

ZÁPADOČESKÁ UNIVERZITA V PLZNI
Fakulta aplikovaných věd
Katedra matematiky

DIPLOMOVÁ PRÁCE

Předpovědi vysokofrekvenčních časových řad pomocí modelů typu ARCH
a metody Monte Carlo

Plzeň 2013

Vypracoval: Bc. Tomáš Pakosta
Vedoucí práce: RNDr. Blanka Šedivá, Ph.D.

Čestné prohlášení

Prohlašuji, že jsem zadanou diplomovou práci vypracoval samostatně s přispěním vedoucího práce. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

Datum: 20. května 2013

.....
podpis

Poděkování

Rád bych na tomto místě poděkoval všem, kteří mi s prací přímo i nepřímo pomohli, protože bez nich by tato práce nevznikla. Především bych chtěl poděkovat vedoucí práce RNDr. Blance Šedivé, Ph.D. za vedení, pomoc a odborné rady při zpracování této práce.

!!! Originál zadání !!!

Abstrakt

Hlavním cílem této práce je získat model typu ARCH pro zvolená a popsaná vysokofrekvenční data, konkrétně data časové řady ceny zlata. Dalším cílem je získat za pomoci vybraného a odhadnutého modelu typu ARCH předpověď pro danou časovou řadu zkonstruovanou pomocí metody Monte Carlo. Metoda Monte Carlo je v této práci také popsána a hlavně její aplikace pro modely typu ARCH. Obsahem práce je také zpracovaný popis zvoleného software Project R, který poslouží jako dobrý nástroj pro zpracování časové řady z pohledu modelování a předpovídání. Project R lze také využít k velmi dobré interpretaci výsledků pomocí programovatelných grafů. V závěrečné části této práce je zpracováno porovnání předpovědí se skutečnými hodnotami v odhadovaném období.

Klíčová slova: časová řada, volatilita, ARCH, GARCH, Project R, Monte Carlo

Abstract

The main goal of this thesis is to obtain a model ARCH type for chosen and described high-frequency data, namely data time series prices of gold. The next target is to obtain forecast of the time series constructed using Monte Carlo methods with the help of selected estimated ARCH type model. Monte Carlo method is also described in this thesis and especially application to ARCH type models. The thesis also include description of the selected software Project R which will serve as a good tool for processing time series from the perspective of modeling and forecasting. Project R would be used for a very good interpretation of the results using programmable graphs. In the final section of this thesis is processed comparing between predictions with the actual values in the estimated period.

Keywords: time series, volatility, ARCH, GARCH, Project R, Monte Carlo

Obsah

1 Úvod	1
2 Časové řady typu ARCH	2
2.1 Finanční časové řady	6
2.2 Charakteristické rysy chování finančních časových řad	6
2.2.1 Předpoklad normality	8
2.2.2 Předpoklad lineární závislosti	10
3 Modely volatility	11
3.1 Základní reprezentace	11
3.2 Lineární modely volatility	12
3.2.1 Modely ARCH (Autoregressive Conditional Heteroscedasticity)	12
3.2.2 Modely GARCH (Generalized ARCH)	14
3.3 Výstavba modelů volatility	15
3.4 Testování podmíněné heteroskedasticity	15
3.5 Volba řádu modelu	16
3.6 Odhad parametrů	17
3.7 Verifikace modelu	18
3.7.1 Kritéria hodnocení modelu	18
3.8 Konstrukce předpovědí	19
4 Monte Carlo	21
4.1 Předpovědi modelů GARCH pomocí MC	21
5 Project R	23
5.1 Počáteční nastavení a načtení dat	23
5.2 Použité nástroje a funkce	25
6 Aplikační část	28
6.1 Základní popis a předzpracování dat	29
6.2 Předpoklad normality	35
6.3 Fáze 1: Odhad úrovněvého modelu	37
6.3.1 Test residuí	40
6.4 Fáze 2: Odhad modelu volatility	42
6.4.1 Test residuí a úprava modelu	44
6.5 Celkový konečný model	47
6.6 Monte Carlo předpovědi	49
6.6.1 Porovnání se skutečností	53
7 Závěr	58
Literatura	59
A Přílohy obsažené na CD	60
B Přílohy tištěné	61

Seznam obrázků

1	Vývoj časové řady logaritmů výnosů ceny zlata	5
2	Simulace modelu ARCH(2)	13
3	Vývoj časové řady zlata za období 9.2.2005-31.10.2012	28
4	Vývoj časové řady zlata a odhad exponenciálního trendu	29
5	Graf ACF pro řadu cen zlata	30
6	Graf PACF pro řadu cen zlata	31
7	Vývoj logaritmů cen zlata a odhad lineárního trendu	32
8	Vývoj časové řady logaritmů výnosů ceny zlata	33
9	Graf ACF logaritmů výnosů časové řady ceny zlata	34
10	Graf PACF logaritmů výnosů časové řady ceny zlata	34
11	caption	35
12	Q-Q graf výnosů	36
13	Řada výnosů a odhad modelu ARMA(1,0)	39
14	Histogram residuí, gaussovský jádrový odhad hustoty rozdělení a teoretické normální rozdělení	40
15	ACF residuí ARMA(1,0) modelu	41
16	PACF residuí ARMA(1,0) modelu	41
17	ACF kvadrátů residuí ARMA(1,0) modelu	42
18	PACF kvadrátů residuí ARMA(1,0) modelu	43
19	Histogram residuí modelu GARCH(1,1), gaussovský jádrový odhad hustoty rozdělení a teoretické normální rozdělení	45
20	Histogram residuí $\{\hat{\varepsilon}_t\}$ modelu GARCH(1,1), gaussovský jádrový odhad hustoty rozdělení a teoretické Studentovo t -rozdělení	47
21	Podmíněný rozptyl odhadnutého modelu GARCH(1,1) a nepodmíněný rozptyl	48
22	Simulované předpovědi podmíněného rozptylu \hat{h}_t na 20 dní dopředu pomocí metody MC	50
23	Simulované předpovědi na 20 dní pro residua $\hat{\varepsilon}_t$ modelu ARMA(1,0) pomocí metody MC	50
24	Simulované hodnoty residuí ARMA modelu a intervaly empirické kvantilové funkce.	52
25	Residua ARMA modelu a empirické intervaly spolehlivosti metody Monte Carlo.	53
26	Residua ARMA modelu a 95% intervaly spolehlivosti pomocí příkazu <i>predict</i>	53
27	Graf odhadů intervalů metody MC (červeně) a pomocí příkazu <i>predict</i> (modře)	55
28	Graf odchylek mezi horními hranicemi intervalových pásů daných metod	56
29	Graf odchylek mezi dolními hranicemi intervalových pásů daných metod	56
30	Graf historické střední hodnoty logaritmů výnosů časové řady ceny zlata	61
31	Graf historického rozptylu hodnoty logaritmů výnosů časové řady ceny zlata	61

Seznam tabulek

1	Základní údaje časové řady cen zlata	29
2	Statistické charakteristiky řady výnosů	35
3	Výsledky testů normality pro řadu výnosů	35
4	Tabulka kritérií pro různé modely typu ARMA	38
5	Tabulka odhadů parametrů pro ARMA(1,0) model	38
6	Výsledky testů normality pro řadu $\{\varepsilon_t\}$	40
7	Hodnoty maximalizované věrohodnostní funkce, AIC a BIC modelů GARCH	43
8	Odhadnuté parametry modelu GARCH(2,1)	43
9	Odhadnuté parametry modelu GARCH(1,1)	44
10	Testy standardizovaných residuí $\{\hat{\varepsilon}\}$ modelu GARCH(1,1)	44
11	Odhadnuté parametry modelu GARCH(1,1) s předpokladem Studentova t -rozdělení pro residua $\{\hat{\varepsilon}\}$	45
12	Testy standardizovaných residuí $\{\hat{\varepsilon}\}$ modelu GARCH(1,1) s předpokladem Studentova t -rozdělení	46

1 Úvod

S pojmem „časová řada“ je svázána spousta teorie, dat, modelů či aplikací. Z tohoto také vyplývá mnoho různých názorů či přístupů, jak se postavit ke zpracování časových řad. Velký zájem, zejména v ekonomii, je kladen na zpracování krátkodobých časových řad, tj. řad, jejichž frekvence hodnot je v obdobích kratších než 1 rok. Zvláštní postavení, v této skupině časových řad, pak mají denní časové řady, u kterých bylo empirickým výzkumem zjištěno, že právě tyto časové řady se vyznačují v čase proměnlivou variabilitou (tzv. volatilita). Tato vlastnost však porušuje předpoklady pro použití klasických lineárních modelů. Počátkem 80. let přišel Robert Engle s novou koncepcí modelování volatility, kde zavádí modely typu ARCH neboli autoregresní modely s podmíněnou heteroskedasticitou (Autoregressive Conditional Heteroscedasticity). Koncepce modelů typu ARCH se postupně rozvíjela v další modely tohoto typu tak, aby co nejvíce reflektovaly různá typická chování určitých druhů vysokofrekvenčních časových řad. Další vývoj modelů typu ARCH pak vychází z myšlenky modifikace předpokladu normality, na kterém je založena většina modelů časových řad.

Cílem této diplomové práce je zpracovat finanční časovou řadu ceny zlata za období 9.2.2005 - 31.10.2012 a určit její předpověď. Dosažením tohoto cíle se předpokládá užití modelů podmíněné heteroskedasticity a užití softwaru Project R, pomocí kterého budou získány všechny výsledky a grafické výstupy. Pro určení charakteru předpovědi časové řady cen zlata bude využita metoda Monte Carlo a na konci práce bude popsáno zhodnocení použití metody Monte Carlo pro předpovídání dalšího vývoje časové řady.

V první části diplomové práce je popsána potřebná teorie, která je využita v aplikační části. Jedná se o popis finančních časových řad, jejich vlastností, předpokladů a charakteristických rysů chování těchto řad. V druhé části jsou popsány samotné modely volatility použité pro zpracování podmíněné heteroskedasticity v časové řadě. K těmto modelům je také popsána jejich výstavba, volba řádu, odhad parametrů, verifikace modelu a předpovědi. Další částí je pak popis metody Monte Carlo, která je dále využita k odhadu předpovědi časové řady ceny zlata. Před zpracováním časové řady ceny zlata je nezbytný popis programu, ve kterém je aplikační část zpracována a tím je Project R. Samotná aplikační část je poslední částí této práce, ve které je vyšetřen průběh časové řady za pomoci popsané teorie v předcházejících částech. Odhadnutý model časové řady cen zlata je ověřen a použit k předpovědím.

Závěrem diplomové práce bude zhodnocení předpovědi pro danou časovou řadu pomocí metody Monte Carlo. Také bude stručně popsáno možné rozšíření této práce.

2 Časové řady typu ARCH

Ekonomické časové řady můžeme popsat jako empirická pozorování v oblasti ekonomiky, která jsou uspořádána v čase, do řady hodnot, směrem od minulosti do přítomnosti.

Pokud ekonomické časové řady budeme rozlišovat z pohledu délky intervalu sledovaných hodnot, zjistíme, že spadají do dvou skupin a těmi jsou dlouhodobé časové řady, které mají sledované hodnoty v ročních nebo delších časových intervalech. Naopak krátkodobé časové řady mají interval sledovanosti kratší než jeden rok. S tímto rozdělením souvisí také tvar ekonomických časových řad. Tvar časové řady bude vyhlazenější, čím delší interval sledovaných hodnot budeme mít. To vyplývá z poznatku, že časové řady mají časově svázané jednotlivé hodnoty. Kdybychom časovou návaznost neuvažovali, pojem časové řady by zcela ztratil smysl a neurčili bychom jejich tvar a charakteristické vlastnosti.

Charakteristickými vlastnostmi ekonomických časových řad jsou: trend, sezónnost, nelinearita, podmíněná heteroskedasticita a společné vlastnosti více časových řad jako je například tzv. společný trend. Tyto vlastnosti se zpravidla nevyskytují ve všech časových řadách současně. Výskyt některé z vlastností závisí na typu časové řady, např. heteroskedasticita, která zde bude později zkoumána se vyskytuje u vysokofrekvenčních řad (spadají pod krátkodobé časové řady), kterými jsou právě například finanční časové řady.

Pro potřeby teoretických modelů lze časovou řadu chápat jako speciální typ náhodného procesu.

Stochastický proces

Mějme v čase uspořádanou řadu náhodných veličin $\{X(s, t), s \in \mathbf{S}, t \in \mathbf{T}\}$, kde \mathbf{S} je výběrový prostor a \mathbf{T} je indexní řada, pak ji nazveme *stochastickým procesem*, kde pro každé $t \in \mathbf{T}$ je $X(\cdot, t)$ náhodná veličina definovaná na výběrovém prostoru \mathbf{S} a pro každé $s \in \mathbf{S}$ je $X(s, \cdot)$ realizace stochastického procesu definovaná na indexní řadě \mathbf{T} (tj. uspořádaná řada čísel, z níž každé odpovídá jedné hodnotě indexní řady). Budeme předpokládat, že indexní řada je řadou celých čísel, tj. $\mathbf{T} = \{0, \pm 1, \pm 2, \pm 3, \dots\}$ a že uvažované stochastické procesy mají nespojitý charakter. Jelikož časovou řadu lze chápat jako realizaci stochastického procesu, budou se dále stochastické procesy označovat jako $\{X_t, t = 0, \pm 1, \pm 2, \pm 3, \dots\}$, resp. $\{X_t\}$ a jejich realizace, tj. časové řady jako X_t [1].

Časová řada je tedy systémem náhodných veličin $\{X_1, X_2, \dots\}$ (resp. $\{X_1, X_2, \dots, X_T\}$, $t = 1, 2, \dots, T$) a $\{x_1, x_2, \dots\}$ (resp. $\{x_1, x_2, \dots, x_T\}$, $t = 1, 2, \dots, T$) jsou jednotlivé realizace časové řady.

Časovou řadu X_t nazveme *bílým šumem*, pokud X_t je posloupnost nekorelovaných stejně rozdělených náhodných veličin s nulovou střední hodnotou a s konečným rozptylem σ^2 . Bílý šum¹ označíme $WN(0, \sigma^2)$. Má-li X_t navíc normální rozdělení $N(0, \sigma^2)$, nazývá se X_t *Gaussovský bílý šum*.

Stacionarita

Striktně stacionárním procesem nazveme takový stochastický proces, jestliže pro jakoukoliv konečnou indexní část t_1, t_2, \dots, t_n z $\mathbf{T} = 0, \pm 1, \pm 2, \pm 3, \dots$ a jakékoliv reálné číslo k , pro které $t_i + k \in \mathbf{T}$, $i = 1, 2, \dots, n$, platí:

$$F(X_{t_1}, X_{t_2}, \dots, X_{t_n}) = F(X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}),$$

kde $F(\cdot)$ je sdružená distribuční funkce. Striktní (silná, kompletní) stacionarita tedy požaduje, aby libovolná skupina náhodných veličin z časové řady měla v případě libovolného časového posunu stejné rozdělení jako neposunutá [1].

¹White Noise.

Časovou řadu $\{X_t : t \in \mathbf{T}\}$ pak nazveme slabě stacionární, jestliže platí:

- (i) $EX_t^2 < \infty$,
- (ii) $EX_t = \mu$ pro všechna $t \in \mathbf{T}$ a nějaké $\mu \in \mathbb{R}$,
- (iii) $cov(X_p, X_r) = \gamma(|p - r|)$ pro všechna $p, r \in \mathbf{T}$.

Bod (iii) znamená, že autokovarianční funkce X_t závisí pouze na vzdálenosti indexů p a r . U stacionárních časových řad proto autokovarianční funkci píšeme jen s jedním argumentem, který udává tuto vzdálenost: $\gamma(p, r) \stackrel{\text{ozn.}}{=} \gamma(l)$ pro $l = |p - r|$.

Striktně stacionární proces, který má první dva obecné momenty konečné, je také slabě stacionární proces. Stochastický proces se označuje jako normální nebo gaussovský, jestliže jeho sdružené pravděpodobnostní rozdělení je normální. Striktní a slabá stacionarita jsou pak ekvivalentními pojmy, neboť normální rozdělení je jednoznačně charakterizováno prvními dvěma obecnými momenty. V dalších částech se zabýváme gaussovskými stochastickými procesy, protože většina zde uvedených výsledků z těchto procesů vychází [1].

Autokorelační funkce (ACF)

Pro stacionární stochastický proces X_t lze vyjádřit autokovarianční funkci mezi veličinami X_t a X_{t-k} takto:

$$\gamma_k = cov(X_t, X_{t-k}) = E[(X_t - \mu)(X_{t-k} - \mu)], \quad k = 0, 1, 2, \dots, T - 1.$$

Autokorelační funkci (ACF) pak jako:

$$\rho_k = \frac{cov(X_t, X_{t-k})}{\sqrt{D(X_t)}\sqrt{D(X_{t-k})}} = \frac{\gamma_k}{\gamma_0},$$

kde vzhledem ke stacionaritě procesu $var(X_t) = X_{t-k} = \gamma_0$. V případě stacionárního stochastického procesu má autokorelační funkce následující vlastnosti:

- (i) $\rho_0 = 1$,
- (ii) $|\gamma_k| \leq \gamma_0; |\rho_k| \leq 1$ pro všechna k ,
- (iii) $\gamma_k = \gamma_{-k}$ a $\rho_k = \rho_{-k}$ pro všechna k .

Autokovarianční a autokorelační funkce je tedy symetrická kolem $k = 0$. Tato vlastnost vyplývá ze skutečnosti, že časová vzdálenost dvojic veličin X_t a X_{t+k} , X_{t-k} a X_t je stejná. Z tohoto důvodu a z důvodu (i) je autokorelační funkce vyjadřována pouze pro $k > 0$. Graf autokorelační funkce se nazývá korelogram.

Výběrová autokorelační funkce

Ve většině případů jsou parametry μ, γ_0 a ρ_k neznámé. Za předpokladu stacionarity může být střední hodnota procesu μ odhadována prostřednictvím výběrového průměru a rozptyl procesu γ_0 může být odhadován pomocí výběrového rozptylu:

$$\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t, \quad s^2 = \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2,$$

kde T je počet hodnot časové řady. Odhad ρ_k je dán výběrovou autokorelací se zpožděním k :

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^T (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2}, \quad k = 1, 2, \dots, T - 1.$$

Parciální autokorelační funkce (PACF)

Korelace mezi veličinami X_t a X_{t-k} může být způsobena jejich korelací s veličinami $X_{t-1}, X_{t-2}, \dots, X_{t-k+1}$. Parciální autokorelace podávají informaci o korelaci veličin X_t a X_{t-k} očištěnou o vliv veličin ležících mezi nimi. Parciální korelaci se zpožděním k vyjadřuje parciální regresní koeficient ϕ_{kk} v autoregresi k -tého řádu:

$$X_t = \phi_{k1}X_{t-1} + \phi_{k2}X_{t-2} + \dots + \phi_{kk}X_{t-k} + e_t,$$

kde veličina e_t je nekorelovaná s veličinami X_{t-j} , $j \geq 1$. Vynásobíme-li obě strany předchozí rovnice X_{t-j} , potom střední hodnota této rovnice má tvar:

$$\gamma_j = \phi_{k1}\gamma_{j-1} + \phi_{k2}\gamma_{j-2} + \dots + \phi_{kk}\gamma_{j-k}$$

takže platí:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{j-k}$$

a pokud dosadíme za $j = 1, 2, \dots, k$, získáme systém rovnic. Dále pak postupně dosazujeme za $k = 1, 2, \dots$ a použitím Cramerova pravidla získáme $\phi_{11}, \phi_{22}, \dots, \phi_{kk}$. Například ϕ_{kk} vypadá takto:

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

Výběrová parciální autokorelační funkce

Nahrazením ρ_i odhadem $\hat{\rho}_i$ v předcházejícím vztahu pro ϕ_{kk} získáme výběrovou parciální funkci $\hat{\phi}_{kk}$. Namísto komplikovaných počítání determinantů navrhl Durbin (1960) rekurzivní vztah [1]:

$$\hat{\phi}_{kk} = \frac{\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1j} \hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1j} \hat{\rho}_j}$$

$$\hat{\phi}_{kj} = \hat{\phi}_{k-1j} - \hat{\phi}_{kk} \hat{\phi}_{k-1k-j}, \quad j = 1, 2, \dots, k-1.$$

Výběrová ACF a výběrová PACF tedy slouží jako odhad pro ACF a PACF. Podle [3] lze ukázat, že za určitých předpokladů má rozdíl $\hat{\rho}_k - \rho_k$ asymptoticky normální rozdělení. Toho je samozřejmě možné využít k testování hypotézy $H_0 : \rho_k = 0$ pro konkrétní $k \geq 1$. Pokud data odpovídají realizaci Gaussovského bílého šumu, pak $\hat{\rho}_k$ leží s pravděpodobností 95% ($\alpha = 5\%$) v intervalu $(u_{\frac{\alpha}{2}}/\sqrt{n}, u_{1-\frac{\alpha}{2}}/\sqrt{n})$, kde u_p je kvantil normálního normovaného rozdělení. Tento interval lze nazvat intervalem spolehlivosti pro výběrové ACF a PACF. Leží-li koeficienty ACF a PACF v tomto intervalu, pak je lze považovat za nulové.

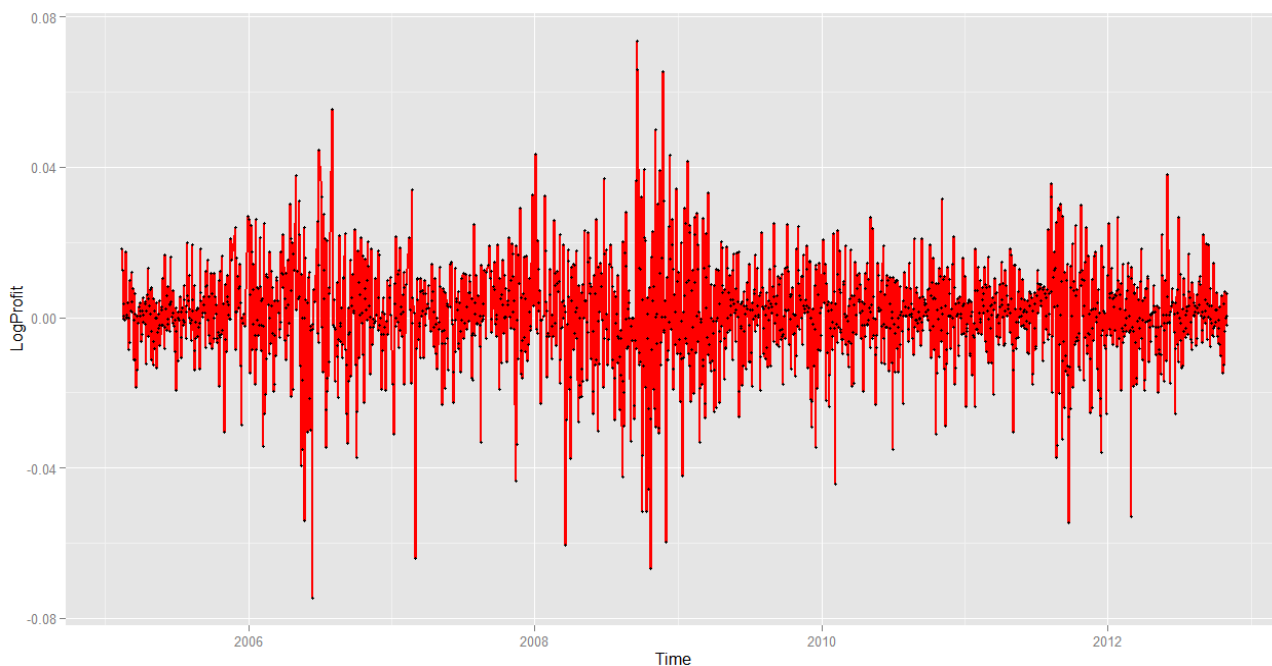
Podmíněná heteroskedasticita

V analýzách zabývajících se finančními časovými řadami, se často vychází z jednoho ze základních předpokladů, tj. že logaritmy tzv. výnosů (koeficienty růstu) mají normální rozdělení s konstantní střední hodnotou a konstantním rozptylem v čase. Logaritmování se provádí z toho důvodu, že ceny nemohou být záporné a předpokládá se tedy jejich logaritmicko-normální rozdělení.

Ze skutečnosti vyplývá, že časové řady logarimů výnosů mají rozdělení špičatější s „tlustšími“ konci než je normální rozdělení. Tato vlastnost může být způsobena charakteristickým rysem jejich chování, resp. chováním logaritmů jejich koeficientů růstu. Variabilita (volatilita) těchto časových řad se v průběhu času mění, období s vysokou variabilitou střídají období s variabilitou nižší. To vyplývá z rostoucí a klesající nejistoty na trhu. Můžeme si představit, že logarimus výnosu má normální rozdělení s rozptylem, který se v čase mění. Tuto vlastnost nazveme *podmíněnou heteroskedasticitou*.

Nepodmíněné rozdělení logaritmů výnosů je pak směsicí normálních rozdělení, kde ta s malým podmíněným rozptylem koncentrují výnosy v blízkosti střední hodnoty a naopak ta s velkým podmíněným rozptylem posouvají výnosy do konců rozdělení. Výstupem je nepodmíněné špičaté rozdělení s „tlustými“ konci.

Například z obrázku logaritmů výnosů (1), který je uveden také v kapitole 6.1, je patrné, že vysoká volatilita přechází do nízké volatility postupně a stejně tak naopak. To znamená, že volatilita určité úrovně se udržuje po určitý čas, volatilita v čase t tedy závisí na volatilitě v čase $t - 1$.



Obr. 1: Vývoj časové řady logaritmů výnosů ceny zlata

Podmíněnou heteroskedasticitu můžeme vyjádřit jako:

$$(\ln X_t - \ln X_{t-1})^2 = \alpha + \rho(\ln X_{t-1} - \ln X_{t-2})^2 + \varepsilon_t. \quad (1)$$

Pokud je parametr $\rho = 0$, pak podmíněná heteroskedasticita v časové řadě není. Naopak pokud $\rho \neq 0$ a zároveň $|\rho| \leq 1$, pak podmíněná heteroskedasticita v časové řadě existuje. Čím je parametr $|\rho|$ blíže k jednotce, tím významnější je autokorelace kvadrátů logaritmů [1].

2.1 Finanční časové řady

Finanční časové řady patří mezi vysokofrekvenční časové řady, které spadají pod krátkodobé časové řady, jejichž frekvence sledování je kratší než jeden rok.

Sledovaným prvkem finančních časových řad je základní informace finančních trhů, kterou je cena. Může to být například cena akcie, cena měny, cena dluhopisu. Základní finanční časové řady vycházejí tedy z cen dosahovaných na veřejných trzích nebo charakterizují ceny a jejich vývoj.

Základním rysem finančních časových řad je vysoká frekvence zaznamenávaných hodnot. Tyto hodnoty jsou zaznamenávány nejčastěji v měsíční, týdenní nebo denní frekvenci, ale mohou být zaznamenávány například na burzách i v hodinových nebo minutových intervalech. Vliv na dynamiku takto rychlé frekvence záznamu dat mají jak systematické faktory (výrazně se projevuje trendová a cyklická složka), tak nesystematické faktory, což má za následek jejich vysokou a proměnlivou variabilitu.

Z následující kapitoly vyplyne, že pro investora poskytuje výnos stejnou informaci jako samotná cena aktiva a navíc jej lze interpretovat v procentech, což může být jasnějším ukazatelem výhodnosti dané investice. Proto se při zpracování finančních časových řad většinou namísto cen aktiv pracuje s jejich výnosy.

2.2 Charakteristické rysy chování finančních časových řad

Základní a primární hypotézou o chování finančního trhu je hypotéza efektivního trhu². V literatuře lze nalézt například formulaci efektivního trhu takovou, že pokud ceny plně zahrnují očekávání a informace všech účastníků trhu, jsou jejich změny nepredikovatelné.

S hypotézou efektivního trhu souvisí zřejmě nejstarší model „chování“ cen aktiv. Tento model je známý pod jménem martingál a lze ho popsat například takto: *Jestliže P_t je cena aktiva v čase t , potom očekávaná cena v čase $t + 1$ je cena v čase t , za podmínky znalosti všech cen aktiva v minulosti, tj. v čase $t - 1, t - 2, \dots$* Z hlediska tvorby předpovědí model martingálu implikuje, že nejlepší (podle minimální střední čtvercové chyby) předpovědi zítřejší ceny je cena dnešní [1].

Představme si časovou řadu cen aktiva jako realizaci stochastického procesu $\{P_t\}$, tzn. řadu náhodných veličin uspořádaných v čase, pak martingál vyjádříme jako:

$$E[P_{t+1}|P_t, P_{t-1}, \dots] = P_t, \quad (2)$$

nebo také jako:

$$E[P_{t+1} - P_t|P_t, P_{t-1}, \dots] = 0, \quad (3)$$

kde $E[.]$ je podmíněnou střední hodnotou.

Dále model martingálu předpokládá, že nepřekrývající se cenové změny ve všech časových posunech (vpřed i vzad) nejsou korelované, tj. jsou lineárně nezávislé. Avšak mohou být stále nelineárně závislé. Tato závislost se může projevit tak, že cenové změny nejsou v čase konstantní. Dříve se předpokládalo, že martingál je nutnou podmínkou efektivního trhu. V současnosti se v souvislosti s podmínkou efektivního trhu hovoří nejen o nepredikovatelnosti, ale také o vztahu očekávané cenové změny a rizika (rozptylu). Z hlediska investora může jistě být zajímavé, že očekávaná cenová změna je kladná, i když je zatížená vysokým rizikem. Ale naopak z hlediska trhu je budoucí cena stále prakticky nepredikovatelná, a jedná se tedy o efektivní trh. Martingál klade restrikcii

²První formulace Bachelier (1900), Cowles (1933).

pouze na očekávanou cenovou změnu nikoliv na riziko, nemůže proto být nutnou podmínkou efektivního trhu, jelikož mohou nastat situace, kdy lze trh stále považovat za efektivní, ale nejedná se již o model martingálu.

Martingál můžeme také vyjádřit vztahem:

$$P_t = P_{t-1} + a_t, \quad (4)$$

kde a_t se označuje jako přírůstek martingálu nebo také jako diference martingálu. Tato forma zápisu je podobná modelu náhodné procházky.

Zde se na rozdíl od martingálu předpokládá, že $\{a_t\}$ je proces bílého šumu³. V tomto procesu jsou náhodné veličiny nejen nekorelované, ale také stejně rozdělené s nulovou střední hodnotou a konstantním rozptylem. Můžeme vycházet z představy, že rozdělení, které mají tyto náhodné veličiny, je normální, tj. $a_t \sim N(0, \sigma_a^2)$.

Toto je sice zřejmé, ale je zde zásadní nedostatek. Cena aktiva nemůže být menší než nula, minimální dosažitelný relativní přírůstek ceny neboli minimální dosažitelný jednoduchý výnos aktiva je tedy:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = -1. \quad (5)$$

Protože náhodná veličina mající normální rozdělení může nabývat jakéhokoliv reálného čísla a ze vztahu (4) vyplývá, že cena aktiva P_t má normální rozdělení, není jeho dolní mez, a tedy ani dolní mez jednoduchého čistého výnosu zaručena.

Těmto problémům lze předejít úvahou, že jednoduché výnosy aktiva definované jako:

$$R_t + 1 = \frac{P_t}{P_{t-1}} \quad (6)$$

a nazývané také jako koeficienty růstu ceny aktiva, by měly mít rozdělení nezáporné náhodné veličiny. Pro tento případ se logicky nabízí rozdělení logaritmicko-normální⁴.

Tedy za předpokladu, že jednoduchý výnos $R_t + 1$ má logaritmicko-normální rozdělení, pak jeho logaritmus:

$$r_t = \ln(R_t + 1) = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln P_t - \ln P_{t-1} = p_t - p_{t-1} \quad (7)$$

má normální rozdělení. A výnos aktiva za k období od času $t-k$ do času t lze vyjádřit jako součin k jednoduchých výnosů aktiva:

$$\begin{aligned} R_t(k) + 1 &= (R_t + 1) \cdot (R_{t-1} + 1) \cdot \dots \cdot (R_{t-k+1} + 1) = \\ &= \frac{P_t}{P_{t-1}} \cdot \frac{P_{t-1}}{P_{t-2}} \cdot \frac{P_{t-2}}{P_{t-3}} \cdot \dots \cdot \frac{P_{t-k+1}}{P_{t-k}} = \frac{P_t}{P_{t-k}}. \end{aligned} \quad (8)$$

Za předpokladu logaritmicko-normálního rozdělení jednoduchých výnosů má také celý výnos stejné rozdělení. Jeho logaritmická transformace má normální rozdělení a je rovna součtu k logaritmovaných jednoduchých výnosů [1], tj.:

$$r_t(k) = r_t + r_{t-1} + r_{t-2} + \dots + r_{t-k+1}. \quad (9)$$

Logaritmické výnosy jsou v aplikacích finančních časových řad často používány a tak si lze pro upřesnění namísto X_t uvažovat r_t .

³Obvykle se také vychází ze silnějšího předpokladu, že $\{a_t\}$ je proces striktního bílého šumu, ve kterém jsou náhodné veličiny nezávislé a stejně rozdělené s nulovou střední hodnotou a konstantním rozptylem.

⁴Logaritmus náhodné veličiny s logaritmicko-normálním rozdělením má normální rozdělení.

2.2.1 Předpoklad normality

Jak bylo uvedeno v předcházející kapitole základním předpokladem, ze kterého se často vychází je, že logaritmy výnosů mají normální rozdělení s konstantní střední hodnotou μ a konstantním rozptylem σ_r^2 , tj. předpoklad $r_t \sim N(\mu, \sigma_r^2)$. O tomto rozdělení víme, že je symetrické a že jeho šikmost je rovna nule a jeho špičatost je rovna číslu tři⁵.

Avšak ve většině případů je odhad špičatosti logaritmu výnosů finančních časových řad větší než tři, to znamená, že skutečné rozdělení logaritmu výnosů je špičatější než normální rozdělení, tzn. že rozdělení výnosů má tedy sklon k *leptokurticitě*, neboli výnosy blízko střední hodnoty a výnosy velmi vysokých záporných a kladných hodnot se objevují častěji než by předpokládala normalita. Tato vlastnost bude demonstrována na reálných datech v aplikační části. Další odchylkou (i když ne tak výraznou) může být nesymetričnost rozdělení výnosů, tj. koeficient šikmosti není nulový. Z toho vyplývá, že kladné a záporné výnosy se nevyskytují se stejnou četností. Tento jev byl pozorován u většiny aktiv kromě směnných kurzů [1].

Není žádným překvapením, že i přes uvedené nedostatky je předpoklad normálního rozdělení výnosů poměrně běžný, protože ve prospěch předpokladu normálního rozdělení jednoznačně hovoří příznivé statistické vlastnosti, například v podobě snadného testování hypotéz nebo při odhadu parametrů pomocí metody maximální věrohodnosti.

K těmto faktům pak někteří autoři poukazují na to, že i kdyby např. denní výnosy nebyly normálně rozdělené, pak tvrzení centrální limitní věty (rozdělení součtu velkého množství stejně rozdělených a nezávislých náhodných veličin s konečným rozptylem je přibližně normální) a z logaritmu vztahu (8) vyplývá, že řady výnosů s nižší frekvencí záznamu (měsíční, roční) konvergují k normálnímu rozdělení⁶. Rychlost konvergence ale není příliš vysoká, i když jsou splněny všechny předpoklady centrální limitní věty. Přesto se však i ekonomové shodují, že pro delší frekvence záznamu cen aktiv se rozdělení výnosů blíží k normálnímu rozdělení (Například v [13], [12]). Naopak pro denní a menší frekvence výnosů už podobná shoda neexistuje a nejčastěji se uvádí srovnání například se Studentovo t -rozdělením⁷.

K-S test a Lilliefors normality test

Oba tyto testy patří do kategorie testů dobré shody, pomocí kterých se zkoumá průběh celé distribuční funkce. Test Kolmogorova-Smirnova⁸ je obecný pro jakýkoliv typ rozdělení a Lillieforsova modifikace se používá pro test normálního rozdělení dat, kdy parametry neznáme a musíme je odhadnout přímo z dat pomocí výběrového průměru a rozptylu (využívá tedy jiných kritických hodnot než K-S test). K-S test je založen na výpočtu suprema vzdálenosti empirické a teoretické distribuční funkce a jeho následném porovnání s kritickou hodnotou K-S testu $D_n(\alpha)$, např. v [4].

Testujeme hypotézu takovou, že náhodný výběr X_1, \dots, X_n pochází z rozložení s distribuční funkcí $\phi(x)$. Nechť $F_n(x)$ je výběrová distribuční funkce.

Testová statistika:

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - \phi(x)|.$$

⁵Vzorci pro šikmost a špičatost a jejich bodové odhady jsou uvedeny např. v [1, str. 20].

⁶Tento fakt je ověřen např. v práci [14].

⁷Také například Lévyho stabilní rozdělení, rozdělení s proměnlivým rozptylem.

⁸Například zde: http://en.wikipedia.org/wiki/Kolmogorov-Smirnov_test.

Nulovou hypotézu zamítáme na hladině významnosti α , když $D_n \geq D_n(\alpha)$, kde $D_n(\alpha)$ je tabelovaná kritická hodnota⁹. Pro $n \geq 30$ lze $D_n(\alpha)$ aproximovat výrazem:

$$\sqrt{\frac{1}{2n} \ln \frac{2}{\alpha}}.$$

Nulová hypotéza musí specifikovat distribuční funkci zcela přesně, tj. včetně všech jejích parametrů. Pokud je neznáme a odhadujeme je (Lillieforsův test), pak se nám mění i rozložení testové statistiky D_n .

Pro hladinu významnosti 0.05 (resp. 0.01) má příslušná kritická mez asymptoticky platnou hodnotu 1,358/ \sqrt{n} (resp. 1,628/ \sqrt{n}). Lilliefors pak dále tuto hodnotu upravil pro případ testu normality rozdělení dat, kdy se nejdříve odhadnou teoretické parametry normálního rozdělení pomocí průměru a výběrové směrodatné odchylky. Asymptoticky platná kritická hodnota je pak 0,89/ \sqrt{n} (1,04/ \sqrt{n}) pro hladinu významnosti 0.05 (0.01).

Shapiro-Wilk normality test

Ve statistice se Shapiro-Wilkův test používá pro testování hypotézy takové, že náhodný výběr X_1, \dots, X_n pochází z normálního rozdělení s blíže nespecifikovanými parametry μ a σ^2 , $N(\mu, \sigma^2)$. Test je založen na porovnání bodů sestrojeného Q-Q grafu a regresní přímky proložené těmito body.

Testovací statistika:

$$W = \frac{b^2}{S^2} = \frac{\left(\sum_{i=1}^k a_{n-i+1} (y_{n-i+1} - y_i) \right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

kde y_i jsou uspořádané hodnoty náhodného výběru X_1, \dots, X_n , a_{n-i+1} jsou tabelizované váhy, \bar{y} je výběrový průměr a $k = \frac{n}{2}$ je-li n sudé, resp. $k = \frac{n-1}{2}$, je-li n liché. Pokud je výsledek testu (W) blízký jednotce (pokud $W = 1$, pak jde o dokonalou shodu), p-value není dostatečně nízké a tedy hodnota testové statistiky nepřekročí tabelovanou kritickou hodnotu Shapiro-Wilkova testu, pak není možno zamítnout H_0 předpokládající normální rozdělení, resp. čím je hodnota testovací statistiky blíže k jednotce, tím je lepší shoda mezi teoretickým a empirickým rozložením.

Podle [10] postupujeme tímto způsobem:

1. Hodnoty jednotlivých pozorování seřadíme od nejmenší po největší tak, abychom získali uspořádaný výběr $y = (y_1, \dots, y_n)$, kde $y_1 \leq \dots \leq y_n$
2. Vypočteme $S^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i^2) - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2$.
3. Vypočteme b podle vzorce $b = a_n(y_n - y_1) + \dots + a_{k+2}(y_{k+2} - y_k)$
4. Vypočteme testovou statistiku $W = \frac{b^2}{S^2}$
5. Hodnotu testové statistiky porovnáme s tabelovanou kritickou hodnotou Shapiro-Wilkova testu a učiníme závěr o zamítnutí, resp. nezamítnutí nulové hypotézy na hladině významnosti α .

⁹Například zde: http://www.kmt.zcu.cz/person/Kohout/info_soubory/letnise/tabulky.htm, odkaz platí i pro tabelizované váhy a kritické hodnoty Shapiro-Wilk testu.

Jarque-Bera test

Tento test je testem dobré shody založený na porovnávání šikmosti a špičatosti s normálním rozdělením.

Testové kritérium:

$$JB = \frac{n}{6} \left(SK^2 + \frac{1}{4} (KU - 3)^2 \right),$$

kde SK je šikmost, KU je špičatost a n je počet pozorování.

Kritická hodnota:

$$\chi_{1-\alpha}^2(\nu),$$

kde $\nu = 2$ je počet stupňů volnosti. Pokud je $JB < \chi_{1-\alpha}^2$, pak je přijímána hypotéza o tom, že data se řídí normálním rozdělením¹⁰.

2.2.2 Předpoklad lineární závislosti

V klasických analýzách finančních časových řad předpokládáme, že logaritmy výnosů jsou nekorelované stejně rozdělené náhodné veličiny s nulovou střední hodnotou a konstantním rozptylem (proces bílého šumu), nebo nezávislé stejně rozdělené náhodné veličiny s nulovou střední hodnotou a konstantním rozptylem (proces striktního bílého šumu).

Jednou ze společných vlastností logaritmů výnosů časových řad je proměnlivá variabilita neboli volatilita. Variabilita se mění ve velmi krátkých časových úsecích (pak se mezi nimi mohou objevit výnosy, které můžeme považovat za extrémní hodnoty) nebo může zůstat po určité době téměř stabilní a pak se mění, tzn. variabilita se mění ve shlucích.

Z proměnlivé variability logaritmů výnosů vyplývají další zajímavé poznatky. Zjistilo se například, že variabilita může mít vliv na úroveň a sílu autokorelace v časových řadách výnosů. Zajímavé je, že bylo zjištěno, že vysoká variabilita následuje po záporném výnosu¹¹.

Uvedené vlastnosti finančních časových řad a samozřejmě i jiné vlastnosti vyplývají z ekonomické podstaty trhu. Zároveň také z chování a jednání ekonomických subjektů (investorů).

Klasické lineární modely však předpokládají pouze jeden typ závislosti (korelační závislost) a proto nemohou zachytit charakteristické rysy analyzovaných časových řad¹².

K odhalení lineárních závislostí lze využít následující test, který se také používá k verifikaci modelů AR, MA, ARMA.

Portmanteauv test (Ljung-Boxova statistika)

Nechť X_1, \dots, X_n jsou pozorované hodnoty časové řady $\{X_t\}$ s autokorelační funkcí $\rho(\cdot)$. Za platnosti hypotézy:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0$$

platí, že:

$$Q_k = n(n+2) \sum_{i=1}^k \frac{\hat{\rho}_i}{n-i} \xrightarrow{\zeta} \chi_k^2, \quad k \in \mathbb{N},$$

kde $\xrightarrow{\zeta}$ značí konvergenci v distribuci pro $n \rightarrow \infty$. Číslo k udává počet korelačních koeficientů, které testujeme. Výsledky ze simulačních studií ukázaly, že optimální volbou k je přibližně $\ln(n)$ [2].

¹⁰Například zde: http://en.wikipedia.org/wiki/Jarque-Bera_test.

¹¹Více informací je uvedeno v knize [1, str. 25], kde je tato skutečnost otestována.

¹²Jiné typy závislostí umožňují uvažovat třída nelineárních modelů.

3 Modely volatility

Ještě před popisem samotných modelů volatility by bylo dobré se seznámit s některými důležitými pojmy, které jsou obecně platné pro modelování různých typů ekonomických a finančních časových řad. Základem pro podávání informace o charakteru stochastického procesu je autokorelační funkce a parciální autokorelační funkce (např. modely stacionárních časových řad třídy AR, MA a ARMA jsou charakteristické specifickou formou autokorelační funkce a parciální autokorelační funkce, takže odhady těchto funkcí lze použít pro identifikaci modelu) [1].

3.1 Základní reprezentace

Základními charakteristikami náhodných veličin stochastického procesu jsou nepodmíněná a podmíněná střední hodnota a nepodmíněný a podmíněný rozptyl.

Uvažujme například proces AR(1) ve formě:

$$X_t = \phi_1 X_{t-1} + a_t, \quad (10)$$

kde $|\phi_1| < 1$ a $\{a_t\}$ je proces bílého šumu s nulovou střední hodnotou a rozptylem σ_a^2 .

Tento proces lze podle [1] vyjádřit ve formě:

$$X_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \phi_1^3 a_{t-3} + \dots \quad (11)$$

Z tohoto vyjádření je zřejmé, že $E(X_t) = 0$. Podmíněná střední hodnota je střední hodnota veličiny X_t za předpokladu, že náhodné veličiny nabyly konkrétních hodnot v časech $t-1, t-2, \dots$. Podmíněná střední hodnota závisí na volbě podmínky, a je tedy její funkcí, tj. *funkce regresní*. V případě procesu AR(1) je podmínkou určitá hodnota náhodné veličiny v čase $t-1$, $E(X_t|X_{t-1}) = \phi_1 X_{t-1}$. Podmíněná střední hodnota veličiny X_t je tedy závislá na čase (proměnlivá v čase).

Ze vztahu (11) vyplývá, že nepodmíněný rozptyl veličiny X_t je neměnný v čase, tj. $\text{var}(X_t) = \frac{\sigma_a^2}{1-\phi_1^2}$. Podmíněný rozptyl se označuje jako *funkce skedastická* a její průběh charakterizuje měnlivost rozptylu veličiny X_t v závislosti na hodnotách veličin X_{t-1}, X_{t-2}, \dots . Ze vztahu (10) vyplývá, že podmíněný rozptyl veličiny X_t je $\text{var}(X_t|X_{t-1}) = \sigma_a^2$ a je také v čase neměnný (označuje se jako funkce homoskedastická, v případě proměnlivého podmíněného rozptylu se označuje jako funkce heteroskedastická). Všechny tyto vlastnosti jsou charakteristické pro modely typu AR, MA a ARMA.

Modely volatility vycházejí z představy, že např. model (10) lze modifikovat do tvaru:

$$X_t = \phi_1 X_{t-1} + \varepsilon_t, \quad (12)$$

kde $|\phi_1| < 1$ a $\{\varepsilon_t\}$ je podmíněně heteroskedastický proces s podmíněnou střední hodnotou $E(\varepsilon_t|\Omega_{t-1}) = 0$ a podmíněným rozptylem $\text{var}(\varepsilon_t|\Omega_{t-1}) = E(\varepsilon_t^2|\Omega_{t-1}) = h_t$, kde Ω_{t-1} je relevantní minulá informace až do času $t-1$. Tyto požadavky splňuje například model procesu ε_t ve tvaru:

$$\varepsilon_t = e_t + h_t^{1/2}, \quad (13)$$

kde veličiny procesu $\{e_t\}$ jsou nezávislé náhodné veličiny s nulovou střední hodnotou a jednotkovým rozptylem a h_t je podmíněný rozptyl. Pokud je rozdělení náhodné veličiny e_t normované normální, za podmínky informace, která je k dispozici v čase $t-1$, tj. $e_t \sim N_{t-1}(0, 1)$, potom je rozdělení náhodné veličiny X_t také normální, za podmínky informace, která je k dispozici v čase $t-1$, ale s podmíněným rozptylem, který se mění v závislosti na čase, tj. $X_t \sim N_{t-1}(0, h_t)$.

3.2 Lineární modely volatility

Modely volatility poprvé popsal Engle(1982)¹³. Tyto modely a modely z nich odvozené se označují jako lineární, protože jsou charakteristické tím, že podmíněný rozptyl je lineární funkcí veličin $\varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \dots, \varepsilon_{t-q}^2$.

V dalších částech budou popsány jednotlivé modely volatility, které formulují vývoj podmíněného rozptylu h_t v čase a jejich úplný tvar se bude vždy řídit podle modifikace zvoleného úrovnového modelu, tj. například (12), kde $\varepsilon_t \sim N(0, h_t)$, jak bylo uvedeno v předchozí kapitole, proto jsou formulovány pouze vývoje podmíněného rozptylu.

3.2.1 Modely ARCH (Autoregressive Conditional Heteroscedasticity)

ARCH(1)

Podmíněný rozptyl tohoto modelu má tvar:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2, \quad (14)$$

kde $\omega > 0$ a $\alpha_1 \geq 0$, aby bylo zaručeno, že podmíněný rozptyl je kladné číslo. Dále platí, že když $\alpha_1 = 0$, pak je podmíněný rozptyl v čase konstantní a proces $\{\varepsilon_t\}$ se označuje za podmíněně homoskedastický. Pokud k oběma stranám rovnice (14) přičteme ε_t^2 a odečteme h_t , dostaneme autoregresivní tvar modelu:

$$\varepsilon_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \nu_t, \quad (15)$$

kde $\nu_t = \varepsilon_t^2 - h_t = h_t(e_t^2 - 1)$ a $\varepsilon_t \sim N(0, 1)$.

Podmíněná a nepodmíněná střední hodnota procesu $\{\nu_t\}$ je nulová (tj. proces není autokorelovaný). Z teorie autoregresivních procesů vyplývá, že proces ARCH(1) je stacionární v kovariancích, je-li $\alpha_1 < 1$, pak nepodmíněná střední hodnota procesu ε_t^2 , to je nepodmíněný rozptyl procesu ε_t , má tvar:

$$\text{var}(\varepsilon_t) = \frac{\omega}{1 - \alpha_1}. \quad (16)$$

Je tedy konstantní v čase a proces ε_t je nepodmíněně homoskedastický, autokorelační funkce je ve zpoždění k rovna α_1^k .

Prostřednictvím modelu ARCH lze zachytit shluky volatility v časové řadě. To vyplývá z autoregresivního tvaru modelu (15). Je-li ε_{t-1} v absolutní hodnotě vysoké, potom lze očekávat, že ε_t v absolutní hodnotě bude vysoké. Tedy vysoké (resp. nízké) hodnoty časové řady budou následovány vysokými (resp. nízkými) hodnotami jakéhokoliv znaménka.

Tento model umožňuje také zachytit vyšší špičatost pravděpodobnostního rozdělení.

ARCH(q)

Podmíněný rozptyl tohoto modelu má tvar:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2. \quad (17)$$

Podmínky $\omega > 0$ a $\alpha_i \geq 0$ pro $i = 1, 2, \dots, q$ zaručují kladný rozptyl. Opět lze převést vztah (17) na autoregresivní tvar (konkrétně je to tvar modelu AR(q) procesu $\{\varepsilon_t^2\}$):

$$\varepsilon_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \nu_t, \quad (18)$$

¹³<http://finance.martinsewell.com/arch-garch/Engle2001.pdf>

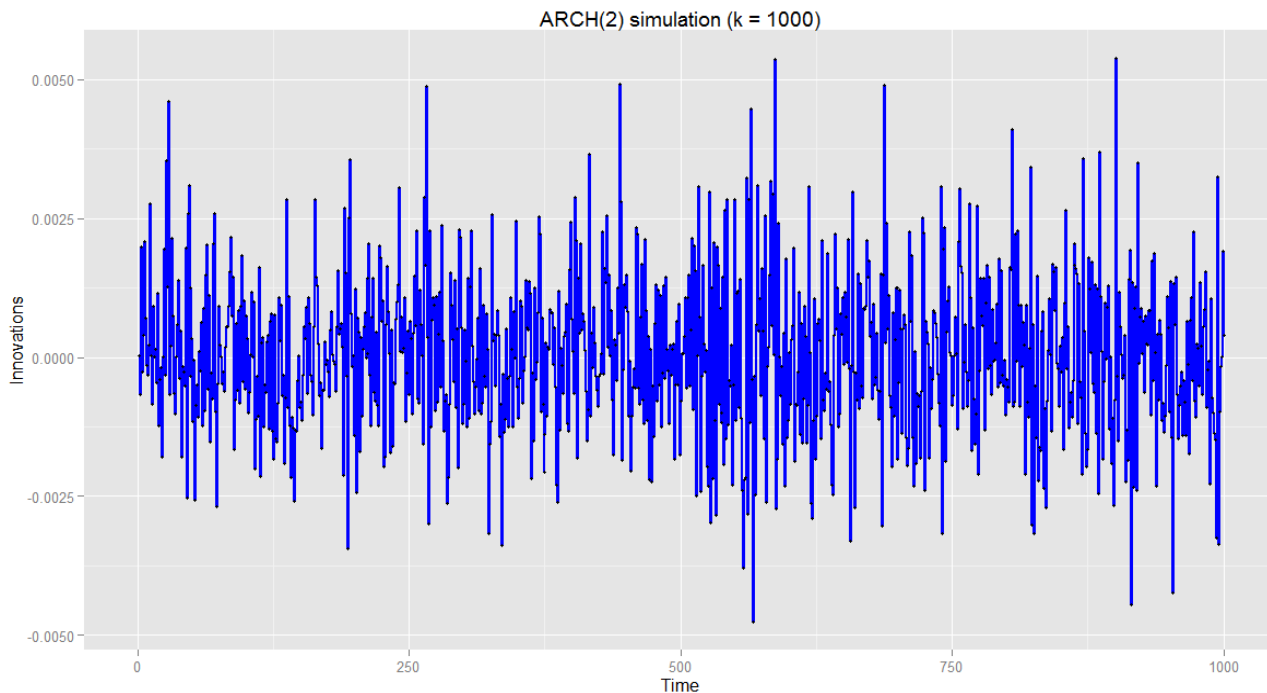
kde $\nu_t = \varepsilon_t^2 - h_t$.

Nepodmíněný rozptyl procesu $\{\varepsilon_t\}$ má tvar:

$$\text{var}(\varepsilon_t) = \frac{\omega}{1 - \alpha_1 - \dots - \alpha_q}, \quad (19)$$

a to znamená, že je konstantní v čase a proces $\{\varepsilon_t\}$ je nepodmíněně homoskedastický.

Následující obrázek ukazuje příklad (jedné realizace) nasimulovaného modelu ARCH(2) s nastavenými parametry $\alpha_1 = 0.15$, $\alpha_2 = 0.34$.



Obr. 2: Simulace modelu ARCH(2)

R code:

```

1 # Arch simulace
2 k = 1000
3
4 # ARCH(2) - use default omega and specify alpha, set beta=0!
5 spec = garchSpec(model = list(alpha = c(0.15, 0.34), beta = 0))
6 sim = garchSim(spec, n = k)
7
8 # Graf simulace ARCH(1)
9 ggplot(data.frame(Time=seq(1:k), Innovations = sim$garch), aes(Time, Innovations))
10   + geom_line(colour="blue", size = 1)
11   + geom_point(colour = "black", size = 1)
12   + labs(title = "ARCH(2)_simulation_(k=_1000)")

```

3.2.2 Modely GARCH (Generalized ARCH)

GARCH(1,1)

Při modelování časových řad, u kterých je třeba použít modely ARCH(q) s vysokým číslem q, vzniká problém s odhadováním velkého množství parametrů, umocněný existencí řady omezujících podmínek¹⁴. Tomuto problému se můžeme vyhnout rozšířením modelu ARCH o zpožděný podmíněný rozptyl¹⁵.

Rozšíření modelu ARCH(1) o podmíněný rozptyl v prvním zpoždění má model podmíněného rozptylu tvar:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}. \quad (20)$$

Podmínky $\omega > 0$, $\alpha_1 > 0$ a $\beta_1 \geq 0$ zaručují kladný podmíněný rozptyl. Model (20) se označuje jako GARCH(1,1) a lze ho použít tam, kde by bylo vhodné volit model ARCH s mnoha zpožděními. Abychom blíže nastínilí myšlenku modelů GARCH, přepíšme vztah (20) takto:

$$\begin{aligned} h_t &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \\ &= \omega + (\alpha + \beta) h_{t-1} + \alpha (\varepsilon_{t-1}^2 - h_{t-1}). \end{aligned} \quad (21)$$

Podmíněný rozptyl v této podobě se pak rovná váženému součtu rozptylu h_{t-1} předpovězeného v minulém období a neočekávaného minulého šoku $\varepsilon_{t-1}^2 - h_{t-1}$. Parametr α měří vliv tohoto šoku na předpověď pro další období, $(\alpha + \beta)$ pak představuje rychlost, s jakou vliv šoku v následujících obdobích vymizí. Čím více se $(\alpha + \beta)$ blíží jedné, tím je čas dozrívání šoku delší.

Pokud přičteme ε_t^2 k oběma stranám modelu (20) a odečteme h_t , lze tento model přepsat do tvaru modelu ARMA(1,1):

$$\varepsilon_t^2 = \omega + (\alpha_1 + \beta_1) \varepsilon_{t-1}^2 + \nu_t - \beta_1 \nu_{t-1}, \quad (22)$$

kde $\nu_t = \varepsilon_t^2 - h_t$. Jestliže $\alpha_1 + \beta_1 < 1$, pak z tohoto zápisu vyplývá, že model GARCH(1,1) je stacionární v kovariancích. Nepodmíněný rozptyl procesu $\{\varepsilon_t\}$ má tvar:

$$\text{var}(\varepsilon_t) = \frac{\omega}{1 - \alpha_1 - \beta_1} \quad (23)$$

a je tedy konstantní v čase a proces $\{\varepsilon_t\}$ je nepodmíněně homoskedastický.

Podmínka $\alpha_1 > 0$ zaručuje, aby nenastala situace, že $\alpha_1 = 0$. Důvod této podmínky vyplývá z modelu (22). Nebylo by pak možné identifikovat parametr β_1 ¹⁶.

Bollerslev (1988) také odvodil rekurentní vztah pro autokorelační funkci procesu ε_t^2 :

$$\rho_1 = \alpha_1 + \frac{\alpha_1^2 \beta_1}{1 - 2\alpha_1 \beta_1 - \beta_1^2}, \quad (24)$$

$$\rho_k = (\alpha_1 + \beta_1)^{k-1} \rho_1, \quad k = 2, 3, \dots \quad (25)$$

Hodnoty autokorelační funkce s rostoucím zpožděním k exponenciálně klesají. Rychlost klesání těchto hodnot závisí na hodnotě součtu $\alpha_1 + \beta_1$.

Pokles autokorelační funkce je velmi pozvolný pokud se tento součet blíží k hodnotě jedna. Hodnoty parciální autokorelační funkce se chovají obdobně. Obecně tedy lze konstatovat, že tvar ACF a PACF procesu ε_t^2 odpovídá tvaru těchto funkcí modelu ARMA(p,q).

¹⁴Např. nenulovost podmíněného rozptylu, stacionarita.

¹⁵Toto rozšíření navrhl Bollerslev (1986).

¹⁶Podrobněji viz.[1, str. 167].

GARCH(p,q)

Podmíněný rozptyl modelu GARCH má obecný tvar:

$$h_t = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i}. \quad (26)$$

Podmínky $\omega > 0$, $\alpha_i > 0$ pro $i = 1, 2, \dots, q$ a $\beta_i \geq 0$ pro $i = 1, 2, \dots, p$ zaručují kladný podmíněný rozptyl. Model (26) se označuje jako GARCH(p,q) a lze ho přepsat do tvaru:

$$\varepsilon_t^2 = \omega + \sum_{i=1}^m (\alpha_i + \beta_i) \varepsilon_{t-i}^2 - \sum_{i=1}^p \beta_i \nu_{t-i} + \nu_t, \quad (27)$$

kde $\nu_t = \varepsilon_t^2 - h_t$. Nepodmíněný rozptyl procesu $\{\varepsilon_t\}$ je:

$$\text{var}(\varepsilon_t) = \frac{\omega}{1 - \alpha(1) - \beta(1)}, \quad (28)$$

kde

$$\alpha(1) = \sum_{i=1}^q \alpha_i \quad \text{a} \quad \beta(1) = \sum_{j=1}^p \beta_j.$$

Další typy a modifikace

V literatuře lze nalézt další typy těchto modelů. Zde jsou uvedeny jen ty nejdůležitější pro aplikační část této práce. Mimo jiné byly vyvinuty další modifikace těchto modelů pro zachycení nejrůznějších typů asymetrických efektů (QGARCH, VS-GARCH nebo ANST-GARCH), které lze nalézt např. v [1].

3.3 Výstavba modelů volatility

Výstavbou modelů volatility rozumíme určitý postup pro správnou aplikaci modelů volatility. Tento postup lze obecně shrnout do těchto kroků:

1. Pro časovou řadu se určí vhodný lineární nebo nelineární úrovnový model.
2. Analyzuje se kvalita úrovnového modelu z hlediska výsledku testu podmíněné heteroskedasticity a normality, případně se otestuje hypotéza podmíněné heteroskedasticity nelineárního typu.
3. Odhadnou se parametry lineárního nebo nelineárního modelu podmíněné heteroskedasticity.
4. Ověří se vhodnost daného modelu diagnostickými testy.
5. Pokud je třeba, model se dále modifikuje.
6. Konečný model se použije k popisným nebo predikčním účelům.

3.4 Testování podmíněné heteroskedasticity

Test podmíněné heteroskedasticity lineárního typu vychází z formulace modelu ARCH. Podmíněný rozptyl h_t modelu ARCH(q) ve tvaru (17) je konstantní, jsou-li parametry odpovídající veličinám $\varepsilon_{t-1}^2, \dots, \varepsilon_{t-q}^2$ rovny nule. Nulová hypotéza, tj. hypotéza podmíněné heteroskedasticity, je $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_q = 0$, alternativní hypotézou je, že alespoň jeden parametr je různý od nuly, tj. $H_1 : \text{non}H_0$.

Test je prováděn v následujících krocích:

1. Odhadnou se parametry lineárního nebo nelineárního úrovnového modelu a získají se rezidua $\hat{\varepsilon}_t$ a reziduální součet čtverců $RSE_0 = \sum \hat{\varepsilon}_t^2$.
2. Konstruuje se regresní model

$$\hat{\varepsilon}_t^2 = \omega + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \alpha_2 \hat{\varepsilon}_{t-2}^2 + \dots + \alpha_q \hat{\varepsilon}_{t-q}^2 + u_t, \quad (29)$$

z kterého se získá reziduální součet čtverců RSE_1 a index determinace R^2 .

3. Testové kritérium LM ve tvaru $T \cdot R^2$ má za předpokladu platnosti nulové hypotézy asymptoticky rozdělení $\chi^2(q)$.
4. F -verze tohoto testového kritéria pro malé výběry má tvar

$$F_{LM} = \frac{(RSE_0 - RSE_1)/q}{RSE_1/(T - q - 1)}, \quad (30)$$

její rozdělení lze za předpokladu platnosti nulové hypotézy aproximovat rozdělením $F(q, T - q - 1)$.

Tento test se často označuje jako ARCH test. Lze jej také interpretovat jako test autokorelace čtverce nesystematické složky.

ARCH-LM test

Pro řadu $\{\varepsilon_t\}$ se zkonstruuje lineární regresní model, kde jako vysvětlující proměnné vystupuje q zpožděných hodnot $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$, tj.:

$$\varepsilon_t = \omega + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \dots + \alpha_q \varepsilon_{t-q} + \nu_t.$$

Dále pak odhadneme parametry modelu a vypočítáme koeficient determinace R^2 . Za platnosti nulové hypotézy:

$$H_0 : \alpha_i = 0 \text{ pro } i = 1, \dots, q$$

má statistika $LM = T \cdot R^2$ asymptoticky $\chi^2(q)$ rozdělení s q stupni volnosti. Tento test lze interpretovat tak, že pokud by pro model ARCH(q) ve tvaru (17) platilo $\alpha_1 = \dots = \alpha_q = 0$, byl by jeho podmíněný rozptyl konstantní.

3.5 Volba řádu modelu

Při výstavbě modelu typu ARCH narazíme na zásadní problém a tím je volba řádu modelu. Tento řád lze přibližně odhadnout z průběhu autokorelační funkce nebo parciální autokorelační funkce řady $\{\varepsilon_t^2\}$. Avšak žádný formální test, podle kterého by bylo možné určit řád modelu není k dispozici.

Obvyklý postup při určování řádu modelu typu ARCH je takový, že se nejprve odhadne model nízkého řádu a poté se tento řád upraví např. podle výsledků statistické významnosti parametrů nebo analýzy standardizovaných reziduí.

Ve velké většině případů jsou postačující modely nízkých řádů např. ARCH(1), ARCH(2) nebo GARCH(1,1), GARCH(2,1) atd.

3.6 Odhad parametrů

Odhady parametrů modelů volatility se obvykle provádějí pomocí metody maximální věrohodnosti¹⁷.

V aplikacích modelu ARCH se běžně uvažuje normované normální rozdělení $\epsilon \sim N(0, 1)$ nebo standardizované Studentovo rozdělení $\epsilon \sim \frac{z_t}{\sqrt{\text{var} z_t}}$, kde $z_t \sim t_\nu$. Je samozřejmě možné využít i jiná rozdělení, například v [5].

Uvažujme tedy model ARCH(m) a $\epsilon \sim N(0, 1)$. Dále pak označme pozorované hodnoty y_1, \dots, y_T náhodných veličin Y_1, \dots, Y_T . Podle vlastností podmíněné hustoty využijeme:

$$\begin{aligned} f(y_1, \dots, y_T | \alpha) &= f(y_T, \dots, y_{m+1} | \alpha, \mathcal{F}_m) \cdot f(y_m, \dots, y_1 | \alpha) = \\ &= f(y_T, \dots, y_{m+2} | \alpha, \mathcal{F}_{m+1}) \cdot f(y_{m+1} | \alpha, \mathcal{F}_m) \cdot f(y_m, \dots, y_1 | \alpha) = \\ &\vdots \\ &= f(y_T | \alpha, \mathcal{F}_{T-1}) \cdots f(y_{m+1} | \alpha, \mathcal{F}_m) \cdot f(y_m, \dots, y_1 | \alpha), \end{aligned}$$

kde $f(y_1, \dots, y_T | \alpha)$ je sdružená hustota veličin Y_1, \dots, Y_T v bodě y_1, \dots, y_T , která je podmíněná vektorem parametrů $\alpha = (\alpha_1, \dots, \alpha_m)$ a $f(y_t | \alpha, \mathcal{F}_{t-1})$ je hustota náhodné veličiny Y_t podmíněná vektorem parametrů α a σ -algebrou $\mathcal{F}_{t-1} = \sigma(Y_{t-1}, \dots, Y_1)$ generovanou veličinami Y_{t-1}, \dots, Y_1 .

Protože $\epsilon \sim N(0, 1)$ a $Y_t = \sigma_t \epsilon_t$, jsou podmíněná rozdělení $Y_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$ a tedy:

$$f(y_t | \mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{y_t^2}{2\sigma_t^2}}.$$

Stále uvažujme model, kde $\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \dots + \alpha_m Y_{t-m}^2$, nebo-li $\sigma_t^2 = \sigma_t^2(\alpha)$. Pak můžeme psát:

$$f(y_1, \dots, y_T | \alpha) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{y_t^2}{2\sigma_t^2}} f(y_m, \dots, y_1 | \alpha).$$

A protože vyjádření hustoty $f(y_m, \dots, y_1 | \alpha)$ by bylo složité, podmíníme sdruženou hustotu $f(y_1, \dots, y_T | \alpha)$ hodnotami y_1, \dots, y_m a budeme je odhadovat společně s α . Dostaneme potom:

$$f(y_1, \dots, y_T | \alpha, y_1, \dots, y_m) = f(y_1, \dots, y_T | \alpha, \mathcal{F}_m) = \prod_{t=m+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{y_t^2}{2\sigma_t^2}}.$$

Metoda maximální věrohodnosti tedy staví na myšlence, že náš výběr z rozdělení s hustotou f je „nejvíce pravděpodobný“, a tedy hustota f je v bodě určeném naším výběrem maximální. Tato hustota závisí na neznámém vektoru α . Hledáme tedy takové jejich hodnoty, aby hustota byla maximální, které jsou pak odhadem metodou maximální věrohodnosti.

Druhým případem je, že ϵ_t má standardizované Studentovo t -rozdělení s ν stupni volnosti. V tomto případě je ϵ_t ve tvaru $\frac{z_t}{\sqrt{\text{var} z_t}}$, kde $z_t \sim t_\nu$. Z vlastností Studentova rozdělení víme, že $\text{var} z_t = \frac{\nu}{\nu-2}$ pro $\nu > 2$. Hustotu ϵ_t lze získat jako hustotu lineární transformace z_t s využitím věty o transformaci náhodných veličin (viz. [6]). Z tohoto vyplývá:

$$f_{\epsilon_t}(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{(\nu-2)\pi}} \left(1 + \frac{x^2}{\nu-2}\right)^{-\frac{\nu+1}{2}},$$

kde $\nu > 2$, $x \in \mathbb{R}$ a $\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy$ pro $x > 0$. Podobně jako v případě $\epsilon_t \sim N(0, 1)$ dostáváme:

$$f(y_1, \dots, y_T | \alpha, \mathcal{F}_m) = \prod_{t=m+1}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{(\nu-2)\pi\sigma_t^2}} \left(1 + \frac{y_t^2}{(\nu-2)\sigma_t^2}\right)^{-\frac{\nu+1}{2}}.$$

¹⁷Například v knize [1, str. 190] je tato metoda podrobně popsána.

Tuto funkci maximalizujeme, pokud je počet stupňů ν zadán a jedinými neznámými parametry jsou α . Pokud není parametr ν zadán, pak je přidán do podmínky podmíněné hustoty a odhadne se spolu s α . V praxi se ukázalo, že odhadnutý počet stupňů volnosti je nejčastěji mezi hodnotami 3 a 6 (uvedeno například v [2]). Pokud bychom potřebovali počáteční nastavení tohoto parametru např. pro nějaký iterační proces, volba tohoto parametru by byla mezi uvedenými čísly.

Výsledný odhad má tvar:

$$\hat{\sigma}_t = \hat{\alpha}_0 + \hat{\alpha}_1 Y_{t-1}^2 + \dots + \hat{\alpha}_m Y_{t-m}^2.$$

Je dobré si také uvědomit, že odhady parametrů dané časové řady pro model volatility mohou být zkreslené. To může být způsobeno např. odlehlými pozorováními, které také mohou zkreslit výsledky testů podmíněné heteroskedasticity. Těmto problémům lze předejít pomocí metod robustních odhadů parametrů, které jsou vzhledem k odlehlým pozorováním robustní¹⁸. Tento přístup však není cílem práce.

3.7 Verifikace modelu

Ověřit platnost a vhodnost modelu pro konkrétní časovou řadu můžeme několika způsoby:

1. Testování nesystematické složky - konkrétně se jedná o testy autokorelace (např. portmanteau uvedený v kapitole 2.2.2), normality (např. Jarque-Bera test uvedený v kapitole 2.2) a podmíněné heteroskedasticity (Ljung-Boxův Q-test¹⁹ nebo ARCH-LM test uvedený v kapitole 3.4)
2. Testování linearity - pro testování lineárního modelu GARCH proti nějakému nelineárnímu modelu volatility lze použít testy jako jsou SB, NSB nebo PSB²⁰. Tyto testy souvisejí s testováním podmíněné heteroskedasticity nelineárního typu.

Velkou vahou ve správném specifikování modelu je to, že pro jeho standardizovaná residua by mělo platit $\hat{e}_t \sim \text{IID}(0, 1)$ ²¹. To lze ověřit právě testem ARCH.

3.7.1 Kritéria hodnocení modelu

Další posouzení vhodnosti určitého modelu lze provést několika způsoby. Jedním z nich je porovnáním kritérií, které určitým způsobem modely hodnotí.

V této části uvedeme kritéria AIC a BIC, která jsou definována jako²²:

$$\begin{aligned} AIC &= -2\ln L(\hat{\varphi}) + 2M \\ BIC &= -2\ln L(\hat{\varphi}) + M\ln(T), \end{aligned}$$

kde $\ln L(\hat{\varphi})$ je hodnota maximalizované logaritmicke věrohodnostní funkce, $M = p + 1$ a p je počet parametrů modelu, T je počet pozorování²³. Bylo dokázáno, že AIC vede k nadhodnocení řádu autoregrese ([1, str. 105], Akaike (1979)).

Vybere se takový model, který vede k minimální hodnotě těchto kritérií.

¹⁸Jedna z metod je uvedena v knize [1, str. 193].

¹⁹Viz. například: http://en.wikipedia.org/wiki/LjungBox_test.

²⁰Tyto testy jsou popsány například v knize [1, str. 187].

²¹IID = independent and identically distributed (nezávislé a stejně rozdělené).

²²Akaikeho informační kritérium (Akaike Information Criterion), Bayesovo informační kritérium (Bayesian Information Criterion)

²³Podle manuálu pro R: <http://tolstoy.newcastle.edu.au/R/e2/help/06/09/1750.html>.

3.8 Konstrukce předpovědí

Pokud bychom uvažovali proces ARMA(P,Q)-GARCH(p,q) ve tvaru:

$$\phi(B)X_t = \Theta(B)\varepsilon_t, \quad \varepsilon_t \sim GARCH(p, q), \quad (31)$$

kde $\phi(B) = 1 - \phi_1 B - \dots - \phi_P B^P$, $\Theta(B) = 1 + \Theta_1 B + \dots + \Theta_Q B^Q$. Označme \hat{X}_{t+k} nejlepší (ve smyslu minimální střední kvadratické chyby) k -krokovou předpověď tohoto procesu konstruovanou v čase t , tj. $\hat{X}_{t+k} = E_t(X_{t+k})$. Lze ukázat, že podmíněná heteroskedasticita nemá na bodové předpovědi \hat{X}_{t+k} vliv. Naopak se ale projeví na velikosti chyb předpovědí a předpovědních intervalů.

Pro výpočet kvadratické chyby předpovědi budeme předpokládat, že kořeny polynomu $1 - \phi_1 z - \dots - \phi_P z^P$, $z \in \mathbb{C}$ leží vně jednotkového kruhu.

Pak můžeme vyjádřit proces X_t v kauzálním tvaru:

$$X_t = \frac{\Theta(B)}{\phi(B)}\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}.$$

Tento proces v čase $t+k$ je možné zapsat ve formě:

$$X_{t+k} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t+k-j},$$

protože $E_t(\varepsilon_{t+k-j}) = 0$ pro $j < k$, nejlepší předpověď \hat{X}_{t+k} je tedy:

$$\hat{X}_{t+k} = \sum_{j=k}^{\infty} \psi_j \varepsilon_{t+k-j}.$$

Chybu předpovědi můžeme vypočítat jako:

$$r_{t,k} = X_{t+k} - \hat{X}_{t+k} = \sum_{j=0}^{k-1} \psi_j \varepsilon_{t+k-j}$$

a minimální nepodmíněná střední kvadratická chyba (Mean Square Error, MSE) predikce je potom:

$$MSE(\hat{X}_{t+k}) = E(r_{t,k}^2) = \text{var}(r_{t,k}) = \sigma_\varepsilon^2 \sum_{j=0}^{k-1} \psi_j^2, \quad (32)$$

kde σ_ε^2 značí nepodmíněný rozptyl procesu ε_t .

$$MSE_t(\hat{X}_{t+k}) = E_t(r_{t,k}^2) = \sum_{j=0}^{k-1} \psi_j^2 E_t(h_{t+k-j}), \quad (33)$$

kde $E_t(h_{t+k-j})$ je nejlepší předpověď podmíněného rozptylu h_{t+k-j} . Při výpočtu této předpovědi můžeme postupovat následovně: nechť h_t je podmíněný rozptyl procesu GARCH(p,q) ve tvaru:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 h_{t-1} + \dots + \beta_p h_{t-p}.$$

Hodnotu tohoto podmíněného rozptylu v čase $t+k$ můžeme zapsat jako:

$$h_{t+k} = \omega + \alpha_1 \varepsilon_{t+k-1}^2 + \dots + \alpha_q \varepsilon_{t+k-q}^2 + \beta_1 h_{t+k-1} + \dots + \beta_p h_{t+k-p},$$

potom nejlepší předpověď $\hat{h}_{t+k} = E_t(h_{t+k})$ má tvar:

$$\begin{aligned}\hat{h}_{t+k} &= \omega + \alpha_1 \hat{\varepsilon}_{t+k-1}^2 + \cdots + \\ &\alpha_q \hat{\varepsilon}_{t+k-q}^2 + \beta_1 \hat{h}_{t+k-1} + \cdots + \beta_p \hat{h}_{t+k-p},\end{aligned}$$

kde

$$\begin{aligned}\hat{h}_{t+j} &= E_t(h_{t+j}) = \hat{\varepsilon}_{t+j}^2 = E_t(\varepsilon_{t+j}^2) & j \geq 1, \\ \hat{h}_{t+j} &= E_t(h_{t+j}) = h_{t+j} & j \leq 0, \\ \hat{\varepsilon}_{t+j} &= E_t(\varepsilon_{t+j}^2) = \varepsilon_{t+j}^2 & j \leq 0.\end{aligned}$$

Dále budeme uvažovat předpověď vypočítanou na základě modelu GARCH(1,1). Podmíněný rozptyl h_t je:

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1},$$

předpověď \hat{h}_{t+k} má potom tvar:

$$\hat{h}_{t+k} = \omega + \alpha \hat{\varepsilon}_{t+k-1}^2 + \beta \hat{h}_{t+k-1},$$

a opakovanou substitucí lze získat:

$$\begin{aligned}\hat{h}_{t+k} &= \omega \sum_{j=0}^{k-1} (\alpha + \beta)^j + (\alpha + \beta)^{k-1} \alpha \varepsilon_t^2 + (\alpha + \beta)^{k-1} \beta h_t \\ &= \omega \sum_{j=0}^{k-1} (\alpha + \beta)^j + (\alpha + \beta)^{k-1} h_{t+1}.\end{aligned}$$

Jestliže je proces GARCH(1,1) stacionární, tzn. že platí $\alpha + \beta < 1$ a $\sigma_\varepsilon^2 = \omega / (1 - \alpha - \beta)$, lze předchozí tvar vyjádřit jako:

$$\hat{h}_{t+k} = \sigma_\varepsilon^2 + (\alpha + \beta)^{k-1} (h_{t+1} - \sigma_\varepsilon^2)$$

a pokud $k \rightarrow \infty$, předpověď podmíněného rozptylu konverguje k nepodmíněnému rozptylu σ_ε^2 (více například v [1] nebo [9]).

4 Monte Carlo

Metoda Monte Carlo je poměrně široká skupina numerických výpočetních metod, která je založena na využití náhodných veličin a teorie pravděpodobnosti. Můžeme říci, že jde o simulaci pomocí stochastických metod, které využívají pseudonáhodná čísla. Tato metoda je široce využívána např. pro simulace experimentů, počítání integrálů, výpočtů čísla π , řešení diferenciálních rovnic atd.

Simulování je často jednoduchou cestou k řešení velmi složitých, ale praktických úloh. Jinak řečeno základním principem metody Monte Carlo je simulace velkého množství možných budoucích stavů, jejichž počet může jít až do stovek tisíců. Následně se pak provádí analýza hlavních charakteristik výsledků. Tato metoda spadá do skupiny metod, které jsou využívány v případech, kdy matematické řešení problému je buď příliš složité nebo jej dokonce nelze nalézt. Do skupiny úloh řešitelných pomocí metody Monte Carlo můžeme zařadit téměř libovolnou funkci, která převádí vstupní veličiny X na výstupní Y . Co je potřeba znát je distribuční funkce vstupních veličin X , pomocí které se generují náhodné vstupní veličiny a tak zaznamenáváme výstupní veličiny. Po dostatečném²⁴ počtu opakování, lze pomocí statistické analýzy výstupů odhadnout hledaný parametr.

Pro Monte Carlo existují dvě varinaty:

- Geometrický model - v tomto případě se při výpočtu nepoužívá model reálného děje (př. výpočet integrálu),
- Analogový model - musíme znát pravděpodobnostní rozdělení zkoumaných jevů. Opakovanou simulací pak získáváme výsledky (náhodné veličiny).

Pro generátory náhodných čísel, které se využívají při této metodě, je potřeba využít nějaký hardwarový generátor. Ty jsou většinou založeny na nějakých náhodných procesech (např. se využívá aktuální datum a čas, měření nějakého šumu, např. polovodičů či kolísání napětí, rozpad atomového jádra izotopu). Metodu Monte Carlo lze využít v mnoha aplikacích a modelech.

O této metodě je sepsáno mnoho článků a knih v různých aplikacích či verzích. Další informace o této metodě lze nalézt např. v [7] nebo [8].

4.1 Předpovědi modelů GARCH pomocí MC

V této části bude stručně popsáno, jakým způsobem lze pomocí metody Monte Carlo předpovědět budoucí hodnotu časové řady modelované pomocí GARCH modelu.

Uvedmě nejdříve, co v tomto případě předchází použití metody MC:

- Pro stochastickou časovou řadu máme již odhadnutý a verifikovaný model podmíněné střední hodnoty (ARMA) a odhad $\hat{\varepsilon}_t$,
- Pro řadu residuí $\hat{\varepsilon}_t$ máme odhadnutý a verifikovaný model podmíněného rozptylu (GARCH),
- Pro standardizovaná residua $\hat{\varepsilon}_t$ modelu GARCH jsme určili jakým rozdělením se řídí (nejčastěji normální nebo studentovo rozdělení).

V tuto chvíli je vše připraveno pro aplikaci metody Monte Carlo.

²⁴Zde se objevuje problém, co znamená dostatečný. Obecně a logicky však platí, že přesnost odhadu pomocí metody Monte Carlo roste s počtem opakování.

Jak už bylo uvedeno výše, metoda je založena na generování náhodných čísel a budeme tedy postupovat v následujících krocích:

- Budeme tedy generovat náhodná čísla R , resp. pravděpodobnosti, v intervalu $(0, 1)$.
- Dále budeme potřebovat modelovou distribuční funkci $p = F(\hat{\epsilon}_t)$, resp. inverzní distribuční funkci neboli kvantilovou funkci $\hat{\epsilon}_t = F^{-1}(p)$ rozdělení residuí $\hat{\epsilon}_t$ (většinou kvantilová funkce normálního nebo studentova rozdělení).
- Pro generovanou náhodnou pravděpodobnost R určíme pomocí kvantilové funkce první simulovanou hodnotu $\hat{\epsilon}_{t+1}^1 = F^{-1}(R)$.
- Tím získáme jednu simulovanou předpověď $\hat{\epsilon}_{t+1}^1 = F^{-1}(R)$ pro standardizovaná residua GARCH modelu. Pokud uvažujeme např. model GARCH(1,1), pak simulovanou předpověď pro $t + 1$ podmíněného rozptylu bude $\hat{h}_{t+1} = \hat{\omega} + \hat{\alpha}_1 \hat{\epsilon}_t + \hat{\beta}_1 \hat{h}_t$.
- Poté lze určit jednu simulovanou předpověď pro $\hat{\epsilon}_{t+1}^1 = \hat{\epsilon}_{t+1} \cdot \sqrt{\hat{h}_{t+1}}$, tj. simulované residuum ARMA modelu v čase $t + 1$.
- Tento postup pak opakujeme k -krát, abychom dostali k náhodných simulací předpovědi v čase $t + 1$, tj. $\hat{\epsilon}_{t+1}^1, \dots, \hat{\epsilon}_{t+1}^k$ a dále pak $\hat{\epsilon}_{t+1}^1, \dots, \hat{\epsilon}_{t+1}^k$.

Pro delší předpověď až do času $t + h$ lze postupovat krok po kroku, kdy popsany postup opakujeme, ale až poté, co dokončíme simulaci do času $t + h$. Tedy v momentě, kdy pomocí náhodného čísla získáme odhad $\hat{\epsilon}_{t+1}^1$, vygenerujeme další náhodnou pravděpodobnost R , pro kterou určíme $\hat{\epsilon}_{t+2}^1 = F^{-1}(R)$ a také určíme \hat{h}_{t+2} za pomoci $\hat{\epsilon}_{t+1}^1$, který jsme odhadli v přechozím kroku. Pomocí těchto simulovaných odhadů můžeme dále vypočítat $\hat{\epsilon}_{t+2}^1$.

Takto bychom pokračovali až do času $t + h$ a poté celý tento postup opakovali a generovali řadu $\hat{\epsilon}_{t+1}^2, \dots, \hat{\epsilon}_{t+h}^2$, resp. $\hat{\epsilon}_{t+1}^2, \dots, \hat{\epsilon}_{t+h}^2$. Podle zvolené hodnoty k počtu simulací pak postup opakujeme k -krát a tedy poslední získanou řadou simulovaných hodnot bude $\hat{\epsilon}_{t+1}^k, \dots, \hat{\epsilon}_{t+h}^k$, resp. $\hat{\epsilon}_{t+1}^k, \dots, \hat{\epsilon}_{t+h}^k$.

5 Project R

R je jazyk a prostředí pro statistické výpočty a grafiku. Je to GNU projekt, který je podobný jazyku S a prostředí, které bylo vyvinuto v Bell laboratories (dříve AT & T, nyní Lucent Technologies) Johnem Chambersem a kolegy.

Projekt R nabízí širokou řadu statistických funkcí (lineární a nelineární modelování, klasické statistické testy, analýzy časových řad, klasifikace, clustering, ...) a grafické techniky, a je vysoce rozšiřitelný. Jazyk S je často preferovaným prostředkem pro výzkum statistické metodiky a R poskytuje Open Source cestu k účasti na této činnosti.

Jednou ze silných stránek R je snadnost, s níž mohou být produkovány dobře navržené a publikovatelně kvalitní grafy, včetně matematických symbolů a vzorců.

R je k dispozici jako Free software v souladu s podmínkami GNU General Public License (Free Software Foundation) ve formě zdrojového kódu. Běží na nejrůznějších platformách UNIX a podobných systémech (včetně FreeBSD a Linux), Windows a MacOS.

R je integrovaná sada softwarových zařízení pro manipulaci s daty, výpočty a grafické zobrazení. Zahrnuje tedy:

- efektivní zpracování dat a skladovací prostory,
- sadu operátorů pro výpočty polí, zejména `matic`,
- velkou, soudržnou, integrovanou kolekci dílčích nástrojů pro analýzu dat,
- grafické zařízení pro analýzu dat a zobrazení buď na obrazovce nebo pro tisk,
- dobře vyvinutý, jednoduchý a efektivní programovací jazyk, který zahrnuje podmínky, smyčky, uživatelem definované rekurzivní funkce a vstupní a výstupní zařízení.

Termín „prostředí“ je určen k charakterizování plně plánovaného a soudržného systému, než přidavného narůstání velmi specifických a neflexibilních nástrojů, jak je tomu s jinými softwary pro analýzu dat.

Mnoho uživatelů smýšlí o R jako o statistickém systému. Vývojáři dávají přednost spíše pojmu prostředí, ve kterém jsou prováděny statistické metody. R může být rozšířen (snadno) pomocí balíčků. Existuje osm balíčků dodávaných s distribucí R a mnoho dalších je k dispozici prostřednictvím internetových stránek skupiny CRAN obsahujících velmi širokou řadu moderních metod statistiky.

R má svou vlastní dokumentaci ve formátu LaTeX-like, který se používá k zásobování komplexní dokumentace, a to jak on-line formátů tak tištěných podobách [18].

5.1 Počáteční nastavení a načtení dat

Ze stránek <http://www.r-project.org/> lze stáhnout instalační soubor, který nám umožní nainstalovat Project R²⁵. Po instalaci spustíme Project R a do hlavní konzole napíšeme příkazy:

R code:

```
1 install.packages("Rcmdr")
2 library(Rcmdr)
```

²⁵Podrobnější informace a návody lze nalézt například na stránkách <http://www.karlin.mff.cuni.cz/~kulich/vyuka/Rdoc/index.html>.

Prvním příkazem nainstalujeme balíček, který umožní použití uživatelsky příjemnější konzole. Druhým příkazem tuto konzoli spustíme. V této konzoli načteme vytvořený soubor v projektu R, který umožní zpracování časové řady. Pomocí sekvence příkazů `"File" - "Open script file" - ".. \GARCH" - MainGARCH.R` otevřeme zdrojový soubor, který je na CD přiloženém k této diplomové práci.

V první části tohoto skriptu je načtení balíčků potřebných v dalších částech skriptu a také nastavení pracovního adresáře a načtení dat.

R code:

```

1 # Instalace použitých balíčků
2 install.packages("tseries") # balíček pro časové řady
3 install.packages("FinTS") # balíček pro časové řady
4 install.packages("nortest") # balíček pro testy normality
5 install.packages("ggplot2") # balíček pro lepší grafy
6 install.packages("xtable") # balíček pro sázení tabulek do LaTeXu
7 install.packages("lmtest") # balíček pro Likelihood ratio test
8 install.packages("fGarch") # balíček pro modely typu ARCH
9 install.packages("reshape") # balíček pro funkci melt na zpracování dat
10 install.packages("quantreg") # balíček pro funkci quantile
11 install.packages("scales") # balíček pro vykreslování v ggplot2 funkce geom_quantile
12
13 # Načtení použitých knihoven
14 library(xtable) # knihovna pro sázení tabulek do LaTeXu
15 library(fGarch) # knihovna pro modelování volatility
16 library(lmtest) # knihovna pro Likelihood ratio test
17 library(FinTS) # knihovna s nástroji pro finanční časové řady
18 library(nortest) # knihovna pro testy normality
19 library(tseries) # knihovna pro časové řady
20 library(ggplot2) # knihovna pro grafický nástroj
21 library(reshape)
22 library(quantreg)
23 library(scales)
24 library(relimp, pos=4)
25 require(graphics)
26 require(grid)
27
28 # Nastavení adresáře s daty
29 getwd() # zjištění aktuálního pracovního adresáře
30 setwd("C:/Users/pakosta/Documents/R") # nastavení pracovního adresáře
31
32 # Nastavení cest k použitým funkcím
33 filepath <- "../functions/" # nastavení cesty k funkcím,
34 # složka s funkcemi musí být v aktuálním pracovním adresáři
35 source(paste(filepath, "QQplot.r", sep = ""))
36 source(paste(filepath, "qpacf.r", sep = ""))
37 source(paste(filepath, "qacf.r", sep = ""))
38 source(paste(filepath, "NKH.r", sep = ""))
39 source(paste(filepath, "NKHst.r", sep = ""))
40
41 # Příprava dat
42 # Načtení dat
43 data <- read.csv(file = "gold.csv", head = FALSE, sep = ";")
44
45 # Rozdělení dat do skupin
46 # Převod formátu datumů na formát datumů v R
47 dates = rev(as.Date(data$V1, format = "%d.%m.%Y"))
48 serie = rev(data$V2) # uložení dat do pomocné proměnné

```

Jak je vidět z příkazů první části skriptu, data jsou načtena ze souboru `.csv`, který je také na CD přiloženém k této práci. Soubor obsahuje 2 sloupce s datумы (formát `DD.MM.RRRR`) a cenami zlata v daných časech (s desetinnou tečkou). Dále je pak nastavena cesta do vlastního pracovního adresáře, který je zvolen pomocí příkazu `setwd` a také obsahuje složku s dalšími funkcemi, které hlavní skript využívá.

5.2 Použité nástroje a funkce

V této části budou uvedeny některé důležité balíčky a nejdůležitější použité funkce. K těmto vybraným funkcím budou vždy stručně napsané základní informace. Pokud požadujeme detailnější informace o balíčku nebo funkci, v příkazovém řádku programu napíšeme před název balíčku nebo funkce otazník (např. `?FinTS` nebo `?shapiro.test`). Chybějící balíčky lze doinstalovat pomocí již uvedené funkce `install.packages("balicek")`.

Balíček „**tseries**“ - pro analýzu časových řad a výpočetní finančnictví.

Podrobné informace dostupné na: <http://cran.r-project.org/web/packages/tseries/tseries.pdf>

Balíček „**FinTS**“ - R-kový protějšek knihy Analysis of Financial Time Series, 2nd ed. (Wiley), od autora Tsay (2005). Zahrnuje soubory dat, funkce a soubory skriptů potřebné k práci s některými příklady.

Podrobné informace dostupné na: <http://math.uncc.edu/~zcaif/FinTS.pdf>

Balíček „**nortest**“ - soubor 5 testů pro složené hypotézy normality.

Podrobné informace dostupné na: <http://cran.r-project.org/web/packages/nortest/nortest.pdf>

Balíček „**ggplot2**“ - je kreslicí systém pro R, který používá jen dobré vlastnosti ze základní R-kové grafiky a balíčku Lattice grafiky. Tento balíček poskytuje silný model pro grafiku, který umožňuje lehce vytvářet komplexní vícevrstvou grafiku.

Podrobné informace dostupné na: <http://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf>

Balíček „**fGarch**“ - Prostředí pro výuku "Finanční inženýrství a výpočetní finance".

Podrobné informace dostupné na: <http://cran.r-project.org/web/packages/fGarch/fGarch.pdf>

Další použité balíčky lze dohledat na stránkách: <http://cran.r-project.org/>.

V následující části budou uvedeny funkce vytvořené v programu R, které jsou použity v hlavním výpočtovém skriptu a které již v dalších částech budou jen volány. V aplikační části této práce budou uváděny další R kódy s příkazy a funkcemi, které byly využity k řešení, výpočtu nebo sestavení grafu. Pokud budou pro další výsledky použity stejné příkazy nebo funkce, které již byly uvedeny a jsou tedy použity pouze pro jiná data, pak už R kód nebude uváděn. Celý zdrojový kód pro získání všech výsledků lze nalézt na CD, které je přílohou této práce.

Funkce pro vytvoření Q-Q grafu:

R code:

```

1 # Q-Q graf
2 QQplot <- function(y){
3   x <- qnorm(ppoints(y))
4   d <- data.frame(x = x, y = sort(y))
5   xx <- qnorm(c(0.25, 0.75))
6   yy <- quantile(y, c(0.25, 0.75))
7   slope <- diff(yy) / diff(xx)
8   int <- yy[1L] - slope * xx[1L]
9   p <- ggplot(d, aes(x = x, y = y)) + geom_point() + geom_hline(yintercept = 0)
10      + xlab('Theoretical Quantiles') + ylab('Sample Quantiles')
11      + geom_abline(intercept = int, slope = slope, color = 'blue')
12   return(p)
13 }
```

Funkce pro vytvoření ACF grafu:

R code:

```

1 # ACF
2 qacf <- function(x, conf.level = 0.95, max.lag = NULL, min.lag = 0, title = "") {
3   ciline <- qnorm((1 - conf.level) / 2) / sqrt(length(x))
4   bacf <- acf(x, plot = FALSE, lag.max = max.lag, type = "correlation")
5   bacfdf <- with(bacf, data.frame(lag, acf))
6   if (min.lag > 0) { bacfdf <- bacfdf[-seq(1, min.lag), ] }
7   significant <- (abs(bacfdf[, 2]) > abs(ciline))^2
8   bacfdf <- cbind(bacfdf, significant)
9   q <- qplot(lag, acf, data = bacfdf, geom = "bar", stat = "identity",
10            position = "identity", ylab = "Autocorrelation", main = title, fill = factor(significant))
11     + geom_hline(yintercept = -ciline, color = "blue", size = 0.2)
12     + geom_hline(yintercept = ciline, color = "blue", size = 0.2)
13     + geom_hline(yintercept = 0, color = "red", size = 0.3)
14     + scale_fill_hue(name = paste("Significant_at_the\n", conf.level, "level"),
15                      breaks = 0:1, labels = c("False", "True"))
16     + theme(legend.justification = c(1,1), legend.position = c(1,1))
17   return(q)
18 }

```

Funkce pro vytvoření PACF grafu:

R code:

```

1 # PACF
2 qacf <- function(x, conf.level = 0.95, max.lag = NULL, min.lag = 0, title = "") {
3   ciline <- qnorm((1 - conf.level) / 2) / sqrt(length(x))
4   bacf <- acf(x, plot = FALSE, lag.max = max.lag, type = "correlation")
5   bacfdf <- with(bacf, data.frame(lag, acf))
6   if (min.lag > 0) { bacfdf <- bacfdf[-seq(1, min.lag), ] }
7   significant <- (abs(bacfdf[, 2]) > abs(ciline))^2
8   bacfdf <- cbind(bacfdf, significant)
9   q <- qplot(lag, acf, data = bacfdf, geom = "bar", stat = "identity",
10            position = "identity", ylab = "Autocorrelation", main = title, fill = factor(significant))
11   q <- q + geom_hline(yintercept = -ciline, color = "blue", size = 0.2)
12     + geom_hline(yintercept = ciline, color = "blue", size = 0.2)
13     + geom_hline(yintercept = 0, color = "red", size = 0.3)
14     + scale_fill_hue(name = paste("Significant_at_the\n", conf.level, "level"), breaks = 0:1, labels = c("False", "True"))
15     + theme(legend.justification = c(1,1), legend.position = c(1,1))
16   return(q)
17 }

```

Funkce pro vytvoření grafu histogramu, jádrového odhadu funkce hustoty a hustoty odhadnutého teoretického normálního rozdělení:

R code:

```

1 NKH <- function(x, title = "") {
2   # Jádrový odhad distribuční funkce
3   kernel = density(x, bw = "nrd0", adjust = 1, kernel = c("gaussian"), weights = NULL, window = kernel, width,
4                  give.Rkern = FALSE, cut = 3, na.rm = FALSE)
5
6   xfit <- seq(min(x), max(x), length=length(kernel$y))
7   yfit <- dnorm(xfit, mean = mean(x), sd = sd(x))
8
9   df <- data.frame(type = factor( rep(c("kernel_(gaussian)", "Teoretical_normal"),
10                                each=length(kernel$x))), LogP = c(kernel$y, yfit), axisX = c(kernel$x, kernel$x))
11
12 # vykreslení jadroveho odhadu pomoci ggplot2
13 H = ggplot(data.frame(x = x), aes(x = x))
14 H = H + geom_histogram(aes(y = ..density..), colour=I('#FFFFFF'), binwidth = density(x)$bw)
15     + xlab('Data(LogProfit)') + ylab('Density')
16     + geom_line(data = df, aes(x=axisX, y=LogP, colour=type), size = 1)
17     + geom_polygon(data = df, aes(x=axisX, y=LogP, fill=type, group=type), alpha = 0.2)
18     + theme(legend.justification = c(1,1), legend.position = c(1,1))
19   return(H)
20 }

```

V práci je také použita vytvořená funkce pro vykreslení histogramu, jádrového odhadu funkce hustoty a hus-

toty odhadnutého teoretického studentovo rozdělení. Tato funkce je podobná jako předcházející, ale je modifikovaná právě pro studentovo rozdělení:

R code:

```

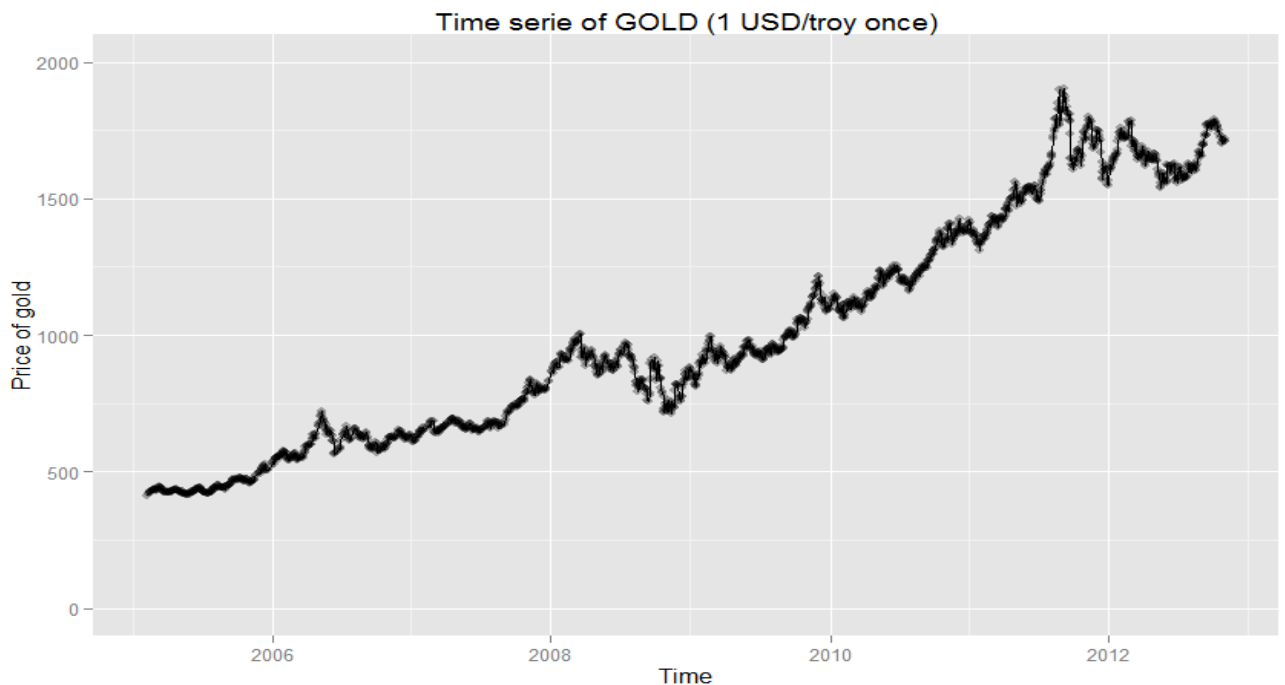
1 NKHst <- function(x, t1, t2, t3, title = "") {
2   # Jádrový odhad distribuční funkce
3   kernel = density(x, bw = "nrd0", adjust = 1, kernel = c("gaussian"), weights = NULL, window = kernel,
4                 give.Rkern = FALSE, cut = 3, na.rm = FALSE)
5
6   xfit <- seq(min(x), max(x), length=length(kernel$y))
7   yfit <- dstd(xfit, t1, t2, t3)
8
9   df <- data.frame(type = factor(rep(c("kernel_(gaussian)", "Theoretical_student"), each=length(kernel$x))),
10                    LogP = c(kernel$y, yfit), axisX = c(kernel$x, kernel$x))
11
12   H = ggplot(data.frame(x = x), aes(x = x))
13   H = H + geom_histogram(aes(y = ..density..), colour="I(#FFFFFF)", binwidth = density(x)$bw)
14   H = H + xlab("Data(LogProfit)") + ylab("Density")
15   H = H + geom_line(data = df, aes(x = axisX, y = LogP, colour = type), size = 1)
16   H = H + geom_polygon(data = df, aes(x = axisX, y = LogP, fill = type, group = type), alpha = 0.2)
17   H = H + theme(legend.justification=c(1,1), legend.position=c(1,1)) # pozice legendy
18
19   return(H)
20 }
```

6 Aplikační část

Doposud probranou teorii lze aplikovat na vysokofrekvenční časové řady, kterými jsou např. devizové kurzy, komodity, akcie a jiné. Aplikační část této práce bude zaměřena na zpracování vývoje komodity zlata²⁶. Investiční zlato je kupované jako výhodná investice na ochranu proti znehodnocení úspor. Obchoduje se zejména jako zlaté slitky, cihly nebo mince s vysokou ryzostí kovu (až 999/1000), přičemž v ČR je dodávka investičního zlata osvobozena od DPH²⁷.

Investiční zlato se obchoduje ve standardizovaných hmotnostech, jednou z nich je například 1 oz (trojská unce = 31,103 g). Zlato patří k jedné z neatraktivnějších investic, protože pro mnohé investory má nemalý význam, jako diverzifikační nástroj či jako uchovatel hodnoty v dobách nejistoty nebo při očekávání vysoké inflace, kterou by zlato mělo zahrnout ve své ceně. V [15] je zpracována analýza této řady a možné vlivy na vývoj ceny zlata, které se však ukázaly jako nevýznamné.

Data ceny zlata jsou stažena z [11] a obsahují údaje od 9.2.2005 do 31.10.2012. Zahrnují tak celkem 1858 pozorování s denní frekvencí, ve smyslu čistě obchodních dnů. Kvůli existenci neobchodních dnů²⁸ nemohou být časové okamžiky pořízení jednotlivých pozorování stejně vzdáleny a proto bude vždy existovat nakumulovaná informace po průběhu neobchodních dnů, která by teoreticky měla ovlivnit počátek obchodování zvýšenou volatilitou. Ve skutečnosti ovšem není zvykem zveřejňovat kurzotvorné informace v neobchodních dnech a tento efekt by tak měl být potlačen [16].



Obr. 3: Vývoj časové řady zlata za období 9.2.2005-31.10.2012

²⁶Informace o investičním zlatě: http://wiki.kurzy.cz/Investicni_zlato/.

²⁷Podrobnosti v §92 „Zvláštní režim pro investiční zlato“ zákona o DPH č. 235/2004 Sb. <http://zakony.kurzy.cz/235-2004-zakon-o-dani-z-pridane-hodnoty-dph/paragraf-92/>.

²⁸Svátky či víkendy.

R code:

```

1 # Vykreslení časové řady
2 qplot(dates, serie, geom = "line", xlab = "Time", ylab = "Price_of_gold",
3       main = "Time_serie_of_GOLD_(1_USD/troy_ounce)", ylim = c(0,2000))
4 + geom_point(colour = "black", size = 2, alpha = 0.3)

```

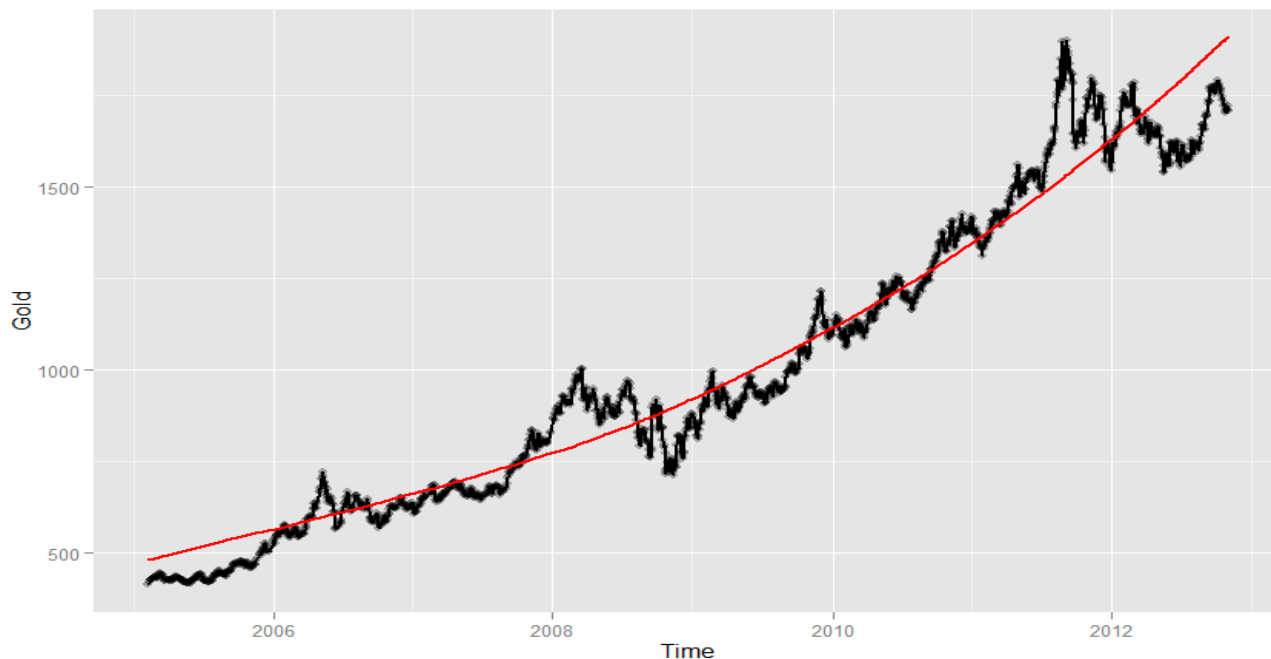
6.1 Základní popis a předzpracování dat

V této části budou popsány základní charakteristiky zpracovávaných dat ceny zlata.

Rok	Min.	25% Kvantil	Medián	Průměr	75% Kvantil	Max.	Počet	Sm. odchylka
2005	412.9	426.8	438.2	446.5	463.0	527.9	210	25.921
2006	528.0	571.4	612.0	606.5	634.6	721.4	220	40.336
2007	607.2	658.8	677.1	700.6	743.2	837.2	208	59.356
2008	712.3	828.0	884.9	872.3	921.8	1008.0	237	68.617
2009	812.7	921.3	949.4	975.1	1016.0	1216.0	258	86.073
2010	1063.0	1139.0	1217.0	1229.0	1319.0	1426.0	254	99.088
2011	1312.0	1436.0	1546.0	1574.0	1699.0	1903.0	257	149.915
2012	1539.0	1613.0	1655.0	1663.0	1721.0	1792.0	214	65.962
2005-2012	412.9	657.8	933.2	1025.0	1382.0	1903.0	1858	418.524

Tab. 1: Základní údaje časové řady cen zlata

Časová řada ceny zlata je nestacionární, má totiž určitý trendový vývoj v čase a jak je z obrázku 4 vidět, mohlo by se jednat o exponenciální trend. Pomocí autokorelační funkce (ACF), resp. parciální autokorelační funkce (PACF)²⁹, lze potvrdit, že časová řada je nestacionární.



Obr. 4: Vývoj časové řady zlata a odhad exponenciálního trendu

²⁹Podle předchozí kapitoly 2.

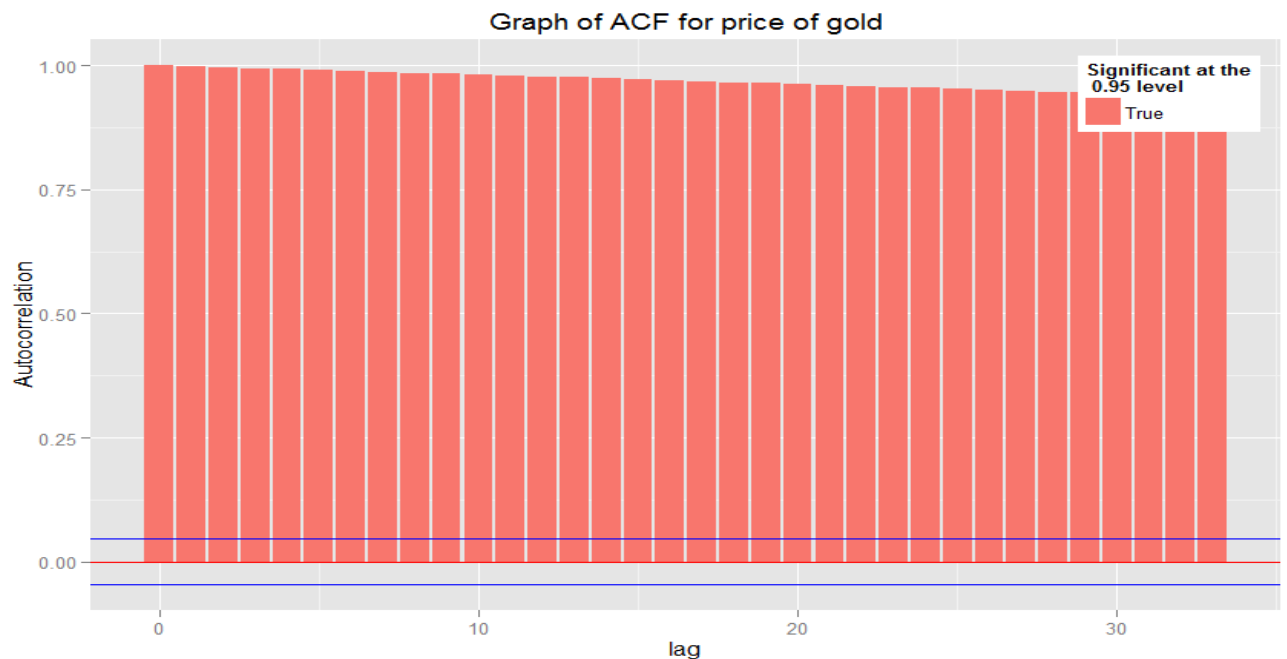
R code:

```

1 # Získání základních statistických charakteristik
2 summary(serie) # informace o datech
3 sqrt(var(serie)) # informace o směrodatné odchylce
4 length(serie) # informace o délce řady
5
6 # Řada cen zlata + exponenciální trend
7 ts = serie
8 tr <- seq(from = 1, to = length(ts), by = 1) #cas
9 yt <- ts * log(ts)
10 xt1 <- tr
11 xt2 <- tr * ts
12 summary(a<-lm(formula = yt~xt1 + xt2 -1))
13 exp(a$coef[1])
14 exp(a$coef[2])
15 expT = exp(a$coef[1])*(exp(a$coef[2])^tr)
16
17 df <- data.frame(x = dates, trend = expT)
18 G = ggplot(data.frame(Time = dates, Gold = serie), aes(Time,Gold), ylim = c(0,2000))
19   + geom_line(colour = "black", size = 1)
20   + geom_point(colour = "black", size = 2, alpha = 0.3)
21   + geom_line(aes(dates, trend), data = df, colour = "red", size = 1)
22 print(G)

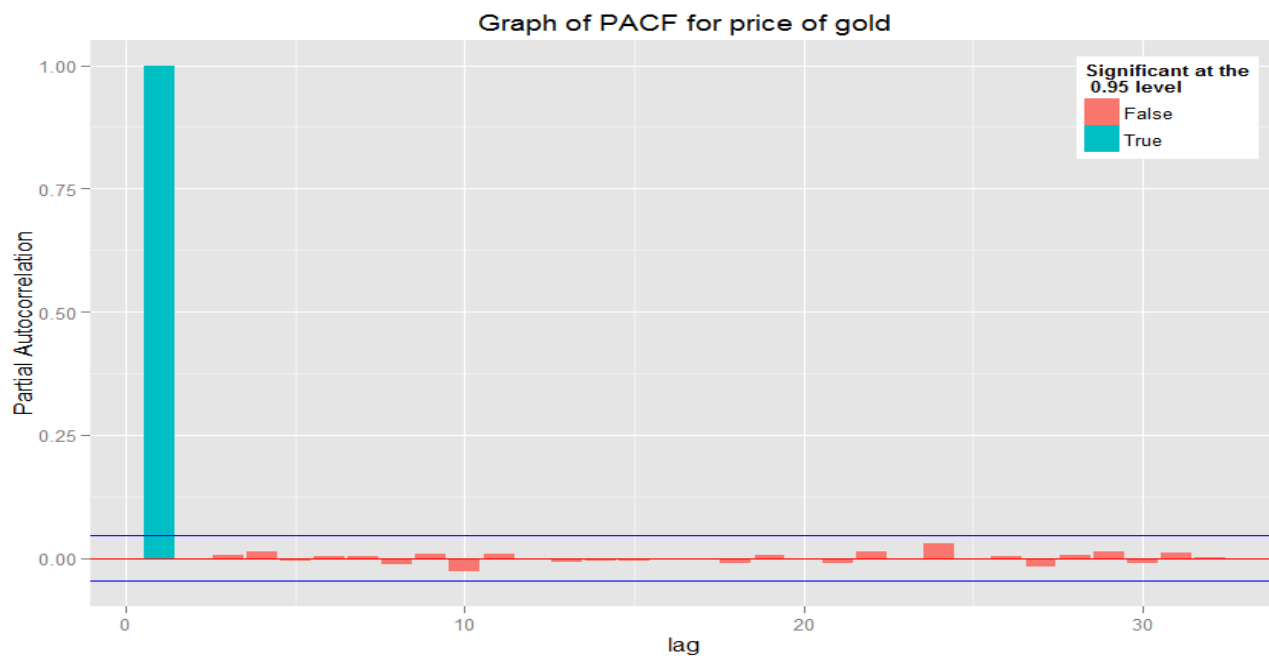
```

Charakter ACF na obrázku 5 podporuje tvrzení, že řada cen zlata je nestacionární. Graf ACF ukazuje velmi pozvolně klesající průběh autokorelační funkce, resp. koeficientů korelace, které přesahují hranici statistické významnosti³⁰.



Obr. 5: Graf ACF pro řadu cen zlata

³⁰Pravděpodobnostní 95% interval $(-1.96/\sqrt{n}, 1.96/\sqrt{n})$ uvedený v kapitole 2.



Obr. 6: Graf PACF pro řadu cen zlata

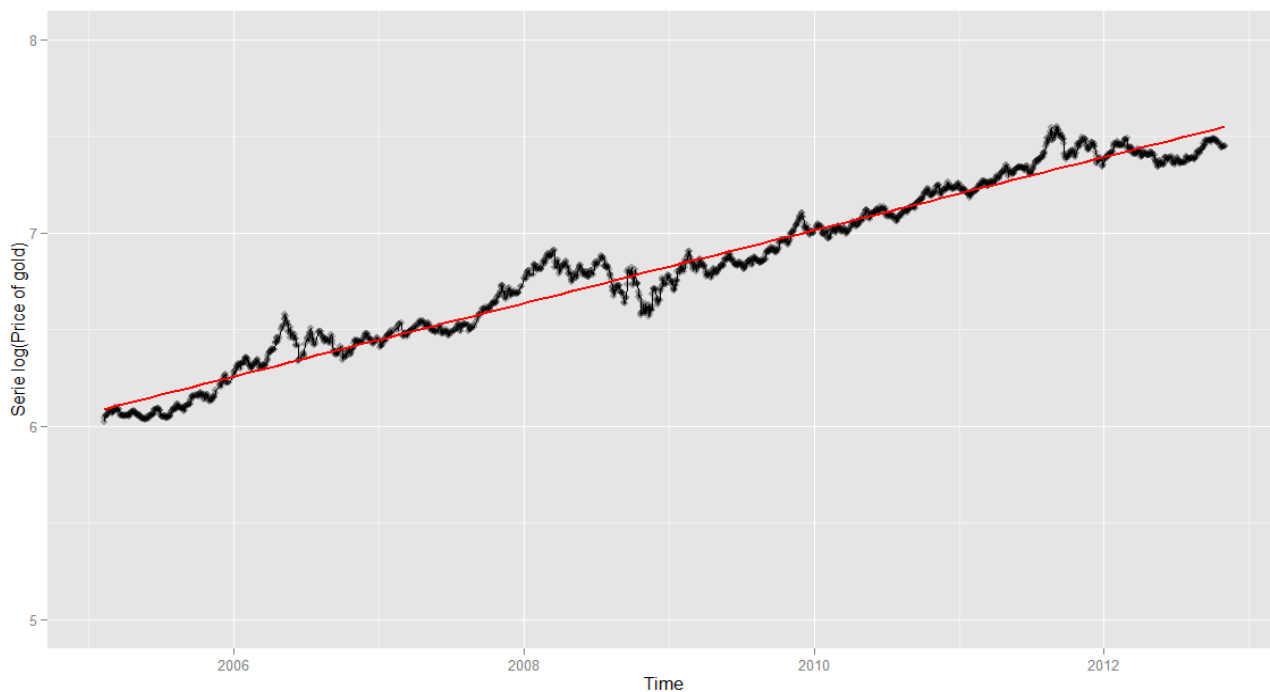
R code:

```

1 #ACF a PACF původních dat
2 ACF1 = acf(serie, lag.max = NULL, type = "correlation",
3           plot = FALSE, na.action = na.fail)
4 PACF1 = acf(serie, lag.max = NULL, type = "partial",
5            plot = FALSE, na.action = na.fail)
6 LAGACF1 = length(ACF1$lag) # nastavení délky zpoždění
7 LAGPACF1 = length(PACF1$lag) # nastavení délky zpoždění
8
9 # vykreslení ACF pomocí ggplot2
10 source("../Functions//qacf.r")
11 G = qacf(serie, conf.level = 0.95, max.lag = LAGACF1,
12         title = "Graph_of_ACF_for_price_of_gold")
13 print(G)
14
15 # vykreslení PACF pomocí ggplot2
16 source("../Functions//qpacf.r")
17 G = qpacf(serie, conf.level = 0.95, max.lag = LAGPACF1,
18         title = "Graph_of_PACF_for_price_of_gold")
19 print(G)

```

Nestacionární finanční časovou řadu ceny zlata lze podle vzorce³¹ $r_t = \ln P_t - \ln P_{t-1} = p_t - p_{t-1}$ transformovat na stacionární. Získáme tak řadu logaritmických výnosů, která má v tomto případě intuitivnější význam než jednoduchá diferencovaná řada. Nejprve tedy logaritmováním linearizujeme předpokládaný exponenciální trend a dostáváme tak řadu logaritmů cen zlata, která v čase vypadá následovně.



Obr. 7: Vývoj logaritmů cen zlata a odhad lineárního trendu

R code:

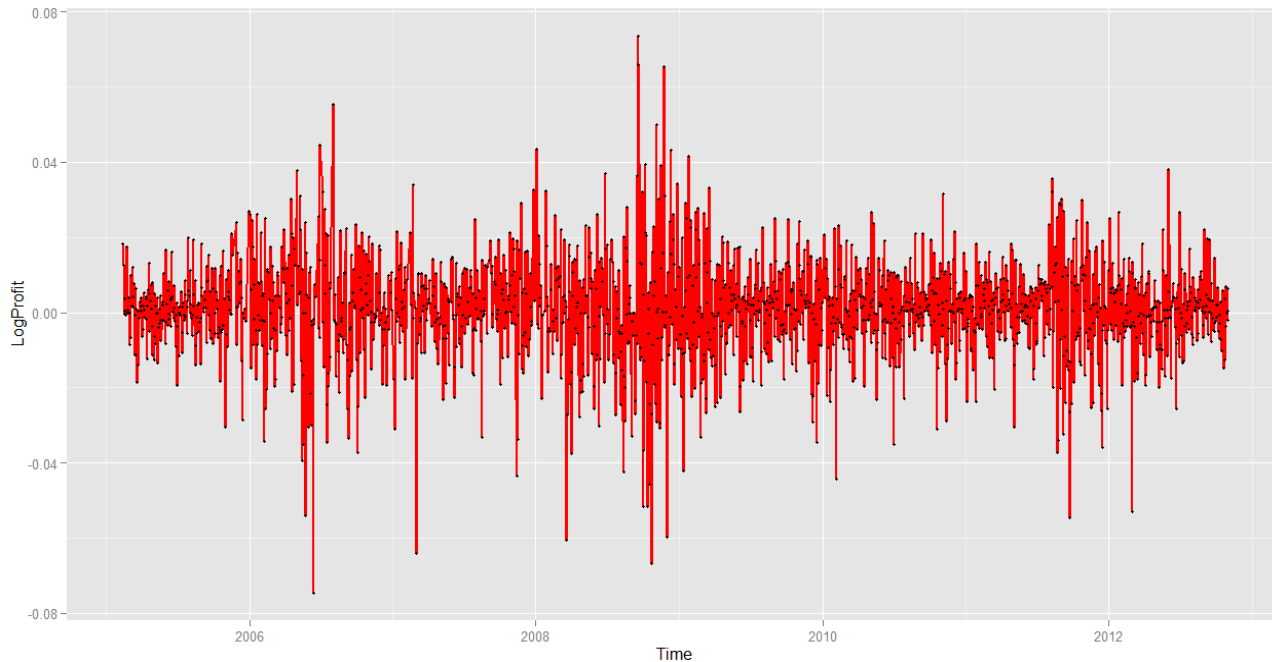
```

1 # Logaritmování původní časové řady
2 logSerie = log(serie) # logaritmy dat
3
4 # Graf logaritmů cen zlata + lineární trend
5 G = ggplot(data.frame(Time = dates, serie = logSerie), aes(Time, serie))
6   + geom_point(colour = "black", size = 2, alpha = 0.3)
7   + geom_line()
8   + geom_smooth(method = "lm", se = FALSE, colour = "red", size = 1)
9   + scale_fill_discrete("") + ylim(c(5,8)) + ylab("Serie_log(Price_of_gold)")
10 print(G)

```

³¹Uvedený v kapitole 2.2.

Po logaritmické transformaci lze diferencí 1. řádu získat logaritmy výnosů původní časové řady, které v čase vypadají takto:



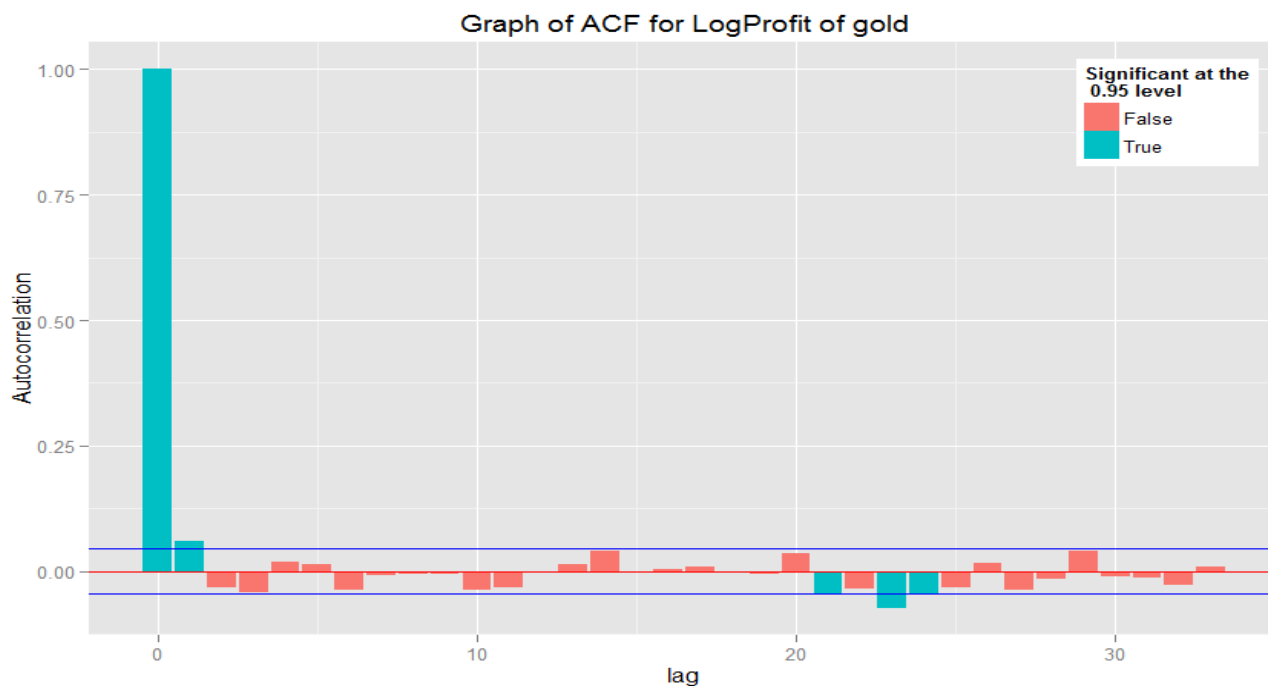
Obr. 8: Vývoj časové řady logaritmů výnosů ceny zlata

R code:

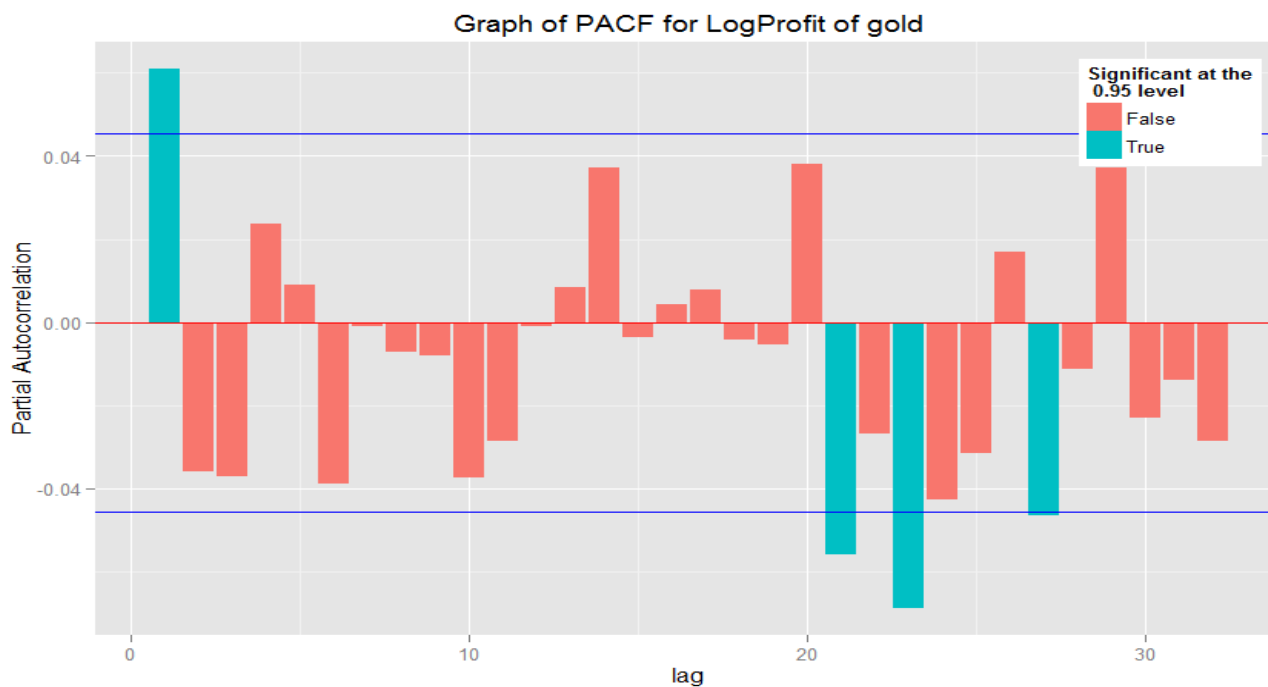
```

1 # Dokončení převodu na stacionární časovou řadu (diference 1. řádu)
2 lags = 1 # nastavení řádu difference
3 LogProfit = diff(logSerie, lag=lags) # výnosy logaritmů dat
4
5 # Graf logaritmů výnosů
6 G = ggplot(data.frame(Time=dates[2:length(dates)],LogProfit), aes(Time,LogProfit))
7   + geom_line(colour = "red", size = 1)
8   + geom_point(colour = "black", size = 1)
9 print(G)

```



Obr. 9: Graf ACF logaritmů výnosů časové řady ceny zlata



Obr. 10: Graf PACF logaritmů výnosů časové řady ceny zlata

Řada logaritmů výnosů zlata (dále už jen výnosů) můžeme již považovat za stacionární, což je patrné z přecházejících obrázků 8, 9, 10. Autokorelaci již neprokazuje ani sestavená ACF a PACF, kde je řada výnosů charakteristická nesystematickými výkyvy kolem nulové hodnoty. Zároveň můžeme na obrázku 8 pozorovat proměnlivou volatilitu (variabilitu), která naznačuje měnící se rozptyl.

6.2 Předpoklad normality

Modelování časových řad je často svazováno s předpokladem, že data (resp. residua dat) pocházejí z normálního rozdělení (nebo alespoň se k němu blíží).

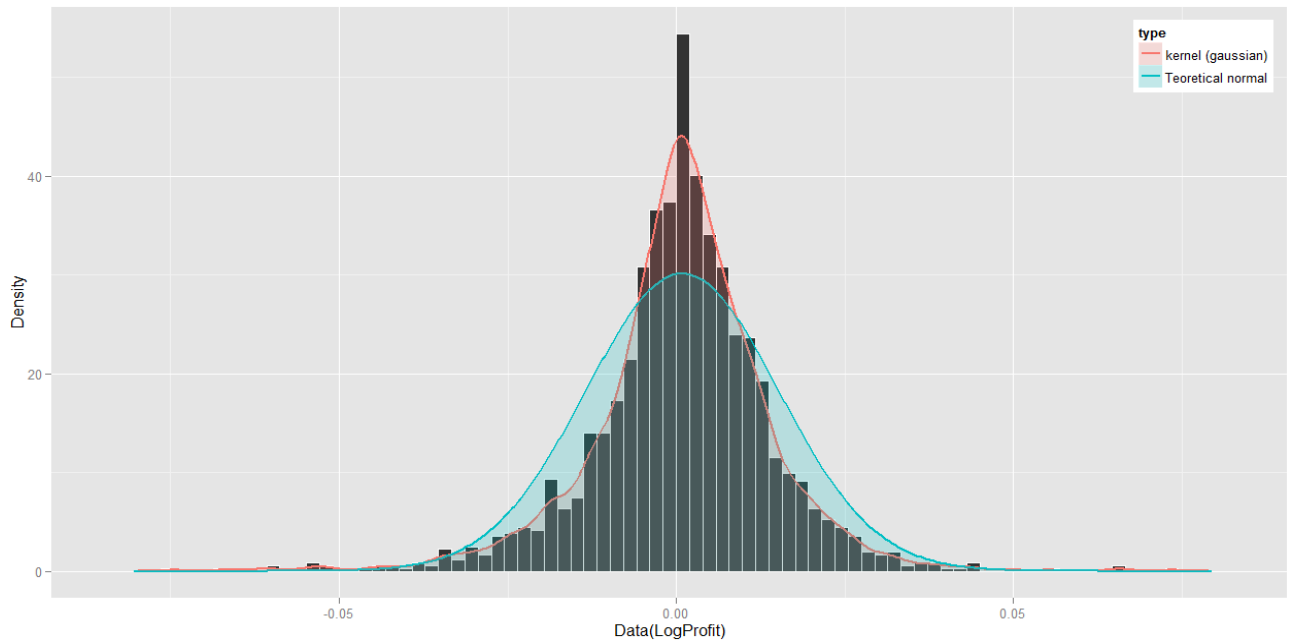
Rok	Počet	Průměr	Sm. odchylka	Šikmost	Špičatost	Min.	Max.
2005	209	0.00109	0.00836	-0.15630	1.54873	-0.03036	0.02695
2006	220	0.00095	0.01624	-0.56442	2.38897	-0.07450	0.05535
2007	208	0.00128	0.01201	-1.10897	4.68733	-0.06414	0.03391
2008	237	0.00024	0.02001	0.04469	1.83305	-0.06683	0.07355
2009	258	0.00085	0.01266	-0.13740	0.51992	-0.04202	0.04166
2010	254	0.00102	0.01023	-0.56905	2.11572	-0.04424	0.03167
2011	257	0.00037	0.01264	-0.52707	1.67968	-0.05448	0.03555
2012	214	0.00045	0.00967	-0.41322	4.83308	-0.05293	0.03802
2005-2012	1857	0.00077	0.01324	-0.35826	3.76726	-0.07450	0.07360

Tab. 2: Statistické charakteristiky řady výnosů

Liliefors test	$D = 0.076$	p-value < $2.2 \cdot 10^{-16}$
Shapiro-Wilk test	$W = 0.9528$	p-value < $2.2 \cdot 10^{-16}$
Jarqu-Bera test	$\chi^2 = 1142.172$	p-value < $2.2 \cdot 10^{-16}$

Tab. 3: Výsledky testů normality pro řadu výnosů

Tabulka 3 ukazuje testy normality popsané v kapitole 2.2, které neprokazují normalitu výnosů. Odlišnost od normálního rozdělení ukazuje také tabulka 2, kde šikmost a špičatost vychází odlišně od normálního rozdělení (a to i v dílčích rocích), které má šikmost rovnou nule a špičatost rovnou třem.

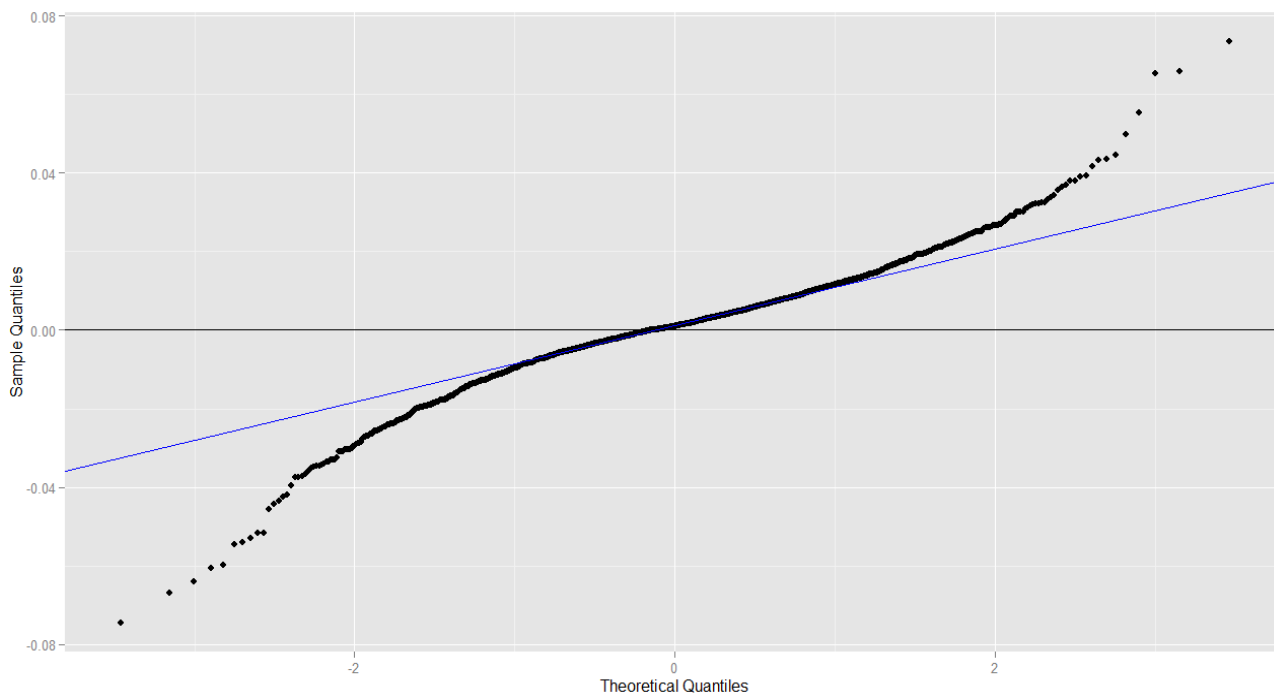


Obr. 11: Histogram výnosů, gaussovský jádrový odhad hustoty rozdělení a teoretické normální rozdělení³²

³²Základní informace o odhadech jádrových hustot lze nalézt např. v disertační práci Mgr. Martina Řezáče, Ph.D., dostupné z: http://is.muni.cz/th/20411/prif_d/disertace.pdf nebo také v diplomové práci Karolíny Mladé, dostupné z: http://is.muni.cz/th/270467/prif_m/diplomkaMlada.pdf.

Na předchozím obrázku je pak histogram výnosů a odhadnuté rozdělení výnosů pomocí gaussovského jádrového odhadu a také teoretické normální rozdělení s parametry odhadnutými z dat výnosů. Z tohoto porovnání jsou patrné vlastnosti finančních časových řad popsané v kapitole 2.2.

Histogram³³ na obrázku 11 ukazuje větší špičatost oproti normálnímu rozdělení a těžké konce³⁴, které ukazuje také Q-Q graf na následujícím obrázku 12. Tyto obrázky a tabulky demonstrují, jak se data výnosů liší od normálního rozdělení s odhadnutými parametry střední hodnoty a rozptylu z dat výnosů, ale také indikují použití modelů volatility.



Obr. 12: Q-Q graf výnosů

R code:

```

1 # Testy normality
2 x = LogProfit
3
4 TSstat = FinTS.stats(x); TSstat
5 LLT = lillie.test(x); LLT
6 SWT = shapiro.test(x); SWT
7 JBT = jarque.bera.test(x); JBT
8
9 # Grafické testy normality logaritmů výnosů (LogProfit)
10
11 # Jadrový odhad distribuční funkce
12 source('./functions//NKH.r')
13 G = NKH(LogProfit)
14 print(G)
15
16 # QQ graf logaritmů výnosů
17 source('./functions//QQplot.r')
18 G = QQplot(LogProfit)
19 print(G)

```

³³Četnosti histogramu vyplývají z funkce v R. Zde je implementováno rule-of-thumb (pravidlo palce) pro volbu šířky pásma gaussovského jádrového odhadu hustoty. (Silvermanovo "pravidlo", Silverman (1986, strana 48, kapitola (3.31))).

³⁴Nebo také tlusté konce, anglicky: heavy tails, fat tails.

6.3 Fáze 1: Odhad úrovněvého modelu

V dalších částech již začíná samotné modelování časové řady, které se bude více méně řídit podle kapitoly 3.3. Nejprve se pokusíme z dat výnosů odstranit lineární závislosti pomocí tzv. „úrovněvého modelu“. Z předcházejících obrázků lze vidět, že úroveň časové řady se v čase mění a proto lze použít jeden z lineárních modelů ARMA(p,q)³⁵, které jsou založeny na podmínce, že se sice podmíněná střední hodnota mění v čase, podmíněný rozptyl je však konstantní, což neodpovídá realitě. Tento problém lze vyřešit právě uvažovanými modely volatility v další části. Podstatným rysem této koncepce je, že se nemění původní požadavek normality.

Opodstatnění použití ARMA modelu vyplývá již z obrázku 10 na straně 34, kde PACF výnosů ukazuje porušení nekorelovanosti už při $k = 1$ (dále pak $k = 21, 23, 27$). Tento fakt lze ještě ověřit například pomocí Portmentau testu. Řada výnosů obsahuje 1857 hodnot a přirozený logaritmus tohoto počtu je 7.5267. Z toho plyne použití Portmentau testu pro $k = 8$ korelačních koeficientů. Výsledkem testu je p -hodnota rovna 0.04551 a to je na hranici zamítnutí (na hladině významnosti $\alpha = 5\%$). Hypotézu o nekorelovanosti tedy zamítáme a pomocí ARMA modelu se pokusíme závislosti odstranit.

R code:

```
1 # Test autokorelace výnosů (LogProfit)
2 x = LogProfit
3 # Nastavení hodnoty zpoždění podle vzorce log(n)
4 k = round(log(length(x)))
5 # Portmentau test – hypotéza o nekorelovanosti
6 Q = Box.test(x, lag = k, type = 'Ljung'); Q
```

Řád modelu ARMA lze určit několika způsoby. Jedním z nich je určení řádu podle tvaru ACF a PACF dat výnosů. Možné tvary ACF a PACF a k nim určené řady modelu jsou³⁶:

- ARMA(1,0): ACF exponenciálně klesá k nule, PACF má pouze jeden vrchol na pozici 1.
- ARMA(2,0): ACF má sinusový průběh nebo několik exponenciálních poklesů, PACF má vrcholy na pozicích 1 a 2. Na dalších pozicích nejsou již korelace.
- ARMA(0,1): ACF má pouze jeden vrchol na pozici 1, PACF jde exponenciálně k nule.
- ARMA(0,2): ACF má dva vrcholy na pozici 1 a 2, bez dalších korelací, PACF má sinusový průběh, nebo exponenciálně tlumený průběh.
- ARMA(1,1): ACF má exponenciální pokles začínající na pozici 1, PACF má oscilující pokles začínající na pozici 1.
- ARMA(1,1): ACF má oscilující pokles začínající na pozici 1, PACF má exponenciální pokles začínající na pozici 1.
- ARMA(p,q): ACF přímý nebo oscilující pokles k nule začínající na pozici q . PACF přímý nebo oscilující pokles k nule začínající na pozici p .

Pokud se podíváme na ACF a PACF na obrázku 9 a 10 (na straně 34), lze vidět, že tvary těchto funkcí odpovídají spíše modelu ARMA(1,1).

³⁵Často je využíván pouze jednoduchý model $X_t = \mu + \varepsilon_t$, zvláště pak u finančních časových řad.

³⁶Například zde: www.statsoft.cz/file1/PDF/newsletter/2012_07_23_ARMA.pdf.

V následující tabulce budou shrnuty výsledky odhadnutých ARMA(p,q) modelů. Požadujeme co nejjednodušší model, proto bude využit model maximálně do řádu ARMA(1,1). Kritéria $\ln L(\hat{\varphi})$, AIC a BIC jsou spočtena podle kapitoly 3.7.1.

model	$\ln L(\hat{\varphi})$	AIC	BIC
ARMA(1,0)	5400.62	-10795.23	-10778.65
ARMA(0,1)	5400.78	-10795.97	-10778.97
ARMA(1,1)	5400.87	-10793.74	-10771.64

Tab. 4: Tabulka kritérií pro různé modely typu ARMA

Podle AIC kritéria lze volit model ARMA(0,1), který vede k nejnižší hodnotě AIC a stejně tak podle BIC kritéria.

Volba konečného modelu bude však z jiného pohledu. Protože se kritéria jednotlivých modelů nijak významně od sebe neliší, můžeme využít model ARMA(1,0), kvůli jeho snadné interpretovatelnosti. Obecný model ARMA(1,0) tedy vypadá takto:

$$X_t = \mu + \phi X_{t-1} + \varepsilon_t,$$

kde X_t jsou výnosy, $|\phi| < 1$ a $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ je proces bílého šumu se střední hodnotou 0 a rozptylem σ_ε^2 .

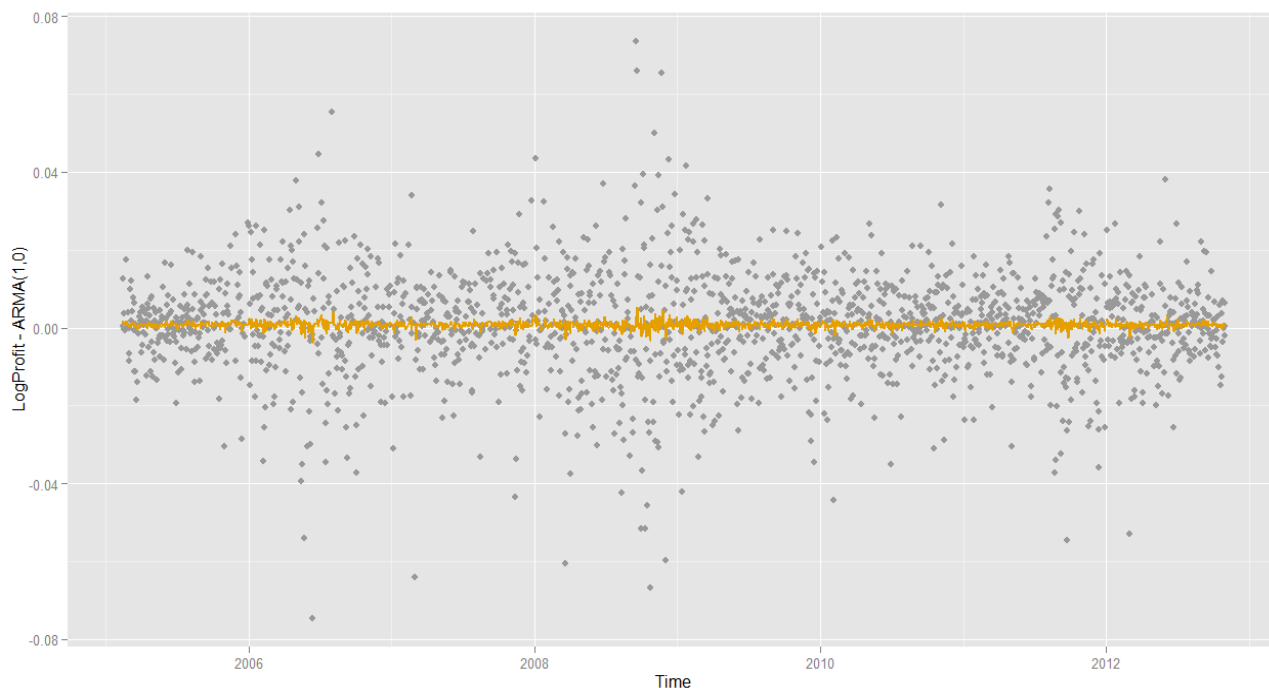
Odhady parametrů modelu ARMA(1,0) jsou³⁷:

	Estimate	Std. Error	t-value	$Pr(> t)$
μ	-0.0007192	0.0003071	-2.342	0.01920 *
ϕ	0.0608666	0.0231723	2.627	0.00862 **

Tab. 5: Tabulka odhadů parametrů pro ARMA(1,0) model

V předcházející tabulce si můžeme všimnout, že parametry sice jsou statisticky významné, ale dalo by se zde uvažovat o přímém použití modelů typu ARCH, díky nepřiliš vysoké významnosti parametrů u tohoto konkrétního modelu ARMA(1,0). V tomto případě bylo však rozhodnuto o ponechání modelu ARMA(1,0), který by mohl mít svá určitá ekonomická opodstatnění, která však v této práci nebudeme rozebírat ani podrobně zkoumat.

³⁷Významnost parametrů: '***' pro interval (0, 0.001), '**' pro interval (0.001, 0.01), '*' pro interval (0.01, 0.05), '.' pro interval (0.05, 0.1), '' pro interval (0.1, 1).



Obr. 13: Řada výnosů a odhad modelu ARMA(1,0)

R code:

```

1 #ARMA model
2 x = LogProfit
3 p = 1
4 q = 0
5
6 ARMAm = arma(x, order = c(p, q), include.intercept = TRUE)
7 summary(ARMAm)
8
9 # M = počet_parametrů+1, tj. konstanta+p+q+1
10 M = 1 + p + q + 1
11
12 # AIC kriterium
13 ARMAmAIC = summary(ARMAm10)$aic
14
15 # Max. log-likelihood kriterium
16 ARMAmLLH = (summary(ARMAm10)$aic - 2 * M) / (-2)
17
18 # BIC kriterium
19 ARMAmBIC = ((summary(ARMAm10)$aic - 2*M)) + M * log(length(x))
20
21 # Příprava dat pro graf
22 df <- data.frame(dates = dates[(max(p,q) + 2):length(dates)],
23                 LogP = LogProfit[(max(p,q) + 1 + lags):length(dates) - lags],
24                 Res = ARMAm$residuals[(max(p,q) + 1 + lags):length(dates) - lags],
25                 arma = ARMAm$fitted.value[(max(p,q) + 1 + lags):length(dates) - lags])
26
27 # Graf ARMA modelu
28 G = ggplot(df, aes(x = dates))
29   + xlab('Time')
30   + ylab('LogProfit--ARMA(1,0)')
31   + geom_point(aes(y = LogP), size = 2, colour = "#999999")
32   + geom_line(aes(y = arma), size = 1, colour = "#E69F00")
33 print(G)

```

6.3.1 Test residuí

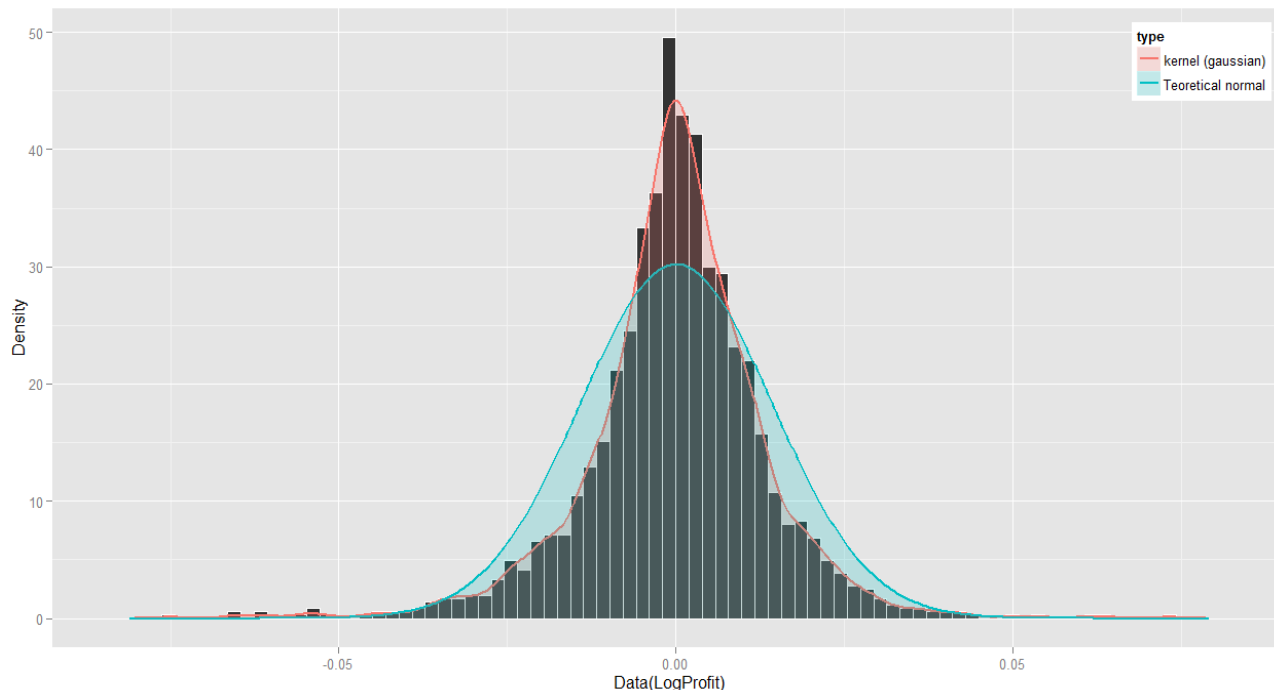
Po odhadu úrovnového modelu následuje test residuí $\hat{\varepsilon}_t$, které by v případě, že v modelu není objevena podmíněná heteroskedasticita, měly vést k procesu gaussovského bílého šumu. Následující údaje výsledků testů normality, autokorelace a homoskedasticity pro řadu $\{\hat{\varepsilon}_t\}$ potvrzují fakt zmíněný na začátku kapitoly 6.3, že tyto modely neuvažují nekonstantnost rozptylu. Díky tomu je vidět porušení předpokladu gaussovského bílého šumu residuí ARMA modelu.

Liliefors test	$D = 0.0734$	p-value < $2.2 \cdot 10^{-16}$
Shapiro-Wilk test	$W = 0.9545$	p-value < $2.2 \cdot 10^{-16}$
Jarqu-Bera test	$\chi^2(2) = 1086.963$	p-value < $2.2 \cdot 10^{-16}$

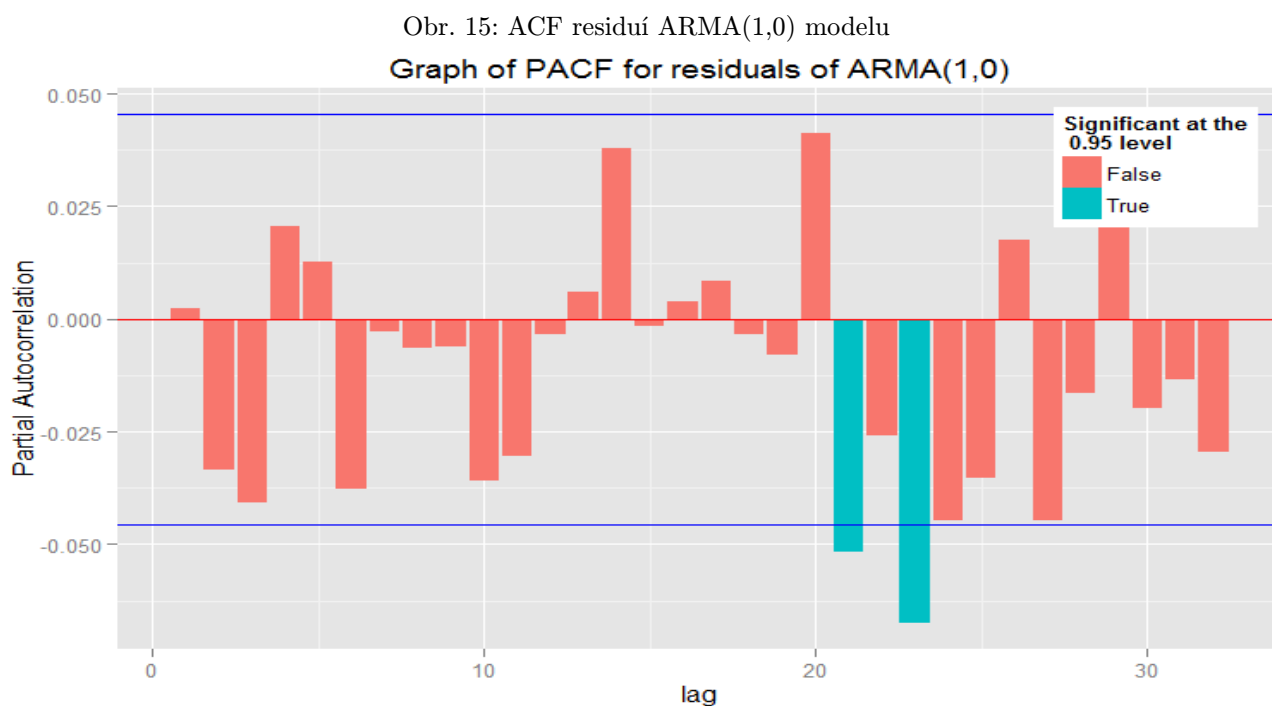
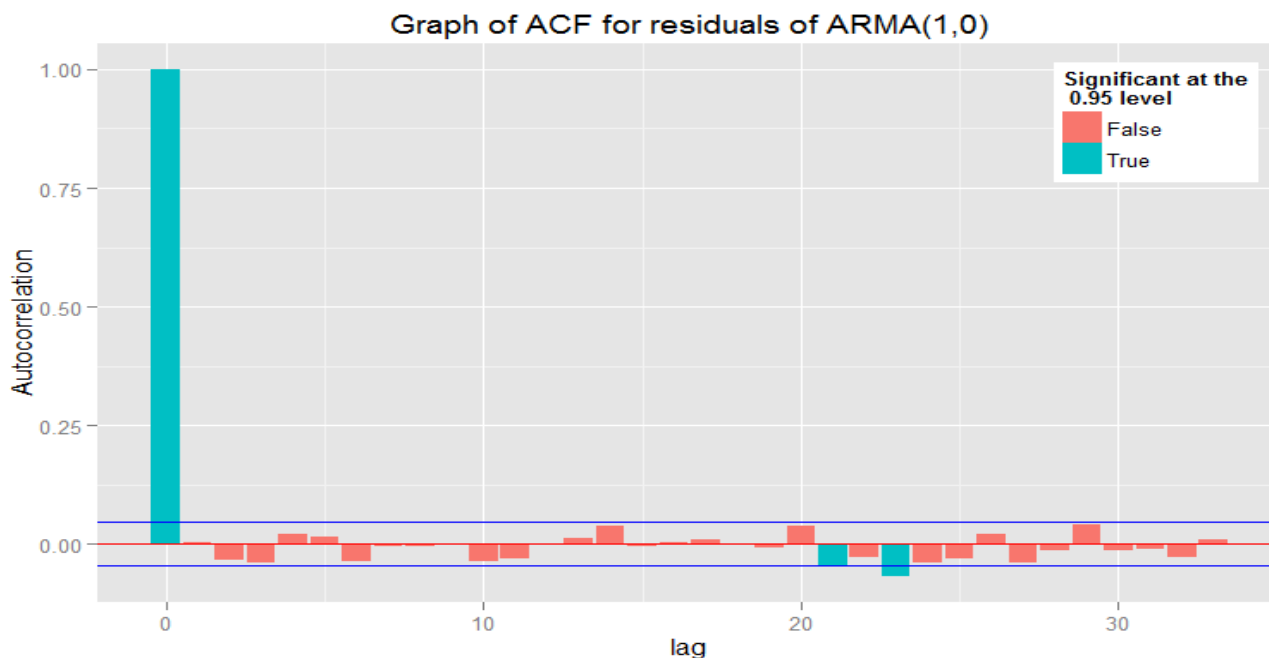
Tab. 6: Výsledky testů normality pro řadu $\{\varepsilon_t\}$

Testy normality residuí ARMA(1,0) modelu zamítají hypotézu o normalitě dat (tabulka 6). Na grafu 14 je vidět, že residua mají rozdělení špičatější a také s těžšími konci.

Při porovnání následujících obrázků ACF a PACF residuí modelu ARMA(1,0) a obrázků ACF a PACF logaritmu výnosů z kapitoly 6.1 (obrázek 9 a 10), můžeme vidět, že některé závislosti se podařilo odstranit. Pro ověření můžeme opět použít Portmanteau test z kapitoly 2.2.2. V tomto případě však s modifikací - testová statistika $Q(k)$ v případě modelu ARMA(p,q) konverguje v distribuci k χ^2 -rozdělení o $k - (p + q)$ stupních volnosti, tj. $\chi^2_{m-(p+q)}$ -rozdělení. Tím se zahrne odhadovaný počet parametrů [2]. Podle Portmanteauova testu ($k = 8 - (1 + 0) = 7$), který dává p -hodnotu 0.2295, tak lze přijmout hypotézu o nekorelovanosti residuí ARMA(1,0) modelu.



Obr. 14: Histogram residuí, gaussovský jádrový odhad hustoty rozdělení a teoretické normální rozdělení



Posledním krokem je pak test podmíněné heteroskedasticity³⁸ pomocí ARCH-LM testu (popsaný v kapitole 3.4), který vrátí téměř nulovou p -hodnotu $< 2.2 \cdot 10^{-16}$ ($\chi^2(8) = 148.54$). Zamítáme tedy hypotézu o nepřítomnosti podmíněné heteroskedasticity.

³⁸Testování tzv. ARCH efektu.

R code:

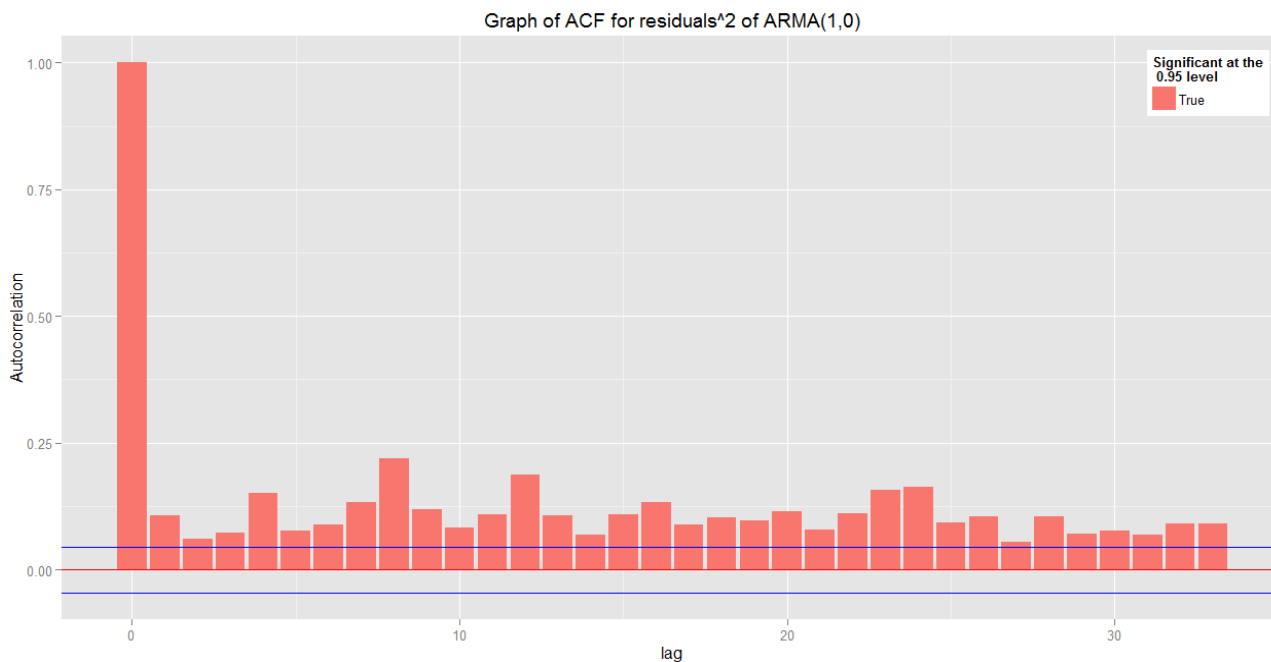
```

1 # Test podmíněné heteroskedasticity výnosů (LogProfit)
2 x = LogProfit
3 # Nastavení hodnoty zpoždění podle vzorce log(n)
4 k = round(log(length(x)))
5 # Arch-LM test
6 AT = ArchTest(x, lags = k); AT

```

6.4 Fáze 2: Odhad modelu volatility

Stejně jako v předcházející části, kde byla odstraněna měnící se úroveň časové řady výnosů, zde se budeme zabývat odstraněním variability z residuů ARMA modelu, která se také v čase mění. V této fázi je tedy potřeba zvolit model volatility. Jako první se nabízí využití modelu ARCH(q). Použití tohoto modelu však často vede k výběru modelu s vysokým řádem q . Proto lze využít modelu volatility typu GARCH(p,q). V praxi bývají často využity tyto modely s nízkým řádem. Z tohoto důvodu bude použit model GARCH nejvýše do řádu GARCH(2,2). K určení řádu modelu GARCH můžeme využít ACF a PACF kvadrátů residuů ARMA(1,0) modelu, které jsou na následujících obrázcích.



Obr. 17: ACF kvadrátů residuů ARMA(1,0) modelu



Obr. 18: PACF kvadrátů residuí ARMA(1,0) modelu

Předcházející obrázky naznačují řád modelu GARCH(1,1), avšak řád můžeme určit přesněji pomocí hodnoty Akaikeho (AIC) a Bayesova informačního kritéria (BIC) a hodnoty maximalizované logaritmické věrohodnostní funkce. Tyto hodnoty jsou uvedeny v následující tabulce pro různá p a q modelu GARCH.

	GARCH(1,1)	GARCH(2,1)	GARCH(1,2)	GARCH(2,2)
$\text{Ln } L(\hat{\varphi})$	5558.165	5559.668	5558.052	5559.992
AIC	-5.98617	-5.98671	-5.98497	-5.98598
BIC	-5.97724	-5.97480	-5.97306	-5.97110

Tab. 7: Hodnoty maximalizované věrohodnostní funkce, AIC a BIC modelů GARCH

Podle kritéria BIC v předcházející tabulce volíme model GARCH(1,1). Podle AIC bychom volili model GARCH(2,1) a pokud budeme dále zkoumat tento model, zjistíme, že v tomto případě má model parametry α_1 a α_2 statisticky nevýznamné (na hladině významnosti $\alpha = 5\%$, resp. $\alpha = 1\%$).

	Odhad	Sm. odchylka	t-stat.	p -hodnota
ω	$1.960 \cdot 10^{-06}$	$7.076 \cdot 10^{-07}$	2.770	0.00561 **
α_1	$1.883 \cdot 10^{-02}$	$1.484 \cdot 10^{-02}$	1.269	0.20443
α_2	$3.583 \cdot 10^{-02}$	$1.799 \cdot 10^{-02}$	1.992	0.04637 *
β_1	$9.344 \cdot 10^{-01}$	$1.145 \cdot 10^{-02}$	81.579	$< 2 \cdot 10^{-16}$ ***

Tab. 8: Odhadnuté parametry modelu GARCH(2,1)

To nám napovídá k použití modelu GARCH(1,1), jak už ukazuje kritérium BIC, které má větší vypovídací schopnost než AIC podle kapitoly 3.7.1. Volíme tedy model GARCH(1,1) jako finální model volatility. Kritéria však neříkají nic o tom, jaké předpoklady splňují residua modelu nebo jestli jsou parametry modelu statisticky významné.

	odhad	sm. odch.	t-stat.	p-hodnota
ω	$1.684 \cdot 10^{-06}$	$5.811 \cdot 10^{-07}$	2.897	0.00376 **
α_1	$4.792 \cdot 10^{-02}$	$7.197 \cdot 10^{-03}$	6.658	$2.77 \cdot 10^{-11}$ ***
β_1	$9.426 \cdot 10^{-01}$	$8.645 \cdot 10^{-03}$	109.027	$< 2 \cdot 10^{-16}$ ***

Tab. 9: Odhadnuté parametry modelu GARCH(1,1)

Podle tabulky 9 jsou parametry v pořádku, všechny jsou statisticky významné a pro model mají smysl.

R code:

```

1 # Odhad parametrů GARCH
2 GARCHm = garchFit(~ garch(1,1), data = ARMAe, include.mean = FALSE, cond.dist = "norm")
3 # Hodnota maximalizované logaritické věrohodnostní funkce
4 GARCHm@fit$llh
5 # Hodnota Akaikeho (AIC) informačního kriteria
6 GARCHm@fit$ics[1]
7 # Hodnota Bayesova (BIC) informačního kriteria
8 GARCHm@fit$ics[2]
9
10 summary(GARCHm)
11
12 # Všechny informace (testy residuí, odhady parametrů,...) jsou obsaženy v:
13 GARCHm@fit
14 summary(GARCHm)

```

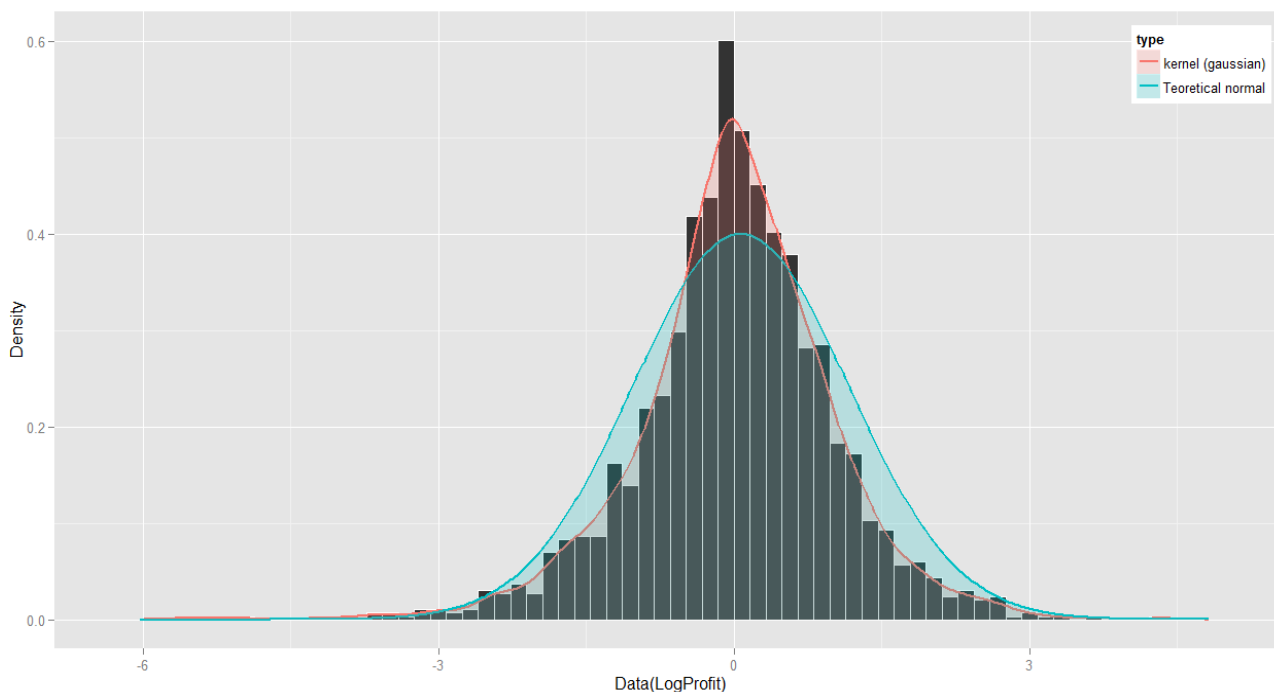
6.4.1 Test residuí a úprava modelu

Pro ověření vhodnosti modelu GARCH(1,1) lze využít standardizovaných reziduí $\hat{\epsilon}_t = \hat{\epsilon}_t / \sqrt{\hat{h}_t}$, pro které by mělo platit $\hat{\epsilon}_t \sim N(0, 1)$. Následující tabulka ukazuje, že model GARCH(1,1) uspěl při modelování volatility, resp. testy ukázaly, že v řadě residuí $\hat{\epsilon}_t$ již není podmíněná heteroskedasticita prokázána.

			Statistic	p-Value
Jarque-Bera Test	$\hat{\epsilon}$	χ^2	448.17190	0
Shapiro-Wilk Test	$\hat{\epsilon}$	W	0.97587	0
Ljung-Box Test	$\hat{\epsilon}$	$Q(10)$	10.79942	0.37336
Ljung-Box Test	$\hat{\epsilon}$	$Q(15)$	12.24231	0.66060
Ljung-Box Test	$\hat{\epsilon}$	$Q(20)$	15.59231	0.74157
Ljung-Box Test	$\hat{\epsilon}^2$	$Q(10)$	4.82837	0.90234
Ljung-Box Test	$\hat{\epsilon}^2$	$Q(15)$	8.87266	0.88408
Ljung-Box Test	$\hat{\epsilon}^2$	$Q(20)$	9.45254	0.97705
LM Arch Test	$\hat{\epsilon}$	TR^2	6.09475	0.91123

Tab. 10: Testy standardizovaných residuí $\{\hat{\epsilon}\}$ modelu GARCH(1,1)

Testy normality residuí (Jarque-Bera test, Shapiro-Wilk test popsaný v kapitole 2.2) však ukazují nesplnění podmínky normality. Tento fakt můžeme ilustrovat také na následujícím grafu histogramu residuí $\hat{\epsilon}_t$.



Obr. 19: Histogram residuí modelu GARCH(1,1), gaussovský jádrový odhad hustoty rozdělení a teoretické normální rozdělení

Jak je vidět rozdělení hodnot residuí sestaveného modelu GARCH(1,1) je špičatější s těžšími konci. V tomto případě by model nebyl vyhovující a v hodně případech se tento fakt přehlíží nebo je znám, ale přesto se model použije pro predikci.

Nyní máme tedy dvě možnosti dalšího postupu. Buď model zamítneme a pokusíme se upravit parametry, tak aby model splnil tyto podmínky³⁹ (nebo lze použít úplně jiný model) a nebo se pokusíme najít jiné rozdělení, kterým by se residua tohoto modelu mohla řídit. Provedeme tedy odhad modelu GARCH(1,1) znovu s předpokladem Studentova t -rozdělení (podle kapitoly 3.6).

	Odhad	Sm. odchylka	t-stat.	p -hodnota
ω	$1.240 \cdot 10^{-06}$	$6.207 \cdot 10^{-07}$	1.998	0.0457 *
α_1	$4.944 \cdot 10^{-02}$	$8.985 \cdot 10^{-03}$	5.502	$3.76 \cdot 10^{-8}$ ***
β_1	$9.467 \cdot 10^{-01}$	$9.239 \cdot 10^{-03}$	102.466	$< 2 \cdot 10^{-16}$ ***
ν	4.719	$0.572 \cdot 10^{-01}$	8.255	$2.22 \cdot 10^{-16}$ ***

Tab. 11: Odhadnuté parametry modelu GARCH(1,1) s předpokladem Studentova t -rozdělení pro residua $\{\hat{\epsilon}\}$

R code:

```

1 # Odhad modelu GARCH s předpokladem studentova t-rozdělení
2 GARCHm = garchFit(~ garch(1,1), data = ARMAe, include.mean = FALSE, cond.dist = "std")
3 # Residua modelu
4 GARCHe = residuals(GARCHm, standardize = TRUE)

```

³⁹V tomto případě je tato možnost vyloučena. Žádná ze základních úprav parametrů (přidávání, ubírání) ať už ARMA modelu nebo GARCH modelu nevede ke zlepšení či splnění podmínek normality standardizovaných residuí. Změnou počtu parametrů (nehledě na jejich významnost) dosáhneme minimálních změn.

V tomto případě byl společně s parametry modelu GARCH(1,1) odhadován také parametr ν , což je tzv. *shape* parametr, který je odhadován jako počet stupňů volnosti t -rozdělení. Všimněme si také, že tento parametr skutečně vyšel v mezích (3, 6) uvedených v kapitole 3.6.

			Statistic	p-Value
Ljung-Box Test	$\hat{\epsilon}$	$Q(10)$	10.81694	0.37196
Ljung-Box Test	$\hat{\epsilon}$	$Q(15)$	12.29773	0.65637
Ljung-Box Test	$\hat{\epsilon}$	$Q(20)$	15.58943	0.74175
Ljung-Box Test	$\hat{\epsilon}^2$	$Q(10)$	4.80247	0.90398
Ljung-Box Test	$\hat{\epsilon}^2$	$Q(15)$	8.89128	0.88313
Ljung-Box Test	$\hat{\epsilon}^2$	$Q(20)$	9.47362	0.97674
LM Arch Test	$\hat{\epsilon}$	TR^2	5.97691	0.91724

Tab. 12: Testy standardizovaných residuí $\{\hat{\epsilon}\}$ modelu GARCH(1,1) s předpokladem Studentova t -rozdělení

Testy standardizovaných residuí prokázaly nekorelovanost a nepřítomnost podmíněné heteroskedasticity⁴⁰. Ovšem je opět nutné ověřit shodu s předpokládaným rozdělením daných residuí. Vhodným testem je například Kolmogorovův-Smirnovův test pro rovnost distribučních funkcí (popsaný v kapitole 2.2). Výsledkem K-S testu je p -hodnota rovna 0.9133. Tedy hypotéza o tom, že standardizovaná residua pocházejí ze studentova t -rozdělení, s počtem stupňů volnosti $\nu = 4.87399$, není zamítnuta a model lze považovat za adekvátní.

R code:

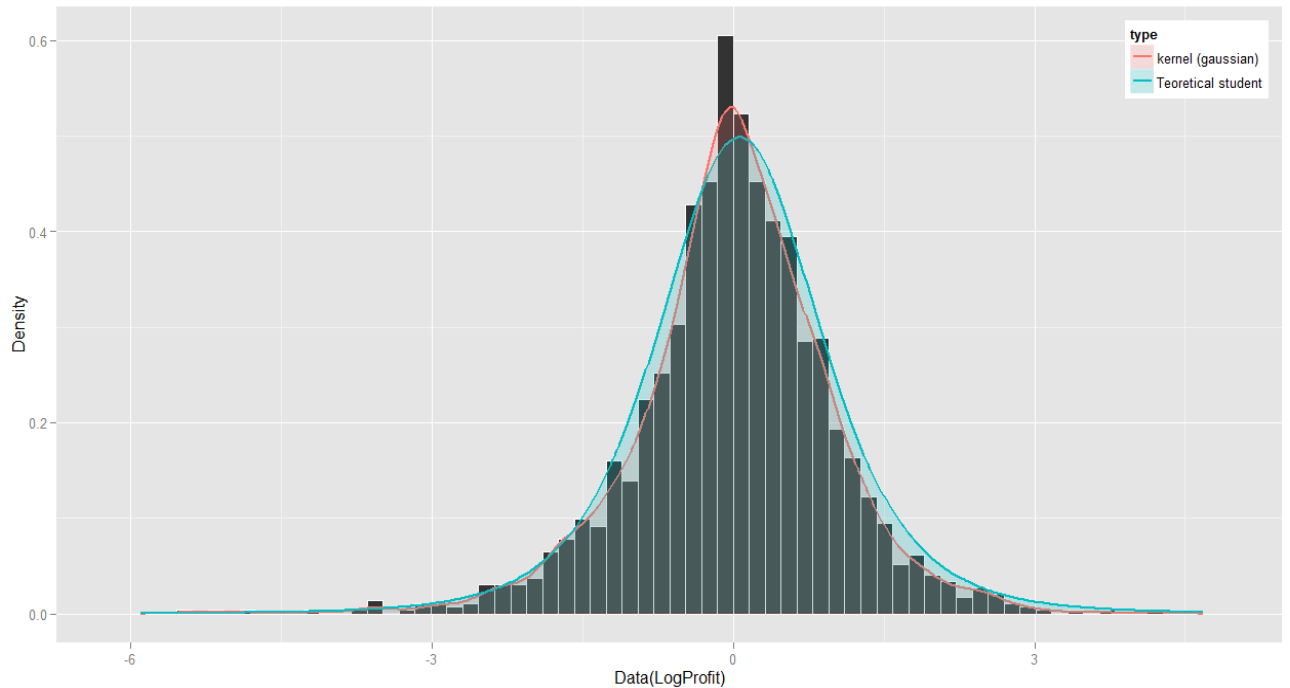
```

1 # Test t-rozdělení standardizovaných residuí
2 # Parametry
3 t1 = 0
4 t2 = 1
5 t3 = GARCHm@fit$coef[4]
6 # K-S test
7 ks.test(GARChE, dstd(t1, t2, t3))

```

⁴⁰Také bychom tento fakt mohli opět ověřit graficky pomocí ACF a PACF standardizovaných residuí a kvadrátů těchto residuí.

Výsledek Kolmogorova-Smirnova testu ještě ilustrujeme následujícím obrázkem grafu odhadu hustoty standardizovaných residuí a t -rozdělení s počtem stupňů volnosti 4.87399.



Obr. 20: Histogram residuí $\{\hat{\epsilon}_t\}$ modelu GARCH(1,1), gaussianový jádrový odhad hustoty rozdělení a teoretické Studentovo t -rozdělení

R code:

```
1 # Jádrový odhad distribuční funkce (student)
2 source("../functions//NKHst.r")
3 G = NKHst(GARCHE, t1, t2, t3)
4 print(G)
```

6.5 Celkový konečný model

ARMA-GARCH je konečným modelem, kterým se rozumí klasický lineární model ARMA proces s podmíněně heteroskedstickými inovacemi. Při konstrukci tohoto modelu se nejprve odhadnul daný ARMA model, pro odstranění lineárních závislostí, a pro jeho residua se použil jeden z modelů podmíněné heteroskedasticity (GARCH).

Model ARMA(1,0)-GARCH(1,1), resp. AR(1)-GARCH(1,1) má tvar:

$$\begin{aligned} X_t &= \mu + \phi X_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \sqrt{h_t} \epsilon_t, \\ h_t &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \end{aligned}$$

kde X_t jsou logaritmy výnosů ceny zlata, $|\phi| < 1$, $\epsilon_t \sim IID(0, 1)$, $\omega > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$. Pod označením IID⁴¹ náhodných veličin ϵ_t se v tomto případě skrývá standardizované Studentovo t -rozdělení s odhadnutým parametrem $\nu = 4.719$ stupňů volnosti [17].

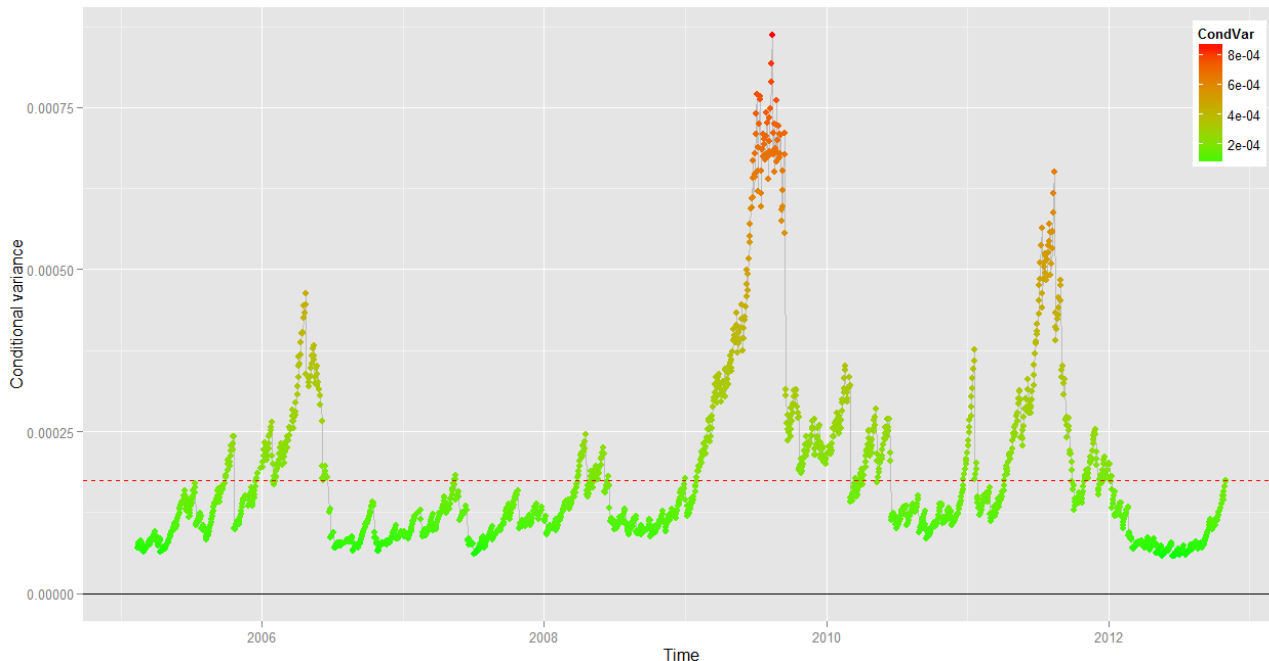
⁴¹Independent and Identically Distributed, tj. nezávislé a stejně rozdělené.

Model s odhadnutými parametry:

$$X_t = 0.0007122 + 0.0608662X_{t-1} + \hat{\varepsilon}_t,$$

$$\hat{\varepsilon}_t = \sqrt{\hat{h}_t}\hat{\varepsilon}_t,$$

$$\hat{h}_t = 0.00000124 + 0.04944\hat{\varepsilon}_{t-1}^2 + 0.9467\hat{h}_{t-1}.$$



Obr. 21: Podmíněný rozptyl odhadnutého modelu GARCH(1,1) a nepodmíněný rozptyl

R code:

```

1 # Vykreslení podmíněného rozptylu
2
3 df <- data.frame(dates = dates[(length(dates)-length(GARCHm@h.t)+1):length(dates)],
4                 CondVar = GARCHm@h.t)
5 {
6 G = ggplot(df, aes(x = dates))
7 G = G + xlab('Time') + ylab('Conditional variance')
8   + geom_line(aes(y=CondVar),size = 0.5, colour = "grey")
9   + geom_point(aes(y=CondVar, colour = CondVar), size = 2)
10  + scale_colour_gradient(low = "green", high = "red")
11  + theme(legend.justification=c(1,1), legend.position=c(1,1))
12  + geom_hline(yintercept=0, color="black")
13  + geom_hline(yintercept=var(ARMAe), linetype='dashed', color="red")
14 }
15 print(G)

```


6.6 Monte Carlo předpovědi

V této části se pokusíme blíže analyzovat charakter předpovědi časové řady pomocí odhadnutého modelu z předchozích částí a k tomu využít metodu Monte Carlo. Postup aplikace této metody byl již popsán v kapitole 4.1. Tento postup tedy převedeme do jazyka R:

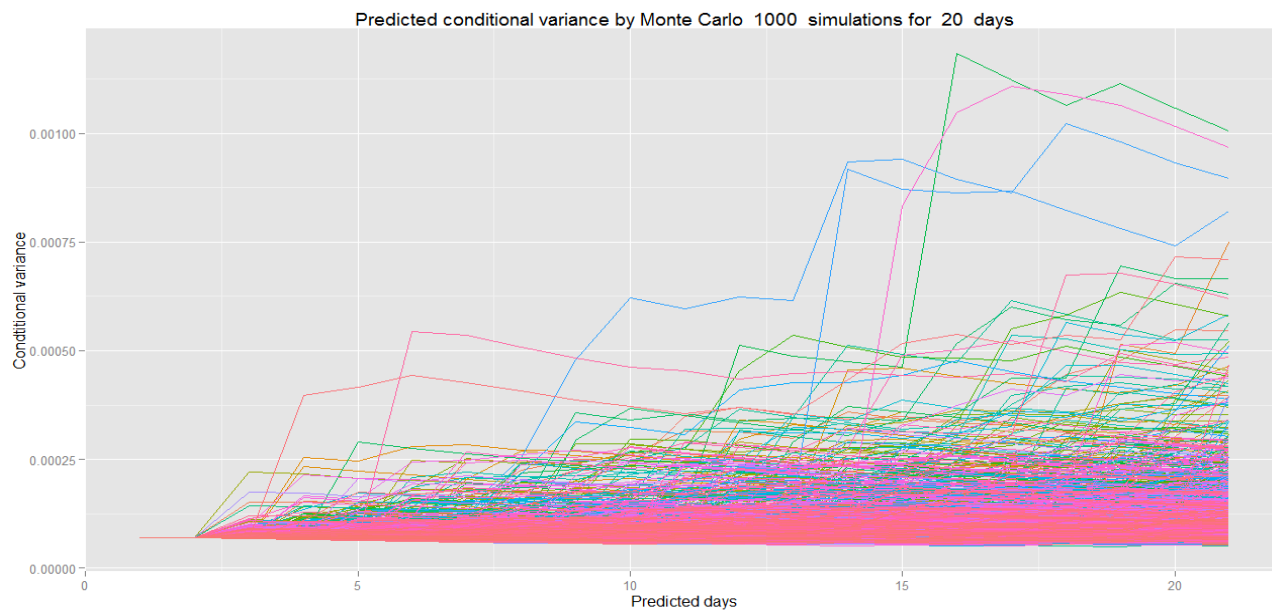
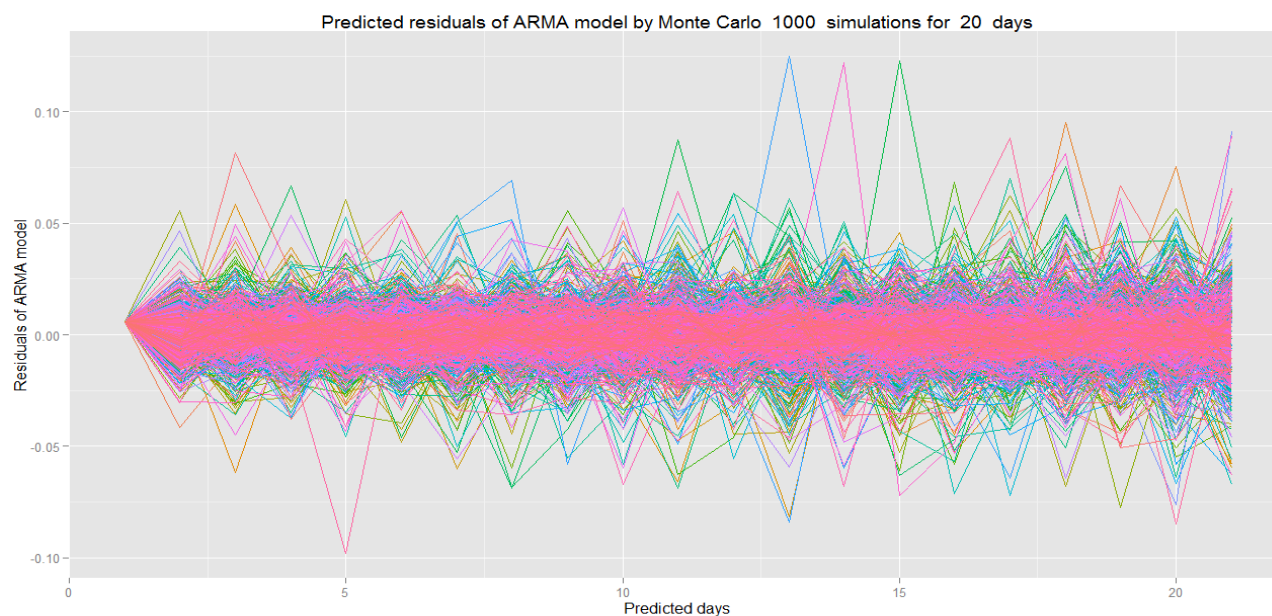
```

R code:
1 # MONTE CARLO
2 # Parametry
3 n.days = 20 # počet odhadovaných časových jednotek dopředu
4 t3 # odhadnutý stupeň volnosti pro t-rozdělení
5 k = 1000 # počet simulací
6
7 # Inicializace k metodě Monte Carlo
8 h.t = 0
9 h.t = GARCHm@h.t[length(GARCHm@h.t)] # hodnota podm. rozptylu v čase t
10
11 e = 0
12 e = ARMAe[length(ARMAe)] # hodnota residua ARMA modelu v čase t
13
14 logProfit = 0
15 logProfit = LogProfit[length(LogProfit)] # hodnota residua logaritmu výnosů v čase t
16
17 i=0
18 MCsim = seq(1:(n.days+1))
19 HT = seq(1:(n.days+1))
20 MClogProfit = seq(1:(n.days+1))
21
22 # Iterace metody Monte Carlo
23 for (i in 1:k) {
24   j=0
25   for (j in 1:n.days) {
26     R = runif(k); # generování náhodného čísla
27     MC = qt(R, df=t3); # hodnoty kvantilové funkce t-rozdělení pro simulované náhodné pravděpodobnosti
28     # simulovaný podm. rozptyl
29     h.t[j+1] = GARCHm@fit$coef[1] + GARCHm@fit$coef[2]*(e[j]^2) + GARCHm@fit$coef[3]*h.t[j]
30     e[j+1] = MC*sqrt(h.t[j+1]) # simulované residuum ARMA modelu
31     logProfit[j+1] = ARMAm$coef[2] + ARMAm$coef[1]*logProfit[j] + e[j+1] # simulované logaritmy výnosů
32   }
33   # Uložení i-té iterace do pom. proměnné
34   MCsim = data.frame(MCsim, e)
35   HT = data.frame(HT, h.t)
36   MClogProfit = data.frame(MClogProfit, logProfit)
37 }

```

Nasimulovaná předpověď pro podmíněný rozptyl s odhadnutými parametry z předchozí kapitoly a simulovaná residua ARMA modelu lze vidět na následujících obrázcích, kde první hodnota je vždy hodnota odhadovaného modelu časové řady zlata z času t , což je čas poslední známé hodnoty časové řady cen zlata.

Dále byl nastaven počet simulovaných hodnot na $k = 1000$ a počet předpovídaných dnů na $h = 20$.

Obr. 22: Simulované předpovědi podmíněného rozptylu \hat{h}_t na 20 dní dopředu pomocí metody MCObr. 23: Simulované předpovědi na 20 dní pro residua $\hat{\varepsilon}_t$ modelu ARMA(1,0) pomocí metody MC**R code:**

```

1 # Graf MC předpovědí pro podmíněný rozptyl
2 A = melt(HT, id = 'HT', variable_name = 'series')
3 G = ggplot(A, aes(HT, value)) + geom_line(aes(colour = series))
4   + theme(legend.position = "none")
5   + xlab("Predicted_days")
6   + ylab("Conditional_variance")
7   + labs(title = paste("Predicted_conditional_variance_by_Monte_Carlo_", k, "_simulations_for_", n.days, "_days"))
8 print(G)
9
10 # Graf MC předpovědí pro residua ARMA modelu analogicky

```

Z předcházejících obrázků je patrné, že simulace metody Monte Carlo vlastně vymezují oblast, ve které by se mohla nová budoucí data pohybovat. Toho lze využít několika způsoby. Zde bude uveden způsob určení intervalů pomocí empirických kvantilových funkcí. Obecně lze kvantilovou funkci (resp. inverzní distribuční funkci) napsat jako:

$$Q(p) = F^{-1}(p) \quad \text{nebo také} \quad x_p = F^{-1}(p),$$

kde pro danou pravděpodobnost p lze pomocí kvantilové funkce nalézt odpovídající hodnotu p -kvantilu x_p náhodné veličiny X (tj. platí $x_p = F^{-1}(p)$ a $F(x_p) = p$). Inverzní funkci lze vytvořit pouze ke spojitě rostoucí distribuční funkci, resp. pro spojitá rozdělení⁴². Empirické kvantilové funkce lze získat z n dat vzestupným tříděním jejich hodnot x_i a vytvořením tzv. schodovité funkce, kde osa y pak udává hodnoty x_i a osa x pravděpodobnost $1/n$ jednotlivých $x_i, i = 1, \dots, n$.

R code:

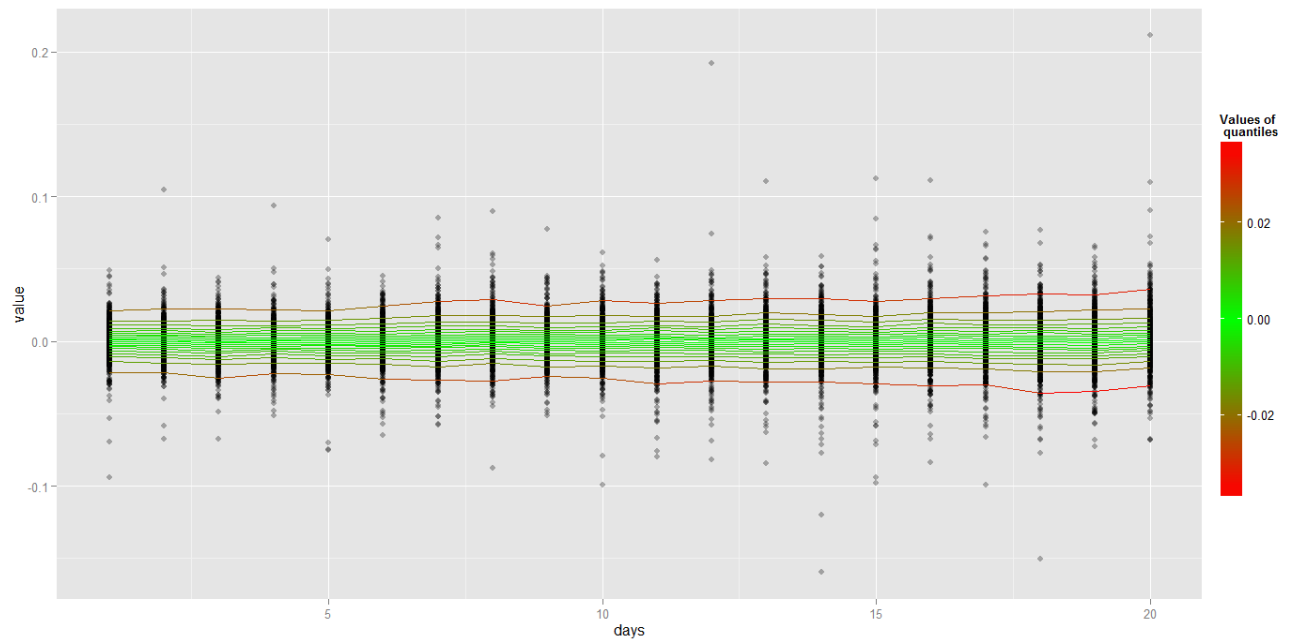
```

1 # Empirické intervalové odhady pro simulaci předpovědi MC
2 empInt = 0
3 i = 0
4 j = 0
5 Means = 0 # pro ukládání průměrů simulací v daných dnech
6 MCSim1 = data.frame(days = seq(1:(n.days))) # pomocná proměnná pro graf
7
8 for (i in 1:n.days){
9   P=0
10      for (j in 1:k) {
11        P[j] = MCSim[i+1, j+1]
12        MCSim1[i,j+1] = MCSim[i+1, j+1]
13      }
14      Means[i] = mean(P)
15      P = ecdf(P) # empirical cumulative distribution function
16      if (i==1) {
17        empInt = quantile(P, probs = c(seq(alpha/2, 0.5, 0.05), rev(1-seq(alpha/2, 0.5, 0.05))),
18                          na.rm = FALSE, names = TRUE, type = 1)
19      }
20      else {
21        empInt = data.frame(empInt, empInt = quantile(P, probs = c(seq(alpha/2, 0.5, 0.05),
22                          rev(1-seq(alpha/2, 0.5, 0.05))), na.rm = FALSE, names = TRUE, type = 1))
23      }
24 }

```

⁴²Pro případy diskrétních a smíšených rozdělení je třeba definici kvantilové funkce zobecnit: $Q(p) = \inf x | F(x) \geq p$.

Následující graf ukazuje simulované hodnoty v jednotlivých dnech a 97.5%, 92.5%,..., 7.5%, 2.5% kvantily zkonstruované pomocí empirických kvantilových funkcí, vytvořených výše uvedeným postupem. Pro simulovaná data tedy byla vytvořena empirická kvantilová funkce v jednotlivých budoucích dnech a z těchto funkcí pak bylo možné získat jednotlivé procentní meze, které tvoří předpovědní intervaly.



Obr. 24: Simulované hodnoty residuí ARMA modelu a intervaly empirické kvantilové funkce.

R code:

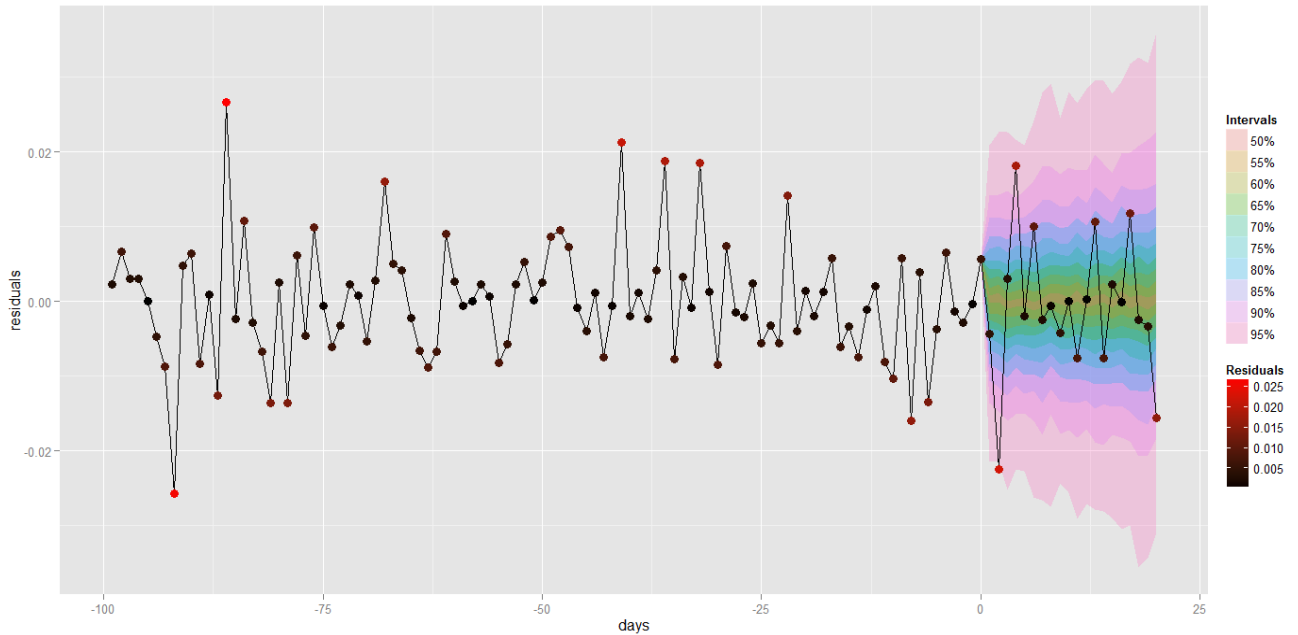
```

1 # Příprava pro graf
2 P = names(quantile(P, probs = c(seq(alpha/2, 0.5, 0.05), rev(1-seq(alpha/2, 0.5, 0.05))),
3           na.rm = FALSE, names = TRUE, type = 1))
4
5 empInt = melt(empInt);
6 empInt = data.frame(day = rep(seq(1:n.days), each = length(P)), quantiles = rep(P, n.days), empInt)
7 names(empInt) = c("days", "percentage", "variable", "Quantiles")
8
9 # Graf empirických intervalů vytvořených simulacemi
10 A = melt(MCsim1, id = 'days', variable_name='series')
11 G = ggplot(data = A, aes(days, value, group = series))
12   + geom_point(col = "black", alpha = 0.3)
13   + geom_line(data = empInt, aes(days, Quantiles, group = percentage, colour = Quantiles))
14   + scale_colour_gradient2(name = "Values_of\n_n_quantiles_",
15                             midpoint = 0, low = "red",
16                             mid = "green", high = "red",
17                             guide = "colourbar")
18   + theme(legend.key.height = unit(2, "cm"))
19 print(G)

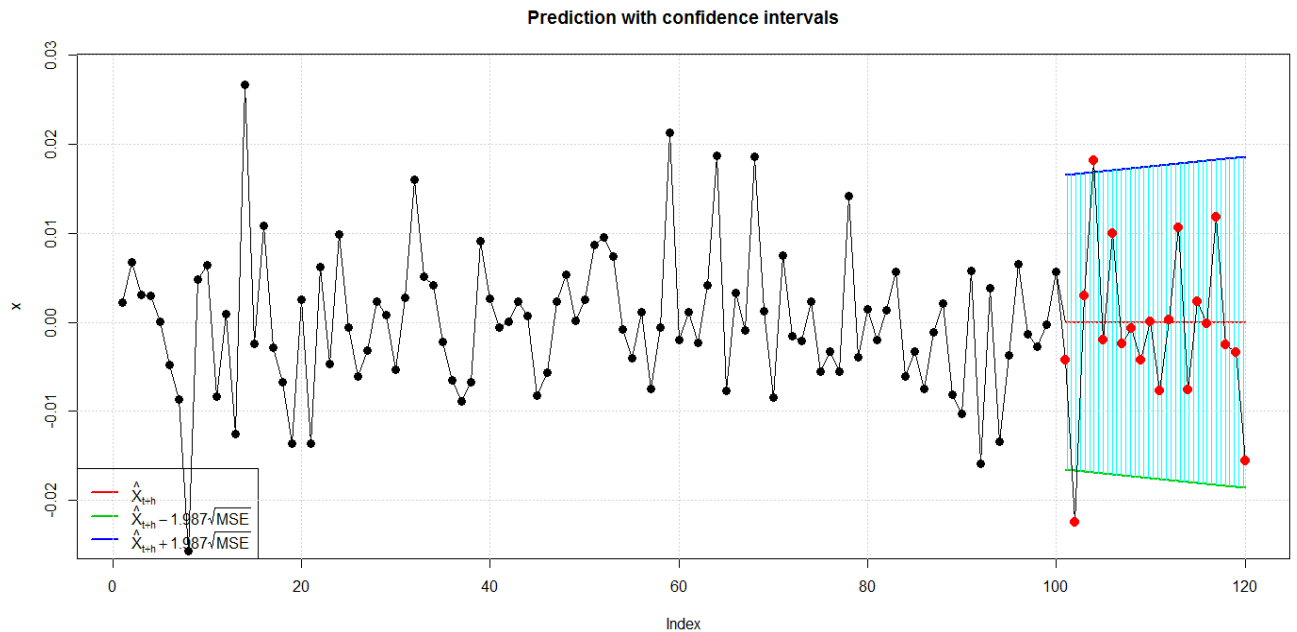
```

6.6.1 Porovnání se skutečností

V této fázi porovnáme předpovědi sestavené pomocí metody Monte Carlo se skutečností. V projektu R načteme a přidáme další data do časové řady a opět provedeme několik prvních kroků, které jsou stejné jako na počátku zpracování původní časové řady. Jedná se tedy o zlogaritmování a diferencování nově přidaných dat, odstranění lineárních závislostí pomocí již odhadnutého ARMA(1,0) modelu. Tím získáme nová residua $\varepsilon_{t+1}, \varepsilon_{t+2}, \dots, \varepsilon_{t+h}$ (resp. jejich odhady), kde $h = 20$.



Obr. 25: Residua ARMA modelu a empirické intervaly spolehlivosti metody Monte Carlo.



Obr. 26: Residua ARMA modelu a 95% intervaly spolehlivosti pomocí příkazu *predict*.

Předcházející graf na obrázku 25 ukazuje, jak dobře metoda Monte Carlo vymezila předpovědní oblast pro skutečné hodnoty. Pro porovnání pak graf na obrázku 26 ukazuje předpovědní interval sestavený pomocí příkazu *predict*, který užívá metodu předpovědi popsanou v kapitole 3.8.

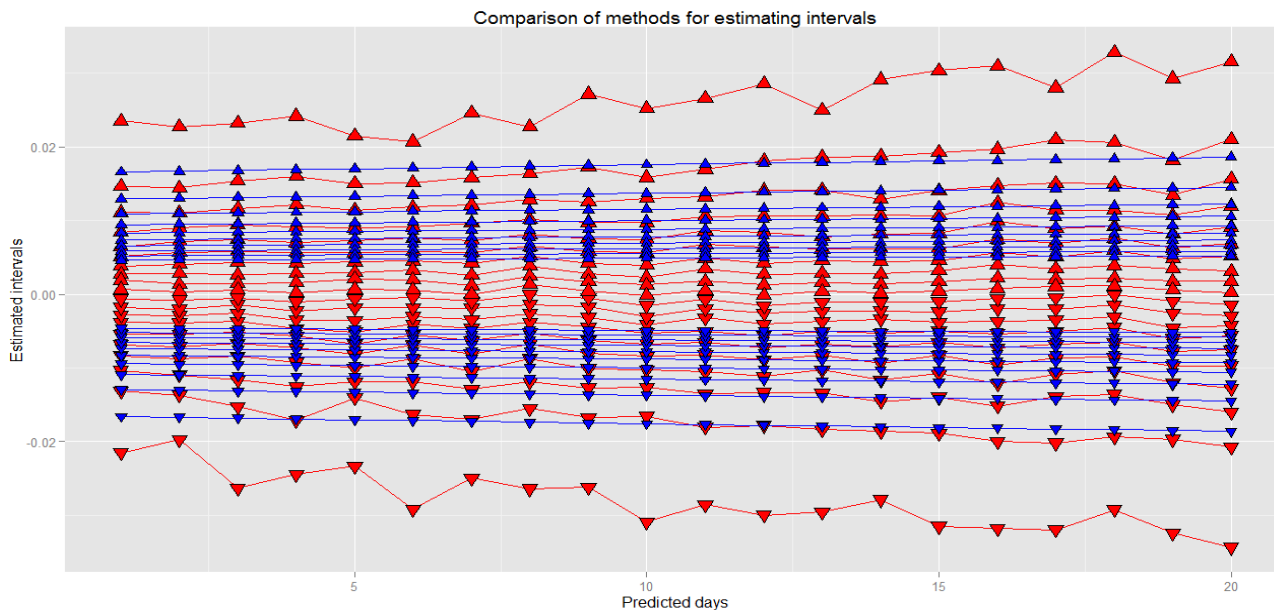
R code:

```

1 # Příprava nových dat pro srovnání predikce
2
3 # Načtení dat s přidávanými hodnotami (+20 dní)
4 data2 <- read.csv(file = "goldPred.csv", head = FALSE, sep = ";")
5
6 # Rozdělení dat do skupin
7 dates2 = rev(as.Date(data2$V1, format = "%d.%m.%Y"))
8 serie2 = rev(data2$V2)
9
10 # Převod na stacionární časovou řadu
11 logSerie2 = log(serie2) # logaritmy dat
12 lags2 = 1 # nastavení řádu diference
13 LogProfit2 = diff(logSerie2, lag = lags2) # logaritmy výnosů
14
15 # Určení residuí ARMAe pomocí již odhadnutého modelu ARMA pro další doplněná data
16 model = 0
17 i = 0
18 a = 100 # počet posledních známých dat
19
20 # Příprava pro Graf Predict
21 for (i in 2:length(LogProfit2)) {
22   model[i] = ARMAm$coef[2] + ARMAm$coef[1] * LogProfit2[i]
23 }
24 ARMAePred = LogProfit2[(max(p,q)+1):length(LogProfit2)] - model[1:(length(model)-1)]
25 K = ARMAePred[(length(ARMAePred)-(a+n.days)):length(ARMAePred)] # výběr pouze posledních a+n.days dat
26
27 # Graf předpovědi pomocí "predict" + porovnání se skutečností
28 m = seq(1:(a+n.days))
29 predict(GARCHm, n.ahead = n.days, plot = TRUE, mse = "cond", nx = a, conf = (1-alpha))
30 lines(K[2:(a+n.days+1)])
31 points(K[2:(a+1)], col = "black", pch = 20, cex = 1.8)
32 points(m[(a+1):(a+n.days)],K[(a+2):(a+n.days+1)], col = "red", pch = 20, cex = 2)
33
34 # Příprava pro Graf Monte Carlo
35 K = ARMAePred[(length(ARMAePred)-(n.days)):length(ARMAePred)] # výběr nových (n.days) hodnot residuí
36 K = data.frame(days = seq(1:n.days), real = K[2:length(K)]) # uložení do datového rámce s příslušnými dny
37 M = ARMAe[(length(ARMAe)-a):length(ARMAe)] # výběr pouze (a) posledních hodnot residuí
38 A3 = data.frame( days = rev(A$days[1:(length(A2)*n.days)]),
39                 gr = rep(0, length(A2)*n.days),
40                 residuals = rev(rep(Mean, length(A2))),
41                 real = rep(rev(K$real), length(A2)),
42                 Intervals = rev(rep(rev(A2), each = n.days)),
43                 Intervals2 = (rep(rev(A2), each = n.days)),
44                 Upper = rev(melt(A1[length(A1):((length(A1)-1)/2+2)])$value),
45                 Lower = rev(melt(A1[2:((length(A1)-1)/2+1)])$value))
46 A4 = with(A3, A3[order(days), ]) # seřazení podle dnů
47 i=0
48 for (i in 1:a) {
49   A4 = data.frame( days = c(rep(-i+1, length(A2)), A4$days),
50                   gr = c(rep(1, length(A2)), A4$gr),
51                   residuals = c(rep(M[length(M)-i+1], length(A2)), A4$residuals),
52                   real = c(rep(M[length(M)-i+1], length(A2)),A4$real),
53                   Intervals = rep(levels(A4$Intervals),(length(P)+i)),
54                   Intervals2 = rev(rep(levels(A4$Intervals2), (length(P)+i))),
55                   Upper = c(rep(M[length(M)-i+1], length(A2)), A4$Upper),
56                   Lower = c(rep(M[length(M)-i+1], length(A2)), A4$Lower))
57 }
58
59 # Graf (A4)
60 G = ggplot(A4, aes(x = days, y = residuals))
61   + geom_ribbon(aes(ymin = Lower, ymax = Upper, fill = Intervals,
62                   group = Intervals2), alpha = 0.25)
63   + geom_line(aes(days, real), col = "black")
64   + geom_point(aes(days, real, colour = abs(real)), size = 3)
65   + scale_colour_gradient(name = "Residuals", low = "black", high = "red" )
66 print(G)

```

Při porovnání předešlých obrázků 25 a 26, které ukazují posledních 100 + 20 hodnot residuů ARMA modelu zpracované časové řady, lze vidět, že 95% intervaly se u těchto dvou metod sestavení předpovědi liší. A protože příkaz *predict* neumí vykreslit graf s více intervaly, podívejme se na porovnání všech předpovědních intervalů daných metod v jiném grafu.



Obr. 27: Graf odhadů intervalů metody MC (červeně) a pomocí příkazu *predict* (modře)

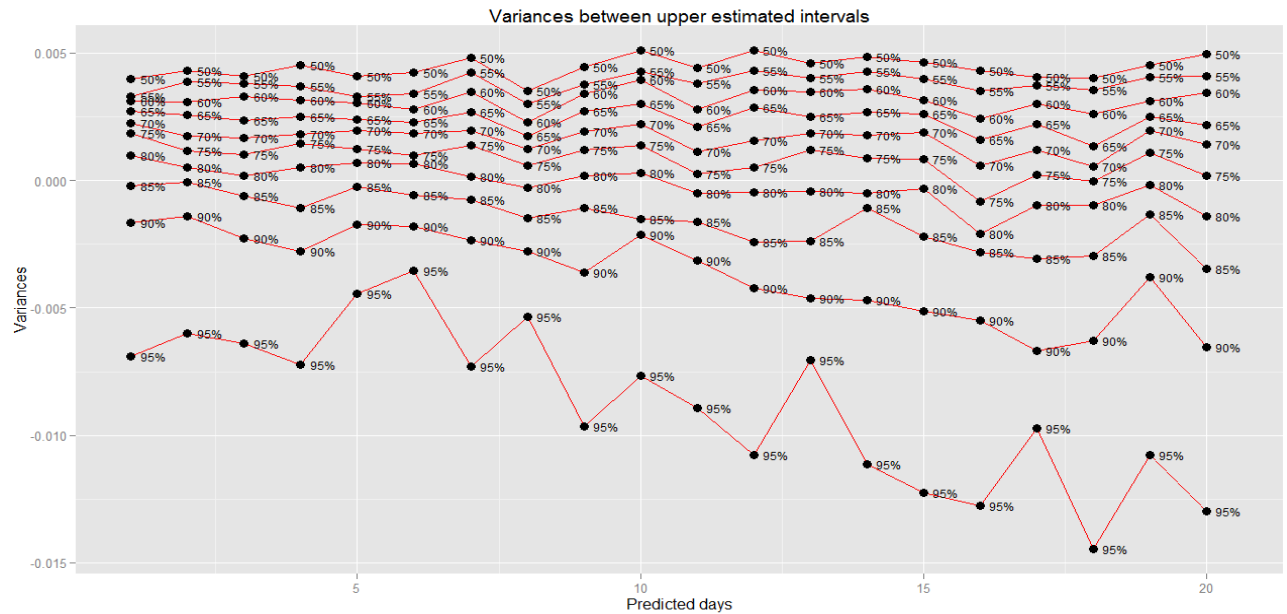
R code:

```

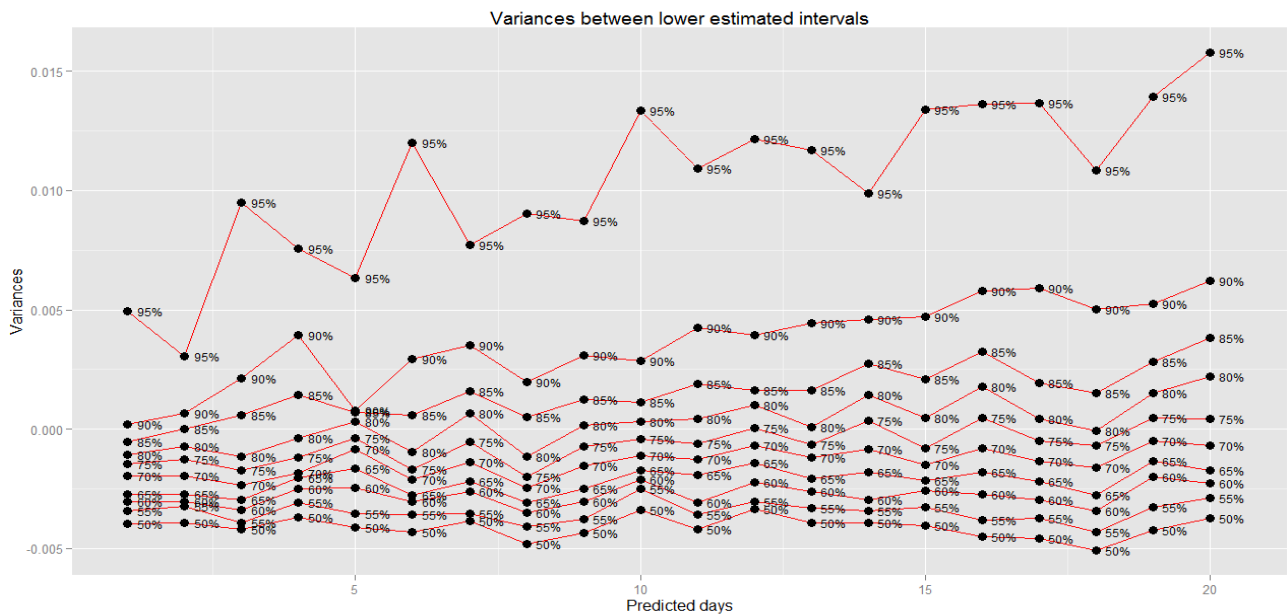
1 # Porovnání predikčních metod pro danou simulaci
2 # Načtení dat (data simulace byla upravena v excelu a znovu načtena - v příloze na CD)
3 intervals <- read.csv(file = "intervals.csv", head = TRUE, sep = ";")
4
5 # Graf empirických intervalů
6 GA = ggplot(intervals, aes(x = days, y = lowerInterval2, group = Intervals))
7   + scale_shape_identity()
8   + geom_line(col = "red")
9   + geom_point(aes(shape = 25), fill = "red", size = 4)
10  + geom_line(aes(y = upperInterval2), col = "red")
11  + geom_point(aes(y = upperInterval2, shape = 24), fill = "red", size = 4)
12  + geom_line(aes(y = lowerInterval), col = "blue")
13  + geom_point(aes(y = lowerInterval, shape = 25), fill = "blue", size = 3)
14  + geom_line(aes(y = upperInterval), col = "blue")
15  + geom_point(aes(y = upperInterval, shape = 24), fill = "blue", size = 3)
16 print(GA)

```

Z předchozího grafu vidíme jasné rozdíly mezi výsledky odhadu předpovědních intervalových mezí obou metod. Graf tedy ukazuje jak velké oblasti pokrývají jednotlivé pásy obou metod při dané $\alpha = 5\%, 10\%, \dots, 50\%$. Z tohoto grafu však poměrně těžko posoudíme konkrétní rozdíly při dané α . Lepší interpretaci umožní následující grafy.



Obr. 28: Graf odchylek mezi horními hranicemi intervalových páسů daných metod



Obr. 29: Graf odchylek mezi dolními hranicemi intervalových páсů daných metod

R code:

```

1 # Graf odchylek mezi danými metodami intervalových odhadů
2 G = ggplot(intervals, aes(x = days, y = varUpper, group = Intervals))
3   + scale_shape_identity()
4   + geom_line(col = "red")
5   + geom_point(size = 3)
6   + geom_text(aes(x = days+0.4, label = Intervals), size = 3)
7 print(G)

```


Předchozí graf 28 ukazuje odchylky mezi jednotlivými horními hranicemi pásů intervalových odhadů. Naopak graf 29 ukazuje odchylky mezi dolními hranicemi pásů intervalových odhadů. Pokud se zaměříme na odchylky, které jsou v grafu nulové, resp. blízké nule, můžeme určit pro jaké pravděpodobnostní pásy dávají obě metody podobné výsledky. Pokud je odchylka kladná, pak metoda Metoda Carlo dává větší horní hranici a tedy širší intervalový pás. Pro zápornou odchylku platí, že metoda Monte Carlo dává nižší horní hranici a tedy užší intervalový pás pro predikci. Pro dolní intervalovou hranici to platí zrcadlově, záporná odchylka říká, že metoda Monte Carlo dává širší pás a kladná odchylka užší pás pro dané pravděpodobnosti. Pokud se podíváme na pásy v čase $t + 1$, pak je z grafů odchylek patrné, že obě metody dávají podobné výsledky pro 85% pás v horní hranici intervalového odhadu a pro 90% pás v dolní hranici intervalového odhadu. Čím delší je předpověď, tím se hodnoty intervalových pásů obou metod potkávají v užších pravděpodobnostních pásech. V čase $t + 20$ jsou hodnoty intervalových odhadů podobné při 75% pravděpodobnosti horní i dolní hranice předpovědních pásů.

V tomto případě si však uvědomme, že tyto výsledky platí pro danou simulaci metody Monte Carlo.

7 Závěr

Cílem mé diplomové práce bylo stručně popsat základní charakteristiku finančních časových řad, modelů typu ARCH a popis aplikace těchto modelů, které jsou využity v aplikační části. Před zpracováním aplikační části jsem si vybral konkrétní časovou řadu ceny zlata, na kterou jsem aplikoval uvedené postupy a metody k získání výsledků ke všem cílům této práce. Dále jsem si vybral software *Project R*, který jsem s jeho vlastnostmi také popsal v teoretické části. V tomto softwaru jsem zpracoval všechny postupy a metody aplikační části a zároveň jsem ho také použil pro zpracování a interpretaci výsledků.

V aplikační části jsem stručně popsal charakteristiku časové řady ceny zlata za období 9.2.2005 - 31.10.2012 a provedl předzpracování dat jako přípravu pro modelování této časové řady, ze kterého vyšly najevo některé typické vlastnosti finančních časových řad. Prvním krokem při modelování této řady bylo stanovení úrovněového modelu ARMA, druhým krokem stanovení modelu GARCH pro residua modelu ARMA. Při aplikaci těchto modelů jsem postupoval podle známých metod popsaných v textu práce, kde jsou popsány i postupy pro výběr a ověření modelů. Výsledkem zpracované finanční řady ceny zlata je, že tuto řadu lze modelovat pomocí modelu ARMA(1,0)-GARCH(1,1).

Dalším cílem této práce bylo určení předpovědí pro zvolenou časovou řadu a k tomuto cíli využít metodu Monte Carlo. Základní princip této metody jsem také popsal v teoretické části a zároveň nastínil jakým způsobem je tuto metodu možné využít k předpovědím při modelování časové řady pomocí modelů typu ARCH. Předpovědní intervaly sestavené pomocí metody Monte Carlo pro časovou řadu ceny zlata jsou uvedeny v poslední části této práce a jsou porovnány se skutečnými hodnotami, pro které byly předpovědní intervaly odhadovány. Toto porovnání ukázalo, jak dobře dokáže tato metoda odhadnout předpovědi, ale také jak dobře byl celkový model pro časovou řadu odhadnut. Nové skutečné hodnoty nejsou nijak významně vychýleny z předpovědních intervalů sestavených pomocí metody Monte Carlo a můžeme tak tvrdit, že sestavený model a metodu odhadu přepovědi lze v tomto případě využít k dalšímu předpovězení vývoje časové řady ceny zlata.

V této části bych chtěl také ohodnotit software *Project R*, který jsem si vybral a tímto způsobem také otestoval jeho využití k modelování finančních časových řad. Pomocí tohoto software je možné poměrně snadno připravit model pro určitou časovou řadu. Výsledky lze jednoduchým způsobem získat pomocí řady příkazů a funkcí, ale také mnoha manuálů a rad, které lze snadno získat z již vytvořené komunity okolo tohoto softwaru. *Project R* také slouží jako dobrý a pěkný nástroj pro interpretaci výsledků ve formě grafů, které lze naprosto jakkoliv naprogramovat a nadefinovat podle představ každého uživatele.

Způsobů, jak řešit modelování časových řad je obecně mnoho a proto tato práce obsahuje mnoho potenciálu k dalšímu rozšíření či zkoumání. Například by bylo možné zkoumat a využít některé nelineární modely volatility a porovnat výsledky s lineárními modely volatility. Poměrně užitečným rozšířením by bylo zkoumání stability parametrů daných modelů pro časovou řadu ceny zlata nebo také citlivosti výsledků na změnu počtu simulací u metody Monte Carlo. Dalším rozšířením práce by také mohlo být zpracování dalších jiných typů předpovědí a porovnání výsledků s předpovědmi pomocí metody Monte Carlo.

Literatura a zdroje

- [1] J. Arlt, M. Arltová: *Finanční časové řady*. Grada Publishing a.s., Praha, 2003. ISBN 80-247-0330-0.
- [2] R. S. Tsay: *Analysis of financial time series*. Wiley, New York, 2002.
- [3] Z. Prášková: *Základy náhodných procesů II*. Karolinum, Praha, 2007.
- [4] J. Hendl: *Přehled statistických metod*. Portál, s.r.o. Praha, 2009.
- [5] J. D. Hamilton: *Time Series Analysis*. Princeton univerzity Press, Princeton, 1994.
- [6] J. Anděl: *Matematická statistika*. Státní nakladatelství technické literatury, Praha, 1985.
- [7] M. H. Kalos, P. A. Whitlock: *Monte Carlo Methods*. Second, Revised and Enlarged Edition, WILEY-VCH, Weinheim, 2008.
- [8] J. Mun: *Modelling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization Techniques*. John Wiley & Sons, Inc., Hoboken, 2006.
- [9] R. T. Baillie, T. Bollerslev: *Prediction in Dynamic Models with Time-Dependent Conditional Variances*, Journal of GARCH, 1992, 52:91-113. Dostupné z: https://www.msu.edu/user/baillie/J_Econometrics.1992.pdf
- [10] Shapiro, S.S., Wilk, M.B.: *Biometrika*. An Analysis of Variance Test for Normality (Complete Samples). 1965, roč. 52, č. 3/4, s. 591-611. Dostupné z: <http://www.jstor.org/stable/2333709>.
- [11] *KURZY CZ* [online]. [cit. 2012-12-08]. Dostupné z: <http://www.kurzy.cz/komodity/index.asp?A=5&idk=87&od=29.9.2003&curr=USD&unit=&lg=1&MAXROWS=50>
- [12] E. F. Fama: *The Behaviour of Stock-Market Prices*. Journal of Business, 1965 Volume 38. Dostupné z: http://www.ifa.com/Media/Images/PDFfiles/FamaThe_Behavior_of_Stock_Market_Prices.pdf
- [13] R. Cont: *Empirical properties of asset returns: stylized facts and statistical issues*. Quantitative Finance, 2001 Volume 1. Dostupné z: <http://www.cmap.polytechnique.fr/~rama/papers/empirical.pdf>
- [14] *Analýza finančních časových řad typu ARCH*. Plzeň, 2010. Dostupné z: <https://stag-ws.zcu.cz/ws/services/rest/kvalifikacniprace/downloadPraceContent?adipIdno=36997>. Bakalářská práce. Západočeská univerzita. Vedoucí práce RNDr. Blanka Šedivá, Ph.D.
- [15] *Analýza vývoje ceny ropy, zlata, vybraných blue chips a akciových indexů*. Brno, 2011. Dostupné z: http://is.muni.cz/th/206665/esf_m/Diplomka4.pdf. Diplomová práce. Masarykova univerzita. Vedoucí práce Ing. Gabriela Oškrdalová.
- [16] *Struktura a vlastnosti modelu GARCH(1,1)*. Praha, 2009. Dostupné z: http://is.muni.cz/th/206665/esf_m/Diplomka4.pdf. Bakalářská práce. Vysoká škola ekonomická v Praze. Vedoucí práce Ing. Jan Pígl.
- [17] *Studentovo t-rozdělení a jeho aplikace*. Olomouc, 2008. Dostupné z: <http://mant.upol.cz/soubory/OdevzdanePrace/B09/b09-13-aw.pdf>. Bakalářská práce. Univerzita palackého. Vedoucí práce RNDr. Karel Hron, Ph.D.
- [18] *The R Project: for Statistical Computing*. [online]. [cit. 2013-01-05]. Dostupné z: <http://www.r-project.org/>.

A Přílohy obsažené na CD

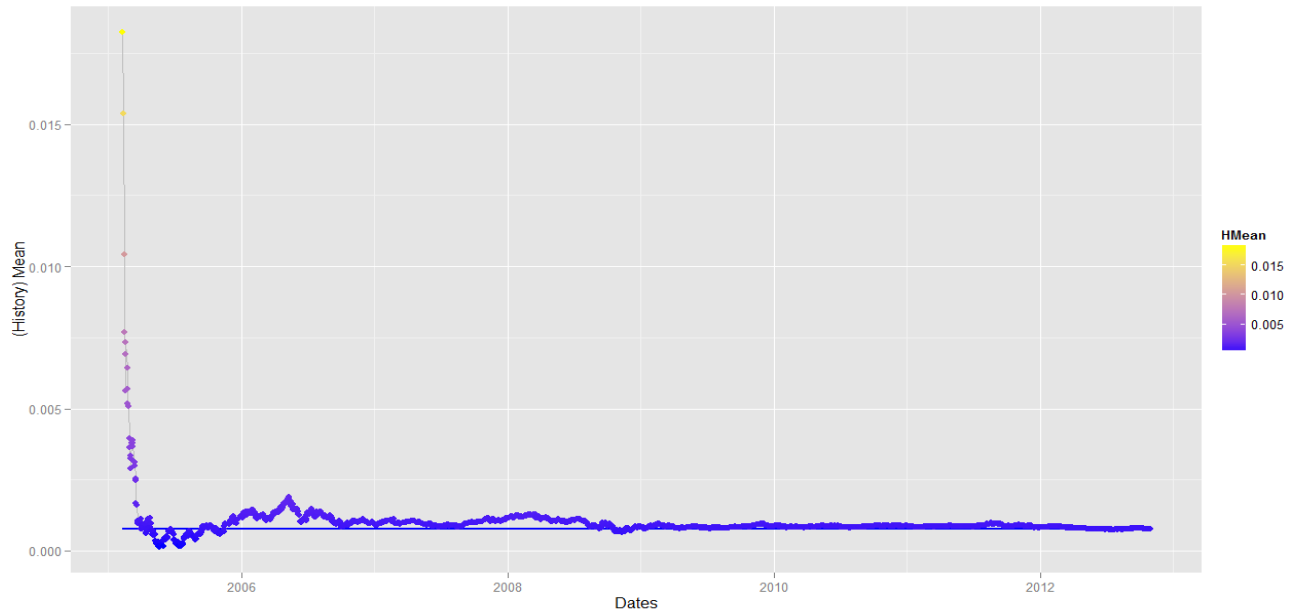
P1 DPFIS_PAKOSTA_A10N0162P.pdf - soubor s kompletním textem diplomové práce v elektronické podobě.

P2 DATA - složka obsahující všechna použitá data.

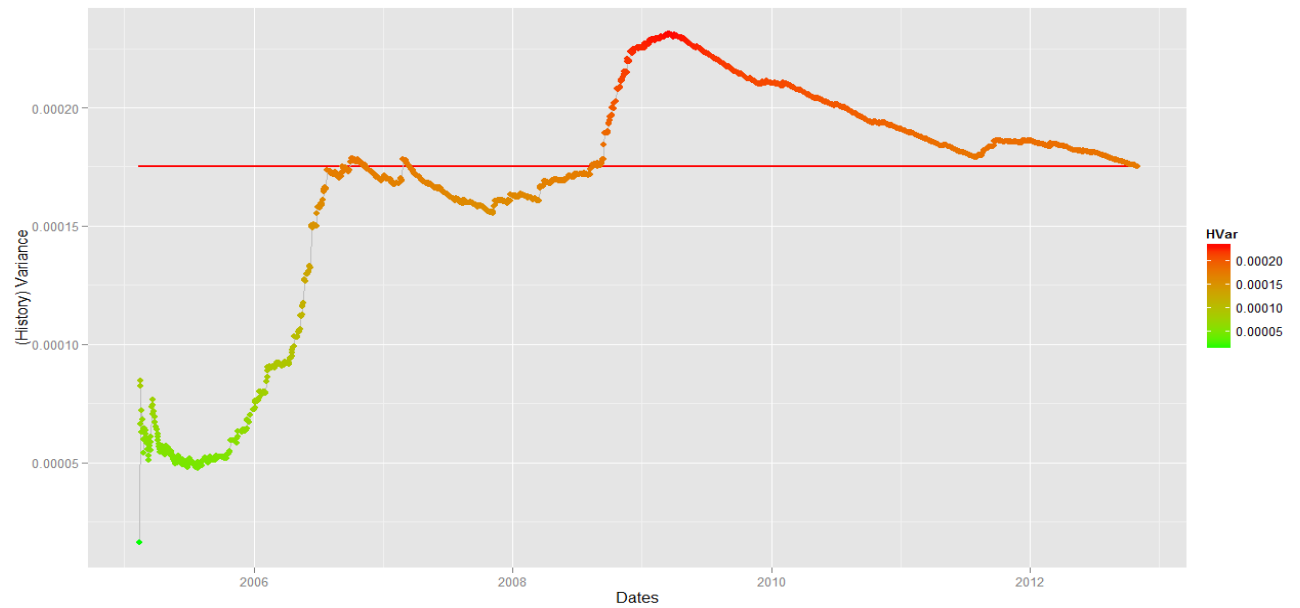
P3 R - složka obsahující všechny zdrojové soubory, které obsahují celé zpracování časové řady v projektu R.

P4 LATEX - složka obsahující zdrojový kód textu práce v \LaTeX u.

B Přílohy tištěné



Obr. 30: Graf historické střední hodnoty logaritmu výnosů časové řady ceny zlata



Obr. 31: Graf historického rozptylu hodnoty logaritmu výnosů časové řady ceny zlata