: Safety and quality perspectives', *Food Control*, Vol. 39, (2014) pp. 172–184.

Bai, J., Zhang, C. and Jiang, J. 'The role of certificate issuer on consumers' willingness-to-pay for milk traceability in China', *Agricultural Economics*, Vol. 44(4-5), (2013) pp. 537–544.

Boxall, P. C. and Adamowicz, W. L. 'Understanding heterogeneous preferences in random utility models: A latent class approach', *Environmental and Resource Economics*, Vol. 23(4), (2002) pp. 421–446.

Brooks, K. and Lusk, J. L. 'Stated and revealed preferences for organic and cloned milk: Combining choice experiment and scanner data', *American Journal of Agricultural Economics*, Vol. 92(4), (2010) pp. 1229–1241.

Cheng, L., Yin, C. and Chien, H. 'Demand for milk quantity and safety in urban China: Evidence from Beijing and Harbin', *Australian Journal of Agricultural and Resource Economics*, Vol. 59(2), (2015) pp. 275–287.

Costanigro, M., Deselnicu, O. and Kroll, S. 'Food beliefs: Elicitation, estimation and implications for labeling policy', *Journal of Agricultural Economics*, Vol. 66(1), (2015) pp. 108–128.

Dhar, T. and Foltz, J. D. 'Milk by any other name ... consumer benefits from labeled

milk', *American Journal of Agricultural Economics*, Vol. 87(1), (2005) pp. 214–228.

Fujii, S. 'Environmental concern, attitude toward frugality, and ease of behavior as determinants of pro-environmental behavior intentions', *Journal of Environmental Psychology*, Vol. 26(4), (2006) pp. 262–268.

Fuller, F., Huang, J., Ma, H. and Rozelle, S. 'Got milk? The rapid rise of China's dairy sector and its future prospects', *Food Policy*, Vol. 31(3), (2006) pp. 201–215.

Garrod, G., Ruto, E., Willis, K. and Powe, N. 'Investigating preferences for the local delivery of agri-environment benefits', *Journal of Agricultural Economics*, Vol. 65(1), (2014) pp. 177–190.

He, H., Kukar-Kinney, M. and Ridgway, N. M. 'Compulsive buying in China: Measurement, prevalence, and online drivers', *Journal of Business Research*, Vol. 91, (2018) pp. 28–39.

Hughner, R. S., McDonagh, P., Prothero, A., Shultz, C. J. and Stanton, J. 'Who are organic food consumers? A compilation and review of why people purchase organic food', *Journal of Consumer Behaviour*, Vol. 6(2-3), (2007) pp. 94–110.

Jedidi, K. and Jagpal, S. 'Willingness to pay: Measurement and managerial implications' In *Handbook of Pricing Research in Marketing* (Cheltenham, England, Edward Elgar Publishing, 2009).

Jia, N., Li, L., Ling, S., Ma, S. and Yao, W. 'Influence of attitudinal and low-carbon factors on behavioral intention of commuting mode choice cross-city study in China', *Transportation Research Part A: Policy and Practice*, Vol. 111, (2018) pp. 108–118.

Jin, S., Zhang, Y. and Xu, Y. 'Amount of information and the willingness of consumers to pay for food traceability in China', *Food Control*, Vol. 77, (2017) pp. 163–170.

Kildegaard, H., Olsen, A., Gabrielsen, G., Møller, P. and Thybo, A. 'A method to measure the effect of food appearance factors on children's visual preferences', *Food Quality and Preference*, Vol. 22(8), (2011) pp. 763–771.

Lancaster, K. J. 'A new approach to consumer theory', *Journal of Political Economy*, Vol. 74(2), (1966) pp. 132–157.

Liu, R., Pieniak, Z. and Verbeke, W. 'Consumers' attitudes and behaviour towards safe food in China: A review', *Food Control*, Vol. 33(1), (2013) pp. 93–104.

Magnusson, M. K., Arvola, A., Hursti, U.-K. K., Åberg, L. and Sjödén, P.-O. 'Choice of organic foods is related to perceived consequences for human health and to environmentally friendly behaviour', *Appetite*, Vol. 40(2), (2003) pp. 109–117.

McFadden, D. 'Econometric models for probabilistic choice among products', *Journal of Business*, Vol. 53(3) (1980) pp. S13–S29.

McFadden, D. 'Economic choices', *American Economic Review*, Vol. 91(3), (2001) pp. 351–378.

McFadden, D. 'The measurement of urban travel demand', *Journal of Public Economics*, Vol. 3(4), (1974) pp. 303–328.

Michaelidou, N. and Hassan, L. M. 'The role of health consciousness, food safety concern and ethical identity on attitudes and intentions towards organic food', *International Journal of Consumer Studies*, Vol. 32(2), (2008) pp. 163–170.

Miklavec, K., Pravst, I., Grunert, K. G., Klopčič, M. and Pohar, J. 'The influence of health claims and nutritional composition on consumers' yoghurt preferences', *Food Quality and Preference*, Vol. 43, (2015) pp. 26–33.

Nasir, V. A. and Karakaya, F. 'Underlying motivations of organic food purchase intentions', *Agribusiness*, Vol. 30(3), (2014) pp. 290–308.

Olynk, N. J., Wolf, C. A. and Tonsor, G. T. 'Production technology option value: The case of rbST in Michigan', *Agricultural Economics*, Vol. 43, (2012) pp. 1–9.

Pouta, E., Heikkilä, J., Forsman-Hugg, S., Isoniemi, M. and Mäkelä, J. 'Consumer choice of broiler meat: The effects of country of origin and production methods', *Food Quality and Preference*, Vol. 21(5), (2010) pp. 539–546.

Rao, V. R. *Applied Conjoint Analysis* (Berlin, Heidelberg, Springer, 2014).

Rosseel, Y. 'Lavaan: An R package for structural equation modeling', *Journal of Statistical Software*, Vol. 48(2), (2012) pp. 1–36.

Sarrias, M. and Daziano, R. 'Multinomial logit models with continuous and discrete individual heterogeneity in R: The gmnl package', *Journal of Statistical Software*, Vol. 79(2), (2017) pp. 1–46.

Sheremet, O., Healey, J. R., Quine, C. P. and Hanley, N. 'Public preferences and willingness to pay for forest disease control in the UK', *Journal of Agricultural Economics*, Vol. 68(3), (2017) pp. 781–800.

Train, K. E. *Discrete Choice Methods with Simulation* (Cambridge: Cambridge University Press).

Wang, J., Chen, M. and Klein, P. G. 'China's dairy united: A new model for milk production', *American Journal of Agricultural Economics*, Vol. 97(2), (2015) pp. 618–627.

Wolf, C. A., Tonsor, G. T. and Olynk, N. J. 'Understanding U.S. consumer demand for milk production attributes', *Journal of Agricultural and Resource Economics*, Vol. 36(2), (2011) pp. 326 –342.

Wu, L., Wang, H., Zhu, D., Hu, W. and Wang, S. 'Chinese consumers' willingness to pay for pork traceability information: the case of Wuxi', *Agricultural Economics*, Vol. 47(1), (2016) pp. 71–79.

Wu, L., Yin, S., Xu, Y. and Zhu, D. 'Effectiveness of China's organic food certification policy: Consumer preferences for infant milk formula with different organic certification labels', *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, Vol. 62(4), (2014) pp. 545–568.

Xu, L. and Wu, L. 'Food safety and consumer willingness to pay for certified traceable food in China', *Journal of the Science of Food and Agriculture*, Vol. 90(8), (2010) pp. 1368–1373.

Yamamoto, Y., Iwamoto, H., Masuda, K. and Managi, S. 'Valuing the influence of underlying attitudes and the demand for organic milk in Japan', *Agricultural Economics*, Vol. 39(3), (2008) pp. 339–348.

Yoo, C. W., Parameswaran, S. and Kishore, R. 'Knowing about your food from the farm to the table: Using information systems that reduce information asymmetry and health risks in retail contexts', *Information & Management*, Vol. 52(6),

(2015) pp. 692–709.

Yu, X. 'Productivity, efficiency and structural problems in Chinese dairy farms', *China Agricultural Economic Review*, Vol. 4(2), (2012) pp. 168–175.