

SOUDOBÉ POSTUPY PROGNÓZOVÁNÍ KRÁTKÝCH ČASOVÝCH ŘAD UKAZATELŮ AGRÁRNÍHO SEKTORU

RECENT APPROACHES TO FORECASTING SHORT TIME SERIES OF AGRICULTURAL INDICATORS

Bohumil Kába

Anotace:

V ekonomické praxi je často nutné analyzovat časové řady s malým počtem pozorování, které mnohdy vykazují velmi nepravidelný průběh. Zpracování takových řad je obtížné a standardní analytické postupy zde zpravidla selhávají. V daném příspěvku jsou prezentovány některé postupy, které statistická metodologie nabízí pro modelování a extrapolaci krátkých a nestabilních časových řad. Efektivnost těchto procedur a relace mezi jejich interpolačními a extrapolačními vlastnostmi jsou dokumentovány výsledky empirické srovnávací studie, realizované na souboru vybraných zemědělských ukazatelů kandidátských zemí EU.

Summary:

Some economic time series are very short and irregular. In such situations, the identification and selection of an adequate forecasting model can be very difficult. This paper describes some recent methods and models for short time series forecasting. The results of a comparative study implemented on a set of selected agricultural indicators in candidate countries of the European Union are given.

Key words:

time series, forecasting model, forecast, combined forecasts

Klíčová slova:

časová řada, prognostický model, předpověď, kombinované předpovědi

Úvod

Pro tvorbu předpovědí v jednorozměrných časových řadách je nutné mít k dispozici dostatečné množství relevantních údajů v žádoucí kvalitě a struktuře. Délka disponibilní časové řady je fundamentálním kriteriem výběru adekvátního prognostického modelu, neboť různé modely se liší ve svých požadavcích na minimální potřebnou délku řady. Mnohé časové řady ukazatelů agrárního sektoru bývají poměrně krátké a nestabilní a jejich modelování respektive prognózování pomocí standardních modelů (např. klasických trendových modelů nebo Boxových – Jenkinsových modelů) může být velmi problematické. V daném příspěvku budou uvedeny některé flexibilní prognostické techniky, doporučované pro časové řady s malým počtem pozorování a jejich efektivnost bude ilustrována výsledky empirické komparativní studie.

Cíl a metodika

Cílem dané studie byla identifikace modelů, vhodných pro tvorbu předpovědí v krátkých časových řadách ukazatelů agrárního sektoru (obsahujících cca 10 pozorování) a posouzení prognostické efektivity těchto modelů. Pozornost byla věnována zejména modelu náhodné procházky (DeLurgio 1998) a vybraným typům modelů exponenciálního vyrovnávání (Cipra 1986), které kladou velmi malé nároky na délku analyzované řady. Použitelnost uvedených modelů pro modelování a extrapolaci reálných časových řad byla empiricky ověřována na souboru ročních časových řad celkové produkce zrnin, brambor, cukrovky, zeleniny, ovoce, hovězího masa, mléka, vepřového masa a drůbeže. Tyto řady se týkaly vybraných kandidátských zemí Evropské unie (České republiky, Estonska, Lotyšska, Litvy a Slovenska) a zahrnovaly období 1992 – 2002 respektive 1993 – 2002. Disponibilní řady tedy měly 10 respektive 11 pozorování a jejich stochastická struktura byla poměrně komplikovaná (60 % řad bylo nestacionárních, u 42 % řad byla identifikována odlehlá respektive extrémní pozorování).

Pro analýzu stacionárních časových řad (respektive řad s indukovanou stacionaritou) byl použit Brownův model jednoduchého exponenciálního vyrovnávání. U nestacionárních řad (s různými druhy lokálních trendů) byly aplikovány Brownův model dvojitého exponenciálního vyrovnávání, Holtův model exponenciálního vyrovnávání, exponenciální vyrovnávání s tlumeným trendem a model náhodné procházky s posunem (SAS/ETS User's Guide 1995).

Pomocí všech modelů byly konstruovány předpovědi s tříčlenným horizontem. Prognostická kvalita modelů byla retrospektivně posuzována prostřednictvím střední absolutní procentuální chyby MAPE (DeLurgio 1998).

Výsledky

Rozbor předpovědních chyb MAPE ukázal, že při tvorbě extrapolačních předpovědí krátkých časových řad ukazatelů rostlinné a živočišné produkce lze nejlepší výsledky přisoudit zejména jednoduchému exponenciálnímu vyrovnávání a modelu náhodné procházky. S jistým odstupem byly tyto modely následovány modelem dvojitého exponenciálního vyrovnávání a modelem exponenciálního vyrovnávání s tlumeným trendem, jejichž předpovědní chyby byly obvykle velmi podobné a na akceptovatelné úrovni. Méně úspěšný byl Holtův model, který na zlomy trendu reagoval s dosti velkým zpožděním.

Vybrané výsledky, které charakterizují úspěšnost modelů ve fázi extrapolace, jsou shrnuty v tabulce 1. Tato tabulka prezentuje nejlepší identifikovaný předpovědní model pro každou kandidátskou zemi EU, která byla zařazena do analýzy a souhrnné statistické charakteristiky jeho předpovědních chyb MAPE, jmenovitě aritmetický průměr, medián, dolní a horní kvartil.

Tabulka1. Charakteristiky hodnot MAPE pro nejlepší individuální modely

Země	Model	Průměr	Medián	Dolní kvartil	Horní kvartil
Česká republika	B1	10,76	8,96	5,93	14,31
Estonsko	B1	9,73	10,30	3,62	17,63
Lotyšsko	B1	10,34	11,49	4,63	13,24
Litva	B1	10,56	6,97	3,92	13,12
Slovensko	RW	15,04	13,18	8,39	23,09

B1 – jednoduché exponenciální vyrovnávání

RW – model náhodné procházky

Při tvorbě extrapolačních předpovědí bylo obvykle možné identifikovat větší počet adekvátních modelů, které poskytovaly pro tytéž řady různé předpovědi. Lze přepokládat, že každá metoda obsahuje potenciálně užitečnou informaci a prezentuje různé dílčí aspekty globálního analyzovaného procesu. Jeví se tedy účelným vytvářet tzv. kombinované předpovědi, které agregují informaci z jednotlivých metod. V této studii byly kombinované předpovědi konstruovány ve formě prostého aritmetického průměru individuálních předpovědí a byly uskutečněny experimenty s různými variantami kombinovaných předpovědí (různé počty předpovědí, kombinace různých typů modelů).

Souhrnné výsledky ukazatele MAPE kombinovaných modelů jsou prezentovány v tabulce 2, kde C2 respektive C3 jsou kombinované předpovědi dvou respektive tří nejlepších individuálních modelů (tzn. modelů, které u dané řady poskytly předpovědi s nejnižšími relativními chybami). Tabulka uvádí pro každý soubor řad sledovaných kandidátských zemí EU nejlepší nalezený kombinovaný model a jeho odpovídající souhrnné charakteristiky MAPE.

Tabulka 2. Charakteristiky hodnot MAPE pro nejlepší kombinované modely

Země	Model	Průměr	Medián	Dolní kvartil	Horní kvartil
Česká republika	C3	6,47	7,41	3,22	9,98
Estonsko	C2	7,01	6,20	4,19	10,17
Lotyšsko	C2	7,75	7,23	4,80	9,43
Litva	C2	6,98	6,01	3,04	10,72
Slovensko	C2	10,44	8,12	6,39	15,31

Diskuse

Předpovědi, které jsou zatíženy předpovědními chybami MAPE menšími než 15 %, bývají konvenčně považovány za velmi kvalitní. Prezentované výsledky ukázaly, že u všech dílčích souborů řad jednotlivých sledovaných zemí tomuto kritériu vyhovovaly zejména předpovědi, získané pomocí jednoduchého exponenciálního vyrovnávání a modelu náhodné procházky. Tyto modely lze pokládat za robustní prognostické nástroje, které u většiny časových řad analyzovaného typu obvykle poskytují kvalitní předpovědi i když v individuálních případech mohou být předstiženy efektivnějšími postupy tvorby extrapolačních předpovědí.

Výrazného zlepšení kvality předpovědí se dosáhlo kombinováním individuálních předpovědí. Z výsledků je patrné, že kombinované modely předčily nejlepší individuální modely prakticky ve všech sledovaných charakteristikách. V neúspěšnějších kombinacích byly nejčastěji zastoupeny modely jednoduchého exponenciálního vyrovnávání, náhodné procházky a exponenciálního vyrovnávání s tlumeným trendem. Pokud jde o počet modelů zařazených do nejlepších kombinací, byla nejvíce frekventována kombinace dvou individuálních modelů, v souboru řad vztahujících se k České republice se osvědčila kombinace tří individuálních modelů.

U velmi krátkých řad nelze prognostickou efektivnost modelu posuzovat pomocí retrospektivních hodnot MAPE. V takových situacích se doporučuje užívat pro evaluaci modelu celé referenční období řady. Proto je účelné zkoumat relace mezi interpolačními a extrapolačními charakteristikami analyzovaných modelů. Z dané studie vyplynulo, že pro nejlepší výše zmíněné individuální modely lze usuzovat na středně silnou pozitivní korelaci

mezi interpolačními a předpovědními chybami. Z chování modelu ve fázi interpolace je tedy možné orientačně odhadovat jeho chování ve fázi extrapolace.

Závěry

Výsledky rozsáhlé empirické analýzy ukázaly, že i pro krátké a nestabilní časové řady mohou být identifikovány velmi kvalitní prognostické modely. Pro extrapolace ročních časových řad ukazatelů rostlinné a živočišné produkce pěti uvažovaných kandidátských zemí EU lze doporučit zvláště modely s "krátkou pamětí" – jednoduché exponenciální vyrovnávání a model náhodné procházky, které při tvorbě předpovědí využívají zejména nejnovějších pozorování časové řady. Tyto modely prokázaly dostatečnou flexibilitu, statistickou robustnost i schopnost přizpůsobovat se případným strukturálním změnám. Je však třeba poznamenat, že předpovědní efektivnost uvedených procedur je podmíněna dodržáním všech zásad výstavby prognostických modelů a respektováním vzájemné provázanosti etap identifikace modelu, odhadu jeho parametrů i konečného ověření zkonstruovaného modelu pomocí adekvátních diagnostických metod.

V reálných situacích obvykle nelze exaktně identifikovat "optimální" prognostický postup. Je tedy účelné doplnit předpovědi z individuálních modelů tvorbou kombinovaných předpovědí. Tato prognostická procedura často zvyšuje přesnost předpovědi a redukuje variabilitu předpovědi. Jak naznačily prezentované výsledky, postačuje zpravidla kombinovat dva případně tři nejlepší individuální modely.

Literatura

Cipra, T. (1986): Analýza časových řad s aplikacemi v ekonomii, SNTL/ALFA, Praha
DeLurgio, S. A. (1998): Forecasting Principles and Applications, Irwin/McGraw-Hill, Boston
SAS/ETS User's Guide (1995), Version 6, SAS Institute Inc.

Kontaktní adresa:

Doc. RNDr. Bohumil Kába, CSc.
Katedra statistiky PEF ČZU
165 21 Praha 6 – Suchbátka
tel.: 224 382 236
e-mail: kaba@pef.czu.cz