Ekonomická fakulta
Faculty of Economics
Jihočeská univerzita v Českých Budějovicích
University of South Bohemia in České Budějovice

Jihočeská univerzita v Českých Budějovicích

Ekonomická fakulta

Katedra aplikované matematiky a informatiky

Bakalářská práce

Analýza vybraných ekonomických časových řad

Vypracovala: Michaela Bednářová

Vedoucí práce: RNDr. Jana Klicnarová, Ph.D.

České Budějovice 2014

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH
Fakulta ekonomická
Akademický rok: **2012/2013**

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení:	**Michaela BEDNÁŘOVÁ**
Osobní číslo:	**E11028**
Studijní program:	**B6208 Ekonomika a management**
Studijní obor:	**Účetnictví a finanční řízení podniku**
Název tématu:	**Analýza vybraných ekonomických časových řad**
Zadávající katedra:	**Katedra aplikované matematiky a informatiky**

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

Cílem práce je naučit se využívat metody časových řad pro vyhodnocování vybraných časových řad. Studentka bude mít za úkol analyzovat vybrané časové řady - dle oboru svého zájmu. Zvolené časové řady nejprve prostuduje teoreticky a stanoví cíl analýzy. Na základě charakteru časových řad a cíle analýzy zvolí vhodné metody pro studium těchto řad, které nastuduje a aplikuje na zkoumané časové řady. Závěrem práce bude shrnutí a komentář dosažených výsledků.

Metodický postup:

1. Studentka se seznámí s problematikou časových řad, především dekompozicí časových řad, sezónním očišťováním a modelováním trendu.
2. Studentka si zvolí ekonomické časové řady pro analýzu.
3. Nastudované postupy a modely aplikuje na vybraná data s cílem najít modely, které nejlépe vysvětlují chování vybraných časových řad.
4. Závěrem studentka získané výsledky řádně interpretuje a okomentuje.

Rozsah grafických prací:	**cca 10 stran**
Rozsah pracovní zprávy:	**40 - 50 stran**
Forma zpracování bakalářské práce:	**tištěná**

Seznam odborné literatury:

1. **KLUFOVÁ, Renata, Michael ROST a Jana KLICNAROVÁ.** *Modelování regionálních procesů.* **1. vyd. Praha: Alfa nakladatelství, 2012, 247 s. Ekonomie studium. ISBN 978-808-7197-530.**
2. **CIPRA, Tomáš.** *Analýza časových řad s aplikacemi v ekonomii.* **1. vyd. Praha: SNTL - Nakladatelství technické literatury, 1986, 246 s.**
3. **CIPRA, Tomáš.** *Finanční ekonometrie.* **1. vyd. Praha: Ekopress, 2008, 538 s. ISBN 978-80-86929-43-9.**
4. **HAMILTON, James D.** *Time series analysis.* **Princeton: Princeton University Press, 1994, xiv, 799 s. ISBN 06-910-4289-6.**
5. **MONTGOMERY, Douglas C, Cheryl L JENNINGS a Murat KULAHCI.** *Introduction to time series analysis and forecasting.* **Hoboken, N.J.: Wiley-Interscience, c2008, xi, 445 s. ISBN 04-716-5397-7.**

Vedoucí bakalářské práce:	**RNDr. Jana Klicnarová, Ph.D.** Katedra aplikované matematiky a informatiky
Datum zadání bakalářské práce:	**2. ledna 2013**
Termín odevzdání bakalářské práce:	**15. dubna 2014**

doc. Ing. Ladislav Rolínek, Ph.D.
děkan

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH
EKONOMICKÁ FAKULTA
Studentská 13 (1)
370 05 České Budějovice
L.S.

prof. RNDr. Pavel Tlustý, CSc.
vedoucí katedry

V Českých Budějovicích dne 27. března 2013

Poděkování

Děkuji vedoucí mé bakalářské práce RNDr. Janě Klicnarové, Ph.D. za vedení a cenné metodické připomínky a rady při zpracování této práce.

Prohlášení k bakalářské práci

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

Datum

Podpis studenta

Obsah

1. Úvod

Cílem mé bakalářské práce je na základě analýzy časových řad, týkajících se vývoje mezd v určitém odvětví, zjistit chování těchto hodnot řady v čase. Ze statistického úřadu jsem si vybrala časovou řadu zachycující vývoj průměrných mezd v odvětví peněžnictví a pojišťovnictví. Na těchto datech budu tedy dále analyzovat jejich vývoj, trend, sezónnost a hledat vhodný regresní model. Na základě vytvořeného modelu se následně pokusím predikovat jejich budoucí hodnoty.

V první části bakalářské práce vysvětlím pojem časová řada, vyjmenuji jejich typy, popíši jejich složky a způsoby zpracování dat. Zmíním také, jaký význam má analýza časových řad a vyjmenuji některé specifické problémy, které by mě během analýzy mohly potkat. Další část věnuji základním postojům k analyzování dat, jako jsou například spektrální analýza, Boxova-Jenkinsova metodologie, apod. V neposlední řadě také popíši jednu z nejdůležitějších částí týkající se měr vhodností modelu, na základě jejichž výpočtu budu po dosazení potřebných parametrů do modelu schopna ověřit jeho vypovídací schopnost a následně předpovídat budoucí hodnoty. Dále také vysvětlím, jak časové řady můžeme očistit, nebo-li z nich eliminovat například trend, sezónnost apod.

Pro analyzování dat použiji software Statistica a Excel. V Excelu provedu pomocí bodových grafů subjektivní analýzu, kde se na základě korelačního pole pokusím vybrat co možná nejvhodnější trendovou křivku, která bude moje data prokládat nejlépe. Zvolím i možnost zobrazení indexu determinace, pomocí kterého svou hypotézu poté ověřím. Ve Statistice otestuji, zda řada vykazuje nějakou sezónnost pomocí F-testu a dále budu data analyzovat funkcí vícenásobné regrese a jednoduché lineární regrese. Prostřednictvím těchto metod získám potřebné parametry k určení a vytvoření vhodného modelu, na základě kterého budu schopná následně data vyhodnocovat do budoucna.

2. Literární rešerše

2.1. Časové řady

Data, která potřebujeme analyzovat různými statistickými metodami, jsou většinou ve formě tzv. časových řad. Můžeme je najít nebo získat v mnoha oblastech vědeckých, ale i jiných disciplín.

Pod pojmem časová řada si můžeme představit chronologicky uspořádaná sledování hodnot nějaké náhodné veličiny. Časovou řadou mohou být například indexy spotřebitelských cen za určité časové období, počty přijatých žáků na základní školy, počet narozených dětí v České republice, ceny akcií na burze, údaje o denní návštěvnosti kulturních památek, průměrné mzdy zaměstnanců, apod. Data ve tvaru časových řad, neboli časově posbíraná data, najdeme například ve vědě, technice, ekonomii, zemědělství, průmyslu, ale i v mnoha dalších oblastech všude kolem nás (Klufová, Rost, Klicnarová, 2012).

Jak také napsal Montgomery, Jennings a Kulahci (2008) ve své knize, analýzy časových řad jsou významné pro předpovídání a plánování důležitých rozhodnutí v určitých oblastech. Například v oblastech operačního managementu (předpověď velikosti produkce, …), financí (předpověď příjmů z investic, …), průmyslu, ekonomiky (předpovědi významných ekonomických ukazatelů, …) a mnoha dalších.

2.1.1. Typy časových řad

Podle Klufová, Rost, Klicnarová (2012) rozlišujeme dva základní typy časových řad, a to spojité a diskrétní časové řady:

- *„Spojité časové řady jsou řady, kde hodnoty časové řady známe a můžeme měřit v každém okamžiku“* (např. měření elektrického napětí na určitém místě),
- „*Diskrétní časové řady jsou naproti tomu řady, jejichž hodnoty známe jen v určitých nespojitých časových bodech*“ (např. teploty o šesté hodině, počet šťastně provdaných párů za daný rok, …).

Dále můžeme rozlišit i nízkofrekvenční časové řady (např. počty narozených dětí), vysokofrekvenční časové řady (např. ceny akcií, napětí sítě po sekundách), nebo krátkodobé (periodicita časové řady je kratší než jeden rok, pozorování jsou týdenní, měsíční) a dlouhodobé časové řady (periodicita časové řady je delší než 1 rok).

2.2. Význam analýzy časových řad

Při analyzování časové řady je naším cílem konstrukce modelu, který nejlépe popisuje, jaké chování můžeme od zkoumané časové řady očekávat. Pomocí vytvořeného modelu poté pochopíme chování a mechanismus časové řady. Ta může vykazovat pravidelné cykly, trend apod. Pomocí takové konstrukce můžeme také zjistit, jak a čím jsou hodnoty naší časové řady ovlivňovány a naopak, na čem pravděpodobně nezávisejí. Což nám pomůže částečně ovládat a tím i optimalizovat chování časové řady, výběrem a volbou vhodných statistických testů. Můžeme se také konkrétněji zaměřit na sezónní a cyklické složky, vývoj řady, a nebo její trend. V jiném případě můžeme řadu od těchto složek očistit a soustředit se na studování očištěné řady. Model, který následně získáme, můžeme využít k předpovídání budoucích hodnot této řady. Nesmíme ale zapomenout, že hodnoty časových řad jsou pouze náhodně pozorované veličiny. Z toho vyplývá, že námi vytvořený nejvhodnější model není schopen určit přesnou budoucí hodnotu, ale dá nám pouze její odhad (Rost, Klufová, Klicnarová, 2012).

Analýza časových řad, včetně předpovídání jejich budoucího chování, se stává jednou z nejdůležitějších oblastí v rozvoji současné statistiky. Hlavním důvodem rostoucího významu této disciplíny je fakt, že se úspěšně vyrovnává s popisem dynamických systémů, se kterými často přicházíme do styku (Cipra, 1986).

2.3. Specifické problémy analýzy časových řad

Nyní se stručně zmíním o některých problémech, které jsou spojovány se specifickým uspořádáním dat, a které by mohly narušit výzkum časové řady. Jedním z problémů je například volba časových bodů pozorování, kde nemáme možnost výběru frekvence pozorování zkoumaných veličin (tzn. data jsme již získaly naměřená v určité frekvenci a bylo by obtížné, v některých případech i nemožné, tyto data dopočítávat v jiných časových bodech a zbytečně je „přehušťovat“). V každém případě je naším úkolem volit pozorování se stejným intervalem. Jako dalším oříškem k rozuzlení se může stát problém s kalendářem. Tyto problémy jsou většinou způsobené přírodou nebo je za ně zodpovědný sám člověk. Patří sem například různá délka kalendářních měsíců, jiný počet víkendů a pracovních dnů v měsíci atd. V případě, že chceme s těmito daty pracovat, je potřeba tuto časovou řadu nejprve standardizovat.

V datech, která použiji k analýze, by se podle jejich charakteristiky žádný z těchto problémů neměl vyskytovat, proto nebudu jednotlivé části rozebírat dále a odkáži na odbornou publikaci, kde Cipra (1986) popisuje tuto problematiku podrobněji.

2.4. Základní postoje k analýze časových řad

Volba vhodné metody k analyzování konkrétní časové řady závisí na několika faktorech. Nyní zmíním ty nejdůležitější. Jedním z nich je *účel analýzy*, kde důležitým rozhodnutím je, jaký bude cíl naší analýzy, neboť různé metody výpočtů nás dovedou k různým předpovědím. Dalším faktorem je *typ časové řady*. Typem, nebo-li charakterem časové řady, můžeme rozumět délku časové řady, ale také předpokládané rozdělení náhodných veličin, které tvoří časovou řadu, a jejich vzájemný vztah. V neposlední řadě je důležitým rysem i *zkušenost a znalost analytika*, který rozbor časové řady provádí a jeho počítačové a programové vybavení (Cipra, 1986 & Klufová, Klicnarová, Rost, 2012).

Jak jsem již uvedla výše, k analyzování časových řad je velmi důležitá zkušenost a znalost analytika. Je mnoho analýz, kde neexistuje jednoznačný postup a důležité je pouze jeho subjektivní posouzení. Je zcela na něm, který postup si zvolí, a k jakým výsledkům se poté dostane. Jako první krok v analýze se proto doporučuje grafické znázornění zkoumané časové řady. Na základě získaného grafu teprve následuje rozhodnutí o dalším postupu. Z hodnot zanesených do grafu se analytik pokouší určit, zda průběh časové řady obsahuje nějaký trend, periodu či sezónnost, nebo naopak, zda se jedná o stacionární řadu (tj. střední hodnota a rozptyly jsou stále stejné). Dále postupuje analytik podle příslušných metod. Jsou to například:

- Dekompozice časových řad
- Spektrální analýza
- Boxova-Jenkinsova metodologie

O těchto metodách se zmíním pouze okrajově. Globálně jsou tyto statistické testy analytiky často využívané, ale pro moji práci nejsou tolik významné. Pravděpodobně je nevyužiji, protože spektrální analýza se zabývá především cyklickou složkou, která je v mé zkoumané řadě zanedbatelná. V této práci se ale budu zabývat především dekompozicí časových řad, jelikož mým cílem je analyzovat vývoj časové řady.

2.4.1. Spektrální analýza

Pomocí spektrální analýzy můžeme identifikovat cyklické složky. Tato analýza zkoumá časovou řadu jako součet goniometrických funkcí (tj. sinusových a cosinusových funkcí s různými frekvencemi) a bílého šumu. Naším cílem při této metodě je rozkladem nalézt hodnoty (frekvence), které jsou v analyzované řadě nejvíce významné. Ty získáme pomocí speciálních nástrojů, tzv. periodogramu a spektrální hustoty (Rost, Klicnarová, Klufová, 2012 & Cipra, 1986).

2.4.2. Boxova – Jenkinsova metodologie

Podle Cipry (1986) je úkolem Boxovo – Jenkinsovy metodologie rozbor časové řady, která obsahuje navzájem závislá (korelovaná) pozorování. Zkoumá právě vztahy mezi jednotlivými náhodnými pozorováními, která jsou navzájem závislá (tj. základem konstrukce modelu je reziduální složka časové řady a její korelační analýza).

Některé z modelů, jak opět uvedl Cipra (2008), jsou například proces klouzavých součtů MA (q) řádu q, který patří mezi nejjednodušší. Další tzv. autoregresní proces AR (p) řádu p, nebo smíšený proces ARMA (p, q) řádu p a q.

2.5. Míry vhodnosti modelu

Jak už zmínili Rost, Klicnarová, Klufová (2012), pro posouzení vhodnosti zvoleného modelu je důležité si uvědomit, co je cílem naší analýzy. Ať je to vysvětlení naměřených hodnot, predikce budoucích stavů, nebo vysvětlení mechanismu modelu a jeho následné řízení. U každého z cílů budeme preferovat jiný model. Existuje několik možností, jak sestrojit model zkoumané časové řady a jak tuto řadu analyzovat. Poté pochopitelně při různých postupech dosáhneme odlišných výsledků.

Zapletal (2000) ve své publikaci napsal: „*Základním kritériem pro rozhodování o vhodném typu trendové funkce by měla být kritéria věcně ekonomická*“. Čímž myslel, že by funkce měla být vybrána podle věcného hlediska analýzy určitého ekonomického jevu. Při této analýze by vlastně posuzoval, zda je funkce rostoucí nebo klesající, konvexní nebo konkávní atd.

Závěrem potřebujeme tedy najít míru vhodnosti modelu. I když si stanovíme, co bylo cílem našeho modelování, nemůžeme úplně jednoznačně posoudit jeho vhodnost. Zdrojem chyb vyskytujících se v předpovědích časových řad bývá většinou reziduální složka, kterou nelze předpovědět. Musíme tedy dávat pozor na to, abychom konstruovali modely, ve kterých je rozptyl reziduální složky minimální. Někdy může být velký rozptyl reziduální složky dán nevhodně zvoleným modelem, kde mohl být zapomenut nějaký podstatný člen. Míru vhodnosti modelu můžeme zjistit na základě následujících vzorců:

- Součet čtvercových chyb

$$SSE = \sum_{t=n+1}^{n+h} (y_t - \hat{y}_t)^2 = \sum_{t=n+1}^{n+h} e_t^2. \tag{1}$$

- Střední čtvercová chyba

$$MSE = \frac{1}{h} \sum_{t=n+1}^{n+h} (y_t - \hat{y}_t)^2 = \frac{1}{h} \sum_{t=n+1}^{n+h} e_t^2. \tag{2}$$

- Střední absolutní chyba

$$MAE = \frac{1}{h} \sum_{t=n+1}^{n+h} |y_t - \hat{y}_t| = \frac{1}{h} \sum_{t=n+1}^{n+h} |e_t|. \tag{3}$$

- Střední absolutní procentní chyba

$$MAPE = \frac{100}{h} \sum_{t=n+1}^{n+h} \left| \frac{y_t - \hat{y}_t}{y_t} \right|. \tag{4}$$

Klufová, Rost, Klicnarová (2012) také tvrdí, že: „*Míra SSE roste s každým bodem predikce či pozorování, proto se někdy používá míra MSE, jejíž velikost není ovlivněna délkou predikovaného úseku (tzv. vylepšená míra SSE). Míry MAE a MAPE se využívají především v případech s odlehlými pozorováními, neboť nepenalizují tak silně velkou chybu předpovědi“.*

2.6. Predikce budoucích hodnot

Předpovídání budoucích hodnot a analyzování chování řady je jedním z nejčastějších cílů analýzy časových řad a následně neopomenutelné tvorby nejvhodnějšího modelu. Tyto předpovědi jsou nadále využívány v mnoha analýzách, úvahách o vývoji určitého stavu a především při zkoumání vztahů mezi dvěma veličinami.

Musíme ovšem brát v potaz zmínku Čermákové a Střelečka (1995), že hodnoty vypočítané na základě naší predikce jsou zatíženy chybou. Vypočítané hodnoty tedy nebudou vykazovat úplně přesné informace, ale pouze jejich pravděpodobný odhad. Odhady a predikce lze provádět dvěma způsoby:

- Bodový odhad

Postup je zde takový, že podle určitého pravidla vypočítáme požadované číslo, které předpokládáme jako nejvíce se blížící odhadované hodnotě. Postup výpočtu není náročný a je obvykle používaný. Jedinou nevýhodou je fakt, že nelze zajistit stoprocentní přesnost tohoto odhadu.

- Intervalový odhad

Intervalový odhad vznikne zkonstruováním intervalu, který zahrnuje hledanou hodnotu našeho základního souboru dat. Volíme zde spolehlivost (1- α), kde hodnota $1 - \alpha$ je koeficient spolehlivosti a α je hladina významnosti. Tuto hodnotu si obvykle volíme sami. Volíme tedy $\alpha = 0{,}05$ nebo $\alpha = 0{,}01$. Při tomto odhadu, již podle odborné publikace Čermákové a Střelečka (1995), můžeme naši předpověď považovat za přesnější.

K tomuto tématu ve svém díle Montgomery, Jennings a Kulahci (2008) vymezili několik základních kroků, ke správnému modelování a předpovídání budoucích hodnot časových řad. Doporučují postupovat například podle následujících bodů: zobrazení časové řady a určení jejích základních funkcí s ohledem na možná odlehlá pozorování; eliminace trendových a sezónních složek pomocí diferenciace nebo tvorby vhodného modelu; tvorba a následný výběr modelu vhodného k dalším předpovědím; ověření vhodnosti vybraného modelu; zájem o rozdíly mezi původní řadou a hodnotami řady, které byly předpovězeny a mnoho dalších užitečných poznámek, které by analytik ve své práci neměl opomenout a měl by se jimi řídit.

2.7. Dekompozice časových řad

Dekompozicí časové řady rozumíme rozklad časové řady na jednotlivé složky. Montgomery, Jennings a Kulahci (2008) ve své knize mj. také uvádějí, že kromě „klasického“ přístupu k dekompozici časových řad pomocí trendových a sezónních složek, existuje mnoho dalších algoritmů nahrazujících tuto dekompozici. V základním matematickém modelu pro tuto dekompozici ale předpokládáme, že řada může obsahovat trend (T), sezónní složku (S), cyklickou složku (C) a náhodnou (reziduální, nesystematickou) složku ($\mathcal{E}$).

Cílem dekompozice časových řad je nalézt vhodné parametry pro funkci trendu, popř. sezónnosti a cykličnosti tak, aby nám po odečtení těchto modelových složek z časové řady zbyla jenom reziduální složka, která bude již pouze bílým šumem, popřípadě stacionární časovou řadou. Nyní popíši, co pod pojmy trend, sezónnost, cyklicita a reziduální složka rozumíme:

Trend (T)

Vyjadřuje dlouhodobé změny ve vývoji hodnot analyzovaného ukazatele (typ, tvar, funkci), dlouhodobého růstu či poklesu jeho střední hodnoty. Podle Artla a Artlové (2007) je trend způsoben faktory, které v dlouhém období působí stejným směrem (například technologie výroby, různé podmínky, stanovy). V případě, že z dlouhodobého hlediska nedochází k žádné změně a hodnoty ukazatele dané časové řady kolísají kolem určité úrovně, mluvíme o konstantním trendu. Trend může dále být rostoucí nebo klesající, strmý, mírný, ale může se i s postupem času měnit.

Sezónní složka (S)

Vyjadřuje pravidelně se opakující odchylky od trendové složky, které se odehrávají v rámci nějakého časového období (nejčastěji roku, týdne, apod.). Typicky tyto změny souvisejí se změnou ročního období, lidskými zvyky, průběhem pracovního týdne apod.

Cyklická složka (C)

Vyjadřuje kolísání okolo trendu, tedy nějaké pravidelné dlouhodobější fáze vzrůstu a poklesu. Délky cyklů bývají různé.

Náhodná (reziduální) složka (Ɛ)

Je část řady, která zbyde po eliminaci trendové, sezónní a cyklické složky a je tvořena náhodnými pohyby, které už svou povahou nejsou systematické. Jsou to nepodchytitelné a nepopsatelné vlivy působící na časovou řadu. (Klufová, Rost, Klicnarová, 2012 & Hindls, Hronová, Novák, 2000).

Máme dva základní způsoby rozkladu časových řad do složek, které jsem uvedla výše. Je to dekompozice aditivní a dekompozice multiplikativní:

Aditivní dekompozice

Tato dekompozice vznikne součtem všech složek časové řady. Tyto složky jsou vyjádřeny ve stejných jednotkách, tj. v jednotkách, ve kterých je udávána pozorovaná hodnota časové řady Y_t. Její využití je především tam, kde rozmanitost hodnot časové řady je v čase téměř neměnná. Používáme následující vyjádření:

$$Y_t = T_t + S_t + C_t + \varepsilon_t \tag{5}$$

Multiplikativní dekompozice

Multiplikativní rozklad je druhý případ pro rozložení náhodné veličiny do stejných složek jako v případě aditivního modelu. Jen sčítání je zde nahrazeno násobením. V tomto modelu má stejné jednotky jako pozorování časové řady Y_t pouze složka trendu (T). Ostatní složky jsou bezrozměrné, udávají pouze relativní změnu (Klufová, Rost, Klicnaroví, 2012).

Použití této dekompozice je podle Arlta, Arltové a Rublíkové (2002) vhodné v případě, že rozmanitost hodnot řady v čase roste, nebo se mění. Postupujeme podle následujícího vzorce:

$$Y_t = T_t * S_t * C_t * \varepsilon_t \tag{6}$$

kde:

Y_t … náhodná veličina uskutečňující se v čase t;

T_t … trendová složka této veličiny;

S_t … sezónní složka této veličiny;

C_t … cyklická složka této veličiny;

ε_t … náhodná složka této veličiny v čase t.

Dekompozice je v dnešní době statistiky využívaná především proto, že umožňuje zjistit určité chování vývoje analyzovaného jevu. Dále poskytuje možnost odstranění sezónní složky z časové řady, díky kterému můžeme současně porovnávat několik dalších časových řad ve stejný okamžik. V neposlední řadě odstraněním trendu budeme schopni lépe a výrazněji analyzovat a modelovat sezónnost. Nebo přesněji predikovat budoucí hodnoty a popisovat vývoj řady.

2.7.1. Metody eliminace trendu

Jak napsal Cipra (2008), v souvislosti s eliminací trendu můžeme mluvit také o vyrovnávání nebo vyhlazování časové řady, protože při tomto postupu potlačujeme sezónní a náhodné fluktuace.

K trendu můžeme přistupovat ze dvou směrů. Klufová, Klicnarová, Rost (2012) uvádí, že máme dva základní přístupy. *Globální* (známější), který využívá především regresní analýzu a předpokládá, že pro naší zkoumanou řadu existuje právě jedna funkce, s jejíž pomocí jsme schopni vyjádřit funkci trendu. Druhý přístup, *lokální,* nám říká, že také existuje funkce k popsání trendu, ale její parametry jsou platné pouze lokálně, což znamená, že se postupně mění.

Jednoduchým příkladem eliminace trendu je *subjektivní metoda průměrování horních a dolních bodů zvratu*. Zde nejprve spojíme lomenými čarami horní body zvratu a pak dolní body zvratu (tj. lokální maxima a minima dané trajektorie), a pro každý časový moment vyznačíme křivkou v obrázku střed vzdáleností mezi horní a dolní, námi vytvořenou, lomenou čárou. Metoda je nazývána subjektivní z toho důvodu, že není nikde přesně určeno, které body lze považovat za horní a dolní body zvratu.

Trend můžeme ale popsat i analyticky a to jednoduchými křivkami. Podle Cipry (2008) typ nejvhodnější matematické křivky pro danou časovou řadu určujeme nejčastěji na základě předběžného grafického záznamu řady, nebo na základě předpokládaných vlastností trendové složky.

Nejpoužívanější trendové křivky:

- Konstantní trend

$$Tr_1 = \beta_0\,, \qquad t = 1, \ldots, n, \tag{7}$$

- Lineární trend

$$Tr_1 = \beta_0 + \beta_1 t, \qquad t = 1, \ldots, n, \tag{8}$$

- Kvadratický trend

$$Tr_1 = \beta_0 + \beta_1 t + \beta_2 t^2, \qquad t = 1, \dots, n, \tag{9}$$

- Exponenciální trend

$$Tr_1 = \alpha\beta^t, \qquad t = 1, \dots, n \quad (\beta > 0) \tag{10}$$

Najdeme-li vhodnou trendovou funkci (křivku), potom pomocí MNČ odhadneme hodící se model. Na základě výpočtů vybereme ten, který bude mít podle metody nejmenších čtverců nejnižší hodnotu.

Volba vhodného modelu trendu

Kromě znalosti předpokládaných vlastností trendové funkce z vizuálního rozboru a grafického posouzení řady, můžeme například podle Pavelky a Klímy (2000) použít také testy založené na jednoduchých charakteristikách časové řady. Při závěrečném rozhodování o vhodném typu trendové funkce je důležité věnovat pozornost hodnotě koeficientu vhodnosti modelu (MNČ). Vybereme tu funkci, která bude mít tuto hodnotu nejnižší. V následující tabulce č. 1 můžeme vidět, jak lze pomocí jednoduchého testu dojít k závěru a překontrolovat tak pravdivost řešené hypotézy:

Tabulka 1: tabulka běžných trendových funkcí (testy)

Trend	Test trendových funkcí
Lineární	První diference přibližně konstantní
Kvadratický	Druhá diference přibližně konstantní
Exponenciální	Koeficient růstu (tj. podíly následujících hodnot) přibližně konstantní
Modifikovaný exponenciální	Podíly následujících hodnot jsou přibližně konstantní
Logistický	Podíly následujících prvních diferencí převrácených hodnot jsou přibližně konstantní
Gompertzův	Podíly následujících prvních diferencí zlogaritmovaných hodnot jsou přibližně konstantní

Zdroj: Pavelka, Klímek (2000) & Klicnarová, Rost, Klufová (2012)

Analýzu trendu můžeme ale provést i pomocí dalších adaptivních přístupů k trendové složce, také nazývaných mechanickým vyrovnáním časových řad, a to analýzou klouzavých průměrů, nebo častěji používaným exponenciálním vyrovnáním. Tyto adaptivní přístupy můžeme obecně charakterizovat pro jejich schopnost pracovat se systematickými složkami, které mění v čase svůj globální charakter. Takže pro ně nelze použít konkrétní matematickou křivku s konstantními parametry (Cipra, 2008).

Metoda klouzavých průměrů

Využití těchto průměrů závisí na trendu, který předpokládáme u konkrétní časové řady. Podstatou podle Hindlse, Hronové a Nováka (2000) je, že posloupnost původních empirických hodnot nahradíme řadou průměrů vypočítaných přímo z těchto pozorování. Při metodě klouzavých průměrů nejprve řadu rozdělíme na menší časové úseky a poté na těchto částech odhadujeme lokální trendy určitého stupně. Bohužel ale přicházíme o vyrovnané hodnoty pozorování na konci a na začátku řady.

Pokud předpokládáme lineární trend, Klicnarová, Klufová a Rost (2012) uvádějí ve své publikaci, že nám postačí k vyrovnání časové řady pouze *prostý klouzavý průměr* liché délky. Pokud je náš trend jiný než lineární, využijeme *vážené klouzavé průměry*, kde pro různé trendy najdeme různé váhy vhodných klouzavých průměrů, viz. dále například také Cipra (2008).

Exponenciální vyrovnání

Vyrovnaná hodnota je zde zvláštním příkladem klouzavého průměru, kdy všechny doposud pozorované hodnoty vyrovnané řady vážíme do minulosti exponenciálně klesajícími vahami (Cipra, 2008).

Analytik se následně sám může rozhodnout, a vybrat pro něj nejvhodnější způsob eliminace trendu. Jelikož budu svá data analyzovat pomocí regrese, budu tedy přistupovat globálně a předpokládat, že pro zkoumanou řadu existuje právě jedna funkce, s jejíž pomocí budu schopná dále vyjádřit hledanou funkci trendu. Ve své práci provedu analýzu trendu pomocí subjektivní metody průměrování horních a dolních bodů zvratu. Kde na základě grafického záznamu popíši trend analyticky pomocí jednoduchých křivek.

2.7.2. Metody eliminace sezónnosti

Nyní se budu zabývat metodami eliminujícími složku, která popisuje periodické změny v časové řadě odehrávající se pravidelně, nebo přibližně podobně v průběhu každého kalendářního roku. Cílem těchto metod je odstranit z časové řady pravidelné výkyvy. Tato analýza se často provádí až po dekompozici dat od trendu. Sezónní složku odhadujeme pomocí průměru (Cipra, 2008 & Klufová, Rost, Klicnarová, 2012).

Test sezónnosti:

Pokud potřebuji statisticky analyzovat, zda časová řada vykazuje konstantní sezónnost, potom podle Klufová, Rost, Klicnarová (2012), budu testovat nulovou hypotézu:

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0,$$

$$H_A: non\ H_0$$

Tuto hypotézu a sezónnost následně otestuji pomocí níže přiloženého vzorce:

$$\mathrm{F} = \frac{\frac{\mathrm{S_b}}{\mathrm{p}-1}}{\frac{\mathrm{S}_r}{(\mathrm{p}-1)(\mathrm{k}-1)}}, \tag{11}$$

kde:

β_p … sezónní faktory v analyzovaném modelu;

p … délka periody;

k … počet cyklů k pozorování;

S_b … celkový součet čtverců;

S_r … reziduální součet čtverců.

Jednotky, ve kterých jsou sezónní faktory měřeny závisí na tom, zda je příslušná dekompozice aditivní nebo multiplikativní. V případě aditivní dekompozice se sezónní faktory měří ve stejných jednotkách jako daná časová řada. U multiplikativní dekompozice jsou faktory bezrozměrné veličiny a je pro ni charakteristické, že výkyvy se s rostoucím trendem zvětšují a to i v případě, že se multiplikativní sezónní faktory v jednotlivých sezonách pravidelně opakují. Na rozdíl od aditivní dekompozice, kde monotonie trendu nic neovlivňuje (Cipra, 2008).

2.8. Charakteristika dat

Vzhledem k mému studijnímu oboru jsem si pro svou práci vybrala ekonomickou časovou řadu znázorňující čtvrtletní data průměrných měsíčních mezd v České republice od 1. čtvrtletí roku 2000 do 2. čtvrtletí roku 2013. Budu tedy analyzovat průměrné mzdy v konkrétním odvětví peněžnictví a pojišťovnictví. Zdrojem těchto dat je Český statistický úřad.

2.8.1. Průměrná hrubá měsíční mzda

Nyní podrobně popíši, co by si každý z nás měl pod pojmem průměrná hrubá měsíční mzda představit.

Jak je uvedeno v metodice Českého statistického úřadu (2013): "P*růměrná hrubá měsíční mzda představuje podíl mezd bez ostatních osobních nákladů připadající na jednoho zaměstnance evidenčního počtu za měsíc. Do mezd se zahrnují základní mzdy a platy, příplatky a doplatky ke mzdě nebo platu, prémie a odměny, náhrady mezd* *a platů, odměny za pracovní pohotovost a jiné složky mzdy nebo platu, které byly v daném období zaměstnancům zúčtovány k výplatě“.*

„Jedná se tedy o hrubé mzdy, tj. před snížením o pojistné na všeobecné zdravotní pojištění a sociální zabezpečení, zálohové splátky daně z příjmů fyzických osob a další zákonné nebo se zaměstnancem dohodnuté srážky“.

2.8.2. Analýza vývoje průměrných mezd

Jelikož data, která analyzuji ve své práci, se týkají průměrných mezd, uvádím pro přiblížení charakteristiky a problematiky tohoto oboru nejdůležitější údaje z analýzy, kterou provedl Luboš Marek (2010) na průměrných mzdách za celou Českou republiku.

Posoudil tedy mzdy a jejich historická data z hlediska jejich vývoje, z hlediska kvantilových měr, zajímal se také o vývoj variability mezd a jejich změn v čase. Kde zvyšující se variabilita říká, jak se změní diferenciace mezd v čase.

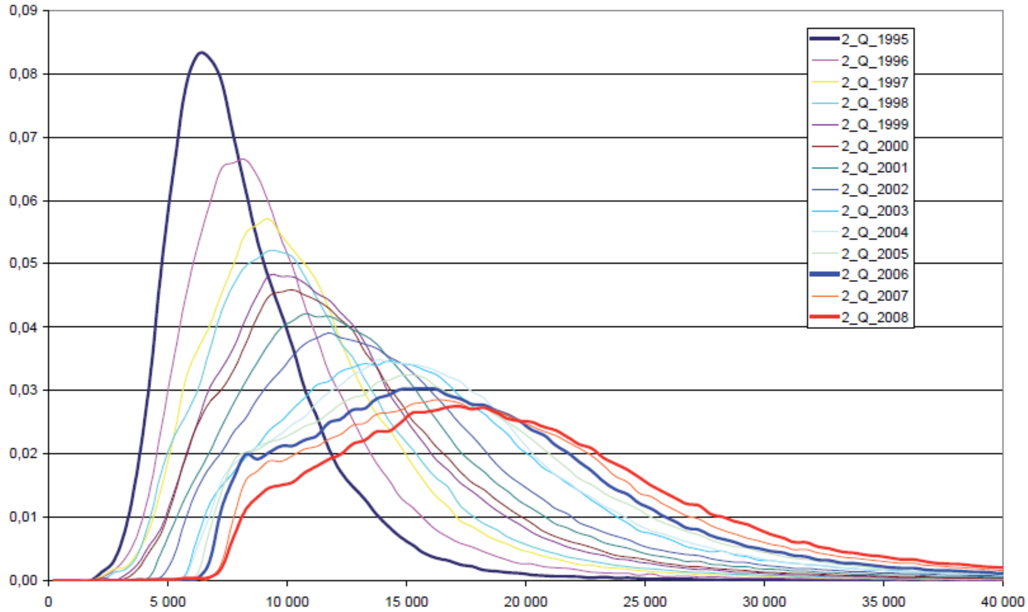
Jeho analýza zahrnovala pouze data za druhé čtvrtletí příslušného roku. Argumentem výběru konkrétně tohoto čtvrtletí byl fakt, že toto období má v celém roce nejstabilnější fond pracovní doby (tzn. nejméně svátků a jiných rušivých efektů analýzy). Ve své práci bral všechny zaměstnané jako rovnocenně hodnocené osoby (tj. údaje jsou propočítány jako „nepřepočtené“, tedy fyzické osoby). Musíme tedy brát na vědomí, že z důvodu tohoto způsobu výběru, se mohou v analýze projevit nějaké nežádoucí jevy ve formě nepřesností v předpovědích a ostatních analýzách.

Podle Marka (2010), který analyzoval mzdy za celou Českou republiku od roku 1995 do roku 2008 pomocí základních statistických charakteristik, se obecně průměrná mzda ve všech odvětvích vyvíjela následovně:

Četnost mezd

V první fázi došel k závěru, že během let se mzdy změnily ve svém rozložení, a výrazně se svojí výší od sebe v čase vzdalují. Tj. růst absolutní výše mezd a výrazný skok v jejich variabilitě, který je zobrazen grafem v následujícím obrázku č. 1, kde různobarevné křivky znázorňují, jak se od sebe mzdy v jednotlivých letech liší. Můžeme soudit, že mezi lety 1995 a 1996 byl skok obrovský, na rozdíl od let 2007 a 2008, kde je vzdálenost téměř zanedbatelná.

Obrázek 1: rozdělení četnosti mezd v čase



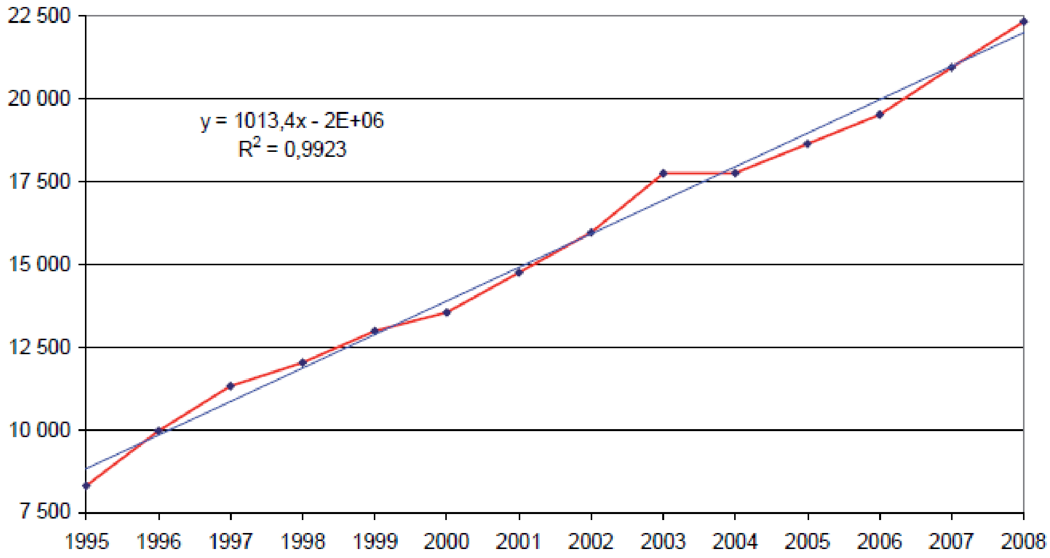
Zdroj: Marek (2010, VŠE)

Vývoj průměrné mzdy

Následující obrázek č. 2, znázorňující vývoj průměrné mzdy vyrovnaný lineární trendovou přímkou, nás informuje grafem o několika skutečnostech.

Z grafu je možné vyčíst, že průměrná mzda trvale roste a její vývoj je proložen lineární přímkou. Tato přímka je pro analýzu vhodná, což je patrné z hodnoty indexu determinace R = 0,9923, který nám ukazuje hodnotu dost vysokou na to, abychom nemuseli hledat více vyhovující model. Dále pomocí tohoto grafu, rovnice a trendové přímky, jsme schopni předpovědět průměrnou výši mzdy pro další období. Ovšem důležité je tuto predikovanou hodnotu pokládat pouze za orientační analytickou předpověď. Skutečná hodnota může v budoucnosti být ovlivněna různými neočekávanými ekonomickými faktory.

Obrázek 2: vývoj průměrné mzdy

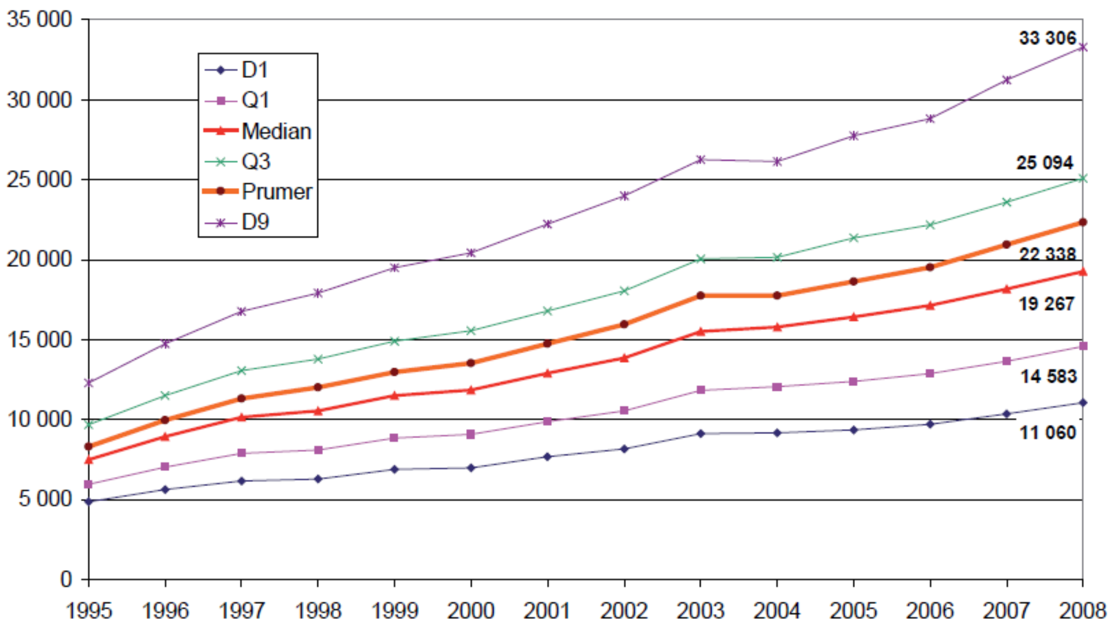


Zdroj: Marek (2010, VŠE)

Kvantilové míry mezd

Dalších několik zajímavých skutečností se můžeme dozvědět z grafu na obrázku č. 3, na základě kterého Marek (2010) objasnil problematiku kvantilů a průměrů v průměrných mzdách. Zde, pro pochopení, kvantil $\widetilde{x_p}$ znaku X můžeme například podle Čermákové a Střelečka (1995) vysvětlit následující definicí: *„Kvantilem nazveme a zvolíme číslo tak, aby většina hodnot znaku byla menších než* $\widetilde{x_p}$ *a* $100(1-P)\%$ *hodnot znaku bylo větších než* $\widetilde{x_p}$*“.* Dále mezi ostatní známé kvantily můžeme zařadit například tzv. kvartily: x_{25} (dolní kvartil), x_{50} (medián), x_{75} (horní kvartil). Tato tři čísla rozdělují systematicky uspořádanou řadu hodnot na čtyři zhruba stejně velké úseky, tj. na části se zhruba stejnými rozměry.

Obrázek 3: kvantily a průměry mezd



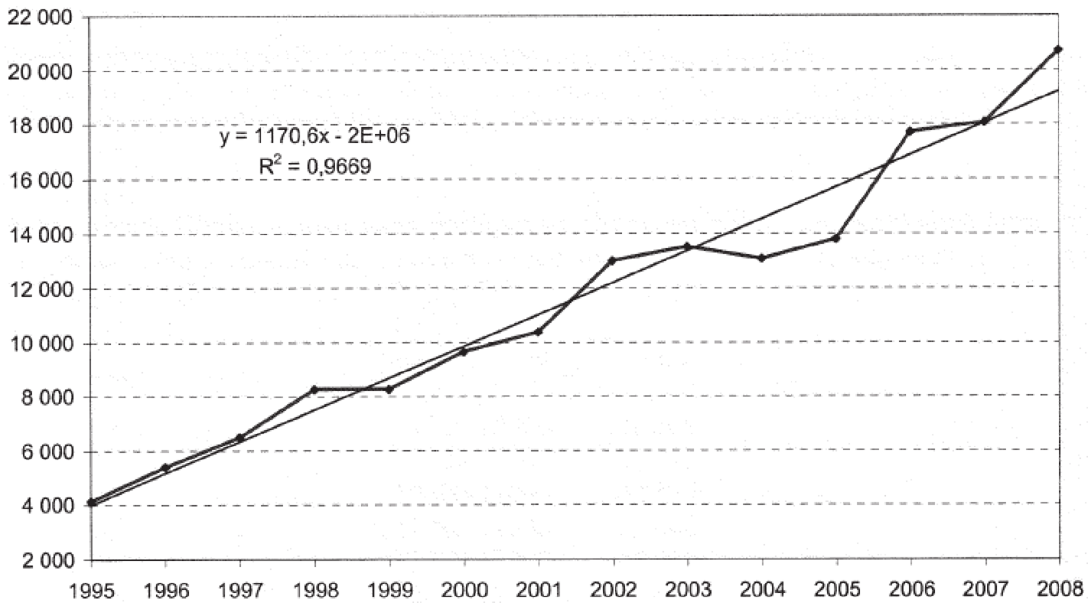
Zdroj: Marek (VŠE)

Červeně znázorněná křivka mediánu (tj. 50-ti procentní kvantil), která rozděluje naši množinu hodnot na dvě stejně velké části, nám říká, že polovina analyzovaných mezd je menší než 19 267 Kč (tj. zhruba dvě třetiny zaměstnanců mají nižší mzdu než je celostátní průměr) a rozdíl mezi průměrnou mzdou a mediánem mezd se v čase nepatrně zvětšuje. Aritmetický průměr máme o 3 071 Kč větší než medián.

Variabilita mezd

Pro variabilitu mezd, nebo-li jejich proměnlivost, měřenou pomocí směrodatné odchylky, můžeme stanovit podobné závěry jako pro průměrnou mzdu. Řekneme tedy, že variabilita (proměnlivost) mezd je v čase trvale rostoucí, z čehož můžeme usuzovat následnou zvětšující se diferenciaci mezd. Vývoj variability v čase je opět vyjádřen lineární přímkou jako nejvhodnější možnou trendovou křivkou. Mzdy tedy v čase rostou lineárním trendem a vzdalují se svou velikostí jedna od druhé viz. následující obrázek č. 4:

Obrázek 4: variabilita mezd



Zdroj: Marek (2010, VŠE)

Shrnutí

Podle grafů z výše zmíněných obrázků můžeme podle Marka (2010) učinit několik závěrů:

- Rozdělení četnosti mezd se mezi roky 1995 – 2008 výrazně změnilo jak svým tvarem, tak zároveň došlo ke změně variability, polohy, šikmosti a špičatosti rozdělení mezd,
- Průměrné mzdy jsou v čase rostoucí podle lineární trendové přímky,
- Lze sestrojit analytickou předpověď pro další rok, ale je potřeba dbát na další faktory, které by mohli předpověď výrazně odlišit od skutečnosti,

- Roste variabilita mezd a s ní i související zvyšující se diferenciace mezd.

Nyní bude mým úkolem vypracovat obdobnou analýzu, zaměřující se ovšem na zkoumání vývoje řady mezd v jednom konkrétním odvětví, a následně predikovat hodnoty v následujících obdobích.

3. Metodika

Účelem této práce je na základě využití metod časových řad pro vyhodnocování dat, analyzovat vybranou časovou řadu a to tak, že bude zjištěn trend a vývoj této řady. Dále stanovena a otestována sezónnost, pokud se v této řadě vyskytuje. Závěrem bude na základě získaných skutečností vytvořen vhodný model, který popisuje variabilitu nezávisle proměnných s největší přesností. Ten bude ověřen patřičnými statistickými testy, aby poté mohla být provedena predikce budoucího vývoje velikosti průměrné mzdy. Data použitá k analýzám v praktické části bakalářské práce jsou aktualizována a byla získána ze statistického úřadu.

V následující části bakalářské práce se budu tedy zabývat otázkou chování dat ve vybrané časové řadě. Podle charakteru vybraných dat využiji následující postupy a metody. Pomocí subjektivní analýzy dat a zobrazením časové řady na korelačním poli bodového grafu, se pokusím co nejpřesněji vysvětlit chování řady a odůvodnit zobrazovaný vývoj, který si již delší dobu podle informací statistického úřadu udržuje pozvolný růst lineárním trendem. Uvidíme ale, jak se na datech projeví skutečnost výskytu hospodářské nerovnováhy v určitém časovém úseku této analyzované řady. Poté pomocí jednoduché lineární regrese provedu F-test sezónnosti, a pokusím se tím prokázat hypotézu, o předpokládaném výskytu sezónnosti v této časové řadě. Z důvodu čtvrtletního měření dat zvolím pro získání potřebných parametrů sezónnosti a trendu do modelu statistický test vícenásobné regrese. Díky tomuto testu budu moci do modelu zahrnout i mezičtvrtletní změny působící na velikost mzdy. Rozhodující ve volbě vhodného modelu pro mě ale bude tzv. index determinace, jehož hodnota vyjadřuje jak vhodný je model k vysvětlení analyzovaných dat. Říká nám, z kolika procent vysvětlují nezávisle proměnné právě jednu závisle proměnnou veličinu. Budu tedy vytvářet různé modely tak dlouho, dokud hodnota indexu nebude co nejblíže 1. Poté na základě modelu, který bude nejlépe vysvětlovat chování časové řady, provedu bodový odhad a stanovím vývoj mzdy ve vybraném odvětví v dalších obdobích.

4. Vlastní práce

K analýze do praktické části mé bakalářské práce jsem si vybrala čtvrtletní data průměrných mezd z odvětví peněžnictví a pojišťovnictví za posledních 12 let. Čtvrtletní data jsem zvolila z důvodu větší vypovídací schopnosti analýzy a přesnější predikce budoucích hodnot. Odvětví jsem vybírala na základě vlastních zájmů. Velký význam při výběru hrál jak můj studijní obor na Jihočeské Univerzitě, tak případná volba budoucího povolání právě v tomto odvětví, které co se ohodnocení týče, se jeví velice lukrativním.

4.1. Základní práce s daty

Nejprve jsem se zabývala zjišťováním největšího nárůstu a poklesu velikosti mezd mezi lety 2000 a 2013. Pomocí programu Excel a funkce MAX a MIN jsem vypočítala, že největšího nárůstu dosáhla průměrná mzda v odvětví peněžnictví a pojišťovnictví v prvním čtvrtletí roku 2008. Mzda tu vzrostla o celých 13 425 Kč. Naopak největší propad jsem zaznamenala hned v následujícím, druhém, čtvrtletí roku 2008. Zde se mzda snížila o 11 977 Kč. Tato neočekávaná a velká změna, která viditelně rozhodila plynulý vývoj dat, byla pravděpodobně způsobena propukající ekonomickou krizí.

Dále jsem pomocí první diference zjišťovala průměrnou mezičtvrtletní změnu bazického indexu (tj. průměrný nárůst a pokles velikosti mezd mezi jednotlivými obdobími). V následující tabulce č. 2 uvádím hodnoty průměrného zvýšení nebo snížení mezd mezi jednotlivými čtvrtletími.

Zvláštním případem je zde fakt, že mzda mezi čtvrtým a prvním čtvrtletím vzrůstala o téměř stejnou částku jako v období ze třetího do čtvrtého čtvrtletí, kde je tento jev spíše očekávaný (vánoční benefity, 13. platy, atd.). Proto se později v této práci nad touto situací zamyslím.

Tabulka 2: průměrné mezičtvrtletní změny velikosti mezd

Čtvrtletí	Mezičtvrtletní změna (Kč)
1.	3 624
2.	- 3 010
3.	- 2 244
4.	3 806

Zdroj: vlastní zpracování

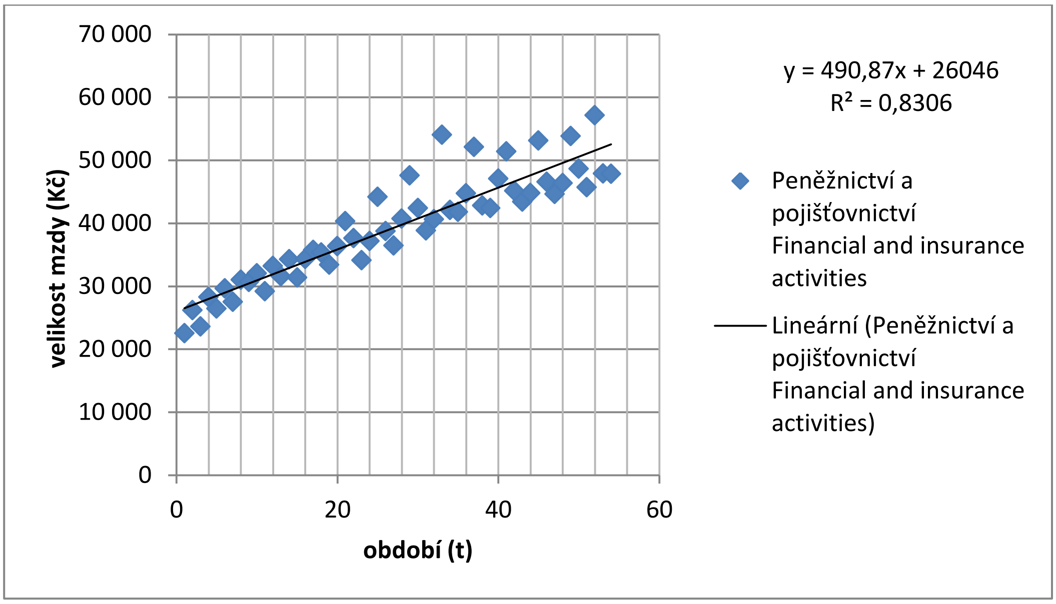
V tabulce si můžeme tedy všimnout, že v prvních čtvrtletích mezi roky 2000 a 2013 mzda průměrně vzrůstala o 3 624 Kč, v druhých a třetích čtvrtletích výrazně klesala o částku větší než 2 000 Kč a ve čtvrtých čtvrtletích opět rostla průměrně o 3 806 Kč.

4.2. Výběr regresního modelu

Pro správné analyzování jsem si data časové řady nejprve zobrazila v grafu, abych na jejich základě mohla vyčíst několik důležitých skutečností o zkoumané řadě a následně najít nejvhodnější model, který ji popisuje.

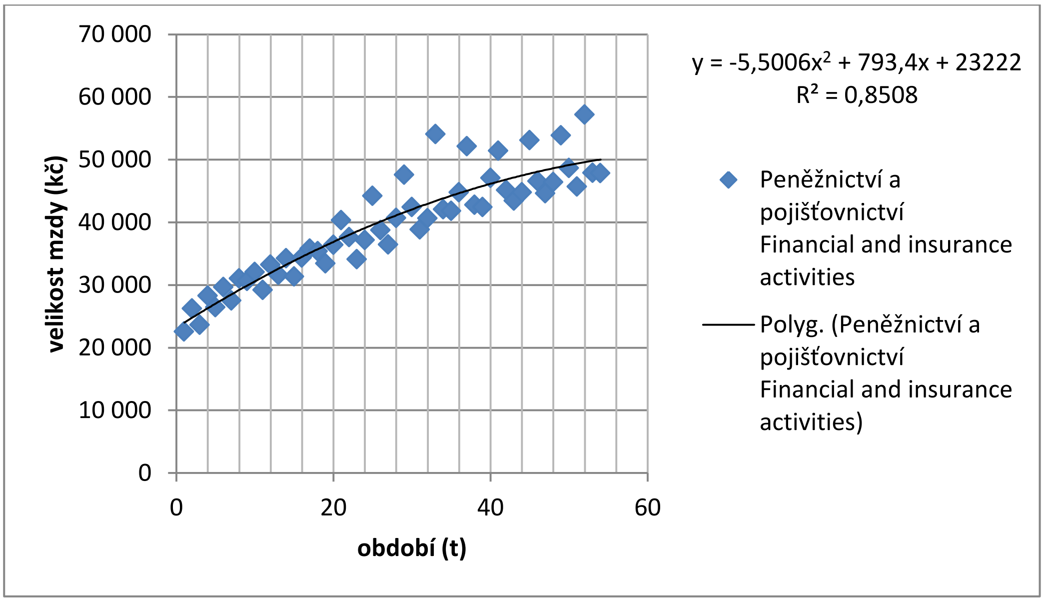
Danou časovou řadu jsem poté proložila různými křivkami (viz. bodový graf 1,2 a 3) tak, abych mohla zkoumat možnou závislost mezi náhodnými veličinami na korelačním poli a zjistit tak, jaký druh závislosti bude pro mé zkoumané veličiny nejvhodnější. Hledala jsem tedy vztah mezi dvěma proměnnými, který jsem vymodelovala na korelačním poli. Jako závisle proměnná „*Y*“ v této analýze bude průměrná mzda v peněžním odvětví a nezávisle proměnná „*X*“ bude čas.

Graf 1: mzdy (peněžnictví a pojišťovnictví) proložené lineárním trendem



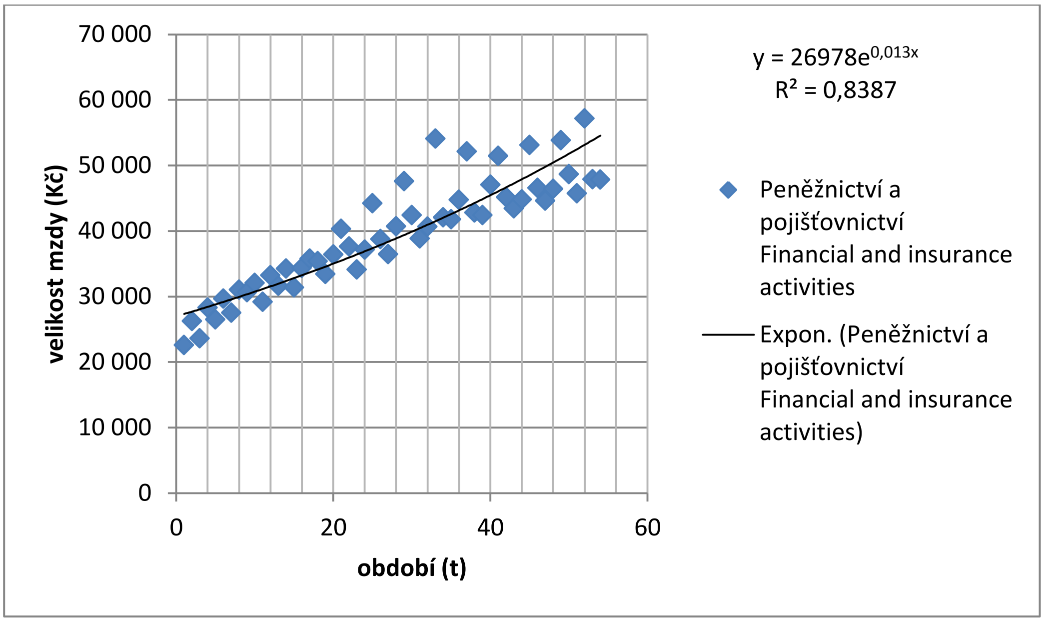
Zdroj: vlastní zpracování

Graf 2: mzdy (peněžnictví a pojišťovnictví) proložené kvadratickým trendem



Zdroj: vlastní zpracování

Graf 3: mzdy (peněžnictví a pojišťovnictví) proložené exponenciálním trendem



Zdroj: vlastní zpracování

Na zobrazených grafech jsem provedla subjektivní analýzu. Postup nebyl náročný. Pomocí dvou křivek, z nichž první jsem vedla dolními body zvratu a druhou horními body zvratu této řady, mi vznikl obrazec. V jeho středu, tedy mezi křivkou spojující horní body zvratu a mezi křivkou spojující dolní body zvratu, jsem následně vedla další křivku, vyjadřující již hledaný vývoj. Tato třetí křivka byla výsledkem subjektivní analýzy. Na základě jejího průběhu, ale i podle obecných údajů o průměrných mzdách, se mohlo zdát, že analyzovaná data se vyvíjejí lineárně. Tato myšlenka je ovšem pouze mým subjektivním názorem, a z důvodu nedostatečných zkušeností s touto analýzou je možné, že se mohu mýlit. Proto se pokusím vše ověřit statistickým testem.

Mým dalším úkolem proto bylo zjistit hodnoty koeficientů β_i. Tedy pomocí metody nejmenších čtverců najít takovou křivku, kde bude součet čtverců minimální. Pokud bych předpokládala nejjednodušší model s jednou závisle proměnnou a jednou nezávisle proměnnou použila bych lineární model:

$$Y = \beta_0 + \beta_1 x \tag{12}$$

kde:

Y … studovaný model;

β_0 … b_0 jako absolutní člen;

β_1 … b_1 jako lineární člen.

a postupovala bych ve výpočtu podle následujících vzorců:

absolutní člen

$$b_0 = \frac{\sum y_i - b_1 \sum x_i}{n} \tag{13}$$

lineární člen

$$b_1 = n \frac{\sum y_i x_i - \sum y_i \sum x_i}{n \sum x_i^2 - (\sum x_i)} \tag{14}$$

Avšak podle zobrazených grafů jsem vybrala jako nejvhodnější k modelaci závislosti a predikci Y na základě znalosti X model kvadratický. U tohoto modelu se data nejméně vzdalovala od křivky, která je prokládala, a nesnižovala tak příliš pravděpodobnost chybné predikce. Z toho vyplívá, že nemohu postupovat ve výpočtu na základě lineárního modelu, jak jsem předpokládala podle subjektivní analýzy. Vybrala jsem model kvadratický, v jehož případě by zápis vzorce a výpočet koeficientů byl početně náročnější. Proto jsem se rozhodla koeficienty později zjistit pomocí software Statistica a Excel.

Nyní jsem tedy pomocí software Statistica a Excel porovnala, zda mnou vybraný kvadratický model je opravdu nejvhodnější pro vysvětlení zkoumaných dat. Zásadním v této analýze pro mě byl koeficient determinace R^2, nebo-li míra vhodnosti modelu, jehož hodnota říká, z kolika procent dovedou nezávisle proměnné v modelu vysvětlit variabilitu té závisle proměnné. V následující tabulce č. 3 zobrazuji hodnoty tohoto koeficientu vypočítané pro jednotlivé regresní modely.

Tabulka 3: hodnoty koeficientu determinace pro výběr regresního modelu

model	**hodnota koeficientu R^2 (%)**	**hodnota koeficientu upraveného R^2 (%)**
Lineární	0,8306	0,8273
Kvadratický	0,8508	0,8449
Exponenciální	0,8387	-

Zdroj: vlastní zpracování

Podle hodnot v tabulce jsem dokázala, že kvadratický model bude opravdu moje data prokládat nejlépe a zvolím proto, jako nejvhodnější regresní model, model kvadratický. Ten bude variabilitu závislé proměnné Y vysvětlovat pomocí X z 85 %.

Z důvodu náročnějšího postupu při počítání koeficientů β_0 a β_1 zjistím jejich hodnotu pomocí statistických testů. Nyní přikládám výslednou rovnici, která nejlépe popisuje závislost mezi proměnnými:

$$y = 23\,222 - 5{,}50x^2 + 793{,}4x \quad (15)$$

Další krok by se měl týkat testu, který zjistí, zda je model významný jako celek. Můžeme ho nazvat také test významnosti koeficientu determinace. Potřebujeme zde prokázat, že alespoň jedna nezávisle proměnná (vysvětlující) vysvětluje závisle proměnnou, které také říkáme vysvětlovaná. Protože tento fakt nedokážeme jednoznačně potvrdit ani přes příznivou hodnotu koeficientu determinace budu testovat hypotézu:

$$H_0\colon \beta_1 = \beta_2 = \cdots = \beta_p = 0,$$

$$H_A\colon non\ H_0$$

V mém modelu tedy platí vzorec:

$$\frac{n-m+1}{m} * \frac{R^2}{1-R^2} \sim F\,(m, n-m-1). \quad (16)$$

Do rovnice jsem dosadila za jednotlivé neznámé následující údaje, kde:

$$n\ (počet\ pozorování) = 54,$$

$$m\ (délka\ periody) = 4,$$

$$R^2 = 0{,}8508,$$

$$F\ (4; 54 - 4 - 1).$$

Poté vyšla hodnota $F = 5{,}7024$, která je opravdu z intervalu (4; 49). Na základě tohoto výsledku mohu tedy zamítnout nulovou hypotézu ve prospěch alternativní a potvrdit, že některá z vysvětlujících proměnných skutečně vysvětluje vysvětlovanou proměnnou.

I v software Statictica jsem byla schopná si tuto skutečnost ověřit pomocí výpočtu jednoduché lineární regrese. Tento test jsem provedla na hladině významnosti 95 %. Následně jsem na základě vypočítané p-hodnoty byla schopná odvodit, zda jsou jednotlivé regresory významné. Pokud byla hodnota p-value menší než 0,05 % mohla jsem konstatovat, že regresor má v modelu nějaký význam.

Obrázek 5: jednoduchá lineární regrese, kvadratický model_p-hodnota

Výsledky regrese se závislou proměnnou : Prom1 (Tabulka1)
R= ,92239888 R2= ,85081969 Upravené R2= ,84496949
F(2,51)=145,43 p<0,0000 Směrod. chyba odhadu : 3336,4

N=54	b*	Sm.chyba z b*	b	Sm.chyba z b	t(51)	p-hodn.
Abs.člen			23221,86	1414,118	16,42145	0,000000
Prom2	1,473046	0,220237	793,40	118,623	6,68846	0,000000
V2**2	-0,579432	0,220237	-5,50	2,091	-2,63095	0,011233

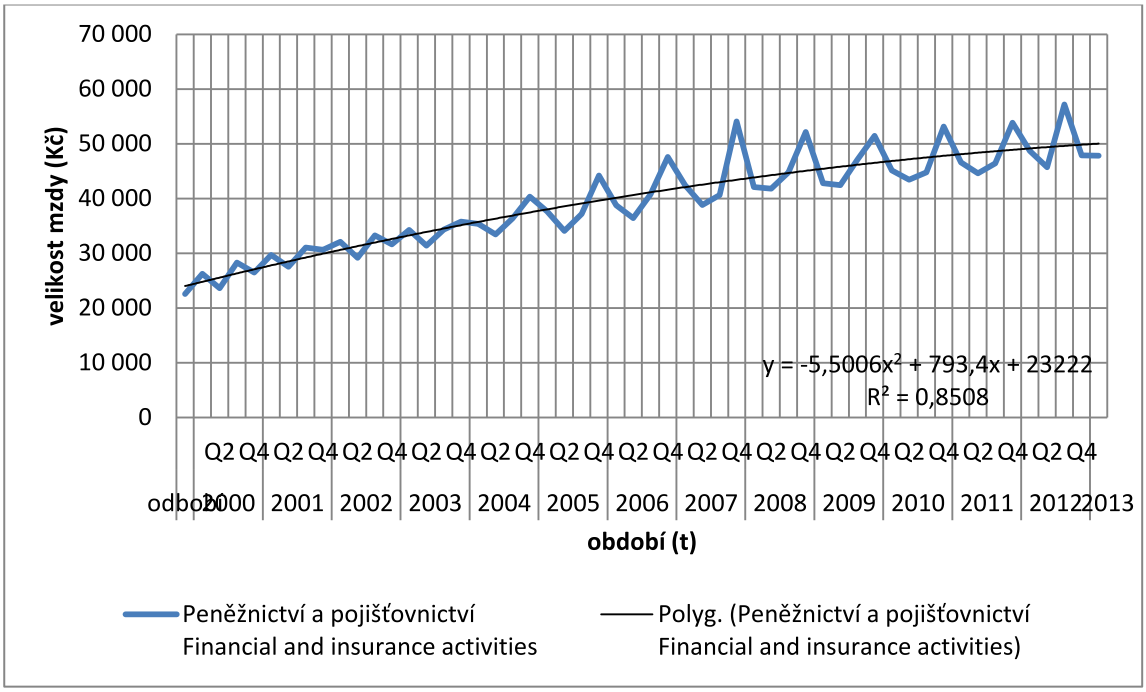
Zdroj: vlastní zpracování

Podle přiloženého obrázku vidíme, že p-hodnota je u všech členů, včetně členu absolutního, významná. Mohu tedy zamítnout nulovou hypotézu o nevýznamnosti regresorů, ve prospěch hypotézy alternativní a usoudit, že mzda se vyvíjela v průběhu let kvadratickým trendem. V další analýze budu tedy předpokládat kvadratický vývoj trendové přímky.

4.3. Sezónnost v časové řadě

V této kapitole se budu věnovat sezónnosti a způsobům, jak z řady tyto pravidelné výkyvy eliminovat. Pro přehlednost jsem přiložila graf zachycující výraznou sezónnost v časové řadě.

Graf 4: peněžnictví a pojišťovnictví_sezónnost



Zdroj: vlastní zpracování

V závislosti na průběhu grafu je ale viditelné, že přibližně do roku 2008 se mzda vyvíjela aditivně:

$$Y_t = T_t + S_t + \varepsilon_t \tag{17}$$

a poté, pravděpodobně po vyústění ekonomické krize, se řada chovala spíše multiplikativně:

$$Y_t = T_t * S_t * \varepsilon_t. \tag{18}$$

Tímto se ale zabývat prozatím nebudu. Budu předpokládat aditivní vývoj.

Nyní se zaměřím na otázku výskytu sezónnosti v této časové řadě. Pro ověření správnosti mé hypotézy provedu na analyzované řadě F-test sezónnosti tak, abych mohla statisticky prokázat, zda se v řadě opravdu vykazuje sezónnost.

Moje nulová hypotéza bude ve tvaru:

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0,$$

a alternativní ve tvaru:

$$H_A: non\ H_0$$

Poté budu pro výpočet testové statistiky F, součtu čtverců S_b a reziduálního součtu čtverců S_r postupovat následovně:

$$F = \frac{\frac{S_b}{p-1}}{\frac{S_k}{(p-1)(k-1)}}, \tag{19}$$

Kde:

$$S_b = k \sum_{j=1}^{p} (\bar{y}_j - \bar{y})^2 \tag{20}$$

$$S_r = \sum_{i=1}^{k} . \sum_{j=1}^{p} (y_{ij} - \bar{y})^2 - p \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2 - k \sum_{j=1}^{p} (\bar{y}_j - \bar{y})^2 \tag{21}$$

a platí

$$F \sim F_{p-1,(p-1)(k-1)} \tag{22}$$

kde:

F … testová statistika;

S_b … celkový součet čtverců;

S_r … reziduální součet čtverců;

p … délka periodického cyklu;

k … počet pozorování.

Opět pomocí software Statistica jsem provedla pro kontrolu test sezónnosti s pravděpodobnostním koeficientem α = 0,05 a soustředila jsem se na výslednou hodnotu p-value. Hodnota p-value v mém případě vyšla výrazně nižní než α.

Víme, že pokud nám tato hodnota vyjde menší než alfa, potom můžeme zamítnout nulovou hypotézu ve prospěch hypotézy alternativní a sezónnost bude tedy prokazatelná. Naopak, pokud nám hodnota p-value vychází větší než alfa, potom sezónnost v časové řadě prokazatelná není.

Obrázek 6: sezónnost_F-test

Efekt	Jednorozměrné testy významnosti pro peněžnictví (Tabulka1) Sigma-omezená parametrizace Dekompozice efektivní hypotézy SČ	Stupně volnosti	PČ	F	p
Abs. člen	43520360E+10	1	8,225201E+10	11628,31	00000000
čtvrtletí	06043870E+08	3	6,916209E+07	9,78	74077034
rok	92582650E+09	13	2,539587E+08	35,90	00000000
Chyba	93954700E+08	37	7,073428E+06		

Zdroj: vlastní zpracování

Na časové řadě průměrných mezd v odvětví peněžnictví a finančnictví můžeme podle výsledné hodnoty p-value (ve sloupečku označeném „p“) na obrázku č. 6 zamítnout nulovou hypotézu ve prospěch alternativní a prokázat tím přítomnost sezónnosti.

4.4. Vícenásobná regrese – model

Nyní budu pomocí vícenásobné regrese hledat nejvhodnější model vysvětlující data s největší přesností. Vícenásobná regrese je použita z důvodu čtvrtletně nasbíraných dat. Budu analyzovat data dvěma způsoby tak, že nejprve analyzuji řadu jako celek, a poté pouze její druhou polovinu. Pokusím se tak zamezit nepřesnostem ve výpovědi indexu determinace způsobeným historickými daty.

4.4.1. Model pro celou časovou řadu

Zkoumaná časová řada je podle mých statistik nejpravděpodobněji charakterizovaná kvadratickým vývojem trendu a pravidelnou čtvrtletní sezónností, kterou jsem si ověřila F-testem. Nyní, na základě hypotézy pro můj model budu hledat jednotlivé parametry trendu a sezónnosti a poté, pomocí těchto parametrů, zjistím sezónně očištěné hodnoty.

Budu předpokládat základní aditivní model ve tvaru:

$$Y_t = \alpha_0 + \alpha_1 * t + \alpha_2 * t^2 + \beta_1 u_1(t) + \beta_2 u_2(t) + \beta_3 u_3(t) + \beta_4 u_4(t) + \varepsilon_t, \tag{23}$$

kde:

Y_t … analyzovaný model;

α_0 … absolutní člen;

$\alpha_1 * t$ … odhadovaný parametr lineárního trendu;

$\alpha_2 * t^2$ … odhadovaný parametr kvadratického trendu;

β_t … odhadované parametry sezónnosti;

$u_1(t)$ … proměnné;

ε_t … náhodná složka (bílý šum).

Následně bude mým cílem sestavit takový model, ve kterém bude rozptyl reziduí minimální. Pokud jsem ke své analýze použila čtvrtletní data, počet sezón a tedy i proměnných se bude rovnat 4. Musím dávat pozor a první model sestavím pouze se třemi proměnnými tak, aby regresní složky v tomto modelu nebyly lineárně závislé. Hodnoty těchto potřebných proměnných získám pomocí software Statistica. Poté se teprve budu soustředit na sestavení modelu se čtyřmi sezónami. Ty získám tak, že nejprve vypočítám průměr všech parametrů sezónnosti, včetně čtvrté, prozatím nulové hodnoty, které jsem získala ze software Statistica. Ten následně postupně odečtu od jednotlivých parametrů sezónnosti. Součet všech čtyř nových parametrů sezónnosti je nyní roven nule. Nesmíme také opomenout upravit absolutní člen. Ten změníme přičtením průměrné hodnoty původních parametrů sezónnosti.

Jednotlivé hodnoty potřebných parametrů jsem získala pomocí vícenásobné regrese v programu Statistica. Zobrazuji je v následujících tabulkách:

Tabulka 4: parametry trendu (1Q2000 – 2Q2013)

Parametry trendu	
Lineární parametr	822,18
Kvadratický parametr	-6,02

Zdroj vlastní zpracování

Tabulka 5: parametry sezónnosti_3 proměnné (1Q2000 – 2Q2013)

Období	Parametry sezónnosti
1. Čtvrtletí	2816,85
2. Čtvrtletí	-684,03
3. Čtvrtletí	-3314,44
4. čtvrtletí	0
Absolutní člen	23194,15

Zdroj vlastní zpracování

Tyto hodnoty tvoří následující model se třemi proměnnými u_p:

$$Y_t = 23194 + 822t - 6{,}02t^2 + 2817u_1 - 684u_2 - 3314u_3 \tag{24}$$

Dalším krokem byla úprava těchto parametrů pomocí průměru původních hodnot parametrů, jak jsem uváděla dříve, a následné vytvoření nové rovnice se čtyřmi proměnnými u_p. Tabulka č. 6 zobrazuje nově upravené proměnné, které jsem poté dosadila do rovnice.

Tabulka 6: parametry sezónnosti_4 proměnné (1Q2000 – 2Q2013)

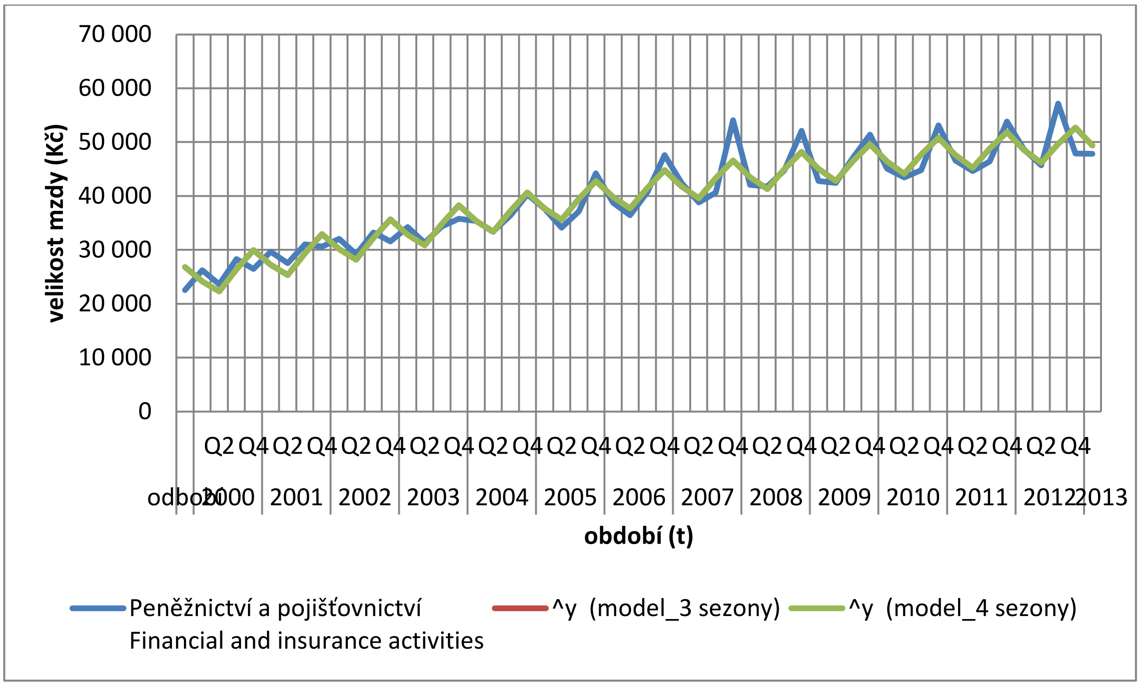
období	**Parametry sezónnosti**
1. čtvrtletí	3112,26
2. čtvrtletí	-388,62
3. čtvrtletí	-3019,04
4. čtvrtletí	295,40
Absolutní člen	22 898,75

Zdroj vlastní zpracování

Potom nový model vypadá takto a dokáže, podle výsledku vícenásobné regrese, vysvětlit variabilitu mezd z 92 %:

$$Y_t = 22899 + 822t - 6{,}02t^2 + 3112u_1 - 389u_2 - 3019u_3 + 295u_4 \tag{25}$$

Graf 5: původní data + regresní modely (1Q2000 – 2Q2013)



Zdroj: vlastní zpracování

V přiloženém grafu č. 5 můžeme vidět porovnání původních dat zobrazených modrou křivkou s vytvořenými modely, které jsou označeny červenou a zelenou barvou. Tyto dva modely jsou totožné, vykazují pravidelnou čtvrtletní sezónnost a pomocí nich budu schopná v další kapitole této práce predikovat budoucí hodnoty vývoje mezd v odvětví.

Dále tento graf také přehledně zobrazuje odhadované hodnoty modelu. V jeho průběhu si ovšem můžeme všimnout dvou výraznějších skoků, které se vypínají nad novou předpovědí. Tento viditelný zvrat byl pravděpodobně způsoben v prvním případě propukající ekonomickou krizí v zemi roku 2008 a druhý případ podle posledních informací zapříčinila daňová reforma, a to zrušením daňových stropů na konci roku 2012.

4.4.2. Daňová reforma – řešení problému s daty

Nová vyhláška podle expertů zvýšila průměrnou mzdu. To se na mých datech značně projevilo na přelomu let 2012 a 2013. V tomto místě časové řady průměrná mzda za čtvrté čtvrtletí měla tendenci již několik let být o pár tisíc nižší než průměrná mzda za první čtvrtletí roku. To se ovšem na základě vyhlášky výrazně změnilo, především v mém zkoumaném odvětví průměrné mzdy v peněžnictví a pojišťovnictví, když na konci roku 2012 mzda najednou výrazně vzrostla v porovnání se začátkem roku 2013.

Tato situace tak zapříčinila na datech náhlý zvrat a nyní způsobuje vyšší riziko nepřesnosti odhadu na základě mnou právě získaného modelu. Koeficient determinace v mých propočtech má sice podle své hodnoty vysokou vypovídací schopnost, ale to může být způsobeno hodnotami v první polovině řady. Ta se podle grafu od roku 2000 do konce roku 2007 vyvíjí víceméně v pravidelných intervalech se stejnými odchylkami. Druhá polovina už tak pravidelný vývoj nemá. Proto je možné, že v tomto případě je vysoká hodnota koeficientu determinace zapříčiněná převážně vývojem první části časové řady.

Tuto hypotézu si nyní otestuji. Vypočítám hodnotu koeficientu determinace z již získaného modelu, ale pro období od prvního čtvrtletí roku 2008 do druhého čtvrtletí roku 2013, na základě následujícího vzorce:

$$R^2 = \frac{ESS}{TSS}, \tag{26}$$

potom

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2, \tag{27}$$

$$ESS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2, \tag{28}$$

kde:

R^2 … koeficient determinace;

TSS … celkový (úplný) součet čtverců;

ESS … vysvětlovaný součet čtverců;

y_i … původní data;

$\bar{y}$ … průměrná hodnota;

$\hat{y}_i$ … předpovídaná hodnota.

Jelikož zrušení daňových stropů je pouze nárazová situace, pokusila jsem se nebrat tento fakt v potaz a svá data upravila tak, abych co nejlépe zakryla tento jednorázový výkyv, u kterého nepředpokládáme další vliv a opakování.

Do vzorce jsem tedy dosadila data s upravenými hodnotami a vliv daňové reformy jsem nebrala v potaz. Po dosazení jsem získala následující hodnoty, které zobrazuji v tabulce č. 7.

Tabulka 7: koeficient determinace, TSS, ESS

TSS	426 782 802,95
ESS	186 307 497,43
Koeficient determinace $\boldsymbol{R^2}$ (%)	0,437

Zdroj: vlastní zpracování

Ukázalo se, že mé tvrzení je pravdivé. Koeficient determinace pro druhou polovinu vypovídá o variabilitě dat pouze ze 44 %. Z toho vyplývá, že původní vysoká hodnota koeficientu (asi 92 %), která vypovídá o spolehlivosti předpovědí na základě tohoto modelu, je tvořena převážně z dat v první části analyzované řady.

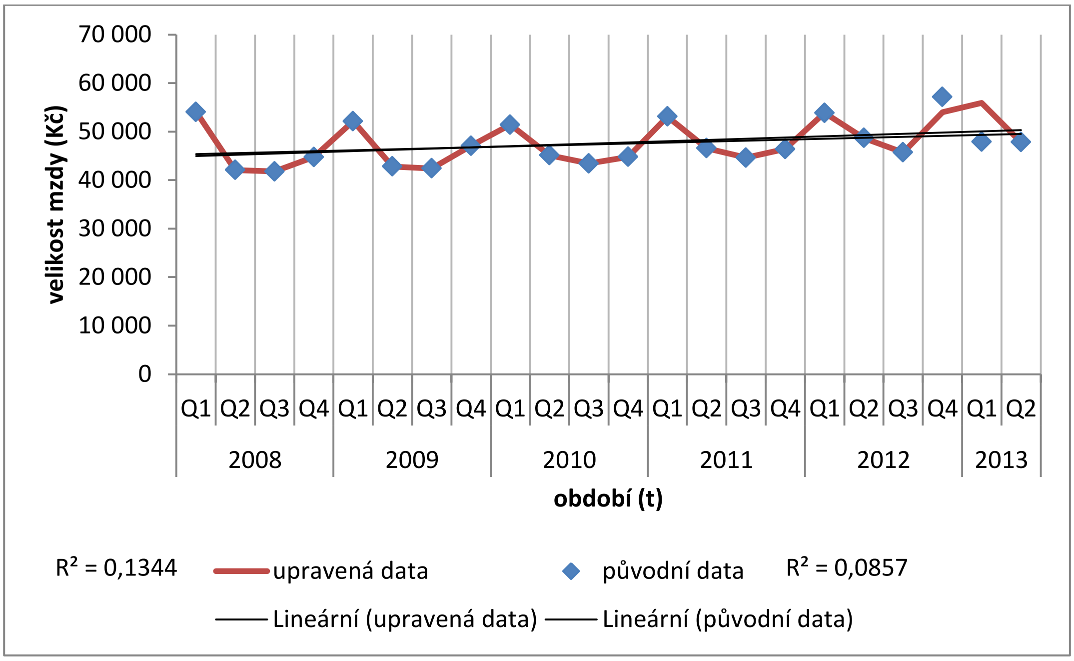
4.4.3. Model po rozdělení časové řady na dvě části

Na základě výsledků mé předchozí hypotézy jsem se tedy rozhodla zkoumanou řadu rozdělit na dvě části. Historická data časové řady pro mě v tuto chvíli, kdy mým cílem je predikovat budoucí stav, nejsou natolik důležitá. Z tohoto důvodu nyní vytvořím nový model pro aktuálnější data od roku 2008 do druhého čtvrtletí roku 2013.

Zvolím identický postup pro tvorbu tohoto modelu jako v předchozím případě, kde jsem tvořila model pro řadu jako celek. Tedy na základě grafu zvolím vhodné proložení trendovou přímkou tak, aby co možná nejlépe vysvětlovala zobrazovaná data.

To v tomto případě ale nebude tak jednoduché z důvodu vysoce kolísavé sezónnosti druhé části řady. Data se pravděpodobně prozatím nestihla po neočekávaném zvratu v ekonomické situaci státu vrátit do normálu. Pro zjednodušení a zpřesnění jsem se z této části řady snažila odstranit alespoň jeden výrazný výkyv na přelomu let 2012 a 2013 a model vytvořila pro upravenou řadu. Přiložený graf č. 6 zobrazuje vliv drobné korekce na data a chování řady. Původní data z ČSÚ jsou představována modrými body a vývoj řady upravené vyobrazuje červená křivka.

Graf 6: původní a upravený model (1Q2008 - 2Q2013)



Zdroj: vlastní zpracování

Grafické zobrazení i vypočítaná statistika R^2 nám ukazují, že drobná korekce hodnot výrazněji změnila vypovídací schopnost pro analýzu. Podle velikosti koeficientu determinace, jehož vypovídací hodnota pro původní data byla 8,6 %, jsem mohla po úpravě hodnot řady sestavit o něco přesnější a vhodnější regresní model vypovídající o řadě a jejím trendu z 13,4 %.

Vícenásobnou regresí v programu Statistica jsem získala potřebné parametry pro sestavení modelu s upravenými hodnotami dvou kvartálů. Model jsem proložila lineárním trendem.

Tabulka 8: parametry sezónnosti (1Q2008 - 2Q2013)

sezóna	Parametry sezónnosti	Upravené parametry sezónnosti
1. čtvrtletí	6292,08	6073,5
2. čtvrtletí	-1886,03	-2104,62
3. čtvrtletí	-3531,72	-3750,3
4. čtvrtletí	0	-218,58

Zdroj: vlastní zpracování

Tabulka 9: parametry trendu (1Q2008 - 2Q2013)

Parametr trendu	
Lineární trend	272,28

Zdroj: vlastní zpracování

V tabulce č. 8 uvádím v druhém sloupci parametry sezónnosti vypočítané statistickou analýzou. Po vypočítání průměru z těchto dat v druhém sloupci jsem mohla vyjádřit upravené parametry sezónnosti do sloupce třetího. Tedy odečtením celkové průměrné hodnoty z jednotlivých řádků druhého sloupce dojdeme k výsledku, který zobrazují data ve sloupci posledním.

Vytvářím tedy následující model představován křivkou, která je podle analýzy vícenásobné regrese v software Statistica schopná vysvětlovat zkoumaná data s přesností 88 %. Hodnota absolutního členu vychází 44 131 a průměrná hodnota sezónních parametrů je 218,58. Absolutní člen po úpravě má tedy hodnotu 44 350.

Nejprve model s původními daty:

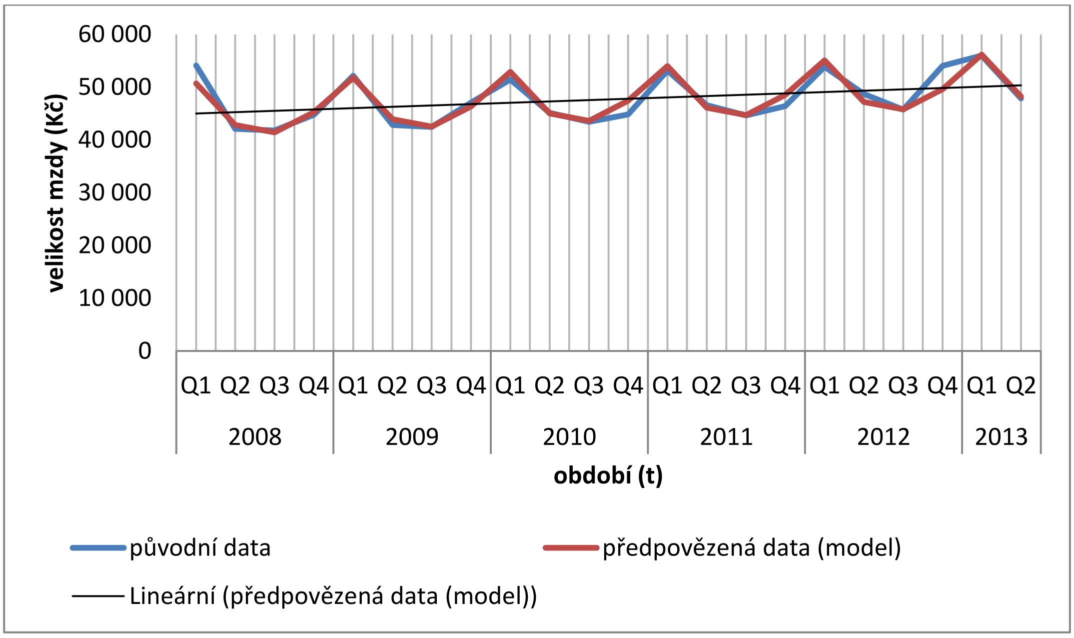
$$Y = 44131 + 272t + 6292u_1 - 1886u_2 - 3532u_3, \tag{29}$$

a po úpravě:

$$Y = 44350 + 272t + 6074u_1 - 2105u_2 - 3750u_3 - 219u_4 \tag{30}$$

Na základě tohoto modelu mohu nyní predikovat data. Koeficient determinace má dostatečně vysokou hodnotu pro další práci a analyzování. Není tedy potřeba hledat další model. Nejprve ale zobrazuji porovnání původních dat z ČSÚ s daty vytvořenými na základě tohoto modelu.

Graf 7: původní data + regresní modely (1Q2008 – 2Q2013)



Zdroj: vlastní zpracování

4.5. Predikce budoucích hodnot

Nyní se zaměřím na bodovou předpověď a odhad parametru β pro velikost mzdy v následujících obdobích. Mým bodovým odhadem bude následně hodnota, kterou vypočítám ze získaného modelu vícenásobné regrese po dosazení určitých regresorů. Omezím se pouze na předpověď šesti období z důvodu možné nepřesnosti, která se s každým dalším obdobím může zvyšovat, a výsledky jsou potom nepřesné. Tyto nepřesnosti, jak už jsem uváděla dříve ve své bakalářské práci, jsou způsobené mou prací s náhodnou veličinou.

Pro analyzování dat v této kapitole mé bakalářské práce jsem si již připravila a vypočítala několik potřebných údajů v předchozích kapitolách. Zkoumaná data ve své práci jsem rozdělila na dvě části. Nejprve jsem provedla analýzu dat v rozpětí 1Q2000 –

2Q2013 a pokoušela se získat nejvhodnější model, který by celou tuto řadu vysvětloval co nejlépe, abych na jeho základě mohla predikovat budoucí hodnoty. Jako druhou variantu jsem zvolila rozdělení řady na dvě poloviny a tvořila jsem model pouze pro druhou část dat, tj. rozpětí od 1Q 2008 – 2Q 2013.

Stěžejním bodem v těchto analýzách pro mě byla hodnota koeficientu determinace. Jak už jsem uvedla výše ve své práci, platí zde tvrzení: Čím vyšší hodnota koeficientu R^2, tím lepší model, který je schopen přesněji popsat a vysvětlit chování časové řady a následně předpovídat její budoucí vývoj.

Tabulka 9: koeficienty determinace - predikce

Koeficient determinace R^2	**Hodnota (%)**
R^2 – 1Q 2000 – 2Q 2013	43,65
R^2 – 1Q 2008 – 2Q 2013	87,8

Zdroj: vlastní zpracování

Na základě výsledných hodnot koeficientu determinace z obou modelů, zobrazených v tabulce č. 9, který jsem si vypočítala speciálně pro druhou část řady v období 1Q 2008 – 2Q 2013, mé rozhodnutí bylo, že predikci budoucích hodnot provedu na základě druhého modelu s hodnotou koeficientu R^2 = 87,8 %.

Budu tedy predikovat budoucí hodnoty časové řady, která zobrazuje vývoj hodnot průměrných mezd v odvětví peněžnictví a pojišťovnictví od roku 2000 do druhého čtvrtletí roku 2013. Omezím se na predikci pěti až šesti období. Abych mohla tvrdit, že tyto budoucí hodnoty jsou mým co nejpřesnějším odhadem.

Postupovala jsem dosazováním potřebných neznámých do vytvořeného modelu, na základě kterého mi poté vycházela data vyjadřující vývoj mzdy v příštích obdobích. V následující tabulce zobrazuji jednotlivé výsledné hodnoty mezd pro každé čtvrtletí v šesti dalších obdobích.

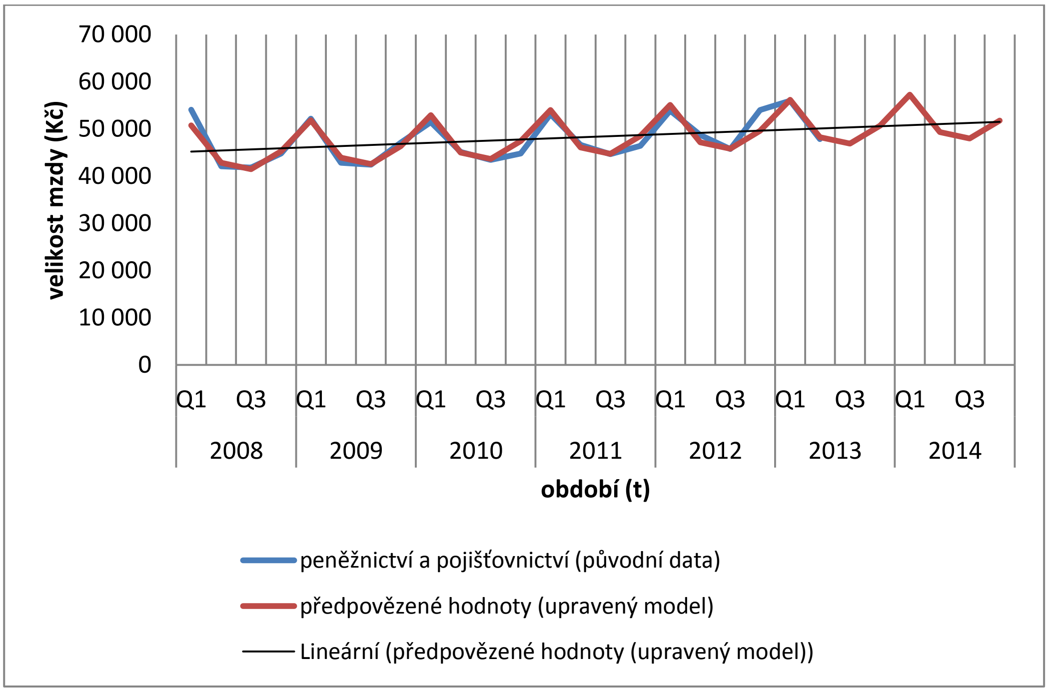
Tabulka 10: predikované hodnoty

Období	Velikost mzdy (Kč)
3Q 2013	46 862
4Q 2013	50 666
1Q 2014	57 230
2Q 2014	49 324
3Q 2014	47 951
4Q 2014	51 755

Zdroj: vlastní zpracování

A přikládám graf č. 8 zobrazující tyto budoucí hodnoty v závislosti a návaznosti na původních analyzovaných datech ze statistického úřadu. Modrá křivka zde zobrazuje průběh původních analyzovaných dat, které jsem získala z Českého statistického úřadu a červeně je zobrazen záznam modelu vytvořeného pomocí regresní analýzy, včetně predikovaných hodnot.

Graf 8: predikované hodnoty



Zdroj: vlastní zpracování

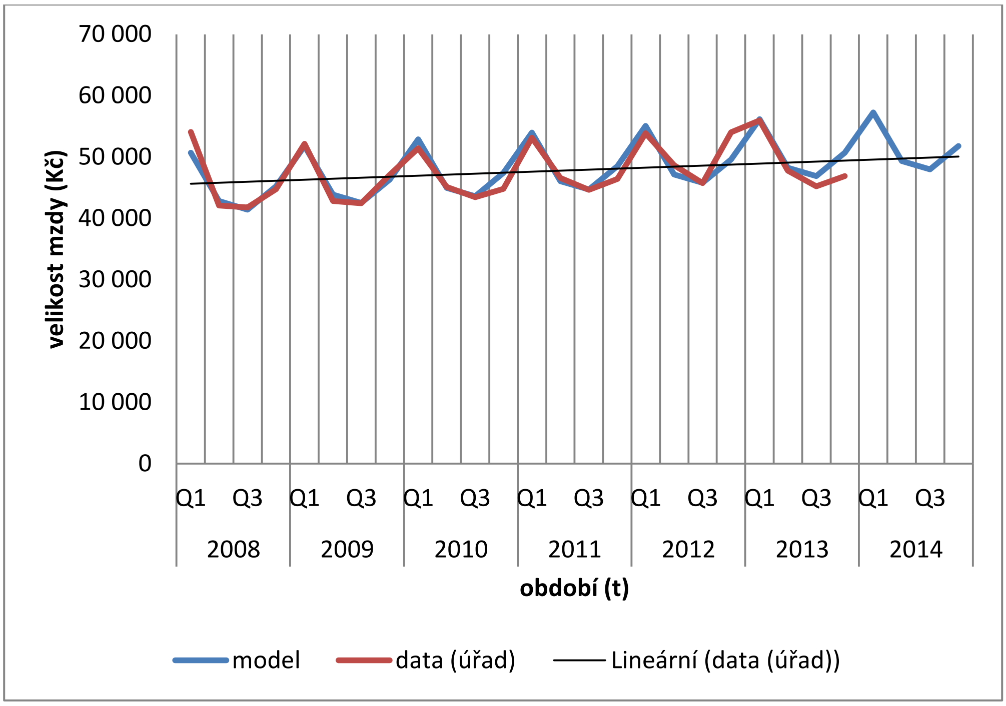
Shrnutí a závěrečný komentář

Podle velikosti vypočítaných dat mohu nyní soudit, že mzda se bude nadále vyvíjet stejným předpokládaným trendem s lineárním průběhem. Nejvyšší hodnoty bude částka mzdy dosahovat vždy v prvním čtvrtletí roku. Poté bude následující dvě čtvrtletí mírně klesat a ve čtvrtém čtvrtletí se částka začne pomalu stupňovat svou velikostí až k nejvyššímu bodu.

Jelikož čas postoupil a od doby, kdy jsem časovou řadu začala analyzovat, uběhl už nějaký čas, dohledala jsem na Českém statistickém úřadě aktualizaci mnou analyzovaných dat, abych následně mohla pro zajímavost porovnat alespoň část mé predikce se skutečnými hodnotami.

Proto jsem pro porovnání mých predikovaných hodnot s predikovanými hodnotami ČSÚ přiložila graf, který zobrazuje nejaktuálnější naměřené hodnoty statistického úřadu. Ten, k podivu ještě dnes, nazývá hodnoty roku 2012 a 2013 jako předběžné hodnoty a v grafu jsou představovány červenou křivkou. Můj analyzovaný model i s dalšími predikovanými hodnotami je zobrazen křivkou modrou.

Graf 9: predikce + aktualizovaná data z ČSÚ



Zdroj: vlastní zpracování

Statistický úřad z časového důvodu ještě neuvádí přesná stoprocentní data, ani data pro rok 2014. V grafu č. 9 ale můžeme vidět, že mzda se podle mého modelu i podle modelu ČSÚ vyvíjí lineárním trendem, který má tendenci mírného růstu. Nepatrný rozdíl nastává mezi 3Q 2013 a 4Q 2013, kde mzda podle statistiky ČSÚ nerostla na přelomu čtvrtletí tak prudce. Stále ovšem musíme brát v potaz, že naše dedukce přesně nevypovídají o skutečnosti, jelikož hodnoty, které jsem vypočítala, jsou prozatím pouhým odhadem.

5. Závěr

V první části mé práce jsem se věnovala časovým řadám, jejich popisu, významu a dělení. Popsala jsem jejich využití a způsoby, jakými lze takováto data zobrazená v řadě analyzovat. Stručně jsem objasnila pojmy spektrální analýza nebo Boxova-Jenkinsova metodologie. Zaměřila jsem se na míry vhodnosti modelu a způsob predikce budoucích hodnot. V další části jsem také vysvětlila způsoby dekompozice časových řad a charakterizovala jsem použitá data.

Praktickou část mé bakalářské práce jsem věnovala analyzování průměrných mezd v odvětví peněžnictví a pojišťovnictví pomocí regresního přístupu. Analyzovaná data mi poskytl Český statistický úřad, který eviduje a následně zveřejňuje jednotná a základní statistická data určená především pro veřejnost.

Začala jsem základními pracemi s daty. Zjistila jsem hodnoty nejnižších a naopak nejvyšších mezd ve zkoumaném období, následně jsem hledala největší a nejmenší mezičtvrtletní nárůst mezd apod. V této poslední analýze základních prací s daty jsem objevila některé nejasnosti, které ovšem byly následně vysvětleny náhlou právní úpravou v daňové problematice. Dalším úkolem bylo najít nejvhodnější regresní model popisující data co nejpřesněji. Na základě zobrazeného korelačního pole dat, které jsem proložila křivkou, se poté jako nejvhodnější jevil polynomický trend druhého řádu. Předpokládala jsem tedy kvadratický trend, čtvrtletní sezónnost, a protože všechny členy vyšly na základě statistických testů významné, mohla jsem na základě tohoto odhadu získat potřebné parametry trendu a sezónnosti. Pomocí nich jsem následně vytvořila model. Ten vypovídal o variabilitě proměnných z 92 %. Z důvodu pochybností o nerovnoměrném vývoji grafu a možnosti malé vypovídací schopnosti řady v její druhé části, jsem ale usoudila, že bude vhodné analyzovanou řadu rozdělit na dvě poloviny. Vypočítala jsem tedy koeficient determinace pouze pro druhou část řady s výsledkem pouhých 44 %. Následně jsem vytvořila samostatný nový model pouze pro druhou polovinu dat, kde jsem předpokládala, na rozdíl od předcházejícího modelu trend lineární. Zde už jeho vypovídací schopnost byla v odpovídající velikosti asi 88 %. Nesmím ovšem zapomenout na drobnou korekci dat, kterou jsem provedla. V analyzované časové řadě se totiž vyskytly nepochopitelné výkyvy, objevené již při základní práci s daty, které po dohledání podrobnějších informací o vývoji mzdy, byly způsobeny daňovou reformou na přelomu let 2012 a 2013 a vyvolaly v mé analýze znatelné nepřesnosti. Z důvodu nepředpokládaného opakování této situace jsem se však rozhodla nebrat tento fakt v potaz a nepřesnosti upravila tak, jako by k nim ve skutečnosti nedošlo.

V poslední části práce pro mě tento fakt představoval důležité rozhodnutí ve výběru modelu pro predikci. K další analýze jsem si samozřejmě vybrala model s vyšší vypovídací schopností, která byla v takové velikosti, že nebylo již dále nutné hledat vhodnější model. Pomocí modelu vícenásobné regrese, ovšem pouze pro druhou část řady, jsem určila predikce jednotlivých průměrných čtvrtletních mezd.

Na základě grafů jsem následně porovnala predikované hodnoty s původními hodnotami a hodnotami aktualizovanými ČSÚ, které jsem si dohledala s odstupem času. Zkontrolovala jsem tak své predikce s předběžnými hodnotami zveřejněnými statistickým úřadem. Ty se ve srovnání od sebe výrazně nevzdalovaly. Mohla jsem tedy konstatovat, že mnou vybraný regresní model byl zvolen správně a na jeho základě jsem předpověděla odpovídající výsledné hodnoty.

Summary

The goal of this bachelors work was to analyse the selected time series by using the methods of evaluating data. These time series describe the average wage in the finance and insurance industry. With the help of regression analysis to determine and adjust the trend, identify and set seasonality (you can also say make trend and seasonal decomposition). Another task was to make an appropriate model, which will describe the analysed data the most accurately. Than in the end to verify all of these calculations and facts by special statistical tests, to be able to make a prediction about the future development of the average wages.

However at first, the readers of this work were familiarized with the issues of the time series. I described their use and the ways in which can the timely collected data be analysed, and I gave some examples of the time series. I also mentioned the characteristic of the analysed data, which were found on Czech statistical office.

In the end of this work I summarized all the results and answers based on the discussion in this bachelors work. By executing all of my tasks I can say that I fulfilled my bachelor work.

Key words:

Time series, methods of evaluating, regression analysis, decomposition of time series, seasonality, trend, prediction, calculations.

Přehled použité literatury

Marek, L. (2010). Analýza vývoje mezd v ČR v letech 1995–2008. Politická ekonomie, 58(2), 186-206.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). Introduction to time series analysis and forecasting (Vol. 526). Hoboken, New Jersey: John Wiley & Sons.

Cipra, T. (2008). Finanční ekonometrie. Praha: Ekopress.

Cipra, T. (1986). Analýza časových řad s aplikacemi v ekonomii. Praha: SNTL/ALFA.

Hindls, R., Hronová, S., & Novák, I. (2000). Metody statistické analýzy pro ekonomy. 2. přeprac. vyd. Praha: Management press.

Klufová, R., Rost, M., & Klicnarová, J. (2012). Modelování regionálních procesů. Praha: Alfa Nakladatelství.

Průměrná mzda a evidenční počet zaměstnanců (2013) Metodika in: Český statistický úřad [online]. Dostupné z: (http://www.czso.cz/csu/redakce.nsf/i/pmz_m).

Pavelka, F., & Klímek, P. (2000). Aplikovaná statistika. Vysoké učení technické, Fakulta managementu a ekonomiky ve Zlíně.

Arlt, J., & Arltová, M. (2007). Ekonomické časové řady: Vlastnosti, metody modelování, příklady a aplikace. Praha: Grada.

Arlt, J., Arltová, M., & Rublíková, E. (2002). Analýza ekonomických časových řad s příklady. Praha: Oeconomica.

Zapletal, J. (2000). Úvod do analýzy ekonomických časových řad. Brno: PC-DIR Real.

Čermáková, A., Střeleček, F. (1995). Statistika I. České Budějovice: Jihočeská univerzita.

Poznámky z předmětu SMAC (viz. http://moodle.ef.jcu.cz/)

Seznam obrázků

Seznam grafů

Seznam tabulek

Seznam příloh