Enhancing Sustainability Marketing Strategies in Online Transactions: A Categorical Factorization Approach

Zoe Ziqi Zhong*

Department of Management, London School of Economics and Political Science, London WC2A 2AE, United Kingdom

Guanzhi Wang

School of Business, University of Leicester, Leicester LE2 1RQ, United Kingdom

ABSTRACT

The rapid advancement of internet technology has revolutionised business operations, particularly in the realm of online purchasing. Companies are increasingly embracing online platforms to showcase their products and services, as consumers are drawn to the cost-effectiveness and convenience of online shopping. This study employs Categorical Factor Analysis (FA) to analyse click data from JD.com, with the objective of investigating the relationship between online consumer characteristics and purchase history. The primary focus is to identify the influential factors and latent variables that impact the average transaction value of online orders. The findings of this research highlight the pivotal role played by sustainability marketing strategies, along with other factors such as cognitive quality, financial capacity, group demand levels, and online shopping convenience. Interestingly, the study reveals that the utilization of coupons, often associated with sustainability marketing efforts, may have unintended consequences on overall transaction values. Each identified factor demonstrates a significant 99% level of significance ($p < 0.01$) within the regression analysis. This study contributes valuable insights into online consumer behaviour, emphasising the importance of sustainability marketing strategies in shaping consumer choices and guiding companies in the development of environmentally conscious pricing and allocation strategies.

CCS CONCEPTS

• **Applied computing**; • **Electronic commerce**; • **Online shopping** → Computing methodologies; • **Machine learning**; • **Machine learning approaches**; • **Factorization methods** → Human-centered computing; • **Human computer interaction (HCI)**; • **Empirical studies in HCI**;

KEYWORDS

Consumer Behaviour, Digital Marketing, E-Commerce, Promotional Strategies

*Corresponding author's e-mail: Z.Zhong6@lse.ac.uk

1 INTRODUCTION

The expansive capability of the Internet has achieved global ubiquity, reshaping its primary focus from mere information exchange to a powerful platform that not only facilitates business transactions and shopping experiences but also serves as a catalyst for sustainability-driven endeavors [1]. The evolution of Internet technology has effectively transcended the temporal and spatial limitations associated with traditional models, solidifying the Internet as the predominant marketplace where sustainability marketing plays an increasingly pivotal role. Beyond its convenience and efficiency, the Internet has become the preferred platform for a diverse array of activities, including sustainable shopping, banking, and bill payments. From a commercial perspective, the Internet has revolutionized how businesses market, sell, and engage with customers, with sustainability marketing strategies gaining prominence. The surge in online shopping, especially for sustainable products, can be attributed to its multifaceted client benefits and strengthened security protocols [2], underscoring the pressing need for a comprehensive understanding by both vendors and consumers to optimize their online spending while aligning with sustainability principles.

JD.com, a technological pioneer in China, specializes not only in supply chain solutions and logistics but also places a significant emphasis on sustainability and green initiatives. Their commitment extends to digital transformation through the integration of advanced technologies like 5G, AI, big data, cloud computing, and IoT, while concurrently incorporating sustainability marketing into their business model. This seamless technological integration has empowered JD.com to amass a substantial wealth of user-centric data, encompassing user demographics, order details, and access modalities, offering valuable insights into the variables that influence order values, particularly those associated with sustainability-driven choices. This study postulates a correlation between these influential factors and order transaction amounts, employing Factor Analysis (FA) to elucidate this relationship and identify predominant features, including sustainability marketing strategies. Consequently, the insights gleaned from this research can prove instrumental not only in the formulation of pricing strategies and the enhancement of average transaction values on e-commerce platforms but also in the development of precise predictive models that incorporate sustainability as a critical factor.

This paper is organized as follows: The subsequent section provides a contextual background, offering an insight into the realm of online purchasing with a sustainability marketing perspective. The third section outlines the methodology, presenting the analytical processes, research findings, and regression analyses of the

study, highlighting the significance of sustainability in the results. The final section concludes the paper, drawing inferences from the study conducted, emphasizing the pivotal role of sustainability marketing in shaping online consumer behavior and offering recommendations for sustainable e-commerce strategies.

2 BACKGROUND

2.1 Online Shopping and the Role of Sustainability Marketing

China's inaugural online purchase in 1998 marked the inception of its remarkable online shopping journey. As of 2020, China's annual online market sales had soared to 10.8 trillion yuan, constituting a substantial 21.9% of the total societal sales, with projections indicating a further increase to 1.122 trillion yuan in 2021 [3]. On a global scale, online shopping has transcended periodic activity to become an integral part of daily life, culminating in a staggering $4.28 trillion in sales in 2020, with expectations of further growth to $5.4 trillion in 2022. The onset of the COVID-19 pandemic further solidified the indispensability of online shopping due to widespread quarantine and lockdown measures [4].

The concept of online shopping, initially conceived in the mid-1990s by visionaries Doddy and Davidson (1967), has now materialized as a fundamental aspect of modern consumerism, thanks to the continuous evolution of Internet technology. This transformation has not only reshaped consumer habits but also revolutionized sales methodologies, necessitating a shift from traditional seller-centric strategies to a more holistic approach centered around consumer behavior and data-driven insights [5]. It is within this evolving landscape that sustainability marketing finds its foothold, as consumers increasingly seek ethically and environmentally responsible products and businesses. This evolution presents not only challenges but also substantial opportunities to analyze customer data and identify influential factors in online sales, with sustainability marketing strategies playing a pivotal role in shaping purchasing decisions. The next section outlines the methodology used in this study to pinpoint the determinants of online transaction value, with particular attention to the sustainability aspect.

2.2 Factor Analysis (FA)

Factor Analysis (FA) is a statistical approach employed to explore the underlying constructs affecting observed variables and is pivotal for dimensionality reduction [6]. It provides profound insights into correlations among variables, integrating several correlated variables into fewer factors, revealing how independent factors control intricate relationships among measured indicators [7].

The two types of factor analysis, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) have different targets. Particularly, this study focuses on EFA, which attempts to discover the latent structure impacting the observed variables. Let ξ_j be latent variables (i.e. factors) and x_i be the observed variables (i.e. items) that acts as measures of factors. By conventional standardisation of the factors, they are related such that $X = +$ in the form of matrices, of which each row is

$$x_i = \sum_{j=1}^{q} \lambda_{ij}\xi_j + \delta_i,$$

where λ_{ij} is the factor loading for the i-th item and j-th factor, and δ_i is normally distributed with variance called unique variance, denoted by $\delta_i \sim N(0, \theta_{ii})$ for $i = 1, \ldots, p$. To conduct the EFA, firstly calculate the correlation matrix of $x_1, \ldots, x_p$. The most commonly used is Pearson's correlation, expressed as a square matrix

$$R = (r_{k, l})_{k, l=1, \ldots p},$$

where

$$r_{k, l} = \frac{\sum_{n=1}^{N} (x_{k, n} - \overline{x_k})(x_{l, n} - \overline{x_l})}{\sqrt{\sum_{n=1}^{N} (x_{k, n} - \overline{x_k})^2} \sqrt{\sum_{n=1}^{N} (x_{l, n} - \overline{x_l})^2}} \quad (1)$$

and $\bar{x}_i$ is the mean over N observed value of each entry, $i = k, l$ from 1, $\ldots p$. But in this article an alternative correlation will be introduced specifically for the existence of categorical observed variables. Then choose a factor extraction, such as principal component, maximum likelihood (ML), principal axes factor and generalised least squares, to estimate the factor loadings and the unique variances and to choose the factors accounting for the variance the most. Lastly, rotate the factors obtained to find a more interpretable factor solution.

This paper will compare the result of categorical EFA to achieve better the research aim, which is to derive the most influential characteristics of the online transaction value.

3 METHODOLOGY AND RESULTS

3.1 Data Source and Variables Selection

The analysis utilizes datasets from JD.com, provided for the 2020 MSOM Data Driven Research Challenge, detailed by [8]. These datasets, from March 2018, encapsulate the entire customer experience cycle from product browsing to reception, focusing on a specific consumable category, and are devoid of major promotions or holidays influences. They are structured into seven tables detailing SKUs, users, orders, clicks, among others.

In summary, the user table describes 457, 298 users, showcasing a predominance of female customers (over 60%) and majorly comprises individuals aged 16-45, making up around 75% of the users. The order table consists of 486, 928 unique orders, and the click table records over 20 million click instances from 2.5 million users, across various platforms.

For analyzing transaction volumes, understanding user shopping behavior is crucial. The relevant variables are grouped logically to understand correlations between them and identify key contributors to shopping behavior. The twelve selected variables are categorized into order-related, user characteristics related to online shopping accounts, and those related to demographics like age and education level. Refer to Table 1 for detailed groupings and meanings of selected variables.

3.2 Pearson's vs Polychoric Correlation

This section aims to assess the appropriateness of Pearson's and polychoric correlations for EFA involving a mixture of numerical and categorical variables.

Table 1: List of variables

Group	Variables	Meanings
Key Measure	*final_transaction*	The final transaction volume of the user paid for the item in the order.
Order Feature	*direct_discount_rate*	The discount rate when the user directly purchased the product in that order.
	coupon_discount_rate	The discount rate applied when the user used a coupon in that order.
	bundle_discount_rate	The discount rate applied when the products provide a bundle discount in that order.
	PC	Whether the order was completed on the PC channel (coded as 1) or the Mobile channel (coded as 0).
User's Online	*plus*	Whether the user is plus membership or not.
Shopping	*user_level*	A number ranging from 1~5 indicated how much is the total
Characteristics		purchase of that user in the past.
	purchase_power	The purchasing power estimation of that user
User's Personal	*age_level*	Age level of that user, there are five age levels: "16~25", "26~35",
Characteristics		"36~45", "46~55" and ">= 56", coded from 1 to 5 respectively.
	female	Whether the user is female or not.
	single	Whether the user is single or not.
	education	Indicating the education level of that user, ranged from 1: less than high school to 4: holds a graduate degree.
	city_level	Indicating the city development level of the user according to the order address, ranging from 1: rural small cities to 5: the most developed cities such as Beijing and Shanghai.

Table 2: Matrix of Pearson's correlation

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) direct_discoun~e	1.000											
(2) coupon_discoun~e	-0.065	1.000										
(3) bundle_discoun~e	-0.106	-0.021	1.000									
(4) PC	-0.018	-0.004	-0.000	1.000								
(5) plus	-0.018	0.065	0.012	0.025	1.000							
(6) user_level	-0.030	0.025	0.022	0.086	0.456	1.000						
(7) purchase_power	-0.020	0.009	0.016	0.022	0.217	0.330	1.000					
(8) age_level	0.004	0.035	-0.003	0.106	0.071	0.218	0.062	1.000				
(9) female	0.009	0.000	0.008	-0.061	-0.199	-0.255	-0.122	-0.023	1.000			
(10) single	-0.002	-0.052	-0.010	-0.043	-0.183	-0.299	-0.111	-0.353	-0.013	1.000		
(11) education	-0.040	0.015	-0.000	0.070	0.344	0.480	0.162	0.117	-0.087	-0.176	1.000	
(12) city_level	-0.037	-0.015	0.001	0.060	0.099	0.174	0.058	0.051	-0.007	-0.030	0.192	1.000

Pearson's Correlation-Based Principal Component EFA (Unrotated): The initial phase in dimensionality reduction involves analyzing the correlation among data. Table 2 represents the correlation matrix based on Equation 1) for the variables. Given the dataset involves millions of clicks, even correlations above 10% are significant and marked.

Observations: Variables related to users' online shopping and personal features are notably correlated, hinting at personal characteristics as potential significant determinants of online shopping behavior, aligning with [9]. They likely fall under the same factor. 'User level' and 'Purchase power' display correlations with numerous other variables, reflecting their information-rich nature and suitability as key consumer feature measures, corroborating their significance as per JD.com.

Table 3: KMO and Bartlett's test

Bartlett test of sphericity	
Chi-square	1.24e+06
Degrees of freedom	66
p-value	0.000
(H0: variables are not intercorrelated)	
Kaiser-Meyer-Olkin Measure of Sampling Adequacy	
KMO	0.712

Table 3 illustrates the Mayer-Kaiser-Olkin (KMO) score at 0.712 (>0.5), signifying the acceptability of factor analysis for this dataset. The results of Pearson's Correlation-Based Principal Component EFA (Unrotated) are not shown because they were not ultimately used in this study.

Pearson's Correlation-Based Principal Component EFA (Unrotated): Polychoric correlation is advantageous over Pearson's correlation in factor analysis as it considers the latent variables' underlying distribution leading to categorical ones, assuming a multivariate Normal distribution, and remains robust even when this assumption is violated [10]. This has been substantiated by studies such as [11], demonstrating its superior performance in ML EFA over Pearson's, providing more precise structure estimations under varying conditions.

In this subsection, factor analysis utilizes a polychoric correlation matrix, depicted in Table 4, to integrate categorical variables more effectively.

The correlations amplify in comparison to Pearson's, for instance, 'user_level' and 'plus' is 0.456 with Pearson's and 0.654 with polychoric. Importantly, polychoric correlation maintains continuous numerical variables intact, offering insight into the latent distribution for the categorical variables. This is coherent with our endeavor to comprehend the latent distribution for categorical variables while preserving the integrity of continuous numerical variables. Based

on this matrix, EFA was reiterated using principal-component factors with eigenvalues in Table 4.

3.3 Rotation Selection

Factor Analysis results can be rotated to yield better interpretability. The initial EFAs were unrotated, serving as reference for evaluating subsequent rotational choices: varimax and promax.

Varimax Rotation (Polychoric correlation-based Principal Component EFA): Varimax, an orthodox rotation, assumes factors are orthogonal and hence uncorrelated. It optimizes variance within each column of the factor loadings matrix, distinguishing variables by factors. Varimax is prevalent due to its computational simplicity, with a 47% application rate in studies over six years [12]. However, its simplicity can sometimes compromise the interpretability of the results [13]. The results of Polychoric correlation-based Principal Component EFA (Varimax rotated) are not shown because they were not ultimately used in this study.

Promax Rotation (Polychoric correlation-based Principal Component EFA): Promax, an oblique rotation, allows for correlated variables and is purported to yield simpler structures. Its results are depicted in Table 5 and Figure 1.

Comparing the unrotated, varimax and promax scatter plots and loading tables reveals the superior interpretability of rotated results. Promax rotations are chosen for subsequent discussions due to their allowance for correlations, aiding in more nuanced interpretations. The aim is not just data reduction, but also summarizing key influences. Factor 1 is indicative of consumer cognitive quality and spending power; Factor 2, influenced by age and marital status, represents average consumer demand [14]; Factor 3 encapsulates online shopping convenience, contingent on purchase platform and city development level [15]; Factor 4 is synonymous with the impact of general marketing, contrasting the applications of direct and bundle discounts; and Factor 5 represents the additional effort consumers exert for discounts through coupons. The interpretations of all five factors are consolidated in Table 6.

Table 4: Polychoric Correlation Matrix

	di-rect_	coupon_	bun-dle_	PC	plus	user_level	pur-chase_	age_level	educa-tion	city_level	female	single
direct_	1											
coupon_	-0.065	1										
bundle_	-0.106	-0.021	1									
PC	-0.845	-0.011	-0.0	1								
plus	-0.026	0.083	0.016	0.078	1							
user_level	-0.032	0.027	0.024	0.218	0.654	1						
pur-chase_	-0.242	0.013	0.021	0.064	0.391	0.412	1					
age_level	0.002	0.041	0.001	0.253	0.148	0.305	0.104	1				
educa-tion	-0.044	0.018	0.0	0.192	0.543	0.585	0.214	0.182	1			
city_level	-0.039	-0.017	0.001	0.157	0.148	0.203	0.073	0.072	0.232	1		
female	0.012	0.001	0.012	-0.175	-0.34	-0.361	-0.284	-0.038	-0.135	-0.01	1	
single	-0.003	-0.066	-0.012	-0.127	-0.314	-0.392	-0.166	-0.524	-0.257	-0.046	-0.021	1

Table 5: Factor loadings and unique variances for Polychoric correlation-based Principal Component EFA (Promax rotated)

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Uniqueness
direct_dis~e	0.010	0.119	-0.229	-0.613	-0.335	0.386
coupon_dis~e	0.067	0.048	-0.087	-0.091	0.876	0.238
bundle_dis~e	0.011	0.075	-0.150	0.844	-0.214	0.268
PC	-0.007	0.202	0.626	0.013	-0.245	0.485
plus	0.849	0.030	-0.034	-0.012	0.117	0.267
user_level	0.783	0.194	0.109	0.010	-0.008	0.230
purchase_p~r	0.665	-0.006	-0.199	0.070	-0.026	0.563
age_level	-0.013	0.841	0.070	0.006	-0.039	0.270
education	0.589	0.089	0.295	-0.032	0.081	0.454
city_level	0.058	-0.134	0.778	-0.065	0.027	0.398
female	-0.621	0.282	-0.013	0.022	0.205	0.563
single	-0.169	-0.826	0.121	-0.016	-0.095	0.250
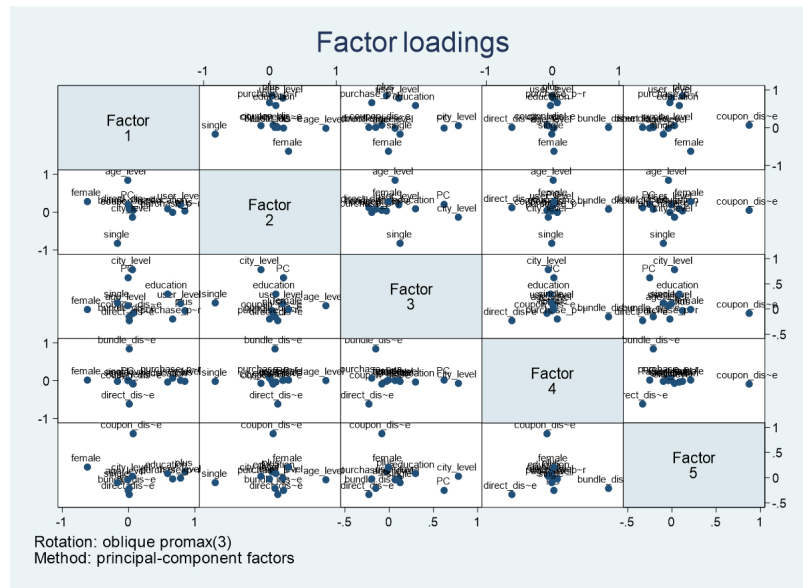
Figure 1: Factor loadings for Polychoric correlation-based Principal Component EFA (Promax rotated)

Table 6: Polychoric Correlation Matrix

Factor	Heavy Loading Variables (loading >0.6)	Categorisation
1	*plus*, *user_level*, *purchase_power*, *female*	General cognitive quality and spending power index of the consumer
2	*age_level*, single	The average consumer demand index for online shopping influenced by age and marital status
3	*PC*, *city_level*	Consumer's online shopping convenience index influenced by city class and ordering platform
4	*direct_discount_rate*, *bundle_discount_rate*	The power of marketing campaign
5	*coupon_discount_rate*	The measure of extra effort for getting a discount: by using a coupon

Table 7: Polychoric Correlation Matrix

	(1) transaction_value	(2) transaction_value	(3) transaction_value	(4) transaction_value	(5) transaction_value
f1	6.444***	11.644***	11.217***	11.77***	12.171***
	(.089)	(.109)	(.109)	(.109)	(.108)
f2		-4.049***	.339***	1.082***	1.026***
		(.047)	(.085)	(.087)	(.086)
f3			1.956***	.077	2.028***
			(.031)	(.054)	(.055)
f4				.906***	.993***
				(.021)	(.021)
f5					-1.347***
					(.005)
_cons	76.069***	72.712***	65.88***	68.865***	66.455***
	(.227)	(.225)	(.248)	(.258)	(.256)
Observations	1149088	1149088	1149088	1149088	1149088

Robust standard errors are in parentheses
*** p<.01, ** p<.05, * p<.1

3.4 Regression Analysis

Regression analyses were performed to validate the extracted factors and investigate their influence on the transaction values of online purchases (Table 7). Linear regressions on five factors were conducted, with final transaction value as the dependent variable. All factors were significant at the 99% level (p<0.01).

Examination of the regression coefficients yields definitive conclusions. Elevated average consumer perceptions, purchasing power, and average consumer demand of a consumer group significantly impact spending amounts. Further, a consumer is likely to spend more in a more amenable shopping environment [f3]. Marketing campaigns [f4] do foster spending, albeit with less efficacy compared to the prior mentioned factors. Notably, the presence of a coupon discount [f5] seemingly detracts from a consumer's willingness to pay, aligning with the findings of [16]. This suggests two plausible explanations; the required additional effort to obtain the coupon or the perceived lack of cost-effectiveness when the coupon is unavailable, causing consumers to abandon purchases.

4 CONCLUSION

Online shopping has become the preferred choice for consumers, a trend that has accelerated significantly during the pandemic. This research, which leveraged unique datasets from JD.com and employed PCA and categorical FA, has delved into the factors influencing online consumer behavior and their impact on transaction values. One of the principal findings of this study is the significance of sustainability marketing strategies, which emerged as a key factor affecting consumer choices and transaction values. All factors included in the final model exhibit a high level of significance at 99% (p<0.01).

The implications of this research extend to guiding future investigations into the influence of these factors and providing strategic recommendations for e-commerce platforms. It emphasizes the importance of improving the environmental convenience offered by online platforms and suggests caution when implementing marketing campaigns, especially those involving coupon issuance. The potential negative effects of coupons on consumer purchase intentions, which could outweigh the benefits of discounts, underscore the need for a more nuanced approach to sustainability marketing.

However, it is essential to acknowledge the limitations of this study, particularly the unexplored interactions among variables due to constraints in factor analysis and the imprecise classification of factors based on existing literature and data distribution. Future research endeavors can focus on refining factor classification using multi-dimensional datasets and conducting lab experiments to uncover causal relationships, ultimately providing a deeper understanding and practical insights for optimising e-commerce strategies with sustainability marketing at the forefront.

REFERENCES

[1] McKinsey Global Institute. 2016. DIGITAL GLOBALIZATION: THE NEW ERA OF GLOBAL FLOWS.
[2] Kim, Young Eun, and Youn-Kyung Kim. 2004. Predicting online purchase intentions for clothing products. In European Journal of Marketing 38, 7, 2004, 883–97.
[3] Jiang, Yawei, Zhenhua Cai, and Huali Cai. 2022. Big Data Research on China's Service Industry Under the COVID-19 Epidemic. In 2021 International Conference on Big Data Analytics for Cyber-Physical System in Smart City, 2022, 1007–13.
[4] Gopinath, Vidya. 2020. CONSUMER BEHAVIOR TRENDS DURING COVID-19 PANDEMIC. In IJSDR2009050 International Journal of Scientific Development and Research (2020).
[5] Nasidi, Qaribu Yahaya, Muhamad Fazil Ahmad, Jamilu Abdulkadir, and Murtala Garba. 2022. Analysing the Mediating Effect of Social Media on Online Shopping Using Partial Least Square. In Online Journal of Communication and Media Technologies 12, 2 (2022), e202213.
[6] Joliffe, I.T., and B.J.T. Morgan. 1992. Principal component analysis and exploratory factor analysis. In Statistical Methods in Medical Research 1, 1 (1992), 69–95.
[7] Stewart, David. 2001. Difference between Principal Components and Factor Analysis. In Journal of Consumer Psychology 10, 1/2 (2001), 75–76.
[8] Shen, Max, Christopher S. Tang, Di Wu, Rong Yuan, and Wei Zhou. 2020. JD.com: Transaction Level Data for the 2020 MSOM Data Driven Research Challenge. In Manufacturing & Service Operations Management (2020).
[9] Park, Chung-Hoon, and Young-Gul Kim. 2003. Identifying key factors affecting consumer purchase behavior in an online shopping context. In International Journal of Retail & Distribution Management 31, 1 (2003), 16–29.

[10] Coenders, G., Saris, W., Satorra, A. 1997. Alternative approaches to structural equation modeling of ordinal data: a Monte Carlo study. In Struct. Equ. Model. 4, 4 (1997), 261–282.

[11] Holgado–Tello, F.P., Chacón–Moscoso, S., Barbero–García, I., Vila–Abad, E. 2010. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. In Quality & Quantity 44, 1 (2010), 153-166.

[12] Osborne, J.W., Costello, A.B., Kellow, J.T. 2008. Best Practices in Exploratory Factor Analysis. In J.W. Osborne (Ed.), Best Practices in Quantitative Methods (2008), 205-213. Thousand Oaks, CA: Sage Publishing.

[13] Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., Strahan, E.J. 1999. Evaluating the use of exploratory factor analysis in psychological research. In Psychological Methods 4, 3 (1999), 272.

[14] Forsythe, Sandra M., Bailey, Anne W. 1996. Shopping Enjoyment, Perceived Time Poverty, and Time Spent Shopping. In Clothing and Textiles Research Journal 14, 3 (1996), 185–91.

[15] Jun, Seung-Pyo, and Do-Hyung Park. 2016. Consumer information search behavior and purchasing decisions: Empirical evidence from Korea. In Technological Forecasting and Social Change 107 (2016), 97–111.

[16] Gao, F., & Chen, J. 2014. The Role of Discount Vouchers in Market with Customer Valuation Uncertainty. In Production and Operations Management 24, 4 (2014), 665–679.