

ZHODNOCENÍ MODELOVÁNÍ VYBRANÝCH ČASOVÝCH ŘAD NA AGRÁRNÍM TRHU V SYSTÉMU SAS

EVALUATION OF THE SELECTED TIME SERIES MODELLING OF AGRARIAN MARKET IN SAS SYSTEM

Zdeněk Louda

Anotace:

Systém SAS je jedním z nejrozšířenějších systémů pro statistickou analýzu dat na světě, oblast analýzy časových řad je jedním z nejužívanějších statistických nástrojů. Příspěvek by měl seznámit s výsledky analýz a modelování vývoje vybraných cen na agrárním trhu za použití tohoto systému a nastínit tudíž možnosti pro analýzu a prognózování časových řad v systému SAS.

Klíčová slova:

systém SAS, analýza časových řad, vývoj cen, model, agrární trh.

Abstract:

System SAS is one of the most popular systems in the world for statistical data analysis, time series analysis is the most widely used among the statistical tools. Report should acquaint with results of modelling the trend in selected prices on agrarian market with utilization of this system and hence outline possibilities for analysis and forecasting of time series in system SAS.

Keywords:

System SAS, time series analysis, trend in prices, model, agrarian market.

ÚVOD

Snaha zkoumat minulý vývoj nejrozličnějších společenských jevů, poučit se z historie a z daných zkušeností a poznatků těžit v budoucnosti, jsou známy již od pradávna. V současné společnosti je tento letitý princip často reprezentován analýzou časových řad a jejich prognózováním do budoucna a tato oblast statistických analýz je jedním z nejužívanějších statistických nástrojů v nejrozličnějších sférách lidského života. Během několika posledních desetiletí se chopily statistických analýz softwarové firmy a neustálým iterativním procesem zlepšování a inovací vytvořily produkty, z nichž jedním z nejrozšířenějších je dnes systém SAS.

CÍL A METODIKA

Cílem příspěvku je zmapovat a ověřit možnosti systému SAS v oblasti modelování a analýz časových řad, neboť analýza časových řad bude jednou ze stěžejních oblastí statistických analýz v disertační práci s názvem „Návrh a ověření metod statistické analýzy pro podnikatelské a marketingové rozhodování na trhu s agrárními produkty“.

Na konkrétním příkladu časové řady (měsíční ceny zem. výrobců tříděných vajec konzumních (v Kč/ks) v období leden 1993 - říjen 2003) by měla být nastíněna procedura

tvorby nejvhodnějšího modelu vývoje zmíněné řady (a případné prognózování jejího vývoje), a dále by měl být systém SAS využit pro analýzu dalších vybraných cen agrárního trhu.

Analyzovaná data vybraných měsíčních časových řad pocházejí ze Situačních a výhledových zpráv – Brambory, Cukrovka a cukr, Drůbež, Hovězí maso, Mléko, Obiloviny, Olejníky, Ovoce, Vepřové maso a Zelenina vydaných Ministerstvem zemědělství České republiky (viz. Literatura). Příspěvek předpokládá teoretické znalosti pokročilých modelů časových řad.

Prostředí systému SAS pro analýzu časových řad

Velkou uživatelskou výhodou statistického prostředí modulu ETS programového balíku SAS je, že uživateli prakticky stačí pouze zadat vstupní data řádně připravená pro počítačové procedury. Uživatel může ještě subjektivně rozdělit řadu na testovací a kontrolní část, ale jinak veškerá výpočtová složitost statistických postupů odpadá, a uživatel se musí probrat jen značným množstvím výstupů, kterou si však může podle svého uvážení „korigovat“.

Modul ETS nabízí možnost diagnostikovat časovou řadu z hlediska přítomnosti trendu, sezónnosti a nutnosti logaritmické transformace dat. Uživatel tedy nemusí pracně vybírat ze všech modelů nabízených programem, ale pouze z těch, které vyhovují zjištěné diagnóze.

Možnostmi hodnocení kvality vytvořených modelů jsou míry hodnocení kvality prognóz „ex post“, graf reziduí, grafy ACF a PACF a otestování významnosti jednotlivých parametrů modelu. Další možností při modelování a prognózování je kombinace vypočtených modelů, ovšem kombinace více jak dvou modelů se nedoporučuje, neboť s rostoucím počtem parametrů těchto modelů klesá interpretovatelnost výsledného kombinovaného modelu. Pro prognostické využití vytvořených modelů nabízí tento modul graf vytvořeného modelu včetně prognóz na určitý počet období do budoucnosti (uživatelem předem stanovených) a tomu odpovídající tabulku prognóz, vše včetně pásu spolehlivosti.

VÝSLEDKY

Nejprve je poměrně detailně nastíněna procedura nalezení nejvhodnějšího modelu vybrané časové řady, v dalším textu jsou pak vyhodnoceny výsledky modelování všech vybraných časových řad trhu s ovocem a zeleninou.

Postup tvorby modelu vybrané časové řady v modulu SAS/ETS

Prvním krokem analýzy časové řady v systému SAS je vizuální posouzení dané řady. K demonstraci postupu byla vybrána časová řada CZV tříděných vajec konzumních (v Kč/ks) v období leden 1993 - říjen 2003 včetně prognózy vývoje na dalších 12 měsících.

Vybraná časová řada již na základě vizuálního posouzení vykazuje značnou sezónnost (body znázorňující původní naměřené hodnoty v levé části výsledného modelu v Grafu č. 2), která je prokázána i zmiňovanou diagnostikou časové řady, stejně jako nepřítomnost trendu v časové řadě a nerelevantnost logaritmické transformace dat. Na základě této diagnózy jsou vybrány 4 vyhovující modely (viz. Obrázek č. 1), z nichž by se mělo podařit vybrat relativně nejlepší model.

Obrázek č. 1: Modely vyhovující diagnostice časové řady a jejich MAPE

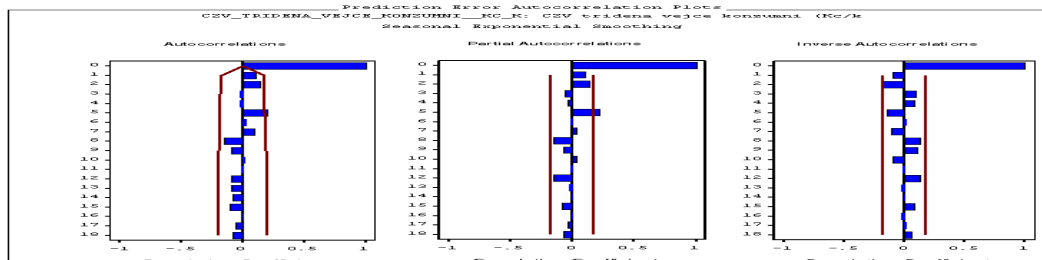
Forecast		Mean Absolute Percent Error
Model	Model Title	
☒	Seasonal Exponential Smoothing	3.07994
☐	ARIMA(2,0,0)(1,0,0)s	3.73749
☐	Seasonal Dummy	12.46666
☐	ARIMA(0,1,1)s NOINT	13.01845

Daný soubor „diagnostice časové řady vyhovujících“ modelů se z hlediska hodnocení

kvality modelu hodnotí pomocí klasických měř hodnocení modelů, nejvhodnější a nejčastěji používanou je pak hodnota střední absolutní procentuální chyby MAPE.

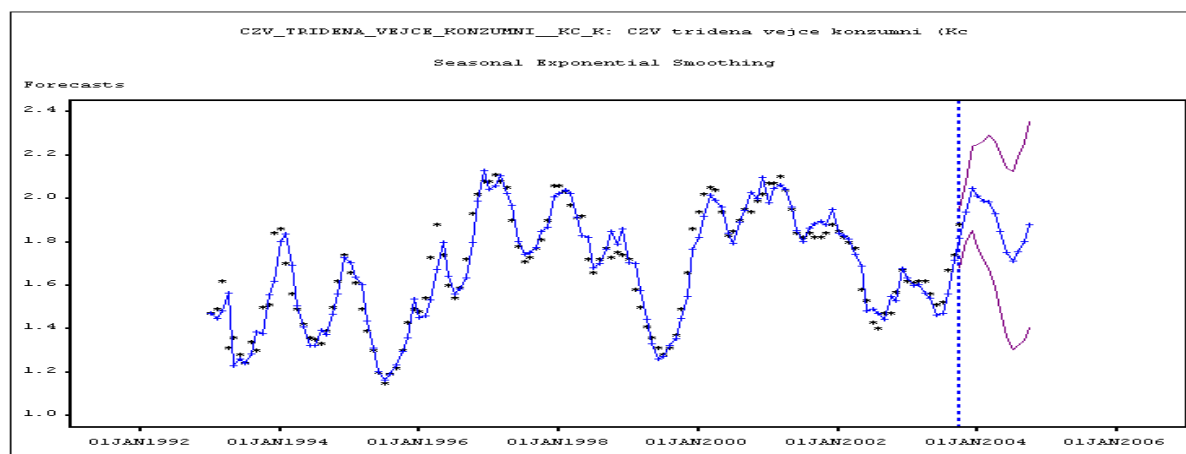
Dalším zmiňovaným hodnotícím kritériem kvality vytvořených modelů je analýza autokorelačních a parciálních autokorelačních funkcí modelů. Tomuto kritériu ze všech čtyř modelů vyhovuje pouze jeden, a to model Seasonal Exponential Smoothing – viz. Graf č. 1.

Graf č. 1: ACF a PACF nejlépe vyhovujícího modelu (Seasonal Exponential Smoothing)



Parametry modelu jsou otestovány jako statisticky významné, model je tedy sledován jako nejvhodnější, a tudíž i konečný (viz. Graf č. 2).

Graf č. 2: Model (Seasonal Exponential Smoothing) vývoje CZV tříděných vajec konzumních včetně prognózy na dalších 12 měsíců (v Kč/ks)



Vzhledem k častému použití daných modelů pro predikci výše sledovaného znaku není nutné rozepisovat funkce vypočtených modelů (často jsou jejich zápisy značně složité a v publikacích různých autorů popsány různě), jak již bylo řečeno, statistickou významnost jejich parametrů modul SAS/ETS testuje a uživatel nepotřebuje prakticky ve většině případů znát funkce příslušných modelů.

Vyhodnocení konstrukce modelů 36 vybraných časových řad v modulu SAS/ETS

Pro tvorbu modelů časových řad v systému SAS bylo vybráno 36 časových řad cen komodit agrárního trhu. Pro každou sledovanou časovou řadu byla provedena v modulu ETS diagnostika řady a byly vypočteny všechny modely, které odpovídají diagnostickým závěrům (v průměru bylo spočteno 8,14 modelů pro každou řadu). Ve sledovaném souboru časových řad byla ve 20 případech diagnostikována možnost a ve 12 případech nutnost transformace původních dat, ve 23 případech byl zjištěn trend a v 16 případech indikována sezónnost časové řady (viz. Tabulka č. 1). Dále byly pro každý model zkoumané časové řady spočteny základní charakteristiky kvality modelu (MAPE, MAE, R^2 , MSE a RMSE), z nichž stěžejní pro výběr nejvhodnějšího modelu byla nejpoužívanější - hodnota MAPE. Za nejvhodnější byl ve většině případů (u 30 komodit) vybrán některý z modelů exponenciálního vyrovnávání (nejčastěji modifikace sezónního exponenciálního vyrovnávání - viz. Tabulka č. 1).

Tabulka č. 1: Vyhodnocení modelů vybraných časových řad SC na trhu s ovocem a zeleninou

komodita	diagnostika časové řady			počet vyhov. modelů	nejvhodnější model	MAPE	vhodnost modelu vzhledem k ACF a PACF
	transf.	trend	sez.				
pomeranče	možná	ano	ne	10	Damped Trend Exponential Smoothing	4,99976	ano
Citróny	možná	ne	ano	8	Log Seasonal Exponential Smoothing	2,28303	ne
Banány	ne	ne	ne	4	Simple Exponential Smoothing	10,36405	ano
Kiwi	možná	ano	ano	20	Winters Method - Additive	5,29004	ne
Jablka	možná	ne	ne	4	Simple Exponential Smoothing	6,49563	ne
Broskve	ano	ano	ano	10	Log Linear Trend with Seasonal Terms	13,78763	ano
stolní hrozny	možná	ano	ano	10	Log-Winters Method - Additive	9,40191	ano
bíle hl. zelí	možná	ne	ano	8	Log Seasonal Exponential Smoothing	6,91884	ne
Cibule	možná	ano	ne	10	Damped Trend Exponential Smoothing	7,67023	ano
květák	možná	ne	ano	8	Log Seasonal Exponential Smoothing	7,43272	ne
Mrkev	možná	ne	ano	8	Log Seasonal Exponential Smoothing	6,18603	ano
okurky salátové	možná	ne	ano	8	Log Seasonal Exponential Smoothing	12,44184	ano
Papriky	ne	ne	ano	4	Seasonal Exponential Smoothing	7,13163	ano
Rajčata	možná	ne	ano	8	Log Seasonal Dummy	10,32547	ne
pšenice krm.	možná	ano	ano	20	Log Winters Method - Additive	1,91263	ano
ječmen slad.	možná	ano	ano	20	Winters Method - Additive	1,71161	ano
Kukuřice krm.	možná	ano	ne	10	Random Walk with Drift	3,29205	ano
repkový olej sur.	možná	ano	ne	10	Damped Trend Exponential Smoothing	0,67853	ano
repkový olej rafi.	možná	ano	ne	10	Damped Trend Exponential Smoothing	1,1577	ano
Brambory konzumní	možná	ne	ano	8	Log Seasonal Exponential Smoothing	8,84704	ne
bramborový škrob	ne	ano	ne	5	Damped Trend Exponential Smoothing	0,81205	ne
krystalový cukr	ano	ano	ne	5	Log Damped Trend Exponential Smoothing	7,04952	ano
krystalový cukr	ne	ano	ne	5	Random Walk with Drift	2,50696	ano
Melasa	možná	ano	ne	10	Damped Trend Exponential Smoothing	2,66626	ano
Krkovice	ne	ano	ne	5	Random Walk with Drift	3,82177	ano
plec bez kosti	ne	ano	ne	5	Damped Trend Exponential Smoothing	1,83544	ano
pečené s kosti čerstvá	ne	ano	ano	10	Winters Method - Additive	1,61574	ano
mléko kravské I+Q	ne	ano	ne	5	Damped Trend Exponential Smoothing	1,02981	ano
mléko polotučné trv.	ne	ano	ne	5	Random Walk with Drift	1,75837	ano
mléko polotučné trv.	ne	ano	ne	5	Damped Trend Exponential Smoothing	0,82671	ano
jat. býci A	možná	ne	ne	4	Log Simple Exponential Smoothing	1,83087	ne
jat. tele savé A	ne	ano	ne	5	Damped Trend Exponential Smoothing	3,06449	ano
hov. m. před. b. kosti	možná	ano	ne	10	Log Damped Trend Exponential Smoothing	1,72727	ano
hov. m. před. s kostí	ne	ne	ne	2	Simple Exponential Smoothing	1,09605	ne
tříděná vejce konz.	ne	ne	ano	4	Seasonal Exponential Smoothing	3,07994	ano
tříděná vejce (Kc/ks)	ne	ano	ano	10	Linear Trend with Autoregressive Errors	2,83533	ano

26 výsledných modelů vybraných časových řad bylo shledáno posouzením grafů funkcí ACF a PACF jako vhodné (viz. poslední sloupec v Tabulce č. 1). V konfrontaci se standardně používaným kritériem kvality modelu a jeho vhodnosti pro případné prognózování budoucího vývoje sledované proměnné (hodnota MAPE menší nebo rovna 5% resp. 10%) nevyhovovaly pouze 4 z 36 výsledných modelů (viz. předposlední sloupec Tabulky č. 1). Výsledkem kombinace obou zmíněných kritérií je 23 vhodných výsledných modelů pro popis vývoje sledované proměnné a použitelných pro případnou predikci řady do budoucna (viz. tučně vtištěné komodity ve stejnojmenném sloupci Tabulky č. 1).

DISKUSE

Nabídka modelů, jež jsou k dispozici po diagnóze časové řady v modulu SAS/ETS, velkou měrou závisí na vstupních datech. Velkým problémem který znepokojuje zejména v přípravné fázi zpracovávání disertační práce, je datová základna. Data jsou často neúplná, obtížně „dohledatelná“, mnohdy nedostupná či téměř nedobytná, v oblasti analýzy časových řad se navíc negativně projevuje „mládí“ České republiky, kdy máme k dispozici jen velmi krátké časové řady, které nejsou zkresleny „centrálně plánovacím hospodářstvím“ před rokem 1989. Je však nutno brát tuto situaci jako objektivní fakt a tak k ní také přistupovat.

ZÁVĚR

Systém SAS je špičkovým světovým produktem nejen pro statistické analýzy a modul ETS je jedním z nejpovedenějších a nejpropracovanějších modulů systému a je velkým pomocníkem pro jakékoli zkoumání a modelování časových řad. Dobré zvládnutí práce s tímto modulem i celým systémem je důležitým předpokladem pro zdárné zpracovávání disertační práce.

Značná nestabilita faktorů a vlivů na agrárních trzích velmi často způsobuje nepředvídatelné výkyvy ve vývoji sledovaných veličin. To se pak projevuje zejména v nemožnosti toto chování popsat vhodným modelem a tudíž v obtížné predikovatelnosti budoucího vývoje sledované proměnné, což bylo ověřeno i provedeným výzkumem, kdy pouze u necelých dvou třetin ze zkoumaných veličin byl nalezen vyhovující model pro popis jejich vývoje a případné předpovědi jejich chování v budoucnu.

Literatura:

- Arlt, J: Moderní metody modelování ekonomických časových řad, Praha, Grada Publishing, 1999
- Arlt, J; Arltová, M; Rublíková, E: Analýza ekonomických časových řad s příklady, VŠE, Praha, 2002
- Cipra, T: Analýza časových řad s aplikacemi v ekonomii, Praha, SNTL, 1986
- Hindls, R; Hronová, S; Novák, I: Metody statistické analýzy pro ekonomy, Praha, Management Press, 2000
- Hindls, R; Hronová, S; Seger, J: Statistika pro ekonomy, Praha, Professional Publishing, 2003
- Svatošová, L; Hříbal, J; Volma, M: Systém SAS – příručka pro uživatele, ČZU, Praha, 2000
- Mze ČR: Situační a výhledová zpráva - Brambory, říjen 2003
- Mze ČR: Situační a výhledová zpráva - Cukrovka, cukr, prosinec 2003
- Mze ČR: Situační a výhledová zpráva - Drůbež, prosinec 2003
- Mze ČR: Situační a výhledová zpráva - Hovězí maso, červen 2003
- Mze ČR: Situační a výhledová zpráva - Mléko, srpen 2003
- Mze ČR: Situační a výhledová zpráva - Obiloviny, září 2003
- Mze ČR: Situační a výhledová zpráva - Olejniný, leden 2004
- Mze ČR: Situační a výhledová zpráva - Ovoce, říjen 2002
- Mze ČR: Situační a výhledová zpráva - Vepřové maso, prosinec 2003
- MZe ČR: Situační a výhledová zpráva - Zelenina, prosinec 2002

Kontaktní adresa autora:

Ing. Zdeněk Louda

Katedra statistiky, Provozně ekonomická fakulta, Česká zemědělská univerzita v Praze,
Kamýcká 129, Praha 6 – Suchbátka, 16521,
telefon: 2 2438 3246, email: louda@pef.czu.cz,