Vysoka škola ekonomická v Praze

Fakulta finance a účetnictví

Katedra bankovnictví a pojišťovnictví

Studijní obor: Finanční inženýrství

Diplomová práce

Modelování volatility na vysokofrekvenčních datech

Autor: Bc. Philipp Buev

Vedoucí: prof. RNDr. Jiří Witzany, Ph.D.

Rok obhajoby: 2020

Čestné prohlášení

Prohlašuji, že diplomovou práci na téma „Modelování volatility na vysokofrekvenčních datech“ jsem vypracoval samostatně a veškerou použitou literaturu a další prameny jsem řádně označil a uvedl v přiloženém seznamu.

V Praze dne Philipp Buev

Poděkování

Velmi rád bych chtěl poděkovat vedoucímu mé diplomové práce panu prof. RNDr. Jiřímu Witzanymu, Ph.D. za jeho ochotu, čas, vstřícný přístup, odborné připomínky a cenné rady. Dále bych chtěl poděkovat svým rodičům, prarodičům a kamarádům za podporu během studia.

Abstrakt

Tato diplomová práce se zabývá modelováním volatility na vysokofrekvenčních datech. V dané práci jsou aplikovány čtyři typy HAR modelů: HAR-RV, HAR-RV-J, HAR-Q a HAR-Q-J. Analýza se provádí na 5minutové časové řadě indexu Moskevské burzy (MOEX). Hlavním cílem dané diplomové práce je výběr nejlepšího modelu pro modelování a předvídání volatility na finančních trzích. Dalším cílem práce je zjistit, zda rozšíření základních typů HAR modelů o realizovanou quarticity a o regresory skoků má pozitivní vliv na predikční schopnosti modelů. Volatilita se bude odhadovat na následující den a na následující týden. V aplikační části této práce budou porovnány předpovědní schopnosti HAR modelů pomocí RMSE, indexu determinace a Mincer – Zarnowitz regrese, a to jak na in-sample, tak i na out-sample datech. Výsledky provedené analýzy ukazují, že rozšíření HAR modelů o regresory skoků zlepšuje předpovědi volatility na následující den, však rozšíření o realizovanou quarticity nemá pozitivní vliv na výkonnost modelů. V případě predikcí volatility na následující týden nelze jednoznačně určit, zda realizovaná quarticity zlepšuje predikční schopnosti HAR modelů. V případě rozšíření modelů o regresory skoků je vidět zlepšení předpovědních schopností modelů při predikci volatility na následující týden. HAR-RV-J model dosáhl nejlepších výsledků při denním horizontu předpovědi. Při týdenním horizontu předpovědi nejlepším modelem je také HAR-RV-J.

Klíčová slova: realizovaná volatilita, vysokofrekvenční data, bipower variance, realizovaná variance, realizovaná quarticity, HAR modely

Abstract

This master thesis deals with volatility modeling on high-frequency data. There are four types of HAR models applied: HAR-RV, HAR-RV-J, HAR-Q and HAR-Q-J. The analysis is carried out on a 5-minute time series of the Moscow Stock Exchange Index (MOEX). The main aim of the thesis is to select the best model for modeling and forecasting volatility in financial markets. Another goal of the thesis is to find out if the extension of basic types of HAR models by realized quarticity and by jump variables has a positive effect on the predictive capabilities of models. Volatility will be estimated for the next day and the following week. In the application part of this thesis, the predictive capability of HAR models will be compared using RMSE, determination index and Mincer - Zarnowitz regression on both in-sample and out-sample data. The results of the analysis show that the extension of HAR models by jump variables does improve the volatility forecasts for the following day, but the extension by realized quarticity does not have positive impact to the performance of model. In the case of volatility predictions for the following week, it cannot be unambiguously determined if the realized quarticity improves predictive ability of HAR models. If models are extended by jump variables, we can see an increase in the performance of the models in volatility prediction for the following week. The HAR-RV-J model achieved the best results at the daily forecast horizon. Also, at the weekly forecast horizon, the best model is HAR-RV-J.

Key words: realized volatility, high-frequency data, bipower variation, realized variation, realized quarticity, HAR models

Seznam použitých zkratek

ARCH – Autoregressive conditional heteroscedasticity

BV – Bipower variation

GARCH – Generalized Autoregressive conditional heteroscedasticity

HAR – Heterogeneous Autoregressive

IV – Integrated variation

MOEX - Moscow Stock Exchange Index

OLS – Ordinary least squares

QV – Quadratic variation

RMSE – Root mean square error

RQ – Realized quarticity

RV – Realized variance

SV – Stochastic volatility

THT – Heterogeneous market hypothesis

TQ – Tripower quarticity

Obsah

Úvod 1

1. Volatilita 3

1.1 Definice volatility 3

1.2 Spojitý bez-arbitrážní proces vykonávaný cenou finančního aktiva 4

1.3 Dekompozice výnosů 4

1.4 Vlastnosti volatility 5

1.5 Kvadratická variace 7

1.6 Realizovaná volatilita a její vlastností 10

1.7 Efekt fungování tržní mikrostruktury 11

2. HAR modely a detekce skoků 13

2.1 Bipower variance 13

2.2 Teorie heterogenních trhů (THT) 16

2.3 HAR-RV model 17

2.4 HAR-RV-J/HAR-RV-CJ modely 18

2.5 HAR-Q a HAR-Q-J modely 20

2.6 Porovnání predikčních schopností modelů 21

2.7 Jiné metody modelování volatility na vysokofrekvenčních datech 22

3. Aplikace modelů 25

3.1 Analýza a popis dat 25

3.1.1 Vysokofrekvenční data 25

3.1.2 Popis dat 26

3.2 Aplikace modelů 32

3.2.1 HAR-RV model 32

3.2.2 HAR-RV-J model 35

3.2.3 HAR-Q model 38

3.2.4 HAR-Q-J model 40
3.3 Porovnání modelů 43
3.4 Modelování předpovědí na různých délkách trénovacích dat 45
3.4.1 Year-by-Year odhad parametrů modelů 45
3.4.2 Porovnání predikčních schopností modelů na různých délkách in-sample dat 47
3.5 Predikce agregované volatility na následující týden 50
Závěr 55
Použitá literatura 57
Seznam tabulek a grafů 61

Úvod

Volatilita na finančních trzích byla v posledních desetiletích pro výzkumníky i odborníky velmi zajímavým jevem. Ve financích je volatilita často definována jako směrodatná odchylka. Volatilita výnosu aktiva ve finanční ekonomii hraje ústřední roli, protože se používá při oceňování aktiv, řízení rizik, v portfolio managementu atd.

Cena finančního aktiva je modelována pomocí stochastického procesu. Nejběžnějším stochastickým procesem používaným při modelování cen aktiv je geometrický Brownův pohyb. Tento zjednodušený přístup při modelování cen finančních aktiv však vyžaduje více předpokladů, které se ve skutečnosti ukázaly jako nepravdivé. Mimo jiné se předpokládá konstantní očekávaný výnos a konstantní volatilita. Kromě toho se předpokládá log-normální rozdělení výnosů finančních aktiv. Ve skutečnosti empirické výzkumy ukázaly, že výnosy mají jiná rozdělení než log-normální. Rozdělení výnosů mají těžké konce a jsou zešikmené doleva. Těžké konce lze na trzích pozorovat jako neobvykle velké pohyby cen aktiv, označované jako skoky. Tyto skoky jsou spojeny s neočekávanými informacemi, které se objevují na trhu.

Ve mnoha modelech se předpokládá, že je volatilita v průběhu času vždy konstantní, nicméně mezi finančními akademiky a odborníky z praxe je všeobecně známo, že se volatilita v průběhu času významně mění a samotná volatilita není přímo pozorovatelná. Lze ji odhadnout pouze v rámci konkrétního modelu. Turbulentní období během finanční krize 2007-2009 pouze potvrdilo, že je potřeba modelovat volatilitu jako časově měnlivou proměnnou.

Technologický pokrok posledních dekád umožnil zpracovávat a využívat vysokofrekvenční data k modelovaní volatility. Využití vysokofrekvenčních dat značně zlepšilo výkonnost modelů volatility. Proto daná diplomová práce se výhradně zaměřuje na modelování volatility na vysokofrekvenčních datech.

V dané práci budou analyzovány čtyři HAR modely: HAR-RV, HAR-RV-J, HAR-Q a HAR-Q-J. Tyto modely budou odhadnuty na 5minutové časové řadě indexu Moskevské burzy (MOEX) za období od 1. června 2005 do 31. května 2019. Hlavním cílem této práce je výběr nejlepšího modelu pro modelování a predikci volatility. Dalším cílem je zjistit, zda přidání realizované quarticity a regresorů skoků přispívá ke zlepšení predikčních schopností HAR modelů.

Tato práce je rozdělena do tří částí. V první části jsou popsány nejdůležitější vlastnosti volatility a její využití. Také jsou definovány pojmy realizovaná volatilita a kvadratická variace. Druhá část se věnuje HAR modelům, teorii heterogenních trhů a problematice detekce skoků. V této části je uveden základní HAR-RV model a jeho rozšíření o regresory skoků a o realizovanou quarticity. Třetí část diplomové práce se zabývá odhadem a testováním vybraných k analýze modelů. V rámci těchto modelů se volatilita odhaduje na následující den a na následující týden. Kromě toho se ve třetí části zkoumá dynamika parametrů HAR modelů. Na konci třetí kapitoly se porovnávají predikční schopnosti odhadnutých modelů. Porovnání modelů bude provedeno pomocí RMSE, indexu determinace a Mincer-Zarnowitz regrese. Výkonnost modelů se bude zkoumat jak na in-sample, tak i na out-sample datových vzorcích.

1. Volatilita

1.1 Definice volatility

Ve financích volatilita je chápána jako míra kolísání ceny určitého finančního aktiva nebo tržního indexu. Nejčastěji se jedná o směrodatnou odchylku výnosů finančního aktiva. Kromě toho volatilitu lze chápat jako míru rizikovosti finančních nástrojů v určitém časovém období, proto volatilita je zásadní pro oceňování aktiv, zajištění nebo řízení rizik. Volatilita podkladového aktiva je jednou z klíčových proměnných při stanovení cen derivátů. Potřeba přesně prognózovat volatilitu z ní učinila jeden z ústředních problémů finanční ekonometrie.

Existují dva hlavní druhy volatility: implikovaná volatilita a realizovaná volatilita. Implikovaná volatilita se často používá v modelech oceňování opcí. Jedná se o volatilitu implikovanou tržní cenou opce na základě modelu oceňování opcí. Jinými slovy, jedná se o tržní hodnocení budoucí volatility. Naproti tomu realizovaná volatilita měří, co se ve skutečnosti stalo v minulosti. V této diplomové práci se výhradně zaměříme na realizovanou volatilitu.

Volatilita, kterou pozorujeme na finančních trzích, je pouze jednou realizací stochastického procesu. Tento stochastický proces je skrytý, neznáme data generující proces, což značně komplikuje předvídání volatility.

Stochastický proces, vykonávaný logaritmickým výnosem finančního aktiva, je možné modelovat dvěma různými způsoby: bud` jako proces spojitý nebo diskrétní. Pokud se jedná o spojitý proces, tento proces se dá formulovat pomocí stochastické diferenciální rovnice. Bez vlivu skoků tento proces bude vypadat následovně:

$$dp_t = \mu_t dt + \sigma_t dW_t$$

Kde p_t je logaritmus ceny finančního aktiva, μ_t je střední hodnota, σ_t je směrodatná odchylka a W_t je Wienerův proces.

Pokud se bude uvažovat vliv skoků:

$$dp_t = \mu_t dt + \sigma_t dW_t + \xi_t dq_t$$

Kde ξ_t je proces potenciální velikosti skoků, q_t je Poissonův proces nabývající hodnot 0;1 (0 v případě absence skoků, 1 v případě přítomnosti skoků).

1.2 Spojitý bez-arbitrážní proces vykonávaný cenou finančního aktiva

Předpokládejme, že p_t je logaritmus ceny finančního aktiva definovaný na celém pravděpodobnostním prostoru (Ω, F, P), kde Ω je množina elementárních jevů, F je $\sigma-$algebra na Ω a P je pravděpodobnost na F. P_t se vyvíjí ve spojitém čase v intervalu $[0;T]$, kde T je konečné číslo a $\{F_t\}_{t\in[0,T]} \subseteq F$ představuje přirozenou filtraci. Všechny historické informace o cenách finančních aktiv a ostatních proměnných jsou známé v čase t. Logaritmický výnos ve spojitém čase v intervalu $[t-h, t]$ je definován následovně:

$$r_{t,h} = p_t - p_{t-h}$$

Kde p_t je logaritmická cena finančního nástroje v čase t, p_{t-h} je logaritmická cena v čase $t-\ h$.

Dále lze odvodit kumulativní výnos:

$$r_{t,t} = p_t - p_0$$

Z výše uvedených rovnic plyne následující vztah:

$$r_{t,h} = r_t - r_{t-h}$$

Předpokládáme, že proces ceny finančního nástroje zůstává téměř striktně pozitivní a je konečný, aby byly p_t a r_t dobře definovány na intervalu $[0;T]$ (Andersen a kol., 2010). Kromě toho počet skokových bodů v r_t procesu je počitatelný v časovém intervalu $[0;T]$.

1.3 Dekompozice výnosů

Pokud uvažujeme, že proces vykonávaný logaritmickými výnosy r_t je bez-arbitrážní a má konečnou střední hodnotu, proces vykonávaný logaritmickou cenou p_t musí patřit do skupiny semi-martingalů. Tyto procesy umožnují jedinečnou dekompozici, jak uvádí stochastická integrální teorie (Protter, 1992). Martingal je proces, pro který platí, že v každém okamžiku realizované sekvence očekávaná hodnota další realizace nezávisí na hodnotách minulých realizací, ale je stejná současné hodnotě. Semi-martingal je definován jako proces, který může být rozložen jako suma lokálních martingalů a *cádlág* (proces s konečnou variací) (Andersen a kol., 2003).

Každý bez-arbitrážní proces vykonávaný logaritmickou cenou finančního aktiva lze jednoznačně rozdělit na sumu konečné variace, na předvídatelnou střední hodnotu logaritmického výnosu μ_t a na lokální martingal M_t:

$$r_t = p_t - p_0 = \mu_t + M_t = \mu_t + M_t^C + \Delta M_t$$

Kde lokální martingal M_t lze rozložit na lokální martingal s nekonečnou variací M_t^C a na martingal skoků (jump martingale) ΔM_t. Predikovatelný proces vykonávaný střední hodnotou logaritmického výnosu μ_t je možné dále rozložit na dvě části: spojitý proces μ_t^C a proces skoků $\Delta\mu_t$. Z výše uvedené rovnice je vidět, že proces vykonávaný logaritmickým výnosem finančního aktiva má stejné důležité vlastnosti jako proces vykonávaný cenou p_t. I když konečná variace a střední hodnota μ_t jsou předvídatelné, tento proces může být stochastický a může zahrnovat skoky, ale spojitá komponenta μ_t^C musí mít hladký průběh. Kromě toho, pokud se skok objeví v predikovatelné střední hodnotě logaritmického výnosu μ_t, musí být simultánní skok v martingalu skoků ΔM_t. Tudíž mohou existovat dva druhy skoků v procesu vykonávaném logaritmickým výnosem aktiva: nepredikovatelné skoky v případě $\Delta\mu_t = 0$, $\Delta M_t \neq 0$ a predikovatelné v případě $\Delta\mu_t \neq 0$. Často nepredikovatelné skoky mohou být způsobeny neočekávanými informacemi jak na makro, tak i na mikro úrovně. Predikovatelné skoky se objevují v tom případě, když se očekávaná informace stává na trhu dostupná. Pokud je nejistota ohledně toho, kdy ten skok nastane, skok by neměl být považován za předvídatelný. Spojitý vývoj střední hodnoty logaritmického výnosu μ_t bude spojitým při úplné absenci predikovatelných skoků.

1.4 Vlastnosti volatility

Zkoumání statistických vlastností finančních časových řad odhalilo řadu zajímavých faktů, které jsou společné pro širokou škálu finančních trhů a nástrojů. V této podkapitole budou uvedeny nejdůležitější vlastnosti volatility.

Na časových řadách výnosů finančních aktiv často lze pozorovat jev, který se nazývá shlukování volatility (Mandelbrot, 1963). Velké změny cen mají tendenci se shlukovat dohromady. Jinými slovy se na trhu střídají období vysoké a nízké volatility.

Fenoménem shluků volatility se zabývalo mnoho ekonomů. Za účelem úspěšného modelování podmíněné volatility byly vyvinuty modely ARCH, GARCH a později GARCH modely byly rozšířeny o různé další specifikace.

GARCH, FIGARCH a modely stochastické volatility pouze navrhují statistické konstrukce pro modelování volatility, ale neposkytují žádná ekonomická vysvětlení toho, proč dochází ke shlukování volatility. Heterogenita v časovém měřítku agenta se považuje za možné vysvětlení příčin shlukování volatility. Investoři s investiční strategií s dlouhodobým horizontem přirozeně pozorují dlouhodobé chování cen finančních aktiv, zatímco se obchodnici snaží využit krátkodobé cenové výkyvy. Heterogenita a s ní související HAR modely budou podrobněji probrány v kapitole č.2.

S pojmem shlukování volatility je úzce spojen pojem mean – reversion. Mean – reversion je vlastnost volatility, která spočívá v tom, že se dlouhodobě volatilita vrací na úroveň svého historického průměru. Ale na trhu můžou přetrvávat období vysoké (nebo nízké) volatility velmi dlouho. Tato skutečnost je způsobena tím, že autokorelační struktura volatility má dlouhou paměť. Kromě toho autokorelační koeficienty s rostoucím zpožděním klesají pomalým tempem, což je typické pro procesy s dlouhou pamětí.

Další vlastností volatility, která je spojená s procesy s dlouhou pamětí, je pozitivní korelace volatility s objemem odchodů provedených na trhu (Lobato a Velasco, 2000).

Empirické analýzy finančních dat vedou k leptokurtickému rozdělení nadměrných výnosů (tedy volatility), které je charakteristické tlustými konci a nadbytečným kurtózysem (Campbell a kol., 1997). Volatilita je charakteristická tím, že má nadměrnou špičatost (kurtosis). Koeficient špičatosti jako statistická charakteristika je velmi důležitý pro investory. Podle hodnoty špičatosti lze poznat pravděpodobnost nastání velkých změn ceny finančních aktiv. Získávání informace o tvaru pravděpodobnostního rozdělení je důležitý krok v modelování cen finančních aktiv. Špičatost měří, jak velký vliv na rozptyl mají odlehlé hodnoty (tails of distribution). Většina investorů je risk – free, což znamená, že investoři preferují finanční aktiva, jejichž pravděpodobnostní rozdělení mají nižší hodnotu koeficientu špičatosti. Jinými slovy čím nižší je koeficient špičatosti, tím je méně pravděpodobně, že se výnosy odchýlí od svých středních hodnot. Špičatost normálního rozdělení je 0. Koeficient šikmosti (skewness) je statistická vlastnost popisující nesymetrii pravděpodobnostního rozdělení náhodné veličiny. V případě pozitivní šikmosti a nadměrné špičatosti je více pravděpodobně, že výnosy finančního aktiva budou pozitivní. Naopak, pokud šikmost bude záporná a špičatost kladná je velmi pravděpodobně, že investoři dosáhnou záporných výnosů (Ivanovski, Narasanov a Ivanovska, 2015).

Další důležitou vlastností je asymetrie volatility (Leverage effect). Leverage effect spočívá v tom, že volatilita reaguje na negativní šok mnohem větším růstem než na pozitivní šok, který je stejně velký jako negativní (Black, 1976). Na akciových trzích volatilita vykazuje silnou negativní korelaci s tržními výnosy, ale na komoditních trzích lze pozorovat kladnou korelaci (Kristoufek, 2014). Asymetrická volatilita, která se projevuje na finančních trzích, vedla k odvození asymetrických modelů podmíněné volatility. V asymetrických modelech pozitivní informace (šok) a negativní informace působí odlišným způsobem na budoucí vývoj volatility. V důsledků existence leverage efektů bylo vyvinuto několik rozšíření GARCH modelů. Dále budou uvedeny nejznámější z těchto modelů. Jeden z prvních modelů popisující asymetrické dopady na budoucí vývoj volatility byl navržen Nelsonem (1991). To byl EGARCH(1,1)[1]. Tento model funguje s přirozeným logaritmem podmíněného rozptylu a normovanými rezidui, má následující podobu:

$$log\left(\sigma_{t|t-1}^2\right) = \alpha_0 + \alpha\left(|z_{t-1}| - E(|z_{t-1}|)\right) + \gamma z_{t-1} + \beta log\left(\sigma_{t-1|t-2}^2\right)$$

Kde $z_t = \sigma_{t|t-1}^{-1}\epsilon_t$

Výhoda použití logaritmu podmíněného rozptylu spočívá v tom, že $log\left(\sigma_{t|t-1}^2\right)$ nikdy nemůže nabývat záporných hodnot. Kromě toho v případě, že se v modelu objeví záporné parametry, není třeba je omezovat.

Model GIR(p,q) je nazván podle svých autorů (Glosen, Jagannathan a Runkle, 1993).

GIR(p,q) má následující tvar:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i\, r_{t-i}^2 + \sum_{j=1}^{p} \beta_j\, \sigma_{t-j}^2 + \sum_{i=1}^{q} \gamma_i\, I_{t-i} r_{t-i}^2$$

Tento model se snaží zachytit leverage efekt pomocí zavedení parametru γ_i, který zachycuje záporné výnosy. Parametr γ_i určuje dodatečný vliv záporného výnosu r_{t-i}^2 na podmíněný rozptyl v čase t. I_{t-i} nabývá hodnotu 1, pokud $r_{t-i}^2 < 0$, a hodnotu 0, je-li $r_{t-i}^2 > 0$.

1.5 Kvadratická variace

V této podkapitole se soustředíme na tom, jak se chová složka martingalu v rozkladu procesu vykonávaného logaritmickou cenou finančního aktiva. Tento rozklad byl proveden

[1] exponenciální GARCH

v podkapitole č.1.4. Bohužel nejsme schopni pozorovat proces M_t (lokální martingal), protože nemáme k dispozici spojitá data. Z toho důvodu používáme charakteristiky diskrétní variace, které reprezentují variační proces přes diskrétní časový interval. Spojitý rozklad provedený v podkapitole č.1.4 se mění na disktrétní formu:

$$r_{t,h} = \mu_{t,h} + M_{t,h} = \mu_{t,h} + M_{t,h}^C + \Delta M_{t,h}$$

Kde $\mu_{t,h} = \mu_t - \mu_{t-h}$, $M_{t,h}^C = M_t^C - M_{t-h}^C$ a $\Delta M_{t,h}^C = \Delta M_t - \Delta M_{t-h}$.

Pokud r_t je semi-martingal, můžeme definovat proces kvadratické variace $[r,r]_t$:

$$[r,r]_t = r_t^2 - 2\int_0^t r_{s-}\, ds$$

Z výše uvedené rovnice je zřejmé, že pokud proces s konečnou variací μ_t je spojitý, kvadratická variace procesu μ_t se rovná nule. Z toho vyplývá, že predikovatelná komponenta μ_t nemá žádný vliv na hodnotu kvadratické variace logaritmického výnosu r_t. Dále definujme kvadratickou variaci logaritmického výnosu:

$$QV_{t,h} = [r,r]_t - [r,r]_{t-h} = [M^C, M^C]_t - [M^C, M^C]_{t-h} + \sum_{t-h<s\leq t} \Delta M_s^2$$

$$QV_{t,h} = [r,r]_t - [r,r]_{t-h} = [M^C, M^C]_t - [M^C, M^C]_{t-h} + \sum_{t-h<s\leq t} \Delta r_s^2$$

Obecně se v literatuře věnované oceňování finančních derivátů předpokládá, že spojité modely pracují se spojitým časovým intervalem a příslušné difúzní procesy jsou představovány ve formě stochastických diferenciálních rovnic.

Pro každý spojitý proces vykonávaný logaritmickou cenou lze udělat dekompozici, která bude mít následující tvar (Protter, 1992):

$$r_{t,h} = \mu_{t,h} + M_{t,h} = \int_{t-h}^t \mu_s ds + \int_{t-h}^t \sigma_s dW_s$$

Kde μ_s je integrovatelný, predikovatelný stochastický proces s konečnou variací, σ_s je striktně pozitivní $cádlág$, pro který platí:

$$P\left[\int_{t-h}^t \sigma_s^2 ds < \infty\right] = 1$$

Kde W_t je Brownův pochyb.

Uveďme několik příkladů vlastností diskutovaných výše. Začneme ze speciálního případu Black-Scholse modelu oceňování opcí (Black – Scholse, 1973). V této specifikaci modelu střední hodnota je konstantní *($\mu_t = \mu$)*, martingal skoků chybí ($\Delta M_t = 0$ *)* a spojitý martingal M_c je standartní Brownův pochyb. Model má následující podobu:

$$dp_t = \mu dt + \sigma dW_t$$

Tudíž kvadratická variace v časovém intervalu $[t-h, t]$ je:

$$QV_{t,h} = \int_{t-h}^{t} \sigma^2 ds = \sigma^2 h$$

Z výše uvedené rovnice vyplývá, že se variace logaritmických výnosu v časovém intervalu h nemění. Dále uveďme Jump-diffusion model (Merton, 1976):

$$dp_t = \left(\mu - \lambda\bar{\xi}\right)dt + \sigma dW_t + \xi_t dq_t$$

Kde q_t je Poissonův proces, který není korelován se standartním Brownovým pohybem W_t, λ je konstantní intenzita skoků, ξ označuje velikost skoků a má normální rozdělení s parametry $\left(\xi, q_\xi^2\right)$.Tento difúzní proces má kvadratickou variaci v časovém intervalu $[t-h, t]$:

$$QV_{t,h} = \int_{t-h}^{t} \sigma_s^2 \, ds + \sum_{t-h \le s \le t} J_s^2 = \sigma^2 h + \sum_{t-h \le s \le t} J_s^2$$

Kde $J_t = \xi_t dq_t \neq 0$ je jenom v případě, když v procesu jsou skoky. V daném modelu kvadratická variace je také konstantní, ale liší se ve složce $\sum_{t-h \le s \le t} J_s^2$.

Poslední model, který je představen v této podkapitole, je obecný Jump-diffusion model:

$$dp_t = \mu_t dt + \sigma_t dW_t + \xi_t dq_t$$

Kvadratická variace modelu uvedeného výše je:

$$QV_{t,h} = \int_{t-h}^{t} \sigma_s^2 \, ds + \sum_{t-h \le s \le t} J_s^2$$

Kvadratická variace obecného Jump-diffusion modelu má dvě složky: integrovaný rozptyl (integrated variance) a variace skoků (jump variation). Integrovaný rozptyl má následující podobu:

$$IV_{t,h} = \int_{t-h}^{h} \sigma_s^2 \, ds$$

Variace skoků je:

$$\sum_{t-h \leq s \leq t} J_s^2$$

1.6 Realizovaná volatilita a její vlastnosti

Daná podkapitola je věnována definici a základním vlastnostem realizované volatility. Dostupnost vysokofrekvenčních dat umožnuje měřit kvadratickou variaci poměrně jednoduchým způsobem: pomocí realizovaného rozptylu (Andersen, 2003). Definice realizovaného rozptylu je následující:

$$\widehat{\mathrm{RV}_{\mathrm{t,h}}} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{r}^2_{\mathrm{t-h+}\left(\frac{\mathrm{i}}{\mathrm{n}}\right)\mathrm{h}}$$

Kde n značí počet pozorování v časovém intervalu $[t-h, t]$. Realizovaný rozptyl je nestranný a konzistentní estimátor variace procesu vykonávaného logaritmickým výnosem finančního aktiva (Barndor-Nielsen a Shephard, 2001). Matematicky se to dá zapsat níže uvedeným způsobem:

$$E\left[RV_{t,h} \middle| F_t\right] = E\left[M_{t,h}^2 \middle| F_t\right] = E\left[\widehat{RV_{t,h}} \middle| F_t\right]$$

Kde $RV_{t,h}$ značí ex-ante variaci procesu vykonávaného logaritmickým výnosem. Daná rovnice ukazuje, že ex-post realizovaný rozptyl $\widehat{RV_{t,h}}$ je nestranný estimátor očekávané ex-ante variace logaritmického výnosu finančního aktiva $(RV_{t,h})$.

$$plim_{n\to\infty} \widehat{RV_{t,h}} = RV_{t,h}$$

Z výše uvedené rovnice je vidět, že realizovaný rozptyl konverguje k variaci logaritmických výnosů podle pravděpodobnosti. Z této rovnice je také vidět, že při zvyšujícím se vzorku realizovaný rozptyl bude konzistentním estimátorem v libovolném časovém úseku $[t - h, t], h > 0$.

Aby nedocházelo k nejasnostem, které mohou vyplývat z použitých značení, upřesníme vztahy mezi výše uvedenými veličinami. Kvadratická variace QV měří rozptyl výnosů (RV), ale kvadratická variace není přímo pozorovatelná, proto skutečné měření RV je provedeno pomocí estimátoru kvadratické variace, kterým je realizovaný rozptyl $\widehat{RV_{t,h}}$. Realizovaný rozptyl měří nejen variabilitu procesu vykonávaného logaritmickým výnosem, ale také umožnuje specifikovat pravděpodobnostní rozdělení logaritmických výnosů, které je nutné znát při modelování a prognózování volatility daného procesu.

Dále definujme realizovanou volatilitu (realized volatility):

$$\widehat{RVR_{t,h}} = \sqrt{\widehat{RV_{t,h}}}$$

Kde $\widehat{RV_{t,h}}$ je realizovaný rozptyl.

1.7 Efekt fungování tržní mikrostruktury

Teoreticky, pokud máme konzistentní realizovaný rozptyl $\widehat{RV_{t,h}}$ a výnosy finančního aktiva, které mají normální rozdělení, můžeme odhadovat variaci výnosů snadným způsobem. Existují však v praxi dva problémy, které komplikují využití výhodných výsledků konvergence. Zaprvé, je potřeba mít k dispozici spojitá data, aby byla splněna podmínka konvergence $\widehat{RV_{t,h}}$ k $RV_{t,h}$. Nicméně, jsme schopni pozorovat pouze diskrétní ceny finančních aktiv. To znamená, že se při modelování objeví diskretizační chyba. Za druhé, v praxi jsou pozorování výnosů finančních aktiv kontaminovány efekty fungování tržní mikrostruktury (bid-ask spready, bid-ask bounce). Nemůžeme agregovat data za použití velmi krátkých časových intervalů. Pokud se nechceme dopustit odhadů s velkým vychýlením, musíme volit takové časové intervaly, které dostatečně zpřesňují konečný odhad. V případě využití velmi krátkých časových intervalů odhadovaná variance bude vychýlená. Kromě toho vychýlenost roste s rostoucím počtem pozorování. Byly provedeny rozsáhlé studie (Zhang a kol., 2005, Andersen a Benzoni, 2007), které se snažily najít optimální časové intervaly pro agregaci dat. Výše uvedené autoři dospěli k závěru, že se optimální časový interval nachází v rozmezí od 5 do 30 minut.

Využití 5 (nebo více) minutových intervalů přináší další problém: vyřazení velkého objemu informace. Například, máme k dispozici data evidované každou vteřinu a chceme agregovat data s využitím 5minutových intervalů, abychom redukovali efekt tržní mikrostruktury. V tomto případě ztrácíme 299 záznamů. Pokud se jedná o vysoce likvidní akcie s vyšší evidovací frekvencí, ztrácíme ještě větší objem dostupné informace.

Existují však jiná řešení problému efektu tržní mikrostruktury než vyřazení velkého objemu dat. Ve své studii (Zhang, 2005) navrhl dvoustupňový estimátor realizované variance. Dvoustupňový estimátor ($TSRV$) se počítá následujícím způsobem. V případě, že máme vteřinová data, vytvoříme dílčí vzorky stejné velikosti, kde první dílčí vzorek zahrnuje první pozorování za určitou frekvenci. Ve druhém dílčím vzorku jsou druhá pozorování za stejnou frekvenci. Pokud chceme agregovat data za 5minutový časový interval[2], dostaneme 300[3] dílčích vzorků stejné velikosti, kde první dílčí vzorek obsahuje 1, 301, 601, … pozorování, druhý vzorek 2, 302, 602, … pozorování atd. Dále spočítáme realizovaný rozptyl pro každý dílčí vzorek, dostaneme 300 odhadů realizovaného rozptylu. Tyto odhady zprůměrujeme a dostaneme zprůměrovaný estimátor realizované variance. Zprůměrovaný estimátor je lepší než standartní estimátor, protože využívá k odhadu realizované variance všechna dostupná pozorování, ale je pořad vychýleným estimátorem kvůli přítomnosti efektu tržní mikrostruktury. Existují další řešení problému vyřazení velkého objemu informací, ale v dané práci se jimi zabývat nebudeme[4].

[2] Použití jiných časových intervalů je také možné.
[3] Počet dílčích vzorků závisí na délce zvolené frekvence. Například, u třicetiminutového časového intervalu bude 1800 dílčích vzorků.
[4] Další řešení daného problému lze najít ve studiích Zhang (2006), Barndorfff-Nielsen a kol. (2008).

2. HAR modely a detekce skoků

2.1 Bipower variance

Daná kapitola je věnována problému identifikace skoků, HAR (Heterogeneous Autoregressive) modelu a jeho různým specifikacím.

Jedním z cílů při práci s HAR modely je identifikace a oddělení skoků v procesu vykonávaném logaritmickým výnosem finančního aktiva. Jak bylo uvedeno v minulé kapitole, kvadratická variace může být rozložena na dvě části: integrovaný rozptyl, který je latentní, a variaci skoků. Existuje možnost, jak identifikovat skoky ve vysokofrekvenčních datech (Barndorff-Nielsen a Shephard, 2006). K tomu lze použít dva estimátory kvadratické variace. První estimátor obsahuje jak integrovaný rozptyl, tak i variaci skoků, druhý estimátor (bipower variance) má pouze integrovaný rozptyl. Pomocí rozdílu těchto dvou estimátorů je možné identifikovat prezenci skoků v procesu vykonávaném logaritmickým výnosem aktiva. Dále definujme bipower varianci. Bipower variance je estimátorem kvadratické variace, který obsahuje pouze integrovaný rozptyl.

$$\widehat{BV_{t,h}^{1}} = \mu_1^{-2} * \sum_{i=2}^{n} \left| r_{t-h+\left(\frac{i-1}{n}\right)h} \right| \left| r_{t-h+\left(\frac{i}{n}\right)h} \right|$$

Kde $\mu_1 = \sqrt{\frac{2}{\pi}}$.

Dokonce i v přítomnosti skoků bipower variance konverguje k integrovanému rozptylu (Barndorff-Nielsen a Shephard, 2004).

$$\widehat{BV_{t,h}^{1}} \rightarrow \int_{t-h}^{h} \sigma_s^2 \, ds$$

Existuje další modifikace bipower variance (Andersen a kol., 2010). Tento estimátor je robustní vůči některým efektům tržní mikrostruktury.

$$\widehat{BV_{t,h}} = \mu_1^{-2} \frac{n}{n-2} \sum_{i=3}^{n} \left| r_{t-h+\left(\frac{i-2}{n}\right)h} \right| \left| r_{t-h+\left(\frac{i}{n}\right)h} \right|$$

Kde $\mu_1 = \sqrt{\frac{2}{\pi}}$.

$BV_{t,h}$ také konverguje k integrovanému rozptylu. Tudíž máme $BV_{t,h}$, která je konzistentním estimátorem integrovaného rozptylu, a $RV_{t,h}$, která je konzistentním estimátorem kvadratické variace. Proto můžeme definovat konzistentní estimátor variace skoků jako rozdíl mezi realizovaným rozptylem (RV_{th}) a bipower variancí (Witzany, 2013):

$$plim_{n\to\infty}\left(\widehat{RV_{t,h}} - \widehat{BV_{t,h}}\right) = \sum_{l=1}^{N_t} J_{t,h,l}^2$$

Kde N_t značí počet nenulových skoků v časovém intervalu $[t-h, t]$. Nicméně v praxi existují dva problémy spojené s využitím daného estimátoru. Zaprvé, může docházet k záporným výsledkům, které nemají žádnou ekonomickou interpretaci. Proto se v praxi daný estimátor upravuje následujícím způsobem:

$$\sum_{l=1}^{N_t} J_{t,h,l}^2 = max\left[\widehat{RV_{t,h}} - \widehat{BV_{t,h}}, 0\right]$$

Zadruhé, estimátor bude identifikovat nenulové skoky příliš často, i když identifikované skoky budou velmi malé. Proto je potřeba rozlišovat významné a nevýznamné skoky. Existuje statistický test, který pomáhá řešit daný problém (Andersen a kol.,2010). Testuje se nulová hypotéza o žádných skocích během dne.

$$Z_{t,h} = \frac{\dfrac{\widehat{RV_{t,h}} - \widehat{BV_{t,h}}}{\widehat{RV_{t,h}}}}{\sqrt{\left(\left(\frac{\pi}{2}\right)^2 + \pi - 5\right) max\left(1, \dfrac{\widehat{TQ_{t,h}}}{\left(\widehat{BV_{t,h}}\right)^2}\right)}}$$

Kde $Z_{t,h}$ má asymptotické normální rozdělení, $\widehat{TQ_{t,h}}$ je tripower quarticity, která je definována následujícím způsobem:

$$\widehat{TQ_{t,h}} = n\mu_{\frac{4}{3}}^{-3}\left(\frac{n}{n-4}\right)\sum_{i=5}^{n}\left|r_{t-h+\left(\frac{i-4}{n}\right)h}\right|^{\frac{4}{3}}\left|r_{t-h+\left(\frac{i-2}{n}\right)h}\right|^{\frac{4}{3}}\left|r_{t-h+\left(\frac{i}{n}\right)h}\right|^{\frac{4}{3}}$$

Kde $\mu_{\frac{4}{3}} = 2^{\frac{2}{3}}\frac{\Gamma\left(\frac{7}{6}\right)}{\Gamma\left(\frac{1}{2}\right)}$.

Kromě toho byl navržen jednoduchý způsob, jak oddělit variaci skoků (jump variation) od integrovaného rozptylu (Andersen a kol., 2007).

Variace skoků je následující:

$$J_{t,h} = I\left(Z_{t,h} > \Phi_\alpha\right)\left(\widehat{RV_{t,h}} - \widehat{BV_{t,h}}\right)$$

Integrovaný rozptyl:

$$C_{t,h} = I\left(Z_{t,h} \leq \Phi_\alpha\right)\widehat{RV_{t,h}} + I\left(Z_{t,h} > \Phi_\alpha\right)\widehat{BV_{t,h}}$$

Kde $I(\quad)$ je funkce indikátoru, Φ_α je hodnota distribuční funkce normálního rozdělení.

Dále existuje způsob identifikace simultánních skoků v procesech vykonávaných cenami finančních aktiv (Gilder a kol., 2014).

$$\sum_{j=1}^{N} I\{Jump_{t,i,j} > 0\} \geq 2 \quad \Rightarrow Cojump$$

$$\sum_{j=1}^{N} I\{Jump_{t,i,j} > 0\} \leq 1 \quad \Rightarrow No \quad Cojump$$

Kde $I\{Jump_\{t,i,j\} > 0\}$ je funkce indikátoru. Pokud I má hodnotu 1, znamená to, že byl identifikován skok v ceně aktiva j během časového intervalu i dne t. Co-jump je identifikován, pokud se aspoň u dvou z N aktiv objeví simultánní skok.

Dále je potřeba definovat pojem integrovaná quarticity (integrated quarticity):

$$IQ_{t,h} = \int_{t-h}^{h} \sigma_s^4 \, ds$$

Kde $t - h$ je časový interval. V případě absence skoků realizovaný rozptyl má následující rozdělení:

$$RV_{t,h} \sim N\left[IV_{t,h}, 2IQ_{t,}\right]$$

Pokud v časové řadě nejsou žádné skoky, integrovanou quarticity lze odhadnout pomocí realizované quarticity:

$$\widehat{RQ_{t,h}} = \sum_{i=1}^{n} r^4_{t-h+\left(\frac{i}{n}\right)h}$$

V tom případě, že časová řada bude obsahovat skoky, lze použit tripower quarticity a *medRQ* estimátor (Andersen a kol., 2014) k odhadu integrované quarticity. Tripower quarticity je definována výše.

$MedRQ$ vypadá následovně:

$$MedRQ = \frac{3\pi}{9\pi + 72 - 52\sqrt{3}} \frac{n^2}{n-2}$$

2.2 Teorie heterogenních trhů (THT)

Teorie heterogenních trhů (Heterogeneous Market Hypothesis) je relativně nový přístup k teorii finančních trhů, který je založen na principu heterogenity jednotlivých účastníků trhu (Muller a kol., 1997). Různé účastníky finančního trhu obchodují v různých časových periodách, což vede k asymetrickému šíření volatility v různých časových intervalech. THT úzce souvisí s Fraktální tržní hypotézou (Fractal Market Hypothesis), která je alternativní teorií k široce používané hypotéze efektivních trhů. Dle teorie heterogenních trhů se volatilita skládá z různých složek. Tento pohled na vícesložkovou strukturu volatility vychází spíše z heterogenní povahy příchozích informací než z heterogenního chování různých účastníků finančního trhu (Seďa, 2012). Daná hypotéza vysvětluje pozorovanou pozitivní autokorelaci mezi volatilitou a počtem účastníků trhu. Pokud finanční trh má homogenní povahu, autokorelace mezi volatilitou a počtem účastníků trhu je záporná, tj. čím více je na trhu obchodníků, tím rychleji cena konverguje ke své reálné tržní hodnotě. Na rozdíl od homogenních obchodníků, heterogenní obchodníci mají různé preference, odlišně se chovají v různých situacích, tím tvoří volatilitu na trhu.

Příčiny heterogenity tržních účastníků jsou různé. To jsou různé časové horizonty obchodování, informační asymetrie, institucionální omezení, geografické umístění, odlišná averze k riziku, různá doba reakce na ekonomické zprávy atd. Důsledkem heterogenity je skutečnost, že různí účastníci trhu jsou ochotni obchodovat se stejnými aktivy za různou cenu a zadávat nákupní/prodejní příkazy v odlišných tržních situacích.

V empirické časti dané diplomové práce budeme pracovat pouze s heterogenitou vyplývající z rozdílu v různých časových horizontech obchodování, protože jej lze snadno zachytit pomocí zpoždění logaritmických výnosů, zatímco jiné příčiny heterogenity jsou těžce kvantifikovatelné. Nejčastěji se rozlišují tři typy účastníků trhu podle frekvence jejich aktivity (Corsi, 2004). Prvním druhem účastníků trhu jsou obchodníci s vysokou vnitrodenní frekvencí. To jsou dealeři, tvůrci trhu a spekulanti. Druhým typem jsou účastníci, kteří realizují obchody s týdenní periodicitou: například, portfolio manažeři. Třetím typem zpravidla jsou

centrální banky, podílové fondy, pojišťovny a ostatní instituce s měsíční frekvencí obchodování. Každý z těchto druhů účastníků je zodpovědný za jinou složku volatility (parciální volatilitu). Jednotlivé složky volatility mají podobnou strukturu jako $AR(1)$ proces (další pozorování je závislé na předchozím). Tyto parciální volatility se vzájemně ovlivňují. Volatilita má kaskádní strukturu. Krátkodobá volatilita je ovlivňována střednědobou volatilitou, dlouhodobá volatilita má vliv na střednědobou, ale opak neplatí. Tento fenomén je vysvětlen tím, že účastnici trhu obchodující vnitrodenně sledují střednědobou a dlouhodobou volatilitu, ale obchodníci s dlouhodobým investičním horizontem nevěnují pozornost krátkodobé volatilitě.

2.3 HAR-RV model

Jednoduchý HAR-RV model předpokládá absenci skoků ve spojitém stochastickém procesu vykonávaném logaritmickou cenou finančního aktiva:

$$dp_t = \mu_t dt + \sigma_t dW_t + \xi_t dq_t$$

HAR-RV model má kaskádní strukturu, kterou lze popsat následujícími třemi rovnicemi:

$$\widetilde{\sigma_{t+1m}^{(m)}} = \widetilde{\alpha^{(m)}} + \widetilde{\beta^{(m)}} RV_t^{(m)} + \widetilde{\epsilon_{t+1m}^{(m)}}$$
$$\widetilde{\sigma_{t+1w}^{(w)}} = \widetilde{\alpha^{(w)}} + \widetilde{\beta^{(w)}} RV_t^{(w)} + \widetilde{\gamma^{(w)}} E_t\left[\widetilde{\sigma_{t+1m}^{(m)}}\right] + \widetilde{\epsilon_{t+1m}^{(w)}}$$
$$\widetilde{\sigma_{t+1d}^{(d)}} = \widetilde{\alpha^{(d)}} + \widetilde{\beta^{(d)}} RV_t^{(d)} + \widetilde{\gamma^{(d)}} E_t\left[\widetilde{\sigma_{t+1w}^{(w)}}\right] + \widetilde{\epsilon_{t+1d}^{(d)}}$$

Kde $\widetilde{\sigma_t^{(d)}}$,$\widetilde{\sigma_t^{(w)}}$,$\widetilde{\sigma_t^{(m)}}$ značí denní, týdenní a měsíční parciální rozptyly, $RV^{(d)}$, $RV^{(w)}$, $RV^{(m)}$ jsou denní, týdenní a měsíční realizované rozptyly, $\widehat{\epsilon^{(d)}}$, $\widehat{\epsilon^{(w)}}$, $\widehat{\epsilon^{(m)}}$ jsou náhodné chyby třech časových horizontů, α, β, γ jsou parametry modelu. V tomto modelu nepozorovaný parciální rozptyl závisí na ex-post realizovaném rozptylu stejné frekvence a na očekávané hodnotě parciálního rozptylu s nižší frekvencí obchodování. Také HAR model lze zapsat v logaritmické a odmocninové podobě. V aplikační části této práce se zaměříme na analýzu HAR modelů v odmocninové specifikaci.

Dále provedeme rekurzivní substituci. Dostaneme model v následující podobě:

$$\sigma_{t+1d}^{(d)} = \alpha + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \widetilde{\epsilon_{t+1d}^{(d)}}$$

Ve výše uvedené rovnici jsou na pravé straně tři pozorované proměnné, ale na levé straně rovnice je jedna latentní proměnná. Jinými slovy se rozptyl (v daném případě denní) modeluje

jako lineární funkce zpožděných realizovaných rozptylů, které jsou agregovány přes různé časové horizonty. Jelikož známe, že realizovaný rozptyl je estimátorem latentního rozptylu, platí vztah:

$$\sigma_{t+1d}^{(d)} = RV_{t+1d}^{(d)} + \epsilon_{t+1d}^{(d)}$$

Po substituci dostaneme výslednou regresní rovnici:

$$RV_{t+1d}^{(d)} = \alpha + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \epsilon_{t+1d}$$

Týdenní a měsíční realizované rozptyly se počítají jako průměrné hodnoty denního rozptylu za poslední týden (5 dnů) a za poslední měsíc (22 dnů):

$$RV_t^{(w)} = \frac{1}{5}\left(RV_t^{(d)} + RV_{t-1}^{(d)} + \cdots + RV_{t-4}^{(d)}\right)$$

$$RV_t^{(m)} = \frac{1}{22}\left(RV_t^{(d)} + RV_{t-1}^{(d)} + \cdots + RV_{t-21}^{(d)}\right)$$

Logaritmická verze HAR-RV modelu vypadá následovně:

$$log\left(RV_{t+1d}^{(d)}\right) = \alpha + \beta^{(d)} log\left(RV_t^{(d)}\right) + \beta^{(w)} log\left(RV_t^{(w)}\right) + \beta^{(m)} log\left(RV_t^{(m)}\right) + \epsilon_{t+1d}$$

Logaritmická transformace modelu zajištuje, aby závislé proměnné nabývaly pouze kladných hodnot, což je výhodou dané specifikace modelu. Ekonomická interpretace výsledků však je obtížnější ve srovnání s ne – logaritmickou specifikací.

V první specifikaci HAR-RV modelu (Corsi, 2004) se předpokládalo, že rezidua mají $iid\ N(0,1)$ rozdělení. Později empirické studie (Andersen a kol., 2010) prokázaly, že rezidua mají problém heteroskedasticity. I když nehomogenita rozptylu je závažný problém, parametry modelu jsou pořad nevychýlené, jenom standardní chyby nejsou správné.

2.4 HAR-RV-J/HAR-RV-CJ modely

Jak bylo zmíněno v minulé kapitole HAR-RV model předpokládá, že proces vykonávaný logaritmickou cenou je spojitý. HAR-RV-J model (Andersen a kol., 2007) je rozšířením původního HAR-RV modelu. V daném modelu se předpokládá, že se proces vykonávaný cenou skládá ze spojité a nespojité částí:

$$dp_t = \mu dt + \sigma dW_t + \xi_t dq_t$$

Do regresní rovnice HAR-RV modelu přidáme vysvětlující proměnné – skoky (J_t):

$$RV_{t+1d}^{(d)} = \alpha + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \beta_J^{(d)} J_t^{(d)} + \beta_J^{(w)} J_t^{(w)} + \beta_J^{(m)} J_t^{(m)} + \epsilon_{t+1}$$

Předpokládá se, že dodatečné regresory ($J_t^{(d)}$,$J_t^{(w)}$,$J_t^{(m)}$) přispějí ke lepšení predikčních schopností původního HAR-RV modelu tím, že pochyby skoků budou odděleny od plynulého spojitého vývoje. Varianci skoků je možné agregovat na různých časových horizontech, stejně jak to bylo u realizované variance (realizované volatility). Pokud při modelování realizovaného rozptylu rozhodneme použit denní, týdenní a měsíční variance skoků, do HAR-RV-J modelu se přidají 3 dodatečné vysvětlující proměnné.

Při zápisu daného modelu v logaritmické formě se vyskytuje drobný problém. Pokud v procesu vykonávaném logaritmickou cenou jsou dny s absencí skoků, vysvětlující proměnná bude mít nulovou hodnotu, což je problém. Proto lze regresní rovnici lehce upravit následujícím způsobem:

$$log\left(RV_{t+1d}^{(d)}\right) = \alpha + \beta^{(d)} log\left(RV_t^{(d)}\right) + \beta^{(w)} log\left(RV_t^{(w)}\right) + \beta^{(m)} log\left(RV_t^{(m)}\right) + \beta_J^{(d)} log\left(1 + J_t^{(d)}\right) + \epsilon_{t+1}$$

Existuje další modifikace HAR modelu, je to HAR-RV-CJ model (Andersen a kol., 2007). Tento model je založen na dekompozici realizovaného rozptylu na dvě komponenty: spojitou část a na skoky. Zaprvé, je třeba spočítat průměrné komponenty:

$$J_t^{(w)} = \frac{1}{5}\left(J_t^{(d)} + J_{t-1}^{(d)} + \cdots + J_{t-4}^{(d)}\right)$$

$$J_t^{(m)} = \frac{1}{22}\left(J_t^{(d)} + J_{t-1}^{(d)} + \cdots + J_{t-21}^{(d)}\right)$$

$$C_t^{(w)} = \frac{1}{5}\left(C_t^{(d)} + C_{t-1}^{(d)} + \cdots + C_{t-4}^{(d)}\right)$$

$$C_t^{(m)} = \frac{1}{22}\left(C_t^{(d)} + C_{t-1}^{(d)} + \cdots + C_{t-21}^{(d)}\right)$$

Komponenty J_t a C_t jsou definovány v podkapitole č.2.1. Regresní rovnice HAR-RV-CJ vypadá následovně:

$$RV_{t+1d}^{(d)} = \alpha + \beta_C^{(d)} C_t^{(d)} + \beta_C^{(w)} C_t^{(w)} + \beta_C^{(m)} C_t^{(m)} + \beta_J^{(d)} J_t^{(d)} + \beta_J^{(w)} J_t^{(w)} + \beta_J^{(m)} J_t^{(m)} + \epsilon_{t+1}$$

Logaritmická verze daného modelu je podobná logaritmické specifikaci HAR-RV-J modelu:

$$log\left(RV_{t+1d}^{(d)}\right) = \alpha + \beta_C^{(d)} log\left(C_t^{(d)}\right) + \beta_C^{(w)} log\left(C_t^{(w)}\right) + \beta_C^{(m)} log\left(C_t^{(m)}\right)$$
$$+ \beta_J^{(d)} log\left(1 + J_t^{(d)}\right) + \beta_J^{(w)} log\left(1 + J_t^{(w)}\right) + \beta_J^{(m)} log\left(1 + J_t^{(m)}\right) + \epsilon_{t+1}$$

2.5 HAR-Q a HAR-Q-J modely

Beta koeficienty v HAR modelech jsou ovlivněny chybami měření realizované volatility (realizovaného rozptylu). Proto byl vyvinut HAR-Q model (Bollerslev, Patton a Quaedvieg, 2015), který je efektivním rozšířením původního HAR-RV modelu. HAR-Q (quarticity) model přímo upravuje beta koeficienty v poměru k velikosti odpovídajících chyb měření. Pokud chyby měření budou relativně vysoké, beta parametry by měly být nižší. V případě nízkých chyb měření, parametry by měly být vyšší, aby se zachytila perzistence realizovaného rozptylu. HAR-Q-F(Full) má následující podobu:

$$RV_t^{(d)} = \alpha + \left(\beta^{(d)} + \beta_{1Q} RQ_{t-1}^{(d)\left(\frac{1}{2}\right)}\right) RV_{t-1}^{(d)} + \left(\beta^{(w)} + \beta_{2Q} RQ_{t-1}^{(w)\left(\frac{1}{2}\right)}\right) RV_{t-1}^{(w)}$$
$$+ \left(\beta^{(m)} + \beta_{3Q} RQ_{t-1}^{(m)\left(\frac{1}{2}\right)}\right) RV_{t-1}^{(m)} + \epsilon_t$$

Kde $RQ_{t-1}^{(d)}$, $RQ_{t-1}^{(w)}$, $RQ_{t-1}^{(m)}$ jsou denní, týdenní a měsíční realizované quarticity. Týdenní a měsíční realizované quarticity se počítají jako průměrné hodnoty denní quarticity za poslední týden (5 dnů) a za poslední měsíc (22 dnů):

$$RQ_t^{(w)} = \frac{1}{5}\left(RQ_t^{(d)} + RQ_{t-1}^{(d)} + \cdots + RQ_{t-4}^{(d)}\right)$$
$$RQ_t^{(m)} = \frac{1}{22}\left(RQ_t^{(d)} + RQ_{t-1}^{(d)} + \cdots + RQ_{t-21}^{(d)}\right)$$

Velikost chyb měření realizovaného rozptylu klesá s rostoucím časovým horizontem. Proto úprava beta parametru u denně zpožděného realizovaného rozptylu *($RV_{t-1}^{(d)}$)* je důležitější než úpravy týdenního a měsíčního parametrů. To také znamená, že při odhadu standartního HAR-RV modelu část váhy bude přesunuta od „hlučného“ denního zpoždění k méně „hlučným“ týdennímu a měsíčnímu zpoždění.

HAR-Q model má pouze upravený beta parametr u denně zpožděného realizovaného rozptylu:

$$RV_t^{(d)} = \alpha + \left(\beta^{(d)} + \beta_{1Q} RQ_{t-1}^{(d)\left(\frac{1}{2}\right)}\right) RV_{t-1}^{(d)} + \beta^{(w)} RV_{t-1}^{(w)} + \beta^{(m)} RV_{t-1}^{(m)} + \epsilon_t$$

HAR-Q model je schopen přidat větší váhu denně zpožděnému realizovanému rozptylu (*RQ* je nízká) při relativně přesném měření *RV*, a snížit váhu při relativně špatném měření *(RQ* je vysoká).

Pokud do regresní rovnice HAR-Q modelu přidáme další vysvětlující proměnnou – skoky (J_t), dostaneme HAR-Q-J model v podobě:

$$RV_t^{(d)} = \alpha + \left(\beta^{(d)} + \beta_{1Q} RQ_{t-1}^{(d)\left(\frac{1}{2}\right)}\right) RV_{t-1}^{(d)} + \beta^{(w)} RV_{t-1}^{(w)} + \beta^{(m)} RV_{t-1}^{(m)} + \beta_J^{(d)} J_{t-1}^{(d)} + \beta_J^{(w)} J_{t-1}^{(w)} + \beta_J^{(m)} J_{t-1}^{(m)} + \epsilon_t$$

V empirické části diplomové práce budou analyzovány oba HAR-Q a HAR-Q-J modely.

2.6 Porovnání predikčních schopností modelů

Predikování různých ekonomických indikátorů (HDP, inflace, volatilita atd.) je jedním z hlavních cílů ekonometrie. Existuje mnoho různých přístupů k porovnání predikčních schopností ekonometrických modelů, však v analytické části dané diplomové práce budou použity tři přístupy: odmocnina ze střední kvadratické chyby (*RMSE*[5]), index determinace běžně označovány jako R^2 a Mincer-Zarnowitz regrese.

RMSE je veličina, která vyjadřuje přesnost predikce modelu. Daná veličina měří odmocninu průměru čtverců rozdílů mezi predikovanými a skutečnými hodnotami:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}$$

Kde *n* je počet generovaných predikcí, *Y* je vektor skutečných hodnot, $\widehat{Y}$ je vektor odhadů. V praxi *RMSE* je téměř vždy pozitivní (nikoli nula). Odmocnina ze střední kvadratické chyby je mírou kvality ekonometrického modelu. Hodnoty *RMSE* blíže nule jsou lepší.

[5] Root mean squared error

Index determinace (R^2) je mírou kvality ekonometrického modelu, který nabývá hodnoty v intervalu [0,1]. Pokud R^2 má hodnotu 0, to znamená, že model nepřináší žádnou informaci. Pokud R^2 je 1 – predikce ekonometrického modelu jsou dokonalé.

Index determinace vypadá následovně:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \left(Y_i - \widehat{Y}_i\right)^2}{\sum \left(Y_i - \overline{Y}_i\right)^2}$$

Kde *RSS* je residual sum of squares, *TSS* je total sum of squares.

Mincer-Zarnowitz (1969) navrhli zajímavý přístup hodnocení predikční schopnosti modelu. Základní myšlenkou daného přístupu je provedení regrese (s konstantou) mezi skutečnými a predikovanými hodnotami. Pokud model je správný, konstanta by se měla rovnat nule, β koeficient jedné a index determinace (R^2) by měl dosahovat vysokých hodnot. Například, pokud uvažujeme HAR-RV model, rovnice Mincer-Zarnowitz regrese vypadá následovně:

$$RV_t^{(d)} = \alpha + \beta \widehat{RV_t^{(d)}} + \epsilon_t$$

Kde $RV_t^{(d)}$ je skutečný denní realizovaný rozptyl v čase t a $\widehat{RV_t^{(d)}}$ je predikovaná hodnota realizovaného rozptylu za základě odhadnutého modelu. Rovnice Mincer-Zarnowitz regrese pro logaritmickou a odmocninovou specifikace modelu by vypadala podobně.

2.7 Jiné metody modelování volatility na vysokofrekvenčních datech

V současné době modelování a prognózování volatility hraje na finančních trzích důležitou roli. Kvůli tomu, že se volatilita využívá při oceňování derivátů (Hull, 2012), při konstrukci VaR (Value at risk), Expected Short fall apod., hodně ekonomů se snažilo vytvořit co nejlepší model, který by sloužil ke správnému modelování a předvídání volatility. Tato podkapitola shrnuje nejznámější metody modelování volatility finančních aktiv na vysokofrekvenčních datech.

ARIMA-RV a ARFIMA-RV modely

Modely typu ARIMA a ARFIMA se dají použit pro modelování realizované volatility (Andersen a kol., 2003). Metodologie modelů ARIMA reprezentuje standardní přístup používaný pro modelování podmíněné střední hodnoty časových řad s lineární závislostí (Arlt a Arltová, 2009). ARIMA model kombinuje AR(p)[6] a MA(q)[7] procesy pro zachycení

[6] Autoregresní proces p-tého řadu
[7] Proces klouzavých průměrů řadu q

vztahů autokorelační a parciální autokorelační funkcí stochastického procesu. Tento stochastický proces může být jak stacionární (integrovaný řadu $d = 0$), tak i nestacionární (integrovaný řadu $d \geq 1$). Tento model dále lze rozšířit tak, aby parametr d (integrovaný řad) nabýval neceločíselných hodnot. Daný proces se nazývá ARFIMA a je frakcionálně diferencovaným procesem. Modely typu ARFIMA zachycují dlouhou paměť v autokorelační struktuře volatility, která je pomalu klesající.

ARIMA(p,d,q) s ARFIMA(p,d,q) modely se dají zapsat níže uvedeným způsobem:

$$\phi(L)(1-L)^d(RV(t,\Delta) - \mu_{RV}) = \psi(L)z_t$$

Kde RV je realizovaná volatilita, μ_{RV} je nepodmíněná střední hodnota realizované volatility, d je parametr diferencí (fraktální), L je operátor zpoždění, z_t je proces bílého šumu a $\phi(L)$, $\psi(L)$ jsou polynomiálními operátory AR a MA, které mají následující podobu:

AR operátor $\phi(L)$ je: $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$

MA operátor $\psi(L)$ je: $\psi(L) = 1 - \psi_1 L - \psi_2 L^2 - \cdots - \psi_q L^q$

V modelech ARIMA-RV a ARFIMA-RV parametr diferencí d stanovuje délku paměti v autokorelační struktuře realizované volatility. Pokud $d = 0$, dané modely se redukují na ARMA-RV(p,d) model, který je stacionární a má krátkou paměť. Pokud $d = 1$, tyto modely odpovídají ARIMA-RV(p,1,q) modelu, který je nestacionární a má nekonečnou paměť.

Realizovaná volatilita v modelech stochastické volatility

Realizovaná volatilita, bipower volatilita a další power-variation estimátory mohou být použity v modelech stochastické volatility (Shephard, 2005). Dané estimátory poskytují relativně přesné odhady volatility, však kvůli existenci efektu tržní mikrostruktury odhady parametrů modelů realizované volatility mohou být zkreslené. Modely stochastické volatility řeší tento problém[8].

Základní SV-RV model používá realizovanou varianci jako dodatečný zdroj informací při odhadu modelu log-SV (Takahashi a kol., 2009). K odhadu modelu se používají dvě časové řady: časová řada realizované variance $RV(t)$ a časová řada logaritmických výnosů $r(t)$.

[8] Více o modelech stochastické volatility lze najít v disertační práci Fičura (2018).

SV-RV model lze zapsat následujícím způsobem:

$$r(t) = \mu + \sigma(t)\epsilon(t)$$

$$h(t) = \alpha + \beta h(t-1) + \gamma\epsilon_v(t)$$

$$log\big(RV(t)\big) = h(t) + \sigma_{RV}\epsilon_{RV}(t)$$

Kde $h(t) = log\big(\sigma^2(t)\big)$ a $\epsilon(t)$, $\epsilon_v(t), \epsilon_{RV}(t)$ mají normální rozdělení se střední hodnotou 0 a rozptylem 1.

Existují rozšíření SV-RV modelů[9]. Původní model se dá rozšířit přidáním skoků do rovnice výnosů (Fičura a Witzany, 2015).

[9] Více o rozšířeních SV-RV modelu lze najít v studiích Fičura a Witzany (2015) a v Fičura a Witzany (2017).

3. Aplikace modelů

V dané kapitole modely prognózování volatility budou aplikovány na časovou řadu indexu Moskevské burzy cenných papírů (MOEX). Hlavním cíle této analýzy je posouzení predikčních schopností modelů a výběr nejvhodnějšího modelu pro předvídání volatility. K analýze jsou vybrány následující modely: HAR-RV, HAR-RV-J, HAR-Q a HAR-Q-J. Budeme se zabývat jak prognózováním realizované volatility na následující den, tak i předvídáním agregované volatility na delší časové horizonty, konkrétně na 1 týden. Všechny modely budou analyzovány v ne-logaritmické podobě. Pomocí vybraných HAR modelů budeme modelovat a prognózovat realizovanou volatilitu. Daná specifikace modelů byla vybrána z důvodů jasnější interpretace parametrů. Veškerá ekonometrická analýza byla provedena v softwaru R.

3.1 Analýza a popis dat

V této podkapitole se zaměříme na popis dat, neboť podrobná znalost analyzovaných dat je nutným předpokladem k lepšímu porozumění čtenářem empirické části dané diplomové práce. Jak již bylo zmíněno v úvodu, v dané práci se používají vysokofrekvenční data pro modelování volatility. K analýze byla vybrána časová řada indexu Moskevské burzy (MOEX) za periodu od 1. června 2005 do 31. května 2019. Index Moskevské burzy se skládá z 50 nejvíce likvidních akciových titulů. Index se počítá každou vteřinu.

3.1.1 Vysokofrekvenční data

Nejdříve začneme z obecného popisu problematiky práce s vysokofrekvenčními daty. Ekonometrická analýza vysokofrekvenčních dat je stále docela nová oblast finanční ekonometrie. Samozřejmě technologický pokrok hraje klíčovou roli v dostupnosti a schopnosti zpracování dat daného typu. Potřeba využití vysokofrekvenčních dat vyplývá ze snahy obchodníků mít co nejvíce informací pro svá rozhodnutí. Na jedné straně čím více informace máme, tím lépe můžeme pochopit fungování finančních trhů. Na druhé straně zpracování vysokofrekvenčních data je složitá úloha, jejíž řešení může vest k různým výsledkům závislým na učiněných rozhodnutích. Existuje několik problémů spojených se zpracováním vysokofrekvenčních dat. Zaprvé, prvotní záznamy obsahují chyby. Tyto chyby mohou být způsobeny překlepy při zaznamenání nebo konverzí dat. Zpravidla počet chyb je přímo úměrný intenzitě obchodování. Za druhé, realizované obchody mohou být zrušeny nebo odloženy, což také způsobuje chyby v záznamech. Prvotní *tick* data jsou asynchronní, neboť se finanční aktiva s vyšší likviditou obchodují s vyšší frekvencí než aktiva s nižší likviditou. Kromě toho se *tick*

frekvence u jednoho finančního aktiva mění během dne. Občas jsme schopni identifikovat odlehlá pozorování a odstranit je beze změny statistických vlastností dat. Ale častěji není možné identifikovat odlehlá pozorování za účelem jejich vyřazení. Pokud algoritmus očišťování vysokofrekvenčních dat je příliš přísný, lze ztratit významné množství informace. V opačném případě je dost pravděpodobně, že zpracovaná data budou obsahovat příliš hodně chyb a budou nepoužitelná. Zpracování vysokofrekvenčních dat je komplexní úloha a neexistuje žádný jediný algoritmus, který by se dalo použit pro očišťování *tick* dat všech finančních aktiv. Je potřeba provést analýzu dat a pak upravit algoritmus vhodným způsobem[10].

3.1.2 Popis dat

Vysokofrekvenční data používaná v dané práci již byla odfiltrována a synchronizována dle stejného času. Jedná se o 5minutová data od 1. června 2005 do 31. května 2019, což je ekvivalentně 3 504 obchodním dnům. Graf č.1 zobrazuje cenový vývoj MOEX indexu.

Graf 1 Cenový vývoj MOEX indexu za období (2005-2019)

Zdroj: vlastní zpracování

Z grafu č.1 je vidět, že cena indexu prudce klesla v roce 2009 kvůli finanční krizi. Takové silné cenové výkyvy mohou mít určitý dopad na výsledky provedené analýzy v kapitole č.3.

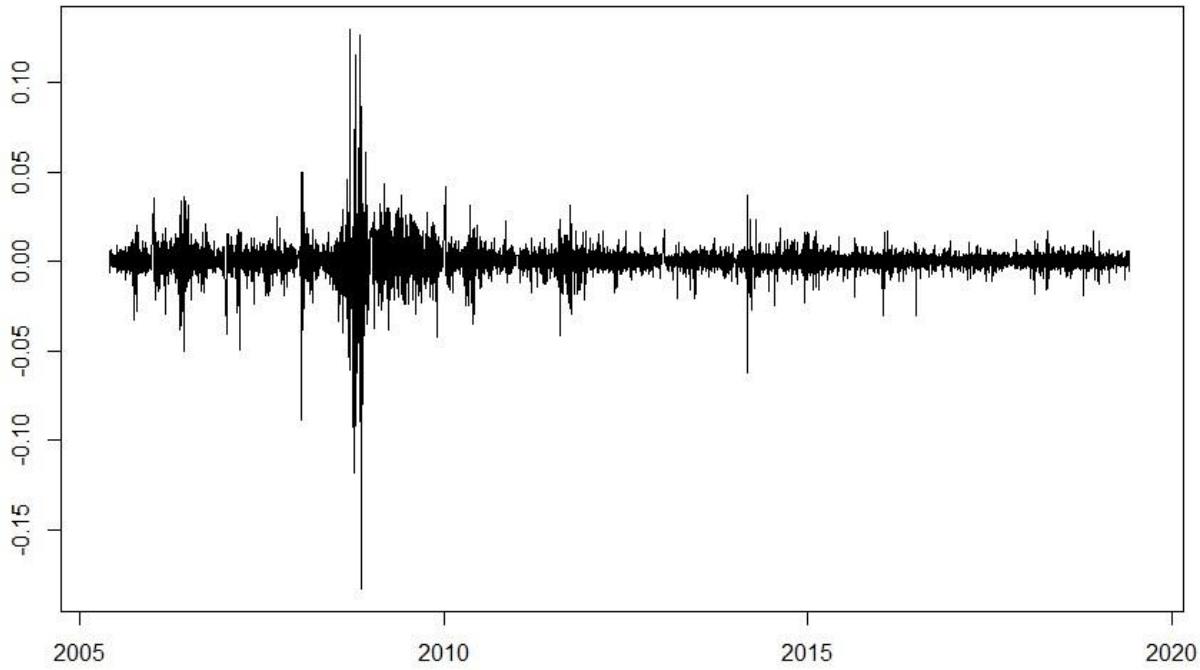
[10] Více o zpracování vysokofrekvenčních dat Falkenberry (2002).

Náš datový vzorek pokrývá periodu finanční krize 2007-2009 a také periodu před a po krizi. Ceny jsou synchronizovány podle obchodních hodin na Moskevské burze, od pondělí do pátku od 10:00 do 18:45 hodin. Celkem za danou periodu je k dispozici 372 254 pozorování. Nejprve naše časová řada byla transformována na časovou řadu logaritmických výnosů. Logaritmické výnosy $r_{t,h}$ byly spočítány dle vztahu:

$$r_{t,h} = ln\left(\frac{P_t}{P_{t-h}}\right)$$

Kde P_t je vnitrodenní pozorování ceny v čase t, h je 5minutový interval mezi dvěma po sobě jdoucími pozorováními. Grafu č.2 znázorňuje vývoj logaritmických výnosů.

Graf 2 Vývoj logaritmických výnosů za období (2005-2019)



Zdroj: vlastní zpracování

Z grafu č.2 je zřejmé, že minimální hodnota logaritmických výnosů byla dosažena v roce 2009. Pak od roku 2009 lze pozorovat relativně stabilní vývoj časové řady výnosů. Kromě toho z grafu č.2 lze pozorovat fenomén shlukování volatility (viz kapitola 1.4). Je vidět, jak se na trhu střídají období vysoké a nízké volatility.

V tabulce č.1 jsou představovány popisné statistiky logaritmických výnosů MOEX indexu.

Tabulka 1 Popisné statistiky logaritmických výnosů MOEX indexu za období (2005-2019)

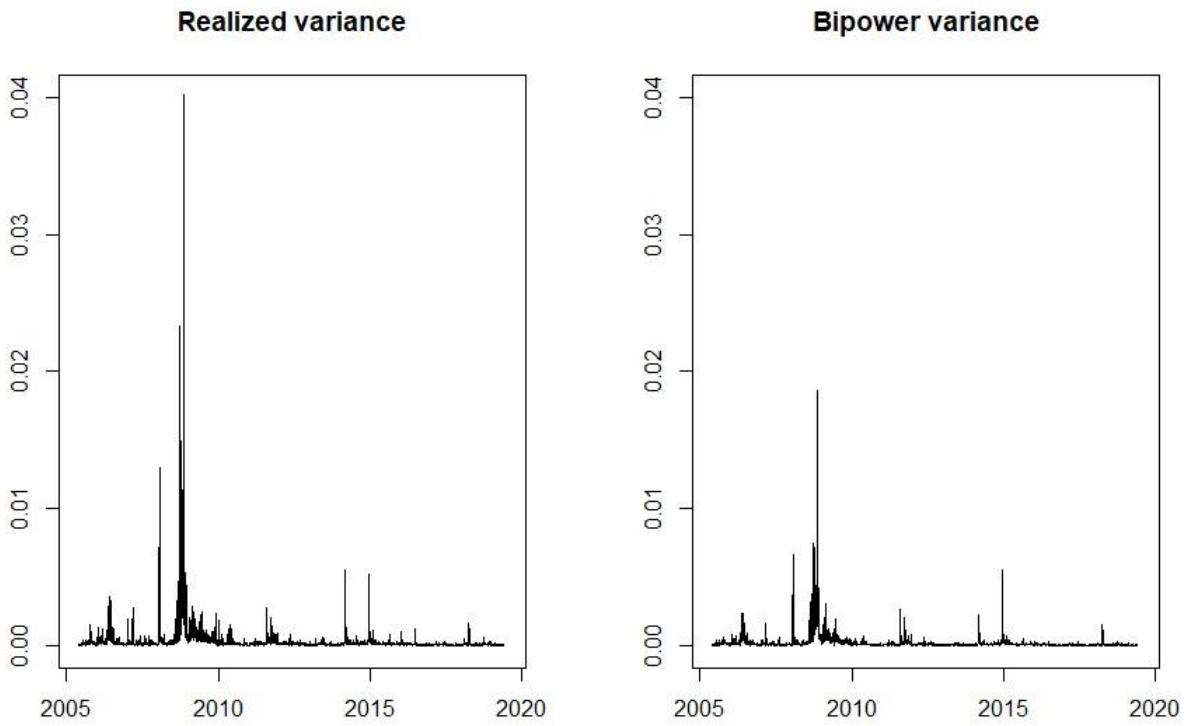
Min. hodnota	-0.18308752
1.Kvantil	-0.00049100
Medián	0.00000000
Střední hodnota	0.00000399
3.Kvantil	0.00051254
Max. hodnota	0.12972261
Šikmost	-2.64855600
Špičatost	871.4288

Zdroj: vlastní zpracování

Z tabulky č.1 je vidět, že logaritmické výnosy mají relativně vysoké minimální (-0.18308752) a maximální (0.12972261) hodnoty. Kromě extrémních hodnot, lze pozorovat lehce pozitivní střední hodnotu (0.00000399) a negativní šikmost (-2.648556), což je charakterním rysem akcií. Negativní hodnota šikmosti signalizuje o tom, že levý konec distribuce logaritmických výnosů je delší než její pravý konec, což znamená, že výrazná část distribuce je koncentrována v její pravé části. Také výnosy mají extrémně vysokou špičatost (871.4288), z čeho je vidět, že se výnosy relativně silně odchylují od své průměrné hodnoty.

V souladu s definicemi, které byly uvedeny v kapitolách č.1 a č.2, spočítané logaritmické výnosy MOEX indexu lze použit k výpočtu realizované variance, bipower variance, realizované quarticity a tripower quarticity. Na grafu č.3 jsou znázorněny vypočtené realizovaná variance a bipower variance.

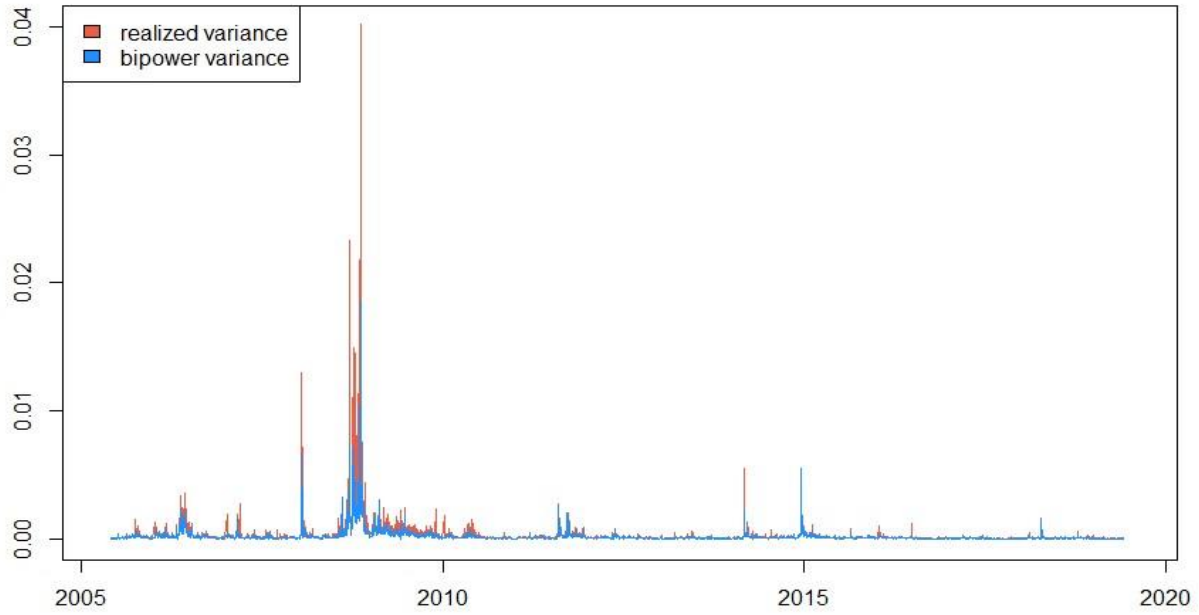
Graf 3 Realizovaná variance a bipower variance



Zdroj: vlastní zpracování

Je opět vidět, že v období finanční krize (2007-2009) realizovaná variance a bipower variance dramaticky vzrostly. Také lze pozorovat určité rozdíly ve vývoji realizované variance a bipower variance. Tyto rozdíly jsou znázorněny pomocí grafu č.4.

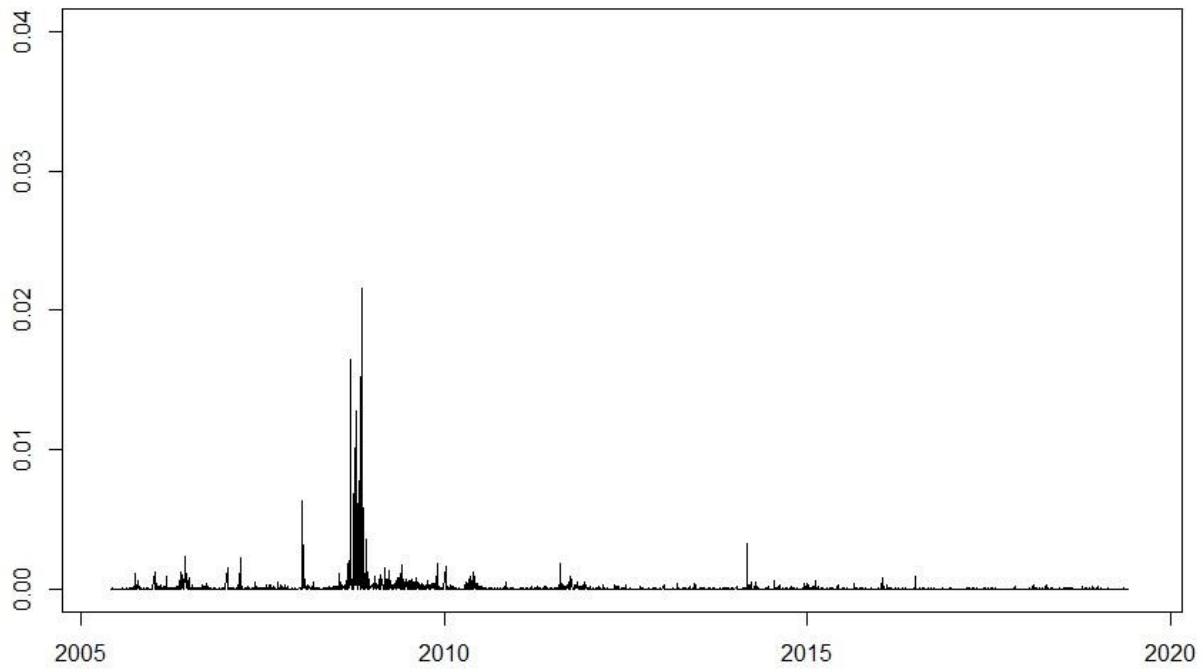
Graf 4 Porovnání Realizované variance vs Bipower variance



Zdroj: vlastní zpracování

Jak již bylo zmíněno v kapitole č.2.1 variance skoků může být odhadnuta pravě pomocí rozdílu mezi realizovanou variancí a bipower variancí. Pokud vypočtený estimátor neupravíme, dostaneme záporné hodnoty skoků, což nemá žádné ekonomické vysvětlení. Proto je potřeba upravit daný estimátor o záporné hodnoty. Kromě toho skoky mohou být identifikovány příliš často, proto je potřeba rozlišovat významné a nevýznamné skoky. Graf č.5 zobrazuje varianci skoků za analyzované období. Při zkoumání skoků je potřeba mít na paměti, že se analýza provádí na dlouhém období, proto lze očekávat poměrně velké množství skoků.

Graf 5 Variance skoků



Zdroj: vlastní zpracování

Je vidět, že se největší skoky objevovaly během finanční krize. Pak od roku 2010 až do konce sledované periody velikost skoků byla relativně malá. Odhad variance skoků byl proveden za účelem použití při konstrukci HAR-RV-J a HAR-Q-J modelů. Dále byly vypočítány základní popisné statistiky použitých estimátorů. V tabulce č.2 jsou shrnuty popisné statistiky realizované variance, bipower variance a variance skoků.

Tabulka 2 Popisné statistiky RV, BV, J

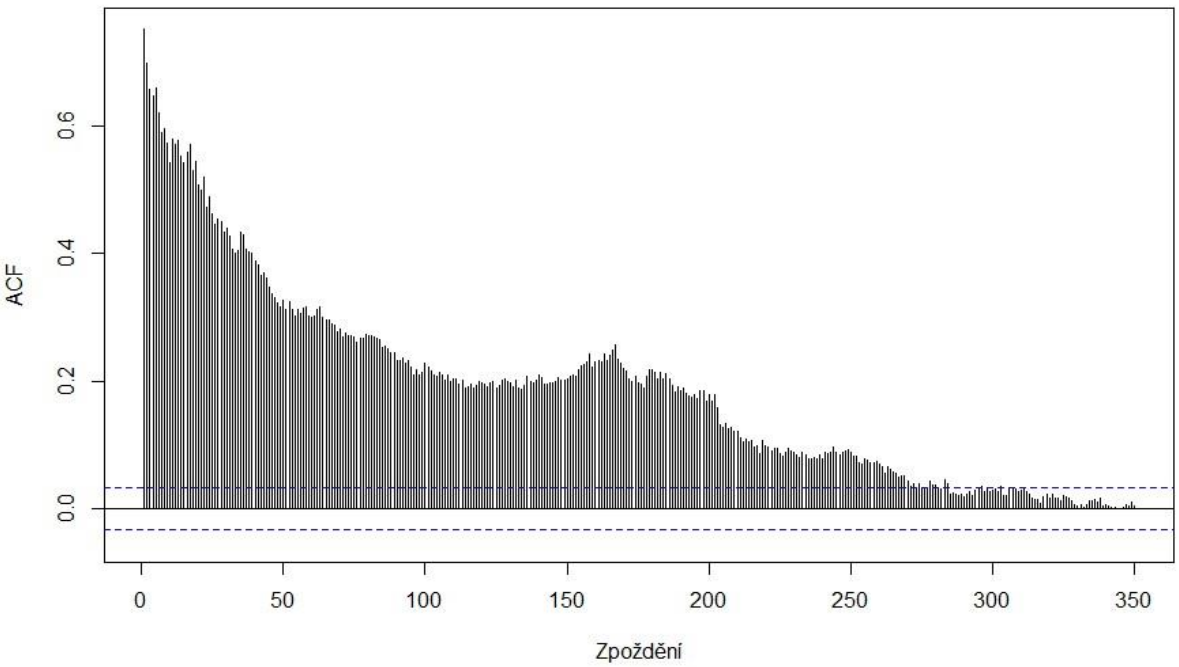
	RV	*BV*	*J*
Min. hodnota	0.00000427	0.00000000	0.00000000
1.Kvantil	0.00006486	0.00004952	0.00000000
Medián	0.00010978	0.00008428	0.00001551
Střední hodnota	0.00031182	0.00019976	0.00010892
3.Kvantil	0.00023004	0.00016750	0.00004970
Max. hodnota	0.04014860	0.01857292	0.02157567
Šikmost	18.59129	15.89608	19.08811
Špičatost	500.169	429.1585	452.6756

Zdroj: vlastní zpracování

Z tabulky č.2 je vidět lehce pozitivní střední hodnotu u realizované variance, bipower variance a variance skoků. Hodnota špičatosti je extrémně velká stejně jako v případě logaritmických výnosů.

Graf č.6 znázorňuje autokorelační strukturu realizované volatility vypočítanou pomocí agregace 5minutových logaritmických výnosů.

Graf 6 Autokorelační funkce realizované volatility



Zdroj: vlastní zpracování

Z grafu je vidět, že autokorelace klesá pomalým tempem, což je typické pro procesy s dlouhou pamětí[11]. Proto lze říci, že realizovaná volatilita vykazuje dlouhou paměť. To je hlavním důvodem využití HAR modelů.

3.2 Aplikace modelů

V dané podkapitole bude provedena aplikace čtyř vybraných modelů: HAR-RV, HAR-RV-J, HAR-Q a HAR-Q-J. Tyto modely budou analyzovány jak na in-sample, tak i na out-sample datech. In-sample vzorek je od 1. června 2005 do 16. června 2015. Out-sample vzorek u všech čtyř modelu je stejné délky – 1000 pozorování, což odpovídá periodě od 17. června 2015 až do 31. května 2019.

3.2.1 HAR-RV model

Začneme původním HAR-RV modelem, který lze zapsat v následující podobě:

$$RV_t^{(d)} = \alpha + \beta^{(d)} RV_{t-1}^{(d)} + \beta^{(w)} RV_{t-1}^{(w)} + \beta^{(m)} RV_{t-1}^{(m)} + \epsilon_t$$

Tento model má 3 vysvětlující proměnné: denní, týdenní a měsíční realizované volatility. K odhadu modelu využijme váženou metodu nejmenších čtverců (WLS)[12] z toho důvodu, že se základní metoda nejmenších čtverců (OLS)[13] příliš zaměřuje na období vysoké volatility a přidává menší váhu periodám s nižší volatilitou. Úroveň volatility se značně mění během zkoumaného období, což signalizuje o přítomnosti problému heteroskedasticity. K tomu, abychom byli schopni odhadnout HAR-RV model metodou WLS, je potřeba nejdříve odhadnout model základní metodou nejmenších čtverců (OLS).

$$\widehat{\beta_{OLS}} = (X^T X)^{-1} X^T Y$$

Kde $\widehat{\beta_{OLS}}$ je matice odhadnutých β koeficientů OLS metodou, X je matice vysvětlujících proměnných (realizované denní, týdenní, měsíční volatility), Y je matice vysvětlované proměnné. Dále sestrojíme váhy jako inverzní hodnoty vyrovnaných hodnot.

$$w_t = \frac{1}{\overline{y}_t}$$

Kde y_t jsou vyrovnané hodnoty.

[11] Procesy s krátkou pamětí vykazují exponenciální pokles autokorelace.
[12] Weighted Least Squares
[13] Ordinary Least Squares

Dalším krokem je provedení vážené metody nejmenších čtverců.

$$\widehat{\beta_{WLS}} = (X^T W X)^{-1} X^T W Y$$

Kde matice vah W je diagonální matice.

Tabulka 3 HAR-RV za období (2005-2015) in sample

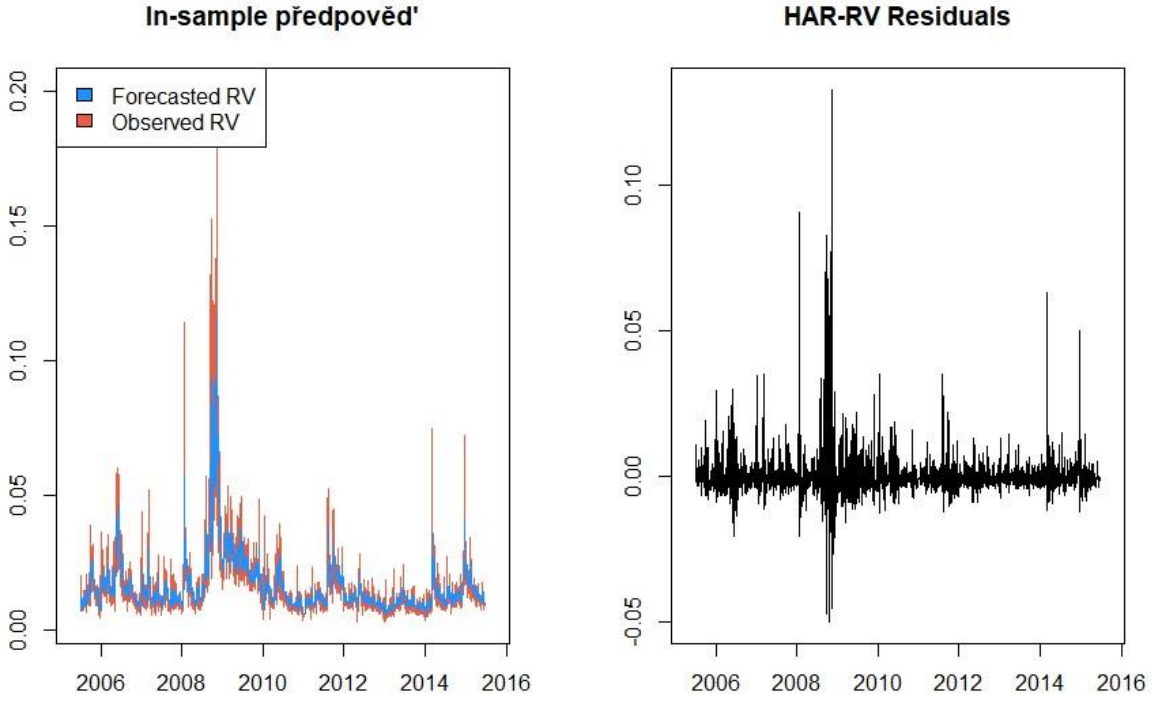
	Estimate	Std. error	p value
α	0.0010341	0.0002948	0.000460
$\beta^{(d)}$	0.3171432	0.0234064	0.000000
$\beta^{(w)}$	0.3854718	0.0390859	0.000000
$\beta^{(m)}$	0.2320957	0.0342570	0.000000
R^2		0.6169	
SSR		0.007761	
$RMSE$		0.007754955	
N		2482	

Zdroj: vlastní zpracování

Tabulka č.3 zahrnuje odhadnuté koeficienty, směrodatné chyby a *p hodnoty* HAR-RV modelu. Je vidět, že všechny parametry modelu jsou významné na 5 % hladině spolehlivosti. Koeficient determinace je měřen ve vztahu k realizované volatilitě a má hodnotu 0.6169, což je akceptovatelná hodnota vzhledem k délce analyzované časové řady. Model obsahuje zpožděné vysvětlující proměnné, největší vliv v modelu je u realizované volatility agregované na týdenním horizontu ($\beta^{(m)}$=0.3854718).

Levá část grafu č.7 porovnává predikované hodnoty realizované volatility na in-sample datech se skutečnými hodnotami realizované volatility. Je vidět, že v období finanční krize 2007–2009 predikované hodnoty realizované volatility jsou mnohem nižší než její skutečné hodnoty. Pravá část grafu č.7 zobrazuje residua HAR-RV modelu.

Graf 7 HAR-RV model (in-sample)



Zdroj: vlastní zpracování

Proveďme analýzu predikční schopnosti modelu na out-sample datech. Jedna se o rolling předpověď s horizontem předpovědi 1 den dopředu. Vždy budeme předvídat realizovanou volatilitu na následující den.

Tabulka 4 HAR-RV model out-sample predikce

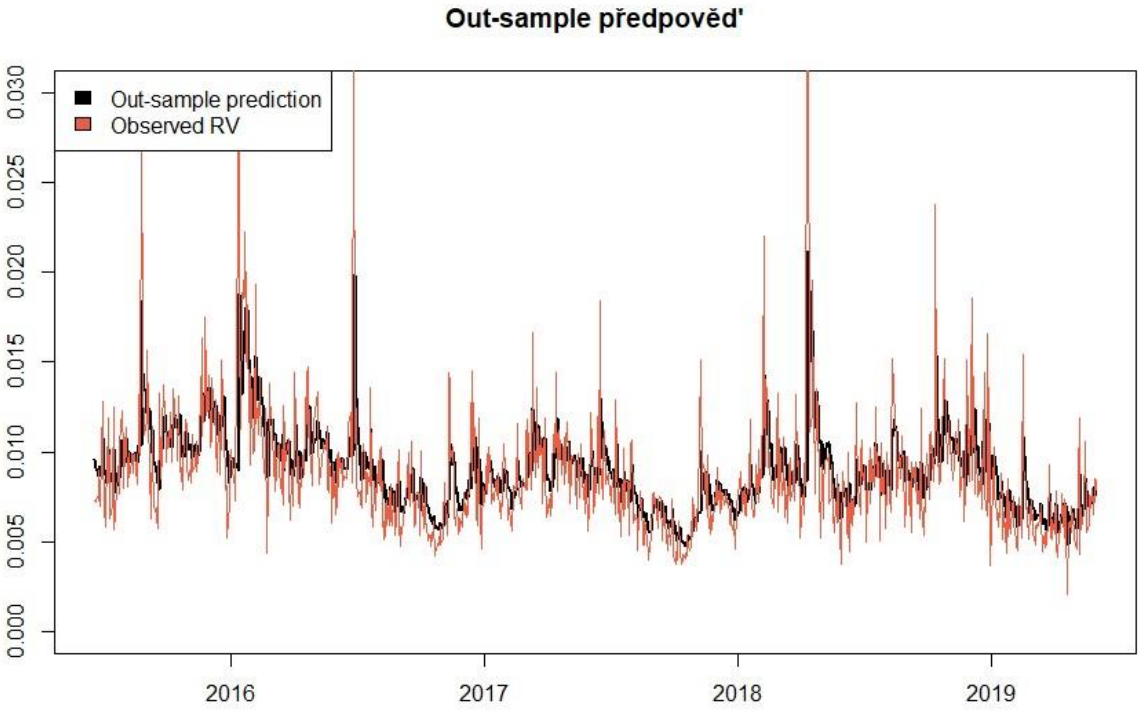
R^2	0.3199
RMSE	0.002744451

Zdroj: vlastní zpracování

Hodnota (0.3199) koeficientu determinace klesla v porovnání s hodnotou R^2 modelu odhadnutého na in-sample datech.

Grafu č.8 zobrazuje porovnání out-sample predikce HAR-RV modelu s pozorovanou realizovanou volatilitou.

Graf 8 HAR-RV model out-sample predikce vs pozorovaná RV



Zdroj: vlastní zpracování

Na grafech č.7 a č.8 je vidět, že HAR-RV model není schopen modelovat vývoj volatility kompletně. Daný model systematicky nadhodnocuje realizovanou volatilitu.

3.2.2 HAR-RV-J model

Dalším analyzovaným modelem je HAR-RV-J. Daný model obsahuje 5 vysvětlujících proměnných: denní, týdenní, měsíční realizované volatility a denní, týdenní volatility skoků. Tento model lze zapsat ve formě:

$$RV_t^{(d)} = \alpha + \beta^{(d)} RV_{t-1}^{(d)} + \beta^{(w)} RV_{t-1}^{(w)} + \beta^{(m)} RV_{t-1}^{(m)} + \beta_J^{(d)} J_{t-1}^{(d)} + \beta_J^{(w)} J_{t-1}^{(w)} + \epsilon_t$$

Tento model původně obsahoval 6 vysvětlujících proměnných, ale při odhadu se proměnná měsíční volatility skoků jevila jako nevýznamná. Proto bylo rozhodnuto ji vynechat.

Tabulka č.5 obsahuje odhad modelu HAR-RV-J za období 2005-2015.

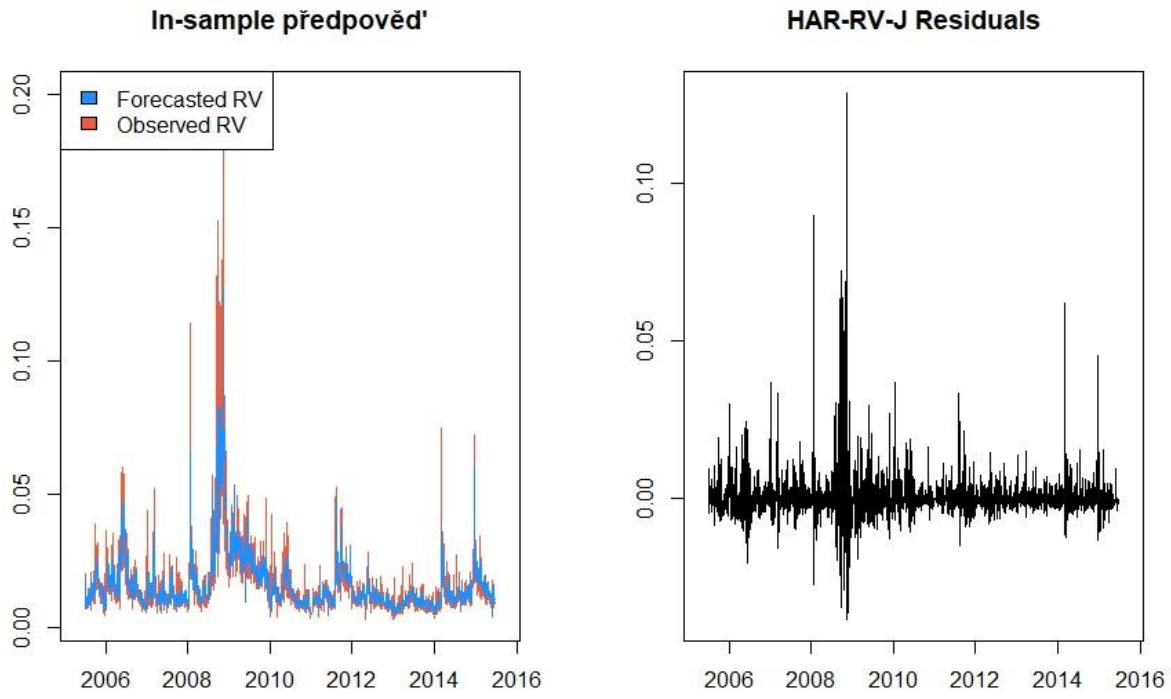
Tabulka 5 HAR-RV-J za období (2005-2015) in sample

	Estimate	Std. error	p value
α	0.0002570	0.0003166	0.41713
$\beta^{(d)}$	0.6847110	0.0341914	0.00000
$\beta^{(w)}$	0.1712160	0.0545368	0.00171
$\beta^{(m)}$	0.2059213	0.0332005	0.01394
$J^{(d)}$	-0.8188473	0.0576183	0.00000
$J^{(w)}$	0.3523444	0.1132143	0.00188
R^2		0.6490	
SSR		0.007432	
$RMSE$		0.007422757	
N		2482	

Zdroj: vlastní zpracování

Z tabulky č.5 je vidět, že pouze konstanta není významná na 5 % hladině spolehlivosti. Ostatní regresory jsou významné na 5 % hladině spolehlivosti. Daný model lépe modeluje vývoj realizované volatility v porovnání s modelem HAR-RV. Je to vidět podle hodnoty indexu determinace, která u HAR-RV-J je vyšší (0.6490).

Graf 9 HAR-RV-J model (in-sample)



Zdroj: vlastní zpracování

Na grafu č.9 je vidět, že HAR-RV-J systematicky podhodnocuje realizovanou volatilitu během sledované periody.

Dále je potřeba otestovat model na out-sample datech. Tabulka č.6 zahrnuje R^2 a *RMSE* po provedení předpovědi na out-sample datech.

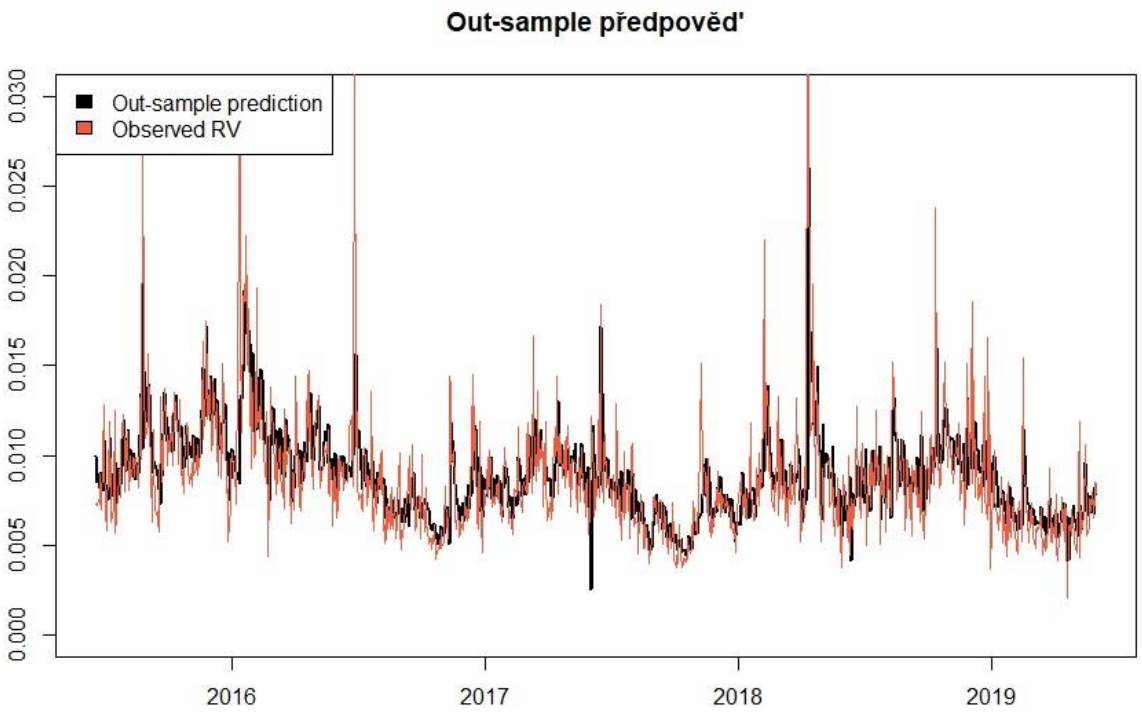
Tabulka 6 HAR-RV-J model out-sample predikce

R^2	0.3339
RMSE	0.002716129

Zdroj: vlastní zpracování

Z tabulky č.6 je vidět, že hodnota R^2 prudce klesla v porovnání s indexem determinace vypočteném na in-sample datech. Graf č.10 zobrazuje vývoj realizované volatility a její predikcí na out-sample vzorku.

Graf 10 HAR-RV-J model out-sample predikce vs pozorovaná RV



Zdroj: vlastní zpracování

Pokud porovnáme graf č.8 a graf č.10, lze konstatovat, že HAR-RV-J lépe zachycuje skutečný vývoj realizované volatility. Například, ve březnu roku 2018 skutečná hodnota realizované volatility dramaticky vzrostla. Tento prudký růst lépe odhadl HAR-RV-J model.

3.2.3 HAR-Q model

Dále provedeme odhad HAR-Q modelu. Tento model budeme analyzovat ve zkrácené podobě, která vypadá následující:

$$RV_t^{(d)} = \alpha + \left(\beta^{(d)} + \beta_{1Q} RQ_{t-1}^{(d)\left(\frac{1}{2}\right)}\right) RV_{t-1}^{(d)} + \beta^{(w)} RV_{t-1}^{(w)} + \beta^{(m)} RV_{t-1}^{(m)} + \epsilon_t$$

HAR-Q model má 4 vysvětlující proměnné. V daném modelu je použita agregace realizované quarticity na denním časovém horizontu. V tabulce č.7 jsou uvedeny odhady koeficientů, R^2, *SSR* a *RMSE* HAR-Q modelu.

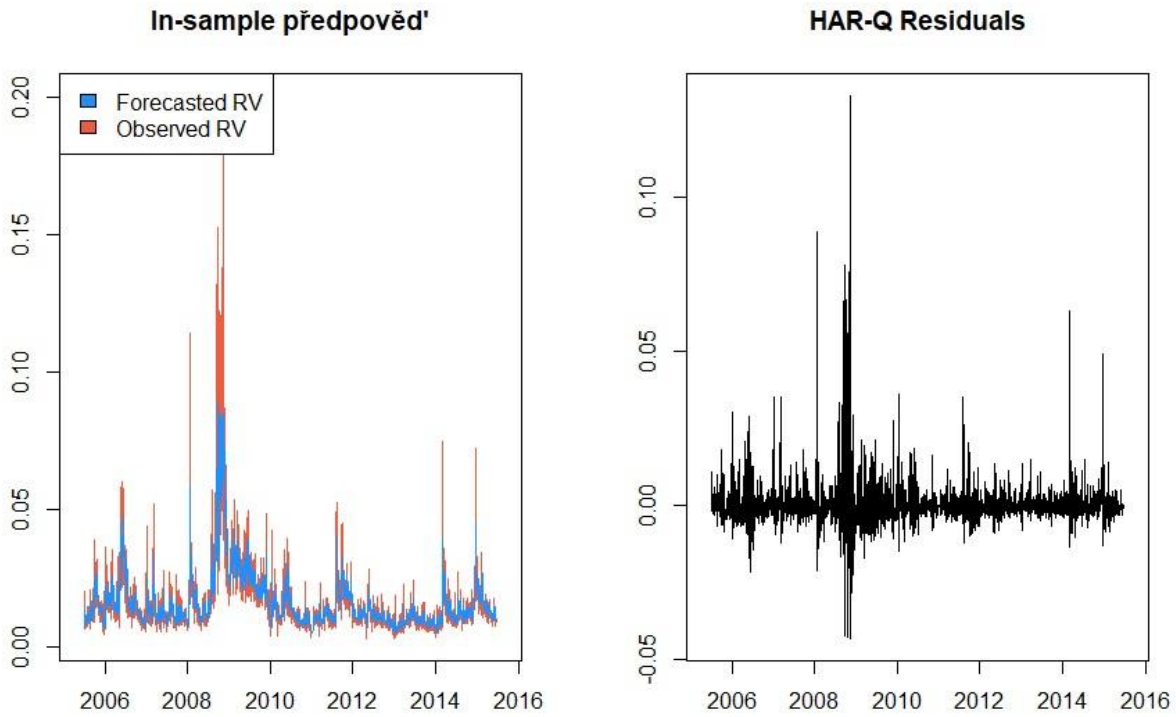
Tabulka 7 HAR-Q za období (2005-2019)

	Estimate	Std. error	p value
α	0.0004272	0.0003056	0.162
$\beta^{(d)}$	0.4181452	0.0275619	0.000000
$\beta^{(w)}$	0.3353199	0.0394343	0.000000
$\beta^{(m)}$	0.2247207	0.0339672	0.000000
β_{1Q}	-1.5721988	0.2317085	0.000000
R^2		0.6239	
SSR		0.004692	
RMSE		0.007683874	
N		2482	

Zdroj: vlastní zpracování

Podle *p hodnoty* z tabulky č.7 lze říci, že pouze konstanta není významná, ostatní vysvětlující proměnné jsou statisticky významné na 5 % hladině spolehlivosti. Index determinace (0.6239) je nižší než u HAR-RV-J modelu, ale je vyšší než u původního HAR-RV modelu.

Graf 11 HAR-Q model (in-sample)



Zdroj: vlastní zpracování

Z grafu č.11 je vidět, že odhadnutý HAR-Q není schopen kompletně zachytit vývoj realizované volatility. Je to zřejmé zejména v periodách velkých skoků. Proveďme out-sample predikce na testovacích datech. Tabulka č.8 zachycuje výsledky out-sample predikce.

Tabulka 8 HAR-Q model out-sample predikce

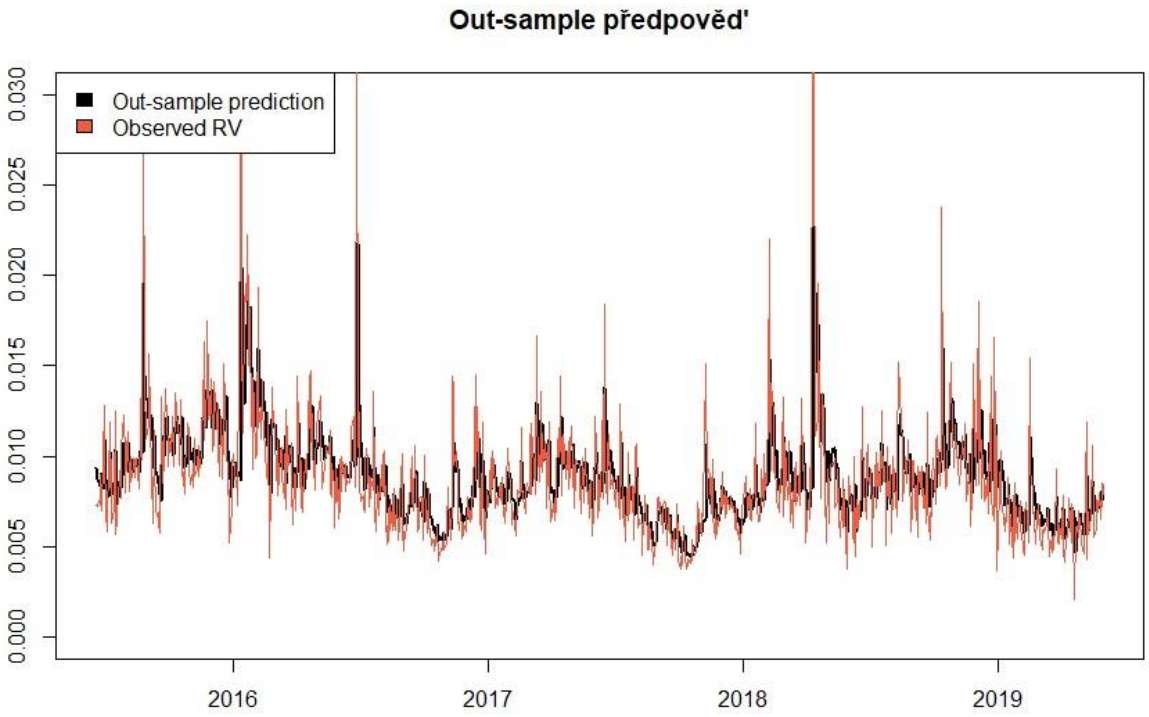
R^2	0.3162
RMSE	0.002751954

Zdroj: vlastní zpracování

Z tabulky č.8 je vidět, že hodnota indexu determinace, spočítaného na out-sample datech, je nižší než hodnota R^2 HAR-RV modelu (viz tabulka č.4). Z čeho lze udělat závěr, že rozšíření původního HAR-RV modelu o realizovanou quarticity nemá pozitivní vliv na výkonnost modelu. Potvrzují to i hodnoty *RMSE*.

Graf č.12 porovnává out-sample predikce se skutečnými pozorováními realizované volatility.

Graf 12 HAR-Q model out-sample predikce vs pozorovaná RV



Zdroj: vlastní zpracování

V periodách nízké volatility HAR-Q model nadhodnocuje realizovanou volatilitu (například období 2017-2018), na jinou stranu daný model lépe odhaduje volatilitu při prudkých vzrůstech její úrovně (například leden 2016 a březen 2018).

3.2.4 HAR-Q-J model

Posledním modelem vybraným k analýze je HAR-Q-J model. Předpokládá se, že HAR-Q-J model bude mít lepší predikční schopnosti než model HAR-Q, protože daný model umožnuje zachytit volatilitu skoků pomocí dodatečných vysvětlujících proměnných. Cílem je také zjistit, zda HAR-Q-J bude mít lepší výkonnost i v porovnání s HAR-RV-J, kvůli možnosti upravit beta koeficienty v poměru k velikosti odpovídajících chyb měření. HAR-Q-J model má následující podobu:

$$RV_t^{(d)} = \alpha + \left(\beta^{(d)} + \beta_{1Q} RQ_{t-1}^{(d)\left(\frac{1}{2}\right)}\right) RV_{t-1}^{(d)} + \beta^{(w)} RV_{t-1}^{(w)} + \beta^{(m)} RV_{t-1}^{(m)} + \beta_J^{(d)} J_{t-1}^{(d)} + \beta_J^{(w)} J_{t-1}^{(w)} + \beta_J^{(m)} J_{t-1}^{(m)} + \epsilon_t$$

Tento model má 7 vysvětlujících proměnných. V HAR-Q-J modelu byla provedena agregace realizované volatility a skoků na denním, týdenním a měsíčním časových horizontech.

Realizované quarticity byla agregována na denním časovém horizontu. Tabulka č.9 znázorňuje odhad analyzovaného modelu na in-sample datech.

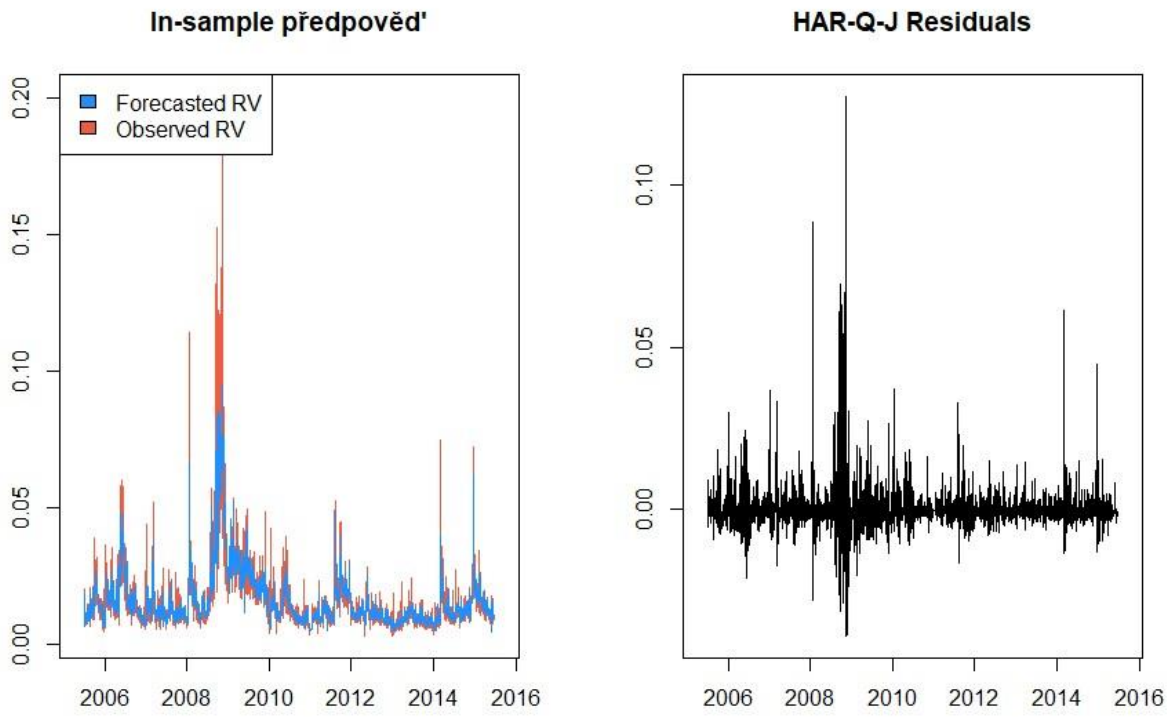
Tabulka 9 HAR-Q-J za období (2005-2015)

	Estimate	Std. error	p value
α	0.0002009	0.0003633	0.5803
$\beta^{(d)}$	0.7529590	0.0363922	0.0000
$\beta^{(w)}$	0.1376995	0.0592161	0.0201
$\beta^{(m)}$	0.1317526	0.0636260	0.0385
β_{1Q}	-1.2232958	0.2316613	0.0000
$J^{(d)}$	-0.7957116	0.0574830	0.0000
$J^{(w)}$	0.3361014	0.1362001	0.0137
$J^{(m)}$	0.2737338	0.2001293	0.1715
R^2		0.6539	
SSR		0.007393	
$RMSE$		0.007380902	
N		2482	

Zdroj: vlastní zpracování

Z tabulky č.9 lze říci, že konstanta a proměnná měsíční skoky nejsou statisticky významné. Ostatní vysvětlující proměnné jsou významné na 5 % hladině spolehlivosti. Dále lze říci, že hodnota indexu determinace (0.6539) u HAR-Q-J modelu je nejvyšší ze všech modelů vybraných k analýze. Graf č.13 ve své levé části znázorňuje porovnání in-sample předpovědi se skutečnými hodnotami realizované volatility. Také v pravé části grafu jsou vykreslena residua.

Graf 13 HAR-Q-J model (in-sample)



Zdroj: vlastní zpracování

Z grafu č.13 je vidět, že HAR-Q-J model nejlépe modeluje vývoj realizované variance jak v období náhlých vzrůstů, tak i na nízké úrovně realizované volatility. Otestujme výkonnost modelu na out-sample datech.

Tabulka 10 HAR-Q-J model out-sample predikce

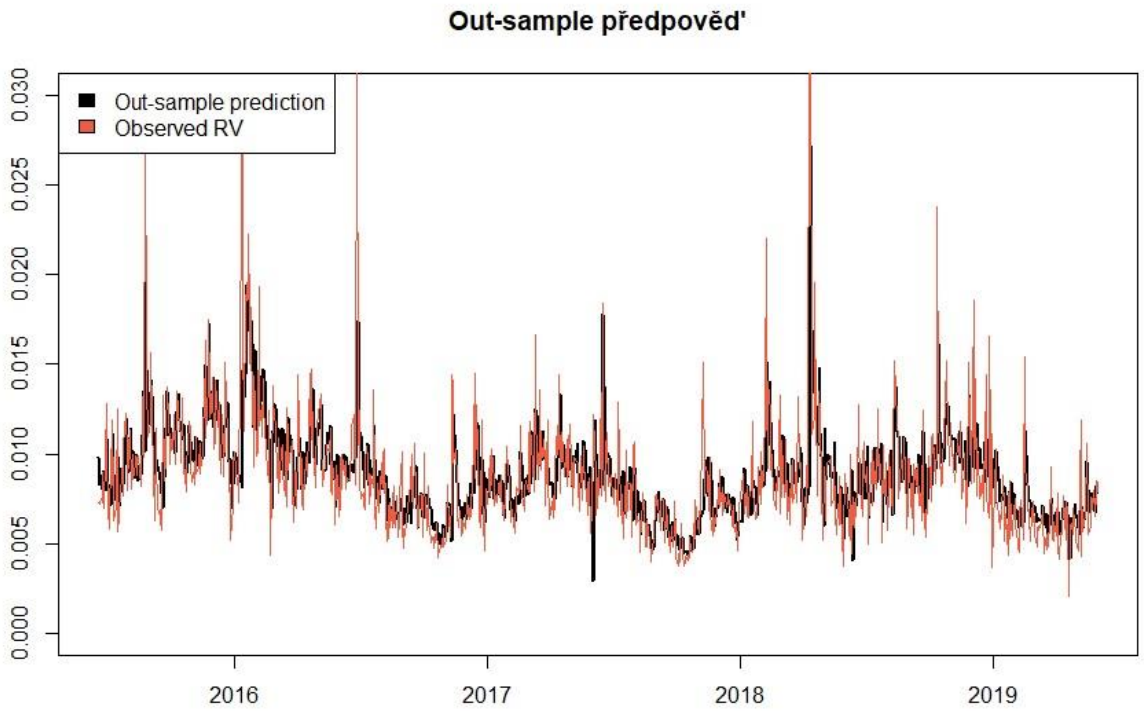
R^2	0.3245
RMSE	0.002735156

Zdroj: vlastní zpracování

Index determinace klesl ze své původní hodnoty (0.6539) na nižší hodnotu (0.3245). Lze říci, že taková hodnota indexu determinace je akceptovatelným výsledkem pro out-sample predikce. Pokud porovnáme hodnoty z tabulek č.10 a č.6 lze konstatovat, že dle hodnot indexu determinace a *RMSE* HAR-RV-J model demonstruje lepší výkonnost než HAR-Q-J.

Dále vykreslíme out-sample predikce oproti skutečným hodnotám realizované volatility.

Graf 14 HAR-Q-J model out-sample predikce vs pozorovaná RV



Zdroj: vlastní zpracování

Z grafu č.14 je vidět, že HAR-Q-J relativně přesně odhaduje realizovanou volatilitu i na out-sample datech. Daný model poskytuje akceptovatelné výsledky jak na nízké úrovně realizované volatility, tak i na vysoké. V následující podkapitole všechny čtyři odhadnuté HAR modely budou porovnány mezi sebou podle stanovených kritérií (podkapitola č. 2.3).

3.3 Porovnání modelů

Výkonnost modelů bude analyzována jak na in-sample, tak i na out-sample datech. Nejdříve proveďme porovnání predikčních schopností modelů pomocí Mincer – Zarnowitz regrese.

V tabulce č.11 jsou uvedeny výsledky Mincer – Zarnowitz regrese spočítané pro modely, které byly odhadnuty na in-sample datech.

Tabulka 11 Mincer – Zarnowitz regrese, in-sample data

	HAR-RV	*HAR-RV-J*	*HAR-Q*	*HAR-Q-J*
α	0.0000	0.0000	0.0000	0.0000
β	1.0000	0.9999	1.0000	1.0000
R^2	0.6169	0.6490	0.6239	0.6539

Zdroj: vlastní zpracování

Jak je vidět z tabulky č.11, hodnoty β koeficientů všech modelů jsou blízké jedné. Hodnoty konstanty α jsou velice blízké nule. Nejlepší hodna indexu determinace je u HAR-Q-J modelu. K tomu abychom mohli udělat jednoznačný závěr, který model má lepší predikční schopnosti, je potřeba porovnat modely na out-sample datech.

Následující tabulka č.12 zobrazuje výsledky Mincer – Zarnowitz regrese spočítané pro modely, které byly odhadnuty na out-sample datech.

Tabulka 12 Mincer – Zarnowitz regrese, out-sample data

	HAR-RV	*HAR-RV-J*	*HAR-Q*	*HAR-Q-J*
α	0.000816	0.0015769	0.0014425	0.0017994
β	0.872887	0.8001246	0.8163702	0.7778082
R^2	0.3369	0.3599	0.3364	0.3561
SSR	0.002713	0.002665	0.002714	0.002673

Zdroj: vlastní zpracování

Z tabulky č.12 je patrné, že hodnota indexu determinace klesla. Nejlepší hodnota R^2 je u HAR-RV-J modelu. Hodnoty konstanty α jsou u všech modelů blízké nule. β koeficienty se u všech modelů výrazně odchylují od cílové hodnoty 1. Nejhorší hodnota β koeficientu je u HAR-Q-J modelu. Nejbližší hodnotu β koeficientu má HAR-RV model, však tento model má nižší hodnotu indexu determinace se srovnání s HAR-RV-J a HAR-Q-J modely.

Z tabulky č.12 je vidět, že u všech analyzovaných HAR modelů poklesla predikční schopnost v porovnání s odhady na in-sample datech, což je běžným výsledkem. Při přechodu na out-sample data předpovědní schopnosti modelů jsou téměř vždy horší.

Na základě Mincer-Zarnowitz regrese na in-sample datech lze říci, že HAR-Q-J má nejlepší výkonnost. Avšak v případě out-sample predikcí HAR-RV-J model demonstruje nejlepší výsledky.

Dále posoudíme predikční schopnosti modelů dle dosažené hodnoty indexu determinace a *RMSE*. Tabulka č.13 zobrazuje hodnoty R^2 a *RMSE*, které byly spočítány na out-sample datech.

Tabulka 13 Porovnání predikčních schopností modelů podle R^2 a *RMSE* (out-sample)

	HAR-RV	*HAR-RV-J*	*HAR-Q*	*HAR-Q-J*
R^2	0.3199	0.3339	0.3162	0.3245
RMSE	0.002744451	0.002716129	0.002751954	0.002735156

Zdroj: vlastní zpracování

Z tabulky č.13 je vidět, že podle hodnot R^2 a *RMSE* nejlepší výkonnost má HAR-RV-J model. Dále porovnáme modely bez regresorů skoků. HAR-Q model má nižší R^2 a vyšší *RMSE* ve srovnání s HAR-RV modelem. Model HAR-Q-J má nižší index determinace a vyšší *RMSE* než HAR-RV-J. Z těchto výsledků lze udělat závěr, že realizovaná quarticity nepřispívá ke zlepšení predikčních schopností HAR modelů na denním horizontu předpovědi. Také je vidět, že přidání regresorů skoků do modelů zlepšuje výkonnost modelů.

3.4 Modelování předpovědí na různých délkách trénovacích dat

3.4.1 Year-by-Year odhad parametrů modelů

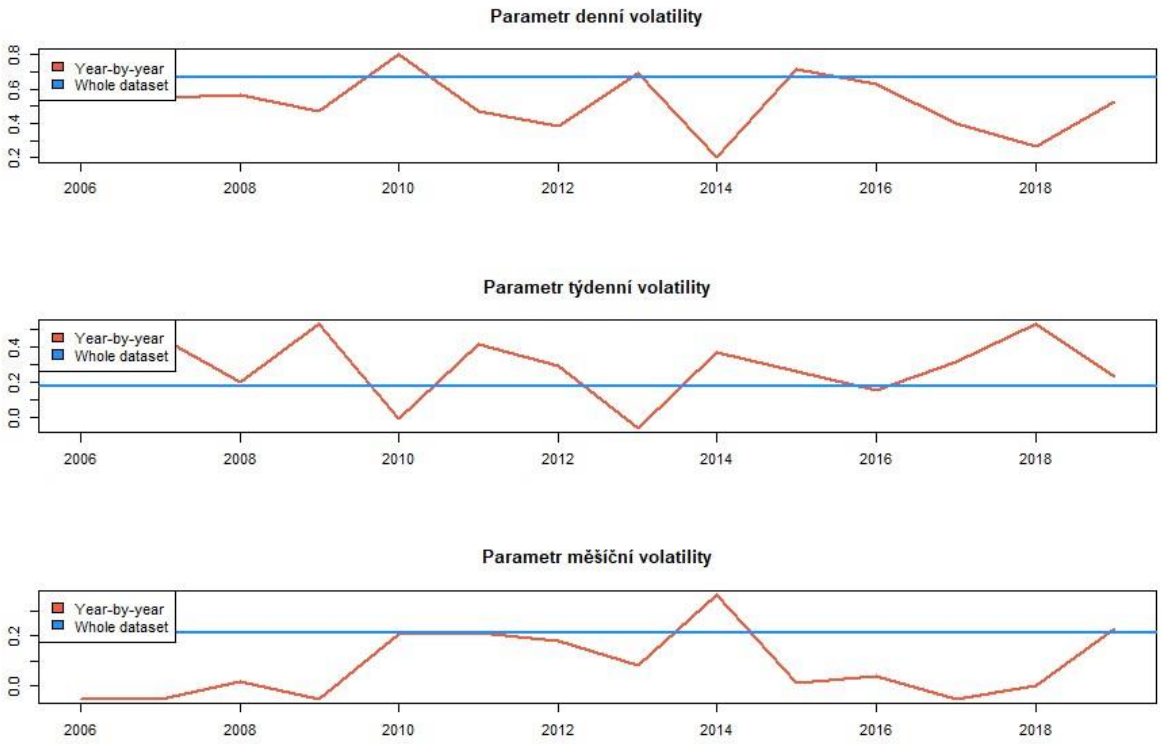
Důvodem, proč se zaměřuje na analýzu dynamiky parametrů HAR modelů, je předpoklad časové stability parametrů při použití OLS metody. Pokud odhadované parametry nabývají různých hodnot na některých časových intervalech, dostaneme ¨průměrný¨ odhad těchto parametrů než při odhadu na celém datovém vzorce. Tudíž předpovědi volatility budou pouze ¨průměrné¨. Samozřejmě, odhad parametrů na delších časových řadách zlepší kvalitu předpovědí, však dané předpovědi nebudou nejlepší možné.

V této části se omezíme na HAR-RV-J a na HAR-Q-J modely ze dvou důvodů. Zaprvé, výsledky vývojů parametrů se u modelů bez dodatečných regresorů skoků neliší od modelů se skoky. Zadruhé, jelikož cílem této práce je zjistit, zda realizovaná quarticity přispívá ke zlepšení predikčních schopností HAR modelů, bylo rozhodnuté vybrat HAR-Q-J model jako zástupce skupiny HAR-Q modelů.

Původní datový vzorek byl rozdělen do 14 podskupin stejné velikosti. Každá podskupina obsahuje přibližně 25 000 logaritmických výnosů indexu Moskevské burzy (MOEX) za období jednoho roku. Rozhodli jsme ukázat dynamiku pouze parametrů realizované volatility $\beta^{(d)}$, $\beta^{(w)}$, $\beta^{(m)}$, protože se tyto parametry mění výrazněji než parametry komponentů skoků a realizované quarticity.

Grafy č.15 a 16 znázorňují dynamiku parametrů HAR-RV-J a HAR-Q-J modelů. Modrá čára ukazuje hodnotu parametru (denní, týdenní a měsíční) realizované volatility odhadnuté na celém datovém vzorce. Červená čára znázorňuje vývoj parametrů odhadnutých na každé ze 14 podskupin.

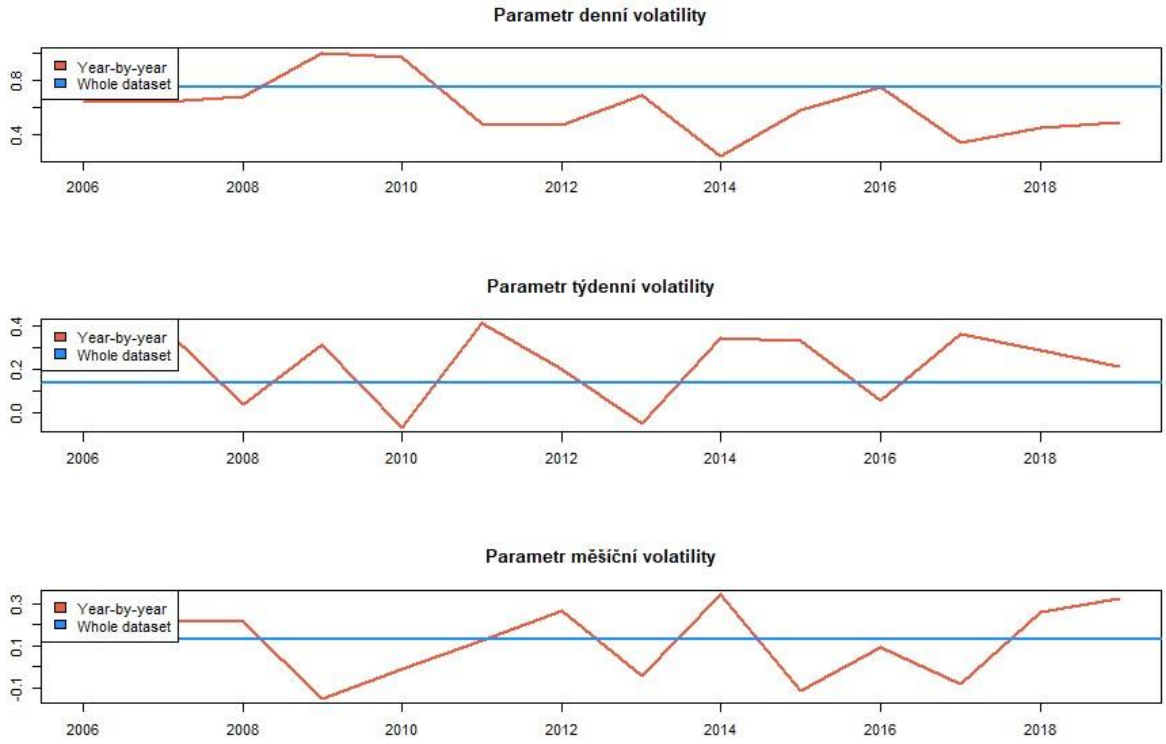
Graf 15 HAR-RV-J Year-by-Year odhad parametrů modelů



Zdroj: vlastní zpracování

Z grafu č.15 je vidět, že se parametry mění výrazně. Nejvíce se mění parametry denní a týdenní volatility. Navíc lze rozpoznat, že se parametry denní a měsíční volatility mění stejným směrem. Z toho lze udělat závěr, že obvykle realizovaná volatilita je především řízena denní a měsíční realizovanou volatilitou, tj. tržní účastníci s denní a měsíční periodicitou obchodování jsou převážně odpovědní za změny ve volatilitě.

Graf 16 HAR-Q-J Year-by-Year odhad parametrů modelů



Zdroj: vlastní zpracování

Graf č.16 ukazuje, že se paramenty denní, týdenní a měsíční realizované volatility mění stejně výrazně. Lze říci, že realizovaná volatilita je řízena denní, týdenní a měsíční volatilitami přibližně ve stejném poměru.

Z předchozích výsledků lze tvrdit, že se parametry zkoumaných HAR modelů výrazně mění v čase, což má za následek omezenou schopnost vylepšit výkonnost modelů přidáním dalších pozorování. Navíc tato dynamika negativně ovlivňuje predikční schopnosti HAR modelů.

3.4.2 Porovnání predikčních schopností modelů na různých délkách in-sample dat

V této podkapitole se zkoumají predikční schopnosti HAR modelů, které jsou odhadnuty na různých délkách in-sample dat. Ze stejných důvodů jako v podkapitole č.3.4.1 se omezíme na dva modely HAR-RV-J a HAR-Q-J. Také nejhorší a nejlepší předpovědi obou modelů budou znázorněny pomocí grafů.

Původní soubor dat byl rozdělen na dvě části. Jedna část představuje přibližně poslední rok časové periody (out-sample), druhá se skládá ze zbývající části datového souboru (in-sample). Parametry HAR modelů byly postupně odhadnuty na posledním roce in-sample dat, na

posledních dvou let in-sample dat atd. Také parametry modelů byly odhadnuty na celém in-sample vzorce.

Pak jsme provedli out-sample předpovědi realizované volatility na každé podskupině in-sample dat. Tabulka č.14 zahrnuje hodnoty indexu determinace a *RMSE* spočítané pro každou out-sample předpověď.

Tabulka 14 Porovnání predikčních schopností modelů na různých délkách in-sample dat

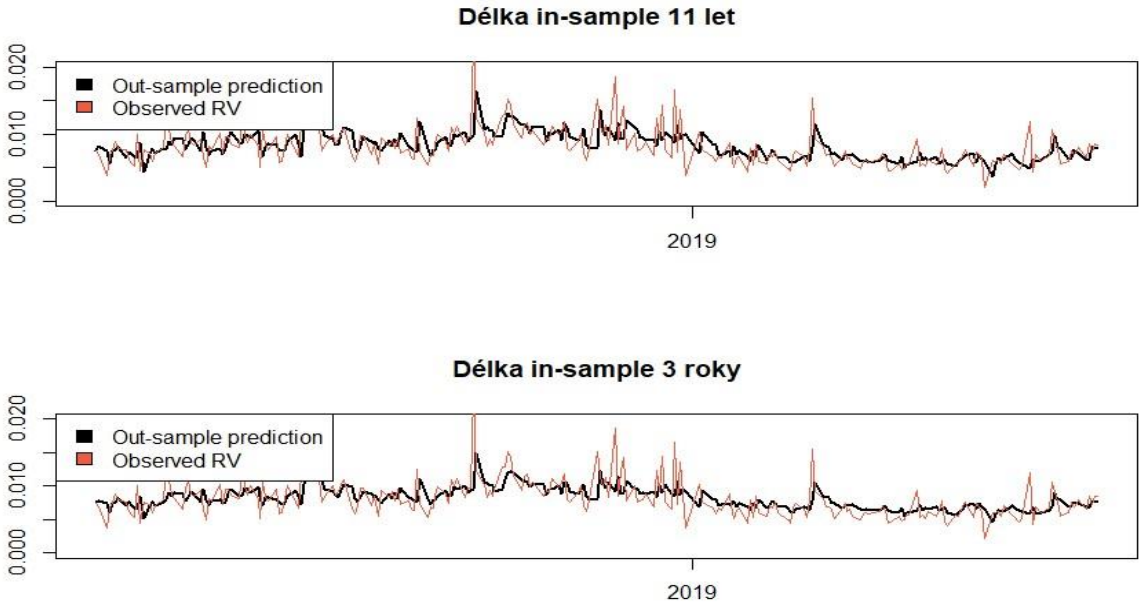
	HAR-RV-J		*HAR-Q-J*	
	R^2	*RMSE*	R^2	*RMSE*
1	0.2468	0.002354	0.2337	0.002374
2	0.2738	0.002311	0.2626	0.002329
3	0.2843	0.002294	0.2812	0.002299
5	0.2596	0.002334	0.2660	0.002324
7	0.2559	0.002340	0.2644	0.002326
9	0.2456	0.002356	0.2488	0.002351
11	0.2370	0.002369	0.2054	0.002418
Whole	0.2402	0.002364	0.2135	0.002405

Zdroj: vlastní zpracování

Z tabulky č.14 lze tvrdit, že nejlepší predikční schopnosti demonstrují HAR modely, které byly odhadnuty na posledních 3 letech in-sample dat. Je to vidět, jak dle hodnoty indexu determinace, tak i dle hodnoty *RMSE*. Nejhorší předpověď u obou modelů dle obou kritérií byla provedena na posledních 11 letech in-sample dat.

Graf č.17 zobrazuje nejhorší a nejlepší out-sample předpověď HAR-RV-J modelu. V horní části grafu je vykreslena nejhorší predikce, v dolní je nejlepší předpověď.

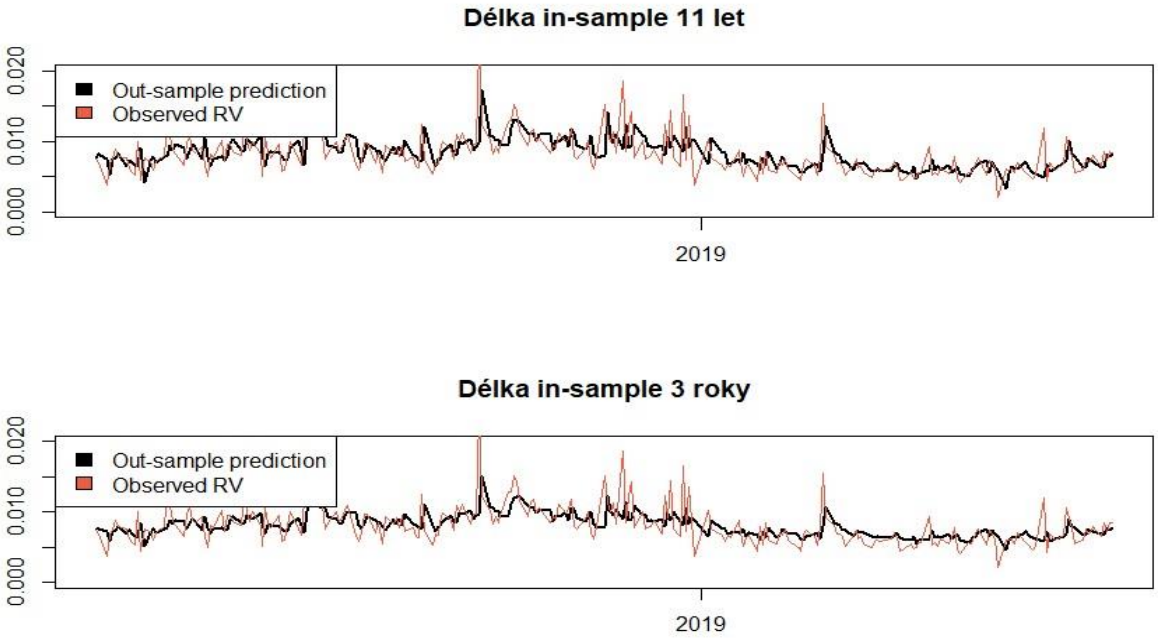
Graf 17 Nejhorší a nejlepší out-sample předpověď HAR-RV-J modelu



Zdroj: vlastní zpracování

Z grafu č.17 je vidět, že HAR-RV-J model odhadnutý na posledních 3 letech in-sample dat lépe předpovídá realizovanou volatilitu zejména v období skoků. Graf č.18 porovnává nejlepší a nejhorší out-sample predikce HAR-Q-J modelu. Dle hodnot indexu determinace a *RMSE* lze tvrdit, že HAR-RV-J lépe modeluje realizovanou volatilitu.

Graf 18 Nejhorší a nejlepší out-sample předpověď HAR-Q-J modelu



Zdroj: vlastní zpracování

Důvodem, proč HAR modely demonstrují nejlepší predikční schopnosti na posledních 3 letech in-sample vzorce, může být skutečnost, že od roku 2010 nedocházelo k velkým skokům v logaritmických výnosech. Modely odhadnuté na delších časových periodách mají horší výkonnost právě kvůli tomu, že datový soubor zahrnuje období extrémně vysoké volatility logaritmických výnosů (období světové finanční krize 2007-2009).

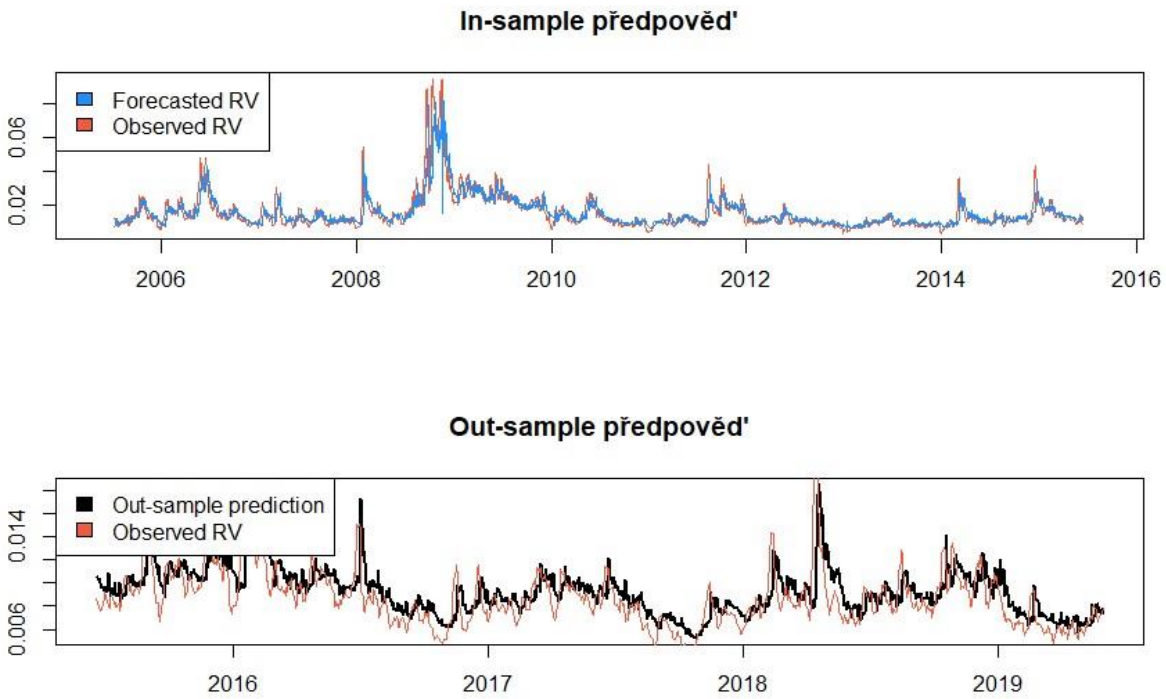
3.5 Predikce agregované volatility na následující týden

V předchozích kapitolách jsme se věnovali predikcím realizované volatility v horizontu jednoho dne dopředu. V dané podkapitole budeme analyzovat předpovědní schopnosti čtyř HAR modelů na delší období. Budeme prognózovat agregovanou volatilitu v horizontu jednoho týdne.

K dispozici máme stejnou časovou řadu jako v případě predikcí na jeden den dopředu. Také používáme stejnou metodiku out-sample predikcí a stejná kritéria při porovnání předpovědních schopností HAR modelů. Graf č.19 zobrazuje in-sample a out-sample předpovědi realizované volatility na základě odhadnutého HAR-RV modelu.

Z dolní části grafu č.19 je patrné, že HAR-RV model vždy nadhodnocuje realizovanou volatilitu při out-sample predikci. Je to nejlépe vidět v období od roku 2015 do roku 2018.

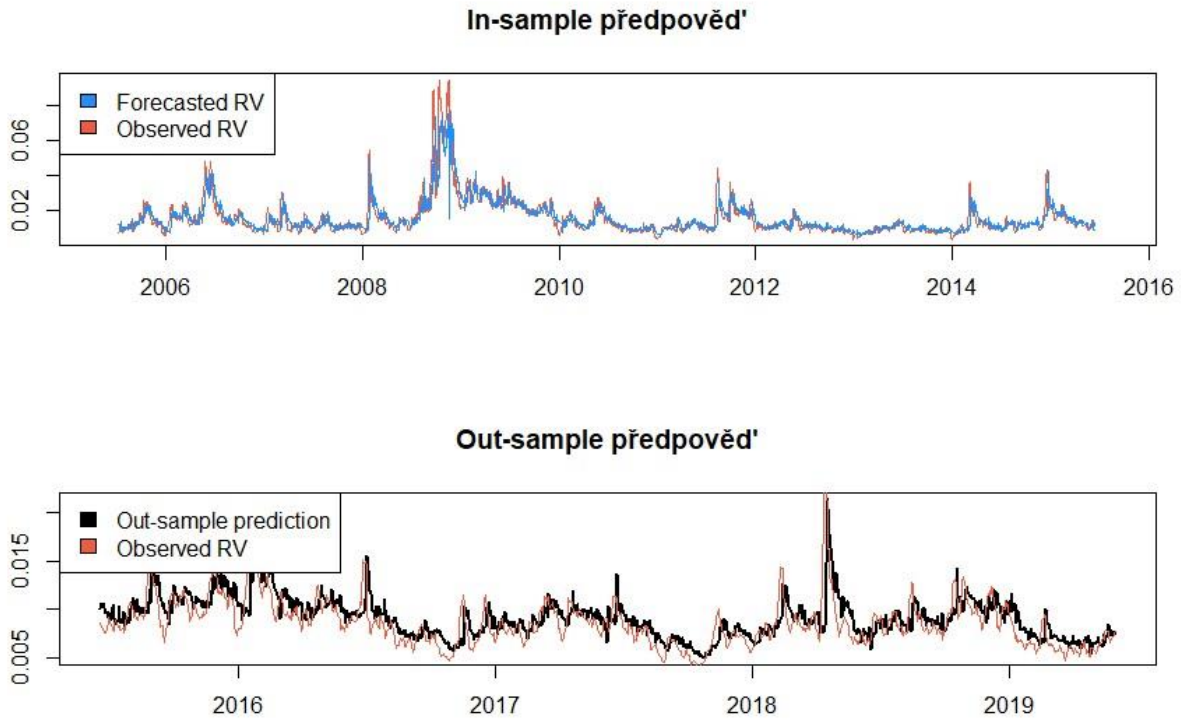
Graf 19 HAR-RV model in-sample a out-sample předpovědi (týdenní horizont)



Zdroj: vlastní zpracování

Grafu č.20 znázorňuje porovnání in-sample a out-sample předpovědí se skutečnými hodnotami realizované volatility. Z grafu je vidět, že HAR-RV-J model lépe modeluje realizovanou volatilitu než základní HAR-RV model, však daný model nepatrně nadhodnocuje skutečnou realizovanou volatilitu.

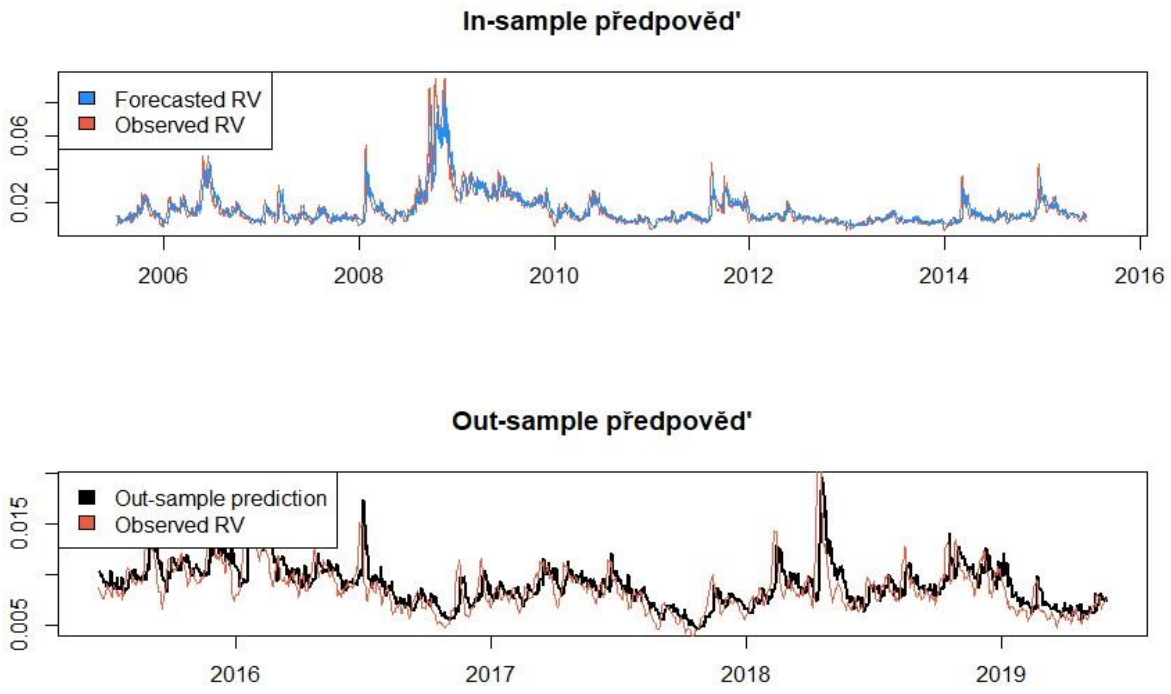
Graf 20 HAR-RV-J model in-sample a out-sample předpovědi (týdenní horizont)



Zdroj: vlastní zpracování

Graf č.21 zobrazuje výsledky HAR-Q modelu. Z grafu není patrné, zda přidáním realizované quarticity k původnímu HAR modelu jsme zlepšili kvalitu předpovědi realizované volatility. Tento model nepatrně nadhodnocuje skutečný vývoj realizované volatility.

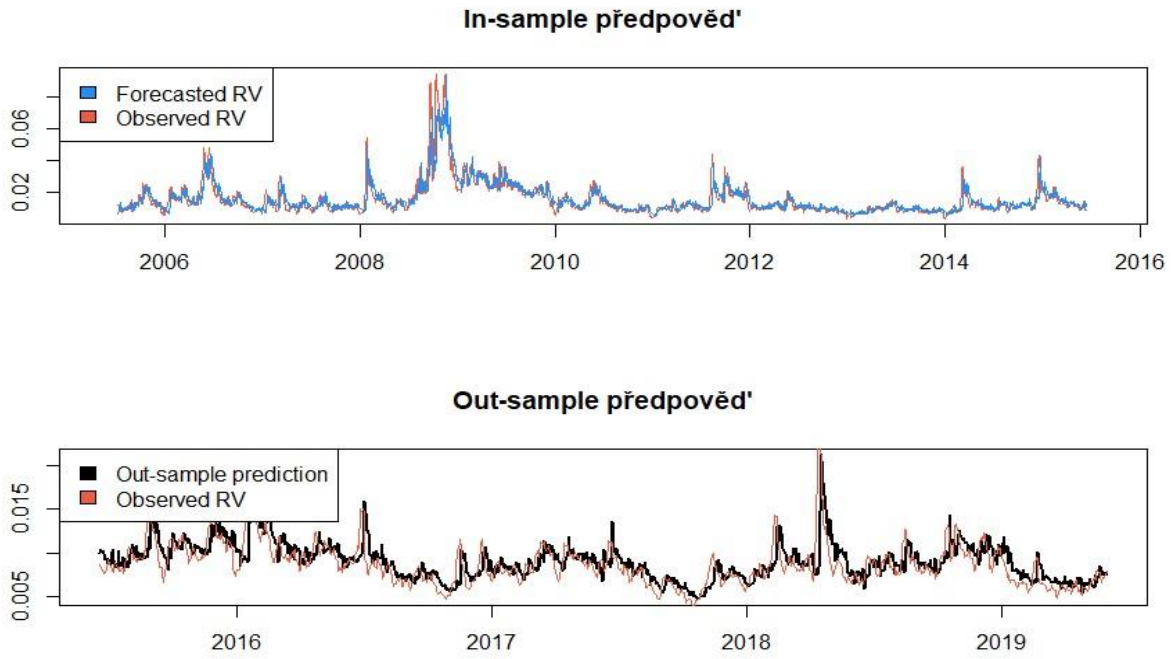
Graf 21 HAR-Q model in-sample a out-sample předpovědi (týdenní horizont)



Zdroj: vlastní zpracování

Dále se podíváme na graf č.22, který zobrazuje in-sample a out-sample předpovědi realizované volatility na základě odhadnutého HAR-Q-J modelu.

Graf 22 HAR-Q-J model in-sample a out-sample předpovědi (týdenní horizont)



Zdroj: vlastní zpracování

Z grafu č.22 je patrné, že HAR-Q-J relativně dobře modeluje vývoj realizované volatility. Abychom mohli vybrat nejlepší model pro modelování agregované volatility v horizontu jednoho týdne, musíme spočítat Mincer-Zarnowitz regrese, R^2 a *RMSE*.

Dále se podíváme na výsledky Mincer-Zarnowitz regrese, které shrnuje tabulka č.15.

Tabulka 15 Mincer – Zarnowitz regrese, out-sample data (týdenní horizont)

	HAR-RV	*HAR-RV-J*	*HAR-Q*	*HAR-Q-J*
α	0.0013281	0.0015858	0.0020148	0.0019808
β	0.7970175	0.7748453	0.7406355	0.7433843
R^2	0.4047	0.4196	0.3996	0.4104
SSR	0.001932	0.001908	0.001940	0.001923

Zdroj: vlastní zpracování

Mincer-Zarnowitz regrese byla spočítána na out-sample datech. Z tabulky č.15 je vidět, že se konstanta α u všech modelů přibližně rovná nule. β koeficienty u všech modelů se odchylují od jedné. Kromě toho hodnoty indexu determinace u HAR-RV-J a HAR-Q-J modelů jsou vyšší než u ostatních modelů.

Dle výsledků Mincer-Zarnowitz regrese lze říci, že realizovaná quarticity nepřispívá ke zlepšení predikčních schopností HAR modelů, ale rozšíření modelů o dodatečné regresory skoků má pozitivní vliv na výkonnost modelů. Také lze konstatovat, že HAR-RV-J je nejlepším z analyzovaných modelů.

V tabulce č.16 se nacházejí výsledky porovnání predikčních schopností modelů podle R^2 a *RMSE*. *RMSE* je spočítána na out-sample datech.

Tabulka 16 Porovnání predikčních schopností modelů podle R^2 a *RMSE* (týdenní horizont)

	HAR-RV	*HAR-RV-J*	*HAR-Q*	*HAR-Q-J*
R^2*(in-sample)*	0.7030	0.7111	0.7161	0.7227
R^2*(out-sample)*	0.3308	0.3477	0.3340	0.3436
RMSE	0.002046453	0.002020445	0.00204147	0.00202669

Zdroj: vlastní zpracování

Z tabulky č.16 je patrné, že nejvyšší hodnotu indexu determinace na out-sample datech má HAR-RV-J model, také tento model má nejnižší hodnotu *RMSE*. Nejhorším modelem je HAR-RV podle obou kritérií jak na in-sample, tak i na out-sample datech. Pokud porovnáme

hodnoty indexu determinace z tabulky č.13 a z tabulky č.16 uvidíme, že R^2 dosahují vyšších hodnot, než v případě predikce realizované volatility na následující den. Daný jev lze pozorovat docela často (Poon a Granger, 2003). Pokud porovnáme modely s dodatečným regresorem realizované quarticity s modely bez dané vysvětlující proměnné, je vidět, že dle stanovených kritérií HAR-Q model je lepší než HAR-RV model, však HAR-Q-J model je horší než HAR-RV-J model. Když porovnáme modely rozšířené o regresory skoků s modely bez nich, lze jednoznačně tvrdit, že dodatečné regresory skoků zlepšují predikční schopnosti modelů.

Dle výsledků dosažených v této podkapitole lze říci, že model HAR-RV-J má nejlepší předpovědní schopnosti v předvídaní agregované volatility na 1 týden dopředu ve srovnání s ostatními analyzovanými modely. Kromě toho nelze jednoznačně určit, zda realizovaná quarticity přispívá ke zlepšení predikčních schopností modelů.

Závěr

Hlavním cílem této diplomové práce byl výběr nejlepšího modelu pro modelování a předvídání volatility na vysokofrekvenčních datech. V práci byly představeny čtyři HAR modely: základní HAR-RV model, HAR-RV-J model, který rozšiřuje původní model o regresory skoků, HAR-Q model upravující beta koeficienty v poměru k velikosti odpovídajících chyb měření a HAR-Q-J, který je rozšířením modelu HAR-Q o regresory skoků. Výše uvedené modely byly nejdříve popsány z teoretického hlediska v kapitole č.2. Pak v kapitole č.3 tyto modely byly aplikovány na časovou řadu indexu Moskevské burzy (MOEX). Byl proveden odhad volatility na následující den. Podkapitola č.3.3 shrnuje výsledky predikčních schopností modelů při odhadu volatility na následující den. Dle hodnot indexu determinace R^2, *RMSE* a Mincer-Zarnowitz regrese na out-sample datech nejlepším modelem je HAR-RV-J model. Pak čtyři HAR modely byly využity i pro konstrukci dlouhodobých předpovědí na následující týden. Nejlepší výsledky byly dosaženy HAR-RV-J modelem i v případě dlouhodobých předpovědí. Nejhorších výsledků dosáhl HAR-RV model při konstrukci dlouhodobých předpovědí jak na in-sample, tak i na out-sample datech. Při konstrukci předpovědí volatility na následující týden se ukázal význam kaskádních modelů s dlouhou pamětí. V případě předpovědí v horizontu jednoho týdne podíl variability závislé proměnné je vysvětlen lépe než u předpovědí volatility na následující den, což koresponduje s výsledky jiných prací. V podkapitole č.3.4 byl proveden Year-by-Year odhad parametrů modelů za účelem posouzení dynamiky parametrů v čase. Z výsledků odhadů parametrů je možné konstatovat, že se parametry modelů mění v čase relativně silně. Dále byly porovnány predikční schopnosti dvou HAR modelů na různých délkách dat. Nejlepší výsledky byly dosaženy u modelů odhadnutých za periodu posledních 3 let in-sample dat.

Výsledky dosažené v aplikační části diplomové práce jsou v souladu s teoretickými přístupy, které byly popsány v kapitolách č.1 a č.2.

Dle mého názoru hlavní a vedlejší cíle této diplomové práce byly splněny. Lze tvrdit, že ze skupiny HAR modelů je nejlepším modelem HAR-RV-J. Také z výsledků provedené analýzy je vidět, že rozšíření modelů o regresory skoků přispívá ke zlepšení predikčních schopností HAR modelů jak při krátkodobých, tak i při dlouhodobých predikcích. Avšak rozšíření modelů o realizovanou quarticity nemá pozitivní vliv na výkonnost modelů v případě krátkodobých předpovědí. Při konstrukci předpovědí na následující týden, nelze jednoznačně určit, zda realizovaná quarticity zlepšuje predikční schopnosti HAR modelů.

Z relativně nízkých hodnot indexu determinace je vidět, že podíl variability závislé proměnné HAR modelů není vysvětlen perfektně. Lze říci, že v případě předpovědí v horizontu jednoho dne, všechny čtyři HAR modely neúplně zachycují dynamiku realizované volatility. Možným řešením tohoto problému může být využití State Space HAR modelu, který byl navržen Bekiermanem a Mannerem (2018). Tento model umožnuje lépe zachytit dynamiku volatility pomocí zavedení časově měnícího autoregresního parametru.

Kromě toho při zpracování původních vysokofrekvenčních dat je důležité správně vybrat algoritmus očišťování dat. Při příliš přísném algoritmu lze ztratit významné množství informace potřebné ke správnému odhadu modelů. Pokud algoritmus bude příliš mírný, odhady volatility budou vychýlené.

Použitá literatura

1. ANDERSEN, T. and BENZONI, L., (2007): "Realized Volatility. In T. Andersen, R. Davis, J. Kreiss, and T. Mikosch (Eds.)," Volatility. In J. Birge and V. Linetsky (Eds.), Handbook of Financial Time Series. Springer Verlag.
2. ANDERSEN, T. G., BONDARENKO, O. and GONZALEZ-PEREZ, M. T., (2015): "Exploring return dynamics via corridor implied volatility." Review of Financial Studies, pp. 2902-2945.
3. ANDERSEN, T., BOLLERSLEV, T. and HUANG, X., (2010): "A reduced form framework for modeling volatility of speculative prices based on realized variation measures." Journal of Econometrics 160(1): pp. 176-189.
4. ANDERSEN, T., BOLLERSLEV, T., DIEBOLD, F. and LABYS, P., (2003): "Modeling and Forecasting Realized Volatility." Econometrica 71(2): pp. 579-625.
5. ANDERSEN, T.G., BENZONI, L., LUND, J., (2002). “An Empirical Investigation of Continuous-Time Equity Return Models“, The Journal of Finance, June 2002, Vol.57, No.3, pp.1239-1284
6. ANDERSEN, T.G., BOLLERSLEV, T. and DIEBOLD, F.X., (2007). "Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility", Review of Economics and Statistics, November 2007, Vol. 89, No. 4, p.701-720.
7. ANDERSEN, T.G., BOLLERSLEV, T., DIEBOLD, F.X. and LABYS, P., (1999): "Understanding, Optimizing, Using and Forecasting Realized Volatility and Correlation." Working Paper.
8. ANDERSEN, T.G., BOLLERSLEV, T., DIEBOLD, F.X., LABYS, P., (2003). “Modeling and Forecasting Realized Volatility”, Econometrica, March 2003, Vol. 71, No. 2, p. 579-625
9. ANDERSEN, T.G., BOLLERSLEV, T., CHRISTOFFERSEN, P.F. and DIEBOLD, F.X., (2005). "Volatility Forecasting", National Bureau of Economic Research, March 2005, NBER Working Paper No. 11188.
10. ANDERSEN, T.G., DOBREV, D. and SCHAUMBURG, E., (2014). "A Robust Neighborhood Truncation Approach to Estimation of Integrated Quarticity", Econometric Theory, February 2014, Vol. 30, No.1, pp.3-59.
11. ARLT, Josef, ARLTOVÁ, Markéta, (2009): "Ekonomické časové řady." 1. vyd. Praha: Professional publishing, 275 s. ISBN 978-80-86946-85-6.

12. BARNDORFF-NIELSEN, O. E., HANSEN, P.R., LUNDE, A., SHEPARD, N., (2008). "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise", Econometrica, November 2008, Vol.76, p.1481–1536.
13. BARNDORFF-NIELSEN, Ole E. and SHEPHARD Neil, (2002). "Econometric analysis of realized volatility and its use in estimating stochastic volatility models", Journal of the Royal Statistical Society: Series B, Vol.64, No.2, pp. 253-280.
14. BARNDORFF-NIELSEN, Ole E. and SHEPHARD Neil, (2004). "Power and Bipower Variation with Stochastic Volatility and Jumps", Journal of Financial Econometrics, 2004, Vol.2, No.1, p.1-48
15. BARNDORFF-NIELSEN, Ole E. and SHEPHARD, Neil, (2001): "Non-gaussian Orstein-Uhlenbeck based models and some of their uses in financial economics." Journal of the Royal Statistical Society, Series B (63): pp. 167-241.
16. BEKIERMAN, Jeremias, MANNER, Hans, (2018). "Forecasting realized variance measures using time-varying coefficient models", International Journal of Forecasting, Vol.34, No.2, April-June 2018, pp. 276-287
17. BLACK, Fischer, (1976): "Noise." Journal of Finance.
18. BLACK, Fischer, SCHOLES, Myron, (1973). "The Pricing of Options and Corporate Liabilities", Journal of Political Economy, May-June 1973, Vol.81, No.3, pp.637-654
19. BOLLERSLEV, T., PATTON, A. J., QUAEDVLIEG, R., (2016). "Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting", Journal of Econometrics, May 2016, Vol.192, No.1, pp.1-18.
20. CAMPBELL, J.Y., LO, A.W. and MACKINLAY, A.C., (1997): "The Econometrics of Financial Markets." Princeton.
21. CORSI, Fulvio, (2004). "A Simple Long Memory Model of Realized Volatility", Institute of Finance, University of Lugano, Switzerland, August 2004, Working paper, p.1-31
22. FALKENBERRY, T. N., (2002): "High Frequency Data Filtering." Technical report.
23. FIČURA, Milan and WITZANY, Jiří, (2015): "Using high-frequency power-variation estimators in the Bayesian estimation of Stochastic-Volatility Jump-Diffusion models." The 9th International Days of Statistics and Economics, Prague, September, 2015, pp. 423-434, ISBN 978-80-87990-09-3.
24. FIČURA, Milan and WITZANY, Jiří, (2017): "Estimation of SVJD Models with Bayesian Methods." Advanced methods of computational finance, Prague: Oeconomica, 2017, pp.161-204, ISBN 978-80-245-2207-4.

25. FIČURA, Milan, (2018): "Modelling and Forecasting of Stochastic Volatility and Jumps." Dissertation Thesis, University of Economics in Prague.
26. GILDER, D., SHACKLETON, M. B. and TAYLOR, S. J., (2014): "Cojumps in stock prices: Empirical evidence." Journal of Banking & Finance 40(C): pp. 443-459.
27. HULL, John, (2012): "Options, futures, and other derivatives." Boston: Prentice-Hall, ISBN 978-0-13-216494-8.
28. IVANOVSKI, Zoran, STOJANOVSKI, Toni and NARASANOV, Zoran, (2015): "Volatility and Kurtosis of Daily Stock Returns at MSE." UTMS Journal of Economics 6 (2): pp.209–221.
29. KRISTOUFEK, Ladislav, (2014): "Commodity futures and market efficiency." Energy Economics 42, pp. 50-57.
30. LOBATO, Ignacio and VELASCO, Carlos, (2000): "Long Memory in Stock-Market Trading Volume." Journal of Business & Economic Statistics, 18, (4): pp. 410-27.
31. LUX, T. and MARCHESI, M., (2000): "Volatility Clustering in Financial Markets: A Microsimulation of Interacting Agents." International Journal of Theoretical and Applied Finance, 3, pp. 675-702.
32. MANDELBROT, B., (1963): "The Variation of Certain Speculative Prices." The Journal of Business, Vol. 36, No. 4.
33. MINCER, Jacob, ZARNOWITZ, Victor, (1969). “The Evaluation of Economic Forecasts“, Economic Forecasts and Expectations, J. Mincer (Ed.), National Bureau of Economic Research (NBER), New York, 1969, ISBN:0-870-14202-X, pp.1-47.
34. NELSON, Daniel, (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach." Econometrica, 59, pp. 347-370.
35. POON, S.H., GRANGER, C.W.J, (2003): "Forecasting financial market volatility: A review." Journal of Economics Literature, 41, 2, pp. 478-539.
36. PROTTER, P., (1992): "Stochastic integration and differential equations: A new approcah." New York: Springer-Verlag.
37. SEĎA, P., (2012): "Performance of Heterogeneous Autoregressive Models of Realized Volatility: Evidence from U.S. Stock Market." World Academy of Science, Engineering and Technology.
38. SHEPHARD, Neil, (2005): "Stochastic volatility: selected readings." Oxford: Oxford University Press, ISBN 0-19-925720-5.

39. TAKAHASHI, M., OMORI, Y., WATANABE, T., (2009): "Estimating Stochastic Volatility Models Using Daily Returns and Realized Volatility Simultaneously", Computational Statistics & Data Analysis, April 2009, Vol.53, No.6, pp.2404-2426.
40. WITZANY, Jiří, (2013): "Financial Derivatives. Valuation, Hedging and Risk Management." 1. vyd. Praha: Nakladatelství Oeconomica, 374 s., ISBN 978-80-245-1980-7.
41. ZHANG, L., (2006): "Effcient estimation of stochastic volatility using noisy observations: A multi-scale approach." Bernoulli (12): pp. 1019-1043.
42. ZHANG, L., MYKLAND, Per A. and AIT-SAHALIA, Y., (2005). “A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data“, Journal of the American Statistical Association, December

Seznam tabulek a grafů

Seznam tabulek

Tabulka 1 Popisné statistiky logaritmických výnosů MOEX indexu za období (2005-2019) .28
Tabulka 2 Popisné statistiky RV, BV, J 31
Tabulka 3 HAR-RV za období (2005-2015) in sample.....33
Tabulka 4 HAR-RV model out-sample predikce.....34
Tabulka 5 HAR-RV-J za období (2005-2015) in sample.....36
Tabulka 6 HAR-RV-J model out-sample predikce 37
Tabulka 7 HAR-Q za období (2005-2019).....38
Tabulka 8 HAR-Q model out-sample predikce 39
Tabulka 9 HAR-Q-J za období (2005-2015) 41
Tabulka 10 HAR-Q-J model out-sample predikce.....42
Tabulka 11 Mincer – Zarnowitz regrese, in-sample data 43
Tabulka 12 Mincer – Zarnowitz regrese, out-sample data 44
Tabulka 13 Porovnání predikčních schopností modelů podle R^2 a RMSE (out-sample).....45
Tabulka 14 Porovnání predikčních schopností modelů na různých délkách in-sample dat.....48
Tabulka 15 Mincer – Zarnowitz regrese, out-sample data (týdenní horizont) 53
Tabulka 16 Porovnání predikčních schopností modelů podle R^2 a RMSE (týdenní horizont) 53

Seznam grafů

Graf 1 Cenový vývoj MOEX indexu za období (2005-2019) 26
Graf 2 Vývoj logaritmických výnosů za období (2005-2019).....27
Graf 3 Realizovaná variance a Bipower variance 29
Graf 4 Porovnání Realizované variance vs Bipower variance.....29
Graf 5 Variance skoků.....30
Graf 6 Autokorelační funkce realizované volatility 31
Graf 7 HAR-RV model (in-sample).....34
Graf 8 HAR-RV model out-sample predikce vs pozorovaná RV 35
Graf 9 HAR-RV-J model (in-sample) 36

Graf 10 HAR-RV-J model out-sample predikce vs pozorovaná RV 37
Graf 11 HAR-Q model (in-sample).....39
Graf 12 HAR-Q model out-sample predikce vs pozorovaná RV 40
Graf 13 HAR-Q-J model (in-sample).....42
Graf 14 HAR-Q-J model out-sample predikce vs pozorovaná RV.....43
Graf 15 HAR-RV-J Year-by-Year odhad parametrů modelů.....46
Graf 16 HAR-Q-J Year-by-Year odhad parametrů modelů 47
Graf 17 Nejhorší a nejlepší out-sample předpověď HAR-RV-J modelu 49
Graf 18 Nejhorší a nejlepší out-sample předpověď HAR-Q-J modelu 49
Graf 19 HAR-RV model in-sample a out-sample předpovědi (týdenní horizont).....50
Graf 20 HAR-RV-J model in-sample a out-sample předpovědi (týdenní horizont).....51
Graf 21 HAR-Q model in-sample a out-sample předpovědi (týdenní horizont) 52
Graf 22 HAR-Q-J model in-sample a out-sample předpovědi (týdenní horizont).....52