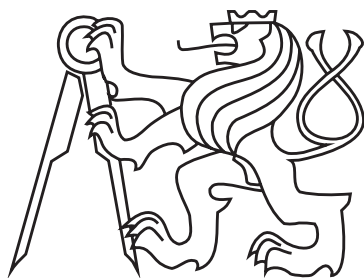


ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA ELEKTROTECHNICKÁ
KATEDRA EKONOMIKY, MANAŽERSTVÍ A HUMANITNÍCH
VĚD



Modelování a předpověď tržních cen elektrické energie
Modelling and forecast of market prices of electricity

DIPLOMOVÁ PRÁCE

Studijní program: Elektrotechnika, energetika a management

Studijní obor: Ekonomika a řízení energetiky

Vedoucí práce: doc. RNDr. Kateřina Helisová, Ph.D.

Bc. Martin Dolenský

Praha, 2018

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Dolenský** Jméno: **Martin** Osobní číslo: **420288**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávající katedra/ústav: **Katedra ekonomiky, manažerství a humanitních věd**
Studijní program: **Elektrotechnika, energetika a management**
Studijní obor: **Ekonomika a řízení energetiky**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Modelování a předpověď tržních cen elektrické energie

Název diplomové práce anglicky:

Modelling and forecast of market prices of electricity

Pokyny pro vypracování:

1. Nastudování různých metod statistického modelování cen elektrické energie na krátkodobém trhu.
2. Zpracování reálných dat o cenách elektřiny.
3. Aplikace modelů a metod předpovědí cen elektrické energie na reálná data a statistické vyhodnocení jejich přesnosti.
4. Vyvození ekonomických závěrů, využitelnost předpovědi cen elektrické energie v praxi.

Seznam doporučené literatury:

Weron, R.: Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach. John Wiley & Sons, 2006. ISBN 13 978-0-470-05753-7.

Jméno a pracoviště vedoucí(ho) diplomové práce:

doc. RNDr. Kateřina Helisová, Ph.D., katedra matematiky FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **26.01.2018** Termín odevzdání diplomové práce: **25.05.2018**

Platnost zadání diplomové práce: **30.09.2019**

doc. RNDr. Kateřina Helisová, Ph.D.
podpis vedoucí(ho) práce

podpis vedoucí(ho) ústavu/katedry

prof. Ing. Pavel Ripka, CSc.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

Datum převzetí zadání

Podpis studenta

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne 18. 5. 2018

Podpis autora práce

Poděkování

Tímto bych chtěl poděkovat paní doc. RNDr. Kateřině Helisové, Ph.D., vedoucí mé diplomové práce, za cenné rady a trpělivé vedení, které mi pomohly práci zdárně dokončit. Dále děkuji své rodině za podporu během studia.

Abstrakt

Tato práce se zabývá krátkodobým predikváním spotových cen elektrické energie. Práce obsahuje přehled přístupů k predikování cen elektřiny. Jsou zde popsány vybrané metody z kategorie statistických přístupů, konkrétně z Boxovy–Jenkinsovy metodologie a metodologie GARCH. Vybrané metody jsou použity k aplikaci na reálná data, která zahrnují spotové ceny elektrické energie na denním trhu OTE za rok 2017. Na závěr byla ověřena předpovědní schopnost sestrojených modelů.

Klíčová slova: cena elektrické energie, spotová cena, predikce, předpověď, Box-Jenkins, GARCH, ARIMA, ARIMA-G

Abstract

The work presented here concerns itself with short-term forecasting of spot electricity prices in the day-ahead market. This work presents an overview of the approaches used for forecasting of electricity prices. Moreover, the methods chosen for forecasting are described, containing approaches based on the Box–Jenkins methodology and the GARCH methodology. The chosen methods are then applied to real-world data which consists of spot electricity prices in the OTE day-ahead market. In conclusion, the effectiveness of forecasting models is verified.

Keywords: electricity price, spot electricity price, forecast, forecasting, Box-Jenkins, GARCH, ARIMA, ARIMA-G

Obsah

1	Úvod	17
2	Princip a fungování trhu s elektřinou	21
2.1	Druhy trhů a jejich fungování	21
2.1.1	Blokový trh	22
2.1.2	Denní trh	22
2.1.3	Vnitrodenní trh	23
2.1.4	Vyrovnávací trh s regulační energií	24
2.2	Stanovení ceny elektřiny	24
2.2.1	Marginální cena, MCP	24
2.2.2	Marginální vs. diskriminatorní trh	25
3	Metodologie	27
3.1	Používané metody pro předpověď cen	27
3.1.1	Dělení metod dle časového horizontu předpovědi	27
3.1.2	Dělení metod dle použité metodologie	28
3.2	Boxova–Jenkinsova metodologie	32
3.2.1	Stacionarita	32

3.2.2	Autokovarianční a autokorelační funkce a jejich odhad	33
3.2.3	Parciální autokorelační funkce a její odhad	34
3.3	Proces klouzavých průměrů MA	35
3.4	Autoregresní proces AR	36
3.5	Smíšený proces ARMA	38
3.6	Aplikace Boxovy–Jenkinsovy metodologie	39
3.6.1	Identifikace modelu	39
3.6.2	Odhad modelu	41
3.6.3	Diagnostika modelu	41
3.7	Proces ARIMA	47
3.8	Proces SARIMA	48
3.9	Modely GARCH	49
3.9.1	ARCH model	50
3.9.2	Testy na podmíněnou heteroskedasticitu	51
3.9.3	GARCH modely	52
3.10	Posuzování přesnosti předpovědi	52
3.10.1	Mean Error (ME)	52
3.10.2	Mean Percentage Error (MPE)	53
3.10.3	Mean Absolute Error (MAE)	53
3.10.4	Mean Absolute Percentage Error (MAPE)	53
3.10.5	Root Mean Square Error (RMSE)	54
4	Realizace predikce, numerické výsledky	55
4.1	Data	55

<i>OBSAH</i>	11
4.1.1 Změny zimního a letního času, korekce datasetu	55
4.1.2 Sezónnost	56
4.2 Realizace modelování a předpovědi	56
4.2.1 Identifikace a konstrukce adekvátního modelu	56
4.2.2 Predikce na základě vybraného modelu	57
4.2.3 Realizace modelů ARIMA–GARCH	57
4.3 Výzkum efektivity předpovědi pro první hodinu	58
4.3.1 AIC přesnější než BIC	58
4.3.2 Nevhodnost krátkých učicích období	60
4.4 Výzkum pro 24 souborů dat	61
4.5 Algoritmus předpovědi pro 24 hodin	61
4.6 Vyhodnocení přesnosti předpovědí ARIMA	64
4.7 Algoritmus předpovědi ARIMA–GARCH	70
4.8 Vyhodnocení přesnosti předpovědí ARIMA–G	71
5 Závěr	79

Seznam obrázků

1.1	Počty a druhy vědeckých prací publikovaných na téma predikování cen elektrické energie v letech 2000 až 2016. [12]	18
1.2	Rozdělení vědeckých prací o predikování cen elektrické energie na základě použitých přístupů. [12]	19
2.1	Princip fungování denního trhu. [20]	23
2.2	Nabídková a poptávková křivka a výsledná marginální cena MCP v bodě jejich průniku. Vlevo pak oboustranná aukce, vpravo jednostranná aukce v systému Single Buyer (SB). [10, 20]	25
3.1	Druhy přístupů k předpovědi cen elektrické energie. [20]	31
3.2	Příklad funkce odezvy na impuls (impulse–response funkce). Vlevo impulse–response funkce procesu MA(3) ($\theta_1 = 0,6; \theta_2 = -0,5; \theta_3 = 0,4$), vpravo pak funkce procesu AR(2) ($\varphi_1 = 0,6; \varphi_2 = 0,2$) s jednotkovým impulsem v čase 0. [8]	42
4.1	Maximální zjištěné chyby pro jednotlivé délky učícího období. Vlevo pro informační kritérium <i>AIC</i> , vpravo <i>BIC</i>	59
4.2	Vlevo průměrná chyba v závislosti na zvolené délce učícího období, vpravo pak maximální zjištěná chyba v závislosti na délce učícího období. Grafy nemají stejné měřítko, i tak je však jednoznačné zhoršování předpovědí se zkracující se učící délkou pod hranicí 50 hodnot.	60
4.3	Nejmenší průměrná <i>RMSE</i> v jednotlivých hodinách při optimálních délkách učícího období a použití informačního kritéria <i>AIC</i> . <i>ADF</i> –test červeně, resp. <i>KPSS</i> –test černě.	63

4.4	<i>RMSE</i> predikčního algoritmu pro třídní předpověď se začátkem ve dnech 200 až 363 (19. 7. až 29. 12. 2017).	64
4.5	Predikce a skutečné hodnoty třídní předpovědi MCP pro 265. až 267. den (22. 9. až 24. 9. 2017), tedy případu s nejmenší chybou. Červeně je znázorněna predikce, černě skutečné hodnoty.	66
4.6	Predikce a skutečné hodnoty třídní předpovědi MCP pro 300. až 302. den (27. 10. až 29. 10. 2017), tedy případu s největší chybou. Červeně je znázorněna predikce, černě skutečné hodnoty.	67
4.7	Predikce a skutečné hodnoty jednodenních předpovědí MCP pro 300., 301. a 302. den (27. 10., 28. 10. a 29. 10. 2017). Červeně je znázorněna predikce, černě skutečné hodnoty.	69
4.8	<i>RMSE</i> predikčního algoritmu pro třídní předpověď se začátkem ve 200. až 363. dni (19. 7. až 29. 12. 2017). Modře jsou znázorněny chyby původního předpovědního algoritmu <i>ARIMA</i> , červeně pak předpovědní chyby algoritmu <i>ARIMA – G</i>	71
4.9	Predikce a skutečné hodnoty třídní předpovědi MCP pro 325. až 327. den (21. 11. až 23. 11. 2017), tedy případu, kdy rozšíření <i>ARIMA – G</i> přineslo největší zlepšení. Červeně je znázorněna predikce <i>ARIMA – G</i> , modře predikce původního algoritmu <i>ARIMA</i> , černě skutečné hodnoty. .	73
4.10	Predikce a skutečné hodnoty třídní předpovědi MCP pro 312. až 314. den (8. 11. až 10. 11. 2017), tedy případu, kdy rozšíření <i>ARIMA – G</i> přineslo největší zhoršení. Červeně je znázorněna predikce <i>ARIMA – G</i> , modře predikce původního algoritmu <i>ARIMA</i> , černě skutečné hodnoty. .	77
4.11	Predikce a skutečné hodnoty třídní předpovědi MCP pro 332. až 334. den (28. 11. až 30. 11. 2017), tedy případu, kdy rozšíření <i>ARIMA – G</i> přineslo druhé největší zhoršení. Červeně je znázorněna predikce <i>ARIMA – G</i> , modře predikce původního algoritmu <i>ARIMA</i> , černě skutečné hodnoty. .	77

Seznam tabulek

4.1	Nejlepší délky učicích období a příslušné průměrné chyby <i>RMSE</i> . Podtrženě jsou uvedeny učicí délky příslušející testu na jednotkový kořen s nižší průměrnou chybou, které spolu s příslušným testem používá predikční algoritmus.	62
4.2	Průřezové statistiky rozptylů modelů <i>ARIMA</i> jednotlivých hodin z třídenního výzkumu <i>RMSE</i>	65
4.3	Parametry předpovědních modelů predikce 265. až 267. dne (22. 9. až 24. 9. 2017), tedy případu s nejmenší chybou.	65
4.4	Parametry předpovědních modelů predikce 300. až 302. dne (27. 10. až 29. 10. 2017), tedy případu s největší chybou.	68
4.5	Parametry předpovědních modelů jednodenní predikce 301. dne (28. 10. 2017).	68
4.6	Parametry předpovědních modelů jednodenní predikce 302. dne (29. 10. 2017).	69
4.7	Parametry předpovědních modelů predikce 325. až 327. dne (21. 11. až 23. 11. 2017), tedy případu, kdy rozšíření <i>ARIMA</i> – <i>G</i> přineslo největší zlepšení. Řádek 2 obsahuje <i>p</i> –hodnoty testu na heteroskedasticitu <i>LM</i> (Lagrange Multiplier) reziduí nerozšířených modelů <i>ARIMA</i> . Pokud byla zamítnuta nulová hypotéza homoskedasticity, byl zvolen model <i>ARIMA</i> – <i>G</i> a řádky 4 a 6 pak obsahují <i>p</i> –hodnoty testů jeho reziduí (<i>Q</i> –testu na nekořelovanost, McLeod–Li testu na remanentní <i>ARCH</i> –efekt). V opačném případě byl zvolen model <i>ARIMA</i> s <i>p</i> –hodnotou <i>Q</i> –testu jeho reziduí na řádce 4 a se směrodatnou odchylkou na řádce 5. Řádek 3 obsahuje řády modelu <i>ARIMA</i> v obou případech.	74

- 4.8 Parametry předpovědních modelů predikce 312. až 314. dne (8. 11. až 10. 11. 2017), tedy případu, kdy rozšíření *ARIMA* – *G* přineslo největší zhoršení. Řádek 2 obsahuje *p*–hodnoty testu na heteroskedasticitu *LM* (Lagrange Multiplier) reziduí nerozšířených modelů *ARIMA*. Pokud byla zamítnuta nulová hypotéza homoskedasticity, byl zvolen model *ARIMA* – *G* a řádky 4 a 6 pak obsahují *p*–hodnoty testů jeho reziduí (Q–testu na nekorelovanost, McLeod–Li testu na remanentní *ARCH*–efekt). V opačném případě byl zvolen model *ARIMA* s *p*–hodnotou Q–testu jeho reziduí na řádku 4 a se směrodatnou odchylkou na řádku 5. Řádek 3 obsahuje řády modelu *ARIMA* v obou případech. 75
- 4.9 Parametry předpovědních modelů predikce 332. až 334. dne (28. 11. až 30. 11. 2017), tedy případu, kdy rozšíření *ARIMA* – *G* přineslo druhé největší zhoršení. Řádek 2 obsahuje *p*–hodnoty testu na heteroskedasticitu *LM* (Lagrange Multiplier) reziduí nerozšířených modelů *ARIMA*. Pokud byla zamítnuta nulová hypotéza homoskedasticity, byl zvolen model *ARIMA* – *G* a řádky 4 a 6 pak obsahují *p*–hodnoty testů jeho reziduí (Q–testu na nekorelovanost, McLeod–Li testu na remanentní *ARCH*–efekt). V opačném případě byl zvolen model *ARIMA* s *p*–hodnotou Q–testu jeho reziduí na řádku 4 a se směrodatnou odchylkou na řádku 5. Řádek 3 obsahuje řády modelu *ARIMA* v obou případech. 76

Kapitola 1

Úvod

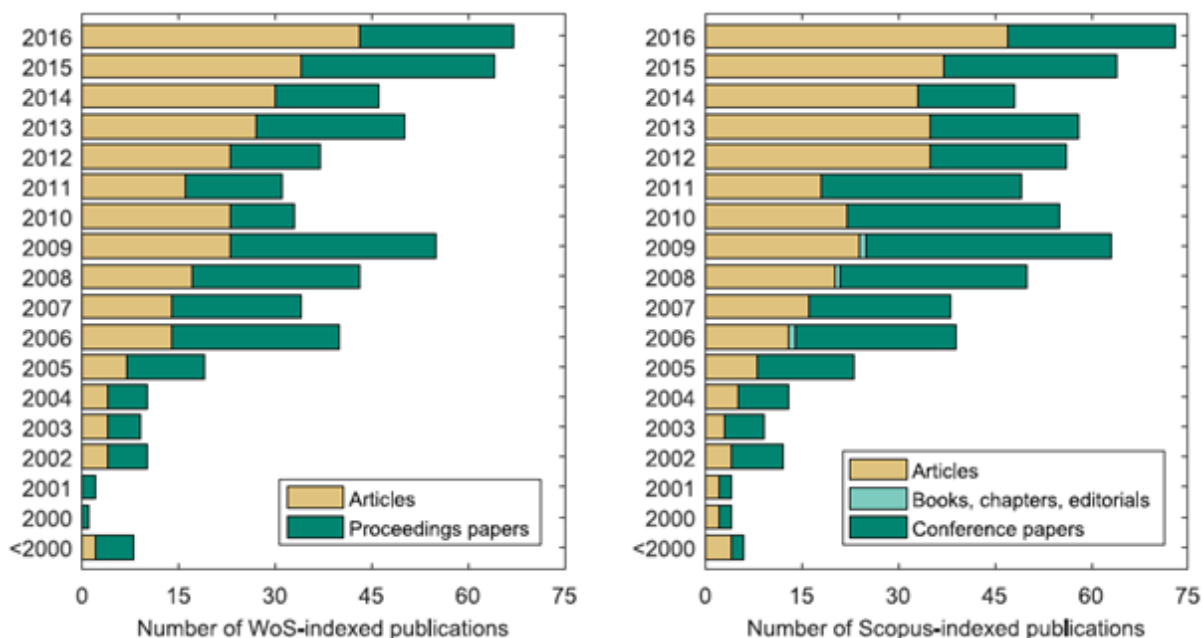
Téma predikování cen elektrické energie je spojeno s procesem liberalizace trhů s elektřinou, který zásadním způsobem změnil tvář a fungování energetických systémů mnoha států po celém světě. Liberalizace trhů s elektřinou je komplexní přeměna majetkových a obchodních vazeb v elektrizační soustavě. Energetické systémy na celém světě byly dříve zpravidla tvořeny vertikálně integrovanými monopoly, které byly zcela vlastněny a bez většího omezení regulovány příslušnými státy. Proces liberalizace, který byl dovršen již ve značné části světa, znamená především privatizaci státních energetických monopolů, rozdělení jednotlivých článků vertikálně integrovaného monopolu a zavedení institucí „obchodníka s elektřinou“ a „velkoobchodního trhu s elektřinou“. Jedná se tedy o zavedení tržních principů do dříve centrálně plánovaného odvětví. [10]

V Evropě, stejně jako v Americe, začal proces liberalizace v 90. letech 20. století. [17, 21] Velká Británie přistoupila k liberalizaci trhu s elektřinou v roce 1990 a samotný proces zahrnoval mimo jiné rozdělení monopolní společnosti dominující výrobě a přenosu elektřiny na více samostatných konkurenčních subjektů (horizontální separace) dezintegraci oblasti výroby a přenosu elektřiny (vertikální separace). Proces liberalizace je ve Velké Británii považován za završený od roku 1996. Ve Skandinávii byl velkoobchodní trh s elektřinou otevřen v Norsku v roce 1992, později se přidalo Švédsko, Finsko a Dánsko. Austrálie otevřela svůj národní trh s elektřinou v roce 1998. Ve stejném období probíhal proces liberalizace i v USA, v roce 1998 byl otevřen trh v Kalifornii. [21] V České republice lze považovat proces liberalizace za završený od roku 2006. [10]

Důvodem potřeby liberalizace trhu s elektřinou byl fakt, že obstarávání elektrizační soustavy systémem státně-monopolních firem již neodpovídalo stupni dosaženého vědeckotechnického vývoje a překázelo rozvoji obchodu s energií. [1, 18] Zavedení „tržní ceny elektřiny“ s sebou mimo jiné přineslo potřebu predikování její výše.

Až na relativně málo výjimek nevycházely do roku 2000 žádné vědecké publikace na

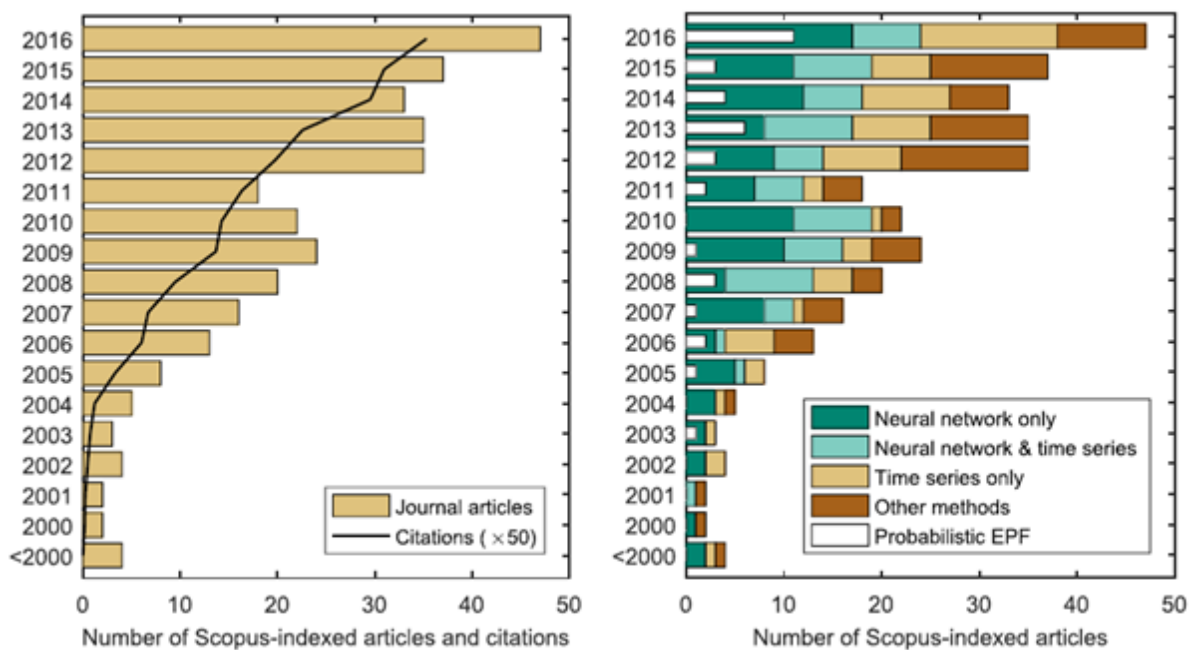
téma predikování tržních cen elektrické energie. [12, 20] Po zvýšeném počtu ročně publikovaných vědeckých prací po přelomu tisíciletí následoval velký nárůst v letech 2005 a 2006, ve kterých oproti roku 2004 došlo nejprve ke zdvojnásobení v roce 2005 a v roce 2006 byl počet publikovaných článků oproti roku 2004 již trojnásobný. Tento zvýšený zájem o téma predikování cen elektrické energie následovaly velké přírůstky ročně publikovaných prací v letech 2009 a 2016. Vývoj publikační činnosti na toto téma je znázorněn na Obr. 1.1 a Obr. 1.2.



Obrázek 1.1: Počty a druhy vědeckých prací publikovaných na téma predikování cen elektrické energie v letech 2000 až 2016. [12]

V současné době existuje rozsáhlý soubor etablovaných metodologií, viz Kapitola 3.1. Novým nastupujícím trendem je pak pravděpodobnostní způsob předpovídání cen elektřiny (probabilistic electricity price forecasting, dále jen PEPF). Ten stojí mj. i za zvýšeným počtem publikovaných vědeckých prací v letech 2014 až 2016. Ačkoliv je vědeckými autoritami metodologii PEPF přisuzován velký potenciál, zdráhají se ji zatím označovat za etablovanou a prověřenou. [12] Vědecké články publikované v posledních letech se však netýkají pouze PEPF. Značnou měrou je konstituují i články, v nichž figurují přístupy bez PEPF – mimo jiné i analýza časových řad.

Tato práce se zaměřuje na předpovídání spotových cen elektrické energie metodologií z kategorie statistických přístupů. Jmenovitě se jedná o autoregresní analýzu časových řad v rámci Boxovy–Jenkinsovy metodologie a metodologie *GARCH*. Práce obsahuje i shrnutí způsobu vzniku zkoumaných spotových cen. Cílem této práce je uvedení metodologie potřebné k realizaci predikce, dále aplikace této metodologie na reálná data a studium efektivnosti této aplikace.



Obrázek 1.2: Rozdělení vědeckých prací o predikování cen elektrické energie na základě použitých přístupů. [12]

Struktura práce je následující: Kapitola 2 obsahuje popis mechanismu stanovení tržní ceny elektrické energie. V kapitole 3 je uveden nejprve souhrn v současnosti užívaných metodologií, přičemž dále následuje konkrétní metodologie použitá k samotné realizaci predikce v této práci. Kapitola 4 obsahuje popis realizované predikce spotové ceny elektřiny a dosažené výsledky.

Kapitola 2

Princip a fungování trhu s elektřinou

2.1 Druhy trhů a jejich fungování

Krátkodobý trh s elektřinou je nástrojem liberalizovaného trhu s elektrickou energií, jehož prostřednictvím je umožněno jeho účastníkům svobodně nakupovat a prodávat elektrickou energii a rozhodovat mimo jiné způsobem anonymní aukce o její ceně.

Na rozdíl od většiny ostatních trhů, na kterých se obchoduje s jinými komoditami, má trh s elektrickou energií určitá specifika. Ta jsou dána zejména formou a vlastnostmi této komodity.

Nejdůležitější specifikum trhu s elektřinou však pramení z jedné její unikátní vlastnosti – skladování elektrické energie je nepraktické a z ekonomického hlediska většinou nesmyslné. Na trhu s elektrickou energií je stěžejní zachování výkonové bilance, tj. okamžitá výroba neboli dodávka se musí rovnat okamžité spotřebě. Nasmlouvané množství elektřiny je třeba příslušnými účastníky dodat či odebrat zcela přesně, a to jak v kvantitách, tak v čase.

Větší nerovnováhy v síti by hrozily ovlivněním parametrů elektřiny, např. nárůstem či poklesem frekvence, v daleko větším rozsahu, než který by postihl původce nerovnováhy. Nezodpovědnost jednoho účastníka trhu by zde tedy mohla přímo znemožnit ostatním účastníkům trhu nakoupit či prodat kvalitní produkt. V extrémních případech by byl navíc možný úplný blackout sítě. [9] V praxi je odchylka běžnou, a někdy i při nejlepším svědomí nevyhnutelnou, součástí obchodování. To ovšem neznamená, že se jedná o stav žádoucí a že za ni její původci nenesou finanční odpovědnost. Zachování výkonové bilance sítě, a tedy i přímé vyrovnávání odchylek zajišťuje provozovatel přenosové soustavy, společnost ČEPS, v rámci systémových služeb. Za každou případnou odchylku od bilance, o kterou se v rámci své činnosti musel ČEPS postarat, pak platí původce této odchylky

v rámci systému zúčtování odchylek provozovaném operátorem trhu OTE. [7]

Neskladovatelnost elektřiny, sezónnost poptávky a závislost poptávky i nabídky na aktuálním počasí, to vše se projevuje poměrně velkými změnami cen elektřiny v čase. Pohyb cen tohoto druhu a rozsahu nelze nalézt na žádném jiném trhu. [20]

Jedinečnost elektřiny se odráží i ve způsobu, jakým se s elektřinou obchoduje. Na organizovaných krátkodobých trzích i burzách s elektřinou se cena a obchodované množství elektrické energie určují vždy s nějakým časovým předstihem, jehož délka závisí na druhu trhu. S elektřinou nelze obchodovat spojitě v reálném čase. Je tomu tak z důvodu možnosti ověření proveditelnosti nasmlouvaných obchodů operátorem trhu z hlediska přenosových možností elektrizační soustavy. Dále tento způsob obchodování umožňuje již diskutované vyhodnocení a zúčtování odchylek (rozdílů mezi nasmlouvanými a opravdu uskutečněnými dodávkami či odběry) účastníků trhu. [20]

Organizovaný krátkodobý trh s elektřinou v ČR se dělí na:

1. blokový trh,
2. denní trh,
3. vnitrodenní trh,
4. vyrovnávací trh s regulační energií. [22]

2.1.1 Blokový trh

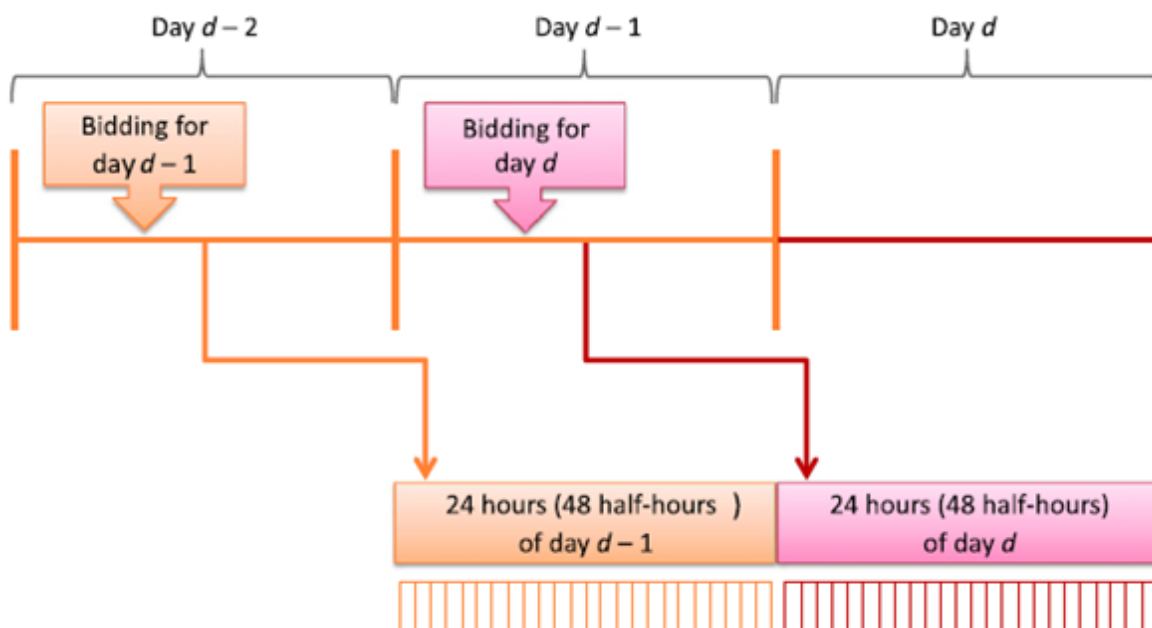
Organizovaný blokový krátkodobý trh s elektřinou slouží k obchodování konstantních hodnot hodinového zatížení v MW v hodinách časového období dvanáctihodinového či čtyřadvacetihodinového bloku. Existují tři druhy bloků, a sice Base (0:00 – 24:00 hod.), Peak (8:00 – 20:00) a Offpeak (0:00 – 8:00; 20:00 – 24:00). [14] Nejmenší jednotkou obchodního bloku na blokovém trhu je konstantní hodnota hodinového výkonu 1 MW v hodinách časového období bloku. Čas otevření blokového trhu je doba 5 dní před dnem dodávky v 9:30, okamžikem ukončení obchodování je pak čas 13:30 hodin dne před dnem dodávky.

2.1.2 Denní trh

Denní trh (DT OTE), umožňuje účastníkům obchodovat elektřinu v každé hodině dne dodávky (den d), přičemž pro každou z 24 hodin dne dodávky je aukčním způsobem separátně určeno množství obchodované elektřiny a cena za jednotkový konstantní výkon.

Nejmenší inkrement výkonu zde na rozdíl od blokového trhu není 1 MW, ale hodnota 0,1 MW. Čas otevření u denního trhu není omezený, uzavírá se v 11:00 hodin dne před dnem dodávky $d - 1$. [14]

Je to právě denní trh, na kterém se stanovuje marginální cena elektrické energie, jejíž predikci se zabývá tato práce. Cena stanovená na denním trhu se rovněž označuje jako „spotová“, stejně jako cena elektrické energie stanovená obdobným způsobem na jiném trhu, který tvoří ekvivalent DT OTE, např. EPEXSpot. Spotová cena na DT OTE je stanovována v eurech. [14]



Obrázek 2.1: Princip fungování denního trhu. [20]

2.1.3 Vnitrodenní trh

Organizovaný vnitrodenní trh s elektřinou je platforma, kde obchodníci sjednávají odběr či výrobu elektřiny již v průběhu obchodního dne d , a de facto slouží k eliminaci odchylky úpravou portfolia. [7] Vnitrodenní trh se otevírá ve dni $d - 1$ po uzavření denního trhu v 15:00 a obchod na něm lze domluvit až jednu hodinu před realizací odběru nebo dodávky ve dni d . Lze na něm tedy obchodovat téměř v reálném čase.

2.1.4 Vyrovnávací trh s regulační energií

Vyrovnávací trh s regulační energií je místem, kde se obchodníci s elektřinou setkávají s provozovatelem přenosové soustavy, společností ČEPS, a.s. Zde již nicméně nedochází k oboustrannému obchodu a nakupujícím je zde pouze společnost ČEPS a.s. Vyrovnávací trh s regulační energií se otevírá hodinu před hodinou dodávky a uzavírá 30 minut před hodinou dodávky. Účelem tohoto trhu je redukovat objem aktivované regulační energie, a tedy snížit náklady na podpůrné služby. [14]

Podpůrné služby i regulační energie nakoupená společností ČEPS na vyrovnávacím trhu pak slouží v rámci systémových služeb k zachování výše diskutované výkonové bilance. [9]

2.2 Stanovení ceny elektřiny

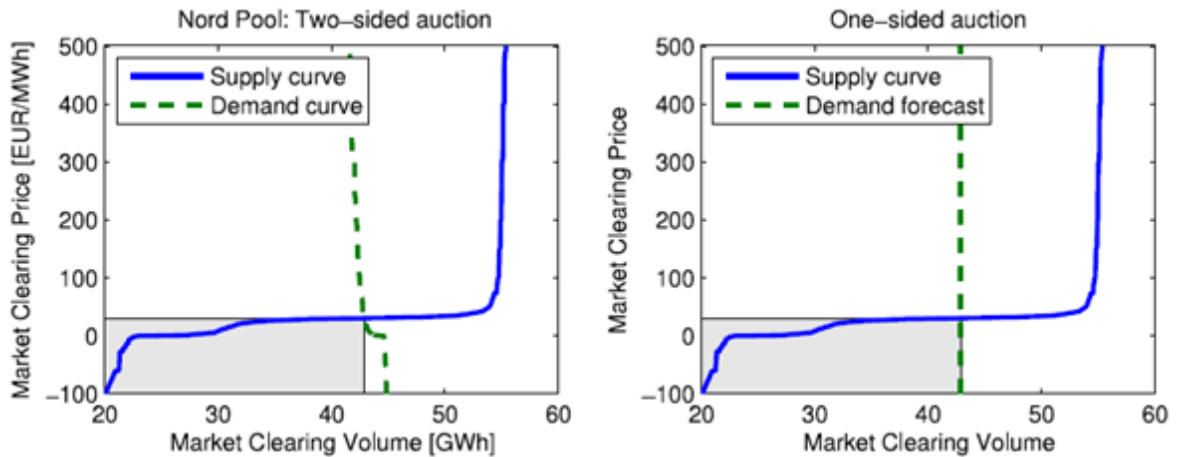
2.2.1 Marginální cena, MCP

Úlohou organizovaného trhu s elektřinou je zajištění styku nabídky a poptávky obchodujících subjektů. Konkrétně na denním, resp. spotovém, trhu se tak děje za pomoci stanovení tzv. marginální ceny (market clearing price, dále jen MCP). Trhy provozuje a zajišťuje nezávislý subjekt, jenž se obchodu nemůže účastnit. V případě denního trhu je to operátor trhu OTE. Marginální cena je na těchto trzích stanovována v aukci probíhající ve dni $d - 1$, v níž mohou registrovaní účastníci trhu překládat své nabídky a poptávky určitého množství elektrické energie za určitou cenu. V okamžiku uzavření trhu nashromáždí operátor trhu předložené nabídky a poptávky a zkonstruuje z nich agregovanou nabídkovou a poptávkovou křivku. Tento proces probíhá totožným způsobem pro všechny realizované aukce, tedy pro každou hodinu dne d . Tyto agregované křivky zachycují kombinace všech v okamžiku uzavření trhu předložených vůlí dodat či odebrat elektřinu v určitém množství a za určitou cenu. Marginální cena je pak stanovena zvlášť pro každou hodinu dne d jako průnik příslušné nabídkové a poptávkové křivky.

Účastníci jsou posléze v rámci každé hodinové aukce rozděleni na úspěšné a neúspěšné. Úspěšní účastníci jsou ti, kteří v aukci nechtěli nakupovat pod nebo prodávat nad konečnou marginální cenou dne d . Neúspěšní účastníci jsou pak všichni ostatní. Úspěšní účastníci budou moci ve dni d nakoupit či prodat elektřinu v souladu se svou vůlí ze dne $d - 1$.

Operátor trhu poté, stále ještě ve dni $d - 1$, oznámí tyto výsledky denního trhu jeho jednotlivým účastníkům. Poskytne jim informace o nabídkových a poptávkových křivkách pro den dodávky, množství zobchodované elektřiny a její cenu, a to pro každou obchodní

hodinu. [22] Účastníci mají díky tomu čas a informace potřebné k tomu, aby mohli vyrovnat svoji pozici na vnitrodenním či vyrovnávacím trhu.



Obrázek 2.2: Nabídková a poptávková křivka a výsledná marginální cena MCP v bodě jejich průniku. Vlevo pak oboustranná aukce, vpravo jednostranná aukce v systému Single Buyer (SB). [10, 20]

Na trhu s elektřinou je přípustné předkládat nabídky nebo poptávky se zápornou cenou. Ty pak reflektují situaci časového období, v němž je velmi nízká poptávka po elektřině a uplné odstavení produkujícího zdroje je neekonomické nebo minimálně méně výhodné než někomu platit za to, že odebere vyrobenou energii. I to je specifíkem trhu s elektřinou.

2.2.2 Marginální vs. diskriminatorní trh

Za zmínku dále stojí, že spotové trhy s elektřinou na světě se dělí na dvě skupiny v závislosti na tom, zda výše nabídky nebo poptávky předložené jeho účastníkem do aukce ve dni $d - 1$ nějak ovlivní konkrétní výši ceny elektřiny, která bude pro tohoto účastníka platit po tom, kdy dojde k uzavření trhu, a je určena marginální cena. Jinak řečeno, spotové trhy se dělí ve způsobu stanovení marginální ceny na „marginální“ a „diskriminatorní“.

Na marginálním trhu je MCP pro všechny úspěšně nabízející a poptávající účastníky trhu stejná. Na marginálním trhu zaplatí všichni poptávající, jejichž předložená nabízená částka za poptávanou elektřinu byla vyšší nebo rovna výsledné marginální ceně prostě výslednou marginální cenu. Pro výrobce platí zrcadlově stejné pravidlo: Výrobci, kteří za dodanou elektřinu požadovali částku nižší či rovnou výsledné MCP, obdrží za jimi poskytnutý jednotkový konstantní hodinový výkon MCP.

Diskriminatorní trh naproti tomu přiřazuje všem úspěšně nabízejícím a poptávajícím

přesně jimi nabízenou, resp. poptávanou, platbu za zobchodovanou elektřinu.

Český trh s elektřinou, stejně jako většina světových trhů, funguje na marginálním principu.

Princip diskriminatorního trhu adoptovala např. Velká Británie.

Kapitola 3

Metodologie

3.1 Používané metody pro předpověď cen

Existuje velké množství přístupů k predikci cen elektřiny. Budeme-li uvažovat i jejich kombinace, je počet variant, které lze uvažovat, opravdu nepřehledný. Existuje nicméně soubor kategorií, do něhož lze nějakým způsobem zařadit všechny existující postupy predikování, přičemž všechny mají své specifické silné a slabé stránky. [20] Přístupy k modelování a předpovídání se dělí z hlediska dvou hlavních kritérií. První takové kritérium je časový horizont předpovědi, druhým kritériem je konkrétní aplikovaná metodologie.

3.1.1 Dělení metod dle časového horizontu předpovědi

Z hlediska časového horizontu je zvykem hovořit o třech druzích predikování cen elektřiny, a sice o krátkodobém (short-term price forecasting, dále jen STPF), střednědobém (medium-term price forecasting, dále jen MTPF) a dlouhodobém (long-term price forecasting, dále jen LTPF) predikování cen elektřiny. Co se konkrétních časových horizontů týče, neexistuje univerzální klasifikace, která by těmto třem kategoriím připisovala konkrétní hodnoty časových rozpětí.

Dlouhodobé předpovědi se používají pro hodnocení investic do zdrojů elektrické energie a uvažované období predikce se zde zpravidla počítá na roky.

Střednědobé predikování se většinou provádí v rámci měsíců a nachází své uplatnění ve výpočtech týkajících se účetních bilancí, oceňování derivátů na elektrickou energii či risk-managementu.

Krátkodobé předpovědi jsou prováděny pro časový horizont několika hodin až několika dnů a využívají se pro plánování výroby a operativní působení na spotovém trhu. [21]

3.1.2 Dělení metod dle použité metodologie

Druhým kritériem klasifikujícím způsoby predikování cen elektřiny je použitá metodologie a podle ní se dělí do následujících skupin [20], jak je znázorněno na Obr. 3.1.

Multi-agent přístupy

Jako multi-agent přístupy se označují predikční modely, které k předpovědi přistupují simulací chodu systému (trhu s elektřinou), na kterém působí agenti (registrovaní účastníci trhu). Marginální cena elektřiny, která je předmětem zkoumání a predikce, je pak určena na základě nabídky a poptávky agentů. [11]

Multi-agent přístupy se vyvinuly z původních předliberalizačních predikčních postupů. V období před liberalizací se k účelu předpovědi cen používaly tzv. Production-cost modely (dále jen PCM). Ty se zpravidla zabývaly střednědobými a dlouhodobými předpověďmi, které byly odvozovány párováním budoucích odhadů poptávky a souboru existujících a plánovaných výrobních kapacit. PCM byly pro původní vertikálně-monopolní systém zcela dostačující. Z production-cost modelů nejprve vznikly tzv. Equilibrium modely (dále jen EM), zvané rovněž modely teorie her. EM zanesly do PCM modelů strategické chování účastníků trhu. Ačkoliv EM byly rovněž nahrazeny, a to agent-based přístupy, stále je vhodné tyto modely teorie her použít pro predikci cen elektřiny na trzích, kde neexistuje žádná historie cen, je známé rozložení poptávky a jsou známy náklady na výrobu elektřiny.

Multi-agent přístupy se vyznačují velkou flexibilitou při zkoumání chování účastníků trhů s elektřinou. Tato přednost je nicméně vykoupena zvýšenými požadavky na přijímání předpokladů napříč celým modelem (hráči, strategie i výplatní matice). Využívají se například při odhadech, zda budou ceny elektřiny vyšší než marginální náklady a jak to ovlivní chování účastníků trhu.

Fundamentální modely

Fundamentální modely se snaží o postulaci technických, meteorologických a ekonomických vztahů přítomných během produkce a obchodování s elektřinou. Po stanovení funkčních závislostí mezi jednotlivými proměnnými (počasí, zátěže, stavy systému) jsou tyto

jednotlivé proměnné samostatně a nezávisle modelovány a predikovány, a to většinou statistickými, neparametrickými přístupy.

Fundamentální modely se typicky používají pro střednědobé horizonty a tím pádem pro oceňování finančních derivátů a risk-management. Představují konkureční přístup k reduced-form modelům, oproti nimž mohou nabídnout detailnější popis fundamentálních hybných sil trhu, avšak za cenu zvýšené komplexity. Jsou velmi náročné na objem vstupních dat. Jejich použití dále vyžaduje přijetí určitých technicko-ekonomických předpokladů, a přesnost jimi předpovídaných cen je tedy velice citlivá na správnost těchto předpokladů.

Reduced-form modely

Reduced-form modely jsou přístupy k predikci cen elektřiny zařazované většinou do kategorie střednědobých přístupů a pro krátkodobé předpovědi se zpravidla nepoužívají. Jejich mechanismus je založený na statistickém zkoumání zpravidla jednorozměrné časové řady vyjadřující vývoj ceny elektrické energie v čase. Komplexní analýzou fungování trhu se narozdíl od fundamentálních modelů nezabývají. Reduced-form, tedy jump-diffusion a Markov regime-switching modely se používají zejména pro oceňování derivátů na elektřinu.

Statistické modely

U statistických modelů je modelování ceny založeno na převážně regresní či autoregresní metodologii. Do statistických modelů časových řad mohou případně vstupovat i vnější neboli exogenní proměnné jako počasí nebo hodinové hodnoty výroby či spotřeby. Statistické, a zejména pak autoregresní, modely jsou téměř výlučně používány ke krátkodobému předpovídání cen elektřiny. Mezi statistické modely patří:

Similar-day predikce se snaží v souboru dat nalézt dny s charakteristikami podobnými předpovídanému dni a předpověď zakládá na historických hodnotách cen právě v těchto dnech. Předpověď pak může tvořit jediná hodnota „podobného dne“ či určitá regrese více hodnot z těchto „podobných dnů“. Charakteristiky sdružující dny mezi sebou jako podobné jsou například den v týdnu či den v roce. Speciálním případem similar-day přístupu je tzv. naivní metoda, která za podobné považuje stejné dny v týdnu v rámci dvou po sobě nadcházejících týdnů. Pondělí by se tedy podle této logiky nemělo příliš lišit od pondělí předcházejícího týdne, úterý od úterý předcházejícího týdne atd.

Regresní modely vyjadřují vztah mezi určitým počtem prediktorních proměnných a na nich závislou kriteriální proměnnou. Zpravidla využívají ke svému fungování metodu nejmenších čtverců, při níž dochází k volbě vysvětlujících, prediktorních, proměnných tak, aby byla suma čtverců rozdílů mezi pozorovanými a předpověděnými hodnotami co nejmenší.

Autoregresní modely časových řad jsou standardní modely časových řad, které berou v úvahu jednak náhodnou povahu předpovídáné proměnné a zároveň její autokorelaci, tj. korelaci hodnot časové řady v určitém časovém odstupu. Sem se řadí modely z Boxovy–Jenkinsovy metodologie, mezi které patří modely *AR* (autoregressive), *MA* (moving average), *ARMA* (autoregressive moving average), *ARIMA* (autoregressive integrated moving average), *SARIMA* (seasonal autoregressive integrated moving average) a jejich modifikace s exogenními vnějšími proměnnými typu X (například $AR - X$, $MA - X$, $ARMA - X$, atd.)

Prahové (Threshold) autoregresní modely se snaží vystihnout „skokové“ chování cen elektřiny přepínáním mezi určitým počtem režimů, které představují různé autoregresní modely, podobně jako Markov regime–switching (dále jen MRS), u reduced–form modelů. Rozdíl oproti MRS nicméně spočívá ve způsobu přepínání mezi režimy. Při použití MRS je proměnná, která specifikuje právě používaný režim, tzv. latentní a celý proces přepínání je založený na pravděpodobnostech. Threshold autoregresní modely naproti tomu přepínají režimy čistě deterministicky na základě hodnoty určité pozorovatelné proměnné. Přepínání probíhá na základě zvolené prahové hodnoty (threshold) a okamžité hodnoty pozorovatelné proměnné. Onou pozorovatelnou proměnnou, jejíž okamžitá hodnota determinuje aktivní režim, bývá velmi často cena samotná.

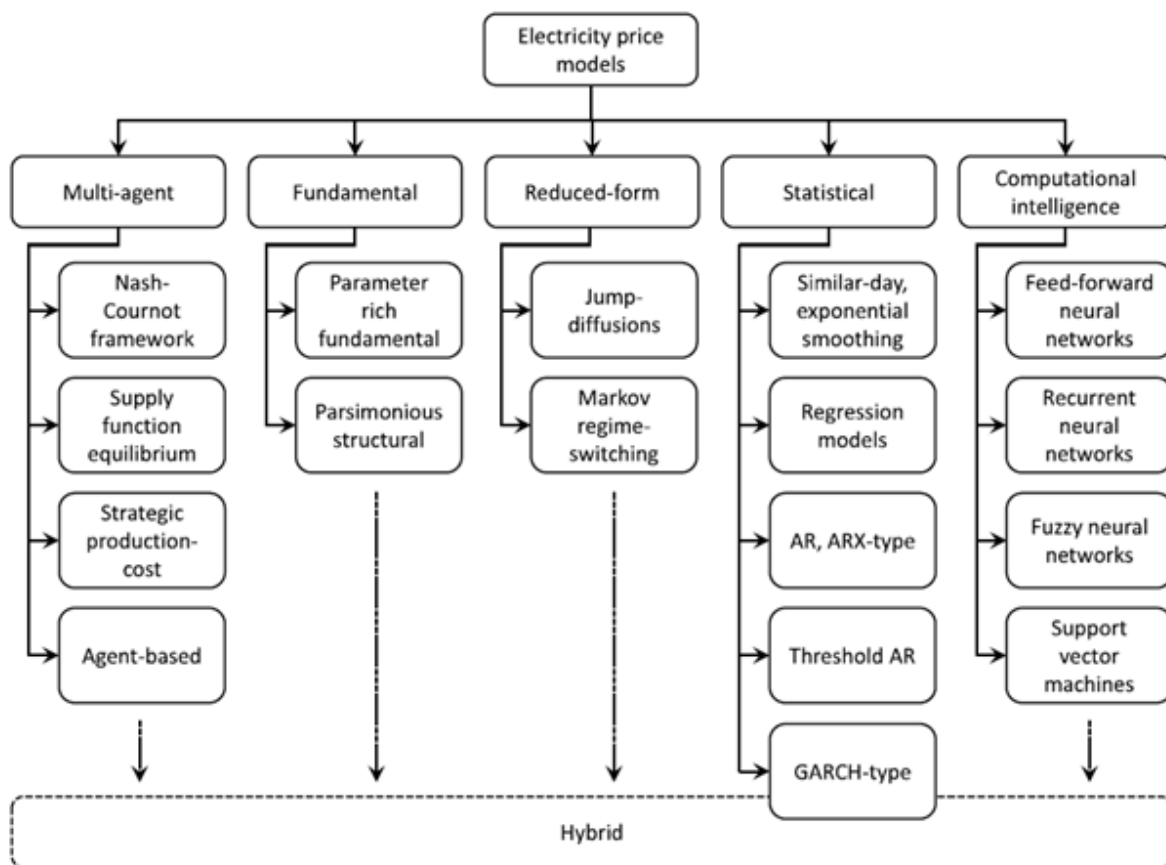
GARCH modely jsou přístupy k predikci, které se snaží řešit problém heteroskedasticity neboli nehomogenitu rozptylů. Autoregresní modely časových řad z Boxovy–Jenkinsovy metodologie vyžadují pro správné fungování alespoň slabou stacionaritu jimi zkoumaných časových řad, což kromě jiného znamená i jejich homoskedasticitu, tedy časovou neměnnost rozptylu jejich hodnot. Reálné časové řady však vykazují proměnný, a tedy nekonstantní, rozptyl, jehož hodnota je navíc závislá na předešlých hodnotách, což se s předpokladem stacionarity neshoduje. Rozptyl je tedy v těchto modelech, jinak podobných modelům *AR*, modelován svým samostatným autoregresním procesem a není uvažován konstatní.

Statistické metody vyžadují korektní numerické odhady, stejně jako kvalitní kalibrační data. K predikci ceny elektřiny v krátkodobém horizontu jsou vhodné díky její sezónnosti. Jejich slabinou je však špatná schopnost modelovat skoky, které jsou společně se sezónností s cenou elektřiny úzce spjaty.

Computational intelligence methods

Metody klasifikované jako Computational intelligence methods (dále jen CIM) jsou metody, které v sobě mohou zahrnovat strojové učení, evoluční algoritmy či fuzzy logiku.

Nejsilnější stránkou modelů CIM je jejich schopnost zvládnout nelinearitu a chování vykazující špičky. Tato jejich schopnost je ovšem zároveň znatelnou slabinou těchto modelů, neboť lepší schopnost přizpůsobit se nelineárnímu chování předpovídané veličiny, alespoň v případě predikce ceny elektrické energie, ještě neznamená lepší schopnost veličinu samotnou přesně předpovídat. Situaci ještě více komplikuje široké množství dostupných CIM přístupů, netrivialita porovnávání jejich přesností, a z toho pramenící složitost hledání optimálního modelu pro popis predikované veličiny.



Obrázek 3.1: Druhy přístupů k předpovědi cen elektrické energie. [20]

3.2 Boxova–Jenkinsova metodologie

Jedním ze způsobů, kterým lze přistoupit k predikci cen elektřiny, je užití Boxovy–Jenkinsovy metodologie. [2] Ta je založena na autoregresních stochastických metodách, jejichž podstatu tvoří autokorelační analýza. Boxova–Jenkinsova metodologie zahrnuje způsoby posuzování časových řad z hlediska přítomné autokorelace, soubor modelů určených k předpovědi budoucích hodnot a doporučení, jaký model je nejlepší pro popis dané časové řady, na základě druhu a míry detekované autokorelace. Zkoumaná řada je v našem případě jednorozměrná časová řada reálných čísel, která představují marginální cenu elektřiny MCP.

Následuje představení Boxovy–Jenkinsovy metodologie, přičemž tato práce vychází z konceptů popsanych v [2].

3.2.1 Stacionarita

Důležitým atributem časové řady pro potřeby Boxovy–Jenkinsovy metodologie je stacionarita. Stacionarita znamená, že je časová řada stochasticky ustálená, a lze ji chápat tak, že stochastické chování časové řady v libovolném okamžiku nezávisí na jeho pozici v čase. Pro potřeby metodologie má význam zejména, je-li řada tzv. slabě stacionární, a splňuje tedy následující podmínky:

1. Střední hodnota řady je konstantní a v čase neměnná:

$$E(y_t) = \mu = konst,$$

2. kovariance časové řady závisí pouze na velikosti časového kroku a je při stejném časovém kroku napříč časem neměnná:

$$cov(y_k, y_t) = E(y_k - \mu)(y_t - \mu) = cov(y_k + h, y_t + h) \quad h \in \mathbb{Z},$$

3. z druhé podmínky ještě vyplývá následující atribut slabě stacionární časové řady, a sice její konstantní rozptyl neboli homoskedasticita:

$$var(y_t) = \sigma_y^2 = konst.$$

Pozn.: U [15] je slabá stacionarita definovaná pouze vlastností 2. Zde, kde vycházím z [2], je tato definice stacionarity přísnější.

V literatuře se slabá stacionarita pro potřeby popisu fungování Boxovy–Jenkinsovy metodologie označuje rovněž zkráceně jako „stacionarita“. [2]

3.2.2 Autokovarianční a autokorelační funkce a jejich odhad

Autokovarianční a z ní vzešlá autokorelační funkce dohromady tvoří ukazatele, podle nichž v Boxově–Jenkinsově metodologii vybíráme z jejích dostupných modelů ten nejvhodnější pro modelování dané časové řady. Představují kvantitativní vyjádření korelovanosti časové řady v čase a lze je vyjádřit následujícím způsobem:

Autokovarianční funkce pro zpoždění k , nebo také „autokovariance pro zpoždění k “, má tvar

$$\gamma_k = \text{cov}(y_t, y_{t-k}) = E(y_t - \mu)(y_{t-k} - \mu), \quad k \in \mathbb{Z}.$$

Autokorelační funkce, zkráceně *ACF* (z angl. autocorrelation function), pro zpoždění k , nebo také „autokorelace pro zpoždění k “, má výraz

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma_k^2}, \quad k \in \mathbb{Z}.$$

Platí zároveň, že obě funkce jsou sudé, tj. $\gamma_k = \gamma_{-k}$ a $\rho_k = \rho_{-k}$.

V praxi se používají pro práci se stacionárními časovými řadami odhady těchto funkcí.

Pro jejich uskutečnění je nejprve zapotřebí bodový odhad střední hodnoty stacionární časové řady

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Dále je pak možné uskutečnit odhad autokovarianční funkce

$$c_k = \frac{1}{n} \sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y}), \quad k = 0, 1, \dots, n-1$$

a odhad autokorelační funkce

$$r_k = \frac{c_k}{c_0}, \quad k = 0, 1, \dots, n-1.$$

V zájmu zachování věcnosti odhadů autokovariance a autokorelace se pro příslušný počet pozorování n zpravidla požaduje, aby $n > 50$, a pro časový krok k platí $k < n/4$.

Důležitou hodnotou společnou pro autokovarianční i autokorelační funkci je tzv. bod useknutí k_0 . Bod useknutí je hodnota $k = k_0$, za kterou začínají být autokovarianční a autokorelační funkce ρ_k nulové.

Pro práci s odhady autokovariance a autokorelace a bodem useknutí je zapotřebí znát maximální vzdálenost odhadované korelační funkce r_k od nuly, abychom mohli s předem danou spolehlivostí o populační korelační funkci prohlásit, že je v tomto k nulová. K tomuto účelu se používá tzv. Bartlettova aproximace. Podle ní platí, že je-li $\rho_k = 0$ pro $k > k_0$, pak za předpokladu asymptotické normality je

$$r_k \sim N\left(0, \sqrt{\frac{1}{n}\left(1 + 2 \sum_{j=1}^{k_0} r_j^2\right)}\right) \quad \text{pro } k > k_0.$$

3.2.3 Parciální autokorelační funkce a její odhad

Důležitým ukazatelem autokorelace stacionárních časových řad je vedle autokovarianční a autokorelační funkce rovněž tzv. parciální autokorelační funkce (*PACF*, z angl. partial autocorrelation function).

Parciální autokorelační funkce, nebo též parciální autokorelace, s časovým krokem k se značí ρ_{kk} , a je definována jako parciální korelační koeficient mezi y_t a y_{t-k} při pevných hodnotách $y_{t-k+1}, \dots, y_{t-1}$. Odhad parciální autokorelační funkce se značí jako r_{kk} . Vezmeme-li např. autoregresní model $AR(k)$, viz (3.1),

$$y_t = \delta + \varphi_{k1}y_{t-1} + \varphi_{k2}y_{t-2} + \dots + \varphi_{kk}y_{t-k} + \epsilon_t,$$

pak odhadem r_{kk} parciální autokorelační funkce ρ_{kk} je bodový odhad $\hat{\varphi}_{kk}$ koeficientu φ_{kk} .

Rozdíl mezi autokorelací a parciální autokorelací je tedy dán tím, že v prosté autokorelaci je současná hodnota y_t ovlivněna i náhodnými složkami hodnot $y_{t-k+1}, \dots, y_{t-1}$, přičemž parciální autokorelace s tímto vlivem nepočítá a uvažuje pouze korelaci dvou náhodných složek, a to u proměnných y_t a y_{t-k} .

V praxi se odhad autokorelační funkce provádí dle vzorce:

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} \quad \text{pro } k > 1,$$

kde

$$r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j} \quad \text{pro } j = 1, \dots, k-1.$$

3.3 Proces klouzavých průměrů MA

Proces klouzavých průměrů řádu q se značí jako $MA(q)$ (z angl. moving average) a má tvar

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} = \theta(B) \epsilon_t,$$

kde $\theta_1, \theta_2, \dots, \theta_q$ jsou jeho parametry a $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ je tzv. operátor klouzavých průměrů.

Proměnná B dále představuje tzv. operátor časového posunu, přičemž pro jeho použití platí, že

$$\begin{aligned} \epsilon_t B &= \epsilon_{t-1}, \\ \epsilon_t B^q &= \epsilon_{t-q}. \end{aligned}$$

Členy $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$ jsou nezávislé, stejně rozdělené, tzv. iid, náhodné veličiny, které dohromady tvoří tzv. bílý šum, pro které platí:

$$\epsilon_t \sim N(0, \sigma^2).$$

Proces $MA(q)$ je vždy stacionární. Výše popsaný proces $MA(q)$ má nulovou střední hodnotu, tedy

$$E(y_t) = E(\epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}) = 0.$$

Rozptyl σ_y^2 procesu $MA(q)$ je v rámci splnění podmínek stacionarity rovněž časově neměnný, tj.

$$\sigma_y^2 = (1 + \theta_1^2 + \dots + \theta_q^2)\sigma^2.$$

Nakonec pro autokorelační funkci procesu $MA(q)$ platí, že

$$\rho_k = \begin{cases} \frac{\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_q\theta_{q-k}}{(1 + \theta_1^2 + \dots + \theta_q^2)\sigma^2} & \text{pro } k = 1, \dots, q, \\ 0 & \text{pro } k > q. \end{cases}$$

3.4 Autoregresní proces AR

Autoregresní proces řádu p se značí jako $AR(p)$ (z angl. autoregression) a má tvar

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \epsilon_t. \quad (3.1)$$

Členy $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$ tvoří stejně jako v případě procesu MA bílý šum, tj. platí:

$$\epsilon_t \sim N(0, \sigma^2).$$

Jinak lze autoregresní proces řádu p vyjádřit ve tvaru

$$y_t - \varphi_1 y_{t-1} - \varphi_2 y_{t-2} - \dots - \varphi_p y_{t-p} = \varphi(B)y_t = \epsilon_t,$$

kde $\varphi_1, \varphi_2, \dots, \varphi_p$ jsou parametry. Výraz $\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ se nazývá autoregresní operátor.

Narozdíl od procesů MA , které jsou pro konečná q stacionární vždy [8], je u procesů $AR(p)$ stacionarita podmíněna tím, že pro všechny kořeny z_1, z_2, \dots, z_p polynomu $\varphi(z)$ platí

$$|z_1|, \dots, |z_p| > 1,$$

tj. leží vně jednotkového kruhu.

Je-li splněna tato podmínka stacionarity, má výše popsaný proces AR střední hodnotu

$$E(y_t) = E(\varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \epsilon_t) = 0,$$

a rozptyl

$$\sigma_y^2 = \frac{\sigma^2}{1 - \varphi_1 \rho_1 - \dots - \varphi_p \rho_p}.$$

Autokorelační funkce procesu $AR(p)$ je dále určena výrazem

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p} \quad \text{pro } k > 0.$$

Pro vyjádření autokorelačních funkcí ρ_1, \dots, ρ_p pomocí parametrů modelu $AR(p)$ $\varphi_1, \dots, \varphi_p$ se používá tzv. soustava Yule–Walkerových rovnic:

$$\begin{aligned} \rho_1 &= \varphi_1 + \varphi_2 \rho_1 + \dots + \varphi_p \rho_{p-1}, \\ \rho_2 &= \varphi_1 \rho_1 + \varphi_2 + \dots + \varphi_p \rho_{p-2}, \\ &\vdots \\ \rho_p &= \varphi_1 \rho_{p-1} + \varphi_2 \rho_{p-2} + \dots + \varphi_p. \end{aligned}$$

Pomocí soustavy Yule–Walkerových rovnic můžeme s pomocí parametrů autoregresního modelu vyjádřit jeho autokorelace či naopak. Soustavu Yule–Walkerových rovnic zapíšeme následujícím způsobem. Nejprve definujme vektory φ a ρ parametrů $AR(p)$ modelu vztahy

$$\varphi = [\varphi_1, \dots, \varphi_p], \quad \rho = [\rho_1, \dots, \rho_p]$$

a matici Yule–Walker autokorelací

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix}.$$

Pak platí, že

$$\rho = \mathbf{R} \varphi$$

tudíž

$$\varphi = \mathbf{R}^{-1} \rho.$$

3.5 Smíšený proces ARMA

Smíšený proces *ARMA* je výsledek spojení metodiky *AR* a *MA* do jednoho modelu. Proces *ARMA* umožňuje modelování autokorelace časových řad ze dvou hledisek. Prvním hlediskem je, jak závisí současné hodnoty časové řady na hodnotách předchozích. Tuto závislost modeluje jeho *AR* část, která vysvětluje současnou hodnotu predikované veličiny na základě určité lineární kombinace zvoleného počtu předchozích pozorování a současné inovace zastoupené bílým šumem. *MA* část modelu pak propůjčuje schopnost lépe modelovat závislost současné hodnoty časové řady na předchozích inovacích.

Smíšený proces *ARMA* řádů p a q , značený jako $ARMA(p, q)$, je definován jako:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

Smíšený proces $ARMA(p, q)$ lze, je-li jeho *MA* proces invertibilní, vyjádřit zkráceným způsobem. Jelikož je ale možné jakýkoliv neinvertibilní *MA* proces vyjádřit invertibilním *MA* procesem, který je do momentů druhého řádu totožný, je podmínka invertibility procesu *MA* pro potřeby Boxovy–Jenkinsovy metodologie vždy splněna. [8] Smíšený proces $ARMA(p, q)$ je tedy možné vždy vyjádřit zkráceně jako

$$\begin{aligned} \varphi(B)y_t &= \theta(B)\epsilon_t \\ y_t &= \varphi(B)^{-1}\theta(B)\epsilon_t = \Psi(B)\epsilon_t. \end{aligned}$$

Podmínka stacionarity procesu *ARMA* se váže na podmínku stacionarity procesu *AR*. Podmínka invertibility je dále splněna vždy, a to díky možnosti vyjádřit jakýkoliv neinvertibilní *MA* proces jako invertibilní při zachování slabé stacionarity.

Střední hodnota smíšeného procesu se odvíjí od procesů *AR* a *MA*, které obsahuje, a na základě jejich nulové očekávané hodnoty je střední hodnota *ARMA* nulová, tj.

$$E(y_t) = E(\varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}) = 0.$$

Autokorelační funkce splňuje diferenciální rovnici

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p} \quad \text{pro } k > q.$$

3.6 Aplikace Boxovy–Jenkinsovy metodologie

Konstrukce modelu z Boxovy–Jenkinsovy metodologie se obvykle provádí ve třech krocích

1. Identifikace modelu: Na základě zkoumaných dat a hodnot z nich odvozených ukazatelů vybereme vhodný model z výše popsaných variant.
2. Odhad modelu: Odhad modelu je odhad konkrátních hodnot parametrů indentifikovaného modelu.
3. Diagnostika modelu: Model z kroku č. 2 může být otestován na hladině významnosti 5 % nebo se jinak posoudí a ověří jeho předpovědní schopnosti.

3.6.1 Identifikace modelu

Identifikace řádu modelu Boxovy–Jenkinsovy metodologie může probíhat dvěma způsoby. Lze ji provést buď subjektivně nebo čistě s pomocí výpočetní techniky.

Identifikace prostřednictvím tvaru autokorelační funkce

První, subjektivní, způsob identifikace řádu příslušného modelu časové řady se uskutečňuje prohlídkou tvarů autokorelační a parciální autokorelační funkce. Na základě vypracovaného tabulkového předpisu [2], který je součástí Boxovy–Jenkinsovy metodologie je poté zvolen vhodný model k popisu časové řady a predikci – $AR(p)$, $MA(q)$ nebo $ARMA(p, q)$.

	$AR(p)$	$MA(q)$	$ARMA(p, q)$
ρ_k	k_0 neexistuje, ρ_k ve tvaru křivky U	$k_0 = q$	k_0 neexistuje, ρ_k ve tvaru křivky U po prvních $p - q$ hodnotách
ρ_{kk}	$k_0 = p$	k_0 neexistuje, ρ_k omezená křivkou U	k_0 neexistuje, ρ_{kk} omezená křivkou U po prvních $p - q$ hodnotách

Identifikace prostřednictvím informačních kritérií

V případě identifikace modelu prostřednictvím informačních kritérií se řády p a q modelu $ARMA(p, q)$ se volí pomocí informačního kritéria AIC (Akaike Information Criterion) či kritéria BIC (Bayes Information Criterion). Varianta $AR(p)$ je pak vyjádřena jako případ smíšeného modelu $ARMA$ s nulovým řádem q , tj. $ARMA(p, 0)$. Model $MA(q)$ je pak zvolen analogickým způsobem, tj. $ARMA(0, q)$.

Informační kritéria jsou metody ohodnocení vhodnosti modelu pro popis empirických dat, které vycházejí z teorie informace. Tato kritéria posuzují modely na základě hodnoty věrohodnostní funkce modelu, počtu parametrů modelu a v případě Bayesova informačního kritéria BIC i podle počtu pozorování dat, ze kterých byl model zkonstruován.

Věrohodnostní funkce je funkce parametru Θ , která vyjadřuje hodnotu pravděpodobnosti, resp. hustoty pravděpodobnosti, při konkrétní realizaci náhodného výběru [5]. Má tvar

$$L(\Theta|\mathbf{x}_i) = f(x_1, x_2, \dots, x_n|\Theta) = f(x_1|\Theta)f(x_2|\Theta)\dots f(x_n|\Theta) = \prod_{i=1}^n f(\mathbf{x}_i|\Theta).$$

Z důvodu nižší výpočetní náročnosti při hledání parametru Θ s nejvyšší věrohodností se věrohodnostní funkce logaritmuje, aby funkce byla vyjádřena součtem místo součinem. Tato funkce se nazývá logaritmická věrohodnostní funkce a je tvaru

$$l(\Theta|\mathbf{x}_i) = \log L(\Theta|\mathbf{x}_i) = \sum_{i=1}^n f(\mathbf{x}_i|\Theta).$$

Logaritmická věrohodnostní funkce je integrální součástí AIC i BIC . AIC vychází z tzv. Kullback–Leiblerovy funkce, která zkoumá ztrátu informace při modelování reality, kritérium BIC je ve svém základu postaveno na Bayesově větě. [3] Oba dva modely k logaritmické věrohodnostní funkci přidávají penalizaci vyššího počtu parametrů modelu. Přidání parametru do modelu totiž prakticky vždy zvýší jeho věrohodnost $L(\Theta|x_i)$, nicméně s sebou přináší problém: Vyšší než nezbytné řády modelu mohou do jeho vlastní struktury započítávat i rezidua v pozorování či jinak nevýznamné jevy, které snižují jeho

schopnost postihnout realitu. Při předpovědi mají poté modely se zbytečně vysokými řády vyšší chybovost, ačkoliv lépe popisovaly minulá data. *AIC* a *BIC* se liší v tom, jakým způsobem penalizují vyšší počet parametrů:

$$AIC = -2 \log L(\Theta | \mathbf{x}_i) + 2m$$

$$BIC = -2 \log L(\Theta | \mathbf{x}_i) + m \ln(n)$$

kde m je počet parametrů modelu a n je počet pozorování. V zásadě se lze pokusit o znázornění metodiky infomačních kritérií takto: Mějme hypotetický skutečný model, který popisuje závislosti v empirických datech z hlediska počtu parametrů naprosto neredundantně. *BIC* nám poté poskytuje konzistentní odhad řádu modelu, který k řádu hypotetického skutečného modelu konverguje skoro jistě. Zároveň však má model odhadnutý kritériem *BIC* vyšší rozptyl, tj. není nutně eficientní. *AIC* naproti tomu klade důraz na eficienci modelu, tedy na nižší rozptyl, na úkor nižší konzistence. [2]

3.6.2 Odhad modelu

Existuje více přístupů k odhadu konkrétních hodnot parametrů daného modelu. K odhadu se většinou nejprve může přistoupit tzv. momentovým způsobem, což je výpočet parametrů na základě vztahů mezi autokorelacemi a jeho parametry. Tyto vztahy bývají povětšinou velmi dobře popsány (např. soustava Yule–Walkerových rovnic) pro celou škálu modelů z Boxovy–Jenkinsovy metodologie. Momentové odhady získané na základě výpočtu z autokorelací však jsou většinou považovány pouze za předběžné a slouží zejména jako počáteční hodnoty iteračních procesů prováděných za použití výpočetní techniky.

V jazyce R lze pro softwarový odhad využít balíček *forecast*, konkrétně funkci *auto.arima*, postavenou na inovačních stavových prostorových modelech a exponenciálním vyhlazování. [6]

3.6.3 Diagnostika modelu

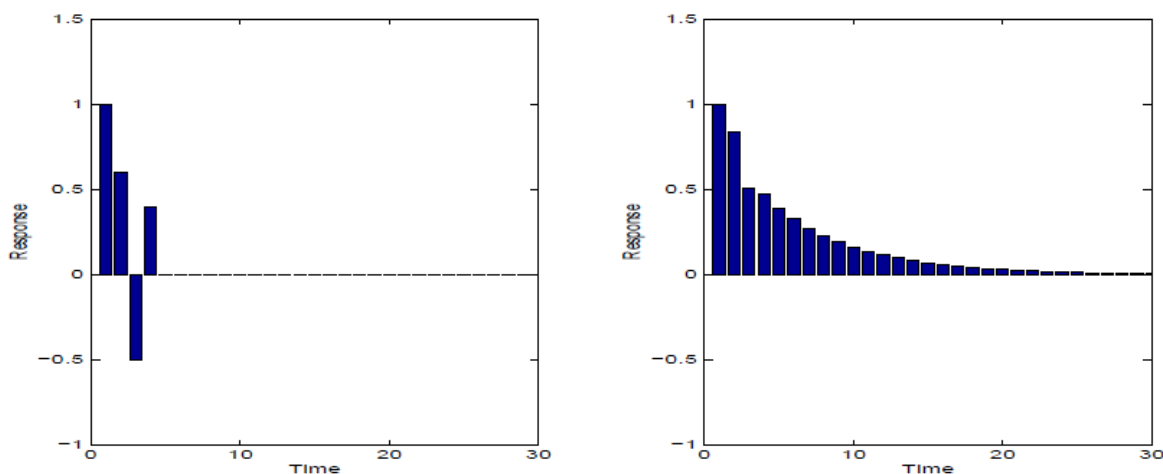
Diagnostika znamená ověření adekvátnosti zkonstruovaného modelu pomocí diagnostických nástrojů. Adekvátností modelu rozumíme kompatibilitu modelu s analyzovanými daty. Boxova–Jenkinsova metodologie oplývá pro ověření adekvátnosti svých modelů propracovanou diagnostikou, která se skládá z následujících přístupů k ověření adekvátnosti.

Kontrola stacionarity

Kontrola stacionarity znamená v tomto případě především ověření dvou aspektů modelu. Prvním je kontrola podmínky stacionarity u odhadnutého AR polynomu, tedy zda jeho kořeny leží v komplexní rovině vně jednotkového kruhu.

Druhým aspektem kontroly stacionarity, jímž je odhadnutý model diagnostikován, je analýza odezvy na impuls (impulse–response). Při této analýze sledujeme, jak se na probíhající odhadu současné hodnoty projeví námi zvolený impuls, tedy je-li modelem absorbován či v něm přetrvává. Velikost zkušebního impulsu bývá většinou standardizována jako jednonásobek či vícenásobek odhadnuté směrodatné odchylky bílého šumu.

Pro potřeby kontroly stacionarity analýzou odezvy na impuls lze zmínit skutečnost, že MA vyjádření (MA representation) jakéhokoliv procesu z Boxovy–Jenkinsovy metodologie je zároveň jeho funkcí odezvy na impuls (impulse–response funkcí). Při splnění určitých podmínek lze $AR(p)$ proces jakéhokoliv řádu p převést na proces MA , získat tím MA vyjádření, a tím i funkci odezvy na impuls. [8] Tato skutečnost je dobře patrná z následujícího obrázku. Pro funkci odezvy na impuls procesu MA vlevo je patrné, že vliv jednotkového impulsu v čase nula na pozdější, historické, hodnoty je roven hodnotě koeficientu pro daný časový okamžik.



Obrázek 3.2: Příklad funkce odezvy na impuls (impulse–response funkce). Vlevo impulse–response funkce procesu $MA(3)$ ($\theta_1 = 0,6; \theta_2 = -0,5; \theta_3 = 0,4$), vpravo pak funkce procesu $AR(2)$ ($\varphi_1 = 0,6; \varphi_2 = 0,2$) s jednotkovým impulsem v čase 0. [8]

Jelikož bílý šum je dle definice posloupnost nezávislých, stejně rozdělených náhodných veličin, znamená to, že má nulovou autokorelaci. Je tudíž zapotřebí prověřit i tento aspekt vypočteného modelu. K tomuto účelu se v praxi používají následující dva přístupy:

Bartlettova aproximace První možností ověření nekorelovanosti vypočteného bílého šumu je již zmíněná Bartlettova aproximace. Pro nulovou hypotézu nekorelovanosti $\rho_k = 0$ pro $k > 0$ zde za předpokladu asymptotické normality platí

$$|r_k(\hat{\epsilon}_t)| \sim N\left(0, 2\sqrt{\frac{1}{n}}\right) \quad \text{pro } k > 0.$$

Q–testy Q–testy souhrně testují významnost prvních K autokorelací vypočteného bílého šumu, přičemž za K se doporučuje volit hodnotu $K \approx \sqrt{n}$, kde n představuje počet pozorování v časové řadě.

Jako Q–statistiky se pak v praxi používají Boxova–Piercova a testově silnější, v této práci používaná Ljungova–Boxova statistika. Kritický obor Ljungovy–Boxovy statistiky, v němž je zamítnuta nulová hypotéza o nekorelovanosti v bílém šumu ve prospěch alternativy, že složky jsou korelovány, je na hladině významnosti α

$$Q = n(n+2) \sum_{k=1}^K \frac{1}{n-k} (r_k(\hat{\epsilon}_r))^2 \geq \chi_{1-\alpha}^2(K-p-q).$$

Problém nestacionarity

Jak již bylo zmíněno, Boxova–Jenkinsova metodologie pro svoje správné použití činí striktní požadavky na stacionaritu modelovaných časových řad. Nestacionarita časových řad působí modelům Boxovy–Jenkinsovy metodologie různé problémy, které znemožňují přesnou předpověď. Odezvy na impuls v nestacionárním prostředí na rozdíl od stacionárního nejsou absorbovány a mohou si zachovat nekonečné trvání. Použití nestacionárních dat dále může vést ke zdánlivé regresi – nalezení klamných nevýznamných regresních vztahů a k narušení pravděpodobnostních rozdělení souvisejících veličin.

Rozlišujeme následující dva typy nestacionarity:

Deterministické nestacionarita Zde je nestacionarita způsobena deterministickým trendem. Deterministický trend, který je také označován jako drift, je možné po jeho identifikaci jednoduše eliminovat a řadu tak stacionarizovat.

Stochastická nestacionarita V tomto případě je nestacionarita způsobena postupnou kumulací bílého šumu. To znamená, že na zkoumanou časovou řadu je superponován

určitý stochastický proces, který způsobuje její nestacionaritu, např. náhodná procházka s driftem, viz níže. Problém stochastické nestacionarity se řeší diferencováním původní časové řady a zkoumáním takto vzniklé řady prvních diferencí. Například u zmíněné náhodné procházky s driftem

$$y_t = \alpha + y_{t-1} + \epsilon_t$$

je možné časovou řadu, v níž je tento destacionarizující prvek obsažen, jednoduše stacionarizovat přechodem k řadě prvních diferencí. První diference rovněž jasně ukazuje, co se přičte v každém časovém kroku k časové řadě oproti předchozí hodnotě. Vidíme tu, že u náhodné procházky s driftem máme řadu jednak s deterministickým trendem α , který představuje jeho lineární směrnici, a zároveň s trendem, který tvoří prosté nasčítávání bílého šumu.

$$\Delta y_t = \alpha + \epsilon_t$$

Přechodem k řadě prvních diferencí tedy získáváme získáme bílý šum posunutý do úrovně α , což je triviálně stacionární řada.

Testy na jednotkový kořen

Charakteristický polynom výrazu z předchozího bodu popisujícího příklad stochastické nestacionarity, tedy náhodné procházky s driftem, jednoznačně vykazuje jednotkový kořen – autoregresní operátor má v tomto případě právě jeden kořen rovný jedné. Přítomnost jednotkového kořenu v charakteristické rovnici odhadnutého modelu popisujícího časovou řadu tedy přímo značí stochastickou nestacionaritu této časové řady. Určení přítomnosti jednotkového kořene je ovšem netriviální, a nesmí docházet k záměně nestacionárních modelů s pravým jednotkovým kořenem se stacionárními modely s téměř jednotkovým kořenem. Za tímto účelem se k prověřování přítomnosti jednotkového kořene používají následující testy:

Dickeyův–Fullerův test Dickeyův–Fullerův, zkráceně *DF*–test, je zavedený a rozšířený přístup k testování na jednotkový kořen. V praxi se používá ve třech variantách označovaných jako τ –testy.

1. Jednostranný test náhodné procházky proti stacionárnímu *AR*(1) procesu, též zvaný τ –test, který testuje

$$H_0 : y_t = y_{t-1} + \epsilon_t \quad H_1 : y_t = \varphi_1 y_{t-1} + \epsilon_t \quad \text{pro } \varphi_1 < 1.$$

2. Jednostranný test náhodné procházky proti stacionárnímu $AR(1)$ procesu s nenulovou úrovní τ_μ -test, který testuje

$$H_0 : y_t = y_{t-1} + \epsilon_t \quad H_1 : y_t = \alpha + \varphi_1 y_{t-1} + \epsilon_t \quad \text{pro } \varphi_1 < 1.$$

3. Jednostranný test náhodné procházky proti stacionárnímu $AR(1)$ procesu s lineárním trendem τ_τ -test, který testuje

$$H_0 : y_t = y_{t-1} + \epsilon_t \quad H_1 : y_t = \alpha + \beta t + \varphi_1 y_{t-1} + \epsilon_t \quad \text{pro } \varphi_1 < 1.$$

Agregovaně lze všechny tři varianty DF -testu zapsat jako

$$H_0 : \Delta y_t = \Psi y_{t-1} + \epsilon_t \quad \text{pro } \Psi = 0, \tag{3.2}$$

kde alternativní hypotéza závisí na hodnotě parametrů α a β , tj.

$$H_1 : \Delta y_t = \alpha + \beta t + \Psi y_{t-1} + \epsilon_t \quad \text{pro } \Psi < 0.$$

V tomto agregovaném zápisu platí $\Psi = \varphi_1 - 1$, $\alpha = \beta = 0$ pro τ -test a $\beta = 0$ pro τ_μ -test.

Testová statistika pro všechny tři varianty DF -testu je poměr

$$DF = \frac{\hat{\Psi}}{\hat{\sigma}(\hat{\Psi})}$$

testující významnost regresního parametru Ψ v modelu 3.2. Samotný DF -test přitom naráží na skutečnost, že testová statistika DF nemá klasické t -rozdělení jako v případě běžného t -poměru. Důsledkem tohoto faktu je, že pro zamítnutí nulové hypotézy $\Psi = 0$ potřebujeme významnější hodnotu t -poměru. Kritické hodnoty jsou nicméně pro tento případ již vyzkoumány a stanoveny. Při rozhodování o platnosti hypotéz je však zapotřebí brát tento fakt na zřetel. Na tuto skutečnost je poukazováno mj. i při zápisu t -testu, kde se k jeho symbolu připisuje hvězdička, tedy

$$DF \leq t_{1-\alpha}^*(n),$$

kde $t_{1-\alpha}^*(n)$ je kvantil modifikovaného t -rozdělení o n stupních volnosti.

Rozšířený Dickeyův–Fullerův test Nerozšířený DF -test lze použít pouze tehdy, je-li reziduální složka bílý šum – jinými slovy, proces závisle proměnných Δy_t nesmí obsahovat autokorelovanost. Nedodržení tohoto předpokladu s sebou nese následek zvýšení chyby prvního druhu pro uvažovaný DF -test, tj. pravděpodobnost zamítnutí platné H_0 je pak větší než α . Rozšířený DF -test (augmented DF -test, dále jen ADF -test) však pouze nahrazuje původní nulovou hypotézu novou nulovou hypotézou ve tvaru

$$H_0 : \Delta y_t = \Psi y_{t-1} + \sum_{t=1}^p \gamma_t \Delta y_{t-1} + \epsilon_t \quad \text{pro} \quad \Psi = 0,$$

kde γ_t jsou parametry přidaných autoregresních členů Δy_{t-1} , přičemž úkolem těchto přidaných členů a jejich parametrů je absorbovat dynamickou strukturu obsaženou v závisle proměnné Δy_t . Alternativa je

$$H_1 : \Delta y_t = \alpha + \beta t + \Psi y_{t-1} + \sum_{t=1}^p \gamma_t \Delta y_{t-1} + \epsilon_t \quad \text{pro} \quad \Psi < 0.$$

Všechny ostatní aspekty zůstávají stejné s původním DF -testem. Stanovení řádu p přidaných autoregresních členů se doporučuje výpočetní technikou na základě informačních kritérií.

KPSS–test $KPSS$ -test (Kwaitkovski, Phillips, Schmidt, Schin) reaguje na skutečnost, že DF -test a ADF -test někdy nemají dostatečnou rozlišovací schopnost, aby mohly korektně rozhodnout o stacionaritě časové řady. Nelze-li například zamítnout nulovou hypotézu nulového kořene, může to buď znamenat, že platí nestacionarita, ale stejně tak to může zároveň znamenat, že nemáme dostatek informací pro vyloučení nestacionarity. Z tohoto důvodu vzniknul $KPSS$ -test, který je v podstatě modifikací ADF -testu s tím rozdílem, že hypotézy jsou u něj přesně opačné. Testuje tedy nulovou hypotézu stacionarity proti alternativní hypotéze nestacionarity.

Pro zajištění korektnosti a směrodatnosti závěrů učiněných na základě metodiky DF -testu se doporučuje vždy provádět dva testy, ADF -test a $KPSS$ -test, a to vždy současně. Za směrodatný výsledek lze pak označit pouze případy, kdy v situaci, ve které u jednoho testu H_0 nelze zamítnout, naopak u druhého testu H_0 zamítáme ve prospěch H_1 . Výpočetní software však tento postup často neumožňuje, viz sekce 4.2.1.

3.7 Proces ARIMA

Integrovaný smíšený proces *ARIMA* je aplikací Boxovy–Jenkinsovy metodologie na časové řady se stochastickou nestacionaritou. Oproti standardnímu smíšenému procesu *ARMA*(p, q) v sobě navíc zahrnuje diferenční řád d . Ten vyjadřuje počet diferencování modelované časové řady.

Integrovaný smíšený proces řádů p, d, q , značený jako *ARIMA*(p, d, q), se zapisuje

$$\varphi(B)w_t = \alpha + \theta(B)\epsilon_t, \quad (3.3)$$

kde

$$w_t = \Delta^d y_t$$

představuje d -tou diferenci modelované časové řady y_t a (3.3) je již stacionární model *ARMA*(p, q). Pro vyjádření modelu *ARIMA* lze také zvolit tzv. souhrnný zápis, tedy

$$\varphi(B)\Delta^d y_t = \alpha + \theta(B)\epsilon_t,$$

který obsahuje tzv. zobecněný autoregresní operátor $\varphi(B)\Delta^d$ na levé straně. Zobecněný autoregresní operátor má stejně jako autoregresní operátor u čistě stacionárních modelů p kořenů ležících vně jednotkového kruhu v komplexní rovině. Kromě nich má nicméně zobecněný autoregresní operátor navíc ještě d -násobný jednotkový kořen.

Řád diferencování d v praxi obvykle nepřekračuje dvojku.

Odhad a následné stanovení konkrétního řádu diferencování d pro časovou řadu se provádí následujícími způsoby:

- testy na jednotkový kořen,
- subjektivní prohlídka průběhů řad $y_1, \Delta y_1, \Delta^2 y_1$ a jejich odhadnutých korelogramů a parciálních korelogramů,
- porovnávání výběrových směrodatných odchylek řad $y_1, \Delta y_1, \Delta^2 y_1$,
- aplikace informačních kritérií, modifikovaných pro modely ARIMA.

3.8 Proces SARIMA

Modely *SARIMA*, neboli sezónní modely Boxovy–Jenkinsovy metodologie, jsou aplikací této metodologie určenou speciálně pro stochastické modelování sezónnosti. Jedná se ve své podstatě o modely *ARIMA* s autoregresním operátorem či operátorem klouzavých průměrů, ve kterém je skok časového posunu větší než jedna. Například pro řád sezónnosti $s = 12$ u měsíčních pozorování je příslušný model *SARIMA* popsán jako

$$\Phi(B^{12})\Delta_{12}^D y_t = \Theta(B^{12})\mu_t.$$

Sezónní autoregresní operátor řádu P je pak popsán jako

$$\Phi(B^{12}) = 1 - \Phi_1 B^{12} - \Phi_2 B^{24} - \dots - \Phi_P B^{12P}$$

a pro sezónní operátor klouzavých součtů řádu Q platí

$$\Theta(B^{12}) = 1 - \Theta_1 B^{12} + \Theta_2 B^{24} + \dots + \Theta_Q B^{12Q}.$$

Sezónní modely Boxovy–Jenkinsovy metodologie pak mají zvláštní sezónní diferenční operátor

$$\Delta_{12} = 1 - B^{12}.$$

Sezónní modely disponují navíc ještě jednou zvláštností, a sice, že předpokládají korelaci náhodných složek neboli inovaci v rámci sousedních měsíců. Tato korelace je však předpokládána mezi sousedními měsíci v rámci původní nesezónní řady. Je-li tedy například sezónnost v návaznosti na výše uvedený příklad dvanáctiměsíční, tj. pouze mezi lednovými hodnotami, pak tato korelace v inovacích není předpokládána napříč tímto časovým skokem jednoho roku, nýbrž mezi měsíci leden – únor. Rada takto korelovaných inovací μ_t je modelována vlastním nesezónním modelem *ARIMA* jako

$$\varphi(B)\Delta^d \mu_t = \theta(B)\epsilon_t.$$

Rozvedený zápis úplného modelu *SARIMA*, neboli multiplikativního sezónního procesu řádu $(p, d, q) \times (P, D, Q)_s$ se sezónností s má tvar

$$\varphi(B)\Phi(B^s)\Delta^d \Delta_s^D y_t = \theta(B)\Theta(B^s)\epsilon_t.$$

3.9 Modely GARCH

Modely Boxovy–Jenkinsovy metodologie, jejichž prostřednictvím lze uskutečňovat predikce časových řad, jsou omezeny požadavkem na slabou stacionaritu dat. Toto omezení s sebou zároveň nese skutečnost, že nejsou schopny modelovat časové řady s proměnnou střední hodnotou nebo rozptylem.

Nelineární modely mohou narozdíl od modelů čistě Boxovy–Jenkinsovy metodologie modelovat řady s podmíněnou (na čase závislou) střední hodnotou a podmíněným rozptylem.

Pro potřeby matematického zápisu filosofie nelineárních modelů se pracuje s proměnnou Ω_{t-1} . Ta představuje veškerou informaci o časové řadě známou do času $t - 1$. Jedná se tím pádem o poměrně komplexní množinu složenou ze všech minulých hodnot časové řady y_{t-1}, y_{t-2}, \dots , všech předpovědních chyb či inovací e_{t-1}, e_{t-2}, \dots a vhodných funkcí těchto hodnot. Na čase závislá střední hodnota, stejně jako časové závislý rozptyl, se vyjadřuje nelineárními funkcemi $g(\Omega_{t-1})$ a $h(\Omega_{t-1})$ informace Ω_{t-1} :

$$\mu_t = E(y_t | \Omega_{t-1}) = g(\Omega_{t-1}),$$

$$\sigma_t^2 = h_t = \text{var}(y_t | \Omega_{t-1}) = h(\Omega_{t-1}),$$

kde $h(\cdot) > 0$.

Hodnota časové řady y_t je pak nelineárními modely obecně vyjádřena jako

$$y_t = \mu_t + \sigma_t \epsilon_t = g(\Omega_{t-1}) + \sqrt{h(\Omega_{t-1})} \epsilon_t,$$

kde

$$\epsilon_t \sim N(0, 1). \tag{3.4}$$

Za zmínku stojí rovněž fakt, že náhodné veličiny

$$e_t = \sigma_t \epsilon_t \tag{3.5}$$

jsou nekorelované, ale nejsou nezávislé.

3.9.1 ARCH model

Modely *ARCH* (autoregressive conditional heteroscedasticity, autoregresní podmíněná heteroskedasticita) a *GARCH* (generalized *ARCH*, zobecněný *ARCH*) představují systematický přístup k modelování proměnného rozptylu. Jejich metodologie je postavena na předpokladu, že proměnná volatilita finančních časových řad je kvadratickou funkcí minulých předpovědních chyb či inovací e_t . Tato funkční závislost vychází z předpokladu, že po větších volatilitách lze ve finančních časových řadách zpravidla očekávat volatilitu menší a naopak. Tento jev znamená, že jsou volatilita pozitivně autokorelované a lze je tedy modelovat autoregresním modelem. Na základě rovnic (3.5) a (3.4) lze odvodit, že

$$\sigma_t^2 = \text{var}(e_t | \Omega_{t-1}) = E(e_t^2 | \Omega_{t-1}),$$

což znamená, že rozptyl řady v daném časovém okamžiku je přibližně roven čtverci inovace ve stejném okamžiku. V návaznosti na tuto skutečnost lze tedy rozptyl řady v čase t vyjádřit jako lineární kombinaci čtverců inovací neboli předpovědních chyb a předpis modelu *ARCH* má poté tvar

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_m e_{t-m}^2,$$

přičemž koeficienty α musí splňovat požadavek tzv. regularity. Všeobecně se požaduje pro kladné hodnoty rozptylu σ_t^2 požaduje, aby

$$\alpha_0 > 0, \alpha_1 \geq 0, \dots, \alpha_m \geq 0.$$

Celý model *ARCH* řádu m , *ARCH*(m), je definován jako

$$y_t = \mu_t + e_t, \quad e_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_m e_{t-m}^2.$$

Metodiku a výhodu proměnné volatilita a jejího modelování lze přitom přenést do Boxovy–Jenkinsovy metodologie. Například pro *AR*(p) – *ARCH*(m) proces platí, že

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + e_t, \quad e_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_m e_{t-m}^2.$$

Řád m se identifikuje jako bod useknutí odhadnutého parciálního korelogramu analogicky jako je tomu u modelu *AR*(p). Samotný odhad modelu je otázkou maximalizace logaritmičkových věrohodnostních funkcí, a tedy softwarovou záležitostí, přičemž lze použít například algoritmus *BHHH*.

3.9.2 Testy na podmíněnou heteroskedasticitu

Testování časové řady na *ARCH*-efekt neboli na přítomnost podmíněné heteroskedasticity je určitým způsobem specifické, neboť neprobíhá na časové řadě, kde se přítomnost heteroskedasticity snažíme detekovat. Jedná se totiž o test sériové nezávislosti prováděný na reziduích určitého modelu popisujícího původně testovanou řadu, přičemž se obvykle jedná o model z Boxovy–Jenkinsovy metodologie. Předpokládáme, že lineární sériová závislost v původní testované řadě je eliminována právě modelem, jehož rezidua testujeme. Pokud se tedy v reziduích modelu popisujícího původní časovou řadu vyskytuje sériová závislost, je v původní časové řadě přítomna podmíněná heteroskedasticita. V praxi se k tomuto účelu používají dva testy: [2, 16, 19]

McLeod–Li test

McLeod–Li test (dále jen *ML*-test) vznikl na základě Ljungova–Boxova testu a jedná se ve své podstatě o *Q*-test čtverců reziduí modelu časové řady. Kritický obor Ljungovy–Boxovy statistiky *ML*-testu, v němž je zamítnuta nulová hypotéza o nepřítomnosti podmíněné heteroskedasticity v původní časové řadě ve prospěch alternativy, že časová řada je podmíněně heteroskedastická, je na hladině významnosti α

$$Q = n(n+2) \sum_{k=1}^K \frac{1}{n-k} (r_k(\hat{\epsilon}_r^2))^2 \geq \chi_{1-\alpha}^2(K-p-q).$$

Lagrange Multiplier test

Lagrange Multiplier test (dále jen *LM*-test) řádu m je prováděn na základě regrese

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_m e_{t-m}^2 + \epsilon_t, \quad (3.6)$$

ve které se snažíme vysvětlit hodnotu čtverce současného rezidua lineární kombinací minulých hodnot čtverců reziduí až do řádu m , zcela v duchu modelu *ARCH*(m). Nulová hypotéza je, stejně jako v případě *ML*-testu, že v původní časové řadě podmíněná heteroskedasticita přítomna není a řada je tedy homoskedastická.

Testová statistika *LM*-testu řádu m a její kritický obor, ve kterém zamítáme nulovou hypotézu o nepřítomnosti podmíněné heteroskedasticity v původní časové řadě ve prospěch alternativy, že časová řada je podmíněně heteroskedastická, mají na hladině významnosti α tvar

$$LM = T \cdot R^2 \geq \chi_{1-\alpha}^2(m),$$

kde R^2 je koeficient determinace v modelu 3.6 a T délka časové řady.

3.9.3 GARCH modely

Model *ARCH* vykazuje při své aplikaci určité nedostatky. Mezi ně lze v kontextu praktického zacházení s modelem zařadit zejména vysoký řád m , který model často vyžaduje pro adekvátní popis vývoje volatility. Z toho vzniká problém velkého množství zpracovávaných parametrů společně s nedodržením požadavku nezápornosti těchto parametrů. Navíc není zohledněna asymetrie odchylek, tedy stav, kdy kladné a záporné odchylky e_t mají odlišný vliv na volatilitu.

Tyto nedostatky se staly motivací pro vytvoření modelu *GARCH*. Model *GARCH* řádů (m, s) je definován jako

$$y_t = \mu_t + e_t, \quad e_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i e_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2,$$

přičemž parametry modelu splňují

$$\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, \quad \sum_{i=1}^{\max\{m,s\}} (\alpha_i + \beta_i) < 1.$$

3.10 Posuzování přesnosti předpovědi

K posouzení přesnosti předpovědi určitého modelu se používají ukazatele, jejichž vstupními parametry jsou odhadnuté hodnoty předpovídané veličiny a skutečné, empirické hodnoty. Jednotlivé ukazatele pak mají každý svůj způsob, jakým vyjadřují shodu předpovědi s realitou.

3.10.1 Mean Error (ME)

Mean Error je relativně triviální ukazatel přesnosti předpovědi. Je definován předpisem

$$ME = \frac{1}{n} \sum_{h=1}^n P_h - \hat{P}_h,$$

kde P_h je skutečná cena pro danou hodinu, \hat{P}_h je cena predikovaná modelem a n je počet hodin, pro které se predikuje. Ukazatel ME prakticky nedokáže vyjádřit skutečnou velikost odchylky od předpovědi, neboť se postupně nasčítávané a poté průměrované kladné a záporné chyby vzájemně anulují. Z tohoto důvodu je jeho použitelnost diskutabilní.

3.10.2 Mean Percentage Error (MPE)

Ukazatel MPE je zatížený stejným nedostatkem jako ME a neposkytuje zaručeně relevantní informaci o přesnosti předpovědi. Definován je předpisem

$$MPE = \frac{1}{n} \sum_{h=1}^n \frac{P_h - \hat{P}_h}{P_h}.$$

3.10.3 Mean Absolute Error (MAE)

Mean Absolute Error koriguje nedostatek ME průměrováním absolutních hodnot chyb. Je definován předpisem

$$MAE = \frac{1}{n} \sum_{h=1}^n |P_h - \hat{P}_h|,$$

a tedy představuje průměrnou chybu modelu získanou z absolutních hodnot chyb všech jeho předpovědí.

3.10.4 Mean Absolute Percentage Error (MAPE)

Ukazatel

$$MAPE = \frac{1}{n} \sum_{h=1}^n \frac{|P_h - \hat{P}_h|}{P_h}$$

funguje velice podobně jako MAE , namísto absolutních hodnot chyb však průměruje jejich procentní vyjádření vůči skutečným hodnotám cen.

3.10.5 Root Mean Square Error (RMSE)

Ukazatel

$$RMSE = \sqrt{\frac{1}{n} \sum_{h=1}^n (P_h - \hat{P}_h)^2}$$

je v určitém ohledu podobný ukazateli MAE . Zde se však jedná o odmocněný průměr čtverců chyb, a tím pádem určitou obdobu směrodatné odchylky.

Kapitola 4

Realizace predikce, numerické výsledky

4.1 Data

Zdrojový soubor dat použitý k aplikaci metodologie zahrnuje marginální ceny denního trhu OTE za rok 2017. [13] Jedná se o 8760 hodinových hodnot marginální ceny elektrické energie MCP v EUR pro každou z 24 hodin dne, 1. lednem 2017 počínaje a 31. prosincem 2017 konče.

4.1.1 Změny zimního a letního času, korekce datasetu

Ke změně ze střeoevropského času (SEČ) na střeoevropský letní čas (SELČ) v roce 2017 došlo v neděli 26. 3. tím, že se ve 2:00 posunuly hodiny na 3:00. Tato skutečnost se v původním datasetu projevuje tak, že 26. 3. je zde MCP uvedena pouze pro 23 hodinových hodnot. Analogicky došlo v neděli 29. 10. 2017 ke změně ze střeoevropského letního času (SELČ) na střeoevropský čas (SEČ) tak, že se ve 3:00 posunuly hodiny na 2:00. V původním datasetu tedy neděli 29. 10. 2017 náleželo 25 hodnot MCP, namísto obvyklých 24 hodinových hodnot.

Tato skutečnost byla korigována odpovídajícím zprůměrováním hodnot MCP tak, aby každému dni v souboru dat náleželo 24 hodinových hodnot marginálních cen denního trhu. V neděli 26. 3. byla hodnota MCP pro v ten den neexistující třetí hodinu vytvořena jako aritmetický průměr marginální ceny druhé a čtvrté hodiny. U neděle 29. 10. 2017 byla ze dvou hodnot marginálních cen pro třetí hodinu vytvořena jedna hodnota aritmetickým zprůměrováním.

4.1.2 Sezónnost

Ceny elektrické energie na denním trhu vykazují výraznou sezónnost z hlediska denního času. Ceny elektrické energie ve „špičce“ a mimo ni se výrazně liší. Touto skutečností a jejím dopadem na postup při samotném predikování ceny se zabýval Weron [21], přičemž jednoznačně doporučuje rozdělení souboru všech 8760 po sobě následujících hodin roku na 24 datasetů po 365 hodnotách, které zahrnují hodnoty jednotlivých hodin. Například první dataset tak obsahuje pouze hodnoty MCP pro první hodiny (časové rozmezí 0:00 až 1:00) pro všech 365 dnů v roce. Weron ve své publikaci dále odůvodňuje správnost tohoto postupu odkazem na vědecké práce, které ukazují, že přesnost předpovědi tohoto postupu je vždy výrazně vyšší než při aplikaci statistické autoregresní metodologie na původní nerozdělený dataset.

4.2 Realizace modelování a předpovědi

Za programové prostředí zaštiťující realizaci analýzy a výpočtů byl zvolen programovací jazyk R. V jeho prostředí byla realizována identifikace a konstrukce modelu i samotná předpověď.

4.2.1 Identifikace a konstrukce adekvátního modelu

Výběr optimálního modelu, tedy jeho identifikace i odhad, byly v rámci této práce prováděny pomocí funkce *auto.arima* z knihovny *forecast*. Tato funkce, jejímž vstupním parametrem je časová řada, identifikuje pomocí zvoleného informačního kritéria optimální model vstupní časové řady z Boxovy–Jenkinsovy metodologie. Posuzování možných modelů probíhá pomocí informačního kritéria. Vybrán je ten model, který má oproti ostatním modelům nejlepší, tj. nejnižší, hodnotu informačního kritéria. Hledání probíhá v rámci výchozího nastavení v množině modelů *ARIMA* od řádů p i q rovných jedné až do řádů p i q rovných pěti. Je-li *OSCB* testem [6] zjištěna sezónnost, přesune funkce automaticky hledání do množiny modelů *SARIMA*, přičemž původní rozmezí uvažovaných řádů p i q od 1 do 5 je zachováno, ve stejném rozmezí se odhadují i řády P i Q .

Funkce *auto.arima* při svém hledání detekuje a zohledňuje deterministickou i stochastickou nestacionaritu. V případě deterministické nestacionarity dat předloží funkce *auto.arima* jako svůj výsledek model s driftem, tedy deterministickým lineárním trendem. Stochastická nestacionarita koresponduje s přítomností jednotkového či téměř jednotkového kořenu příslušného autoregresního operátoru modelujícího zpracovávanou časovou řadu. Stochastická nestacionarita se tím pádem ověřuje testy na jednotkový kořen *ADF* či *KPSS*, přičemž funkce *auto.arima* zde nabízí možnost volby. Je-li v časové

řadě s pomocí zvoleného testu detekována nestacionarita, funkce automaticky řadu diferencuje, dokud není stacionární. Poté pokračuje s identifikací a odhadem modelu obvyklým způsobem na stacionarizovaných datech a vzniká model *ARIMA* s nenulovým řádem i . U nestacionarity z hlediska odhadnutého *AR* polynomu odhaduje funkce *auto.arima* modely vždy tak, že jejich kořeny leží v komplexní rovině vně jednotkového kruhu. [6]

Jen poznamenejme, že co se týče testování stochastické nestacionarity testy *ADF* i *KPSS* najednou, kdy je řada prohlášena za nestacionární resp. stacionární na základě shodného vyjádření obou testů (viz sekce 3.6.3), tuto funkcionalitu *auto.arima* nenabízí. Test na jednotkový kořen se tak stává jedním z parametrů důležitých pro optimální postup, viz Kapitola 4.4. Funkce *auto.arima* tedy v rámci své vnitřní komplexní metodologie identifikuje pro danou časovou řadu vhodný model v rámci celé výše popsané Boxovy–Jenkinsovy metodologie, tedy *ARIMA* či *SARIMA*, a zároveň provede jeho číselný odhad.

4.2.2 Predikce na základě vybraného modelu

Pro bodový odhad ceny elektřiny na základě vytvořeného modelu byla použita funkce *forecast*, rovněž z knihovny *forecast*. Vstupním parametrem funkce *forecast* je datový objekt modelu vytvořeného funkcí *auto.arima*. Jelikož jsou *auto.arima* a *forecast* součástí jedné knihovny, není zde vůbec žádný problém s kompatibilitou a je tak možné realizovat predikci při všech variantách vytvořeného modelu. Parametry funkce *forecast* kromě datového objektu modelu zahrnují i požadované nastavení predikování. Pro praktické použití bez hlubšího zasahování do vnitřní metodologie funkce má význam pouze nastavení intervalů spolehlivosti předpovědi. Ty jsou ponechány na výchozích hodnotách 80% a 95%. Samotná komplexní metodologie funkce je postavená na exponenciálním vyhlazování. [6]

4.2.3 Realizace modelů *ARIMA*–*GARCH*

Jelikož knihovna *forecast* nenabízí programový aparát k odhadu a realizaci kombinovaných modelů *ARIMA* – *GARCH* [6], byla k tomuto účelu využita knihovna *rugarch*. [4] Funkce *ugarchspec* z knihovny *rugarch* slouží ke specifikaci modelu, který má být realizován, a umožňuje stanovení požadovaných řádů modelů *ARIMA* i *GARCH*. Prostřednictvím funkce *ugarchspec* je vytvořen datový objekt obsahující informace o požadavcích na realizaci modelu, který je pak předáván dalším funkcím knihovny *rugarch* jako vstupní parametr. Druhou funkcí z této knihovny využívanou v rámci práce je *ugarchfit*. Vstupním parametrem funkce *ugarchfit* je datový objekt vytvořený funkcí *ugarchspec* a časová řada, která má být modelována. Na základě těchto vstupů je funkcí *ugarchfit* naitován model *ARIMA* – *GARCH*. Konečně funkce *ugarchforecast* realizuje samotnou před-

pověď a její jediný potřebný parametr je právě model *ARIMA – GARCH* vytvořený funkcí *ugarchfit*, samozřejmě vedle počtu hodnot, které chceme odhadnutým modelem předpovídat.

4.3 Výzkum efektivity předpovědi pro první hodinu

Postup při vytváření modelů cen elektrické energie má více zásadních parametrů a má-li být v rámci možností optimální, je správný výběr těchto parametrů komplikovaný. Je tedy zapotřebí empirického posouzení správnosti předpovídání napříč těmito parametry, kterými jsou délka tzv. učicího období, druh testu na jednotkový kořen a použité informační kritérium. Délkou učicího období je zde myšlena délka časového rozmezí před začátkem samotné předpovědi, pro které se identifikuje model a odhadují jeho parametry. Délka učicího období může být počet hodnot jednoho z 24 datasetů MCP pro určitou hodinu nebo z hlediska celkové předpovědi pak počet dnů před zahájením předpovědi.

První takový výzkum byl uskutečněn na souboru dat prvních hodin a zahrnoval následující postup. Pro různou délku $l = 30, \dots, 200$ učicího období (tj. počet hodnot, z nichž je model nafitován) byla učiněna předpověď pro danou hodinu v následujících sedmi dnech, tj. pro dny $l + n + 1, l + n + 2, \dots, l + n + 7$ kde $n = 1, \dots, 365 - 7$. To znamená, že počínaje všemi hodnotami od 31. do 358. včetně byla uskutečněna předpověď pro sedm následujících hodnot, tedy pro 31. – 37. hodnotu, 32. – 38. hodnotu atd. Každá předpověď byla prováděna s délkami učicího období 30 až 200 dnů (hodinových hodnot), avšak s ohledem na spodní hranici datasetu. Například pro 31. den tedy byla sedmidenní předpověď provedena pouze jednou, a to s délkou učicího období 30, pro 32. hodinu už byly realizovány předpovědi dvě: jedna s délkou učicího období 30, jehož počátek ležel v 2. hodnotě a konec se nacházel v 31. hodnotě časové řady, druhá s délkou učicího období rovnou 31 hodnotám, jehož počátek ležel v 1. hodnotě časové řady a konec opět v 31. hodnotě. Analogicky se postupovalo u následujících hodnot prvních hodin.

Každá taková předpověď byla realizována jak pro model odhadnutý s pomocí kritéria *AIC*, tak i pro *BIC* model. Pro testování přítomnosti jednotkového kořenu v časové řadě učicího období byl použit *ADF*-test.

Takto uskutečněné šetření odhalilo fakta, na jejichž základě byla provedena rozhodnutí ohledně postupu pro další hodiny.

4.3.1 AIC přesnější než BIC

První skutečnost, která vyšla najevo z výzkumu přesnosti předpovědí, se týká volby informačního kritéria, na základě kterého automatizovaná funkce *auto.arima* posuzuje rela-

tivní vhodnost modelů přicházejících v úvahu. Pro obě informační kritéria v uskutečněném výzkumu souboru dat prvních hodin, tj. AIC i BIC , je délka učícího období s nejnižší průměrnou chybou týdenní předpovědi rovna 124 hodnotám časové řady.

Průměrná chyba $RMSE$ všech předpovědí s touto optimální délkou učícího období rovnou 124 hodinám je rovna

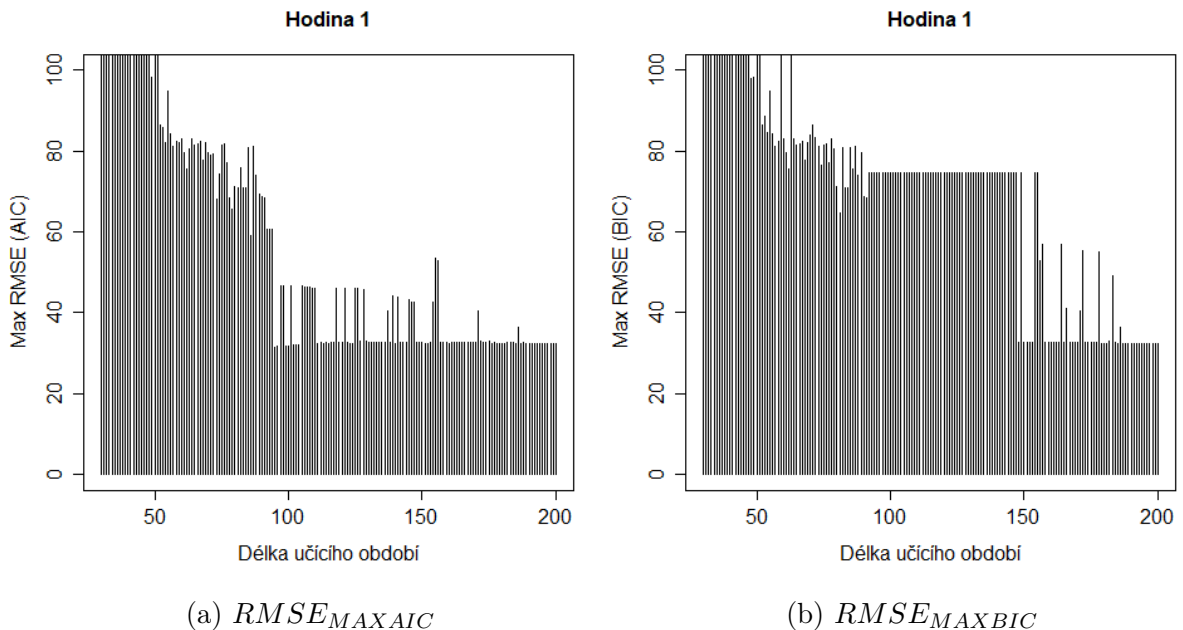
$$RMSE_{AIC124} = 8,09 \quad RMSE_{BIC124} = 8,25 \quad \Delta RMSE_{124} = 0,16.$$

Pro všech 40 698 uskutečněných předpovědí pak průměrná chyba vyšla

$$RMSE_{AIC} = 9,22 \quad RMSE_{BIC} = 9,32 \quad \Delta RMSE = 0,10.$$

V obou případech, tedy jak při optimálním nastavení učícího období, tak celkově, vede použití informačního kritéria AIC na modely s lepší průměrnou předpovědní schopností.

Modely odhadnuté kritériem AIC navíc nepodávaly lepší výkon pouze z hlediska průměrné chyby předpovědi. Z prohlídky grafů zobrazujících maximální zaznamenané chyby $RMSE$ v závislosti na délce učícího období je zjevné, že použití informačního kritéria AIC vedlo k výběru modelů s podstatně nižší možnou maximální chybou pro významnou část zkoumaných délek učícího období.



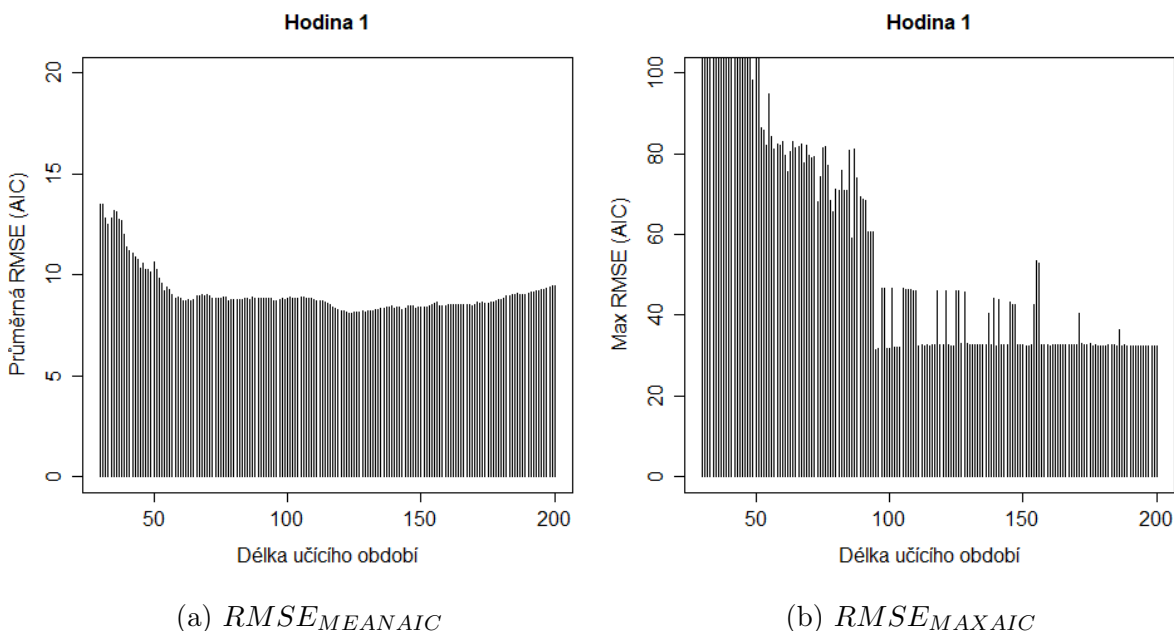
Obrázek 4.1: Maximální zjištěné chyby pro jednotlivé délky učícího období. Vlevo pro informační kritérium AIC , vpravo BIC .

$RMSE$ chyba kritéria AIC vyšla u 37 délek učení z celkového počtu 171 zkoumaných, tedy v necelých 22 % případech, nižší než poloviční $RMSE$ chyba kritéria BIC . Tento jev

byl prevalentní ve střední oblasti délek učícího období, tedy mezi délkami 95 až 149, a je z grafů na Obr. 4.1 dobře patrný. Na základě těchto skutečností bylo pro všechny další výpočty vybráno informační kritérium AIC .

4.3.2 Nevhodnost krátkých učících období

Při posuzování závislosti průměrné předpovědní chyby na délce učícího období vyšla najevo vyšší chybovost u modelů s nižšími délkami učícího období. Jmenovitě se jedná o oblast učících období kratší než 50 dnů (hodinových hodnot). V této oblasti lze pozorovat tendenci zvyšující se předpovědní chyby se zkracující se délkou učícího období. Pod hranicí 50 hodnot lze navíc dále pozorovat extrémní maximální chyby předpovědi, které se nad touto hranicí již nevyskytují, jak ilustruje Obr. 4.2.



Obrázek 4.2: Vlevo průměrná chyba v závislosti na zvolené délce učícího období, vpravo pak maximální zjištěná chyba v závislosti na délce učícího období. Grafy nemají stejné měřítko, i tak je však jednoznačné zhoršování předpovědí se zkracující se učící délkou pod hranicí 50 hodnot.

Na základě této skutečnosti byla v dalším postupu za nejkratší délku učícího období zvolena délka 50 dnů. K tomuto rozhodnutí přispěly ještě další dva důvody: Prvním je doporučení týkající se výpočtů autokorelace, tedy vyšetřování autokovariance a autokorelace časových řad. Tyto výpočty, které jsou základní součástí odhadu modelu, je doporučeno provádět se soubory dat alespoň s 50 hodnotami. [2] Toto doporučení se tedy v realizovaném výzkumu na časové řadě prvních hodin 365 dnů empiricky ukázalo

jako opodstatněné. Druhým důvodem je dále snížení již tak značné výpočetní náročnosti při realizaci výzkumu chybovosti pro 24 souborů dat.

4.4 Výzkum pro 24 souborů dat

Předchozí výzkum se zaměřoval zejména na informační kritéria a volbu délky učícího období v rámci fungování funkce *auto.arima*. Na jeho základě byla přijata rozhodnutí o zúžení zkoumaného souboru délek učícího období, přičemž už dále budou zkoumány pouze délky 50 až 200. Dále byla učiněna volba informačního kritéria *AIC* jako jediného posuzovaného kritéria.

Další realizovaný výzkum, jehož popis následuje, staví na získaných poznacích a byl proveden již pro všech 24 souborů dat s jednotlivými hodinovými hodnotami, tedy i znovu pro první hodinu, neboť má za cíl rozkrýt další souvislosti, které prvním testem objasněny nebyly. Zaměřením dále zůstává zjištění optimální délky učícího období pro všechny hodiny dne a výzkum přesnosti předpovědi. Optimální délka učícího období je zjišťována jak pro *ADF*, tak *KPSS* test, a tím pádem je testována přesnost nejen v závislosti na délce učení, ale i na použitém testu na jednotkový kořen.

Postup aplikovaný na všech 24 hodinových souborů dat byl obdobný jako v případě výzkumu popsaného v předchozí kapitole, tedy počínaje všemi hodnotami od 51. do 359. (včetně) byla uskutečněna předpověď sedmi budoucích hodnot a posouzena její přesnost. Každá předpověď byla prováděna s délkami učícího období 50 až 200, avšak i zde s ohledem na spodní hranici datasetu. Souhrn výsledků je uveden v Tab. 4.1.

Optimální délky učení vyjadřují délky učícího období, se kterými byly dosaženo předpovědí s nejmenší průměrnou chybou. Konkrétní hodnoty těchto průměrných chyb jsou uvedeny ve sloupci „průměrná *RMSE* při optimálním učení.“ V tabulce uvedené hodnoty dále ilustruje Obr. 4.3. Průřezové statistiky obou testů jsou:

$$RMSE_{ADF} : \quad max = 15,42 \quad min = 7,11 \quad mean = 11,58 \quad sd = 2,84$$

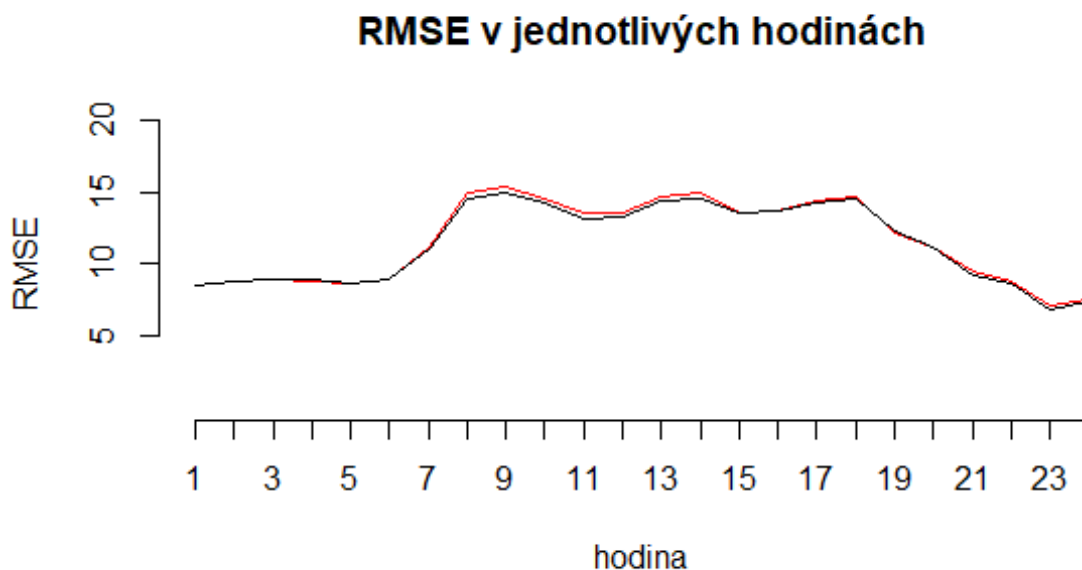
$$RMSE_{KPSS} : \quad max = 15,01 \quad min = 6,79 \quad mean = 11,44 \quad sd = 2,76$$

4.5 Algoritmus předpovědi pro 24 hodin

Algoritmus předpovědi střední hodnoty marginální ceny staví na výsledcích předešlého výzkumu. Ty jsou zde aplikovány ve snaze získat nejlepší možné nastavení parametrů hledání predikčního modelu. Těmi byly délka učícího období, použité informační kritérium

Tabulka 4.1: Nejlepší délky učicích období a příslušné průměrné chyby $RMSE$. Podtrženě jsou uvedeny učicí délky příslušející testu na jednotkový kořen s nižší průměrnou chybou, které spolu s příslušným testem používá predikční algoritmus.

Hodina	Optimální délka učení ADF	Optimální délka učení $KPSS$	Průměrná $RMSE$ při optimálním učení ADF	Průměrná $RMSE$ při optimálním učení $KPSS$
1	<u>124</u>	56	8,56	8,58
2	<u>123</u>	58	8,74	8,82
3	<u>124</u>	61	8,88	8,97
4	<u>125</u>	53	8,80	8,92
5	<u>122</u>	122	8,63	8,71
6	120	<u>77</u>	8,89	8,87
7	120	<u>69</u>	11,12	11,03
8	121	<u>84</u>	14,96	14,57
9	120	<u>81</u>	15,42	15,01
10	92	<u>67</u>	14,48	14,26
11	88	<u>66</u>	13,61	13,14
12	87	<u>66</u>	13,60	13,21
13	191	<u>53</u>	14,74	14,46
14	67	<u>59</u>	14,91	14,47
15	<u>69</u>	56	13,58	13,60
16	<u>199</u>	53	13,65	13,69
17	74	<u>52</u>	14,35	14,25
18	85	<u>54</u>	14,67	14,52
19	<u>85</u>	59	12,13	12,25
20	90	<u>59</u>	11,14	11,11
21	81	<u>74</u>	9,50	9,28
22	91	<u>60</u>	8,85	8,66
23	124	<u>54</u>	7,11	6,79
24	124	<u>61</u>	7,57	7,34



Obrázek 4.3: Nejmenší průměrná $RMSE$ v jednotlivých hodinách při optimálních délkách učícího období a použití informačního kritéria AIC . ADF -test červeně, resp. $KPSS$ -test černě.

a test na jednotkový kořen. Nastavení predikčního algoritmu tedy přímo vychází z Tab. 4.1 nejlepších učících intervalů.

Samostatně pro každou hodinu je tedy nejprve zvolen test na jednotkový kořen. Tato volba je učiněna na základě průměrných hodnot $RMSE$ a je zvolen ten test, pro který je průměrná chyba při jemu náležející nejlepší učící vzdálenosti nižší. V závislosti na tom, jaký je vybrán test na jednotkový kořen, je poté použita i jeho odpovídající optimální délka učícího období. Informační kritérium je vždy AIC .

Pro každou hodinu zvlášť následuje i samotná předpověď s takto zvolenými optimálními parametry. Počet dní, pro který chceme predikovat, odpovídá počtu hodnot, který budeme vypočítávat predikčním algoritmem ze všech 24 hodinových řad. Takto získáme 24 řad predikcí. Řady predikcí již obsahují hodnoty vytvořené predikčním algoritmem. Pořadí hodnoty v řadě predikcí koresponduje s tím, o kolikátý den predikce se jedná. A pořadové číslo samotné řady predikcí určuje hodinu. K určení přesnosti předpovědi jsou pak použity skutečné hodnoty cen z původního datasetu.

4.6 Vyhodnocení přesnosti předpovědí ARIMA

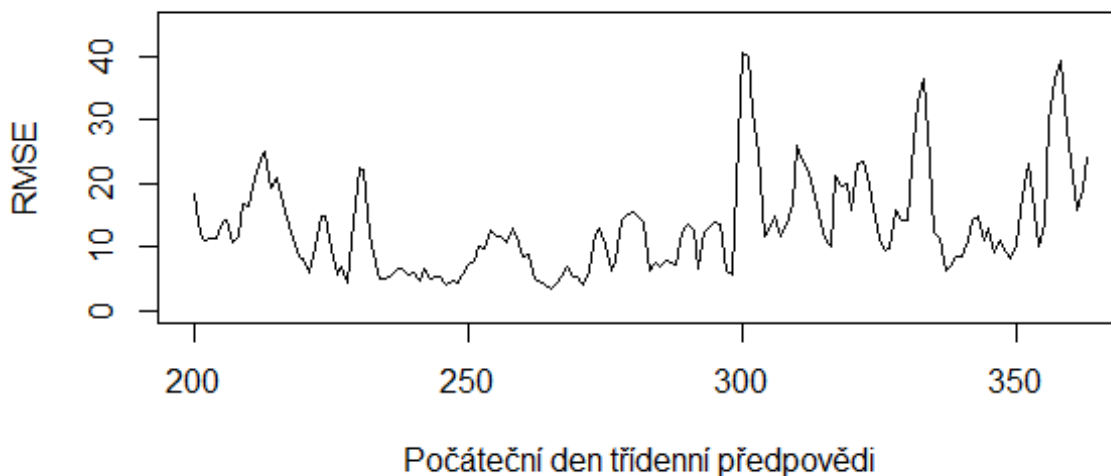
Předpovědní schopnost výše popsaného algoritmu byla ověřena v třídením predikčním intervalu, a to v rámci původního datasetu pro rok 2017, pro který byl prováděn i výzkum optimálního nastavení předpovědních parametrů. V literatuře se většinou předpovědi za účelem posouzení přesnosti realizují na 7 dní. [21] Já jsem se však rozhodl realizovat předpovědi na 3 dny. Důvodem tohoto rozhodnutí bylo, že se třídení předpovědi ukázaly jako vhodnější pro potřeby této práce, zejména z hlediska grafického zobrazování jejich průběhů. Třídení předpovědi zároveň stále disponují dostatečnou délkou pro využití jimi poskytnutých hodnot a posouzení jejich přesnosti. Predikční program (viz příloha) umožňuje realizovat předpověď a posoudit její přesnost pro libovolný počet dnů.

Nejdelší délka učícího období, se kterou pracuje a kterou vyžaduje predikční algoritmus, je 199 hodinových hodnot, viz tabulka výše. Proto byl za výchozí den ověřování predikčního algoritmu vybrán 200. den roku (19. 7. 2017). Od tohoto výchozího dne včetně počínaje, den po dni, až do 363. (29. 12. 2017) dne byla realizována předpověď na tři dny a spočítána její chyba.

Průběh $RMSE$ šetření je k vidění níže, včetně průřezových statistik.

$RMSE$: max = 40,69 min = 3,36 mean = 13,37 sd = 7,7

RMSE třídení předpovědi pro 200. až 363. den roku



Obrázek 4.4: $RMSE$ predikčního algoritmu pro třídení předpověď se začátkem ve dnech 200 až 363 (19. 7. až 29. 12. 2017).

Tabulka 4.2: Průřezové statistiky rozptylů modelů ARIMA jednotlivých hodin z třídeního výzkumu *RMSE*.

Hodina	1	2	3	4	5	6	7	8
mean σ^2	60,18	63,93	65,19	59,92	58,52	79,39	118,06	170,62
sd σ^2	37,64	38,69	38,26	31,35	32,26	60,18	94,22	77,68

Hodina	9	10	11	12	13	14	15	16
mean σ^2	207,03	196,61	175,59	174,69	220,19	221,44	174,81	161,30
sd σ^2	79,75	74,42	71,98	68,70	105,11	99,45	70,83	11,15

Hodina	17	18	19	20	21	22	23	24
mean σ^2	188,37	210,52	144,02	134,86	92,76	80,65	54,80	64,66
sd σ^2	95,44	112,16	78,84	69,75	24,26	25,00	42,63	50,66

Minimální chyba *RMSE* je rovna 3,36, tedy nejlepší předpověď, byla zaznamenána pro 265. až 267. den (22. 9. až 24. 9. 2017). Průběh předpovědi i skutečné hodnoty ceny elektřiny je znázorněn na Obr. 4.5, parametry modelů pro jednotlivé hodiny jsou uvedeny v Tab. 4.3.

Tabulka 4.3: Parametry předpovědních modelů predikce 265. až 267. dne (22. 9. až 24. 9. 2017), tedy případu s nejmenší chybou.

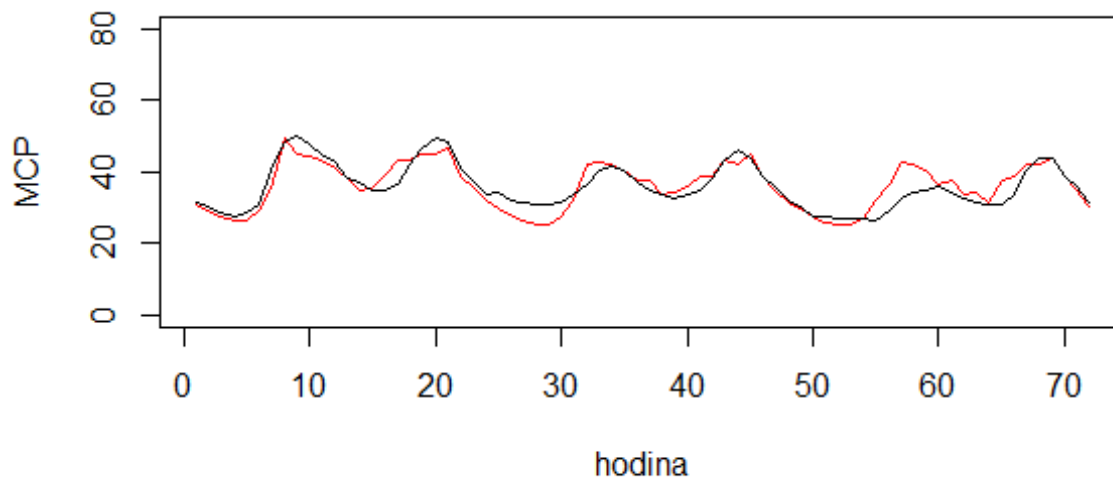
Hodina	1	2	3	4	5	6	7	8
sd	5,07	5,45	5,76	5,76	5,67	5,85	7,18	11,01
<i>p</i> -value Q-test	0,852	0,737	0,754	0,608	0,273	0,222	<0,001	0,003
ARIMA	1, 0, 0	1, 0, 0	1, 0, 0	1, 0, 0	1, 0, 0	1, 0, 0	0, 0, 1	0, 0, 3

Hodina	9	10	11	12	13	14	15	16
sd	12,49	11,78	11,90	12,23	11,54	11,48	12,77	12,23
<i>p</i> -value Q-test	0,034	0,053	0,215	0,189	0,248	0,316	0,055	<0,001
ARIMA	0, 0, 1	0, 1, 2	0, 1, 2	0, 1, 4	1, 0, 0	2, 0, 0	0, 0, 1	3, 0, 2

Hodina	17	18	19	20	21	22	23	24
sd	11,45	10,97	10,26	8,69	9,47	7,53	5,18	5,88
<i>p</i> -value Q-test	0,283	0,598	0,452	0,957	0,936	0,806	0,58	0,985
ARIMA	0, 0, 4	0, 0, 5	1, 0, 0	0, 0, 1	1, 0, 0	0, 0, 0	1, 0, 0	0, 0, 3

Naproti tomu maximální chyba *RMSE* rovna 40,69, tedy nejhorší předpověď, byla zaznamenána pro 300. až 302. den (27. 10. až 29. 10. 2017). Průběhy predikce a skutečných hodnot jsou znázorněny na Obr. 4.6, údaje o modelech použitých pro predikci obsahuje Tab. 4.4.

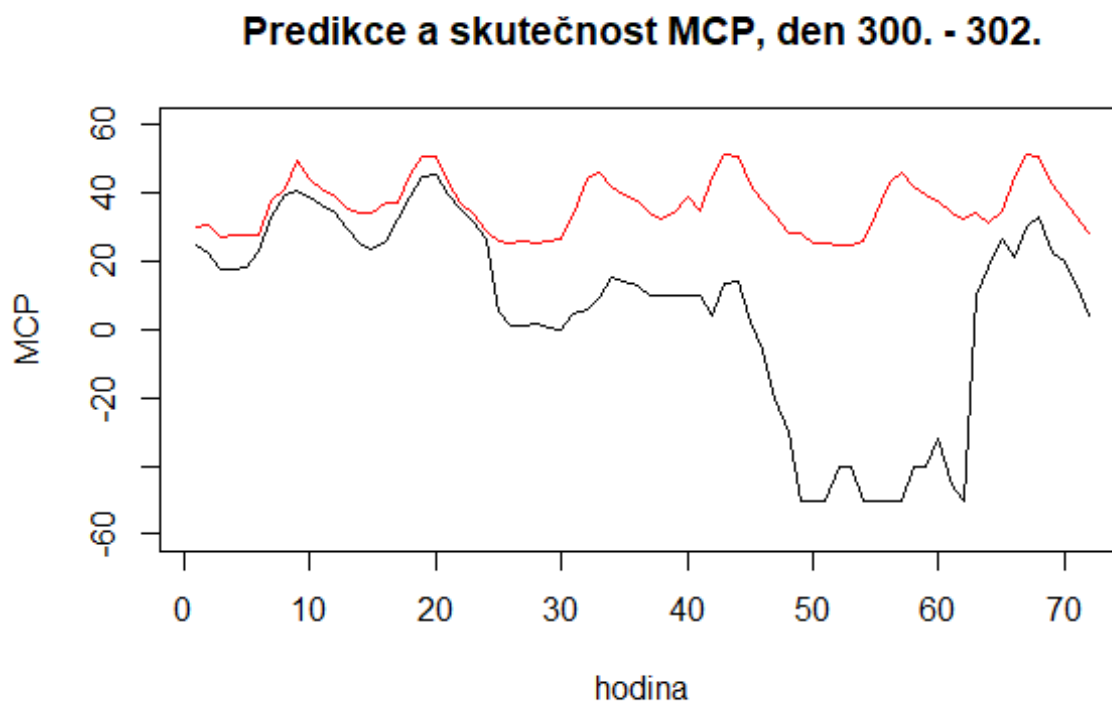
Predikce a skutečnost MCP, den 265. - 267.



Obrázek 4.5: Predikce a skutečné hodnoty třídení předpovědi MCP pro 265. až 267. den (22. 9. až 24. 9. 2017), tedy případu s nejmenší chybou. Červeně je znázorněna predikce, černě skutečné hodnoty.

Jak je patrné z Obr. 4.6, je velká chyba předpovědi způsobena jevem diskutovaným v úvodu této práce, spojeným se zvláštností trhu s elektřinou. Cena obchodované komodity na konci druhého dne předpovědi (28. 10. 2017), jmenovitě v 46. hodině intervalu zájmu, nabyla záporné hodnoty -6 EUR/MWh. Pokles ceny navíc pokračoval dál. Ve 49. hodině intervalu zájmu, tedy v první hodině třetího dne predikce (29. 10. 2017), klesla marginální cena elektřiny na denním trhu na závratných -50 EUR/MWh, minimum skutečné ceny v posuzovaném intervalu. Až v 63. hodině, tedy v 15. hodině třetího dne byla cena elektřiny opět kladná s hodnotou 9,99 EUR/MWh.

Provedeme-li předpověď pouze pro 300. den (27. 10. 2017) a vynecháme z posuzování přesnosti specifický druhý a třetí den (28. 10. a 29. 10. 2017), je chyba předpovědi $RMSE$ rovna 6,70, což je oproti chybě $RMSE$ 40,69 celého třídeního intervalu 300. až 302. dne (27. 10. až 29. 10.) výrazné zlepšení přesnosti, jak je rovněž patrné z Obr. 4.7a. Parametry předpovědních modelů pro samostatnou předpověď 300. dne (27. 10.) se shodují s parametry modelů předpovědi celého třídeního intervalu. Lze je nalézt v Tab. 4.4. Předpovídání ceny elektřiny pro 301. a 302. den (28. 10. a 29. 10.) samostatně dostupnou přesnost v těchto dnech však již takto výrazně nezlepší. U dne 301. (28. 10.) samostatná denní předpověď ještě přinese určité zlepšení v podobě $RMSE$ 31,36 oproti původní chybě 40,69. Denní předpovědi poukazují na to, že je to právě 302. den (29. 10.) se svými extrémními zápornými hodnotami cen, který se zejména podepsal na tak špatném výsledku původní třídení předpovědi. Chyba předpovědi $RMSE$



Obrázek 4.6: Predikce a skutečné hodnoty třídní předpovědi MCP pro 300. až 302. den (27. 10. až 29. 10. 2017), tedy případu s největší chybou. Červeně je znázorněna predikce, černě skutečné hodnoty.

v tomto dni samotném je totiž rovna 58,39. Průběhy těchto jednodenních předpovědí jsou zobrazeny na Obr. 4.7.

Tabulka 4.4: Parametry předpovědních modelů predikce 300. až 302. dne (27. 10. až 29. 10. 2017), tedy případu s největší chybou.

Hodina	1	2	3	4	5	6	7	8
sd	6,49	6,70	7,12	6,77	6,81	7,84	9,05	10,47
p -value Q-test	0,999	0,992	0,724	0,948	0,778	0,97	0,108	0,002
ARIMA	4, 0, 0	3, 0, 2	1, 0, 0	3, 0, 2	1, 0, 0	1, 0, 0	0, 0, 1	1, 1, 2

Hodina	9	10	11	12	13	14	15	16
sd	10,48	8,64	7,54	7,07	7,09	7,32	8,83	12,28
p -value Q-test	0,018	0,932	0,869	0,812	0,592	0,596	0,712	<0,001
ARIMA	0, 1, 2	0, 0, 1	0, 0, 1	0, 0, 1	0, 0, 1	0, 0, 1	0, 1, 1	3, 0, 2

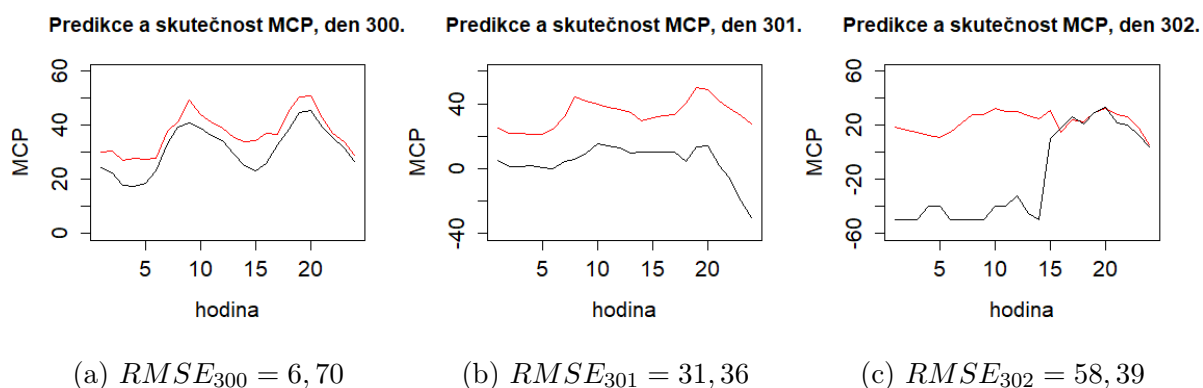
Hodina	17	18	19	20	21	22	23	24
sd	7,31	7,73	9,47	9,23	7,06	6,72	7,41	8,05
p -value Q-test	0,77	0,72	0,089	0,811	0,808	0,994	0,852	0,883
ARIMA	0, 0, 1	0, 1, 2	0, 1, 2	0, 1, 1	1, 0, 0	3, 0, 0	2, 0, 0	2, 0, 0

Tabulka 4.5: Parametry předpovědních modelů jednodenní predikce 301. dne (28. 10. 2017).

Hodina	1	2	3	4	5	6	7	8
sd	6,53	6,73	7,09	6,82	6,83	7,83	9,03	10,35
p -value Q-test	0,99	0,986	0,916	0,948	0,799	0,977	0,09	0,002
ARIMA	3, 0, 0	3, 0, 2	1, 0, 1	3, 0, 2	1, 0, 0	1, 0, 0	0, 0, 1	1, 1, 2

Hodina	9	10	11	12	13	14	15	16
sd	10,51	8,65	7,57	7,09	7,03	7,41	8,71	12,31
p -value Q-test	0,012	0,847	0,791	0,713	0,822	0,436	0,305	<0,001
ARIMA	0, 1, 2	0, 0, 1	0, 0, 1	0, 0, 1	1, 0, 1	0, 0, 1	0, 1, 1	3, 0, 2

Hodina	17	18	19	20	21	22	23	24
sd	7,30	7,63	8,10	9,25	7,07	6,65	7,42	8,02
p -value Q-test	0,758	0,606	0,458	0,802	0,804	0,993	0,816	0,85
ARIMA	0, 0, 1	1, 1, 1	1, 1, 2	0, 1, 1	1, 0, 0	3, 0, 0	2, 0, 0	2, 0, 0



Obrázek 4.7: Predikce a skutečné hodnoty jednodenních předpovědí MCP pro 300., 301. a 302. den (27. 10., 28. 10. a 29. 10. 2017). Červeně je znázorněna predikce, černě skutečné hodnoty.

Tabulka 4.6: Parametry předpovědních modelů jednodenní predikce 302. dne (29. 10. 2017).

Hodina	1	2	3	4	5	6	7	8
sd	6,70	7,09	7,11	7,00	6,89	8,30	8,95	10,93
p -value Q-test	0,959	0,866	0,9	0,827	0,798	0,801	0,092	0,015
ARIMA	4, 0, 0	4, 0, 0	3, 0, 2	3, 0, 2	3, 0, 2	0, 0, 1	0, 0, 1	4, 1, 1

Hodina	9	10	11	12	13	14	15	16
sd	10,97	9,17	8,11	7,66	7,81	7,74	7,75	12,41
p -value Q-test	0,051	0,96	0,984	0,971	0,935	0,999	0,656	<0,001
ARIMA	1, 1, 2	0, 0, 1	0, 0, 1	0, 0, 1	0, 0, 1	1, 0, 1	0, 1, 1	3, 0, 2

Hodina	17	18	19	20	21	22	23	24
sd	7,97	8,85	8,85	9,98	8,42	8,78	10,43	11,06
p -value Q-test	0,976	0,807	0,152	0,871	0,966	0,971	0,995	0,992
ARIMA	0, 0, 1	1, 0, 0	1, 1, 1	1, 0, 0	0, 0, 1	1, 0, 0	1, 0, 0	1, 0, 0

4.7 Algoritmus předpovědi ARIMA–GARCH

Výhody spojené s možností modelování proměnného rozptylu, která u modelů Boxovy–Jenkinsovy metodologie není k dispozici, se staly hlavní motivací pro zavedení metodologie *GARCH* do předpovídání cen elektřiny. Čisté modely *ARCH* či *GARCH* se pro predikování cen elektrické energie neosvědčily jako vhodné, avšak jejich kombinace s modely *ARIMA*, tedy modely *ARIMA – GARCH* (dále jen *ARIMA – G*), jsou již v tomto oboru dlouho etablované. [20, 21] Tato sekce obsahuje popis provedení a výsledky posouzení vhodnosti rozšíření Boxovy–Jenkinsovy metodologie o modely *GARCH*.

Algoritmus předpovědi rozšířený o funkcionalitu modelování heteroskedasticity funguje v základu stejně jako původní algoritmus postavený pouze na Boxově–Jenkinsově metodologii. I zde se podle právě modelované hodiny určí efektivnější test na jednotkový kořen a jemu náležející nejvhodnější učící délka. S použitím informačního kritéria *AIC* je dále pomocí funkce *auto.arima* zkonstruován vhodný model z Boxovy–Jenkinsovy metodologie.

Předpovědní postup modelující heteroskedasticitu pomocí metodologie *GARCH* však k tomuto procesu přidává ověření homoskedasticity reziduí. Za tímto účelem jsou na reziduích modelů provedeny *LM*–test a *ML*–test. Pokud oba tyto testy na hladině významnosti 5 % nezamítnou homoskedasticitu reziduí modelu, je samotná predikce realizována z původního modelu *auto.arima*, u jehož reziduí byla homoskedasticita ověřována. Postup tedy v tomto případě zůstává totožný jako u prvního předpovědního algoritmu.

Zamítne-li však jeden z testů či oba testy nulovou hypotézu homoskedasticity a je-li tedy v reziduích modelu *auto.arima* přítomna heteroskedasticita, je přikročeno k realizaci modelu *ARIMA – G*. Ten je vytvořen v prostředí knihovny *rugarch*. Řády původního *ARIMA* modelu *auto.arima* zůstávají při tvorbě modelu *ARIMA – G* zachovány. Rozptyl nového modelu je proměnný a je popsán modelem *GARCH*(1, 1). Řády modelu *GARCH* byly zvoleny s ohledem na doporučení literatury. [2, 21]

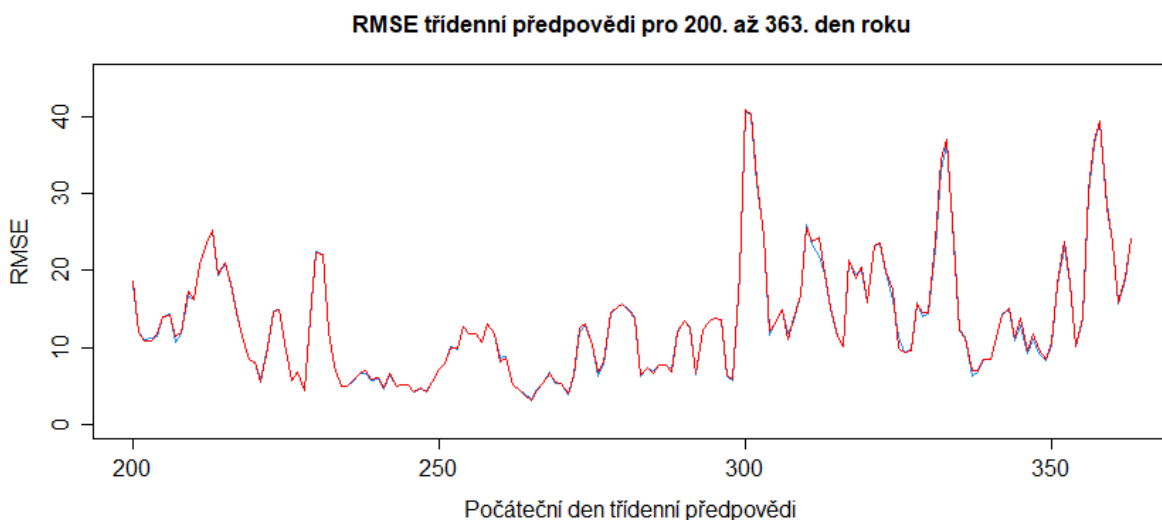
Další změna, kterou s sebou v algoritmu *ARIMA – G* nese detekce heteroskedasticity v reziduích původního *ARIMA* modelu, se týká délky učícího období. Funkce *rugarchfit* vyžaduje pro identifikaci modelu a odhad jeho parametrů délku učícího období alespoň 100 hodnot. Navíc může být u časové řady učícího období detekována stochastická nestacionarita testem na jednotkový kořen. Za této situace může být modelovaná časová řada učícího období až dvakrát diferencována a může tedy dojít k jejímu zkrácení až o 2 hodnoty. Z tohoto důvodu jsou optimální učící délky pro jednotlivé hodiny převzaté z původního nerozšířeného algoritmu *ARIMA* kratší než 102 hodnot pro potřeby algoritmu *ARIMA – G* prodlouženy na 102 hodnot.

U fitovací funkce *rugarchfit* knihovny *rugarch* může nastat situace, že se jí nepodaří úspěšně nafitovat *ARIMA – G* model popisující časovou řadu učícího období. V těchto

případech je pak i přes heteroskedasticitu zjištěnou testy *LM* či *ML* k modelování použit původní model funkce *auto.arima* a řada je modelována čistě Boxovou–Jenkinsovou metodologií.

4.8 Vyhodnocení přesnosti předpovědí ARIMA–G

Pro posouzení přesnosti výše popsaného predikčního algoritmu *ARIMA–G* byl pro tento algoritmus opakovan výše popsaný výzkum chyb třídních předpovědí v intervalu 200. až 363. počátečního dne (19. 7. až 29. 12. 2017). Výsledky tohoto výzkumu jsou zajímavé zejména v porovnání s výsledky původních předpovědí, které byly realizovány bez rozšíření *GARCH*. Na Obr. 4.8 je grafické porovnání chyb těchto dvou předpovědních algoritmů.



Obrázek 4.8: RMSE predikčního algoritmu pro třídní předpověď se začátkem ve 200. až 363. dni (19. 7. až 29. 12. 2017). Modře jsou znázorněny chyby původního předpovědního algoritmu *ARIMA*, červeně pak předpovědní chyby algoritmu *ARIMA–G*.

Průřezové statistiky změn chyb *RMSE* při zavedení *GARCH* jsou

$$\Delta GARCH : \quad \max = 2,45 \quad \min = -1,47 \quad \text{mean} = 0,09 \quad \text{sd} = 0,38$$

Zavedení metodologie *GARCH* do původního předpovědního algoritmu tedy vedlo ke zvýšení chyby *RMSE* až o rozdíl 2,45. Na druhou stranu přineslo modelování proměnného rozptylu v jednom konkrétním případě největší zmenšení *RMSE* o 1,47. V průměru pak rozšíření *GARCH* zvětšilo předpovědní chybu *RMSE* o 0,09. Celkem přineslo zavedení *GARCH* u 164 realizovaných třídních předpovědí zlepšení přesnosti v 70 případech.

V 94 případech byla chyba modelů $ARIMA - G$ větší, než u nerozšířených modelů $ARIMA$.

Samotné průřezové statistiky $RMSE$ obou přístupů jsou

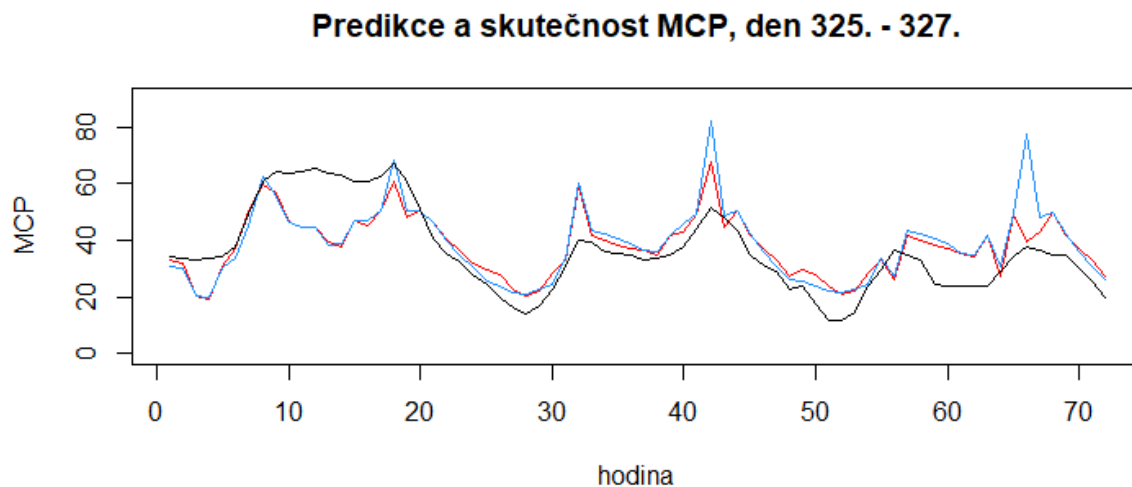
$$\begin{array}{llll} RMSE_{ARIMA-G} : \max = 40,91 & \min = 3,24 & mean = 13,47 & sd = 7,8 \\ RMSE_{ARIMA} : \max = 40,69 & \min = 3,36 & mean = 13,37 & sd = 7,7 \end{array}$$

Maximální hodnota chyby 40,91 rozšířeného algoritmu $ARIMA - G$ byla stejně jako u nerozšířeného algoritmu $ARIMA$ zaznamenána pro období 300. až 302. dne (27. 10. až 29. 10. 2017). Stejně tak minimum $RMSE$ náleželo po rozšíření stejnému období jako u původního postupu, tj. časovému rozmezí 265. až 267. dne (22. 9. až 24. 9. 2017). Rozšíření $ARIMA - G$ tedy dokázalo snížit minimální chybu $RMSE$ na 3,24, ale jeho maximální chyba se zvýšila na 40,69.

Největší zlepšení $RMSE$ rovné 1,47 přineslo rozšíření $ARIMA - G$ v predikčním intervalu 325. až 327. dne (21. 11. až 23. 11. 2017). Původní algoritmus měl chybu předpovědi $RMSE_{ARIMA} = 11,43$, rozšířený algoritmus předpovídal s chybou $RMSE_{ARIMA-G} = 9,96$. Průběhy předpovědi jsou vykresleny na Obr. 4.9, parametry modelů uvedeny v Tab. 4.7.

Nejhorší vliv pak mělo rozšíření $ARIMA - G$ v predikčním intervalu 312. až 314. dne (8. 11. až 10. 11. 2017). $RMSE$ v tomto dni stoupla z původní hodnoty $RMSE_{ARIMA} = 21,84$ na $RMSE_{ARIMA-G} = 24,29$. Průběhy předpovědi obou přístupů a skutečné hodnoty v daném období jsou na Obr. 4.10. Zhoršení predikce lze v případě 312. až 314. dne připsat zejména predikci modelu $ARIMA - G$ pro 11. hodinu předpovědi. Ten předpověděl postupný pokles ceny elektřiny během těchto tří dnů až do záporných hodnot. Parametry předpovědních modelů pro 312. až 314. den jsou uvedeny v Tab. 4.8.

Co se týče chybného poklesu predikované ceny v 11. hodině a jeho vlivu na celkové zhoršení předpovědi, vezměme pro porovnání případ, kdy algoritmus $ARIMA - G$ oproti původnímu algoritmu $ARIMA$ přinesl druhé největší zhoršení. Náleželo-li první největší zhoršení již zmiňovanému rozmezí 312. až 314. dne, pak k druhému největšímu zhoršení při rozšíření původního algoritmu došlo v intervalu 332. až 334. dne (28. 11. až 30. 11. 2017). Jedenáctá hodina zde byla stejně jako v případě největšího zhoršení modelována funkcemi z knihovny *rugarch*. K obdobnému poklesu do záporných hodnot už však v predikci pro tuto hodinu nedošlo, viz Obr. 4.11. Navýšení chyby způsobené rozšířením $ARIMA - G$ zde bylo rovné 1,6, přičemž chyba původního algoritmu byla rovna $RMSE_{ARIMA} = 32,86$ a rozšířeného $RMSE_{ARIMA-G} = 34,46$. Rozdíl $RMSE$ mezi prvním a druhým největším zhoršením činí 0,85 a kromě případu prvního největšího zhoršení již rozšíření $ARIMA - G$ podobný neopodstatněný odklon od původní předpovědi nepredikovalo. Parametry modelů použitých k predikci období 332. až 334. dne jsou uvedeny v Tab. 4.9.



Obrázek 4.9: Predikce a skutečné hodnoty tří denní předpovědi MCP pro 325. až 327. den (21. 11. až 23. 11. 2017), tedy případu, kdy rozšíření $ARIMA-G$ přineslo největší zlepšení. Červeně je znázorněna predikce $ARIMA-G$, modře predikce původního algoritmu $ARIMA$, černě skutečné hodnoty.

V realizovaném vyhodnocení přesnosti vedlo rozšíření $GARCH(1,1)$ standardních modelů $ARIMA$ ke zvýšení přesnosti předpovědí jen v mírné menšině případů, přičemž v mírné většině případů dokonce přesnost snížilo. To může být v rozporu s očekáváním, podle kterého by měla možnost modelování proměnného rozptylu zlepšit předpovědní schopnosti. Dosažený výsledek však odpovídá hodnocení předpovědních schopností kombinovaných modelů $ARIMA-G$ v literatuře, kde existuje mnoho zaznamenaných případů právě jak zlepšení, tak i zhoršení předpovědních schopností v důsledku zavedení tohoto rozšíření. [21]

Tabulka 4.7: Parametry předpovědních modelů predikce 325. až 327. dne (21. 11. až 23. 11. 2017), tedy případu, kdy rozšíření $ARIMA - G$ přineslo největší zlepšení. Řádek 2 obsahuje p -hodnoty testu na heteroskedasticitu LM (Lagrange Multiplier) reziduí nerozšířených modelů $ARIMA$. Pokud byla zamítnuta nulová hypotéza homoskedasticity, byl zvolen model $ARIMA - G$ a řádky 4 a 6 pak obsahují p -hodnoty testů jeho reziduí (Q -testu na nekorelovanost, McLeod-Li testu na remanentní $ARCH$ -efekt). V opačném případě byl zvolen model $ARIMA$ s p -hodnotou Q -testu jeho reziduí na řádku 4 a se směrodatnou odchylkou na řádku 5. Řádek 3 obsahuje řády modelu $ARIMA$ v obou případech.

Hodina	1	2	3	4	5	6
p -value LM $ARIMA$	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
ARIMA	0, 0, 1	0, 0, 1	1, 0, 0	1, 0, 0	0, 0, 1	0, 0, 1
p -value Q-test	0,322	0,228	0,098	0,225	0,198	0,471
sd $ARIMA$	–	–	–	–	–	–
p -value ML $ARIMA - G$	1	0,999	0,363	0,936	0,999	0,959

Hodina	7	8	9	10	11	12
p -value LM $ARIMA$	0,001	0,049	0,001	<0,001	<0,001	<0,001
ARIMA	0, 0, 1	0, 0, 3	0, 0, 1	0, 0, 1	0, 0, 1	0, 0, 1
p -value Q-test	0,001	0,005	0,001	0,008	0,053	0,045
sd $ARIMA$	–	–	–	–	–	–
p -value ML $ARIMA - G$	0,096	<0,001	0,24	0,481	0,605	0,234

Hodina	13	14	15	16	17	18
p -value LM $ARIMA$	<0,001	<0,001	0,533	0,007	0,323	0,036
ARIMA	1, 0, 0	1, 0, 0	0, 1, 2	3, 0, 2	1, 1, 1	2, 1, 3
p -value Q-test	0,029	0,084	0,143	<0,001	0,082	<0,001
sd $ARIMA$	–	–	11,22	–	14	–
p -value ML $ARIMA - G$	0,308	0,957	–	<0,001	–	<0,001

Hodina	19	20	21	22	23	24
p -value LM $ARIMA$	0,039	0,518	0,017	0,001	<0,001	<0,001
ARIMA	1, 0, 0	1, 0, 0	0, 0, 1	0, 0, 1	0, 0, 1	0, 0, 1
p -value Q-test	0,009	0,302	0,35	0,524	0,853	0,814
sd $ARIMA$	–	14,53	–	–	–	–
p -value ML $ARIMA - G$	<0,001	–	0,731	0,999	1	1

Tabulka 4.8: Parametry předpovědních modelů predikce 312. až 314. dne (8. 11. až 10. 11. 2017), tedy případu, kdy rozšíření *ARIMA* – *G* přineslo největší zhoršení. Řádek 2 obsahuje *p*–hodnoty testu na heteroskedasticitu *LM* (Lagrange Multiplier) reziduí nerozšířených modelů *ARIMA*. Pokud byla zamítnuta nulová hypotéza homoskedasticity, byl zvolen model *ARIMA* – *G* a řádky 4 a 6 pak obsahují *p*–hodnoty testů jeho reziduí (*Q*–testu na nekorelovanost, McLeod–Li testu na remanentní *ARCH*–efekt). V opačném případě byl zvolen model *ARIMA* s *p*–hodnotou *Q*–testu jeho reziduí na řádku 4 a se směrodatnou odchylkou na řádku 5. Řádek 3 obsahuje řády modelu *ARIMA* v obou případech.

Hodina	1	2	3	4	5	6
<i>p</i> -value LM <i>ARIMA</i>	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
ARIMA	3, 0, 0	3, 0, 0	1, 0, 0	0, 0, 4	1, 0, 0	0, 0, 1
<i>p</i> -value Q-test	0,105	0,066	0,046	0,422	0,015	0,111
sd <i>ARIMA</i>	–	–	–	–	–	–
<i>p</i> -value ML <i>ARIMA</i> – <i>G</i>	1	0,998	0,91	0,57	0,451	0,964

Hodina	7	8	9	10	11	12
<i>p</i> -value LM <i>ARIMA</i>	0,002	0,039	0,011	<0,001	<0,001	<0,001
ARIMA	2, 0, 2	2, 0, 3	2, 0, 2	0, 0, 1	3, 0, 2	3, 0, 2
<i>p</i> -value Q-test	0,005	0,003	0,067	0,034	<0,001	0,04
sd <i>ARIMA</i>	–	–	–	–	–	–
<i>p</i> -value ML <i>ARIMA</i> – <i>G</i>	0,811	0,899	0,184	0,529	<0,001	0,673

Hodina	13	14	15	16	17	18
<i>p</i> -value LM <i>ARIMA</i>	<0,001	<0,001	0,22	0,003	0,119	0,055
ARIMA	0, 0, 1	0, 0, 1	1, 1, 2	3, 0, 2	0, 0, 1	3, 0, 0
<i>p</i> -value Q-test	0,118	0,1	0,869	<0,001	0,546	0,494
sd <i>ARIMA</i>	–	–	8,39	–	9,43	10,23
<i>p</i> -value ML <i>ARIMA</i> – <i>G</i>	0,937	0,994	–	<0,001	–	–

Hodina	19	20	21	22	23	24
<i>p</i> -value LM <i>ARIMA</i>	0,093	0,335	<0,001	<0,001	<0,001	<0,001
ARIMA	3, 0, 0	2, 0, 3	3, 0, 0	0, 0, 1	1, 0, 0	1, 0, 0
<i>p</i> -value Q-test	0,085	0,87	0,446	0,198	0,618	0,605
sd <i>ARIMA</i>	9,86	10,57	–	–	–	–
<i>p</i> -value ML <i>ARIMA</i> – <i>G</i>	–	–	0,997	1	1	1

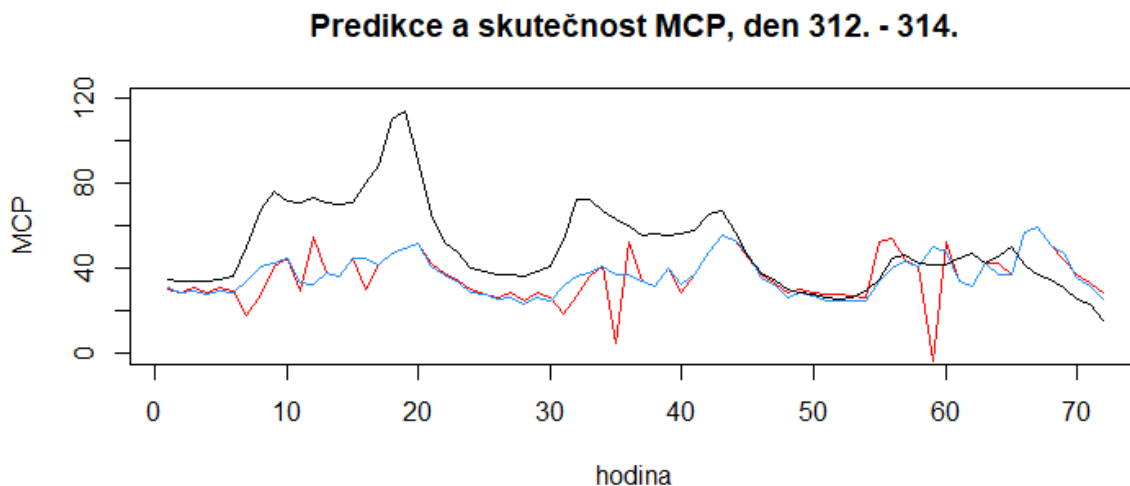
Tabulka 4.9: Parametry předpovědních modelů predikce 332. až 334. dne (28. 11. až 30. 11. 2017), tedy případu, kdy rozšíření $ARIMA - G$ přineslo druhé největší zhoršení. Řádek 2 obsahuje p -hodnoty testu na heteroskedasticitu LM (Lagrange Multiplier) reziduí nerozšířených modelů $ARIMA$. Pokud byla zamítnuta nulová hypotéza homoskedasticity, byl zvolen model $ARIMA - G$ a řádky 4 a 6 pak obsahují p -hodnoty testů jeho reziduí (Q -testu na nekorelovanost, McLeod-Li testu na remanentní $ARCH$ -efekt). V opačném případě byl zvolen model $ARIMA$ s p -hodnotou Q -testu jeho reziduí na řádku 4 a se směrodatnou odchylkou na řádku 5. Řádek 3 obsahuje řády modelu $ARIMA$ v obou případech.

Hodina	1	2	3	4	5	6
p -value LM $ARIMA$	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
$ARIMA$	2, 0, 3	0, 0, 1	3, 0, 2	3, 0, 2	3, 0, 2	0, 0, 1
p -value Q -test	0,7	0,152	0,294	0,522	0,095	0,304
sd $ARIMA$	–	–	–	–	–	–
p -value ML $ARIMA - G$	1	0,997	0,138	0,268	0,1	0,932

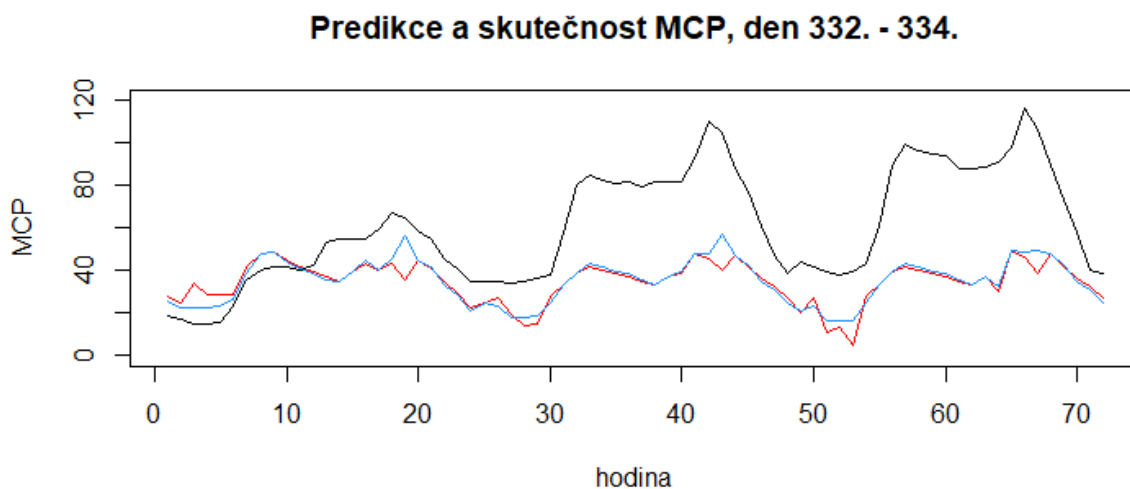
Hodina	7	8	9	10	11	12
p -value LM $ARIMA$	0,002	0,256	0,002	<0,001	<0,001	0,002
$ARIMA$	0, 0, 1	2, 0, 3	0, 0, 1	0, 0, 1	0, 0, 1	0, 0, 1
p -value Q -test	<0,001	0,012	<0,001	0,032	0,134	0,189
sd $ARIMA$	–	15,83	–	–	–	–
p -value ML $ARIMA - G$	0,075	–	0,274	0,577	0,683	0,311

Hodina	13	14	15	16	17	18
p -value LM $ARIMA$	<0,001	<0,001	0,468	0,008	0,783	0,043
$ARIMA$	0, 0, 1	0, 0, 1	0, 1, 2	3, 0, 2	3, 1, 2	1, 1, 1
p -value Q -test	0,145	0,208	0,373	<0,001	0,596	0,017
sd $ARIMA$	–	–	11,95	–	13,52	–
p -value ML $ARIMA - G$	0,59	0,925	–	<0,001	–	<0,001

Hodina	19	20	21	22	23	24
p -value LM $ARIMA$	0,036	0,343	0,038	0,019	<0,001	<0,001
$ARIMA$	2, 1, 3	1, 0, 0	0, 0, 1	0, 0, 1	0, 0, 1	0, 0, 1
p -value Q -test	<0,001	0,343	0,279	0,521	0,808	0,789
sd $ARIMA$	–	15,03	–	–	–	–
p -value ML $ARIMA - G$	<0,001	–	0,818	0,999	1	1



Obrázek 4.10: Predikce a skutečné hodnoty třídní předpovědi MCP pro 312. až 314. den (8. 11. až 10. 11. 2017), tedy případu, kdy rozšíření $ARIMA - G$ přineslo největší zhoršení. Červeně je znázorněna predikce $ARIMA - G$, modře predikce původního algoritmu $ARIMA$, černě skutečné hodnoty.



Obrázek 4.11: Predikce a skutečné hodnoty třídní předpovědi MCP pro 332. až 334. den (28. 11. až 30. 11. 2017), tedy případu, kdy rozšíření $ARIMA - G$ přineslo druhé největší zhoršení. Červeně je znázorněna predikce $ARIMA - G$, modře predikce původního algoritmu $ARIMA$, černě skutečné hodnoty.

Kapitola 5

Závěr

Tato práce poskytla přehled mechanismu vzniku cen elektřiny na spotovém trhu a přístupů k jejich předpovědi. Na základě doporučení z existující literatury a široké případové studie byla pro předpověď byla zvolena Boxova–Jenkinsova metodologie. Samotná realizace studie a předpovědi byla provedena v prostředí statistického softwaru R.

Studie proběhla na datech o hodinových cenách v roce 2017, tedy na 8760 hodnotách. Z důvodu sezónnosti přítomné v cenách elektřiny zahrnoval postup nejprve rozdělení původních dat na 24 samostatných datových souborů, které obsahují ceny elektřiny za rok 2017 pouze pro jednu hodinu dne. Každý z těchto 24 hodinových datasetů tedy obsahuje 365 hodnot. Pro předpověď ceny elektřiny je pak zapotřebí realizovat 24 modelů, přičemž každý z těchto modelů předpovídá průběh ceny elektřiny pro svoji příslušnou hodinu.

Ačkoliv je samotné modelování ceny, tj. identifikace modelu a odhad jeho parametrů, prováděno automatizovaně, je zapotřebí stanovit hned několik důležitých parametrů, na základě kterých může být provedeno. Na nastavení těchto parametrů závisí zásadní měrou přesnost předpovědi modelu, který je s tímto nastavením identifikován a odhadován. Za účelem zjištění optimálního nastavení byl realizován empirický výzkum předpovědních schopností modelů s různými parametry.

Předmětem výzkumu byla nejprve identifikace efektivnějšího informačního kritéria, na základě kterého se posuzují relativní vhodnosti modelů. Ze dvou posuzovaných informačních kritérií *AIC* a *BIC* se jako vhodnější ukázalo kritérium *AIC*.

Dále se výzkum zabýval délkou tzv. učícího období, které představuje počet po sobě následujících hodinových hodnot, resp. dnů, pro které je následně identifikován příslušný predikční model a učiněn jeho odhad. Výzkum vyzkoušel předpovědní schopnosti modelů odhadnutých z široké škály těchto učebních období. Pro každou hodinu, resp. hodinový dataset, byla nalezena délka učebního období se kterou bylo dosaženo nejlepších

průměrných předpovědních výsledků.

Posledním posuzovaným parametrem byl test na jednotkový kořen. Přítomnost jednotkového kořenu v modelu určité časové řady svědčí o přítomnosti stochastické nestacionarity v této řadě. Stochasticky nestacionární data nelze s pomocí Boxovy–Jenkinsovy metodologie správně a spolehlivě modelovat. Tento druh nestacionarity je však při její detekci možné snadno odstranit diferencováním příslušné řady, přičemž detekce se provádí právě testem na jednotkový kořen. V této práci byly používány dva takovéto testy, a sice *ADF* a *KPSS*. Provedený výzkum našel pro každou hodinu vhodnější z těchto dvou kritérií.

Realizovaná předpověď se posléze opírala o tyto poznatky. Nastavení parametrů funkce pro modelování cen elektřiny tedy zahrnovalo volbu informačního kritéria *AIC* pro posuzování relativní vhodnosti modelů a dále volbu testu na jednotkový kořen a délky učebního období v závislosti na předpovídané hodině, tj. volbu testu na jednotkový kořen a učící délky s nejnižší průměrnou chybou.

Předpovědní schopnosti tohoto systému byly empiricky ověřeny realizací třídenní předpovědi pro 200. až 363. den roku (19. 7. až 29. 12. 2017). Nejlepších předpovědních výsledků bylo dosaženo u předpovědi pro 265. až 267. den, tedy dny 22. 9. až 24. 9. 2017. Nejhorší predikční výkon byl pak zaznamenán u předpovědi pro 300. až 302. den, tedy u období 27. 10. až 29. 10. 2017. Pro tyto dvě předpovědi jsou v práci uvedeny grafy zobrazující skutečný a predikovaný průběh ceny elektřiny. Zároveň jsou pro tyto dva případy uvedeny atributy vytvořených modelů, které posloužily k realizaci předpovědi. Těmito atributy jsou směrodatná odchylka modelu, dále p -hodnota Q -testu reziduí tohoto modelu, který ověřuje jejich nekorelovanost, a jeho konkrétní řád *ARIMA*, popisující jeho zařazení v rámci Boxovy–Jenkinsovy metodologie.

Jako největší problém pro přesnost předpovědi realizovaných Boxovou–Jenkinsovou metodologií se ukázalo skokové chování ceny elektřiny, konkrétně nastane-li její rychlý pokles do relativně velmi nízkých záporných hodnot. Modely Boxovy–Jenkinsovy metodologie toto chování dokáží vystihnout jen ztěží či jej nedokáží modelovat vůbec a tento jejich nedostatek je zmiňován i v odborné literatuře.

V další části práce byl predikční postup založený na Boxově–Jenkinsově metodologii rozšířen o modelování proměnného rozptylu *GARCH*. To bylo provedeno testováním reziduí vytvořených modelů testy *LM* a *ML* na přítomnost heteroskedasticity. Pokud heteroskedasticita v reziduích detekována nebyla, předpovídal predikční program stejně jako v případě původního postupu popsaného výše. V případě přítomnosti heteroskedasticity v modelované časové řadě byl pro realizaci predikce příslušné časové řady použit model *ARIMA* – G , kde G označuje model *ARIMA* s proměnným rozptylem popsaným modelem *GARCH*(1, 1).

Posouzení přesnosti rozšířeného předpovědního postupu *ARIMA* – G a jeho porovnání

s původním předpovědním postupem bylo uskutečněno opakováním realizace třídenní předpovědi pro 200. až 363. den roku (19. 7. až 29. 12. 2017) pro postup $ARIMA - G$. Rozšíření modelů Boxovy–Jenkinsovy metodologie o možnost modelování proměnného rozptylu metodologií $GARCH$ vedlo u mírné menšiny realizovaných předpovědí k jejich zlepšení a u mírné většiny k jejich zhoršení, přičemž průměrná chyba v důsledku tohoto rozšíření vzrostla. Tento výsledek je ve shodě s hodnocením modelů $ARIMA - G$ v literatuře, kde je rozšíření G připisováno rovnou měrou jak zlepšování, tak zhoršování předpovědí. [21]

Použitá literatura

- [1] Tomáš Bartoš, Petr Strejček: *Liberalizace elektro-energetického sektoru vybraných zemí EU*.
Dostupné z: <https://energetika.tzb-info.cz/energeticka-politika/8903-liberalizace-elektro-energetickeho-sektoru-vybranych-zemi-eu>, 2012.
- [2] Tomáš Cipra: *Finanční ekonometrie*. 2. vydání, Ekopress, Praha, 2014.
- [3] Frank J. Fabozzi, Sergio M. Focardi, Svetlozar T. Rachev, Bala G. Arshanapalli: *The Basics of Financial Econometrics; Tools, Concepts, and Asset Management Applications*. John Wiley & Sons, New York, 2014.
- [4] Alexios Ghalanos: *Introduction to the rugarch package. (Version 1.3-8)*.
Dostupné z: https://cran.r-project.org/web/packages/rugarch/vignettes/Introduction_to_the_rugarch_package.pdf, 2018.
- [5] Kateřina Helisová: *Matematika pro ekonomiku*. Dostupné z: math.feld.cvut.cz/helisova/01MPE_zapisky.pdf, 2011.
- [6] Robert J. Hyndman, Yeasmin Khandakar: *Automatic Time Series Forecasting: The forecast Package for R*. Journal of Statistical Software, číslo 27, vydání 3, str. 1 – 22, 2008.
- [7] Igor Chemišinec, Jakub Nečesaný: *Základní činnosti OTE, a. s., na trhu s elektřinou a plynem*. Přednáška, Katedra ekonomiky, manažerství a humanitních věd, FEL ČVUT, Praha, 4. prosince 2017.
- [8] Robert M. de Jong: *Lecture 2: ARMA Models*. Ohio State University, Department of Economics. Dostupné z: <http://www.econ.ohio-state.edu/dejong/note2.pdf>.
- [9] Richard Kabele: *Trhy s podpůrnými službami*. Přednáška, Ekofórum, VŠE, Praha, 2. května 2011.
- [10] Milan Kloubec: *Liberalizace vybraných trhů s elektřinou a komparace jejich vývoje*. Disertační práce, FEL ČVUT, Praha, 2014.

- [11] Vladimir S. Koritarov: *Real-World Market Representation with Agents*. IEEE power & energy magazine, číslo 2, vydání 4, str. 39 – 46, 2004.
- [12] Jakub Nowotarski, Rafał Weron: *Recent advances in electricity price forecasting: A review of probabilistic forecasting*. Renewable and Sustainable Energy Reviews, číslo 81, str. 1548 – 1568, 2018.
- [13] OTE, a.s.: *Roční zpráva o trhu s elektřinou*. Dostupné z: http://www.ote-cr.cz/statistika/rocni-zprava/attached/report/2017/-Rocni_zprava_o_trhu_2017_V0.zip, 2017.
- [14] OTE, a.s.: *OTE Product Sheet 2017*. Dostupné z: http://m.ote-cr.cz/about-ote/files-statutory-bodies-and-management/Product_sheet_OTE_EN.pdf, 2017.
- [15] Zuzana Prášková, Petr Lachout: *Základy náhodných procesů I*. 2. vydání, Matfyzpress, Praha, 2015.
- [16] Pär Sjölander: *Stationary Unbiased Finite Sample ARCH-LM Test Procedure*. Applied Economics, číslo 43, vydání 8, str. 1019 – 1033, 2010.
- [17] Aneta Šátralová: *Liberalization Process of the EU Electricity Market*. Diplomová práce, VŠE, Praha, 2011.
- [18] Miroslav Vítek: *Ekonomika dopravních energetických systémů*. 2. vydání, Skripta, ČVUT, Praha, 2008.
- [19] W. Wang, P. H. A. J. M. Van Gelder, J. K. Vrijling, J. Ma: *Testing and modelling autoregressive conditional heteroskedasticity of streamflow processes*. Nonlinear Processes in Geophysics, číslo 12, vydání 1, str. 55 – 66, 2005.
- [20] Rafał Weron: *Electricity price forecasting: A review of the state-of-the-art with a look into the future*. International Journal of Forecasting, číslo 30, vydání 4, str. 1030 – 1081, 2014.
- [21] Rafał Weron: *Modeling and Forecasting Electricity Loads and Prices, A Statistical Approach*. John Wiley & Sons, New York, 2006.
- [22] Vyhláška 408/2015 Sb. o Pravidlech trhu s elektřinou. Dostupné z: <https://www.eru.cz/-/vyhlaska-c-408-2015-sb-o-pravidlech-trhu-s-elektrinou>, 2015.