

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

1/2018

ročník/volume 28

Recenzovaný vedecký časopis so zameraním na prezentáciu moderných štatistických a demografických metód a postupov.

Scientific reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures.

Článok/Article: 2

Typ článku/Type of article: vedecký článok/scientific article

Strany/Pages: 18 – 28

Dátum vydania/Publication date: 15. január 2018/January 15, 2018



Roman PAVELKA
Štatistický úrad SR

MODELOVÁNÍ ČASOVÝCH ŘAD POMOCÍ PROCEDURY EXPAND STATISTICKÉHO SYSTÉMU SAS

MODELING OF TIME SERIES WITH THE SAS PROCEDURE EXPAND

ABSTRAKT

Mezi nejdůležitější operace nad časovými řadami jsou především interpolace časové řady a agregace časové řady. Ke konverzi časových řad z jednoho časového intervalu na jiný (resp. z jedné frekvence pozorování na jinou) a k interpolaci chybějících hodnot v časové řadě je možné použít proceduru EXPAND statistického systému SAS. Pomocí procedury EXPAND je možné časovou řadu rozložit z kratších časových intervalů na časovou řadu s delšími časovými intervaly, případně rozšířit časovou řadu s delšími časovými intervaly na časovou řadu s kratšími časovými intervaly. Například měsíční časové řady mohou být interpolovány ze čtvrtletních časových řad anebo čtvrtletní časové řady mohou být agregovány z časových řad měsíčních. Procedura EXPAND poskytuje také i mnoho dalších transformací časových řad.

ABSTRACT

Among the most important operations over time series belong mainly the interpolation and aggregation of time series. To convert time series from one time period to another (or the frequency of observations to another) and to interpolate the missing values in a time series, the procedure EXPAND of the SAS statistical system could be used. The EXPAND procedure enable the decomposition of short-term time series to long-term time series or extend the long-term time series to short-term ones. For example, monthly time series may be interpolated from quarterly time series or quarterly time series can be aggregated from the monthly time series. This EXPAND procedure also provides many other transformations of time series.

KLÍČOVÁ SLOVA

agregace, časová řada, interpolace, statistický software SAS

KEY WORDS

aggregation, time series, interpolation, SAS statistical software

1. ÚVOD

Programová procedura EXPAND statistického systému SAS¹ je vcelku málo známá, ale velmi výkonná softwarová funkcionality, která obsahuje mnoho nástrojů pro práce s daty časových řad. Procedura EXPAND umožňuje velmi efektivní postupy při zpracování časových řad, bez nichž by programování nad daty časových řad bylo mnohdy velmi komplexní a složité. Procedura EXPAND je částí softwarového modulu systému SAS, který se nazývá *Econometrics and Time Series (ETS)*². Procedura

¹ *Statistický systém SAS (z angl. Statistical Analysis System) je integrovaný systém softwarových produktů vyráběný americkou firmou SAS Institute, Inc. – poznámka autora.*

² *Softwarový modul SAS/ETS poskytuje nástroje pro širokou škálu aplikací v oblasti komerční, vládní i akademické. Hlavní využití postupů SAS/ETS je ekonomická analýza, prognóza, ekonomické a finanční modelování, analýza časových řad, finanční výkazy a manipulace s daty časových řad. Zpravidla se*

EXPAND není dostupná v jiných softwarových modulech statistického systému SAS. V jiných modulech statistického systému SAS není k dispozici řada nástrojů pro práci s daty časových řad, jako je v proceduře EXPAND. Poprvé byla procedura EXPAND zavedena do systému SAS ve verzi 6.0 a neustále se rozvíjela. V současné době je procedura EXPAND součástí modulu ETS ve verzi 14.2 [2].

2. KONCEPCE FUNKCIONALITY PROCEDURY EXPAND

Znalost klíčových momentů koncepce funkcionality procedury EXPAND, které jsou základem toho, jak systém SAS zpracovává data, čas a tzv. datetime³ proměnné⁴, mají centrální význam pro efektivní využití této procedury. Mezi nejdůležitější operace nad časovými řadami, které zajišťuje procedura EXPAND v rámci statistického programového systému SAS, lze jmenovat:

- agregace časové řady z vyšší na nižší vzorkovací frekvenci⁵, resp. z kratších na delší časové intervaly, ve kterých jsou pozorovány hodnoty časové řady;
- interpolaci hodnot časové řady o vyšší frekvenci pozorování z údajů pozorovaných z časové řady při nižší vzorkovací frekvenci;
- nahrazení chybějících hodnot v časové řadě hodnotami odhadnutými pomocí polynomických (spline) funkcí jako výchozí metodou nebo jinými volitelnými interpolačními metodami dokumentovanými v dokumentaci softwarového modulu ETS ve verzi 14.2 statistického systému SAS [3];
- měnit hodnoty časových řad aplikací operátorů transformace na časové řady, a to buď před, nebo po agregaci, resp. interpolaci vstupních časových řad. Mezi širokou škálou transformačních operátorů, které jsou k dispozici v programovém modulu EXPAND, existují například klouzavé součty, klouzavé průměry, operátory zpoždění, dekompozice časové řady a mnoho dalších typů transformačních operátorů;
- měnit charakteristiky pozorování vstupních řad. Pozorování časových řad lze měřit jako hodnoty na počátku měřeného intervalu, na jeho konci, ve středu intervalu nebo jako intervalové průměry nebo součty. Procedura EXPAND dokáže provádět konverze mezi uvedenými případy. Lze tedy konstruovat odhady intervalových průměrů z pozorování na začátku intervalu, odhady hodnot pozorování ze začátku intervalů nebo jeho středu na základě pozorovaných intervalových průměrů, apod.

Koncepci interpolace a agregace v rámci metod a funkcí procedury EXPAND přibližuje obrázek č. 1. Procedura EXPAND umožňuje odhadovat hodnoty časových řad o delších časových intervalech na základě pozorování časových řad s kratšími časovými intervaly a naopak (tedy agregace a interpolace). Pozorování mohou být sledována od sekundových intervalů až do časových intervalů v trvání několika desetiletí.

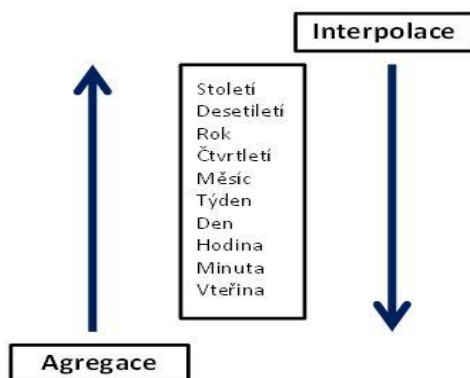
používá tehdy, když zkoumaná data obsahují časové závislosti, souběžné vztahy nebo data jsou výsledkem dynamických procesů (poznámka autora).

³ Jedná se o proměnnou obsahující datum i čas současně – poznámka autora.

⁴ Tyto pojmy jsou podrobně diskutovány v kapitole 4 dokumentu SAS/IETS "SAS/ETS® 14.2 User's Guide, seventh Edition ("Date Intervals, Formats and Functions"), str. 125 – 154.

⁵ Vzorkovací frekvence (dříve vzorkovací kmitočet, anglicky *sampling rate, sample rate* nebo *sampling frequency*) definuje počet vzorků za jednotku času (obvykle za 1 sekundu). Převrácená hodnota vzorkovací frekvence je vzorkovací perioda nebo vzorkovací interval, neboli čas mezi hodnotami řady.

Obrázek č. 1: Agregace a interpolace časové řady pomocí procedury EXPAND



Zdroj: konstrukce podle [7]

K zajištění správného průběhu operací nad časovými řadami pomocí procedury EXPAND je potřebné dodržet 2 důležité zásady kladené na vstupní datové soubory:

- alespoň 1 proměnná v souboru dat musí vykazovat charakter času, data nebo tzv. datetime. Pokud není splněn tento požadavek, procedura EXPAND operaci nad takovými daty nevykoná anebo výsledek operací nad těmito daty nebude správný;
- analyzovaná proměnná, která vystupuje v roli pozorování časové řady, musí být uspořádána ve vzestupném pořadí (od nejnižší hodnoty po nejvyšší) podle časové nebo datové proměnné. Hodnoty časové nebo datové proměnné musí být přítomny pro každý okamžik nebo interval pozorování časové řady. Například mají-li být analyzovány proměnné reprezentující pozorování po dobu 24 měsíců, potom řady musí obsahovat hodnotu datové proměnné pro každý měsíc, a to dokonce, i když 1 nebo více pozorování jsou chybějící.

Výstupem činnosti procedury EXPAND je výstupní soubor dat, ve kterém se nachází transformovaná časová řada. Procedura EXPAND nevytváří ani tištěný ani obrazový výstup.

3. SYNTAX PŘÍKAZU PRO POUŽITÍ PROCEDURY EXPAND V SYSTÉMU SAS

Obecný tvar syntaxe příkazu pro použití procedury EXPAND je následující [4]:

PROCEDURE EXPAND

DATA = vstupující soubor dat **OUT** = vystupující soubor dat

FROM = časový interval **TO** = časový interval

METHOD = konverzní metoda;

BY seskupovací proměnná;

CONVERT původní proměnná = nová proměnná /

OBSERVED = frekvence (charakteristika vstupní řady, charakteristika výstupní řady)

METHOD = transformační metoda

TRANSFORMIN = (transformační operace)

TRANSFORMOUT = (transformační operace);

ID časová (datumová) proměnná;

V uvedeném obecném tvaru příkazu pro využití procedury EXPAND jsou volitelné části příkazu vyznačeny kurzívou. Ostatní parametry příkazu jsou povinné. Přesná syntaxe příkazu včetně popisu je uvedena v dokumentaci systému SAS/ETS [5].

Proceduru volá příkaz PROCEDURE EXPAND. Název vstupního datového souboru identifikuje parametr DATA. Název datového souboru, který obsahuje výsledky činnosti procedury, je udáván v parametru OUT. Parametr FROM udává časový interval pozorování časové řady vstupní, tj. řady, která je vstupní řadou pro operace procedury. Volitelný parametr TO udává časový interval pozorování výstupní časové řady, která je uložena v souboru dat pod názvem v parametru OUT. Není-li parametr TO specifikován, mají pozorování ve výstupní datové sadě stejnou vzorkovací frekvenci jako pozorování ve vstupní datové sadě. Procedura EXPAND podporuje možnost zpracování pozorování po skupinách v datových souborech. Vstupní datový soubor musí být uspořádán podle hodnot proměnné v parametru BY, a současně ve vzestupném pořadí proměnné uvedené v parametru ID.

Příkaz CONVERT identifikuje proměnnou (proměnné) s řadou (řadami) pozorování, nad kterými má procedura EXPAND vykonávat operace, vymezuje požadované operace a popřípadě i název proměnné ve výstupní datové sadě, ve které budou uloženy hodnoty pozorování výstupní časové řady. Pro příkaz CONVERT platí několik obecných pravidel:

- procedura EXPAND může analyzovat pouze numerické proměnné. Vložení znakové proměnné do příkazu CONVERT vygeneruje chybovou zprávu a způsobí, že procedura EXPAND zastaví vykonávání příkazů;
- pro každou analýzu/transformaci časové řady, kterou provádí procedura EXPAND, je vyžadován samostatný příkaz CONVERT;
- do jediné procedury EXPAND lze umístit více příkazů CONVERT;
- do několika příkazů CONVERT ve stejné proceduře EXPAND lze umístit stejnou proměnnou (stejnou časovou řadu). To je užitečné, když je třeba použít různé operátory transformace (viz níže)
- pomocí procedury EXPAND lze analyzovat různé proměnné (různé časové řady) z 1 datového souboru samostatným příkazem CONVERT a
- proměnné ze souboru vstupních dat, které nejsou uvedeny v příkazu BY, CONVERT nebo ID, nejsou kopírovány do souboru vytvořených procedurou EXPAND.

Parametrem ID se identifikuje proměnná ze vstupních dat, která je procedurou EXPAND používána jako index časové řady. Procedura EXPAND předpokládá, že proměnná uvedená v příkazu ID je datová, časová nebo datetime proměnná SAS⁶.

I když vynechání parametru ID v příkazu procedury EXPAND negeneruje chybu, hodnoty pozorování časové řady ve výstupním datovém souboru mohou mít menší smysl. Ve většině situací, kdy je parametr ID v příkazu pro proceduru vynechán, procedura EXPAND předpokládá, že první pozorování výstupní časové řady bylo pořízeno dne 1. ledna 1960. Následující pozorování výstupní časové řady budou poté

⁶ Pozorování časové řady ve vstupním datovém souboru, na kterou má být aplikovány operace procedury EXPAND, musí být uspořádány vzestupně podle pořadí hodnot proměnné umístěné v příkazu ID.

sledována v čase zvyšujícím se o přírůstek definovaný v systémové proměnné INTNX⁷. Hodnoty datové, časové nebo datetime proměnné uvedené v příkazu ID používá procedura EXPAND pro přesné výpočty doby pozorování výstupních časových řad při agregování nebo interpolace dat časových řad.

Pro proceduru EXPAND je výchozím nastavením považovat pozorované hodnoty časové řady na počátku intervalů uvedených v parametrech FROM = a TO =. V reálném světě nemusí být pozorování sledována pouze na počátcích intervalů, což může způsobit velmi odlišné výsledky, zvláště pokud nejsou hodnoty naměřeny na začátku daného intervalu nebo nepředstavují jednu pozorovanou hodnotu pro konkrétní bod v čase. Polohu jednotlivých pozorování časové řady lze v proceduře řídit pomocí parametru OBSERVED v příkazu CONVERT [1]. Pro parametr OBSERVED = existuje 6 možných hodnot, jak je uvedeno v následujícím textu:

BEGINNING	Pozorování na počátku intervalu (mezi pozorováními)
MIDDLE	Pozorování ve středu intervalu (mezi pozorováními)
END	Pozorování na konci intervalu (mezi pozorováními)
TOTAL	Součet hodnot pozorování v intervalu (mezi pozorováními)
AVERAGE	Průměry přes hodnoty pozorování v intervalu (mezi pozorováními)
DERIVATIVE	Numerická derivace vyrovnávající polynomické (spline) funkce ⁸ .

Průměrné hodnoty jsou vyjádřeny v časových jednotkách definovaných volbou FROM = nebo TO =. To znamená, že celková hodnota pro interval se rovná součinu průměrné hodnoty pro interval a šířky intervalu. Pro účely interpolace jsou průměrné hodnoty nejprve převedeny na hodnoty celkové pomocí tohoto předpokladu a pak se interpolované součty převedou zpět na průměry vydělením šířky výstupních intervalů. Pokud jsou vstupní i výstupní řady hodnoty průměrné a nemění se ani jednotky pro časy pozorování, na délce časových intervalů nezáleží. Parametrem OBSERVED = DERIVATIVE se požaduje, aby výstupní řady byly derivace křivky, která vyrovnává pozorování vstupní datové řady metodou kubických polynomických (spline) funkcí.

Způsob, jakým se vstupní časová řada interpoluje na časovou řadu výstupní, je určen pomocí parametru CONVERT [7]. K interpolaci vstupní řady pozorování na výstupní řadu nabízí procedura EXPAND následující metody:

⁷ INTNX je programovací funkce statistického systému SAS, která se používá k operacím s hodnotami data, resp. času. Zvyšuje hodnotu data, resp. času o určitý počet předem specifikovaných intervalů.

⁸ Anglické slovo spline označuje zařízení na kreslení křivek. Při konstrukci trupu lodí a později letadel se používalo pro získání "hladké" křivky, procházející zadanými body, speciální pružné pravítko spline, které se upevnilo v zadaných bodech a svou vlastní pružností se prohnulo tak, že vytvořilo křivku takového tvaru, který vyhovoval konstruktérům; její tvar bylo možno ještě regulovat pomocnými závažími. Při zjednodušení matematické formulace se dospělo k tomu, že taková křivka je matematicky popsána po částech polynomem třetího stupně, který má nespojitosti ve třetí derivaci ve spojích jednotlivých uzlů. Proto dostala název kubický spline. Počátek teorie o splinech tedy spočíval v propojení bodů křivky empiricky zadané několika body v rovině. Přesněji řečeno, je zadáno $k+2$ čísel x_0, \dots, x_{k+1} na ose x a k nim jsou empiricky stanoveny hodnoty $f(x_0), \dots, f(x_{k+1})$ nějaké veličiny. Přirozený požadavek je proložit body $(x_i, f(x_i))$ křivku, která by vyjadřovala (pravděpodobný) průběh oné veličiny na celém intervalu (x_0, x_{k+1}) . Pro tu funkci $f(x)$, kterou máme zkonstruovat, jsou ovšem požadovány některé vlastnosti, které jsou známy o zkoumané veličině, např. spojitost nebo spojitost derivace apod. Tento postup, kterému říkáme interpolace dat, není ovšem úplně zadán. Nejpřirozenější postup je ten, že mezi uzly $(x_i, f(x_i))$ proložíme např. části polynomů a jejich styčné body podrobíme nějaké podmínce. To je princip tzv. interpolačních polynomických (spline) funkcí. Polynomická (spline) funkce stupně 1 je lomená čára, polynomická (spline) funkce stupně 0 je skoková funkce (po částech konstantní).

SPLINE	Vyrovňuje pozorování vstupní časové řady pomocí křivky z kubických polynomických funkcí. Pro okamžiková pozorování (data typu point-in-time) je polynomická křivka omezena procházením daných datových bodů. Pro intervalová nebo průměrná pozorování jsou definované integrály polynomů přes intervaly vstupní časové řady omezeny tak, aby se rovnaly součtům hodnot pozorování v daných intervalech.
JOIN	Vyrovňuje pozorování časové řady propojením po sobě následujících segmentů přímkami. Pro okamžiková pozorování (point-in-time data) spojuje po sobě následující nechybějící vstupní hodnoty přímkami. Pro intervalová nebo průměrná data, jako spojovací body se používají středy intervalů a jejich souřadnice jsou zvoleny tak, aby hodnoty integrálů po částech lineární křivky odpovídaly součtu hodnot pozorování vstupní řady.
STEP	Vyrovňuje pozorování časové řady nespojitou po částech konstantní křivkou (krokovou funkcí). Pro okamžiková vstupní data (typu point-in-time) je výsledná kroková funkce rovna poslední vstupní hodnotě. Pro intervalové celkové nebo průměrné údaje se hodnoty krokové funkce rovnají průměrné hodnotě intervalu.
AGGREGATE	Provádí agregaci časových řad bez interpolace chybějících hodnot.
NONE	Nebude provedena žádná interpolace.

Podrobnější popis je uveden v dokumentaci procedury EXPAND softwarového modulu ETS systému SAS.

4. POUŽITÉ ČASOVÉ ŘADY

Pro účely modelování časových řad pomocí procedury EXPAND byly vybrány ekonomické časové řady, které popisují ekonomické ukazatele produkované státní statistikou. Jedná se o časovou řadu hodnot:

- Dow Jonesův kompozitní index zveřejňovaný na burze Wall Street Journal (označení SNYDJCM) sledovaný denně od 1. ledna 1988 do 5. února 1992;
- peněžního agregátu M1 zahrnující oběživo a jednodenní vklady (označení řady M1), která byla zjišťována s měsíční periodicitou od ledna roku 1980 do ledna roku 1992;
- hrubého domácího produktu (označení řady GDP), který byl sledován se čtvrtletní periodicitou od 1. čtvrtletí roku 1980 po 4. čtvrtletí roku 1998 a
- hodinových počtů průjezdů osobních vozidel na hlavní městské komunikaci (označení řady HOURLY_COUNT), které byly pozorovány od 1. čtvrtletí roku 1980 po 4. čtvrtletí roku 1991.

Vybrané časové řady jsou volně dostupné v knihovně SASHELP statistického systému SAS v datových souborech CITIhour, CITIDAY, CITYMON a CITIQTR. Délka jednotlivých časových řad je určena dostupností údajů v jednotlivých datových souborů knihovny SASHELP. Proměnná identifikující index časové řady v rámci procedury EXPAND je nazvána DATE pro všechny ukázkové příklady.

5. UKÁZKY PRAKTICKÉHO MODELOVÁNÍ ČASOVÝCH ŘAD

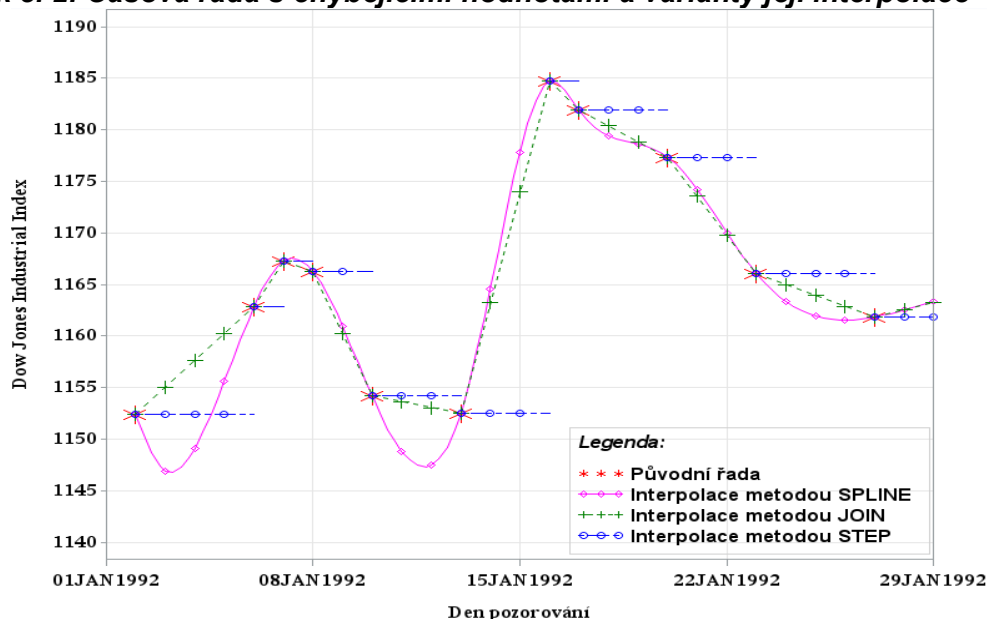
Příklad 1: Použití příkazu procedury EXPAND k interpolaci chybějících hodnot v časové řadě s pozorováními v nerovnoměrných časových intervalech

V časové řadě Dow Jonesova kompozitního indexu zveřejňovaného na burze Wall Street (označení SNYDJCM v datovém souboru CITIDAY v knihovně SASHELP), který byl sledován denně od 1. ledna 1988 do 5. února 1992, byly za měsíc leden 1992 zjištěny chybějící hodnoty. Chybějící hodnoty byly interpolovány pomocí následujícího příkazu:

```
PROCEDURE EXPAND DATA = SASHELP.CITIDAY
  TO = DAY
  OUT = CITIDAY_INTERPOL;
  ID DATE;
  CONVERT SNYDJCM = NEWSNYDJCM;
RUN;
```

Výsledkem působení procedury jsou časové řady, které vznikly různou metodou interpolace. Modelované časové řady vycházející z časové řady na vstupu procedury byly interpolovány hodnotami odhadnutými pomocí polynomických (spline) funkcí, přímkovou interpolací a interpolací po částech konstantní křivkou. Interpolované časové řady (včetně původní neúplné řady pozorování) jsou ilustrovány na obrázku č. 2.

Obrázek č. 2: Časová řada s chybějícími hodnotami a varianty její interpolace



Zdroj: vlastní konstrukce

Příklad 2: Agregování měsíčních pozorování na pozorování čtvrtletní s interpolací chybějících hodnot

Ve vstupní časové řadě peněžního agregátu M1 zahrnující oběživo a jednodenní vklady (označení řady FM1 v datovém souboru CITYMON v knihovně SASHELP), která byla zjišťována s měsíční periodicitou od ledna roku 1980 do ledna roku 1992, byly zjištěny chybějící hodnoty. Agregování vstupní časové řady na řadu se čtvrtletními pozorováními včetně interpolace chybějících hodnot bylo realizováno pomocí následujícího příkazu:

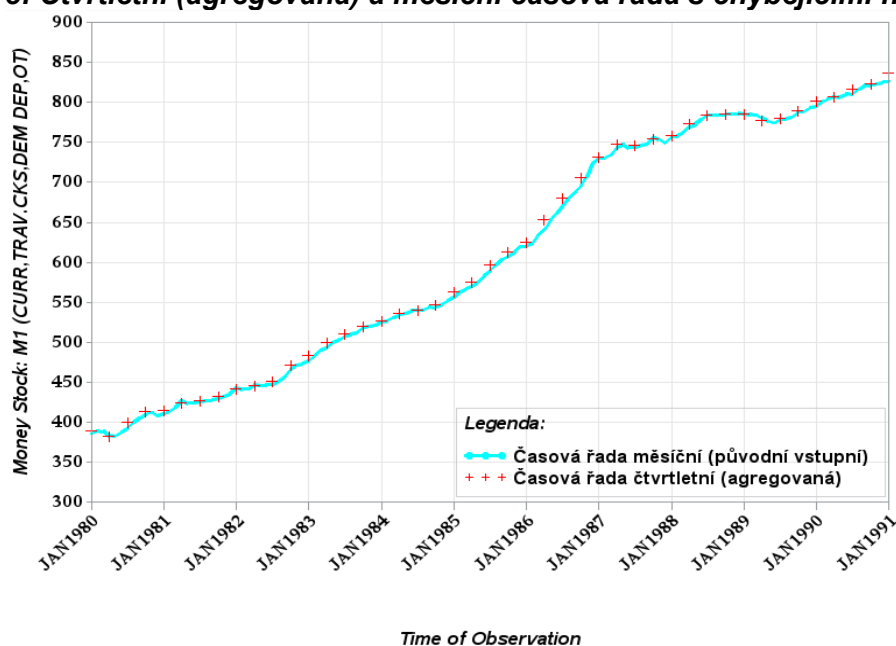

```

PROCEDURE EXPAND
  DATA = SASHELP.CITYMON
  FROM = MONTH TO =QTR
  OUT = QUATERLY;
  ID DATE;
  CONVERT FM1 = FM1_QTR;
RUN;

```

Výsledná časová řada FM1_QTR obsahuje pozorování sledovaná se čtvrtletní periodicitou. Před provedením agregace měsíčních hodnot původní vstupní časové řady byly chybějící hodnoty doplněny pomocí interpolace pomocí polynomických (spline) funkcí. Porovnání agregované čtvrtletní časové řady FM1_QTR a původní měsíční časové řady s chybějícími pozorováními jsou ilustrovány na obrázku č. 3.

Obrázek č. 3: Čtvrtletní (agregovaná) a měsíční časová řada s chybějícími hodnotami



Zdroj: vlastní konstrukce

Příklad 3: Interpolace měsíčních odhadů z časové řady čtvrtletních pozorování

Na základě hodnot časové řady hrubého domácího produktu (označení řady GDP z datového souboru CITIQTR v knihovně SASHELP), které byly pozorovány se čtvrtletní periodicitou od 1. čtvrtletí roku 1980 po 4. čtvrtletí roku 1991, se má pomocí procedury EXPAND odhadnout časová řada měsíčních pozorování hrubého domácího produktu (označená jako GDP_WEEKLY). Interpolace měsíčních odhadů z časové řady čtvrtletních pozorování byla uskutečněna pomocí následujícího příkazu:

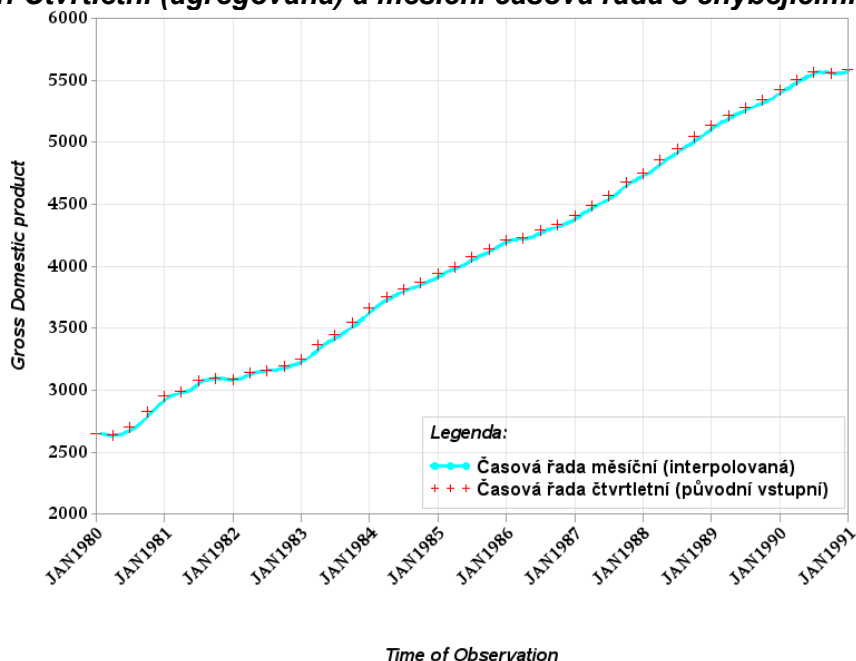
```

PROCEDURE EXPAND
  DATA = SASHELP.CITIQTR
  FROM = QUARTER TO = MONTH
  OUT = MONTHLY;
  ID DATE;
  CONVERT GDP = GDP_MONTHLY / OBSERVED=MIDDLE;
RUN;

```

Výsledná časová řada pozorování hrubého domácího produktu GDP_WEEKLY má týdenní periodicitu zjišťování. Byla použita výchozí metoda interpolace, a to pomocí kubických polynomických funkcí. Parametr OBSERVED v příkazu CONVERT zajistí, že interpolovaná pozorování měsíční časové řady jsou uvažována uprostřed čtvrtletní periody, tedy vždy za každý druhý měsíc čtvrtletí. Porovnání výchozí čtvrtletní časové řady GDP a výsledné časové řady pozorování hrubého domácího produktu GDP_WEEKLY jsou ilustrovány na obrázku č. 4.

Obrázek č. 4: Čtvrtletní (agregovaná) a měsíční časová řada s chybějícími hodnotami



Zdroj: vlastní konstrukce

Příklad 4: Použití transformačních operací v rámci konverze časové řady

Příklad má ilustrovat, jak lze v rámci konverze časové řady (agregace časové řady s hodinovou periodicitou na časovou řadu s týdenní periodicitou) vykonávat transformační operace. Vstupní časová řada je nekompletní, protože se v ní vyskytují i chybějící pozorování. Na základě hodnot časové řady hodinových počtů průjezdů osobních vozidel na hlavní městské komunikaci (označení řady HOURLY_COUNT datového souboru CITIHOURL v knihovně SASHELP), které byly pozorovány od 1. čtvrtletí roku 1980 po 4. čtvrtletí roku 1991, se má pomocí procedury EXPAND odhadnout časová řada týdenních pozorování počtů průjezdů osobních vozidel na hlavní městské komunikaci (označená jako WEEKLY_COUNT) a modifikace této časové řady.

PROCEDURE EXPAND

```

DATA = SASCLASS.CITIHOURL
FROM = OUR TO= WEEK.2
OUT = WEEKLY_COUNT
METHOD = NONE;

ID DATE;
CONVERT HOURLY_COUNT = WEEKLY_COUNT /
TRANSFORMIN = (SETPMISS 0);
CONVERT HOURLY_COUNT = LAG1 /

```

```

      TRANSFORMIN = (SETMISS 0) TRANSFORMOUT = (LAG 1);
CONVERT HOURLY_COUNT = DIF1 /
      TRANSFORMIN = (SETMISS 0) TRANSFORMOUT = (DIF 1);
CONVERT HOURLY_COUNT = MOVESUM3 /
      TRANSFORMIN = (SETMISS 0) TRANSFORMOUT = (MOVSUM 3);
CONVERT HOURLY_COUNT = MOVEAVE3 /
      TRANSFORMIN = (SETMISS 0) TRANSFORMOUT = (MOVAVE 3);
CONVERT HOURLY_COUNT = PCTDIF /
      TRANSFORMIN=(SETMISS 0) TRANSFORMOUT = (PCTDIF 1);
RUN;

```

V příkladu 4 bylo použito více příkazů CONVERT pro výpočet několika klouzavých statistik ze stejné časové řady (stejně proměnné) z datového souboru na vstupu příkazu EXPAND. Parametr TO = WEEK.2 v příkazu EXPAND zabezpečí, že nové časové řady s týdenní periodicitou považují jako začátek týdne pondělí. Parametrem TRANSFORMIN se nejprve vyplní prázdné hodnoty časové řady o denní periodicitě pozorování na hodnotu 0 a následně se provede agregace denních pozorování na časovou řadu pozorování týdenní. Po provedení agregace na časovou řadu s týdenní periodicitou se vykonají operace nastavené parametrem TRANSFORMOUT. Po provedení procedury EXPAND se syntaxí podle Příkladu 4 byly vytvořeny nové proměnné obsahující nové časové řady v souboru dat na výstupu:

WEEKLY_COUNT	Jednoduchý součet hodnot pozorování časové řady od HOUR do WEEK (agregace)
LAG1	Zpětné posunutí jednoduchého součtu hodnot pozorování časové řady od HOUR do WEEK (zpožděná agregace o 1 pozorování)
MOVESUM3	Klouzavý součet jednoduchého součtu hodnot pozorování časové řady od HOUR do WEEK (klouzavé součty agregace o 3 pozorování zpětně)
MOVEAVE3	Klouzavý průměr jednoduchého součtu hodnot pozorování časové řady od HOUR do WEEK (klouzavé průměry agregace o 3 pozorování zpětně)
PCTDIF	Procentuální změna (přírůstek) mezi současným a předcházejícím pozorováním

6. ZÁVĚR

Procedura EXPAND je velmi mocný nástroj, pomocí kterého je možné rychle a snadno provádět potřebné operace na analyzovaných časových řadách před použitím jiných softwarových nástrojů SAS. Široká škála transformačních operátorů zabudovaných do procedury EXPAND ve verzi 6.12 je neustále doplňována i ve verzi 8 a 9. Softwarové nástroje, které procedura EXPAND nabízí uživatelům statistického systému SAS, umožňují velmi efektivně pracovat s daty shromážděnými v časové oblasti.

LITERATURA

- [1] MORGAN, D. P.: The Essential Guide to SAS Dates and Times, Second Edition. Cary, NC: SAS Institute Inc, 2014, xii + 264 s. ISBN 978-1-62959-491-0, s. 206.
 [2] SAS Institute Inc. 2016. SAS/ETS® 14.2 User's Guide. Seventh Edition. Cary, NC: SAS Institute Inc., 2016, x + 4149 s. ISBN (on-line) s. 861 – 910.

- [3] SAS Institute Inc. 2016. SAS/ETS® 14.2 User's Guide. Seventh Edition. Cary, NC: SAS Institute Inc., 2016, x + 4149 s. ISBN (on-line) s. 879 – 882.
- [4] SAS Institute Inc. 2016. SAS/ETS® 14.2 User's Guide. Seventh Edition. Cary, NC: SAS Institute Inc., 2016, x + 4149 s. ISBN (on-line) s. 861 – 910.
- [5] SAS Institute Inc. 2016. SAS/ETS® 14.2 User's Guide. Seventh Edition. Cary, NC: SAS Institute Inc., 2016, x + 4149 s. ISBN (on-line) s. 871 – 876.
- [6] SAS Institute Inc. 2016. SAS/ETS® 14.2 User's Guide. Seventh Edition. Cary, NC: SAS Institute Inc., 2016, x + 4149 s. ISBN (on-line) s. 881 - 884.
- [7] Using proc EXPAND with Time Series Data. In: SAS Conference Proceedings: Western Users of SAS Software 1996, 1996, San Francisco, California. 276 – 279 s. [online]. [cit. 25. 08. 2017]. <<http://www.lexjansen.com/wuss/1996/WUSS96073.pdf>>, s. 276.

RESUME

Time series analysis is an important statistical approach to a systematic assessment of time-related dynamics of variables or time-related relationships among several variables. Among various procedures in SAS/ETS, the PROC EXPAND allows us to transform time series data so that the data can be used for various univariate and multivariate time series analyses. In particular, the PROC EXPAND has a very useful feature that can transform time series data to different frequencies. For example, the monthly time series data can be transformed to higher frequency data, e.g., daily or weekly data, to lower frequency data, e.g., quarterly or annual data. Various multivariate time series approaches, such as transfer function modeling, econometrical modeling (vector autoregression) etc., assume that all the variables in the model are based on the same frequency. For these cases, the PROC EXPAND is an extremely useful tool.

PROFESNÍ ŽIVOTOPIS

Ing. Roman Pavelka, PhD., v letech 1995–2010 pracoval v poradenské společnosti Trexima, s. r. o. Na pozici statistik-analytik se zabýval analýzami zejména mzdových a personálních dat. Podílel se na tvorbě pravidelných statistických přehledů a reportů. Spolupracoval s akademickými pracovišti, agenturami i soukromými subjekty na realizaci a vyhodnocování ad-hoc statistických výzkumů. Oblast jeho vědeckého zájmu představují výběrová šetření, odhady a statistické modely. V letech 2012 až 2013 se zúčastnil zahraniční stáže ve Velké Británii. Od roku 2013 působil v Národnom ústave certifikovaných meraní vzdelávania (NÚCEM), kde zajišťoval statistické vyhodnocování výsledků testování žáků a studentů. Od roku 2015 pracuje v odboru metod statistických zjišťování Štatistického úradu SR.

KONTAKT

Roman.Pavelka@statistics.sk