

ZÁPADOČESKÁ UNIVERZITA V PLZNI

FAKULTA EKONOMICKÁ

Diplomová práce

Ekonomické časové řady

The economic time series

Bc. Josef Šeffl

Plzeň 2019

ZÁPADOČESKÁ UNIVERZITA V PLZNI

Fakulta ekonomická

Akademický rok: 2018/2019

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Josef ŠEFFL**

Osobní číslo: **K17N0110P**

Studijní program: **N6209 Systémové inženýrství a informatika**

Studijní obor: **Systémy projektového řízení**

Název tématu: **Ekonomické časové řady**

Zadávací katedra: **Katedra podnikové ekonomiky a managementu**

Z á s a d y p r o v y p r a c o v á n í :

1. Popište metody analýzy časových řad.
2. Na vybrané časové řady aplikujte popsané metody.
3. Porovnejte dosažené výsledky.
4. Formulujte a diskutujte závěry.

Rozsah grafických prací: **neuveden**
Rozsah kvalifikační práce: **60 - 80 stran**
Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:


- CIPRA, Tomáš. *Finanční ekonometrie*. Praha: Ekopress, 2008. ISBN 978-80-86929-43-9.
- ARLT, Josef, ARLTOVÁ, Markéta. *Ekonomické časové řady*. Praha: Professional Publishing, 2009. ISBN 978-80-86946-85-6.
- KOČENDA, Evžen, ČERNÝ, Alexandr. *Elements of time series econometrics: an applied approach*. Third edition. Prague: Charles University in Prague, Karolinum Press, 2015. ISBN 978-80-246-3199-8.
- ŘEZANKOVÁ, Hana, LÖSTER, Tomáš. *Úvod do statistiky*. Praha: Oeconomica, 2009. ISBN 978-80-245-1514-4.
- KLUFOVÁ, Renata, ROST, Michael, KLICNAROVÁ, Jana. *Modelování regionálních procesů*. Praha: Alfa Nakladatelství, 2012. *Ekonomie studium*. ISBN 978-80-87197-53-0.

Vedoucí diplomové práce: **Ing. Mgr. Milan Svoboda, Ph.D.**
Katedra ekonomie a kvantitativních metod

Datum zadání diplomové práce: **23. října 2018**
Termín odevzdání diplomové práce: **23. dubna 2019**


Doc. Ing. Michaela Krechovská, Ph.D.
děkanka




Doc. PaedDr. Dana Egerová, Ph.D.
vedoucí katedry

V Plzni dne 23. října 2018

Čestné prohlášení

Prohlašuji, že jsem diplomovou práci na téma

„Ekonomické časové řady“

vypracoval samostatně pod odborným dohledem vedoucího diplomové práce a za použití pramenů uvedených v příložené bibliografii.

Plzeň dne

.....

podpis autora

Poděkování

Tímto bych chtěl poděkovat Ing. Mgr. Milanu Svobodovi Ph.D. za metodické vedení práce, vstřícnost a cenné rady, které mi poskytl při zpracovávání této diplomové práce.

Obsah

Úvod.....	7
1 Časové řady	8
1.1 Význam analýzy časových řad.....	9
1.2 Potenciální problémy při zpracování analýzy	9
1.3 Základní charakteristiky časových řad.....	10
1.4 Základní přístupy k analýze časových řad	12
2 Dekompozice časové řady	14
2.1 Modelování trendu	15
2.1.1 Klasické přístupy	15
2.1.2 Adaptivní přístupy	17
2.2 Modelování sezónnosti.....	22
2.2.1 Kvantifikace sezónních výkyvů.....	23
2.2.2 Sezónní očišťování.....	24
3 Box-Jenkinsova metodologie	26
3.1 Stacionarita.....	26
3.2 Autokorelační a parciální autokorelační funkce	27
3.3 Modely (AR, MA, ARIMA a SARIMA).....	30
4 Extrapolace časových řad.....	34
4.1 Extrapolace u dekompozice časových řad	34
4.2 Extrapolace u vybraných metod z Box-Jenkinsovy metodologie	36
4.3 Dummy proměnné.....	37
5 Analýza vybraných časových řad.....	38
5.1 Popis dat.....	38
5.1.1 Tržby za ubytování	38
5.1.2 Hrubá mzda.....	40
5.2 Dekompozice časových řad.....	42
5.2.1 Modelování trendu (ubytování)	42
5.2.2 Modelování sezónnosti (ubytování).....	52
5.2.3 Modelování trendu (hrubá mzda)	54
5.2.4 Modelování sezónnosti (hrubá mzda).....	64
5.3 Box-Jenkinsova metodologie	66
5.3.1 ARIMA (ubytování)	67

5.3.2	ARIMA (mzdy).....	75
5.3.3	SARIMA (mzdy)	79
6	Extrapolace časových řad.....	83
6.1	Extrapolace časové řady „tržby za ubytování“	83
6.1.1	Klasické přístupy – Kvadratický trend	85
6.1.2	Adaptivní přístupy – Trojité exponenciální vyrovnávání	86
6.1.3	ARIMA (1,1,1)	87
6.1.4	ARIMA (1,1,0)	88
6.1.5	SARIMA (1,1,0)	89
6.1.6	SARIMA (0,1,1)	90
6.2	Extrapolace časové řady „hrubá mzda“	92
6.2.1	Klasické přístupy – Kvadratický trend	93
6.2.2	Adaptivní přístupy – Trojité exponenciální vyrovnávání	94
6.2.3	ARIMA (1,1,1)	96
6.2.4	ARIMA (0,1,1)	97
6.2.5	SARIMA (1,1,0)	98
6.2.6	SARIMA (0,1,1)	99
7	Porovnání výsledků	101
7.1	Časová řada „tržby z ubytování“	101
7.2	Časová řada „hrubá mzda“	102
	Závěr	103
	Literatura a další zdroje	105
	Seznam použitých tabulek, grafů a obrázků	107
	Abstrakt.....	111
	Abstract.....	112

Úvod

S časovými řadami, jakožto s hodnotami určité veličiny seřazenými do chronologického sledu se lze setkat téměř ve všech oblastech lidského života. Jednou z nejběžnějších oblastí, kde se časové řady vyskytují, je ekonomie, ve které se časové řady mohou dělit na makroekonomické (státní) a mikroekonomické (podnikové). Důležitost časových řad pro dnešní svět tkví v důkladné analýze minulých i současných dat, která vede k odpovědnému rozhodování. Navíc podnik, který umí předvídat vývoj určitých ekonomických ukazatelů získává konkurenční výhodu.

Práce se zabývá analyzováním vývoje vybraných časových řad z ČSÚ a jejich následným porovnáním se skutečnými daty v roce 2018. Ze statistického úřadu jsou vybrány časové řady zachycující „tržby z ubytování mezi roky 2000 až 2017“ a „vývoj hrubé mzdy mezi roky 2000 až 2017“. Cílem této diplomové práce je aplikace vybraných metod na reálná data a jejich následné porovnání v roce 2018, které určí, jaká vybraná metoda je pro analyzování vybraných časových řad nejvhodnější.

Diplomová práce je rozdělena do sedmi kapitol. V první kapitole jsou vymezeny základní obecné pojmy k časovým řadám. Ve druhé kapitole je popsán dekompoziční přístup. Třetí kapitola popisuje vybrané metody Box-Jenkins. Ve čtvrté kapitole jsou uvedeny postupy extrapolace pomocí dekompozičního a Box-Jenkinsova přístupu. První čtyři kapitoly jsou zpracovány na základě odborné literatury. Pátá kapitola je zaměřena na aplikaci popsaných metod na vybrané časové řady. Pomocí dvou přístupů zmíněných v druhé a třetí kapitole jsou analyzovány vývoje, trendy a sezónnosti časových řad. V šesté kapitole jsou pomocí nejvhodnějších modelů predikovány budoucí hodnoty na dva roky dopředu, kdy predikované hodnoty v roce 2018 jsou porovnány se skutečnými daty. V poslední kapitole je uvedeno porovnání výsledků a jsou určeny nejvhodnější modely pro predikci časových řad.

Pro analýzu dat a provedení všech potřebných výpočtů je použit software Excel a jeho doplňky NumXL a XLSTAT.

1 Časové řady

Časovou řadou lze rozumět určité množství dat zaznamenané v čase. Jinými slovy se jedná o empirická data za určité období seřazena do chronologického sledu. Určitým obdobím se myslí rovnoměrné časové intervaly, kterými jsou například měsíce, čtvrtletí, pololetí či roky. Hančlová a Tvrdý (2003, s. 3) popisují časovou řadu a významnost rovnoměrných intervalů následovně: „*Časovou řadou rozumíme posloupnost hodnot ukazatelů měřených v určitých časových intervalech. Tyto intervaly jsou zpravidla rovnoměrné (ekvidistantní), a proto je můžeme zapsat následujícím způsobem:*

$$y_1, y_2, \dots, y_n \text{ neboli } y_t, t = 1, 2, \dots, n,$$

kde: y_t ... značí analyzovaný ukazatel

t ... časová proměnná s celkovým počtem pozorování n .“

Co se týče historie časových řad, první zkoumání časových řad bylo uskutečněno již dříve (starověké Řecko) a to v oblasti astronomie. Dnes se časové řady využívají téměř ve všech oborech a oblastech, např.: v demografii se může jednat o počet obyvatel v určité zemi, v meteorologii o řadu naměřených teplot, v ekonomii o průměrnou hrubou mzdu či hrubý domácí produkt atd. Pokud jsou empirickými daty hodnoty některého ekonomického ukazatele, zpravidla jsou označeny jako ekonomické časové řady. Nejen ekonomické časové řady lze rozdělit na základě různých hledisek na jednotlivé typy, které jsou uvedeny v následující kapitole (Arlt, Arltová, 2009).

Typy časových řad

Z hlediska potřeb této práce lze časové řady dělit dle následujících hledisek:

- dle rozhodného časového hlediska
 - intervalové časové řady – sledování dat v určitém časovém intervalu, např. k jednomu měsíci (měsíční příjmy, výdaje), ke čtvrtletí (kvartální příjmy, výdaje) nebo k jednomu roku (dovoz České republiky mezi lety 2015 až 2018).
 - okamžikové časové řady – sledování dat v určitém časovém okamžiku, např. inventura k poslednímu dni v měsíci (stav materiálu, zásob) nebo počet zaměstnanců k poslednímu dni v měsíci v roce 2018.

- dle periodicity pozorování
 - krátkodobé časové řady – zaznamenávání dat v periodách maximálně do jednoho roku, tzn. denně, týdně, měsíčně, čtvrtletně, pololetně.
 - dlouhodobé časové řady – zaznamenávání dat v delších periodách, tzn. ročně, po desetiletích.

Časové řady lze rozdělit také na **stochastické** (statistické) a **deterministické**. Stochastické časové řady v sobě obsahují jakousi nejistotu nebo náhodu. Oproti tomu deterministické časové řady lze kvantifikovat a jejich průběh lze vyjádřit matematickým vzorcem (Cipra, 2008).

1.1 Význam analýzy časových řad

Cílem analýzy časových řad je zkonstruovat model, který dokáže co nejlépe prezentovat chování dané časové řady a zjistit, zda se v ní vyskytují nějaké trendy, sezónnosti či cykly. Empirické hodnoty mohou být ovlivňovány spoustou faktorů, díky konstrukci modelu, je možné odhalit o které faktory se jedná a o které nikoli. Pokud je následně vybrán vhodný statistický test, je možné se tyto časové řady pokusit optimalizovat. Z časových řad lze také analyzovat možné příčiny změn chování v minulosti a v neposlední řadě mohou pomoci při predikci budoucího stavu a vývoje. Předpovědi se můžou zdát jako výhodný pohled do budoucnosti, např. předpověď příjmu či předpověď objemu produkce společnosti.

Analýzu časových řad lze v dnešní době úspěšně porovnat s popisem dynamických systémů, z toho důvodu jsou tyto analýzy hojně využívány ve statistice. Vytvořený model ovšem není schopen předpovědět reálné budoucí hodnoty, pouze může vygenerovat odhad, se kterým lze dále pracovat a v budoucnu porovnat, zda byl odhad správný či mylný. Aby byl odhad co nejpřesnější je důležité se vyhnout následujícím potenciálním problémům (Cipra, 1986).

1.2 Potenciální problémy při zpracovávání analýzy

Při zpracovávání analýzy časových řad se lze setkat s problémy, při kterých může dojít k narušení celé analýzy. Je důležité tyto problémy prozkoumat a snažit se jim vyhnout. Problémy mohou být následující:

- **Problémy s kalendářem** – při analýze časových řad lze narazit na problémy, které souvisejí s variabilním počtem dnů v měsíci či počtem pracovních dnů v měsíci. Aby se předešlo těmto problémům, zavedla se metoda očištění tzv. standardní měsíc, který má 30 dní.
- **Problémy s délkou časové řady** – délka časové řady vyjadřuje celkový počet pozorování. Příliš krátká časová řada neumožňuje analyzovat konkrétní časovou řadu (např. model Box-Jenkins vyžaduje časovou řadu o minimálně 50 pozorováních). Stejně tak příliš dlouhá časová řada, u které hrozí změna chování modelu, jelikož by se vývoj dat v dlouhém období mohl výrazně změnit.
- **Problémy s volbou časových bodů pozorování** – při zpracovávání analýz se pracuje především s diskrétními hodnotami. Jelikož velké množství veličin je spojitých, je zapotřebí je nějakým způsobem převést na diskrétní hodnoty. Tzn. diskretizaci spojitě časové řady, např. teplota v 11 hodin dopoledne na určitém místě. Zásadní je dobře zvolený způsob této diskretizace, jelikož může vést ke zkresleným výsledkům (Arlt, Arltová, 2009).

Po přezkoumání nebo eliminaci těchto výše zmíněných potenciálních problémů, lze časové řady začít analyzovat, nejprve je však zapotřebí uvést některé základní charakteristiky časových řad.

1.3 Základní charakteristiky časových řad

Při analýze časových řad je zapotřebí, zjistit charakter procesu, ten vypovídá o tom, co vůbec řada znamená a jaký vývoj reprezentuje. Pro přehlednost se při zpracovávání analýzy využívají především **grafy**, které přehledně ukazují vývoj časové řady nebo **popisné charakteristiky**.

Popisné charakteristiky

Obecně existuje velké množství popisných charakteristik, zde jsou však uvedeny jen charakteristiky, které jsou využity v práci (Hindls, 2007):

- **Charakteristiky polohy**

- **Chronologický průměr**

Diskrétní časové řady se vypočítají jako chronologický průměr. Pro výpočet chronologického průměru je zapotřebí, aby délka mezi jednotlivými časovými okamžiky byla identická. Vzorec pro výpočet **chronologického průměru**:

$$\bar{y} = \frac{\frac{y_1+y_2}{2} + \frac{y_2+y_3}{2} + \dots + \frac{y_{t-1}+y_t}{2}}{t-1}. \quad (1)$$

Pokud délky mezi jednotlivými časovými řadami nejsou identické, použije se vzorec pro výpočet **váženého chronologického průměru**:

$$\bar{y} = \frac{\frac{y_1+y_2}{2} \times d_1 + \frac{y_2+y_3}{2} \times d_2 + \dots + \frac{y_{t-1}+y_t}{2} \times d_{t-1}}{d_1+d_2+\dots+d_{t-1}}, \quad (2)$$

kde: y_t hodnoty ukazatelů v časovém okamžiku t ,
 d_t jednotlivé délky intervalů.

- **Medián**

Jedná se o prostřední hodnotu výběru, při kterém jsou hodnoty seřazeny dle velikosti. Neboli polovina hodnot je menší nebo rovna mediánu a druhá polovina je větší nebo rovna mediánu. V případě, že se tato hodnota nedá určit přesně, počítá se medián z dvou prostředních hodnot jako jejich průměr.

- **Charakteristiky variability**

- **Rozpětí**

Jedná se o rozdíl mezi maximální a minimální hodnotou.

- **Rozptyl**

Rozptyl je vypočten jako součet kvadratických odchylek od průměru, dělený počtem hodnot ve výběru zmenšeným o 1. Tento postup je popsán vztahem:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (3)$$

- **Směrodatná odchylka**

Tento ukazatel variability je počítán jako odmocnina z rozptylu. Výpočet směrodatné odchylky je dán vztahem:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}. \quad (4)$$

- **Charakteristika šikmosti**

Koeficient šikmosti

Koeficient šikmosti udává, jakým způsobem jsou hodnoty rozloženy kolem zvoleného středu, zda jsou souměrně rozloženy nebo zda jsou zešikmeny do strany. Koeficient lze vypočítat dle vztahu:

$$s_m = \frac{1}{n} \sum_{i=1}^n \frac{(y_i - \bar{y})^3}{s^3} \quad (5)$$

- **Charakteristika špičatosti**

Koeficient špičatosti také udává, jakým způsobem jsou hodnoty rozloženy kolem zvoleného středu, nezkuamají se však horizontálně, nýbrž vertikálně. Čím je rozdělení špičatější, tím více hodnot je kolem zvoleného středu. Naopak, čím je rozdělení plošší, tím jsou hodnoty více vychýleny od středu. Koeficient šikmosti je dán tímto vztahem:

$$k_m = \frac{1}{n} \sum_{i=1}^n \frac{(y_i - \bar{y})^4}{s^4}. \quad (6)$$

1.4 Základní přístupy k analýze časových řad

Jako obecný způsob modelování časových řad se považuje jednorozměrný model, který je dán vztahem:

$$y_t = f(t, \varepsilon_t), \quad (7)$$

kde: y_t ... hodnota modelovaného ukazatele v čase t

ε_t ... náhodná složka v čase t .

Kromě jednorozměrných modelů se lze setkat i s modely vícerozměrnými, ty jsou charakteristické tím, že vývoj časové řady není ovlivněn pouze časovým faktorem, ale může na něj působit mnoho jiných faktorů.

Vícerozměrný model lze popsat tímto vztahem:

$$y_t = f(t, g_1, g_2, g_3, \dots, g_n, \varepsilon_t), \quad (8)$$

kde: $g_1, g_2, g_3, \dots, g_n$... faktory, které ovlivňují vývoj ukazatele y .

Analýzu časových řad lze provádět různými metodami, volba jednotlivých metod záleží na různých okolnostech. Jednou z nich je **typ časové řady**, typem časové řady se myslí její charakter, velikost intervalu měření a rozdělení náhodných veličin. Některé metody jsou totiž vhodné jen pro časové řady určitého typu (např. nekonstruuje se časová řada o 20 údajích dle metody Box-Jenkins). Dále záleží také na **znalosti a zkušenosti osoby**, která provádí analýzu. Při analýze by osoba měla efektivně využívat IT vybavení a softwarovou podporu. Právě osoba, která analýzu provádí, má za úkol nejprve zanést do grafu hodnoty časových řad a z výsledku se rozhodne o dalším postupu. Z grafu analytik zjišťuje, zda časová řada obsahuje nějaké trendové, sezónní či cyklické složky.

Po rozhodnutí může využít jedné z metod modelování časových řad. Nejčastěji jsou použity tyto metody:

- dekompozice časové řady,
- Box-Jenkinsova metodologie,
- spektrální analýza,
- neuronové sítě.

Pro účely této práce budou využity pouze metody dekompozice časové řady a Box-Jenkinsova metodologie, jelikož spektrální analýza zkoumá jen cyklickou složku a cílem je analýza vývoje a predikce. Neuronové sítě jsou výpočetním modelem používaným v umělé inteligenci, který přebírá chování pozorovaných struktur. Spektrální analýzou a neuronovými sítěmi se tato práce nezabývá (STATSOFT 2013, Zichová 2011).

2 Dekompozice časové řady

Dekompozice neboli rozklad časové řady na jednotlivé složky vychází z předpokladu, že po rozkladu se lépe identifikuje chování a mechanismus časové řady. Při dekompozici lze časovou řadu rozložit na následující čtyři složky:

- trendová složka (T_t)
- sezónní složka (S_t)
- cyklická složka (C_t)
- náhodná složka (ε_t)

Trendová složka prezentuje dlouhodobé změny v chování časové řady (např. dlouhodobý růst či stagnaci). Naopak **sezónní složka** prezentuje periodické změny v časové řadě, které se uskuteční zpravidla každé období. Může se jednat o změnu ročních období či o lidské zvyky (např. zvýšený prodej kaprů na Vánoce). **Cyklická složka** je vyjádřením nějaké dlouhodobé fluktuace kolem trendu. Může se jednat o různé poklesy či růsty hodnot časové řady. Jedná se o složku, která se pravidelně v časové řadě opakuje, avšak její charakter se může v průběhu doby změnit. Je zapotřebí analyzovat delší časové řady, protože u těch kratších by se cyklická složka nemusela vyskytovat. U některých časových řad je cyklická složka zahrnuta do trendu jako součást. **Náhodná složka** neboli reziduální složka nepatří do systematických složek časových řad, je tvořena pouze náhodnými výkyvy. Vyskytují se zde veškeré chyby měření a vlivy, které působí na časovou řadu a nelze je zařadit do systematické části.

Cílem metody dekompozice je identifikace jednotlivých složek, přičemž časové řady nemusí obsahovat všechny tyto systematické složky.

Důvody použití dekompozice časové řady:

- analýza jednotlivých složek může pomoci odhalit zákonitosti v časové řadě,
- z časové řady lze eliminovat sezónnosti, díky tomu lze porovnat trendy několika časových řad najednou,
- z časové řady lze eliminovat trendy, díky tomu lze lépe modelovat sezónnost,
- lepší predikce jednotlivých složek i vývoje celé časové řady, a to tím způsobem, že se jednotlivé složky sečtou nebo vynásobí, podle toho, o jaký typ dekompozice se jedná (Klufová, 2012).

Jednotlivé typy dekompozice jsou rozděleny podle vztahu a funkcí mezi jednotlivými složkami. Tyto relace lze popsat třemi základními modely (Cipra, 1986):

a) Aditivní model

U aditivního modelu se jednotlivé složky časové řady (trendová, sezónní, cyklická a náhodná složka) sčítají:

$$y_t = T_t + S_t + C_t + \varepsilon_t. \quad (9)$$

b) Multiplikativní model

U multiplikativního modelu se jednotlivé složky časové řady násobí:

$$y_t = T_t S_t C_t \varepsilon_t. \quad (10)$$

c) Smíšený model

Smíšený model je vypočten jako kombinace součtu a součinu jednotlivých složek, např.:

$$y_t = T_t S_t C_t + \varepsilon_t. \quad (11)$$

Následující část se zabývá popisem a identifikací jednotlivých složek.

2.1 Modelování trendu

Modelování trendu a jeho analýza je jednou z nejpodstatnějších částí při dekompozici časové řady. Tvar trendu a jeho charakter mohou mít různou podobu, může být rostoucí či klesající, strmý či plochý, může se i měnit v průběhu doby. Existují dva základní přístupy k eliminaci trendové složky:

- klasické přístupy,
- adaptivní přístupy.

2.1.1 Klasické přístupy

Trendová složka je u klasického přístupu popsána jako jednoduché matematické funkce. Odhad parametrů funkcí se většinou vypočte metodou nejmenších čtverců. Mezi základní matematické funkce, které se často používají patří:

- **Lineární trend**

$$T_t = \beta_0 + \beta_1 t, \quad (12)$$

kde: $t = 1, 2, 3, \dots, n$.

- **Kvadratický trend**

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2, \quad (13)$$

kde: $t = 1, 2, 3, \dots, n$.

- **Exponenciální trend**

$$T_t = \beta_0 \beta_1^t, \quad (14)$$

kde: $\beta_1 > 0$

$t = 1, 2, 3, \dots, n$.

- **Logistický trend**

$$T_t = \frac{\omega}{1 + \beta_0 \beta_1^t}, \quad (15)$$

kde: $\omega > 0$,

$\beta_0 > 1$,

$0 < \beta_1 < 1$,

$t = 1, 2, 3, \dots, n$.

Odhad parametrů u metody logistické lze provádět např. metodou částečných součtů, metodou vybraných bodů nebo metodou diferenčních odhadů parametrů. (Hindls, 2002).

Pro výběr nejvhodnějšího modelu trendové složky z klasických přístupů existují jednoduché orientační testy, které pomohou nejvhodnější variantu identifikovat. Orientační testy vycházejí z charakteristických vlastností jednotlivých křivek.

Tabulka 1 - Přehled orientačních testů

Trend	Orientační test
lineární	první difERENCE jsou přibližně konstantní
kvadratický	druhé difERENCE jsou přibližně konstantní
exponenciální	podíly sousedních hodnot (resp. první difERENCE logaritmů ln) jsou přibližně konstantní

Zdroj: Cipra, 2008

2.1.2 Adaptivní přístupy

Klasické přístupy spočívají především v neměnnosti parametrů v čase, oproti tomu, adaptivní přístupy konstruují trendovou složku jako model, ve kterém se v průběhu času mění jednotlivé parametry. Tato změna parametrů je zapříčiněna předpokladem, že krátké úseky časových řad lze vyrovnat pomocí matematických funkcí.

Mezi nejznámější adaptivní přístupy patří např. metoda klouzavých průměrů a metoda exponenciálního vyrovnávání.

- **Metoda klouzavých průměrů**

Existují různé druhy klouzavých průměrů (prostý, centrovaný, klouzavý), které dokáží částečně odstranit vlivy náhodných veličin a očistit tak časovou řadu. Pro účely této práce budou použity jen prosté a vážené klouzavé průměry, které slouží k vyrovnání sezónnosti.

Vyrovnávání pomocí klouzavých průměrů se provádí tak, že empirická pozorování jsou nahrazena vypočtenými průměry těchto pozorování. Klouzavé průměry proto, že při výpočtu průměrů se vždy „klouže“ o jedno empirické pozorování dopředu, ale počet těchto průměrů je vždy stejný, proto nejstarší průměr z minulého měření je vypuštěn. Neboli klouzavé průměry jsou vypočteny z kratších úseků o délce $d = 2m + 1$ hodnot, na které byla původní časová řada rozdělena.

Pro výpočet výsledných hodnot v podobě **prostého klouzavého průměru** se uvádí následující vztah:

$$y_t = \frac{y_{t-m} + y_{t-m+1} + \dots + y_t + \dots + y_{t+m-1} + y_{t+m}}{2m+1}, \quad (16)$$

U krátkodobých časových řad se lze setkat se situací, kde střední body klouzavých částí nejsou celá čísla a nelze je přímo spojit s původním pozorováním. Tento problém je řešen pomocí **centrovaných klouzavých průměrů**. Centrované klouzavé průměry lze použít i k vyrovnání sezónnosti (Cipra, 1986).

Kromě výše zmíněných typů klouzavých průměrů existují také vážené klouzavé průměry.

Vážené klouzavé průměry se od předchozího typu liší tím, že neklade všem pozorováním stejnou váhu. Každá perioda je vynásobena váhou odpovídající danému

čtvrtletí. Vyrovnané hodnoty pomocí vážených klouzavých průměrů pro **parabolický** trend jsou spočteny následovně:

$$\hat{y}_t = \sum_{i=-p}^p w_i y_{t,i} \quad (17)$$

kde: y_t ... hodnota časové řady v určitém okamžiku,

w ... přiřazená váha ke každé periodě

$$w_i = \frac{3}{4m(m^2-4)} (3m^2 - 7 - 20i^2), \quad (18)$$

kde: $i = -p, \dots, -1, 0, 1, \dots, p$.

Vzorec (17) je nazýván váženým klouzavým průměrem v čase t . Hodnoty w_i vypočtené dle vzorce (18) jsou váhy. Vypočtené váhy musí splňovat podmínku:

$$\sum_{i=-p}^p w_i = 1 \text{ a } w_i = w_{-i} \quad (19)$$

Postup výpočtu vah (18) je určen pro čtyři období, tedy pro čtvrtletní časové řady. Jelikož se jedná o $(m+1)$ členů, za m ve vzorci pro výpočet vah je dosazeno číslo $4+1$ (např. pro měsíční by za m bylo dosazeno $12+1$). Za i jsou ve vzorci dosazeny hodnoty kolem nuly, tedy kolem vyrovnané hodnoty. V tomto případě jsou uvedeny hodnoty $-2, -1, 0, 1, 2$. V případě, že je metoda klouzavých průměrů pro vyrovnání nedostačující, lze použít následující metodu (Montgomery a kol., 2008).

- **Metoda exponenciálního vyrovnávání**

Tato metoda vyrovnávání je založena na aplikaci metody vážených nejmenších čtverců na všechna dostupná pozorování dané časové řady s předpokladem, že váhy jednotlivých pozorování w se směrem do minulosti exponenciálně zmenšují, to znamená, že pozorování, které je blíže časovému okamžiku n má větší vliv na budoucí vývoj než pozorování z delší minulosti.

U tohoto typu vyrovnávání má časová řada následující tvar:

$$w_t = \alpha^t, \quad (20)$$

kde: α ... vyrovnávací konstanta, $0 < \alpha < 1$,

t ... pozorování od časového okamžiku n , $t = 1, 2, 3, \dots, n$.

Parametry modelu lze u modelu exponenciálního vyrovnávání získat také metodou nejmenších čtverců, která je popsána následujícím vztahem:

$$\sum_{k=0}^{n-1} (\hat{y} - y_t)^2 \alpha^t \rightarrow \min, \quad (21)$$

kde: y_t ... hodnota trendové složky v čase t ,

\hat{y} ... hodnota skutečných dat.

Důležité je pozorovat, jak se trendová složka chová v krátkých úsecích časové řady. Pokud je tato složka konstantní, používá se tzv. **jednoduché exponenciální vyrovnávání**. V případě, že by tyto krátké úseky měly lineární trend, použilo by se **dvojitě exponenciální vyrovnávání**. V posledním případě, kdyby trend vykazoval známky kvadratického tvaru, použilo by se **trojitě exponenciální vyrovnávání**. Cílem exponenciálního vyrovnávání je nalezení odhadu parametrů $\beta_0, \beta_1, \beta_2, \dots, \beta_n$, kde vždy záleží na zvoleném typu. Hodnoty odhadů závisí na časovém okamžiku jejich sestavení, jelikož jsou počítány ze všech dostupných dat. Pokud by se přidaly nové pozorování, bylo by zapotřebí znovu odhadovat parametry, což by bylo zbytečně náročné. Proto se využívají „*vyrovnávací statistiky*“, které počítají pouze s počátečními odhady parametrů.

Po seznámení s různými druhy exponenciálního vyrovnávání, budou uvedeny základní vzorce pro dvojitě a trojitě exponenciální vyrovnávání, jelikož jednoduché vyrovnávání nebude v práci použito z důvodu absence konstantní trendové složky. Vzorce budou uvedeny bez odvození výpočtů, případné odvození výpočtu lze nalézt v literatuře, např. Cípra (1986).

Dvojitě exponenciální vyrovnání

Dvojitě exponenciální vyrovnání, někdy označováno jako Brownova metoda, se používá, při předpokladu, že časové řady jsou v krátkých úsecích lineární, vzorec (12):

$$T_t = \beta_0 + \beta_1 t,$$

Počáteční odhad parametrů β_0 a β_1 je regresním proložením přímky prvními pozorováními. Tyto odhady slouží k výpočtu počáteční hodnoty vyrovnávacích statistik:

$$S_0 = b_0(0) - \frac{\alpha}{1-\alpha} b_1(0), \quad (22)$$

$$S_0^{[2]} = b_0(0) - \frac{2\alpha}{1-\alpha} b_1(0). \quad (23)$$

Z těchto vzorců lze vidět, že $b_0(0)$ a $b_1(0)$ jsou počátečními hodnotami pro parametry β_0 a β_1 , kde α představuje vyrovnávací konstantu, jejíž hodnota byla zvolena subjektivně. Vyrovnávací statistiky mají tedy následující podobu:

$$S_t = (1 - \alpha)y_t + \alpha S_{t-1}, \quad (24)$$

$$S_t^{[2]} = (1 - \alpha)S_t + \alpha S_{t-1}^{[2]}, \quad (25)$$

kde: $t = 0, 1, 2, \dots, n$,

S_{n-t} ... jednoduchá vyrovnávací statistika

$S_{n-t}^{[2]}$... dvojitá vyrovnávací statistika.

Konečný výpočet vyrovnané hodnoty je ve tvaru: jednoduchá vyrovnávací statistika násobena dvěma, mínus dvojitá vyrovnávací statistika.

Trojité exponenciální vyrovnání

Trendová složka v krátkých úsecích vykazuje známky kvadratického tvaru (13):

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2,$$

Počáteční odhad parametrů β_0 , β_1 a β_2 je regresním odhadem, který je získán výpočtem metodou nejmenších čtverců z prvotních hodnot, ze kterých jsou vypočteny následující vyrovnávací statistiky:

$$S_0 = b_0(0) - \frac{\alpha}{1-\alpha} b_1(0) + \frac{\alpha(1+\alpha)}{2(1-\alpha)^2} b_2(0), \quad (26)$$

$$S_0^{[2]} = b_0(0) - \frac{2\alpha}{1-\alpha} b_1(0) + \frac{2\alpha(1+2\alpha)}{2(1-\alpha)^2} b_2(0), \quad (27)$$

$$S_0^{[3]} = b_0(0) - \frac{3\alpha}{1-\alpha} b_1(0) + \frac{3\alpha(1+3\alpha)}{2(1-\alpha)^2} b_2(0). \quad (28)$$

Stejně jako u dvojitého exponenciálního vyrovnání, hodnoty $b_0(0)$, $b_1(0)$ a $b_2(0)$ označují počáteční hodnoty pro parametry β_0 , β_1 a β_2 a stejně tak je

subjektivně zvolena vyrovnávací konstanta α . I podle názvu lze určit, že tato metoda používá kromě jednoduché a dvojitě vyrovnávací statistiky také trojitou, která má tvar:

$$S_t^{[3]} = (1 - \alpha)S_t^{[2]} + \alpha S_{t-1}^{[3]}, \quad (29)$$

kde: $t = 0, 1, 2, \dots, n - 1$.

Výpočet vyrovnané hodnoty je znázorněn vztahem:

$$\hat{y}_t = 3S_t - 3S_t^{[2]} + S_t^{[3]}, \quad (30)$$

kde: $t = 0, 1, 2, \dots, n$.

Pro výsledky všech výše uvedených metod je potřeba vytvořit orientační testy či vypočítat statistická kritéria, aby se zjistilo, který model je nejvhodnější.

Orientační testy u klasických přístupů byly založené na základě analýzy grafu časové řady. V tomto případě vzniká riziko, že každý, kdo analýzu provádí, má své subjektivní postoje a může dojít k odlišnému výběru vhodného trendu. K částečnému či úplnému odstranění subjektivity v rozhodování, u obou přístupů, existují statistická kritéria:

Tabulka 2 - Statistická kritéria

Statistická kritéria	Vzorec
střední chyba odhadu „Mean Error“	$ME = \frac{\sum(y_t - T_t)}{n}$
reziduální součet čtverců „Residual Sum of Squares“	$RSS = \sum_{t=1}^n (y_t - T_t)^2$
střední čtvercová chyba odhadu „Mean Squared Error“	$MSE = \frac{\sum(y_t - T_t)^2}{n}$
střední absolutní chyba odhadu „Mean Absolute Error“	$MAE = \frac{\sum y_t - T_t }{n}$
střední procentní chyba odhadu „Mean Percentage Error“	$MPE = \sum_{t=1}^n \left(\frac{y_t - T_t}{y_t} \right) \frac{100}{n}$
Střední absolutní procentní chyba odhadu „Mean Absolute Percentage Error“	$MAPE = \sum_{t=1}^n \left(\frac{ y_t - T_t }{y_t} \right) \frac{100}{n}$

Zdroj: Hindls, 2000

Tato výše uvedená statistická kritéria jsou ukazatelem, který znázorňuje, jak moc je model podobný skutečnosti. Výpočty vycházejí z porovnání teoretických hodnot s těmi empirickými (součet čtverců nebo průměrů odchylek). Nejvhodnějším modelem je ten, který má nejmenší hodnotu zvoleného kritéria. Pokud je zapotřebí získat nejvhodnější model pro predikci, nejčastěji se používá kritérium střední čtvercové chyby odhadu (M.S.E.) a kritérium střední absolutní procentní chyby odhadu (M.A.P.E.). Tato kritéria se používají u klasických i adaptivních přístupů (Cipra, 1986).

2.2 Modelování sezónnosti

Již v předchozích kapitolách je zmínka o existenci periodicky se vyskytujících výkyvů v časové řadě. Sezónní složka znamená periodicky se opakující obousměrné výkyvy řady od trendu. Tyto výkyvy obvykle mají svou periodicitu jeden rok či kratší. U ekonomických časových řad se nejčastěji lze setkat s periodicitou měsíční či čtvrtletní, málokdy je uvedena periodicitu půlroční. Výkyvy jsou způsobené důsledkem přímých i nepřímých příčin, které se každý rok opakují a většinou souvisejí se střídáním ročních období. Lze uvést na příkladu, že lidé kupují v obchodních domech více šály a čepice na podzim a v zimě, než je tomu například v létě.

Oproti předchozím vzorcům, zde bude užitá jiná symbolika, například dříve užívaná posloupnost časové proměnné $t = 1, 2, 3, \dots, n$, nyní tvoří posloupnost sezón v jednotlivých obdobích. Posloupnost časových proměnných bude značena jako $i = 1, 2, 3, \dots, m$ let. Každý tento rok obsahuje r sezón, které se značí j . Navíc zde nebude uvedena cyklická složka, jelikož ji lze pro účely této práce považovat za součást reziduální složky. Následující vztah popisuje trendovou složku.

$$t = T_{ij} = (i - 1)r + j, \quad (31)$$

kde: $i = 1, 2, 3, \dots, m$,

$$j = 1, 2, 3, \dots, r.$$

Model v aditivním tvaru dekompozice by pak vypadal následovně:

$$y_{ij} = T_{ij} + S_{ij} + \varepsilon_{ij}, \quad (32)$$

kde: $i = 1, 2, 3, \dots, m$,

$$j = 1, 2, 3, \dots, r,$$

y_{ij} ... pozorovaný ukazatel v roce i a v sezóně j .

Pokud se tedy předpokládá existence sezónnosti v časové řadě, je zapotřebí tyto výkyvy nejprve kvantifikovat. Následujícím krokem je provedení tzv. sezónního očištění, které má za úkol očistit časovou řadu od sezónnosti, jelikož ta může zkreslovat dynamiku ekonomických jevů. Výsledkem provedení sezónního očištění je umožnění provádět kvalifikovaná srovnání vývoje v jednotlivých obdobích (Hindls 2000, Box a kol. 2008).

2.2.1 Kvantifikace sezónních výkyvů

Sezónnost v aditivním tvaru dekompozice „model konstantní sezónnosti“

U aditivního tvaru lze považovat sezónnost za pravidelnou, v tom případě jsou odchylky sezónních faktorů pro jednotlivé roky konstantní. Model aditivní dekompozice je popsán vztahem (32).

Ačkoli je již známá trendová složka, lze ji očistit od časové řady:

$$y_{ij} - T_{ij} = S_{ij} + \varepsilon_{ij}, \quad (33)$$

kde: $i = 1, 2, 3, \dots, m,$

$j = 1, 2, 3, \dots, r.$

Následující výpočet je založen na vyjádření sezónního faktoru, z již očištěné časové řady od trendu, jako aritmetický průměr ze všech dat, které jsou v j-tém období.

$$I_j = \frac{1}{m} \sum_{i=1}^m (y_{ij} - T_{ij}), \quad (34)$$

kde: $j = 1, 2, 3, \dots, r.$

K pokračování ve výpočtu, je důležité si uvědomit, že se sezónní odchylky v rámci jednoho roku kompenzují, proto je předpoklad, že se jejich roční součet musí rovnat nule neboli výše uvedené výpočty by měly splňovat normalizační pravidlo:

$$\sum_{j=1}^r I_j = 0, \quad (35)$$

kde: $j = 1, 2, 3, \dots, r.$

V případě, že časová řada nevyhovuje normalizačnímu pravidlu, musejí se hodnoty vycentrovat, a to odečtením aritmetického průměru:

$$I_{j(\text{centr.})} = I_j - \bar{I}. \quad (36)$$

Sezónnost v multiplikačním tvaru dekompozice „model proporcionální sezónnosti“

U multiplikačního tvaru je sezónní složka ovlivněna především vývojovými změnami v povaze trendové složky, protože u tohoto tvaru není sezónnost jako samostatná složka, jako je tomu u aditivního tvaru, nýbrž je násobkem trendu. Multiplikační tvar je dán tímto vztahem:

$$y_{ij} = T_{ij}S_{ij}\varepsilon_{ij}, \quad (37)$$

kde: $i = 1, 2, 3, \dots, m,$

$j = 1, 2, 3, \dots, r.$

Dalším krokem, stejně jako u aditivního tvaru, je očištění od trendové složky:

$$\frac{y_{ij}}{T_{ij}} = S_{ij}\varepsilon_{ij}. \quad (38)$$

Odhad sezónních faktorů je proveden obdobně jako u aditivní dekompozice, pomocí vztahu:

$$I_j = \frac{1}{m} \sum_{i=1}^m \frac{y_{ij}}{T_{ij}}, \quad (39)$$

kde: $i = 1, 2, 3, \dots, m,$

$j = 1, 2, 3, \dots, r.$

Sezónní odchylky se v rámci jednoho roku kompenzují, proto je předpoklad, že se jejich roční součet musí rovnat nule. Výše uvedené výpočty by tedy měly splňovat normalizační pravidlo, které je dáno vztahem (35), tedy:

$$\sum_{j=1}^r I_j = 0,$$

Po kvantifikaci výkyvů by mělo následovat jejich očištění od trendové složky, které je uvedeno níže.

2.2.2 Sezónní očišťování

Sezónní očišťování má za úkol očistit časovou řadu od sezónnosti, protože díky ní může být zkreslena dynamika ekonomických jevů. V předchozích případech byla časová řada

očišťována od trendové složky, v této části je uvedeno, jak lze očistit časovou řadu od sezónní složky (Andersen, 2009).

Algoritmus sezónního očišťování je rozdělen do tří základních kroků:

1. výpočet klouzavých průměrů,
2. kvantifikace sezónních faktorů,
3. a očištění od sezónní složky.

První dvě části algoritmu jsou již známy z předešlé části. Poslední část je vypočtena obvyklým způsobem, který uvádí například Hindls a kol. (1997, s. 158): *„od hodnot původní časové řady buď odečteme vždy příslušný rozdílový sezónní faktor (předpokládá-li se model konstantní sezónnosti), anebo hodnoty původní časové řady vydělíme vždy příslušným indexním sezónním faktorem (v modelu proporcionální sezónnosti)“*.

Sezónní složku lze eliminovat také jinými způsoby, a to Hendersonovými filtry, Wintersovo exponenciálním vyrovňáváním, regresními metodami atd. (Cipra, 1986). Pro účely této práce však postačí metody dekompozice časové řady uvedené v kapitole 2 a vybrané metody z metodologie Box-Jenkins, které jsou uvedeny v následující kapitole.

3 Box-Jenkinsova metodologie

Dekompozice časové řady kladla velký důraz na práci se systematickými složkami časové řady a brala jednotlivá pozorování jako navzájem nekorelovaná. Naproti tomu, Box-Jenkinsova metodologie klade důraz na modelování časové řady pomocí **reziduální složky**, která je v tomto případě **korelovaná**. Je zřejmé, že nejvíce pozornosti se věnuje reziduální složce, to ale neznamená, že se nedají modelovat řady s významným trendovým či sezónním charakterem. Pro tyto časové řady jsou použity tzv. trendové modely, které budou uvedeny v kapitole 3.3 (Forbelská, 2009).

Výhody metodologie: díky své flexibilitě a své poměrně rychlé adaptaci na změnu, přináší tato metoda nejlepší výsledky.

Nevýhody metodologie: analyzovaná časová řada musí být dostatečně dlouhá a je zde poměrně špatná interpretace výsledných modelů pro zadavatele.

U metodologie Box-Jenkins je podmínkou, že analyzované časové řady musí být stacionární.

3.1 Stacionarita

Metodologie je sestavena pouze pro stacionární časové řady, proto je zapotřebí některé nestacionární řady převést na stacionární pomocí diferencí či kovariancí. Nejprve, co to vlastně stacionarita je. Jedná se o charakteristickou vlastnost hodnot časových řad vracet se k určité konstantě (Cipra, 1986).

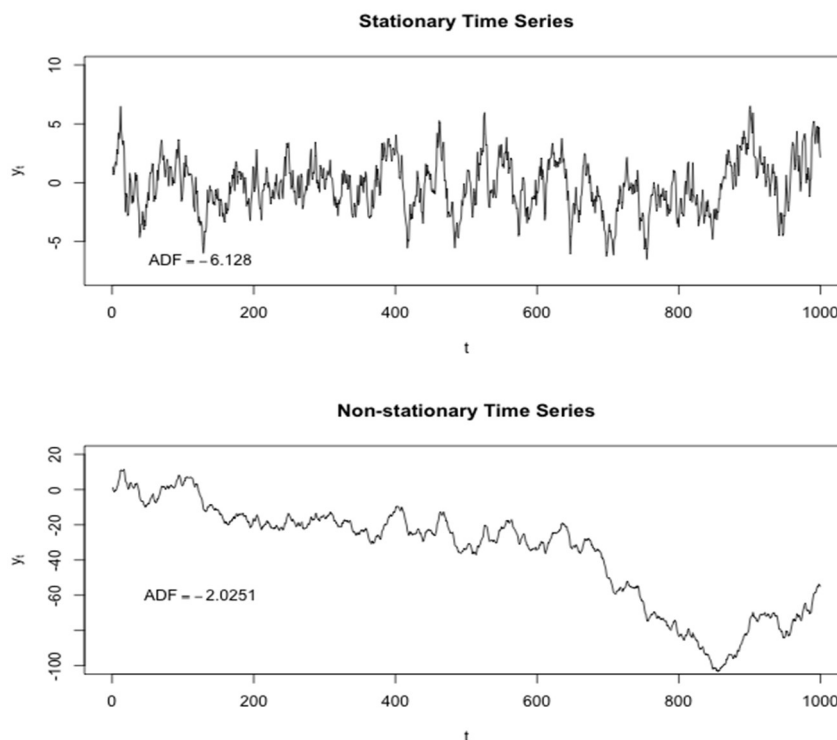
Existují dvě základní koncepce stacionarity, striktní a slabá stacionarita. Striktní stacionarita předpokládá, že pravděpodobnostní rozdělení dvou odpovídajících si vektorů hodnot časové řady je stejné, a to bez ohledu na umístění těchto vektorů. Tzn. pravděpodobnostní rozdělení je neměnné s ohledem na posuny v čase t . Naopak slabá stacionarita se vyznačuje tím, že má v jednotlivých okamžicích stejnou střední hodnotu, rozptyl a kovarianci dvou hodnot (Hamilton 1994, Hendl 2014).

Pro časové řady, které mají vlastnost slabé stacionarity, je vzájemná závislost jednotlivých pozorování popsána pomocí autokorelační či parciální autokorelační funkce.

Dále budeme stacionaritou rozumět **stacionaritu slabou**.

Rozdíl mezi stacionární a nestacionární časovou řadou lze vidět na obrázku č. 1. U prvního případu lze vidět, že hodnoty mají tendenci se vracet k určité konstantě a v druhém případě se hodnoty nevracejí.

Obrázek 1 - Stacionární vs nestacionární časové řady



Zdroj: <http://wikimedia.cz>, 2019

3.2 Autokorelační a parciální autokorelační funkce

Autokorelační funkce (ACF)

V případě stacionárního procesu Y_t lze autokovarianční funkci mezi veličinami y_t a y_{t-k} vyjádřit jako

$$\gamma_k = cov(y_t, y_{t-k}) = E(y_t - \mu)(y_{t-k} - \mu), \quad (40)$$

kde: $k = 0, 1, 2, \dots, n-1$,

a autokorelační funkci (ACF) jako

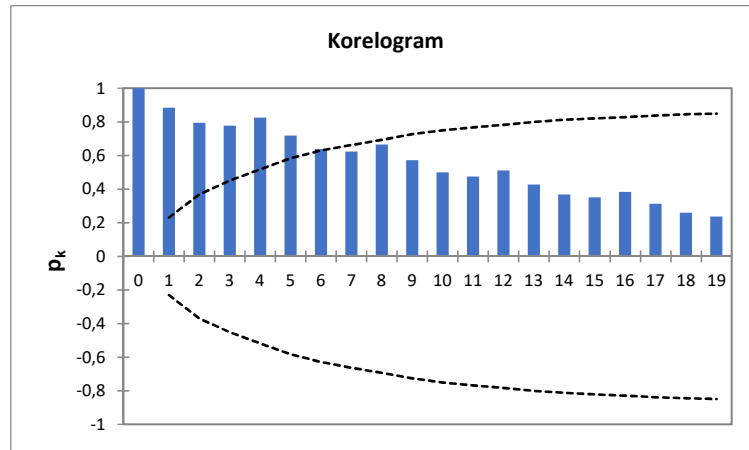
$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{cov(y_t, y_{t-k})}{\sqrt{D(y_t)}\sqrt{D(y_{t-k})}}, \quad (41)$$

kde: vzhledem ke stacionaritě $D(y_t) = D(y_{t-k}) = \gamma_0$,

hodnota $\rho_0 = 1$ a ostatní hodnoty ACF se pohybují v intervalu $\langle -1;1 \rangle$.

Graf autokorelační funkce se nazývá korelogram, jedna z jeho forem je zachycena na obrázku č. 2.

Obrázek 3 - Korelogram



Zdroj: vlastní zpracování, EXCEL 2019

Hodnoty ACF v intervalu $\langle -1;1 \rangle$ jsou neznámé, ale lze je odhadnout přímo z pozorovaných dat. Tedy, odhad r_k autokorelační funkce ρ_k lze získat vztahem:

$$r_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad (42)$$

kde: $k = 1, 2, \dots, n-1$.

Hodnoty ACF prezentují vztah mezi dvěma pozorováními podle toho, jak je ovlivněn pozorováními a náhodnými vlivy mezi těmito dvěma pozorováními. Např. lze uvést, že hodnota autokorelační funkce mezi hodnotami y_3 a y_7 je ovlivněna hodnotami y_4, y_5 a y_6 . Hodnota ACF tedy dává informaci o síle lineární závislosti (Řezanková 2009).

Parciální autokorelační funkce (PACF)

U PACF je většinou korelace mezi dvěma náhodnými veličinami způsobená korelací s veličinou třetí. Parciální autokorelace podává informace o korelaci veličin y_t a y_{t-k} očištěné o vliv veličin, které se nachází mezi nimi. Hodnoty této funkce jsou dány vztahem:

$$\rho_{kk} = \frac{\det P_k^*}{\det P_k}, \quad (43)$$

kde $\det P_k$ je determinant matice

$$P_k = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{pmatrix},$$

a $\det P_k^*$ je determinant matice

$$P_k^* = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_1 \\ \rho_1 & 1 & \dots & \rho_2 \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_k \end{pmatrix}.$$

Výše uvedené matice jsou poměrně totožné až na poslední sloupek, ten se liší. Odhad r_{kk} PACF ρ_{kk} lze získat dosazením odhadnutých hodnot r_k za hodnoty ρ_k ve vztahu (43).

Z výsledných grafů ACF a PACF je nutné určit typ parametrů na základě pravidel odvozených z vlastností ACF a PACF:

- „AR (1): ACF exponenciálně klesá k nule, PACF má pouze jeden vrchol na pozici 1.
- AR (2): ACF má sinusový průběh nebo několik exponenciálních poklesů, PACF má 2 vrcholy na pozicích 1 a 2. Na dalších pozicích nejsou již korelace.
- MA (1): ACF má pouze jeden vrchol na pozici 1, PACF jde exponenciálně k nule.
- MA (2): ACF má dva vrcholy na pozici 1 a 2, bez dalších korelací, PACF má sinusový průběh, nebo exponenciálně tlumený průběh.
- ARMA (1, 1): ACF má exponenciální pokles začínající na pozici 1, PACF má oscilující pokles začínající na pozici 1.
- ARMA (p, q): ACF přímý nebo oscilující pokles k nule začínající na pozici q. PACF přímý nebo oscilující pokles k nule začínající na pozici p“ (STATSOFT, 2019).

Základní procesy a metody, se kterými metodologie pracuje jsou uvedeny v následující kapitole.

3.3 Modely (AR, MA, ARIMA a SARIMA)

Veškerá odvození výpočtů následujících modelů jsou obsažena v literatuře (např. Cipra 1986).

Proces klouzavých součtů MA (q)

Proces klouzavých součtů řádu q je označován procesem, ve kterém hodnoty vysvětlované veličiny v čase t jsou tvořeny lineární kombinací minulých i současných hodnot náhodné veličiny ε_t . Pro názornost jsou v následujícím vztahu zachyceny klouzavé součty prvního řádu.

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}, \quad (44)$$

kde: $\theta < 1$, $\theta_i \dots$ je indikátorem „paměti“ procesu

$y_t \dots$ empirické údaje časové řady,

$\varepsilon_t \dots$ nekorelované náhodné pozorování s σ^2 (rozptyl) = konst. a E (střední hodnota) = 0.

V případě, že θ je nulová, nemá proces MA žádnou paměť, z toho důvodu lze tento proces označit za „bílý šum“. Výše uvedený model se označuje jako MA (1), neboli **model klouzavých součtů prvního řádu**. Tento model patří mezi nejjednodušší, se kterými metodologie pracuje (Cipra, 1986).

Autoregresní proces AR (p)

Kromě modelu klouzavých součtů se používají také modely autoregresní. **Autoregresní model** se označuje AR(p), neboli autoregresní model p-tého řádu a má následující podobu.

$$y_t = y_{t-1} + \varepsilon_t. \quad (45)$$

V této rovnici je y_t vypočteno jako funkce několika předcházejících pozorování, tedy hodnota časové řady v čase t je vytvořena lineární kombinací předcházejících hodnot z minulosti.

Při kombinaci modelu klouzavých součtů (MA) a autoregresního modelu (AR) vzniknou smíšené modely (ARMA, ARIMA).

Model ARMA (p, q)

Jedná se o smíšený model, tedy autoregresní model klouzavých součtů a je popsán následujícím vztahem:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t - \phi_1 a_{t-1} - \phi_2 a_{t-2} - \dots - \phi_q a_{t-q}, \quad (46)$$

kde: p ... počet autoregresních typů,

q ... zpožděné klouzavé průměry.

Při analýze časových řad se může stát, že jednotlivá pozorování nevyhovují předpokladům MA nebo AR modelů, proto mohou být spojena do modelu ARMA. Pokud se bude model označovat jako ARMA (p, 0), jedná se o proces AR(p), stejně je tomu s modelem MA(q). Tvar ACF bude mít v tomto případě tvar exponenciálně klesajících pohybů a exponenciálně klesajících sinusoidních pohybů.

Model ARIMA

Výše uvedené modely a procesy se zabývaly pouze stacionárními časovými řadami, avšak v ekonomické praxi se lze setkat s velkým množstvím časových řad, které jsou tvořené nestacionárními stochastickými procesy. Tato nestacionarita je zpravidla způsobená měnící se střední hodnotou v čase či měnícím se rozptylem. Z toho důvodu jsou zavedeny integrované procesy.

Model ARIMA je autoregresní integrovaný proces klouzavých průměrů řádů p , d a q . Označ. ARIMA (p, d, q), kde d je řád diference. Model se nazývá také sezónním modelem, ve kterém je trendová složka modelována stochasticky. Model s využitím „operátoru zpětného posunu B “ lze popsat následujícím vztahem:

$$\phi_p = (B)(1 - B)^d Y_t = \theta_q(B) \varepsilon_t, \quad (47)$$

položí-li se:

$$w_t = (1 - B)^d Y_t, \quad (48)$$

pak lze w_t označit jako stacionární ARMA (p, q). Operátor zpětného posunu představuje zpoždění o jedno období, díky tomuto operátorovi lze modely s mnoha zpožděními zapsat stručnějším způsobem.

Hodnoty vypočtené pomocí ARIMA (p, d, q) se nemusí centrovat, jelikož platí následující vztah:

$$\Delta^d(Y_t - \bar{Y}) = \Delta^d Y_t. \quad (49)$$

Model ARIMA se používá pro krátkodobou predikci časových řad, a to v případě, pokud jsou výsledné odhady AR nebo MA nepoužitelné pro ekonomické či statistické účely. (Arlt, 2009; Cipra, 1986).

Model SARIMA

Model SARIMA neboli sezónní autoregresní integrovaný proces klouzavých průměrů řádů p, d a q v sobě zahrnuje sezónní autoregresní proces (SAR) a sezónní proces klouzavých průměrů (SMA). U tohoto modelu lze také předpokládat závislost mezi veličinami v jednotlivých sezónách.

SARIMA (0, 0, 0) (P, D, Q), kde P je řád SAR procesu, D je řád sezónní difference a Q je řád SMA procesu, lze zapsat následujícím vztahem:

$$\Phi_p(B^4)(1 - B^4)^D Y_t = \sum_{j=1}^4 \delta_j^* D_{j,t} + \Theta_q(B^4)\varepsilon_t, \quad (50)$$

kde: Φ_p ... parametry SAR procesu

Θ_q ... parametry SMA procesu

D ... řád sezónní difference

Pokud se rovnice (50) dále rozšíří o invertibilní, stacionární a nestacionární nesezónní část, lze získat model SARIMA (p, d, q) (P, D, Q) s konstantou:

$$\Phi_p(B)\varphi_p(B)(1 - B)^d(1 - B^4)^D Y_t = \sum_{j=1}^4 \delta_j^* D_{j,t} + \theta_q(B)\Theta_Q(B^4)\varepsilon_t, \quad (51)$$

kde: θ_q ... parametry sezónního MA procesu

φ_p ... parametr nesezónního AR procesu

$D_{j,t}$... nula-jedničková sezónní pomocná proměnná

První část modelu SARIMA je nesezónní (AR), která je násobena sezónní složkou (AR), dále vynásobená nesezónní diferencí a také sezónní diferencí a daným trendem. Tento součin se rovná součinu nesezónní (MA) a sezónní (MA).

Pokud se do modelu zadají konkrétní údaje, například, pokud model bude ve tvaru SARIMA $(0,1,1) \times (0,1,1)_4$, lze ho zapsat vztahem:

$$(1 - B)(1 - B^4)Y_t = (1 + \theta_1 B)(1 + \Theta_1 B^4)\varepsilon_t. \quad (52)$$

Pokud je časová řada stacionární, je dalším krokem zjištění hodnot řádů p a q pro procesy AR a MA. Tyto řády lze určit z grafu autokorelační a parciální autokorelační funkce (Hindls 2000, Cipra 2008).

4 Extrapolace časových řad

Cílem každé analýzy časových řad je, co možná nejpřesněji, odhadnout budoucí hodnoty daných časových řad pomocí nejvhodnějšího modelu. Předpovědi jsou používány pro rozhodování a plánování vývoje určité veličiny. Odhadnutá doba pomocí nejvhodnějšího modelu nikdy nebude úplně přesná, jelikož se jedná jen o odhad s určitou pravděpodobností. Jde tedy o kvantitativní předpovědní metody.

Extrapolaci časových řad lze provádět dvěma způsoby:

- bodovým odhadem,
- intervalovým odhadem.

Bodový odhad – je metoda predikce, při které se počítá požadované číslo, které je nejbližší odhadované hodnotě. Postup je poměrně lehký a často využívaný. Nevýhodou bodového odhadu je jeho nepřesnost, jelikož výsledek je vždy zatížen určitou chybou.

Intervalový odhad – je metoda predikce, při které se zkonstruuje interval, ve kterém je zahrnuta hledaná hodnota. Zvolí se hranice spolehlivosti $(1 - \alpha)$, kde α vyjadřuje hladinu významnosti. V obvyklých výpočtech se uvádí 5 %. Tato metoda odhadu je přesnější než bodový odhad, jelikož udává, v jakém rozmezí se predikované hodnoty mohou pohybovat.

V následující části budou uvedeny extrapolace pro metody: dekompozice časové řady a metodologie Box-Jenkins, uvedené v kapitolách 2 a 3.

4.1 Extrapolace u dekompozice časových řad

Při dekompozici časových řad se pracuje se systematickými složkami, které se odhadnou do budoucnosti.

U obou typů dekompozice se výpočty budou lišit. U aditivních přístupů se jedná o součet sezónní a trendové složky a u multiplikačních přístupů o součin těchto dvou složek. Liší se i výpočty bodových předpovědí u klasických či adaptivních přístupů, tyto přístupy jsou uvedeny níže (Cipra, 1986).

Klasické přístupy

- **Lineární trend**

Výpočet bodové předpovědi:

$$T_p = b_0 + b_1 t, \quad (53)$$

kde: $t > n$

T_p ... předpověď budoucí hodnoty v čase t .

- **Kvadratický trend**

Výpočet bodové předpovědi:

$$T_p = b_0 + b_1 t + b_2 t^2, \quad (54)$$

kde: $t > n$

T_p ... předpověď budoucí hodnoty v čase t .

- **Exponenciální trend**

Výpočet bodové předpovědi:

$$T_p = b_0 b_1^t, \quad (55)$$

kde: $t > n$

T_p ... předpověď budoucí hodnoty v čase t .

V následující části bude popsána extrapolace pomocí klouzavých průměrů a pomocí exponenciálního vyrovnávání. Obě tyto metody patří do adaptivních přístupů.

Adaptivní přístupy

- **Exponenciální vyrovnávání**

Výpočty a postupy pro předpověď dvojitého a trojitého exponenciálního vyrovnávání jsou následující.

- **Dvojitě exponenciální vyrovnávání**

Výpočet bodové předpovědi:

$$\hat{y}_{t+h}(t) = b_0(t) + b_1(t)h = \left(2 + \frac{(1-\alpha)h}{a}\right)S_t - \left(1 + \frac{(1-\alpha)h}{a}\right)S_t^{[2]}, \quad (56)$$

$$\text{kde: } b_0(t) = S_t - S_t^{[2]},$$

$$b_1(t) = \frac{(1-\alpha)}{a} (S_t - S_t^{[2]}).$$

○ **Trojité exponenciální vyrovnávání**

Výpočet bodové předpovědi:

$$\hat{y}_{t+h} = \frac{1}{2\alpha^2} \left\{ (6\alpha^2 + (1+5\alpha)(1-\alpha)h + (1-\alpha)^2 h^2 S_t - (6\alpha^2 + 2(1+4\alpha)(1-\alpha)h + 2(1-\alpha)^2 h^2 S_t^{[2]} + (2\alpha^2 + (1+3\alpha)(1-\alpha)h + (1-\alpha)^2 h^2) S_t^{[3]} \right\}. \quad (57)$$

Výpočty intervalových předpovědí zde nejsou uvedeny, z důvodu velké obsáhlosti a nepřehlednosti. Veškeré postupy jsou uvedeny v příslušné literatuře (např. Cipra 1986 či Arlt, Arltová 2009).

4.2 Extrapolace u vybraných metod z Box-Jenkinsovy metodologie

• **Model ARIMA**

Tento typ modelu má podobu viz vztah č. 47, avšak tento tvar lze také zapsat tímto způsobem:

$$T_t^* = \Phi_p T_{t-1}^* + \dots + \Phi_p T_{t-p}^* + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \dots - \Theta_q \varepsilon_{t-q}, \quad (58)$$

$$\text{kde: } T_t^* = (1 - B)^d T_t.$$

Výpočet bodové předpovědi je poté:

$$T_{t+1}^* = \Phi_1 T_t^* + \dots + \Phi_p T_{t-p+1}^* + \varepsilon_{t+1} - \Theta_1 \varepsilon_t - \dots - \Theta_q \varepsilon_{t-q+1}, \quad (59)$$

$$\text{kde: } T_{t+1}^* \dots \text{ predikovaná hodnota,}$$

$$\Phi_1, \dots, \Phi_p \text{ a } \Theta_1, \dots, \Theta_q \text{ jsou odhady parametřů.}$$

Ověření vhodnosti použitých modelů pro predikování hodnot lze provést na základě kritérií M.S.E. (střední čtvercová chyba odhadu) a M.A.P.E. (střední absolutní procentní chyba odhadu).

- **Model SARIMA**

Následující vztah poskytuje předpovědní model, ve kterém jsou použity odhadnuté koeficienty modelem SARIMA.

$$\hat{y}_i = \phi_0 + \phi_1 y_{i-1} + \Phi_1 y_{i-4} - \phi_1 \Phi_1 y_{i-5} + \theta_1 \varepsilon_{i-1} + \theta_1 \varepsilon_{i-4} + \theta_1 \theta_1 \varepsilon_{i-5}, \quad (60)$$

kde: \hat{y}_i ... predikovaná hodnota.

Vhodnost modelu SARIMA lze zjistit např. porovnáním sumy odchylek predikovaných hodnot a skutečných hodnot. Model, který bude mít nejmenší sumu odchylek je nejpřesnější (Artl, Artlová 2009).

4.3 Dummy proměnné

Aby bylo možné sestavit model časových řad, je zapotřebí do modelu zakomponovat sezónní složku. Cílová podoba modelu může posloužit např. pro extrapolaci časové řady za použití nejvhodnějších metod. Pro určení co možná nejvhodnějšího výsledného modelu je důležité správně zařadit do modelu sezónní složku. K tomu lze využít tzv. DUMMY proměnné neboli umělé proměnné, které používají kombinaci 0-1 vektorů. Slouží k co nejpřesnějšímu vystihnutí významu regresních koeficientů. Pokud jsou DUMMY proměnné zavedeny do modelu, je zapotřebí se vyvarovat vzniku perfektní multikolinearity. To znamená, že do modelu mohou být zařazeny jen umělé proměnné v počtu $n-1$, viz tabulka č. 3, kde jsou pro první pololetí samé nulové vektory.

Příkladem Dummy proměnných je např. matice proměnných na dvě pololetí:

Tabulka 3 - Ukázka DUMMY

Označení I	I_j	Označení D	DUMMY			
α_1	0,900	D_1	0	0	0	0
α_2	1,100	D_2	0	1	0	0

Zdroj: vlastní zpracování, EXCEL 2019

Předpokladem pro sezónní indexy je, že jejich suma se musí rovnat počtu sezón (v tomto případě dvěma).

5 Analýza vybraných časových řad

5.1 Popis dat

První analyzovanou časovou řadou jsou čtvrtletní **tržby z ubytování** mezi roky 2000 a 2017 v ČR. Druhou analyzovanou časovou řadou je čtvrtletní vývoj **hrubé (nominální) mzdy** mezi roky 2000 a 2017 v ČR. Čtvrtletní data jsou vybrána kvůli větší přesnosti při odhadu a lepší vypovídající schopnosti dat. Tyto dvě časové řady byly vybrány, jelikož tržby za ubytování mohou dobře znázorňovat sezónnost v porovnání s vývojem hrubé mzdy, která nevykazuje až tak velké výkyvy sezónnosti. Navíc hrubá mzda je poměrně sledovanou veličinou v České republice, a proto by bylo dobré analyzovat její chování a zjistit, jak se bude vyvíjet v budoucnu. Cyklická složka v těchto časových řadách není analyzována, a to kvůli poměrně malému množství dat, proto je považována za součást náhodné složky.

5.1.1 Tržby za ubytování

Data v letech 2000-2009 jsou převzata z „*Archiv: Klasifikace NACE Rev. 1.1. (OKEČ 55)*“ a data v letech 2010-2017 jsou převzata z „*NACE Rev. 2 (CZ-NACE 55)*“ z internetových stránek Českého statistického úřadu. Celkové tržby za ubytování vycházejí ze zpracování čtvrtletních výkazů pro ekonomické subjekty ve vybraných odvětvích. Tržby za ubytování jsou uvedeny v milionech Kč.

Tabulka 4 - Časová řada tržeb za ubytování (v mil. Kč)

Období	Tržby	Období	Tržby	Období	Tržby	Období	Tržby
2000 - Q1	2 367	2004 - Q3	3 843	2009 - Q1	4 401	2013 - Q3	10 480
2000 - Q2	2 806	2004 - Q4	3 635	2009 - Q2	5 398	2013 - Q4	9 554
2000 - Q3	2 672	2005 - Q1	3 013	2009 - Q3	5 243	2014 - Q1	7 406
2000 - Q4	2 788	2005 - Q2	4 096	2009 - Q4	5 978	2014 - Q2	10 609
2001 - Q1	2 319	2005 - Q3	3 724	2010 - Q1	6 842	2014 - Q3	11 042
2001 - Q2	3 464	2005 - Q4	3 574	2010 - Q2	9 598	2014 - Q4	10 370
2001 - Q3	2 967	2006 - Q1	3 451	2010 - Q3	10 112	2015 - Q1	7 989
2001 - Q4	2 699	2006 - Q2	4 718	2010 - Q4	9 342	2015 - Q2	11 995
2002 - Q1	2 463	2006 - Q3	4 429	2011 - Q1	7 227	2015 - Q3	12 459
2002 - Q2	3 423	2006 - Q4	4 329	2011 - Q2	10 211	2015 - Q4	11 173
2002 - Q3	2 602	2007 - Q1	3 470	2011 - Q3	10 050	2016 - Q1	9 105
2002 - Q4	2 602	2007 - Q2	4 549	2011 - Q4	9 432	2016 - Q2	12 286
2003 - Q1	2 391	2007 - Q3	4 242	2012 - Q1	7 204	2016 - Q3	13 736
2003 - Q2	3 226	2007 - Q4	4 338	2012 - Q2	10 188	2016 - Q4	12 175
2003 - Q3	3 508	2008 - Q1	3 910	2012 - Q3	10 270	2017 - Q1	9 684
2003 - Q4	3 062	2008 - Q2	4 976	2012 - Q4	9 286	2017 - Q2	14 316
2004 - Q1	3 109	2008 - Q3	4 652	2013 - Q1	7 143	2017 - Q3	14 946
2004 - Q2	4 428	2008 - Q4	4 534	2013 - Q2	10 107	2017 - Q4	13 489

Zdroj: Český statistický úřad, 2019

V této časové řadě je celkem 72 údajů, které budou dále analyzovány. Na grafu č.1 lze vidět, jak časová řada vypadá.

Graf 1 - Časová řada (tržby za ubytování)



Zdroj: vlastní zpracování, EXCEL 2019

Hned na první pohled, lze vidět rozdíl mezi první a druhou částí časové řady. Tento nárůst v roce 2009–2010 může být způsoben odlišnou klasifikací ekonomických činností. Jak bylo již zmíněno, do konce roku 2009 byla použita klasifikace NACE Rev. 1.1. (OKEČ 55) a od začátku roku 2010 je použita klasifikace NACE Rev. 2 (CZ-NACE 55). Tento fakt může mít za následek výrazný růst v tržbách, kvůli možné změně výpočtu. Růst tržeb mohl být také způsoben ekonomickým růstem po ekonomické krizi. Na první pohled lze tvrdit, že dochází ke změně trendu, kde v první části grafu rostou hodnoty lineárně a ve druhé části spíše exponenciálně.

Základní charakteristiky

Tabulka 5 - Základní charakteristiky (ubytování tržby)

Ukazatele	Empirická data	Ukazatele	Empirická data
Stř. hodnota	6628,14	Šikmost	0,56
Chyba stř. hodnoty	431,03	Rozdíl max-min	12627,13
Medián	4847,05	Minimum	2319,36
Směr. odchylka	3657,38	Maximum	14946,49
Rozptyl výběru	13376448,71	Součet	477225,93
Špičatost	-1,03	Počet	72,00

Zdroj: vlastní zpracování, EXCEL 2019

Základní charakteristiky uvedené v tabulce č. 5 jsou vypočítány za pomoci vzorců uvedených v teoretické části. V základních ukazatelích není uveden modus, jelikož nelze určit, protože každá hodnota je zastoupena jen jednou. Ostatní ukazatele jsou poměrně srozumitelné, jen je zapotřebí uvést, co znamenají výsledky špičatosti a šikmosti. Záporný výsledek špičatosti svědčí o tom, že rozložení hodnot v časové řadě má spíše plochý tvar. Kladná hodnota šikmosti naopak svědčí o tom, že časová řada je levostranně asymetrická oproti normálnímu rozdělení.

5.1.2 Hrubá mzda

Tato časová řada obsahuje čtvrtletní data hrubé mzdy mezi roky 2000 až 2017. Hrubá mzda je uváděna v Kč. Data jsou převzata z databáze časových řad ARAD, která je sestavována Českou národní bankou. Avšak v metodice sestavování je uvedeno, že zdrojem dat jsou údaje z Českého statistického úřadu, přesněji data za civilní sektor hospodářství ČR.

Tabulka 6 - Časová řada hrubá mzda (v Kč)

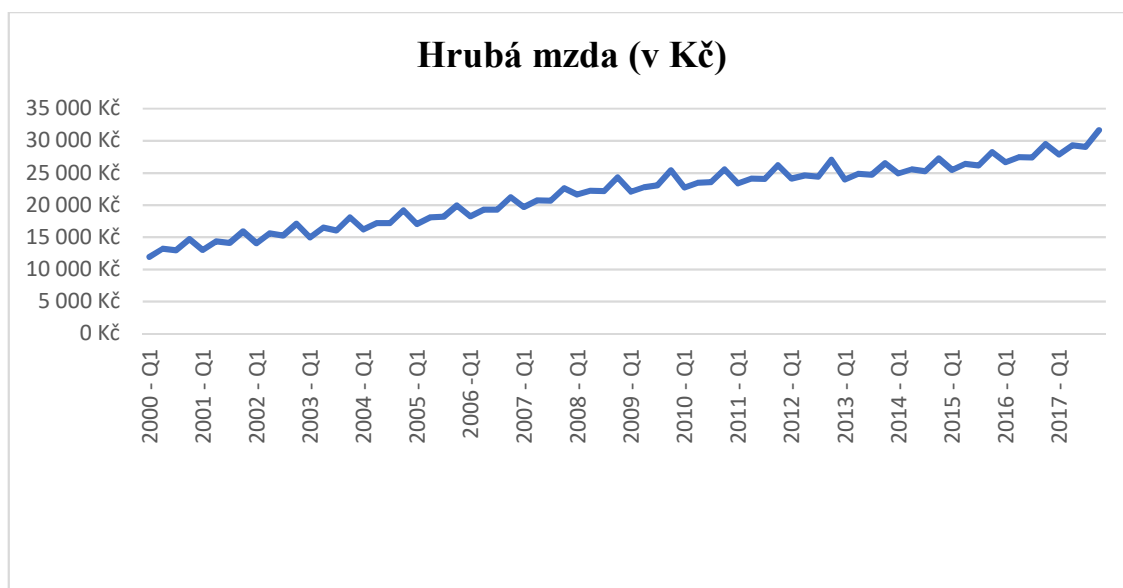
Období	Hrubá mzda	Období	Hrubá mzda	Období	Hrubá mzda	Období	Hrubá mzda
2000 - Q1	11 941 Kč	2004 - Q3	17 190 Kč	2009 - Q1	22 108 Kč	2013 - Q3	24 735 Kč
2000 - Q2	13 227 Kč	2004 - Q4	19 183 Kč	2009 - Q2	22 796 Kč	2013 - Q4	26 525 Kč
2000 - Q3	12 963 Kč	2005 - Q1	17 067 Kč	2009 - Q3	23 091 Kč	2014 - Q1	24 931 Kč
2000 - Q4	14 717 Kč	2005 - Q2	18 112 Kč	2009 - Q4	25 418 Kč	2014 - Q2	25 569 Kč
2001 - Q1	13 052 Kč	2005 - Q3	18 203 Kč	2010 - Q1	22 738 Kč	2014 - Q3	25 279 Kč
2001 - Q2	14 391 Kč	2005 - Q4	19 963 Kč	2010 - Q2	23 504 Kč	2014 - Q4	27 261 Kč
2001 - Q3	14 117 Kč	2006 - Q1	18 270 Kč	2010 - Q3	23 600 Kč	2015 - Q1	25 497 Kč
2001 - Q4	15 908 Kč	2006 - Q2	19 300 Kč	2010 - Q4	25 591 Kč	2015 - Q2	26 408 Kč
2002 - Q1	14 083 Kč	2006 - Q3	19 305 Kč	2011 - Q1	23 372 Kč	2015 - Q3	26 163 Kč
2002 - Q2	15 599 Kč	2006 - Q4	21 269 Kč	2011 - Q2	24 116 Kč	2015 - Q4	28 258 Kč
2002 - Q3	15 268 Kč	2007 - Q1	19 687 Kč	2011 - Q3	24 107 Kč	2016 - Q1	26 683 Kč
2002 - Q4	17 133 Kč	2007 - Q2	20 740 Kč	2011 - Q4	26 211 Kč	2016 - Q2	27 452 Kč
2003 - Q1	14 986 Kč	2007 - Q3	20 721 Kč	2012 - Q1	24 131 Kč	2016 - Q3	27 396 Kč
2003 - Q2	16 529 Kč	2007 - Q4	22 641 Kč	2012 - Q2	24 627 Kč	2016 - Q4	29 491 Kč
2003 - Q3	16 088 Kč	2008 - Q1	21 632 Kč	2012 - Q3	24 439 Kč	2017 - Q1	27 880 Kč
2003 - Q4	18 096 Kč	2008 - Q2	22 246 Kč	2012 - Q4	27 055 Kč	2017 - Q2	29 335 Kč
2004 - Q1	16 231 Kč	2008 - Q3	22 181 Kč	2013 - Q1	23 985 Kč	2017 - Q3	29 058 Kč
2004 - Q2	17 223 Kč	2008 - Q4	24 309 Kč	2013 - Q2	24 877 Kč	2017 - Q4	31 661 Kč

Zdroj: ARAD, 2019

Tato časová řada obsahuje rovněž čtvrtletní údaje za osmnáct let, tedy také 72 hodnot.

Na grafu č. 2 lze vidět, jak časová řada vypadá.

Graf 2 - Časová řada (hrubá mzda)



Zdroj: vlastní zpracování, EXCEL 2019

Oproti předchozí časové řadě, je zde vidět poměrně hladký vývoj a růst hodnot hrubé mzdy. Na první pohled je patrné, že se hrubá mzda zvyšuje každým rokem, avšak v rámci jednoho roku kolísá vlivem různých faktorů.

Základní charakteristiky

Tabulka 7 - Základní charakteristiky (hrubá mzda)

Ukazatele	Empirická data	Ukazatele	Empirická data
Stř. hodnota	21651,65	Šikmost	-0,22
Chyba stř. hodnoty	575,28	Rozdíl max-min	19720
Medián	22689,50	Minimum	11941
Směr. odchylka	4881,42	Maximum	31661
Rozptyl výběru	23828289,07	Součet	1558919
Špičatost	-0,98	Počet	72

Zdroj: vlastní zpracování, EXCEL 2019

Opět lze konstatovat výsledek špičatosti a šikmosti. Výsledná hodnota špičatosti je v tomto případě také záporná, tudíž rozložení hodnot v časové řadě má plochý tvar. Šikmost zde vyšla negativně, což svědčí o pravostranné asymetrii vůči normálnímu rozdělení. Ukazatel modus zde opět není uveden, jelikož ho nelze určit.

5.2 Dekompozice časových řad

5.2.1 Modelování trendu (ubytování)

Klasické přístupy

V následující části jsou popsány vybrané klasické přístupy při dekompozici časové řady „tržby za ubytování“. Nejprve jsou znázorněny jednotlivé grafy pro lineární, kvadratickou, exponenciální a mocninou regresi a pomocí SW Excel jsou vypočteny upravené koeficienty determinace. Výsledné hodnoty jsou uvedeny v tabulce č. 8.

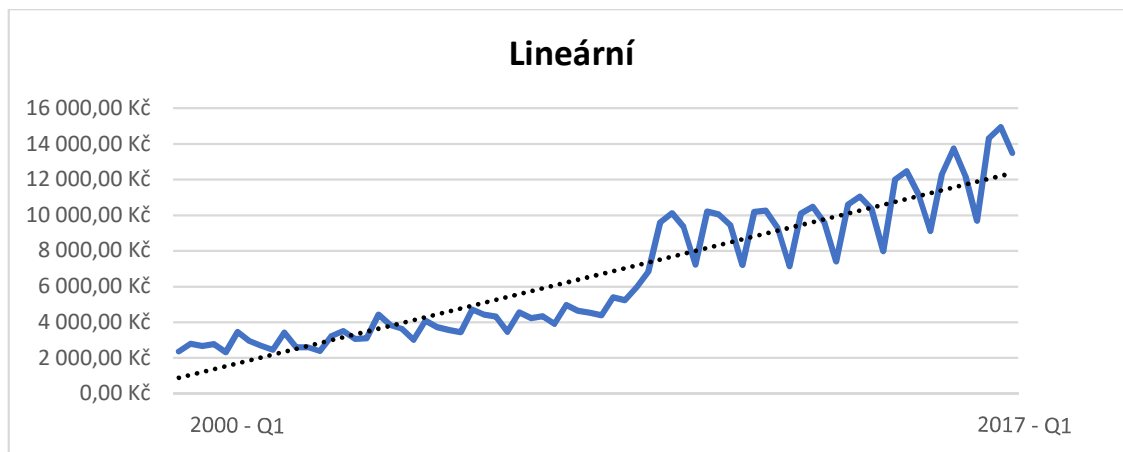
Tabulka 8 – Upravené koeficienty determinace (ubytování)

Regrese	Koeficient determinace (upravený) - R^2
Lineární	85,44%
Kvadratická	88,61%
Exponenciální	90,63%
Logistická	57,72%
Mocninná	69,65%

Zdroj: vlastní zpracování, EXCEL 2019

Na základě výsledků koeficientu determinace byla vyloučena logistická a mocninná regrese, protože měly nejmenší vypovídající hodnotu. U zbývajících regresí je znázorněn jejich průběh na grafu a jsou vypočteny koeficienty b_0 , b_1 a popřípadě b_3 . Pro zjištění kvality vypovídající hodnoty, jsou pro každou regresi vypočteny statistická kritéria reziduálního součtu čtverců (R.S.S.) a střední čtvercové chyby odhadu (M.S.E.).

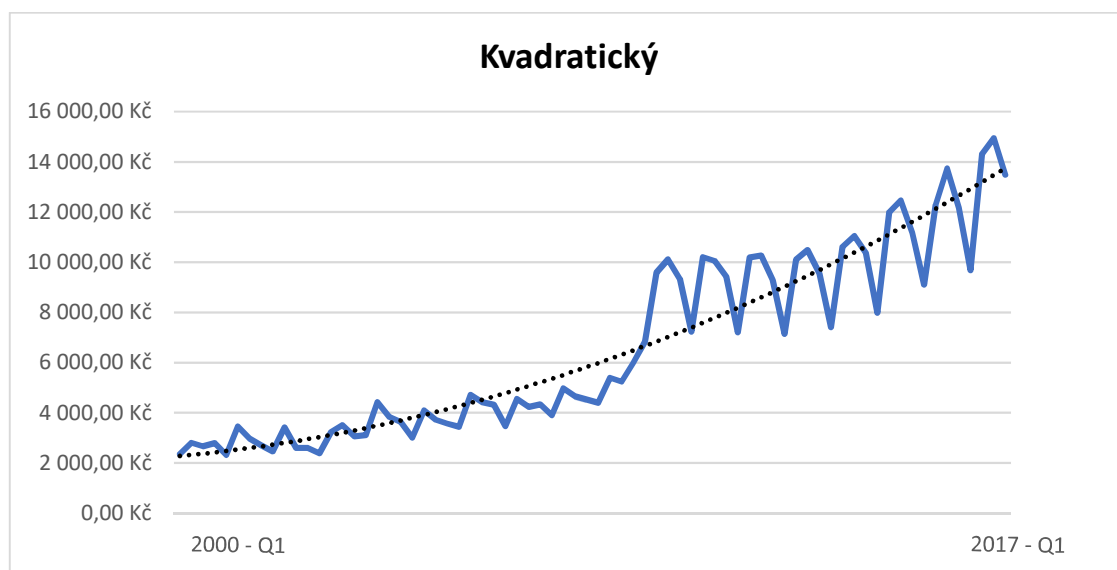
Graf 3 - Lineární trend (ubytování)



Zdroj: vlastní zpracování, EXCEL 2019

Lineární trend má podobu rovnice (12), tedy $T_t = \beta_0 + \beta_1 t$.

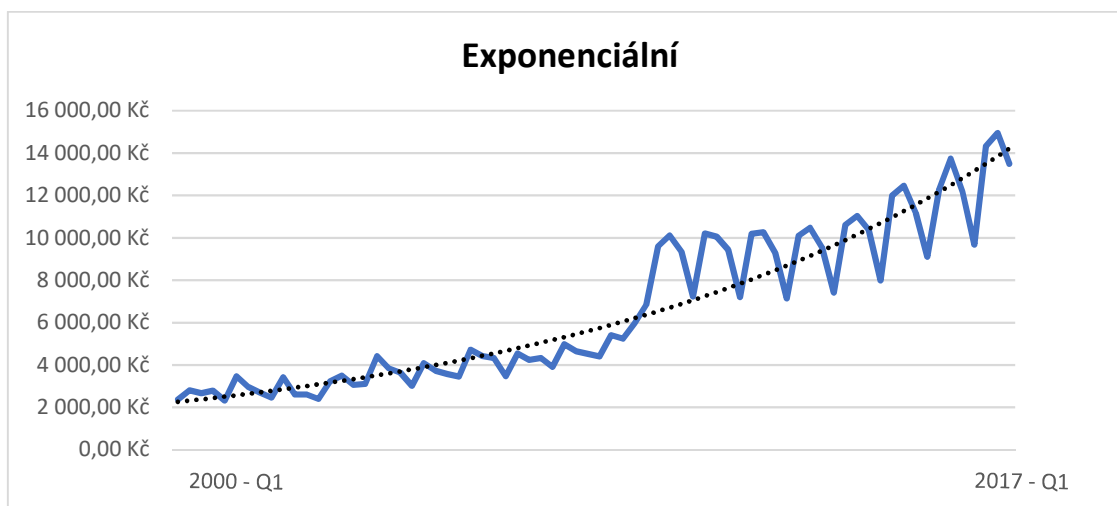
Graf 4 - Kvadratický trend (ubytování)



Zdroj: vlastní zpracování, EXCEL 2019

Kvadratický trend má podobu rovnice (13), tedy $T_t = \beta_0 + \beta_1 t + \beta_2 t^2$.

Graf 5 - Exponenciální trend (ubytování)



Zdroj: vlastní zpracování, EXCEL 2019

Exponenciální trend má podobu rovnice (14), tedy $T_t = \beta_0 \beta_1^t$.

Při pohledu na výše uvedené grafy lze určit, který trend prokládá empirická data nejlépe, v tomto případě je tomu tak u exponenciálního trendu. Avšak pouze vizuální posouzení nestačí, proto jsou dále v programu Excel vypočítány hodnoty koeficientů u každého

druhu trendu, ty jsou také použity při predikci či sestavení výsledného modelu. Výsledné hodnoty jsou uvedeny v tabulce č. 9.

Tabulka 9 - Hodnoty koeficientů b (ubytování)

Lineární	b1	b0	
	161,535	732,125	
Kvadratický	b2	b1	b0
	1,675	39,280	2 239,930
Exponenciální	b2	ln(b1)	b1
	0,026	7,698	2 203,910

Zdroj: vlastní zpracování, EXCEL 2019

Pro zjištění vypovídající schopnosti modelů a kvality jejich predikce, jsou vypočtena níže uvedená statistická kritéria. K výpočtu kritérií bylo zapotřebí vypočítat jednotlivá rezidua pro každý trend a také jejich druhé mocniny. Rezidua jsou vypočtena jako rozdíl empirické hodnoty a hodnoty, která je určena daným trendem. Pro výpočet kritéria R.S.S. byl použit vzorec $RSS = \sum_{t=1}^n (y_t - T_t)^2$ a pro výpočet M.S.E. byl použit vzorec $MSE = \frac{RSS}{n}$. Výsledné hodnoty jsou uvedeny v následující tabulce.

Tabulka 10 - Výsledky kritérií (ubytování)

VÝSLEDKY			
Ukazatel	Lineární	Kvadratický	Exponenciální
R.S.S.	138 274 446,82	108 154 483,83	115 770 148,47
M.S.E.	1 920 478,43	1 502 145,61	1 607 918,73

Zdroj: vlastní zpracování, EXCEL 2019

Nejlepším modelem je ten, který má nejmenší hodnotu zvoleného kritéria. Nejčastěji je použito kritérium střední čtvercové chyby odhadu (M.S.E.). Pokud by se braly v úvahu jen výsledky determinace, vyšel by nejlépe exponenciální trend. Ale v tomto případě je nejmenší hodnota R.S.S. i M.S.E. právě u kvadratického trendu, tudíž by měl být lepší než exponenciální trend.

Výsledný vztah, který popisuje závislost mezi proměnnými zahrnující sezónnost pomocí lineárního a exponenciálního trendu, může být popsán následujícím vztahem:

$$Y_t = (b_0 + b_1 t)(\alpha_2 D_{2x} + \alpha_3 D_{3x} + \alpha_4 D_{4x}) + c_1 e^{c_2 q} (\alpha_2 D_{2q} + \alpha_3 D_{3q} + \alpha_4 D_{4q}) + \varepsilon_{x,q}, \quad (61)$$

kde: D_i ... Dummy proměnná vynásobená příslušným obdobím,
 b_0 a b_1 ... hodnoty lineárních koeficientů,
 c_1 a c_2 ... hodnoty exponenciálních koeficientů,
 α_i ... hodnota sezónního indexu,
 x ... období <1;40>,
 q ... období <41;72>.

Po dosazení:

$$Y_t = (732,12 + 161,53t)(1,092D_{2x} + 0,969D_{3x} + 1,057D_{4x}) + 2203,91e^{0,02589q} (1,092D_{2q} + 0,969D_{3q} + 1,057D_{4q}) + \varepsilon_{x,q}. \quad (62)$$

Tento model je tvořen v první části lineárním trendem a ve druhé části (od roku 2010) je tvořen trendem exponenciálním, z tohoto důvodu je sledované období rozděleno na dvě části, na období x , které je sledováno mezi prvním a čtyřicátým obdobím a na období q , které je sledováno mezi čtyřicátým prvním a sedmdesátým druhým obdobím. Dummy proměnné (tabulka č. 17), které jsou v modelu obsaženy pomáhají lépe zakomponovat sezónní složku.

Adaptivní přístupy

Jak bylo zmíněno v kapitole 2.1.2, adaptivní přístupy konstruují trendovou složku jako model, ve kterém se v průběhu času mění jednotlivé parametry. Tato změna parametrů probíhá za předpokladu, že krátké úseky časových řad lze vyrovnat pomocí matematických funkcí.

Metoda klouzavých průměrů

V rámci metody klouzavých průměrů je na data z časové řady použitý tzv. **vážený klouzavý průměr**, jenž je uveden jako vztah (17). Nejprve však bylo nutné vypočítat hodnoty vah, které jsou potřeba k dalšímu výpočtu. Dle vzorce (18) byly vypočteny následující váhy:

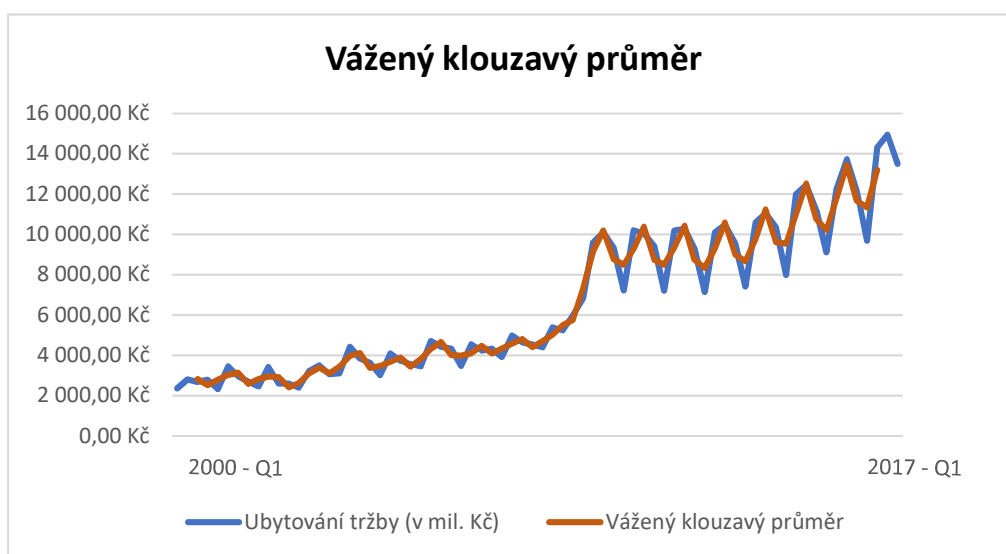
Tabulka 11 - Hodnoty vah (ubytování)

Pomocné výpočty	
váhy W_i	i
-0,0857	-2
0,34286	-1
0,48571	0
0,34286	1
-0,0857	2
1	Kontrola

Zdroj: vlastní zpracování, EXCEL 2019

Výsledné váhy byly použity pro výpočet vyrovnaných hodnot dle vzorce (17). Vyrovnané hodnoty pomocí váženého klouzavého průměru jsou poměrně shodné s daty empirickými, až na některé extrémní hodnoty. Vyrovnané hodnoty v porovnání s empirickými daty lze vidět na následujícím grafu.

Graf 6 - Vážený klouzavý průměr (ubytování)



Zdroj: vlastní zpracování, EXCEL 2019

Jak lze vidět na grafu, u vyrovnaných hodnot chybí první dvě a poslední dvě hodnoty. Jelikož nevyrovnaných hodnot, při metodě váženého klouzavého průměru, bude vždy p (v tomto případě $p = 4$).

Hodnota statistického kritéria **M.S.E. je 365 004,4.**

Metoda exponenciálního vyrovnávání

Při této metodě se počítá se všemi empirickými daty z minulosti. Pro odhad parametrů je použito metody nejmenších čtverců a váhy jednotlivých pozorování se snižují směrem do minulosti. Nejprve je provedeno vyrovnání pomocí dvojitého exponenciálního vyrovnávání.

- **Dvojitě exponenciální vyrovnávání**

U dvojitého exponenciálního vyrovnávání se předpokládá, že trendová složka v krátkých úsecích je popsána lineárním polynomem (12):

$$T_t = \beta_0 + \beta_1 t,$$

K výpočtu počátečních hodnot vyrovnávacích statistik byl nejprve zvolen koeficient „*alfa*“, který má být v rozmezí mezi $\langle 0,7; 1 \rangle$. Poté pomocí vzorců (22) a (23) byly vypočteny počáteční hodnoty.

Po dosazení do vzorců:

$$S_0 = 732,12 - \frac{0,7}{(1 - 0,7)} 161,53$$

$$S_0^{[2]} = 732,12 - \frac{2(0,7)}{(1 - 0,7)} 161,53$$

Výsledné hodnoty počátečních hodnot vyrovnávacích statistik jsou znázorněny v tabulce č. 12.

Tabulka 12 - Počáteční vyr. statistiky dvojité (ubytování)

Koeficient alfa	0,700
Pomocné výpočty	
S_0	355,211
$S_0^{[2]}$	-21,703

Zdroj: vlastní zpracování, EXCEL 2019

Jednoduché a dvojité vyrovnávací statistiky jsou vypočteny pomocí vzorců (24) a (25). Vyrovnané statistiky u dvojitého exponenciálního vyrovnávání spolu s původními hodnotami jsou uvedeny v následující tabulce.

Tabulka 13 - Vyrovnané statistiky dvojité (ubytování)

t	y _t	St	St[2]	t	y _t	St	St[2]
1	2366,64	958,64	272,40	19	3842,66	3579,48	3232,61
2	2806,26	1512,93	644,56	20	3634,86	3596,09	3341,65
3	2671,91	1860,62	1009,38	21	3013,00	3421,16	3365,51
4	2787,99	2138,83	1348,21	22	4095,92	3623,59	3442,93
5	2319,36	2192,99	1601,65	23	3724,21	3653,78	3506,18
6	3463,86	2574,25	1893,43	24	3574,20	3629,9	3543,3
7	2967,32	2692,17	2133,05	25	3450,59	3576,11	3553,14
8	2699,15	2694,27	2301,42	26	4718,09	3918,7	3662,81
9	2463,25	2624,96	2398,48	27	4428,73	4071,71	3785,48
10	3423,17	2864,42	2538,26	28	4328,60	4148,78	3894,47
11	2602,43	2785,83	2612,53	29	3470,03	3945,15	3909,67
12	2602,39	2730,8	2648,01	30	4548,74	4126,23	3974,64
13	2390,90	2628,83	2642,26	31	4241,79	4160,9	4030,52
14	3226,12	2808,01	2691,98	32	4338,16	4214,08	4085,59
15	3507,53	3017,87	2789,75	33	3910,00	4122,85	4096,77
16	3062,23	3031,18	2862,18	34	4976,00	4378,8	4181,38
17	3109,47	3054,66	2919,92	35	4652,00	4460,76	4265,19
18	4428,06	3466,68	3083,95	36	4534,00	4482,73	4330,45
t	y _t	St	St[2]	t	y _t	St	St[2]
37	4401,00	4458,21	4368,78	55	10479,77	9535,4	9155,32
38	5398,00	4740,15	4480,19	56	9553,70	9540,89	9270,99
39	5243,00	4891	4603,43	57	7406,12	8900,46	9159,83
40	5978,00	5217,1	4787,53	58	10609,03	9413,03	9235,79
41	6841,81	5704,51	5062,63	59	11042,21	9901,78	9435,59
42	9597,77	6872,49	5605,59	60	10369,69	10042,16	9617,56
43	10111,57	7844,22	6277,18	61	7988,93	9426,19	9560,15
44	9341,92	8293,53	6882,08	62	11995,37	10196,94	9751,19
45	7227,38	7973,68	7209,56	63	12459,23	10875,63	10088,52
46	10211,39	8645	7640,19	64	11173,37	10964,95	10351,45
47	10050,17	9066,55	8068,1	65	9105,08	10406,99	10368,11
48	9431,74	9176,1	8400,5	66	12285,96	10970,68	10548,88
49	7203,74	8584,4	8455,67	67	13736,19	11800,33	10924,32
50	10187,98	9065,47	8638,61	68	12175,03	11912,74	11220,84
51	10269,61	9426,71	8875,04	69	9683,65	11244,01	11227,8
52	9286,22	9384,56	9027,9	70	14316,09	12165,64	11509,15
53	7143,26	8712,17	8933,18	71	14946,49	12999,89	11956,37
54	10107,17	9130,67	8992,43	72	13488,70	13146,53	12313,42

Zdroj: vlastní zpracování, EXCEL 2019

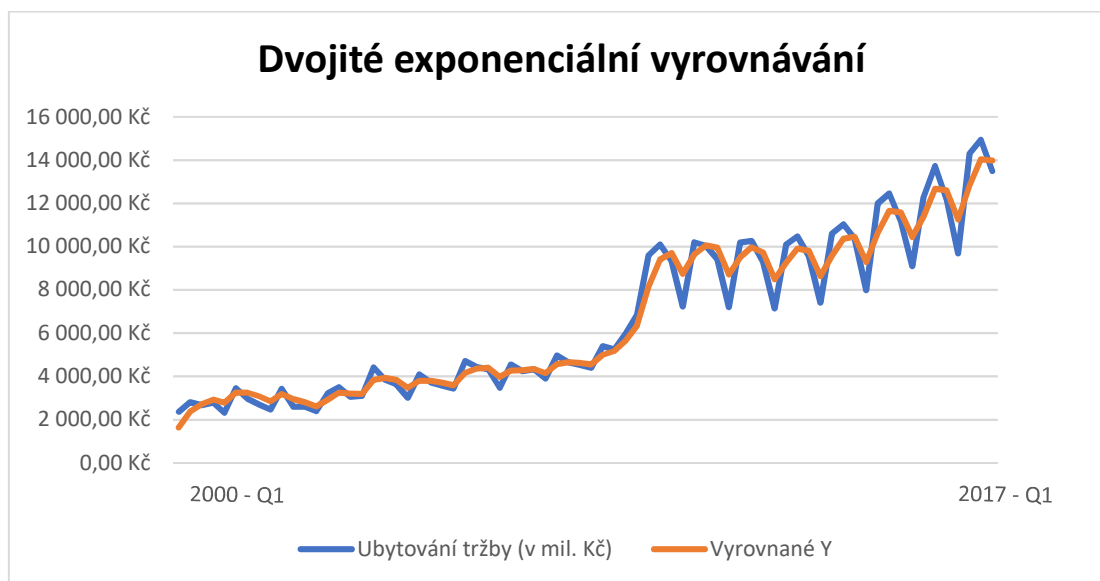
Vyrovnané hodnoty pomocí dvojitého exponenciálního vyrovnávání jsou vypočteny jako dvojnásobek jednoduché statistiky, mínus dvojitá statistika pro jednotlivá období.

Po dosazení hodnot jsou vyrovnané hodnoty vypočteny (např. pro $t=1$) jako:

$$\hat{y}_t = 2(958,64) - 272,4. \quad (63)$$

Porovnání dvojitého exponenciálního vyrovnávání a empirických dat lze vidět na následujícím grafu.

Graf 7 - Dvojité exp. vyrovnávání (ubytování)



Zdroj: vlastní zpracování, EXCEL 2019

Vyrovnané hodnoty pomocí dvojitého exponenciálního vyrovnávání nejsou příliš odpovídající skutečnosti, o čemž hovoří i výsledná hodnota statistického kritéria **M.S.E. 452 474, 17**. Prozatím tedy lze tvrdit, že metoda pomocí vážených klouzavých průměrů je přesnější.

Trojité exponenciální vyrovnávání

U trojitého exponenciálního vyrovnání se předpokládá, že trendová složka v krátkých úsecích je popsána kvadratickým polynomem (13):

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2,$$

K výpočtu počátečních hodnot vyrovnávacích statistik byl opět zvolen koeficient „*alfa*“, který má být v rozmezí mezi $\langle 0,7; 1 \rangle$. Poté byly vypočteny pomocí vzorců (26), (27) a (28) počáteční hodnoty.

Po dosazení do vzorců:

$$S_0 = 2239,93 - \frac{0,7}{(1 - 0,7)} 39,28 + \frac{0,7(1,7)}{2(1 - 0,7)^2} 1,6747$$

$$S_0^{[2]} = 2239,93 - \frac{2(0,7)}{(1 - 0,7)} 39,28 + \frac{2(0,7)(1 + (2(0,7)))}{2(1 - 0,7)^2} 1,6747$$

$$S_0^{[3]} = 2239,93 - \frac{3(0,7)}{(1 - 0,7)} 39,28 + \frac{3(0,7)(1 + (3(0,7)))}{2(1 - 0,7)^2} 1,6747$$

Počáteční hodnoty vyrovnávacích statistik jsou znázorněny v tabulce č. 14.

Tabulka 14 - Počáteční vyrovnávací statistiky trojitě (ubytování)

Koeficient alfa	0,700
Pomocné výpočty	
S_0	2148,365
$S_0^{[2]}$	2056,875
$S_0^{[3]}$	1965,459

Zdroj: vlastní zpracování, EXCEL 2019

Výpočet pokračuje identifikací jednoduché, dvojité a trojitě vyrovnávací statistiky pomocí vzorců (24), (25) a (29), výsledné hodnoty jsou uvedeny v následující tabulce.

Tabulka 15 - Vyrovnané statistiky trojité (ubytování)

t	y _t	St	St[2]	St[3]	t	y _t	St	St[2]	St[3]
1	2366,64	2213,85	2103,97	2007,01	19	3842,66	3581,52	3246,63	3018,23
2	2806,26	2391,57	2190,25	2061,98	20	3634,86	3597,52	3351,90	3118,33
3	2671,91	2475,67	2275,88	2126,15	21	3013,00	3422,17	3372,98	3194,72
4	2787,99	2569,37	2363,92	2197,48	22	4095,92	3624,29	3448,37	3270,82
5	2319,36	2494,37	2403,06	2259,15	23	3724,21	3654,27	3510,14	3342,62
6	3463,86	2785,21	2517,70	2336,72	24	3574,20	3630,25	3546,17	3403,68
7	2967,32	2839,85	2614,35	2420,01	25	3450,59	3576,35	3555,23	3449,15
8	2699,15	2797,64	2669,33	2494,81	26	4718,09	3918,87	3664,32	3513,70
9	2463,25	2697,32	2677,73	2549,68	27	4428,73	4071,83	3786,57	3595,56
10	3423,17	2915,08	2748,93	2609,46	28	4328,60	4148,86	3895,26	3685,47
11	2602,43	2821,28	2770,64	2657,81	29	3470,03	3945,21	3910,24	3752,90
12	2602,39	2755,61	2766,13	2690,31	30	4548,74	4126,27	3975,05	3819,55
13	2390,90	2646,20	2730,15	2702,26	31	4241,79	4160,93	4030,81	3882,93
14	3226,12	2820,17	2757,16	2718,73	32	4338,16	4214,10	4085,80	3943,79
15	3507,53	3026,38	2837,93	2754,49	33	3910,00	4122,87	4096,92	3989,73
16	3062,23	3037,14	2897,69	2797,45	34	4976,00	4378,81	4181,49	4047,25
17	3109,47	3058,83	2946,03	2842,02	35	4652,00	4460,77	4265,27	4112,66
18	4428,06	3469,60	3103,10	2920,35	36	4534,00	4482,74	4330,51	4178,01
t	y _t	St	St[2]	St[3]	t	y _t	St	St[2]	St[3]
37	4401,00	4458,21	4368,82	4235,26	55	10479,77	9535,40	9155,32	8866,01
38	5398,00	4740,15	4480,22	4308,75	56	9553,70	9540,89	9270,99	8987,51
39	5243,00	4891,01	4603,46	4397,16	57	7406,12	8900,46	9159,83	9039,20
40	5978,00	5217,10	4787,55	4514,28	58	10609,03	9413,03	9235,79	9098,18
41	6841,81	5704,52	5062,64	4678,79	59	11042,21	9901,78	9435,59	9199,40
42	9597,77	6872,49	5605,60	4956,83	60	10369,69	10042,16	9617,56	9324,85
43	10111,57	7844,22	6277,18	5352,93	61	7988,93	9426,19	9560,15	9395,44
44	9341,92	8293,53	6882,09	5811,68	62	11995,37	10196,94	9751,19	9502,16
45	7227,38	7973,68	7209,56	6231,05	63	12459,23	10875,63	10088,52	9678,07
46	10211,39	8645,00	7640,19	6653,79	64	11173,37	10964,95	10351,45	9880,08
47	10050,17	9066,55	8068,10	7078,08	65	9105,08	10406,99	10368,11	10026,49
48	9431,74	9176,10	8400,50	7474,81	66	12285,96	10970,68	10548,88	10183,21
49	7203,74	8584,40	8455,67	7769,07	67	13736,19	11800,33	10924,32	10405,54
50	10187,98	9065,47	8638,61	8029,93	68	12175,03	11912,74	11220,84	10650,13
51	10269,61	9426,71	8875,04	8283,46	69	9683,65	11244,01	11227,80	10823,43
52	9286,22	9384,56	9027,90	8506,79	70	14316,09	12165,64	11509,15	11029,15
53	7143,26	8712,17	8933,18	8634,71	71	14946,49	12999,89	11956,37	11307,31
54	10107,17	9130,67	8992,43	8742,02	72	13488,70	13146,53	12313,42	11609,15

Zdroj: vlastní zpracování, EXCEL 2019

Vyrovnané hodnoty trojitého exponenciálního vyrovnávání jsou vypočteny podle vzorce (30), tedy:

$$\hat{y}_t = 3S_t - 3S_t^{[2]} + S_t^{[3]},$$

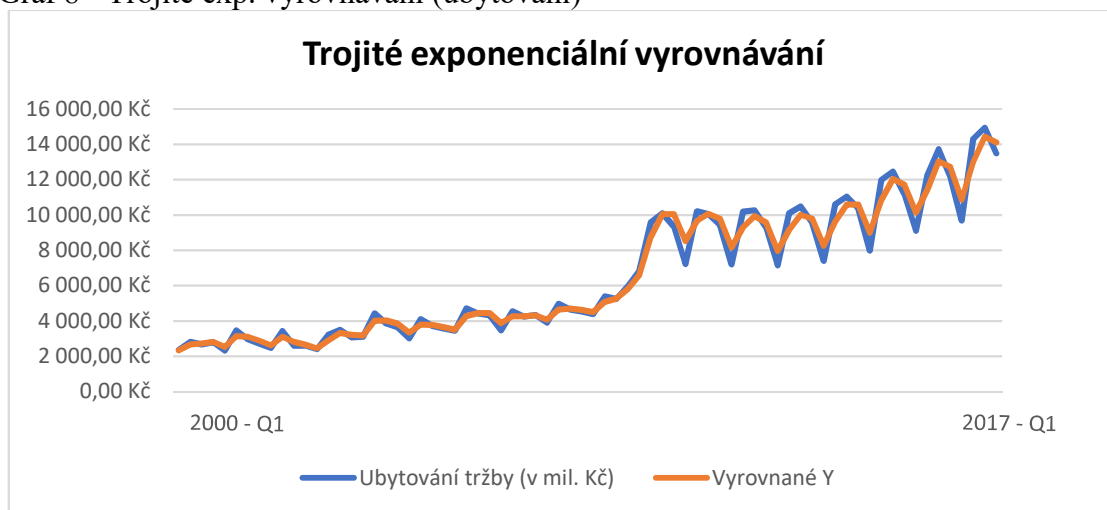
kde: $t = 0,1,2, \dots, n.$

Po dosažení hodnot jsou vyrovnané hodnoty vypočteny (např. pro $t=1$) jako:

$$\hat{y}_t = 3(2213,85) - 3(2103,97) + 2007,01. \quad (64)$$

Porovnání trojitého exponenciálního vyrovnávání a empirických dat lze vidět na následujícím grafu.

Graf 8 - Trojité exp. vyrovnávání (ubytování)



Zdroj: vlastní zpracování, EXCEL 2019

Na grafu lze vidět, že trojité exponenciální vyrovnávání více odpovídá empirickým hodnotám, než je tomu u předchozích metod. Po výpočtu statistického kritéria M.S.E pomocí reziduí., bylo zjištěno, že tento typ vyrovnání je prozatím nejvhodnějším modelem. Hodnota M.S.E. je 277 562, 6.

5.2.2 Modelování sezónnosti (ubytování)

Modelování sezónní složky bylo provedeno v multiplikativním tvaru dekompozice za použití vzorce (37):

$$y_{ij} = T_{ij}S_{ij}\varepsilon_{ij},$$

K výpočtu periodické složky, byl použit tzv. empirický sezónní index. Jelikož jsou data čtvrtletní, délka periody je 4. Je zapotřebí vypočíst tedy 4 indexy pomocí vzorce (39):

$$I_j = \frac{1}{m} \sum_{i=1}^m \frac{y_{ij}}{T_{ij}},$$

Empirické sezónní indexy musí splňovat normalizační pravidlo vzorec (35) a také součet indexů se musí rovnat počtu sezón.

Tabulka 16 - Sezónní indexy (ubytování)

Sezónní indexy	I_j
1. čtvrtletí	0,879
2. čtvrtletí	1,092
3. čtvrtletí	0,969
4. čtvrtletí	1,057
SUMA	4,00

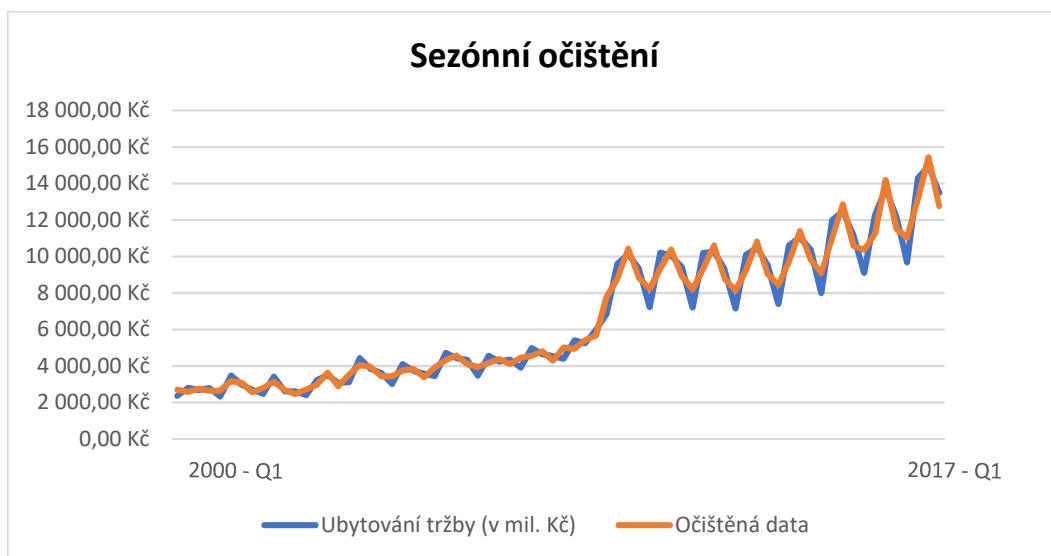
Zdroj: vlastní zpracování, EXCEL 2019

Výsledné hodnoty empirických sezónních indexů značí, že v prvním čtvrtletí jsou tržby za ubytování zpravidla nejmenší. Naopak ve druhém čtvrtletí tržby za ubytování dosahují nejvyšších hodnot za celý rok.

Časovou řadu lze očistit od sezónnosti v multiplikačním tvaru pomocí vztahu: $\frac{y_{i,j}}{I_j}$.

Popisovaná složka časové řady obsahuje pouze trendovou složku a sezónní složku, nikoli však náhodnou složku. Na následujícím grafu je zachyceno porovnání empirických hodnot a sezónně očištěných hodnot.

Graf 9 - Sezónní očištění (ubytování)



Zdroj: vlastní zpracování, EXCEL 2019

Na grafu lze vidět, že očištěné hodnoty jsou podobné empirickým, avšak jsou zde očištěny některé extrémní hodnoty.

Při uvažování zařazení DUMMY proměnných do modelu je použita matice sezónních indexů spolu s „nulovými a jedničkovými“ vektory, ty slouží pro to, aby model uměl vynásobit vektory s odpovídajícím čtvrtletím. Jejich označení je D_1 až D_4 .

Tabulka 17 - DUMMY proměnné (ubytování)

Označení I	I_j	Označení D	DUMMY			
α_1	0,879	D_1	0	0	0	0
α_2	1,092	D_2	0	1	0	0
α_3	0,969	D_3	0	0	1	0
α_4	1,057	D_4	0	0	0	1

Zdroj: vlastní zpracování, EXCEL 2019

Ve výše uvedené tabulce jsou v prvním čtvrtletí všude nuly, proto do výsledného vzorce nebude zahrnuta hodnota sezónního indexu pro první čtvrtletí, to zabrání vzniku multikolinearity.

5.2.3 Modelování trendu (hrubá mzda)

Klasické přístupy

V následující části jsou popsány vybrané klasické přístupy při dekompozici časové řady „hrubé mzdy“. Nejprve jsou znázorněny jednotlivé grafy pro lineární, kvadratickou, exponenciální a mocninou regresi a pomocí SW Excel vypočteny upravené koeficienty determinace. Výsledné hodnoty jsou uvedeny v tabulce č. 18.

Tabulka 18 – Upravené koeficienty determinace (mzdy)

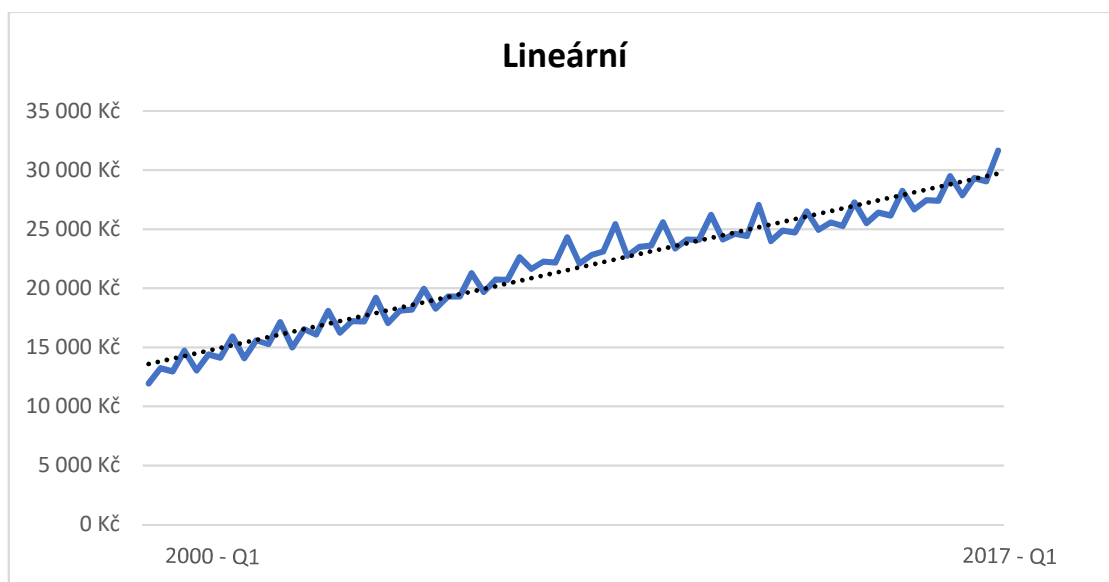
Regrese	Koeficient determinace (upravený) - R^2
Lineární	94,83%
Kvadratická	95,75%
Exponenciální	92,31%
Logistická	83,67%
Mocninná	89,44%

Zdroj: vlastní zpracování, EXCEL 2019

Při pohledu na výsledky koeficientů determinace u této časové řady, je zřejmé, že koeficienty jsou více vyrovnané než u předešlé řady. Avšak dvě regrese budou vyloučeny, jelikož mají nejmenší upravený koeficient determinace, jedná se o regresi logistickou a mocninou. U ostatních regresí bude znázorněn průběh na grafu a vypočítány koeficienty

b_0 , b_1 a popřípadě b_3 . Pro zjištění kvality vypovídající hodnoty, budou pro každou regresi vypočtena statistická kritéria reziduálního součtu čtverců (R.S.S.) a střední čtvercové chyby odhadu (M.S.E.). Je zde použit obdobný způsob znázornění a výpočtů, jako u předešlé časové řady.

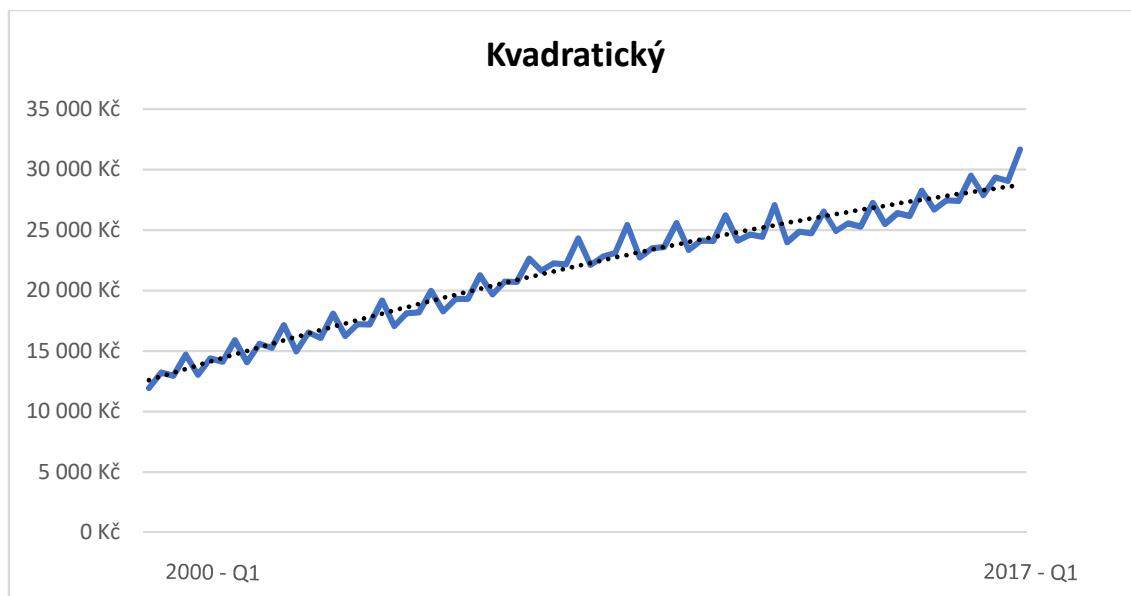
Graf 10 - Lineární trend (mzdy)



Zdroj: vlastní zpracování, EXCEL 2019

Lineární trend má podobu rovnice (12), tedy $T_t = \beta_0 + \beta_1 t$.

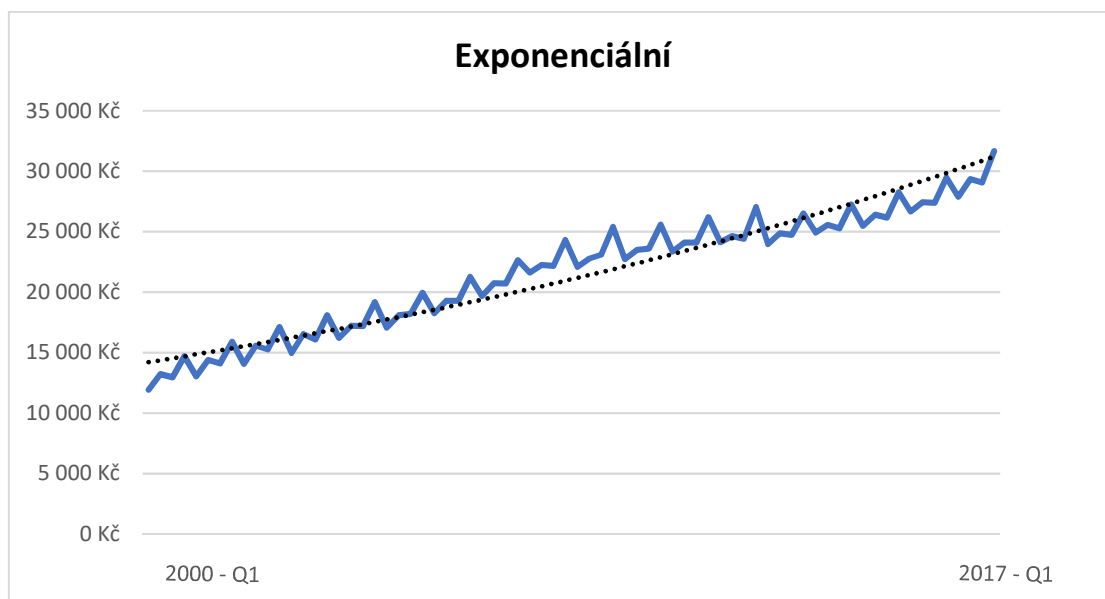
Graf 11 - Kvadratický trend (mzdy)



Zdroj: vlastní zpracování, EXCEL 2019

Kvadratický trend má podobu rovnice (13), tedy $T_t = \beta_0 + \beta_1 t + \beta_2 t^2$.

Graf 12 - Exponenciální trend (mzdy)



Zdroj: vlastní zpracování, EXCEL 2019

Exponenciální trend má podobu rovnice (14), tedy $T_t = \beta_0 \beta_1^t$.

Z těchto grafů a také z výsledku hodnot determinace lze usoudit, že nejlépe prokládá empirická data právě kvadratický trend. Dále byly pomocí programu Excel vypočítány hodnoty koeficientů pro každý trend. Výsledné hodnoty jsou uvedeny v následující tabulce.

Tabulka 19 - Hodnoty koeficientů b (mzdy)

Lineární	b1	b0	
	227,137	13 361,164	
Kvadratický	b2	b1	b0
	-1,201	314,814	12 279,813
Exponenciální	b2	ln(b1)	b1
	0,011	9,551	14 065,733

Zdroj: vlastní zpracování, EXCEL 2019

Ke zjištění, jakou mají modely vypovídající schopnost, a které budou nejlépe predikovat data, jsou vypočtena statistická kritéria reziduální součet čtverců a střední čtvercová chyba odhadu. Pro výpočet kritéria R.S.S. byl použit vzorec $RSS = \sum_{t=1}^n (y_t - T_t)^2$ a pro

výpočet M.S.E. byl použit vzorec $MSE = \frac{RSS}{n}$. Výsledné hodnoty jsou uvedeny v následující tabulce.

Tabulka 20 - Výsledky kritérií (mzdy)

VÝSLEDKY			
Ukazatel	Lineární	Kvadratický	Exponenciální
R.S.S.	87 429 374,36	71 937 735,03	141 529 799,81
M.S.E.	1 214 296,87	999 135,21	1 965 691,66

Zdroj: vlastní zpracování, EXCEL 2019

Nejlepším modelem je ten, který má nejmenší hodnotu zvoleného kritéria. V případě této časové řady vyšly také nejmenší hodnoty u kvadratického trendu. Proto lze tvrdit, že z těchto tří regresí, nejlépe empirická data prokládá kvadratický trend.

Výsledný vztah, který popisuje závislost mezi proměnnými zahrnující sezónnost v kvadratickém trendu, může být popsán následujícím vztahem:

$$Y_t = (b_0 + b_1t + b_2t^2)(\alpha_2D_{2t} + \alpha_3D_{3t} + \alpha_4D_{4t}) + \varepsilon_t, \quad (65)$$

kde: D_{it} ... dummy proměnná vynásobená příslušným obdobím,

b_0, b_1 a b_2 ... hodnoty kvadratických koeficientů,

α_i ... hodnota sezónního indexu

t ... období

Po dosazení:

$$Y_t = (12279,81 + 314,81t + (-1,2t^2))(1,092D_{2t} + 0,969D_{3t} + 1,057D_{4t}) + \varepsilon_t. \quad (66)$$

Model je tvořen dle kvadratického trendu, jelikož tento trend má nejnižší upravený koeficient determinace ze všech matematických funkcí. Do modelu byly začleněny Dummy proměnné z tabulky č. 26.

Adaptivní přístupy

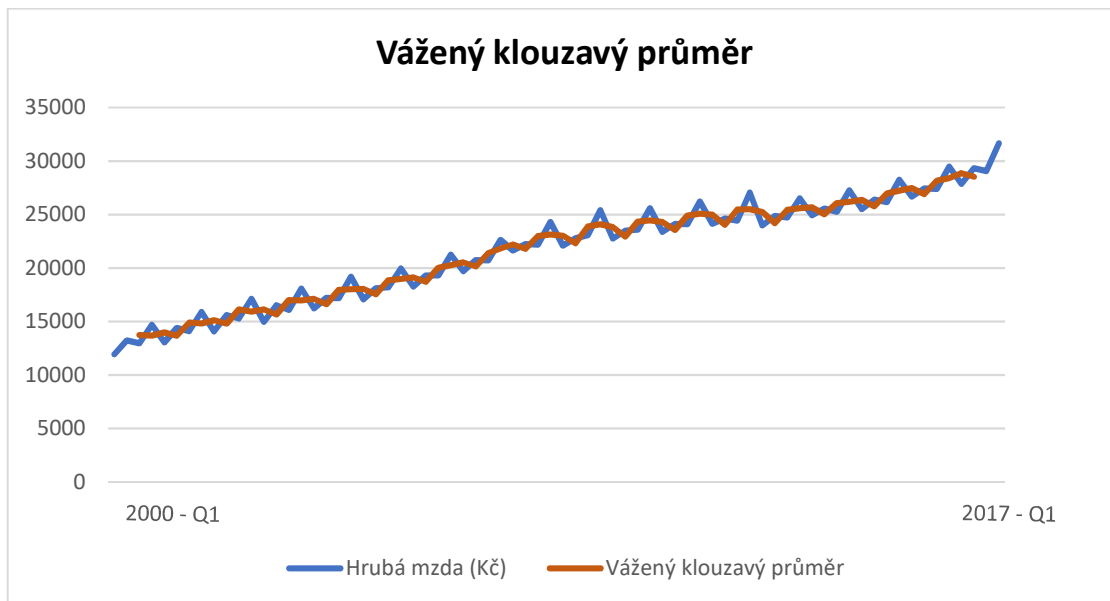
Metoda klouzavých průměrů

I v tomto případě byl použit tzv. vážený klouzavý průměr, který je v práci uveden jako vzorec (17). I u této řady bylo nutné vypočítat hodnoty vah, ale jelikož se jedná také o

čtvrtletní data, váhy zůstaly stejné jako u předchozí časové řady, proto zde již neuvádím výsledné hodnoty, viz tabulka č. 11.

Vyrovnané hodnoty pomocí vážených klouzavých průměrů jsou vypočteny pomocí vzorce (17). Vyrovnané hodnoty spolu s empirickými hodnotami lze vidět na následujícím grafu.

Graf 13 - Vážený klouzavý průměr (mzdy)



Zdroj: vlastní zpracování, EXCEL 2019

Na grafu lze vidět, že křivka vyrovnaných hodnot, pomocí váženého klouzavého průměru, je „osekána“ od sebemenších výkyvů. Stejně jako u předchozí časové řady, i tady chybí první a poslední dvě hodnoty, jelikož se jedná opět o $p=4$.

Hodnota statistického kritéria **M.S.E. je 788 988, 8.**

Dvojitě exponenciální vyrovnávání

Jak již bylo uvedeno, dvojitě exponenciální vyrovnávání je v krátkých úsecích trendu popsáno lineárním polynomem (12). Opět je pro výpočet subjektivně zvolena hodnota koeficientu „*alfa*“, je nutné, aby hodnota byla z rozmezí $<0,7;1>$. Dále jsou vypočteny počáteční hodnoty vyrovnávacích statistik pomocí vzorců (22) a (23).

Po dosazení do vzorců:

$$S_0 = 13361 - \frac{0,7}{(1 - 0,7)} 227,14$$

$$S_0^{[2]} = 13361 - \frac{2(0,7)}{(1 - 0,7)} 227,14$$

Výsledné hodnoty počátečních hodnot vyrovnávacích statistik jsou znázorněny v následující tabulce.

Tabulka 21 - Počáteční vyr. statistiky dvojitě (mzdy)

Koeficient alfa	0,700
Pomocné výpočty	
S_0	12831,010
$S_0^{[2]}$	12301,010

Zdroj: vlastní zpracování, EXCEL 2019

Obdobně jsou vypočteny jednoduché a dvojitě vyrovnávací statistiky za použití vzorců (24) a (25), které jsou uvedeny v následující tabulce.

Tabulka 22 - Vyrovnané statistiky dvojité (mzdy)

t	y _t	St	St[2]	t	y _t	St	St[2]
1	11941	12564,00	12379,91	19	17190	16823,50	16347,90
2	13227	12762,90	12494,81	20	19183	17531,35	16702,93
3	12963	12822,93	12593,25	21	17067	17392,04	16909,67
4	14717	13391,15	12832,62	22	18112	17608,03	17119,17
5	13052	13289,41	12969,65	23	18203	17786,52	17319,38
6	14391	13619,88	13164,72	24	19963	18439,46	17655,40
7	14117	13769,02	13346,01	25	18270	18388,63	17875,37
8	15908	14410,71	13665,42	26	19300	18662,04	18111,37
9	14083	14312,40	13859,52	27	19305	18854,93	18334,44
10	15599	14698,38	14111,17	28	21269	19579,15	18707,85
11	15268	14869,27	14338,60	29	19687	19611,50	18978,95
12	17133	15548,39	14701,54	30	20740	19950,05	19270,28
13	14986	15379,67	14904,98	31	20721	20181,34	19543,60
14	16529	15724,47	15150,82	32	22641	20919,24	19956,29
15	16088	15833,53	15355,64	33	21632	21133,07	20309,32
16	18096	16512,27	15702,63	34	22246	21466,95	20656,61
17	16231	16427,89	15920,20	35	22181	21681,16	20963,97
18	17223	16666,42	16144,07	36	24309	22469,51	21415,64
t	y _t	St	St[2]	t	y _t	St	St[2]
37	22108	22361,06	21699,26	55	24735	24838,61	24780,96
38	22796	22491,54	21936,95	56	26525	25344,53	24950,03
39	23091	22671,38	22157,28	57	24931	25220,47	25031,16
40	25418	23495,37	22558,70	58	25569	25325,03	25119,32
41	22738	23268,16	22771,54	59	25279	25311,22	25176,89
42	23504	23338,91	22941,75	60	27261	25896,15	25392,67
43	23600	23417,24	23084,40	61	25497	25776,41	25507,79
44	25591	24069,37	23379,89	62	26408	25965,89	25645,22
45	23372	23860,16	23523,97	63	26163	26025,02	25759,16
46	24116	23936,91	23647,85	64	28258	26694,91	26039,89
47	24107	23987,94	23749,88	65	26683	26691,34	26235,32
48	26211	24654,86	24021,37	66	27452	26919,54	26440,59
49	24131	24497,70	24164,27	67	27396	27062,48	26627,15
50	24627	24536,49	24275,93	68	29491	27791,03	26976,32
51	24439	24507,24	24345,33	69	27880	27817,72	27228,74
52	27055	25271,57	24623,20	70	29335	28272,91	27541,99
53	23985	24885,60	24701,92	71	29058	28508,43	27831,92
54	24877	24883,02	24756,25	72	31661	29454,20	28318,61

Zdroj: vlastní zpracování, EXCEL 2019

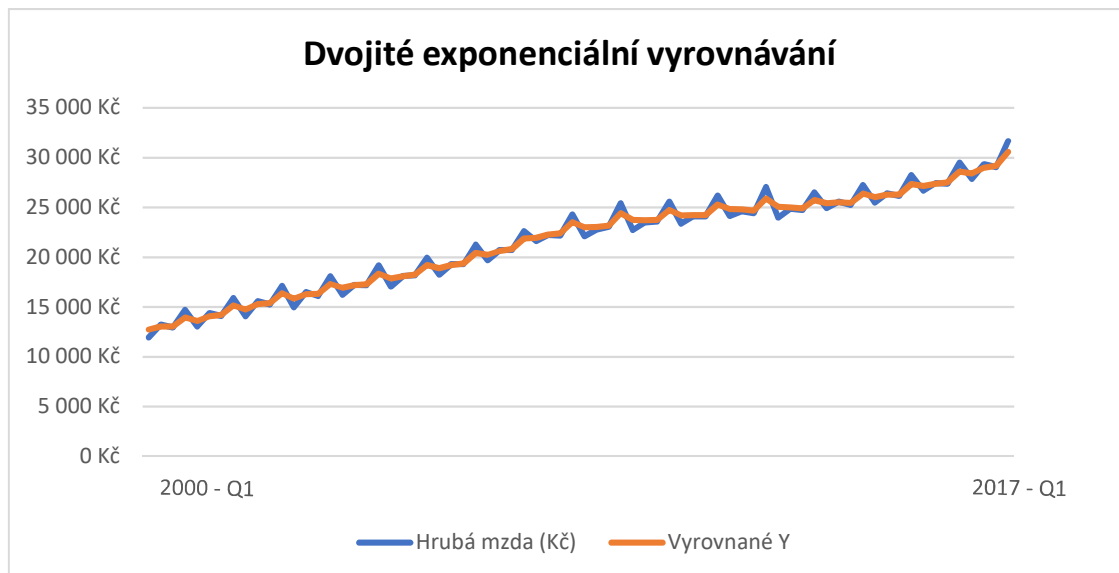
Vyrovnané hodnoty jsou vypočítány za použití vyrovnaných statistik tímto způsobem: dvojnásobek jednoduché vyrovnané statistiky mínus dvojitá vyrovnávací statistika pro jednotlivá období.

Po dosažení vypočtených hodnot jsou vyrovnané hodnoty vypočteny (např. pro t=1) jako:

$$\hat{y}_t = 2(12564) - 12379,91. \quad (67)$$

Pro představu, jak vyrovnané hodnoty vypadají s porovnáním s empirickými daty, je uveden graf č. 14, který tuto situaci vystihuje.

Graf 14 - Dvojitě exp. vyrovnávání (mzdy)



Zdroj: vlastní zpracování, EXCEL 2019

Při pohledu na graf, je křivka vyrovnaných hodnot více vystihující než u metody vážených klouzavých průměrů. Lze vidět, že vyrovnané hodnoty nedosahují takových extrémů jako empirická data. O větší kvalitě této metody hovoří i výsledná hodnota **M.S.E. 328 850, 59**.

Trojité exponenciální vyrovnávání

Trojité exponenciální vyrovnávání je v krátkých úsecích trendu popsáno kvadratickým polynomem (13). Opět byla pro výpočet subjektivně zvolena hodnota koeficientu „alfa“ $\langle 0,7; 1 \rangle$. Výpočet pokračuje identifikací počátečních hodnot vyrovnávacích statistik pomocí vzorců (26), (27) a (28).

Po dosazení do vzorců:

$$S_0 = 12279,81 - \frac{0,7}{(1 - 0,7)} 314,81 + \frac{0,7(1,7)}{2(1 - 0,7)^2} (-1,2011)$$

$$S_0^{[2]} = 12279,81 - \frac{2(0,7)}{(1 - 0,7)} 314,81 + \frac{2(0,7)(1 + (2(0,7)))}{2(1 - 0,7)^2} (-1,2011)$$

$$S_0^{[3]} = 12279,81 - \frac{3(0,7)}{(1 - 0,7)} 314,81 + \frac{3(0,7)(1 + (3(0,7)))}{2(1 - 0,7)^2} (-1,2011)$$

Výsledné hodnoty počátečních vyrovnávacích statistik jsou uvedeny v následující tabulce.

Tabulka 23 - Počáteční vyr. statistiky trojitě (mzdy)

Koeficient alfa	0,700
Pomocné výpočty	
S_0	11545,180
$S_0^{[2]}$	10810,500
$S_0^{[3]}$	10075,770

Zdroj: vlastní zpracování, EXCEL 2019

Poté byly vypočteny jednoduché, dvojité a trojitě vyrovnávací statistiky za použití vzorců (24), (25) a (29). Vyrovnávací statistiky uvádí tabulka č. 24.

Tabulka 24 - Vyrovnané statistiky trojité (mzdy)

t	y _t	St	St[2]	St[3]	t	y _t	St	St[2]	St[3]
1	11941	11663,93	11066,53	10372,99	19	17190	16822,03	16337,84	15769,70
2	13227	12132,85	11386,43	10677,02	20	19183	17530,32	16695,59	16047,47
3	12963	12381,89	11685,07	10979,44	21	17067	17391,32	16904,31	16304,52
4	14717	13082,43	12104,27	11316,89	22	18112	17607,53	17115,27	16547,74
5	13052	13073,30	12394,98	11640,32	23	18203	17786,17	17316,54	16778,38
6	14391	13468,61	12717,07	11963,34	24	19963	18439,22	17653,35	17040,87
7	14117	13663,13	13000,89	12274,61	25	18270	18388,45	17873,88	17290,77
8	15908	14336,59	13401,60	12612,70	26	19300	18661,92	18110,29	17536,63
9	14083	14260,51	13659,27	12926,67	27	19305	18854,84	18333,66	17775,74
10	15599	14662,06	13960,11	13236,70	28	21269	19579,09	18707,29	18055,20
11	15268	14843,84	14225,23	13533,26	29	19687	19611,46	18978,54	18332,20
12	17133	15530,59	14616,84	13858,33	30	20740	19950,02	19269,98	18613,54
13	14986	15367,21	14841,95	14153,42	31	20721	20181,32	19543,38	18892,49
14	16529	15715,75	15104,09	14438,62	32	22641	20919,22	19956,14	19211,58
15	16088	15827,42	15321,09	14703,36	33	21632	21133,06	20309,21	19540,87
16	18096	16508,00	15677,16	14995,50	34	22246	21466,94	20656,53	19875,57
17	16231	16424,90	15901,48	15267,30	35	22181	21681,16	20963,92	20202,07
18	17223	16664,33	16130,34	15526,21	36	24309	22469,51	21415,60	20566,13
t	y _t	St	St[2]	St[3]	t	y _t	St	St[2]	St[3]
37	22108	22361,06	21699,23	20906,06	55	24735	24838,61	24780,96	24569,47
38	22796	22491,54	21936,93	21215,32	56	26525	25344,53	24950,03	24683,64
39	23091	22671,38	22157,26	21497,90	57	24931	25220,47	25031,16	24787,90
40	25418	23495,36	22558,69	21816,14	58	25569	25325,03	25119,32	24887,32
41	22738	23268,16	22771,53	22102,76	59	25279	25311,22	25176,89	24974,19
42	23504	23338,91	22941,74	22354,45	60	27261	25896,15	25392,67	25099,74
43	23600	23417,24	23084,39	22573,43	61	25497	25776,41	25507,79	25222,15
44	25591	24069,37	23379,88	22815,37	62	26408	25965,89	25645,22	25349,07
45	23372	23860,16	23523,97	23027,95	63	26163	26025,02	25759,16	25472,10
46	24116	23936,91	23647,85	23213,92	64	28258	26694,91	26039,89	25642,44
47	24107	23987,94	23749,87	23374,71	65	26683	26691,34	26235,32	25820,30
48	26211	24654,86	24021,37	23568,70	66	27452	26919,54	26440,59	26006,39
49	24131	24497,70	24164,27	23747,37	67	27396	27062,48	26627,15	26192,62
50	24627	24536,49	24275,93	23905,94	68	29491	27791,03	26976,32	26427,73
51	24439	24507,24	24345,33	24037,76	69	27880	27817,72	27228,74	26668,03
52	27055	25271,57	24623,20	24213,39	70	29335	28272,91	27541,99	26930,22
53	23985	24885,60	24701,92	24359,95	71	29058	28508,43	27831,92	27200,73
54	24877	24883,02	24756,25	24478,84	72	31661	29454,20	28318,61	27536,09

Zdroj: vlastní zpracování, EXCEL 2019

Vyrovnané hodnoty trojitého exponenciálního vyrovnávání se vypočtou dle vzorce (30), tedy:

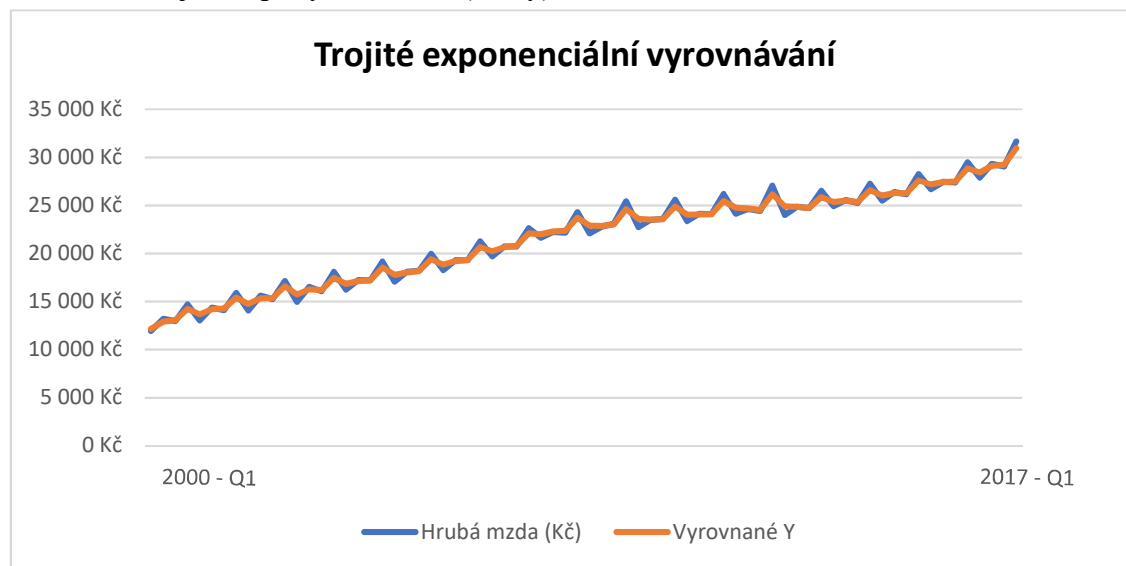
$$\hat{y}_t = 3S_t - 3S_t^{[2]} + S_t^{[3]},$$

Po dosažení vypočtených hodnot jsou vyrovnané hodnoty vypočteny (např. pro $t=1$) jako:

$$\hat{y}_t = 3(11663,93) - 3(11066,53) + 10372,99. \quad (68)$$

Trojité exponenciální vyrovnávání s porovnáním s empirickými daty lze vidět v následujícím grafu.

Graf 15 - Trojité exp. vyrovnávání (mzdy)



Zdroj: vlastní zpracování, EXCEL 2019

Z grafu č. 15 lze vidět, že křivka trojitého exponenciálního vyrovnávání je téměř totožná s empirickými daty nebo se k nim přibližuje. O čemž svědčí výsledek sumy druhých mocnin reziduí, **M.S.E. 209 678, 86**.

5.2.4 Modelování sezónnosti (hrubá mzda)

Pro modelování sezónnosti byl opět použit multiplikativní tvar dekompozice, za použití vzorce (37):

$$y_{ij} = T_{ij}S_{ij}\varepsilon_{ij},$$

K identifikaci sezónní složky byly vypočteny sezónní faktory, které byly vypočteny pomocí vážených klouzavých průměrů v jednotlivých sezónách.

Výpočet čtyř potřebných indexů byl vypočten podle vzorce (39).

Tabulka 25 - Sezónní indexy (mzdy)

Sezónní indexy	I_j
1. čtvrtletí	0,954
2. čtvrtletí	1,031
3. čtvrtletí	0,962
4. čtvrtletí	1,053
SUMA	4,00

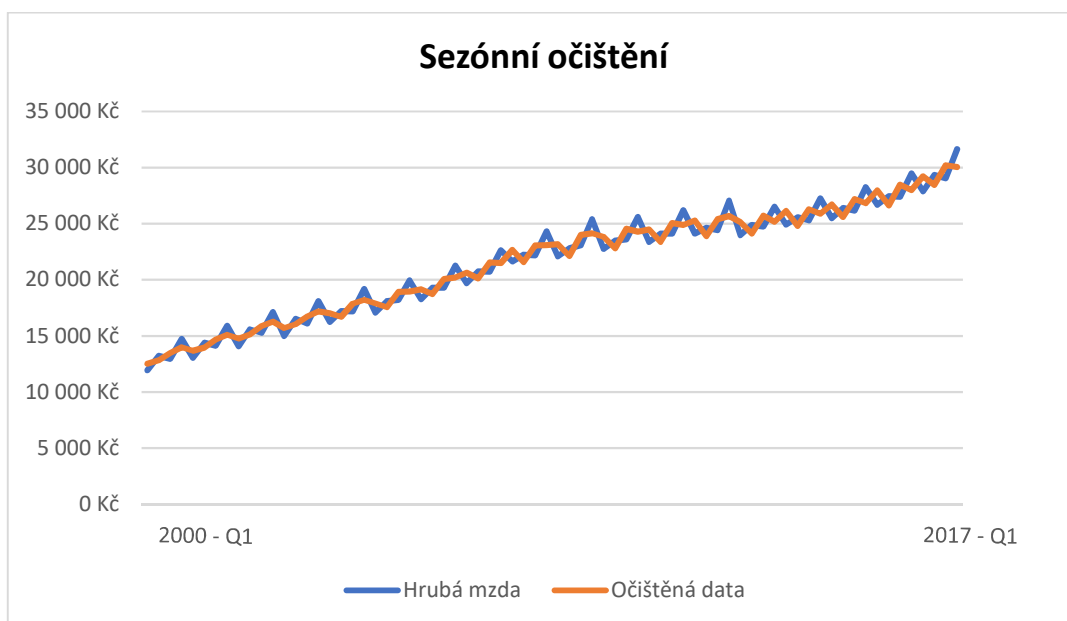
Zdroj: vlastní zpracování, EXCEL 2019

Výsledné hodnoty musí splňovat normalizační pravidlo (35) a také součet indexů se musí rovnat počtu sezón. Obě tato pravidla jsou splněna. Výsledné hodnoty sezónních indexů značí, že nejvyšší mzdy v rámci jednoho roku jsou ve čtvrtém čtvrtletí. Druhé nejvyšší mzdy jsou pak ve druhém čtvrtletí roku. Nejmenších hodnot dosahují mzdy obvykle v prvním čtvrtletí.

Časovou řadu lze očistit od sezónnosti v multiplikativním tvaru dekompozice pomocí vztahu: $\frac{y_{t,j}}{I_j}$. Tato složka časové řady obsahuje také jen trendovou a sezónní složku, bez náhodné složky.

V grafu č. 16 je uvedeno porovnání empirických hodnot s hodnotami sezónně očištěnými.

Graf 16 - Sezónní očištění (mzdy)



Zdroj: vlastní zpracování, EXCEL 2019

Na výše uvedeném grafu lze vidět očištění od extrémů v empirických hodnotách, tedy sezónně očištěnou časovou řadu hrubé mzdy.

Při uvažování zařazení DUMMY proměnných do modelu je opět použita matice sezónních indexů spolu s „nulovými a jedničkovými“ vektory, ty slouží pro to, aby model uměl vynásobit vektory s odpovídajícím čtvrtletím. Jejich označení je D_1 až D_4 .

Tabulka 26 - DUMMY proměnné (mzdy)

Označení I	I_j	Označení D	DUMMY			
α_1	0,954	D_1	0	0	0	0
α_2	1,031	D_2	0	1	0	0
α_3	0,962	D_3	0	0	1	0
α_4	1,053	D_4	0	0	0	1

Zdroj: vlastní zpracování, EXCEL 2019

5.3 Box-Jenkinsova metodologie

Metody Box-Jenkinsovy metodologie zkoumají v časových řadách vztahy mezi náhodnými pozorováními, za předpokladu, že pozorování jsou na sobě závislá. Jelikož obě vybrané časové řady jsou ekonomického typu, je tento předpoklad splněný. V této kapitole budou uvedeny pouze modely ARIMA a SARIMA pro obě časové řady.

K vypracování metody ARIMA byl použit doplněk programu Excel – XLSTAT od společnosti Addinsoft. Ke zpracování metody SARIMA byl použit doplněk programu Excel – NumXL od společnosti Spider Financial. Při analýze časových řad budou, dle výsledků grafů ACF a PACF, použity následující metody:

- ARIMA (1,1,1), zde je zastoupena autoregresní (AR) i integrační (I) část a také klouzavé průměry (MA).
- ARIMA (1,1,0), zde je zastoupena jen autoregresní (AR) a integrační (I) část.
- ARIMA (0,1,1), zde je naopak zastoupena jen integrační část (I) a klouzavé průměry (MA).
- SARIMA (1,1,0), kromě sezónních kolísání je zde zastoupena autoregresní (AR) a integrační (I) část.
- SARIMA (0,1,1), kromě sezónních kolísání je zde zastoupena integrační (I) část a klouzavé průměry (MA).

5.3.1 ARIMA (ubytování)

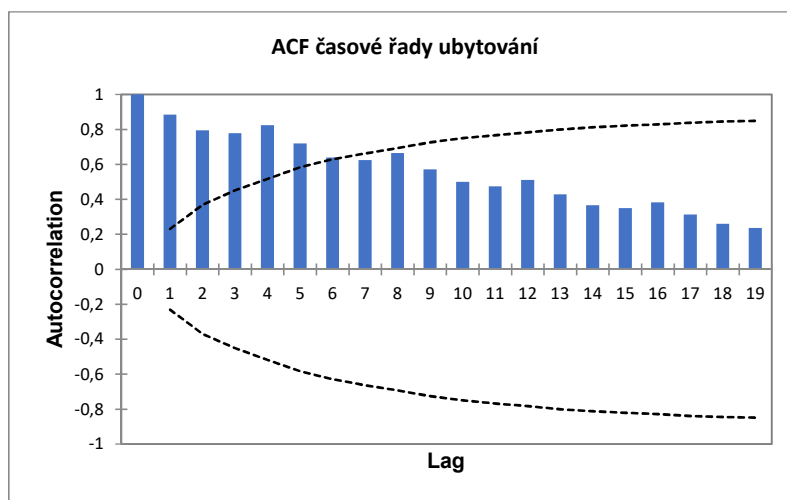
ARIMA

Vzorec tohoto modelu má podobu rovnice (58), tedy

$$y_t = \Phi_p y_{t-1} + \dots + \Phi_p y_{t-p} + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \dots - \Theta_q \varepsilon_{t-q}.$$

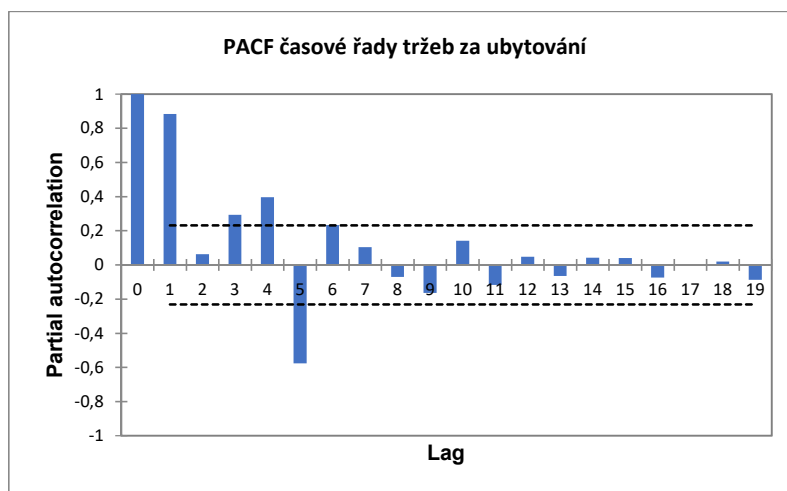
U tohoto tvaru rovnice je důležité nejprve ověřit stacionaritu časové řady tržeb z ubytování pomocí grafu ACF (autokorelační funkce) a PACF (parciální autokorelační funkce).

Graf 17 - ACF časové řady tržeb z ubytování



Zdroj: vlastní zpracování, XLSTAT 2019

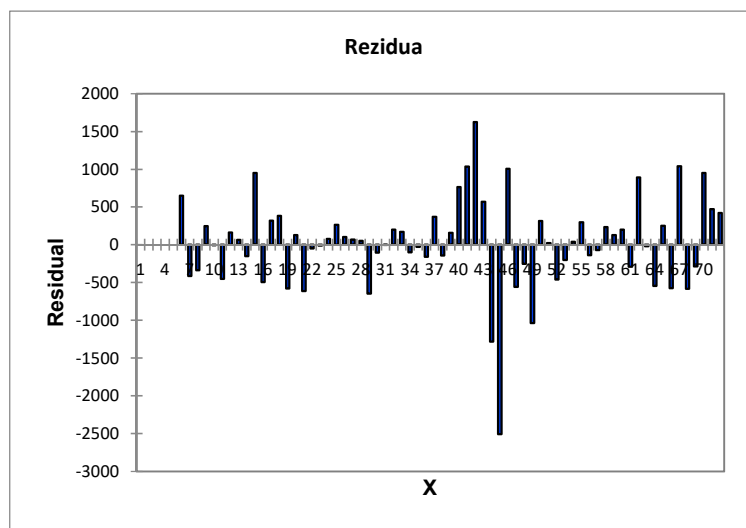
Graf 18 - PACF časové řady tržeb z ubytování



Zdroj: vlastní zpracování, XLSTAT 2019

Z výše uvedených grafů ACF a PACF vyplývá, že lze použít **jednotkové** parametry (např. 1,1,1) u modelu ARIMA, jelikož ACF má exponenciální pokles začínající na pozici 1 a PACF má oscilující pokles začínající na pozici 1. Při zpracování dat, vzniknou rezidua, která jsou prezentována v následujícím grafu.

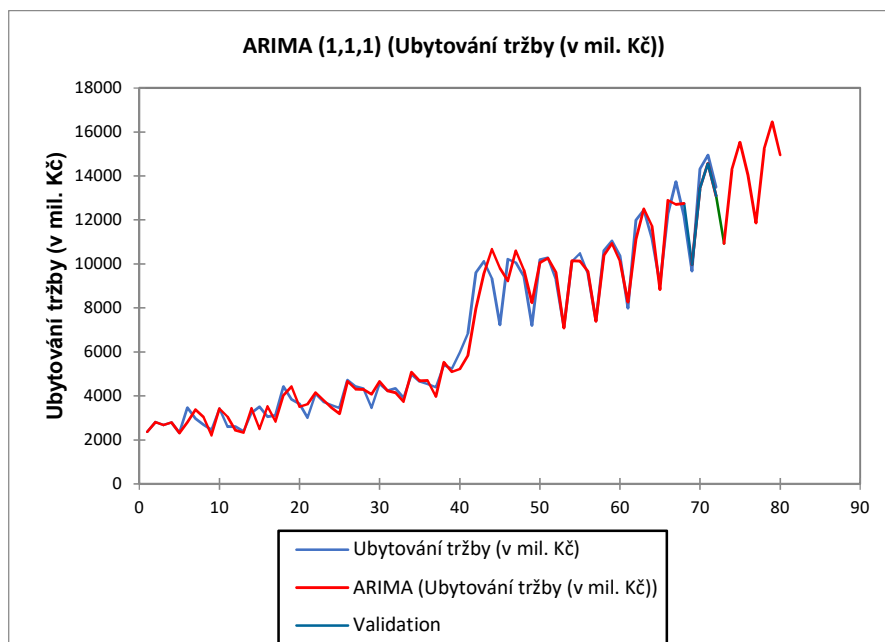
Graf 19 - Rezidua ubytování



Zdroj: vlastní zpracování, XLSTAT 2019

V následujícím grafu je zachyceno porovnání empirických dat spolu s křivkou ARIMA.

Graf 20 – ARIMA (1,1,1) (ubytování)



Zdroj: vlastní zpracování, XLSTAT 2019

Na grafu lze vidět, že tato metoda je poměrně výstižná a dobře popisuje trend časové řady. V následující části budou u každého modelu uvedeny pouze základní informace a zhodnocení pomocí kritéria M.S.E. a M.A.P.E, graf jednotlivých metod bude uveden až v kapitole č. 6 „Extrapolace časových řad“. U nevhodnějšího modelu bude uvedena odpovídající rovnice spolu s vypočtenými hodnotami pomocí dané metody.

Model (1,1,1)

V následující tabulce jsou uvedeny základní informace o modelu ARIMA (1,1,1).

Tabulka 27 - Základní hodnoty ARIMA (1,1,1) ubytování

Parametr	Hodnota	Standardní chyba	Spodní hranice (95%)	Horní hranice (95%)
AR(1)	-0,102	0,758	-1,588	1,383
SAR(1)	-0,058	0,340	-0,725	0,608
MA(1)	0,188	0,743	-1,267	1,644
SMA(1)	-0,286	0,322	-0,918	0,346

Zdroj: vlastní zpracování, XLSTAT 2019

Kritéria hodnocení vhodnosti modelu jsou uvedeny v tabulce č. 28.

Tabulka 28 - Hodnotící kritéria ARIMA (1,1,1) ubytování

Ukazatel	Hodnota
M.S.E.	358 119, 1911
M.A.P.E.	6,5234

Zdroj: vlastní zpracování, XLSTAT 2019

Model (1,1,0)

V následující tabulce jsou uvedeny základní informace o modelu ARIMA (1,1,0).

Tabulka 29 - Základní hodnoty ARIMA (1,1,0) ubytování

Parametr	Hodnota	Standardní chyba	Spodní hranice (95%)	Horní hranice (95%)
AR(1)	0,084	0,127	-0,165	0,333
SAR(1)	-0,311	0,123	-0,552	-0,070

Zdroj: vlastní zpracování, XLSTAT 2019

Kritéria hodnocení vhodnosti modelu jsou uvedeny v tabulce č. 30.

Tabulka 30 - Hodnotící kritéria ARIMA (1,1,0) ubytování

Ukazatel	Hodnota
M.S.E.	358 502,9
M.A.P.E.	6,5218

Zdroj: vlastní zpracování, XLSTAT 2019

Model (0,1,1)

V následující tabulce jsou uvedeny základní informace o modelu ARIMA (0,1,1).

Tabulka 31 - Základní hodnoty ARIMA (0,1,1) ubytování

Parametr	Hodnota	Standardní chyba	Spodní hranice (95%)	Horní hranice (95%)
MA(1)	0,089	0,131	-0,168	0,345
SMA(1)	-0,333	0,117	-0,563	-0,103

Zdroj: vlastní zpracování, XLSTAT 2019

Kritéria hodnocení vhodnosti modelu jsou uvedeny v tabulce č. 32.

Tabulka 32 - Hodnotící kritéria ARIMA (0,1,1) ubytování

Ukazatel	Hodnota
M.S.E.	358 513,4
M.A.P.E.	6,5629

Zdroj: vlastní zpracování, XLSTAT 2019

Na základě výsledků hodnotících kritérií bylo zjištěno, že nejvhodnější model pro popis empirických dat je model ARIMA (1,1,0), jelikož má nejmenší hodnotu M.S.E. a také má nejmenší hodnotu M.A.P.E. Výsledný vztah je popsán:

$$y_t = -0,311 + y_{t-1} + 0,084 (y_{t-1} - y_{t-2}) \quad (69)$$

Část vypočtených hodnot pomocí metody ARIMA (1,1,0) je ilustrativně uvedena v následující tabulce.

Tabulka 33 - Hodnoty ARIMA (1,1,0) ubytování

t	Tržby	ARIMA (1,1,0)	t	Tržby	ARIMA (1,1,0)	t	Tržby	ARIMA (1,1,0)
1	2366,64	2366,64	19	3842,66	4423,12	37	4401,00	3978,25
2	2806,26	2806,26	20	3634,86	3510,90	38	5398,00	5525,82
3	2671,91	2671,91	21	3013,00	3628,97	39	5243,00	5092,20
4	2787,99	2787,99	22	4095,92	4151,14	40	5978,00	5224,31
5	2319,36	2319,36	23	3724,21	3791,77	41	6841,81	5838,03
6	3463,86	2811,20	24	3574,20	3456,93	42	9597,77	7970,28
7	2967,32	3381,25	25	3450,59	3190,30	43	10111,57	9554,55
8	2699,15	3047,95	26	4718,09	4650,03	44	9341,92	10660,79
9	2463,25	2208,32	27	4428,73	4308,27	45	7227,38	9811,28
10	3423,17	3427,16	28	4328,60	4292,20	46	10211,39	9232,51
11	2602,43	3061,07	29	3470,03	4074,81	47	10050,17	10601,12
12	2602,39	2436,16	30	4548,74	4650,64	48	9431,74	9728,08
13	2390,90	2325,59	31	4241,79	4241,79	49	7203,74	8235,36
14	3226,12	3435,27	32	4338,16	4145,83	50	10187,98	10048,99
15	3507,53	2509,91	33	3910,00	3744,74	51	10269,61	10261,49
16	3062,23	3526,90	34	4976,00	5083,27	52	9286,22	9625,94
17	3109,47	2831,88	35	4652,00	4687,55	53	7143,26	7085,89
18	4428,06	4024,71	36	4534,00	4704,42	54	10107,17	10150,58

Zdroj: vlastní zpracování, XLSTAT 2019

SARIMA (ubytování)

Model SARIMA je sezónní autoregresní integrovaný proces klouzavých průměrů řádů p , d a q . V této kapitole jsou uvedeny jen modely, které byly v práci použity spolu s jejich základními veličinami. Model SARIMA má podobu vzorce (51):

$$\Phi_p(B^4)\varphi_p(B)(1-B)^d(1-B^4)^D Y_t = \sum_{j=1}^4 \delta_j^* D_{j,t} + \theta_q(B)\Theta_q(B^4)\varepsilon_t.$$

V následující části budou uvedeny modely SARIMA (1,1,0) a SARIMA (0,1,1), nikoli však model (1,1,1) jelikož po zadání dat do programu, vyšly veškeré veličiny jako konstanty a nešla uskutečnit predikce.

SARIMA (1,1,0)

Výsledky koeficientů pomocí metody SARIMA (1,1,0) jsou uvedeny v následující tabulce.

Tabulka 34 - Koeficienty SARIMA (1,1,0) ubytování

SARIMA(1,1,0)(1,1,0)⁴		
	Parametr	Hodnota
dlouhodobý průměr	μ	0,0018
nesezónní koef.	ϕ_1	-0,0277
sezónní koef.	Φ_1	0,0000
směr. Odchylka residuí	σ	0,1087
pořadí integrace	d	1
délka periody	s	4
pořadí sezónní integrace	D	1

Zdroj: vlastní zpracování, NumXL 2019

V následující tabulce jsou ilustrativně uvedeny hodnoty vypočtené pomocí metody SARIMA (1,1,0) spolu s reálnými daty.

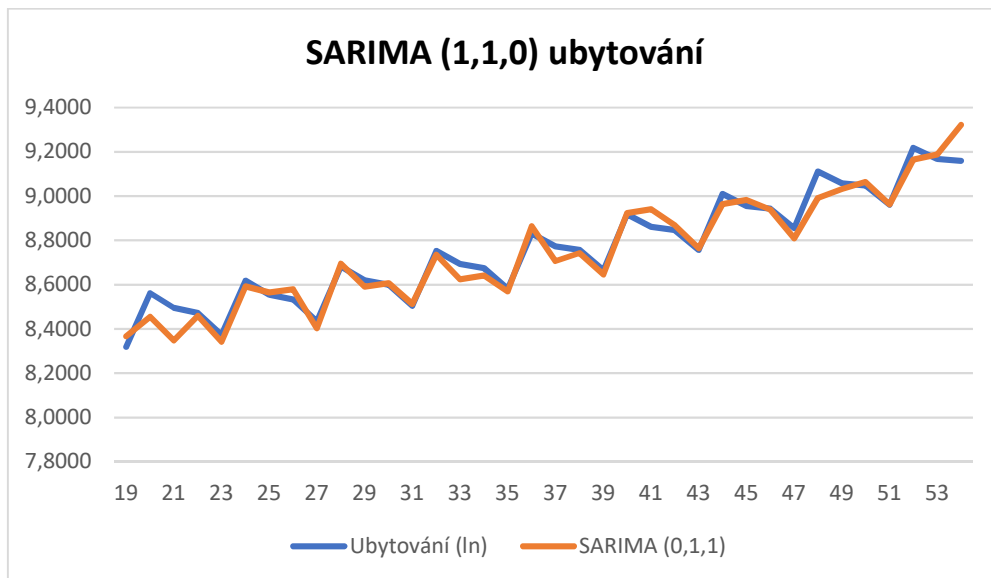
Tabulka 35 - Hodnoty SARIMA (1,1,0) ubytování

t	Ubytování (ln)	SARIMA (1,1,0)	t	Ubytování (ln)	SARIMA (1,1,0)
19	8,3187	8,3226	37	8,7732	8,7708
20	8,5615	8,4815	38	8,7567	8,7069
21	8,4960	8,4463	39	8,6651	8,6214
22	8,4722	8,4592	40	8,9172	8,8223
23	8,3733	8,3759	41	8,8608	8,8856
24	8,6180	8,6230	42	8,8461	8,8313
25	8,5544	8,5519	43	8,7564	8,6753
26	8,5323	8,5226	44	9,0103	9,0518
27	8,4353	8,4527	45	8,9558	8,9424
28	8,6818	8,6752	46	8,9429	8,9390
29	8,6200	8,5713	47	8,8550	8,7369
30	8,5998	8,6824	48	9,1107	9,1263
31	8,5046	8,4151	49	9,0580	9,0739
32	8,7530	8,7094	50	9,0470	9,0370
33	8,6930	8,7896	51	8,9609	8,8572
34	8,6746	8,6638	52	9,2184	9,2247
35	8,5812	8,5046	53	9,1676	9,1501
36	8,8314	8,6258	54	9,1584	9,1695

Zdroj: vlastní zpracování, NumXL 2019

Výše uvedené hodnoty jsou zaneseny do následujícího grafu.

Graf 21 – SARIMA (1,1,0) hodnoty ubytování



Zdroj: vlastní zpracování, NumXL 2019

SARIMA (0,1,1)

Výsledky koeficientů pomocí metody SARIMA (0,1,1) jsou uvedeny v následující tabulce.

Tabulka 36 - Koeficienty SARIMA (0,1,1) ubytování

SARIMA(0,1,1)(0,1,1) ₄		
	Parametr	Hodnota
dlouhodobý průměr	μ	0,0018
nesezónní koef.	θ_1	-0,0109
sezónní koef.	Θ_1	-0,7142
směr. Odchylka residuí	σ	0,0904
pořadí integrace	d	1
délka periody	s	4
pořadí sezónní integrace	D	1

Zdroj: vlastní zpracování, NumXL 2019

V následující tabulce jsou ilustrativně uvedeny hodnoty vypočtené pomocí metody SARIMA (0,1,1) spolu s reálnými daty.

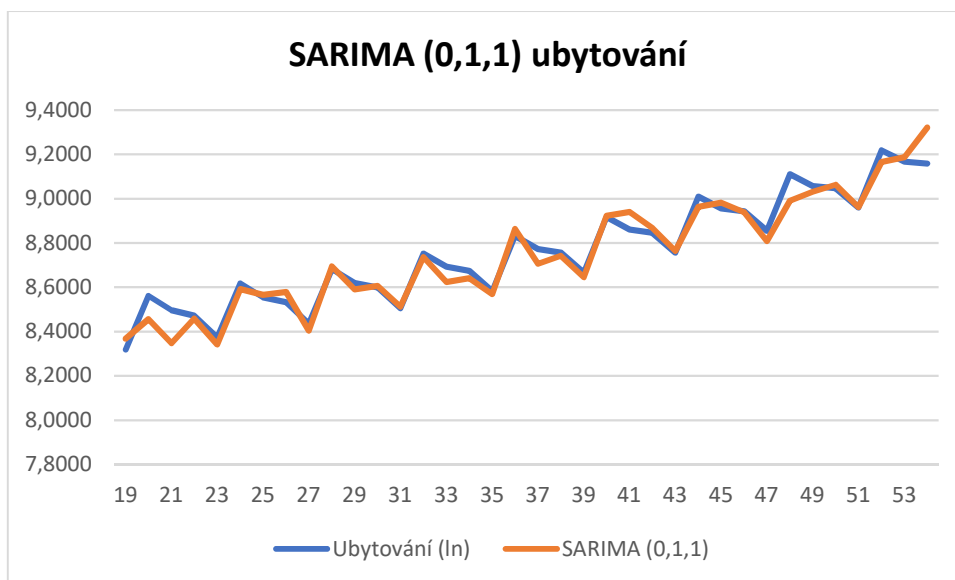
Tabulka 37 - Hodnoty SARIMA (0,1,1) ubytování

t	Ubytování (ln)	SARIMA (0,1,1)	t	Ubytování (ln)	SARIMA (0,1,1)
19	8,3187	8,3675	37	8,7732	8,7064
20	8,5615	8,4555	38	8,7567	8,7423
21	8,4960	8,3475	39	8,6651	8,6459
22	8,4722	8,4595	40	8,9172	8,9234
23	8,3733	8,3418	41	8,8608	8,9408
24	8,6180	8,5917	42	8,8461	8,8686
25	8,5544	8,5655	43	8,7564	8,7640
26	8,5323	8,5793	44	9,0103	8,9633
27	8,4353	8,4034	45	8,9558	8,9826
28	8,6818	8,6951	46	8,9429	8,9392
29	8,6200	8,5908	47	8,8550	8,8094
30	8,5998	8,6064	48	9,1107	8,9905
31	8,5046	8,5123	49	9,0580	9,0317
32	8,7530	8,7359	50	9,0470	9,0631
33	8,6930	8,6234	51	8,9609	8,9621
34	8,6746	8,6408	52	9,2184	9,1651
35	8,5812	8,5686	53	9,1676	9,1881
36	8,8314	8,8640	54	9,1584	9,3213

Zdroj: vlastní zpracování, NumXL 2019

Výše uvedené hodnoty jsou zaneseny do následujícího grafu.

Graf 22 - SARIMA (0,1,1) hodnoty ubytování



Zdroj: vlastní zpracování, NumXL 2019

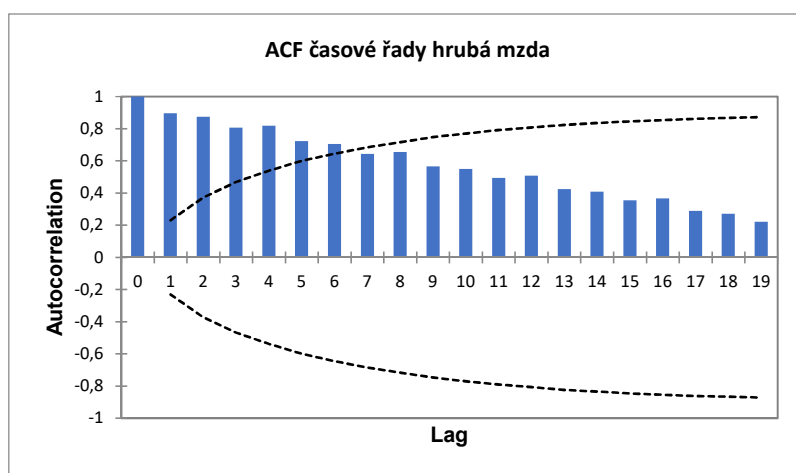
Při porovnání hodnot tržeb za ubytování modelu SARIMA (1,1,0) a modelu SARIMA (0,1,1), nelze na první pohled říct, který model je lepší a více odpovídající, proto budou tyto metody dále porovnány v kapitole č. 6 „Extrapolace časových řad“.

5.3.2 ARIMA (mzdy)

ARIMA

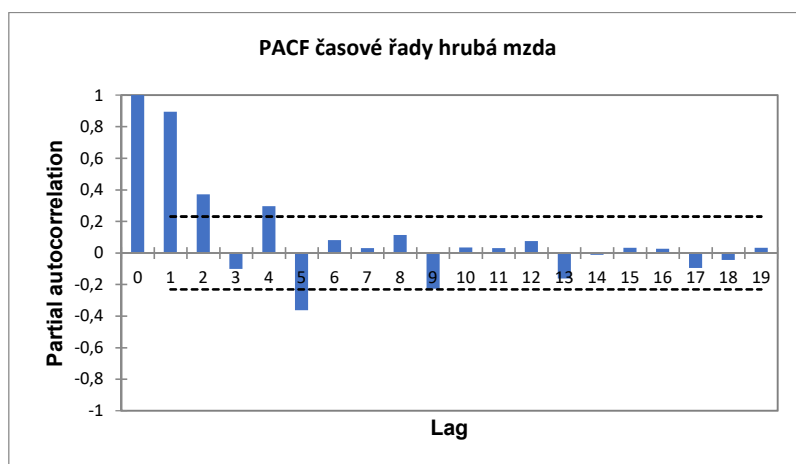
Vzorec tohoto modelu je totožný jako u metody ARIMA (ubytování), tedy vzorec (58). Prvním krokem je také ověření stacionarity časové řady hrubé mzdy pomocí grafu ACF (autokorelační funkce) a PACF (parciální autokorelační funkce).

Graf 23 - ACF časové řady hrubé mzdy



Zdroj: vlastní zpracování, XLSTAT 2019

Graf 24 - PACF časové řady hrubá mzda

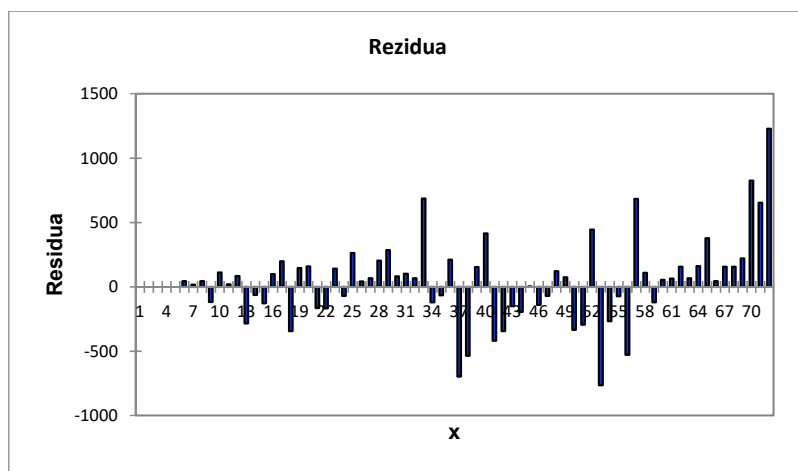


Zdroj: vlastní zpracování, XLSTAT 2019

Z výše uvedených grafů ACF a PACF lze pozorovat nestacionaritu časové řady, stejně jako u předchozí časové řady. Výsledky z grafů opět hovoří o použití „jedničkových“ parametrů pro modely ARIMA a SARIMA (dle pravidel uvedených v kapitole 3.2), jelikož hodnoty v ACF exponenciálně klesají a hodnoty PACF mají oscilující pokles začínající na pozici 1.

Při zpracování dat, vzniknou rezidua, která jsou prezentována v následujícím grafu.

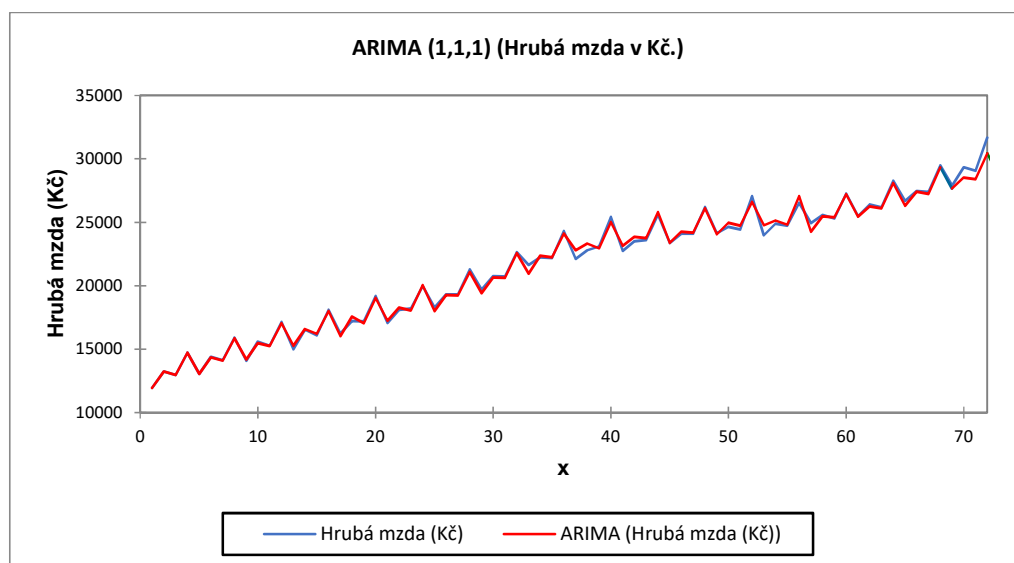
Graf 25 - Rezidua mzdy



Zdroj: vlastní zpracování, XLSTAT 2019

V následujícím grafu je zachyceno porovnání empirických dat spolu s křivkou ARIMA.

Graf 26 – ARIMA (1,1,1) mzdy



Zdroj: vlastní zpracování, XLSTAT 2019

Na grafu lze vidět, že metoda ARIMA (1,1,1) opět vystihuje a vykresluje empirická data až na malé odchylky.

Následující část obsahuje základní data a zhodnocení vybraných typů ARIM. Zhodnocení probíhá pomocí kritéria M.S.E. a M.A.P.E. U nevhodnějšího modelu bude uvedena odpovídající rovnice spolu s vypočtenými hodnotami pomocí dané metody.

Model (1,1,1)

V následující tabulce jsou uvedeny základní informace o modelu ARIMA (1,1,1).

Tabulka 38 - Základní hodnoty ARIMA (1,1,1) mzdy

Parametr	Hodnota	Standardní chyba	Spodní hranice (95%)	Horní hranice (95%)
AR(1)	-0,100	0,197	-0,487	0,286
SAR(1)	0,499	0,131	0,243	0,755
MA(1)	-0,432	0,160	-0,745	-0,119
SMA(1)	-1,000	0,669	-2,312	0,312

Zdroj: vlastní zpracování, XLSTAT 2019

Kritéria hodnocení vhodnosti modelu jsou uvedena v tabulce č. 39.

Tabulka 39 - Hodnotící kritéria ARIMA (1,1,1) mzdy

Ukazatel	Hodnota
M.S.E.	73 605,8
M.A.P.E.	0,8520

Zdroj: vlastní zpracování, XLSTAT 2019

Model (1,1,0)

V následující tabulce jsou uvedeny základní informace o modelu ARIMA (1,1,0).

Tabulka 40 - Základní hodnoty ARIMA (1,1,0) mzdy

Parametr	Hodnota	Standardní chyba	Spodní hranice (95%)	Horní hranice (95%)
AR(1)	-0,407	0,116	-0,635	-0,180
SAR(1)	-0,290	0,119	-0,523	-0,058

Zdroj: vlastní zpracování, XLSTAT 2019

Kritéria hodnocení vhodnosti modelu jsou uvedena v tabulce č. 41.

Tabulka 41 – Hodnoticí kritéria ARIMA (1,1,0) mzdy

Ukazatel	Hodnota
M.S.E.	97 912,4
M.A.P.E.	0,9778

Zdroj: vlastní zpracování, XLSTAT 2019

Model (0,1,1)

V následující tabulce jsou uvedeny základní informace o modelu ARIMA (0,1,1).

Tabulka 42 – Základní hodnoty ARIMA (0,1,1) mzdy

Parametr	Hodnota	Standardní chyba	Spodní hranice (95%)	Horní hranice (95%)
MA(1)	-0,426	0,102	-0,627	-0,226
SMA(1)	-0,472	0,176	-0,816	-0,127

Zdroj: vlastní zpracování, XLSTAT 2019

Kritéria hodnocení vhodnosti modelu jsou uvedena v tabulce č. 43.

Tabulka 43 - Hodnoticí kritéria ARIMA (0,1,1) mzdy

Ukazatel	Hodnota
M.S.E.	87 407, 69
M.A.P.E.	0,9204

Zdroj: vlastní zpracování, XLSTAT 2019

Na základě výsledných hodnoticích kritérií bylo zjištěno, že nejvhodnější model pro popis empirických dat je model ARIMA (1,1,1), protože vykazuje nejmenší hodnotu M.S.E. a také má nejmenší hodnotu M.A.P.E. Výsledná rovnice tohoto modelu může být popsána vztahem:

$$y_t = 0,499 - 0,1(y_{t-1} - y_{t-2}) + \varepsilon_t - 0,432\varepsilon_{t-1} - 1\varepsilon_{t-2}. \quad (70)$$

Část vypočtených hodnot pomocí metody ARIMA (1,1,1) je ilustrativně uvedena v následující tabulce.

Tabulka 44 - Hodnoty ARIMA (1,1,1) mzdy

t	Hrubá mzda	ARIMA (1,1,1)	t	Hrubá mzda	ARIMA (1,1,1)	t	Hrubá mzda	ARIMA (1,1,1)
1	11941	11941,00	19	17190	17044,59	37	22108	22805,65
2	13227	13227,00	20	19183	19023,73	38	22796	23332,90
3	12963	12963,00	21	17067	17231,48	39	23091	22936,43
4	14717	14717,00	22	18112	18280,55	40	25418	25002,22
5	13052	13052,00	23	18203	18062,39	41	22738	23155,51
6	14391	14346,29	24	19963	20034,27	42	23504	23847,58
7	14117	14099,15	25	18270	18007,35	43	23600	23749,79
8	15908	15863,43	26	19300	19258,87	44	25591	25786,38
9	14083	14200,95	27	19305	19239,63	45	23372	23371,23
10	15599	15486,66	28	21269	21066,37	46	24116	24255,48
11	15268	15249,31	29	19687	19401,73	47	24107	24177,06
12	17133	17048,03	30	20740	20659,38	48	26211	26089,49
13	14986	15269,39	31	20721	20619,89	49	24131	24056,91
14	16529	16590,87	32	22641	22574,93	50	24627	24959,90
15	16088	16215,74	33	21632	20946,13	51	24439	24734,16
16	18096	17996,10	34	22246	22366,35	52	27055	26610,41
17	16231	16032,07	35	22181	22245,45	53	23985	24748,70
18	17223	17566,58	36	24309	24097,83	54	24877	25143,78

Zdroj: vlastní zpracování, XLSTAT 2019

5.3.3 SARIMA (mzdy)

Analýza dat pomocí metody SARIMA byla opět provedena v programu NumXL. Základní model SARIMY má opět podobu vzorce (51). V následující části je provedena analýza dat pomocí metod SARIMA (1,1,0) a SARIMA (0,1,1), tyto metody byly zvoleny na základě výsledků ACF a PACF. Data v těchto modelech pak obsahují, kromě sezónních kolísání, také autoregresní část, integrační část nebo klouzavé průměry.

SARIMA (1,1,0)

Výsledky koeficientů pomocí metody SARIMA (1,1,0) jsou uvedeny v následující tabulce.

Tabulka 45 - Koeficienty SARIMA (1,1,0) mzdy

SARIMA(1,1,0)(1,1,0) ₄		
	Parametr	Hodnota
dlouhodobý průměr	μ	-0,0003
nesezónní koef.	ϕ_1	-0,4687
sezónní koef.	Φ_1	0,0000
směr. Odchylka residuí	σ	0,0144
pořadí integrace	d	1
délka periody	s	4
pořadí sezónní integrace	D	1

Zdroj: vlastní zpracování, NumXL 2019

V následující tabulce jsou ilustrativně uvedeny hodnoty vypočtené pomocí metody SARIMA (1,1,0) spolu s reálnými daty.

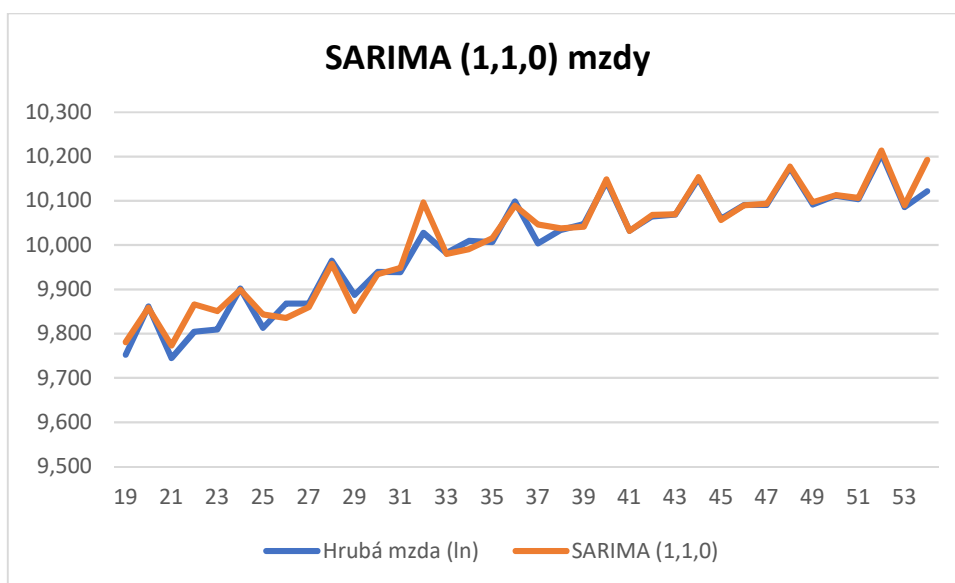
Tabulka 46 - Hodnoty SARIMA (1,1,0) mzdy

t	Hrubá mzda (ln)	SARIMA (1,1,0)	t	Hrubá mzda (ln)	SARIMA (1,1,0)
19	9,752	9,781	37	10,004	10,046
20	9,862	9,859	38	10,034	10,038
21	9,745	9,774	39	10,047	10,041
22	9,804	9,866	40	10,143	10,149
23	9,809	9,851	41	10,032	10,032
24	9,902	9,899	42	10,065	10,068
25	9,813	9,843	43	10,069	10,069
26	9,868	9,835	44	10,150	10,153
27	9,868	9,860	45	10,059	10,056
28	9,965	9,958	46	10,091	10,090
29	9,888	9,851	47	10,090	10,093
30	9,940	9,934	48	10,174	10,177
31	9,939	9,948	49	10,091	10,097
32	10,028	10,097	50	10,112	10,113
33	9,982	9,980	51	10,104	10,106
34	10,010	9,991	52	10,206	10,214
35	10,007	10,015	53	10,085	10,090
36	10,099	10,089	54	10,122	10,192

Zdroj: vlastní zpracování, NumXL 2019

Výše uvedené hodnoty jsou zaneseny do následujícího grafu.

Graf 27 - SARIMA (1,1,0) hodnoty mzdy



Zdroj: vlastní zpracování, NumXL 2019

SARIMA (0,1,1)

Výsledky koeficientů pomocí metody SARIMA (0,1,1) jsou uvedeny v následující tabulce.

Tabulka 47 - Koeficienty SARIMA (0,1,1) mzdy

SARIMA (0,1,1)(0,1,1)4		
	Parametr	Hodnota
dlouhodobý průměr	μ	-0,0003
nesezónní koef.	ϕ_1	-0,4442
sezónní koef.	Φ_1	-0,2345
směr. Odchylka residuí	σ	0,0137
pořadí integrace	d	1
délka periody	s	4
pořadí sezónní integrace	D	1

Zdroj: vlastní zpracování, NumXL 2019

V následující tabulce jsou ilustrativně uvedeny hodnoty vypočtené pomocí metody SARIMA (0,1,1) spolu s reálnými daty.

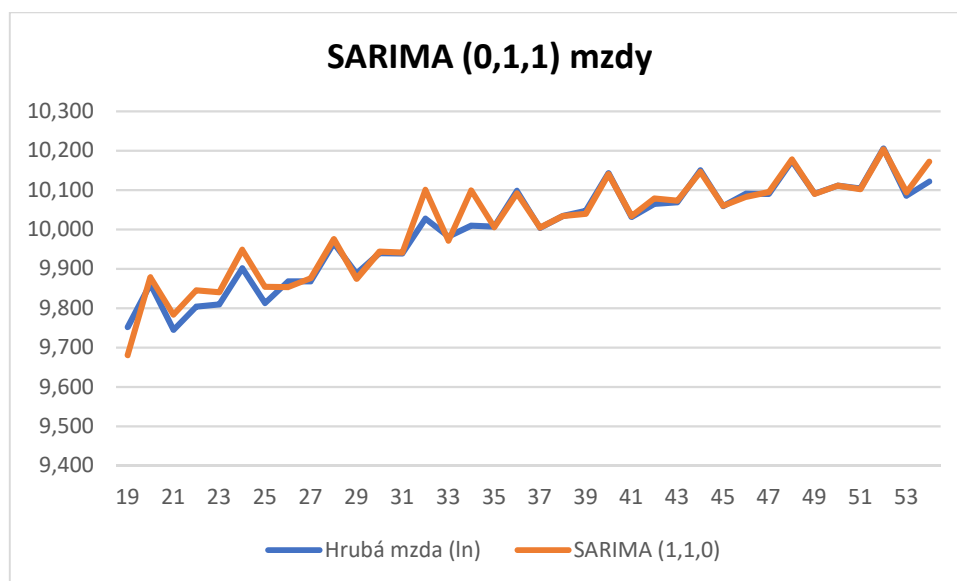
Tabulka 48 - Hodnoty SARIMA (0,1,1) mzdy

t	Hrubá mzda (ln)	SARIMA (1,1,0)	t	Hrubá mzda (ln)	SARIMA (1,1,0)
19	9,752	9,681	37	10,004	10,005
20	9,862	9,879	38	10,034	10,034
21	9,745	9,784	39	10,047	10,039
22	9,804	9,846	40	10,143	10,141
23	9,809	9,841	41	10,032	10,034
24	9,902	9,949	42	10,065	10,079
25	9,813	9,854	43	10,069	10,073
26	9,868	9,854	44	10,150	10,146
27	9,868	9,876	45	10,059	10,060
28	9,965	9,976	46	10,091	10,083
29	9,888	9,875	47	10,090	10,095
30	9,940	9,944	48	10,174	10,178
31	9,939	9,941	49	10,091	10,090
32	10,028	10,101	50	10,112	10,111
33	9,982	9,972	51	10,104	10,102
34	10,010	10,099	52	10,206	10,204
35	10,007	10,005	53	10,085	10,094
36	10,099	10,092	54	10,122	10,172

Zdroj: vlastní zpracování, NumXL 2019

Výše uvedené hodnoty jsou zaneseny do následujícího grafu.

Graf 28 - SARIMA (0,1,1) hodnoty mzdy



Zdroj: vlastní zpracování, NumXL 2019

Při porovnání hodnot hrubé mzdy modelu SARIMA (1,1,0) a modelu SARIMA (0,1,1), nelze na první pohled říct, který model je lepší a více odpovídající, proto budou tyto metody porovnány při predikci budoucích hodnot.

O predikci neboli extrapolaci časových řad je následující kapitola.

6 Extrapolace časových řad

Předchozí kapitoly byly o tom, jak dobře model popisuje data, avšak je zapotřebí se zaměřit také na to, jak se data budou vyvíjet v budoucnu. Proto se tato kapitola bude zabývat predikcí obou časových řad za pomoci již zmíněných metod. Predikce je zaměřená jen na následující 2 roky, a to z důvodu, že případná změna metodiky výpočtu dat může ovlivnit jejich hodnotu nebo je mohou ovlivnit některé ekonomické výkyvy.

Analýza časových řad probíhala mezi prvním čtvrtletím roku 2000 až čtvrtým čtvrtletím 2017. Mezitím, než byla tato práce napsána, byl na stránkách Českého statistického úřadu kompletně doplněn i rok 2018, proto tedy lze předpovězené hodnoty porovnávat s rokem 2018. Právě suma odchylek predikovaných hodnot od empirických hodnot může být dalším indikátorem, jak jsou modely spolehlivé pro extrapolaci.

V práci byly použity jak bodové, tak intervalové odhady. Na základě extrapolací kritéria M.S.E. bylo zjištěno, který model by měl být nejvhodnější pro extrapolaci dat. Již v předchozích kapitolách byly připraveny a vypočítány některé potřebné údaje u obou časových řadám. Extrapolace budou opět provedeny pro každou časovou řadu zvlášť. Nejprve je uvedena časová řada „tržby za ubytování“.

6.1 Extrapolace časové řady „tržby za ubytování“

Pro použití nejvhodnějšího modelu pro predikci dat, je zapotřebí vybrat model s nejnižším extrapolacním kritériem M.S.E., proto jsou v následující tabulce uvedeny hodnoty tohoto kritéria u všech použitých metod.

Tabulka 49 - Hodnoty M.S.E. (ubytování)

Klasické + aditivní přístupy	
Metoda	Hodnota M.S.E.
Lineární trend	1 920 478,43
Kvadratický trend	1 502 145,61
Exponenciální trend	1 607 918,73
Klouzavé průměry	365 004,43
Dvojité exp. vyrovnávání	452 474,17
Trojité exp. vyrovnávání	277 562,60
Box-Jenkins	
ARIMA (1,1,1)	358 119,19
ARIMA (1,1,0)	358 502,91
ARIMA (0,1,1)	358 513,42

Zdroj: vlastní zpracování, Excel 2019

Na základě výsledků extrapolčního kritéria, jehož výpočet je upřesněn v tabulce č. 2, bylo zjištěno, že u klasických přístupů se jako nejvhodnější model jeví trend kvadratický. U adaptivních přístupů se nejlépe jeví trojité exponenciální vyrovnávání. Proto budou pro predikci použity tyto modely.

U modelu Box-Jenkins jsou hodnoty velmi podobné, proto bude extrapolace hodnot uvedena pro všechny typy ARIMY a také pro oba druhy SARIMY. Jelikož hodnoty M.S.E. u modelů ARIMA vyšly podobně, budou uvedeny navíc porovnání ukazatele hodnoty M.A.P.E. pro každý model.

Tabulka 50 – Hodnoty M.A.P.E. (ubytování)

Box-Jenkins	
Metoda	Hodnota M.A.P.E.
ARIMA (1,1,1)	6,5234
ARIMA (1,1,0)	6,5218
ARIMA (0,1,1)	6,5629

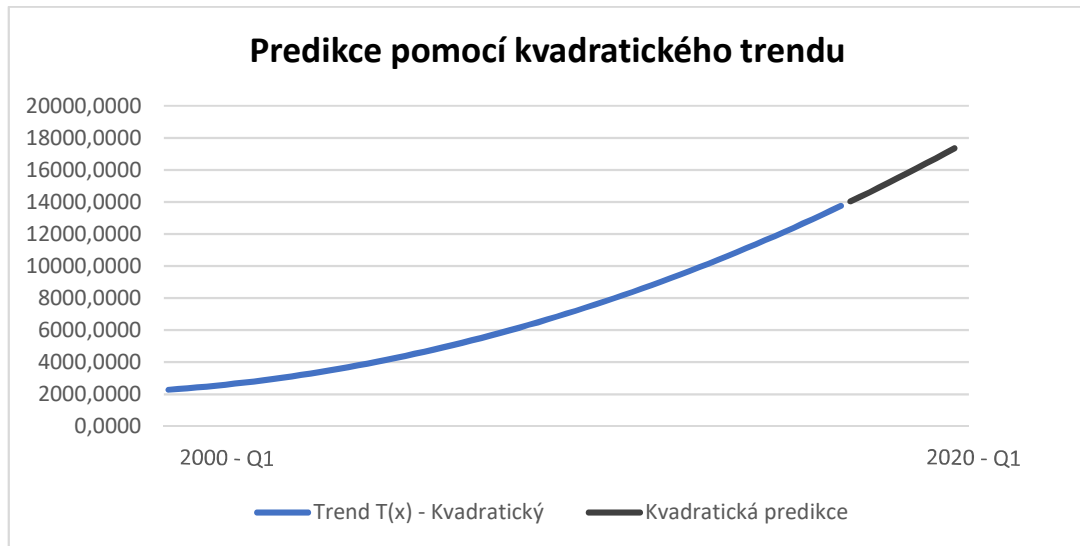
Zdroj: vlastní zpracování, XLSTAT 2019

Hodnoty ukazatele M.A.P.E. vyšly nejmenší u ARIMY (1,1,1) a ARIMY (1,1,0), proto bude predikce provedena pomocí těchto dvou metod, jelikož třetí metoda vždy bude mít větší absolutní procentní chybu odhadu.

6.1.1 Klasické přístupy – Kvadratický trend

Z klasických přístupů byl vybrán kvadratický trend, jelikož jeho extrapolační hodnota M.S.E. nabývala nejmenší hodnoty. Byla tedy provedena predikce na základě kvadratického trendu.

Graf 29 - Kvadratická predikce (ubytování)



Zdroj: vlastní zpracování, Excel 2019

Výše uvedená predikce byla vypočtena na základě kvadratického trendu pomocí vzorce (54):

$$T_p = b_0 + b_1t + b_2t^2,$$

Za t byla dosazena čísla, která označují nadcházející období (73 – 80), jelikož empirických dat bylo 72. Následující tabulka zachycuje porovnání skutečných hodnot a predikovaných hodnot v roce 2018.

Tabulka 51 - Srovnání hodnot kvadr. trend (ubytování)

Predikované hodnoty	Skutečné hodnoty	
	Kvadratický trend	
14 031,96 Kč	2018 - Q1	10 355,72 Kč
14 317,42 Kč	2018 - Q2	14 430,30 Kč
14 606,23 Kč	2018 - Q3	15 089,80 Kč
14 898,39 Kč	2018 - Q4	14 976,00 Kč

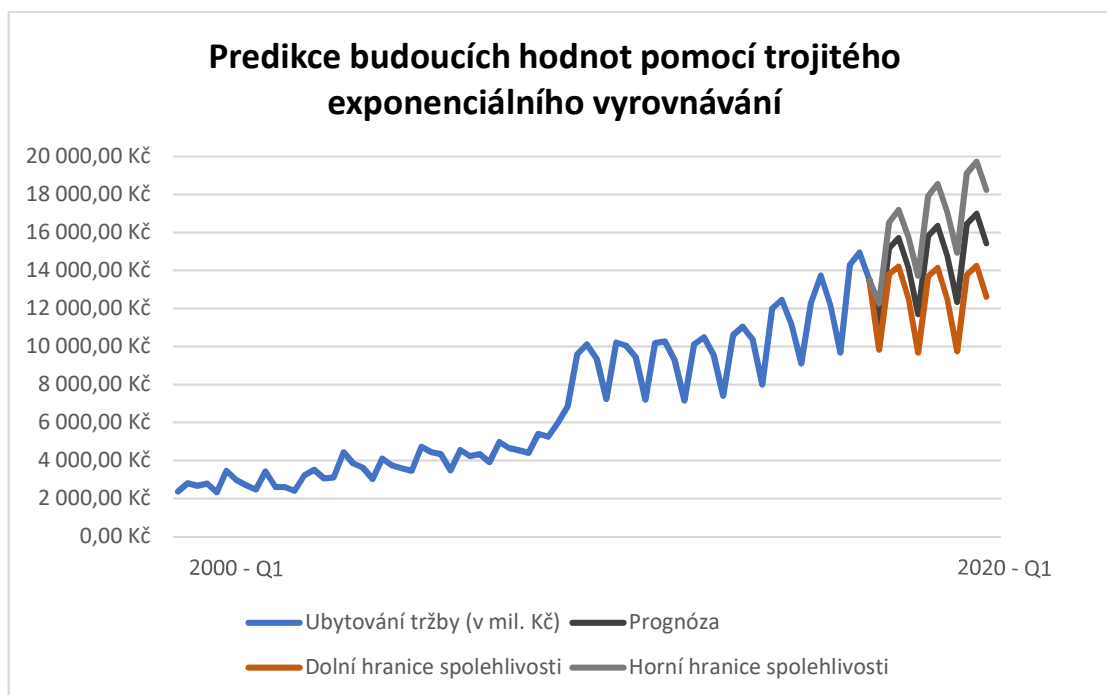
Zdroj: vlastní zpracování, Excel 2019

Jelikož je predikce dle kvadratického trendu poměrně nepřesná, bude vyzkoušena predikce dat pomocí adaptivního přístupu.

6.1.2 Adaptivní přístupy – Trojité exponenciální vyrovnávání

U adaptivních přístupů vyšla nejmenší hodnota extrapolačního kritéria u trojitého exponenciálního vyrovnávání. Predikce byla provedena v excelu pomocí funkce FORECAST.ETS, která spočítala bodové předpovědi dle vzorce (57). Za použití další funkce FORECAST.ETS.CONFINT byla vypočtena intervalová předpověď neboli dolní a horní mez. Následující graf zachycuje predikci hodnot pomocí trojitého exponenciálního vyrovnávání.

Graf 30 - Predikce trojité exp. vyrovnávání (ubytování)



Zdroj: vlastní zpracování, Excel 2019

Na predikovaných hodnotách lze vidět, že by se měly dále vyvíjet ve stejném trendu jako doposud, avšak jen mezi horní a dolní mezí. Pro tento interval byla použita hladina významnosti $\alpha = 0,05$. Na grafu lze vidět 95% interval spolehlivosti pro predikované hodnoty.

Následující tabulka zachycuje porovnání predikovaných hodnot v roce 2018 a jejich skutečné hodnoty.

Tabulka 52 - Srovnání hodnot trojitě exp. vyrovnávání (ubytování)

Predikované hodnoty	Skutečné hodnoty	
	Trojitě exp. vyr.	
11 055,09 Kč	2018 - Q1	10 355,72 Kč
15 161,97 Kč	2018 - Q2	14 430,30 Kč
15 702,89 Kč	2018 - Q3	15 089,80 Kč
14 130,91 Kč	2018 - Q4	14 976,00 Kč

Zdroj: vlastní zpracování, Excel 2019

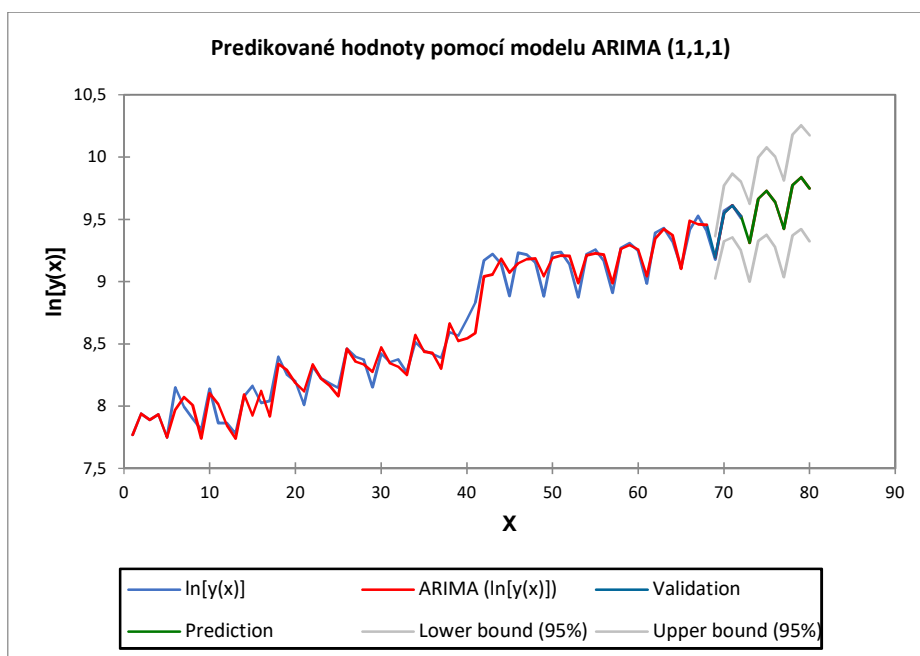
Z tabulky č. 52 lze usoudit, že predikované hodnoty jsou blíže ke skutečnosti, než je tomu u kvadratického trendu, navíc tato data v sobě zachycují také sezónní kolísání.

V následující části jsou hodnoty predikovány pomocí Box-Jenkinsovy metodologie.

6.1.3 ARIMA (1,1,1)

Časová řada „tržby za ubytování“ byla predikována opět o 8 kroků dopředu. Predikce byla vytvořena také v programu XLSTAT. Bylo zapotřebí nastavit počet kroků a označit hodnoty, jež byly zlogaritmovány.

Graf 31 - Predikce ARIMA (1,1,1) ubytování



Zdroj: vlastní zpracování, XLSTAT 2019

Na grafu lze vidět původní data spolu s vykreslenými daty pomocí modelu ARIMA (1,1,1), na které navazuje bodová předpověď. Tato předpověď je mezi horní a dolní mezí,

které byly nastaveny na $\alpha = 0,05$, jedná se tedy o 95% interval spolehlivosti. Porovnání predikovaných dat s reálnými daty v roce 2018 je uvedeno v následující tabulce.

Tabulka 53 - Srovnání hodnot ARIMA (1,1,1) ubytování

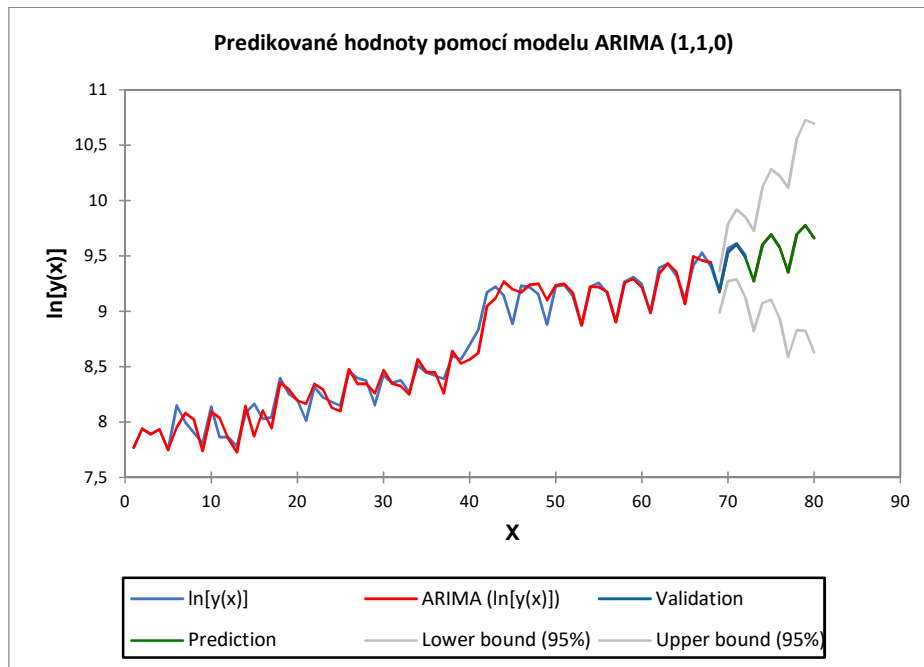
Predikované hodnoty	Skutečné hodnoty	
	ARIMA (1,1,1)	
10 889,71 Kč	2018 - Q1	10 355,72 Kč
14 268,19 Kč	2018 - Q2	14 430,30 Kč
15 398,07 Kč	2018 - Q3	15 089,80 Kč
13 983,94 Kč	2018 - Q4	14 976,00 Kč

Zdroj: vlastní zpracování, XLSTAT 2019

6.1.4 ARIMA (1,1,0)

Analýza časové řady pomocí metody ARIMA (1,1,0) byla provedena obdobně v programu XLSTAT. Graf vývoje trendu spolu s předpovědí včetně odhadnutého intervalu lze vidět na následujícím grafu.

Graf 32 - Predikce ARIMA (1,1,0) ubytování



Zdroj: vlastní zpracování, XLSTAT 2019

Na uvedeném grafu lze vidět původní data spolu s vykreslenými daty pomocí modelu ARIMA (1,1,0), na které navazuje bodová předpověď. Horní a dolní mez je opět

nastavena na 95% interval spolehlivosti. Porovnání predikovaných dat s reálnými daty v roce 2018 je uvedeno v následující tabulce.

Tabulka 54 - Srovnání hodnot ARIMA (1,1,0) ubytování

Predikované hodnoty	Skutečné hodnoty	
ARIMA (1,1,0)		
10 930,54 Kč	2018 - Q1	10 355,72 Kč
14 289,29 Kč	2018 - Q2	14 430,30 Kč
15 531,43 Kč	2018 - Q3	15 089,80 Kč
14 032,61 Kč	2018 - Q4	14 976,00 Kč

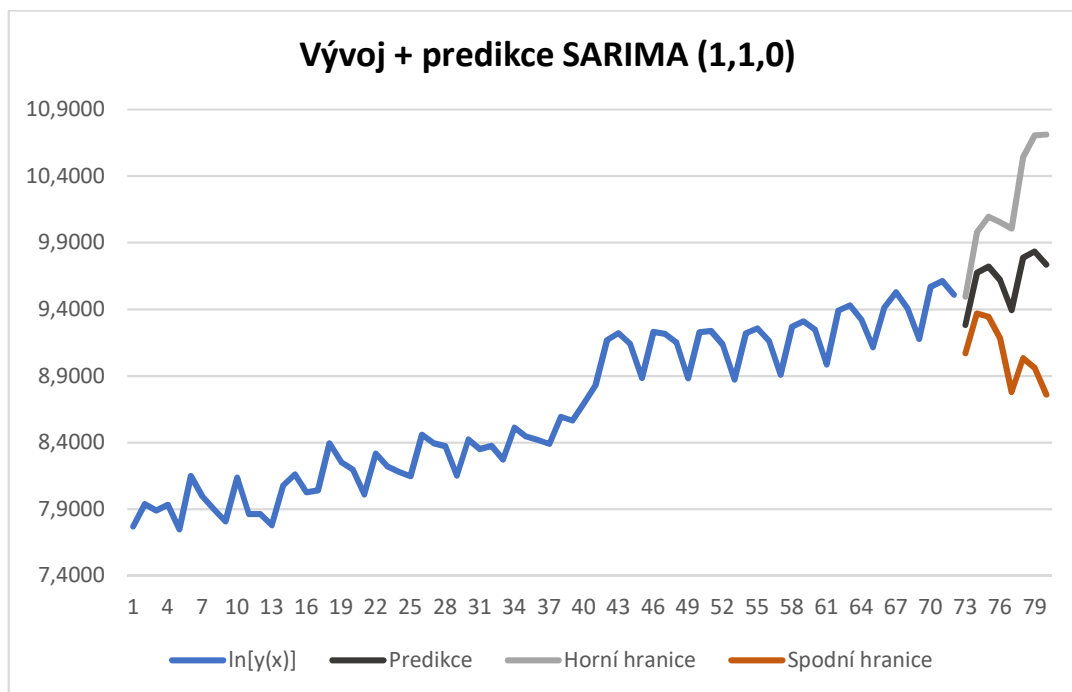
Zdroj: vlastní zpracování, XLSTAT 2019

Při porovnání předpovězených hodnot u obou výše uvedených metod, je uvažována také suma odchylek od skutečných hodnot, k získání informací o tom, která metoda je přesnější v porovnání s rokem 2018. Nejprve byly vypočteny odchylky od skutečných hodnot u obou metod a poté z nich pomocí funkce ABS byly udělané absolutní hodnoty, které byly sečteny. Tento výsledek ukázal, že i přesto, že model ARIMA (1,1,0) má menší ukazatel M.A.P.E. než metoda ARIMA (1,1,1), mají predikované hodnoty větší odchylku od skutečných hodnot. Suma absolutních odchylek v roce 2018 u modelu ARIMA (1,1,1) vyšla **1996,44** a suma absolutních odchylek u modelu ARIMA (1,1,0) vyšla **2100,85**. Proto lze tvrdit, že model ARIMA (1,1,1), jenž je zastoupený všemi částmi, je pro predikci časové řady „tržby za ubytování“ vhodnější.

6.1.5 SARIMA (1,1,0)

Tento model je predikován opět na 8 období, tedy na 2 roky dopředu. K predikci byl použit program NumXL, ve kterém byl označen typ modelu a do predikce byl nastaven počet kroků. Výsledek lze vidět na následujícím grafu, který zachycuje vývoj časové řady do konce roku 2017 a následně na něj navazují predikované hodnoty spolu s horní a dolní mezí.

Graf 33 - Predikce SARIMA (1,1,0) ubytování



Zdroj: vlastní zpracování, NumXL 2019

Horní a dolní mez je opět nastavena na 95% interval spolehlivosti. Porovnání predikovaných dat s reálnými daty v roce 2018 je uvedeno v následující tabulce.

Tabulka 55 - Srovnání hodnot SARIMA (1,1,0) ubytování

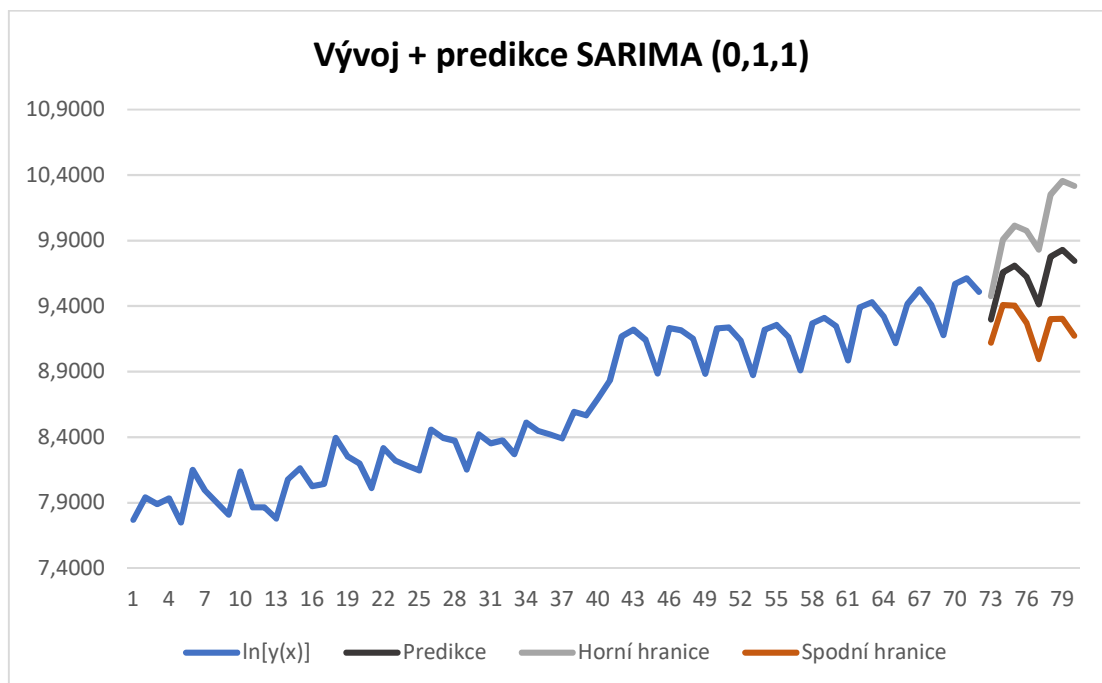
Predikované hodnoty	Skutečné hodnoty	
SARIMA (1,1,0)		
10 994,45 Kč	2018 - Q1	10 355,72 Kč
15 626,97 Kč	2018 - Q2	14 430,30 Kč
16 257,38 Kč	2018 - Q3	15 089,80 Kč
14 799,58 Kč	2018 - Q4	14 976,00 Kč

Zdroj: vlastní zpracování, NumXL 2019

6.1.6 SARIMA (0,1,1)

Analýza časové řady pomocí metody SARIMA (0,1,1) byla obdobně provedena v programu NumXL. Graf vývoje časové řady spolu s předpovědí včetně odhadnutého intervalu lze vidět na následujícím grafu.

Graf 34 - Predikce SARIMA (0,1,1) ubytování



Zdroj: vlastní zpracování, NumXL 2019

Stejně jako v předchozím případě je pro horní a dolní interval spolehlivosti nastaveno $\alpha = 0,05$. Porovnání predikovaných dat s reálnými daty v roce 2018 je uvedeno v následující tabulce.

Tabulka 56 – Srovnání hodnot SARIMA (0,1,1) ubytování

Predikované hodnoty SARIMA (0,1,1)	Skutečné hodnoty	
	11 430,54 Kč	2018 - Q1
14 959,78 Kč	2018 - Q2	14 430,30 Kč
15 569,69 Kč	2018 - Q3	15 089,80 Kč
14 421,26 Kč	2018 - Q4	14 976,00 Kč

Zdroj: vlastní zpracování, NumXL 2019

Při porovnání předpovězených hodnot u obou výše uvedených metod, je vypočítána suma odchylek od skutečných hodnot. Postup výpočtu je stejný jako v předchozím případě u metod ARIMA. Nejprve jsou vypočteny odchylky od skutečných hodnot u obou metod SARIMA a poté jsou z nich pomocí funkce ABS vypočteny absolutní hodnoty, které se sečtou.

Hodnota sumy absolutních odchylek u modelu SARIMA (1,1,0) vyšla **3179,39** a hodnota u modelu SARIMA (0,1,1) vyšla **2638,93**. Z těchto výsledků lze tvrdit, že vhodnějším modelem pro predikci časové řady „tržby z ubytování“ je rozhodně SARIMA (0,1,1).

6.2 Extrapolace časové řady „hrubá mzda“

K výběru nejvhodnějšího modelu pro predikci časové řady „hrubá mzda“ slouží následující tabulka, ve které jsou uvedeny hodnoty extrapolčního kritéria M.S.E. Tyto hodnoty M.S.E. byly vypočítány vzorcem, který je upřesněn v tabulce č. 2.

Tabulka 57 – Hodnoty M.S.E. (mzdy)

Klasické + aditivní přístupy	
Metoda	Hodnota M.S.E.
Lineární trend	1 214 296,87
Kvadratický trend	999 135,21
Exponenciální trend	1 965 691,66
Klouzavé průměry	788 988,82
Dvojitě exp. vyrovnávání	328 850,59
Trojitě exp. vyrovnávání	209 678,86
Box-Jenkins	
ARIMA (1,1,1)	73 605,84
ARIMA (1,1,0)	97 912,44
ARIMA (0,1,1)	87 407,69

Zdroj: vlastní zpracování, XLSTAT 2019

Při pohledu na výsledné hodnoty lze tvrdit, že z klasických přístupů prokládá empirická data nejlépe kvadratický trend. U adaptivních přístupů je nejvhodnější metodou trojitě exponenciální vyrovnávání, jelikož, čím více exponenciálních vyrovnávání je, tím přesnější data lze predikovat. Hodnoty M.S.E. u metodologie Box–Jenkins již nejsou tak podobné jako u předchozí časové řady. Nejlépe zde vyšla metoda ARIMA (1,1,1) a nejhorší metodou je ARIMA (1,1,0). Aby se zjistilo, které metody ARIMA jsou nejpřesnější, jsou v následující tabulce uvedeny hodnoty absolutních procentních chyb odhadu (M.A.P.E.).

Tabulka 58 – Hodnoty M.A.P.E. (mzdy)

Box-Jenkins	
Metoda	Hodnota M.A.P.E.
ARIMA (1,1,1)	0,8520
ARIMA (1,1,0)	0,9778
ARIMA (0,1,1)	0,9204

Zdroj: vlastní zpracování, XLSTAT 2019

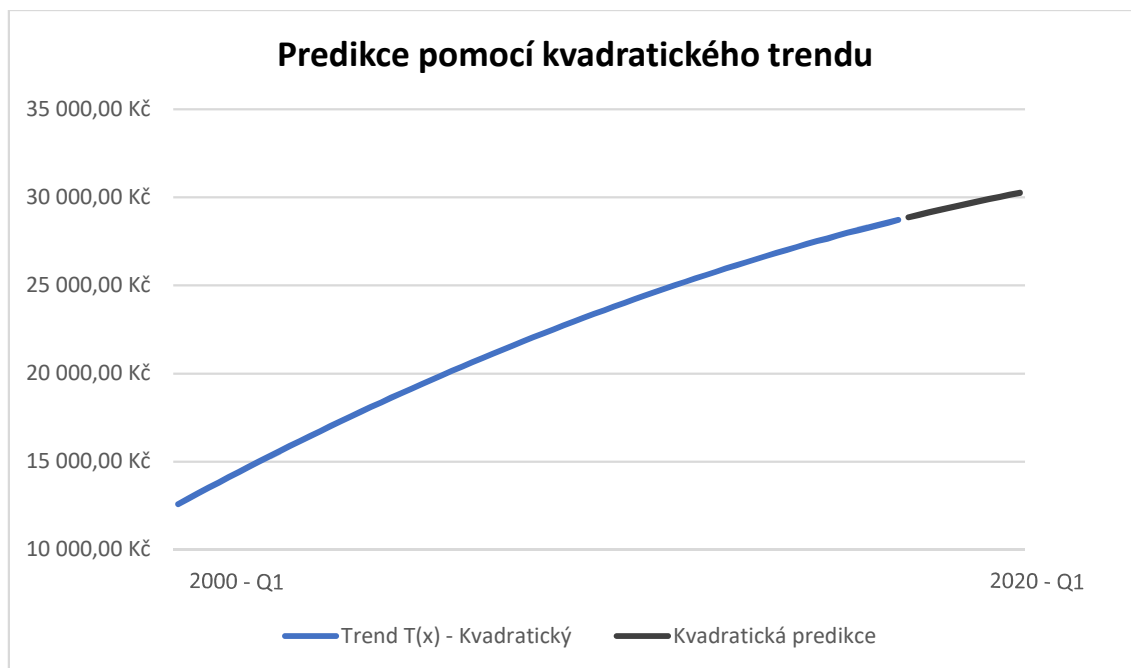
U časové řady „hrubá mzda“ jsou na první pohled menší absolutní procentuální chyby odhadu, to je způsobené tím, že tato časová řada není tak kolísavá jako „tržby za ubytování“. Nejmenší hodnota je opět u metody ARIMA (1,1,1).

Pro predikci budou tedy použity modely ARIMA (1,1,1) a ARIMA (0,1,1), jelikož zbývající třetí metoda by neměla lepší vypovídající hodnotu než tyto dvě.

6.2.1 Klasické přístupy – Kvadratický trend

Jako nejvhodnější u klasických přístupů vyšel trend kvadratický, jehož hodnota M.S.E. je nejmenší. Predikce dle kvadratického trendu je na následujícím grafu.

Graf 35 – Kvadratická predikce (mzdy)



Zdroj: vlastní zpracování, Excel 2019

Tento trend byl vypočten opět za pomoci vzorce (54), kde do vzorce byly doplněny hodnoty následujících období, tj. období 73, období 74, ..., 80. Na základě kvadratické predikce vyšly hodnoty, které jsou prezentovány a porovnány v následující tabulce.

Tabulka 59 – Srovnání hodnot kvadr. Trend (mzdy)

Predikované hodnoty	Skutečné hodnoty	
Kvadratický trend		
28 866,14 Kč	2018 - Q1	30 285,32 Kč
29 004,55 Kč	2018 - Q2	31 866,07 Kč
29 140,56 Kč	2018 - Q3	31 516,92 Kč
29 274,17 Kč	2018 - Q4	32 643,67 Kč

Zdroj: vlastní zpracování, Excel 2019

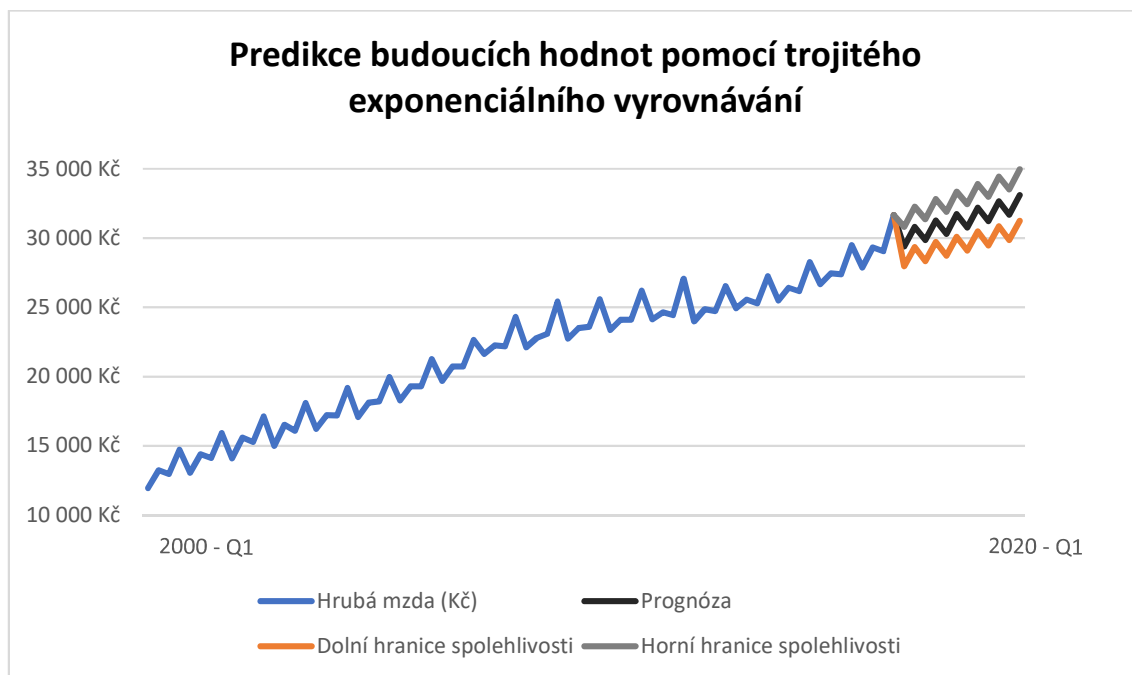
V tomto případě lze vidět, že hodnoty predikované kvadratickým trendem neodpovídají skutečnosti, což je z velké míry způsobené tím, že se na začátku roku 2018 zvedaly mzdy, a to do modelu nelze zahrnout, proto je třeba říci, že žádné metody, které zde jsou uvedeny, nejsou tak přesné ani spolehlivé, aby zachytily nahodilé trendy zvyšování mezd. Avšak přesnější by měla být následující metoda.

6.2.2 Adaptivní přístupy – Trojitě exponenciální vyrovnávání

Co se týče výsledků extrapolčních kritérií, nejmenší hodnotu vykazovalo opět trojitě exponenciální vyrovnávání. Při predikci touto metodou byl použit následující postup: nejprve se pomocí funkce FORECAST.ETS spočítaly bodové předpovědi dle vzorce (57). Za použití další funkce FORECAST.ETS.CONFINT byla vypočtena intervalová předpověď neboli dolní a horní mez.

V následujícím grafu jsou zachyceny predikované hodnoty pomocí trojitěho vyrovnávání.

Graf 36 – Predikce trojitého exp. Vyrovnávání (mzdy)



Zdroj: vlastní zpracování, Excel 2019

Na grafu lze vidět, že hodnoty hrubé mzdy budou dále růst v rozmezí mezi horní a dolní hranicí. Pro tento interval byla použita hladina významnosti $\alpha = 0,05$. Na grafu lze vidět 95% interval spolehlivosti pro predikované hodnoty.

Hodnoty predikované pomocí trojitého exponenciálního vyrovnávání v porovnáním se skutečnými hodnotami v roce 2018 jsou uvedeny v tabulce č. 60.

Tabulka 60 – Srovnání hodnot trojité exp. Vyrovnávání (mzdy)

Predikované hodnoty	Skutečné hodnoty	
Trojité exp. vyr.		
29 391,81 Kč	2018 - Q1	30 285,32 Kč
30 810,37 Kč	2018 - Q2	31 866,07 Kč
29 851,76 Kč	2018 - Q3	31 516,92 Kč
31 270,33 Kč	2018 - Q4	32 643,67 Kč

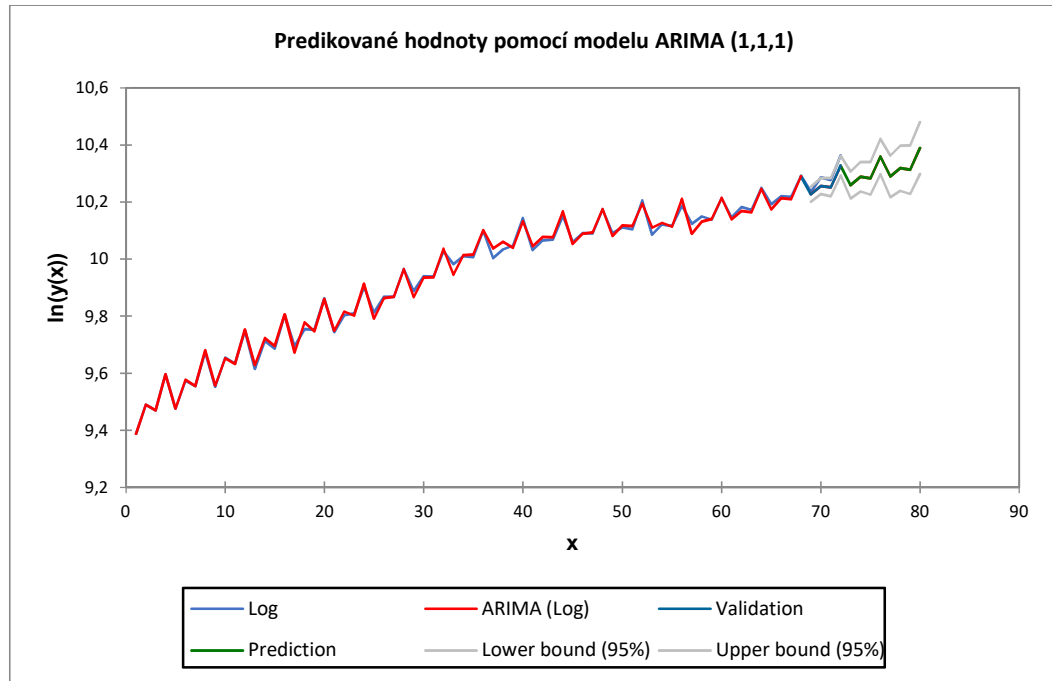
Zdroj: vlastní zpracování, Excel 2019

Hodnoty uvedené v tabulce jsou již přesnější, než hodnoty u kvadratického trendu a je v nich také zanesena sezónnost.

6.2.3 ARIMA (1,1,1)

Následující predikce pomocí metody ARIMA (1,1,1) byla vytvořena v programu XLSTAT, kde byl nastaven počet budoucích kroků a označeny hodnoty. Graf vývoje trendu spolu s předpovědí včetně odhadnutého intervalu lze vidět na grafu č. 37.

Graf 37 – Predikce ARIMA (1,1,1)



Zdroj: vlastní zpracování, XLSTAT 2019

Graf prezentuje původní data spolu s vykreslenými daty pomocí modelu ARIMA (1,1,1). Dále na tato data navazuje předpověď a horní a dolní hranice. Tyto hranice jsou také nastaveny jako 95% interval spolehlivosti. Aby bylo možné porovnat skutečná data s predikovanými, byly do tabulky zaneseny reálné hodnoty časové řady a predikované hodnoty v roce 2018.

Tabulka 61 – Srovnání hodnot ARIMA (1,1,1) mzdy

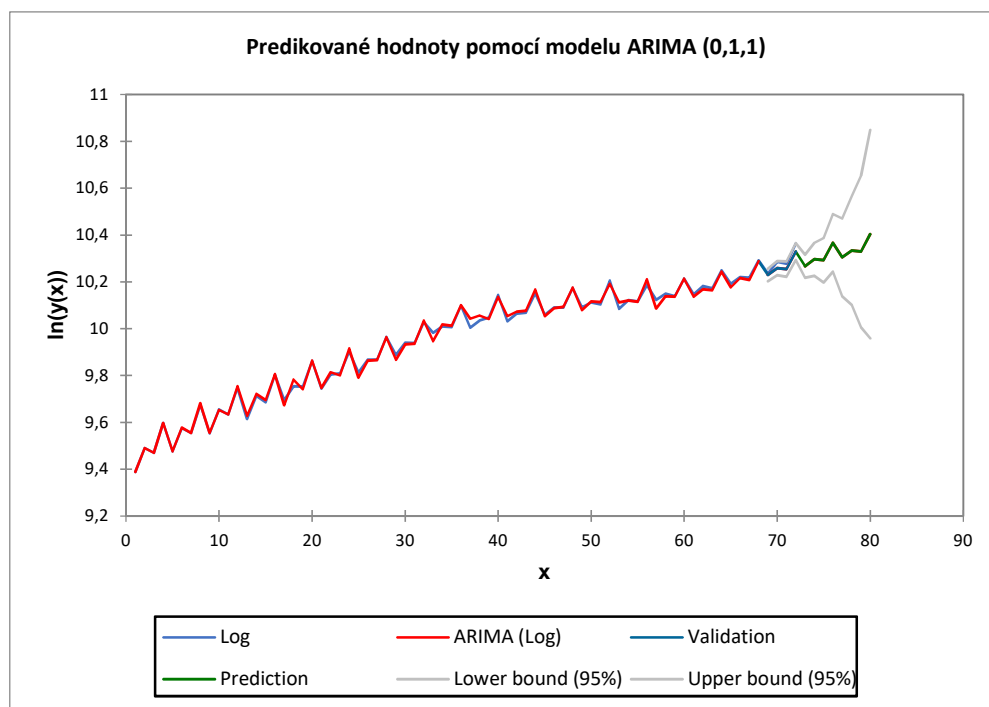
Predikované hodnoty	Skutečné hodnoty	
ARIMA (1,1,1)		
28 545,55 Kč	2018 - Q1	30 285,32 Kč
29 427,95 Kč	2018 - Q2	31 866,07 Kč
29 299,57 Kč	2018 - Q3	31 516,92 Kč
31 294,17 Kč	2018 - Q4	32 643,67 Kč

Zdroj: vlastní zpracování, XLSTAT 2019

6.2.4 ARIMA (0,1,1)

Analýza časové řady metodou ARIMA (0,1,1) byla vytvořena také v XLSTAT. Graf vývoje trendu spolu s předpovědí včetně odhadnutého intervalu lze vidět na následujícím grafu.

Graf 38 – Predikce ARIMA (0,1,1) mzdy



Zdroj: vlastní zpracování, XLSTAT 2019

Z grafu lze vidět, že na křivku skutečných hodnot „log“ navazuje předpověď spolu s hranicemi spolehlivosti, které jsou nastaveny na 95% interval spolehlivosti. Porovnání predikovaných dat s reálnými daty v roce 2018 je uvedeno v následující tabulce.

Tabulka 62 – Srovnání hodnot ARIMA (0,1,1) mzdy

Predikované hodnoty	Skutečné hodnoty	
	ARIMA (0,1,1)	
28 649,34 Kč	2018 - Q1	30 285,32 Kč
29 435,58 Kč	2018 - Q2	31 866,07 Kč
29 297,19 Kč	2018 - Q3	31 516,92 Kč
31 373,58 Kč	2018 - Q4	32 643,67 Kč

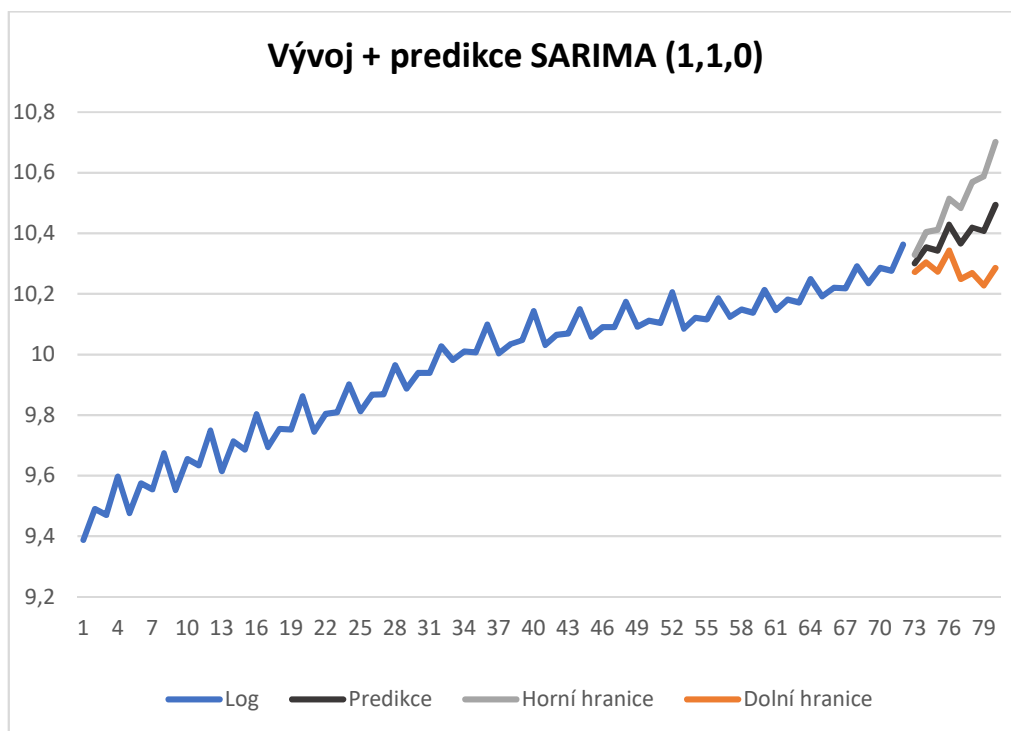
Zdroj: vlastní zpracování, XLSTAT 2019

Při porovnání predikovaných hodnot pomocí ARIMA (1,1,1) a ARIMA (0,1,1) se skutečnými hodnoty, je uvažována také suma odchylek od skutečných hodnot. To pomůže k lepší identifikaci, který model je pro predikci přesnější. Nejprve byly vypočteny odchylky od skutečných hodnot v každém období a poté se z nich pomocí funkce ABS vytvořily absolutní hodnoty. Výsledky těchto hodnot, pro ARIMA (1,1,1) **7744,73** a pro ARIMA (0,1,1) **7556,29**, vypovídají o tom, že menší sumu odchylek má metoda ARIMA (0,1,1), avšak tento výsledek není relevantní, kvůli již zmíněnému nahodilému zvyšování mezd. Při pohledu na **meziroční difference** časové řady „hrubá mzda“ je zřejmé, že hrubá mzda meziročně rostla do roku 2015 v řádech stovek, avšak od roku 2015 zvyšování mzdy bylo v řádech tisíců, a to kvůli rostoucí ekonomice České republiky.

6.2.5 SARIMA (1,1,0)

Model je predikován opět na 8 období. Pro predikci pomocí metody SARIMA (1,1,0) byl použit opět program NumXL. Vývoj časové řady spolu s navazující predikcí této řady je uveden v následujícím grafu.

Graf 39 – Predikce SARIMA (1,1,0) mzdy



Zdroj: vlastní zpracování, NumXL 2019

Horní a dolní mez jsou opět nastaveny na $\alpha = 0,05$, jedná se tedy o 95% interval spolehlivosti. Porovnání predikovaných dat s reálnými daty v roce 2018 je uvedeno v následující tabulce.

Tabulka 63 – Srovnání hodnot SARIMA (1,1,0) mzdy

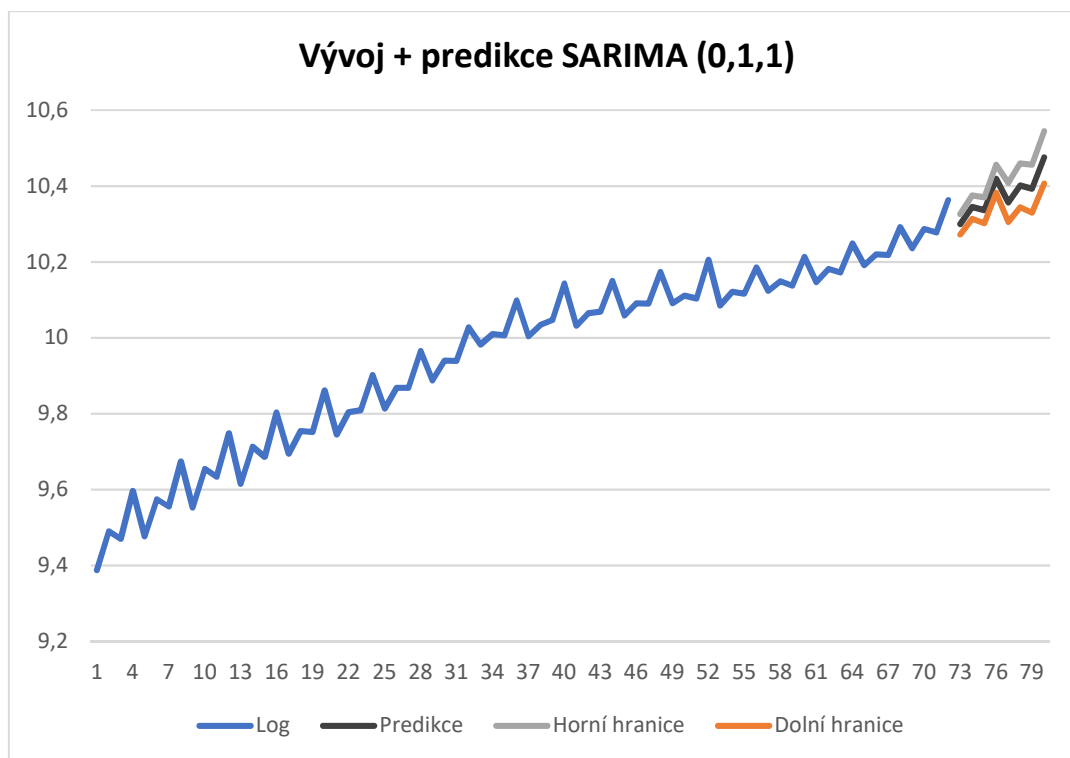
Predikované hodnoty SARIMA (1,1,0)	Skutečné hodnoty	
	29 811,88 Kč	2018 - Q1
31 378,50 Kč	2018 - Q2	31 866,07 Kč
31 049,18 Kč	2018 - Q3	31 516,92 Kč
33 676,70 Kč	2018 - Q4	32 643,67 Kč

Zdroj: vlastní zpracování, NumXL 2019

6.2.6 SARIMA (0,1,1)

Graf vývoje časové řady spolu s předpovědí včetně odhadnutého intervalu lze vidět na následujícím grafu.

Graf 40 – Predikce SARIMA (0,1,1) mzdy



Zdroj: vlastní zpracování, NumXL 2019

Podobně jako u předchozího případu, byla hodnota intervalu spolehlivosti nastavena na 95 %. Porovnání predikovaných hodnot se skutečnými hodnotami v roce 2018 je uvedeno v následující tabulce.

Tabulka 64 – Srovnání hodnot SARIMA (0,1,1) mzdy

Predikované hodnoty SARIMA (0,1,1)	Skutečné hodnoty	
	29 791,58 Kč	2018 - Q1
31 090,51 Kč	2018 - Q2	31 866,07 Kč
30 854,82 Kč	2018 - Q3	31 516,92 Kč
33 337,14 Kč	2018 - Q4	32 643,67 Kč

Zdroj: vlastní zpracování, NumXL 2019

U porovnávaných predikovaných hodnot u obou výše uvedených metod, se opět bere v úvahu suma odchylek od skutečných hodnot. Postup výpočtu je již uveden u předešlých metod. Hodnota sumy absolutních odchylek u modelu SARIMA (1,1,0) vyšla **2461,79** a hodnota u modelu SARIMA (0,1,1) vyšla **2624,86**. Z těchto výsledků lze tvrdit, že vhodnějším modelem pro predikci časové řady „hrubé mzdy“ je rozhodně SARIMA (1,1,0).

7 Porovnání výsledků

V této kapitole jsou u obou časových řad porovnány nejuvhodnější metody, které byly vypočteny na základě analýzy časových řad, za použití dekompozičního přístupu a vybraných přístupů z metodologie Box-Jenkins.

7.1 Časová řada „tržby z ubytování“

Při porovnání dekompozičního přístupu s metodami ARIMA a SARIMA (tabulka č. 49), bylo nejprve zjištěno, že tuto časovou řadu nejlépe vystihuje dekompoziční metoda trojitěho exponenciálního vyrovňování, jelikož u ní vyšla nejmenší hodnota ukazatele M.S.E.

U metody SARIMA byly za rok 2018 vypočteny sumy odchylek predikovaných hodnot od reálných hodnot. Pomocí tohoto výpočtu, bylo zjištěno, že suma odchylek u metody trojitěho exponenciálního vyrovňování vyšla **2889,22**, u metody SARIMA (0,1,1) vyšla **2638,93** a u metody SARIMA (1,1,0) vyšla **3179,4**.

Tabulka 65 - Porovnání odchylek (ubytování)

	Metody		
	Trojitě exponenciální vyrovňování	SARIMA (0,1,1)	SARIMA (1,1,0)
Abs. odchylka (2018-Q1)	699,37	1074,82	638,73
Abs. odchylka (2018-Q2)	731,67	529,48	1196,67
Abs. odchylka (2018-Q3)	613,09	479,89	1167,58
Abs. odchylka (2018-Q4)	845,09	554,74	176,42
SUMA	2889,22	2638,93	3179,40

Zdroj: vlastní zpracování, Excel 2019

Tento výsledek je známkou toho, že pro modelování vývoje tržeb za ubytování a jejich predikci pro rok 2018 je lepší model SARIMA (0,1,1), který je uveden v kapitole 3.3.

Lze konstatovat, že model SARIMA (0,1,1) si pomocí diferencování poměrně dobře poradil s trendem analyzované časové řady, i přes poměrně prudký nárůst tržeb v roce 2009-2010.

7.2 Časová řada „hrubá mzda“

Porovnání použitých metod při analýze této časové řady je obdobně provedeno na základě srovnání hodnot M.S.E. (tabulka č. 57). Při analýze bylo zjištěno, že obecně lepší metody pro modelování časové řady „hrubé mzdy“ vycházejí z metodologie Box-Jenkins. Nejmenší hodnotu extrapolacího kritéria M.S.E. a kritéria M.A.P.E. vykazuje model ARIMA (1,1,1).

U metody SARIMA byly za rok 2018 vypočteny sumy odchylek predikovaných hodnot od reálných hodnot. Pomocí tohoto ukazatele, bylo zjištěno, že suma odchylek u metody ARIMA (1,1,1) vyšla **7744,74**, u metody SARIMA (1,1,0) vyšla **2461,79** a u metody SARIMA (0,1,1) vyšla **2624,86**.

Tabulka 66 - Porovnání odchylek (mzdy)

	Metody		
	ARIMA (1,1,1)	SARIMA (1,1,0)	SARIMA (0,1,1)
Abs. odchylka (2018-Q1)	1739,77	473,44	493,74
Abs. odchylka (2018-Q2)	2438,12	487,57	775,56
Abs. odchylka (2018-Q3)	2217,35	467,74	662,10
Abs. odchylka (2018-Q4)	1349,50	1033,03	693,47
SUMA	7744,74	2461,79	2624,86

Zdroj: vlastní zpracování, Excel 2019

Z výše uvedené tabulky je zřejmé, že nejlepší metodou pro modelování vývoje hrubé mzdy v ČR a její predikci pro rok 2018, je metoda SARIMA (1,1,0), která je uvedena v 3.3.

Závěr

V diplomové práci byla analyzována časová řada „tržby za ubytování“ mezi roky 2000 až 2017 a časová řada „hrubá mzda“ mezi roky 2000 až 2017. V obou případech se jedná o intervalovou časovou řadu se čtvrtletní periodicitou. V prvních čtyřech kapitolách byly definovány základní pojmy a popsány vybrané metody pro analýzu a také pro predikci časových řad. V praktické části pak byl použit software Excel spolu s doplňky NumXL a XLSTAT k aplikaci vybraných metod na časové řady.

Cílem diplomové práce byla aplikace vybraných metod na reálná data a porovnání výsledků, pro kterou časovou řadu je vhodnější, která metoda. Protože teorie časových řad je velice rozsáhlá a oblast ekonomických časových řad je v neustálém rozvoji, byly zvoleny jen metody, které lze poměrně snadno použít v praxi. Na obě analyzované časové řady byly použity totožné metody a přístupy, aby se výsledné hodnoty daly lépe interpretovat a porovnat. Při zpracovávání analýz bylo zjištěno, že nezáleží jen na výběru statistického přístupu, ale také na samotné ekonomické podstatě problému, který je řešen, jelikož někdy je dynamický systém popsán nejlépe jednoduchým modelem a jindy zase velice složitým modelem. Přístupy a metody, které byly v práci použity jsou především modelové – prezentují, jak vhodně využít postupy pro analýzu časových řad k praktickému řešení v oblasti makroekonomie („hrubé mzdy“) nebo mikroekonomie („tržby za ubytování“) v určitém podniku.

V diplomové práci byly nejčastěji pro porovnání přístupů a metod použita hodnotící kritéria M.S.E. (střední čtvercová chyba odhadu), M.A.P.E. (střední absolutní procentní chyba odhadu), koeficient determinace a suma odchylek predikovaných hodnot od reálných. Toto poslední kritérium bylo použito především v kapitole „Extrapolace časových řad,“ ve které byly predikované hodnoty pomocí nejvhodnějších metod porovnávány s reálnými hodnotami v roce 2018.

Nejvhodnější metodou pro analýzu časové řady „tržby za ubytování“ z klasických přístupů je kvadratický trend, z adaptivních přístupů trojitě exponenciální vyrovnávání a z metodologie Box-Jenkins je nejvhodnější SARIMA (0,1,1). Nejvhodnější a nejpřesnější metodou pro predikci roku 2018 u této časové řady vyšla metoda SARIMA (0,1,1). Nejvhodnější metodou pro analýzu časové řady „hrubá mzda“ je z klasických přístupů kvadratický trend, z adaptivních přístupů trojitě exponenciální vyrovnávání a

z metodologie Box-Jenkins je nejvhodnějším modelem SARIMA (1,1,0). Nejvhodnější a nejpřesnější metodou pro predikci budoucích hodnot u této časové řady vyšla metoda SARIMA (1,1,0). Z těchto výsledků lze vidět, že pro každou časovou řadu je nutné použít různé metody analýzy, jelikož se každá časová řada chová odlišně a může být zastoupena různými složkami.

Literatura a další zdroje

Literatura

- ANDERSEN, Torben G. *Handbook of financial time series*. Berlin: Springer, c2009.
- ARLT, Josef a Markéta ARLTOVÁ. *Ekonomické časové řady*. Praha: Professional Publishing, 2009. ISBN 978-80-86946-85-6.
- ARLT, Josef a Markéta ARLTOVÁ. *Ekonomické časové řady: [vlastnosti, metody modelování, příklady a aplikace]*. Praha: Grada, 2007. ISBN 978-80-247-1319-9.
- BOX, George E. P., Gwilym M. JENKINS a Gregory C. REINSEL. *Time series analysis: forecasting and control*. 4th ed. Hoboken, N.J.: John Wiley, c2008. ISBN 978-0-470-27284-8.
- CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii: celostátní vysokoškolská učebnice pro stud. matem.-fyz. fakult studijních oborů I I Fyzikálně matematické vědy*. Praha: Státní nakladatelství technické literatury, 1986.
- CIPRA, Tomáš. *Finanční ekonometrie*. Praha: Ekopress, 2008. ISBN 978-80-86929-43-9.
- FORBELSKÁ, Marie. *Stochastické modelování jednorozměrných časových řad*. Brno: Masarykova univerzita, 2009. ISBN 978-80-210-4812-6.
- HAMILTON, James D. *Time series analysis*. Princeton, N.J.: Princeton University Press, c1994. ISBN 978-0-691-04289-3.
- HENDL, Jan. *Statistika v aplikacích*. Praha: Portál, 2014. ISBN 978-80-262-0700-9.
- HINDLS, Richard, Ilja NOVÁK a Jara KAŇOKOVÁ. *Metody statistické analýzy pro ekonomy*. Praha: Management Press, 1997. ISBN 80-859-4344-1.
- HINDLS, Richard, Ilja NOVÁK a Stanislava HRONOVÁ. *Metody statistické analýzy pro ekonomy*. 2. přeprac. vyd. Praha: Management Press, 2000. ISBN 80-726-1013-9.
- HINDLS, Richard. *Statistika pro ekonomy*. 8. vyd. Praha: Professional Publishing, 2007. ISBN 978-80-86946-43-6.

- KLUFOVÁ, Renata, Michael ROST a Jana KLICNAROVÁ. *Modelování regionálních procesů*. Third edition. Praha: Alfa Nakladatelství, 2012. Ekonomie studium. ISBN 978-80-87197-53-0.
- KOČENDA, Evžen a Alexandr ČERNÝ. *Elements of time series econometrics: an applied approach*. Third edition. Prague: Charles University in Prague, Karolinum Press, 2015. ISBN 978-80-246-3199-8.
- KOZÁK, Josef, Josef ARLT a Richard HINDLS. *Úvod do analýzy ekonomických časových řad*. Praha: Vysoká škola ekonomická, 1994. ISBN 80-707-9760-6.
- MONTGOMERY, Douglas C., Cheryl L. JENNINGS a Murat KULAHCI. *Introduction to time series analysis and forecasting*. Hoboken, N.J.: Wiley-Interscience, c2008. ISBN 978-0-471-65397-4.
- ŘEZANKOVÁ, Hana a Tomáš LÖSTER. *Úvod do statistiky*. Praha: Oeconomica, 2009. ISBN 978-80-245-1514-4.
- ZICHOVÁ, Jitka. *Non-negative time series and their applications*. Praha: Matfyzpress, 2011. ISBN 978-807-3781-477.

Webové stránky

- ARAD [online]. [cit. 2019-04-11]. Dostupné z: <https://www.cnb.cz/docs/ARADY/HTML/index.htm>
- Český statistický úřad [online]. [cit. 2019-04-15]. Dostupné z: https://www.czso.cz/csu/czso/casove_rady
- HANČLOVÁ, Jana a Lubor TVRDÝ. *Úvod do analýzy časových řad* [online]. 2003 [cit. 2019-03-29]. Dostupné z: https://www.fd.cvut.cz/department/k611/PEDAGOG/VSM/7_AnalyzaCasRad.pdf
- Statsoft [online]. 2013 [cit. 2019-03-21]. Dostupné z: http://www.statsoft.cz/file1/PDF/newsletter/2013_04_02_StatSoft_Casove_rady_v_softwaru_STATISTICA.pdf
- Wikimedia [online]. [cit. 2019-04-15]. Dostupné z: https://www.wikimedia.cz/wiki/Hlavn%C3%AD_strana

Seznam použitých tabulek, grafů a obrázků

Tabulky

Tabulka 1 - Přehled orientačních testů	16
Tabulka 2 - Statistická kritéria.....	21
Tabulka 3 - Ukázka DUMMY	37
Tabulka 4 - Časová řada tržeb za ubytování (v mil. Kč)	38
Tabulka 5 - Základní charakteristiky (ubytování tržby)	39
Tabulka 6 - Časová řada hrubá mzda (v Kč)	40
Tabulka 7 - Základní charakteristiky (hrubá mzda).....	41
Tabulka 8 – Upravené koeficienty determinace (ubytování).....	42
Tabulka 9 - Hodnoty koeficientů b (ubytování)	44
Tabulka 10 - Výsledky kritérií (ubytování)	44
Tabulka 11 - Hodnoty vah (ubytování).....	46
Tabulka 12 - Počáteční vyr. statistiky dvojité (ubytování).....	47
Tabulka 13 - Vyrovnané statistiky dvojité (ubytování)	48
Tabulka 14 - Počáteční vyrovnávací statistiky trojité (ubytování)	50
Tabulka 15 - Vyrovnané statistiky trojité (ubytování).....	51
Tabulka 16 - Sezónní indexy (ubytování).....	53
Tabulka 17 - DUMMY proměnné (ubytování).....	54
Tabulka 18 – Upravené koeficienty determinace (mzdy).....	54
Tabulka 19 - Hodnoty koeficientů b (mzdy)	56
Tabulka 20 - Výsledky kritérií (mzdy)	57
Tabulka 21 - Počáteční vyr. statistiky dvojité (mzdy).....	59
Tabulka 22 - Vyrovnané statistiky dvojité (mzdy)	60
Tabulka 23 - Počáteční vyr. statistiky trojité (mzdy).....	62
Tabulka 24 - Vyrovnané statistiky trojité (mzdy).....	63
Tabulka 25 - Sezónní indexy (mzdy).....	65
Tabulka 26 - DUMMY proměnné (mzdy).....	66
Tabulka 27 - Základní hodnoty ARIMA (1,1,1) ubytování.....	69
Tabulka 28 -Hodnotící kritéria ARIMA (1,1,1) ubytování	69
Tabulka 29 - Základní hodnoty ARIMA (1,1,0) ubytování.....	69
Tabulka 30 - Hodnotící kritéria ARIMA (1,1,0) ubytování	70

Tabulka 31 - Základní hodnoty ARIMA (0,1,1) ubytování.....	70
Tabulka 32 - Hodnotící kritéria ARIMA (0,1,1) ubytování	70
Tabulka 33 - Hodnoty ARIMA (1,1,0) ubytování	71
Tabulka 34 - Koeficienty SARIMA (1,1,0) ubytování.....	72
Tabulka 35 - Hodnoty SARIMA (1,1,0) ubytování.....	72
Tabulka 36 - Koeficienty SARIMA (0,1,1) ubytování.....	73
Tabulka 37 - Hodnoty SARIMA (0,1,1) ubytování.....	74
Tabulka 38 - Základní hodnoty ARIMA (1,1,1) mzdy.....	77
Tabulka 39 - Hodnotící kritéria ARIMA (1,1,1) mzdy.....	77
Tabulka 40 - Základní hodnoty ARIMA (1,1,0) mzdy.....	77
Tabulka 41 – Hodnotící kritéria ARIMA (1,1,0) mzdy	78
Tabulka 42 – Základní hodnoty ARIMA (0,1,1) mzdy	78
Tabulka 43 - Hodnotící kritéria ARIMA (0,1,1) mzdy.....	78
Tabulka 44 - Hodnoty ARIMA (1,1,1) mzdy	79
Tabulka 45 - Koeficienty SARIMA (1,1,0) mzdy	79
Tabulka 46 - Hodnoty SARIMA (1,1,0) mzdy.....	80
Tabulka 47 - Koeficienty SARIMA (0,1,1) mzdy	81
Tabulka 48 - Hodnoty SARIMA (0,1,1) mzdy.....	81
Tabulka 49 - Hodnoty M.S.E. (ubytování)	84
Tabulka 50 – Hodnoty M.A.P.E. (ubytování).....	84
Tabulka 51 - Srovnání hodnot kvadr. trend (ubytování)	85
Tabulka 52 - Srovnání hodnot trojitě exp. vyrovnávání (ubytování)	87
Tabulka 53 - Srovnání hodnot ARIMA (1,1,1) ubytování	88
Tabulka 54 - Srovnání hodnot ARIMA (1,1,0) ubytování	89
Tabulka 55 - Srovnání hodnot SARIMA (1,1,0) ubytování	90
Tabulka 56 – Srovnání hodnot SARIMA (0,1,1) ubytování.....	91
Tabulka 57 – Hodnoty M.S.E. (mzdy).....	92
Tabulka 58 – Hodnoty M.A.P.E. (mzdy).....	93
Tabulka 59 – Srovnání hodnot kvadr. Trend (mzdy).....	94
Tabulka 60 – Srovnání hodnot trojitě exp. Vyrovnávání (mzdy).....	95
Tabulka 61 – Srovnání hodnot ARIMA (1,1,1) mzdy.....	96
Tabulka 62 – Srovnání hodnot ARIMA (0,1,1) mzdy.....	97
Tabulka 63 – Srovnání hodnot SARIMA (1,1,0) mzdy.....	99

Tabulka 64 – Srovnání hodnot SARIMA (0,1,1) mzdy.....	100
Tabulka 65 - Porovnání odchylek (ubytování)	101
Tabulka 66 - Porovnání odchylek (mzdy)	102

Grafy

Graf 1 - Časová řada (tržby za ubytování).....	39
Graf 2 - Časová řada (hrubá mzda).....	41
Graf 3 - Lineární trend (ubytování)	42
Graf 4 - Kvadratický trend (ubytování)	43
Graf 5 - Exponenciální trend (ubytování).....	43
Graf 6 - Vážený klouzavý průměr (ubytování).....	46
Graf 7 - Dvojité exp. vyrovnávání (ubytování)	49
Graf 8 - Trojité exp. vyrovnávání (ubytování)	52
Graf 9 - Sezónní očištění (ubytování).....	53
Graf 10 - Lineární trend (mzdy)	55
Graf 11 - Kvadratický trend (mzdy)	55
Graf 12 - Exponenciální trend (mzdy).....	56
Graf 13 - Vážený klouzavý průměr (mzdy).....	58
Graf 14 - Dvojité exp. vyrovnávání (mzdy)	61
Graf 15 - Trojité exp. vyrovnávání (mzdy).....	64
Graf 16 - Sezónní očištění (mzdy).....	65
Graf 17 - ACF časové řady tržeb z ubytování	67
Graf 18 - PACF časové řady tržeb z ubytování	67
Graf 19 - Rezidua ubytování.....	68
Graf 20 – ARIMA (1,1,1) (ubytování)	68
Graf 21 – SARIMA (1,1,0) hodnoty ubytování.....	73
Graf 22 - SARIMA (0,1,1) hodnoty ubytování	74
Graf 23 - ACF časové řady hrubé mzdy	75
Graf 24 - PACF časové řady hrubá mzda	75
Graf 25 - Rezidua mzdy.....	76
Graf 26 – ARIMA (1,1,1) mzdy	76
Graf 27 - SARIMA (1,1,0) hodnoty mzdy.....	80
Graf 28 - SARIMA (0,1,1) hodnoty mzdy.....	82

Graf 29 - Kvadratická predikce (ubytování).....	85
Graf 30 - Predikce trojité exp. vyrovnávání (ubytování).....	86
Graf 31 - Predikce ARIMA (1,1,1) ubytování.....	87
Graf 32 - Predikce ARIMA (1,1,0) ubytování.....	88
Graf 33 - Predikce SARIMA (1,1,0) ubytování.....	90
Graf 34 - Predikce SARIMA (0,1,1) ubytování.....	91
Graf 35 – Kvadratická predikce (mzdy)	93
Graf 36 – Predikce trojitého exp. Vyrovnávání (mzdy)	95
Graf 37 – Predikce ARIMA (1,1,1)	96
Graf 38 – Predikce ARIMA (0,1,1) mzdy	97
Graf 39 – Predikce SARIMA (1,1,0) mzdy	98
Graf 40 – Predikce SARIMA (0,1,1) mzdy	99

Obrázky

Obrázek 1 - Stacionární vs nestacionární časové řady.....	27
Obrázek 2 - Korelogram	28

Abstrakt

ŠEFFL, Josef. *Ekonomické časové řady*. Plzeň, 2019. 110 s. Diplomová práce. Západočeská univerzita v Plzni. Fakulta ekonomická.

Klíčová slova: časové řady, dekompozice časových řad, metodologie Box-Jenkins, analýza časových řad, extrapolace.

Tato diplomová práce je zaměřena na analýzu vybraných časových řad „tržby za ubytování“ mezi roky 2000-2017 a „hrubá mzda“ mezi roky 2000-2017. Cílem diplomové práce je aplikace vybraných metod na reálná data a porovnání výsledků. Použité metody vychází z dekompozičního přístupu a z přístupu Box-Jenkins. Pro výběr vhodného předpovědního modelu pro obě časové řady jsou zkonstruovány předpovědi pro rok 2018 a následně porovnány se skutečnými hodnotami. V konečné fázi jsou metody a modely mezi sebou porovnány a jsou vybrány ty nejvhodnější pro modelování obou časových řad.

Abstract

ŠEFFL, Josef. *The economic time series*. Plzeň, 2019. 110 s. Diploma Thesis. University of West Bohemia. Faculty of Economics.

Key words: time series, time series decomposition, Box-Jenkins methodology, time series analysis, extrapolation.

The diploma thesis is focused on analysis of selected time series „sales for accommodation“ between 2000-2017 and „gross wage“ between 2000-2017. The aim of this thesis is to apply selected methods to real data and compare results. The selected methods are from decomposition approach and from the Box-Jenkins methodology. To select the appropriate prediction model for both time series, there are constructed forecasts for 2018 and then compared to actual values. In the final phase, the methods and models are compared with each other and the most suitable models are selected for modeling both time series.