

Predikcia časových radov

BC. Pavol Loffay¹

21. novembra 2015

Abstrakt: Práca spracováva predikciu časových radov prevzatých z monitorovacieho systému Hawkular^a. Tento systém dokáže monitorovať Java aplikácie, alebo fyzický stroj na ktorom je spustený. Z množiny pozorovaných metrík som vybral nasledujúce: vyťaženie Java hromady (heap), miesto na disku a počet voľných databázových spojení. Tieto časové rady som analyzoval a cieľom bolo zostaviť model, ktorý najlepšie popisoval priebeh danej časovej rady. V práci som postupoval podľa Box – Jenkinsových metód.

Kľúčové slová: Časová rada; ARIMA; Hawkular; ACF

JEL klasifikácia: C53

^aDostupné na <http://www.hawkular.org>

1 Úvod

[2] [1]Povšimnite si prvého odstavce, který není podle standardních pravidel L^AT_EXu odsazen. Všechny ostatní odstavce již odsazeny jsou; jsou formátovány automaticky, proto jen vždy začněte nový odstavec buď odřádkováním, nebo příkazem \par. Mezi odstavci je nastavené automatické řádkování 6 bodů, proto prosím nepoužívejte žádné další příkazy na řádkování. Všimněte si ještě tohoto *zvýrazněného* slova a také tohoto **velmi důležitého** slova (ke zvýraznění použijte příslušné styly).

Většina článků začíná stručným představením toho, jaké otázky jsou řešeny, co vás k jejich řešení vedlo a shrnuje dosažená empirická zjištění. Úvod by měl být většinou psán jednoduchým „netechnickým“ jazykem s minimem odborných ekonomických a statistických výrazů. I laický čtenář by tak měl obecně pochopit problém a získané závěry, které jsou v projektu či příspěvku řešeny. Právě na tomto místě je nejlepší příležitost k tomu, vyzdvihnout originalitu a přínos vašeho článku, tedy čím je zajímavý a proč má cenu ho vůbec číst. Není od věci přehledně představit i obsah jednotlivých částí zprávy.

Na úvod ještě uveďme, že zde navrhované názvy nejsou nijak závazné a můžete jednotlivé kapitoly nazvat i jinými, originálnějšími názvy, které lépe vystihují daný problém.

¹Masarykova univerzita, Fakulta informatiky, obor: Service Science Management Engineering, p.loffay@mail.muni.cz

2 Ekonomický model

Tato část obsahuje formální popis teoretického modelu. Obvykle je techničtěji založená a využívá se zde více jazyk ekonomie a matematiky. V této části je možné se zaměřit na posluchače, kteří jsou experty v daném oboru. V této části je věnován prostor specifikaci používaného ekonomického modelu a definování ekonomických proměnných.

Ekonomický model může být mnohdy rozsáhlý a komplikovaný. Vaším úkolem je vysvětlit model zcela jasně, a to co nejstručněji a nejjednodušším způsobem. Není třeba používat ryze technického žargonu. Kde je to možné, snažte se používat jednoduchých a výstižných pojmů a obrátů namísto zbytečně komplikovaných výrazů. Vaším cílem je ukázat kvalitu vašich myšlenek, nikoli šíři a rozsah vaší slovní zásoby.

Vhodnou součástí vaší práce je seznámení čtenářů s aktuální úrovní výzkumu daného tématu. Není ovšem nutné představovat přehled veškeré literatury, která se k danému tématu vztahuje. Na odbornou literaturu se můžete odkazovat pomocí příkazu `\cite{label}`, například „Enders [?] uvádí ...“ nebo „Podle [?] můžeme ...“. Zkuste si odpovědět na otázky: Věnoval se tomuto konkrétnímu tématu již někdo před vámi? Pokud ano, jaké jsou závěry jeho práce? Bude vaše práce něčím daný problém rozšiřovat nebo jenom replikujete dané metody pro jiná data?

Na závěr této sekce byste si měli stanovit jednu až dvě pracovní hypotézy, které budete v práci testovat a které by měly být nejdůležitějším prvkem vaší práce. Nesnažte se stanovovat moc obecné hypotézy (např. jestli globalizace má vliv na ekonomický růst²) nebo naopak takové hypotézy, u kterých je na první pohled patrné, že platí/neplatí (např. velká auta mají větší spotřebu jako malá auta). Příkladem vhodné hypotézy v případě, že se vaše práce bude věnovat problému spropitného, je například, jestli muži dávají v průměru vyšší spropitné než ženy a jaké je případná velikost tohoto rozdílu.

3 Ekonometrický model

V této části byste měli čtenáři představit ekonometrické nástroje, které využijete při své analýze. Není nutné uvádět ty techniky, které jsou známy průměrnému vzdělanému člověku daného oboru. V případě předmětu *Základy ekonometrie* tedy není nutné vysvětlovat čtenáři, co je to lineární regrese a že se její koeficienty odhadují pomocí metody nejmenších čtverců. Uvádějte jen ty metody, které nejsou v daném oboru běžné.

Jelikož se předpokládá, že v této sekci naplno využijete sazby matematiky, kterou nabízí L^AT_EX, obsahuje tento balíček tyto předdefinované prostředí a funkce. Pomocí příkazů `\Cbb`, `\Rbb`, `\Zbb`, `\Nbb` můžete vysázet tučné symboly pro číselné obory \mathbb{C} , \mathbb{R} , \mathbb{Z} , \mathbb{N} .

Dále pomocí příkazu `\e{x}` lze vysázet e^x (všimněte si rozdílu mezi e^x a e^x), lze použít i L^AT_EXový příkaz `\exp{x}`, jež po vysazení vypadá takto $\exp x$. Především při sazbě integrálu můžete použít příkaz `\dif{x}` – viz rozdíl $\int f(x)dx$ a $\int f(x)dx$.

Při maximalizačních či minimalizačních úlohách můžete využít předdefinovaných operátorů `\argmax` nebo `\argmin`, které lze kombinovat např. i s předchozími příkazy, viz vzorec 3.1. Dalším předdefinovaným operátorem je `\sqn`, který vysází operátor signum sgn .

²Takové obecné hypotézy stejně potřebují konkretizovat ve smyslu, co je myšleno globalizací a co ekonomickým růstem.

$$\beta = \arg \max_{x \in \mathbb{R}} \int_0^x \sin t e^{-t} dt \quad (3.1)$$

Tato šablona obsahuje také podporu základním statistickým funkcím. Najdete zde příkaz `\E{X}`, který sází střední hodnotu $E(X)$. Dále příkaz `\Var{X}` na vysázení rozptylu $\text{Var}(X)$, příkaz `\Cov{X}{Y}` na vysázení kovariance $\text{Cov}(X, Y)$ a příkaz `\Cor{X}{Y}` na vysázení $\text{Cor}(X, Y)$.

Příkaz `\vec{}` je modifikovaný tak, aby místo šipky nad svým argumentem, tento argument vysázel tučně, např. tučné α, β, γ oproti normálnímu α, β, γ .

Pro opravdové matematické „fajnšmekry“ jsou v šabloně nachystány i prostředí veta, definice a pozn. Lze použít i L^AT_EXovské prostředí proof.

Věta 3.1 (Moje). *Moje věta.*

Důkaz. Zde je důkaz předchozí věty, jehož konec poznám pomocí čtverečku napravo. □

1 Definice. Moje definice.

Poznámka. Moje poznámka.

4 Data

Touto částí začíná část s empirickými výsledky vaší práce. Měla tvořit hlavní a nejdůležitější část vaší práce. Rozsahem by měla odpovídat přibližně polovině rozsahu celé práce. V této části byste měli nejprve představit data, se kterými pracujete. Někdy se v této části zmiňuje i program, popř. jeho balíčky, ve kterém jste data zpracovávali. Dále byste měli uvést odhady modelu, případně zkontrolovat i splnění jeho předpokladů. Jako poslední přichází na řadu interpretace výsledků, což je nejzásadnější část celé práce, proto byste ji měli věnovat zvýšenou pozornost. Zkuste si odpovědět na otázky: Jsou odhadnuté koeficienty významné nebo jsou p-hodnoty jen těsně nad zvolenou hladinou významnosti (při menším počtu pozorování je vhodné pracovat s větší hladinou významnosti)? Jsou výsledky v souladu s vaší představou? Je model správně specifikovaný? Zvolili jste správný datový vzorek? Zamítáte či nezamítáte hypotézy stanovené na začátku práce?

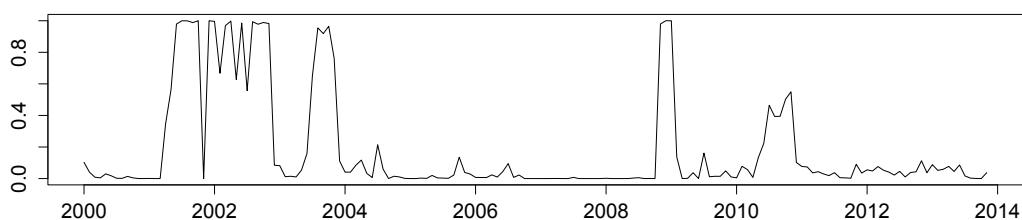
Jelikož pracujete s reálnými daty³, je potřeba s nimi čtenáře seznámit. Měli byste začít s tím, že uvedete zdroj, odkud vaše data pocházejí. Pokud data pocházejí z dotazníkového šetření, můžete uvést podobu dotazníku v příloze A společně s popisem cílové skupiny, která na tento dotazník odpovídala. V případě, že jste museli nějakým způsobem vaše data transformovat, nezapomeňte tuto skutečnost do práce také napsat.

Pro vyšší názornost je vhodné data vizualizovat, viz Obrázek 1.

Standardně se v podobných člancích vyskytuje i tabulka se základními popisnými charakteristikami, viz Tabulka 1⁴.

³Doporučujeme si přečíst přílohu D učebního textu, ve kterém se blíže seznámíte, jak pracovat s datovými databázemi.

⁴Tabulky uvedené v této šabloně jsou pouze příklady, jak může formátování tabulek vypadat a nemusíte se tohoto formátování striktně držet. Může volit vlastní formát, kde se preferuje co nejjednodušší použití čar (pro přehlednost tabulek).



Obr. 1: Pod obrázkem by měl následovat popis grafu společně s uvedením **zdroje**.

Tabuľka 1: Pod tabulkou by měl následovat popis tabulky společně s uvedením **zdroje**.

Proměnná	Průměr	Sm. odchylka	Minimum	Maximum
Proměnná1	10	1,2	7	12
Proměnná2	5	1	3	4,7
⋮	⋮	⋮	⋮	⋮

5 Odhadnutý model

Všechny statistické metody, které využíváte, jsou zkonstruované na základě jistých předpokladů. Zde přichází vhodné místo, kdy je potřeba splnění těchto předpokladů testovat. Testování některých předpokladů můžete rozdělit do dalších podsekcí (pomocí příkazu `\subsection{}` pro číslovanou variantu a `\subsection*{}` pro nečíslovanou variantu), např. pro ilustraci: Odhad lineární specifikace a Odhad log-lineární specifikace.

Pro vysázení těchto předpokladů můžete využít například i seznamy.

Nečíslovaný seznam:

- jedna
- dva
- tři

Číslovaný seznam:

1. jedna
2. dva
3. tři

Dále byste měli v této části přepsat do přehledné tabulky odhady koeficientů společně s jejich směrodatnými hodnotami a jejich významností, viz Tabulka 2. Nedoporučuje se kopírovat výsledky v podobě „screenshotu“, který poté vložíte jako obrázek do textu. Program *Gretl* nabízí jako jeden z výstupů \LaTeX , který vám vygeneruje automaticky kód, který poté stačí jen zkopírovat do šablony.

6 Interpretace výsledků

Jakmile máte odhadnutý model, který splňuje všechny předpoklady kladené na tento model, měli byste být schopni vhodně interpretovat dosažené výsledky vzhledem ke stanoveným pracovním hypotézám na začátku práce.

Na tomto místě je vhodné zdůraznit, že neexistují „dobré“ ani „špatné“ empirické výsledky. V mnoha případech může nastat, že empirické výsledky nepotvrdí původní představu výzkum-

Tabuľka 2: Pod tabuľkou by měl následovat popis tabulky společně s uvedením **zdroje**.

Proměnná	Model A	Model B
Proměnná1	10,56841 *** (2,54851)	
Proměnná1Mod		5,87787 *** (1,00155)
Proměnná2	50,56984 *** (5,00527)	47,59663 *** (4,55953)
Proměnná3	20,96541 * (3,64164)	25,05194 ** (3,70647)
Proměnná4	17,87406 (40,22301)	113,33046 * (7,29106)
Koeficient determinace	0,822118	0,842665
F test	21,78750 ***	20,56845 ***
Durbin–Watson	0,007214	0,26546
Test Jarque–Bery	10,3663 ***	9,59123 ***

Pozn. v závorkách jsou uvedeny směrodatné odchylky odhadu parametrů, které jsou odhadovány s využitím robustních odhadů (Neweyho-Westův HAC estimátor);

*, **, *** označuje hladinu významnosti 10 %, 5 %, 1 %.

níka a ten může být zklamaný. V ideálním světě přichází výzkumník s novou teorií, provede empirickou práci a ta mu potvrdí jeho domněnku. Reálný svět je jiný. Proměnnou, která by měla být statisticky významná, identifikuje náš model jako statisticky nevýznamnou nebo koeficienty, které měly být kladné, mohou být někdy záporné. Je třeba mít na paměti, že zjištění, že nějaká teorie nepopisuje realitu dostatečně přesně, je stejně hodnotné jako zjištění, že tato teorie funguje velice dobře. Nemusíme proto zoufat, když nějaké takové zdánlivě špatné výsledky dostaneme. Vždy si ovšem musíme zkontrolovat (i v případě výsledků souhlasících s nějakou teorií), že jsme postupovali korektním způsobem. Nejhorším přístupem je nějakým způsobem „šolichat“ s daty či metodami, aby to „vyšlo pěkně“ (a navíc tyto pochybné metody zatajovat a tvářit se, že jsme postupovali korektně).

Empirické výsledky mohou být někdy zavádějící. Jeden statistický test může identifikovat jednu věc, druhý naopak věc zcela opačnou. Vysvětlující proměnná, která je v jednom modelu nevýznamná, může v dalším modelu vyjít jako významná. V takovém případě toho moc neuděláme a tyto výsledky ve větší počestnosti zveřejníme (nevybereme si jen ten výsledek, který se nám hodí) a pokusíme se v rámci možností porozumět tomu, proč takováto nejasnost vzniká a jaké může mít logické vysvětlení.

7 Záver

Podakovanie

Na záver by som chcel poďakovať Ing. Danielovi Němcovi, Ph.D. za návrh na vypracovanie tejto témy a za veľmi príjemné a užitočné konzultácie. Ďalej by som chcel poďakovať Ester Železnákovéj za gramatickú korektúru textu.

Literatúra

- [1] Brockwell, P.; Davis, R.: *Time Series: Theory and Methods*. Praha: Springer-Verlag New York, 2009, ISBN 978-0-387-97429-3.
- [2] Tomáš, C.: *Finanční ekonometrie*. Ekopress, 2008, ISBN 978-80-86929-43-9.

A Přílohy

Zde můžete uvést přílohy.