ANALÝZA JEDNÉ MOŽNOSTI APROXIMACE EMPIRICKÉ S-KŘIVKY

Bohdan Linda
Katedra matematiky, FES, Univerzita Pardubice

One of many curves apposite trend of time series is Gompertz curve. This curve is used above all in case, when the growth rate in observed period isn't constant. The contribution deals with the evaluation of the different methods its coefficients.estimation.

Pro současné období jsou charakteristické rozsáhlé inovace výrobků i výrobních technologií a zavádění nových výrobků na trh. To s sebou přináší problémy jak technického, tak i ekonomického charakteru. Jedním z problémů, který je nutno řešit, je problém odbytu nově zaváděných a inovovaných výrobků. Při formulaci prognóz se často používají trendové křivky typu S.

Cílem tohoto příspěvku je ukázat vyrovnání časové řady jednou ze složitějších trendových funkcí typu S – Gompertzovou křivkou, která znázorňuje závislost růstu sledovaného jevu na čase. Využívá se například ke znázornění průběhu zájmu o nový výrobek. Po zahájení výroby stoupá zájem nejdříve pomalu, pak se prudce zrychluje a za inflexním bodem se opět zpomalí a asymptoticky se blíží k hodnotě k. Podobně můžeme pomocí Gompertzovy křivky analyzovat rozvoj služeb, dopravy, výnosů v zemědělství, vývoz určitého produktu, některé biologické jevy, vhodným způsobem charakterizuje i množství výrobků v užívání.

Gompertzova křivka má tvar

$$T_t = k \cdot \alpha^{\beta^t} \quad , \; k>0, \; 0<\alpha<1, \;\; 0<\beta<1.$$

Metody odhadu koeficientů Gompertzovy křivky jsou náročnější vzhledem k tomu, že vyjádření křivka není lineární v parametrech a ani ji nelze vhodným způsobem linearizovat, jako je tomu například u exponenciální funkce. Bez použití počítačové techniky nebylo možné řešit řadu problémů obecnějšího charakteru (např. proložení křivky řadou neekvidistantních bodů metodou nejmenších čtverců lineárních odchylek).

Gompertzova křivka patří do skupiny asymetrických S-křivek. Těžiště jejích hodnot v čase je až za inflexním bodem. Křivka má horní asymptotu rovnou konstantě k, která udává horní mez např. prodeje nebo spotřeby při jejím nasycení pro $t \to \infty$. Parametr β je mírou strmosti křivky, čím menší je hodnota tohoto parametru, tím strměji křivka roste. Parametr α se odhaduje pomocí podílu T_0/k, kde T_0 můžeme interpretovat jako počet výrobků v používání v čase 0 a k je mezní hodnota čili stav nasycení.

Pro odhady parametrů Gompertzovy křivky existují různé metody. Většina z nich je založena na logaritmické transformaci její rovnice, kdy pak dostaneme rovnici modifikovaného exponenciálního trendu

$$\ln T_t = \ln k + \beta^t \ln \alpha$$

Funkce není lineární v parametrech, proto nemůžeme k odhadu jejích parametrů použít metodu nejmenších čtverců v její analytické formě. Musí být použita jiná odhadová technika. Pro porovnání byly vybrány tři různé metody: metoda vybraných bodů, metoda dílčích průměrů a metoda částečných součtů.

Tyto metody byly testovány na náhodně generovaných údajích. Předpokládali jsme, že odchylky mají normální rozložení pravděpodobností $N(0,\sigma)$ a volili jsme různé hodnoty směrodatné odchylky. Bylo provedeno více než sto simulací pro různé hodnoty parametrů α, β a k. Jako kriteria pro rozhodování, které z odhadů parametrů Gompertzovy křivky jsou nejkvalitnější, jsme používali reziduální součet čtverců. Za vhodnější se zpravidla považuje ten odhad regresní funkce, kdy reziduální součet čtverců dosahuje nižších hodnot.

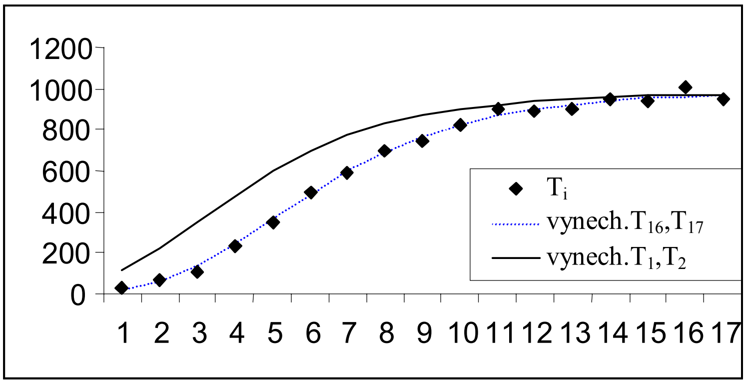
Pro ilustraci jsou uvedeny výpočty parametrů získané jmenovanými metodami z jedné série generovaných dat. Uvažujeme v tomto případě Gompertzovu křivku tvaru $T=980\cdot 0{,}002^{0{,}7^t}$. Generované odchylky mají N(0,25) rozložení pravděpodobností.. Hodnoty pro sedmnáct členů časové řady jsou zapsány v následující tabulce:

t_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
T_i	28	66	105	236	348	496	593	696	748	825	904	894	897	953	936	1002	951

Metoda částečných součtů:

Při odhadu parametrů metodou částečných součtů rozdělíme soubor sedmnácti pozorování na tři stejně velké části o pěti pozorováních. Protože máme počet pozorování, který není dělitelný třemi, musíme dva členy časové řady vynechat. V literatuře se doporučuje vynechat první resp. první dva členy, protože se jedná o nejstarší měření s nejmenší vypovídací hodnotou. Tvar Gompertzovy křivky je ale takový, že první členy mají velký význam a jejich vynecháním dostáváme nevhodné odhady. Naopak hodnoty poslední již konvergují ke konstantě k a na tvar odhadované křivky nemají významný vliv. Proto v součtech S_1, S_2, S_3 vynecháváme poslední dvě měření a počítáme $S_1=\sum_{i=1}^{5}\ln T_i$, $S_2=\sum_{i=6}^{10}\ln T_i$, $S_3=\sum_{i=11}^{15}\ln T_i$.

Následující obrázek ukazuje rozdíl ve tvaru Gompertzovy křivky při výpočtu částečných součtů s vynecháním prvních dvou členů a s vynecháním posledních dvou členů.



Obr.1:Tvary křivky s vynecháním prvních dvou členů a posledních dvou členů časové řady

Rovnice odhadnutého trendu metodou částečných součtů je $\hat{T}_t=\hat{k}\cdot a^{b^t}=984{,}782\cdot 0{,}00426^{0{,}7107^t}$.

Pro součet čtverců platí: $S_e = 6\,449{,}1$.

Rovnice odhadnutého trendu metodou dílčích průměrů je $\hat{T}_t=\hat{k}\cdot a^{b^t}=950{,}1846\cdot 0{,}004996^{0{,}7068^t}$.

$S_e = 14\,113{,}5$.

Rovnice odhadnutého trendu metodou vybraných bodů je $\hat{T}_t=\hat{k}\cdot a^{b^t}=969{,}3758\cdot 0{,}0073^{0{,}7209^t}$.

$S_e = 12\,068{,}7$.

Výsledky výpočtů pro jednu skupinu generovaných dat jsou shrnuty do tabulky:

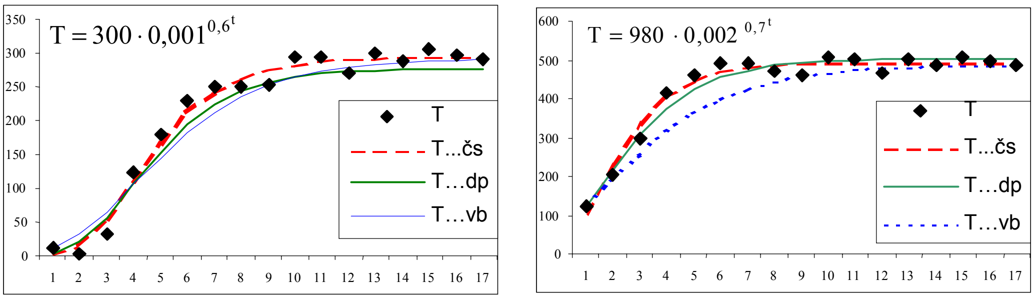
dané parametry	odhad-met. částeč. součtů	odhad-met. dílčích prům.	odhad-m. vybraných bodů
k=980	984,78	950,18	969,38
a=0,002	0,0043	0,002	0,0073
b=0,7	0,7107	0,7068	0,7209
součet čtverců	6 449,1	14 113,5	12 068,7

Na základě uvedených výsledků lze říci, že nejlepší odhad parametrů Gompertzovy křivky pro danou časovou řadu jsme získali metodou částečných součtů s vynecháním posledních dvou členů časové řady.

Následující tabulka ukazuje hodnoty reziduálních součtů čtverců, jakožto kritéria úspěšnosti odhadů, pro odhady uvedenými třemi metodami při různých hodnotách parametrů Gompertzovy křivky i při různých směrodatných odchylkách chyb. Nejnižší hodnota reziduálního součtu čtverců je vždy vyznačena tučně.

		částečné součty	dílčí průměry	vybrané body
k=300	σ=1	**14,80958747**	20,72359285	45,47363339
a=0,001		**13,31573142**	28,82861443	23,88060213
b=0,6		**46,08446121**	331,7592288	240,7245849
	σ=5	**312,0382608**	562,5222631	561,2189737
		373,0615111	531,6270401	1367,582185
		372,2283046	725,9200707	1163,381402
	σ=10	**1398,556885**	2856,655272	4365,851026
		1123,677283	2419,723854	1868,501024
		1939,174308	2098,461238	8475,237112
	σ=13	**1985,622435**	2927,446271	3524,189716
	σ=15	**3008,914031**	7484,672146	9337,942666
k=980	σ=5	**501,4020622**	1069,388689	866,7057044
a=0,002	σ=10	**2050,769405**	4438,097249	3478,521601
b=0,7	σ=20	**4609,299165**	8865,356311	8240,740254
	σ=25	**6449,080946**	14113,54158	12068,66035
	σ=29	**11572,8397**	14910,05904	15809,66324
k=500	σ=5	344,7207591	**297,2763902**	3440,239894
a=0,05	σ=10	**2263,32308**	4889,120606	8951,391576
b=0,5	σ=20	**5571,959623**	8071,831192	37705,49241
	σ=12	**1654,786612**	5144,675684	19960,7151
	σ=18	21545,67118	**3683,496389**	10717,61023

Na následujících obrázcích (obr.2a, 2b) jsou znázorněny trendové funkce, které byly získány po výpočtu parametrů jednotlivými metodami (čs – částečných součtů, dp – dílčích průměrů, vb- vybraných bodů).



Obr.2a, 2b:Tvary křivky při odhadech parametrů různými metodami

Bylo provedeno více než sto simulací a v 93% případů byly získány nejlepší výsledky rovněž metodou částečných součtů (na tento závěr ukazuje i výše uvedená tabulka). Ve výsledcích se neprojevila ani hypotéze, že výběr metody je závislý na konkrétních hodnotách parametrů k, a, b, proto se můžeme domnívat, že metoda částečných součtů je nejvhodnější z uvedených metod pro odhad parametrů Gompertzovy křivky.

V praxi již bylo ověřeno, že prognózy provedené pomocí Gompertzovy křivky na základě znalosti počátečních hodnot časové řady poskytují velmi dobré výsledky. Znalost této problematiky pak umožňuje při zavádění nových nebo inovovaných výrobků na trh předpovídat na základě znalostí počátečních hodnot časové řady průběh rozvoje tohoto výrobku. Je pak podle toho možné usměrňovat výrobu, import i export tak, aby nedocházelo k poruchám způsobeným nedostatkem nebo přebytkem výrobku na trhu.

Literatura:

1. Cipra,T.: Analýza časových řad s aplikacemi v ekonomii, SNTL/ALFA, Praha 1986
2. Chajdiak,J.,Rublíková,E.,Gudába,M.: Štatistické metódy v praxi, Statis, Bratislava 1998
3. Korόny,S.: Úvod do analýzy časových radov s aplikáciou na mesačné časové rady, Ekomstat, Tr.Teplice 1996
4. Kubanová,J.:Matematická statistika. Univerzita Pardubice 1999.
5. Kubanová,J.: Trend úmrtnosti v České republice. . In: Sborník 7. demografické konference s mezinárodní účastí na téma Demografické, zdravotné a sociálno-ekonomické aspekty úmrtnosti. Trenčianske Teplice 1999, s.86 – 90, ISBN 80-88946-00-X
6. Pacáková,V.:Aplikovaná poistná štatistika. Vysokoškolská učebnica, FHI EU Bratislava, Vydavatĺstvo Ekonóm 1998.
7. Řezanková,H.: Metody pro získávání znalostí z dat, Štatistické metódy v praxi, Bratislava 1998
8. Vojtková,M.: Odhad parametrov logistickej trendovej funkcie pomocou regresie, Ekomstat, Tr. Teplice 1990
9. Wei,W.,S.:Time series analysis, Univariate and multivariate methods, Addison-Wesley Publishing Company, California 1990

Kontaktní adresa:

doc. RNDr. Bohdan Linda, CSc., KM FES Univerzita Pardubice
Studentská 84, 532 10 Pardubice
☎ 040 6036020
e-mail: Bohdan.Linda@upce.cz

Recenzoval: doc.Ing.Jan Čapek,CSc., Katedra informačních systémů, FES, UPa