WCIT 2010

Retail core banking services e-banking client cluster identification

Ivan Soukal[a]*, Martina Hedvicaková[a]

[a]*University of Hradec Králové, Faculty of informatics, Department of economics*
Rokitanského 62, Hradec Králové 500 03, Czech republic

Abstract

This paper is focused on retail core banking services market in the Czech Republic. There are presented the data analysis outcomes from expert web-system for client's bank costs computation and real-time bank offer comparison. Using nonhierarchical cluster analysis there are identified 5 client profiles by specific frequencies of bank services usage. Paper reveals that among the e-banking population there are expectable clusters so the unexpected cluster small businessmen that are not supposed to use retail products and cluster of strong desk usage preference clients by contrast to activated e-banking.

ⓒ 2010 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
Selection and/or peer-review under responsibility of the Guest Editor.

Keywords: Banking; retail core banking services; e-banking; cluster analysis; client

1. Introduction

This paper is focused on the retail core banking services market (thereinafter only as RCBS abbreviation), which concern a great majority of consumers in European Union market (thereinafter only as EU abbreviation). These services include account administration, cash utilization, realizations of money transfers and additional services. Various studies assigned, or elaborated, by European Committee or European Union Directorate-General for Health and Consumers Protection [1], [2], [3] show the fundamental problem in this market from the point of view of the consumer. This problem is low transparency and very difficult product comparison of offers of individual banks with respect to client costs (generally included in the term bank charges), which restricts the principle of the invisible hand of the market. Opacity of offers of individual banks and impossibility to compare them, easily and clearly, is one of the key manifestations of asymmetry of information in the market. This problem was observed across the EU.

As the very response in the Czech Republic (thereinafter only as CZ abbreviation) to this situation the web-based RCBS expert system called The Bank charge calculator (thereinafter only as Calculator) was created as an independent project of the RCBS market focused web bankovnipoplatky.com. This paper analyses Calculator respondent's data using cluster analysis based on k-means algorithm. Clusters were computed using statistical software IBM PASW 18 (formerly known as SPSS).

---

* Ivan Soukal. Tel.: +420-493-332-363;
E-mail address: ivan.soukal@uhk.cz.

2. Market situation and solution in the Czech Republic

The SEPA monitoring study focused only on imposition of charges on electronic banking confirmed how difficult it is to ascertain client costs of these services because of unclear and nontransparent tariff structures of bank products. One of the main findings was that with regard to 69% of offers of bank institutions, even after studying these offers thoroughly, experts had to ask the banks subsequently about the exact amount of charges for the individual services [1]. The consequence of the intransparency and lack of clarity was found out in the study of UK Office of Fair Trading and is the following. Customers basically do not know how much they pay for bank services, not even after they were already charged for the observed services [2]. On the basis of recommendation by the company elaborating the study, the EU commissioner for interior market Charles McCreevy, before the end of his term of office, took charge of several most crucial issues. The major issue for Central and Eastern European countries was to create an on-line accessible and easy to use comparison service, which enables customers to choose the bank most suitable for their needs from the perspective of costs.

As the very response in the Czech Republic there was created The Calculator. This service compares costs for clients in individual offered accounts on the basis of their RCBS use and recommends the lowest costs products. Knowledge base of the Calculator contains the tariff data of 12 banks (more that 98% of the RCBS market in the CZ) and their 45 accounts. Client only fills the detailed form concerning the usage of specific services – 52 questions in total (25 questions with attached sub questions and three additional questions). All fills are saved so the data for this study were acquired from 12 012 respondents who used the form of the Calculator in the time period from 12th October 2009 to 30th July 2010. The main tool for further help to choose the optimal cost account requires the identification of the most important RCBS client profiles. Because of specific respondent data gathering there has to be limited the population for which the client profiles will be identified. The client profiles in this paper are computed for and from e-banking client population (non-e-banking respondents were excluded from analysis). Analysis method of K-means was chosen by recommendation of [4], [5], [6] due to number of respondents and variable scale.

3. Cluster analysis results

3.1. Treatment including the extreme clients

Within this clustering, 8 804 respondents per 21 variables were evaluated by the method listwise. Data validation of this treatment did not filtered respondents who are not supposed to be retail clients. They are using retail accounts for their small business or as self-employed although it is not allowed. Still this phenomenon is not that rare and it resulted in wide tariff change focused on very high RCBS usage frequencies e.g. in CZ well known low-cost mKonto account offered by mBank Commerzbank at the beginning of the second quarter of 2010. With regard of the presence of those respondents, it was not possible to use the methodology of the source [7] with the number of clusters 5 and a survey of the optimum number of clusters had to be made on the basis of maximization of the global rule G_5 [4], [6]. The highest G_5 value was reached at number of clusters 6.
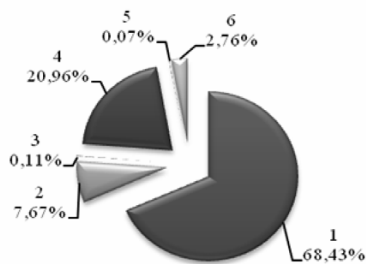
In the Table 1. there are shown values of cluster centroid for each of a typical client behavior pattern. The graph (Fig 1.) then shows members count of the individual clusters. It is apparent that, within the analysis of the extreme clients data, low-element clusters with minimum share in the whole were also formed (both clusters are special form of an extreme client differed by desk preference and ATM withdrawals abroad). For this reason, only 4 clusters will be described closely:

Fig 1. Shares of computed clusters, source: own research

Table 1. Centroid values for each cluster, source: own research

Variable/cluster	1	2	4	6
Domestic ATM withdrawal, own bank	2,6	3,1	3,7	4,2
Domestic ATM withdrawal, foreign bank	,7	,5	1,2	1,8
Abroad ATM withdrawal, own bank	,1	,2	,3	,3
Abroad ATM withdrawal, foreign bank	,1	,1	,3	,3
Incoming payment from foreign bank	2,2	2,2	3,6	21,5
Incoming payment from own bank	1,0	1,2	2,1	11,8
Direct payments to own bank at desk	,1	1,1	,1	,3
Direct payments to own bank Internet	1,5	,9	4,2	13,6
Direct payments to foreign bank at desk	,1	1,9	,1	,1
Direct payments to foreign bank Internet	3,3	1,2	6,8	18,0
Standing orders to own bank at desk	,1	1,8	,0	,1
Standing orders to own bank Internet	,6	,3	2,8	2,1
Standing orders to foreign bank at desk	,1	3,1	,0	,0
Standing orders to foreign bank Internet	1,8	,4	5,2	3,3
Encashment to own bank at desk	,0	,8	,0	,0
Encashment to own bank Internet	,2	,1	1,4	,6
Encashment to foreign bank at desk	,0	1,4	,0	,1
Encashment to foreign bank Internet	,6	,2	2,3	1,1
Cash deposit at desk	,3	,6	,6	2,2
Cash withdrawal at desk	,2	,7	,3	1,2
Cash back	,1	,1	,3	,3

- The average client, 68,4% – cluster 1 is major group of the e-banking clients population. It shares common frequency of ATM withdrawals with the others clusters (approximately 3 times from client's own bank and once from foreign bank). Typical for this client is preference of electronic banking usage both for direct payments, standing orders and encashment. At the desk communication and services are used rarely for money transfers, approximately once per year, which corresponds to the value 0,1 and it is the same as ATM withdrawals abroad. There it is possible to interpret 0,1 value as that approximately one from 10 clients in the cluster executes a payment operation at desk once per month.
- The average client with "at the desk services" preference, 7,6% – cluster 2 is almost 10 times smaller than the major cluster. The typical client of this cluster differs from the average one in the lower preference of internet banking and significantly higher preference (almost 9 times higher preference) of carrying out money transfers at the desk. This client does not differ much from the major profile concerning the frequency of RCBS usage (incoming payments, ATM withdrawals, direct payments etc.). Still it has been to a certain extent surprise that there are clients with active e-banking and strong desk preference.
- The active client, 20,9% – cluster 4 includes clients whose preferences regarding the information channel are very close to the cluster 1. The difference is the frequency of RCBS usage per month which is approximately 2–3times higher (concerning direct payments, standing orders and encashment). Higher frequency was found at the services of incoming payments (2 times higher) and also ATM withdrawals from both own and foreign bank (higher approximately by 50%).
- The extreme client, 2,7% – cluster 6 includes clients who are obviously not supposed to use retail accounts. There is high probability that who this profile is consisted of self-employed, small business or well known half-legal internet auction businessmen without license. This cluster differs from previous profiles by more than 10times higher frequency of incoming and outgoing payments per month (34 incoming payments, 38 outgoing payments). As a communication channel The Internet is strongly preferred. At the desk services are used especially for cash deposits and a bit less for cash withdrawals. Those services have surpassed the major cluster at the desk usage more than 6times and so the cluster with at the desk preference (more than 2 times).
- Clusters 3 and 5, 0,2% together, are low-element clusters with centroids, or it can be said mean values are close to the extreme client profiles ones. Their activity is even higher than in cluster 6, another difference of the cluster 4 was use of ATM abroad.

The cluster analysis have shown that the majority of the population with active e-banking have stronger internet preference, however, there still is a not completely insignificant group of clients who, though having active e-banking, prefer to settle their bank operations at desk. The second surprising outcome is that almost all profiles, excluding the extreme one, have almost identical preference of using ATM almost 3 times a month. This analysis has also proved that in The CZ there are some retail products used for small business. At last but not least this analysis has computed very exact frequencies of specific client profile RCBS usage that weren't available for public yet because understandable lack of willingness of the banks to share business related data.

3.2. Treatment excluding the extreme clients

Within framework of this clustering, 6 943 respondents per 21 variables were analysed using the listwise method. Data validation of this treatment used more strict data filtering to guarantee that source data won't include extreme clients this paper described before. Exclusion of extreme clients allowed usage of source [7] method based on 5 clusters.

Fig 2. Shares of computed clusters, source: own research

Table 2. Centroid values for each cluster, source: own research

Variable/cluster	1	2	4	6
Domestic ATM withdrawal, own bank	3,1	2,5	2,8	2,1
Domestic ATM withdrawal, foreign bank	1,0	,4	,8	,4
Abroad ATM withdrawal, own bank	,4	,1	,1	,1
Abroad ATM withdrawal, foreign bank	,4	,1	,1	,1
Incoming payment from foreign bank	3,1	1,7	2,2	1,4
Incoming payment from own bank	2,3	,8	,7	,7
Direct payments to own bank at desk	,1	1,1	,0	,1
Direct payments to own bank Internet	3,6	,5	1,5	,8
Direct payments to foreign bank at desk	,1	1,7	,0	,1
Direct payments to foreign bank Internet	4,4	,7	4,0	1,3
Standing orders to own bank at desk	,0	1,6	,0	,1
Standing orders to own bank Internet	2,5	,2	,8	,3
Standing orders to foreign bank at desk	,0	2,7	,0	,2
Standing orders to foreign bank Internet	3,0	,3	3,0	,6
Encashment to own bank at desk	,0	,8	,0	,1
Encashment to own bank Internet	1,3	,1	,2	,1
Encashment to foreign bank at desk	,0	1,2	,0	,1
Encashment to foreign bank Internet	1,6	,1	1,1	,2
Cash deposit at desk	,6	,5	,2	,3
Cash withdrawal at desk	,3	,5	,1	,2
Cash back	,3	,1	,1	,1

In the table there are shown values of cluster centroid for each of a typical client behavior pattern. The graph then shows members count of the individual clusters. As it can be seen on the graph attached, there has been computed one extremely low-member cluster. Because of its insignificant influence there will be described in detail 4 clusters only:

- The average client, 39,3% – cluster 3 is major group of the e-banking client population. It shares common frequency of ATM withdrawals with the others clusters (approximately 3 times from client's own bank and once from foreign bank). Typical for this client is preference of electronic banking usage both for direct payments, standing orders and encashment. Usage of at the desk services is sparse. Usage of these services is consisted of cash deposit, cash withdrawal only once or twice per year (this interpretation can be reversed as in the previous analysis, that is one from ten clients from this group uses an at desk cash withdrawal once per month).

- The active client, 16,3% – cluster 1 is a group of the more active clients, where, compared to the average client, the frequency of incoming payments is 2 times higher. Usage of services direct payments to own bank, cash ATM withdrawal from foreign bank, cash deposit or withdrawal and standing orders to own bank is 3 times higher. Concerning other services, this profile is similar to the average client and this client also shares with the average client the preference of the communication channel of e-banking. This client also has the highest frequency of ATM use abroad, although it is only 3 withdrawals per year on average.

- The average client with "at the desk services" preference 8,3% – cluster 2 is smaller than the cluster with desk preference clients from previous clustering. RCBS usage frequency of money transfers and incoming payments are very similar to the average client profile, the difference is that realization almost always occurs at the desk. This difference from the average client cluster can be noticed e.g. at almost 5 times higher frequency of cash deposits and withdrawals at the desk.

- The passive client, 35,9% – cluster 5 includes clients with low frequencies of monthly usage of all the monitored services. It can be noticed e.g. on services of money transfers and incoming payments, where this client profile receives only 2 payments per month and carries out only 2 direct payments and one standing order. All transfer

services are done via internet. Compared to the average client profile, this cluster also has three times lower month frequency of ATM withdrawals from own bank and frequency of ATM withdrawal from foreign bank is lower by half. This client could be called low-cost client.

Compared to the previous analysis (that included extreme clients), this treatment allowed cluster analysis to achieve better differentiation. It allowed the k-means algorithm to separate passive clients from the cluster of average clients. In the first clustering, passive clients had an effect up the profile of the average client. The major cluster (almost 70%) then has been split up to two smaller clusters that can be described in greater detail. Cluster member count is almost equal (the difference passive-average client is only approximately 5%). It is true that statistically this split did not have strong impact on average, but from the marketing point of view, it is a crucial to know the difference between the so-called low-cost client and average client with, however, 2–3 times higher use of services. Compared to the first treatment active client cluster's share is here lower by 4%. The main reason is more strict data validation. The second treatment also raised the share of the client profile using mainly desk operations but an increase was less than 1%. The main benefit of the second more detailed treatment lies in the separation of passive clients from those average ones.

3.3. Next possible use of clustering results

The very identification of client clusters within the population of RCBS clients with e-banking is beneficial both public and scientific readers. The professional level can, on the basis of the existing situation describe here (the understandable impossibility to obtain non-aggregated data from banks about the RCBS usage) use client clusters as a basic tool for monitoring the price level or client cost level of RCBS. The only what has to be done is to regularly compute using the Calculator costs for each of the profiles. This figure will be followed by share of each bank on the market (using data from annual bank reports of from Calculator question which bank client currently uses). Using e.g. weighed sum approach there can be determined average level of RCBS client costs. In this case the weights will be share of clusters on the first level of computation and on the second level there will be used bank market shares. The public reader might use computed profiles as a guide what bank is the best for his or hers usage pattern. Client just identifies (based on month RCBS usage frequencies) what profile describes his or hers needs best. So it can be used also as a tool for reducing the RCBS market asymmetry of information. This approach allows better functionality of the principle of the invisible hand of the market more efficient than e.g. legal ban of charging of some services. This fact was proved in CZ during November 2009 after the legal harmonization with the Directive 2007/64/ES (the prohibition certain charges for monthly client informing about the account balance, card cancellation etc.). The result was not the decrease of charges, banks showed very fast and easy adaptation when new charges were introduced to regain profit lost because of mentioned charge ban [7].

4. Conclusion

Using non-aggregated data concerning RCBS e-banking client's month usage there was computed k-means cluster analysis. The main part included two treatments of clustering of 8 804 respondents and of 6 943 respondents. First cluster analysis identified a major cluster with almost 68,5% share (for details of RCBS usage please see Table 1.), which represented the average client. Analysis confirmed open secret that some small businessman and self-employed use retail products although it is not allowed. This cluster size was about 2,7 % and RCBS usage frequencies as high as more than 30 incoming and outgoing payments per month. The second non-expected outcome was cluster of clients with active e-banking and significant preference of using at the desk services (7,6 % of respondents). Besides these clusters one more profile was identified with the size of 20,9 % – active client. This client shares the preference of Internet as the communication channel, however his or her monthly use of RCBS is 2–3 times higher compared to the average client.

The second treatment analyzed strictly filtered data (extreme usage of RCBS was removed). Within this analysis, the group of passive clients (35,9% share) was separated from the major cluster. It differs from the average client (39,3% share) by lower activity by 50 %, Internet preference is the same. This preference is also common for active client (16,3 % share). This cluster differs from major population by 2–3 times higher activity. The last group of clients (8,3 % share) showed at the desk preference and activity similar to the average client. The results can be used as a tool for further research of the price level/client's costs level of RCBS. The second possible use is as a tool for

swift RCBS provider comparison when there are computed certain costs for each of retail account monitored by Calculator project.

5. Acknowledgement and funding

References

1. European Union Directorate-General for Health and Consumers Protection. SEPA monitoring study. 2008, [online], URL: http://ec.europa.eu/consumers/rights/docs/SEPA_monitoring_study.pdf, cit. 22.8.2010

2. European Union Directorate-General for Health and Consumers Protection. Retail financial services to the consumer markets scorecards. 2009, [online], URL: http://ec.europa.eu/consumers/rights/docs/swd_retail_fin_services_en.pdf, cit. 22.8.2010

3. Commission of the European communities. Study on the Data collection for prices of current accounts provided to consumers. 2009, [online], URL: http://ec.europa.eu/consumers/strategy/docs/prices_current_accounts_report_en.pdf, cit. 24.8.2010

4. A.D. Gordon, Classification, 2nd edition, Chapman & Hall/CRC, London, 1999.

5. S. Brian et al., Cluster analysis, 4th edition, Arnold Hodder headline group, London, 2001.

6. P. Hebák, Vícerozměrné staistické metody, 3. díl, (translated as Multivariate statistical methods, 3rd volume), Infromatorium, Prague, 2005

7. I. Soukal, Directive 2007/64/ES legal harmonization impacts on CBS client's costs in the ČR, Proceedings of the conference Hradec economical days 2010, Hradec Králové, 2008