

# SHLUKOVÁ ANALÝZA POPTÁVKOVÉ STRANY TRHU ZÁKLADNÍCH BANKOVNÍCH SLUŽEB

Jan Draessler, Ivan Soukal, Martina Hedvičáková

## Úvod

Tento článek je zaměřen na trh základních bankovních služeb fyzickým osobám pro nepodnikatelské účely, které se běžně označují termínem retail core banking services (dále v textu pouze jako RCBS), či retail day-to-day services. Tento trh je fundamentální součástí retailového bankovníctví vytvářející kolem 2 % HDP v rámci Evropské Unie (dále v textu pouze jako EU). Jeho případná nestabilita nebo neefektivnosti brzdí ekonomický rozvoj i díky tomu, že do retailového bankovníctví spadá více než polovina veškeré činnosti v bankovníctví. I proto je věnováno stále více pozornosti jeho nedokonalostem. Zelená kniha [5] a odvětvové šetření [3] zaměřené na RCBS trh (v návaznosti na Bílou knihu z roku 2005 a na následné studie) jasně vyjmenovávají problémy, se kterými se tato oblast potýká. Jedna skupina problémů je přímo spjata s rozhodováním zákazníků retailového bankovníctví, které je mimo jiné omezováno informační asymetrií. Informační asymetrie je sice přítomna prakticky ve všech tržních strukturách světové ekonomiky, ovšem příčiny, míra, důsledky a možnosti řešení se liší. V námi sledované oblasti je tedy negativním dopadem zmíněné asymetrie postižena primárně strana poptávky. Tato asymetrie pramení primárně z netransparentních cen, složitosti produktů a nemožnosti snadné meziproduktové komparace. Kupříkladu během studie [4] muselo být téměř 70 % bank dodatečně kontaktováno z důvodu nedostatečných nebo nejasných tarifních informací. Celkově je snaha spotřebitele získat přehled o cenách na RCBS trhu časově velmi náročným úkolem nejen vzhledem k objemu informací, či případným nekompletním informacím. Úkol je dále znesnadněn, především v posledních letech, také díky tarifním změnám zahrnujícím tzv. product-tying, volně přeloženo jako propojování produktů. Výsledkem snahy takového spotřebitele by byly potom pouze extrémně vysoké náklady obětované příležitosti.

Jednou z možností, jak řešit problém asymetrie informací a přidruženého problému nízké mobility klientů, je také nástroj zmíněný a doporučený

pro střední Evropu včetně České republiky (dále v textu pouze jako ČR) v [2], kterým je komparační internetová stránka. V ČR ovšem nevznikla komparační stránka ve formě tabulky srovnávající dostupnost služeb a jejich zpoplatnění u jednotlivých bank, ale primitivní expertní systém nazvaný Kalkulátor bankovních poplatků (dále v textu pouze jako Kalkulátor). Tento systém na základě uživatelské způsoby užití propočítá poplatky a doporučí nejvhodnější produkt, prozatím pouze, na základě kritéria ceny a spotřebitelem poptávaných služeb. Tento způsob má oproti pouhé srovnávací tabulce několik výhod, ale po stránce výzkumu je nejvýznamnější, že díky tomu systému je možné získat dostatečné velké vzorek z určité populace. Tento vzorek může být užít k dostatečně podrobnému popsání poptávkové strany pro další výzkumné účely i pro dodatečné zmírnění negativních dopadů asymetrie informací na stranu spotřebitelů. V rámci našeho výzkumu, zaměřeného právě na asymetrii informací v bankovníctví, bude užito tohoto zdroje dat pro identifikaci hlavních vzorců užití RCBS, resp. rozčlenění zkoumaných prvků do co nejvíce kompaktních a dobře separovaných shluků. Výsledky analýzy nebudou samoúčelné, či pouze pro deskriptivní účely. Bude jich využito v dalším výzkumu, ale také pro ještě snadnější produktovou komparaci, za účelem další redukce asymetrie informací na trhu RCBS v rámci ČR.

## 1. Cíl a metodika výzkumu

Cílem výzkumu je klasifikace klientů českých bankovních institucí a vymezení základních skupin ve vztahu k transakcím na klientských účtech. Článek navazuje a doplňuje text [16], kdy metodika předpokládala použití metody k-means za účelem rozčlenění respondentůvých odpovědí z pilotního provozu Kalkulátoru a nastínění dalších možností využití výstupů systému. Jeden z problémů, který se vyskytl v publikovaném textu, je asymetrie mezi počtem prováděných transakcí a koncentrací klientů s nízkým počtem transakcí a nízkým obrátem. Tato asymetrie měla za následek vytvoření jedno-

ho hlavního clusteru obsahujícího 68 % objektů, a několika menších, málo významných clusterů zahrnujících zejména klienty s extrémně vysokými počty transakcí a vysokým obrátem na účtu. Hledání vhodného počtu clusterů bylo přitom určováno na základě kritéria  $G_s$ . [9]

## 1.1 Sběr dat

Sběr dat je proveden v rámci projektu Kalkulátor bankovních poplatků, který je umístěn na serveru bankovnipoplatky.com. Je tedy nutno předpokládat, že šetření se omezuje na klienty s přístupem k internetu, a na určité úrovni informační gramotnosti. Formulář navíc žádným způsobem neověřuje jedinečnost respondentů, nelze se tedy vyvarovat opakovanému přístupu téhož respondenta. Přesto, že nelze v tomto ohledu považovat vzorek za reprezentativní v rámci celé populace klientů bankovních institucí, mají data svoji vypovídací hodnotu a vzhledem k charakteru a rozsahu pořízených údajů, jde o unikátní soubor dat, která prakticky nelze získat jiným způsobem. Navíc vzhledem k přidané hodnotě v podobě odhadu poplatků lze očekávat, že zájemce o kalkulaici zadá do kalkulátoru pravdivé hodnoty.

## 1.2 Příprava dat

Data pořízená pomocí online formuláře obsahují množství nedostatků v úplnosti dat. Přesto, že se formulář stále vyvíjí, prozatím se nepodařilo prosadit ověření platnosti vložených dat. Data tak obsahují množství chybných údajů a nevyplněných hodnot. Součástí přípravy dat bylo nutno provést

- 1) Eliminaci vybraných proměnných. Z důvodu velkého rozsahu sledovaných údajů, bylo provedeno vyloučení proměnných, které odpovídají nevyužívaným či málo využívaným službám. Lze předpokládat, že takové služby jen málo charakterizují standardní chování klientů, a navíc by použití takové proměnné ve shlukové analýze mohlo mít za následek vytváření umělých clusterů, kde rozhodujícím kritériem by byly právě málo využívané služby.
- 2) Odstranění záznamů s neplatnými hodnotami. Množství záznamů v databázi obsahovalo neplatné hodnoty či nevyplněné údaje. Z databáze byly odstraněny záznamy obsahující textové hodnoty v numerických proměnných. Jejich oprava by v některých případech sice

možná, nicméně však náročná a jejich ztráta v souboru navíc neznamená významnou změnu z hlediska struktury klientů. Současně jsou ze souboru odstraněny záznamy, které neodpovídají transakcím na běžných účtech, protože v rámci analýzy v [16] bylo zjištěno, že RCBS produkty jsou využívány i k podnikatelským účelům. To je nejen v rozporu se zaměřením aktuálního výzkumu, ale též ve většině případů i se smluvním vztahem banka–retailový klient. S ohledem na tento fakt byla vygenerována četnostní tabulka ke každé kvantitativní proměnné a individuálně posouzena tzv. cutoff četnost.

- 3) Náhrada chybějících hodnot. Literatura obvykle v tomto případě doporučuje náhradu průměrnou hodnotou proměnné, náhradu náhodným číslem s rozdělením proměnné, či náhradu dle regresního modelu v závislosti na ostatních proměnných, eventuálně objekt z datové matice vypustit. [9] V našem případě jde o specifický případ, kdy bylo nutno k jednotlivým proměnným přistupovat individuálně s přihlédnutím k jejich charakteru a významu v modelu. Je třeba vzít v úvahu fakt, že vynechání hodnoty v dotazníku může mít různé důvody a je tedy třeba rozlišit, zda je hodnota vynechána, protože klient službu nevyužívá, nebo zda je hodnota vynechána z důvodu opomenutím, či z jiného důvodu.
  - a) Náhrada hodnoty nulou. I když je tato operace nestandardní, má v tomto případě své opodstatnění. Lze očekávat, že pokud klient využívá pro běžné transakce např. telebankingu, nevyužívá pro tytéž transakce pobočku. Vzhledem k tomu je vynechání považováno za vyjádření: „Službu neuvžívám.“ a tudíž měly by chybějící být hodnoty nahrazeny nulou. Některé služby v Kalkulátoru, jako kroky VIII a IX (viz podkapitola 2.1) jsou evidentně doplňkové služby, kde by respondenta zadávání nulových hodnot do každé kolonky pouze zbytečně zdržovalo.
  - b) Regresní model. U proměnných, kde nelze použít náhradu nulou, je použit odhad na základě jednoduchého regresního modelu, kde vysvětlující proměnná bude stanovena z korelační matice. Použití vícerozměrné regrese je v našem případě neúčelné.

- c) Vyloučení záznamu. V případě, že počet zazdaných hodnot je nedostatečný, je záznam ze souboru odstraněn.

- hodnota vlastních čísel by měla být minimálně 1,0;
- popsána variabilita by měla činit alespoň 70 % původní variability;
- počet komponent by měl být výrazně nižší, než je původní počet proměnných [9], [13].

### 1.3 Transformace proměnných

Jak již bylo zmíněno v úvodu kapitoly, výsledky publikované v článku [16] se potýkaly s problematikou členění klientů v oblastech extrémních hodnot. Použití vhodné transformace umožní symetrizací rozdělení proměnných a potlačení významu vzdáleností u vysokých hodnot jednotlivých proměnných tak, aby při použití shlukové analýzy docházelo k detailnějšímu posouzení rozdílů v rámci většinové skupiny klientů aplikované na jednotlivé proměnné.

Část ze sledovaných proměnných, zahrnutých do analýzy, má rozdělení výrazně zešikmené doleva k nulové hodnotě, zpravidla s jediným vrcholem. U těchto proměnných se vhodnou ukázala být logaritmická transformace ve tvaru

$$X^0 = \ln(X+1) \quad (1)$$

### 1.4 Eliminace multikolinearity

Vzhledem k charakteru a počtu analyzovaných proměnných lze očekávat vzájemnou závislost jednotlivých proměnných. Multikolinearita by mohla velmi významně ovlivnit výslednou kvalitu shlukování a zařazení jednotlivých prvků do výsledných clusterů. Pro eliminaci multikolinearity je v práci použito metody hlavních komponent. Metoda hlavních komponent umožňuje diagnostiku vzájemné závislosti a náhradu stávajícího systému navzájem závislých proměnných novým systémem navzájem nekorelovaných komponent za současného snížení dimenze vektorového prostoru. Analýza je v rámci produktu IBM PASW 18 (dříve jako SPSS) zahrnuta pod Faktorovou analýzu.

Poněvadž nelze předpokládat rovnocennou váhu jednotlivých proměnných, kde variabilita některých proměnných se výrazně liší od ostatních, byla zvolena extrakce založená na korelační matici. Použití korelační matice při extrakci sice může komplikovat interpretaci komponent, v našem případě však vlastní interpretace komponent není tak zásadní.

Při určování optimálního počtu nových proměnných z nerotovaného řešení se opíráme o obvykle doporučovaná kritéria, založená na hodnotách vlastních čísel a celkové variabilitě popsané komponentami:

Vzhledem k tomu, že sběr dat je realizován dotazníkovým šetřením, lze předpokládat buď vyšší počet komponent, či nižší úroveň popsané variability původního systému.

### 1.5 Shluková analýza

Cílem celé práce je určení výrazných skupin zákazníků bankovních ústavů podle toho, jak využívají jednotlivé služby na svých klientských účtech. Vzhledem k tomuto cíli byla jako metoda zvolena shluková analýza, která se v praxi užívá pro segmentaci různých ekonomických subjektů, např. [12] v návaznosti na [11]. Ve srovnání s textem [16], je v této práci použita dvojkroková shluková analýza, která dovoluje místo eukleidovské vzdálenosti užití i věrohodnostní míry.

Dvojkroková shluková analýza je založena na použití CF (cluster feature) charakteristik, která je základem algoritmu BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) [17]. Algoritmus BIRCH postupuje ze základního členění podshluků v listech CF stromu tradiční metodou hierarchického shlukování [9]. Použitím CF charakteristik, reprezentujících jednotlivé skupiny dat abstrahuje metoda od jednotlivých případů uvnitř skupiny, což snižuje nároky na výpočetní výkon počítače, zejména u velmi rozsáhlých souborů.

Algoritmus dvojkrokového shlukování v PASW umožňuje dvě metody měření vzdáleností. Kromě tradiční eukleidovské míry, použitelné pouze pro spojitě proměnné, je do algoritmu implementována míra označená jako log-likelihood, založená na poklesu věrohodnostní funkce při slučování shluků, která je za určitých předpokladů na rozdělení proměnných použitelná pro spojitě i kategoriální proměnné. Tato metodika byla představena v [10] a dále popsána např. v [14]. Součástí algoritmu je navíc možnost nastavení automatické volby počtu shluků, přitom k odhadu optimálního počtu shluků je možné použít jedné ze dvou charakteristik; BIC (Bayesian's information criterion) a AIC (Akaike's information criterion). V rámci tohoto textu jsme použili míru nepodobnosti na

věrohodnostním poměru (log-likelihood) s automatickou volbu clusterů při použití BIC charakteristiky. Hlavním kritériem pro určení výsledného počtu shluků bylo, zda lze každý shluk odůvodnit a identifikovat s konkrétním chováním klientů, tj. zda se každý nově přidáný shluk významně liší od ostatních v určité charakteristické kategorii proměnných. Posouzení odlišnosti je provedeno na základě subjektivního kritéria použitelnosti. Redukce šumu nebyla využita vzhledem k riziku možného potlačení sice malého, ale validního shluku klientů preferujících vedle elektronického bankovníctví též zřizování služeb na pobočce.

## 2. Zdroj dat

### 2.1 Poslání systému Kalkulátor a proces získávání dat

Systém Kalkulátor byl vytvořen v rámci projektu Bankovnípoplatky.com, jehož cíle jsou dle zakladatelů kultivace vztahu banka–klient a klient–banka, snižování bankovních poplatků a zvyšování „bankovní“ vzdělanosti. Posláním Kalkulátoru je být snadno dostupným, nezávislým, zdarma poskytovaným nástrojem pro srovnání nabídek bank působících na území ČR. Snadná dostupnost zdarma je plněna fungováním internetového rozhraní Kalkulátoru 24 hodin denně (kromě plánovaných odstavek) bez nutnosti poplatků, či registrací. Nezávislost nejen na bankách, ale i na jiných komerčních nebo veřejných subjektech je pečlivě plněna a dohlíženo v rámci celého projektu Bankovnípoplatky.com. Jedinou reklamou je Google kontextová reklama, kde reklamní odkazy generuje robot a nelze proto články a aktivity Bankovnípoplatky.com podezřívát z nedostatečné objektivnosti. Srovnání nabídek bank působících v ČR je plněno sledováním 48 produktů u 14 bank. Sice v ČR působí více bank, resp. poboček zahraničních bank, ale pokud se tedy angažují na RCBS trhu, je jejich tržní podíl extrémně nízký. To souvisí také s tržním modelem RCBS v ČR, který je při výpočtu Herfindahl-Hirschmanovým indexem trhem oligopolním (výpočet nebyl založen jako obvykle na bilančních sumách, ale na mnohem vhodnějším ukazateli pro hodnocení RCBS trhu – počty klientů). Výše popsaným posláním je naplněna jedna z klíčových myšlenek, resp. doporučení EU pro tento trh, vydaná ještě před koncem funkčního období Charlese McCreevyho na postu euro-

komisaře pro vnitřní trh, který věnoval retailovému bankovníctví velikou pozornost (zodpovědný např. za vypracování [5]).

Získávání dat je přímo propojeno s užitím Kalkulátoru spotřebitelem. Uživatelské rozhraní je dostupné na URL: <http://www.bankovnipoplatky.com/kalkulator.html>, kde je postupně odpovídáno na otázky ohledně využívání RCBS vyplňováním webového dotazníku. Kalkulátor je na základě logické posloupnosti rozčleněn na devět částí, které celkem obsahují 54 otázek (25 otázek s přidruženými podotázkami, 2 doplňkové otázky a dotaz zodpovídaný formou textového pole) vztahen k měsíčnímu užívání RCBS:

- I. účet,
- II. výpisy,
- III. karetní služby,
- IV. přímé bankovníctví,
- V. platby – jednorázové,
- VI. platby – trvalé příkazy,
- VII. platby – povolení k inkasu (včetně služby soustředěné inkaso plateb obyvательства),
- VIII. hotovostní operace,
- IX. další služby.

Na závěr Kalkulátor obsahuje dvě otázky na aktuálně využívanou banku a typ účtu. Všechna využití Kalkulátoru jsou ukládána do databáze odpovědí respondentů a na základě SQL dotazů mohou být generovány do tabulek. Sběr je dat je anonymní a není tedy potřeba brát v potaz zákon č. 101/2000 Sb. Do databáze se neukládají IP adresy a po spotřebiteli, resp. respondentovi není požadován žádný identifikační údaj. Všechny prvky (jednotliví respondenti) jsou jednoznačně identifikovány pořadovým identifikátorem a dodatečně časem odeslání požadavku na výpočet.

Tento způsob sběru dat přináší svá omezení a rizika, která budou komentována v následující části. Přesto ve srovnání s jinými výzkumy založenými na primárních subjektivních datech se u dat v Kalkulátoru dá očekávat vyšší vypovídací hodnota, resp. vyšší věrnost odpovědí, než je u dat získaných přímým dotazováním v terénu (vlak, ulice) nebo sběrem za užití prostředků dálkové komunikace (e-mail, telefon). Na rozdíl od těchto výzkumů totiž vyplněním formuláře Kalkulátoru a následným výpočtem se tvoří přidaná hodnota především samotnému respondentovi. Respondent užívá Kalkulátor za účelem získání nákladového a funkčního přehledu na RCBS trhu kritériem svého individuálního způsobu využívání

RCBS. Potom je tedy primárně v respondentově zájmu zadat co nejpřesnější a co nejpravdivější údaje. Ve výzkumech za užití prostředků dálkové komunikace není klient motivován k přesným a úplným odpovědím, protože jakákoliv jeho odpověď je zaznamenána bez zpětné vazby směrem k respondentovi. Faktor vnitřní motivace respondenta při této metodě sběru dat působí významně na to, aby respondent věrně vyplnil vše, co je pro něj podstatné a tím zároveň pomohl následnému výzkumu poskytnutím pravdivých a úplných dat.

## 2.2 Zdrojová data a vymezení populace

Fyzickým zdrojem dat pro výzkum byla databáze uložených odpovědí respondentů Kalkulátoru stručně popsaná v předchozí části. Z hlediska marketingového výzkumu se jedná o data:

- vicerozměrná – je sledováno 54 proměnných o užívání RCBS vyplňované respondentem a 2 proměnné přiřazené systémem, z toho:
  - 43 intervalových proměnných – četnosti měsíčního užívání RCBS případně částky se službou spojené,
  - 6 dichotomických proměnných – přítomnost určitých služeb (karetní služby, e-banking apod.),
  - 4 nominální proměnné – forma a frekvence zasílání výpisů, aktuálně užívaná banka a aktuálně užívaný produkt,
  - 1 textová proměnná – textové pole pro zaplání připomínek a návrhů pro Kalkulátor,
  - 2 ordinální proměnné – systémové proměnné přidělování pořadového čísla jednotlivým respondentům a čas, kdy bylo provedeno uložení odpovědi.
- primární – byla získána přímo od spotřebitelů RCBS,
- subjektivní – data vychází z úsudku spotřebitele o vlastním užívání RCBS.

Ve sledovaném období 2. 1. 2010 – 31. 12. 2010 činil počet užití Kalkulátoru, resp. respondentů 17 906. Metoda sběru dat je rozhodně vzdálena od optimální metody prostého náhodného výběru respektujícímu v rámci skupin strukturu populace (po stránce geografické, věkové, příjmové apod.). Vzhledem k tomu nelze považovat zkoumaný výběr z populace za reprezentativní pro populaci bankovních klientů ČR i přes vysoký

počet respondentů. Jedna ze základních disproporcí mezi výběrem a reálnou bankovní populací je v zastoupení účtů s aktivovaným přístupem k PC na celku. Dle posledního šetření v rámci [6] je relativní četnost tohoto znaku v RCBS populaci ČR 69 %. V databázi Kalkulátoru je podíl prvků s tímto znakem vyšší než 84 %. Expertním odhadem výzkumného týmu je (vzhledem k neexistenci dat pro srovnání nelze ověřit testováním hypotéz), že zkoumaný výběr obsahuje zkreslení směrem k mírně aktivnějším klientům s preferencí internetového bankovníctví. Naopak se výsledky odchyľují od pasivních klientů s pobočkovými preferencemi (např. důchodci, klienti s nedůvěrou v bezpečnost internetového bankovníctví). Celkově mezi hlavní limitující faktory (kromě geografických a demografických faktorů) determinující populaci respondentů Kalkulátoru patří přístup na Internet. Tento předpoklad je výrazněji limitujícím především ze socio-demografického hlediska. Přesto vzhledem k tomu, že v roce 2009 činil podíl populace v produktivním věku, která v tomto roce komunikovala přes Internet 92 % [7], nejde o velmi výrazné omezení. Ve skupině s důchodovým věkem je toto číslo ostře nižší.

Vzhledem k výše prezentovaným faktům se analýza vztahuje pouze na klienty s aktivovaným přístupem k účtu přes PC, resp. internetovým bankovníctvím. Všechny následné výsledky a závěry jsou tedy platné pouze pro tuto populaci v rámci ČR v produktivním věku 16–74 let.

## 3. Realizace metodiky

### 3.1 Příprava surových dat

V rámci zdrojových dat byly nejprve vydefinovány typy proměnných, resp. škály dle typů proměnných (viz podkapitola 2.2) a také jejich povolené hodnoty, konkrétně vyřazení prvků se zápornými hodnotami proměnných kromě průměrného zůstatku, vyřazení prvků s hodnotami intervalových proměnných vyjadřujících četnost užití jednotlivých RCBS nad 60 měsíčně. Následně bylo odstraněno přes 2 700 prvků, kde odpověď na otázku aktivního internetového bankovníctví byla negativní nebo byla uložena vůbec. Na zbylých 15 000 prvcích bylo následně řešeno, které proměnné se budou účastnit analýzy. Ohledně kategoriálních proměnných bylo rozhodnuto, že nemají pro analýzu většího významu. Pro kategori-

ální – dichotomické proměnné ohledně přímého bankovníctví byly již v rekódovací fázi prvky filtrovány, resp. zvoleny jen ty s aktivovaným internetovým bankovníctvím. Kvalitativní proměnné ohledně formy a frekvence zasilání výpisů nemají na celkové zpoplatnění RCBS ze strany bank také vliv a to od listopadu 2009 vzhledem k přijetí taxativního výčtu služeb se zakázaným zpoplatněním ze strany bank v rámci harmonizace práva ČR s právem EU (podrobněji viz [15]). Dvojice kategoriálních proměnných týkající se aktuální banky a aktuálně užívaného produktu nebyla zvolena, z důvodu pozdějšího narušení shlukování, kdy např. klienti poptávající shodně nízkonákladové RCBS volí v drtivé většině mezi 3mi účty – tedy u klientů se shodným vzorcem chování je přítomen rozdílný znak způsobující rozpor ve shlukování.

Poslední nekvantitativní proměnná zadávaná respondentem je textové pole se zprávou nebo doporučeními pro tým Kalkulátoru. Většinou respondentů není využívána. Ohledně proměnných kvantitativních byla šetřena četnost nul, či nevyplněných hodnot, tedy bylo šetřeno využívání dané služby. Je sice pravdou, že vyřazením minimálně užívaných proměnných je možné přijít o identifikaci málo početné, ale velmi specifické skupiny prvků, ovšem je nutné respektovat fakt, že shluková analýza je velmi citlivá na nevýznamné znaky [9]. Z šetření četností jasně vyplynulo, že služby s komunikačním kanálem telebanking a sběrný box nejsou vhodné pro další analýzu, protože se nevyužívaly v téměř 99 % případů. Stejně tak byly vyřazeny služby spojené s výměnou hotovosti, příjmem nadměrného vkladu v hotovosti a výdej hotovosti v rámci služby cash-back, která je sice v západní Evropě téměř denně užívanou službou, ale RCBS spotřebitel v ČR často ani netuší, k čemu slouží, či kde je poskytována. Dále byla vyřazena část spojená s užíváním ATM vlastní a ATM cizí banky v zahraničí spolu s přidruženou otázkou na průměrnou vybíranou částku v zahraničí. Závěrem tedy bylo zvoleno 19 proměnných pro další analýzu, resp. jako budoucí identifikátory vzorců využívání RCBS. Pro přesný výčet proměnných, které byly do další analýzy využity, shlédněte prosím tabulku v podkapitole 3.3.

Volba proměnných byla následována rekódováním, které mělo několik fází, podrobněji viz Metodika. Ohledně specifické metody vyplňování prázdných polí nulou je konkrétním příkladem

zřídka využívané služby např. užívání pobočky pro zřízení inkasní platby, sběrný box nebo příjem mincí. Z předchozích výzkumů na datech pilotního provozu Kalkulátoru [16] bylo identifikováno, že většinu respondentů zajímá, jinými slovy je využívána, pouze necelá polovina otázek, resp. služeb. Z tohoto důvodu byly prázdné odpovědi nahrazeny nulou, kromě otázek, kde je evidentní, že nevyplnění nemůže být bráno jako „nevyužívám“ (např. dotazy na minimální obrát a průměrný zůstatek).

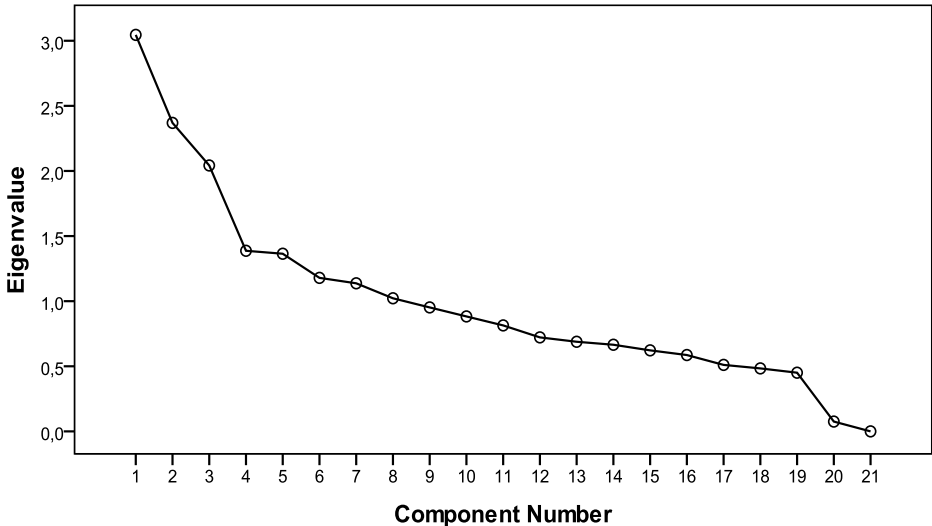
Pro doplnění chybějící dat v proměnné průměrná částka výběru z ATM cizí banky v případě, že respondentem byla četnost výběrů, byla užita nelineární regrese z důvodu snahy o zachování struktury dat. Výběr proměnné pro regresi proběhl na základě výpočtu koeficientů korelace mezi proměnnou Výběr\_ATM\_cizí\_ČR\_částka a ostatními proměnnými. Nejvyšší koeficient korelace vykazovala proměnná Výběr\_ATM\_cizí\_ČR. Ostatní koeficienty byly ostře nižší a pouze Výběr\_ATM\_vlastní\_ČR a JPÚ\_do\_cizí\_Inet (JPÚ je nadále užívanou zkratkou pro jednorázový příkaz k úhradě) vykazovaly velmi slabou korelaci -0,12 a 0,13. Výpočtem byly zjištěny předpokládané hodnoty a doplněny do souboru. V rámci rekódování bylo dále ověřováno, zda v případě, že respondent deklaruje užívání RCBS spojených s platebními kartami, také opravdu jednu z karet zvolil (jsou sledovány druhy kreditní, debetní embosovaná, debetní neembosovaná). Konkrétně se jedná o služby výběry z bankomatu (dále v textu pouze jako ATM) vlastní nebo cizí banky a cash-back.

Tím byla ukončena rekódovací fáze. Jejím důsledkem bylo, že následně užitě metody (metoda hlavních komponent a shluková analýza), pro něž bylo zvoleno nastavení listwise, nebraly v potaz více než 6 000 prvků, resp. byly prováděny na přesně 8 549 prvcích. Číslo 6 000 se může zdát vysoké, ovšem je nutné zohlednit potřebu listwise principu (vysvětlen např. v [9]) a také velmi specifickou formu získávání dat prostřednictvím Internetu.

### 3.2 Transformace a snížení dimenze

Vzhledem k vysoké dimenzionalitě (19 proměnných) a jednoho z předpokladů pro následnou shlukovou analýzu ve smyslu nekorelovanosti proměnných, bylo nutné posoudit vzájemné vztahy mezi proměnnými. Z korelační matice bylo zřejmé, že mezi některými proměnnými lze před-

Obr. 1: Scree plot graf



Zdroj: vlastní výzkum

pokládat významnou míru závislosti. Například u proměnných JPÚ\_do\_vlastní\_Inet-JPÚ\_do\_cizí\_Inet a také u již dříve zmíněných proměnných ohledně četnosti výběru z ATM cizí banky a průměrné vybírané částky dosahovaly korelační koeficienty poměrně výrazných hodnot. Z tohoto důvodu bylo nutné přistoupit k redukci multikolinearity viz metodika. Podíly na vysvětleném rozptylu dat je možné sledovat zde ve scree plot grafu (obr. 1).

Jak je vidět z grafu, tak při uplatnění pravidla o zlomu scree plot grafu např. v [9], [13] by měly být do dalších analýz vybrány čtyři komponenty z důvodu evidentního zlomu v příspěvku ke kumulativní vysvětlené variabilitě, který se nachází mezi čtvrtou a pátou komponentou. Tímto zlomem končí pomyslná lineární funkce s negativní směrnici. Druhá, tentokrát s mírnější směrnici, je viditelná od šesté do dvanácté komponenty. Přestože je zlom evidentní, je nutné zohlednit nejen scree plot graf a jeho, jak oba zmíněné zdroje akcentují, často velmi individuální interpretaci. Z toho důvodu byla užita pravidla popsaná v metodice. Jednotlivé hodnoty transformovaných proměnných byly uloženy v rámci komponent a tím fází řešící předpoklady pro shlukovou analýzu ve smyslu rozdělení a nekorelovanosti byly splněny.

### 3.3 Shlukování

Myšlenka na vytvoření určitého počtu klientských profilů byla převzata z marketingových výzkumů prováděných v letech 2005–2009 firmou Capgemini v rámci dokumentů World retail banking report, pro podrobnější přehled o cíli analýzy, zkoumané populaci a výsledcích shlédněte prosím první kapitolu v [1]. V rámci těchto výzkumů bylo ze začátku užíváno dělení na tři základní profily, dle frekvence užití – tedy méně aktivní, aktivní a velmi aktivní. V posledním dokumentu v roce 2009 (World retail banking report byl vydán i za rok 2010, ovšem RCBS sekce již chybí a více prostoru se věnovalo tématu hypotečních a pojistných trhů) bylo užito již dělení i na základě komunikačního kanálu, resp. klient s pobočkovou a klient s internetovou preferencí. Prvotní předpoklad našeho výzkumu se ohledně pravděpodobného výsledku odvíjel od tohoto rozčlenění a bylo tedy očekáváno vytvoření tří shluků klientů s preferencí internetového bankovníctví.

Vzhledem k rozměru úlohy a především zkoumanému výběru o počtu prvků 8 549 bylo uvažováno pouze o nehierarchických metodách nebo modifikovaných hierarchických. Zvolena byla modifikovaná hierarchická metoda využívající CF strom v implementovaná v PASW 18 v rámci dvoukrokové

shlukové analýzy. Vstupními proměnnými do analýzy bylo osm komponent z předchozích fází výpočtu.

Výsledkem výpočtu bylo rozřazení prvků do tří shluků, což se shodovalo s úvodním předpokladem inspirovaným výzkumy [1]. V rámci jednotlivých shluků byl následně vypočítán vektor průměrných hodnot původních proměnných, resp. centroid shluku vyjádřený v původní škále. Centroidy byly poté analyzovány a interpretovány viz následující kapitola. Následně bylo překročeno i analýze s pevným počtem shluků: čtyři a pět shluků. Pro všechny analýzy se užilo stejného nastavení (kromě již zmíněného pevného počtu shluků). Hodnocení kvality rozkladu siluetou je sice o něco nižší než u rozkladu s optimálním počtem shluků dle BIC. Rozdíl ale není natolik závažný, a případný vyšší počet smysluplných shluků více odpovídá specifickým požadavkům

kladeným na analýzu problému. Vzhledem k tomu bylo za uspokojivé řešení přijata, vedle analýzy se třemi shluky, také varianta se čtyřmi shluky. Dodatečný čtvrtý shluk byl jednoduše interpretovatelný, jasně odlišitelný a nebyl nízkoprvkovým shlukem (na tvorbu těchto shluků je upozorováno např. v [9], kde při vysokém počtu prvků a nadbytečně vysokém počtu shluků hrozí tvorba nízkoprvkových shluků složených z odlehklých pozorování prakticky bez interpretačního významu). Tato situace ovšem nenastala u varianty s pěti shluky. Nově vzniklý shluk sice nebylo možné z hlediska velikosti považovat za zcela zanedbatelný, nicméně jeho interpretace byla poměrně komplikovaná a odlišená od nejbližších skupin nevýznamně. Závěrem této kapitoly je tabulka centroidů z obou přijatých variant shlukování, interpretace viz následující kapitola s výsledky.

**Tab. 1: Charakteristické užívání RCBS v rámci jednotlivých shluků (Part 1)**

Shluk	Varianta A (3 shluky)			Varianta B (4 shluky)			
	1.	2.	3.	1.	2.	3.	4.
Počet prvků shluku	5 276	2 922	351	2 295	3 000	2 908	346
Relativní podíl shluku	62 %	34 %	4 %	27 %	35 %	34 %	4 %
Minimální_obrat_částka	21 120	23 165	17 409	23 212	19 458	23 213	17 578
Průměrný_zůstatek_částka	22 103	18 212	23 428	23 870	20 657	18 258	23 723
Výběr_ATM_vlastní_ČR	2,9	2,8	3,1	2,8	3,0	2,8	3,1
Výběr_ATM_cizí_ČR	0	1,7	0,5	0	0	1,7	0,6
Výběr_ATM_cizí_ČR_částka	0	2 755	815	0	0	2 763	827
Příchozí_platba_z_cizí	2,2	2,5	2,0	2,1	2,3	2,5	2,0
Příchozí_platba_z_vlastní	0,9	1,1	1,0	0,4	1,3	1,1	1,0
JPÚ_do_vlastní_pobočka	0	0	0,3	0	0	0	0,3
JPÚ_do_vlastní_inet	1,8	2,0	1,8	1,0	2,4	2,0	1,8
JPÚ_do_cizí_pobočka	0	0	0,3	0	0	0	0,3
JPÚ_do_cizí_inet	3,8	4,5	3,2	4,0	3,6	4,5	3,2
TPÚ_do_vlastní_pobočka	0	0	0,2	0	0	0	0,2
TPÚ_do_vlastní_inet	0,9	0,9	0,7	0,1	1,6	0,9	0,8
TPÚ_do_cizí_pobočka	0	0	0,2	0	0	0	0,2
TPÚ_do_cizí_inet	2,4	2,7	1,7	2,4	2,4	2,8	1,7



Tab. 1: Charakteristické užívání RCBS v rámci jednotlivých shluků (Part 2)

	Varianta A (3 shluky)			Varianta B (4 shluky)			
Inkaso_do_vlastní_pobočka	0	0	0,2	0	0	0	0,2
Inkaso_do_vlastní_inet	0,3	0,3	0,2	0,1	0,5	0,3	0,2
Inkaso_do_cizí_pobočka	0	0	0,3	0	0	0	0,3
Inkaso_do_cizí_inet	0,9	0,9	0,5	0,9	0,9	0,9	0,5
Vklad_hotovost_pobočka	0,2	0,3	0,5	0,1	0,3	0,3	0,5
Výběr_hotovost_pobočka	0,1	0,1	0,3	0	0,1	0,1	0,3

Zdroj: vlastní výzkum

Vysvětlivky k tabulce (tyto zkratky budou dále užívány i v textu):

- všechny údaje se vztahují k měsíčnímu užívání,
- všechny částky jsou uváděny v Kč,
- syntax kvantitativních proměnných vztahených k četnosti užití je – „typ služby“\_“banka kam je služba platbou směřována“\_“komunikační kanál užitý pro zadání služby bance respondent“.
- užité zkratky v tabulce:
  - ATM – bankomat,
  - JPÚ – jednorázový příkaz k úhradě,
  - TPÚ – trvalý příkaz k úhradě,
  - Inet – Internetové bankovníctví,
  - Inkaso – inkasní služby včetně služby soustředěné inkaso plateb obyvatelstva (SIPO), kde je jedno inkasní spojovací číslo počítáno jako jedna operace nezávisle na tom, kolik operací je pod tímto spojovacím číslem prováděno.

## 4. Shrnutí výsledků

### 4.1 Identifikované profily užití

V předchozí tabulce s centroidy jednotlivých shluků jsou v první variantě shlukování, kde byla nastavena automatická volba počtu shluků, identifikovány tři profily:

- Tento profil byl označen jako průměrný klient internetového bankovníctví, resp. mainstream tvořící více než 62 % celku. Jeho preference tohoto informačního kanálu je velmi výrazná. Tento klient nevybírá z ATM cizí banky a lze tedy tvrdit, že v rámci těchto klientů byla poplatková diverzifikace, resp. postupně zdra-

žování výběrů z ATM cizích bank, úspěšná. Podíl u převodových služeb mezi odesláním prostředků na účty vlastních a cizích bank je přibližně 1:2. Celkově jsou v průměru měsíčně provedeny přibližně tři výběry z ATM, podáno téměř šest JPÚ, tři TPÚ, jedna inkasní platba a přibližně jednou za pět měsíců vklad nebo výběr hotovosti na pobočce.

- Tento profil byl označen jako aktivnější klient, tvořící třetinu celku. S průměrným klientem sdílí silnou preferenci ohledně komunikačního kanálu. Hlavním rozdílem je vyšší četnost vybírání z ATM a to i cizí banky. Jeho aktivita je pouze mírněji vyšší v JPÚ. V ostatních operacích se s drobnou tolerancí s průměrným klientem téměř shoduje.
- Tento profil byl označen jako průměrný klient s pobočkovou preferencí. Sice jde o relativně malý shluk (4 %), přesto ho nelze označit za nízkoprvkový shluk složený pouze z odlehklých pozorování bez interpretační hodnoty. Jeho aktivita je prakticky shodná s průměrným klientem, ale zásadním rozdílem je volba komunikačního kanálu, kde je preferována občasná návštěva pobočky banky v průměru jednou za tři měsíce. Existuje tedy skupina klientů s aktivním internetovým bankovníctvím, kterými je stále i pro běžné operace občasně preferována návštěva pobočky.

U druhé varianty shlukování by se mohlo zdát, že dodatečná analýza nepřinesla tolik nového. Přesto s ohledem k především následným poplatkům se nové rozčlenění může projevit. Došlo totiž k oddělení pasivnějšího klienta od mainstreamového shluku. Tím byl tento shluk v četnos-

tech posunut směrem k aktivnějšímu klientovi, až byl tento rozdíl prakticky smazán. Odlišujícím kritériem se poté stala volba mezi službami s napažením na vlastní, nebo cizí banku. Většina bank rozlišuje, kam a odkud byly prostředky odeslány, resp. při operacích v rámci svých poboček je poplatek za služby nižší nebo nulový. Záleží ovšem na dalších faktorech jako cenová politika (tradiční versus nízkonákladové banky) a komunikační kanál užití při zřízení služby (zřízení na pobočce versus elektronické bankovníctví). Ve variantě se čtyřmi shluky byly identifikovány následující profily:

1. Tento profil byl označen jako pasivnější klient internetového bankovníctví se svým 27% podílem. Celkově se od průměrného klienta (shluk č. 2) liší nižší frekvencí užití JPÚ, tak TPÚ. Tento také prakticky nenavštěvuje pobočku kvůli vkladu, či výběru hotovosti – tyto služby se využívají méně jak jedenkrát ročně.
2. Tento profil byl označen jako průměrný klient internetového bankovníctví, resp. mainstream tvořící více než 35 % celku. V rámci této varianty shlukování by mohl být označen za klienta „s preferencí vlastní banky“. Toto označení by odpovídalo nejen výběrům z ATM pouze vlastní banky, ale též tomu, že má nejvyšší podíl v platebních operacích směřovaných do vlastní banky. Stejně jako v předchozí analýze je i zde jasná preference internetového bankovníctví. V průměru tedy měsíčně provede přibližně tři výběry z ATM, šest JPÚ, čtyři TPÚ, jedno inkaso a přibližně jednou za tři měsíce navštíví pobočku kvůli vkladu hotovosti, výběry hotovosti na pobočce jsou ještě méně četnější.
3. Tento profil byl označen jako průměrný vicebankovní klient, tvořící třetinu celku. S průměrným klientem sdílí silnou preferenci ohledně komunikačního kanálu. Hlavním rozdílem je vyšší četnost vybirání z ATM a to i cizí banky a to přibližně dva výběry měsíčně. Dalším rozdílem je nejvyšší podíl operací směřovaných do cizí banky na rozdíl klienta průměrného. Rozdíl v aktivitě, resp. četnostech užití, jsou nízké kolem jednoho užití za dva měsíce.
4. Tento profil byl označen jako průměrný klient s pobočkovou preferencí. Velikostí i objemem prováděných operací prakticky odpovídá posledního profilu v tříshlukové variantě.

## 4.2 Další výzkum

Jak již bylo naznačeno v předchozích fázích článku, tak tato analýza není pouze izolovanou snahou o deskripci poptávkové strany RCBS trhu v ČR. Zpětně mohou být shluky užity např. tím, že bude pravidelně zveřejňováno, jaké náklady by byly u jednotlivých bankovních produktů. Klient banky by pouze na začátku identifikoval, který shluk je mu nejbližší a nadále by už jen v případě potřeby sledoval tabulku s náklady svého shluku místo zdoluhavého vyplňování formuláře Kalkulátoru, či ještě mnohem zdoluhavějšího pročítání sazebníků bank. Míra redukce asymetrie informací by tedy byla pouze závislá na míře shody individuálního vzorce užití klienta s nejbližším vzorcem reprezentovaným centroidem některého ze shluků. Pravidelného sledování zpoplatnění se dá následně využít při snaze o monitorování cenové hladiny RCBS služeb.

Dále by bylo možné zohlednit i jeden z předpokladů poněkud odlišněji zaměřeného výzkumu [8], že v centrální Evropě (ČR, Slovensko, Maďarsko, Polsko) existuje významná podobnost ve spotřebních zvycích a potřebách, pohledu na roli peněz a v konečném důsledku také na chování bankovních klientů. V tom případě by se výsledky z výzkumů na klientech ČR mohli s určitou mírou odchylky převzít jako vzorec s platností pro výše zmíněnou čtveřici. Vzhledem ke snahám EU v rámci propojování a redukce bariér národních trhů, též deklarovaných v [5] a [3] by pak identifikace nejčastějších klientských profilů a následné informování o jejich zpoplatnění v rámci centrální Evropy bude nejen zdrojem dat pro komparaci národních RCBS trhů, ale především usnadnění výběru z nabídek bank na mezinárodním trhu centrální Evropy.

Pokud bychom se zaměřili pouze na ČR, skýtají analyzovaná data možnost odpovědi na další otázku ohledně chování RCBS klientů. Jednou z těchto otázek je testování hypotéz ohledně aktivit klientů nízkonákladového bankovníctví v porovnání s klienty tzv. velké trojice Česká spořitelna a.s., Komerční banka a.s., Československá obchodní banka a.s. Dále může být opět zkoumána asymetrie informací, resp. nepřímo měřena zjišťováním měrou adverzního výběru vzhledem k proměnné celkové náklady a poptávané služby.

## Závěr

V rámci článku byl komentován přínos analýzy strany poptávky na trhu RCBS a následně analýza

provedena za účelem identifikace hlavních vzorců užití RCBS. Je sice pravdou, že výzkum založený na subjektivních datech ze systému Kalkulátor má své limity, protože klient nemusí sám objektivně posoudit své RCBS zvyklosti, přesto není jiné volby, pokud je potřeba obsáhlý výběr z populace bez nutnosti platit komerční sběr dat. Totiž vzhledem k bankovnímu tajemství není možné získat neagregovaná data a i po anonymizaci dat je jejich získání nereálné z důvodu ochrany tentokrát obchodního tajemství, protože každou bankou je z pochopitelných důvodů považována znalost svého tržního podílu jako součást jejich know-how. Zpracovávání agregovaných dat, jejichž pravděpodobnější získání je již vyšší, nemá pro podrobnější výzkum většího významu. Průzkumy přístupné veřejnosti ohledně průměrného klienta založené právě na agregovaných datech bank se konaly od roku 2000 již několikrát, ale vzhledem k variabilitě v populaci, resp. náhlylosti průměru k vychýlení, mají spíše orientační charakter.

V rámci populace ČR klientů s aktivním internetovým bankovníctvím, v produktivním věku 16–74 let s pravděpodobným alespoň minimálním zájmem o své RCBS byl shlukovou analýzou zkoumán, po předchozích nutných přípravách surových dat, výběr o velikosti 8 549 prvků. V tomto výběru byly identifikovány v první variantě výpočtu tři shluky a ve druhé variantě čtyři shluky. Průměrný klient první varianty tvořil více než 62 % celku. Tento klient nevybírá z ATM cizí banky a lze tedy tvrdit, že v rámci těchto klientů byla poplatková diverzifikace, resp. postupné zdražování výběrů z ATM cizích bank, úspěšná. Podíl u převodových služeb mezi odesláním prostředků na účty vlastních a cizích bank je přibližně 1:2. Celkově jsou v průměru za měsíc provedeny 3 výběry z ATM, téměř šest JPÚ, tři TPÚ, jedna a přibližně jednou za pět měsíců návštěva pobočky banky za účelem výběru či vkladu hotovosti. Ve druhé variantě byl průměrný klient rozdělen na shluk s vyšším poměrem směřování služeb do vlastní banky a na shluk se směřováním služeb do banky cizí. Oba shluky byly velikostně srovnatelné kolem 34 %. Vzhledem k tomu, že vzorek, jak bylo zmíněno v úvodu práce, nelze považovat za reprezentativní, neodráží velikost shluků přesné rozložení klientů v populaci RCBS klientů ČR. Lze však předpokládat, že vytvořené profily poměrně dobře popisují chování klientů s aktivním elektronickým bankovníctvím bez ohledu na velikost skupin.

Výsledky mohou být dále užity za účelem snížení asymetrie informací, resp. nákladů na získání přehledu o potenciálních cenách na trhu RCBS, což je jeden z problémů přímo, či nepřímo akcentovaných orgány EU v rámci dokumentů [5], [4], [2], [3].

Článek byl zpracován v rámci projektu specifického výzkumu „Adverzní výběr v prostředí retailového bankovníctví“ číslo 2105 z výdajů na výzkum a vývoj Ministerstva školství, mládeže a tělovýchovy ČR.

## Literatura

- [1] Capgemini worldwide. *World retail banking report 2009* [online]. Capgemini, 2009. [cit. 2011-08-06]. 56 s. (PDF). Dostupné z: <[http://www.at.capgemini.com/m/at/tl/World\\_Retail\\_Banking\\_Report\\_2009.pdf](http://www.at.capgemini.com/m/at/tl/World_Retail_Banking_Report_2009.pdf)>.
- [2] Commission of the European communities. *Retail financial services to the consumer markets scorecards*. [online]. Brussels: Commission of the European communities, 2009-09-22 [cit. 2011-07-28]. 33 s. (PDF). Dostupné z: <[http://ec.europa.eu/consumers/rights/docs/swd\\_retail\\_fin\\_services\\_en.pdf](http://ec.europa.eu/consumers/rights/docs/swd_retail_fin_services_en.pdf)>.
- [3] Commission of the European communities. *Sector Inquiry under Article 17 of Regulation (EC) No 1/2003 on retail banking (Final Report)*. [online]. Brussels: Commission of the European communities, 2007-01-31. [cit. 2011-07-26]. 11 s. (PDF). Dostupné z: <[eur-lex.europa.eu/LexUriServ/site/en/com/2007/com2007\\_0033en01.pdf](http://eur-lex.europa.eu/LexUriServ/site/en/com/2007/com2007_0033en01.pdf)>.
- [4] Commission of the European communities. *SEPA monitoring study*. [online]. Brussels: Commission of the European communities, 2008-08-18. [cit. 2011-07-28]. 115 s. (PDF). Dostupné z: <[http://ec.europa.eu/consumers/rights/docs/SEPA\\_monitoring\\_study.pdf](http://ec.europa.eu/consumers/rights/docs/SEPA_monitoring_study.pdf)>.
- [5] Commission of the European communities. *The green Paper on Retail Financial Services in the Single Market*. [online]. Brussels: Commission of the European communities, 2007-04-30. [cit. 2011-07-26]. 20 s. (PDF). Dostupné z: <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2007:0226:FIN:EN:PDF>>.
- [6] Česká národní banka. *Zpráva o výkonu dohledu nad finančním thém*. [online]. Praha: Česká národní banka, 2011-06-09. [cit. 2011-08-03]. 137 s. (PDF). Dostupné z: <<http://www.cnb.cz/>>

- miranda2/export/sites/www.cnb.cz/cs/dohled\_financni\_trh/souhrnne\_informace\_fin\_trhy/zpravy\_o\_vykonu\_dohledu/download/dnft\_2010\_cz.pdf>. ISBN 978-80-87225-31-8.
- [7] Český statistický úřad. *Internet a komunikace* [online]. Praha: Český statistický úřad, 2010-08-12 [cit. 2011-08-06]. Dostupné z: <[http://www.czso.cz/csu/redakce.nsf/i/internet\\_a\\_komunikace](http://www.czso.cz/csu/redakce.nsf/i/internet_a_komunikace)>.
- [8] DVOŘÁK, P., HANOUSEK, J. Paying for Banking Services: What Determines the Fees?. *CERGE-EI Working Paper Series* [online]. 2009, č. 388. [cit. 2011-08-06]. Dostupné z: <[www.cerge-ei.cz/pdf/wp/Wp388.pdf](http://www.cerge-ei.cz/pdf/wp/Wp388.pdf)>. ISSN 1211-3298.
- [9] HEBÁK, P. a kol. *Vícerozměrné statistické metody (3)*. 1. vyd. Praha: Informatorium, 2005. 255 s. ISBN 80-7333-039-3.
- [10] CHIU, T. D. et al. A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*. San Francisco, 2001.
- [11] KRAFTOVÁ, I., KUBANOVÁ, J. Využití shlukové analýzy při deskripci firem regionu s akcentem na jejich produktivitu a kapitálovou sílu. *E+M Ekonomie a Management*. 2003, roč. 5, č. 1, s. 87-92. ISSN 1212-3609.
- [12] KRAFTOVÁ, I., KUBANOVÁ, J. Využití shlukové analýzy při deskripci firem regionu s akcentem na jejich velikostní charakteristiky. *E+M Ekonomie a Management*. 2002, roč. 4, č. 4, s. 89-94. ISSN 1212-3609.
- [13] MELOUN, M., MILITKÝ, J., HILL, M. *Počítačová analýza vícerozměrných dat v příkladech*. 1. vyd. Praha: Academia, 2005. 449 s. ISBN 80-200-1335-0.
- [14] ŘEZANKOVÁ, H. a kol. *Shluková analýza dat*. 2. rozšířený vyd. Praha: Professional publishing, 2009. 218 s. ISBN 978-80-86946-81-8.
- [15] SOUKAL, I. Dopady harmonizace systému klientských nákladů CBS v ČR se Směrnici 2007/64/ES. In *Hradecké ekonomické dny 2010: sborník příspěvků díl II. z vědecké konference Ekonomický rozvoj a management regionů*. Hradec Králové: Gaudeamus, 2010. s. 105–108. ISBN 978-80-7435-041-2.
- [16] SOUKAL, I., HEDVIČÁKOVÁ, M. Retail core banking services e-banking client cluster identification. *Procedia Computer Science Journal*. 2010, Vol. 2, Iss. 3, s. 1205-1210. ISSN 1877-0509.
- [17] ZHANG, T., RAMAKRISHAN, R., LIVNY, M. BIRCH: An Efficient Data Clustering Method for Very Large Databases, In *Proceedings of the ACM SIGMOD Conference on Management of Data*. Montreal, Canada. 1996.

**Mgr. Jan Draessler, Ph.D.**

Univerzita Hradec Králové

Fakulta informatiky a managementu

Katedra informatiky a kvantitativních metod

Jan.Draessler@uhk.cz

**Ing. Ivan Soukal**

Univerzita Hradec Králové

Fakulta informatiky a managementu

Katedra ekonomie

ivan.soukal@uhk.cz

**Ing. Martina Hedvičáková, Ph.D.**

Univerzita Hradec Králové

Fakulta informatiky a managementu

Katedra ekonomie

martina.hedvicakova@uhk.cz

Doručeno redakci: 28. 6. 2011

Recenzováno: 29. 8. 2011, 3. 9. 2011

Schváleno k publikování: 20. 9. 2011

**ABSTRACT****CLUSTER ANALYSIS OF THE DEMAND SIDE OF THE RETAIL CORE BANKING SERVICES MARKET****Jan Draessler, Ivan Soukal, Martina Hedvičáková**

*This paper is focused on the retail core banking services market. The demand side of this market is affected by the negative effects of the information asymmetry. European Union authorities are aware of that as they monitor the situation as well as they set out possible solutions that should reduce the main sources of asymmetry – offer opacity and the lack of offer comparison tools. In the Czech Republic there was introduced independent comparison tool project in the 2010. This system's database holds more than 15,000 answers about the day-to-day usage. This serves as a data source for our analysis. The aim is to classify the Czech working age e-banking clients, i.e. to identify basic retail core banking services usage patterns. As preparations there were performed verification-validation phase, logarithmical transformation and dimension reduction by principal component analysis. The two-step cluster analysis was performed on 8,549 members classified by 19 variables concerning the type of the service, moth usage frequency or average amount of money, communication channel used to order the service and the bank the money were transferred to. Analysis indentified as an optimal number of clusters 3. There were identified mainstream client, more active client and the client with the "at the desk" preference. Sub-optimal, but still acceptable, solution was consisted of 4 clusters: mainstream client with his or hers own bank preference, mainstream client with other bank preference, less active client and "at the desk" preference client. Identified clusters don't just describe the demand side as standard marketing research output but they can be used again as an information asymmetry reduction tool.*

**Key Words:** cluster analysis, information asymmetry, retail core banking services market.

**JEL Classification:** C38, G21, D82.