

Vybrané soudobé postupy analýzy časových řad zemědělských ukazatelů

Selected modern methods of agricultural time series analysis

Bohumil Kába

Annotation: The paper deals with comparative analysis of univariate models of agricultural time series. The main attention is focused in specification of separate forecasting models, examination the accuracy of various forecasting methods and aggregating information by combining forecasts from two or more forecasting methods. The explanation is supplemented by the selected numeric results of the extensive empirical study.

Anotace: Příspěvek shrnuje výsledky komparační analýzy jednorozměrných modelů časových řad zemědělských ukazatelů. Hlavní pozornost byla soustředěna na specifikaci jednotlivých modelů, porovnání přesnosti stanovených prognóz, posouzení relací mezi interpolačními a extrapolacími vlastnostmi zkoumaných modelů a tvorbu agregovaných předpovědních modelů. Výklad je doplněn výsledky rozsáhlé empirické studie.

Key words: time series, forecasting model, accuracy of forecasts, combined forecasts, statistical package

Klíčová slova: časová řada, předpovědní model, přesnost předpovědí, kombinované předpovědi, statistický paket

Úvod

Soudobá metodologie analýzy a prognózování časových řad zemědělských ukazatelů se zejména v několika posledních desetiletích velmi intenzivně rozvíjí a doplňuje jak o nové, často velmi sofistikované postupy, tak o modifikace stávajících technik. Praktická využitelnost těchto postupů se musí opakovaně a systematicky verifikovat a zkoumat na reálných časových řadách, neboť modely těchto řad nemají univerzální ani trvalý charakter a musí být konstruovány v souladu s vlastnostmi studovaných ukazatelů. Dosavadní empirické studie([2], [3]), orientované na prozkoumání vlastností reálných časových řad zemědělských ukazatelů České republiky a na posouzení použitelnosti adekvátních analytických respektive prognostických modelů se týkaly relativně malého počtu řad určitého typu i omezeného počtu modelů. Je tedy účelné pohlížet na ně jako na jisté pilotní studie a doplnit je podstatně

obsáhlejší analýzou. Daný příspěvek stručně prezentuje některé výsledky, jež byly získány aplikací jednorozměrných modelů časových řad na souboru 360 řad z oblasti agropotravinářského komplexu České republiky. Daný soubor obsahoval 160 časových řad ukazatelů zjišťovaných s roční periodicitou, 100 řad ukazatelů zjišťovaných se čtvrtletní periodicitou a 100 řad ukazatelů zjišťovaných s měsíční periodicitou.

Cíl a metodika

V úvodu zmíněné dílčí studie naznačily, že zemědělské časové řady mají obvykle značně komplikovanou pravděpodobnostní strukturu, která velmi ztěžuje tvorbu extrapolačních předpovědí. Ukázaly rovněž, že při prognózování časových řad zemědělských ukazatelů se mohou úspěšně uplatnit některé adaptivní modely exponenciálního vyrovnávání a technika tzv. kombinovaných předpovědí. Ta je založena na agregování individuálních předpovědí, poskytnutých jednotlivými prognostickými metodami. Cílem prezentované studie je verifikace respektive korektura a doplnění těchto předběžných závěrů. Hlavní důraz byl kladen na prozkoumání vlastností kombinovaných modelů, zejména posouzení, které metody a jaký počet individuálních prognostických technik kombinovat, jakým způsobem je agregovat a jaké je aplikační využití kombinovaných předpovědí.

Naplnění uvedených cílů vyžadovalo, aby do zkoumání byly zahrnuty velmi různorodé časové řady, lišící se délkou referenčního období i počátky předpovědí. Pro analýzu a prognózování časových řad tohoto typu je třeba disponovat velmi širokým okruhem modelů. Z těchto důvodů byl pro zmíněné zpracování zvolen programový systém SAS, který je pokládán za jeden z nejdokonalejších a nejobsažnějších statistických programových paketů. Jeho komponenta TSFS (Time Series Forecasting System) má ve své základní nabídce několik desítek jednorozměrných modelů časových řad – zejména širokou škálu adaptivních modelů exponenciálního vyrovnávání, Box – Jenkinsovy modely, ale též vybrané analytické trendové modely a triviální modely, konstruuující předpovědi pomocí aritmetického průměru originálních hodnot respektive aritmetického průměru logaritmů originálních hodnot.

Komponenta TSFS umožňuje konstruovat nejen jednotlivé prognostické modely, ale též různé kombinované předpovědi. V dané studii bylo experimentováno s různými variantami kombinovaných předpovědí a na základě zjištěných výsledků bylo rozhodnuto vytvářet agregované předpovědi vždy ze všech metod, specifikovaných pro příslušnou řadu a dále z různého počtu nejlepších (ve smyslu vhodného selekčního kritéria) předpovědí.

Výsledky

Disponibilní časové řady byly nejprve vyrovnány pomocí modelů, implementovaných v komponentě TSFS systému SAS. Protože by soubor konstruovaných modelů byl neúměrně rozsáhlý, byl jeho rozsah redukován pomocí speciálních diagnostických testů ([4]), jež z nabízených modelů vybírají pouze ty, které jsou vhodné pro analytické zpracování dané řady a generování předpovědí. V dané studii byly nejčastěji vybrány tyto modely:

- Brownův model dvojitého exponenciálního vyrovnávání (BE)
- Holtův dvouparametrický model exponenciálního vyrovnávání (HE)
- model exponenciálního vyrovnávání s tlumeným trendem (EDT)
- model náhodné procházky (RW)

a dále dva analytické trendové modely – lineární trendový model (LT) a exponenciální trendový model (ET). U sezónních časových řad se též dílčím způsobem uplatnily Wintersovy modely exponenciálního vyrovnávání a Box – Jenkinsovy modely. Vzhledem k jejich malému počtu však výsledky těchto modelů nejsou prezentovány.

Všechny odhadnuté modely pak byly využity pro tvorbu předpovědí s tříletným časovým horizontem. Vzhledem k různorodosti analyzovaných řad byla pro komparaci interpolačních i extrapolacních vlastností modelů použita střední absolutní procentuální chyba (Mean Absolute Percentage Error = MAPE).

Nejlepší interpolační výsledky byly dosaženy aplikací modelu exponenciálního vyrovnávání s tlumeným trendem. Při vyrovnání časových řad s roční periodicitou získávání údajů činila průměrná hodnota MAPE pro tento model 8,21 %, při vyrovnání čtvrtletních respektive měsíčních řad pak tyto průměrné hodnoty byly 3,57 % respektive 2,36 %.

Při tvorbě předpovědí se z individuálních modelů velmi osvědčil model náhodné procházky a modely exponenciálního vyrovnávání (zejména Brownův model dvojitého exponenciálního vyrovnávání a model exponenciálního vyrovnávání s tlumeným trendem). Kombinované předpovědi byly vytvářeny ve dvou verzích – ve formě prostého aritmetického průměru individuálních předpovědí a ve formě váženého aritmetického průměru se systémem tzv. regresních vah. U všech typů řad se jako efektivnější varianta, poskytující přesnější předpovědi, jevila metoda prostého aritmetického průměru.

Vybrané výsledky, jež charakterizují úspěšnost modelů ve fázi extrapolace, jsou shrnuty v následující tabulce. Tato tabulka uvádí typ řady, výsledky nejlepšího individuálního modelu a pro srovnání též výsledky nejlepšího analytického trendového modelu a dále výsledky nejlepšího kombinovaného modelu.

Tabulka 1

Souhrnné charakteristiky ukazatele MAPE

Typ řady	model	průměr	medián	IQR
Roční	RW	7,40	5,35	5,10
	LT	24,95	18,94	22,92
	C2	4,55	3,68	3,55
Čtvrtletní	RW	3,95	2,66	3,01
	LT	8,28	5,70	7,37
	C2	2,16	1,63	1,49
Měsíční	BE	4,79	2,51	2,73
	LT	13,14	11,00	11,05
	C2	1,67	1,12	2,24

(Symbolem C2 je v této tabulce označena kombinace dvou nejlepších individuálních prognostických modelů ve formě prostého aritmetického průměru, IQR znamená interkvartilové rozpětí, tzn. robustní charakteristiku variability)

Diskuse

Na základě dosažených výsledků lze konstatovat, že tvorbu kvalitních předpovědí budoucího vývoje časových řad mnoha důležitých zemědělských ukazatelů lze zakládat na relativně jednoduchých modelech časových řad. Složitě strukturované velké množství zemědělských časových řad, které často vykazují velkou variabilitu, zlomy trendu případně odlehlá pozorování a o jejichž průběhu máme pouze omezenou informaci, zřejmě velmi dobře odpovídá naivní model náhodné procházky. Tento model lze doporučit především při tvorbě předpovědí ročních respektive čtvrtletních časových řad. Kvalitou extrapolacních vlastností se tomuto modelu blíží i modely exponenciálního vyrovnávání, zejména model exponenciálního vyrovnávání s tlumeným trendem a Brownův model dvojitého exponenciálního vyrovnávání. Velmi neúspěšné při tvorbě extrapolacních předpovědí byly – v reálné praktické činnosti velmi frekventované – analytické trendové modely.

Významného zlepšení kvality předpovědí se u všech typů analyzovaných řad podařilo dosáhnout aplikací kombinovaných předpovědí. Ukázalo se, že technika agregovaných předpovědí ve srovnání s nejlepšími individuálními prognostickými modely snížila předpovědní chyby v průměru o cca 10 – 11 %. V tabulce 1 jsou prezentovány výsledky, dosažené kombinací dvou nejlepších individuálních modelů. Velmi podobné výsledky přinesly i kombinace tří, čtyř respektive pěti nejlepších prognostických modelů (kombinace

většího počtu individuálních modelů již vedly k významně vyšším předpovědním chybám). S ohledem na princip parsimonie modelu je tedy zřejmě možné doporučit tvorbu kombinovaných předpovědí pouze na základě dvou nejlepších individuálních modelů.

Ve studii byla rovněž věnována pozornost relaci mezi interpolačními a extrapolačními vlastnostmi studovaných modelů. K tomuto účelu byly pro všechny disponibilní časové řady vypočteny koeficienty korelace pořadí modelů při interpolaci a extrapolaci a vyhodnoceny. Průměr těchto koeficientů pro roční řady činil 0,196, pro čtvrtletní řady 0,179 a pro měsíční řady 0,470. Odpovídající hodnoty mediánů byly 0,288, 0,345 a 0,600. Lze tedy konstatovat, že z chování modelu ve fázi interpolace lze pouze velmi orientačně usuzovat na chování ve fázi extrapolace a tedy na kvalitu předpovědí na tomto modelu založených.

Závěr

Výsledky dané studie jsou do značné míry konsistentní s pracovními hypotézami, naznačenými v metodické části příspěvku. Vzhledem ke značnému rozsahu analyzovaného souboru časových řad lze validitu závěrů, které pro tvorbu extrapolačních předpovědí budoucího vývoje zemědělských ukazatelů doporučují zejména užití modelů exponenciálního vyrovnávání respektive modelu náhodné procházky, pokládat za vysokou. Významným závěrem obecného charakteru je rovněž konstatování, že technikou kombinování předpovědí lze výrazně snížit riziko předpovědní chyby a omezit variabilitu těchto chyb. Pro agregaci předpovědí lze doporučit prostý aritmetický průměr individuálních předpovědí a jeho výpočet postačí založit na dvou nejlepších individuálních prognostických modelech.

Literatura:

1. Forecasting Examples for Business and Economics Using the SAS System, SAS Institute, Inc., Cary, NC, USA, 1996
2. Kába, B.: Konstrukce kombinovaných předpovědí časových řad ekonomických ukazatelů, In: Zborník vedeckých prác z MVD 97, Nitra, 1997
3. Kába, B., Svatošová, L.: Prognostické modely spotřeby potravin, In: Zborník vedeckých prác z MVD 2000, Nitra, 2000
4. SAS/ETS User's Guide, Version 6, SAS Institute, Inc., Cary, NC, USA, 1993