

MOŽNOSTI VYUŽITÍ PŘÍSTUPU INDEXOVÁNÍ LATENTNÍ SÉMANTIKY PŘI PŘEDPOVÍDÁNÍ FINANČNÍCH KRIZÍ

Petr Hájek, Unicorn College s.r.o., Vysoká škola ekonomická v Praze; **Michal Střížík**, **Pavel Praks**, VŠB – Technická univerzita Ostrava; **Petr Kadeřábek**, Univerzita Karlova v Praze*

1. Úvod

Finanční krize rozhodně nejsou pouze fenoménem moderní doby. Krize likvidity zažil starověký Řím, středověká království i mnohé moderní ekonomiky operující s celou řadou sofistikovaných finančních produktů a disponující stále účinnějšími ekonomicko-politickými nástroji.

Finanční krize přitom nemusejí mít nutně podobu krizí měnových. Projevují se i jako krize dluhové a stabilita světových financí byla několikrát v minulosti vážně ohrožena také krizemi bankovními (např. Baring cisis v roce 1890, či 80. léta 20. století v USA). Zatím největší bankrot státu v historii zažila Argentina v roce 2002, když vyhlásila platební neschopnost nad zhruba 160 mld. USD. Pokud dojde ke ztrátě důvěry, investoři požadují vrácení své půjčky, a kapitálovým odlivem tak často způsobí velké znehodnocení měny.

Právě vážné negativní sociálně-ekonomické dopady zpravidla doprovázející finanční krize jsou jedním z hlavních důvodů, proč se problematice lokálních i globálních finančních krizí, studiu jejich kauzality a mechanismů, jejich modelovému popisu a především pak otázce možné predikce věnuje stále více jak teoretických, tak praktických ekonomů. Pochopení finančních krizí může usnadnit budoucí intervence vlády či monetárních autorit cílené k eliminaci případných negativních dopadů na reálnou ekonomiku.

Problematika finančních krizí je studována po celém světě, ale také i v ČR, významnější články v této oblasti jsou přiblíženy v Hájek (2007). V českém prostředí vznikly např. následující studie: Dvořák (2004, 2007), Pazour (2008, 2008), Helísek (2004) nebo také Hájek (2005).

Problematika finančních krizí je neobyčejně komplexní. Proto je velmi obtížné krize úspěšně předpovídat a dosud jen velmi málo studií přineslo uspokojivé výsledky využitelné v praxi. Téměř všechny studie věnované této problematice sledovaly vývoj makroekonomických agregátů, jež jsou ve finální podobě mnohdy dostupné se zpožděním i několika měsíců a v každé zemi mohou být jinak definovány, což ovlivňuje interpretaci při jejich srovnávání. Snad největším problémem je samotná frekvence jednotlivých dat časových řad používaných ekonomických veličin: předpovídání založená na čtvrtletních datech prakticky pozbývají smyslu právě kvůli zpoždění a revizím různých ukazatelů. U měsíčních dat není situace výrazně lepší.

* Článek vychází z výsledků uvedených v Hájek (2007). Výzkum byl finančně podpořen MŠMT ČR (projekt MSM 6198910007).

Náš příspěvek se řadí k nemnoha těm, které testují přístup predikce krizí založený na analýze časových řad mikroekonomických ukazatelů s denní frekvencí. Pojem finanční krize je v naší studii chápán v širším pojetí: jako termín postihující různé podoby krizí měnových, bankovních, dluhových, jejich kombinace a jiné vážné problémy zasahující hospodářství konkrétní země, protože deflace cen aktiv může způsobit kromě krize měnové také krizi bankovní a v souvislosti s těmito typy krizí i krizi dluhovou.

Článek má následující strukturu: v kapitole 2 uvedeme současné přístupy používané k předpovídání finančních krizí, v kapitole 3 představíme ve stručnosti metodu indexování latentní sémantiky. Kapitola 4 se zabývá definicí hypotézy, použitými daty a postupem řešení. Dosažené výsledky jsou diskutovány v kapitole 5. Závěr a možnosti budoucího výzkumu jsou nastíněny v kapitole 6.

2. Současné přístupy používané v problematice predikce finančních krizí

Krize mohou mít mnoho různých příčin a tyto mohou být reprezentovány různými proměnnými. Některé z nich je možné na základě historických studií generalizovat jako tzv. vysvětlující proměnné. Řadí se k nim měnový kurz, devizové rezervy, domácí inflace nebo tempo růstu úvěrů.

Nejčastěji používané přístupy k predikci krizí jsou (např. podle analýzy provedené v Hájek, 2007) signální přístup skupiny Kaminské (Kaminsky, Lizondo, Reinhart, 1998), metody nelineární regrese používající logitové či probitové funkce (Berg, Pattillo, 1998) a Markovův přepínací model (Abiad, 2003). Alternativních a méně prozkoumaných metod však existuje jistě mnoho – je možné mezi ně zahrnout např. neuronové sítě (Fioramanti, 2006) či přístupy fraktální analýzy (Agaev, Kuperin, 2004).

Kaminsky et al. (1998) shrnuli výsledky 28 studií věnovaných měnovým krizím posledních 20 let. Abiad (2003) uvádí shrnutí dalších 30 studií realizovaných po analýze Kaminsky et al. (1998). Berg a Patillo (1998) srovnali výsledky signálního přístupu a modelů nelineární regrese a zjistili, že tyto modely překonaly alternativní měřítko zranitelnosti, jako jsou spready dluhopisů nebo úvěrové ratingy.

Signální přístup (Kaminsky et al., 1998) je schopen úspěšně předpovědět přibližně 70 % krizí, ale za cenu velkého množství falešných signálů (60 %). Problémy, s nimiž se potýká, vyplývají z jeho binární podstaty: vynechávání určitých informací, vnášení umělé korelace do modelu, problém definice krize, různé možnosti výběru proměnných a jejich zpracování.

Metody nelineární regrese, k nimž se řadí logitové a probitové modely (Kumar et al., 2002; Manasse et al., 2003), předpovídají správně 74–89 % krizí za cenu 14–21 % falešných signálů. Jsou tedy ve srovnání se signálním přístupem úspěšnější, i když pracují s podobnými omezeními.

Přístupy zjišťující míru zranitelnosti země, resp. vážnost krize při jejím spuštění (Bussière a Mulder, 1999; Tornell, 1999), jsou určitou modifikací metod nelineární regrese; opět tak pracují s podobnými omezeními. Tyto metody nesledují načasování krize, ale stanovení doby, kdy je země ke krizi náchylná. Jedná se sice o logický předpoklad, ale problém nastává s první i druhou generací krizí, kdy k útoku dojde

v prvním momentu, kdy může být úspěšný, a tedy údaje zpravidla indikující větší zranitelnost země nemusí být v době útoku vůbec zřetelné. Problém nastává již se samotným stanovením, jaká hodnota té které proměnné je už riziková. Dosavadní úspěšnost těchto ze skupiny zmiňovaných metod predikce je zatím nejmenší.

Metoda kontinuálního krizového indexu Petera Vlaara (Vlaar, 2000) usiluje o kombinaci signálního přístupu s regresními metodami motivovanou eliminací nevýhod jednotlivých přístupů, jako je chybějící informace ukazující na vážnost krize (takže vyšší index značí větší krizi) nebo potřeba velkého množství krizí přináší výsledky odpovídající „spojení“ uvedených metod. Ačkoliv se zhoršilo procento falešných signálů v porovnání s regresními metodami a došlo k jeho přiblížení mírně k metodám signálním, výrazně vylepšil předpovědní schopnost v porovnání se signální metodou směrem k metodám nelineární regrese. Při správném určení zhruba 9 z 10 krizí tak Vlaar dosáhl více než třetinového určení klidných období jako krizových. Na rozdíl od studie Andrewa Berga a Catherine Patillo (1998) se navíc jeho modelu, který dosáhl výrazně nižších hodnot poměru šumu k signálu (bohužel však zde odlišná konstrukce metod nedovoluje přímé srovnání), daří správně předpovídat lépe mimo výběrový soubor, který sloužil pro stanovení parametrů modelu.

Snad ještě lepší výsledky než nelineární regresní metody umožňuje Markovův přepínací model (Abiad, 2003). Jeho výhodou je, že nepracuje s apriorní definicí krize. Na rozdíl od předchozích metod nevynechává informace o délce trvání krizí a dynamice časových řad. Tato metodologie by určitě měla být v budoucnu důkladněji prozkoumána a otestována.

Velmi zajímavé výsledky dávají první použití metod umělé inteligence. Neuronové sítě, které nepotřebují téměř žádné předpoklady k použití, mohou být velmi dobrým predikčním nástrojem. Jsou schopné správně předpovídat řádově stejně jako metoda Petera Vlaara, avšak s falešnými signály v řádu procent. Bohužel s výhodami přináší také velké nevýhody, jako je možnost přespecializace na trénovací data (overfitting), nemají uzavřený tvar řešení a možnosti jejich podob jsou téměř nekonečné. Interpretovat jednotlivé parametry modelu (význam jednotlivých proměnných) je navíc prakticky nemožné. Tyto nevýhody degradují metodu na pouhé využití při předpovídání, protože pro tvůrce politik nebude přinášet užitečnou informaci o zdrojích křehkosti.

Velmi pozoruhodné výsledky přináší také fraktální analýza (Agaev a Kuperin, 2004), která si své místo při analýze finančních časových řad pomalu hledá. Tento nelineární přístup v sobě zahrnuje mnoho metod, pomocí kterých je možné přistoupit k analýze finančních časových řad. Autoři ve zmíněné studii použili metodu hölderových exponentů, která ukázala nejpřesněji ze všech ostatních na velké propady na měnových i akciových trzích a jeví se pro předpovídání velmi nadějně. Metoda neumožňuje vysvětlit, proč dojde k propadu, pouze indikuje, že k němu dojde. U metody je nutné analyzovat vliv šesti arbitrárně zvolených vstupních parametrů, které do modelu vstupují spolu s časovou řadou (např. typ regrese a délka období, přes které se vybraná regrese vypočítaných hölderových exponentů počítá).

Každý předpovědní model se musí primárně soustředit na co nejvyšší poměr signálu k šumu. Množství falešných signálů u predikčních metod dosud představuje závažný problém, stejně jako zpravidla používaná arbitrární¹ definice krize.

1 K problematice definování krizí viz Hájek (2007), kde je provedeno srovnání více než 10 přístupů.

Předmětem diskuzí zůstává např. i otázka, zda do analýzy zahrnout též neúspěšné útoky (odvrácené včasným zásahem autorit). Proto se jeví smysluplným hledání přístupů, jež *ex-ante* definici krize nevyžadují a přicházející krizi jsou schopné indikovat z vývoje časových řad ekonomických veličin. Jako perspektivní se v tomto směru jeví Markovův přepínací model² a některé metody umělé inteligence.³ Dalším problémem, se kterým se studie zpravidla potýkají, je samotný předvýběr makroekonomických údajů; mnoho z nich ani není příliš spolehlivých (např. v méně rozvinutých ekonomikách neexistují tržně stanovené úrokové sazby a autoritami vyhlášené nebývají spolehlivé, to se týká i HDP, nezaměstnanosti aj.). Z uvedených důvodů je patrné, že pozornost by se měla soustředit na mikroekonomické údaje např. o zadluženosti nebo investičních aktivitách firem. Ty však nejsou standardními cestami dostupné.

Dostupnost stále výkonnější výpočetní techniky dnes umožňuje analyzovat finanční časové řady též pomocí metod, které dosud nenáležely k tradičním ekonomickým přístupům. Jedním z nich je i přístup vyhledávání informací (Grossman a Frieder, 2000). Na rozdíl od studií zabývajících se analýzou izolovaných časových řad ve smyslu jejich rozkladu na cyklickou složku, sezónní složku a trend (Hamilton, 1994), algoritmus vyhledávání informací je aplikován na problém nalezení „podobných“ časových řad (Das a Gunopulos, 2003). Náš příspěvek demonstruje možnost vyhledávání podobných časových řad finančních ukazatelů pomocí metody indexování latentní sémantiky (Latent Semantic Indexing; LSI) a její aplikaci na rozpoznávání křehkých období ekonomik, v nichž lze očekávat spuštění finanční krize.

3. Metoda indexování latentní sémantiky

Metody numerické lineární algebry, zvláště singulární rozklad matic (Singular Value Decomposition; SVD), jsou používány pro vyhledávání informací ve strategii Latent Semantic Indexing (Berry et al, 1995, 1999). Metoda LSI byla původně aplikována jako efektivní nástroj pro vyhledávání textových dokumentů. Může být chápána jako varianta vektorového modelu s aproximací originálních dat pomocí SVD nebo jiných numerických metod (Berry et al, 1999). Numerické experimenty přitom ukazují, že metody redukce dimenzionality problému (dimenze dat) mohou přinést dvě následující výhody:

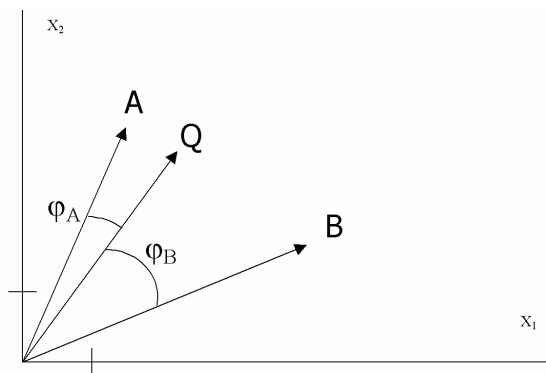
(i) filtrování šumu a (ii) přirozené shlukování dat „které si jsou významově podobné“:

Metody lineární algebry lze s úspěchem využít v nejrůznějších aplikacích vyhledávání informací, např. pro rozpoznávání a rekonstrukce tváří (Miller, 2004), pro makroekonomickou analýzu (Dvořák et al., 2004), pro analýzu hydrochemických dat (Praus et al., 2007) a pro biometrickou identifikaci osob pomocí fotografií očních duhovek (Praks et al. 2007).

Metoda LSI zahrnuje dva kroky: (i) SVD rozklad matice dokumentů, kdy SVD je použito pro odstranění šumu a redundantních informací z dat; a (ii) výpočet koeficientu podobnosti mezi transformovanými (extrahovanými) daty.

2 Např. Jeanne, Masson (1998), Fratzscher (2000), Gomez-Puig, Montalvo (1997), Engel, Hakkio (1994), Abiad (2003), Martinez-Peria (2002), Cerra, Saxena (2002).

3 V oblasti neuronových sítí např. Ashok Nag a Amit Mitra (1999), Peltonen (2001), Peltonen (2006), Fioramanti (2006), Apoteker (2000), Apoteker, Barthelemy (2000).



Na obrázku č. 1 symboly A, B, Q reprezentují vektory (dokumenty). Symboly φ_A a φ_B označují úhly mezi A, Q, respektive mezi B, Q. Dokument A je více podobný dotazu Q než dokument B, neboť $\varphi_A < \varphi_B$. Malý úhel mezi dokumenty je ekvivalentní velké podobnosti.

3.1. SVD rozklad matice dokumentů

Nechť symbol A označuje reálnou $m \times n$ matici. Cílem SVD je nalézt rozklad

$$A = USV^T, \quad (1)$$

kde S je nezáporná diagonální matice $m \times n$, která obsahuje na své diagonále tzv. singulární čísla. Matice U , resp. V^T , označují $m \times m$, resp. $n \times n$, ortogonální matice: $U^T = U^{-1}$, $V^T = V^{-1}$. Sloupce matice U , resp. matice V^T , jsou nazývány levými, resp. pravými, singulárními vektory. SVD rozklad může být proveden tak, že singulární čísla jsou na diagonále seříděna sestupně. Výpočet rozkladu (1) je časově náročný pro velké matice. Navíc se ukazuje, že výpočet singulárních čísel blízkých nule a jim příslušejících singulárních vektorů může zhoršovat kvalitu vyhledávání. Proto je v reálných výpočtech uvažováno pouze k největších singulárních čísel a jim příslušejících singulárních vektorů. V tomto případě je mnohorozměrný prostor redukován do k -dimenzionálního prostoru

$$A_k = U_k S_k V_k^T, \quad (2)$$

kde symbol U_k označuje $m \times k$ matici obsahující prvních k sloupců matice U , symbol S_k označuje $k \times k$ diagonální matici obsahující prvních k singulárních čísel a konečně symbol V_k označuje $n \times k$ matici, která vznikla vybráním prvních k sloupců matice V . Sloupce matice V_k^T označují v našem případě transformované (redukované) dokumenty originální databáze. Jinými slovy, SVD umožňuje aproximovat matici A vzhledem ke sloupcovým vektorům a k -tá aproximace (A_k) matice A je sestavena vybráním pouze k -prvních singulárních čísel matice S , zatímco zbývající jsou zanedbány.

Aproximací matice A maticí A_k může vzniknout chyba, která klesá se zvyšující se hodnotou k , a opačně, se snižujícím se k nabývá chyba vyšší hodnoty. Při aproximacích maticemi A_k nízkých hodnot k se do sloupcových vektorů (vektory dokumentů – časových řad) promítají hodnoty z jiných sloupců dle vazeb mezi nimi. Právě tato vlastnost

umožňuje zachytit skryté vazby mezi dokumenty a termy. Optimální výkon LSI (tedy volbu vhodného k) je třeba ověřit experimentálně.

Takto rozšířený vektorový model, uvažující případné skryté vazby mezi dokumenty a termy, již lze použít pro efektivní vyhledávání. Po výpočtu singulárního rozkladu se získávání relevantních dokumentů pomocí uložené LSI databáze provádí výpočtem kosinové vzdálenosti mezi dotazovacím vektorem q a sloupcovými vektory redukované matice A_k .

Algoritmus pro výběr optimálního počtu singulárních čísel a vektorů není znám (Berry et al, 1995, 1999). Z tohoto důvodu je počet vlastních čísel a vlastních vektorů zvolen experimentálně.

3.2. Výpočet koeficientu podobnosti mezi transformovanými daty

Existuje mnoho přístupů pro výpočet podobnosti mezi dvěma vektory (Das, 2003; Grossman et al, 2000). Jedním z nich je i tzv. kosinová podobnost, kdy podobnost mezi dvěma vektory (dokumenty) je vyjádřena

$$\cos \varphi_j = \frac{(q, D_j)}{\sqrt{(q, q)} \sqrt{(D_j, D_j)}} \quad (3)$$

Symboly q , resp. D_j , označují transformovaný dotaz, resp. transformovaný dokument, a $1 \leq j \leq n$. Geometrický význam kosinové podobnosti je demonstrován ve 2D vektorovém prostoru na obrázku 1.

Procedura *lsi* vrací uživateli vektor koeficientu podobnosti (3), ve kterém i -tý element obsahuje hodnotu, kterou interpretujeme jako „míru“ významové podobnosti (podobnost) mezi i -tým dokumentem a dotazem. („Míra významové podobnosti“ bude dále v textu uváděna zkráceně jako „podobnost“.) Algoritmus LSI byl implementován v programovém prostředí MATLAB (The Math Works, Ltd.), umožňujícím výpočet několika singulárních čísel a vektorů matice A .

4. Testovaná hypotéza, použitá data a postup řešení

Metoda LSI již byla částečně testována jako pomocný nástroj při predikcích finančních krizí založených na analýzách *makroekonomických* dat (Střížik, 2004; Dvořák et al., 2004; Bednářová, 2005). V rámci námi provedené alternativní studie byla ale tato technika aplikována na *mikroekonomické* údaje zprostředkované reakcemi investorů, jež se promítají do akciových indexů daných států. Tyto údaje jsou totiž dostupné s denní frekvencí, a v porovnání s makroekonomickými ukazateli typu HDP nebo bilance běžného účtu tak mohou reflektovat očekávání o budoucím vývoji ekonomiky téměř okamžitě po zveřejnění konkrétních zpráv (např. o nezaměstnanosti, inflaci).

Testovanou hypotézou tedy byl předpoklad, že *akciové trhy reflektují přicházející měnovou krizi, resp. turbulenci v měnové politice, kterou může být depreciace měny nebo růst úrokových sazeb*. Předmětem studia je hypotéza, zda vývoj cen akciového indexu může signalizovat vývoj v měnové oblasti, nikoliv kauzality mezi vývojem cen akcií a měnovým kurzem (zda propad akcií později způsobí růst sazeb či znehodnocení měny anebo očekávané znehodnocení měny a růst sazeb způsobí de facto zpětně mnohem dřívejší pokles akciových trhů).

4.1. Ověření hypotézy

Tabulka 1:

Použité časové řady

Země	Index akciového trhu (hodnoty Close)	Kurz	Úroková sazba
Argentina	Buenos Aires SE, Merval Index, Close, ARS	USD/ARS	Interbank Rates, BAIBOR O/N, Fixing, ARS
Austrálie	S&P/ASX, 20, Index, Total Return, Close, AUD	USD/AUD	Deposit Rates, 1 Month, Close, AUD
Belgie	BXS, All Share Index, Total Return, Close, EUR	USD/BEF	Deposit Rates, 3 Month, Close, EUR
Brazílie	Sao Paulo SE, Bovespa Total, Close, BRL	USD/BRL	Policy Rates, SELIC Rate, BRL
ČR	Prague SE, PX Index, Close, CZK	USD/CZK	Interbank Rates, PRIBOR 1 month, Fixing, CZK
Dánsko	Copenhagen SE, KAX Index, Close, DKK	USD/DKK	Interbank Rates, CIBOR 1 month, Fixing, DKK
Egypt	EFG, Hermes Index, Close, EGP	USD/EGP	HSBC's Emerging Sovereign Debt Index, Yield, Egypt
Filipíny	Philippine SE, Composite Index, Close, PHP	USD/PHP	Interbank Rates, PHIBOR 1 month, Fixing, PHP
Finsko	Finland HEX, Helsinki Stock Market Indice	USD/FIM	Interbank Rates, EURIBOR 1 month, Fixing, EUR
Hong Kong	Hang Seng, Index Futures 1-Pos, Close, HKD	USD/HKD	Interbank Rates, HIBOR 1 month, Fixing, HKD
Chile	Santiago SE, IGPA General Index, Close, CLP	USD/CLP	Interbank Rates, Average O/N Rate, CLP
Indie	Bombay SE, 100 Index, Close, INR	USD/INR	India T-BILL N3 1 month Yield
Indonésie	HSBC Indonesia Index, Price Return, USD	USD/IDR	Interbank Rates, JIBOR 1 month, Fixing, IDR
Itálie	Milan SE, Mta, MIB Index, Close, EUR	USD/ITL	Interbank Rates, BBA LIBOR 1 month, Fixing, EUR
JAR	FTSE/JSE, All Share Index, Close, ZAR	USD/ZAR	Treasury Bills, 3 Month Tender, Yield, Close, ZAR
Jižní Korea	KSE, KOSPI 200 Index, Close, KRW	USD/KRW	Interbank Rates, Overnight Call Rate, KRW
Malajsie	KLSE, Composite Index, Close, MYR	USD/MYR	Interbank Rates, KLIBOR 1 month, Fixing, MYR
Mexiko	HSBC Mexico Index, Price Return, USD	USD/MXN	Mexico Bank Deposit Rate 90 day
Německo	Germany DAX, Close	USD/DEM	Interbank Rates, BBA LIBOR 1 month, Fixing, EUR
Nigérie	Nigeria, Nigeria SE, All-Share Index, Close, NGN	USD/NGN	Nigeria Interbank offer rate NIBOR 7 day
Pákistán	HSBC Pakistan Index, Price Return, USD	USD/PKR	Deposit Rates, 1 Month, Close, PKR
Peru	Lima SE, General Index, Close, PEN	USD/PEN	Interbank Interest Rate, PEN
Polsko	Warsaw SE, WIG Index, Close, PLN	USD/PLN	Interbank Rates, WIBOR 1 month, Fixing, PLN
Rumunsko	Vanguard, VAB General Index, Close, RON	USD/RON	Interbank Rates, BUBID 1 Month, Fixing, ROL
Rusko	RTS, Index (RTSI), Close, USD	USD/RUB	Brady Bonds, M-Fin 5, Price, Close, USD
Singapur	SGX, Straits Times Index, Close, SGD	USD/SGD	Deposit Rates, 3 Month, Close, SGD
Slovensko	Bratislava SE, SAX Index, Close, SKK	USD/SKK	Interbank Rates, BRIBOR O/N, Fixing, SKK
USA	DJI Averages, Price Return, Close, USD	EUR/USD	Government Benchmarks, 1 year, Yield, Close, USD
Srí Lanka	Colombo SE, All Share Index, Close, LKR	USD/LKR	Deposit Rates, 1 Month, Close, LKR
Španělsko	Madrid SE, General Index, Close, EUR	USD/ESP	Deposit Rates, 3 Month, Close, EUR
Švédsko	Stockholm SE, OMXS30 Index, Close, SEK	USD/SEK	Government Benchmarks, 2 year, Yield, Close, SEK
Švýcarsko	Swiss Market Index (SMI), Close, CHF	USD/CHF	Interbank Rates, BBA LIBOR 1 month, Fixing, CHF
Turecko	ISE, National-100 Index, Close, TRY	USD/TRY	Policy Rates, Central Bank O/N Borrowing Rate
Velká Británie	FTSE, 100, Index, Price Return, Close, GBP	USD/GBP	Interbank Rates, BBA LIBOR 1 month, Fixing, GBP
Venezuela	HSBC Venezuela Index, Price Return, USD	USD/VEB	Venezuela Overnight Rate
Zimbabwe	ZSE, Industrial Index, Close, ZWD	USD/ZWD	Zimbabwe Interbank Rate

Kurz byl použit spotový (hodnoty Close)

Výhodou použitého přístupu LSI je, že použitá data nemusí být předem upravována (normována), neboť bylo použito automatické normování. Výpočet koeficientu podobnosti mezi daty byl proveden kosinovou podobností, která automaticky provedla normování výsledných podobností do intervalu $<-1;1>$ (viz sekce 3.2). Díky tomuto postupu jsme se vyhnuli dalšímu náročnému předzpracování dat (mj. přepočtu kurzů a přepočtu úrokových sazeb), protože normalizace dat nemění strukturu těchto dat, pouze normu reprezentovaných vektorů, nikoliv však úhel (podobnost) mezi těmito

vektory. Např. různé tržní úrokové sazby, které jsou vysoce korelovány, ovlivnily pouze délku daného vektoru (např. o velikost rizikové premie), nikoliv úhel mezi těmito vektory, tedy podobnost nebyla použitím jiné tržní úrokové sazby ovlivněna (neboli srovnání výsledků u zemí, kde u jedné byla použita např. interbank overnight rate a u jiné depozitní, nezkrsluje výsledky).

Navíc i přesto, že měnové kurzy vůči hlavní rezervní měně USD nebyly předem jakkoliv upraveny, ukazují dosažené výsledky na vhodnost použitého postupu. V reálném světě může být krize také importována, proto nás zajímal i vliv událostí ovlivňujících měnový kurz ze strany USA. Z tohoto důvodu jsme se rozhodli pro použití běžných kurzů a nikoliv kurzů upravených tak, abychom tyto externí vlivy minimalizovali.

Základní datový soubor pro testování byl připraven z časových řad dostupných za období 1985–2007 (tabulka 1). Každému dni následujícímu po 1. lednu 1985 byly přiřazeny odpovídající údaje z časových řad směnného kurzu dané měny vůči USD, hlavního indexu akciového trhu a úrokových sazeb.⁴ Tímto způsobem bylo vytvořeno vždy 1 pozorování (obchodní den). Pokud byly hodnoty hlavního indexu akciového trhu pro danou zemi dostupné pouze za velmi krátký časový úsek (tj. řádově několik let) a byla přitom k dispozici delší časová řada indexu burzy počítaného jinou institucí než burzou (např. investiční bankou), byla použita tato časová řada jako proxy vývoje akciového trhu.⁵ Jelikož trhy obchodují jen v některých dnech a metoda LSI standardně pracuje s identickými formáty dat, byla do testování zahrnuta pouze pozorování (obchodní dny) s alespoň jedním dostupným sledovaným ukazatelem, přičemž chybějící údaje byly lineárně interpolovány. Takto byly ošetřeny víkendy, svátky a jiné dny, kdy trhy neobchodovaly; tyto dny nebyly pro testování uvažovány. Poté byla pro každé jednotlivé datum (obchodní den) od ledna 1985 do března 2007 u každé sledované země vygenerována časová řada sestávající z 64 následných pozorování hlavního indexu akciového trhu.⁶ Uvedeným způsobem bylo vytvořeno pro každou zemi několik tisíc časových řad, které byly podrobeny vlastní analýze.⁷

Jako dotaz v rámci aplikace LSI metody byla zvolena časová řada hlavního indexu akciového trhu obsahující 64 pozorování odpovídajících obchodním dnům právě předcházejícím propuknutí finanční krize v Thajsku dne 2. 7. 1997, kdy došlo k deprecaci tamní měny o 18,03 %. Potvrzení hypotézy vyžadovalo, aby po časové řadě, která byla pro konkrétní zemi metodou LSI vyhledána jako nejvíce podobná dotazu, bezprostředně následovalo období (časový úsek o 128 pozorováních),⁸ kdy došlo ke znehodnocení dané měny anebo k růstu úrokových sazeb; a opačně pro řady vyhledané jako nejméně podobné dotazu. Období 128 pozorování tedy považujeme za tzv. křehké období, kdy očekáváme, že může být krize spuštěna náhodnou událostí.

4 Byly sledovány tržní sazby; výjimku tvořilo Turecko a Brazílie, kde údaje tržní sazby nebyly dostupné v dostatečně dlouhém časovém období, a tak byly použity sazby vyhlášené centrální bankou.

5 Toto byl případ několika jihoamerických zemí a také např. Singapuru, kde byl použit Straits Times Index kompilovaný vydavatelstvím novin.

6 Šedesát čtyři po sobě jdoucích hodnot, kterými trh zavíral obchodní den. Tento počet obchodních dní přibližně odpovídá jednomu čtvrtletí.

7 V drtivé většině případů bylo dostupných cca 3–5 tisíc časových řad (např. u Argentiny 3455, Austrálie 5088, Brazílie 3580, Egypta 3651). U vyspělých zemí se počet řad často blížil šesti tisícům.

8 Vývoj měnového kurzu a úrokových sazeb tedy byl expertně interpretován v souvislosti s podobností k dotazu.

5. Výsledky a diskuze

Testování byly podrobeny údaje z časových řad indexů akciových trhů celkem 36 zemí. Pro interpretaci odpovědí obdržených na základě aplikace metody LSI bylo v prvním kroku nutné zvolit relevantní hodnotu koeficientu k , udávajícího počet singulárních čísel (a tedy velikost konečné dimenze problému po redukci). Jak již bylo uvedeno, volba hodnoty k je arbitrární a provádí se zpravidla na základě experimentů. Praks et al. (2007), Krátký (2002), Střížík (2004) a Bednářová (2005) aplikovali metodu LSI v různých vědních disciplínách a jako relevantní volili $k = 10$. Po rozsáhlých numerických experimentech byla tato hodnota též zvolena v rámci zde uváděných testů.

Metodou LSI byly jednotlivé časové řady základního datového souboru ve formě odpovědi na dotaz seřazeny v pořadí od časové řady nejvíce podobné dotazu po časovou řadu nejméně podobnou dotazu. Pořadí těchto řad bylo analyzováno pro každou jednotlivou zemi zvlášť.⁹ Protože byly řady v rámci základního datového souboru generovány po jednotlivých po sobě jdoucích dnech, bylo častým jevem, že se mezi sousedními odpověďmi vyskytovaly časové řady s počátky navzájem se lišícími pouze o několik málo dnů. V takovém případě byla pro interpretaci zvolena z příslušných odpovědí vždy odpověď nejvíce podobná dotazu.¹⁰

Postup interpretace výsledků vyhledávání je ilustrován na příkladu Velké Británie, případu výjimečném množstvím falešných signálů, které byly na základě aplikace metody LSI generovány, třebaže mezi nejvíce podobnými časovými řadami bylo téměř přesně identifikováno období krize ERM. Níže uvedené identifikátory časových řad vygenerovaných ze základního datového souboru obsahují informaci o zkoumané zemi (1. a 2. znak; např. „UK“ pro Velkou Británii), sledovaném ukazateli (3. a 4. znak; „SE“ pro Stock Exchange – hlavní index akciového trhu, „IR“ pro úrokové sazby, „FX“ pro měnový kurz dané měny vůči USD), počátku časové řady o 64 pozorováních (5. až 15. znak).¹¹ Identifikátor je zakončen příslušnou hodnotou koeficientu podobnosti, uvedenou v exponenciálním tvaru, kdy číslice se znaménkem za písmenem e označují dekadický exponent. Koeficient podobnosti nabývá hodnot z intervalu $\langle -1; +1 \rangle$ a lze jej interpretovat následovně: s klesající hodnotou koeficientu podobnosti klesá podobnost mezi odpovědí (vyhledanou časovou řadou) a dotazem.

V časovém úseku následujícím za koncem řady „UKSE08-May-2002 8.933296e-001“, nejvíce podobné dotazu, měna zhodnotila o 4,5 %. Sazby (United Kingdom, Interbank Rates, BBA LIBOR 1 month, Fixing, GBP) klesaly. Tento výsledek hypotézu nepodporuje. Za druhou dotazu nejvíce podobnou řadou „UKSE13-Jul-1990 8.905759e-001“ měna oslabilo o 12 %, což hypotézu podporuje. Navíc jen mezi 91. a 122. pozorováním měna depreciovala o 14,8 %. Sazby spíše klesaly ze 14 % k 12,5 %. V období následujícím za koncem třetí dotazu nejvíce podobné řady „UKSE23-Jul-1998 8.399127e-001“ libra v jasném trendu znehodnotila o 4,9 %. Sazby klesaly ze 7,5 %

9 Například při dotazu předkrizového vývoje na akciovém trhu v Thajsku vůči Argentině byla matice se singulárními čísly počítána pouze na množině Argentiny, nikoliv na všech dostupných zemích najednou. Takto bylo docíleno vyšší přesnosti výpočtů.

10 Pokud například dvě řady nejvíce podobné dotazu začínaly s odstupem jednoho dne a počátek třetí nejvíce podobné řady byl posunut o rok, pak druhá nejvíce podobná časová řada nebyla uvažována.

11 Tato řada byla srovnávána s dotazem Thajska – s předkrizovým obdobím. Datum tedy ukazuje na začátek této řady, za kterou následuje řada o 128 pozorováních.

k 5,3 %. Měnová krize ERM byla identifikována čtvrtou nejvíce podobnou řadou „UKSE12-Jun-1992 7.933816e-001“, a to z hlediska identifikace počátku krize téměř přesně: hned ve 3. pozorování za dotazem (řada 128 pozorování započala 10. 9. 1992) je zřejmá první výraznější devalvace o 3,35 %, o další tři dny později Velká Británie opustila ERM. Tři ze čtyř časových řad nejvíce podobných dotazu hypotézu podporují.

V časovém úseku následujícím za koncem řady „UKSE27-Dec-1988 -8.346157e-001“, nejméně podobné dotazu, měna znehodnotila o 11 % a ve stejném období (mezi 36. a 60. pozorováním) vzrostly sazby. Za řadou „UKSE01-Apr-1997 -7.697942e-001“ mezi 20. a 90. pozorováním byla libra slabší oproti zbývajícím hodnotám v řadě 128 pozorování cca o 6 %. Sazby rostly z 6,7 % na 7,5 %. V období následujícím za koncem řady „UKSE17-Nov-1993 -7.568404e-001“ sazby klesaly a měna zhodnochovala. Dvě ze tří nejméně podobných řad jsou tedy v rozporu s hypotézou.

Dosažené výsledky jsou u Velké Británie smíšené a není možné jednoznačně konstatovat, že vývoj ve Velké Británii podporuje hypotézu, přestože byla nalezena krize ERM. Také devalvace z roku 1990 byla identifikována jako u jiných států tehdejšího EHS. Na druhou stranu počet tří falešných signálů (1 u nejvíce podobné a 2 u nejméně podobných časových řad) je v takto malém vzorku relativně vysoký.

Analýza výsledků vyhledávání časových řad metodou LSI je přehledně shrnuta v tabulce 2. Rozboru bylo podrobeno 250 z celkového počtu více než 150 tisíc metodou LSI testovaných řad. Celkem 205 řad podporovalo hypotézu (správně předpovězené signály), 41 odporovalo hypotéze (předpovězeny falešné signály) a 4 řady nebyly interpretovatelné. Mezi nejvíce podobnými řadami činilo množství falešných signálů 17 %, u nejméně podobných řad 15 %. Pokud zemi v analyzovaném časovém úseku postihla významná měnová krize a současně byly pro dané období dostupné časové řady sledovaných ukazatelů, byla tato událost ve většině případů identifikována mezi 6 časovými řadami, které byly metodou LSI vyhledány jako řady nejvíce podobné dotazu.

Výsledky vyhledávání je třeba interpretovat s ohledem na očekávání trhu, která bývají jiná u zemí v měnové unii a jiné u rozvíjejících se zemí. Mezi časovými řadami identifikovanými metodou LSI jako nejvíce podobné předkrizovému vývoji v Thajsku se nacházelo nezanedbatelné množství falešných signálů, kdy vyhledané časové řady buď reprezentovaly období předcházející krizím o více než půl roku nebo kdy vývoj trhu nebyl determinován očekáváním měnových turbulencí, ale jinými, s měnovou politikou nesouvisejícími, důvody.

Na základě diskutovaných výsledků lze soudit, že akciové trhy jsou v principu schopné rozpoznat přicházející problémy v měnové politice. Testovaná hypotéza tedy byla potvrzena. Metodu LSI aplikovanou v této práci na vyhledání úseků časových řad indexů akciových trhů podobných krizovému vzoru není možné chápat jako nástroj schopný detekovat finanční krize univerzálně a jednoznačně. Lze ji však doporučit pro vyhledávání křehkých období ekonomik, kdy může dojít ke spuštění krize. Jak vyplývá z kapitol 3 a 4.1, je samozřejmé, že toto období může být v dalších studiích definováno různými způsoby (např. v délce či v kombinaci s jinými časovými řadami).

Tabulka 2

Výsledky pro úlohu vyhledávání definovanou dotazem Thajsko.

Země	Řady nejvíce podobné dotazu			Řady nejméně podobné dotazu		Poznámka
	potvrzující hypotézu	? ¹²	odporující hypotézu	potvrzující hypotézu	odporující hypotézu	
Argentina	5	0	0	5	0	1 problém s načasováním
Austrálie	1	0	2	3	0	
Belgie	3	0	0	3	0	
Brazílie	6	0	0	4	0	
ČR	3	0	1	1	1	1 problém s načasováním
Dánsko	3	0	2	1	2	
Egypt	2	1	1	3	0	
Filipíny	2	0	1	3	0	téměř přesně načasována asijská krize
Finsko	2	0	1	3	0	1 problém s načasováním
Hong Kong	2	0	1	3	0	
Chile	3	0	0	3	0	přesně načasována prudká depreciace
Indie	2	1	0	1	2	
Indonésie	3	0	1	3	1	byla předpovězena krize
Itálie	2	0	1	3	0	přesné zachycení krize ERM
Jihoafrická rep.	3	0	2	2	1	
Jižní Korea	3	0	0	3	0	předpovězena asijská krize
Malajsie	1	0	2	2	1	správně nalezená krize
Mexiko	5	0	0	4	1	1 problém s načasováním
Německo	3	0	1	3	0	
Nigérie	2	1	0	2	1	
Pákistán	3	1	0	2	1	
Peru	3	0	0	3	0	
Polsko	2	0	1	3	0	
Rumunsko	6	0	0	5	0	1 problém s načasováním
Rusko	3	0	0	3	0	téměř přesně zachycena krize
Singapur	1	0	2	3	0	předpovězena asijská krize
Slovensko	3	0	0	3	0	
Srí Lanka	3	0	0	3	0	
Španělsko	5	0	1	5	0	
Švédsko	2	0	1	3	0	
Švýcarsko	3	0	0	2	1	
Turecko	5	0	0	3	0	
USA	2	0	1	3	0	
Velká Británie	3	0	1	1	2	přesně předpovězena krize ERM
Venezuela	3	0	0	0	3	
Zimbabwe	3	0	0	2	1	většina řad s indexem podobnosti větším než 0,8
Celkem	106	4	23	99	18	
Poměr v %	80	3	17	85	15	

12 Otazník symbolizuje nemožnost jednoznačné interpretace časové řady z důvodu fixace měnového kurzu (nebo jeho velmi malých změn) v kombinaci s nedostupnými úrokovými sazbami

Jak je z textu (kapitola 4) patrné, netestovali jsme, zda pokud v nějaké zemi, ve které došlo ke krizi, ji metoda našla, nýbrž hypotézu, zda předkrizový vývoj v dané zemi, pokud byl velmi podobný předkrizovému období v Thajsku, vedl k očekávané depreciaci anebo růstu tržních sazeb v konkrétní zemi. Jinými slovy, testovali jsme, zda následovala depreciace nebo růst tržních úrokových sazeb v případě velké podobnosti s předkrizovým obdobím v Thajsku. Z tohoto důvodu není relevantní srovnání s dalšími studiemi zabývajícími se předpovídáním měnových či finančních krizí. Použitý princip a dosažené výsledky pouze ukazují na využitelnost metody LSI v dalším ekonomickém výzkumu – při tvorbě modelů předpovídání krizí, které by pak již bylo možné srovnat s jinými studiemi (tyto by se lišily např. v použití různých proměnných, délek časových řad, nastavení konstanty k , kombinace panelových dat jako jednoho obrazu dotazu). Úloha by pak samozřejmě musela být jinak definována a musela by existovat referenční množina krizí, které musí být nějak definovány. Naopak, pro účely našeho příspěvku nemusela být krize definována, protože byla použita prakticky jediná thajská krize, která je za krizi obecně považována.

V budoucnu by bylo zajímavé vyhledávat podobnosti nejen mezi časovými řadami stejné délky, ale taktéž vyhledávat časové řady, které jsou podobné danému (kratšímu) vzoru. Výsledky vyhledávání by mohly být reprezentovány shluky, kdy by se v takovém shluku nacházely vždy podobné časové řady. Ukazuje se, že pro shlukování lze s výhodou použít tzv. Fiedlerův vektor (Fiedler, 1973). Výhodou těchto shlukovacích technik je skutečnost, že není vyžadována informace o výsledném počtu shluků a shlukovací metody pracují zcela automaticky (tzv. unsupervised clustering).

6. Závěr

Metoda indexování latentní sémantiky (Latent Semantic Indexing; LSI) jako aplikace singulárního rozkladu matic (Singular Value Decomposition; SVD) byla poprvé použita pro vyhledání úseků časových řad indexů akciových trhů podobných časové řadě indexu akciového trhu odpovídající období právě předcházejícímu dni 2. 7. 1997, kdy v Thajsku byla 18,03% měnovou depreciací spuštěna vážná globální finanční krize. Metoda LSI byla aplikována v rámci testování hypotézy, že akciové trhy mají dostatečnou schopnost reflektovat přicházející měnovou krizi, resp. turbulenci v měnové politice, kterou mohou být depreciace měny nebo růst úrokových sazeb. Třebaže se mnohé ze sledovaných zemí od Thajska ekonomicko-politicky do značné míry lišily a liší, byly pomocí metody LSI indikovány nadcházející měnové problémy ve většině případů a v několika z nich dokonce s přesností jediného dne. Metoda LSI může představovat v rámci predikce finančních krizí jeden z účinných doplňkových nástrojů pro specifikaci a analýzu křehkých období, kdy krize mohou být spuštěny náhodnou událostí.

Bezpečná predikce finančních krizí metodou, která by s vysokou pravděpodobností (hraničící téměř s jistotou) umožnila detekovat nebezpečí krize, však zůstane patrně ještě dlouhou dobu otázkou, na kterou budou ekonomové hledat odpověď. Finanční krize totiž představují neobyčejně komplexní problém a bude velice obtížné zachytit všechny potenciálně významné faktory a vazby mezi nimi tak, aby na jejich základě navržený predikční model dokázal finanční krize předví-

dat univerzálně a s dostatečnou spolehlivostí. Kromě kvantifikovatelných ukazatelů (makroekonomické či mikroekonomické proměnné) mohou totiž vývoj ekonomik ovlivnit i faktory kvalitativní povahy, jako jsou např. psychologické aspekty spojené s rozhodováním investorů při existenci asymetrické informace či stádního chování. Takové aspekty mají tím větší vliv, čím méně informací je známo o rozsahu konkrétních problémů, jako je tomu v rámci probíhající realitní krize v USA. Pokud by se podařilo i faktory tohoto typu v budoucnosti kvantifikovat a zahrnout do modelu a pokud by byl k dispozici dostatečně rozsáhlý soubor časových řad charakterizujících vývoj ekonomik, mohly by se metody numerické lineární algebry typu LSI, umožňující filtraci šumu a přirozené shlukování „významově si navzájem podobných“ dat, stát velmi užitečným pomocníkem při predikcích finančních krizí. Prezentované výsledky aplikace metody LSI na vyhledávání časových řad mikroekonomických ukazatelů podobných krizovému vzoru jsou toho dokladem.

Literatura

- ABIAD, A. 2003. Early Warning Systems: A Survey and a Regime Switching Approach. [IMF Working Paper WP/03/32]. Washington D.C.: IMF, 1998.
- AGAEV, I. A.; KUPERIN, YU. A. 2004. *Multifractal Analysis and Local Hoelder Exponents Approach to Detecting Stock Markets Crashes*. Saint-Petersburg : State University, 2004.
- APOTEKER, T.; BARTHELEMY S. 2000. Genetic Algorithms and Financial Crises in Emerging Markets. [TAC Financial]. Saint Hilaire des Landes: May 2000.
- APOTEKER, T. 2000. An example of the measures provided by Riskmonitor on Turkey. [TAC Financial]. Saint Hilaire des Landes: June 2000.
- BEDNÁŘOVÁ, P. 2005. Možnosti využití speciálních kvantifikačních metod v ekonomické analýze. [Diplomová práce]. Jindřichův Hradec: FM VŠE v Praze, 2005.
- BERG, A.; PATILLO, C. 1998. Are Currency Crises Predictable? A Test. [IMF Working Paper 98/154]. Washington D.C.: IMF, 1998.
- BERG, A.; BORENSZTEIN, E.; PATILLO, C. 2005. Assessing Early Warning Systems: How Have They Worked in Practice? *IMF Staff Papers*. 2005, vol. 52, no 3.
- BERRY, W. M.; DRMAČ, Z.; JESSUP, J. R. 1999. Matrices, Vector Spaces, and Information Retrieval. *SIAM Review*. 1999, 41, s. 336–362.
- BERRY, W. M.; DUMAIS, S. T.; O'BRIEN, G. W. 1995. Using linear algebra for intelligent information retrieval. *SIAM Review*. 1995, 37 s. 573–595.
- BUSSIÈRE, M.; MULDER, CH. 1999. Political Instability and Economic Vulnerability. [IMF Working Paper No. WP/99/46]. Washington D.C.: IMF, April 1999.
- DAS, G.; GUNOPULOS, D. 2003. Time Series Similarity and Indexing. In Nong Ye (ed.) *Handbook on Data Mining*. Lawrence Erlbaum Associates, 2003.
- DVOŘÁK, P.; STŘÍŽÍK, M.; PRAKS, P.; PUDIL, P.; ŠUMPÍKOVÁ, M.; LEŠETICKÝ, O. 2004. The feasibility of using special quantitative methods for prediction of currency crises. Příspěvek prezentovaný na International Scientific Conference: European Finance – Theory, Politics and Practice, Universita Matěja Bela, Banská Bystrica, Slovensko, s. 1–26, 2004, ISBN 80-8055-968-6 (CD-ROM).
- DVOŘÁK, P. 2004. *Finanční krize jako globální problém a možnosti jejího vzniku v ČR*. Studie Národohospodářského ústavu Josefa Hlávky. Studie č. 2/2004.
- DVOŘÁK, P. 2007. Monetární teorie cyklu, dluhový problém a finanční krize. *Politická ekonomie*. 2007, roč. 55, č. 2, s. 183–205.

- FIORAMANTI, M. 2006. *Predicting Sovereign Debt Crises Using Artificial Neural Networks: A Comparative Approach*. Trieste : University of Trieste, 2006.
- GOLDSTEIN, M.; KAMINSKY, G. L.; REINHART, C. M. 2000. *Assessing Financial Vulnerability: An Early Warning System for Emerging Markets*. Washington D.C. : Institute for International Economics. 2000.
- GROSSMANN, D. A.; FRIEDER, O. 2000. *Information retrieval: Algorithms and heuristics*. Norwell (MA) : Kluwer Academic Publishers, 2000.
- HÁJEK, P. 2007. Možnosti využití netradičních kvantitativních metod při předpovídání finančních krizí [Disertační práce]. Praha: Vysoká škola ekonomická v Praze, Katedra měnové teorie a politiky, 2007.
- HÁJEK, P. 2005. Financial Crisis Predictions: Specification of Pre-crisis Periods in Turkey, Argentina and Thailand. *Acta Oeconomica Pragensia*. 2005, roč. 13, č. 1, s. 57–69. ISSN 0572-3034.
- HAMILTON, J. 1994. *Time series analysis*. Princeton (NJ) : Princeton University Press, 1994.
- HELÍSEK, M. 2004. *Měnová krize (empirie a teorie)*. 1. vyd. Praha : Professional Publishing 2004. 180 s. ISBN 80-86419-82-7.
- KAMINSKY, G. L.; LIZONDO, S.; REINHART, C. M. 1998. Leading indicators of currency crisis [IMF Staff Papers 45]. Washington D.C. : IMF, 1998.
- KRÁTKÝ, M. 2002. *Využití SVD pro indexování latentní sémantiky*. Ostrava : VŠB Technická univerzita Ostrava, 2002
- KUMAR, M.; MOORTHY, U.; PERRAUDIN, W. 2002. Predicting Emerging Market Currency Crashes [IMF Working Paper 02/07]. Washington, D.C. : International Monetary Fund, 2002.
- FIEDLER, M. 1973. Algebraic connectivity of Graphs. *Czechoslovak Mathematical Journal*. 1973, č. 23 (98).
- MANASSE, P.; ROUBINI, N.; SCHIMMELPFENNIG, A. 2003. Predicting Sovereign Debt Crises [IMF Working Paper WP/03/221]. Washington, D.C.: International Monetary Fund, 2003.
- MULLER, N.; MAGAIA, L.; HERBST, B. M. 2004. Singular Value Decomposition, Eigenfaces and 3D Reconstructions. *SIAM Review*. 2004, vol. 46, no. 3, s. 518–545.
- PAZOUR, M. 2004. Nové metodologické přístupy k tvorbě empirických modelů měnových krizí. *Politická ekonomie*. 2004, roč. 52, č. 3, s. 375–388.
- PAZOUR, M. 2008. Stanovení náchylnosti ekonomiky k nadměrným tlakům na měnový kurs. *Politická ekonomie*. 2008, roč. 56, č. 5, s. 598–620.
- PRAKS, P.; MACHALA, L.; SNÁŠEL, V. 2007. On SVD-free Latent Semantic Indexing for Iris Recognition of Large Databases. In V. A. Petrushin; L. Khan (Eds.) *Multimedia Data mining and Knowledge Discovery*. London : Springer-Verlag, 2007. s. 472–486. ISBN 1-84628-436-8. http://dx.doi.org/10.1007/978-1-84628-799-2_24
- PRAUS, P.; PRAKS, P. 2007. Information retrieval in hydrochemical data using the latent semantic indexing approach. *Journal of Hydroinformatics*. 2007, vol 9, no 2, s. 135–143. London : IWA Publishing. ISSN: 1464-7141.
- STRÍŽÍK, M. 2004. Možnosti využití speciálních kvantitativních metod při predikci finančních krizí [Diplomová práce]. Jindřichův Hradec: FM VŠE v Praze, 2004.
- TORNELL, A. 1999. Common Fundamentals in the Tequila and Asian Crises [Working Paper 7139]. Cambridge (MA): NBER, 1999.
- VLAAR, P. J. G. 2000. Early Warning Systems for Currency Crises [Papers 671]. Amsterdam: De Nederlandsche Bank, 2000.

POSSIBILITIES OF FINANCIAL CRISES FORECASTING WITH LATENT SEMANTIC INDEXING

Petr Hájek, Unicorn College s.r.o., V Kapslovně 2767/2, CZ – 130 00 Praha 3; University of Economics in Prague, Faculty of Finance and Accounting, nám. W. Churchilla 4, CZ – 130 67 Praha 3 (hajekp@gmail.com); **Michal Strížik**, VŠB – Technical University of Ostrava, Faculty of Safety Engineering, Lumírova 13, CZ – 700 30 Ostrava (michal.strizik@vsb.cz); **Pavel Praks**, VŠB – Technical University of Ostrava, Department of Mathematics, 17. listopadu 15, CZ – 708 33 Ostrava (pavel.praks@vsb.cz); **Petr Kadeřábek**, Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies, Smetanovo nábřeží 995/6, CZ – 110 01 Praha 1 (petrkaderabek@seznam.cz).

Abstract

Our study contains application of Latent Semantic Indexing on financial crises prediction. Hypothesis to test was that equity markets are able to predict even sharp changes in monetary policy during a quarter ahead of such a change (which was searched during two quarters that followed). This hypothesis, tested on sample of 36 countries between years 1985 and 2007, has been confirmed according to interest rate and foreign exchange expert interpretation.

The studied application of LSI even though it timed several crises on their exact start day is not suitable for financial crises prediction but can be recommended for specification and analysis of fragile countries which are or could be prone to a crisis.

Keywords

dimension reduction, financial crises, information retrieval, Latent Semantic Indexing (LSI), linear algebra, multidimensional data, Singular Value Decomposition (SVD), stock market,

JEL Classification

C61, E44, E47