Journal of Business Research 69 (2016) 2101–2107

Contents lists available at ScienceDirect

Journal of Business Research

Consumer behavior on cashback websites: Network strategies☆

María Teresa Ballestar [a], Pilar Grau-Carles [b], Jorge Sainz [b,*]

[a] Open University of Catalonia, Tibidabo Avenue 39-43, Barcelona 08035, Spain
[b] Applied Economics Department, Rey Juan Carlos University, Paseo de los Artilleros, s.n., Madrid 28032, Spain

ARTICLE INFO

Article history:
Received 1 August 2015
Received in revised form 1 October 2015
Accepted 1 November 2015
Available online 30 December 2015

Keywords:
Cashback
Electronic commerce
Affiliation marketing
PLS-SEM

ABSTRACT

The size of cashback sites, both in terms of users and business, has grown considerably over the last decade. This article presents a complete analysis of the behavior of the users of the webs both in terms of transactions, and navigation and registration on cashback sites by using a large sample of one of the largest European sites. The study also presents a first analysis on the structure of the sites. An analysis using Partial Least Squares Structural Equation Modelling shows that the volume of the user's network, the diversification of the navigation, and the size of the transactions are relevant to the decision of the consumer and to his or her engagements on the affiliate merchants. These results represent a first step on the understanding of these marketing strategies and open new areas of research.

1. Introduction

The introduction of the information and communication technologies (ICTs) has rapidly transformed traditional commerce. This situation has affected the strategies of merchants to rapidly advertise their products and has transformed traditional marketing strategies to adapt them to the new environment. One of these strategies, cashback mechanisms, is significantly growing as a marketing instrument, transforming the traditional mail-in rebates in a new internet service. Cashback mechanisms also allow for development of a new business, the cashback webs. These sites aggregate merchants that offer those rebates, facilitating the transaction to the consumer. Thus, these webs share the revenue of the merchant's profit obtained through the advertisement transaction in an effort to increase sales and attract consumers' attention.

The size of the market has been rapidly growing. Different estimates show that these sites accumulate at least 100 million users in Europe and North America, generating a global business of 2.500 million dollars. Top sites, as rebates in the US, recently acquired by the Japanese giant Rakuten, Topcashback, Quidco, Fanli, or Beruby, generate thousands of daily transactions and generate consumers important rebates that can amount to hundreds of dollars every year.

This growth has also attracted the interest of academics on the area, although it is still a fairly new and open field. Most of the initial literature on the area focuses on modelling the cashback rebates deriving them from the traditional brick and mortars counterparts. Jain (2007) focuses on the business models of the search engines as profit producers to merchants and how to share of that surplus. Chen, Ghosh, Mcafee, and Pennock (2008) study the short- and long-term properties of the cashback rebates on conversion rates and profitability. Relevant literature (and patents) also exists on the mathematical properties of the consumer networks on cashback sites (Fu, Chen, Qin, & Guo, 2013), trying to explode its profitability, or defining the optimal strategy of the merchants on those sites (Ho, Ho, & Tan, 2013). Empirically, the only relevant study is by Vana, Lambrecht, and Bertini (2015), focusing on the profitability of cashback sites, and pointing out that cashback payments increase the likelihood of repeated purchases and their amount.

Cashback behavior study closely builds upon affiliate marketing literature. As Duffy (2005) points out, the irruption of e-commerce introduced a new system with no rules that requires research, particularly, on the creation of communities of affiliates both in large and small networks of sites, offering a win–win relationship for both sides of the market. Rust and Chung (2006) show different possible relations among services in an information economy, whereas Libai, Biyalogorsky, and Gerstner (2003) and Homburg, Droll, and Totzek (2008) focus on "how," describing the economics of the different affiliation methods and their profitability, and the benefits in the short and long term for the companies.

☆ The authors are grateful to contributions from participants in the Multidisciplinary Seminar at the Internet Interdisciplinary Institute, IN3 (Barcelona, Spain) and in the Applied Economics Seminar at the URJC and ESIC (Madrid, Spain). The authors give special thanks to Joan Torrent from UOC and IN3 and Angel Díaz from Rey Juan Carlos University for carefully reading this study and providing valuable suggestions. This research was partially supported by grants from the Research Project-Ministerio Economía y Competitividad ECO2012-32554.

* Corresponding author.
E-mail addresses: mballestard@uoc.edu (M.T. Ballestar), pilar.grau@urjc.es (P. Grau-Carles), jorge.sainz@urjc.es (J. Sainz).

http://dx.doi.org/10.1016/j.jbusres.2015.12.015
0148-2963/

The literature on customer rebates (coupons) is relevant to this analysis. Shaffer and Zhang (1995) show that coupons can serve as a strategic tool to keep consumer loyalty and attraction. In a connected world, this advantage translates on mass customization, which can serve to improve profitability in the market allowing for enduring relations between companies and costumers (Ansari & Mela, 2003). Incentives help to establish the right motivation for customers, as empirical (Byers, Mitzenmacher, & Zervas, 2012) and theoretical (Miller, Hofstetter, Krohmer, & Zhang, 2011) research shows.

This research tries to enlarge this literature by using the database of one of the largest cashback sites in Europe, which concentrates on its profitability to consumers. Out of a large and detailed database of the site, this study focuses on a window of the data to analyze the structure of their online behavior and their decisions by using a version of the structural equation model (SEM). The results shed light on the profiles of the site consumers and on their profitability, premiering on the area and opening new areas of research on the marketing of the sites and consumer value.

2. Theoretical questions

Cashback websites are a performance-based marketing strategy where a portal rewards one or various users by the lead/visit that they realize to the web of their affiliates. The relevant agents are (1) the cashback portal—which presents the offer to the costumers, (2) that portal's network of affiliate merchants, and (3) the consumers and their network of affiliate consumers (i.e., those who entered in the portal through their recommendation or the recommendation of their referees—up to the second level).

The business model of the cashback websites works as an affiliate model, where both the recommender and the referee benefit from the referee's transactions in the network (up to the second degree: affiliates and the affiliates of the affiliates). This system implies that customers financially benefit from the transactions made by themselves and their network of affiliates.

Transactions of type click/visit and search generate cashback without making a financial outlay by the user, conversely to the purchase type or others, which usually requires the acquisition of a product or service. In this situation, and as Zaglia (2013) points out, the interaction among consumers reinforces the use of the brand, which in the current model allows testing the relationship between the size of the consumer network and the benefits of cashback, because these benefits come mainly from customer activity.

H1. The size of the network of costumer (up to the second level) is relevant in the total number of transactions of type click/visit.

H2. The size of the network of costumer (up to the second level) positively influences obtaining economic benefit.

Consumers derive utility from its shopping experience (Lee & Tan, 2003). The number of stores and the number of different categories within those stores for which the user has performed transactions are the relevant factors in consumption diversification in *cashback* portals, even considering that one category can include several shops.

H3. Diversification in consumption leads to an increase in the number of transactions of type click/lead and search made by users and their network of users.

H4. Diversification in consumption positively influences economic benefits of the customer.

Finally, the higher the transaction volume of such click/visit or registration the user and network of users make, the greater the economic benefit the user obtains (Vana et al., 2015).

H5. Volume and type of transactions are crucial to consumer financial benefits in terms of cashback.

3. Method

3.1. Data collection

The present study uses information stored in the data warehouse of one of the largest cashback sites in Continental Europe (the Site from now on), currently present in fourteen countries, with more than two million customers, commercial agreements with 4332 stores and more than €5 million in cashback a year. This research uses a representative sample that allows building and efficiently evaluating the model. Its structure is the following.

3.1.1. Stores

The Site had agreements with 1373 stores during the sample period, although customer activity concentrated in 75% of them, generating more than €800 million in cashback. Customers can perform three kinds of activities to generate *cashback* in their accounts: one-click interaction or visit, registration, and purchase. Each store can generate cashback from one type of activity in the same period, but they can switch the activity depending on the business strategy:

1) One-click interaction or visit activity can be of several types such as: watching videos, visiting websites, becoming a fan in social networks, fulfilling surveys, using search engines.
2) Registration consists in becoming an identified user on a new website.
3) Purchase occurs when a customer buys a product or service in a store by accessing from the cashback platform.

Data shows that, by the end of 2014, the 50 stores with the highest cashback had the split by customer's activity as follows: 72% generated cashback by purchase, 20% by one-click interaction or visit, and only 8% by registration. However, looking at the store's relevance by activity, the split changes as follows: 68% did not stand out in any activity, whereas 20% were relevant in purchases, 6% in one-click interaction or visit, 4% in offers, and 2% in registration. Finally, customers gave to these stores an average punctuation of 3.95 over 5, suggesting a quite remarkable global satisfaction.

3.1.2. Categories

The study allocates all stores offering their products or services in the *cashback* platform in a category and subcategory according to the following criteria: (1) The type of products or services the store offers when users get the cashback by making a purchase; (2) The type of activity customers can perform to get the cashback, such as one-click interaction or visit, registration and purchase.

Once selected the 50 most relevant stores in terms of cashback volume, the number of categories reduces to 6, in which shopping and travel are again the most important categories, concentrating 70% of stores and 61% of cashback generated. In addition, the subcategories belonging to these stores reduce to 9 in shopping and 3 in travel.

3.1.3. Customer's navigation

Total customer transactions were 25.6 million in 2014. In the representative sample, this figure goes to 1.6 million, about 6.5% of the year. 18,520 different users performed these transactions, an average of 89 per customer in a single month. Users and their network were very active in categories in which users receive a cashback without directly purchasing goods or services, with ratios reaching 101 transactions/customer per month, whereas customers are much less active in categories that demand payouts, with ratios around 1.2–1.6 transactions/customer per month.

Regarding customer activity in the different categories according to their transactions, 99.8% of the monthly transactions correspond to registration and navigation through non-transactional categories, whereas the remaining 0.20% of the monthly transactions are from shopping and travel categories. From the point of view of the generated cashback,

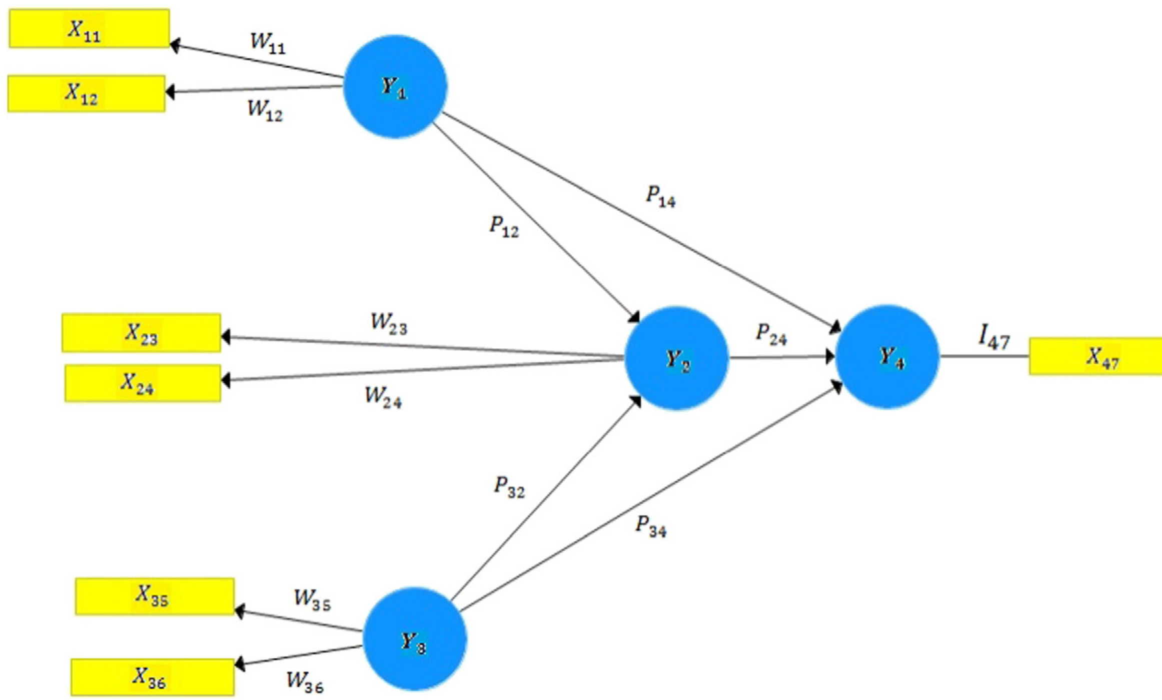

Fig. 1. Structural diagram of the PLS-SEM model showing the relationship between latent variables displayed in circles and observed variables displayed in boxes.

67.3% transactions come from shopping and travel and the rest from the registration and navigation.

As mentioned above, users receive cashback not just for their own activity but also for their network's activity in the platform. Once again, the source of the customer's activity (direct activity or network activity) varies according to the point of view. In terms of transactions, 44% of them come from users' own activity, whereas 56% come from their network activity. Users directly generate 84.2% of the cashback, whereas users' network generates 15.8%.

3.1.4. Users/customers

The sample consists of 18,520 users who transacted over the chosen period. 87.7% of the transactions during that period involved the 50 most relevant cashback-generating stores.

In socio-demographic terms, the gender ratio is skewed with more men (63%), with an average age of 34.3, than women (37%), with an average age of 36.2. The sample shows a big concentration of users regarding residence, with 40% of them living in urban regions.

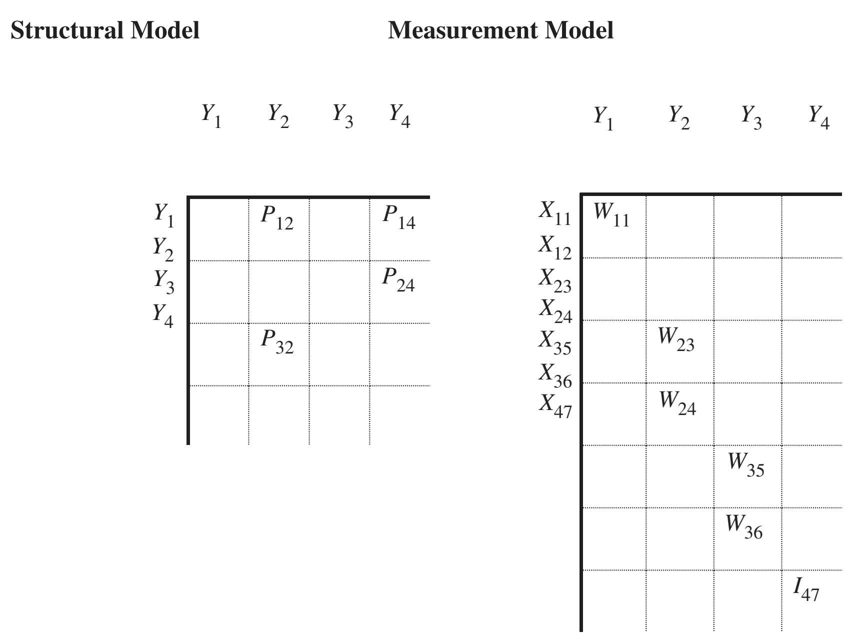

Fig. 2. Diagrams of the structural model showing the assumptions of latent variables and measurement models showing relationships between latent and observed variables.

Table 1
List of latent variables and indicator variables in the PLS-SEM model.

Latent variable		Item variables	Description
Y_1	Consumption diversification	n_distinct_cat	Numerical discrete variable which contains the number of distinct categories where the customer has made transactions.
Y_2		n_distinct_widget	Numerical discrete variable which contains the number of distinct stores where the customer has made transactions
	Transactionality level	n_ope_click	Numerical discrete variable which contains the number of one-click interaction or visit transactions made by the customer in the platform.
		n_ope_registros	Numerical discrete variable which contains the number of registration transactions made by the customer in the platform
Y_3	Recommendation capability	user_children_size	Numerical discrete variable which contains the number of first level friends of the customer who belong to his network.
		user_network_size	Numerical discrete variable which contains the total number of friends (first and second level of friends) of the customer who are his full network
Y_4	Economic benefit	sum_uscom_network_amount	Numerical continuous variable which contains the total *cashback* generated by the user's network in the platform. As explained above, the total network is built by first and second level user's friends.

3.2. Input data processing

The data processing used the software SAS, simplifying all the data in a single table that aggregates information at customer level, comprising 18,250 records of 28 variables, of which 7 are relevant for the model. These variables describe customer activities relevant to their economic benefit:

1) n_distinct_cat: Numerical discrete variable that contains the number of distinct categories in which the customer has made transactions.
2) n_distinct_widget: Numerical discrete variable that contains the number of distinct stores in which the customer has made transactions.
3) n_ope_click: Numerical discrete variable that contains the number of one-click interaction or visit transactions the customer makes in the platform.
4) n_ope_registros: Numerical discrete variable that contains the number of registration transactions the customer makes in the platform.
5) user_children_size: Numerical discrete variable that contains the number of first-level friends of the customer who belong to his or her network.
6) user_network_size: Numerical discrete variable that contains the total number of friends (first and second level of friends) of the customer full network.
7) sum_uscom_network_amount: Numerical continuous variable that contains the total *cashback* a user's network generates in the platform. The total network comprises first- and second-level user's friends.

4. Theoretical model

This research uses Partial Least Squares Structural Equation Modelling (PLS-SEM) to discover which latent variables have the highest influence on customer economic benefits. Wold (1966), Chin (1998), and Lohmöller (1989) originally developed PLS-SEM. This method allows to model complex relationships among multiple variables. Researchers often use this approach to confirm relationships among variables. The advantage of SEM-based procedures over other techniques such as factor, discriminant, and principal component analysis is the flexibility of this method to relate theory and data, that is, to model relationships among variables, to construct unobservable latent variables, and to statistically test assumptions against empirical data. Therefore, research is increasing using this technique in marketing

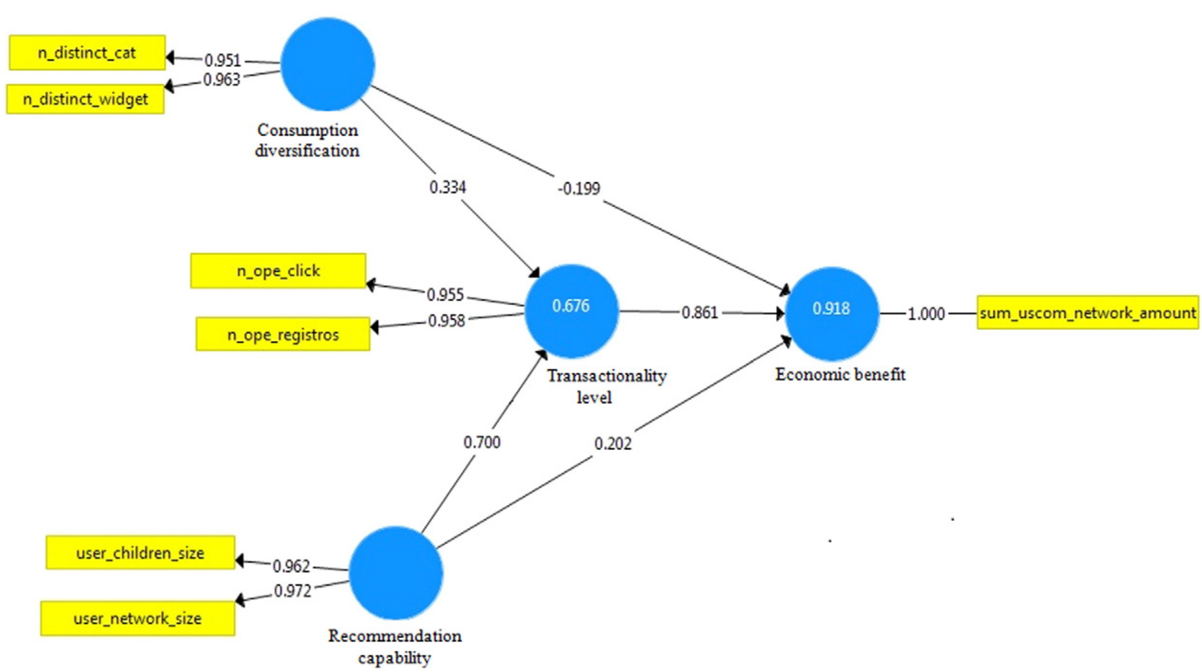

Fig. 3. Path diagram for the structural part of the fitted model.

Table 2
Results of the outer loadings indicating the correlation between the latent variables and the observed variables.

	Economic benefit	Recommendation capability	Consumption diversification	Transactionality level
n_distinct_cat			0.95	
n_distinct_widget			0.96	
n_ope_click				0.96
n_ope_registros				0.96
sum_uscom_network_amount				
user_children_size		0.966		
user_network_size	*Single item construct*	0.976		

and management or organizational research (Hair, Sarstedt, Ringle, & Mena, 2012; Henseler, Ringle, & Sinkovics, 2009).

PLS path model consists of three components: The structural model, the measurement model and the weighting scheme. Fig. 1 shows the structural diagram of the PLS-SEM model. The circles represent the latent variables in the structural model. Both exogenous variables Y_1 and Y_3 establish effect relationships or hypothesis P_{mn} with endogenous variables Y_2 and Y_4. These effect relationships are of direct type when established with latent variable Y_2 but both of direct and indirect types when established with latent variable Y_4. Rectangles represent the observable variables or indicator variables X_{mq} in the measurement model. This is a reflective measurement model because the relationships, called loadings, go from the latent variable to a pair of indicator variables except for the latent variable Y_4 with a single indicator variable. Indeed, one of the most relevant advantages of PLS-SEM model is the possibility of working with latent variables with single indicator variables.

Therefore, Fig. 2 shows the composition of the structural and measurement models according to this structural diagram.

Finally, Table 1 shows a detailed description of the items and latent variables of the PLS-SEM model. The input data for the model is a table consisting of 7 variables (one per each indicator variable of the measurement model) and one record per user, containing consequently 18,250 records.

5. Empirical analysis and results

The aim of the PLS-SEM algorithm is to maximize the percentage of explained variance of the latent endogenous variables and providing the path coefficients and the proportion of explained variance of the endogenous latent variables. Fig. 3 shows the estimation results, produced through SmartPLS, and the value of the estimated measurement model.

Because of the lack of a criterion to evaluate the global model's quality, the study evaluates the measurement model (reflective measurement model in this research) and the structural model as in Hair et al. (2012).

5.1. Measurement model evaluation

For the validation of a reflective measurement model, the study validates the following criteria: reliability, internal consistency reliability, convergent validity, and discriminant validity.

5.1.1. Reliability indicator

This indicator assesses the value and significance of the measurement or outer model coefficients. The outer loadings of Table 2 indicate the correlation between the latent variables and each of the indicators or observed variables. Their values in this model are much higher than the required 0.71 for acceptation in an exploratory model (Hair et al., 2012).

5.1.2. Internal consistency reliability

Two complementary criteria assess the internal consistency of each latent variable of the model: the Cronbach's alpha (a conservative measure because of its characteristics and limitations) and composite reliability. The Cronbach's alpha provides an estimation of the reliability based on the correlations between indicator variables assuming that all of them are equally reliable, whereas composite reliability is a criterion that measures the internal consistency using the outer loadings. Values of both criteria go from 0 to 1 and their cut-off value is 0.6. The first two columns of Table 3 show the results of both criteria. Those results are consistent with the fact the model has high internal consistency because all the values of Cronbach's alpha and Composite reliability are higher than 0.9.

5.1.3. Convergent validity

Convergent validity assesses the correlation among indicator variables of the same latent variable. Due to the fact that this is a reflective measurement model, these correlations should be rather high. One of the most suitable indicators to determine if the model accomplishes the convergent validity criteria is the average variance explained (AVE). The third column of Table 3 shows that all of them are higher than 0.5 which means that each latent variable is able to explain more than 50% of the variance of its indicator variables.

5.1.4. Discriminant validity

Discriminant validity evaluates how different are the latent variables among them, given that each one has to explain a different phenomenon in the model. Two methods measure discriminant validity: the Cross-Loadings method and the Fornell–Larcker criterion. The Cross-Loadings method uses outer loadings of each indicator variable in their respective latent variables. The method confirms discriminant validity, as in this model, if the outer loadings are higher than their outer loadings calculated with other latent variables. Table 4 shows the results of the method showing that all the bold values are greater than the other values in the same column.

The Fornell–Larcker criterion consists in the comparison between the square root of AVE for each latent variable and the rest of correlations among latent variables. This method confirms the discriminant validity if the former are greater than the latter. Table 5 shows discriminant validity exists in the model: bold values are greater than those located both in vertical and horizontal.

5.2. Structural model

When the valuation of the measurement model is satisfactory, the next stage consists of assessing the structural model. To do so, the study uses the following criteria: R^2 values for endogenous latent variables, effect size f^2, predictive relevance Q^2, evaluation of the structural model path coefficients, and multicollinearity model evaluation.

Table 3
Results for the Cronbach's alpha, composite reliability and average variance explained (AVE).

	Cronbach's alpha	Composite reliability	AVE
Economic benefit	*Single Item Construct*	*Single Item Construct*	*Single Item Construct*
Recommendation capability	0.93	0.97	0.93
Consumption diversification	0.91	0.96	0.92
Transactionality level	0.91	0.96	0.92

Table 4
Results of the cross-loading validation method.

	Economic benefit	Recommendation capability	Consumption diversification	Transactionality level
n_distinct_cat	0.19	0.14	**0.95**	0.40
n_distinct_widget	0.23	0.17	**0.96**	0.45
n_ope_click	0.87	0.67	0.52	**0.96**
n_ope_registros	0.90	0.77	0.34	**0.96**
sum_uscom_network_amount	**1.00**	0.82	0.22	0.93
user_children_size	0.71	**0.96**	0.17	0.69
user_network_size	0.86	**0.97**	0.15	0.76

Bold values show the outer loading of the indicator variable with their respective latent variable. For the model validation they should be higher than the outer loading calculated with other latent variables.

Table 5
Results of the Fornell–Larcker validation criterion showing correlation among latent variables.

	AVE	SQRT (AVE)	Economic benefit	Recommendation capability	Consumption diversification	Transactionality level
Economic benefit	1	1.00	*Single item construct*			
Recommendation capability	0.93	0.97	0.82	**0.97**		
Consumption diversification	0.92	0.96	0.22	0.16	**0.96**	
Transactionality level	0.92	0.96	0.93	0.75	0.45	**0.96**

Bold values are the SQRT(AVE), showing higher values than the correlation among the rest of the latent variables.

5.2.1. R^2 values for endogenous latent variables

The R^2 indicates the percentage of variance of the latent endogenous variables that the latent exogenous variables explain. In this research, R^2 is 0.68 for the “transactionality level” and 0.20 for “economic benefit,” which is a high value.

5.2.2. Effect size f^2 and predictive relevance Q^2

Effect size f^2 is complementary to R^2. f^2 measures how relevant is each exogenous latent variable for explaining their respective endogenous latent variables. Table 6 shows the results of the calculations of the effect size, with values going from $f^2 = 0.20$ to very high $f^2 = 2.95$.

The Blindfolding yields Q^2 values, which are also complementary to $R^2$and useful to measure the predictive relevance of the model. Table 7 shows that these values are higher than 0.5, meaning that exogenous latent variables have a large predictive relevance for the endogenous latent variables.

5.2.3. Evaluation of the structural model path coefficients

Structural model path coefficients represent the relationship between the latent variables. Path coefficients with absolute value less than 0.1 may indicate “small” relationship, values around 0.3 indicate a “medium effect,” and values greater than 0.5 indicate a “large effect.” Additionally, bootstrap yields the p-values to test the coefficients' significance. Table 8 shows the results of the estimated path coefficients together with their t-statistics and p-values, confirming a strong evidence of relationship between path coefficients rejecting the null hypothesis of no relationship in all cases.

Regarding the coefficient values, a strong relationship exists (path coefficient value of 0.86) between customers' “transactionality level” in the platform and their obtained “economic benefit” in terms of *cashback*, meaning that the higher the number of transactions of type click/visit or registration users accumulate in their account, from their own activity or their network, the greater the economic benefit obtained. This result strongly supports H5.

A strong relationship exists between users' “recommendation capability” and their “transactionality level” in the *cashback* platform showing a path coefficient of 0.7. This result means that, in this business model, a user can accumulate more *cashback* transactions when the network is wider, both first and second level. This result confirms H1.

A medium relationship exists between customers' “consumption diversification” and their “transactionality level” in the *cashback* platform, with a path coefficient of 0.33, meaning than more diverse operations, both in number of product categories and in stores, increase the probability that the transactional type increases. This finding confirms H3.

A moderate relationship exists between the users' “recommendation capability” and their obtained “economic benefit” in terms of cashback (path coefficient of 0.2). This result means that the user recommendation capability has a moderate influence on the economic benefit in the user's account, providing moderate support for H2.

A medium to small negative relationship exists between the customer's “consumption diversification” and their obtained “economic benefit” in terms of *cashback*, with coefficient of −0.2. This result implies that a user with greater diversification in the operations, both in terms of product and service categories, is more likely to get less economic benefit *cashback* than a user with a more concentrated activity. This finding is against H4, which assumes that diversification in consumption has positive influence in consumer economic benefit.

5.2.4. Multicollinearity

Another aspect to take into account is the analysis of the absence of multicollinearity in the model, either among observable variables in the measurement model or among latent variables in the structural model. The Variance Inflation Factor (VIF) quantifies the severity of multicollinearity in a least squares regression analysis, providing an index that measures how much the variance of an estimated regression coefficient increases because of collinearity. Following the usual guidelines to interpret VIF results of Table 9, all the values are greater than 1 and smaller than 5, meaning that data present moderate correlation.

Table 6
Results for the effect size f^2.

	Economic benefit	Transactionality level
Recommendation capability	0.20	1.47
Consumption diversification	0.36	0.34
Transactionality level	2.95	

Table 7
Results of the predictive relevance with Q^2 values.

	Q^2 value
Economic benefit	0.76
Transactionality level	0.60

Table 8
Statistical analysis of path coefficients.

	Original sample (O)	Sample mean (M)	Standard error (STERR)	t statistics (|O/STERR|)	p-Values
Recommendation capability → Economic benefit	0.20	0.17	0.0747	2.73*	0.01
Recommendation capability → Transactionality level	0.70	0.67	0.127	5.85*	0.00
Consumption diversification → Economic benefit	−0.20	−0.20	0.04	5.26*	0.00
Consumption diversification → Transactionality level	0.33	0.36	0.08	4.10*	0.00
Transactionality level → Economic benefit	0.86	0.87	0.09	10.18*	0.00

Notes:
* Significant at 0.05 level.

Table 9
VIF for latent and observable variables.

VIF	Economic benefit	Transactionality level		VIF
Economic Benefit			n_distinct_cat	3.263
Recommendation capability	2.540	1.027	n_distinct_widget	3.263
Consumption diversification	1.371	1.027	n_ope_click	3.230
Transactionality level	3.090		n_ope_registros	3.230
			sum_uscom_network_amount	1.000
			user_children_size	4.095
			user_network_size	4.095

6. Conclusions and suggestions for future research

Cashback is a relevant marketing strategy in electronic commerce; cashback increases the engagement of the customers and attracts interest on the sites regardless of the category. Despite this relevance and the increasing numbers of customer engagement and the growing size of its business, little academic literature exists on this topic. This research studies cashback by disentangling some questions of consumer behavior.

By using Partial Least Squares-Structural Equation Model, the empirical analysis shows that customers' recommendation effect is relevant; as the size of the customer's network increases, the total volume of transactions in the customer's account also increases, resulting in higher economic benefit for all the consumers.

The analysis also shows that different categories have different responses from consumers; customers are more likely to engage with the brand through activities that generate cashback without making any purchase or money investment, allowing for an increase on activity, transactions, and economic benefit of customers and their network affiliates.

The analysis also shows that the cashback the costumers/users generate is quite relevant, representing 84.2% of the total amount generated during the period of study. This area is open for further research, including the analysis of the different socioeconomic categories, focusing on the present value of the costumers and deriving algorithms for individual analysis and customization. This study also paves the way to further research on the modellization of consumer behavior, disentangling their decisions on transactions, navigation paths, repeated visits, etc. The use of the data is relevant both academically and for the business, because the data allows for an easy estimation of the monetization of the sites.

References

Ansari, A., & Mela, C.F. (2003). E-customization. *Journal of Marketing Research*, *40*(2), 131–145.

Byers, J.W., Mitzenmacher, M., & Zervas, G. (2012). The groupon effect on yelp ratings: A root cause analysis. *Proceedings of the 13th ACM conference on electronic commerce* (pp. 248–265) ACM.

Chen, Y., Ghosh, A., Mcafee, R.P., & Pennock, D. (2008). Sharing online advertising revenue with consumers. In C. Papadimitrou, & S. Zhang (Eds.), *Internet and network economics* (pp. 556–565). Berlin: Springer Verlag.

Chin, W.W. (1998). The partial least squares approach to structural equation modeling. *Modern Methods for Business Research*, *295*(2), 295–336.

Duffy, D.L. (2005). Affiliate marketing and its impact on e-commerce. *Journal of Consumer Marketing*, *22*(3), 161–163.

Fu, T., Chen, Y., Qin, Z., & Guo, L. (2013). Percolation on shopping and cashback electronic commerce networks. *Physica A: Statistical Mechanics and its Applications*, *392*(12), 2807–2820.

Hair, J., Sarstedt, M., Ringle, C., & Mena, J. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science*, *40*(3), 414–433.

Henseler, J., Ringle, C.M., & Sinkovics, R.R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing (AIM)*, *20*, 277–320.

Ho, Yi-Chun (Chad), Ho, Yi-Jen Ian, & Tan, Yong (2013). Online Cashback Pricing: A New Affiliate Strategy for E-Business. *Proceedings of the 34th International Conference on Information Systems (ICIS)*.

Homburg, C., Droll, M., & Totzek, D. (2008). Customer prioritization: Does it pay off, and how should it be implemented? *Journal of Marketing*, *72*(5), 110–130.

Jain, K. (2007). The good, the bad, and the ugly of the search business. *Fourth bi-annual conference on the economics of the software and internet industries*. Toulouse, France: Toulouse School of Economics.

Lee, K.S., & Tan, S.J. (2003). E-retailing versus physical retailing: A theoretical model and empirical test of consumer choice. *Journal of Business Research*, *56*(11), 877–885.

Libai, B., Biyalogorsky, E., & Gerstner, E. (2003). Setting referral fees in affiliate marketing. *Journal of Service Research*, *5*(4), 303–315.

Lohmöller, J.B. (1989). *Latent variable path modeling with partial least squares*. Heidelberg: Physica-Verlag Heidelberg.

Miller, K.M., Hofstetter, R., Krohmer, H., & Zhang, Z.J. (2011). How should consumers' willingness to pay be measured? An empirical comparison of state-of-the-art approaches. *Journal of Marketing Research*, *48*(1), 172–184.

Rust, R.T., & Chung, T.S. (2006). Marketing models of service and relationships. *Marketing Science*, *25*(6), 560–580.

Shaffer, G., & Zhang, Z.J. (1995). Competitive coupon targeting. *Marketing Science*, *14*(4), 395–416.

Vana, P., Lambrecht, A., & Bertini, M. (2015). *Cashback is cash forward: Delaying a discount to increase future spending*. Available at SSRN: http://ssrn.com/abstract=2588611.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. *Multivariate analysis*, *1*, 391–420.

Zaglia, M.E. (2013). Brand communities embedded in social networks. *Journal of Business Research*, *66*(2), 216–223.