

Úvod do analýzy časových řad

Obsah

1 Úvod.....	2
2 Teoretické základy pro analýzu časových řad.....	2
2.1 Základní pojmy.....	2
2.1.1 Druhy časových řad.....	2
2.1.2 Grafická analýza.....	2
2.1.3 Popisné charakteristiky.....	4
2.2 Základní úpravy časových řad.....	6
2.2.1 Doplnění chybějících hodnot.....	7
2.2.2 Transformace měřítka a kombinace časových řad.....	7
2.2.3 Časový posun.....	7
2.2.4 Sezónní diference.....	7
2.2.5 Kumulativní součet.....	8
2.2.6 Vyhlazování časových řad.....	8
2.3 Problémy časových řad.....	9
2.4 Metody analýzy časových řad.....	10
3 Příklad č. 1: Analýza míry nezaměstnanosti v okrese Karviná.....	20
3.1 Grafická a statistická deskripce.....	20
3.2 Očistění časové řady od sezónních vlivů.....	22
3.3 Tvorba modelu – Exponenciální vyrovnání druhého stupně.....	23
3.4 Výsledná predikce.....	25
4 Příklad č. 2: Postup analýzy ukazatelů na úrovni obce.....	27
4.1 Deskripce statistická a grafická.....	27
4.2 Analýza vztahů.....	30
5 Závěr.....	30
Literatura.....	30

1 Úvod

Cílem analýzy časových řad je většinou konstrukce vhodného modelu. Sestrojení „dobrého“ modelu nám zpravidla umožní porozumět mechanismu, na jehož základě vznikají hodnoty časové řady, a pochopit podmínky a vazby, které působí na vznik těchto hodnot. Na základě změn těchto podmínek či vazeb lze simulovat jejich vliv působící změny ve vývoji časové řady.

Dalším cílem je využití těchto získaných poznatků při předpovědi budoucího chování. Používané postupy jsou založeny na principu, že "historie se opakuje". Tento předpoklad bývá v praxi splněn s různou přesností, a proto je vhodné u vyhlazování a předpovědi v časových řada uvádět i spolehlivost získaných výsledků a hodnotit úspěšnost predikce.

2 Teoretické základy pro analýzu časových řad

2.1 Základní pojmy

Časovou řadou rozumíme posloupnost hodnot ukazatelů, měřených v určitých časových intervalech. Tyto intervaly jsou zpravidla rovnoměrné (ekvidistantní), a proto je můžeme zapsat následujícím způsobem:

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

kde y značí analyzovaný ukazatel, t je časová proměnná s celkovým počtem pozorování n .

2.1.1 Druhy časových řad

Časové řady členíme *podle charakteru ukazatele*:

- *okamžikové* - hodnota ukazatele k určitému okamžiku t (např. počet evidovaných uchazečů),
- *intervalové* - velikost sledovaného ukazatele závisí na délce intervalu, za který je sledován (např. měsíční náklady na rekvalifikace).

Podle *druhu ukazatelů* rozlišujeme časové řady obsahující:

- *absolutní* ukazatele (očistěné),
- *odvozené* ukazatele (součtové, poměrové).

2.1.2 Grafická analýza

Analýza časových řad se v současnosti provádí výhradně na počítačích pomocí vhodného softwaru. Velká většina statistických a ekonometrických softwarů má algoritmy těchto analýz zabudované ve svých standardních nabídkách. Bohužel program EXCEL mezi ně nepatří, proto se budeme muset věnovat relativně jednoduchým algoritmům, které lze vysvětlit. Pro pokročilejší analýzy časových řad doporučujeme statistické softwary: SPSS, STATISTICA, S+.

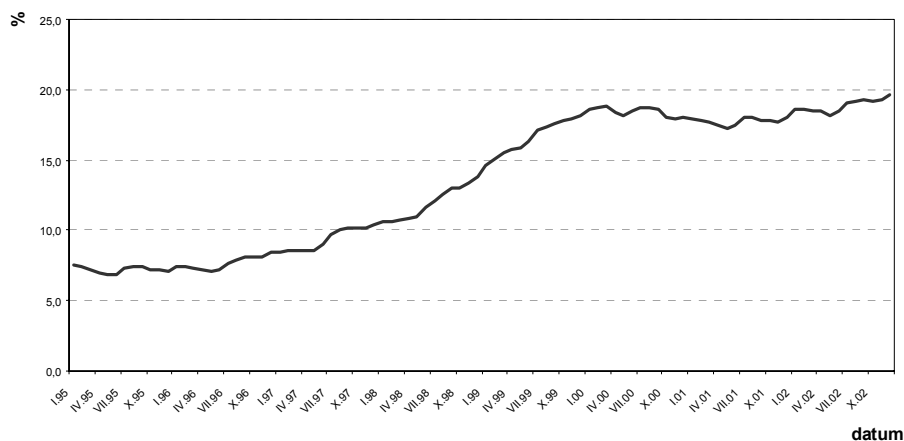
V programu EXCEL je nejvhodnější datovou strukturou pro časové řady standardní **datová matice** ve které je první řádek tvořen krátkým názvem proměnné a potom následují naměřené hodnoty. Jeden řádek datové matice obsahuje pozorování v jednom časovém okamžiku. Hodnoty jsou seřazeny podle času, vzestupně. Ukázka datové matice v EXCELu uvádí tabulka 1, která zahrnuje vývoj měsíční míry nezaměstnanosti v Karvině (%) – u_{KI} za období leden 1995 – březen 1996.

Tab. 1: Datová matice vývoje měsíční míry nezaměstnanosti v Karviné (u_{KI} v %)

Datum	t	Rok	Měsíc	u_{KI}
I.95	1	1995	1	7,53
II.95	2	1995	2	7,38
III.95	3	1995	3	7,18
IV.95	4	1995	4	7,00
V.95	5	1995	5	6,84
VI.95	6	1995	6	6,91
VII.95	7	1995	7	7,30
VIII.95	8	1995	8	7,37
IX.95	9	1995	9	7,42
X.95	10	1995	10	7,18
XI.95	11	1995	11	7,19
XII.95	12	1995	12	7,10
I.96	13	1996	1	7,40
II.96	14	1996	2	7,37
III.96	15	1996	3	7,29

Kromě proměnné t výše definované se obvykle používají další **časové proměnné** dle typu časových řad. Pokud pracujeme s ročními údaji je vhodné zavést další proměnou rok . U čtvrtletních dat kromě proměnné r i proměnou q , jenž nabývá hodnot 1 až 4 podle čtvrtletí. A analogicky postupujeme i u měsíčních údajů. Vedle těchto numerických proměnných se používá v programu EXCEL i proměnná ve formátu datum např. ve tvaru $I.99$ pro grafické znázornění časových řad.

Pro **zobrazení časových řad** a jejich prvotní analýzu slouží *spojnicové grafy*. Vodorovná osa u těchto grafů zaznamenává časovou proměnnou a na svislé ose se zobrazují hodnoty ukazatele časové řady y_t . Příkladem spojnicového grafu vývoje míry nezaměstnanosti v okrese Karviná v letech 1995 až 2002 je obr. 1.



Obr. 1: Vývoj míry nezaměstnanosti v okrese Karviná

Spojnicový graf může zahrnovat i více časových řad, avšak měřítko na svislé ose je stejné.

Dalším důležitým grafem v EXCELu je *graf XY bodový*, který sleduje vývoj časové řady y_t na vývoji hodnot časové řady x_t tzn., že znázorní bod se souřadnicemi $[x_t, y_t]$ pro každý časový okamžik t . Tento typ grafu je vhodný u regresní analýzy.

2.1.3 Popisné charakteristiky

Charakteristiky polohy (průměry)

Při práci s časovými řadami je někdy důležité zjistit jejich průměrné hodnoty:

- *prostý aritmetický průměr* $\bar{y} = \frac{\sum_{t=1}^n y_t}{n}$;
- *vážený aritmetický průměr* $\bar{y} = \frac{\sum_{t=1}^n v_t y_t}{\sum_{t=1}^n v_t}$, kde v_t je váha ukazatele y_t v čase t ;
- *vážený chronologický průměr* $\bar{y}_{ch} = \frac{\frac{y_1 + y_2}{2} d_2 + \frac{y_2 + y_3}{2} d_3 + \dots + \frac{y_{n-1} + y_n}{2} d_n}{d_2 + d_2 + \dots + d_n}$, kde d_t je délka jednotlivých časových intervalů.

Charakteristiky variability

Nejdůležitější míry variability ve statistice patří rozptyl a směrodatná odchylka:

- *rozptyl* je aritmetickým průměrem kvadrátů odchylek od aritmetického průměru:

$$s_y^2 = \frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y})^2 ;$$

- *směrodatná odchylka* je odmocninou z rozptylu $s_y = \sqrt{s_y^2} = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y})^2}$.

Míry dynamiky

Jednoduché míry dynamiky časových řad umožňují charakterizovat jejich základní rysy chování. Mezi základní míry dynamiky časové řady y_t patří:

- *absolutní přírůstek (první diference)* $\Delta y_t = y_t - y_{t-1}$ a *průměrný absolutní přírůstek*

$$\bar{\Delta} = \frac{\sum_{t=2}^n \Delta y_t}{n-1} = \frac{y_n - y_1}{n-1} ;$$

- *koeficient (tempo) růstu* $k_t = \frac{y_t}{y_{t-1}}$, kde $t = 2, \dots, n$, a *průměrný koeficient růstu* $\bar{k} = \sqrt[n]{k_2 \cdot k_3 \cdots k_n} = \sqrt[n]{\frac{y_n}{y_1}}$;
- *meziroční koeficient růstu* např. v případě čtvrtletní časové $k_{(4),t} = \frac{y_t}{y_{t-4}}$, kde $t = 5, 6, \dots, n$;
- *relativní přírůstek* $\delta_t = \frac{\Delta y_t}{y_{t-1}} = \frac{y_t - y_{t-1}}{y_{t-1}} = \frac{y_t}{y_{t-1}} - 1$ a průměrný relativní přírůstek $\bar{\delta} = \bar{k} - 1$.

Korelace

Korelace vyjadřuje relativní míru závislosti ve vzájemném vývoji dvou časových řad např. y_t a x_t

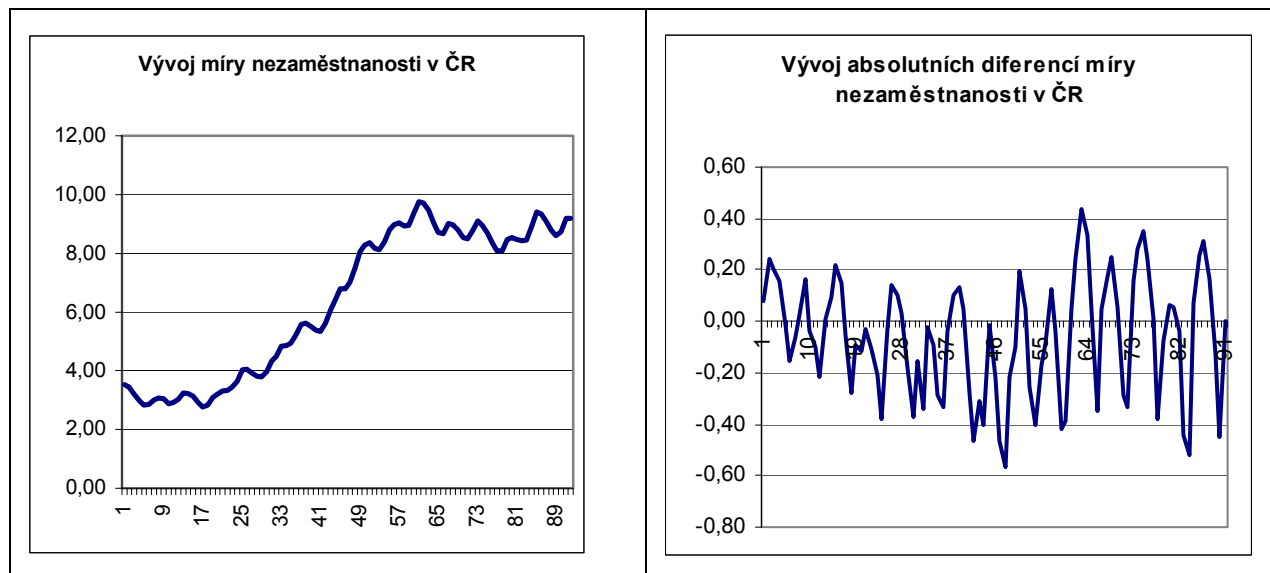
a je dána vztahem $s_{xy} = \frac{\sum_{t=1}^n (x_t - \bar{x}) \cdot (y_t - \bar{y})}{s_x \cdot s_y} \in \langle -1; 1 \rangle$. Hodnoty korelace blíží se ke

hraniční hodnotě -1 vyjadřují, že obě sledované časové řady mají zcela opačný směr v jejich časovém vývoji. Hodnoty s_{xy} blíží se k 1 prozrazují, že časové řady x a y se vyvíjí téměř shodně s hlediska stejných směrů pohybu a vykazují stejnou relativní míru ve vzájemném vývoji.

Stacionární a nestacionární časová řada

Chování časové řady může ze *statistického hlediska* buď podléhat změnám v průměru či variabilitě (řada *nestacionární*), nebo být stále stejná (řada *stacionární*). Zhruba řečeno to znamená, že u stacionární řady nejsme schopni na základě zjištěných statistických parametrů, jako jsou aritmetický průměr hodnot nebo jejich rozptyl, schopni odlišit jeden úsek řady od druhého. Nestacionární řada naopak vykazuje změny v chování: například aritmetický průměr hodnot ze začátku řady je signifikantně jiný než průměr členů na konci (o takové řadě říkáme, že vykazuje *trend*).

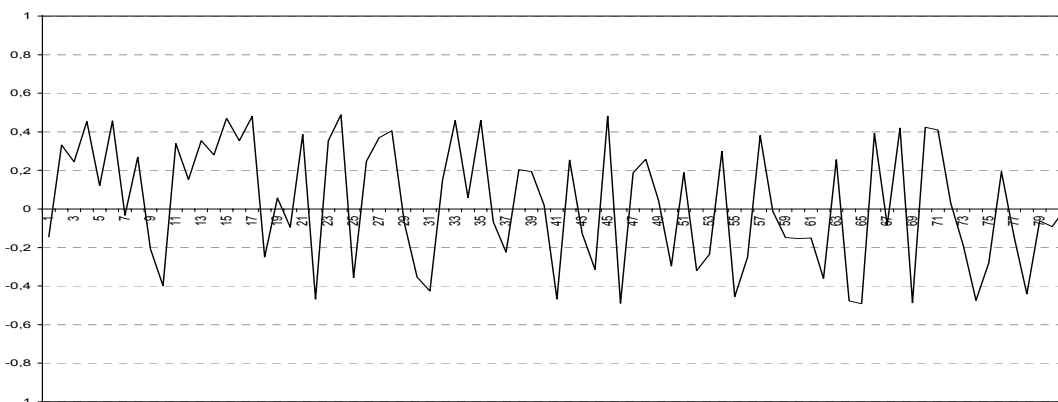
Stacionární chování je podstatným předpokladem některých typů analýz. Je pak třeba stacionaritu testovat a řadu případně vhodným způsobem transformovat s cílem odstranění nestacionarity.



Obr. 2: Vývoj měsíční míry nezaměstnanosti v ČR od roku 1995 do poloviny roku 2002

V grafu je vyobrazen průběh typické nestacionární časové řady, vykazující rostoucí trend, sezónní vlivy v průběhu každého roku a s časem rostoucí rozptyl (sezónní odchylky od průměru se stále zvětšují). Taková řada nevykazuje žádnou časovou změnu parametrů, protože její obecný člen nezávisí ani na čase, ani na předchozích členech řady.

V literatuře se i -tý člen časové řady s charakterem nezávislých realizací normálně rozložené náhodné veličiny se střední hodnotou $\mu = 0$ a konstantním rozptylem označuje jako *bílý šum*. Taková řada je svým způsobem „nejnáhodnější“ ze všech „rozumných“ časových řad, protože o jejím příštím členu v podstatě nevíme na základě předchozího průběhu víc, než že půjde o „nějaké číslo kolem nuly“. Název bílý šum vznikl z toho, že tato časová řada obsahuje rovnoměrný podíl frekvenčních složek všech vlnových délek podobně jako bílé světlo obsahuje složky všech barev spektra.



Obr. 3: Bílý šum

2.2 Základní úpravy časových řad

V další části jsou shrnuty nejčastější transformace či úpravy výchozí časové řady. Mnohé softwarové produkty zahrnují moduly pro tyto automatické výpočty.

2.2.1 Doplnění chybějících hodnot

V časové řadě může některé pozorování chybět a bývá někdy nutné je před zahájením dalších výpočtů doplnit. Doplněné údaje samozřejmě nejsou plnohodnotné a jejich přítomnost snižuje věrohodnost analýzy. Podle účelu transformace lze postupovat některým z následujících přístupů:

- Nahradit chybějící hodnoty *nulami*. Tento způsob lze doporučit tehdy, nevíme-li o řadě nic anebo jen to, že její průměrný člen by měl být nulový (tak tomu bývá např. u naměřených odchylek od nějaké očekávané hodnoty řízeného procesu).
- Nahradit chybějící hodnoty nějakou centrální charakteristikou souboru naměřených hodnot, konkrétně jeho *aritmetickým průměrem* nebo *mediánem*. Lze přitom brát centrální charakteristiku buď celého souboru, nebo pouze okolních bodů.
- Nahradit chybějící hodnotu *lineární interpolací* mezi sousedními body. Hodí se pro řady, které vykazují výraznou setrvačnost.
- Nahradit chybějící hodnoty *trendem* v celém souboru, získaném regresí vhodné křivky.
- Nahradit chybějící hodnoty odhadem založeným na známém či odhadnutém *modelu chování* procesu.

2.2.2 Transformace měřítka a kombinace časových řad

Nelineární transformace měřítka časové řady se používá především pro potlačení či zmírnění nestacionarity řady v případě, kdy např. s rostoucími hodnotami řady roste i rozptyl členů. Pak může logaritmování nebo odmocnění tento problém potlačit. Po provedení analýzy se k původnímu měřítku vrátíme zpětnou transformací: v případě logaritmování je to transformace exponenciální funkcí, v případě odmocnění transformace umocněním.

Někdy bývá vhodné zkombinovat několik časových řad např. jejich sečtením nebo vydělením jedné řady druhou (vypočítáním poměru).

2.2.3 Časový posun

Časový posun znamená vytvoření časové řady opožděné resp. předbíhající časovou řadu, ale jinak s ní totožnou. Představuje to vlastně posunutí časové řady „dopředu“ případně „dozadu“ oproti původní časové řadě. Nově vytvořené proměnné mají ovšem na začátku, resp. na konci tolik chybějících hodnot, o kolik kroků se posun prováděl.

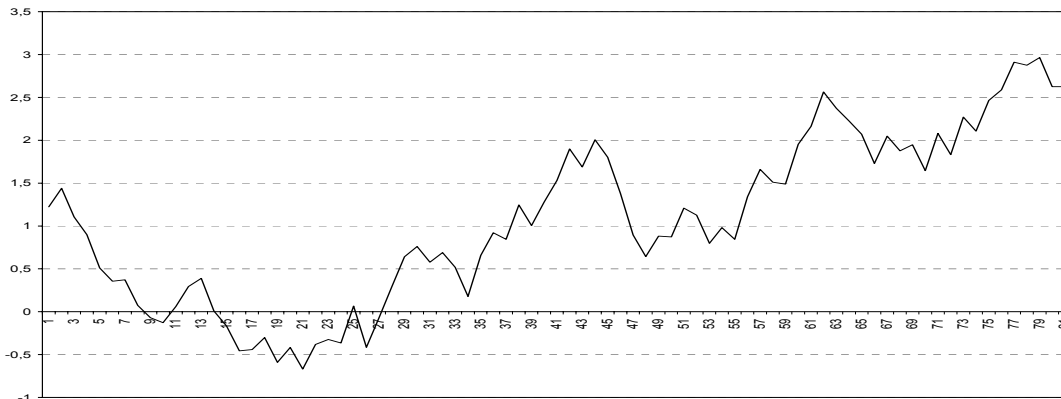
2.2.4 Sezónní diference

Sezónní diference je diference mezi okamžiky, vzdálenými o celistvý násobek délky periody. Například u dat s intervalem jeden měsíc, u nichž definujeme roční sezónní cyklus, se sezónní diference 1. řádu počítá jako rozdíl údaje z letošního ledna minus údaje z loňského ledna, z letošního února minus loňského února atd. Diference vyjadřuje velikost změny, ke které došlo mezi dvěma časovými okamžiky měření. Je-li kladná, řada v daném čase roste, je-li záporná, řada klesá. **Diferencí se data zbavují lineárního trendu, sezónní diferencí sezónních vlivů.**

2.2.5 Kumulativní součet

Opačnou operací k diferenci je kumulativní součet časové řady. Jeho hodnota se rovná součtu všech hodnot od počátku řady až po daný okamžik. Postupnou aplikací diference a kumulativního součtu získáme původní řadu opožděnou o jeden časový interval a zvětšenou nebo zmenšenou o nějakou konstantu.

Důležitou časovou řadou je řada vzniklá kumulativním součtem bílého šumu. Říká se jí náhodná procházka, protože nikdy nelze předvídat, zda tato funkce se obrátí vzhůru nebo dolů. Někdy je též nazývána procházkou „opilého námořníka“. Podle zákonitosti náhodné procházky by se měli řídit např. ceny akcií na burze. Náhodná procházka je hladší nežli bílý šum, jelikož integrace potlačuje vyšší frekvenční složky a zvýrazní nižší frekvence.



Obr. 4: Náhodná procházka tj. nepredikovatelná časová řada

2.2.6 Vyhlazování časových řad

Pokud je některá veličina měřena v příliš krátkých časových intervalech, může se stát, že následující členy se neliší téměř ničím jiným, než nahodilými odchylkami, jakýmsi šumem, který se přičítá ke „správné“ hodnotě sledované veličiny. Pokud lze předpokládat, že tato nahodilá chyba očekávanou hodnotu jednou zvětší a jindy zase zmenší (její střední hodnota je nulová) a jednotlivé chyby nejsou vzájemně závislé (tj. nekorelované), můžeme pak očekávat, že zpřůměrováním několika po sobě následujících pozorování budou se chyby mít tendenci navzájem rušit, zatímco skutečná sledovaná hodnota procesu tím vynikne. Na tomto pozorování jsou založeny metody vyhlazování časových řad.

- *Středové klouzavé průměry*: hodnota je nahrazena aritmetickým průměrem sebe a nejbližších předchozích pozorování, ležících nejdále do dané časové vzdálenosti.
- *Klouzavé průměry z předchozích hodnot*: hodnota je nahrazena aritmetickým průměrem sebe a nejbližších předchozích pozorování.
- *Klouzavé mediány*: hodnota je nahrazena mediánem sebe a nejbližších pozorování, ležících nejdále do dané časové vzdálenosti.

Jednou z aplikací těchto metod je také vyhlazení sezónních vlivů, pokud jako rozpětí zadáme délku jedné periody. V případě měsíčních dat s roční periodicitou je rozpětí 12.

2.3 Problémy časových řad

Při zpracování dat ve formě časové řady se potýkáme s množstvím problémů, které jsou právě pro časové řady specifické. Jedná se především o problémy:

- s volbou časových bodů:
 - okamžikové,
 - intervalové;
- s kalendářem:
 - různá délka měsíců,
 - různý počet víkendů v měsíci ,
 - různý počet pracovních dnů v měsíci ,
 - pohyblivé svátky;
- s délkou časových řad;
- nesrovnatelností dat.

Diskrétní časové řady obsahují pozorování v určitých nespojitých časových bodech a mohou vzniknout třemi způsoby: buď přímo *diskrétní svou povahou*, nebo vznikají *diskretizací spojité časové řady*, případně *agregací* či *průměrováním* hodnot za dané časové období.

Problémy s **kalendářem** znamenají – různá délka kalendářních měsíců, různý počet pracovních dní v měsíci, pohyblivé svátky (např. velikonoce). Tyto nepravidelnosti mohou mít překvapivé následky, avšak je možné je „očistit“ od těchto problémů:

- např. vyrovnání různého počtu dní v měsíci:

$$y_t^{(\text{očistěná})} = y_t \frac{\bar{p}_t}{p_t},$$

kde y_t – hodnota očišťovaného ukazatele,

p_t – počet pracovních dní v měsíci t ,

\bar{p}_t - průměrný počet pracovních dní v měsíci za rok (30,42) či jiný základ např. 30 dní.

- Některé krátkodobé nepravidelnosti v kalendáři mohou být odstraněny pomocí agregace např. použijeme-li čtvrtletně agregované hodnoty místo původních měsíčních údajů.

Problémy s délkou časových řad souvisí s počtem pozorování při analýze časových řad, ale je nezbytné respektovat i vnitřní strukturu řady. Na jedné straně některé analýzy časových řad vyžadují určitou minimální délku řady (např. Boxův-Jenkinsův přístup předpokládá minimálně 50 pozorování), na straně druhé u velice dlouhých časových řad je nebezpečí, že v průběhu tohoto časového období se mění charakteristiky modelu a tudíž vnitřní struktura generující řadu se stává s rostoucí délkou obtížně modelována v případě modelů předpokládající stabilní chování parametrů.

Problémy s nesrovnalostí jednotlivých měření souvisí s výběrovým vzorkem a zároveň reprezentativností tohoto vzorku i s hlediska časového vývoje.

V případě možné volby časových bodů pozorování sledujeme cíl našeho zkoumání, možnosti periodicity původní časové řady, změny ve vývoji a vnitřní strukturu časové řady. Při analýza

časové řady bychom měli vycházet minimálně ze 30 pozorování, což je např. v případě ročních ukazatelů problematické. Rovněž bychom měli respektovat ekvidistantní tj. (stejně vzdálené) časové body.

2. 4 Metody analýzy časových řad

Výběr metody analýzy časových řad závisí na řadě faktorů, ke kterým patří:

- účel analýzy (např. rozpoznání mechanismu generování hodnot časové řady a předpovídání jejího budoucího vývoje)
- typ časové řady,
- zkušenosti statistika,
- dostupná databáze,
- softwarové a hardwarové vybavení.

Základní metody a postupy k analýze časových řad:

- dekompozice časové řady,
- Boxova-Jenkinsova metodologie,
- lineární dynamické modely,
- spektrální analýza časových řad.

Dekompoziční metoda rozkládá časovou řadu na trendovou, cyklickou, sezónní a nesystematickou složku a zabývá se identifikací i modelováním zejména systematických složek, především trendové a sezónní složky.

Boxova-Jenkinsova metodologie bere v úvahu při konstrukci modelu časové řady reziduální složku, která může být tvořena korelovanými (závislými) náhodnými veličinami. Boxova-Jenkinsova metodologie tedy nejen může zpracovávat časové řady s navzájem závislými pozorováními, ale dokonce těžiště jejich postupů spočívá právě ve vyšetřování těchto závislostí neboli tzv. korelační analýze. Kombinují se autoregresivní modely $AR(p)$ s modely klouzavých průměrů reziduální složky $MA(q)$. V případě nestacionární časové řady se provádí stacionarizace např. diferencováním a zjišťuje se řád s parametrem d . Výsledný model se potom označuje jako $ARIMA(p,d,q)$, v případě sezónních vlivů SARIMA modely.

Lineární dynamické modely jsou zpravidla příčinné (kauzální) modely, kde je vysvětlovaná proměnná y_t vysvětlována vývoje svých zpožděných hodnot či dalších vysvětlujících faktorů. Rozdíl od modelu Box-Jenkinse spočívá v tom, že zde kromě popisované časové řady a bílého šumu vystupují ještě další časové řady – příčinné faktory.

Spektrální analýza časových řad má na rozdíl od předcházejících tří případů odlišný přístup spočívající v tom, že se zkoumaná časová řada považuje za směs sinusových a kosinových křivek s různými amplitudami a frekvencemi. Často se rovněž hovoří o tzv. fourierovské analýze. Pomocí speciálních statistických nástrojů se zjišťuje obraz o intenzitě zastoupení jednotlivých frekvencí v časové řadě (tzv. spektrum řady).

Dále bude pozornost věnována dekompozici časové řady.

Dekompozice časové řady

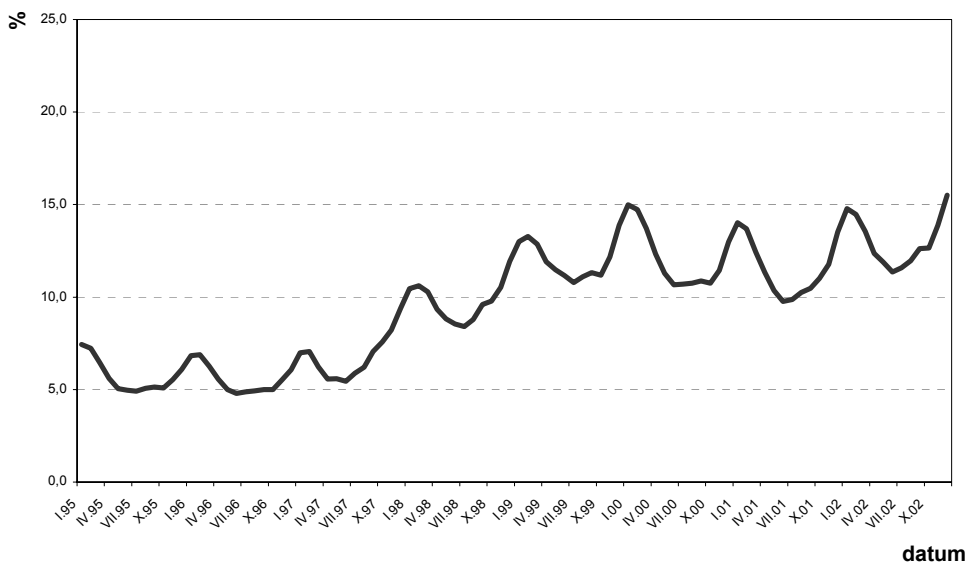
Při klasické analýze časových řad se vychází z předpokladu, že každá časová řada může obsahovat čtyři složky:

- a) trend (Tr_t),
- b) sezónní složku (Sz_t),
- c) cyklickou složku (C_t),
- d) náhodnou složku (E_t).

Provádění rozkladu (**dekompozice**) si klade za cíl snadněji identifikovat pravidelné chování časové řady než původní nerozložené řady.

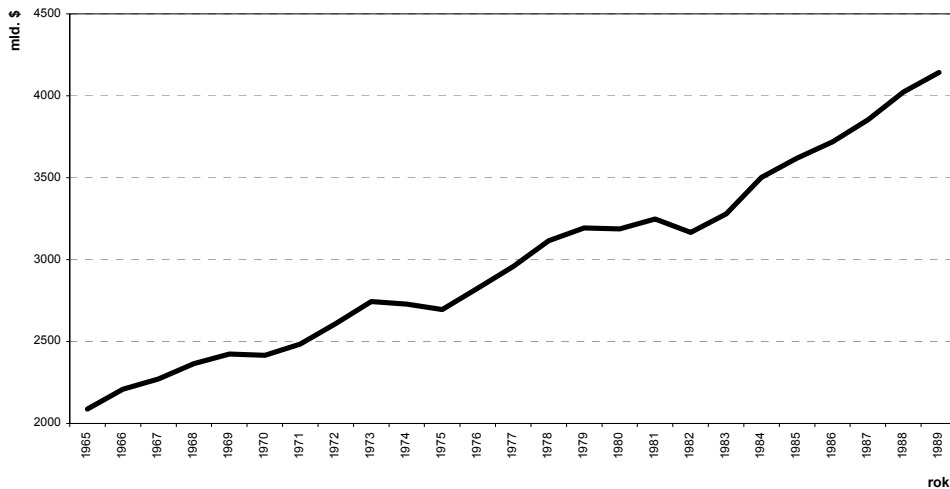
Trend vyjadřuje obecnou tendenci vývoje zkoumaného jevu za dlouhé období. Je výsledkem dlouhodobých a stálých procesů. Trend může být rostoucí, klesající nebo může existovat řada bez trendu. Trendová složka se většinou modeluje pomocí matematických křivek.

Sezónní složka je pravidelně se opakující odchylka od trendové složky. Perioda této složky je menší než celková velikost sledovaného období. Rovněž se tato složka může měnit svůj charakter.



Obr. 5: Vývoj míry nezaměstnanosti v okrese Znojmo

Cyklická složka udává kolísání okolo trendu v důsledku dlouhodobého cyklického vývoje, kdy dochází ke střídání fází růstu a poklesu. Jednotlivé cykly se vytvářejí za zpravidla období delší než jeden rok a mohou mít nepravidelný charakter tzn. různou amplitudu. Cykly jsou v ekonomických časových řadách způsobeny ekonomickými i neekonomickými faktory a často jsou obtížně pozorovatelné. V posledních letech se věnuje pozornost zejména technologickým, inovačním či demografickým cyklům. Obrázek 6 zobrazuje vývoj cyklické složky pro hrubý národní produkt v USA.



Obr. 6: Vývoj hrubého národního produktu USA a letech 1965 - 1989 (reálné ceny k roku 1982 v mld. \$)

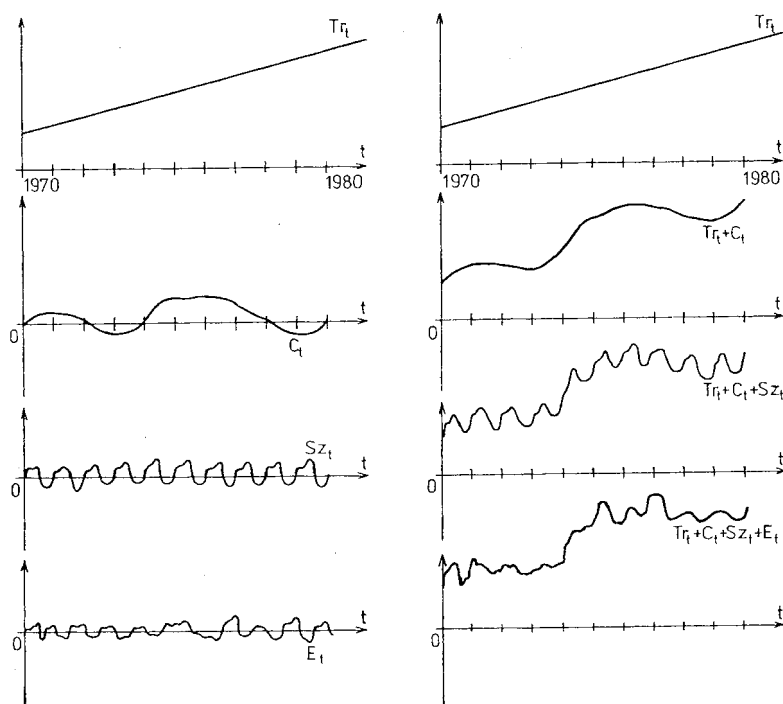
Náhodná (stochastická) složka vyjadřuje nahodilé a jiné nesystematické výkyvy (např. chyby měření). Předpokládá se, že náhodná složka je tvořena tzv. **bílým šumem** s normálním rozdělením. Pod pojmem bílý šum rozumíme nekorelované (vzájemně nezávislé) náhodné veličiny s nulovou střední hodnotou a konstantním rozptylem.

Vlastní dekompozice časové řady může zhrnovat formu aditivní nebo multiplikativní.

Aditivní dekompozice má tvar :

$$y_t = Tr_t + C_t + Sz_t + E_t .$$

Při aditivním tozkladu jsou jednotlivé složky uvažovány ve svých skutečných absolutních hodnotách a jsou měřeny v jednotkách řady y_t . Na obrázku č. 7 je schematicky znázorněn příklad dekompozice aditivní formy časové řady.



Obr. 7: Aditivní dekompozice časové řady

Multiplicativní forma má tvar:

$$y_t = Tr_t \cdot C_t \cdot Sz_t \cdot E_t.$$

Po aditivní dekompozici jsou jednotlivé složky časové řady ve stejných měrných jednotkách jako původní řada. Aditivní dekompozice se používá v případě, že variabilita hodnot časové řady je přibližně konstantní v čase.

Po multiplikativní dekompozici je trendová složka časové řady ve stejných měrných jednotkách jako původní časová řada, ale ostatní složky (cyklická, sezónní, nesystematická) jsou v relativním vyjádření. Tento způsob dekompozice se používá v případě, že variabilita časové řady roste v čase, nebo se v čase mění.

Na jedné straně kladou dekompoziční metody pozornost zejména na *systematické složky* časové řady a předpokládá se, že jednotlivá pozorování jsou navzájem nekorelována. V tomto případě je matematickým nástrojem v dekompozičních metodách zejména **regresní analýza**.

Analýza trendu

Trend v časových řadách je možné popsat pomocí *trendových funkcí* a *klouzavých průměrů*. Modelování trendu pomocí trendových funkcí se používá v případě, kdy trend odpovídá určité funkci např. lineární, kvadratické, exponenciální, S-křivky apod. Modelování trendu pomocí

klouzavých průměrů se používá, jestliže je vývoj časové řady v důsledku silného vlivu nesystematické složky nerovnoměrný nebo má extrémní hodnoty.

Při modelování trendu pomocí **trendových funkcí** se vychází z následujících předpokladů:

- Časová řada y_t je pro $t = 1, 2, \dots, n$ uspořádaná posloupnost hodnot v čase t , které získáme měřením určitého ukazatele ve stejně dlouhých časových intervalech.
- Časovou řadu y_t je možné zapsat ve tvaru $y_t = Y_t + E_t$, kde Y_t představuje teoretický model systematické složky vývoje ekonomického ukazatele Y v čase t a E_t vyjadřuje nesystematickou složku. Tato nesystematická složka má charakter *bílého šumu* (nulová střední hodnota, konstantní rozptyl, vzájemná lineární nezávislost), který se navíc řídí *normálním rozdělením*.
- V analýze časových řad lze vyjádřit $Y_t = f(t)$. Pokud se jedná pouze o časovou řadu s trendovou složkou, potom funkce f je trendová funkce. Je-li v časové řadě rovněž sezónní složka nebo cyklická složka, potom je Y_t kompozicí modelů těchto složek.

Existují dva základní přístupy k eliminaci trendu (vyrovnání, vyhlazení časové řady), kdy se odstraňují sezónní, cyklické a náhodné fluktuace:

- *klasické postupy* eliminace trendu (matematické analytické přístupy),
- *adaptivní postupy*, které automaticky reagují na případné změny v charakteru trendu (např. na změny ve směrnici lineárního trendu).

Matematické analytické přístupy zahrnují metody, při nichž se snažíme popsat trend analyticky některou jednoduchou křivkou. Po odhadu parametrů této křivky lze potom konstruovat bodovou nebo intervalovou předpověď za předpokladu, že charakter trendové funkce se nemění. Při tomto postupu se předpokládá, že analyzovaná časová řada má tvar: $y_t = Tr_t + E_t$, nebo byla na tento tvar převedena.

Základní trendové funkce pro $t = 1, 2, \dots, n$:

- **Konstantní trend** má tvar $Tr_t = \beta_0$,

odhad trendu je $\hat{y} = \hat{Tr} = \bar{y}$, odhad rozptylu nesystematické složky je $s_E^2 = s_y^2$.

- **Lineární trendová funkce** $Tr_t = \beta_0 + \beta_1 t$,

Odhad lineárního trendu je $\hat{y} = \hat{Tr} = \hat{\beta}_0 + \hat{\beta}_1 t$, odhad rozptylu nesystematické složky je

$$s_E^2 = \frac{1}{n-2} \sum_{t=1}^n (y_t - \hat{y}_t)^2.$$

- **Kvadratická trendová funkce (parabola)** má tvar $Tr_t = \beta_0 + \beta_1 t + \beta_2 t^2$,

odhad trendu je $\hat{y}_t = \hat{Tr}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2$, odhad rozptylu nesystematické složky je

$$s_E^2 = \frac{1}{n-3} \sum_{t=1}^n (y_t - \hat{y}_t)^2.$$

- **Exponenciální trendová funkce** má tvar $Tr_t = \beta_0 \beta_1^t$, kde parametry $\beta_0, \beta_1 > 0$ se odhadují metodou nejmenších čtverců, protože trendová funkce se po logaritmické úpravě převede na lineární funkci.

Odhad trendu je $\hat{y} = \hat{Tr}_t = \hat{\beta}_0 \hat{\beta}_1^t$ a odhad rozptylu nesystematické složky upravené po logaritmické transformaci je $s_E^2 = \frac{1}{n-2} \sum_{t=1}^n (y_t - \hat{y}_t)^2$.

- **S-křivka má tvar** $Tr_t = e^{(\beta_0 + \beta_1 \frac{1}{t})}$, který se po logaritmické transformaci dá převést na tvar hyperboly $\ln Tr_t = \beta_0 + \beta_1 \frac{1}{t}$. Parametry odhadujeme opět metodou nejmenších čtverců.

Odhad trendu je $\hat{y}_t = \hat{Tr}_t = e^{(\hat{\beta}_0 + \hat{\beta}_1 \frac{1}{t})}$ a odhad rozptylu nesystematické složky je po linearizaci logaritmováním $s_E^2 = \frac{1}{n-2} \sum_{t=1}^n (y_t - \hat{y}_t)^2$.

- **Modifikovaný exponenciální trend** má tvar $Tr_t = \gamma + \beta_0 \beta_1^t$, kde $\beta_0 < 0$, $0 < \beta_1 < 1$ a $\gamma > 0$. Konstanta γ je asymptotou (úroveň saturace, hladinou nasycení), ke které trend časové řady pro $t \rightarrow \infty$ konverguje. Přírůstek exponenciálního trendu β_1 je pomalejší, než přírůstek lineárního trendu. Modifikovaný exponenciální trend je populární v marketingu. Je to však nelineární funkce, kterou není možné linearizovat žádnou transformací, a proto se její parametry odhadují iterativními metodou. Tyto metody vyžadují výpočet počátečních odhadů parametrů funkce, které se dají získat např. metodou částečných součtů nebo metodou vybraných bodů.

- **Logistický trend** je uveden ve tvaru Pearlovy-Reedovy trendové funkce $Tr_t = \frac{1}{\gamma + \beta_0 \beta_1^t}$,

jejíž inverzní funkce $\frac{1}{Tr_t} = \gamma + \beta_0 \beta_1^t$ má tvar modifikovaného exponenciálního trendu.

Parametry se po inverzní transformaci odhadují stejným způsobem, jako pro modifikovaný exponenciální trend.

- **Gompertzův trend** má tvar $Tr_t = \gamma \beta_0^{\beta_1^t}$, resp. $Tr_t^* = \gamma^* + \beta_0^* \beta_1^t$. Křivka má horní asymptotu $\gamma^* = \ln \gamma$ a vyjadřuje hranici nasycení pro $t \rightarrow \infty$. Parametry původního trendového modelu se po transformaci odhadují jako u modifikovaného exponenciálního trendu nebo jednoduchého exponenciálního trendu.

Předpovídání pomocí trendových funkcí

Jedním ze základních účelů modelování časových řad v čase $t=1, 2, \dots, n$ je využití těchto modelů, v případě jejich statistické významnosti, k předvídání např. předpověď extrapolací. *Extrapolaci*

se rozumí kvantitativní odhady budoucích hodnot časové řady, které vznikají prodloužením vývoje z minulosti a přítomnosti do budoucnosti s horizontem $t=n+1, n+2, \dots, T$, za předpokladu, že se tento vývoj nezmění. Extrapoláční předpovědi rozdělujeme na bodové a intervalové.

Bodová předpověď – *extrapolace „ex ante“* se určuje v čase $t=n$ do okamžiku $t=T$ a označuje se $\hat{y}_n(T-n)$. Horizontem předpovídání se rozumí počet období $(T-n)$ od bodu $t=n$ do budoucnosti. $(1-\alpha)\cdot 100\%$ interval předpovědi (např. 95%) je interval, ve kterém se s pravděpodobností $(1-\alpha)\cdot 100\%$ (např. 95%) nachází skutečná hodnota y_T tzn.

$\hat{y}_n(T-n) \pm t_{1-\alpha/2}(n-(l+1)) \cdot s_p$, kde $t_{1-\alpha/2}(n-(l+1))$ je $(1-\alpha)\cdot 100\%$ kvantil Studentova rozdělení s $n-(l+1)$ stupni volnosti, kde $(l+1)$ je počet odhadnutých parametrů v polynomiálních funkcích, s_p je směrodatná chyba předpovědi v horizontu $(T-n)$.

Když určujeme extrapolace, ta se předpokládá, že vybraný model je správný a skutečné parametry modelu se v čase nemění. V mnoha situacích jsou tyto předpoklady nereálné, protože proces, který generuje vývoj časové řady se mění v čase. Čím je horizont předpovědi delší, tím je možné očekávat větší chyby předpovědi.

Chyba předpovědi při extrapolaci je dána vztahem:

$\hat{E}_T = y_T - \hat{y}_n(T-n)$, kde $\hat{y}_n(T-n)$ je bodová předpověď v čase T a y_T je skutečná hodnota v čase T . Chybu předpovědi lze rozložit na dvě složky:

$\hat{E}_T = (y_T - Y_T) + (Y_T - \hat{y}_n(T-n))$, kde $(y_T - Y_T)$ je chyba způsobená volbou modelu (předpokládá se správná volba tj. tato složka = 0) a $(Y_T - \hat{y}_n(T-n))$ je chyba způsobená odhadem parametrů modelu.

Příklad bodové a intervalové předpovědi pro lineární trendovou funkci:

- bodová předpověď : $\hat{y}_n(T-n) = \hat{\beta}_0 + \hat{\beta}_1(T)$,
- $(1-\alpha)\cdot 100\%$ předpovědní interval : $\hat{y}_n(T-n) \pm s_E \cdot t_{1-\alpha/2}(n-2) \sqrt{1 + \frac{1}{n} + \frac{(T-\bar{t})^2}{n(n^2-1)/12}}$,

kde s_E je směrodatná odchylka reziduí.

Při výběru trendové funkce je nutné respektovat :

- graf časové řady resp její transformace,
- interpolační kritéria (směrodatná odchylka reziduí, koeficient determinace, koeficient autokorelace reziduí, testy parametrů),
- extrapoláční kritéria (průměrné charakteristiky chyb předpovědi „ex post“, graf předpověď-skutečnost).

Grafická analýza slouží k předběžnému výběru vhodné trendové funkce:

- kolísá-li řada prvních diferencí okolo nuly, volíme konstantní trend;
- kolísá-li řada prvních diferencí kolem nenulové konstanty, použijeme lineární trend;

- jestliže má řada prvních diferencí přibližně lineární trend a řada druhých diferencí konstantní trend, volíme kvadratický trend;
- kolísá-li řada koeficientů růstu nebo řada prvních diferencí okolo nenulové hodnoty, volíme jednoduchý exponenciální trend;
- jestliže má řada $\ln y_t$ přibližně hyperbolický průběh, volíme S-křivku;
- jestliže řada podílů sousedních diferencí $(y_t - y_{t-1})/(y_{t-1} - y_{t-2})$ kolísá okolo nenulové konstanty, volíme modifikovaný exponenciální trend;
- jestliže řada podílů sousedních diferencí $(\ln y_t - \ln y_{t-1})/(\ln y_{t-1} - \ln y_{t-2})$ kolísá okolo nenulové hodnoty, volíme Gompertzovu křivku.

Interpolační kritéria zkoumají charakter rozdílů skutečných hodnot y_t a vyrovnaných hodnot \hat{y}_t . Mezi míry přesnosti vyrovnaní náleží následující charakteristiky reziduí:

- součet čtvercových chyb (*Sum of Squared Error*) $SSE = \sum_{t=1}^n E_t^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2$,
- průměrná (střední) čtvercová chyba $= MSE = \frac{SSE}{n}$ a
- průměrná absolutní chyba $= MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$.

Klasická analýza časových řad předpokládá, že trendová funkce má v čase konstantní parametry. V delším časovém období je tento předpoklad nereálný, proto je vhodné využívat adaptivní techniky, jako je metoda klouzavých průměrů a exponenciální vyrovnavání.

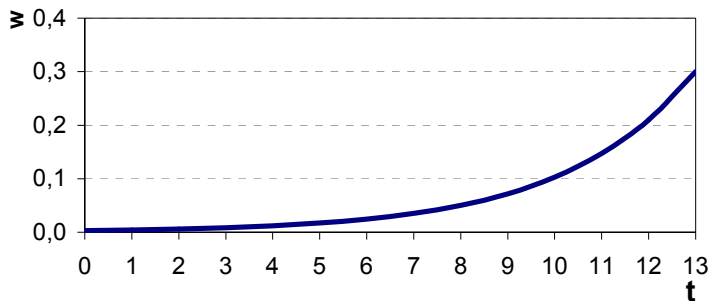
Klouzavý průměr

Metoda klouzavých průměrů se zakládá na myšlence, že časovou řadu y_t pro $t=1, 2, \dots, n$ rozdělíme na kratší časové úseky o počtu hodnot $2m+1$, na kterých odhadujeme lokální polynomicke trendy určitého stupně. Např. konstantní trend se popisuje polynomem nultého stupně, lineární trend polynomem prvního stupně. První část časové řady má $2m+1$ hodnot, které označujeme $y_1, y_2, \dots, y_{2m+1}$, z nich odhadneme parametry lokálního trendu vhodným polynomem a vypočítáme jeho odhad $\hat{T}r_{m+1}$, stejný polynom odhadneme na druhé skupině hodnot řady, $y_2, y_3, \dots, y_{2m+2}$ a vypočítáme odhad lokálního trendu $\hat{T}r_{m+2}$, tímto klouzavým způsobem pokračujeme až do konce časové řady.

V sezónních časových řadách se trendová složka odhaduje pomocí centrovaných klouzavých průměrů, protože délka klouzavé části je sudé číslo.

Exponenciální vyhlazování

Je vhodné zejména pro krátkodobou predikci trendů. Tato technika, nenáročná na čas a teoretické znalosti, rozvíjí myšlenku vyhlazování pomocí klouzavých průměrů. Metoda exponenciálního vyrovnávání je založena na všech předchozích pozorováních, přičemž jejich váha (w) směrem do minulosti klesá podle **exponenciální** funkce: $w_t = (1 - \alpha) \cdot \alpha^{n-t}$ (viz obrázek 8), kde n je počet pozorování a α je vyrovnávací konstanta v intervalu (0;1).

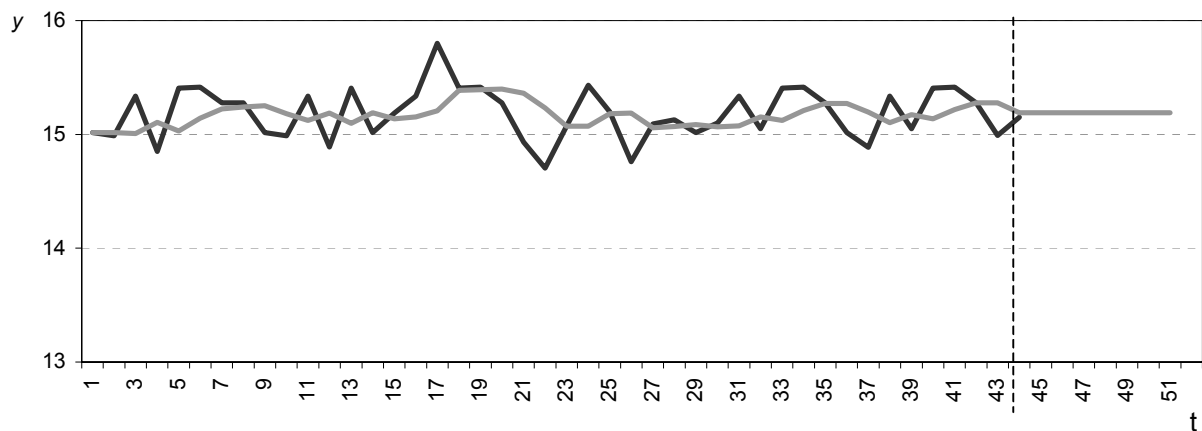


Obr. 8: Vývoj váhy (w) dle exponenciální funkce v čase (t) pro $\alpha = 0,7$ a $n = 13$.

Intenzita zapomínání, vyjádřená velikostí alfy, se stanoví na základě charakteru časové řady. Hledá se taková hodnota α , u které je nejmenší SSE příp. MSE . V programu EXCEL se pro hledání extrému funkcí používá nástroj Řešitel.

Exponenciální vyrovnání prvního stupně.

Tento nejjednodušší způsob vyrovnání lze použít pouze na časové řady, které nevykazují žádný trend, avšak při analýze ukazatelů trhu práce se většinou nepoužívá, ale uvádíme ho pro pochopení složitějších formy. U časových řad, které jsou ve tvaru $y_t = Tr_t + E_t$, lze v případě konstantního trendu nahradit trendovou složku (Tr) konstantou, tj. $Tr_t = \beta_0$. Úkolem je tedy nalézt odhad parametru β_0 , který se v tomto případě rovná vyrovnané hodnotě \hat{y} . Vyrovnaná časová řada se vypočítá podle následujícího rekurentního vzorce: $\hat{y}_t = (1 - \alpha) \cdot y_t + \alpha \cdot \hat{y}_{t-1}$. Pokud se α blíží k hodnotě 1 tak roste vliv minulých pozorování. Pro hledání vhodné α se většinou doporučuje interval $(0,7;1)$. Výše uvedený vzorec lze přepsat i do následujícího tvaru: $\hat{y}_t = \hat{y}_{t-1} + (1 - \alpha) \cdot (y_t - \hat{y}_{t-1})$, který vysvětluje vytváření nové vyrovnané hodnoty z předchozí vyrovnané hodnoty, „opravenou“ o chybu danou rozdílem mezi skutečnou a předcházející vyrovnanou hodnotou. Problémem rekurentních vzorců je stanovit odhad vyrovnané hodnoty pro $t = 1$, kterou neznáme. Existují sice algoritmy jak tuto hodnotu stanovit, ale nejjednodušší je aproximovat ji skutečnou hodnotou v čase $t = 1$. Metoda exponenciálního vyrovnání brzy na tuto nepřesnost „zapomene“, tzn. že po případném počátečním odklonu se vyrovnané hodnoty brzy přiblíží k naměřeným pozorováním. Příklad exponenciálního vyrovnání ukazuje na simulovaných datech obrázek 9.



Obr. 9: Exponenciální vyrovnání prvního stupně s predikcí.

Exponenciální vyrovnání druhého stupně

Dvojité exponenciální vyrovnání používáme v případě, kdy lze předpokládat, že v krátkém období bude mít trendová složka lineární formu: $Tr_t = \beta_0 + \beta_1 \cdot t$. Předpoklad linearity v krátkém období je v praxi velice rozšířen. Postup si předvedeme na následujícím příkladu.

3 Příklad č. 1: Analýza míry nezaměstnanosti v okrese Karviná

V této části si ukážeme typický postup při analýze časových řad z následnou predikcí na příkladu vývoje míry nezaměstnanosti v okrese Karviná v období 1995 –2002.

Budou sledovány následující kroky analýzy časové řady:

1. grafická a statistická deskripce,
2. očištění časové řady od sezónních vlivů,
3. tvorba modelu (exponenciální vyrovnání druhého stupně) s predikcí,
4. konstrukce výsledné predikce.

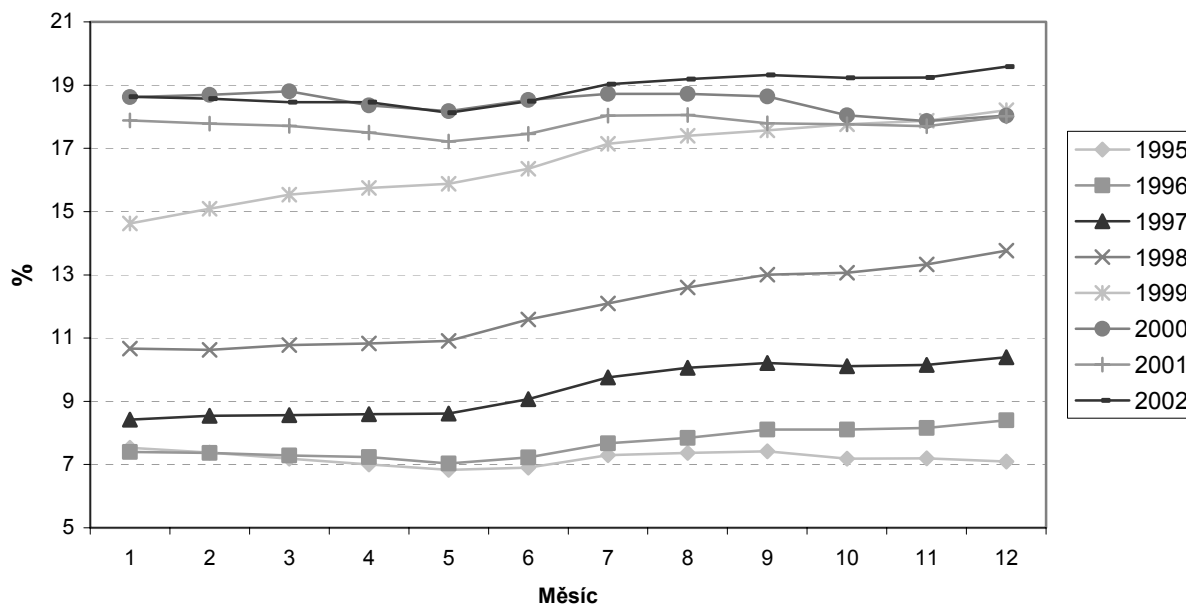
3.1 Grafická a statistická deskripce

Kromě klasického zobrazení časové řady, jak jej můžete vidět na obr č. 1. Je vhodné pro stanovení sezónnosti provést restrukturalizaci dat pomoci kontingenční tabulky. Novou datovou matici je zapotřebí vytvořit tak, aby roky byly ve sloupcích a měsíce v řádcích, viz následující tabulka 2.

Tab. 2: Restrukturovaná datová matice.

Měsíc	Rok							
	1995	1996	1997	1998	1999	2000	2001	2002
1	7,53	7,40	8,42	10,67	14,63	18,62	17,88	18,63
2	7,38	7,37	8,55	10,62	15,09	18,69	17,78	18,57
3	7,18	7,29	8,56	10,78	15,53	18,80	17,71	18,46
4	7,00	7,24	8,59	10,82	15,75	18,36	17,49	18,46
5	6,84	7,04	8,61	10,91	15,88	18,18	17,22	18,12
6	6,91	7,23	9,07	11,59	16,36	18,53	17,46	18,49
7	7,30	7,67	9,76	12,10	17,15	18,73	18,04	19,03
8	7,37	7,85	10,06	12,60	17,39	18,72	18,06	19,19
9	7,42	8,11	10,22	13,00	17,57	18,64	17,79	19,32
10	7,18	8,10	10,11	13,07	17,76	18,05	17,76	19,23
11	7,19	8,16	10,15	13,33	17,87	17,87	17,70	19,24
12	7,10	8,40	10,39	13,76	18,21	18,04	18,02	19,58

Z této tabulky pak lze vytvořit následující graf 10, ve kterém lze pozorovat sezónnost, která se neprojevuje tak výrazně jak např. u okresu Znojmo (viz. obr 5). Je vidět že charakter sezónnosti je u okresu Karviná ovlivněn především od 5. měsíce nárůstem počtu absolventů. Z důvodu struktury zaměstnanosti v okrese se zde neprojevuje typická variabilita způsobená sezónními pracemi.



Obr. 10: Vývoj míry nezaměstnanosti v okrese Karviná v jednotlivých měsících.

Statistickou deskripci dat a to jak časových, tak i průřezových provádíme v programu EXCEL pomocí analytického nástroje *Popisná statistika*¹. Standardní výstup získáme zatržením položky Celkový přehled. Pokud provedeme statistickou deskripci vývoje míry nezaměstnanosti pro okres Karviná v letech 1995 až 2002 dostáváme hodnoty, které jsou uvedeny v následující tabulce.

¹ Tento nástroj se vyvolává položkou analýza dat, která je umístěna v Menu (Nástroje-Analýza dat), pokud chybí v nabídce Analýz dat je nutné do instalovat doplněk Analytické nástroje (Nástroje –Doplňky).

Tab. 3: Popisná statistika míry nezaměstnanosti okresu Karviná v období 1995-2002

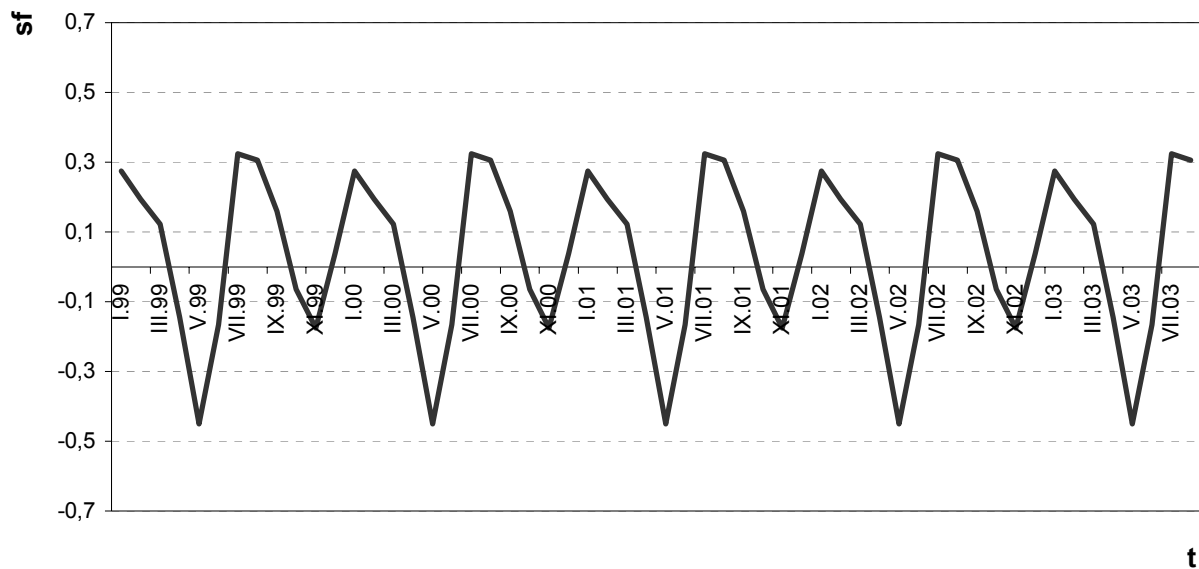
u_KI	Popis
Stř. hodnota	13,48 Průměr.
Chyba stř. hodnoty	0,48 Slouží k výpočtu intervalů spolehlivosti pro průměr.
Medián	14,19 Hodnota, která rozděluje soubor na dvě poloviny. Není citlivá na extrémní hodnoty jako průměr.
Modus	#N/A Nejčastější hodnota, používá se především u kategorizovaných dat.
Směr. odchylka	4,74 Měří variabilitu neboli heterogenost dat.
Rozptyl výběru	22,46 Směrodatná odchylka na druhou.
Špičatost	-1,73 Měří koncentraci hodnot kolem střední hodnoty. U normálního rozdělení se pohybuje kolem nuly.
Šikmost	-0,16 Měří symetrii rozložení hodnot. U normálního rozdělení se přibližuje 0.
Rozdíl max-min	12,75
Minimum	6,84
Maximum	19,58
Součet	1293,61
Počet	96

Tyto charakteristiky mají význam především při porovnávání časových řad jednotlivých regionů.

Nejdůležitější částí této etapy je rozhodnutí o délce analyzovaného období a výběru vhodné metody. Jak již bylo dříve řečeno v daném období nesmí dojít ke změně charakteru časové řady. Rovněž pozorování musí být metodicky srovnatelná. Na druhé straně je požadováno co největší délka časové řady. Pokud pracujeme s měsíčními údaji měla by mít časová řada minimálně 30 pozorování. Některé metody ale požadují daleko větší počet pozorování. Pro naši další analýzu z hlediska vývoje míry nezaměstnanosti použijeme období od roku 1999 do roku 2002, tj. 48 pozorování. Časová řada v tomto období vykazuje mírnou sezónní variabilitu, která má aditivní charakter tzn. neroste její variabilita se zvyšující se mírou nezaměstnanosti. Pozor aditivní forma není pravidlem u míry nezaměstnanosti.

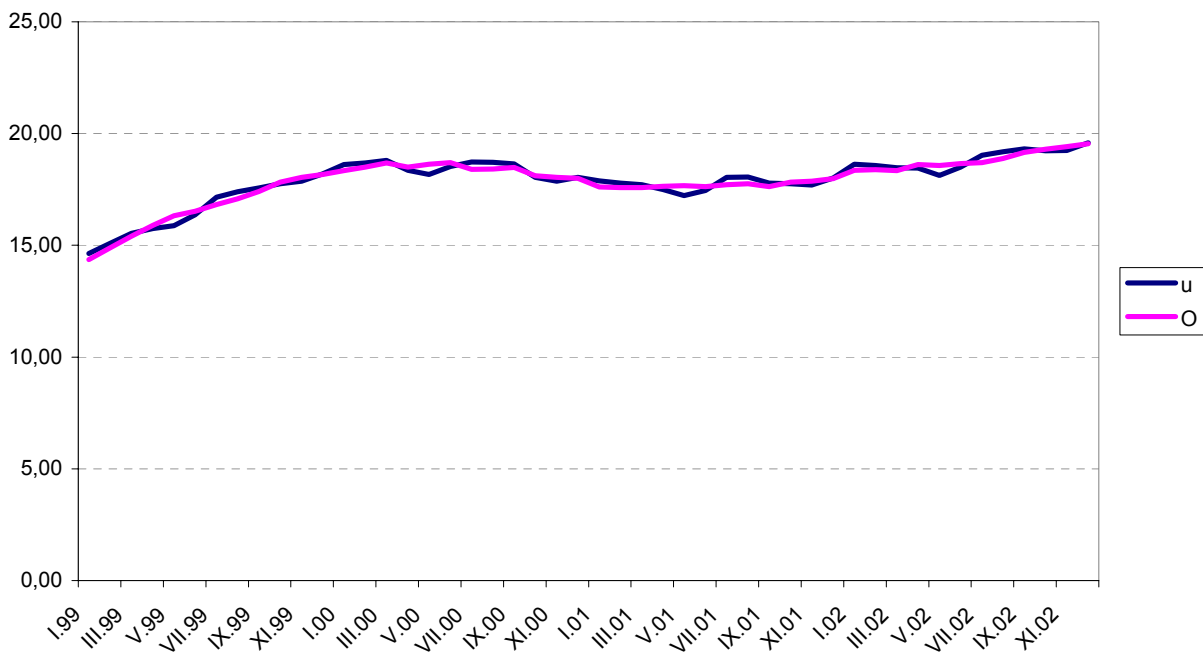
3.2 Očištění časové řady od sezónních vlivů

Pro očištění časové řady musíme nejdříve stanovit sezónní faktor v aditivní formě jak bylo výše uvedeno. Tento faktor stanovíme pomocí centrovaných ročních klouzavých průměrů, které odečteme od skutečně naměřených hodnot. Poté vypočítáme průměrnou měsíční odchylku a tím získáme požadovaný sezónní faktor. Sezónní faktor je pro všechny roky stejný, tato pravidelnost je v níže uvedeném grafu velice zřetelná.



Obr. 11: Vývoj sezónního faktoru (*SF*) v okrese Karviná

V následujícím kroku odečteme tento sezónní faktor od naměřených hodnot. Takto očištěné údaje se hodí nejen do regresních modelů, ale i pro metodu exponenciálního vyrovnání druhého stupně. Na druhou stranu některé metody si sezónnost analyzují svými postupy např. Wintersova metoda nebo SARIMA modely.



Obr. 12: Původní (*u*) a očištěná (*O*) míra nezaměstnanosti v okrese Karviná

3.3 Tvorba modelu – Exponenciální vyrovnání druhého stupně

Při exponenciální vyrovnání druhého stupně postupujeme následujícím způsobem:

1. Stanovíme velikost α v intervalu (0;1) v našem případě $\alpha = 0,7$
2. Nejdříve si stanovíme jednoduchou vyrovnávací statistiku $S1$ podle následujícího vztahu:

$$S1_t = (1 - \alpha) \cdot y_t + \alpha \cdot S1_{t-1}.$$

3. Dále si určíme dvojitou vyrovnávací statistiku $S2$:

$$S2_t = (1 - \alpha) \cdot S1_t + \alpha \cdot S2_{t-1}$$

4. Pro vyrovnanou časovou řadu pak platí vztah:

$$\hat{y}_t = (2 + (1 - \alpha) / \alpha) \cdot S1_t - (1 + (1 - \alpha) / \alpha) \cdot S2_t$$

Hodnoty $S1$ a $S2$ v čase $t = 1$ aproximujeme hodnotou y_1 . Výsledky těchto kroků shrnuje tabulka 4.

Tab. 4: Výpočet $S1$ a $S2$ pro $\alpha = 0,7$

Datum	t	\hat{y}_t	S1	S2
I.99	1	14,356	14,356	14,356
II.99	2	14,678	14,517	14,404
III.99	3	15,165	14,785	14,518
IV.99	4	15,715	15,116	14,698
V.99	5	16,261	15,479	14,932
VI.99	6	16,654	15,793	15,190
VII.99	7	17,014	16,102	15,464
VIII.99	8	17,332	16,398	15,744
IX.99	9	17,658	16,701	16,031
X.99	10	18,044	17,037	16,333
XI.99	11	18,348	17,340	16,635
XII.99	12	18,542	17,589	16,921
I.00	13	18,709	17,815	17,189
II.00	14	18,851	18,020	17,439
III.00	15	18,999	18,219	17,673

5. Vypočteme míru odlišnosti vyrovnaných hodnot od skutečných např. MSE . V našem případě $MSE = 0,0376$
6. Provedeme optimalizace α pomocí Řešitele tak, že budeme minimalizovat hodnotu MSE . Řešitel v našem případě stanovil optimální velikost $\alpha = 0,568$ pro $MSE = 0,0302$.
7. Odhadnuté konstanty vypočteme následovně:

$$\beta_0(t) = 2 \cdot S1_t - S2_t, \beta_1(t) = \frac{1 - \alpha}{\alpha} (S1_t - S2_t), \text{ tzn. } \beta_0(48) = 19,55, \beta_1(48) = 0,142.$$

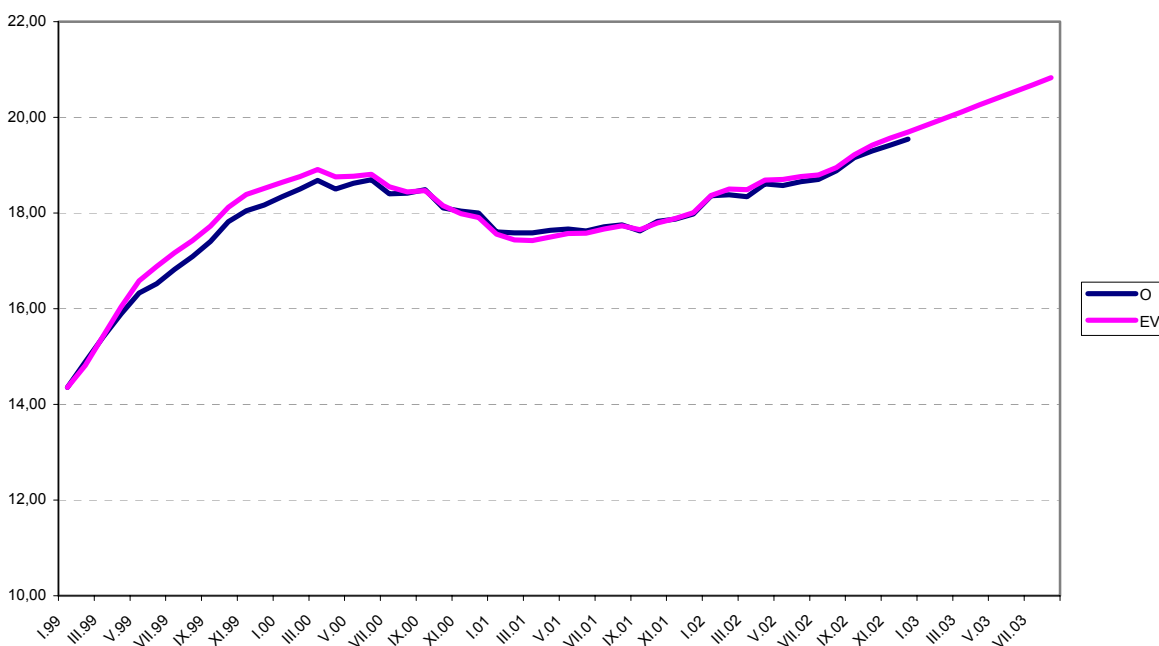
8. A nakonec se provedeme bodovou predikci pomocí odhadnutých konstant dle následujícího vztahu :

$$\hat{y}_t = \beta_0(n) + \beta_1(n)(t - n + 1) \text{ kde } n \text{ je počet naměřených hodnot.}$$

Tab. 5: Predikce očištěných hodnot

t	$t-n+1$	\hat{y}_t
49	2	19,833
50	3	19,975
51	4	20,117
52	5	20,259
53	6	20,401
54	7	20,543
55	8	20,684
56	9	20,826

Porovnání očištěné míry nezaměstnanosti s vyrovnanými hodnotami a jejich predikcí s délkou horizontu 8 měsíců zobrazuje následující graf.

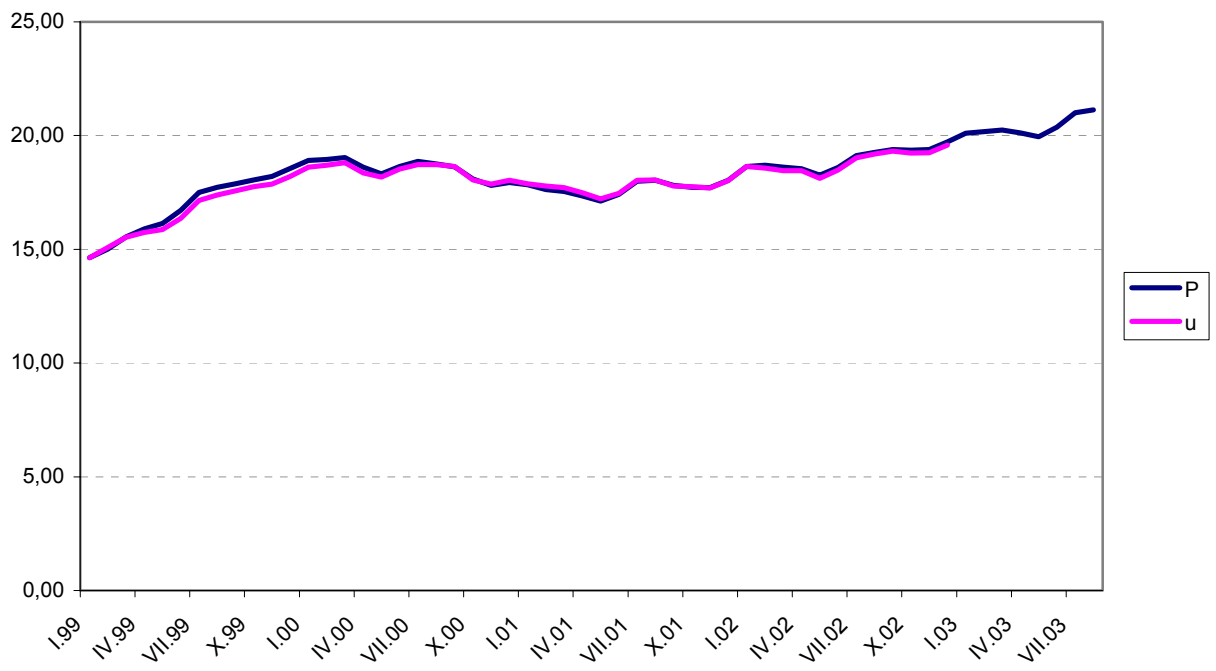


Obr. 13 Dvojité exponenciální vyrovnání (EV) očištěné míry nezaměstnanosti v okrese Karviná s predikcí.

Intervaly spolehlivosti, které jsou obvyklým doplňkem této predikce, se v EXCELU komplikovaně počítají, proto případné zájemce odkazují na publikaci T. Cipry (1986, str. 71).

3.4 Výsledná predikce

Posledním krokem je přičtení sezónního faktoru k predikci trendové složky. Výsledná bodová predikce je uvedena na obr. 14.



Obr. 14: Původní míra nezaměstnanosti (*u*) a její výsledná bodová predikce (*P*).

4 Příklad č. 2: Postup analýzy ukazatelů na úrovni obce

Časové řady na úrovni obce jsou nestabilní, a proto nelze aplikovat standardní postupy analýzy časových řad. Zaměříme tedy pouze na vyhodnocení vhodnosti okresní predikce pro jednotlivé obce. Postup je v prvním kroku analogický jako v předcházejícím případě:

- statistická a grafická deskripce,
- analýzu vztahů mezi okresní mírou nezaměstnanosti a mírou nezaměstnanosti v jednotlivých obcích pomocí koeficientu korelace.

4.1 Deskripce statistická a grafická

Konstrukce popisné statistiky se provede stejně jako v předchozím příkladě. Výstupy EXCELU je vhodné transformovat do podoby tabulky 6. Pro lepší interpretaci je možné seřadit obce např. podle výše průměru. Za období 1995 - 2002 (tj. 96 měsíčních pozorování) je nejvyšší průměr míry nezaměstnanosti ve městě Karviná. Opačný extrém vykazují obce Těrlicko a Chotěbuz. Nejvyšší variabilitu byla shledána pro obec Horní Suchá a nejnižší pro obec Chotěbuz.

Tab. 6: Statistická deskripce míry nezaměstnanosti pro jednotlivé obce okresu Karviná i pro okres jako celek

Území	Stř. hodnota	Chyba stř. hod.	Medián	Modus	Směr. odchylka	Rozptyl výběru	Špičatost	Šikmost	Rozdíl max-min	Minimum	Maximum	Součet	Počet
KARVINÁ	16,18	0,56	17,50	8,70	5,46	29,80	-1,714	-0,209	14,6	8,1	22,7	1553,6	96
ORLOVÁ	15,93	0,56	16,40	21,10	5,49	30,17	-1,751	-0,129	14,6	8,2	22,8	1528,9	96
PETŘVALD	14,82	0,61	15,70	21,10	6,02	36,23	-1,702	-0,204	16,1	6,2	22,3	1422,8	96
DOUBRAVA	13,98	0,66	16,25	19,10	6,44	41,44	-1,638	-0,295	18,4	4,4	22,8	1341,9	96
ČESKÝ_TĚ	13,55	0,53	15,90	18,30	5,18	26,87	-1,568	-0,400	14,0	5,6	19,6	1301,0	96
Okres_KI	13,48	0,48	14,19	#N/A	4,74	22,46	-1,732	-0,158	12,7	6,8	19,6	1293,6	96
HORNÍ_SU	13,24	0,67	13,10	5,40	6,58	43,29	-1,726	-0,046	17,7	4,4	22,1	1271,5	96
BOHUMÍN	12,46	0,47	13,15	6,30	4,56	20,81	-1,625	-0,181	13,1	5,9	19,0	1196,3	96
HAVÍŘOV	12,21	0,57	12,90	18,40	5,59	31,25	-1,774	-0,119	13,9	4,8	18,7	1171,7	96
RYCHVALD	11,08	0,44	11,45	15,80	4,35	18,93	-1,677	-0,127	12,0	4,8	16,8	1063,8	96
STONAVA	10,51	0,47	11,35	4,50	4,64	21,50	-1,446	-0,126	14,1	3,4	17,5	1008,9	96
ALBRECHT	9,82	0,49	10,30	15,00	4,77	22,72	-1,650	-0,100	13,2	3,3	16,5	942,4	96
DĚTMAROV	9,69	0,50	9,80	14,90	4,87	23,72	-1,721	-0,067	13,5	2,9	16,4	930,4	96
PETROVIC	9,38	0,40	9,80	6,50	3,93	15,45	-1,717	-0,091	11,1	3,7	14,8	900,4	96
DOLNÍ_LU	8,99	0,46	11,05	3,10	4,54	20,58	-1,761	-0,231	12,2	2,8	15,0	863,5	96
CHOTĚBUZ	8,41	0,37	9,05	3,70	3,59	12,86	-1,074	0,141	13,1	2,7	15,8	807,5	96
TĚRLICKO	7,92	0,41	8,10	2,20	4,06	16,50	-1,579	-0,190	11,7	2,1	13,8	760,0	96

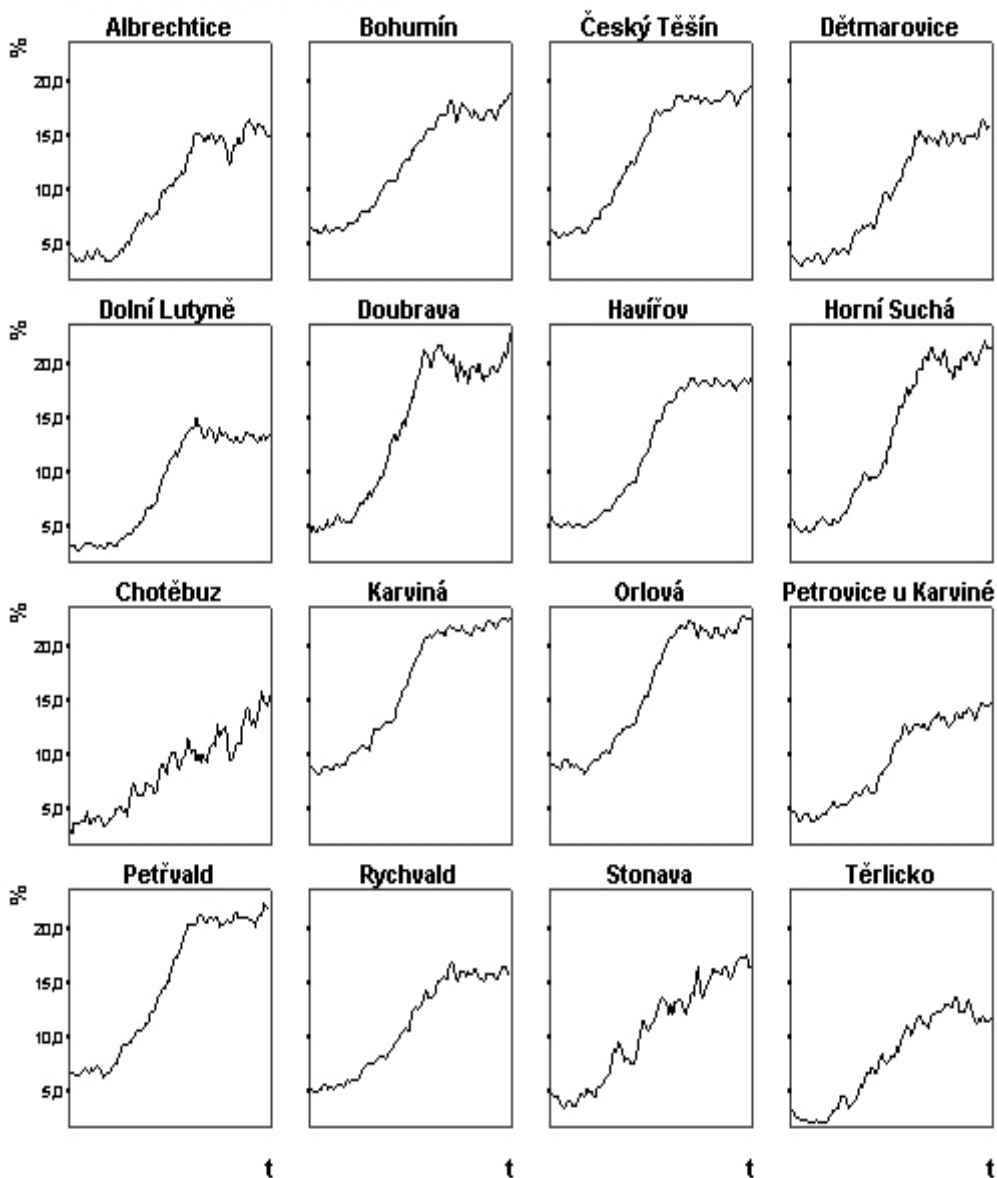
Pro správnou interpretaci výše uvedené tabulky je zapotřebí nutně znát i velikost obcí. Pokud analyzujeme trh práce lze použít počet ekonomicky aktivních obyvatel.

Tab. 7: Velikost obce dle počtu ekonomicky aktivních obyvatel k 3.3.2001

Obec	Počet ekonom. aktivních obyvatel
Havířov	40986
Karviná	31798
Orlová	17621
Český Těšín	13231
Bohumín	11808
Rychvald	3217
Petřvald	3078
Petrovice u Karviné	2339
Dolní Lutyně	2295
Horní Suchá	2071
Albrechtice	2017
Těrlicko	1967
Dětmárovice	1802
Stonava	822
Doubrava	803
Chotěbuz	452

Pozn.: Zdroj dat SLDB, 2001.

Dále je nezbytné provést i grafickou deskripci, která nám pomáhá identifikovat extrémní hodnoty. Při porovnávání ukazatelů je nutné zachovat ve všech grafech stejné měřítko.



Obr. 15: Vývoj míry nezaměstnanosti v obcích v okrese Karviná v období 1995 až 2002

Z výše uvedeného obrázku lze určit odlišný charakter vývoje míry nezaměstnanosti v malých obcích jako je Chotěbuz, Stonava, Těrlicko a další.

4.2 Analýza vztahů

Analýza vztahů mezi mírou nezaměstnanosti v okrese a v jednotlivých obcích se provádí pomocí koeficientu korelace, který měří lineární závislost tohoto vzájemného vztahu. Velikost koeficientu korelace výrazně závisí na výběru časového úseku. Ve druhém sloupci tabulky 8 jsou uvedeny koeficienty korelace pro celé analyzované období 1995 - 2002 a ve třetím sloupci jsou korelace vypočítané ze zkráceného období 2000 až 2002. Jelikož poslední pozorování mají nejsilnější vliv na budoucí vývoj je vhodnější použít koeficienty korelace ze zkráceného období jako míru významnosti okresní predikce pro danou obec. Čím je hodnota tohoto koeficientu korelace blíže 1, lze předpokládat, že vývoj míry nezaměstnanosti v obci bude kopírovat vývoj míry nezaměstnanosti v okrese.

Tab. 8: Korelace s mírou nezaměstnanosti v okrese Karviná

Obec	1995-2002	2000-2002
Orlová	0,999	0,880
Karviná	0,998	0,757
Český Těšín	0,986	0,749
Dětmarovice	0,993	0,714
Doubrava	0,986	0,694
Horní Suchá	0,991	0,659
Albrechtice	0,990	0,649
Petřvald	0,997	0,644
Bohumín	0,993	0,636
Chotěbuz	0,939	0,522
Petrovice u Karviné	0,992	0,506
Dolní Lutyně	0,990	0,306
Rychvald	0,992	0,279
Havířov	0,995	0,234
Stonava	0,965	0,141
Těrlicko	0,978	-0,518

5 Závěr

Předložený pracovní materiál se snaží ve stručné podobě zachytit jednoduché metody analýzy časových řad, které lze aplikovat na regionálních trzích práce v České republice s využitím všeobecně dostupného programu EXCEL. Je nutné podotknout, že výše uvedený software není vhodný pro sofistikované analýzy časových řad, a proto se v těchto souvislostech jeví nejvhodnější metoda analýzy časových řad – metoda exponenciálního vyrovnání druhého stupně. Dalším cílem tohoto materiálu bylo využití těchto modelů pro predikce ukazatelů trhů práce na úrovni okresů případně obcí za předpokladu jejich vhodnosti k předvídání.

Tento materiál byl sestaven na základě níže uvedené literatury a zkušeností autorů.

Literatura

ARLT, J a kol. (2002) *Analýza ekonomických časových řad s příklady*. Praha: VŠE. 148 s.

CIPRA, T. (1986). *Analýza časových řad s aplikacemi v ekonomii*. Praha: SNTL. 248 s.

SC&C (1999). *Úvod do analýzy časových řad v SPSS*. Praha: SC&C. 38 s.