

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH

Ekonomická fakulta

Bakalářská práce

2013

Tomáš Šmarda

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH

Ekonomická fakulta

KATEDRA APLIKOVANÉ MATEMATIKY A INFORMATIKY

Studijní program: B1103 Aplikovaná matematika

Studijní obor: Finanční a pojistná matematika

Matematické modelování ceny zlata

Vedoucí bakalářské práce:
doc. RNDr. Tomáš Mrkvička, Ph.D.

Autor:
Tomáš Šmarda

Prohlašuji, že svoji bakalářskou práci jsem vypracoval samostatně, pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47 zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

Ve Velešíně 19. dubna 2013

.....

Tomáš Šmarda

Poděkování

Tímto bych rád poděkoval vedoucímu mé bakalářské práce doc. RNDr. Tomáši Mrkvičkovi, Ph.D. za cenné rady a připomínky, které mi pomohly při zpracování daného tématu.

Děkuji své rodině za jejich dosavadní podporu při studiu.

Obsah

1.	ÚVOD	3
2.	ZÁKLADY BOX-JENKINSOVI METODOLOGIE	5
2.1.	Stacionarita	6
2.2.	Autokorelační funkce, ACF	7
2.3.	Parciální autokorelační funkce, PACF.....	8
2.4.	Autoregresní proces, AR(p).....	8
2.5.	Proces klouzavých průměrů, MA(q).....	11
2.6.	Smíšené modely ARMA	12
2.7.	Transformace časových řad	12
2.8.	Integrovaný smíšený model, ARIMA.....	13
3.	MODELY VOLATILITY	14
3.1.	ARCH(1).....	14
3.2.	GARCH(1,k).....	15
4.	APLIKACE POPSANÝCH MODELŮ	16
4.1.	ARIMA	17
4.1.1.	Identifikace modelu	19
4.2.	GARCH.....	22
4.2.1.	Odhad modelů.....	25
4.2.2.	Předpověď	30
4.3.	Porovnání predikčních intervalů	30
5.	ZÁVĚR	31

6.	SUMMARY	33
7.	ZDROJE	34
8.	SEZNAM OBRÁZKŮ	35
9.	SEZNAM TABULEK	36
10.	PŘÍLOHY	37

1. Úvod

Matematické modelování se pomocí matematických funkcí a rovnic snaží popsat a patřičně zjednodušit pozorované jevy, které se odehrávají kolem nás. Matematické modely nachází své uplatnění nejen v přírodních vědách, jako je meteorologie, geologie, fyzika ale i ve vědách společenských, jako ekonomie, sociologie, psychologie atd.

Konkrétně v ekonomii, či vědách ekonomii blízkých se matematické modely zaměřují například na operační management, kdy se používají matematické modely pro popis nabídky nebo poptávky po zboží a službách a následnému plánování výroby. V řízení rizik se investoři snaží modelovat a předpovídat výnosy svých investic. Tyto investice kromě akcií, dluhopisů a komodit mohou zahrnovat i různé deriváty, pro jejichž správné oceňování se používá právě matematických modelů. Vlády a jiné centrální instituce analyzují makroekonomické ukazatele HDP, nezaměstnanosti, sledují vývoj populace a na základě těchto modelů a předpovědí přizpůsobují svoji fiskální nebo monetární politiku. Toto je výčet pouze několika případů, kde je možné využití matematického modelování.

Už od samého počátku lidské civilizace sloužilo zlato jako peníze. Postupně, spolu se stříbrem, vytlačilo z trhu ostatní používaná platidla. V dnešní době bylo zlato a takzvaný Zlatý standard (1815 - 1914) nahrazeno ničím nekrytými Fiat měnami vydávanými centrálními bankami. Tento fakt je poměrně důležitý a zajímavý, protože pokud by stále platil Zlatý standard ve své původní podobě, tak bychom žádnou cenu zlata modelovat nemohli.

V první části této práce si stručně představíme Box-Jenkinsovu (1970) metodologii a základní modely, s nimiž pracuje, MA, AR, jejich kombinace ARMA, ARIMA.

Druhá část nám poskytne úvod k modelům volatility, které v posledních letech zaznamenaly bouřlivý rozvoj a v současné době jsou často využívány pro řízení rizika ve finančních institucích. V této části si představíme základní model volatility prezentovaný modelem ARCH (Engle, 1982) a později zjednodušený modelem GARCH (Bollerslev, 1986).

Ve třetí části se pokusím časovou řadu ceny zlata modelovat pomocí zpracované teorie. Cenu budu modelovat pomocí modelu ARIMA a volatilitu modelem GARCH. Získané modely následně využijeme k předpovědi budoucí ceny zlata o 1 krok dopředu.

K výpočtům budu používat statistický software R[7] a MS Excel. Veškeré obrázky a tabulky pochází z vlastních výpočtů.

2. Základy Box-Jenkinsovi metodologie

V moderní statistické analýze časových řad má významné postavení Box-Jenkinsova metodologie, která je považována za základ dnešního matematického modelování časových řad.

Dekompoziční metody spočívají v rozkladu řady na její jednotlivé složky. Box-Jenkinsova metodologie považuje za hlavní prvek při konstrukci matematických modelů nesystematickou (reziduální) složku, která může být tvořena navzájem korelovanými náhodnými veličinami. Box-Jenkins dokonce z tohoto předpokladu vychází a při konstrukci modelů jsou zkoumány právě tyto vazby. Jednou z podmínek, abychom vůbec mohli tyto postupy aplikovat, je požadavek na délku analyzované řady. Ta musí obsahovat alespoň 50 pozorování.

Box-Jenkinsova metodologie pracuje s těmito základními modely:

- Model klouzavých součtů (MA - Moving average model)
- Model autoregresní (AR - Autoregressive model)
- Kombinace klouzavých součtů a autoregresního modelu (např. ARMA, ARIMA)

Směšené modely jsou značně flexibilní a umožňují nám modelovat i nestacionární řady, které rychle mění trend a jen těžko bychom je modelovali pomocí dekompozičních metod.

Hlavní tři kroky, jak postupovat při výstavbě modelů této metodologie jsou:

- Identifikace modelu
- Odhad parametrů modelu
- Ověření a případná úprava modelu

Jako nevýhody této metodologie Cipra[2] uvádí:

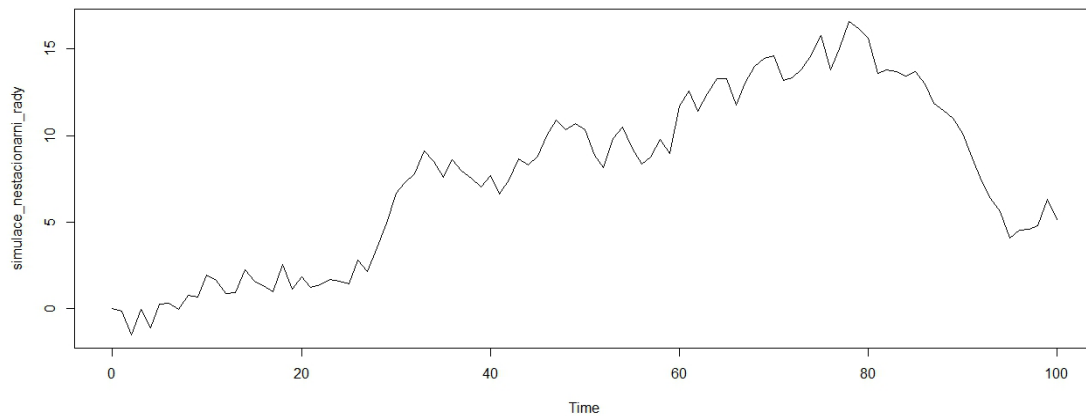
- Nutnost dostatečně dlouhé řady (minimálně 50 a více hodnot).
- Složitá interpretace výsledků, pokud model obsahuje více parametrů.

2.1. Stacionarita

¹Základem pro analýzu časových řad v Box-Jenkinsově metodologii je stacionarita. Striktně stacionární je proces, ve kterém je statistické chování konkrétní řady invariantní vůči časovým posunům. To znamená, že střední hodnota, rozptyl a autokorelační struktura se nám v čase nemění. Tento předpoklad je velmi těžké empiricky ověřit a tak v praxi častěji předpokládáme slabou stacionaritu. Proces je slabě stacionární, pokud střední hodnota a rozptyl je v čase invariantní a kovariance dvou proměnných závisí pouze na vzdálenosti, která tyto dva body odděluje. Ze slabé stacionarity vyplývá, že data v grafu časové řady budou kolísat okolo konstantní úrovně s konstantním rozptylem.

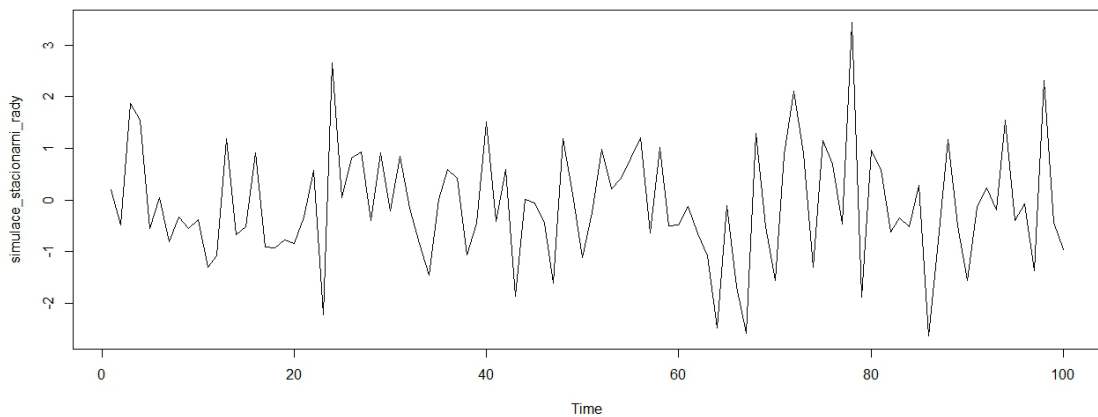
„V Box-Jenkinsově metodologii lze modelovat pouze řady stacionární, přičemž pomocí různých transformací lze většinu nestacionárních řad převést na stacionární.“ [Cipra, 1986 s. 102]. Některé základní transformace řad si ukážeme v kapitole 2.7.

Pro názornost jsem v programu R nasimuloval nestacionární a stacionární řadu.



Obr. 1: Nestacionární řada

¹ Tuto kapitolu jsem zpracoval podle [6].



Obr. 2: Stacionární řada

2.2. Autokorelační funkce, ACF

Pokud budeme uvažovat slabě stacionární proces $\{y_t\}$, lze vyjádřit odhad autokorelační funkce jako:

$$\widehat{\rho}_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=k}^T (y_t - \bar{y})^2} \quad (1)$$

Vlastnosti ACF nám umožňují pozorovat korelaci mezi jednotlivými proměnnými $\{y_t\}$ a $\{y_{t-k}\}$ též časové řady, kde intervalu k říkáme zpoždění. ACF je pro všechna k symetrická okolo 0 a nabývá hodnot od -1 do 1. Jednotlivé odhady ACF můžeme vykreslit do takzvaného korelogramu. V literatuře se často uvádí jako minimální počet pozorování $T > 50$ a k okolo $T/4$, ale ve většině případů postačuje vykreslit ACF pro prvních alespoň 20 až 30 hodnot.

ACF má v Box-Jenkinsově metodologii důležité postavení. Její průběh nám naznačuje, jaký řád modelu MA máme pro modelování řady použít. Při zkoumání ACF se hlavně snažíme najít nulové body, tzn. body, za kterými je ACF nulová a příslušné hodnoty zpoždění k již nevykazují autokorelaci.

Nulové body, takzvané body useknutí se určují pomocí směrodatné odchylky, která se nejčastěji odhaduje Bartlettovou aproximací².

² Bartlettovu aproximaci můžeme nalézt například v [2].

2.3. Parciální autokorelační funkce, PACF

Parciální autokorelační funkce je dalším z důležitých nástrojů Box-Jenkinsovi analýzy. Korelace mezi dvěma proměnnými často způsobuje třetí proměnná, která je ovlivňuje. PACF zkoumá autokorelaci mezi dvěma pozorováními $\{y_t\}$ a $\{y_{t-k}\}$, ovšem očištěné o vliv hodnot ležících mezi nimi $\{y_{t-k+l}\}$. Způsob výpočtu odhadu PACF už není tak jednoduchý jako u ACF a nebudeme ho zde uvádět³.

Stejně jako ACF i PACF se používá k identifikaci příslušného modelu, konkrétně AR. Dále se snažíme určit body useknutí, za kterými je PACF nulová. Pomocí Quenouilleovi aproximace tyto body můžeme najít.⁴

2.4. Autoregresní proces, AR(p)

Proces AR(1)

Tento model můžeme vyjádřit jako

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \quad (2)$$

Kde $\{\varepsilon_t\}$ označuje proces bílého šumu. Což je proces IID veličin⁵. Model (2) můžeme přepsat pomocí operátoru zpětného posunutí následovně

$$(1 - \phi_1 B)y_t = \varepsilon_t \quad (3)$$

Operátor zpětného posunutí představuje zpoždění o jeden krok. Platí tedy, že

$$By_t = y_{t-1} \quad (4)$$

³ Způsob výpočtu PACF můžeme nalézt například v [4].

⁴ Quenouilleovu aproximaci můžeme nalézt například v [2].

⁵ Independent and Identically Distributed. Nezávisle a identicky rozdělené náhodné veličiny. Za bílý šum můžeme označit proces s obvykle nulovou střední hodnotou, konstantním rozptylem. Bílý šum poznáme podle nulových hodnot ACF a PACF.

Proces AR(2)

Tento model obsahuje druhý parametr ϕ

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \quad (5)$$

AR(2) při použití operátoru zpětného posunutí

$$(1 - \phi_1 B - \phi_2 B^2) y_t = \varepsilon_t \quad (6)$$

Proces AR(p)

Rovnici modelu AR(p) zapíšeme podobně jako předchozí dva modely

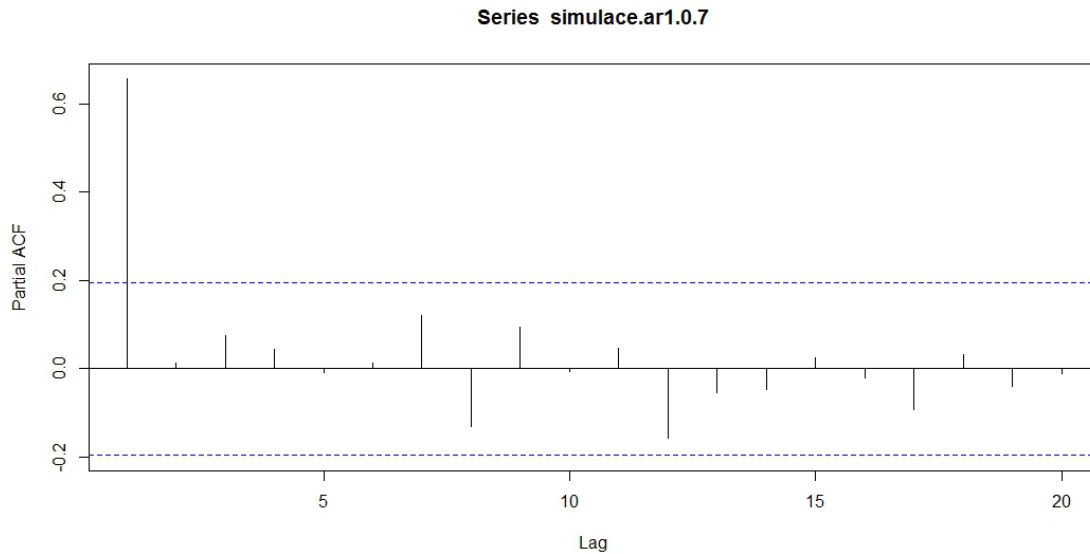
$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (7)$$

S použitím operátoru zpětného posunutí

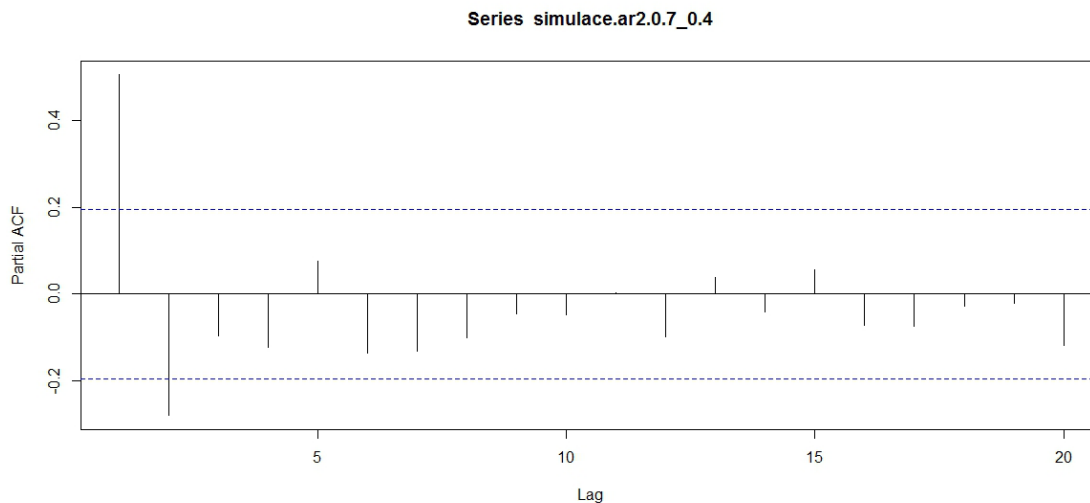
$$\phi_p(B) y_t = \varepsilon_t \quad (8)$$

Existují dva přístupy pro určení řádu p . Identifikace řádu AR pomocí průběhu PACF nebo pomocí nějakého informačního kritéria. My se zaměříme na identifikaci řádu p pomocí PACF.

V programu R jsem nasimuloval dvě časové řady o délce 100 pozorování (Příloha Obr. 15 a Obr. 17) typu AR(p), které pro názornost budou stačit.



Obr. 3: PACF řady simulované procesem $y_t = 0,7y_{t-1} + \varepsilon_t$



Obr. 4: PACF řady simulované procesem $y_t = 0,7y_{t-1} - 0,4y_{t-2} + \varepsilon_t$

Jak můžeme vidět na jednotlivých korelogramech, PACF klesá k nule po prvních p odhadech PACF. Body useknutí nám označují vodorovné čerchované čáry. Hodnoty mezi nulou a touto hranicí považujeme za nulové a nevýznamné. Na obrázku (3) se jedná o proces s jedním kladným parametrem. První řada má významnou pouze první hodnotu, která je kladná. Všechny další hodnoty jsou nulové.

Na druhém obrázku (4) vidíme statisticky významné pouze dvě hodnoty. První je kladná, druhá záporná stejně jako parametry tohoto modelu.

2.5. Proces klouzavých průměrů, MA(q)

Proces MA(1)

Základní model MA(1) zapíšeme takto

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (9)$$

Stejně jako proces AR, můžeme i tento vyjádřit s operátorem zpětného posunutí

$$y_t = (1 - \theta_1 B) \varepsilon_t \quad (10)$$

Proces MA(2)

Model MA se dvěma parametry

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \quad (11)$$

Opět vyjádření pomocí operátoru zpětného posunutí

$$y_t = (1 - \theta_1 B - \theta_2 B^2) \varepsilon_t \quad (12)$$

Proces MA(q)

Vyjádření modelu MA(q)

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (13)$$

Model MA řádu q zapsaný pomocí operátoru zpětného posunutí

$$y_t = \theta_1(B) \varepsilon_t \quad (14)$$

Při určování řádu q modelu MA postupujeme obdobně jako při určování řádu p u AR modelu. Jediný rozdíl je, že teď zkoumáme ACF místo PACF. Sledujeme jednotlivé hodnoty a jejich významnost. Po prvních q hodnot je ACF nulová a my tak můžeme určit řád našeho modelu, viz příloha Obr. 20, 21, 22.

2.6. Smíšené modely ARMA

V některých aplikacích modely AR a MA selhávají, protože k popisu časové řady potřebují mnoho parametrů. Tomu se dá předejít použitím smíšeného modelu ARMA(p,q). V podstatě kombinují modely AR a MA do jednoho. Model ARMA je také používán v modelech podmíněné heteroskedasticity. Například model GARCH může být za jistých okolností považován za model ARMA aplikovaný na podmíněný rozptyl.

Základní model ARMA(1,1) má tvar

$$(1 - \phi_1 B)y_t = (1 - \theta_1 B)\varepsilon_t \quad (15)$$

Obecný model ARMA(p,q) zapíšeme takto

$$\phi_p(B)y_t = \theta_q(B)\varepsilon_t \quad (16)$$

V praxi často vystačíme s modely nízkých řádů, většinou $p, q \leq 2$. Určení řádu modelu podle ACF a PACF není při použití modelu ARMA jednoduché. Většinou nám totiž korelogramy mnoho neřeknou a musíme zkusit nějaký model, který posléze příslušně upravíme. Nejvhodnější model pak posoudíme podle informačního kritéria (např. AIC, BIC atd.)

2.7. Transformace časových řad

Doposud jsme se zabývali modely, které jsou stacionární. V další kapitole si představíme model ARIMA, který nám umožňuje modelovat i nestacionární časové řady. Ty je však zapotřebí před samotnou analýzou příslušně upravit do stacionárního tvaru. V krátkosti si představíme dvě transformace⁶, používané v praktické části.

⁶ Některé další možné transformace lze najít v [2], [4].

Diferencování

Diferencování nám umožňuje odstranit z časové řady trendovou složku, a tak modelovat i nestacionární procesy. První diference můžeme zapsat takto

$$\Delta y_t = y_t - y_{t-1} \quad (17)$$

Logaritmická transformace

Logaritmování řady se používá před diferencováním. Ceny totiž nemohou být záporné, což se o diferencích říci nedá. Tato transformace má dva velké přínosy. Částečně nám stabilizuje rozptyl a diferencováním logaritmované řady získáme aproximované logaritmy výnosů $\{r_t\}$.

$$\ln(r_t) = \ln\left(\frac{y_t}{y_{t-1}} - 1\right) \cong \ln(y_t) - \ln(y_{t-1}) \quad (18)$$

2.8. Integrovaný smíšený model, ARIMA

Model ARIMA(p,d,q) můžeme vyjádřit jako stacionární model ARMA aplikovaný na časovou řadu diferencí řádu d

$$\phi_p(B) \Delta^d y_t = \theta_q(B) a_t \quad (20)$$

3. Modely volatility

V modelech Box-Jenkisovi metodologie se pracuje s předpokladem, že náhodná složka procesu je bílý šum o nulové střední hodnotě a konstantním (nepodmíněným) rozptylu. V praxi, ve finančních řadách obzvlášť, se však setkáváme s nekonstantním v čase proměnlivým (podmíněným) rozptylem. Tuto vlastnost dokážou modely volatility zachytit. Dále řady často vykazují leptokurtické rozdělení, charakteristické svou vyšší špičatostí a „těžšími konci“. Pákový efekt způsobuje vyšší nárůst volatility při cenovém poklesu než při cenovém vzestupu. Časté je také shlukování volatility, kdy velké cenové výkyvy následují opět velké výkyvy a naopak [3]. Modely typu ARCH poprvé představil roku 1982 Robert F. Engle III. V roce 2003 dostal za popsání tohoto modelu Nobelovu cenu za ekonomii.

3.1. ARCH(1)

Při modelování podmíněného rozptylu se vychází z tohoto vztahu

$$y_t = \mu + e_t \quad (21)$$

Kde $\{\mu\}$ značí nějaký model, například ARIMA a $\{e_t\}$ je v tomto případě

$$e_t = \sqrt{h_t} \varepsilon_t, \quad \{\varepsilon_t\} \sim IID(0,1) \quad (22)$$

Náš model volatility ARCH(1) má tvar

$$h_t = \omega + \alpha_1 e_{t-1}^2 + \dots + \alpha_l e_{t-l}^2 \quad (23)$$

Modely ARCH se od svého vzniku značně rozšířili a vzniklo mnoho modifikací.

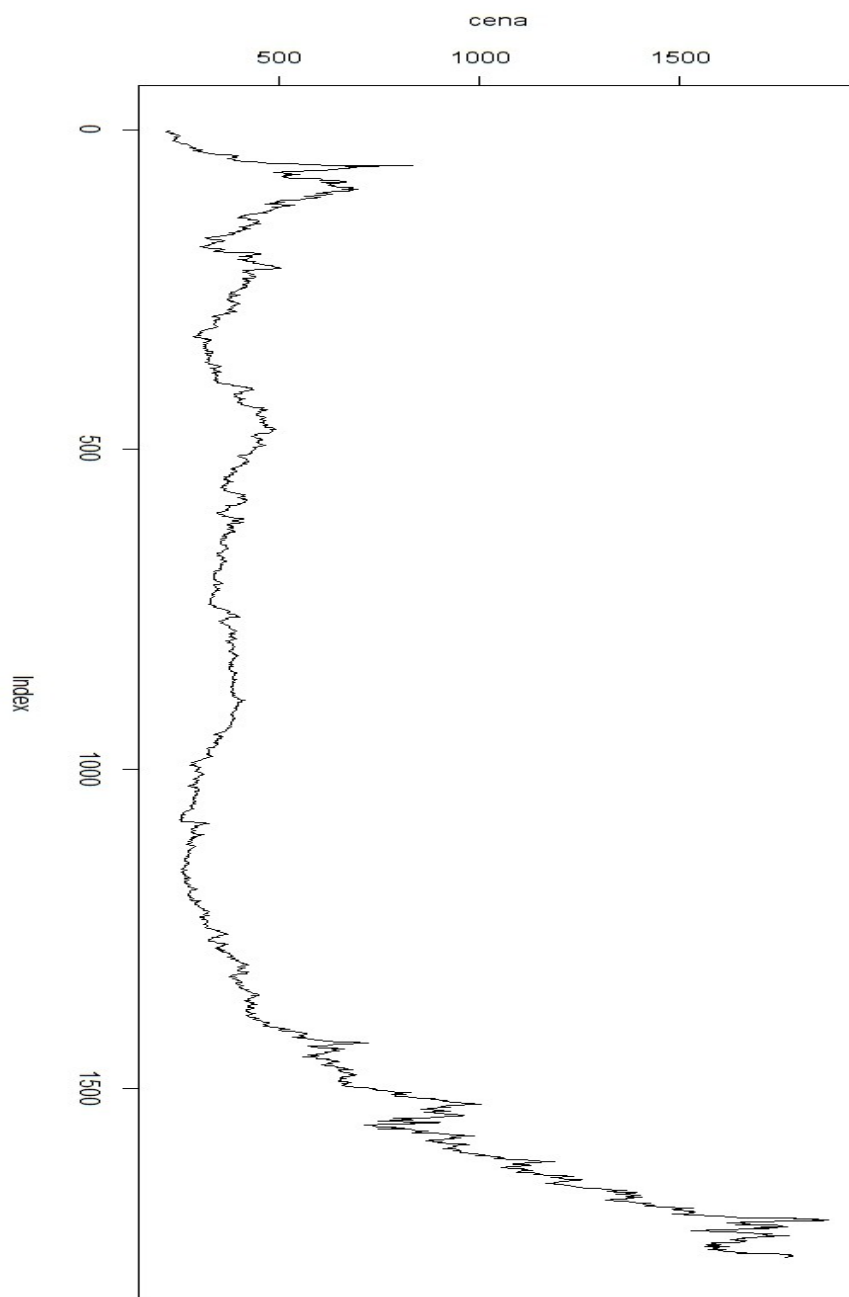
3.2. GARCH(1,k)

Tento model v roce 1986 specifikoval Tim Bollerslev. Aplikace modelu ARCH vede často k nutnosti odhadnout velké množství parametrů. Generalizovaný model, kromě minulých hodnot $\{e_i\}$, počítá také s minulými hodnotami $\{h_t\}$. V praxi si vystačíme s modely nízkých řádů jako (1,1), (2,1), atd.

$$h_t = \omega + \sum_{i=1}^l \alpha_i e_{t-i}^2 + \sum_{j=1}^k \beta_j h_{t-j} \quad (24)$$

4. Aplikace popsaných modelů

Pro praktickou část bakalářské práce jsem zvolil časovou řadu týdenních zavíracích cen zlata⁷ v USD za trojskou unci⁸. Veškerá používaná data, jsou obsažena na příloženém CD, stejně tak všechny provedené výpočty v programech Excel a R.



Obr. 5: Časová řada ceny zlata od 29. 12. 1978 do 28. 9. 2012.

⁷ Zdroj: <http://www.gold.org/>

⁸ 1 oz = 31,1034768 gramů

4.1. ARIMA

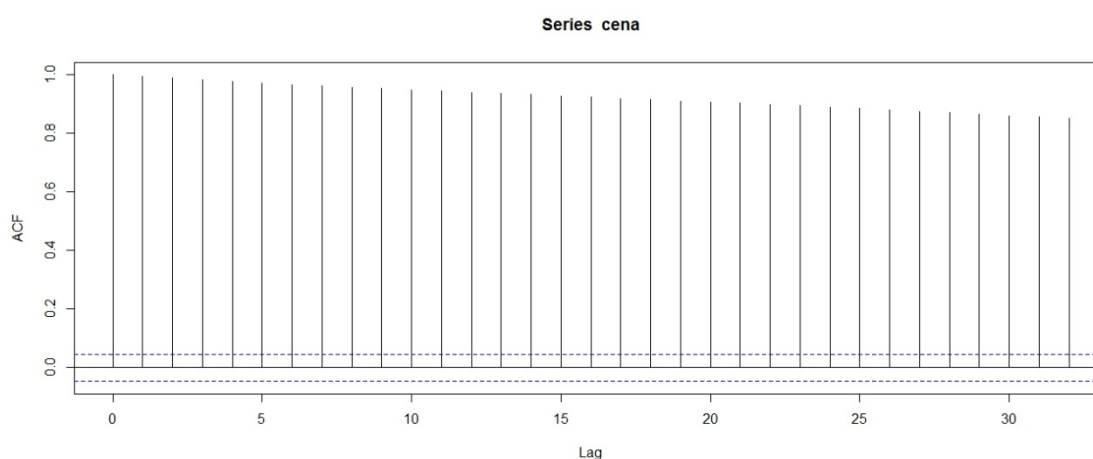
V této části budeme pracovat s řadou o délce 1762 pozorování od 29. 12. 1978 do 28. 9. 2012. Poslední hodnotu ceny z 5. 10. 2012 vynecháme pro porovnání s předpovědí. Už od pohledu na obrázek (5) můžeme předpokládat, že řada není stacionární. Budeme ji tedy muset vhodně transformovat.

Průběh autokorelační funkce nám potvrzuje výše zmíněnou domněnku, že se jedná o nestacionární řadu. Na obrázku (6) vidíme ACF, jak nám klesá velmi pozvolně lineárně, a zároveň je první hodnota PACF (obr. 7) blízká jedné. Nejprve tedy řadu logaritmuje a poté diferencujeme, abychom dosáhli stacionarity. Zvolíme diferencování řádu 1. S vyššími řády diferencování, dochází k nárůstu rozptylu a k přediferencování řady⁹. Můžeme si všimnout větších rozptylů pro druhou ($\hat{\sigma}_{\Delta^2}^2$) a třetí diferenci ($\hat{\sigma}_{\Delta^3}^2$).

$$\hat{\sigma}^2 = 0.000712545$$

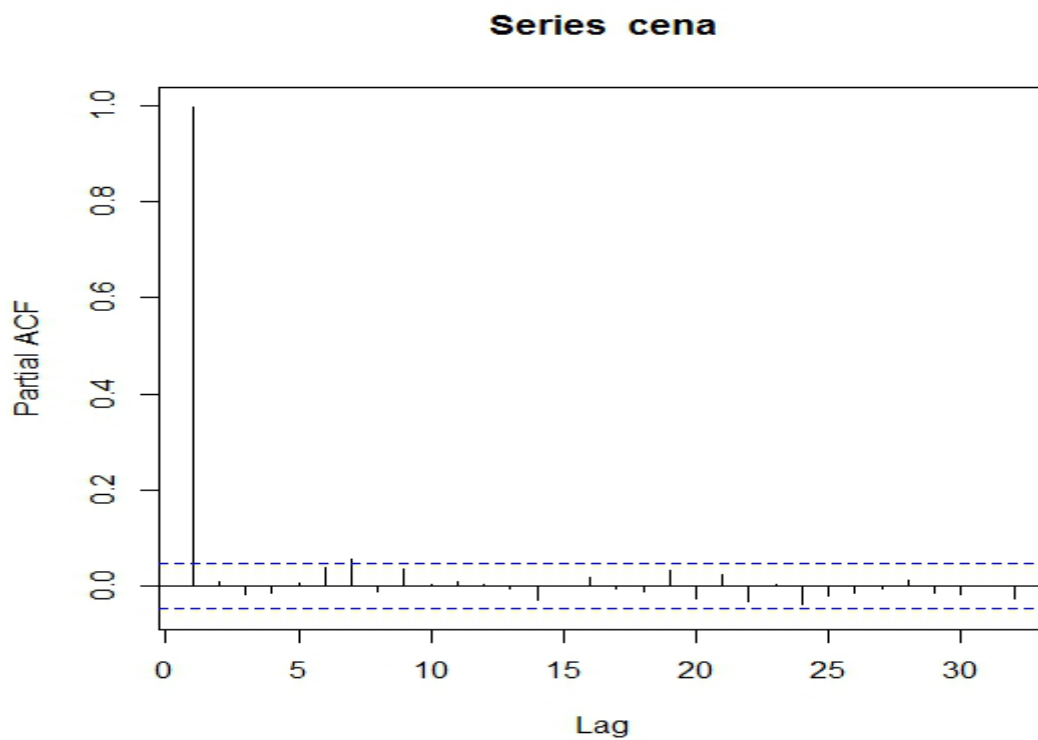
$$\hat{\sigma}_{\Delta^2}^2 = 0.001437771$$

$$\hat{\sigma}_{\Delta^3}^2 = 0.004396908$$



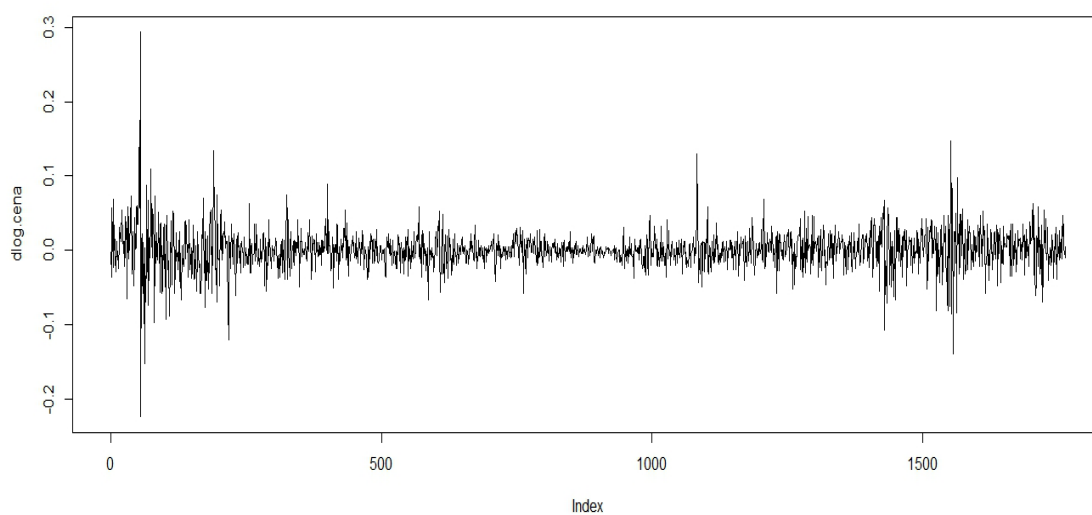
Obr. 6: ACF ceny zlata

⁹ O určení řádu diferencování se můžeme dočíst v [2]



Obr. 7: PACF ceny zlata

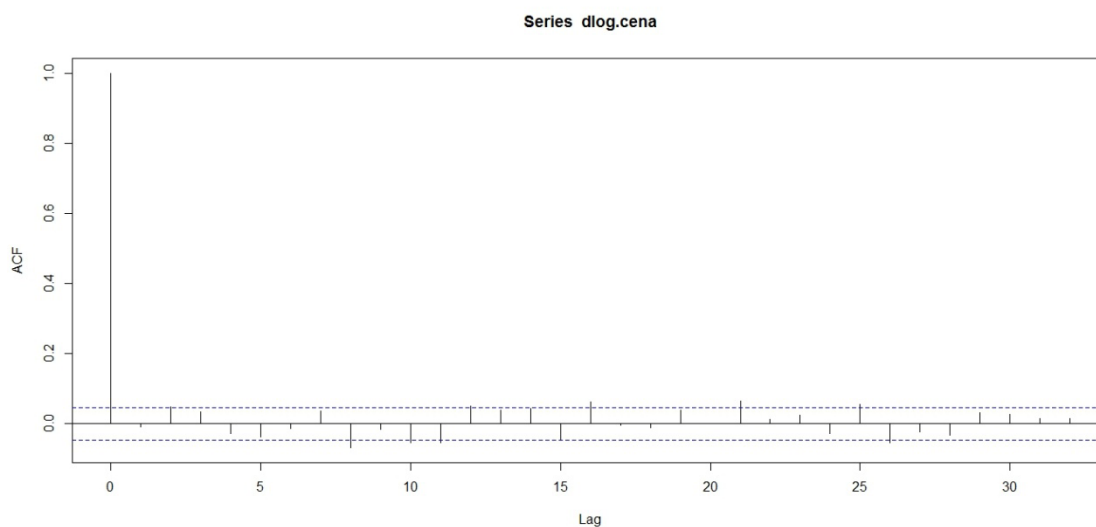
Po logaritmování a diferencování máme na obrázku (8) stacionární časovou řadu ceny zlata připravenou k dalšímu kroku.



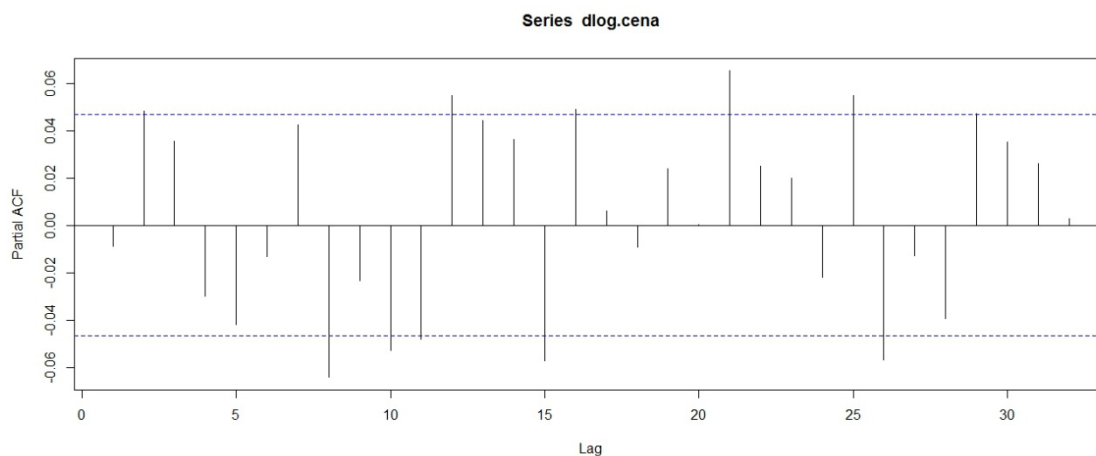
Obr. 8: Logaritmovaná cena zlata a její první diference

4.1.1. Identifikace modelu

V této části bychom se měli pokusit identifikovat řady procesů AR a MA. Pokud si ovšem vykreslíme funkce ACF a PACF, zjistíme, že všechna zpoždění naší řady jsou nulová. I když si můžeme všimnout některých nenulových hodnot, budeme tuto řadu považovat za bílý šum. Musíme mít na paměti, že pracujeme pouze s odhady těchto funkcí a jako takové jsou zatíženy statistickou chybou.



Obr. 9: ACF diferencí logaritmované ceny zlata



Obr. 10: PACF diferencí logaritmované ceny zlata

Podle průběhu ACF a PACF tedy můžeme říci, že řada neobsahuje žádný z procesů AR nebo MA, a proto by bylo vhodnější použití jiného, jednoduššího modelu. Časovou řadu se pokusíme popsat růstovým modelem náhodné procházky¹⁰.

Model si můžeme představit takto:

$$\ln(y_t) - \ln(y_{t-1}) = \mu$$

Kde $\{\mu\}$ je průměr řady diferencí logaritmovaných cen zlata.

Řadu tedy popíšeme jako:

$$\ln(y_t) - \ln(y_{t-1}) = 0,001170689$$

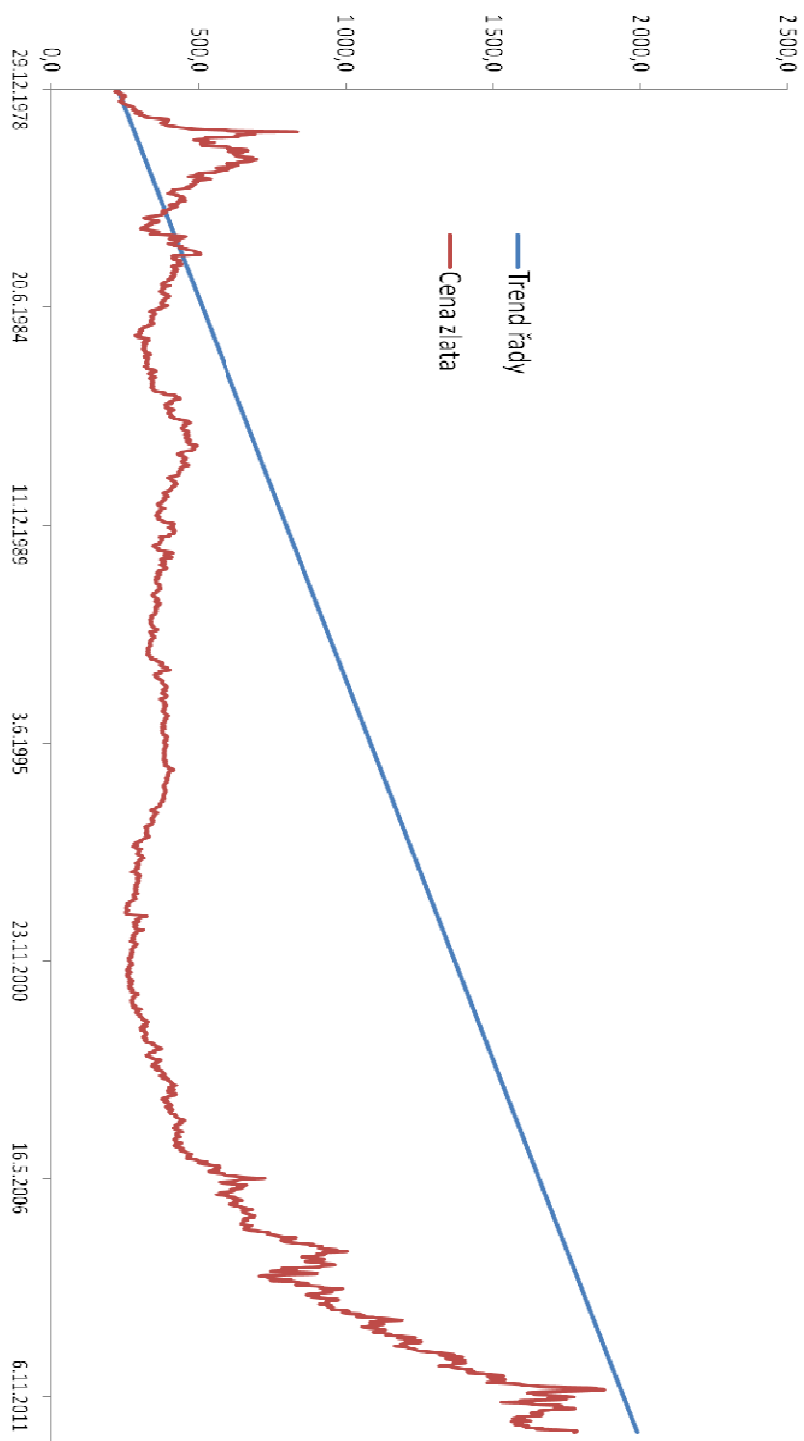
Předpovědi naší řady pomocí modelu:

$$\ln(y_{t+1}) = \ln(y_t) + 0,001170689$$

$$y_{t+1} = e^{(\ln(y_t)+0,001170689)}$$

Pro předpověď jsem vypustil poslední hodnotu (28. 9. 2012) abych mohl porovnat předpověď se skutečností. Růstovým modelem náhodné procházky nelze ze své podstaty v krátkodobém horizontu předvídat. Vypočteme tedy predikční intervaly a v kapitole 4.3 je porovnáme s následujícím modelem. Dále si pro představu vykreslíme trend řady daný načítáním střední hodnoty diferencí (Obr. 11).

¹⁰ Více o náhodné procházce např. v [1]



Obr. 11: Trend řady

4.2. GARCH

V této části se budu snažit aplikovat model volatility, konkrétně model GARCH(p, q) na stejnou časovou řadu ceny zlata od 29. 12. 1978 do 28.9 2012 o délce 1762 pozorování. Při pohledu na původní řadu a tvary funkcí ACF a PACF ji musíme nejprve stacionarizovat. To provedeme opět pomocí logaritmování a následnou diferencí. Stacionární řadu můžeme vidět na obrázku (8). Při podrobnějším zkoumání řady můžeme vypořadovat podmíněný rozptyl, tzn. období ve, kterém dochází k vysokému kolísání ceny. Můžeme si tak všimnou zvýšené volatility na začátku řady. Jedná se například o nafouknutí a splasknutí bubliny okolo roku 1980, kterou způsobila ropná krize a následná inflace. Další významný výkyv ceny můžeme pozorovat z nedávné doby, kdy na přelomu roku 2007 a 2008 došlo k hypoteční krizi v USA a následné světové finanční krizi.

Řada tedy neobsahuje žádnou systematickou složku, pokud by obsahovala, museli bychom ji z řady odstranit a dále pracovat s jejími rezidui.

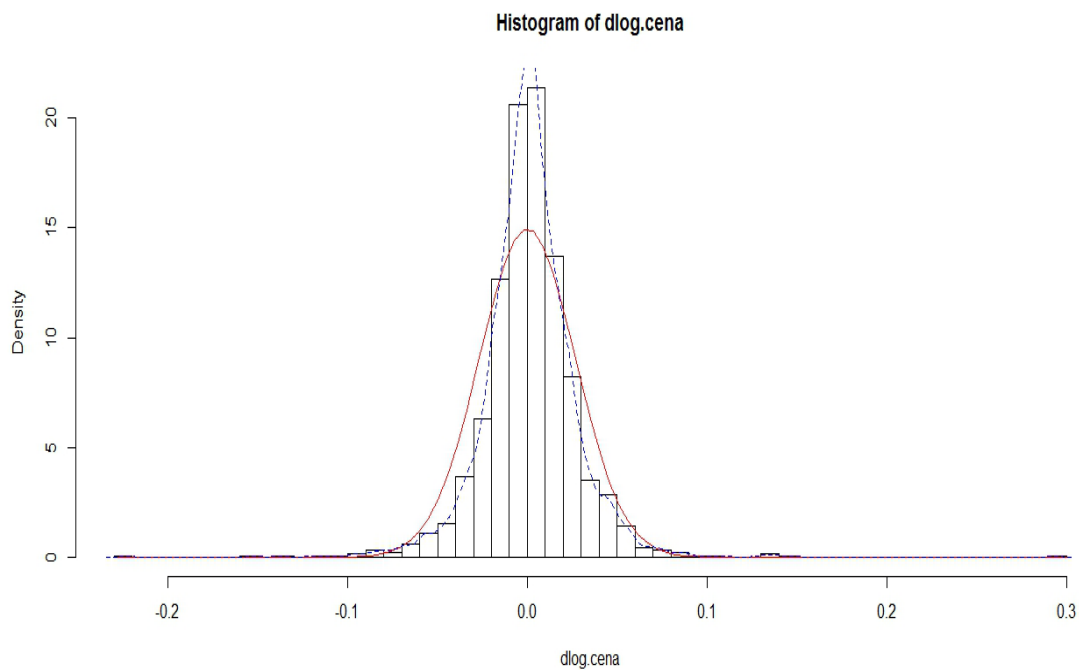
Tento náš předpoklad můžeme ověřit pomocí portmanteau Q-testu v programu R, kde nám p-hodnota nevyvrátila H_0 , že řada neobsahuje korelace systematické složky oproti $H_A = non H_0$.

```
Výstup z R: > Auto.Q(dlog.cena)
```

```
$Pvalue
```

```
[1] 0.8900425
```

Na obrázku (12) vidíme histogram skutečného rozdělení logaritmů výnosů řady (modře) a proložení normálním rozdělením (červeně). Můžeme vypořadovat špičatější rozdělení s těžšími konci typické pro finanční řady.



Obr. 12: Histogram logaritmů výnosů

Řadu otestujeme Jarque-Bera testem, kdy zamítáme $H_0 = o$ normalitě rozdělení

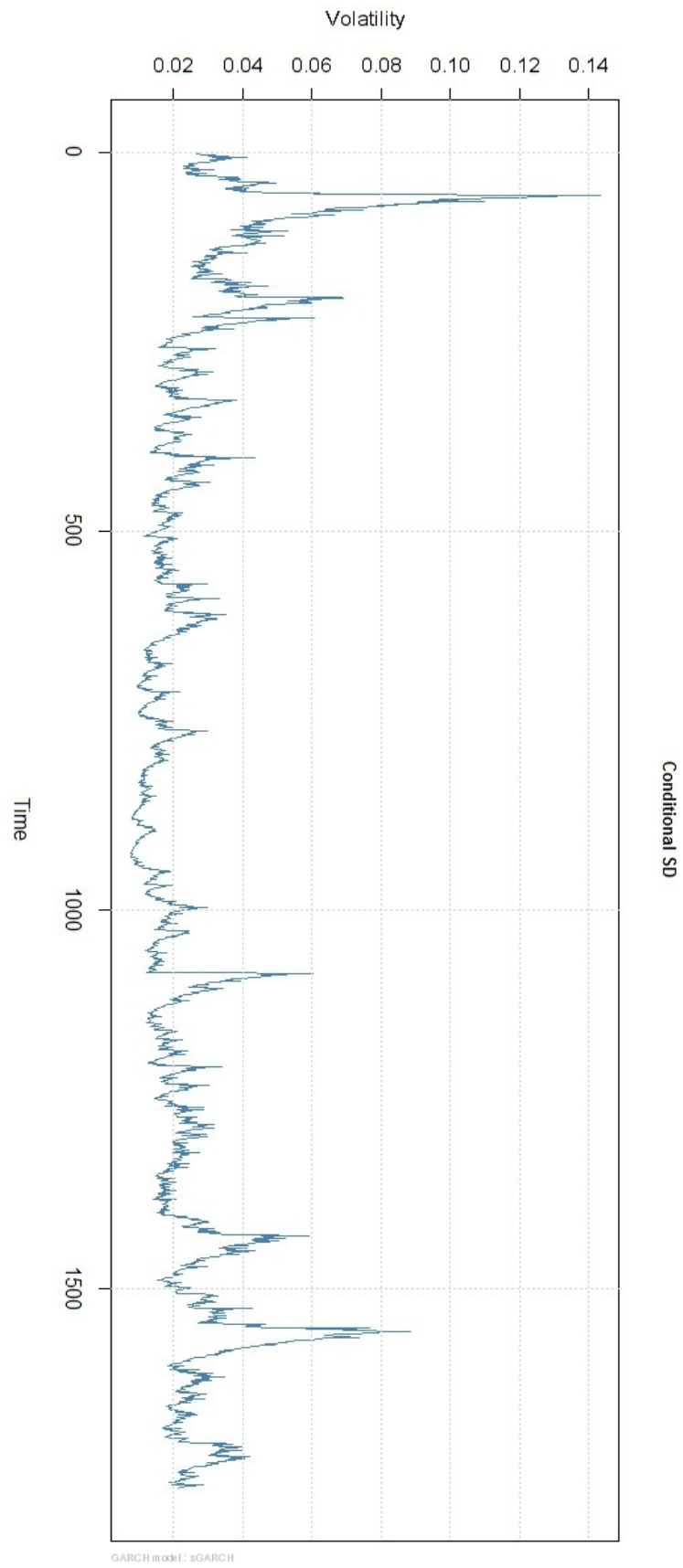
Výstup z R: `> jarque.bera.test(dlog.cena)`

Jarque Bera Test

data: dlog.cena

X-squared = 14415.04, df = 2, p-value < 2.2e-16

Na obrázku (13) můžeme vidět podmíněný rozptyl logaritmů výnosů a lépe tak pozorovat kolísání výnosů, největší okolo let 1980 a 2007-2008.



Obr. 13: Podmíněný rozptyl diferencí logaritmované ceny zlata

Že časová řada vykazuje podmíněnou heteroskedasticitu, se můžeme ujistit použitím ARCH-LM testu

```
Výstup z R: > ArchTest(dlog.cena, lags= 5)

ARCH LM-test; Null hypothesis: no ARCH effects

data: dlog.cena

Chi-squared = 291.1567, df = 5, p-value < 2.2e-16
```

Z výstupu vidíme nízkou p-hodnotu. Zamítáme tak nulovou hypotézu, o konstantním rozptylu.

4.2.1. Odhad modelů

Dále modelujeme řadu pomocí GARCH. Jak bylo popsáno v kapitole (3.2) použijeme modely nízkých řádů GARCH(1,1), GARCH(1,2), GARCH(2,1), kde hlavním kritériem pro výběr modelu bude Akaikeho informační kritérium (AIC).

GARCH (1,1)

Výstup z R: GARCH(1,1)

	Estimate	Std. Error	t value	Pr(> t)
omega	7.183e-06	2.562e-06	2.804	0.00505 **
alpha1	1.504e-01	1.953e-02	7.705	1.31e-14 ***
beta1	8.506e-01	1.787e-02	47.599	< 2e-16 ***

Standardised Residuals Tests:

	Statistic	p-Value
Jarque-Bera Test R Chi ²	181.059	0
Shapiro-Wilk Test R W	0.9863728	7.150675e-12
Ljung-Box Test R Q(10)	15.18437	0.1254834
Ljung-Box Test R Q(15)	22.56689	0.09377866
Ljung-Box Test R Q(20)	30.0206	0.06952061
Ljung-Box Test R ² Q(10)	20.57162	0.02428683
Ljung-Box Test R ² Q(15)	28.20833	0.02030077
Ljung-Box Test R ² Q(20)	30.77005	0.05830026
LM Arch Test R TR ²	22.94184	0.02822381

Information Criterion Statistics:

AIC	BIC	SIC	HQIC
-4.845093	-4.835776	-4.845098	-4.841650

Tento model nebude vhodný. Ljung-Box Test prokazuje autokorelovanost reziduí a ARCH-LM test potvrzuje další ARCH efekt.

GARCH(2,1)

Výstup z R: GARCH(2,1)

	Estimate	Std. Error	t value	Pr(> t)
omega	7.227e-06	2.802e-06	2.579	0.0099 **
alpha1	1.507e-01	2.548e-02	5.917	3.28e-09 ***
alpha2	1.000e-08	3.593e-02	0.000	1.0000
beta1	8.502e-01	2.540e-02	33.477	< 2e-16 ***

Standardised Residuals Tests:

	Statistic	p-Value
Jarque-Bera Test R Chi ²	180.9078	0
Shapiro-Wilk Test R W	0.9863903	7.31298e-12
Ljung-Box Test R Q(10)	15.23891	0.1235914
Ljung-Box Test R Q(15)	22.62785	0.09236734
Ljung-Box Test R Q(20)	30.11872	0.06795267
Ljung-Box Test R ² Q(10)	20.44918	0.02527929
Ljung-Box Test R ² Q(15)	28.07832	0.02108388
Ljung-Box Test R ² Q(20)	30.63038	0.060263
LM Arch Test R TR ²	22.87927	0.02876854

Information Criterion Statistics:

AIC	BIC	SIC	HQIC
-4.844245	-4.831824	-4.844255	-4.839655

Tento model je ještě horší než GARCH (1,1). Parametr α_2 je v modelu nevýznamný. AIC kritérium je větší než u předchozího modelu, kvadráty reziduí jsou korelována a ARCH-LM test stále vykazuje podmíněný rozptyl v reziduích.

GARCH(1,2)

Výstup z R: GARCH(1,2)

	Estimate	Std. Error	t value	Pr(> t)
omega	8.708e-06	3.217e-06	2.707	0.00680 **
alpha1	1.987e-01	2.568e-02	7.738	9.99e-15 ***
beta1	3.366e-01	1.187e-01	2.835	0.00458 **
beta2	4.663e-01	1.111e-01	4.198	2.70e-05 ***

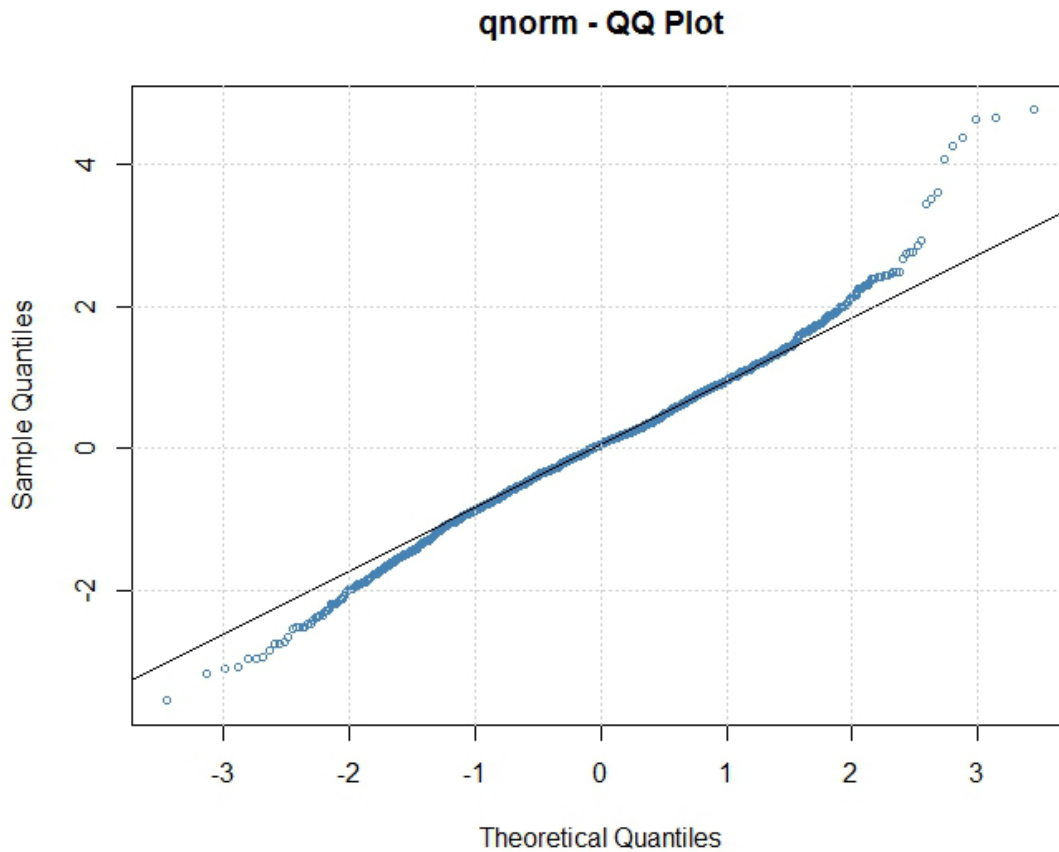
Standardised Residuals Tests:

	Statistic	p-Value
Jarque-Bera Test R Chi ²	169.4747	0
Shapiro-Wilk Test R W	0.9865907	9.466931e-12
Ljung-Box Test R Q(10)	13.99226	0.1733451
Ljung-Box Test R Q(15)	21.68063	0.1164558
Ljung-Box Test R Q(20)	28.84018	0.09095425
Ljung-Box Test R ² Q(10)	10.18614	0.4243171
Ljung-Box Test R ² Q(15)	17.61658	0.2833593
Ljung-Box Test R ² Q(20)	19.49926	0.4896186
LM Arch Test R TR ²	11.87471	0.4557939

Information Criterion Statistics:

AIC	BIC	SIC	HQIC
-4.848063	-4.835641	-4.848073	-4.843473

Tento model GARCH(1,2) budeme považovat za adekvátní a použijeme ho pro predikci. Všechny parametry vychází významné. Diagnostické testy potvrdili nekorelovanost reziduí, stejně tak nebyl prokázán další ARCH efekt v reziduích. Jarque-Bera a Shapiro-Wilk testy prokázaly nenormální rozdělení reziduí. Na obrázku (14) tak můžeme pozorovat odchýlení od normálního rozdělení v podobě těžších konců.



Obr. 14: Rozdělení reziduí GARCH(1,2)

Výsledný model Garch(1,2)

$$h_t = 0,000008708 + 0,1987e^2 + 0,3366h_{t-1} + 0,4663h_{t-2}$$

O tomto modelu můžeme říci, že dopad nových informací $\{e^2\}$ není až tak důležitý a model závisí spíše na minulém rozptylu $\{h_{t-1}\}, \{h_{t-2}\}$.

4.2.2. Předpověď

Předpověď pro 1 krok kupředu

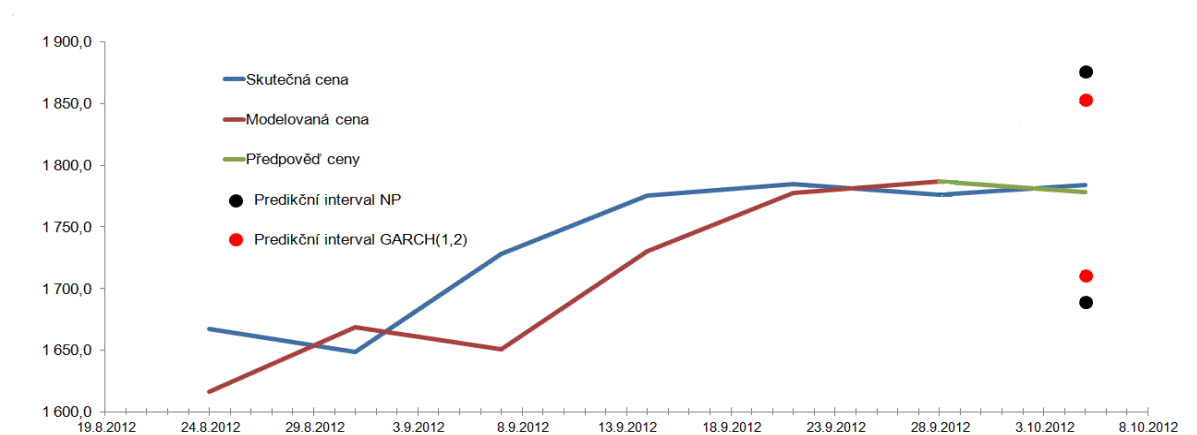
```
Výstup z R:      > predict(Garch12)
                 meanError      meanForecast      standardDeviation
                 0             0.02038191      0.02038191
```

4.3. Porovnání predikčních intervalů

Předpověďenou hodnotu budoucího rozptylu jsem zkopíroval do MS Excel a tam vypočítal a vykreslil porovnání predikčních intervalů našich dvou modelů, obrázek (15). Vidíme, že použitím modelu volatility můžeme zpřesnit předpovědi.

	Náhodná procházka		GARCH(1,2)	
	Horní interval	Dolní interval	Horní interval	Dolní interval
5. 10. 2012	1875,779577	1689,420717	1 852,7176605	1710,449976

Tabulka 1: Porovnání predikčních intervalů



Obr. 15: Porovnání predikčních intervalů

5. Závěr

Původně jsem zamýšlel na cenu zlata aplikovat nejprve model Box-Jenkinsovi metodologie ARIMA a v druhé části tento model použít jako základ modelu volatility pro výpočet bodové předpovědi. Porovnat predikční intervaly těchto dvou modelů a otestovat je. Díky použití modelu GARCH bychom měli získat přesnější predikční intervaly, než s použitím modelu ARIMA.

Nepodařilo se prokázat korelaci v časové řadě výnosů ceny zlata, a tak jsme nemohli použít model ARIMA. Data se zdají naprosto náhodná a nepredikovatelná. Časovou řadu jsem tedy modeloval pomocí procesu náhodné procházky. Vypočítal jsem predikční intervaly pro náhodnou procházku o jeden krok kupředu. Celkový trend řady, daný střední hodnotou diferencí, celkem přesně vystihuje rostoucí průběh ceny zlata.

V modelování podmíněné volatility už jsem byl úspěšnější. Pomocí statistických testů byla prokázána podmíněná heteroskedasticita v týdenních cenách zlata spolu s nenormálním rozdělením. Postupným zkoušením jsem zjistil, že model GARCH(1,2) bude nejvhodnější pro modelování a predikci podmíněného rozptylu této řady. K tomuto výsledku jsem dospěl na základě informačních kritérií a porovnáním jednotlivých statistických testů modelů. Bylo prokázáno, že standardizovaná rezidua modelu GARCH(1,2) už neobsahují žádný další ARCH efekt, stejně tak nebyla vyvrácena hypotéza o nekorelovanosti reziduí. Pokud bych měl k dispozici model ARIMA, mohl jsem vypočítat i bodovou předpověď. Takto si musíme vystačit pouze s predikčním intervalem vypočteným na základě předpovědi podmíněného rozptylu modelu GARCH(1,2).

Vypočtené intervaly jsem porovnal a zjistil, že predikční interval pro GARCH(1,2) je „menší a užší“ než u náhodné procházky. Dostali jsme tak přesnější předpověď, než kdybychom pracovali pouze s modely, předpokládajícími konstantní rozptyl. Výsledky tohoto modelu, bychom tak dále mohli použít pro zpracování celkové analýzy pro prodej či nákup zlata.

Zajímavé by jistě bylo pokračovat v modelování volatility ceny zlata dalšími a složitějšími modely, které postupem času vznikly. Těchto modelů existuje nepřeberné množství, například: IGARCH, Taylor/Schwert, A-GARCH, NA-GARCH, V-GARCH,

Thr-GARCH, GJR-GARCH, log-GARCH, ANST-GARCH, GARJI atd. Zajímavé by jistě byly i vícerozměrné modely časových řad nebo zkoumání korelací a vzájemných vlivů souvisejících řad (cena stříbra, velikost měnových agregátů atd.)

6. Summary

Keywords: Mathematical modelling, Box-Jenkins, volatility model, GARCH, gold price, random-walk process, R-Project.

This bachelor thesis is focused on mathematical modelling of gold price. After introducing basic models of Box-Jenkins methodology and models of conditional heteroskedasticity, the time series was modelled by random walk process. The time series has not constant variance, neither normal distribution, therefore for the second model was considered volatility model GARCH. Using the volatility model GARCH is possible to refine prediction intervals for forecast one step ahead. Results were calculated in R software and MS Excel.

7. Zdroje

- [1] ARLT, Josef. Moderní metody modelování ekonomických časových řad. 1.vyd. Praha: Grada Publishing, 1999, 307 s. ISBN 80-716-9539-4.
- [2] CIPRA, Tomáš. Analýza časových řad s aplikacemi v ekonomii. Praha: SNTL, 1986.
- [3] LIŠČINSKÝ, Zdeněk. Jednorozměrné ARCH modely a jejich vybrané ekonomické aplikace. Brno, 2008. Dostupné z: http://is.muni.cz/th/63335/prif_m/arch_modely.pdf. Diplomová práce. MASARYKOVA UNIVERZITA. Vedoucí práce Vítězslav Veselý.
- [4] MOTGOMERY, Douglas C., Cheryl L. JENNINGS a Maurat KULAHCI. Introduction to Time Series Analysis and Forecasting. Hoboken, New Jersey: John Wiley, 2008. ISBN 9780471653974.
- [5] ŠTĚRBA, Filip. Modely typu ARCH a jejich využití k modelování volatility měnových kurzu. [online]. 2007 [cit. 2013-04-15]. Dostupné z: <http://panda.hyperlink.cz/cestapdf/pdf07c3/sterba.pdf>
- [6] TSAY, Ruey S. Analysis of financial time series. New York: Wiley, 2002, xii, 448 p. ISBN 04-714-1544-8.
- [7] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- [8] Wuertz Diethelm, Chalabi Yohan with contribution from Michal Miklovic, Chris Boudt, Pierre Chausse and others (2012). fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling. R package version 2150.81. <http://CRAN.R-project.org/package=fGarch>

8. Seznam obrázků

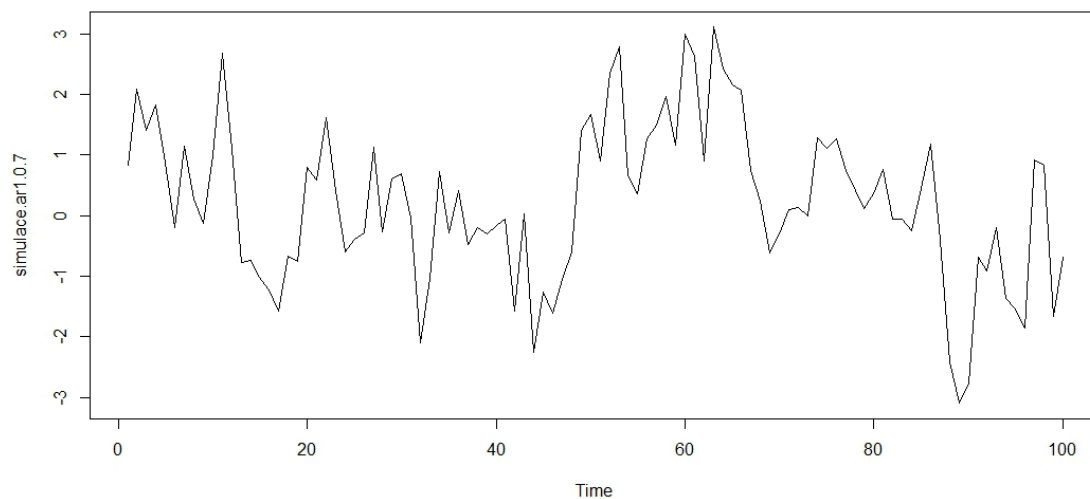
Obr. 1: Nestacionární řada	6
Obr. 2: Stacionární řada	7
Obr. 3: PACF řady simulované procesem	10
Obr. 4: PACF řady simulované procesem	10
Obr. 5: Časová řada ceny zlata od 29. 12. 1978 do 28. 9. 2012.	16
Obr. 6: ACF ceny zlata	17
Obr. 7: PACF ceny zlata	18
Obr. 8: Logaritmovaná cena zlata a její první diference.....	18
Obr. 9: ACF diferencí logaritmované ceny zlata	19
Obr. 10: PACF diferencí logaritmované ceny zlata.....	19
Obr. 11: Trend řady	21
Obr. 12: Histogram logaritmů výnosů	23
Obr. 13: Podmíněný rozptyl diferencí logaritmované ceny zlata	24
Obr. 14: Rozdělení reziduí GARCH(1,2)	29
Obr. 15: Porovnání predikčních intervalů.....	30
Obr. 16: Simulovaná řada AR(1)	37
Obr. 17: ACF simulované řady AR(1)	37
Obr. 18: Simulovaná řada AR(2)	38
Obr. 19: ACF simulované řady	38
Obr. 20: Simulovaná řada MA(1)	39
Obr. 21: PACF simulované řady MA(1)	39
Obr. 22: ACF simulované řady	40

9. Seznam tabulek

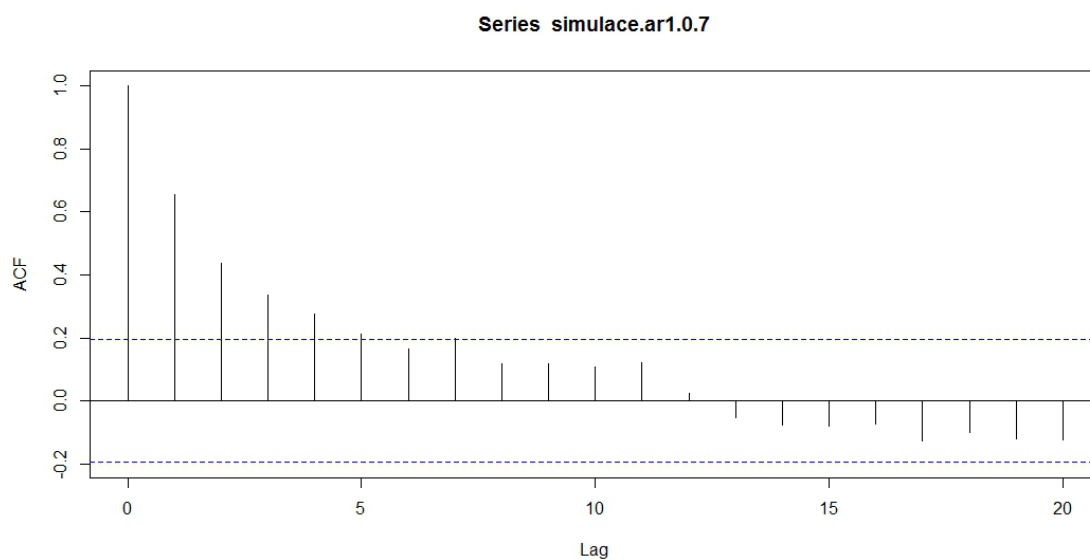
Tabulka 1: Porovnání predikčních intervalů.....	30
---	----

10. Přílohy

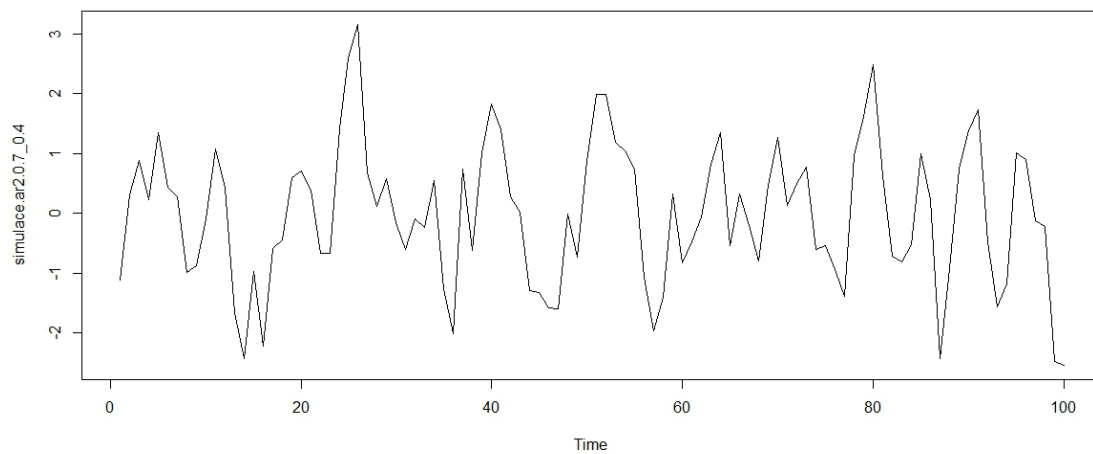
Tištěné



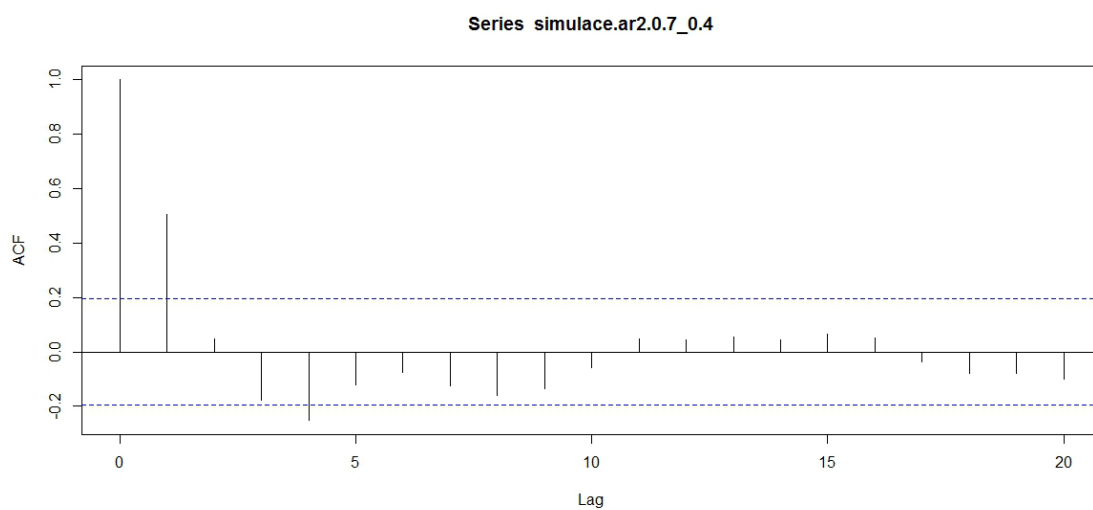
Obr. 16: Simulovaná řada AR(1) $y_t = 0,7y_{t-1} + \varepsilon_t$



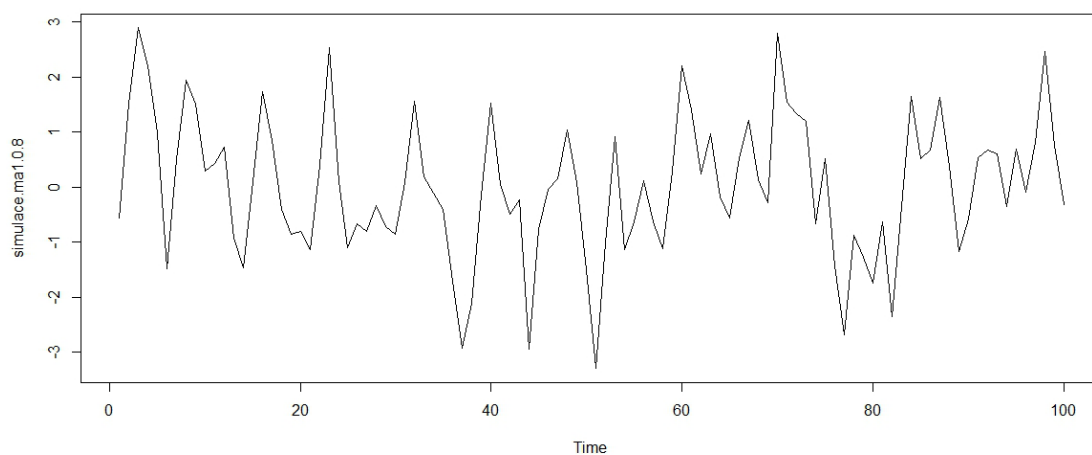
Obr. 17: ACF simulované řady AR(1) $y_t = 0,7y_{t-1} + \varepsilon_t$



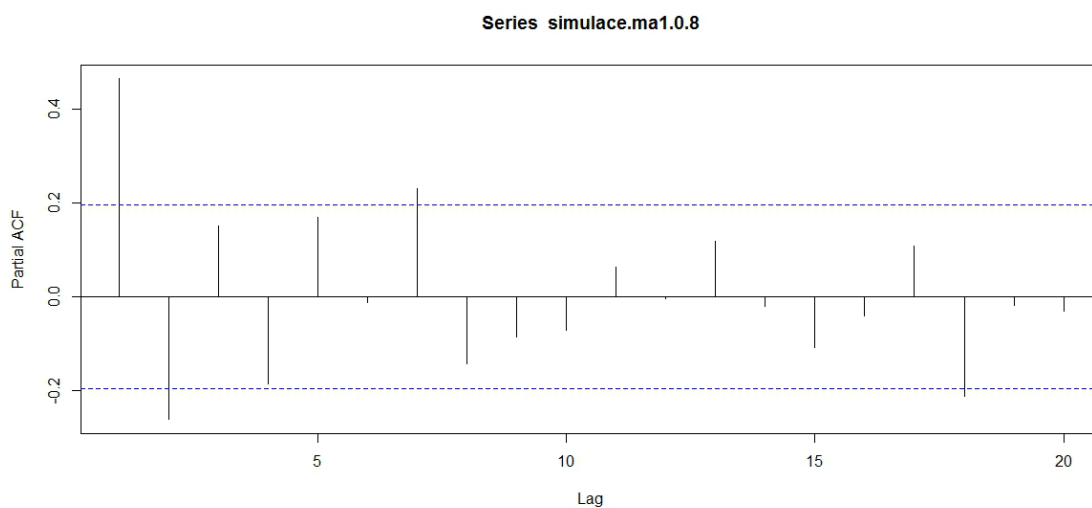
Obr. 18: Simulovaná řada AR(2) $y_t = 0,7y_{t-1} - 0,4y_{t-2} + \varepsilon_t$



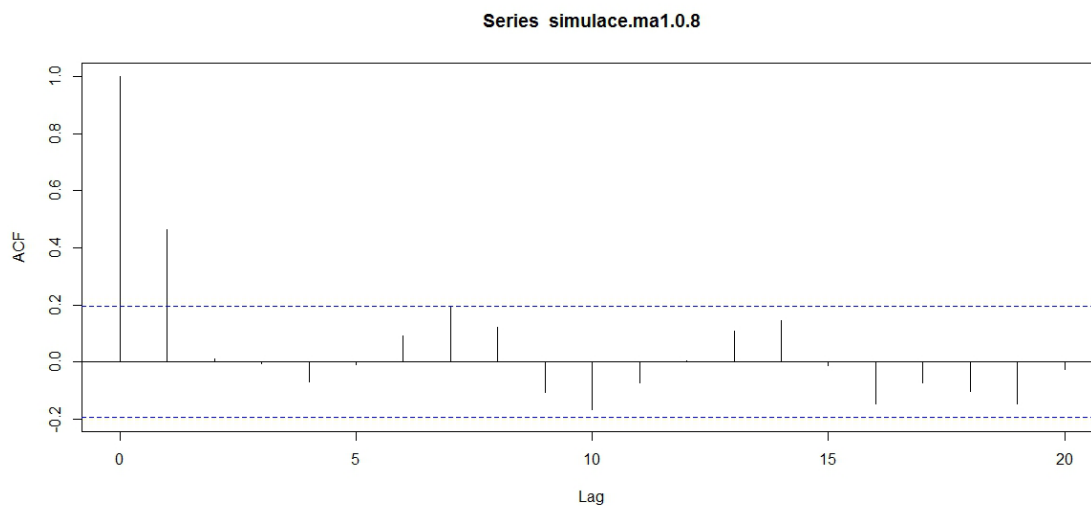
Obr. 19: ACF simulované řady $y_t = 0,7y_{t-1} - 0,4y_{t-2} + \varepsilon_t$



Obr. 20: Simulovaná řada MA(1) $y_t = \varepsilon_t - 0,8\varepsilon_{t-1}$



Obr. 21: PACF simulované řady MA(1) $y_t = \varepsilon_t - 0,8\varepsilon_{t-1}$



Obr. 22: ACF simulované řady $y_t = \varepsilon_t - 0,8\varepsilon_{t-1}$

Obsažené na CD

kod_R.txt – Některé příkazy používané při práci v programu R

zlato.xls – Soubor MS Excel s některými výpočty a časovou řadou ceny zlata

BP_Smarda.pdf – Soubor s bakalářskou prací v elektronické podobě