

Předpovídejte snadno a rychle

Newsletter Statistica ACADEMY

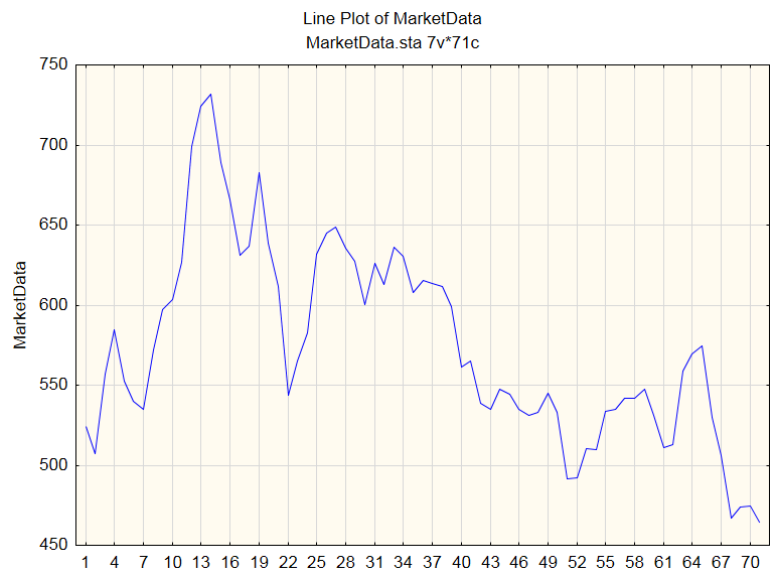


Téma: Časové řady, exponenciální vyrovnávání
Typ článku: Příklad

Dnes se budeme zabývat situací, kdy chceme předpovídat, jak se bude v čase vyvíjet průběh nějaké veličiny. Můžete si pod tím představit například vývoj kurzu, ceny akcií, počtů objednávek daného produktu v dalších obdobích,...

Zadání

Máme k dispozici historii, jak se daná časová řada vyvíjela a chceme odhadnout (předpovědět), jaký bude její průběh dále. My konkrétně chceme předpovídat vývoj časové řady vpravo. Jedná se o reálnou časovou řadu, v našem případě vývoj ceny mědi (řada je to měsíční počínaje lednem 2010 a konče listopadem 2015).

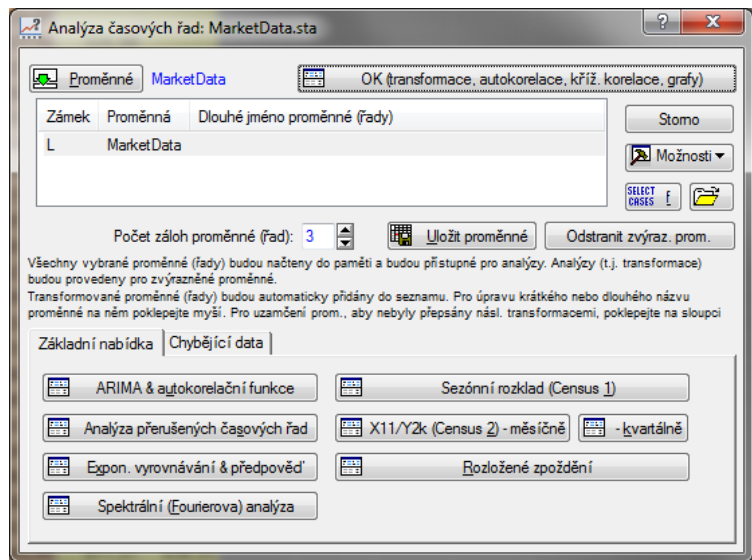


Jak to budeme řešit

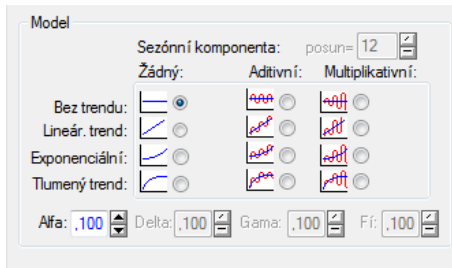
Tento příklad a jeho řešení bude zvláštní v tom, že se vůbec nebudeme zabývat teorií okolo modelů, které použijeme, prostě tupě použijeme metody, které jsou dostupné, a uvidíme, jak to dopadne. Pokud chcete řešit příklad s námi, můžete si data stáhnout [zde](#).

Postup

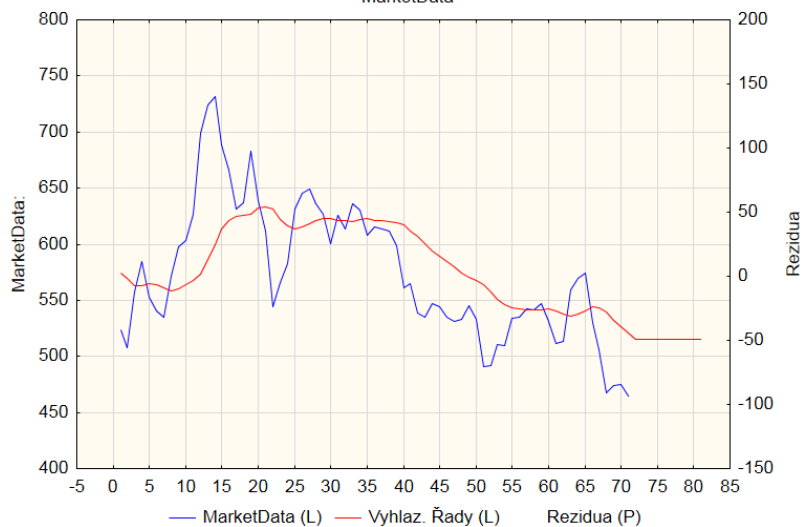
Řešíme predikci časové řady, je tedy jasné, že budeme hledat vhodné metody v sekci **Statistiky->Pokročilé modely->Časové řady/Predikce**. Zkusíme, co se dá vytěžit z metod exponenciálního vyrovnávání. Nejprve vybereme proměnnou pod tlačítkem **Proměnné**, vybraná proměnná se objeví v okně pod tímto tlačítkem:



Zkusíme tedy Exponenciální vyrovnávání, tlačítko **Expon. vyrovnávání & předpověď**. Zde na začátku nebudeme nic nastavovat a klikneme přímo na tlačítko vpravo nahoře **Shrnutí**. Ve výstupech máme porovnání červeného modelu a reálných modrých dat (zelená rezidua jsme v rámci zpřehlednění výstupu smazali). Vidíme, že takovýto model nám moc stačit nebude. Toto byl tedy nejjednodušší model:



Exponenc. vyrovnáv. : S0=574,6
Bez trendu, žádná sezóna, Alfa= ,100
MarketData

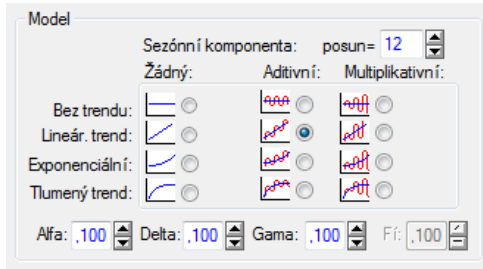


Zkusíme tedy něco zajímavějšího.

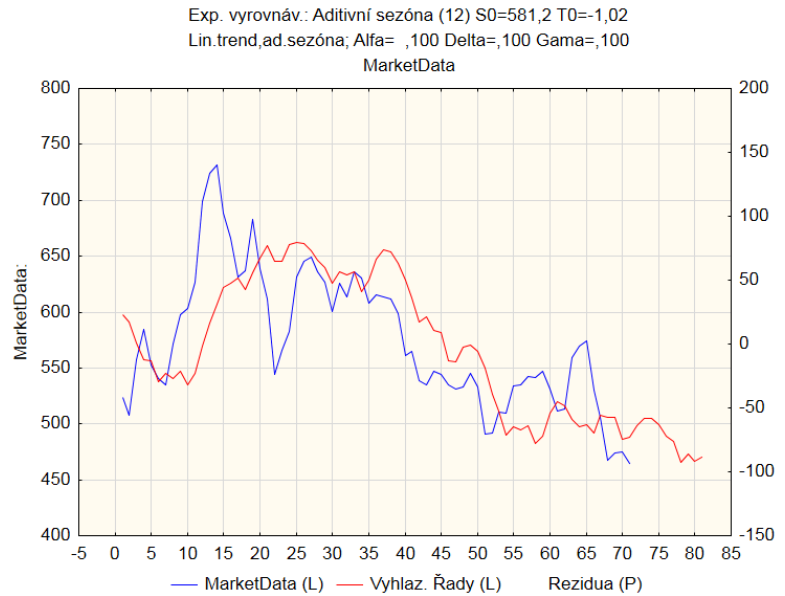
Před dalším pokusem nezapomeňte označit v okně s řadami opět výchozí řadu, toto okno totiž určuje, se kterou řadou právě pracujete (vyskytly se nám zde totiž i dvě nové řady pro předpovědi a rezidua exponenciálního vyrovnávání).

Zámek	Proměnná	Dlouhé jméno proměnné (řady)
L	MarketData	
	MarketData	Exp.vyrov. ;
	MarketData	Exp.vyrov. rezid.;

Druhý pokus. Dobře, vybereme tedy nějaký složitější model, asi to bude jedna z dalších možností v sekci model. Zvolme tedy například toto:



Klikneme na tlačítko v pravém horním rohu. To už vypadá lépe, ale stále to není nic, s čím bychom mohli být spokojeni. Z okének i výstupu vidíme, že model používá nyní tři parametry místo jednoho v prvním pokusu modelování. Nicméně tyto parametry jsou nastaveny pevně tak, jak jsme si je zvolili v dialogu modelu. Opět zvolíme *MarketData* jako výchozí řadu a jdeme na třetí modelování.



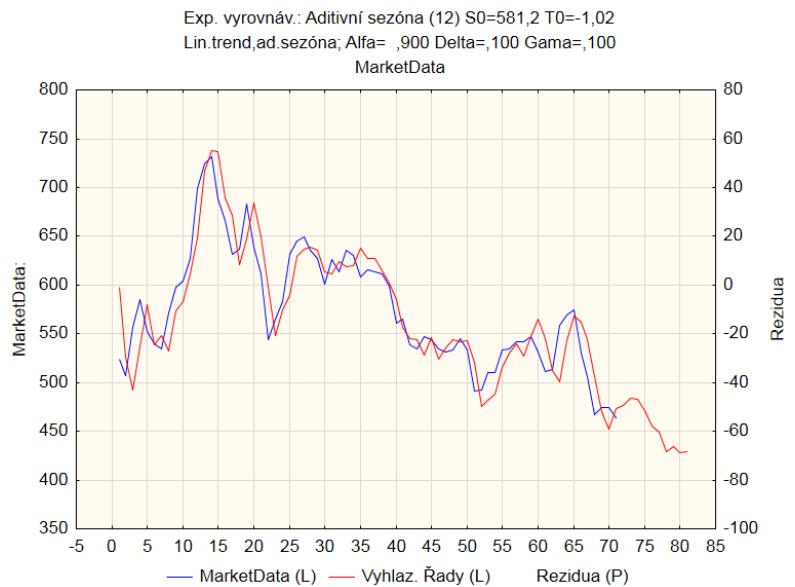
Všimli jsme si, že parametry modelů jsou nastaveny „natvrdo“ na hodnoty 0,1, kdo ale řekl, že tyto parametry jsou nejlepší? To by šlo jistě vylepšit vhodnější volbou těchto parametrů. K tomu slouží záložky **Síťové hledání** a **Automatické hledání**. Zkusíme síťové hledání, které zkouší různé hodnoty parametrů na nějaké nadefinované mřížce hodnot, výstup nám ukáže tabulku s nejlepšími modely, které říťové hledání našlo. Zmáčkneme tlačítko **Provést síťové hledání**. Ve výstupech vidíme, že na prvním místě mají parametry jiné hodnoty, než které jsme měli nastaveny. Vrátime se tedy na záložku **Základ** a zadáme tyto nejlepší parametry ze síťového hledání.

Mřížkové hledání parametrů (nejmenší abs. chyby jsou zvýrazněn (MarketData.sta)
 Model: Lineár. trend,ad.sezóna (12); S0=581,2 T0=-1,02
 MarketData

Model Číslo	Alfa	Delta	Gama	Prům. Chyba	Průměr a Chyba	Suma Mocniny	Průměr Mocniny	Prům. % Chyba	Průměr a % chyba
649	0,900000	0,100000	0,100000	-0,572322	21,07869	51339,39	723,0901	-0,170940	3,698542
658	0,900000	0,200000	0,100000	-0,571703	21,18347	51612,39	726,9351	-0,170768	3,716636
667	0,900000	0,300000	0,100000	-0,570849	21,28455	51906,41	731,0762	-0,170535	3,734070
676	0,900000	0,400000	0,100000	-0,569777	21,38200	52219,42	735,4848	-0,170246	3,750850
685	0,900000	0,500000	0,100000	-0,568505	21,47586	52549,80	740,1381	-0,169906	3,766987
694	0,900000	0,600000	0,100000	-0,567051	21,56620	52896,26	745,0177	-0,169519	3,782487
703	0,900000	0,700000	0,100000	-0,565432	21,65307	53257,78	750,1095	-0,169089	3,797361
712	0,900000	0,800000	0,100000	-0,563663	21,73655	53633,59	755,4027	-0,168621	3,811618
721	0,900000	0,900000	0,100000	-0,561761	21,81669	54023,15	760,8894	-0,168119	3,825267
650	0,900000	0,100000	0,200000	-0,442435	21,42703	54028,26	760,9614	-0,121425	3,759208

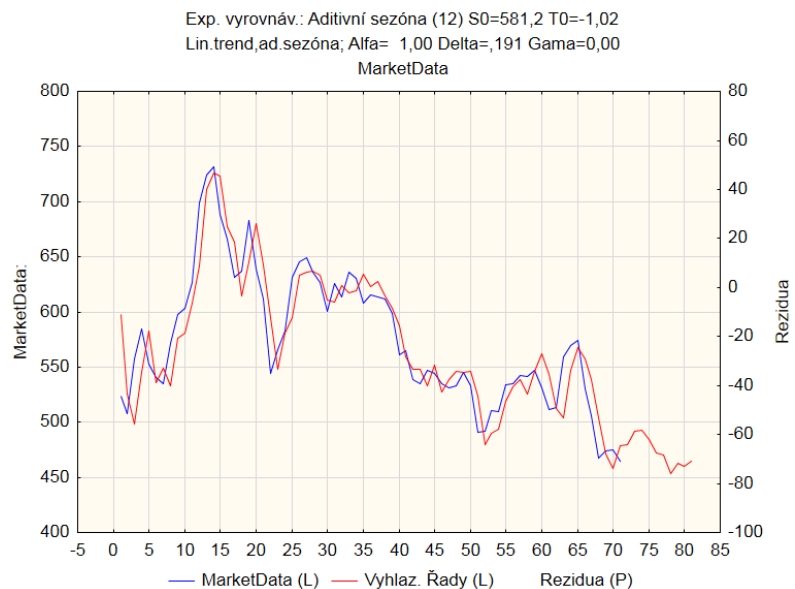


Tento výsledek již vypadá velmi slibně a asi bychom se s ním i spokojili.



Kdybychom zkusili automatické hledání, pak bude výsledný model takovýto:

Na první pohled se výsledky modelů téměř shodují, rozdíl je v tom, že automatické hledání hledalo i v prostoru mimo definovanou mříž, můžeme si všimnout, že hodnota parametru Gama je 0, což znamená, že tento parametr vlastně nepoužíváme.

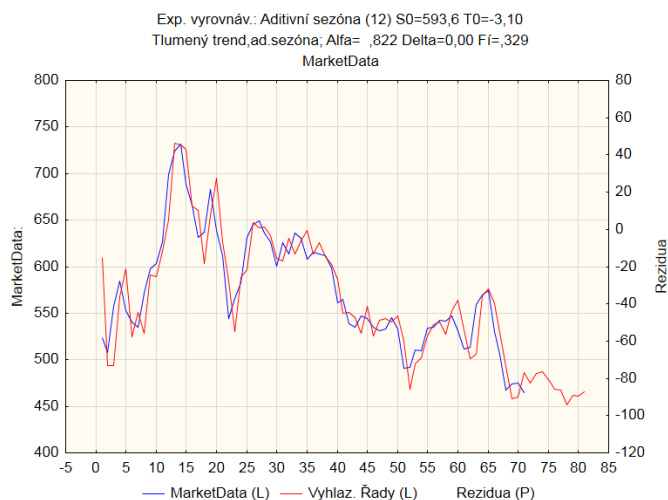
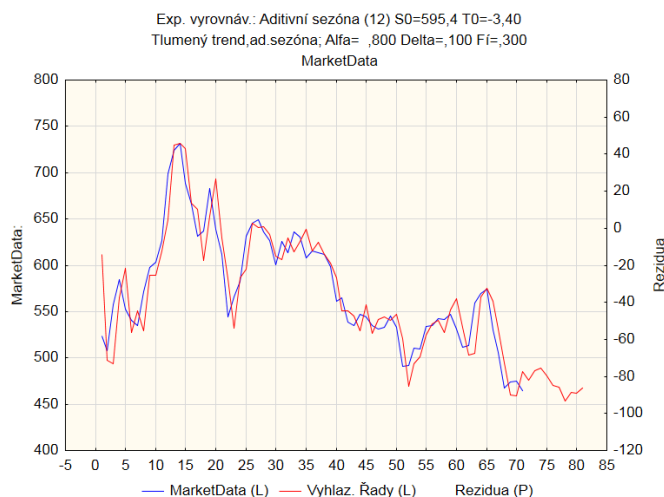


Můžeme se pokusit ještě vylepšit náš model vyzkoušením jiných modelů, můžeme se také trochu orientovat z piktogramů u modelů, které trochu něco říkají o tom, jak může řada v takovémto případě vypadat.

V zásadě při hledání můžete bez jakékoli znalosti teorie zkusit postupně proklikat všechny modely a sítově či automaticky hledat nejlepší parametry a poté vybrat model, který nejrozhodněji odpovídá datům.

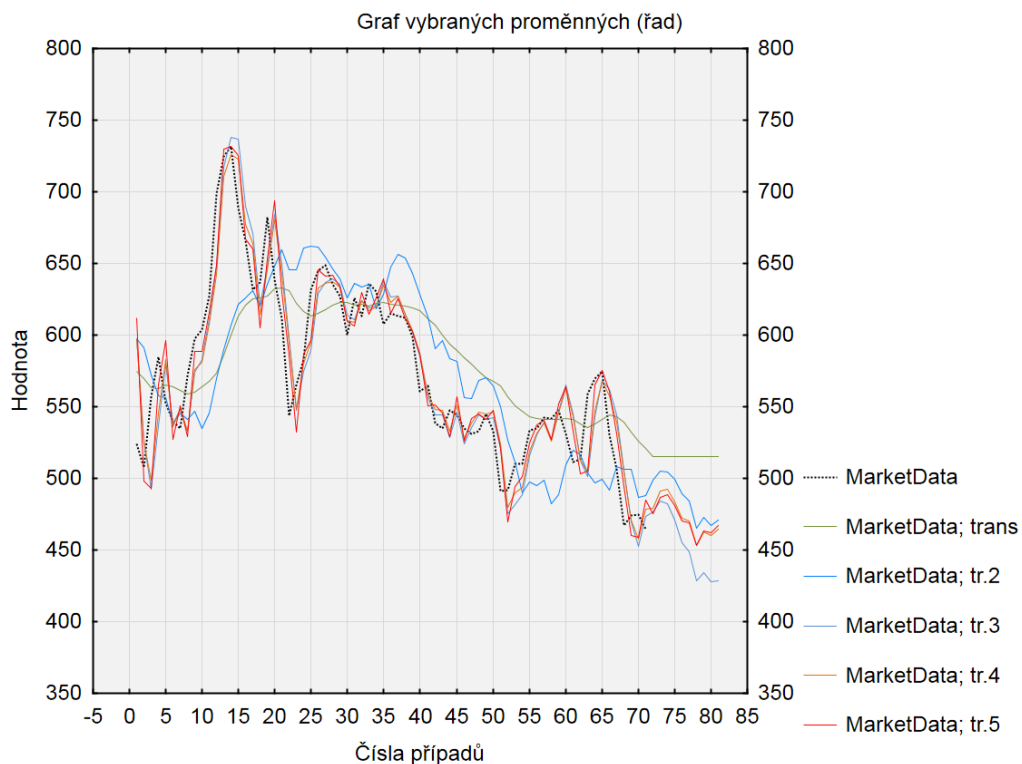
Zkusme ještě model s parametrem Fí:





Závěr

Na následujícím obrázku se můžeme podívat na srovnání všech zmíněných modelů postupně, jak šly za sebou. Je vidět, že jsme byli schopni najít velmi pěkný predikční model, aniž bychom použili jakékoli znalosti teorie. V grafech vidíme také predikce modelů. Kterýkoli z modelů 3, 4, 5 se zdá být dostatečně dobrý.



O problematice časových řad jsme také psali v článku:

[Možnosti vyhodnocení časových řad v Statistica](#)

[Časová řada v modulu Neuronové sítě](#)

[Pokročilá práce s časovou řadou – odhad řádu ARMA](#)