

PREDIKOVÁNÍ CEN AKCIÍ V KOMERČNÍ BANCE

PREDICTING STOCK PRICES IN COMMERCIAL BANK

Robert Zeman

*Vysoká škola technická a ekonomická v Českých Budějovicích,
Katedra ekonomiky a managementu
zeman@mail.vstecb.cz*

Jaroslav Stuchlý

*Vysoká škola technická a ekonomická v Českých Budějovicích,
Katedra přírodních věd
stuchly@mail.vstecb.cz*

Abstrakt:

V článku jsou stručně popsány základní metody a modely používané pro predikci cen akcií v Komerční bance (dále KB) prováděných pomocí analýzy časových řad. Je použita časová řada 1588 konečných cen akcií KB na burze cenných papírů v Praze a jednotlivé modely jsou aplikovány na určení bodových i intervalových předpovědí cen těchto akcií na 12 následných obchodních dní. Výsledky jsou porovnány pomocí celkových a následných měr přesnosti a je vybrána optimální aplikace na vybranou časovou řadu.

Klíčová slova:

Analýza časových řad, předpovědi, predikční modely, ceny akcií, Komerční banka, porovnání předpovědí

Abstract:

The aim of the work is to describe the various ways of modeling the time series analysis that reflect the evolution of stock prices in Commercial Bank (further CB) in order to obtain effective predictive models. It is used the time series of 1588 final prices CB shares on stock exchanges in Prague and the models are applied to determine point and interval forecasts of the price of those shares on the successive 12 trading days. The results are compared using total and successive rates of accuracy and it is selected optimum application on the selected time series.

Key words:

time series analysis, forecasting, predictive models, stock quotes, Commercial Bank, comparing predictions

JEL: G1

1 Úvod

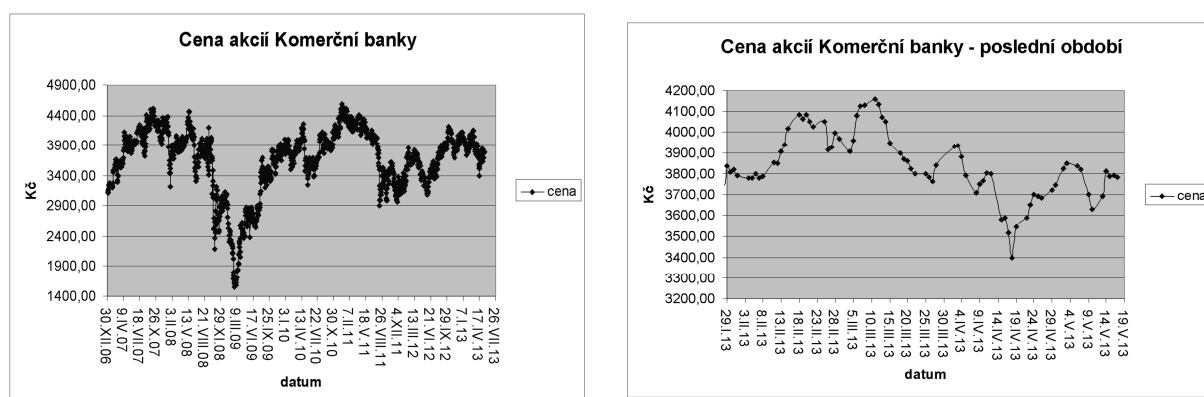
Cílem práce bude popsat jednotlivé možnosti modelování časových řad popisujících vývoj cen akcií za účelem získání efektivních predikčních modelů, aplikovat tyto výsledky na časovou řadu konečných cen akcií KB z jednotlivých obchodních dnů na Pražské burze cenných papírů, porovnat

přesnost předpovědí vypočítaných pomocí jednotlivých modelů, vybrat model, který bude dávat pro daná data nejpřesnější předpovědi a ověřovat splnění podmínek pro adekvátní použití těchto metod.

2 Metody - Materiály a metodika

Použijeme historická data z kurzovních listů KB za období od 2. 2. 2007 do 29. 4. 2013. Dostaneme odtud proměnné: konečná cena akcií (*cena*), celkový objem obchodů (*objem*), data obchodních dnů – očíslováme je do proměnné obchodní dny ($t = 1, 2, \dots, 1588$). Zde dny s chybějícími daty vynecháváme. Data s použitými proměnnými lze najít na webové stránce https://is.vstecb.cz/auth/dok/rfmgr.pl?upload_id=4380872 v datovém souboru koban.xlsx. Abychom mohli provést i následné hodnocení přesnosti předpovědí, doplníme data ještě o období od 30. 4. 2013 do 15. 5. 2013 (12 obchodních dní, $t = 1589, \dots, 1600$). Znázornění doplněných dat je na grafu 1.

Graf 1: Graf doplněné časové řady



K analýzám jsme použili obvyklou metodiku analýzy časových, která je obecně popsána např. v Arlt (2002). Numerické a grafické výsledky jsme získali použitím programu Statgraphics Plus pro Windows 3.0, PcGive 2.10 a Excel.

Modely používané k získávání předpovědí můžeme třídít podle různých hledisek. Nejčastěji používané třídění je na ekonometrické modely, modely založené na trendové analýze, modely založené na časových řadách a kvalitativní předpovědní modely. Předpovědními modely založenými na trendových funkcích pro akcie Erste Bank jsme se zabývali v článku Stuchlý, Zeman (2013). Ukázali jsme, že při velké variabilitě dat jsou tyto předpovědi zatíženy velkými chybami. V tomto článku se budeme zabývat jen předpověďmi založenými na modelech časových řad, které dávají podstatně přesnější výsledky. Kvalitu modelu budeme posuzovat pomocí známých měr přesnosti předpovědi MSE (střední kvadratická chyba), MAE (střední absolutní chyba), MAPE (střední absolutní procentuální chyba) nebo i jiných měr, zavedených např. v Arlt (2002).

Pro získávání předpovědí existují různé metody a modely. Neexistuje ale nic takového jako je nejlepší předpovědní model, použitelný ve všech situacích. Kvalita a vhodnost použitého modelu závisí na vlastnostech použité časové řady, tj. zejména na tom, kolik má řada členů, jaká je jejich variabilita, zda obsahuje trendovou složku (není stacionární), sezónní nebo cyklickou složku, je-li mezi členy řady sériová závislost (autokorelace), apod. Pro každou časovou řadu proto vybíráme nevhodnější předpovědní metodu, popř. můžeme také výsledky z jednotlivých modelů kombinovat.

Do programu Statgraphics jsme načítli výše zmíněná data a v nabídce Analýza časových řad – Předpovídání jsme v položce Porovnání modelů vybírali modely poskytující nejpřesnější výsledky. Jde o tyto modely: model neměnných hodnot (náhodné procházky), klouzavých průměrů, exponenciálního vyrovnávání, exponenciálně vážených klouzavých průměrů a nakonec Box-Jenkinsovy modely (ARIMA-modely). Uvedeme si stručný popis odpovídajících metod a jejich aplikační výsledky.

Podrobnější popis těchto modelů a určování bodových a intervalových předpovědí najdeme např. v Cipra (1986), Stuchlý (2004), Arlt, Arltová (2003), Petrášková (2006), Makridatis, Wheelwright (1989), Grifits, Hill, Judge (1992).

3 Model neměnných hodnot

Předpovědi založené pouze na nejčerstvějších pozorováních proměnné jsou často označovány jako "naivní předpovědi". Naivní modely se vyznačují značnou jednoduchostí, proto je můžeme použít jako srovnávací odhady předpovědí, dříve než uijeme složitější model.

Model neměnných hodnot (model náhodné procházky) je nejjednodušší naivní předpovědní model, používaný u časové řady, která nevykazuje trend (stacionární časová řada). Předpověď se uskutečňuje na i období dopředu a její hodnota je rovna poslední aktuální hodnotě (nejnovější hodnotě časové řady), tedy: $F_{t+i} = X_t$, kde F_{t+i} je předpověď na období $t+i$ (tedy i období dopředu), X_t je poslední známá hodnota časové řady.

Pro časovou řadu s lineárním trendem lze použít model s proporční změnou $F_{t+1} = X_t + p(X_{t+1} - X_t)$, kde p je vhodně zvolená konstanta z intervalu $(0;1)$, nebo časovou řadu očistit od trendu diferencováním. Pro časovou řadu s exponenciálním trendem lze použít model procentuální změny $F_{t+1} = (1+k) X_t$, kde $k = X_t / X_{t-1}$.

Z grafu 1 vidíme, že použitá časová řada přibližně nemá trend (je stacionární). Proto předpověď provedeme pomocí modelu neměnných hodnot na časové řadě konečných cen akcií Komerční banky (soubor koban.xlsx).

Porovnání předpovědí se skutečnými 12 hodnotami a použití výše uvedených měr přesnosti je provedeno ve 2. sloupci tabulky 1. V posledním sloupci jsou míry přesnosti počítané pomocí programu Statgraphics z celé časové řady původních hodnot – míry označené hvězdičkou.

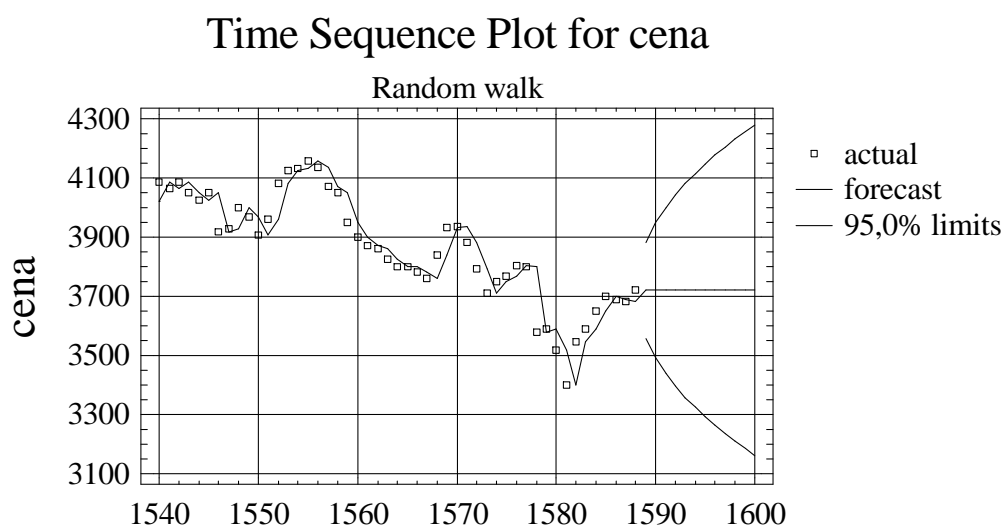
Míry v levé části tabulky 1. představují následné hodnocení přesnosti předpovědí. Zde pro porovnání byly použity nové hodnoty časové řady neobsažené v původní časové řadě (získané z internetu). Uvedené hodnoty měr přesnosti mají samy o sobě jen malou vypovídací schopnost. Mohou být ovšem použity k porovnání s výsledky jiných metod a na jejich základě může pak být zvolena optimální předpovědní metoda pro konkrétní časovou řadu. Vzhledem k tomu, že nově získaných hodnot je použito jen omezené množství, mohou být výsledky takového porovnávání jednotlivých metod zkrácené. Častěji se proto používají míry přesnosti vypočtené z původní časové řady. Tyto míry založené na dlouhé původní časové řadě vypočtené pomocí programu Statgraphics (označené hvězdičkou) jsou v pravé části tabulky. Hvězdičkované míry přesností jsou nižší, což plyne z většího množství použitých dat pro jejich výpočet a ze značného kolísání následných hodnot cen akcií. Při vyhodnocování optimální předpovědní metody budeme proto více přihlížet k hvězdičkováným mírám přesnosti.

Tabulka 1: Naivní předpověď metodou neměnných hodnot a míry přesnosti předpovědi

Datum	Čas t	Předpovězená hodnota	Skutečná hodnota	Absolutní chyba předpovědi	Absolutní chyba předpovědi ²	Procentní chyba v %
30. 4. 2013	1589	3720	3744	24	576	0,64102564
2. 5. 2013	1590	3720	3825	105	11025	2,74509804
3. 5. 2013	1591	3720	3850	130	16900	3,37662338
6. 5. 2013	1592	3720	3838	118	13924	3,07451798
7. 5. 2013	1593	3720	3819	99	9801	2,59230165
9. 5. 2013	1594	3720	3700	20	400	0,54054054
10. 5. 2013	1595	3720	3630	90	8100	2,47933884
13. 5. 2013	1596	3720	3693	27	729	0,73111292
14. 5. 2013	1597	3720	3810	90	8100	2,36220472
15. 5. 2013	1598	3720	3785	65	4225	1,71730515
16. 5. 2013	1599	3720	3791	71	5041	1,87285677
17. 5. 2013	1600	3720	3782	62	3844	1,63934426
MAD	75,08333				MAD*	58,3749
MSE	6888,75				MSE*	6769,68
MAPE v %	1,981022				MAPE* v %	1,67941

Statgraphics počítá pro předpovědi i predikční intervaly. Graf 2 znázorňuje bodové i intervalové predikce.

Graf 2: Předpovědi pomocí modelu neměnných hodnot (náhodné procházky)



4 Model klouzavých průměrů

Předpokládejme nejdříve, že uvažovaná časová řada nevykazuje trend ani sezónní složku. K předpovědi její hodnoty pro čas $t+1$ můžeme potom použít klouzavý průměr jejích posledních m pozorování, tj.

$$\hat{y}_{t+1} = MA(m)$$

kde $MA(m)$ je prostý klouzavý průměr počítaný z m členů podle vzorce

$$MA(m) = \frac{1}{m} (y_t + y_{t-1} + y_{t-2} + \dots + y_{t-m+1}).$$

Stejnou předpověď můžeme použít i na další období dopředu. Určení klouzavé délky m je třeba provést zkusmo (s využitím nahoře uvedených měř).

Má-li uvažovaná řada lineární trend, potom ho nejdříve odstraníme výpočtem prvních diferencí $\Delta y_t = y_t - y_{t-1}$. Časová řada 1. diferencí pak nemá trend a pro čas $t+1$ můžeme určit její předpověď

$\Delta \hat{y}_{t+1}$ pomocí klouzavého průměru řady Δy_t . Předpověď původní řady pro čas $t+1$ je

$$\hat{y}_{t+1} = \Delta \hat{y}_{t+1} + y_t.$$

Podobně, má-li uvažovaná časová řada kvadratický trend, použijeme řadu jejich druhých diferencí $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = y_t - 2y_{t-1} + y_{t-2}$, která již nevykazuje trend a pro čas $t+1$ lze pomocí jejího klouzavého průměru určit předpověď $\Delta^2 \hat{y}_{t+1}$. Předpovědi původní řady pro čas $t+1$ je

$$\hat{y}_{t+1} = \Delta^2 \hat{y}_{t+1} + 2y_t - y_{t-1}.$$

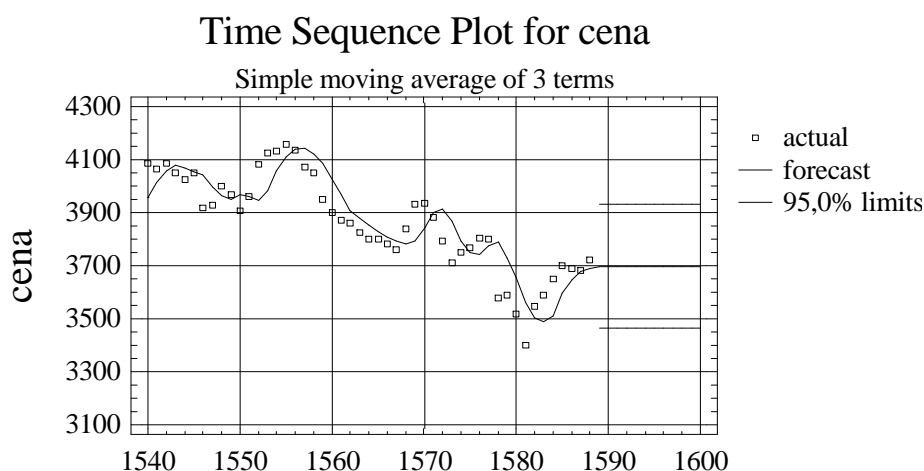
V Statgraphics se přesvědčíme, že při přepokládaném konstantním trendu nejlepší výsledky dávají trojčlenné klouzavé průměry. Předpovědi jsou opět konstantní a nabývají hodnoty 3697,33 Kč. Shrnutí výsledků provedeme do tabulky 2. a grafu 3.

Toto vyrovnání je méně přesné a používáme ho obvykle, když počítáme jen jednu předpověď.

Tabulka 2: Předpovědi metodou trojčlenných klouzavých průměrů a míry přesnosti

Datum	Čas t	Předpovězená hodnota	Skutečná hodnota	Absolutní chyba předpovědi	Absolutní chyba předpovědi ²	Procentní chyba v %
30. 4. 2013	1589	3697,33	3744	46,67	2178,0889	1,246527778
2. 5. 2013	1590	3697,33	3825	127,67	16299,6289	3,337777778
3. 5. 2013	1591	3697,33	3850	152,67	23308,1289	3,965454545
6. 5. 2013	1592	3697,33	3838	140,67	19788,0489	3,665190203
7. 5. 2013	1593	3697,33	3819	121,67	14803,5889	3,185912543
9. 5. 2013	1594	3697,33	3700	2,67	7,1289	0,072162162
10. 5. 2013	1595	3697,33	3630	67,33	4533,3289	1,854820937
13. 5. 2013	1596	3697,33	3693	4,33	18,7489	0,117248849
14. 5. 2013	1597	3697,33	3810	112,67	12694,5289	2,957217848
15. 5. 2013	1598	3697,33	3785	87,67	7686,0289	2,316248349
16. 5. 2013	1599	3697,33	3791	93,67	8774,0689	2,470852018
17. 5. 2013	1600	3697,33	3782	84,67	7169,0089	2,238762559
MAD	86,863				MAD*	58,2828
MSE	9771,7				MSE*	6730,36
MAPE v %	2,2857				MAPE* v %	1,67561

Graf 3: Předpovědi pomocí modelu trojčlenných klouzavých průměrů



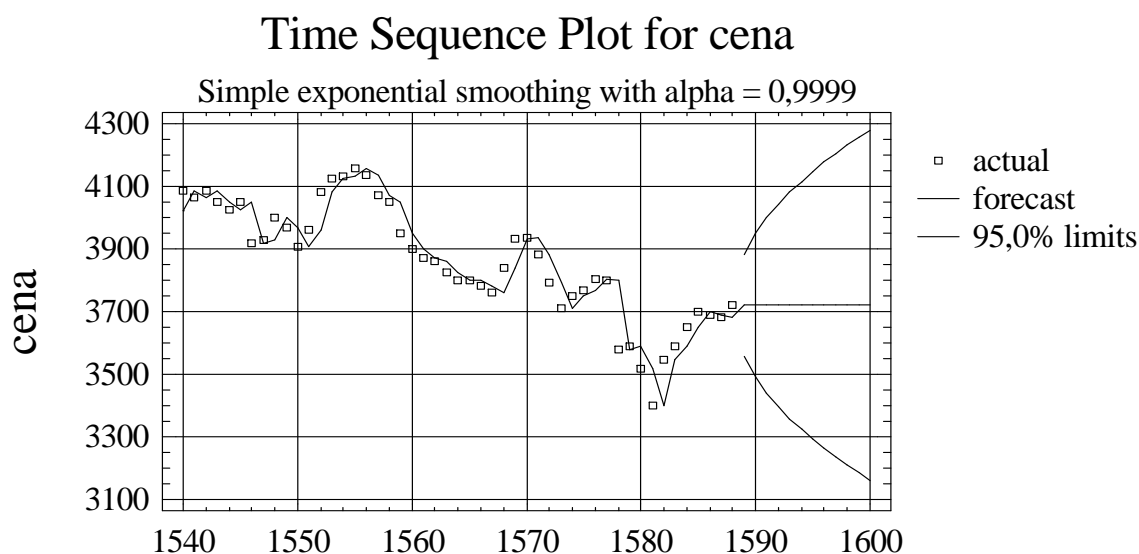
5 Exponenciální vyrovnání

Nejjednodušší tvar má model, nemá-li časová řada trend. Potom používáme tzv. Brownovo prosté exponenciální vyrovnávání, kdy předpovídanou hodnotu \hat{y}_{t+1} časové řady v čase $t+1$ dostaneme jako vážený průměr pozorované hodnoty y_t řady v čase t a předcházející předpovědi \hat{y}_t této řady v čase t , tj. výpočet provádíme podle rekurentního vzorce $\hat{y}_{t+1} = \alpha y_t + (1-\alpha) \hat{y}_t$, kde α je tzv. vyrovnávací konstanta splňující podmínku $0 \leq \alpha \leq 1$. Tuto hodnotu můžeme použít i jako předpověď \hat{y}_{t+T} na T období dopředu. Za počáteční předpověď klademe obyčejně $\hat{y}_1 = y_1$. Optimální volbu vyrovnávací konstanty necháme na programu Statgraphics. Označíme-li vyrovnanou hodnotu řady v čase t symbolem S_t , potom platí $S_t = \alpha y_t + (1-\alpha) S_{t-1}$, kde S_{t-1} je vyrovnaná hodnota řady v čase $t-1$. Má-li uvažovaná časová řada lineární trend, potom používáme tzv. exponenciální vyrovnávání druhého stupně, které spočívá v tom, že vyrovnané hodnoty znovu vyrovnáváme. Má-li časová řada obecnější trend, potom používáme tzv. trojnásobné exponenciální vyrovnávání (viz např. Stuchlý 2004).

Program Statgraphics vybírá jako optimální metodu exponenciálního vyrovnání prosté exponenciální vyrovnání s vyrovnávací konstantou 0,9999 a určí předpovědní hodnotu.

Tabulka pro předpověď metodou prostého exponenciálního vyrovnání a následné míry přesnosti předpovědi jsou stejné jako v tabulce 1. Míry přesnosti předpovědi založené na původní časové řadě vypočtené pomocí programu Statgraphics jsou $MAE^* = 58,375$, $MSE^* = 6774$, $MAPE^* = 1,67941$. Znamená to, že obecně by tyto předpovědi měli vyjít jen mírně přesnější než v modelu náhodné procházky.

Graf 4: Předpovědi pomocí modelu prostých klouzavých průměrů



6 Exponenciálně vážené klouzavé průměry (Holtovo vyrovnání)

Jde o velmi efektivní předpovědní EWMA-metodu použitelnou pro proces se sezónními složkami. Je známa též pod názvem Wintersovo exponenciální vyrovnávání. Vychází z multiplikačního modelu časové řady s lineárním trendem násobeným sezónní složkou. V každém časovém bodě je zvlášť odhadována průměrná úroveň (absolutní člen) a přírůstek trendu (směrnice) a dále sezónní faktor a potom kombinací těchto složek dostaneme celkovou předpověď. Vzorce jsou uvedeny v Stuchlý (2004). Při určování vyrovnávacích konstant postupujeme podobně jako u Brownova exponenciálního vyrovnávání. Pokud se v časové řadě nevyskytuje sezónní složka (tj. $F_t = 1$), potom tato metoda je také známa pod názvem Holtovo exponenciální vyrovnávání. V našem případě se jedná o časovou řadu bez sezónní složky a výsledky získané Holtovým vyrovnáním v Statgraphics zapíšeme opět do tabulky 3. A grafu 5.

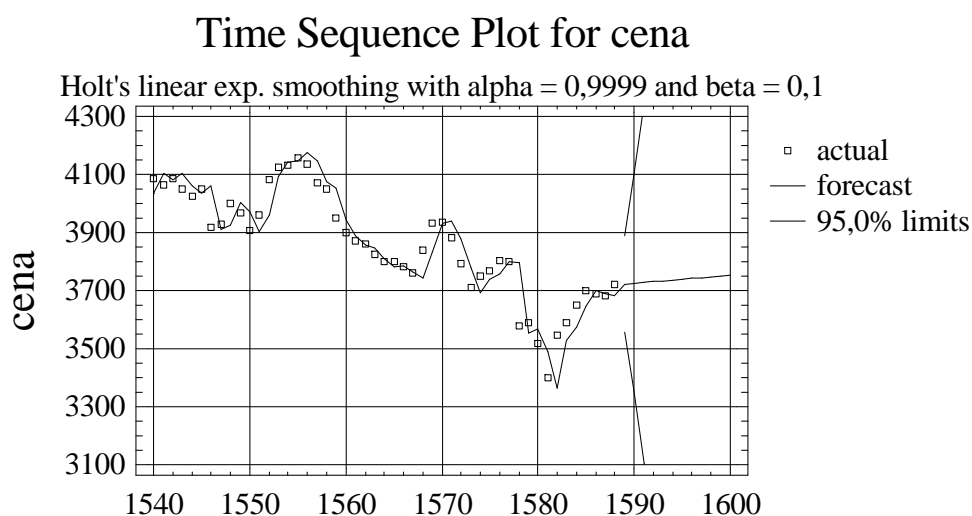
Holtovo vyrovnání dává obecně přes svoji komplikovanost pro použitou časovou řadu méně přesnou předpověď než model náhodné procházky i než model prostého exponenciálního vyrovnání. Jen počítá více předpovědí, ale tyto jsou pro větší odlehlost od zadaných dat zatíženy velkou predikční chybou, která se rychle s časem zvyšuje. Následné hodnocení je však lepší než u modelu náhodné procházky.

Tabulka 3: Předpověď metodou Holtova vyrovnání a míry přesnosti předpovědi

Datum	Čas t	Předpovězená hodnota	Skutečná hodnota	Absolutní chyba předpovědi	Absolutní chyba předpovědi ²	Procentní chyba v %
30. 4. 2013	1589	3722,72	3744	21,28	452,8384	0,5683761
2. 5. 2013	1590	3725,44	3825	99,56	9912,1936	2,6028758
3. 5. 2013	1591	3728,17	3850	121,83	14842,5489	3,1644156
6. 5. 2013	1592	3730,89	3838	107,11	11472,5521	2,7907764
7. 5. 2013	1593	3733,61	3819	85,39	7291,4521	2,2359256
9. 5. 2013	1594	3736,34	3700	36,34	1320,5956	0,9821622
10. 5. 2013	1595	3739,06	3630	109,06	11894,0836	3,0044077

13. 5. 2013	1596	3741,78	3693	48,78	2379,4884	1,3208773
14. 5. 2013	1597	3744,51	3810	65,49	4288,9401	1,7188976
15. 5. 2013	1598	3747,23	3785	37,77	1426,5729	0,9978864
16. 5. 2013	1599	3749,95	3791	41,05	1685,1025	1,0828277
17. 5. 2013	1600	3752,67	3782	29,33	860,2489	0,7755156
MAD	66,9158				MAD*	60,467
MSE	5652,22				MSE*	7226,91
MAPE v %	1,77041				MAPE* v %	1,73671

Graf 5: Předpovědi pomocí modelu Holtova vyrovnání



7 Box-Jenkinsonovy modely (ARIMA-modely)

Box-Jenkinsonova metodologie (BJ-metodologie) předpovídá budoucí hodnoty řady pouze z minulého vývoje časové řady. Klade důraz na pravděpodobnostní a stochastickou analýzu vlastní ekonomické časové řady založenou na filosofii "necht' data hovoří sama za sebe".

BJ-metodologie používá různé předpovědní modely. Zavedeme si třídu nejčastěji používaných autoregresních integrovaných modelů klouzavého průměru označovanou jako ARIMA(p,d,q)-modely, která obsahuje všechny časové řady tvaru

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q},$$

kde prvních $p+1$ členů nazýváme autoregresní část modelu, posledních $q+1$ členů nazýváme částí klouzavého průměru, $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ náhodné složky časové řady, $\phi_0, \phi_1, \phi_2, \dots, \phi_p$ autoregresní parametry, $\theta_1, \theta_2, \dots, \theta_q$ parametry klouzavého průměru.

Abychom vybrali nejlepší ARIMA-model pro daná data, musíme nejdříve odstranit z časové řady trend pomocí vhodné transformace. Nejčastěji stačí použít vhodný stupeň diferencování analyzované časové řady. Potřebný stupeň diferencování je zde označen d . Základním problémem při výběru ARIMA(p,d,q)-modelu je výběr parametrů p, d, q . Při $d = 0$ hovoříme o ARMA(p,q)-modelu (uvažovaná řada je stacionární, tj. trend se nemění s časem a není ji proto třeba diferencovat); při $q = 0$ hovoříme o autoregresním modelu AR(p,d) a při $p = 0$ o modelu čistého klouzavého průměru MA(d,q). Při používání ARIMA-modelu předpokládáme také, že v časové řadě není sezónnost. Existuje i obecnější BJ-model, který zahrnuje kromě trendu i sezónnost, známý pod názvem SARIMA.

Aplikace BJ-metodologie zahrnuje následující kroky:

1. Identifikace modelu.
2. Odhad parametrů modelu.
3. Diagnostika modelu.
4. Výpočet předpovědí.

Identifikací modelu rozumíme výběr správné kombinace hodnot parametrů modelu p , d a q . Používáme k tomu autokorelační a parciální autokorelační funkce – ACF a PACF (viz Cipra 1986 nebo Arlt 2003), jejich odhady a grafy (korelogramy).

Při výběru modelu se zhruba řídíme těmito pravidly:

- a) Model klouzavého průměru $MA(q)$ je charakterizovaný tím, že až po q posunutích ACF výrazně klesá a tím, že PACF postupně klesá k nule.
- b) Autoregresní model $AR(p)$ je charakterizovaný tím, že až po p posunutích PACF výrazně klesá a postupným klesáním ACF.
- c) Je-li výsledek nejednoznačný v tom smyslu, že oba průběhy autokorelační i parciální autokorelační funkce vykazují významné hroty, použijeme smíšený ARMA nebo ARIMA model.

Odhad autoregresních parametrů a parametrů klouzavého průměru provádíme např. v Statgraphics nelineární metodou nejmenších čtverců nebo pomocí jiného programu metodou maximální věrohodnosti.

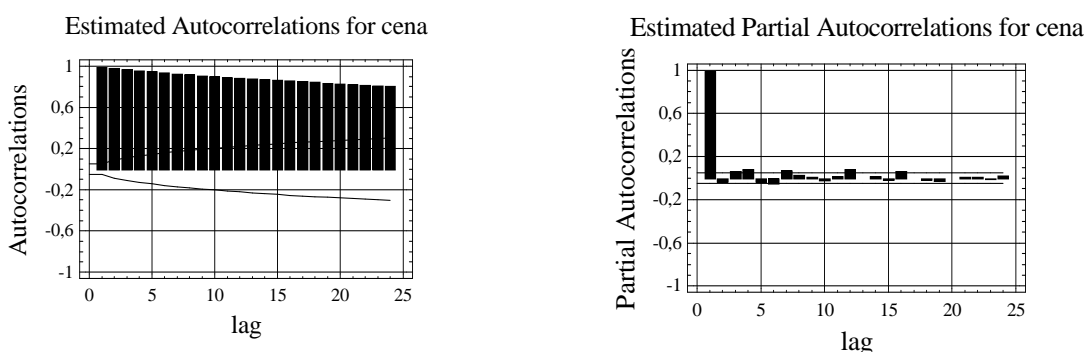
Při diagnostice modelu testujeme významnost odhadnutých parametrů, popř. modelu jako celku a zkoumáme pomocí reziduální analýzy rezidua odhadnutého modelu (jejich náhodnost a sériovou nezávislost, tj. zda rezidua představují tzv. bílý šum).

Splňuje-li odhadnutý model všechny tyto požadavky, můžeme ho použít na předpovědi. V opačném případě se musíme vrátit k identifikaci a použitý model změnit.

V případě naší časové řady se jedná o stacionární řadu bez sezónní složky. K identifikaci modelu znázorníme ACF a PACF v grafu 5.

Z grafu vidíme, že můžeme použít model $AR(1)$ nebo $AR(3)$. Z porovnání výsledků v programu Statgraphics trochu lepší výsledky dává model $AR(3)$. Odhadnutý model má tvar

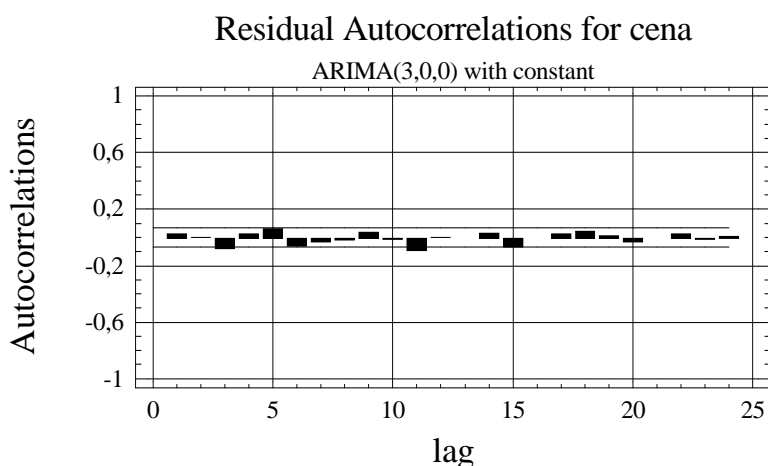
Graf 6: ACF a PACF pro časovou řadu konečných cen akcií Komerční banky



Sumář k modelu ARIMA

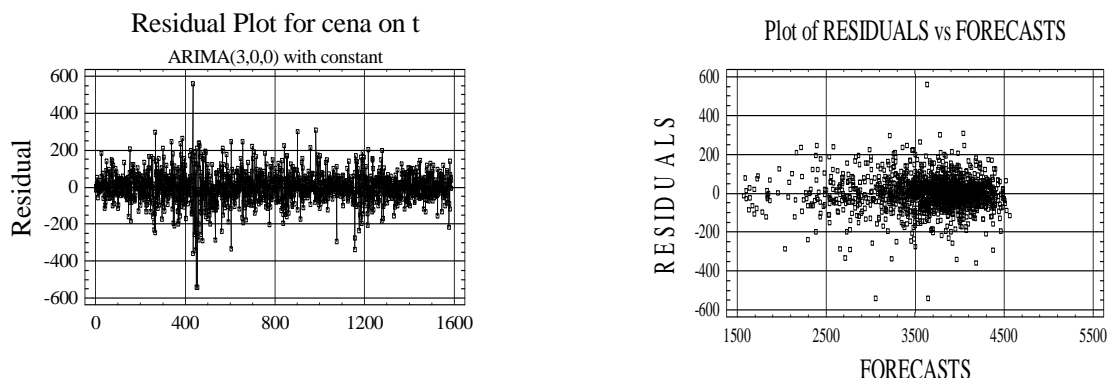
Parametr	odhad	stand. chyba	t - statistika	p-hodnota
AR(1)	1,00558	0,0250846	40,0870	0,000000
AR(2)	-0,0753821	0,0360227	-2,09263	0,036382
AR(3)	0,0637005	0,0250883	2,53905	0,011115
Průměr	3158,45	80,3099	39,3283	0,000000
Konstanta	19,2755			

Graf 7: Graf reziduální ACF



Všechny odhadnuté regresní parametry jsou statisticky významné. Provedeme ještě další diagnostiku modelu. Box-Pierceův test v programu Statgraphics ($T = 56,59$, p -hodnota = 0,000041) zamítá hypotézu o nevýznamné autokorelaci. Potvrzuje to i graf 7, kde některé hodnoty reziduálních autokorelačních koeficientů mírně převyšují meze konfidenčních oblastí. Jde o slabou ($r_{11} = -0,0929$), ale významnou lineární závislost původní na některých posunutých řadách reziduí. Test heteroskedasticity Statgraphics nezahrnuje. Provedme proto tento test alespoň graficky. V grafu 8 je znázorněna závislost reziduí na pořadovém číslu měření a závislost reziduí na vyrovnaných hodnotách cen. Přesnější testování heteroskedasticity reziduí vzhledem k vyrovnaným hodnotám cen provedeme v Excelu pomocí Goldfeldova – Kvantova testu. Po vynechání asi 15% prostředních hodnot dostaneme $F = 1,908 > F_{0,95}(674,674) = 1,135$ (p -hod. = 0). Znamená to, že v modelu byla prokázána významná heteroskedasticita reziduí. Znamená to, že odhady získané obyčejnou metodou nejmenších čtverců (OMNČ) ztrácí své optimální vlastnosti. Jsou jen asymptoticky nestranné a konzistentní, ale nejsou vydatné. Ale Statgraphics neodhadoval tento model OMNČ, ale byla použita obecná metoda nejmenších čtverců. Navíc prokázaná mírná autokorelace znamená, že zvolený model nemusí být zcela adekvátní. Avšak žádný jiný z BJ-modelů odhadnutých v Statgraphics nedává lepší výsledky a proto zůstáváme zatím při porovnání výsledků získaných pomocí Statgraphics u tohoto modelu.

Graf 8: Závislost reziduí na číslu měření a reziduí na vyrovnaných hodnotách



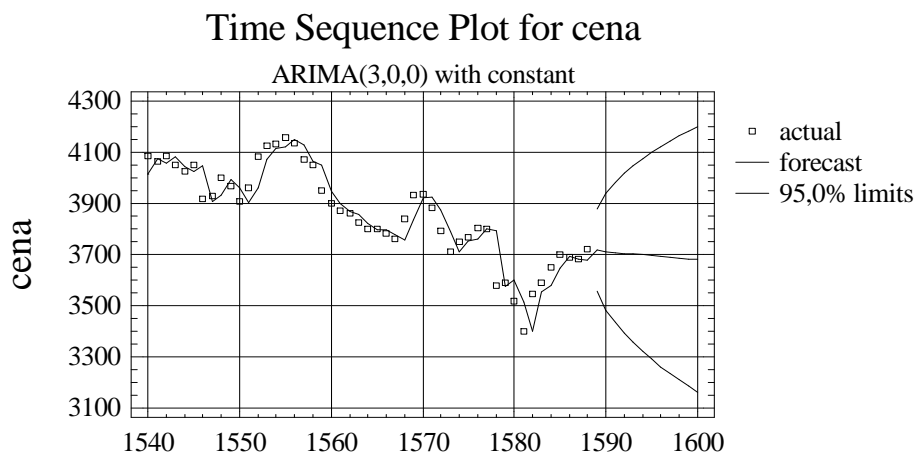
Předpovědi a míry jejich přesností uvádíme v tabulce 4.

Tabulka 4. Předpovědi a jejich míry přesnosti pro vyrovnaní modelem AR (3)

Datum	Čas t	Předpovězená hodnota	Skutečná hodnota	Absolutní chyba předpovědi	Absolutní chyba předpovědi ²	Procentní chyba v %
30. 4. 2013	1589	3717,53	3744	26,47	700,6609	0,7069979
2. 5. 2013	1590	3711,67	3825	113,33	12843,6889	2,9628758
3. 5. 2013	1591	3708,38	3850	141,62	20056,2244	3,6784416
6. 5. 2013	1592	3705,36	3838	132,64	17593,3696	3,4559666
7. 5. 2013	1593	3702,19	3819	116,81	13644,5761	3,0586541
9. 5. 2013	1594	3699,03	3700	0,97	0,9409	0,0262162
10. 5. 2013	1595	3695,9	3630	65,9	4342,81	1,815427
13. 5. 2013	1596	3692,78	3693	0,22	0,0484	0,0059572
14. 5. 2013	1597	3689,68	3810	120,32	14476,9024	3,1580052
15. 5. 2013	1598	3686,6	3785	98,4	9682,56	2,5997358
16. 5. 2013	1599	3683,54	3791	107,46	11547,6516	2,8346083
17. 5. 2013	1600	3680,5	3782	101,5	10302,25	2,6837652
MAD	85,47				MAD*	60,467
MSE	9599,31				MSE*	7226,91
MAPE v %	2,24889				MAPE* v %	1,73671

I když to následné hodnocení nepotvrzuje (míry jsou počítané jen z malého množství dat), obecně ze všech dat vychází všechny míry přesnosti nejnižší. Grafické znázornění bodových i intervalových předpovědí je na grafu 9.

Graf 9: Předpovědi počítané v Statgraphics pomocí modelu AR(3)



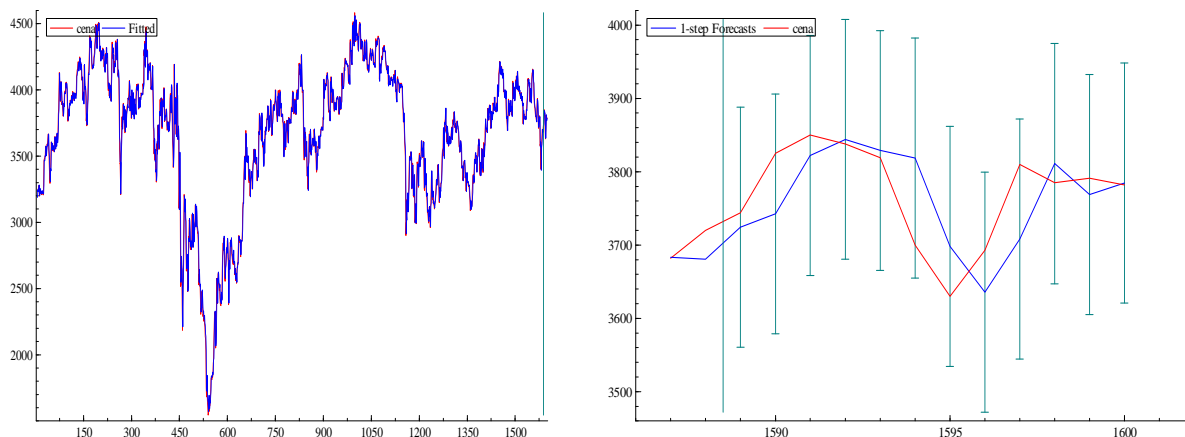
Pokusíme se ještě výsledky zlepšit tím, že odhadneme model AR(3) metodou RALS v programu PcGive 2.10 (popis používání programu – viz Lejnarová, Ráčková, Zouhar 2009).

Pro odhadované parametry dostáváme:

	Coefficient	Stand.chyba	t-hod.	p-hod.
cena_1	1.21504	0.08832	13.8	0.000
cena_2	-0.472218	0.07960	-5.93	0.000
cena_3	0.246455	0.05083	4.85	0.000
Konstanta	39.6017	15.49	2.56	0.011
ro_1	-0.186138	0.08995	-2.07	0.039

V posledním řádku jsou uvedeny údaje pro odhad a autokorelačního koeficientu mezi cenou a 1x posunutou cenou. Autokorelace byla výrazně redukována. Test heteroskedasticity provedeme s využitím Whiteova testu v tomto programu. Dostáváme $\chi^2(6) = 26.134$, p-hod. = 0.0002. Znamená to, že multikolinearita byla mírně redukována. Vyrovnané hodnoty a předpovědi jsou znázorněny na grafu 10. Předpovědi jsou zde porovnány se skutečnými hodnotami cen akcií.

Graf 10: Vyrovnané hodnoty a předpovědi modelu AR(3) získané zobecněnou metodou RALS



V následující tabulce 5 jsou uvedeny v dřívě použité formě předpovědi a jejich míry přesnosti. Vidíme, že kvalita zejména následných odhadů (na levé straně tabulky) se použitím metody RALS velmi výrazně zlepšila.

Tabulka 5: Předpovědi a jejich míry přesnosti pro vyrovnání modelem AR(3) metodou RALS

Datum	Čas t	Předpovězená hodnota	Skutečná hodnota	Absolutní chyba předpovědi	Absolutní chyba předpovědi ²	Procentní chyba v %
30. 4. 2013	1589	3724,322134	3744	19,67786579	387,218402	0,525584022
2. 5. 2013	1590	3742,487664	3825	82,51233552	6808,285513	2,157185242
3. 5. 2013	1591	3822,099387	3850	27,90061279	778,4441939	0,724691241
6. 5. 2013	1592	3844,081169	3838	6,08116872	36,980613	0,158446293
7. 5. 2013	1593	3828,931492	3819	9,931492182	98,63453697	0,260054784
9. 5. 2013	1594	3818,525585	3700	118,5255855	14048,31441	3,203394202
10. 5. 2013	1595	3698,139812	3630	68,1398118	4643,033952	1,877129802
13. 5. 2013	1596	3635,856075	3693	57,1439246	3265,428118	1,547357828
14. 5. 2013	1597	3708,330179	3810	101,6698214	10336,75258	2,668499249
15. 5. 2013	1598	3810,969372	3785	25,96937188	674,408276	0,686112864
16. 5. 2013	1599	3768,682575	3791	22,31742463	498,0674419	0,588694925
17. 5. 2013	1600	3784,629927	3782	2,629927316	6,916517688	0,069538004
MAD	45,2083				MAD*	58,4343188
MSE	3465,21				MSE*	6656,71078
MAPE v %	1,20556				MAPE* v %	1,68387887

8 Porovnání přesnosti předpovědi získaných pomocí jednotlivých modelů

Porovnání přesností modelů je v následujícím přehledu.

 Data proměnné: cena
 Počet pozorování = 1588

Modely

-
- (A) Náhodná procházka
 - (B) Prosté klouzavé průměry ze 3 členů
 - (C) Prosté exponenciální vyrovnání s $\alpha = 0,9999$
 - (D) Holtovo lineární vyrovnání s $\alpha = 0,9999$ a $\beta = 0,1$
 - (E) Model AR(3) odhadnutý zobecněnou MNČ v Statgraphics
 - (F) Model AR(3) odhadnutý metodou RALS v PcGive

Míry přesnosti předpovědí

Model	MSE	MAE	MAPE	ME
(A)	6769,68	58,3749	1,67941	0,37870
(B)	6730,36	58,2828	1,67561	0,71251
(C)	6774,00	58,375	1,67941	0,37874
(D)	7226,91	60,467	1,73671	-0,03327
(E)	6730,36	58,2828	1,67561	3,43492
(F)	6656,71	58,4343	1,68388	0,00000

Tabulka porovnává výsledky z šesti různých předpovědních modelů. Podíváme-li se na chybové statistiky, nejmenší střední kvadratickou chybu (MSE) při odhadu ceny má model (F). Nejmenší střední absolutní chybou (MAE) má model (E). Nejmenší střední absolutní procentuální chyba (MAPE) má opět model (E). Podle těchto výsledků je optimálním modelem model (E) nebo model (F). V následném hodnocení přesnosti je výrazně lepší model (F).

Všechny uvedené modely vykazují na 5% hladině významnosti významnou heteroskedasticitu a autokorelaci. Proto byla v posledním modelu použita k odhadu metoda, která redukuje autokorelaci.

9 Výsledky práce

Přínosem článku je aplikace předpovědních metod a postupů na konkrétních datech KB na burze cenných papírů v Praze. Zejména výsledky z posledního modelu velmi dobře kopírují skutečné hodnoty akcií. Jako nejoptimálnější model pro tato data se jeví model ARIMA(3,0,0), i když jeho použití není zcela bez problémů. Podle nejčastěji používané střední kvadratické míry přesnosti, bych jejich použití seřadil v tomto pořadí:

- ARIMA(3,0,0),
- model neměnných hodnot,
- model jednoduchého exponenciálního vyrovnání,
- model Holtova vyrovnání,
- model vyrovnání trojčlennými klouzavými průměry.

Model ARIMA je nekomplikovanější a vhodný i pro dlouhodobé předpovědi pomocí velmi dlouhých časových řad.

Model neměnných hodnot (náhodné procházky) je nejjednodušší a vhodný jen pro krátkodobé předpovědi.

10 Diskuze

Při použití uvedených metod přihlížíme jen k minulosti časové řady a nezjišťujeme, proč takové hodnoty řada nabyla. Tyto hodnoty jsou obvykle ovlivňovány i množstvím dalších ekonomických, sociálních, společenských a jiných vlivů, které nejsou v použitých modelech brány do úvahy. Proto je třeba brát výsledky předpovědí vždy s určitou rezervou a porovnávat je s předpověďmi kvalitativními, získanými týmy kvalifikovaných a zkušených odborníků nebo s výsledky ekonometrických analýz, které zahrnují do modelu více vysvětlujících proměnných. Jako příklad zde uvádíme propad cen akcií v letech 2008-2009 a jejich návrat na předkrizové ceny patrný z grafu č. 1, který popisuje předseda představenstva KB Henri Bonnet. „Komerční banka čelila hospodářskému poklesu poprvé ve své historii jako banka plně soukromá. Svou schopnost odolat ekonomické krizi opírá o dobře nastavený obchodní

model a strategii přijatou po privatizaci.“ Užitečné je i ze získaných předpovědí počítat průměry. Pokud by byla k dispozici data o těchto vlivech, bylo by vhodnější používat spíše ekonometrické modely založené na regresní analýze. Jako námět pro další výzkum bude užitečné řešit problémy s heteroskedasticitou, což znamená zabývat se některými modely volatility (např. ARCH nebo GARCH modely- viz Engle 1982).

Literatura

- [1] ARLT, J. a kol. *Analýza ekonomických časových řad a příklady*. Praha: VŠE, 2002. ISBN 80-245-0777-3.
- [2] ARLT, J., ARLTOVÁ, M. *Finanční časové řady*. Praha: Grada, 2003. ISBN 80-247-0333-0.
- [3] CIPRA, T. *Analýza časových řad s aplikacemi v ekonomii*. Praha: SNTL a Alfa, 1986.
- [4] ENGLE, R. F. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation. *Econometrica* 50, 987-1007, 1982.
- [5] GRIFFITS, W. E., HILL, R. C., JUDGE, G. G. *Learning and Practicing Econometrics*. N. York: John Wiley, 1992. ISBN 0-471-51364-4.
- [6] LEJNAROVÁ, Š., RÁČKOVÁ, A., ZOUHAR, J. *Základy ekonometrie v příkladech*. Praha: Nakladatelství Oeconomica, 2009. ISBN 978-80-245-1564-9.
- [7] MAKRIDATIS, S., WHEELWRIGHT, S. *Forecasting Methods for Management*. New York: John Wiley, 1989. ISBN 0-471-60063-6.
- [8] PETRÁŠKOVÁ, V. *Prognostické modely v oblasti modelování finančních časových řad*. Dizertační práce. Praha: Česká zemědělská univerzita, 2006.
- [9] STUHLÝ, J. *Statistické metody pro manažerské rozhodování*. Jindřichův Hradec: VŠE, Fakulta managementu, 2004. ISBN 80-245-0153-8.
- [10] STUHLÝ, J., ZEMAN, R. *Analýza závislostí a předpovědi v modelech cenných papírů*. Soubor příspěvků z celoškolského semináře „Tvůrčí činnost jako proces vnímání, poznání a seberealizace“. České Budějovice: Vysoká škola technická a ekonomická, 2013, s. 336 – 346.
- [11] Výroční zpráva Komerční banky, a.s., str. 8