Series on Advanced Economic Issues
Faculty of Economics, VŠB-TUO

Andrea Kolková

PROGNÓZOVÁNÍ POPTÁVKY S VYUŽITÍM GOOGLE TRENDS DAT

Ostrava, 2025

Andrea Kolková

Department of Finance
Faculty of Economics
VŠB-Technical University Ostrava
17. listopadu 2172/15
708 00 Ostrava-Poruba, CZ

andrea.kolkova@vsb.cz

Recenze

Rastislav Kotulič, University of Presov

Boris Popesko, Tomas Bata University in Zlín

Zoltán Rózsa, Alexander Dubček University of Trenčín in Trenčín

The text should be cited as follows: Kolková, A. (2025). *Prognózování poptávky s využitím Google Trends dat*, SAEI, vol. 70. Ostrava: VSB-TUO.

© VŠB-TUO 2025
Printed in VSB-TUO
Cover design by EkF VSB-TUO

This work is licensed under Creative Commons Uveďte původ 4.0 Mezinárodní.

ISBN 978-80-248-4798-6 (print)

ISBN 978-80-248-4799-3 (on-line)

DOI 10.31490/9788024847993

Předmluva

V současné době už strategie podniku nemůže být jen výsledkem statistické analýzy a stále se opakujících učebnicových metod, jako je SWOT, PEST analýza, Porter a jiné, i když i ty mají samozřejmě v podnikání své opodstatnění. Moderní strategie v sobě musí mít i kus odvahy jít dopředu, objevovat nové cesty. Už Tomáš Baťa se ve svých strategiích a manažerských rozhodnutích opíral o myšlenku: „nebát se budoucnosti a nezbožňovat minulost." Moderní strategické a taktické řízení dnes už musí obsahovat i prognózování. V podnikové ekonomice se o určité předpovědi opírá téměř každé manažerské rozhodnutí. O důležitosti prognózování svědčí i množství odkazů na toto téma na internetu. Zadáme-li do Googlu anglický pojem „forecasting", získáme 638 miliónů nejrůznějších odkazů na prognózy, předpovědi a metody s nimi spojené (Štědroň et al., 2019). Stav k 1.5.2024 je ještě o něco vyšší, a to 898 miliónů výsledků vyhledávání slova forecasting na Google.

Předpovídání budoucích veličin je už řadu let jedním ze základních problémů vědeckého výzkumu v podnikové ekonomice. Historie ukazuje, že lidstvo chtělo od nepaměti nahlédnout pod pokličku budoucnosti, a to nejen v oblasti podnikání. První pokusy o prognózování v oblasti ekonomie nebo podnikání můžeme najít už před naším letopočtem a jsou popsány i v Bibli. V první knize Mojžíšově Josef vykládal sny faraonovi a předpověděl možná poprvé hospodářské cykly. Faraonovi se tehdy zdálo, že z Nilu vystoupilo sedm nápadně krásných a tučných krav, poté za nimi z Nilu vystoupilo sedm krav nápadně ošklivých a hubených. Josef tyto tučné a hubené krávy definoval jako 7 let dobrých a 7 let zlých. O náročnosti prognózování také vypovídá citát židovského proroka Izaiáše, který dokonce už 700 let před naším letopočtem řekl: „o budoucnosti nám povězte, ať poznáme, že jste bohové!" (Bible 21. století, 2009).

V současnosti je mnoho metod prognózování a řada výzkumníků, ať už akademiků nebo vědeckých pracovníků jednotlivých podniků, se věnuje neustálému zdokonalování a vývoji dalších metod. V této publikaci je pojednáno o těchto moderních metodách vyvíjených jak v praxi, tak v akademické sféře.

S čím ovšem výzkumníci často bojují a co je jedním z největších omezení pro rozvoj výzkumné práce v této otázce, jsou data. Podniky často svá data nechtějí poskytovat, a to zejména ze strachu z prozrazení některých obchodních tajemství či taktik. Když už někteří data poskytnou, trvají na jejich nezveřejnění, případně utajení. To může stačit pro diplomovou nebo bakalářskou práci, ale vědecký výzkum má smysl jen tehdy, když je zveřejněn. Těžko si lze představit

smysluplnost vědecké práce, která by nebyla publikována a zůstala „v šuplíku“. Přitom výzkumníkům často vůbec nejde o konkrétní podnik, ale hlavně o vývoj nové metodiky, případně ověření její funkčnosti a srovnání stávajících. Z tohoto pohledu může být východiskem využití Google Trends dat (dále jen GT data), na kterých stojí tato publikace.

Tato data nepodléhají žádnému obchodnímu tajemství, neprozrazují žádnou obchodní taktiku. Dávají nám informaci o vyhledávání pojmů, zboží, případně kategorií atd., jejich popis je opět součástí této publikace. GT data nám nikdo nemůže „zakázat“ používat a jsou veřejně dostupná zdarma. Tato publikace se snaží prokázat, že využití GT dat pro vědecké účely má své opodstatnění. Stejně tak je mohou využívat podniky pro prognózování poptávky, jak už individuální, tak tržní. A součástí této publikace je i popis možnosti využít GT data k odhadu poptávky u konkurence.

Je nutné také poznamenat, že rozvoj vědeckých metod v této oblasti by nebyl dost dobře možný bez adekvátní podpory vhodným počítačovým softwarem. Proto i v této publikaci již nevystačíme s klasickými programy jako Excel, SPSS aj. Zejména metody založené na umělé inteligenci jsou již natolik složité, že vyžadují statistické programovací jazyky. V této publikace je využit jazyk R a Python.

Monografie je zaměřena na skupinu kvantitativních metod prognózování poptávky a je výsledkem několikaletého výzkumu. Opírá se o již publikované výstupy i o nové, dosud nepublikované poznatky. Je výsledkem fascinace aplikovanou matematikou, statistikou i informatikou a přesahu prognózování do běžného ekonomického života.

Monografie je rozdělena na sedm částí. První část se věnuje základním metodologicko-teoretickým východiskům, jako je definování logistiky a popis prognózování poptávky. Pojem poptávka také uvádí do kontextu logistiky a logistického řízení, pro něž je prognózování poptávky stěžejním nástrojem. Další kapitola tvoří základní rámec současného vědeckého výzkumu v kvantitativních metodách prognózování poptávky. Také definuje myšlenkový postup při prognózování a jeho vyhodnocování, používaný jak v odborné literatuře, tak i konkrétně v této publikaci. Třetí kapitola se týká definice GT dat i způsobu jejich tvorby, dále také deskripce a dekompozice dat. V této kapitole se také již věnujeme detailní grafické, statistické analýze i dekompozici dat použitých přímo v této publikaci. Čtvrtá kapitola seznámí čtenáře s kvantitativními metodami prognózování, a to v členění na statistické metody, metody inspirované přírodou, včetně metod umělých neuronových sítí, metody kombinované a hybridní a metody vyvinuté v současné praxi. Na závěr kapitoly jsou popsány míry přesnosti v prognózování použité v této monografii. Některé příklady řešené jazykem R jsou v přílohách doplněny o kódy a čtenář si tak může vyzkoušet uvedené výpočty na vlastních reálných datech. Pro lepší orientaci v textu jsou označení balíčků a funkcí v jazyce R a Python psány kurzívou. Kokrétní kódy jsou pak psány ve fontu Courier New.

Stěžejní jsou kapitoly pátá až sedmá. Ty již aplikují možnosti využití GT na konkrétní prognózu poptávky, a to nejprve individuální poptávky, pak i poptávky po vybraných kategoriích vyhledávání (respektive tržní poptávky). Sedmá kapitola pak popisuje, jak lze využít prognózování na základě GT dat pro analýzu konkurence.

Publikaci lze především doporučit vědeckým pracovníkům, kteří pokračují ve vědeckých otázkách v této monografii nastolených. Také vědcům, kteří pracují na vývoji nových metodologií v oblasti podnikové ekonomiky a potýkají se s problémy nedostatku dat. Publikace může být zajímavá i pro odborníky z praxe, zabývajícími se predikcemi. V neposlední řadě ji mohou využít studenti navazujícího magisterského či doktorského studijního programu jako podkladový materiál ke svým diplomovým či disertačním pracím na toto téma.

Andrea Kolková, Ostrava, květen 2025

Obsah

Podrobný obsah

Seznam vybraných zkratek

AHP metody	Metody analytického hierarchického procesu
AIC	Akaikeovo informační kritérium
ARFIMA	Autoregressive fractionally integrated moving average
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
AutoML	Automatizované Machine Learning
BATS	Box-Cox Transformation, ARMA residuals, Trend and Seasonality
BIC	Bayesovské informační kritérium
EDA	Exploratory data analysis
ETS	Exponential Smoothing
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
MATLAB	MATrix LABoratory
ME	Mean error
MPE	Mean Percentage Error
Nnetar	Neural networks
RMSE	Root mean Squared Error
SAS System	Statistical Analysis System
SPSS	Statistical Package for the Social Sciences
Stata	zkratky slov statistika a data
SWOT	Strengths, Weaknesses, Opportunities, Threats
Tbats	Trigonometric Box-Cox Transformation, ARMA residuals, Trend and Seasonality
USD	Americký dolar
SES	Simple Exponential Smoothing
N	počet hodnot
y	závisle proměnná

$\bar{y}$	střední hodnota závisle proměnné
$\hat{y}$	regresní odhad $i$tého pozorování
R^2	koeficient determinace
MdAPE	Median Absolute Percentage Error
RMSPE	Root Mean Square Percentage Error
RMdSPE	Root Median Square Percentage Error
sMAPE	Symmetric Mean Absolute Percentage Error
sMdAPE	Symmetric Median Absolute Percentage Error
MRAE	Mean Relative Absolute Error
MdRAE	Median Relative Absolute Error
GMRAE	Geometric Mean Relative Absolute Error
NNETAR	Neural Network
ACF/PACF	Autocorrelation function/Partial autocorrelation function
ℓ_t	úroveň časové řady v čase t
b_t	sklon v čase t
s_t	sezónní složka časové řady v čase t
m	počet ročních období v roce
α, β^*, γ a ϕ	vyrovnávací parametry

Kapitola 1

Prognózování poptávky

Poptávka je specifická veličina, jejíž prognóza je pro podnikové řízení zásadní a na kterou se váže řada dalších podnikových činností. Jak uvádí Macurová et al. (2018), schopnost dobře predikovat poptávku je velmi důležitá pro plánování a řízení všech dalších činností v logistickém řetězci. Predikce poptávky (demand forecast) může představovat odhad nejen velikosti, ale i struktury budoucích odbytových požadavků.

V minulém století byl zaznamenán obrovský rozmach prognózování poptávky pomocí kvantitativních metod. V té době vznikaly Brownův model (Brown, 1959) a Holtův model (Holt, 1957). Už v tomto století následovali významní prognostičtí autoři jako Armstrong (2001), Bergmaier et al. (2016), Taylor (2003) a samozřejmě Hyndman & Athanasopoulos (2018). Tématem se také zabývá řada výzkumů i po roce 2020, například Mansoor et al. (2021), Li et al. (2021), Nikolopoulos et al. (2020), Pandey et al. (2021) nebo Guo et al. (2021). Z praxe jsou také známé další výzkumy (Shaub, 2020), (Smyl, 2020). Je zřejmé, že tento vývoj bude dynamicky pokračovat a zájem o prognózování poptávky kvantitativními metodami neustane.

Základním zdrojem informací v této monografii jsou zejména předchozí výzkumy, a to publikace ve vědeckých impaktovaných časopisech (Kolková & Ključnikov, 2022), (Kolkova & Rozehnal, 2022), (Kolková & Navrátil, 2021), (Kolková & Ključnikov, 2021), (Kolkova, 2020), (Navratil & Kolkova, 2019), ale také indexovaných v jiných databázích (Kolková et al., 2022), (Kolkova, 2018b). Monografie čerpá i z výsledků prezentovaných na mezinárodních vědeckých konferencích (Kolková & Macurová, 2022), (Kolková, 2020b), (Kolková, 2019). Samozřejmými dalšími zdroji jsou výsledky ostatních vědeckých pracovníků publikovaných zejména v databázích Web of Science nebo Scopus, všechny jsou uvedeny v seznamu literatury a v relevantních částech knihy.

1.1 LOGISTIKA A PREDIKCE POPTÁVKY

Výraz logistika je odvozen z řeckého slova „logistikon“, které znamená rozum či důmysl. Někteří autoři uvádějí, že by mohl mít původ ve slově „logo“, což znamená také rozum, ale i řešení, slovo, myšlenka, úsudek či zákon. Nejprve byla

využívána logistika jen ve vojenství. Zde to byla nauka o zásobování vojsk, jejich přesunech i ubytování. Z těchto principů se pak přesunula i do běžného podnikání.

Předmětem logistiky jsou v současné teorii fyzické, ale i informační a peněžní toky, které se uskutečňují při uspokojování požadavků po produktech (Macurová et al., 2018). Tokem se rozumí posloupnost stavů pohybu a stavů klidu. Jednotlivé fyzické, informační a peněžní toky jsou na sobě často vzájemně závislé.

Hlavními logistickými aktivitami jsou:

- predikce poptávky,
- navrhování logistického řetězce,
- nákup a zpracování objednávek od zákazníků,
- řízení zásob, manipulace s materiálem,
- plánování a řízení výroby a služeb, balení, skladování, doprava,
- řízení zpětných toků a prodejní podpora.

Tento výčet logistických aktivit zahrnuje ucelený pohled na logistiku, a tedy všechny logistické aktivity v podniku. Takové pojetí nazýváme integrální logistikou a je preferováno i ve výuce na VŠB-TU Ostrava. Logistiku tak můžeme definovat jako disciplínu, která se zabývá celkovou optimalizací, koordinací a synchronizací všech činností.

Důležitý pojem je i logistické řízení. Představuje organizování a usměrňování toků i vykonávání integrační, koordinační a synchronizační funkce za účelem dosažení logistických cílů. Důležitým předpokladem správně fungujících logistických procesů je v moderním pojetí také digitalizace podniku.

Predikce poptávky je tak považována v moderní logistice za jednu z hlavních aktivit, kterými se logistické řízení musí zabývat. Současně s tím jsou kladeny zvýšené požadavky na digitalizaci tohoto prvku logistického řízení.

1.2 TYPOLOGIE LOGISTICKÉ POPTÁVKY

Typologie poptávky je poměrně široká a jednotlivé typy obvykle rozdělujeme podle jejich zásadních charakteristik. Rozlišujeme tak dělení podle vztahů v logistickém řetězci, původu poptávky, podle časového průběhu poptávky, podle toho, kde se v logistickém řetězci sledované firmy nachází, podle typu časové řady, která zobrazuje vývoj hodnot poptávky v minulých obdobích. Většinu těchto členění lze vyčíst z dekompozice časové řady historické poptávky, která je diskutována v kapitole 3.

Podle vztahu v logistickém řetězci můžeme poptávku členit, jak uvádí Macurová et al. (2018), na Business to Business (dále jen B2B) a Business to Costumers (dále jen B2C) poptávku. B2B je poptávka po produktech, které jsou určeny k dalšímu prodeji a k dalšímu zpracování. Poptávka na B2C trzích je poptávka po výrobcích a službách určených koncovým zákazníkům.

Podle původu můžeme poptávku členit na nezávislou, odvozenou a závislou. Nezávislá poptávka, označována také jako stochastická, je typická pro poptávku po konečných výrobcích. Podnik ji nemůže přímo ovlivnit a musí být predikována některou z prognostických metod.

Poptávka odvozená je dána tou částí logistického řetězce, který není přímo spojen s konečným zákazníkem. Jedná se tedy o B2B obchody, kdy dodávající může jednoduše spočítat budoucí poptávané množství. To je dáno smluvním vztahem, případně doplněno o různé marketingové akce, jako množstevní sleva, navýšení dodávek v důsledku promo akcí atd. Tyto výkyvy jsou ale obvykle krátkodobého charakteru a odvozená poptávka bývá velmi předvídatelná a stabilní.

Poptávka závislá může být ještě dále rozdělena na vertikálně závislou a horizontálně závislou. **Vertikálně závislá** je poptávka po součástech jednotlivých finálních produktů. Z těchto součástí se pak finální produkt tvoří. Pokud tedy chápeme poptávku po finálním produktu jako nezávislou a odhadneme-li ji například některou z dále uvedených metod, poptávka po dílech na tento produkt už bude vertikálně závislá, a tedy předvídatelná. Druhou závislou poptávkou je **horizontálně závislá** a ta se týká nikoli dílů, ale komplementů k tomuto výrobku. V dalším textu této publikace se budeme zabývat pouze nezávislou poptávkou.

Podle časového průběhu lze poptávku dělit na spojitou a nespojitou. Spojitá poptávka bývá také někdy označovaná jako stejnosměrná, pravidelná. Požadavky na výdej zde přichází stále a trvale. Samozřejmě zde může existovat výkyv, a to jak sezónní, tak vlivem jednorázových šoků. I zde může existovat trend rostoucí, klesající i stagnující. Nicméně poptávka existuje kontinuálně.

Druhý typ je poptávka nespojitá, označovaná jako nestejnosměrná, nepravidelná či sporadická. Vzniká obvykle u položek, které vykazují závislou poptávku. Ta nevzniká trvale, ale často jen jako důsledek poptávky po jiné části produktu nebo služby. Také mnohdy není spojena s koncovým zákazníkem, ale vzniká v jiných částech logistického procesu. V dalším textu se budeme zabývat pouze poptávkou spojitou.

Členění poptávky podle toho, **kde se v logistickém řetězci podniku nachází**, rozlišuje poptávku primární a sekundární. Primární poptávka je po zbožích a službách daného podniku. Určují ji přímo zákazníci.

Sekundární poptávka je dána trhy B2B. Jedná se o prodej produktu či služby za účelem jejich dalšího prodeje či výroby. Tato poptávka, v případě dalšího prodeje produktů koncovým zákazníkům, předchází poptávku primární. Samozřejmě za předpokladu, že koncový prodejce odhadl správně svou primární poptávku. V dalším textu se budeme zabývat poptávkou primární.

Poslední, zde zmiňované členění poptávky, je podle **typu časové řady**. Z tohoto pohledu může být poptávka s trendem, poptávka stagnující, poptávka sezónní či cyklická a nepravidelná poptávka. Poptávka s trendem je taková, která vykazuje systematický růst či pokles v čase.

Stagnující poptávka se naopak v čase nemění, existuje zde pouze rozptyl kolem střední hodnoty. Ten může být způsoben jak náhodnými vlivy, různými ekonomickými, polickými či sociálními šoky, ale také cyklickými výkyvy. Například pravidelně se opakující cyklus v průběhu roku nebo v ročních obdobích.

Dalším typem je cyklická poptávka. Její specifickou formou je poptávka sezónní, což je cyklická poptávka s roční periodou. To znamená, že vykazuje stejné výkyvy v průběhu jednoho kalendářního roku, například zvýšení prodeje před Vánoci či snížení prodeje v průběhu letních prázdnin.

Samozřejmě tyto poptávky mohou tvořit i kombinace, takže může existovat cyklická poptávka s rostoucím či klesajícím trendem, nebo naopak stagnující poptávka s cyklickými výkyvy. V této monografii bude odhad poptávky podle typu časové řady proveden na základě dekompozice, která je diskutována v kapitole 3.4.

Členění využité v publikaci Fildese a Kolassy (2022) klade důraz na účel prognózování pro řízení, a tak se můžeme setkat s členěním na strategickou úroveň, taktickou úroveň a provozní úroveň. Další možnost rozlišuje poptávku na úrovni trhu, na úrovni řetězce a kanálu, na úrovni prodejny, na úrovni produktu v maloobchodě a další podrobná členění.

V této monografii bude využito pouze členění na **individuální** a **tržní poptávku**. Monografie ve svých dalších částech realizuje prognózu individuální poptávky (kapitola 6) a prognózu tržní poptávky, kterou nalezneme v kapitole 7. Individuální poptávkou pro účely monografie rozumíme poptávku po konkrétním produktu nebo službě. Poptávka tržní je obecná poptávka po produktech nebo službách v rámci dané kategorie u všech prodejců.

1.3 HISTORIE PROGNÓZOVÁNÍ POPTÁVKY

Slovo prognóza pochází z řečtiny a skládá se ze dvou částí: „pro“, což znamená před, a „gnosis“, která vyjadřuje poznání. Prognózu tedy můžeme přeložit jako určitou výpověď o budoucím stavu objektivní reality. Je nutné ji ale odlišit od běžné předpovědi či věštění. Na rozdíl od nich se prognóza obvykle opírá o vědecké poznatky. Počátky prognózování jsou známy z dob starověkého Babylonu, kdy vznikaly tzv. delfy, což byly antické věštírny. Ve 40. letech 20. století pak bylo prognózování rozvíjeno v souvislosti s válečnými konflikty a vojenským plánováním.

Historie prognostiky v českých zemích je poměrně krátká a sahá pouze do 60.–70. let minulého století. Její zařazení mezi samostatné vědecké disciplíny není jednoznačné. Samotné definování prognostiky od svého vzniku je také velmi různorodé.

Například Holcr (1981) definuje prognózu jako formu předpovědi, která splňuje určité požadavky, a to musí obsahovat časový nebo prostorový interval, v kterém se prognózovaný jev uskuteční nebo bude objevený, dále interval musí být konečný a musí existovat principiální možnost apriorního stanovení

pravděpodobnosti uskutečnění prognózovaného jevu, prognózovaný jev také musí být verifikovatelný, a konečně prognóza musí být formulovaná úplně přesně a jednoznačně.

Gál (1999) definuje prognózu jako podmíněnou, o vědecké poznatky se opírající výpověď o budoucnostech nějakého objektu či jevu. Dle Wisniewského (1996) je smyslem prognózování snížení neurčitosti znalostí o budoucnu a manažerům poskytne dodatečné informace, které umožní posoudit alternativní možnosti v kontextu s budoucími podmínkami a vyhodnotit budoucí důsledky současných rozhodnutí. Modernější přístupy k prognózování pak uvádějí definici prognózy jako transformace minulé zkušenosti do očekávané budoucnosti.

Vincur a Zajac (2007) definují prognostiku jako „vědeckou disciplínu, jejímž předmětem je studium technických, vědeckých, ekonomických a sociálních faktorů a procesů, které působí na vývoj objektivní reality světa, a která má za cíl vytvoření představy – prognózy o jeho budoucím stavu vyplývajícím z propojených vlivů těchto faktorů a procesů“.

V současné době je pojem prognózování často nahrazován pojmem prediktivní analýza nebo prediktivní analytika. Je definována dle McGregora (2013) a Siegela (2015) takto: „prediktivní analýza je analytický proces, který začíná výběrem dat, po kterém následuje zkoumání dat, analýza a vizualizace sloužící ke zlepšení obchodních a firemních procesů“. Pomocí prediktivní analytiky je hledán vztah mezi proměnnou, která je známá, a proměnnou, která je predikována na základě dat nasbíraných v minulosti. Jak uvádí McGregor (2013), prediktivní analytika se dotýká více druhů kvantitativní analýzy, a to jak statistiky, tak operačního výzkumu a strojového učení či data miningu. Přidáme-li jako zkoumanou veličinu podnikové faktory, lze označit výstup jako ekonomickou prediktivní analytiku. V této monografii jsou slova predikce a prognózování považována za synonyma.

V ekonomice lze prognózovat celou řadu podnikových veličin. Marček (2016) definuje mnoho oblastí, kde se prognózy podnikových veličin uplatňují. Prognózy tržeb jsou řešeny již v několika publikacích, maloobchodní prodej řešil Alon et al. (2001), Chu a Zhang (2003) nebo Pereira et al. (2016), prodeje osobních automobilů prognózovali Pavelková a Homolka (2018). Tržby v potravinářství byly predikovány i alternativně dle Kolkové (2018a, 2018b). Vedou se diskuze o budoucích směrech výzkumu (Gašparíková, 2007) a samozřejmě jsou tvořeny predikce makroekonomických ukazatelů (Radvanský et al., 2010).

V době covidové pandemie se slovo prognóza objevovalo v médiích téměř denně. Samozřejmě, v této monografii nelze postihnout lékařské prognózy, které zahrnují i řadu jiných faktorů. Nicméně i ekonomické prognózování prošlo velkou obměnou. A v současné době už existují studie, zda ekonomický šok, jakým pandemie bezesporu byla, dokáže zasáhnout do přesnosti jednotlivých prognostických metod (Kolkova & Rozehnal, 2022).

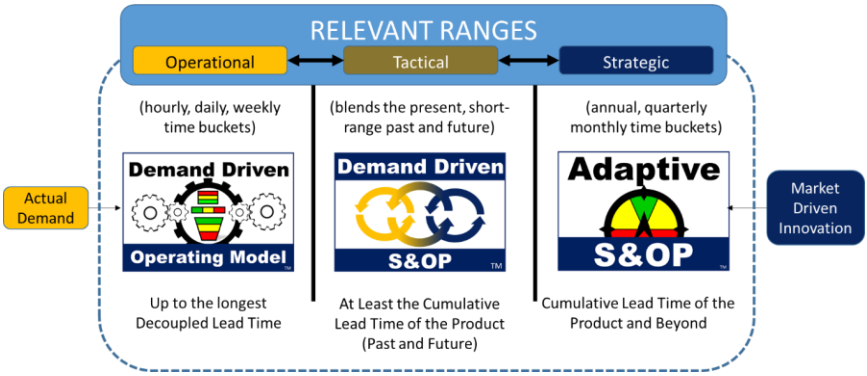
1.4 NOVÉ VÝZVY "POSTPANDEMICKÉ" LOGISTIKY

Je zřejmé, že postcovidové období sebou přineslo nové výzvy a nová zjištění, na které musela logistika reagovat. V předpandemickém období kladly podniky velký důraz na zeštíhlování logistických procesů, aplikovali „Just in Time", a to často beze strachu z případných výpadků. Období pandemie ukázalo, že tento koncept má i svá negativa. Ve zprávě z výzkumného instututu Capgemini (Gya et al., 2020) 57 % organizací hodlá zvýšit své investice do zlepšování odolnosti dodavatelsko-odběratelského řetězce. 66 % oslovených podniků hodlá výrazně změnit logistickou strategii, aby se přizpůsobily novému postcovidovému období. Další „zatěžkávací" zkouškou pro tradiční logistické koncepty byla a je válka na Ukrajině i jinde ve světě. Distribuční kanály, na které byly mnohé podniky zvyklé, se zhroutily a bylo třeba okamžitě měnit taktiku v takové oblasti. To samozřejmě vyvolalo mnohé změny, což klade na nové koncepty v logistice nemalé nároky.

Jednou z možností, jak této nové skutečnosti čelit, je přesnější plánování a právě prognózování poptávky. Celá oblast prognostické analytiky je nyní aplikována do praxe častěji než dříve. Souběžně s tím se zvyšuje důležitost datové analytiky. V současné době také vzniká koncept Demand-Driven Adaptive Enterprise (Ptak, 2018), česky lze tento pojem přeložit jako poptávkou řízený adaptivní podnik. Ten reaguje na nedostatečnost dosavadních konceptů a na vysokou volatilitu logistických toků v současnosti. V tomto konceptu se počítá s poptávkou řízeném provozním modelu, poptávkou je řízen samozřejmě i prodej a provozní plánování. Při naplnění tohoto konceptu už hraje predikce poptávky zásadní roli. Pokud podnik chce pracovat s adaptivními provozními modely, musí se orientovat na poptávkou řízené zásobování.

Poptávkou řízené zásobování klade velké nároky na využívaná data. Přehlednost, přesnost, ale zejména také včasnost získávání dat o materiálových tocích a dodávkách jsou zásadní pro následnou predikci poptávky. Následná výroba musí být dle ní řádně plánována a rozvrhována. Dokonce i dlouhodobé plánování by mělo být v novém pojetí Demand-Driven Adaptive Enterprise řízeno na základě poptávky. Samozřejmě tento koncept musí jít ruku v ruce s automatizací, a to jak v doplňování zásob a v materiálových tocích, ale i s automatizací procesu plánování a rozvrhování odbytu.

Poptávkou řízený adaptivní podnik je model správy a provozu navržený tak, aby se podniky mohly přizpůsobovat neustále se měnícím podmínkám podnikání a nestálému podnikatelskému prostředí. Tento model počítá s nepřetržitým systémem inovací a zpětné vazby. Je tvořen třemi komponenty, a to provozní (operativní) model sestavený na základě predikce poptávky, dále provozní model plánování prodeje a operací taktéž na základě poptávky a následně adaptivní model plánování prodeje a operací. Tento model je tvořen jako obousměrný systém a zahrnuje nyní tolik potřebné zpětné vazby, jak je znázorněno na obrázku 1-1. S&OP představuje plánování prodeje a provozu (Sales and Operations Planing), což je strukturovaný proces plánování, který využívá předpokládanou poptávku a vyžaduje aktivní spolupráci více funkčních týmů podniku (prodej, výroba, správa, řízení obchodu atd.)



Obrázek 1-1 Grafické vyjádření poptávkou řízeného adaptivního podniku

zdroj: (Ptak, 2018)

Další nepopiratelnou výzvou postpandemické logistiky je Omni-kanálová logistika, která je už v současnosti standardem v mnoha podnicích. Zde se jedná o propojení online a offline prodejních postupů, ale také logistických procesů, jako je zásobování, řízení odbytu i sledování materiálových toků. I zde je ovšem obvykle kladen velký důraz na zákaznicky orientované služby, a tedy i v budoucnu poptávka a její predikce budou hrát významnou roli.

Predikování veličin a v podniku stěžejní predikce poptávky je často souhrnně nazývána prediktivní analýza, o které už byla zmínka, a ve výrobních podnicích začíná nacházet své nezanedbatelné místo. Nicméně již nyní se začíná v odborných i podnikových kruzích hovořit o jejím nahrazení analýzou preskriptivní. Preskriptivní analýza využívá kromě standardních nástrojů prediktivní analýzy také simulace, neuronové sítě a strojové učení. V této publikaci bude dále popsána problematika neuronových sítí. Strojovému učení a neuronovým sítím se věnuje například publikace Kolkové a Navrátila (2021). Účelem preskriptivní analýzy je komplexní analytika událostí. Prediktivní analýza si klade za cíl, co a kdy se stane. Preskriptivní bude mít snahu popsat a určit, i proč se to stane. Může tak být nutným pomocníkem při identifikaci budoucích rizik, ohrožení, ale i příležitostí. Stejně tak bude preskriptivní analytika poskytovat i doporučení a zohledňovat variantnost důsledků různých rozhodnutí.

1.5 VARIABILITA DAT A PROGNÓZOVÁNÍ POPTÁVKY

Data samozřejmě také mohou být ovlivněna vnějšími okolnostmi a pro časovou řadu v různých obdobích mohou být velmi odlišná. Touto odlišností se autorka publikace zabývala v článku Cyclo-Mobility as a Contribution to Sustainable Development: Analysis and Forecast of Demand for Bicycles and their Components Before and During the Pandemic Time (Kolková & Macurová, 2022). Byla zde mimo jiné ověřována hypotéza, že pandemické období zvýšilo variabilitu poptávky. Byly aplikována data z konkrétního e-shopu zabývajícího se

prodejem jízdních kol, a to jak výnosy tohoto e-shopu, tak počty prodejů za dané období.

Z výsledků analýzy vyplynulo, že variabilita poptávky se v pandemickém roce 2021 prudce zvýšila. V tom roce byla poptávka více než dvakrát variabilnější než v letech předchozích. Bylo to způsobeno prudkým vzrůstem poptávky v jarních měsících toho roku. Jelikož se poptávka poté ustálila, je výsledná variabilita velká, jak dokládá tabulka 1-1.

Tabulka 1-1 Variabilita denních výnosů a denních prodejů

Rok	Výnosy			Prodeje		
	Průměr	Směrodatná odchylka	Koeficient variability	Průměr	Sm. odchylka	Koeficient variability
2019	138	131 155,91	0,946	34,73	23,15	0,667
2020	190	187 960,54	0,985	51,14	39,38	0,770

Zdroj: Kolková & Macurová, 2022

V článku byla také ověřována hypotéza, že pandemické období změnilo cyklické výkyvy v poptávce během roku. Tuto hypotézu ovšem nebylo možno potvrdit. V roce 2019, 2020 i 2021 v týdenní dekompozici můžeme spatřovat vyšší poptávku uprostřed týdne, případně v jeho druhé půli. Při dekompozici měsíční dokonce můžeme sledovat naprosto shodný cyklus. Poptávka roste druhý a třetí týden v měsíci. To samozřejmě může souviset s výplatními termíny, které v ČR obvykle bývají kolem 10. dne v měsíci. Tuto hypotézu jsme tedy nepotvrdili a lze konstatovat, že pandemické období nezměnilo cyklickou složku poptávky. Lidé si i v pandemii zachovali stejné cyklické nákupní chování.

Dalším významným diskuzním tématem je prognóza poptávky, která z článku vyplynula. Na základě metody ETS bylo na obou skupinách dat potvrzeno, že poptávka bude stagnovat. Z hlediska udržitelnosti by samozřejmě byl vhodnějším výsledkem konstantní nárůst. Je nutno si ale všimnout, že poptávka v předkoronavirovém odvětví byla také stagnující a nyní se ustálil stejný trend, ale na vyšší úrovni. Je tedy možné konstatovat, že pandemie koronaviru trvale zvýšila poptávku po kolech, která v následujících měsících bude mít stagnující trend. Tento článek byl napsán na začátku roku 2022 a další vývoj v roce 2022 tuto prognózu opravdu potvrdil. Variabilita dat tedy byla v tomto článku klíčovým prvkem. Samozřejmě v roce 2024 lze říci, že poptávka po kolech začala klesat. Tak daleko ovšem prognóza nedosahovala.

V článku Hybrid demand forecasting models (Kolková & Rozehnal, 2022) byla data zkoumána v jiném kontextu a to, zda různé modely na datech v období předpandemickém vykazovaly vyšší míru přesnosti než na datech v období během pandemie.

Jako první byla zkoumána hypotéza, že míra přesnosti všech modelů je v době pandemie horší. Tato hypotéza byla potvrzena, což koresponduje i s Cyclo-Mobility as a Contribution to Sustainable Development: Analysis and Forecast of

Demand for Bicycles and their Components Before and During the Pandemic Time (Kolková & Macurová, 2022), a je tedy dáno zejména vysokou variabilitou dat. Druhá hypotéza pak předpokládala, že v předpandemickém období budou hybridní metody nejpřesnější. To se také potvrdilo. Hypotéza třetí pak předpokládala, že v pandemickém období zůstanou nejpřesnější metody statistické, a to právě z důvodu vysoké variability dat, které mohou metody založené na umělé inteligenci ovlivnit více než metody statistické. Tato hypotéza však potvrzena nebyla a z výsledků vyplynul fakt, že i v dynamickém a více variabilním pandemickém období vykazovala největší přesnost metoda ze skupiny metod hybridních.

Tento výsledek lze považovat za klíčový a vyvodit z něj závěr, že má smysl se zabývat novými moderními metodami prognózování i v dynamických obdobích. V roce 2022 a 2023, kdy data byla ovlivněna pandemií koronaviru a následně válkou na Ukrajině, je žádoucí a vhodné používat moderní metody vyvinuté v praxi i v akademické sféře a nezavrhovat machine learning modely a vůbec modely založené na umělé inteligenci.

Kapitola 2

Kvantitativní metody prognózování poptávky ve výzkumu a praxi

V prognostické praxi se uplatňuje v současnosti celá řada různých přístupů a metod. Ty se opírají jak o kvantitativní, tak o kvalitativní metody podnikového výzkumu. Ústředním tématem této monografie jsou proto právě kvantitativní metody prognózování poptávky.

V ekonomickém výzkumu existuje řada kvantitativních metod. Kvantita znamená mnohost, četnost, množství, velikost. Filozoficky je to cokoliv, na co se ptáme otázkou „kolik“. Je to tedy vlastnost, kterou lze změřit a vyjádřit číslem. Z tohoto pohledu můžeme kvantitativní metody kvalifikovat jako metody standardizovaného vědeckého výzkumu, které popisují zkoumanou skutečnost pomocí proměnných, které lze vyjádřit čísly. Dle Bella et al. (2018) „kvantitativní výzkum lze charakterizovat jako projevující určité zájmy, z nichž nejdůležitější jsou: měření, kauzalita, zobecnění a replikace”. Nejjasnější charakteristikou kvantitativních metod je měřitelnost. V kvantitativním výzkumu také vědci často doufají v generování obecně platných pravd, samozřejmě omezených konkrétním kontextem (Bell et al., 2018). V přírodních vědách pak v kvantitativním výzkumu často existuje požadavek replikace (respektive reprodukce) daného experimentu (Chalmers, 2013). Kvantitativní metody nacházejí své uplatnění zejména ve vědách přírodních. Nicméně i vědy humanitní široce používají kvantitativní metody. Například v sociologickém výzkumu mají kvantitativní metody nezastupitelnou roli (Bhattacherjee, 2019). Totéž ale platí i pro výzkum ekonomický, který kvantitativní metody zcela samozřejmě používá ve velké části výzkumných studií. Při prognózování poptávky mají nezastupitelnou roli, a to i přesto, že často bývají doplněny metodami kvalitativními.

2.1 KVANTITATIVNÍ METODY PROGNÓZOVÁNÍ JAKO PŘEDMĚT VĚDECKÉHO VÝZKUMU

O významu používání kvantitativních metod v podnikání svědčí již výzkumy z minulého století (Wisniewski, 1996). Již tehdy byl podíl podniků využívajících kvantitativní metody 66 % s tím, že 24 % firem uvádí, že užitek z těchto metod je velmi vysoký, pouze 0,7 % respondentů v tomto výzkumu uvedlo, že užitek není žádný. V té době z podniků aplikujících kvantitativní metody nejvíce manažerů je používalo k základní a popisné statistice, diskontování cash flow, řízení kvality a zásob. Asi 67 % podniků používalo také rozhodovací analýzy, metody vyrovnání, více než 50 % takových podniků také používalo simulace či regresní analýzy. Lze samozřejmě předpokládat, že s rozvojem výpočetní techniky se využití kvantitativních metod v podnikové ekonomice ještě zvýšilo. S rozvojem výpočetní techniky také roste počet i složitost použitých metod a modelů k prognózování podnikatelských veličin. Dnes již můžeme prognózy postavit na fuzzy logice, umělých neuronových sítích, genetických algoritmech či teorii chaosu (Dostál et al., 2005).

V podnikové ekonomice se v současné praxi i teorii využívá celá řada matematicko-statistických metod. Historické řízení podniků „selským rozumem“ tak již dávno doplňuje, respektive nahrazuje škála metod kvantitativního rozhodování. V podnikové ekonomice pak lze využít takřka všechny metody analýzy dat, jako jsou:

- metoda hlavních komponent,
- faktorová analýza,
- korelační analýza,
- regresní analýza,
- analýza rozptylu,
- analýza kovariance,
- diskriminační analýza,
- logistická regrese, analýza kategoriálních dat,
- shluková analýza,
- vícerozměrné škálování,
- analýza conjoint a celá řada dalších.

Výzkumy, které by se zabývaly přímo prognózováním poptávky v praxi, jsou spíše sporadické. Jak uvádí Fildes et al. (2022), žádné nedávné průzkumy prognostické praxe v maloobchodě nebyly provedeny od doby Petersona (1993), který zjistil omezené použití ekonometrických (kauzálních) metod: nejvíce byl používán manažerský úsudek, následovaný jednoduchými jednorozměrnými metodami. Populární byly zřejmě i testovací trhy a průzkumy. McCarthy et al. (2006) došli k téměř stejnému závěru ve studii, která zahrnovala maloobchod.

Na tyto výzkumy dlouho nebylo navázáno. Ve výzkumu v podnikové ekonomice tedy často chybí zpětná vazba z podniků a výzkumníci často neví, zda jejich výzkum je pro praxi přínosem. Výzkumníci z katedry podnikohospodářské a katedry informatiky Ekonomické fakulty VŠB–Technické univerzity Ostrava se tento nedostatek pokusili v roce 2022 zmírnit a realizovali rozsáhlé dotazníkové šetření v rámci Studentské grantové soutěže. V průběhu února až června 2021 probíhal výzkum (Kolková et al., 2022), který měl odhlalit potřeby firem na poli vědeckého výzkumu v podnikové ekonomice. Cílem tohoto výzkumu je zacílit vědecké aktivity v oblasti kvantitativních metod v podnikové ekonomice přímo na potřeby firem, a to jak velkých, tak malých a středních. Výzkum probíhal formou on-line dotazníkového šetření. V něm bylo osloveno celkem 3150 firem. Návratnost dotazníků však byla jen 124. V případě, že podniky nejsou motivovány (finančně, odměnami, nabídkou spolupráce atd.), je takto nízká návratnost zcela běžná.

Podniky byly rozděleny dle základních charakteristik, a to podle velikosti, zahraničního podílu a sektoru podnikání. Z výsledků vyplynulo, že 70 % podniků opravdu nějakou kvantitativní metodu používá. Ovšem pouze 37 % manažerů odpovědělo, že výsledkům těchto metod vždy naprosto jasně rozumí. Byla také statistickými testy potvrzena významná závislost neporozumění statistickým metodám na velikosti podniku. Významně více nepochopení bylo mezi malými a středními podniky. Závislost na zahraničním podílu nebo sektoru podnikání a porozumění kvantitativním metodám potvrzena nebyla. Síla závislosti, respektive síly asociace byly v tomto výzkumu měřeny několika ukazateli, jako je χ^2, koeficient Lambda, Cramerovo V, Somersovo D, Kendalovo τ_c a koeficient Gamma. Míra závislosti, kterou tento výzkum vyčíslil, je uvedena v tabulce 2-1.

Tabulka 2-1 Výsledek závislostí mezi parametry podniku a využití a porozumění kvantitativním metodám.

	Velikost podniku	**Podíl zahraničního kapitálu**	**Sektor**
Využití kvantitativních metod	Závislost střední	Závislost nízká	Nízká až střední
Porozumění kvantitativním metodám	Závislost triviální až žádná	Závislost triviální až žádná	Nízká až střední

Zdroj: Kolková et al., 2022

Po dotazu manažerům podniků proč kvantitativní metody nevyužívají, označili jako hlavní důvod neexistenci pracovníka, který by data zpracovával, a vysokou cenu outsorcingu těchto služeb. Hlavním důvodem nepoužívání kvantitativních metod u SME pak bylo, že metody jsou příliš akademické a často je jejich využití v praxi nereálné pro svou složitost. Dalším častým důvodem je nedostatek času. V tabulce 2-2 jsou výsledky za všechny podniky dohromady a

výsledným důvodem je opět přílišná akademičnost metod a nereálnost jejich použití v běžné podnikové praxi.

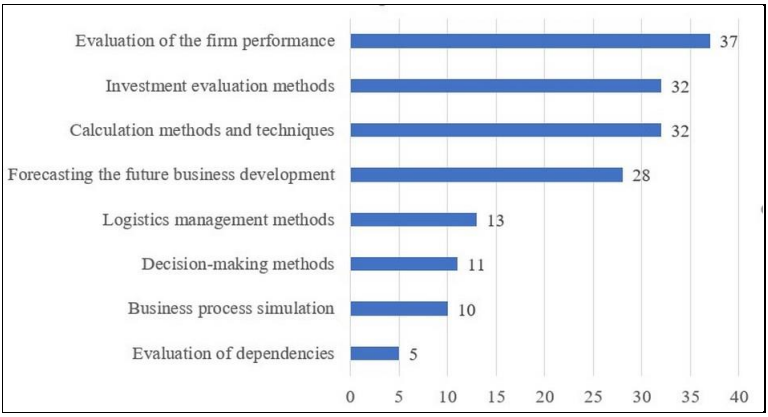
Tabulka 2-2 Důvody nevyužívání kvantitativních metod v podnikové praxi**.**

Výsledky metod jsou nesrozumitelné a složité.	3
Nemáme pracovníka, který by metody uplatňoval v praxi. O outsorcingu v rámci těchto metod neuvažuji z důvodu mé neochoty poskytovat data 3. osobám.	12
Nemáme pracovníka, který by metody uplatňoval v praxi. O outsorcingu v rámci těchto metod neuvažuji z důvodu vysoké ceny těchto služeb.	22
Tyto metody mi připadají pro podnikání zbytečné a neužitečné.	8
Nemám zájem se učit interpretovat nové metody a postupy.	5
Na kvantitativní vyhodnocování svých dat nemám čas.	18
Metody jsou příliš akademické a často je jejich použití v podnikové praxi pro svou složitost nereálné.	27

Zdroj: Kolková et al., 2022

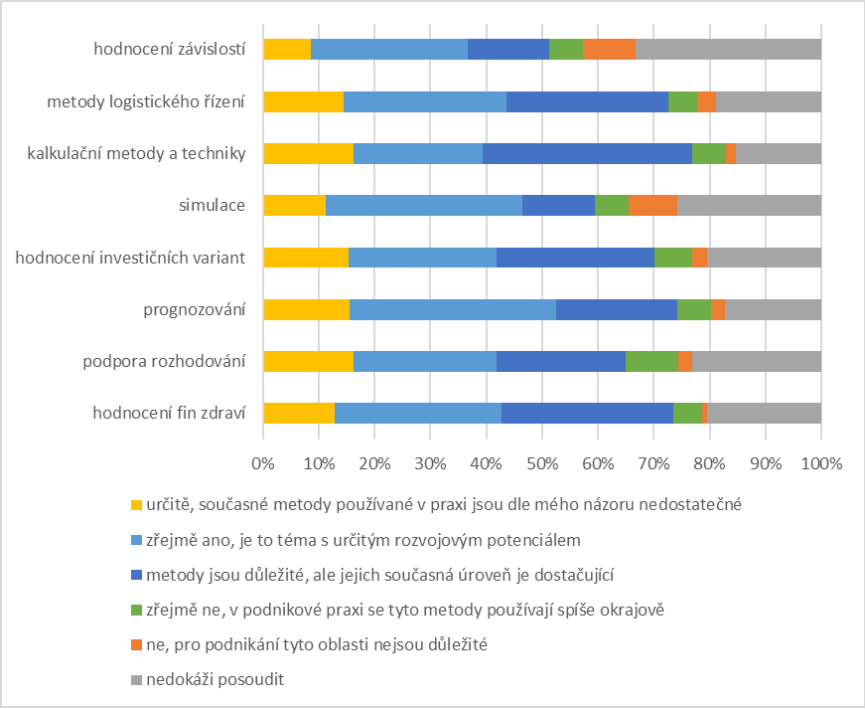
V souladu s tím jsou i výstupy (Fildes et al., 2022). Autoři definují, že navzdory dostupnosti komerčního softwaru s aplikací sofistikovaného modelování není zcela jednoznačné, zda tyto nástroje skutečně poskytují uživatelům výhodu úměrnou vyšším nákladům, které jsou s jejich koupí spojeny. Dalším problémem bývá prognóza nových produktů, která stále často zůstává opřena o kvalitativní metody. Také se ve studii zabývali znalostmi IT odborníků z řad maloobchodu a došli k závěru, že i když jsou vědomosti v oblasti IT vysoké, na informovanost v oblasti prognóz je v maloobchodu kladem příliš malý důraz.

Podniky, a to jak velké, tak SME, jako nejčastější kvantitativní metodu používají hodnocení podnikového výkonu (metody finanční analýzy), metody hodnocení investic a kalkulační metody a techniky. Naproti tomu podnikové simulace nebo hodnocení závislostí podnikových veličin využívá minimum firem. Prognózování je však u většiny firem stále v pozornosti, i když nikoli na jejím vrcholu. Kompletní výsledky výzkumu jsou uvedeny na obrázku 2-1.



Obrázek 2-1 Kvantitativní metody využívané v podnikové praxi.
Zdroj: Kolková et al., 2022

Nicméně na výzkumnou otázku, co podniky považují za nejnutnější k podrobení dalšímu výzkumu, bylo prognózování budoucího vývoje podnikových veličin nejčastější odpovědí, jak uvádí obrázek 2-2. Následují simulační metody a metody logistického řízení. Naopak metody a techniky kalkulací a hodnocení finančního zdraví firmy jsou respondenty hodnoceny jako důležité, ale jejich současná úroveň je postačující. Hodnotíme-li z druhé strany spektra vědeckého potenciálu, tak nejméně potřebné metody jsou dle respondentů hodnocení závislostí a simulace.



Obrázek 2-2 Vědecký potenciál kvantitativních metod (metody, které by dle podniků měly být podrobeny hlubšímu vědeckému zkoumání)
Zdroj: Kolková et al., 2022

Prognózování tak v sobě skrývá velký vědecký potenciál a firmy vyžadují výzkum právě v této oblasti. Na druhou stranu ale chtějí, aby výzkum vedl ke zjednodušení a lehčí aplikovatelnosti i interpretovatelnosti metod. To samozřejmě klade velké nároky na vědecký výzkum v této oblasti. Metody jsou na jednu stranu rozvíjeny a zpřesňovány, dnes je již běžné využití umělých neuronových sítí, hybridních metod a vyšších statistických metod. Ty jsou použity i v této publikaci. Pro větší využití je však nutno je zjednodušit tak, aby byly v praxi použitelné. Tento požadavek může vyřešit specializovaný software, který bude uživatelsky jednoduchý a interpretačně srozumitelný. A samozřejmě aplikace v současných všeobecně známých jazycích jako Python nebo R, které povedou k výpočtům složitých metod na základě jednoduchých předdefinovaných kódů. Výzkum v oblasti prognózování je zcela opodstatněný a podniky jej vyžadují. V současné době se výzkumy často zaměřují pouze na přesnost modelů, nicméně snaha nalézt model časově nenáročný a uživatelsky snadný také přibývá (Kolková, 2022).

2.2 PROGNOSTICKÉ SOUTĚŽE

Dlouholetá snaha nalézt nejlepší prognostický model je deklarována četnými výzkumy. Výzkumy jsou také často srovnávány a konají se různé soutěže, které přispívají ke zvyšování zdravého konkurenčního boje mezi výzkumníky. Časopisy International Journal of Forecasting a Journal of Forecasting také pořádají soutěže v předpovídání časových řad. Nejstarší prognostická soutěž se konala v 70. letech minulého století a je jí označována srovnávací studie technik prognóz časových řad v ekonomice na Nottinghamské univerzitě (Reid, 1969). Vzhledem ke komunikačním možnostem té doby nezahrnovali mnoho účastníků a ani příliš velké prognózy. Ale už v roce 1979 Makridakis a Hibon (1979) analyzovali 111 časových řad a aplikovali mnoho prognostických metod, což vyvolalo ve vědeckém světě poměrně velkou diskuzi. V reakci na tyto ne vždy souhlasné diskuze Madridakis a Hibon vytvořili soutěž zahrnující 1001 časových řad. Do této soutěže, konající se v roce 1982, se již mohlo zapojit více výzkumníků a vědeckou veřejností byla označena jako M-Competition. Hlavními výstupy je zjištění, že statisticky sofistikovanější metody nepřinášejí přesnější prognózy než jednodušší. Dále tvrzení, že hodnoty míry přesnosti různých metod se liší dle toho, jaké kritérium měr přesnosti je použito, což bylo pak analyzováno i jinými výzkumy (Kolková, 2018b). Třetím výstupem z M-Competition je, že kombinace různých metod překonává v průměru použití metod samostatně. A posledním výstupem je poměrně předvídatelný fakt, že výkon jednotlivých metod je závislý na délce horizontu prognózy.

V roce 1993 už proběhla M2-Competition (Makridakis et al., 1993), kterou tvořilo jen 29 sérií, ale s mnohem širšími informacemi. V roce 1998 Makridakis a Hibon vytvořili další soutěž M3-Competition, dle jejich slov poslední pokus rozhodnout, která z metod je nejlepší. V této soutěži byly aplikovány 3003 časové řady z různých oblastí, jako je podnikání, demografie, finance a ekonomie, a zapojilo se 24 výzkumníků. Vítěznou metodou se stal kombinovaný model, nazvaný Theta. S lehkou modifikací (ve formě Thetam) je tato metoda využita i v této publikaci. Výsledky publikovali Makridakis a Hibon (2000) a potvrdili, že M3 je v souladu s výsledky předchozích soutěží kromě hypotézy, že jednoduché metody překonávají složitější, protože zejména Box-Jenkins metoda a ARIMA modely byly mnohem přesnější než jiné.

Další soutěž realizoval Makridakis v roce 2018 s označením M4 Competition. Zde již bylo analyzováno 100 000 časových řad a byly kladeny i nové požadavky na účastníky, kteří například museli zveřejnit své výpočetní kódy na Github. Výsledky soutěže jsou publikovány (Makridakis et al., 2018) a vítěznou metodou je kombinace umělých neuronových sítí a exponenciálního vyrovnání od společnosti Uber. Je tedy zřejmé, že do soutěží tohoto typu se kromě akademiků a vědeckých pracovníků výzkumných institucí v současně době zapojují i vědečtí pracovníci z praxe. Velkým překvapení M4 Competition pak bylo nejen vítězství práce odborníka ze společnosti Uber Technologies (Smyl, 2020), ale i umístění mezi prvními pěti nejpřesnějšími modely i dalších dvou odborníků z praxe, a to Pawlikowského (Pawlikowski & Chorowska, 2020) z firmy Prologistica soft.

Dalším pak Jaganathan et al. (Jaganathan & Prakash, 2020), kteří do svého modelu zahrnuli koncepci předpovědi z podniku Forecast Pro (Business Forecast Systems, Inc., 2018, Forecast Pro TRAC. Version:5.1), ačkoli se soutěže nezúčastnili za podnik, ale jako individuální soutěžící. Výsledky M4 Competition jsou podrobně diskutovány v části kapitoly 4, která se blíže zabývá právě hybridními a kombinovanými modely.

V roce 2020 se konala zatím poslední M5 Competition (Makridakis et al., 2022). Tato soutěž se soustředila na prognózování maloobchodních tržeb, takže byla velice blízká právě prognózování poptávky, které je tématem i této monografie. Kromě toho data zahrnovala i exogenní proměnné, jako jsou informace o cenách nebo speciálních událostech. Hlavním výstupem z této soutěže je, že nejlepší přesnosti dosáhly metody umělých neuronových sítí a metody založené LightGBM, případně jejich kombinace. Umělým neuronovým sítím je věnována část kapitoly 4. LightGBM je zkratka pro light gradient-boosting machine. To je rámec distribuovaný v režimu open source, který slouží pro zvýšení gradientu pro strojové učení. Původně byl vyvinut společností Microsoft. Ukázalo se, že metody umělých neuronových sítí a metody založené LightGBM jsou v prognózování maloobchodu obzvláště vhodné, a to hlavně za předpokladu, že jsou k dispozici další vysvětlující proměnné za účelem zlepšení procesu učení neuronových sítí.

Bohužel, v této soutěži v důsledku předpisů o ochraně soukromí nemohou být zveřejněny informace o zázemí účastníků, takže kromě několika výzkumníků, kteří dali ke zveřejnění souhlas, nelze posoudit, kdo je z akademické sféry a kdo vyvinul svůj model v podnikové praxi.

Zajímavé je národnostní složení účastníků. Většina pocházela z USA (17 %), Japonska (17 %), Indie (10 %), Číny (10 %) a Ruska (6 %) a zbývajících 40 % z 96 dalších zemí. Je tedy velmi zajímavé, že velká, aktivní komunita má zájem o prognózování v rozvinutých i rozvojových zemích, kde lze očekávat, že praktické využití v podnicích je menší. O tom svědčí i první místo v soutěži, které získal Yeon Jun In, vysokoškolský student univerzity Kyung Hee v Jižní Koreji. V soutěži použil stejně váženou kombinaci (aritmetický průměr) různých modelů LightGBM.

Makridakis ale nebyl jediný organizátor soutěží. Paralelně s ním pořádali prognostické soutěže i jiní. Publikovány jsou tak soutěže „Sante Fe Competition“, „Neural Network Competition“, „Kaggle Time Series Competition“, případně „Global Energy Forecasting Competition“.

Tyto výsledky ještě více posílily diskuze o tom, jaké metody jsou pro prognózování nejvhodnější. Například datový vědec Petr Šimeček (2019), který pracoval u Google na vývoji modelů používajících neuronové sítě pro predikci časových řad s více proměnnými, zmínil na PyCon konferenci, která proběhla v roce 2019 v Ostravě, že statistické metody jsou vhodné v případě, že máme několik časových řad s velice dlouhou historií. Naproti tomu prognostické metody založené na hlubokém učení (Deep Learning) jsou vhodné v případě, kdy máme

hodně krátkodobých časových řad. Problém je, že často nevíme, jestli jsou analyzované časové řady dostatečně dlouhé pro správné rozhodnutí o tom, kterou metodu použít. I z toho důvodu se poslední roky na prvních příčkách soutěží M-Competition nebo soutěží na Kaggle umisťují pravidelně hybridní prognostické systémy založené jak na statistických metodách, tak i metodách využívajících strojové učení nebo neuronové sítě, které tvoří několik vrstev, případně hluboké učení, které se od neuronových sítí liší tím, že využívá více vrstev.

Celkově tyto soutěže tvoří historický záznam toho, jak se technologie a s ní i prognózování v čase vyvíjelo. Již nyní skončila M6 soutěž, která trvala celý rok a bude vyžadovat 12 příspěvků od každého účastníka. Nicméně byla možnost věnovat se i prognózám po čtvrtletí. Jako aktivum zde je zvoleno 50 akcií a 50 ETF, takže se soutěž již vzdaluje základnímu tématu této monografie. Cílem M6 soutěže pak bude ověření dvou výzkumných otázek. První bude ověření hypotézy efektivního trhu a druhou určení možné souvislosti mezi přesností a návratností investic. Výsledky budou zveřejněny v nejbližších měsících.

V souvislosti s výsledky posledních soutěží se také kladla otázka na změny v rolích prognostika v současnosti. Zdá se, že budoucnost bude potřebovat odborníky, kteří dokáží ovládat ML modely a nemusí být přímo odborníky na prognózování. Výsledky M4 a M5 soutěží to potvrzují. Mezitím co M4 vyhrál Slawek Smyl, zkušený prognostik společnosti Uber, tak M5 soutěž vyhrál YeonJun In, softwarový odborník bez přílišných znalostí prognózování.

Zůstává však otázkou, zda ML metody definitivně vyřadí ze scény metody statistické. Hlavní výhodou ML modelů oproti statistickým et al.je jejich průměrná vyšší výkonnost a minimální zásahy, které potřebují pro tvorbu prognóz (Makridakis et al., 2022). Nicméně hlavní nevýhodou je obrovský čas, který potřebují k správné implementaci, a školení.

Existují i názory (např. Kolassa, 2022), dle nichž jsou ML metody přeceňovány z hlediska praktické využitelnosti. A faktem je, že pouze 7,5 % týmů v M5 soutěži dokázalo pomocí ML metod překonat benchmark, kterým bylo exponenciální vyrovnání. Jak uvádí Kolassa (2022), investoři očekávají, že datoví vědci budou dělat zázraky. Při srovnání s jednoduchými metodami, které aplikujeme za zlomek nákladů, nejsou však výsledky o tolik lepší.

Je zřejmé, že metody ML již nezmizí a zůstanou bez ohledu na své současné slabiny. V budoucnu se ukáže, zda možnost menšího lidského úsilí a digitalizace pomocí algoritmů, pracující prakticky samostatně, převáží jejich současnou nákladovou náročnost, která časem jistě bude klesat. Nyní je výhodou, že ML metody jsou stále kompatibilní s metodami statistickými. Také je jasné, že statistické metody jsou na světě řadu let a již dosáhly svého vrcholu, a pokud nenastane nějaký ekonomický, politický či sociální šok, jejich přesnost a využitelnost už se nezlepší. Naproti tomu ML metody na základě umělé inteligence jsou v prognostice méně než pět let a očekává se, co přinese jejich budoucnost. Počítače a software se v čase stávají výkonnější a dostupnější, proto

lze očekávat, že metody založené na umělé inteligenci budou nabývat na důležitosti.

V souvislosti s tím se mění i role prognostika. V dobách před několika desítkami let prognostici museli vybírat metody nejvhodnější a nejpřesnější. To je dnes nahrazeno mnohými automatickými funkcemi. Tato automatizace tak mění roli prognostika z někoho, kdo vybírá nejpřesnější model, na někoho, kdo musí správně vyladit algoritmus, který tento výběr provádí. Stále je však potřeba úsudku v případě změn v prodeji, zavádění nových výrobků, poskytovaných akcí apod., což ostatně deklaruje také Kolassa (Kolassa, 2022). Role prognostika jako někoho, kdo pokrývá nové obchodní požadavky a odstraňuje problematické předpovědi, tak zřejmě zůstane. Dá se však říct, že tato role dnes už není ani tak rolí statistika, ale datového vědce s dostatečnými dovednostmi spojenými s umělou inteligencí (Makridakis et al., 2022). Autoři také předpokládají, že postupem času se obory prognostika, statistika a umělá inteligence budou dále propojovat a sjednocovat, až nakonec budou integrovány do jednoho oboru datové vědy.

2.3 PROGNÓZOVÁNÍ POPTÁVKY V PODNIKOVÉ PRAXI

Díky stále dokonalejšímu počítačovému vybavení je možno tvořit modely prognóz přesnějšími, ale také složitějšími. V akademické sféře se však v kontrastu s tím v současnosti častěji rozvíjí myšlenka o nutnosti jednoduchých prognostických modelů určených pro podnikovou praxi, jak uvádí například Zellner (2001) ve svém konceptu KISS (“keep it sophisticatedly simple”) nebo Kesten & Armstrong (2015), kteří tuto teorii potvrzují ve vědecké práci. Důvodem je přílišná složitost modelů, které tak ztrácí schopnost být podkladem pro rozhodování manažerů. Výzkum (Soyer & Hogarth, 2012) dokonce ukázal, že složitější statistické modely jsou mnohdy obtížně pochopitelné i pro akademiky.

Velké podniky po celém světě využívají prognózování samozřejmě pro zlepšení svých obchodních výkonů, snížení nákladů apod. S rostoucí digitalizací a novými formami podnikání, jako je sdílená ekonomika, úloha prognózování neklesá, ale naopak stoupá. S důležitostí prognózování stoupají i nároky na kvalitu a přesnost prognóz. Prognózy se už netýkají pouze základních makroekonomických veličin či poptávky, ale využívají se v celé řadě dalších podnikových procesů.

Spolu s rostoucí složitostí a novostí podnikatelských příležitostí pak rostou také oblasti využití prognózování. Příkladem může být sdílená ekonomika a v rámci ní společnost Uber, která využívá prognózy dokonce ve třech různých oblastech, jak uvádí obrázek 2-3.

- První oblastí jsou předpovědi trhu. Uber předpovídá nejen poptávku, ale i nabídku. Dělá to zejména proto, aby nasměrovali řidiče do oblastí s vysokou poptávkou po službách Uberu.
- Dále Uber využívá prognózy pro vytváření zásobníku pro detekci anomálií v reálném čase.

- Uber také vytváří predikce hardwarové kapacity, což dělají proto, aby předpovídali potřebný výpočetní výkon událostí s vysokou poptávkou. Tento druh prognózy může ušetřit spoustu nákladů.

Výstupem takových prognóz pak jsou hodnoty z širokého spektra veličin. Na příkladu společnosti Uber můžeme vidět, že metody prognózování se již nevyužívají jen na predikci poptávky. Další možnosti využití prognóz dle Bella (Bell, 2018) jsou:

- prognózování nejen poptávky, ale i nabídky,
- prognózování detekce výpadku systému,
- prognózování hardwarové kapacity.

Uber predikuje i nabídku automobilů ke spolupráci v přepravě na základě práce, kterou již nerealizují jen nevytížení taxikáři, ale i běžní řidiči, kteří mají časový prostor pro práci navíc. Jelikož k predikci používá složitých modelů založených na strojovém učení a umělé inteligenci, je pro Uber podstatné i prognózování detekce výpadku systému v reálném čase. Každý výpadek systému je pro tento typ podnikání velkým problémem, a tak se pokouší i o predikce v této oblasti. Nakonec také na základě prognóz plánuje hardwarovou kapacitu, protože tento způsob podnikání se samozřejmě neobjede bez velkých požadavků na výpočetní výkon. Prognózování tak přestává být v rukou podniku jen nástrojem k predikci poptávky, ale také celé řady dalších podnikových faktorů.

Na druhou stranu přibývá podniků, zejména zabývajících se novými technologiemi, které samy vyvíjejí nové modely. Vznikají tak modely právě Uberu, ale také Googlu či Facebooku. Výzkumníci z Uberu, jak již bylo řečeno v předchozí kapitole, zvítězili v M4 Competition, a jelikož svůj model částečně zveřejnili, lze očekávat jeho další využívání. Model využívá dynamický výpočetní grafový systém neuronové sítě, který umožňuje kombinování standardního exponenciálního vyhlazovacího modelu s pokročilými sítěmi s krátkodobou pamětí do společného rámce. Výsledkem je hybridní a hierarchická metoda prognózování. Metoda se vyznačuje dvěma zajímavými vlastnostmi, a to zejména že využívá statistické modely exponenciálního vyhlazování (přesněji Holtův a Holt-Wintersův model s multiplikativní sezónností) v kombinaci s LSTM sítěmi a druhá zvláštnost je použití hierarchické povahy. Ve zvolených datech autor odstranil sezónnost a adaptivně je normalizoval. Metoda generovala přesné předpovědi pro většinu časových řad, zejména však měsíční, čtvrtletní a roční. Přesnost však nebyla nejlepší pro případy denních a týdenních údajů. I přesto výsledná přesnost překonala všechny ostatní soutěžící. Nicméně autor i po ukončení soutěže pracoval na vylepšení modelu (Bandara et al., 2020, nebo Dudek et al., 2022).

Na druhou stranu v podnicích, jak již bylo několikrát řečeno, přetrvává potřeba aplikovat prognostické modely, které jsou snadno uchopitelné a snadno interpretovatelné. Právě na tento požadavek se snaží odpovědět jeden z nových modelů, Prophet. Byl vyvinutý výzkumníky Taylorem a Lethamem ze společnosti

Facebook a publikovaný v roce 2018. Využití tohoto modelu pak bylo zkoumáno i pro jiné prognózy poptávky, než je původně využil Facebook, a to například Navrátilem a Kolkovou, (2019), Kolkovou a Ključnikovem (2022).

V této souvislosti se výzkumníci také zabývají tím, zda by potenciální zvýšení přesnosti mohlo ospravedlnit přidanou složitost a výpočetní požadavky. Problematikou přidané hodnoty přístupů strojového učení právě ve forecastingu se zabývá Gilliland (2020). Dalšími, kteří ostře kritizují komplexní metody prognózování, je Green a Armstrong (2015). Gilliland se ovšem zabývá hlavně tím, co stojí za enormním zlepšením přesnosti. Zde platí jasná úvaha, že přesnost se zlepšuje s tím, jak se prodlužuje výpočetní čas. V tabulce 2-3 je zhodnocení vybraných metod vzhledem k době běhu.

Tabulka 2-3 Výpočetní čas vybraných prognostických metod.

Autor	sMAPE	MASE	Výpočetní čas (min.)
Smyl	11,374	1,536	8056
Montero-Manso et al.	11,72	1,551	46108,3
Legaki & Koutsouri	11,986	1,601	25
Theta	12,309	1,696	12,7
ARIMA	12,669	1,666	3030,9
Naïve2	13,564	1,912	2,9
Naïve1	14,208	2,044	0,2

Source: Gilliland, 2020

Je zřejmé, že dlouhý výpočetní čas nemusí být nutně důvodem pro vyřazení metod a jejich nepoužívání. Pro podniky však výpočetní čas může představovat další náklad. Malé podniky také nemusí disponovat odborníky, kteří dané metody zvládnou aplikovat. Outsourcing pak náklady opět zvyšuje. Samotná přesnější předpověď nemusí firmě přinášet nutně lepší výsledky (dle Gilliand, 2020). Přesnější předpověď musí být také správně interpretována a vést k lepším obchodním rozhodnutím. V podnicích se někdy setkáváme i s tím, že není potřeba zpřesňovat své prognostické modely, protože řada levných a přijatelně spolehlivých metod dostačuje. Využívání složitějších ML modelů v praxi je stále omezené na velké giganty. Proto se řada studií zaměřuje na vícekriteriální zhodnocení modelů (Kolková & Navrátil, 2021). Byla zde zhodnocena kritéria jako výpočetní čas a nároky na počítačové vybavení. Samozřejmě to vše v kontrastu s přesností modelů. Z výsledků této studie vyplynulo, že modely založené na deep learningu měly významně vyšší nároky na počítačové vybavení a výpočetní čas. V některých případech pak byl přijat jednodušší model, protože tato kritéria nedokázala vyvážit přesnost modelů.

2.4 GRAFICKÁ ANALÝZA DAT

Každé kvalitní predikci musí předcházet statistický popis a důkladná analýza kvality dat. Analýza dat obvykle začíná analýzou grafickou. U dat podnikových tomu není jinak. Grafická analýza spočívá ve vizualizaci dat pomocí nejrůznějších typů grafů. Pro analýzu podnikových časových řad je zejména vhodný graf spojnicový, sloupcový či graf krabicový.

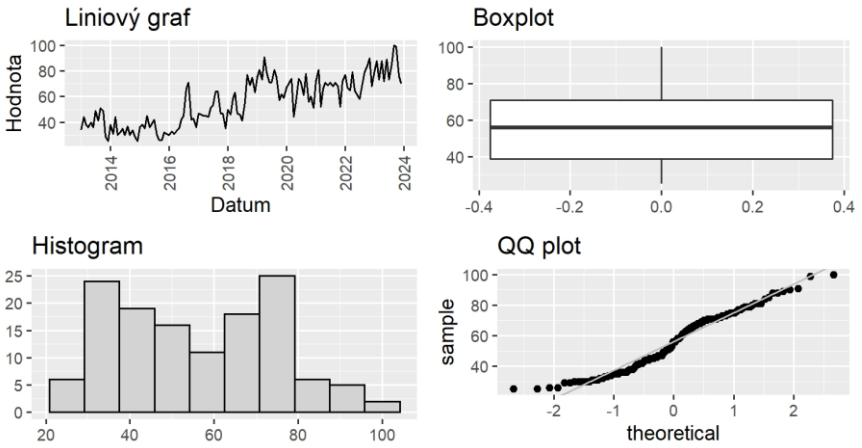
Grafickou analýzu lze realizovat pomocí nejrůznějších programů. Časově nejnáročnější je realizace pomocí programu Excel, ukázka této realizace s komentářem k vhodným typům grafů je v příloze 5. Rychlejším řešením je využití statistických programů jako SPSS, které je taktéž uvedeno v příloze 5. V ní je zobrazena grafická analýza dat Airbnb, která jsou veřejně dostupná na kaggle.com. Tato vizualizace byla využita částečně i v článku autorky Kolkové (Kolková, 2020). Postup při grafické analýze dat v Excelu a SPSS je uveden v příloze 6. Tyto dva programy jsou stále nejpoužívanější i v akademické sféře, nicméně jsou již vytlačovány dalšími, které umožňují složitější statistické výpočty a jednodušší grafické zpracování informací.

Dalšími programy vhodnými k vizualizaci a statistické deskripci dat jsou Statistica, Matematica a jiné. Jsou uvedeny a detailněji popsány v příloze 14. V této monografii se základním statistickým programům již nebudeme věnovat. Grafická i statistická analýza dat je provedena v jazyce R. Nejrychlejším řešením je pak využití programovacích statistických jazyků jako například R, Python. Ukázka vybraných kódů v jazyce R je uvedena v příloze 7. V jazyce R je možno k základní grafické a statistické analýze využít již základních kódů v programu. Vhodnější je ale využití specifických balíčků určených přímo k vizualizaci dat. Jedním z nich je *ggplot2* (Wickham et al., 2020). Pro další studium grafických prvků v jazyce R lze doporučit celou řadu dalších publikací (Venables a Smith, 2020), samozřejmě v prvé řadě samotný manuál k *ggplot2* (Wickham et al., 2020).

Základní grafy časových řad, jak již bylo řečeno, jsou obsaženy v základních funkcích R a lze je vizualizovat pomocí příkazu `plot()`. Ke specifickým grafickým vyjádřením slouží nejrůznější balíčky, které R nabízí. Sezónní grafy můžeme aplikovat pomocí balíčku *tsutils*. Balíčky *feasts* a *brolgar* poskytují grafické prvky pro časové i sezónní grafy. Interaktivní grafy lze vytvořit pomocí balíčku *tsibbletalk*. *Ggseas* je rozšíření *ggplot2* zejména o grafy pro sezónní časové řady. Další možné balíčky jsou například *dCovTS*, *ggseas*, *sugrrants*, *gravitas*, *dygraphs*, *TSstudio*, *ZRA*, *forecast*, *vars* a *fanplot*.

Rovněž Python nabízí jednoduchou vizualizaci už ve svém základu. Také ale poskytuje řadu knihoven určených k vytváření složitějších grafů. Nejznámější a nejpoužívanější knihovny pro vizualizaci jsou *Matplotlib*, *Seaborn*, *Bokeh*, *Plotly*, *Holoviews* nebo *Pygal*.

Grafická analýza dat pro predikci tržní poptávky, stejně jako poptávky po značce je uvedena v příloze 1 respektive 3. Pro velký rozsah není vhodné ji uvést v textu, a proto grafická analýza individuální poptávky po automobilu Toyota Corolla je uvedena na obrázku 3-3.



Obrázek 2-3 Grafická analýza GT dat vyhledávání značky Toyota Corolla
Zdroj: vlastní zpracování

Z grafické analýzy lze odhadovat, že data mají růstový trend a neobsahují žádné odlehlé pozorování. Můžeme také pozorovat určitá cyklická opakování, i když pouze nepravidelná. Také můžeme dedukovat, že zřejmě nebudou náležet do normálního rozložení. Je to patrné už z histogramu, který nemá normální tvar, také QQgraf je příliš zakřiven, což také naznačuje jiné než normální rozložení. Nicméně z grafické analýzy to lze pouze odhadovat. Jaký je skutečný trend a zda v datech jsou nějaké sezónní cykly, přesně odhadneme pouze dekompozicí a normalitu dat musíme otestovat příslušnými testy.

2.4.1 FIRMY ZABÝVAJÍCÍ SE PROGNÓZOVÁNÍM POPTÁVKY

V současné době je řada firem, které se zabývají prognózováním poptávky pro jiné podniky a nabízejí jim vlastní prediktivní softwary. Pro tyto produkty je obvykle typické, že jsou přizpůsobeny využitím v podnicích a jejich základem bývá co největší automatizovanost výpočtu všech prognóz. Prognostické firmy také obvykle staví na tom, že jejich automatické systémy lze používat bez velkých, případně bez jakýchkoli odborných znalostí. Také obvykle poskytují prognózy během několika minut.

V současnosti je lídrem na poli podnikového prognostického výzkumu společnost Uber. Výzkumník z této instituce byl i vítězem M4- Copmpetiton (Smyl, 2020), která již byla popsána v předchozím textu. Pro tuto společnost je stěžejní prognózování zejména poptávky a její podnikatelský úspěch dokládá i úspěšnost tohoto konceptu.

Další z firem úzce se zabývajících prognózováním poptávky je společnost Facebook, která svůj model nazvaný Prophet dokonce zveřejnila. Jeho aplikace na jiných datech než přímo ze společnosti Facebook je například v publikaci Navrátil & Kolková (2019).

Existuje pak řada firem, které nabízejí prognostické služby v rámci své nabídky i jako součást celých ERP systémů. Jedním z nich je například Prologistica Soft (Pawlikowski & Chorowska, 2020), polská společnost, která se zabývá i logistikou a vytvářením ERP systémů komplexně. Výzkumníci této společnosti obsadili 3. místo v M4-competiton. Úspěšný praktický model v této soutěži nabízí i Jaganathan a Prakash (Jaganathan & Prakash, 2020) , kteří, jak už bylo uvedeno, vychází z komerčního produktu Forecast Pro (Business Forecast Systems, Inc., 2018, Forecast Pro TRAC. Version: 5.1.). Business Forecast System je australská firma, která navrhla systém Forecast Pro, jehož výhodou je možnost připojením k jakýmkoli stávajícím systémům nebo databázím.

Je proto záhodno přemýšlet, proč v současné době odborníci z praxe dosahují větších úspěchů v prognostických modelech než akademici. Samozřejmě že soutěž neabsolvovaly všechny firmy, které se prognózováním zabývají. Mnohé považují konkrétní konstrukci modelů za obchodní tajemství. Jiné využívají jen modely již zveřejněné a ke každé prognostické akci přistupují individuálně. Mnohdy se tak ani odborník nedozví, které konkrétní modely používá. V Tabulce 2-5 jsou uvedeny některé nejvýznamější firmy, které se zabývají prognózováním poptávky a produkty i komerčně nabízejí.

Tabulka 2-4 Nejznámější poskytovatelé prognostických modelů

Provider	Model	Method used
ProLogistica Soft	ProLogistica TREND	Naive method, weighted moving average, exponential smoothing, SARIMA, ARIMA X13
IBM	IBM Cognos Analytics	ETS models
Forecast Pro	FP100	Machine learning model, exponential smothing, Box-jenkins method, Dynamic regression, event Models, Multiple-Level Models, Croston's intermittent demand model, moving averages, naive method.
	FP	
	FP TRAC	
Microsoft	Microsof Azure (Stream analytics)	Machine learning model, unspecified statistical methods
	Microsoft Dynamics AX 2012	ARIMA, ARTxp
SAP	GMDH Streamline	Unspecified statistical methods, machine learning model
	SAP F&R	Unspecified statistical methods,regression analysis
	Forecast in BW-Integrated Planning	Average, Floating Average, Weighted Floating Average, Linear Regression, Seasonal linear regression, Simple Exponential Smoothing, Simple exponential smoothing with alpha optimization, Linear exponential, Seasonal Exponential Smoothing, Seasonal trend exponential smoothing, Croston model, Automatic Model Selection
ORACLE and JD EDWARDS	Forecasting methods	Percent Over Last Year, Calculated Percent Over Last Year, Last Year to This Year, Moving Average, Linear Approximation, Least Squares Regression, Second Degree Approximation, Flexible Method, Weighted Moving Average, Linear Smoothing, Exponential Smoothing, Exponential Smoothing with Trend and Seasonality.
EPICOR	Forecasting methods	Weighted Average

Použité zdroje:
https://prologistica.pl/bazawiedzy/prognozowanie-popytu-i-planowanie-sprzedazy.html
https://www.ibm.com/docs/cs/cognos-analytics/11.1.0?topic=d-forecasting
https://www.forecastpro.com/solutions/forecast-pro/forecasting-methods/
https://docs.microsoft.com/cs-CZ/azure/architecture/solution-ideas/articles/demand-forecasting
https://www.microsoft.com/en-us/download/confirmation.aspx?id=43128
https://gmdhsoftware.com/documentation-sl/statistical-forecasting
https://help.sap.com/saphelp_scm700_ehp02/helpdata/en/8d/e2742de6dc455e900652ba8be6338c/frameset.htm

https://help.sap.com/doc/saphelp_nw73ehp1/7.31.19/en-US/4c/af369d0cf30780e10000000a42189b
https://docs.oracle.com/cd/E16582_01/doc.91/e15111/und_forecast_levels_methods.htm#EOAFM00165
https://www.epicor.com/en-us/industry-productivity-solutions/manufacturing/

Zdroj: vlastní zpracování dle uvedených referencí

Často bohužel akademičtí pracovníci nemohou ověřit míry přesnosti některých metod, protože přesné kódy, výpočty a know-how těchto modelů vědecko-výzkumné útvary podniků nezveřejňují. V tom je velká výhoda prognostických soutěží, protože v jejich rámci je povinnost zveřejňovat kódy výpočtů.

Tabulka 2-5 uvádí ty nejznámější, v současné době v praxi používané metody v komerčních produktech. Ve sloupci prvním lze vidět poskytovatele těchto metod, další sloupec tvoří jeho název. V předposledním sloupci je výčet metod, které jsou obvykle pro forecast použity. Poslední sloupec obsahuje odkaz na projektovou dokumentaci k modelu.

2.5 POSTUP PŘI PREDIKCI POPTÁVKY

Stejně jako predikování jakýchkoli jiných podnikových veličin i predikce obsahuje tři základní části, a to:

- nejprve analýzu současného stavu, současné časové řady, zde zejména analýzu již prodávaných produktů, případně nabízených služeb,
- dále samotnou předpověď budoucí poptávky, a to dle předchozího členění kvantitativními či kvalitativními metodami,
- predikce pak musí být ukončena výpočtem chyb predikce a zhodnocením případných opatření k snižování chyb predikce.

K základním požadavkům na predikování poptávky patří zhodnocení variability poptávky, a to základními statistickými veličinami, jako jsou rozptyl, směrodatná odchylka, variační koeficient a další. Samozřejmě je součástí také základní analýza dat, jako jsou grafická analýza, analýza odlehlých či chybějících hodnot apod. Řešení těchto veličin je definováno v kapitole 3.

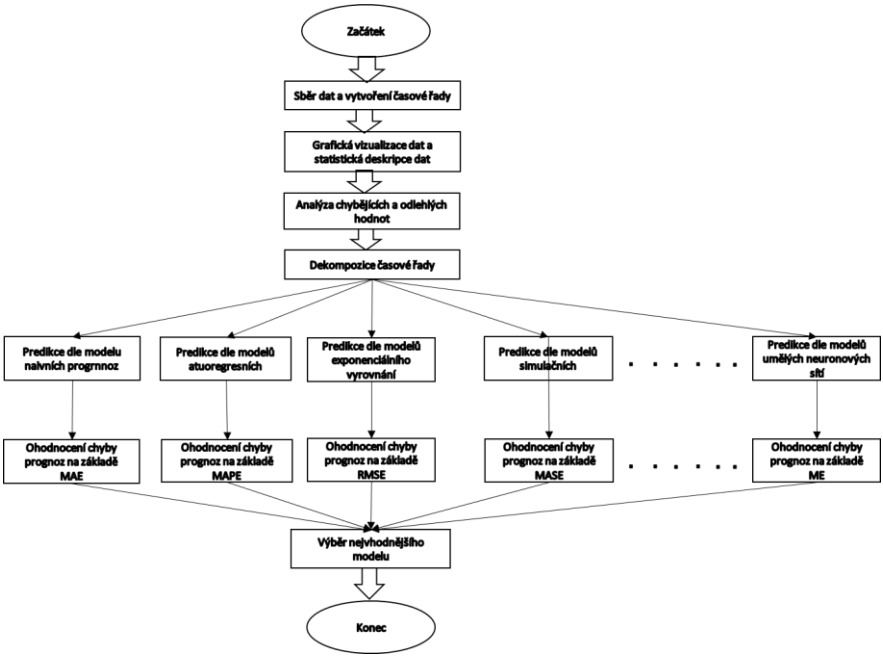
Samotná předpověď poptávky může být realizována na základě nejrůznějších metod. Jak uvádí Macurová et al. (2018), když jsou data o poptávce v minulosti nedostatečná, protichůdná, drahá nebo nevýznamná, je nutné použít metody kvalitativní, které jsou založeny na intuici a zkušenosti a jsou velmi často subjektivní. Důležitá je také analýza chyby predikce poptávky. Chyby predikce je možno definovat na základě několika ukazatelů – ty jsou definovány dále v kapitole 5. Konkrétně chyba v predikci poptávky představuje rozdíl mezi skutečnou poptávkou a její predikcí. Na výši chyby závisí zejména velikost vytvářené pojistné zásoby.

Prognózování pomocí kvantitativních metod v podnikové ekonomice nezahrnuje pouze výpočet samotného modelu prognózy, ale celou řadu dalších kroků, bez kterých je prognóza neúplná. Tyto kroky mají specifickou posloupnost a přímo na sebe navazují.

Jedná se zejména o:

- statistickou deskripci dat a analýzu chybějících a odlehlých hodnot,
- samotnou prognózu na základě vybraných metod,
- ohodnocení chyby predikce,
- výběr metody dle nejnižší chyby.

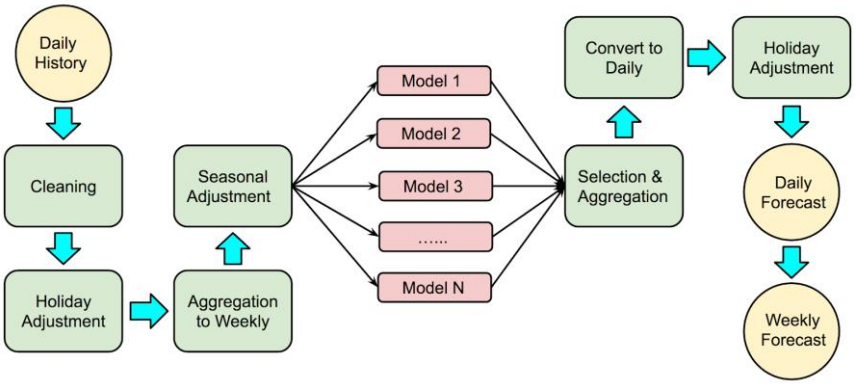
Detailní postup a vazby, které tyto kroky doprovázejí, jsou shrnuty v obrázku 2-3.



Obrázek 2-4 Postup realizace prognóz

Zdroj: vlastní zpracování

Postup, jak lze v různých modelech využít metody prognózování, shrnuje také Tassone & Rohani (2017), výzkumníci společnosti Google. Tento postup je graficky znázorněn na obrázku 2-4.



Obrázek 2-5 Proces prognózování dle výzkumníků v Google
Zdroj: Tassone & Rohani, 2017

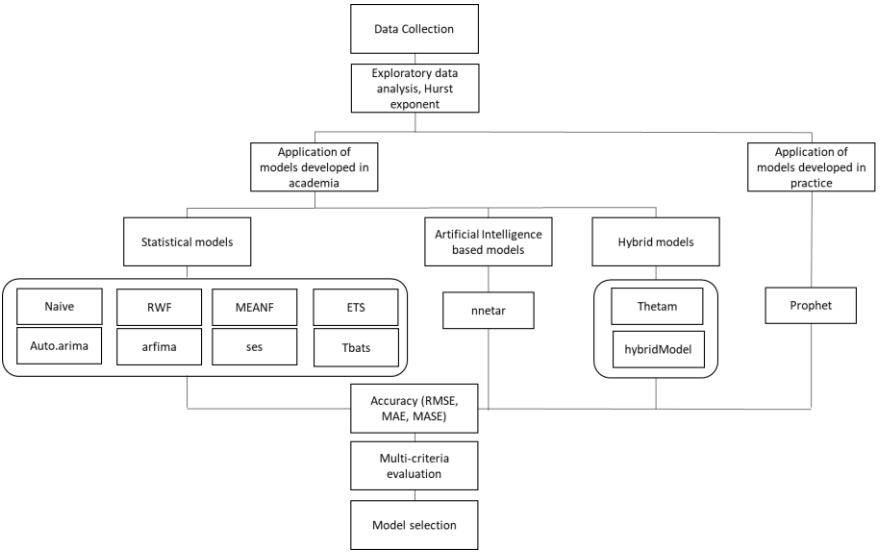
Prognostická procedura i dle těchto autorů začíná získáním a sběrem dat – v tomto grafu jsou preferována denní data. Tato fáze může být velmi náročná a zdlouhavá, může však být pouze využitím big dat, které podnik běžně shromažďuje, jen je prozatím nevyužívá. Další navazující fází je „cleaning“, kdy je potřeba data očistit, upravit odlehlé hodnoty, definovat a vyřešit hodnoty chybějící. Následuje již dekompozice získaných dat, která může být těžištěm a hlavním úkolem prognózování pro daný podnik. Obvykle se data rozkládají na hlavní trend, sezónní složku a náhodnou složku. To ovšem bývá pro podniky dle těchto autorů nedostatečné. Na podnikové procesy mají vliv i jiné prvky, jako jsou dny v týdnu, prázdniny, státní svátky, významné události atd., proto klasická základní dekompozice je často nedostatečná a je nutné ji doplnit o složky těchto prvků. Další etapa už spočívá v aplikaci vybraných modelů prognózování a jejich následné analýze. Výsledkem pak je denní, případně týdenní či měsíční prognóza.

Dalším možným postupem je zařadit pro výběr modelu i **vícekriteriální rozhodování**. Tento postup byl uplatněn například v článku Kolkové a Ključnikova (Kolková & Ključnikov, 2022). Zde byly vybrány již konkrétní prognostické metody, ze kterých se vybíral nejvhodnější model. Tyto metody byly rozděleny na metody vyvinuté v akademické sféře a metody vyvinuté v praxi. Metody vyvinuté v akademické sféře pak byly rozděleny ještě na metody statistické, založené na umělé inteligenci, a hybridní. Následovalo, jako i u jiných postupů, hodnocení míry přesnosti modelů dle vybraných kritérií. Jako inovace v této studii bylo přidáno multi-kriteriální zhodnocení jednotlivých modelů. Jedním z kritérií samozřejmě byla míra přesnosti, nicméně pak byla hodnocena ještě další kritéria, tj. požadavky na znalosti analytiky a na výpočetní výkon počítačů.

Pro vícekriteriální hodnocení byla vybrána Saatyho metoda dle Thomase L. Saaty. Tato metoda byla mnohokrát použita i v českých publikacích (Borovcová & Špačková, 2018), (Peterková, & Franek, 2018) nebo (Mikušová & Čopíková, 2016). Je zřejmé, že kromě kritéria míry přesnosti byla ostatní kritéria silně ovlivněna velikostí podniku. Proto bylo vícekriteriální hodnocení vyčísleno zvlášť

pro velké podniky a zvlášť pro podniky malé a střední (dále SME). Pro SME je zřejmé, že zajištění vhodného pracovníka pro zpracování časových řad bude obtížnější, a proto tomuto kritériu přiřadí větší váhu. Totéž pro SME bude platit i pro zajištění dostatečného výpočetního výkonu.

Další inovací k prognostickému postupu bylo v této studii využití Google Trends dat, která budou diskutována v kapitole 3. Tato data jsou získávána jako souhrnná a lze z nich odvodit poptávku po daném produktu či službě jako celku. Nicméně při využití pro prognózy v konkrétních podnicích je samozřejmě rozdíl, zda se jedná o podnik velký či SME. Za malého a středního podnikatele můžeme dle definice Czechinvest považovat podnik zaměstnávající méně než 250 zaměstnanců a jehož roční obrat nepřesahuje 50 milionů EUR nebo jehož bilanční suma roční rozvahy nepřesahuje 43 milionů EUR. Je tedy zřejmé, že malé a střední podniky tvoří nezanedbatelnou část podnikatelských subjektů. V roce 2017 tvořily 58 % HDP právě malé a střední podniky, v České republice je to jen 40 %. K 31. 12. 2019 však celkový počet podnikatelských subjektů odpovídajících definici malého a středního podnikání dle ministerstva průmyslu a obchodu tvořil 99,8 % všech podnikatelských subjektů (MPO, 2019). Celý postup při prognózování i s vícekriteriálním zhodnocením je uveden v obrázku 2-5.



Obrázek 2-6 Postup prognózy s využitím vícekriteriálního zhodnocení
Zdroj: Kolková & Ključnikov, 2022

Po přihlédnutí k velikosti podniku a zohlednění pomocí vah jednotlivých kritérií to, zda se jedná o SME nebo velký podnik, bylo výzkumem (Kolková & Ključnikov, 2022) potvrzeno, že výběr metod je závislý na velikosti podniku. Výsledná volba metody pro velké podniky je uvedena v Tabulce 2-5 a pro SME v Tabulce 2-6. Lze vidět, že výsledky jsou poměrně rozdílné.

Mezitím co velké podniky jednoznačně preferují model Prophet a přesnost, protože přesnost pro ně je pravděpodobně stěžejním kritériem, u SME je tomu naopak – nejvhodnější stále zůstávají statistické metody, a to zejména pro jejich snadnou aplikovatelnost, respektive nízkou náročnost na výpočetní techniku a lidský kapitál.

Tabulka 2-5 Výsledná volba metody pro velké podniky

Model	Míra přesnosti	Náročnost na znalosti pracovníka	Výpočetní výkon	Celkové pořadí
Statistické modely	0,8416	0,3741	0,2594	1
Modely založené na umělých neuronových sítích	1,6833	0,1870	0,1297	2
Hybridní modely	2,5249	0,0935	0,0648	3
Model vyvinutý v praxi (Prophet)	3,3666	0,2805	0,1945	4

1 …nejhorší, 4… nejlepší

Zdroj: Kolková & Ključnikov, 2022

Tabulka 2-6 Výsledná volba metody pro SME

Model	Míra přesnosti	Náročnost na znalosti pracovníka	Výpočetní výkon	Celkové pořadí
Statistické modely	0,1237	1,8708	1,6343	4
Modely založené na umělých neuronových sítích	0,2475	0,9354	0,8171	2
Hybridní modely	0,3712	0,4677	0,4086	1
Model vyvinutý v praxi (Prophet)	0,4949	1,4031	1,2257	2

1 …nejhorší, 4… nejlepší

Zdroj: Kolková & Ključnikov, 2022

Vícekriteriální zhodnocení bylo uplatněno i v dalších studiích (Kolková & Navrátil, 2021). Zde byla míra přesnosti modelů hodnocena dle dvou kritérií: výpočetní čas a požadavky na výpočetní výkon. Modely byly hodnoceny podle toho, zda predikce předpovídala 90 dní nebo 365 dní dopředu. Výsledky pro obě predikce však byly velmi podobné.

Nejlepší byl model vyvinutý v praxi. Překvapivě velmi úspěšný při vícekriteriálním zhodnocení byl i sezónní naivní model. Je velmi jednoduchý na výpočet, ale jeho přesnost není velká. Když ovšem všechna tři kritéria mají stejnou

váhu, naivní model v ostatních kritériích jednoznačně vede. Proto je jeho úspěšnost na třetím místě.

Naproti tomu modely založené na hlubokém učení, respektive neuronových sítích úspěchy neslavily. Obrovská náročnost na výpočetní čas i výkon ani neodpovídala míře jeho přesnosti. Přesné výsledky této studie jsou uvedeny v tabulce 2-7.

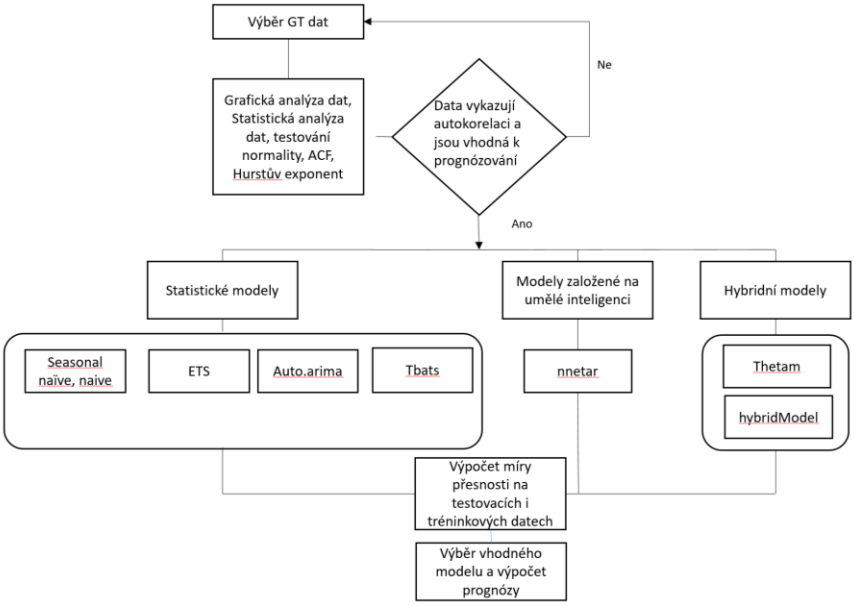
Tabulka 2-7 Výsledky vícekriteriálního hodnocení ve výzkumu

Model (90 days)	MAE	Rank	MSE	Rank	Time	Rank	Computing demand - Rank	Rank Results
Seasonal Naive Model	4,81	5	40,30	6	0,03 s	1	1	13
TBATS Model	2,59	2	10,84	2	461,91 s	4	4	12
Prophet Model	2,51	1	10,63	1	5,18 s	2	2	6
SARIMA (0,1,1) x (1,1,1,7)	2,94	3	12,98	3	17,94 s	3	3	12
Deep Learning Shallow Layers	2,94	3	14,60	4	2550,10 s	5	5	17
Deep Learning Deep Layers	3,94	4	24,09	5	3400,00 s	6	6	21
Model (365 days)	**MAE**	**Rank**	**MSE**	**Rank**	**Time**	**Rank**	**Computing demand - Rank**	**Rank Results**
Seasonal Naive Model	5,38	6	48,50	6	0,04 s	1	1	14
TBATS Model	3,20	1	17,99	2	298,28 s	4	4	11
Prophet Model	3,22	2	17,69	1	5,39 s	2	2	7
SARIMA (0,0,2) x (0,1,1,7)	3,63	3	24,84	3	88,62 s	3	3	12
Deep Learning Shallow Layers	5,27	5	46,33	5	1600,00 s	5	5	20
Deep Learning Deep Layers	4,85	4	37,50	4	3000,00 s	6	6	20

Zdroj: Kolková, Navrátil, 2021

V této práci je aplikován postup dle obrázku 2-6, který začíná výběrem GT dat. Tato data jsou podrobena statistické i grafické analýze. Dále je otestována autokorelace a pro prognózy lze využít pouze data, která mají určitě vzory z minulosti a která se opakují. Proto lze předpokládat i jejich podobný vývoj v budoucnosti, a jsou tedy predikovatelná. Z toho důvodu jsou data bez autokorelace vyloučena a nahrazena jinými.

Další postup je už v souladu se standardním postupem při prognózách, a to výpočet modelů na tréninkových i testovacích datech a jejich následné srovnání pomocí měr přesnosti. Míry přesnosti vypočítané na tréninkových datech jsou v publikaci uvedeny jen pro úplnost a pro srovnání kvality modelu na tréninkových datech a v realitě. Výběr modelu je proveden dle měr přesnosti vypočítaných na testovacích datech.



Obrázek 2-7 Prognostický postup využitý v této monografii
Zdroj: vlastní zpracování

Kapitola 3

Příprava dat pro prognózování

Základem každé prediktivní práce je získání dat. Samozřejmě, pokud realizujeme poptávku individuální, nejkvalitnější data můžeme dostat pouze z konkrétního podniku, pro který predikci či prediktivní analýzu děláme. Pro vědecké účely je často složité se k těmto datům dostat. Podniky obvykle svá data považují za součást obchodního tajemství a vzhledem k tomu, že výsledky vědecké práce musejí být publikovány, a to nejlépe v open source režimu, mají podniky obavu, aby se data nestala předmětem rozhodování konkurence. Vědci v ekonomických tématech jsou tak ve velmi těžké situaci a téměř vždy musí čelit snaze ochránit data (ať už anonymizací či přepočtem dat utajeným koeficientem). Jistou možností je využití dat veřejně dostupných.

Možností získání dat v podnikové ekonomice je několik. Základem získávání dat pro potřeby vědeckého výzkumu v podnikové ekonomice v podmínkách České republiky jsou data zejména z Českého statistického úřadu a primárního výzkumu. Zajímavá data k analýze či predikci lze ale získat i na mezinárodních databázích, jako je eurostat, dále také Kaggle, github, Google. Příklady webových stránek s případnými zdroji pro možné vyzkoušení a ověření metod prognózování, měření přesnosti a komplexní prognózou mohou být:

- https://www.czso.cz/csu/czso/katalog-produktu,
- https://www.kaggle.com/datasets,
- https://ec.europa.eu/eurostat/data/database,
- https://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx,
- https://github.com/collections/open-data.

V této kapitole budou detailně popsána data, která v následujících kapitolách budou predikována. Pro predikci byla vybrána data z kategorií GT dat, a to: automobily, cestování, domácí mazlíčci, domov a zahrada, hobby a volný čas, hry, knihy a literatura, nemovitosti, počítače a elektronika. Pro velký rozsah grafických výstupů je grafická analýza těchto dat uvedena v příloze 1. Taktéž dekompozice těchto dat je uvedena v příloze 2. Pro analýzu konkurence pomocí GT dat jsou vybrána data vyhledávání jednotlivých značek automobilů. Bylo vybráno pět automobilek, a to nejprodávanější značky aut v roce 2022 dle https://www.cebia.cz/novinky/trh-s-automobily/top-100-nejprodavanejsi-auta-

sveta-v-roce-2022. Opět pro velký rozsah jsou výstupy z grafické analýzy v příloze 3 a dekompozice těchto dat v příloze 4. Shrnutí a popis analýz jsou v následujících částech této kapitoly. Samostatná kapitola se také věnuje prognóze individuální poptávky. Pro tuto poptávku bylo vybráno vyhledávání pojmu Toyota Corolla. Deskripce GT dat tohoto vyhledávání je v následujících částech této kapitoly.

3.1 GOOGLE TRENDS DATA

Využití GoogleTrends dat je zajímavým trendem v současnosti, a to jak pro vědecké, tak i podnikové účely. Google Trends data je webová služba, kterou realizuje firma Google. Data jsou poskytována volně a zdarma. Tato služba vznikla 5. srpna 2008 pod názvem Google Insights for Search. Dne 27. září 2012 Google sloučil Google Insights for Search do Trendů Google. Tato služba analyzuje popularitu vyhledávacích dotazů v Google v různých regionech a jazycích. Trendy Google (dále jen GT data) používají grafy k porovnání počtu vyhledávání různých dotazů v čase. Využití GT dat pro prognózování není úplnou novinkou, existuje již celá řada autorů, kteří tato data využívají. Jedním z příkladů je i publikace Kolkové a Ključnikova (Kolková & Ključnikov, 2021). V této studii byla predikována poptávka po knihách poezie v zemích V4. Velice zajímavým zjištěním bylo, že ze zemí V4 Česká a Polská republika jsou země, kde lze očekávat vzrůst poptávky po poezii, na Slovensku a v Maďarsku naopak pokles. Přitom v Polsku začal pozvolný nárůst poptávky až s pandemií, mezitímco růst poptávky po poezii v ČR je na základě dekompozice dat patrný již nejméně od roku 2010.

V posledním desetiletí se využití GT dat prokázalo jako užitečné. Existují studie, které na základě GT dat prognózují poptávku (Kolková & Ključnikov, 2022), zkoumají souvislost mezi daty GT a zaměstnaností nebo nezaměstnaností (Baker & Fradkin, 2017), také spotřebu (Morgavi, 2020), obchod (Gonzales et al., 2020), digitalizaci (Pisu et al., 2020) nebo ceny ve stavebnictví (Cournède et al., 2020). V poslední době byla také použita data Google Trends k posouzení dopadu krize COVID-19 (Abayet al., 2020; Doerr & Gambacorta, 2020).

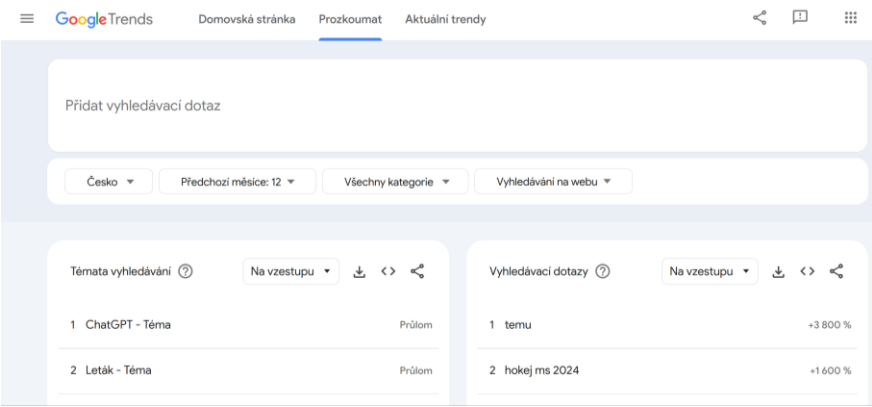
GT data poskytují indexy objemu dat vyhledávání, které měří intenzitu vyhledávání dle vzorce (3.1),

$$\frac{počet\ vyhledávání\ daného\ klíčového\ slova}{celkový\ počet\ vyhledávání} \tag{3.1}$$

a to podle místa a období. Přitom v GT datech lze využívat tyto indexy podle klíčového slova, kategorie klíčových slov nebo tématu. Dotazy založené na klíčových slovech samozřejmě podléhají specifikům daného jazyku. Kategorie a témata jsou harmonizovány napříč jazyky. Proto byly v této práci použity kategorie i individuální vyhledávání. Na GT datech lze nalézt cca 1200 kategorií, které přidělují jednotlivá vyhledávání dle pravděpodobnostního algoritmu (Varian & Choi, 2009).

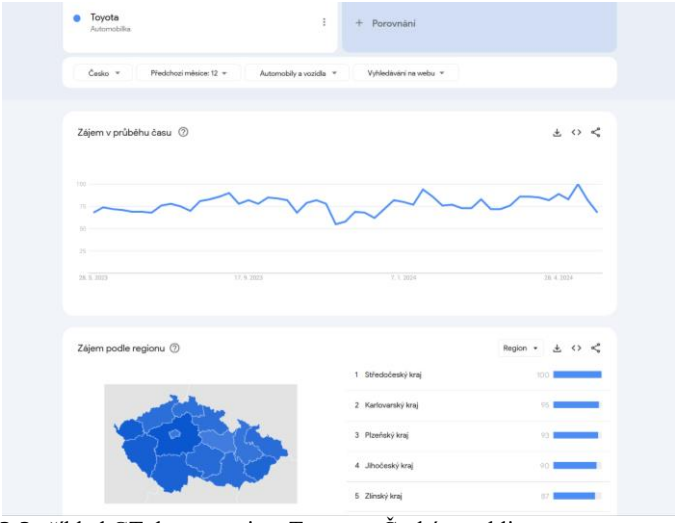
Hlavní nevýhodou GT dat je, že i když jsou sbírána od roku 2004, počet uživatelů vyhledávání Google významně vzrostl, proto relativní intenzita vyhledávání u většiny kategorií klesá. Aby se odstranil tento nedostatek, jsou v práci použita data od roku 2013 do 2023, kdy se dá hovořit o tom, že uživatelská základna je přibližně podobná. Samozřejmostí je, že big data použitá v GT datech je nutno statisticky popsat a případně i zpracovat a řešit případné odlehlé hodnoty či chybějící hodnoty, což je provedeno v následujících částech této kapitoly.

Na stránce https://trends.google.com/ nalezneme několik možností, jak získávat požadované informace. Na své domovské stránce Google nabízí zajímavosti z vyhledávání a několik inspirativních analýz, které ukazují, jak jsou Trendy Google využívány po celém světě zpravodajskými redakcemi, charitativními organizacemi i dalšími subjekty. Na záložce Prozkoumat už můžeme stahovat data, která požadujeme, viz obrázek 3-1. Data jsou tříděna dle jednotlivých zemí. Dále také můžeme využít filtrování dle námi zvoleného časového hlediska. Další možností jsou kategorie (vybrané byly využity i v této publikaci) vytvářené dle podobných pojmů, které jsou na Googlu vyhledávány a pak sdružovány dle významu. Poslední možností filtrování vybraných dat je, zda požadujeme data o vyhledávání na webu, v obrázcích, ve zprávách, na youtube anebo o nákupech prostřednictvím Google.



Obrázek 3-1 Web GT dat pro stahování požadovaných dat
Zdroj: https://trends.google.com/trends/explore?geo=CZ&hl=cs

Dále jsou na stránce k dispozici témata vyhledávání, která můžeme seřadit dle těch, která jsou na vzestupu nebo která jsou vyhledávána nejčastěji. Také konkrétní vyhledávací dotazy – opět na vzestupu nebo nejčastější. Po zadání konkrétního pojmu vyhledávání pak GT data vykreslí graf vývoje v daném časovém horizontu s možností tato data stáhnout. A také graf nejčastějšího vyhledávání ve zvolené oblasti. Na obrázku 3-2 je příklad vyhledávání na webu slova Toyota za posledních 12 měsíců v České republice.



Obrázek 3-2 příklad GT dat pro pojem Toyota v České republice
Zdroj:
https://trends.google.com/trends/explore?cat=47&geo=CZ&q=%2Fm%2F07mb6&hl=cs

Na obrázku vidíme, že vyhledávání po celý rok mělo téměř stagnující trend. Pouze na přelomu roku zaznamenalo mírný pokles. Nejvíce se vyhledával tento pojem ve Středočeském kraji, naopak nejméně v kraji Pardubickém.

3.2 STATISTICKÁ ANALÝZA DAT

V roce 1977 John Turkey vydal knihu s názvem Exploratory Data Analysis (Turkey, 1977), česky je možno název přeložit jako průzkumová (častěji ponecháno slovo explorativní) analýza dat. Pro tuto analýzu se vžilo označení EDA, a i když od té doby uplynulo již mnoho let a byla již řadou publikací doplňována a zpřesňována, zkratka EDA se vžila a používá se dodnes. Slangově bývá dokonce i počešťována a skloňována, a tak věty statistiků „…jdu udělat Edu…“ nebo „…jak ti vyšel Eda…“ jsou běžným hovorovým vyjádřením této analýzy.

Základem pro EDA je vyčíslení základních statistických charakteristik souboru dat. Mezi ně patří zejména míry polohy (střední hodnota, minimum, maximum, medián, modus, kvartily), míry variability (rozptyl, směrodatná odchylka, variační koeficient), míry tvaru (šikmost, špičatost). Tyto hodnoty lze vyčíslit ve všech statistických programech s větší či menší mírou časové náročnosti pro výzkumníka.

Nejpracnější a na čas nejnáročnější je řešení a výpočet EDA v Excelu. Není v něm funkce, která by souhrnně provedla celou analýzu EDA, a tak je nutné každou charakteristiku vypočítat zvláštní funkcí. Nejčastěji používané funkce pro výpočet EDA v Excelu jsou uvedeny v příloze 8. Další možností výpočtu jsou programy jako SPSS, ukázka výpočtu je uvedena v příloze 8, dále také

Mathematika, Statistika atd. Samozřejmě rychlejší výpočet poskytují programy jako R, jehož kódy pro EDA jsou opět uvedeny v příloze 8.

Data v podnikové ekonomice někdy nejsou taková, jaká bychom si přáli, a často se stává, že některé hodnoty jsou vychýlené, extrémní, případně chybí úplně. Proto je nutné na základě EDA provést i další analýzy a následně vyřešit případné problémy plynoucí z jejich existence. Základním problémem dat v podnikové ekonomice jsou tedy chybějící údaje a odlehlé hodnoty.

Chybějící hodnoty (missing values) se i přes velkou snahu výzkumníků mít data kompletní někdy v souboru vyskytují. Zejména jedná-li se o výzkum založený na dotazování, stává se, že respondenti nechtějí odpovídat. Ale také při hodnotách tržeb, cen či jiných kvantitativních podnikových veličin se může stát, že daný údaj byl špatně zapsán, došlo k výpadku elektřiny, systému nebo k lidské chybě a historický údaj se již nepodaří zjistit. Při malých výběrech a malém množství dat je náhrada chybějících dat nevhodná a výsledky by mohla velmi ovlivnit. V takovém případě lze pro výzkum pouze doporučit data doplnit, zvýšit počet dotazovaných či hodnot pozorování. Při velkých výběrech a široké datové matici již lze o možnosti náhrady několika málo chybějících hodnot uvažovat. Přesto je nutno k jejich náhradě přistupovat obezřetně.

Způsobů náhrady chybějící hodnoty je několik. Nejjednodušší je nahradit chybějící hodnotu výběrovým **průměrem**, případně **průměrem** či **mediánem blízkých bodů**. Při znalosti rozdělení zkoumané veličiny lze využít nahrazení **náhodným číslem**, vycházejícím z parametrů daného rozložení. Další možností je využití **regresní funkce** nebo **lineární interpolace**. Řadu z těchto možností nabízí i statistické softwary, nicméně vždy je vhodné zvážit, zda pro výzkum není vhodnější chybějící hodnoty z datové matice zcela vyloučit. Ukázka vyhledání odlehlých a chybějících hodnot v SPSS a v programu R je v příloze 9. V našich datech se chybějící hodnoty nevyskytují, proto není potřeba je řešit.

Po kontrole úplnosti dat by mělo být obvyklé zhodnotit také jejich správnost a konzistentnost. Jinými slovy, hledáme odlehlé hodnoty (outliers), tj. hodnoty, které v podnikové ekonomice jsou vychýleny z různých nestandardních důvodů, například cena odchýlená vlivem jednorázové akce, dávka ve výrobě změněná vlivem vadného materiálu, změny nákupních preferencí daných povětrnostními vlivy či jinak nestandardním počasím. Typickým příkladem, jak je důležité s odlehlými hodnotami pracovat, prezentovaným snad na všech statistických přednáškách, je výpočet průměrné mzdy občanů malé horské vesničky o 20 obyvatelích, kam se právě přistěhoval Bill Gates. Průměrná mzda 20 obyvatel, kdy jeden má měsíční mzdu počítánu v milionech dolarů, jistě nebude mít valnou vypovídací schopnost. V podnikové ekonomice je pak často na zvážení a na zkušenostech výzkumníka, zdali odlehlé hodnoty do výzkumu zahrne či nikoli, v případě průměrné mzdy s Billem Gatesem to jistě nebude vhodné.

Samozřejmostí při přípravě dat je testování normality. Existuje řada testů, které se liší silou, použitím na různých datech a také náročností provedení. Patří mezi ně např. Shapirův-Wilkův, Andersonův-Darlingův, Kolmogorovův-

Smirnovův, Lillieforsův. Testovat normalitu pomocí výpočtů v Excelu je časově poměrně náročná práce, tento postup již nelze doporučit. Naproti tomu testování normality pomocí programů, jako je R či Python aj., je zjednodušeno na zadání jednoho kódu v příkazovém řádku. Testování normality také bývá součástí základu těchto programů, a pokud nevyžadujeme testování pomocí nějakého méně známého testu, nepotřebujeme instalovat žádné balíčky či knihovny.

3.2.1 STATISTICKÁ ANALÝZA DAT PRO PREDIKCI INDIVIDUÁLNÍ POPTÁVKY

Při predikci individuální poptávky po značce automobilu Toyota Corolla, viz tabulka 3-1, je zjištěno, že medián a střední hodnota jsou velmi podobné. Šikmost naznačuje, že by se mohlo jednat o normální rozložení, špičatost je však na pováženou. Ze Shapiro-Wilkinsnova testu vyplývá, že zkoumaná časová řady není normálně rozložena. Jedná se tedy o jiné rozložení.

Tabulka 3-1 Statistiky časové řady vyhledávání pojmu Toyota Corolla

	Toyota Corolla
Mean	55,8712
Median	56
Variance	354,0062
Skewness	0,1702
Kurtosis	1,9533
Shapiro-Wilkinsův test: p-value	0,0003
ACF1	0,8235
Hurstův exponent	1,0617

Zdroj: vlastní zpracování

Koeficient ACF1 je blízký 1. To znamená, že v časové řadě se vyskytuje autokorelace, a je tedy vhodná k predikci, jelikož obsahuje opakující se prvky. Nicméně koeficient je možno použít jen u normálně rozložené časové řady, takže tato informace je zde pouze orientační.

Z toho důvodu byl použit Hurstův exponent. Toto je statistický parametr používaný k měření paměťových vlastností časových řad. Tento koeficient byl pojmenován po britském inženýru a hydrologovi Haroldu Edwinu Hurstovi (Hurst, 1951), který ho poprvé použil k analýze úrovní Nilu. Je definován jako poměr rozsahu fluktuací časové řady v různých časových měřeních k odhadu směrodatné odchylky fluktuací. Tento poměr se poté normalizuje logaritmickým měřítkem počtu pozorování, což umožňuje získat nelineární měření paměti nebo rekurze v časové řadě.

Práce Harolda Hursta je klíčovým zdrojem. Představuje první koeficient, který umožňuje výpočet autokorelací i pro časové řady, které nejsou z normálního rozložení. Hurstův exponent se vztahuje k dlouhodobé korelaci v časové řadě a měří míru paměti nebo rekurze v datech. Jeho výpočet není závislý na předpokladu

normálního rozložení dat. Často se používá k analýze různých typů časových řad, včetně finančních časových řad, hydrologických údajů, biologických dat a dalších. Využívá se k odhadu dlouhodobých trendů a předpovídání budoucího chování časových řad bez ohledu na to, zda jsou data normálně rozdělena nebo ne.

Hodnota Hurstova koeficientu může být v rozmezí od 0 do 1. Hodnota blíže k 0 naznačuje, že časová řada je náhodná nebo má krátkodobou paměť a neprojevuje dlouhodobé trendy nebo vzory. Hodnota blíže k 1 naznačuje, že časová řada má silné dlouhodobé korelace a vykazuje dlouhodobé trendy nebo vzory. Její predikovatelnost je tedy potvrzena.

3.2.2 STATISTICKÁ ANALÝZA DAT PRO PREDIKCI TRŽNÍ POPTÁVKY

Statistická analýza dat pro vyhledávání kategorií tržní poptávky je uvedena v tabulce 3-2.

Tabulka 3-2 Popisná charakteristika dat o vyhledávání zvolených témat

Data	Střední hodnota	Median	Rozptyl	Šikmost	Špičatost
Automobily	83,19008	85	92,50523	-0,56275	2,409227
Cestování	66,38843	66	164,7229	0,462998	2,8127
Domácí mazlíčci	75,34711	76	104,0785	0,224303	2,669966
Domov a zahrada	70,4876	72	155,6853	-0,22831	2,351617
Hobby a volný čas	74,64463	76	87,31433	0,099537	2,646898
Hry	78,95868	80	63,77328	0,17275	2,336901
Knihy a literatura	74,15702	74	119,0335	0,05411	2,602422
Nemovitosti	73,4876	74	122,4353	0,039192	2,77222
Počítače a elektronika	80,04959	80	63,58085	0,04029	2,586284

Zdroj: vlastní zpracování

Z výsledků by se dalo uvažovat, že data jsou normálně rozložena, střední hodnota a medián se příliš neliší. Šikmost se blíží k 0, což svědčí o tom, že mohou být normálně rozložená. Pro špičatost platí, že jsou data z normálního rozložení rovna 3. I zde jsou výsledky blízké 3, takže je možno říci, že jsou normálně rozložená. Tyto prvotní informace z popisné charakteristiky jsou ovšem pouze orientační a pro potvrzení či vyvrácení normality dat je potřeba provést vhodný test. Pokud také chceme predikovat poptávky, je nutné otestovat, zda jsou zvolená data k predikci vhodná. To můžeme udělat pomocí testování autokorelace. Pokud v časové řadě autokorelace bude, je zřejmé, že je možno v časové řadě vysledovat vzory, a je proto vhodná k predikci. Pokud v časové řadě žádné opakující se vzory nejsou, je její predikovatelnost diskutabilní a velmi problematická.

V další části datové analýzy byla tedy zkoumána autokorelace jednotlivých časových řad. Nejprve byl proveden test normality, a to Shapiro Wilkinsův test, viz tabulka 3-3. Na základě testu je patrné, že jen některé časové řady mají P-value větší než 0,05. Je tedy zřejmé, že náleží do normálního rozložení (Hobby a volný čas, knihy a literatura, nemovitosti, počítače a elektronika). Ostatní časové řady normální rozložení nemají (p-value je menší než 0,05). Z toho důvodu může být testování autokorelace pomocí kritéria ACF1 nevhodné.

Tabulka 3-3 Statistiky vyhledávání vybraných témat

	Shapiro-Wilkinsův test p-value	**ACF1**	**Hurstův exponent**
Automobily	0,000114033	0,8016854	1,052874
Cestování	0,021950051	0,7609712	0,7995749
Domácí mazlíčci	0,044796137	0,8815768	1,074766
Domov a zahrada	0,011198374	0,8875027	1,125611
Hobby a volný čas	0,096023621	0,7104613	1,085996
Hry	0,011616112	0,7750617	0,9160354
Knihy a literatura	0,48154676	0,7206558	0,632488
Nemovitosti	0,168998581	0,7150348	1,047134
Počítače a elektronika	0,687313335	0,7434545	0,8354227

Zdroj: vlastní zpracování

Z výsledků je patrné, že dle kritéria ACF1 jsou všechny časové řady poměrně silně pozitivně autokorelované. To naznačuje, že predikovatelnost těchto veličin je prokázána. Z výsledků Shapiro-Wilkinsnova testu vyplývá, že ne všechny časové řady jsou normálně rozložené. Proto byl použit i Hurstův exponent. Na jeho základě opět vidíme, že časové řady jsou predikovatelné, protože se jeho hodnoty blíží 0.

3.2.3 STATISTICKÁ ANALÝZA DAT PRO PREDIKCI POPTÁVKY PO ZNAČCE

V rámci podnikové poptávky predikujeme poptávku ve skupině konkurentů z řady automobilů, data již byly vizuálně posouzena. Nyní je provedena statistická deskripce. Z ní je patrné (viz tabulka 3-4), že střední hodnota a medián jsou podobné. Šikmost je blízká 0 a špičatost blízká 3. Je tedy možné, že data jsou normálně rozložena. Rozptyl je však velmi různý.

Tabulka 3-4 Popisná charakteristika dat jednotlivých značek automobilů

Značka	Střední hodnota	Medián	Rozptyl	Šikmost	Špičatost
Toyota	58,0000	61,0000	61,8000	0,6179	2,8860
Ford	82,0000	88,0000	87,3900	-0,2913	2,1638
Tesla	4,0000	5,0000	5,3330	0,3430	2,3149
Honda	42,0000	51,0000	48,7100	-0,4566	2,2816
Nissan	26,5000	28,0000	28,1200	0,3170	3,4251

Zdroj: vlastní zpracování

V tabulce 3-5 byl proveden test normality, a to Shapiro-Wilkinsův text. Normální rozložení dle tohoto testu mají pouze data automobilek Ford a Nissan. Koeficient ACF u většiny dat nevyšel blízký 1, takže se jeví jako nepredikovatelná, nicméně právě díky nevhodnosti použití testu to není opodstatněné. Hurstův exponent u všech časových řad se blížil 1. Predikovatelnost časových řad je tedy potvrzena a je možno je k predikci použít.

Tabulka 3-5 Statistiky časových řad vyhledávání automobilek dle značek

	Shapiro-Wilkinsův test p-value	ACF1	Hurstův exponent
Toyota	0,0427	0,3988	0,6944622
Ford	0,2298	0,3752	0,9129082
Tesla	0,0083	0,4126	1,039657
Honda	0,0308	0,8863	1,100124
Nissan	0,3912	0,1790	0,8002067

Zdroj: vlastní zpracování

3.3 DEKOMPOZICE ČASOVÝCH ŘAD

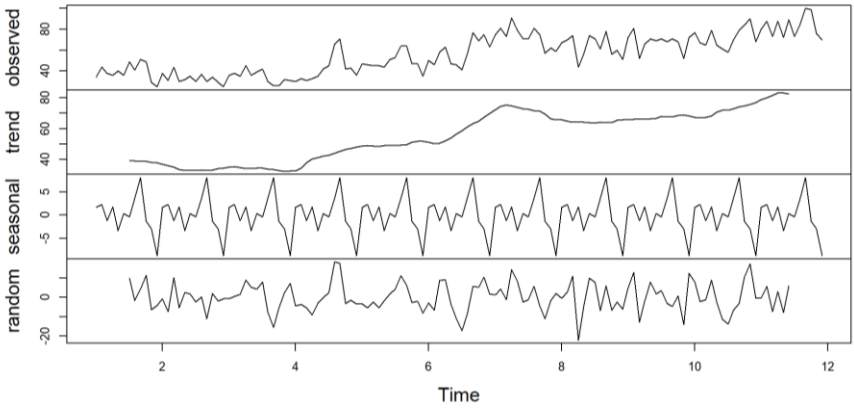
Časové řady je možno rozdělit do několika složek. Když ekonomickou časovou řadu rozložíme, obvykle dostaneme tři komponenty: trend, sezónní složku a složku náhodnou. Můžeme tak získat informace o týdenních či měsíčních přehledech tržeb nebo prodejů. Informace o trendové složce ale zůstává stále zásadní.

Metod dekompozice časových řad je několik. Až do poloviny minulého století byly hojně využívány metody založené na klouzavých průměrech. Dnes je často využívanou metodou klasický rozklad. Klasický rozklad může mít dvě formy – aditivní dekompozici a multiplikativní dekompozici. Tuto formu rozkladu však někteří nedoporučují (Hyndman & Athanasopoulos, 2018), protože předpokládají, že sezónní složka se v letech nemění, což nemusí vždy platit. Také tato metoda není schopna zachytit výkyvy způsobené neočekávanou událostí.

Jako reakce na Hyndmanovu kritiku klasické dekompozice vznikaly další metody a techniky dekompozic, například rozklad X11 popsaný Dagumem & Bianconcinim (2016). Jejich metoda je definována ve dvou podobách. Také je známa metoda X-12 ARIMA z článku Bureaua (Bureau, 2011). Další metoda

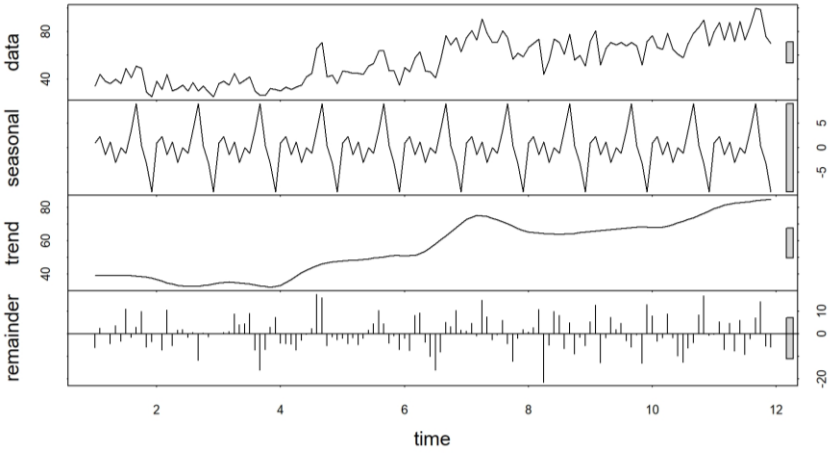
dekompozice se jmenuje SEATS, což je zkratka sezónní extrakce v časových řadách ARIMA. Ta však funguje jen se čtvrtletními a měsíčními údaji. Často ji využívají vládní agentury. Její další rozšíření je označováno TREAMO/SEATS, které definovali Goméz & Maravall (1997). Zajímavou metodou je také STL rozklad dle Clevelanda et al. (1990), což je zkratka pro sezónní a trendový rozklad pomocí Loess (metoda pro odhad nelineárních vztahů), která vznikla až v 90. letech minulého století a je v současnosti nejčastěji používaná.

Pro rozhodnutí, jaká dekompozice bude použita zde, porovnáme dekompozici klasickou (aditivní či multiplikativní) a STL. Na základě automatické funkce *decompoze*, která je součástí balíčku *forecast*, bylo zjištěno, že z klasických dekompozic je vhodnější metoda aditivní dekompozice. Výsledek je na obrázku 3-4.



Obrázek 3-3 Aditivní dekompozice časové řady vyhledávání značky Toyota Corolla. Zdroj: vlastní zpracování

Pro porovnání je na obrázku 3-5 vyobrazena i dekompozice STL. Obě mají různá měřítka. U dekompozice STL je na pravé straně uvedena i lišta, která znázorňuje velikost stejného měřítka. V tomto případě je možno konstatovat, že obě dekompozice přinesly téměř stejné výsledky. Sezónní dekompozice má naprosto stejný výsledek. Trend je v dekompozici pomocí STL více vyhlazený. Náhodná složka je na první pohled u dekompozice aditivní větší.



Obrázek 3-4 STL dekompozice časové řady vyhledávání značky Toyota Corolla.
Zdroj: vlastní zpracování

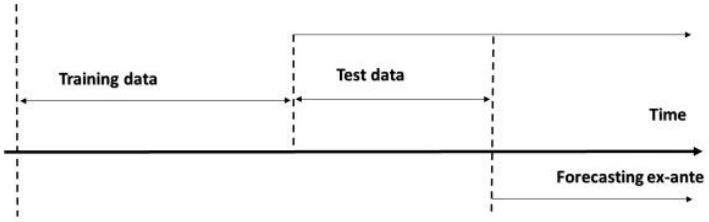
Ačkoli výsledky jsou velmi podobné, ve vědeckých pracích se dnes již více využívá metoda STL. Náhodná složka je u aditivní dekompozice pro naše data větší, a proto i v této publikaci bude využita metoda STL. V dalším postupu byla provedena dekompozice dle STL metody pro vybrané skupiny tržní poptávky, která je pro velký rozsah uvedena v příloze 2. Následně pak také pro všechny značky vozidel – příloha 4.

Kapitola 4

Metody prognózování poptávky

Metod prognózování je v současné době celá řada, a jak již bylo deklarováno v kapitole 1, stále nelze jednoznačně říct, která je nejvhodnější. Historicky nejstarší jsou metody statistické, a proto jsou využívány i v klasických statistických programech jako SPSS, Excel a jiné. Výpočet těchto modelů se dá označit jako nejméně náročný na znalosti prognostika i na výpočetní výkon či výpočetní čas, nikoli ovšem čas prognostika. Další metody definované v této kapitole jsou inspirovány přírodou. Speciálním případem jsou metody založené na umělé inteligenci. Sem patří zejména umělé neuronové sítě. A nejnovější jsou metody kombinované a hybridní, které předchozí dva koncepty kombinují, případně transformují.

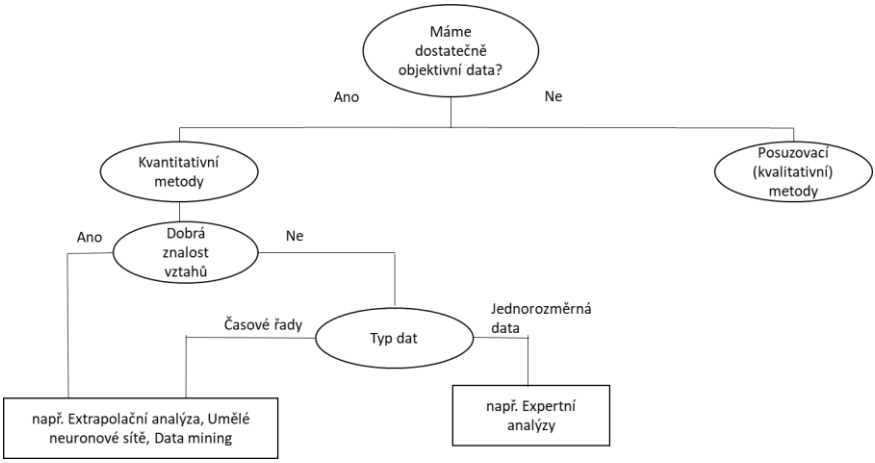
Pro výběr nejvhodnějšího modelu je nutné rozdělit historická data z časové řady na dvě části. A to tréninková data (training data) a testovací data (test data). Tréninková data použijeme k odhadu všech parametrů prognostického modelu a následně použijeme testovací data k samotné prognóze a ověření míry přesnosti daného modelu. Na základě přesnosti výpočtu na testovacích datech pak můžeme vybrat nejvhodnější model. Velikost testovacích dat je obvykle okolo 20 % z celkové časové řady, jak ukazuje obrázek 4-1. Nicméně jak uvádí Hyndman & Athanasopoulos (2018), je to také závislé na tom, jak dlouhou historickou časovou řadu máme a jak dlouhou chceme prognózu.



Obrázek 4-1 Rozdělení dat na tréninková a testovací
Zdroj: vlastní zpracování dle (Hyndman & Athanasopoulos, 2018) a (Marček, 2013))

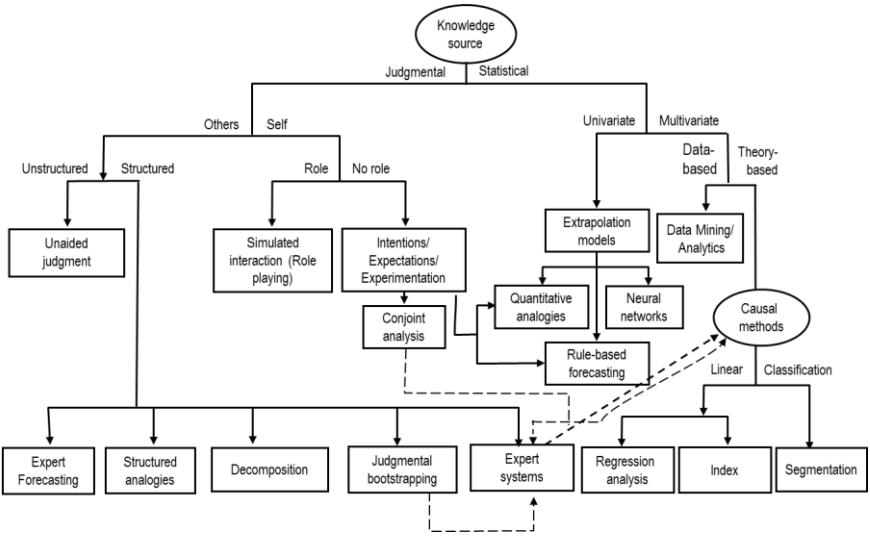
4.1 ZÁKLADNÍ ČLENĚNÍ METOD PROGNÓZOVÁNÍ POPTÁVKY

Metody prognóz lze v obecné rovině členit dle několika hledisek. Nejznámější a nejpoužívanější členění rozděluje metody na kvalitativní a kvantitativní. Miller a Swinehart (2010) kategorizovali metody do tří různých skupin, a to do průzkumných nebo normativních metod, metod založených na expertech, založených na důkazech nebo na předpokladech, třetí skupina členění pak je klasické členění na kvalitativní a kvantitativní metody. Moro a Cortez (2015) klasifikovali metody na kvantitativní, semikvantitativní a kvalitativní metody, toto je členění v současnosti nejpoužívanější. Armstrong a Green člení metody prognózování na posuzovací a kvantitativní s celou škálou dalších dílčích členění. Zjednodušený základní postup pro výběr metod (se zaměřením na metody kvantitativní) je uvedeno v obrázku 4-2. Toto členění vytvořili Armstrong a Green v roce 2007.



Obrázek 4-2 Zjednodušení výběru metod predikcí dle Greena a Armstronga v roce 2007
Zdroj: vlastní zpracování dle forecastingprinciples.com JSA-KCG November 2007

Později autoři (Armstrong & Green, 2014) provedli aktualizaci na současnou podobu, která je zachycena na obrázku 4-3. Toto rozdělení i nadále vychází z členění na metody posuzovací a kvantitativní a znázorňuje i vazby mezi jednotlivými výběry modelů. Metody nazvané jako posuzovací (v originále Jundgmental) je možno ztotožnit s pojmem metoda kvalitativní uváděná i jinými autory.



Obrázek 4-3 Doplněné a přepracované členění prognostických metod dle Armstronga a Greena
Zdroj: Armstrong & Green, 2014

Členění prognóz je celá řada, obvykle je spojeno s konkrétním oborem, pro který je vytvářeno. Členění metod tak reflektuje zejména potřeby jednotlivých prognostiků v závislosti na zvoleném oboru, který je relevantní pro předmět prognózy. Základní rozdělení metod na kvalitativní a kvantitativní ovšem zůstává všem oborům vlastní.

4.1.1 KVALITATIVNÍ METODY

Metody odborného posuzování (Judgmental) lze ve členění dle Mora & Corteze (2015) nazvat kvalitativní. Kvalitativní metody se obvykle neopírají o numerické vyhodnocení dat, ale o odborné posuzování a slovní hodnocení zkoumaných veličin. K těmto metodám patří například expertní panel, kdy skupina odborníků dané organizace, nicméně z různých oblastí studuje a diskutuje danou veličinu (Wisniewski, 1996). Další metodou jsou relevantní stromy (Daim et al., 2006), což je metoda identifikující vývojové fáze, cíle a základní prvky dané podnikové veličiny. Velmi podobnou metodou je také budoucí kolo (futures wheel), kdy zkoumaná událost nebo veličina jsou považovány za jádro nebo střed a události či veličiny, které ji mohou ovlivnit, se považují za hybné síly.

Velmi známou a používanou je metoda SWOT analýza, pomocí níž odborníci identifikují silné nebo slabé stránky, příležitosti a ohrožení společnosti nebo produktů. Vyhledávanou metodou je také rešerše dostupných vědeckých zdrojů (Moro & Cortez, 2015). Na rozhraní mezi metodami kvantitativními a kvalitativními stojí semikvantitativní metody prognózování, které stále můžeme zahrnout dle obrázku 4-2 a 4-3 do metod odborného posuzování.

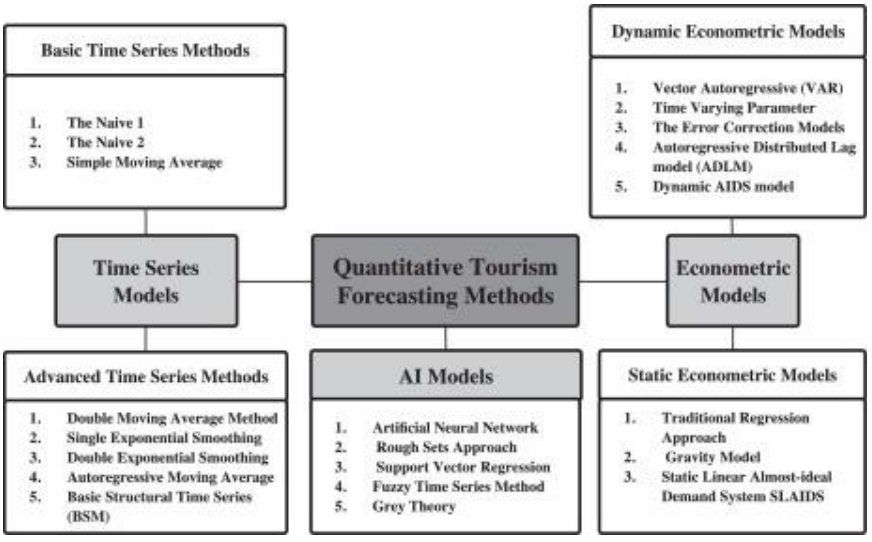
Mezi semikvantitativní metody patří například technologický monitoring, který využívá systematické smyčky k nalezení ideálních podmínek pomocí informací o zpětné vazbě. Další z oblíbených metod je brainstorming, který sbírá představy o budoucnosti jednotlivce nebo skupiny lidí. Je možno sem zařadit i morfologickou analýzu, dotazník/průzkumy, plánování scénářů.

Velmi oblíbenou je také Delphi metoda (Esmaelian et al., 2017), která používá dotazníky v kolech po sobě jdoucích s cílem shromáždit názory co největšího počtu odborníků a dosažení konsensu. Další metodou je mapování zúčastněných stran (Saritas et al., 2013), (Vishnevskiy et al., 2015) využívající statistické techniky k předpovědím toho, kdo jsou zainteresované strany, kde se nacházejí a proč mají zájem o daný produkt či službu atd. Velmi moderní a nová je také metoda dolování dat, kterou využil například Moro a Cortez (2015). Touto metodou se už ale dostáváme na pomezí metod kvantitativních.

4.1.2 KVANTITATIVNÍ METODY PROGNÓZOVÁNÍ

Tyto metody jsou obvykle založeny na matematicko-statistických technikách a numerických výpočtech, jak uvádí Esmaelian et al. (2017), dnes i na složitějších výpočtech na bázi umělé inteligence. Patří mezi ně analýza trendů a trendová extrapolace. Další metodou je například modelování a simulace, kdy předpovídáme budoucí stav veličiny pomocí modelování skutečnosti (Zmeškal et al., 2013). Multi-stage analýza pak kombinuje několik modelů.

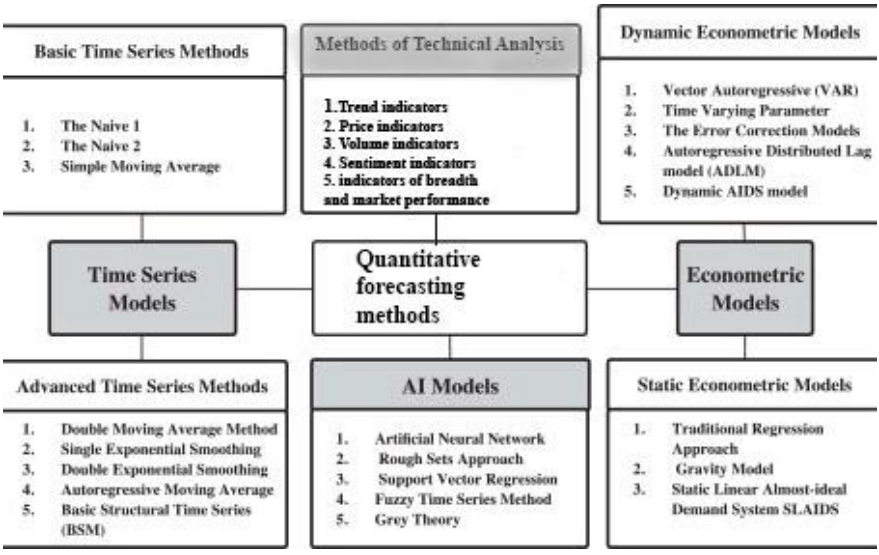
Můžeme sem zařadit i méně známou metodu Future workshop dle Martina (Martino, 2003). A v neposlední řadě je zde i skupina dynamických metod. Ty prognózují na základě dynamických nástrojů, jako jsou například neuronové sítě, fuzzy logika, genetické algoritmy, případně teorie chaosu (Dostál et al., 2005).



Obrázek 4-4 Členění kvantitativních metod prognózování
Zdroj: Esmaelian et al., 2017

Kvantitativní metody můžeme členit dle nejrůznějších hledisek. Jedno ze základních provedl Esmaelian et al. (2017), viz obrázek 4-4. Bylo použito primárně na kvantitativní metody prognózování v oblasti turismu, nicméně lze je zobecnit.

Toto členění je možno opět rozšířit o další metody, případně kombinace. Publikace Kolkové (2018a) na ně navazuje a doplňuje o metody založené na technických indikátorech, viz obrázek 4-5. Metody založené na technických indikátorech jsou experimentem autorky a využila ji i v další publikaci (Kolková & Ključnikov, 2021). Experiment bude ještě podroben hlubšímu zkoumání a využití technických indikátorů v podnikové ekonomice bude také více prozkoumáno. Na obrázku 4-5 je tedy zobrazeno potenciální zařazení těchto ukazatelů mezi standardní metody prognózování.



Obrázek 4-5 Členění metod predikcí dle Kolkové
Zdroj: Kolková, 2018a

Další část této publikace se bude zabývat pouze metodami kvantitativními. Je samozřejmé, že i metody kvalitativní mají v ekonomických disciplínách své opodstatnění, nicméně tato publikace na ně není cílena.

4.2 STATISTICKÉ METODY PROGNÓZOVÁNÍ POPTÁVKY

Jak už bylo uvedeno, nejstarší jsou metody statistické. V podnikové praxi stále mají své velké uplatnění zejména pro menší podniky či podniky nedisponující pracovníky se znalostmi v oblasti datové analytiky. Nejznámější jsou modely založené na exponenciálním vyrovnání a ARIMA modely.

4.2.1 PROGNÓZY NA ZÁKLADĚ EXPONENCIÁLNÍHO VYROVNÁNÍ

Exponenciální vyrovnání se používá na krátkodobé prognózování v různých modifikacích. Prognózy založené na exponenciálním vyrovnání jsou vážené průměry minulých hodnot, přičemž váhy se exponenciálně snižují se stářím použitých dat (Hyndman & Athanasopoulos, 2018). Nebo také jak uvádí Bergmeir et al. (2016) „obecná myšlenka exponenciálního vyrovnání je taková, že nedávná pozorování jsou pro prognózu relevantnější než starší pozorování, což znamená, že by měla mít vyšší váhu než pozorování starší“.

Zkratka ETS vznikla z anglického ExponenTial Smoothing a používá se pro všechny metody. Metody exponenciálního vyrovnání mohou mít ale i své názvy, zejména ty v praxi používané. A to jednoduché exponenciální vyrovnání, Holtova lineární metoda, Aditivní tlumená trendová metoda, aditivní HoltWintersova metoda, Multiplikativní Holt-Wintersova metoda, Holt-Wintersova tlumená metoda. Tyto pojmy se obvykle nepřekládají, v tabulce 4-1 jsou uvedeny v používaných anglických názvech a s označením trendového a sezónního komponentu. Toto členění vychází zTaylora (2003).

Tabulka 4-1 Metody prognóz na základě exponenciálního vyrovnání

(Trend, Sezónnost)	**Metoda**
(N,N)	Simple exponential smoothing
(A,N)	Holt’s linear method
(A_d,N)	Additive damped trend method
(A,A)	Additive Holt-Winters’ method
(A,M)	Multiplicative Holt-Winters’ method
(A_d,M)	Holt-Winters’ damped method

Zdroj: Taylor, 2003

Klasifikace tímto způsobem lze označit za klasifikaci dvoucestnou a je naznačena v tabulce 4-2. Můžeme vybrat trend N (žádný), A (aditivní), A_d (aditivní tlumený). Nebo sezónní komponenty N (žádné), A (aditivní), M (multiplikativní). Taylor (2003) toto členění neoznačoval názvem konkrétních metod, ale vycházel z teoretických dvojic trendu a sezónnosti.

Tabulka 4-2 Dělení metod exponenciálního vyrovnání podle sezónních a trendových komponent dle Taylora

Trendové komponenty	**Sezónní komponenty**		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
A_d (Additive damped)	(A_d,N)	(A_d,A)	(A_d,M)

Zdroj: Taylor, 2003

Samozřejmě lze tyto metody popsat i vzorci jako v tabulce 4-3 dle Hyndmana & Athanasopoulose (2018). Jednotlivé znaky označují ℓ_t úroveň časové řady v čase *t*, b_t znamená sklon v čase *t*, s_t označuje sezónní složku časové řady v čase *t* a *m* označuje počet ročních období v roce, α, β^*, γ a ϕ jsou vyrovnávací parametry.

Tabulka 4-3 Vzorce výpočtu metod exponenciálního vyrovnání

Trend	**Seasonal**		
	N	**A**	**M**
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
A_d	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$

Zdroj: Hyndman a Athanasopoulos, 2018

Tyto metody generují předpovědi bodové, je možno ale předpovídat i predikční intervaly bodových předpovědí. Platí pak, že pro každou metodu existují ještě dva modely, a to s aditivními chybami a s multiplikativními chybami. Pokud jsou použity stejné vyhlazovací parametry, jsou bodové předpovědi obou typů chyb stejné, nicméně predikční intervaly budou různé. Proto je nutné tyto modely rozlišit. Pro rozlišení mezi modelem s aditivními chybami a s chybami multiplikativními přidáme do klasifikace v tabulce 4-3 ještě třetí písmeno, výsledný model pak označujeme ETS, (chyba (error), trend (trend), sezónnost (seasonal)). V tabulce 4-4 je příklad ETS modelů a jejich značení v anglickém jazyce.

Tabulka 4-4 Příklad ETS modelů a jejich značení

Označení	Metoda
ETS(M,N,N)	Simple Exponential Smoothing with Multiplicative Errors
ETS(A,N,N)	Simple Exponential Smoothing with Additive Errors
ETS(A,A,N)	Holt's Linear Method with Additive Errors
ETS(M,A,N)	Holt's Linear Method with Multiplicative Errors

Zdroj: Hyndman & Athanasopoulos, 2018

4.2.2 AUTOREGRESNÍ PROGNOSTICKÉ MODELY

Dalším statistickým přístupem k předpovídání časových řad jsou modely ARIMA (Auto Regresive Integrated Moving Average). Dvojice exponenciálního vyrovnání a ARIMA modely tvoří nejpoužívanější přístupy k predikci časových řad (Hyndman & Athanasopoulos, 2018).

Než přistoupíme k definici ARIMA modelů, je vhodné vysvětlit základní pojmy. Stěžejní ve vysvětlení ARIMA modelů je pochopení pojmu stacionární časová řada. Stacionární časová řada je taková, jejíž vlastnosti nezávisí na době, kdy je sledována. Je tedy zřejmé, že časové řady s trendem a sezónností nejsou stacionární. Příkladem stacionární časové řady je časová řada založená na ryze náhodném procesu, tzv. bílý šum. Někdy ovšem mohou být časové řady matoucí a jak uvádí Hyndman a Athanasopoulos (2018), časová řada s cyklickým chováním, ale bez trendu a sezónnosti je stacionární.

Samozřejmě ale existují způsoby, jak nestacionární řady převést na stacionární, tím je zejména nahrazení hodnot jejich změnami. To je známo jako diferenciace. Diferenciace časových řad, například pomocí logaritmů, může stabilizovat rozptyl časové řady a tím eliminovat trend a sezónnost.

ARIMA modely jsou založeny na myšlence, že hodnoty pozorování veličiny v časové řadě mohou být chápány jako realizace náhodného procesu. ARIMA modely jsou tedy, jak už bylo řečeno, určeny pro časové řady se stochastickým trendem nebo pro takové, které lze převést na stacionární diferencováním (Cipra, 2013).

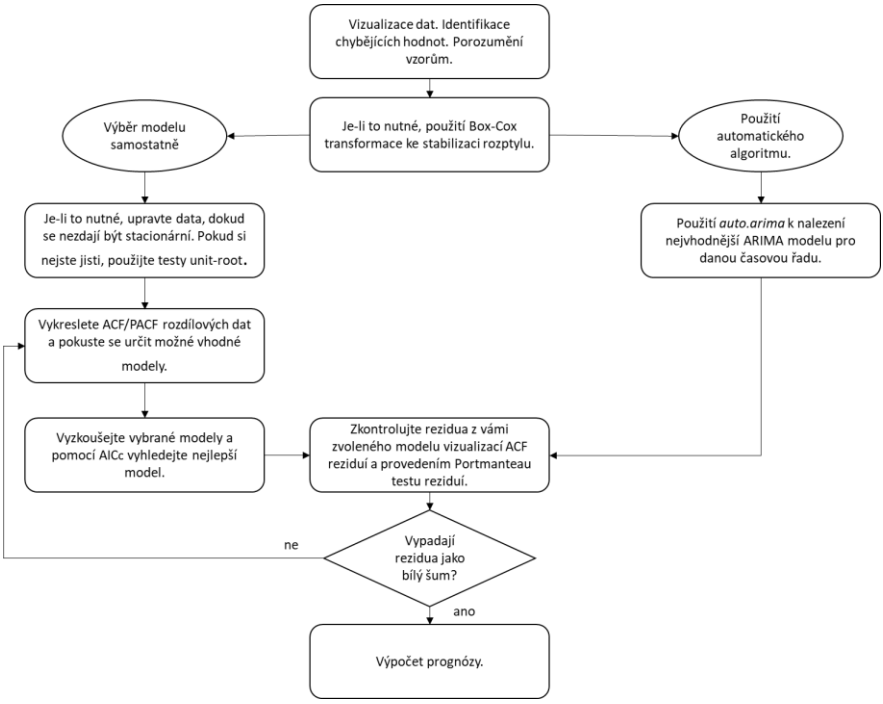
AR (auto regresive) v této zkratce vyjadřuje, že část hodnoty časové řady se dá vysvětlit jako lineární kombinace minulých hodnot. V modelu vícenásobné regrese předpovídáme proměnnou pomocí lineární kombinace veličin. V autoregresním modelu naopak předpovídáme proměnnou pomocí lineární kombinace minulých hodnot proměnné. Termín autoregrese tedy lze označit jako regresi proměnné vůči sobě. Další písmeno v modelu ARIMA je I (integrated) a vyjadřuje onu diferenciaci časové řady, například pomocí logaritmu. Konečně MA (moving average) jsou klouzavé průměry a opírají se o myšlenku, že část rezidua časové řady se dá vysvětlit jako lineární kombinace minulých chyb. Pokud tedy

zkombinujeme diferenciaci s autoregresí a modelem s klouzavým průměrem, dostaneme sezónní model ARIMA.

Modely ARIMA zapisujeme podobně jako ETS s využitím písmen v závorce, tzn. ARIMA (*p,d,q*), kde *p* představuje stupeň autoregresní části, *d* je stupeň diferenciace a *q* je stupeň části klouzavých průměrů. Z tohoto lze vyvodit i zvláštní případy modelu ARIMA. Pokud bychom chtěli jako ARIMA model zapsat bílý šum, pak by to vypadalo takto: ARIMA(0,0,0), náhodná procházka by se zapsala ARIMA(0,1,0), klouzavý průměr by pak byl zapsán jako ARIMA(0,0,*q*).

Kromě nesezónních ARIMA modelů, které byly definovány doposud, rozeznáváme ještě sezónní modely ARIMA, jejichž zápis je pak ARIMA(*P,D,Q*)*m*, kde *m* je počet pozorování za rok. Také se v teorii odlišují velká písmena pro sezónní části modelu a malá písmena znamenají nesezónní model. Zvláštním případem pak mohou být metody ARFIMA, které se odlišují v tom, že parametr *d* může nabývat i necelých hodnot. Dalšími možnými rozšířeními této metody je například SARIMA či SARIMAX.

Pro výpočet těchto modelů je velmi rychlým řešením automatizovaný proces v jazyce R, který vytvořili Hyndman a Athanasopoulos (Hyndman & Athanasopoulos, 2021). Pomocí jednoho kódu *auto.arima* nalezne automatizovaný proces pro zvolená data nejvhodnější model z ARIMA modelů. Zjednodušení celého procesu je deklarováno na obrázku 4-6.



Obrázek 4-6 Zjednodušení ARIMA modelů pomocí automatizovaného *auto.arima*
Zdroj: Hyndman & Athanasopoulos, 2021

V příloze 10 je ukázka výpočtu modelu ETS a ARIMA modelů prostřednictvím programu SPSS, který se používá hlavně k výpočtu těchto dvou modelů. Umožňuje i automatický výběr nejvhodnějšího z nich. Ukázka výpočtu ETS a ARIMA modelů a vybrané kódy v jazyce R jsou v příloze 11. Pro výpočty prognóz již nebude použit tabulkový procesor Excel ani programy jako SPSS, ale jak již bylo deklarováno, pouze statistický jazyk R. I Excel již umožňuje realizaci jednoduchých prognóz, nicméně pro praktické využití v podnicích lze stále rozhodně doporučit použití některého ze statistických programů.

Jazyk R nabízí několik balíčků, pomocí nichž lze provést efektivní prognózy. V této publikaci je využit balíček Hyndmanův z Univerzity v Melbourne a Khandakary z Monash univerzity, významných australských statistiků (Hyndman & Khandakar, 2008). Tento balíček je znám pod jménem *forecast*. V balíčcích jsou již mnohé funkce na výpočet prognóz nadefinované a automatizované. Samozřejmě u všech prognóz, tak jak bylo uvedeno v předchozích kapitolách, je vhodné volit metodu pomocí rozdělení dat na tréninková a testovací a na testovacích datech ověřit míru přesnosti modelu vytvořeného na datech tréninkových.

4.2.3 DALŠÍ MODELY

Další z modelů používaných v této práci je **BATS**. Vychází z modelů předchozích, ale pro svou unikátnost tvoří už samostatnou skupinu. Model BATS představuje zobecnění tradičních modelů sezónních, umožňující více sezónních období (De Livera et al., 2011).

Pojmenování modelu BATS je vlastně zkratka pro klíčová slova z modelů Box-Cox transformace, ARMA model, trend a sezónní komponenty. Toto označení musí být doplněno příslušnými argumenty (ϖ, ϕ, p, q, m_1, $m_2 \ldots\ldots\ldots m_T$) pro definování Box-Cox parametrů a damping parametrů p a q pro vyjádření parametru ARMA modelů a sezónní periodicita je vyjádřena argumenty m_1, $m_2 \ldots\ldots\ldots m_T$.

Velmi zajímavý je také model **Theta**. Byl vytvořen autory Assimakopoulosem a Nikopoulosem v roce 2000 (Assimakopoulos & Nikolopoulos, 2000). Metoda je založena na konceptu modifikace lokálních výkyvů časové řady pomocí koeficientu Theta (označovaného řeckým písmenem θ). Tento koeficient je aplikován na druhou diferenci časové řady dle vztahu

$$X''_{new} = \theta \cdot X''_{data}, \text{ kde} \tag{4.1}$$

$$X''_{data} = X_t - 2 \cdot X_{t-1} + X_{t-2} \tag{4.2}$$

Počáteční časová řada je rozložena na dvě nebo více Theta-linií a každá z nich je pak extrapolována samostatně. Předpovědi jsou pak jednoduše kombinovány. Pro každý horizont prognózy lze použít i jinou kombinaci Theta-linií.

Na základě této myšlenky se pak model Theta vytváří v postupných krocích. První krok je testování sezónnosti na základě t-testu. V dalším kroku dojde k odstranění sezónnosti klasickou multiplikativní dekompoziční metodou. Následuje samotný rozklad na dvě (případně více) Theta-linií. V základním nastavení modelu se pracuje s jednou linií pro $\theta=0$, což představuje lineární regresní přímku a $\theta=2$. Po rozkladu je provedena extrapolace, kdy lineární regresní přímka je extrapolována obvyklým způsobem a druhá linie je extrapolována pomocí jednoduchého exponenciálního vyrovnání. Ve čtvrtém kroku jsou vytvořeny kombinace, a to v základním nastavení se stejnými váhami. Nakonec jsou prognózy resezónizovány. V balíčku *forecastHybrid* je metoda Theta označena jako Thetam a její výpočet je zpřesněn. Model Thetam je také využit v této monografii.

Při dalším rozvoji této metody se předpokládá, že se využije více než dvouliniová metoda, v současnosti ale není jisté, zda použití více linií povede k přesnějším výsledkům. Další možností rozvoje je použití různých vah při kombinaci Theta-linií. Nyní se další výzkum týká využití různých Theta-linií podle kvalitativních a kvantitativních charakteristik dané časové řady.

4.3 PROGNOSTICKÉ METODY INSPIROVANÉ PŘÍRODOU

Výčet prognostických metod rozhodně není úplný a s rozvojem vědění se bude neustále doplňovat, zpřesňovat i upravovat. Je zřejmé, že prognózování v podnikové ekonomice také již nebude pouze doménou ekonomů, ale stále více bude inspirováno i jinými vědními obory. Propojování jednotlivých vědních oborů bude pravděpodobně jednou z nejzajímavějších výzev výzkumu v budoucnosti.

Už Einstein kdysi řekl: „Vidíme překrásně uspořádaný vesmír, který dodržuje spolehlivé zákony, jenže těmto zákonům rozumíme jen mlhavě“. Dnes již některým zákonům možná rozumíme o trochu méně mlhavě, a to i díky Einsteinovi, a stále více vědců také začíná chápat, že žádný vědní obor nestojí samostatně ve vzduchoprázdnu. Věda je propojená síť poznatků, empirie i zkušeností, a i ekonomové mohou hledat a také hledají inspiraci v jiných vědních disciplínách. Prognózování tak může být inspirováno přírodou a přírodními i jinými vědami. Na prognózy můžeme aplikovat Fibonacciho čísla, Elliottovy vlny, teorii her, genetické algoritmy, fuzzy logiku i teorii chaosu. Tento výčet určitě není úplný a než vyjde tato publikace, je pravděpodobné, že někdo objeví další souvislosti ekonomiky s přírodou. Nechme se překvapovat a nepřestávejme žasnout nad dalšími poznatky, které nám vědci v budoucnu představí.

Řešení těchto metod v SPSS, nebo dokonce v Excelu je téměř nereálné nebo příliš časově náročné. Jazyk R a některé jeho balíčky umožňují například výpočet umělých neuronových sítí.

4.3.1 FIBONACCIHO ČÍSLA

Fibonacciho posloupnost čísel je nástroj známý již přes 800 let. Jeho tvůrcem je Leonardo Pisano, který měl přezdívku Fibonacci. Byl to italský matematik, který žil v letech 1175–1250. Často cestoval po světě a z navštívených zemí se vracel se znalostí matematických systémů. Zasloužil se také o konečné zavedení arabské číselné řady místo římské. Je mu připisováno několik matematických objevů. Nejznámějším je definování Fibonacciho posloupnosti. Její zákonitost byly známy již starověkým Egypťanům, Fibonacci byl ovšem první, kdo byl schopen tuto posloupnost popsat.

Základem Fibonacciho práce je objev tzv. zlatého poměru. Tento zlatý poměr je 1:1,618. Na něm je postavena řada přírodních zákonů, a to jak v oblasti živočišné, tak rostlinné říše. Pisano tento svůj objev popsal v knize Liber Abaci (1202). Dnešní matematika uplatňuje i několik dalších Fibonacciho poměrů, které se využívají v nejrůznějších vědních a uměleckých oborech.

Jeden z příkladů je v semenech slunečnice, kde jsou jednotlivá semínka uspořádána do spirál velikosti dvou po sobě jdoucích čísel posloupnosti. Další příklad je možno spatřit ve spirálách zdřevnatělých lístků plodu artyčoku. Vedou dvěma směry a jejich počet kolem stonku opět tvoří dvě po sobě jdoucí čísla Fibonacciho posloupnosti. Podobně se tato vlastnost objevuje u šišek některých jehličnatých stromů (například borovice), i tam většinou počet spirál tvoří Fibonacciho posloupnosti (Pappas, 1989). Zlatý poměr může také měřit otáčky

mezi po sobě rostoucími listy na stonku rostliny, což může platit pro jilm, lípu, buku, lísky, ale například také pro růže (Coxeter 1969; Ball & Coxeter 1987).

Fibonnacciho posloupnost také definuje genealogii včel, stejně jako složení jejich pohlaví. Jeho objev lze nalézt i ve velikosti sousedních vrstev listů zelí, která zachovává opět poměr dvou po sobě jdoucích čísel Fibonnacciho posloupnosti, na poměru velikostí dvou komůrek ulity některých plžů odpovídá poměr dvou po sobě jdoucích čísel Fibonacciho sekvence. Obdobný tvar lze najít u rohů některých druhů čeledi Bovidae. Fibonnacciho posloupnost lze najít u celé řady vyšších rostlin, např. v poměru velikostí jejích listů, nebo úhlu, kterým ze stonku vyrůstají.

Kyž to zobecníme, Fibonacciho posloupnost znamená, že číslo následující je vždy součtem dvou čísel předcházejících, takže je to nekonečná posloupnost čísel začínající 0, 1, 1, 2, 3, 5, 8, 13, 21 atd.

Na finančních trzích se ale v rámci nástrojů technické analýzy uplatňuje spíše logika zlatého poměru. A to tak, že číslo následující je vždy 1,618násobkem čísla předcházejícího a zároveň 0,618násobkem čísla následujícího. Techničtí analytikové pak využívají různé grafické prostředky jako nástroje k predikci kurzů, a to Fibonacciho oblouky (Fibonacci Arcs), Fibonacciho vějíře (Fibonacci Fan Lines), Fibonacciho čáry zpětného návratu (Fibonacci Retracements), Finacciho časové zóny (Fibonacci Time Zones), Fibonacciho expanze (Fibonacci Expansion). Například na forexovém (měnovém trhu) jsou nejznámější cenové úrovně 23,6 %, 32,8 % a 61,8 %, přitom za nejvýznamnější hodnotu je na Forexu označována 61,8 %. Možnosti prognózování na finančních trzích pomocí Fibonacciho čísel analyzovala například Kolková (2017).

4.3.2 TEORIE ELLIOTTOVÝCH VLN

Ralf Nelson Elliott pracoval původně jako účetní na železnici, později pak jako obchodní konzultant. Současně s ekonomikou se zabýval také matematikou, fyzikou a biorytmy. Ve spolupráci s Ch. J. Collinsem publikoval svou teorii vln poprvé v roce 1938 v knize The Wave Principle a pak rozšířil v roce 1946 v knize The Nature's Law: The Secret of the Universe. V roce 1935 předpověděl 13 hodin předem důležité hodinové minimum akcií Industrial a tím se zapsal do historie technické analýzy akciových trhů. Elliott také navázal na matematické postupy Leonarda Fibonacciho. Čísla z Fibonacciho posloupnosti můžeme v Elliottových vlnách nalézt a dalo by se říct, že je jeho základem. Nicméně jak sám Elliott řekl, principy vln objevil v době, kdy o Fibonaccim neměl ani ponětí. Takže se lze domýšlet, že oba velcí matematici došli k podobným závěrům. Jak řekl Casti (2002) "je to jako bychom byli nějak matematikou naprogramováni. Mušle, galaxie, sněhová vločka nebo člověk: všichni jsme vázáni stejným řádem." Na Elliottovy vlny navázal pak Charles Dow svou známou teorií.

Teorie Elliottových vln vychází z předpokladu, že v přírodních jevech existují určité periody. Jelikož hlavním subjektem ekonomiky je člověk, je možné i jeho chápat jako součást přírody. Proto i v ekonomice můžeme nalézt střídání období konjunktury a recese.

Teorie Elliottových vln je primárně založena na psychologii trhu a používá se zejména pro predikci kurzů. Při spekulacích na finančních trzích jsou Elliottovy vlny dodnes využívány (Ribeiro, 2019), (Kostin, 2019) nebo (Maranon, 2018). Pro predikci v podnikové ekonomice se Elliottovy vlny nepoužívají, ale uvažujeme-li, že i podnik není nic jiného než společenstvím lidí (ať už vlastníků, zaměstnanců, tak i zákazníků, dodavatelů či odběratelů, ale i konkurence) a tito lidé jsou opět součástí přírody s jejími cykly, lze i v podniku vysledovat vlny. Aplikace Elliottových vln na predikci podnikových veličin je teoreticky možná.

4.3.3 TEORIE HER

Teorie her je jako vědní disciplína poměrně nová, nicméně její základy se formulovaly snad už od počátku existence člověka. Respektive od okamžiku, kdy si člověk začal uvědomovat, že výsledky rozhodnutí o problémech a situacích, které denně řeší, nemusí být pouze výsledkem jeho aktivity, ale mohou být závislé také na rozhodnutí dalších inteligentních bytostí.

V teorii her jsou hry často prezentovány maticí, jež zobrazuje hráče a možné strategie, které hráč může realizovat. V matici se pak zobrazí zisky z jednotlivých rozhodnutí. Jedním ze základních typů her je vězňovo dilema, odborným jazykem hra s nenulovým součtem. Zde máme dva hráče, což jsou vězni, kteří mají možnost spolupracovat či nespolupracovat s vyšetřovateli. Policisté vězně drží odděleně, a nemohou se tedy domlouvat. Když jeden promluví a druhý ne, vyšetřovatelé označí za viníka druhého. Ten, který promluvil, je pak volný. Když budou oba mlčet, dostanou trest oba, ale kratší, protože jim nelze dokázat vše. Když naopak oba promluví, oba dostanou plný trest. Nejefektivnější rozhodnutí tedy je, když oba budou mlčet, což můžeme označit překvapivě za optimum. Nicméně racionální je pro každého zvlášť spíše situace, kdy budou vypovídat právě z toho důvodu, že se nemohou dohodnout a při mlčení by stále byli pod tlakem možnosti, že druhý naopak vypovídal. Experimentální průzkumy hovoří o tom, že kooperativně (oba mlčí) se chová pouze 40 % hráčů.

První publikované poznatky o teorii her můžeme nalézt u Emila Borela v letech 1921 až 1927. Za zakladatele teorie her je však považován John von Neuman, který o teorii her poprvé promluvil v roce 1928. Jako samostatná matematická disciplína pak teorie her byla definována po vydání rozsáhlé sbírky Theory of Games and Economic Behavior z roku 1944, kde se John von Neumann spojil s Oskarem Morgensternem. Významným vědcem, který se také zasadil o rozvoj teorie her, je samozřejmě také John Forbes Nash, který publikoval řadu výzkumů a 1996 celou knihu o této teorii (Nash, 1996).

Dnes je nejznámější aplikací teorie her v ekonomii a vojenství. Teorii her ale můžeme využít i v jiných oborech, například v sociologii, psychologii, právu, politologii, či dokonce biologii. Logistické modely skladování při kolísavé poptávce řeší například Zhang et al. (2020), problematiku veřejných statků na bázi teorie her Elsenbroich a Payette (2020) nebo ve stejném roce Gimenez a Novo zkoumali úspěch v rodinném podnikání s přihlédnutím na mikroekonomický rámec teorie her. V roce 2019 také doktor Stehel vydal knihu Využití teorie her

při řízení podniku. V ní se věnuje dvěma základním rozhodovacím problémům z hlediska evoluční teorie her, a to problému dodavatelsko-odběratelského řetězce a inovací a imitací.

4.3.4 FUZZY LOGIKA

Fuzzy logika je podobor matematické logiky. Svůj základ spatřuje v teorii fuzzy množin. Fuzzy logiku zavedl roku 1965 Lotfi Zadeh z univerzity v Berkley. Možným prapočátkem fuzzy logiky je paradox sóritás (paradox hromady), který vyslovil Eubúlidos z Milétu (žil v době Aristotelově). Tento paradox začíná na hromadě 1 milionu zrnek písku. Odebráním 1 zrnka hromada ještě nepřestane být hromadou, tedy 999 999 zrnek stále tvoří hromadu. Nicméně budeme-li proces odebrání 1 zrnka opakovat ještě 999 999krát, zbyde 0 zrnek písku, které ještě tvoří hromadu. Podobně je to často vysvětlováno v literaturách na paradoxu holohlavého: vypadne-li nám 1 vlas z hlavy, jsme již holohlaví, nebo ještě vlasatí, a co 2 vlasy, 3 vlasy atd.

Fuzzy logika se tedy opírá o neostré vlastnosti jevů, tak jak už vyplývá z překladu názvu fuzzy: neostrý, rozostřený, nepřesný, opilý…

Lotfi Zadeh, matematik z Ázerbájdžánu, pojmenoval původně svůj přístup jako modifikovanou teorii množin. Klasická teorie množin nám dokáže pouze rozhodnout, zda daný prvek do množiny patří či nikoli, pracuje pouze v množině $\{0, 1\}$. Fuzzy množiny ovšem mohou zahrnout celý interval, to znamená $\langle 0,1\rangle$. Dá se říct, že fuzzy množina je jakousi rozšířenou množinou, protože samozřejmě musí platit, že $\{0,1\} \subset \langle 0,1\rangle$. Fuzzy logika, fuzzy množiny a jejich využití v podnikové praxi jsou i dnes hojně aplikovaná témata vědeckého výzkumu (Zmeškal, 2014), (Wicher et al., 2019). Fuzzy logiku používá například Kataev (Kataev, 2020) k odhadu rizikových faktorů při propagaci nového softwarového produktu.

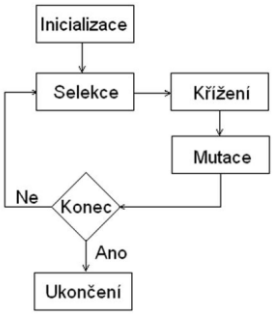
Využití fuzzy logiky v ekonomice je zřejmé. Fuzzy logika dnes proniká do běžného života člověka, můžeme se s ní setkat v řízení automatické pračky, v automatickém ostření fotoaparátu, fuzzy regulaci řízení brzdné soupravy (ABS u auta), a dokonce i v řízení výtahu při dobrzďování dle aktuální hmotnosti.

Na ekonomických fakultách se výzkumu Fuzzy logiky zabývají vědci z nejrůznějších ekonomických sfér. Například aplikací fuzzy-stochastických přístupů ve finančním rozhodování se zabýval Zmeškal v řadě publikací v rámci projektu GA402/02/1046 s rokem ukončení 2004. Ostravská univerzita v současné době má samostatný ústav pro výzkum a aplikace fuzzy modelování a zabývá se projekty jako „Metody fuzzy modelování v analýze forenzních digitálních dat“ nebo „Fuzzy parciální logika“.

4.3.5 GENETICKÉ ALGORITMY

Genetické algoritmy vycházejí z principů evoluční biologie. Techniky zde používané napodobují evoluční principy známé z biologie a jejích pojmů jako dědičnost, křížení, případně přirozený výběr. Genetické algoritmy jsou dnes stále předmětem intenzivního studia, tato poměrně mladá disciplína vznikla teprve v roce 1975 a poprvé je zavedl John Holland na Michigenské univerzitě. Základ pro svou teorii nalezl ve studiu buněčných automatů. Samozřejmě i on vycházel z Darwinovy teorie o vývoji druhů. Dnes genetické algoritmy slouží pro řešení optimalizačních problémů. Nasazují se třeba při řešení složitých problémů, pro které neexistuje použitelný exaktní algoritmus.

Genetický algoritmus je v literaturách popsán na procesech v biologii. První krok je vytvoření nulté generace, v biologii to jsou obvykle náhodně generovaní jedinci. Další je selekce, tento krok v biologii představuje výběr jedinců relevantních pro další množení. Následuje křížení, případně mutace, kde si vybraní rodiče vymění genetický kód a vzniká další generace. Nově vzniklá generace se pak vyhodnotí na základě parametrů určujících kvalitu jedince a následně může nová generace nahradit starou. Celý proces popisuje obrázek 4-7.



Obrázek 4-7 Proces genetického algoritmu
Zdroj: Vlastní zpracování dle (Ganguly, 2017)

Genetické algoritmy v podnikové ekonomice lze využít zejména pro optimalizační problémy. Je to zejména optimalizace výrobního plánu (Ganguly, 2017), výběr vhodné investice, zjišťování finanční situace podniku, umístění logistických center atd. Pro řízení organizačních změn je použil například Arazmjoo (Arazmjoo, 2019). Přímo pro prognózování se genetické algoritmy nepoužívají, ale vzhledem k tomu, že se často používají pro optimalizace fuzzy systémů či optimalizaci parametrů systémů s umělými neuronovými sítěmi, je zřejmé, že jejich využití bude aplikováno i do prognózování.

4.3.6 TEORIE CHAOSU

Slovo „chaos“ pochází z řecké mytologie a označuje nejstarší prvotní božstvo prázdnoty, propasti, temnoty. Aristofanus považuje Chaos za počáteční božstvo, které následně zplodilo ptáky. Podle jiných zdrojů byl první Chronos (čas) a Ananké (nutnost) a ti spolu zplodili Chaos. Básník Quintus označil za dceru boha Chaose bohyni Moiru (osud). Samozřejmě vědecká teorie chaosu z této starořecké mytologie nevychází, nicméně její analogie s vědeckou zkušeností je zajímavá.

Teorie chaosu vychází z fyzikální teorie. Za jejího zakladatele bývá označován Henri Poincaré, který v roce 1900 studoval problém ohybu tří objektů se vzájemnou gravitační silou. Teorii chaosu pak demonstroval E. Lorenz na objevu zvaném Podivný atraktor. Ten pomocí několika jednoduchých rovnic popisuje chování vodního kola, které má místo lopatek děravé nádoby, do nichž přitéká voda a kolem pohybuje. Lorenz očekával, že se kolo bude točit jedním směrem, anebo cyklicky měnit svůj směr, případně že se zastaví. Vodní kolo je však nestabilní a nečekaně mění svůj směr otáčení, který se nedá předpovídat. Tato data zadal Lorenz do počítače, který znázornil graf. Křivka v grafu se nikdy neprotíná a je nekonečná. Většího rozmachu tato teorie pak získala až s rozvojem výpočetní techniky.

Teorie chaosu se v jednoduchosti zabývá chováním nelineárních dynamických systémů, které mohou vykazovat deterministický chaos. V důsledku toho mohou fyzikální systémy vykazovat chaos a jevit se jako náhodné, i když model systému je dobře definovaný a neobsahuje žádné náhodné parametry. Tento systém se využívá k prognózování v atmosféře, tektonice zemských desek, turbulence tekutin. Ale lze ho aplikovat třeba i v psychoterapii, kdy malá změna může způsobit velkou proměnu. Tento efekt může podpořit, ale i zhroutit efektivitu práce v psychoterapii. Samozřejmě ho lze využít i k prognózám vývoje populace a s úspěchem i v ekonomice (Dostál, 2005).

4.4 METODY UMĚLÉ INTELIGENCE

Už od vzniku prvních počítačů se lidstvo snaží vytvořit algoritmus, který by napodoboval činnost lidského mozku. Výsledkem této snahy je pojem umělá inteligence. Pomineme-li rabína Löwa s jeho golemem i Karla Čapka s dramatem R.U.R. a také filozofickou podstatu problematiky myšlení strojů, kterými se zabývali například Pascal, Hobbes a Descartes, datujeme začátky umělé inteligence na počátek 50. let minulého století. S jejími začátky jsou spojena zejména jména jako Alan Turing, William Grey Walter, John McCarthy, ale také Marvin Lee Minsky či následně Alain Colmerauer, Frank Rosenblatt a řada dalších.

Dnes lze umělou inteligenci definovat jako obor informatiky zabývající se tvorbou strojů vykazujících známky inteligentního chování. Definice pojmu „inteligentní chování“ je však stále předmětem diskuze. Oblast umělé inteligence je velmi rozsáhlá a zároveň poměrně mladá vědní disciplína a nemusí zahrnovat pouze problematiku umělých neuronových sítí, ale také genetické algoritmy,

expertní systémy, teorie množin, suport vector regression, fuzzy metody časových řad, grey teorie (Peng et al., 2014) a v roce 2023 i vznik ChatGPT. Wang et al. (2018) do výčtu metod vkládají ještě teorii chaosu. Nelze tedy jednoznačně vyčlenit metody inspirované přírodou a metody umělé inteligence, protože se prolínají. To je samozřejmě logické, protože člověk je součástí přírody a lze ho jako přírodu i chápat.

Metody umělé inteligence hrají dnes již významnou roli v rozhodování, kde je možno využít metody vícekriteriálního rozhodování (Franek, 2016), a umělá inteligence může tvořit podporu AHP metod. Ovšem stále významněji stojí metody umělé inteligence v popředí modelování ekonomických procesů a podstatnou úlohu sehrávají i v oblasti prognózování. Jsou neustále zpřesňovány a doplňovány. Například Wang a Chen (2022) využívají adaptivní neuronové sítě k realizaci předpovědi poptávky v cestovním ruchu. Zároveň se zdokonalováním metod umělých neuronových sítí jsou vytvářeny i metody nové, například long short-term memory, které jsou považovány stále za nejmodernější a jsou nahrazovány stále přesnějšími modely (Anupam & Lawal, 2023). V této práci jsou použity metody umělé inteligence založené na umělých neuronových sítích dle Hyndmana (2021).

4.4.1 NEURONOVÉ SÍTĚ

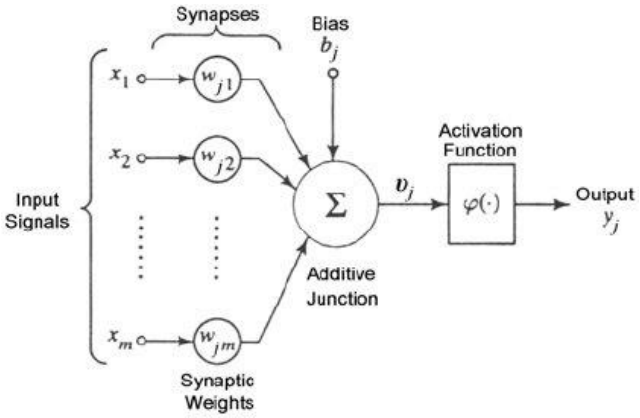
Princip umělé neuronové sítě je inspirován neuronovou sítí v lidském mozku. Její využití je všude tam, kde klasické statistické metody mohou poskytovat nepřesné výsledky. V některých případech prognóz dokonce není ani možné najít matematickou funkci, která by postihla všechny vlivy, jež variabilitu sledované proměnné ovlivňují.

Model neuronu, viz obrázek 4-8, s dopředným šířením signálů se skládá ze tří částí: synaptická spojení, axon a sóma, tzn. tělo buňky. Na základě vah mohou být jednotlivá synaptická spojení potlačena, nebo naopak zvýhodněna. Aktivační funkce zpracuje informace ze synaptických spojení a vygeneruje výstup. Výstup potom přivede výslednou informaci na vstup jiných neuronů.

Výstup neuronu je spočítán ve chvíli, když suma vstupů do neuronu x_i vynásobených jejich konkrétními vahami w_i překročí určitou hodnotu, kterou nazýváme bias. Neuron lze popsat tímto způsobem,

$$y_j = f\left(\sum_{i=1}^{m} x_i \cdot w_{j,i} - b_j\right), \text{ kde} \qquad (4.\ 3)$$

x_i je konkrétní hodnota na i-tém vstupu, $w_{j,i}$ je potom váha tohoto vstupu, b_j je prahová hodnota, m je celkový počet vstupů, f je transformační funkce a y hodnota na výstupu, vše dle logiky na obrázku 4-8.



Obrázek 4-8 Umělá neuronová síť
Zdroj: Haykin, 1994

Problémem prognózování na základě umělých neuronových sítí u tržeb ve službách dle členění NACE se zabývá i předchozí výzkum (Kolková, 2019). V tomto výzkumu byla použita data indexů cen ve službách v běžných cennách od roku 2005 do 2018, a to data měsíční. Na základě těchto dat byla provedena prognóza poptávky pomocí metod ETS, ARIMA, BATS a právě umělých neuronových sítí. Samozřejmě výsledky byly ohodnoceny měrami přesnosti. V rámci tohoto hodnocení se umělé neuronové sítě jevily jako metody nejpřesnější a nejlépe využitelné pro prognózu poptávky. Výzkum v roce 2020 (Kolková, 2020) se zaměřil na využití různých měr přesnosti a také na striktní dodržení rozdělení dat na tréninková a testovací. Zde již zcela jednoznačné výsledky nevyplynuly, i když stále nejvyužívanější metodou pro služby byly metody založené na umělých neuronových sítích.

Umělé neuronové sítě se pro prognózování poptávky již využívají poměrně hodně – viz například dále uvedené publikace (Swaminathan & Venkitasubramony, 2024; Kolková & Ključnikov, 2022; Kolková, 2019). V této monografii je použita funkce *Nnetar* z balíčku *forecast* dle Hyndmana et al. (2021). Jsou to dopředné neuronové sítě s jedinou nabízenou vrstvou a zpožděnými vstupy pro předpovídání jednorozměrných časových řad. Model je průměrem 20 samostatných neuronových sítí. Použití více sítí a jejich průměrování je technika zlepšující stabilitu a přesnost předpovědí. Model je automatizovaný a sám vyhledá nejlepší parametry.

Problematika umělých neuronových sítí je velmi obsáhlá a sama by vystačila na celou další monografii. Zde však byl použit jen základní model. Zpřesnění těchto sítí je ale stále předmětem mnoha vědeckých výzkumů.

4.5 KOMBINOVANÉ A HYBRIDNÍ MODELY

Je zřejmé, že v současné době rostoucí digitalizace a automatizace podnikání nabízí prognózování poptávky i jiné možnosti, než aplikace buď statistických metod, nebo metod založených na umělé inteligenci. Ostatní metody inspirované přírodou v současné praxi logistického prognózování poptávky nenašly své výraznější uplatnění, a tak zůstaly často jen v akademických studiích. Nicméně statistické metody a metody založené na umělé inteligenci nabízí možnosti aplikace nejen samostatně, ale i společně.

První z možností jsou kombinace různých statistických metod. Tyto kombinace mohou být prostým spojením několika metod se stejnou váhou. Případně může být váha určena některou ze sofistikovanějších metod, samozřejmě zase za účelem zpřesnění prognózy. Takto vzniklým modelům říkáme kombinované modely, tzn. modely kombinující výhradně různé statistické přístupy.

Další možností je kombinovat jak statistické metody, tak modely založené na umělé inteligenci do modelu jednoho. Obvykle se používají umělé neuronové sítě ve spojení s nejrůznějšími statistickými modely. Tento přístup už nazýváme hybridní model. Tyto modely tedy kombinují jak statistické přístupy, tak funkce strojového učení, respektive modely založené na umělé inteligenci.

Z výsledků M4 Competition (Makridakis et al., 2018), o kterých již bylo diskutováno v kapitole 1 této monografie, vyplývá, že kombinace několika prognostických metod přináší nejlepší výsledky. Ze 17 metod prezentovaných v této soutěži jako nejpřesnější bylo vyhodnoceno 12 kombinací zejména statistických modelů.

Největším překvapením M4 Competition však byl hybridní přístup, jak je uvedeno v tabulce 6-1. Hybridní model byl také absolutním vítězem této soutěže. Hybridní model v M4 Competition představil Smyl (Smyl, 2020), datový vědec ze společnosti Uber Technologies. Aplikoval hybridní model postavený na umělých neuronových sítích a statistické metodě inspirované exponenciálním vyrovnáním. Vzorce na bázi exponenciálního vyrovnání využil na desezónalizaci a normalizaci časové řady a pokročilé neuronové sítě využil k extrapolaci.

Druhou nejpřesnější metodou dle tabulky 4-5 byla v soutěži ohodnocena metoda, která kombinovala sedm statistických metod a jednu metodu strojového učení. Váhy zde byly počítány pomocí algoritmu strojového učení. Předložila ji španělská univerzita A Coruña a australská Monash University.

Tabulka 4-5 Výsledky M4 Competition

Typ modelu	Autor/autoři	Působiště autora	Pořadí
		Benchmark pro metody	**19**
Hybridní	Smyl, S.	Uber Technologies	**1**
Kombinace	Montero-Manso, P., Talagala, T., Hyndman, R. J. a Athanasopoulos, G.	University of A Coruña a Monash University	**2**
Kombinace	Pawlikowski, M., Chorowska, A. a Yanchuk, O.	ProLogistica Soft	**3**
Kombinace	Jaganathan, S. a Prakash, P.	Individual	**4**
Kombinace	Fiorucci, J. A. a Louzada, F.	University of Brasilia a University of São Paulo	**5**
Kombinace	Petropoulos, F. a Svetunkov, I.	University of Bath a Lancaster University	**6**
Kombinace	Shaub, D.	Harvard Extension School	**7**
Statistická	Legaki, N. Z. a Koutsouri, K.	National Technical University of Athens	**8**
Kombinace	Doornik, J., Castle, J. a Hendry, D.	University of Oxford	**9**
Kombinace	Pedregal, D.J., Trapero, J. R., Villegas, M. A. a Madrigal, J. J.	University of Castilla-La Mancha	**10**
Statistická	Spiliotis, E. a Assimakopoulos, V.	National Technical University of Athens	**11**
Kombinace	Roubinchtein, A.	Washington State Employment Security Department	**12**
Jiná	Ibrahim, M.	Georgia Institute of Technology	**13**
Kombinace	Kull, M., a kolektiv	University of Tartu	**14**
Kombinace	Waheeb, W.	Universiti Tun Hussein Onn Malaysia	**15**
Statistická	Darin, S. a Stellwagen, E.	Business Forecast Systems (Forecast Pro)	**16**
Kombinace	Dantas, T. a Oliveira, F.	Pontifical Catholic University of Rio de Janeiro	**17**
Nejlepší z metod ML	Trotta, B.	Individual	**23**
2 nejlepší z metod ML	Bontempi, G.	Université Libre de Bruxelles	**37**

Zdroj: Makridakis et al., 2018

Zajímavým zjištěním bylo, že z šesti metod, které byly založeny jen na strojovém učení, všechny fungovaly špatně. Dokonce jen jedna překonala benchmark, kterým byla metoda naivní. Z tohoto jasně vyplývá, že přesnost metod založených výhradně na strojovém učení je nízká a jejich využití v praxi tak není opodstatněné. Na druhou stranu ani přesnost jednotlivých samostatných metod statistických není pro praktické využití zajímavá.

První dvě nejúspěšnější metody tvoří hybridní metoda od autora Smyla, nazvaná ES-RNN, a kombinovaná metoda autorů Montero-Manso, P., Talagala, T. S., Hyndman, R. J. a Athanasopoulos, G., kterou autoři nazvali FFORMA. Jediná z volně dostupných metod v jazyce R však je v současné době pouze kombinovaná metoda Davida Shauba, která dosáhla celkově 7. místa, na ročních datech ovšem byla její úspěšnost vyšší a dosáhla třetího místa. Shaub ji dal k dispozici v balíčku *forecastHybrid*. V aplikační kapitole proto bude aplikována z hybridních a kombinovaných metod právě ona.

4.5.1 HYBRIDNÍ METODA ES-RNN

V polovině roku 2016 Smyl se svými spolupracovníky úspěšně vytvořili rozšíření a zobecnění modelů Holt a Holt-Winters (Smyl & Zhang, 2015, a Smyl et al., 2019). Následně pak experimentoval společně s dalšími výzkumníky v používání tohoto modelu společně s modely založenými na umělých neuronových sítích (Smyl & Kuber, 2016). Tyto experimenty využil v M4 Competition a jeho model překonal na ročních datech modely ostatní, nicméně nikoli na datech měsíčních.

Smyl pracuje pro společnost Uber, jejíž podnikání na přesných prognózách stojí. Jak už bylo uvedeno v úvodní kapitole, předpovídají například nabídky řidičů a požadavky zákazníků, a to ve více než 600 městech. Z toho důvodu Uber masivně investuje do zvyšování znalostí svých prognostiků a vytváří špičkové týmy v oblasti prognózování poptávky. Vysoké odborné znalosti pak vedly až k vítězství v M4 Competition tohoto předního vědce v oblasti dat.

Model využívá dynamický výpočetní grafový systém neuronové sítě, který umožňuje spojení standardního exponenciálního vyhlazovacího modelu s pokročilými sítěmi s krátkodobou pamětí do společného rámce. Výsledkem je hybridní a hierarchická metoda prognózování. Vyznačuje se dvěma zajímavými vlastnostmi, a to zejména že využívá statistické modely exponential smoothing (přesněji Holtův a Holt-Wintersův model s multiplikativní sezónností) v kombinaci s LSTM sítěmi a druhá zvláštnost je použití hierarchické povahy.

Metoda generovala přesné předpovědi pro většinu časových řad, zejména však měsíční, čtvrtletní a roční. Přesnost však nebyla nejlepší pro případy denních a týdenních údajů. I přesto výsledná přesnost překonala všechny ostatní autory v M4 Competition. Nicméně autor i po ukončení soutěže pracoval na vylepšení modelu (Bandara et al., 2020, nebo Dudek et al., 2022).

V tomto hybridním modelu, jak již bylo řečeno, kombinuje Smyl vzorce inspirované exponenciálním vyhlazováním, používané k desezónalizaci a normalizaci série, s pokročilými neuronovými sítěmi, využívanými k extrapolaci

řady. V této metodě jsou aplikovány ručně kódované části, kterými jsou vzorce exponenciálního vyrovnání s rekurentními neuronovými sítěmi (odtud označení ES-RNN model). Přesněji řečeno, šlo o hybridní model založený na sloučení Holt-Wintersova modelu a RNN. Část modelu je založena na vzorcích,

$$l_t = \alpha(y_t/S_t) + (1-\alpha)l_{t-1} \tag{4.4}$$

$$S_{t+m} = \gamma(y_t/l_t) + (1-\gamma)S_t \tag{4.5}$$

Výsledná prognóza je pak založena na vzorci,

$$y_{t+1\ldots t+h} = \widehat{RNN(X_t)} \cdot l_t \cdot S_{t+1\ldots t+h}, \text{ kde} \tag{4.6}$$

$$X_t = \frac{y_t}{l_t S_t}. \tag{4.7}$$

Pro své výpočty používá Smyl pravděpodobností programovací jazyk Stan napsaný v C++. Stan funguje pod licencí Nová licence BSD a jeho první vydání je datováno do roku 2012. Jazyk umožňuje integraci s R softwarovým prostředím, programovacím jazykem Python, MATLAB numerickým výpočetním prostředím, programovacím jazykem Julia i integraci se Stata.

V rámci M4 Competition byly používány pouze čistě časové řady bez dalších prvků, které ale v podnikové praxi mohou výrazně ovlivnit výkon modelu. Těmito prvky mohou být jak události typu sportovních, politických či společenských akcí, tak třeba informace o počasí, geografické poloze a podobně, vše dle individuálních aktivit podniků a předmětu podnikání. Hybridní model ES-RNN bude jistě možno těmto atributům přizpůsobit.

Uber v současné době plánuje komplexní zveřejnění svého modelu zahrnutím do knihovny *PyroForecastingLibrary*, a tím zpřístupnění všem uživatelům *Python*. V době realizace této monografie to však ještě k dispozici není, proto výpočet tohoto hybridního modelu v aplikaci není zahrnut.

4.5.2 KOMBINOVANÁ METODA FFORMA

Tato metoda nazvaná FFORMA (Feautre-Based Forecast Model Selection) vychází z meta-learningového přístupu (Montero-Manso et al., 2020). Podstatou je snaha odvodit sadu vah pro kombinaci předpovědí ze souboru metod, které zahrnují naivní metodu, metodu náhodné procházky s trendem, sezónní naivní metodu, Theta metodu, metodu exponenciálního vyrovnání, ARIMA modely, model TBATS, model STLM-AR a metody na základě umělých neuronových sítí.

Všechny tyto metody jsou v balíčku *forecast*, což je asi nejpoužívanější balíček na poli prognózování poptávky v podnikové ekonomice. V této publikaci jsou některé metody použity zvlášť. V balíčku *forecast* využil Montero-Manso tyto příkazy pro jednotlivé metody,

- naivní (*naive*),
- náhodná procházka s driftem (*rwf* s trendem = TRUE),
- sezónní naivní (*snaive*),
- theta metoda (*thetaf*),

- automatizovaný algoritmus ARIMA (*auto.arima*),
- automatizovaný algoritmus exponenciálního vyhlazování (*ets*),
- model TBATS (*tbats*),
- STLM-AR sezónní a trendový rozklad pomocí spraše s AR modelováním sezónně očištěné řady (*stlm* s funkcí modelu *ar*),
- předpovědi časových řad neuronových sítí (*Nnetar*).

Ve všech případech bylo využito výchozí nastavení modelu, jak jej nabízí balíček *forecast*. Jako v téměř každé metodě jsou i zde časové řady rozděleny na tréninkové období a testovací období. Sada funkcí časových řad se počítá z tréninkového období (např. délka časových řad, síla trendu, autokorelace), tyto vstupy tvoří model meta-učení. Každá metoda je přizpůsobena tréninkovému období, prognózy jsou generovány během testovacího období a na tomto období se také počítají chyby předpovědi. Potom z nich lze vypočítat souhrnnou předpověď ztráty z vážené kombinované prognózy pro libovolnou danou sadu vah.

Meta-learningový model se naučí vytvářet váhy pro všechny metody, a to na základě minimalizace míry chyb v prognóze. Výsledná kombinace metod je pak vlastně nalezení funkce, která přiřazuje váhy každé předpovědní metodě s cílem minimalizovat očekávanou ztrátu, která by byla vytvořena, kdyby byly metody vybrány náhodně.

4.5.3 KOMBINOVANÉ MODELY DLE *FORECASTHYBRID*

Hybridní model dle Shauba a Ellise (Shaub & Ellis, 2018), publikovaný v International Journal of Forecasting v roce 2020 (Shaub, 2020), je ve svém základu poměrně jednoduchý. Jedná se o model obsahující tři statistické metody a metodu umělých neuronových sítí v přesně daném poměru (respektive za použití stejných vah). Může se zdát, že tato metoda je velmi jednoduchá a ve věku velkých dat i zastaralá, nicméně výsledky, které poskytuje, tomu neodpovídají.

Zvláště pro roční časové řady je velmi přesná. Pro tyto časové řady v M4 Competition dosáhla dokonce třetího místa. Její hlavní výhoda ovšem spočívá v její nenáročnosti a jednoduchosti na implementaci v podniku. Autor ji volně zpřístupnil v balíčku nazvaném *forecastHybrid* v jazyce R. Tento model pro jeho volné využití bude představen i v kapitole 8.

V této kombinaci byly využity metody *auto.arima*, které představují automatický výběr nejvhodnějších ARIMA modelů. Další použitá metoda byla ETS, model THETA a TBATS, všechny byly diskutovány v kapitole 4.2. Poslední metoda použitá v hybridním modelu dle Shauba byly umělé neuronové sítě, diskutované v kapitole 6.

Všechny komponenty hybridního modelu pocházejí z balíčku *forecast* (Hyndman & Khandakar, 2008). Výsledná pro *forecastHybrid* je pak dána vztahem,

$$f(i) = \sum_{m=1}^{n} w_m \cdot f_m(i), \text{ kde} \tag{4.8}$$

w jsou váhy přiřazené jednotlivé metodě výpočtu, f_m je výsledek dané metody. Váhy je možno určit mnoha způsoby, například subjektivně přiřadit daným uživatelem nebo nejvhodnější je určit pomocí iteračních postupů. V tomto případě pomocí iterací hledáme nejvhodnější kombinace vah, které zajistí nejlepší přesnost. Nicméně to už je výpočetně, časově i znalostně náročnější metoda. Dle zkušeností z M4 Competition dosahovala překvapivě konkurenceschopných výsledků i metoda, kdy váhy byly rozděleny rovnoměrně. Jak již bylo řečeno, hlavní výhodou modelu je jeho jednoduchost a nenáročnost na čas a výpočetní výkon.

Stejně tak nejsou nutné tak velké programátorské znalosti, jako to bývá u jiných hybridních modelů. Například v soutěži M4 Competition, kde se testovalo najednou 23 tisíc ročních časových řad, autorův kód pro jazyk R nepřesáhl 100 řádků. Takový kód pak byl navíc snadno uživatelsky laditelný. Tento přístup by mohl být pro podniky zajímavý právě pro svou rychlost a snadnou variabilnost i pro velká data.

4.6 METODY VYVINUTÉ V PRAXI

V současnosti již je zřejmé, že vývoj nových kvantitativních metod prognózování nebude jen záležitostí akademické komunity, ale stále častěji budou nositeli hlavních inovativních myšlenek odborníci a vědecko-výzkumní pracovníci z praxe. Důkazem toho jsou i poslední M-Competition, kde úspěchy slaví hlavně tito odborníci, ačkoli jsou stále ještě v menšině. Dá se ale předpokládat, že do budoucna bude jejich počet, ale i důležitost vzrůstat.

Již byla popsána jedna z nejúspěšnějších metod vyvinutých v praxi odborníka z firmy Uber (Smyl, 2020). Další metodu představila Prologistica soft, jejíž koncept se opírá o kombinaci statistických metod. Statistické modely jsou vážené dle své výkonnosti. Autoři zde provádí ruční výběr modelů, které mají nejlepší výkon pro každý typ časové řady (Pawlikowski & Chorowska, 2020). Metody, které zvolili, jsou Naive, Naive2, jednoduché exponenciální vyrovnání, exponenciální vyrovnání s automatickou volbou parametrů, opět automatizované damped exponenciální vyrovnání, automatický výběr modelů ARIMA, Theta, který modifikovali na optimalizovaný model Theta a ekonometrické modely (lineární regrese), které byly nastaveny individuálně.

Srihari Jaganathan (Jaganathan & Prakash, 2020) představil svůj model jako individuální. Nicméně vzhledem k tomu, že v kombinovaném modelu použil i komerční softwarový balík Forecast Pro, lze ho jistě zahrnut do modelů vytvořených v praxi. Další modely, které byly využity v této kombinaci, jsou Naïve/snaïve, exponenciální vyrovnání (ETS), dampened ETS, bagged ETS, exponenciální vyrovnání (ES)/komplexní exponenciální vyrovnání (CES) a generalizované exponenciální vyrovnání (GES), model MAPA (Multi-aggregation predikční algoritmus), model THIEF (Temporal Hierarchical Forecasting), ARIMA, Theta, Hybrid Theta, model STL (Seasonal and Trend Decomposition Using Loses Forecast), TBATS, DSHW (Double Seasonal Holt-

Winters), MLP (Multilayer Perceptron)/ ELM (Extreme Learning Machines). Model opět vybírali i subjektivně a primárními faktory v posouzení výběru byly:

- vhodnost metod pro danou frekvenci a předchozí důkazy o jejich výkonu v předchozích soutěžích a srovnávacích testech,
- dostupnost softwaru,
- jejich výkony v předpovědích vzorku v datovém souboru M4 Competition.

Samozřejmě v praxi existuje i řada modelů, které se nezúčastnily soutěží M-Competition. Jedním z dnes velmi používaných modelů je Prophet, který byl vytvořen odborníky z Facebook (Taylor & Letham, 2018). Autoři upozorňují na problémy s vytvářením kvalitních předpovědí, a to velkou spoustu časových řad s různými parametry. A zejména upozorňují na to, že podnikoví odborníci s vysokými odbornými znalostmi prediktivní analytiky jsou dosud poměrně vzácností. Proto jejich hlavní výzvou bylo vytvořit model, který mohou intuitivně upravovat i analytici se základními znalostmi časových řad. O úspěšnosti modelu svědčí i množství citací na Web of Science. Tento model byl mnohokrát aplikován i na data, která byla zcela odlišná od jeho původního použití. Pro prognózování poptávky byl využit v několika publikacích (Kolková & Navrátil, 2021), (Kolková & Ključnikov, 2022) nebo byl použit pro prognózy snižování energie a tvorby elektronického odpadu při těžbě bitcoinů (Jana et al., 2022), pro predikci produkce ropy (Ning et al., 2022). Také další autoři napsali zajímavé studie ověřující schopnosti modelu Prophet (Wang et al., 2022; Ning et al., 2022; nebo Basak et al., 2022).

Prophet je publikovaný v režimu „open source library“ v jazyce R i Python. To je zřejmě hlavní důvod jeho velkého rozšíření mezi akademiky i ostatními výzkumnými pracovníky. Umožnilo to model lépe ověřit na velké skupině dat, a to velmi různorodých jak oborově, tak i místně. Poskytuje možnost provádět predikce s výraznou přesností. Jeho hlavní předností je však možnost zahrnout do výpočtů prognózy poptávky také prázdniny, státní svátky a jiné události, které na poptávku mohou mít významný vliv. Prophet pracuje se třemi hlavními komponenty, a to trend, sezónalita, a právě již zmiňované prázdniny:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t, \text{ kde} \tag{4.9}$$

g(t) je trendová funkce, *s(t)* jsou cyklické změny poptávky, *h(t)* je prázdninový efekt a ϵ_t vyjadřuje chybu, kterou model nedokázal popsat. Prophet tak vyhovuje individuálním požadavkům každého podniku. Do moderních prognostických modelů dnes již podniky vyžadují vložení i státních svátků, případně jakýchkoliv jiných významných dnů pro upřesnění predikce. Především se pak nabízí nutnost implementace období kolem Vánoc, případně i událostí, které se pojí s daným podnikáním, jako jsou světové sportovní akce (mistrovství, olympiáda) nebo regionální akce typu koncertů, výstav apod. To může být pro využití v podnikové sféře velmi zajímavé. Každý podnik je ovlivněn jinými svátky a jinými událostmi a jejich selekce může významně zjednodušit prognózování.

Jsou samozřejmě firmy, které svůj software nezveřejňují, ale využívají ho komerčně. Mezi ně patří produkty, které jsou implementovatelné do SAP, čímž mohou usnadnit uživatelům jejich obsluhu. Mezi ně patří Streamline (je k zakoupení na webu: https://gmdhsoftware.com/cs/) a také doplňky v SAP. Dalším komerčním produktem je IBM®Cognos Analytics sloužící ke zjišťování a modelování trendů, sezónnosti a časových závislostí v datech. V produktu IBM®Cognos Analytics, tak jako v ostatních komerčních programech, lze prognózovat pomocí automatizovaných nástrojů, které modelují data závislá na čase. Automatizovaný výběr modelu a vyladění usnadňuje použití prognóz. Tyto produkty tak umožňují aplikaci i bez znalosti modelování časových řad. V produktu IBM®Cognos Analytics je automatizována i příprava dat. Pokud systém rozpozná vložená data jako časovou řadu, celý proces přípravy dat i jejich dekompozice je automatizuje. Pro prognózu pak využívá jen modely exponencitálního vyrovnání dle IBM®Cognos Analytics (2021).

Jeden z velice známých komerčních produktů je Forecast Pro (2022). Tento produkt už nabízí větší škálu použitých metod, a to včetně strojového učení i dynamické regrese. Společnost Forecast pro však ke koupi nabízí tři typy programů, které se mohou v použití modelů lišit. Nejdražší program je v současné době (počátek roku 2023) oceněn částkou 9 995 USD za roční licenci, údržbu i podporu. Následné roky už pak představují nižší částky. Placený model Forecast pro byl využit v některých časových řadách i ve studii Jaganathana a Prakashe (Jaganathan & Prakash, 2020).

Samozřejmě ani firma Microsoft nezůstala v tvorbě prognostických nástrojů pozadu a nabízí produkt Microsoft Azure Machine Learning s nástrojem Stream Analytic (docs.microsoft.com, 2022; Intruction demand forecasting, 2022). Microsoft (pomocí Event Hubs) umožňuje shromažďování dat v reálném čase, následně Stream Analytic data agreguje a upravuje, pomocí strojového učení se následně vybírá a spouští prognostický model. Ke konečné vizualizaci výsledků je používán další Microsoft produkt – Power BI. Microsoft umožňuje nejen vygenerovat statistickou základní předpověď, která je založena na historických datech, ale následně také použije dynamickou sadu dimenzí prognózy. Samozřejmostí je pak vizualizace trendů poptávky, intervalů spolehlivosti a úpravy prognózy. Také je možno použít upravené prognózy v dalších procesech plánování, které jsou součástí jiných produktů Microsoft. Samozřejmostí je odstranění odlehlých hodnot a měření přesnosti modelů. Další, méně známé produkty jsou například Oracle and JD Edvards nebo Epicor.

4.7 METODY POUŽITÉ V TÉTO MONOGRAFII

V této monografii budou ze skupiny statistických metod využity *naive, snaive, ses, auto.arima, ETS*. Metoda naivní (*naive*) prognózuje následující hodnoty jen jako poslední sledované. Je tedy pro praktické využití logicky poměrně nedostatečná. Obvykle slouží jen jako metoda srovnávací. Porovnáme tyto metody s ostatními, a pokud ostatní v přesnosti nepřekonají metodu *naive*, jsou naprosto nevhodné. Protože data vykazují i sezónnost, je použita i metoda *snaive*. Zde je predikovaná

hodnota sledované veličiny pro každý bod v budoucnosti stanovena jako poslední známá hodnota ve stejné sezóně. Obě metody v monografii slouží k porovnání s ostatními.

Další ze skupiny metod exponenciálního vyrovnání je *ses* (Simple Exponential Smoothing). Tato metoda nevyužívá žádný trend ani sezónnost a je v této monografii opět jen doplňkem k ostatním, kdy se nepředpokládá, že by mohla být nejvhodnější. Ale je to metoda, která je stále v praxi využívána a může být označena jako nejvhodnější.

Metody ETS jsou v této monografii popsány v 4.2.1 dle vzorců definovaných v tabulce 4-3. Pro výpočet je využit jazyk R, a to dle balíčku *forecast* (Hyndman & Khandakar, 2008). Modely ARIMA jsou definovány v kapitole 4.2.2. V této monografii je využita automatická funkce *auto.arima* (Hyndman & Khandakar, 2008). Postup automatického výpočtu v této funkci je definován na obrázku 4-6.

Další z použitých metod je TBATS, která je zkratkou klíčových slov z modelů Box-Cox transformace, ARMA model, trend a sezónní komponenty. Tato metoda je blíže popsána v kapitole 4.2.3. V této monografii je využit výpočet dle De Livery et al. (et al.2011).

Modely založené na umělých neuronových sítích jsou definovány v kapitole 4.4.1. V této monografii je použita funkce *Nnetar* z balíčku *forecast* (Hyndman et al., 2021). Jsou to dopředné neuronové sítě s jedinou nabízenou vrstvou a zpožděnými vstupy pro předpovídání jednorozměrných časových řad. Model je průměrem 20 samostatných neuronových sítí. Použití více sítí a jejich průměrování je technika zlepšující stabilitu a přesnost předpovědí. Model je automatizovaný a sám vyhledá nejlepší parametry. Modely *Thetam* a *forecastHybrid* jsou definovány v kapitole 4.5.3 a jsou vyčísleny v jazyce R s využitím balíčku *forecastHybrid* dle (Shaub, 2020).

V monografii bude nejprve vyčíslena prognóza a následně zhodnocena dle měr přesnosti na datech tréninkových. Spolehlivost modelu je pak ověřena na datech testovacích. Vše v souladu s postupem uvedeným na obrázku 2-6.

Kapitola 5

Míry přesnosti prognóz

V případě, že je možno prognózovat hodnoty více metodami, je vhodné zvolit tu, která poskytuje nejmenší chyby, případně volit i dle jiných kritérií. Chybovost se hodnotí v období známých hodnot, tedy v období ex-post prognózy. Pro vyhodnocení pak platí, že v případě, že zvolená veličina hodnotící chybovost je 0, pak je prognóza bezchybná. Některé míry chybovosti v případě kladné chyby označují model jako podhodnocující skutečnost, a naopak v případě záporné chyby model skutečnost nadhodnocuje.

V souvislosti s přesností metod je vedena nejživější diskuze o tom, zda jsou přesnější metody založené na umělé inteligenci, nebo statistické modely (Kolkova, 2020). Spiliotis et al. (2020) uvádějí, že některé metody strojového učení poskytují lepší předpovědi, a to i z hlediska přesnosti. K podobným závěrům dospěl i Cuhadar (Cuhadar, 2020), který při prognózování cestovního ruchu v Chorvatsku dospěl k názoru, že metody založené na umělých neuronových sítích pracují vždy přesněji ve srovnání se statistickými metodami. Stejně i Pereira a Cerqueira (Pereira & Cerqueira, 2021) při prognózování poptávky po hotelových službách dospěli k názoru, že použití modelů strojového učení může ve srovnání s tradičními metodami exponenciálního vyrovnání snížit střední kvadratickou chybu až o 54 % pro jednodenní horizont prognózy a až o 45 % pro 14denní horizont prognózy. V jiné studii (Balaji Prabhu & Dakshayini, 2020) autoři docházejí k závěru, že model vícenásobné lineární regrese a model umělé neuronové sítě jsou oba užitečným, spolehlivým a poměrně účinným nástrojem pro optimalizaci účinků predikce poptávky při řízení nabídky sklizně potravin tak, aby uspokojivě odpovídala společenským potřebám. Další studie (Abbasimehr et al., 2020) porovnávala metody umělé inteligence se statistickými. Metoda vícevrstvé sítě LSTM byla nejpřesnější, druhou nejpřesnější metodou byly umělé neuronové sítě, nicméně už na třetím místě byla označena jako nejpřesnější metoda ETS.

Na druhou stranu se autoři zabývají otázkou, zda rozdíly mezi přesností metod založených na umělé inteligenci a statistických metod nesouvisí se zkoumanou časovou řadou. Eikeland et al. (2021) docházejí ve své studii k závěru, že statistické modely dosahují vyšší přesnosti na delších předpovědních

horizontech ve srovnání s neuronovými sítěmi, zatímco přístupy strojového učení fungují lépe v predikci v kratších časových intervalech. Naopak Marček (Marček, 2019) ukázal, že všechny modely, které použil (ARIMA, NN, SVM), jsou vhodné jako predikční metody pro použití v prognostických systémech, které běžně předpovídají hodnoty proměnných na konkurenčních energetických trzích. Jiný článek (Bui et al., 2020) doporučuje nové metody filtrování dat, aby se zvýšila míra přesnosti modelů založených na umělé neuronové síti. Zejména v datech rozvojových zemí může docházet k náhlým změnám poptávky. Z výsledků je patrné, že přesnost uvedených modelů může být významně zlepšena díky navržené metodě statistického filtrování dat.

5.1 VÝPOČTY MĚR PŘESNOSTI

Základní metodou měření míry přesnosti je stále R^2. Obvykle ho nazýváme koeficient determinace. Je vyjádřen vztahem,

$$R^2 = \frac{\sum_{p=1}^{N}(y_p - \widehat{y_p})^2}{\sum_{p=1}^{N}(y_p - \bar{y})^2}, \text{ nebo} \tag{5.1}$$

$$R^2 = 1 - \frac{\sum_{p=1}^{N} {e_p}^2}{\sum_{p=1}^{N}(y_p - \bar{y})^2}, \text{ kde} \tag{5.2}$$

N je počet hodnot, y zvolená závisle proměnná, $\bar{y}$ střední hodnota závisle proměnné a $\hat{y}$ je regresní odhad $i$tého pozorování.

Statistika R^2 je ovšem obecně považována za poměrně špatnou míru prediktivních schopností. Armstrong (2001) upozornil na šest studií o použití R^2 a nalezl relevantně malý vztah s přesností prognózy. Proto jsou zavedeny i jiné statistiky, které jsou svou jednoduchostí, užitečností, ale i snadnější pochopitelností užívanější než R^2. Novější metody jsou diskutovány v článku Chen et al. (2017).

Metody měření přesnosti lze rozdělit do mnoha skupin. Tyto skupiny se obvykle neuvádějí v českém jazyce a odborníci je často definují pouze anglickými názvy. V této publikaci tedy budou také pro členění použity anglické výrazy. Hyndman a Athanasopoulos (2018) člení metody výpočtu měr přesnosti dle anglických označení na Scale-Dependent Measures, Measures Based on Percentage Errors, Scaled Errors, ve článku Hyndmana a Koehlera (2006) je přesnost počítána i pomocí dalších skupin, kterými jsou Measures Based on Relative Errors a Relative Measures.

Mezi Scale-Dependent Errors můžeme zařadit ukazatele Mean Error (dále jen ME), Mean Scaled Error (dále MSE), Root Mean Squared Error (dále RMSE) a Mean Absolute Error (dále MAE). Tyto míry přesnosti jsou závislé na měřítku použitých dat a neměly by být používány v porovnání souborů dat s různými měřítky. Lze je vyčíslit následujícími způsoby:

$$MSE = \frac{1}{M}\sum_{p=n+1}^{N}\left(y_p - \widehat{y_p}\right)^2 \tag{5.3}$$

Vhodnější ovšem je použít standardní odchylku, kterou je RMSE dle vztahu

$$RMSE = \sqrt{MSE} \tag{5.4}$$

MSE i RMSE mají stejnou jednotku jako původní časová řada. Ve stejných jednotkách je také deklarován ukazatel MAE, který vyjadřuje průměrnou odchylku skutečných hodnot od prognózovaných. Místo průměrné odchylky je také možno použít medián. Oba výpočty pak popíšeme vztahy

$$MAE = \frac{1}{M}\sum_{p=n+1}^{N}\left|y_p - \widehat{y_p}\right| \text{ nebo } MAE = mean|e_t| \tag{5.5}$$

$$MdAE = median\left(\left|y_p - \widehat{y_p}\right|\right) \tag{5.6}$$

Parametry měr přesnosti založené na tzv. percentage errors jsou již na měřítku nezávislé, což může být jejich velkou výhodou a jejich nejčastější použití je právě k porovnávání měr přesnosti napříč různými datovými sadami. Mezi nejčastější ukazatele tohoto typu patří Mean absolute percentage error (dále jen MAPE):

$$MAPE = \frac{1}{M}\sum_{p=n+1}^{N}\frac{\left|y_p - \widehat{y_p}\right|^2}{y_p} \tag{5.7}$$

V literatuře se často také setkáváme s druhou možností výpočtu, a to se zjednodušením vzorců použitím tohoto ukazatele:

$$p_t = \frac{100e_t}{y_t} \tag{5.8}$$

pak můžeme vzorec například MAPE zjednodušit na zápis

$$MAPE = mean(|p_t|). \tag{5.9}$$

5.2 VÝBĚR METOD MĚŘENÍ PŘESNOSTI

Předcházející ukazatele měřící míry přesnosti jsou běžně používané a ve vědeckých článcích často aplikované. Jednoznačná odpověď, která z uvedených měr přesnosti je nejlepší a nejpřesnější, neexistuje. Odborná veřejnost stále vede vědeckou rozpravu na toto téma. Nicméně předchozí výzkumy ukazují, že využití více různých měr přesností příliš nevede k přijetí jiných modelů.

Mimo jiné se tím zabývá i výzkum (Kolková, 2020), kde jsou na příkladu prodejů v oblasti služeb aplikovány předchozí míry přesnosti, a to pouze s malými rozdíly ve výsledné intepretaci. Míry přesnosti byly aplikovány na 32 typů služeb v členění dle NACE a na modely bylo aplikováno 6 měr přesnosti (ME, RMSE, MAE, MPE, MAPE MASE). Byly vyčísleny prognostické modely na bázi exponenciálního průměru ARIMA, BATS a modely založené na umělých neuronových sítích. Z 32 typů služeb pouze u dvou došlo k situaci, že různé míry přesnosti vedou k přijetí různých výsledků. V tabulce 5-1 jsou uvedeny výsledky tohoto výzkumu pro dvě služby, u kterých vyšly různé metody dle různých měr přesnosti. Pro provozování gastronomických služeb pomocí ukazatele MPE byly nejvhodnější metody ETS, ostatní ukazatele jako nejpřesnější metodu určily umělé

neuronové sítě. V případě realitní činnosti vyšla nejpřesněji metoda umělých neuronových sítí dle všech kritérií. Zajímavé však bylo, že když bychom umělé neuronové sítě z rozhodování vyloučili, podle dvou kritérií by byla vybrána metoda BATS a dle ostatních ARIMA. Ve všech zbývajících 30 typech služby k žádným disproporcím dle zvoleného kritéria míry přesnosti k rozdílům nedocházelo.

Tabulka 5-1 Volba modelu u vybraných typů služeb na základě různých kritérií měr přesnosti

	ME	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**
Provozování gastronomických služeb	**-0,0056**	**0,6856**	**0,4469**	**0,0099**	**0,4585**	**0,0737**
	Nnetar	Nnetar	Nnetar	ETS	Nnetar	Nnetar
Realitní činnost	**0,1271**	**9,8630**	**7,0229**	**-0,3179**	**6,4335**	**0,3474**
	ARIMA*	BATS*	ARIMA*	BATS*	ARIMA*	ARIMA*
	Nnetar	Nnetar	Nnetar	Nnetar	Nnetar	Nnetar

** nejlepší míra přesnosti po vyloučení modelu umělých neuronových sítí*

Zdroj: Kolková, 2020

Z tohoto článku lze tedy vyvodit závěr, že existuje pouze malé procento prognóz, jejichž výsledek je ovlivněn výběrem kritéria měr přesnosti. U drtivé většiny prognóz je zcela irelevantní, jaká míra přesnosti se využije, většinou vedou všechny k přijetí stejného nejpřesnějšího modelu. To bude ověřeno i na prognóze individuální poptávky a pak následně aplikováno na ostatní prognózy.

5.3 ALTERNATIVNÍ MÍRY PŘESNOSTI

Jelikož tedy výzkumníci stále hledají to nejlepší, co by obsáhlo všechny parametry a umožnilo srovnání všech modelů, jsou dále vyvíjeny i další míry přesnosti. Mohou jimi být

$$Median\ Absolute\ Percentage\ Error(\ MdAPE) = median(|p_t|),\quad (5.10)$$

$$Root\ Mean\ Square\ Percentage\ Error\ (RMSPE) = \sqrt{mean(p_t^2)},\quad (5.11)$$

$$Root\ Median\ Square\ Percentage\ Error\ (\ RMdSPE) = \sqrt{median(p_t^2)}.\quad (5.12)$$

Nevýhodou těchto modelů ovšem může být fakt, že pozitivní chyby mají větší váhu než negativní. Toto slabé místo měr vedlo k použití symetrických měr přesností, které pak byly i pojmenovány jako symetrické:

$$Symmetric\ Mean\ Absolute\ Percentage\ Error\ (sMAPE) = mean(200|Y\_t - F\_t\ |/(Y_t + F_t)\quad (5.13)$$

$$Symmetric\ Median\ Absolute\ Percentage\ Error\ (sMdAPE) = median(200|Y\_t - F\_t\ |/(Y_t + F_t) \quad (5.14)$$

Scaled errors jsou založeny na MAE. Metodou, kterou lze do této skupiny zařadit, je mean absolute scaled error (dále jen MASE). Tyto metody navrhli v článku Hyndman a Koehler (2006) jako obecné měřítko přesnosti prognóz.

$$MASE = mean(|q_t|), \text{ kde} \quad (5.15)$$

$$q_j = \frac{e_j}{\frac{1}{T-1}\sum_{t=2}^{T}|y_t - y_{t-1}|} \quad (5.16)$$

Měření založené na relativních chybách je jakýsi alternativní způsob měření chybovosti. Spočívá v tom, že každou chybu dělí chybou získanou pomocí jiné standardní metody měření. Dle Hyndmana a Koehlera (2006) lze využít parametry

$$Mean\ Relative\ Absolute\ Error\ (MRAE) = mean(|r_t|) \quad (5.17)$$

$$Median\ Relative\ Abolute\ Error\ (MdRAE) = median(|r_t|) \quad (5.18)$$

$$Geometric\ Mean\ Relative\ Absolute\ Error\ (GMRAE) = gmean(|r_t|), \text{ kde} \quad (5.19)$$

$$r_t = \frac{e_t}{e_t^*} \quad (5.20)$$

e_t^* je pak predikční chyba získaná metodou, která představuje benchmark.

Další možností, jak měřit míry přesnosti založené na relativních chybách, je hodnota MaxAPE. Tato míra je taktéž vyjádřena v procentech a představuje největší předpokládanou chybu. Na jejím základě můžeme získat představu o nejhorším možném scénáři naší předpovědi.

MaxAE je největší předpokládaná chyba. Je vyjádřená ve stejných jednotkách jako časová řada závisle proměnných. Stejně jako MaxAPE umožní představit si nejhorší možný scénář.

Namísto relativních chyb můžeme použít relativní měřítka a tím získáme další skupinu metod. Hyndman a Koehler (2006) upozorňuje na parametr

$$RealMAE = \frac{MAE}{MAE_b} \quad (5.21)$$

Tímto alternativním způsobem lze také konstruovat metody na bázi ostatních ukazatelů jako RMSE, MdAE, MAPE atd. Tyto alternativní metody se v běžné prognostické praxi ovšem využívají spíše vzácně.

V prognostických modelech je také používán pojem Normalizované BIC, respektive normalizované Bayesovské informační kritérium. Je to obecná míra celkového přizpůsobení modelu, která se pokouší vysvětlit jeho složitost. Toto kritérium je založeno na střední čtvercové chybě a zahrnuje i penalizaci za počet parametrů v modelu délky časové řady. Tím odstraňuje výhodu modelů s více parametry a lze tak porovnávat výsledné statistiky i napříč různými modely.

5.4 NEJPOUŽÍVANĚJŠÍ MÍRY PŘESNOSTI

V současnosti nejpoužívanější metody dle Hyndmana a Athanasopoulose (2018) jsou MAE a RMSE. Metoda MAE je velmi snadná na pochopení i interpretaci, avšak vede k preferencím prognóz mediánu. Naproti tomu RMSE povede ke zvolení metody založené na prognóze střední hodnoty. To je dnes asi hlavní důvod k širokému využití RMSE, a to i přes její obtížnější interpretovatelnost (Hyndman & Athanasopoulos, 2018). Nejčastěji používané modely jsou uvedeny v tabulce 5-2.

Tabulka 5-2 Nejčastěji používané míry přesnosti

Zkratka	Výpočet	Název anglicky
ME	$ME = mean(e_t)$	Mean Error
RMSE	$RMSE = \sqrt{mean(e_t^2)}$	Root Mean Squared Error
MAE	$MAE = mean(\lvert e_t \rvert)$	Mean Absolute Error
MPE	$MPE = mean(p_t)$	Mean Percentage Error
MAPE	$MAPE = mean(\lvert p_t \rvert)$	Mean Absolute Percentage Error
MASE	$MASE = mean(\lvert q_t \rvert)$	Mean Absolute Scaled Error

Zdroj: vlastní zpracování dle Wang et al. (2018), Hyndman & Athanasopoulos (2018), (Marček, 2016)

Realizace výpočtů v programu Excel je založena na individuálním výpočtu jednotlivých vzorců. Tento způsob je pro praktické využití značně zatěžující jak časově, tak z hlediska rizika lidské chyby. V programu SPSS je výpočet již více automatizován a program obsahuje funkce pro výpočet základních měr přesnosti. Postup a možnosti využití tohoto programu jsou v příloze 12. V tomto programu lze vypočítat tyto míry přesnosti:

- stacionární R^2,
- R^2,
- RMSE,
- MAPE,
- MAE,
- MaxAPE,
- MaxAE,
- normalizované BIC,
- Ljung- Box statistiky.

V jazyce R lze vyčíslit v podstatě všechny v současnosti známé míry přesnosti. V této monografii je pro výpočet využit opět balíček *forecast*, případně *forecastHybrid* a *forecTheta*. Ukázka kódů k výpočtu je uvedena v příloze 13.

Kapitola 6

Využití GT dat při prognózování individuální poptávky

Pro výpočet prognózy individuální poptávky byla využita GT data vyhledávání kategorie Toyota Corolla. Na základě těchto dat tedy budeme prognózovat poptávku po tomto automobilu v České republice. Data byla popsána v kapitole 3 a bylo zjištěno, že obsahují vzorce a že jejich prognózování je opodstatněné. Data nebyla nijak upravována a byly ponechány i případné odlehlé hodnoty, chybějící hodnoty data neobsahovala. Z dekompozice provedené v kapitole 3.4 vyplynulo, že data mají jasný rostoucí trend a vykazují silnou sezónnost.

Pro prognózu byl využit, stejně jako u ostatních výpočtů, statistický program R a Rstudio. Byly využity celé řady balíčků zveřejněných akademiky po celém světě, zejména pak balíčky *forecast* dle Hyndmana (Hyndman & Khandakar, 2008) a *ggplot2* (Wickham et al., 2020). Další balíčky jsou pojmenovány u jednotlivých výpočtů, zejména *fotecastHybrid, forecTheta*. K dílčím výpočtům byl použit i Python s knihovnami *Pandas, math, numpy.*

Jako benchmark byla zvolena metoda naive. Platí zásada, že když metoda překoná benchmark, tj. velmi jednoduchou metodu, která nepotřebuje žádné složitější výpočty, tak může být přijata a aplikována v praxi. V této monografii jsou metody s benchmarkem srovnány, nicméně pak je vybrána jedna nejpřesnější dle zvolených kritérií měr přesnosti. A to i přesto, že benchmark překonalo více metod.

6.1 VÝBĚR NEJPŘESNĚJŠÍHO MODELU

Při prognózování byla data rozdělena na tréninková a testovací. Na tréninkových datech byl vytvořen příslušný model a testovací data vyzkoušela, jak je model přesný. Tréninková data tvořila 80 % dat a testovací 20 % dat, jak doporučují Hyndman a Athanasopoulos (Hyndman & Athanasopoulos, 2021). Pro měření přesnosti bylo využito opět více měr přesnosti, tj. ME, RMSE, MAE, MPE, MAPE, které byly definovány v kapitole 5.1. Pro výběr nejvhodnějšího modelu

prognózy byly vybrány metody *naive* jako benchmark, dále *sezónní naivní model*, *arima modely dle auto.arima*, *model ETS, BATS*, *Thetam*, model založený na umělých neuronových sítích *Nnetar* a hybridní model *forecastHybrid*. Výsledky měření přesnosti těchto modelů jsou uvedeny v tabulce 6-1.

Tabulka 6-1 Míry přesnosti modelů na tréninkových a testovacích datech

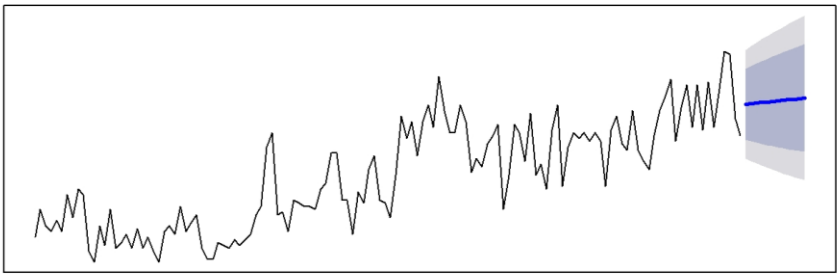
		ME	RMSE	MAE	MPE	MAPE
naive	Training	0,3238	10,3859	7,9429	-1,4782	16,3150
	Test	8,1538	14,2775	11,2308	8,5089	13,9097
snaive	Training	0,3238	10,3859	7,9429	-1,4782	16,3150
	Test	8,1538	14,2775	11,2308	8,5089	13,9097
ses	Training	0,7303	9,4415	7,3583	-1,4246	15,1234
	Test	7,2037	13,7570	10,7922	7,2306	13,4591
auto.arima	Training	1,0432	9,2113	7,0027	-0,7661	14,3885
	Test	8,1002	14,2456	11,2051	8,4372	13,8828
ETS	Training	0,5814	9,4883	7,4007	-1,5138	15,2514
	Test	7,2116	13,7611	10,7959	7,2412	13,4629
TBATS	Training	1,0945	9,4684	7,3744	-0,5961	15,0448
	Test	7,2235	13,7674	10,8014	7,2572	13,4685
Thetam	Training	0,7303	9,4415	7,3583	-1,4246	15,1234
	Test	4,0692	11,6342	9,2086	3,2061	11,7882
nnetar	Training	0,0006	6,1760	4,6531	-1,9961	9,8551
	Test	5,8560	13,0848	10,2065	5,4219	12,8710
forecast Hybrid	Training	0,6946	8,7485	6,8052	-1,3776	13,9764
	Test	6,4859	13,2339	10,3207	6,3056	12,9211

Zdroj: vlastní zpracování

Z výsledků vyplynulo, že nejpřesnější metoda pro zvolená data je Thetam. Ta na testovacích datech vykázala ve všech parametrech nejvyšší přesnost. Při zvoleném benchmarku, kterým byla metoda naive, by však mohly být použity všechny metody, protože všechny benchmark překonaly. Schopnost modelu adekvátně předpovídat hodnotíme dle dat testovacích. Samozřejmě na datech tréninkových i jiné modely poskytují lepší výsledky, nicméně to nic nevypovídá o tom, zda model předpoví správně i na dalších datech. Například metoda Nnetar je na tréninkových datech zdaleka nepřesnější. To je u metod založených na umělých neuronových sítích zcela běžné a bylo to potvrzeno i předchozími výstupy autorky (Kolková & Ključnikov, 2022). Také je z hodnot patrné, že podobné výsledky poskytovaly všechny míry přesnosti a použitím jiné míry přesnosti by nebyl vybrán jiný model. Zvýšení počtu měr přesnosti tudíž nemusí mít vliv na změnu zvoleného modelu.

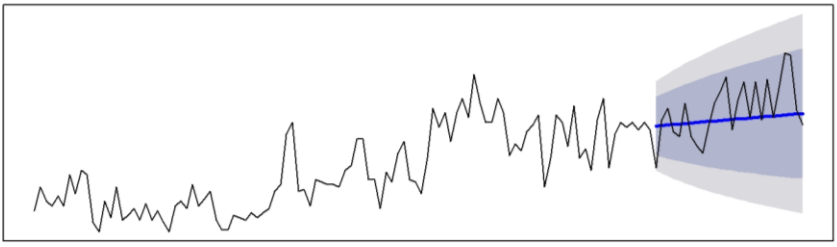
6.2 PROGNÓZA DLE NEJPŘESNĚJŠÍHO MODELU

Pro samotnou prognózu byla tedy vybrána metoda Thetam. Na obrázku 6-1 vidíme výsledky prognózy.



Obrázek 6-1 Prognóza poptávky po automobilu Toyota Corolla dle modelu Thetam
Zdroj: vlastní zpracování

Z obrázku je patrné, že poptávka zřejmě bude mít mírně rostoucí trend. A lze tedy očekávat v následujícím roce její zvýšení . Obrázek 6-1, který ukazuje prognózu do dalšího roku, je možné doplnit obrázkem prognózy, která byla vytvořena pouze na datech tréninkových, to znamená o 20 % kratších. Tato prognóza je na obrázku 6-2 a vidíme, že i zde byl prognózován mírný nárůst zájmu o automobil tohoto typu a skutečná data tomu opravdu odpovídají. Všechny reálné hodnoty GT dat se nacházejí v pásmu 80 % prognózy.



Obrázek 6-2 Prognóza hodnot zájmu o automobil na tréninkových datech s vyznačením skutečných dat.
Zdroj: vlastní zpracování

Grafické vyjádření je vhodné doplnit i číselnými údaji. Je nutné podotknout, že zde se nejedná o přesný budoucí poptávaný počet automobilů Toyota Corolla (takové informace by se daly prognózovat jen velmi ztěžka). Jedná se ale o nárůst zájmu, v tomto případě měřený na základě vyhledávání. Hodnoty zájmu (100 je největší zájem, 0 žádný zájem) dle vyhledávání jsou uvedeny v tabulce 6-2. Point Forecast představuje nejpravděpodobnější hodnoty budoucího vyhledávání (v grafu modrá čára). Hodnoty Lo80 a Hi80 udávají pásmo, kde se hodnoty budou pohybovat s 80% pravděpodobností. Hodnoty Lo95 a Hi95 pak tvoří pásmo, ve

kterém se hodnoty budou pohybovat s pravděpodobností 95 %. Na obrázku 6-1 a 6-2 jsou tato pásma tvořena odstíny šedé, tmavší pro 80% pásmo a světlejší pro 95% pásmo.

Tabulka 6-2 Hodnoty prognózy vyhledávání vozidla Toyota Corolla

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
01.01.2024	81,22849	68,57258	93,8844	61,87295	100,584
01.02.2024	81,43088	68,05771	94,80405	60,97839	101,8834
01.03.2024	81,63327	67,57941	95,68714	60,13974	103,1268
01.04.2024	81,83567	67,13258	96,53875	59,34924	104,3221
01.05.2024	82,03806	66,71323	97,36289	58,60076	105,4754
01.06.2024	82,24045	66,31815	98,16276	57,88939	106,5915
01.07.2024	82,44285	65,94468	98,94101	57,21108	107,6746
01.08.2024	82,64524	65,59065	99,69983	56,5625	108,728
01.09.2024	82,84763	65,25421	100,44106	55,94081	109,7545
01.10.2024	83,05003	64,93379	101,16627	55,34363	110,7564
01.11.2024	83,25242	64,62803	101,8768	54,76888	111,736
01.12.2024	83,45481	64,33578	102,57384	54,21478	112,6948

Zdroj: vlastní zpracování

Z vyčíslených dat tedy jednoznačně vyplývá, že hodnota zájmu o automobil, měřený počtem vyhledávání na Google, bude převyšovat poloviční popularitu pojmu. Lze tedy očekávat, že poptávka po těchto automobilech nebude nižší než v předchozích měsících. To je pro firmu jistě pozitivní zpráva. V dalším výzkumu bychom pak mohli na skutečných datech poptávky provést dekompozici a zjistit tak, v kterých měsících bude poptávka nejvyšší a kdy nižší, a připravit se tak na vývoj nákupů. Dále na skutečných datech by bylo možno s určitou pravděpodobností i předpovědět počet koupených automobilů, pokud by podnik tuto informaci potřeboval. Skutečné nákupy ovšem obvykle jsou o něco zpožděné oproti vyhledávání pojmu. Tomuto jevu se více věnuje behaviorální ekonomie.

Kapitola 7

Využití GT dat při predikci tržní poptávky

Dalším alternativním způsobem využití GT dat je prognózování kategorií vyhledávání. Tato informace může být zajímavá pro start-up firmy nebo pro subjekty, kteří o podnikání teprve uvažují. Touto prognózou dojdou vlastně k předpovědi tržní poptávky, pokud provedeme lehké zjednodušení a kategorii budeme považovat za trh konkrétních produktů či služeb. Kategorie, jak je sdružují analytici v Googlu, jsou tvořeny vyhledáváním podobných produktů a služeb. Z tohoto pohledu tvoří vyhledávání jakési části tržní poptávky.

Z předchozí kapitoly, jak už bylo řečeno, také vyplynul fakt, že není důležité hodnotit více různých kritérií měr přesnosti. V naprosté většině případů vedou k přijetí stejného modelu. Proto v následujících dvou kapitolách bude využito už pouze jedno kritérium, a to RMSE.

7.1 VÝBĚR NEJPŘESNĚJŠÍCH MODELŮ

Nejprve bude v souladu s postupem v kapitole 2.4 nalezen nejvhodnější model pro GT data vyhledávání kategorií. Výsledky měr přesnosti RMSE vypočítané na tréninkových datech jsou uvedeny v tabulce 7-1.

Tabulka 7-1 RMSE modelů pro GT data vyhledávání kategorií na tréninkových datech.

	Naive	Seasonal	ETS	ARIMA	BATS	NNETAR	hybrid Model	Thetam
Automobily	4,727	4,727	4,542	4,383	4,560	4,335	4,417	4,542
Cestování	7,264	7,264	7,214	5,397	5,322	0,004	5,002	7,213
Domácí mazlíčci	3,519	3,519	3,277	3,275	3,282	2,901	3,118	3,275
Domov a zahrada	4,217	4,217	4,002	3,786	4,007	3,792	3,832	4,000
Hobby a volný čas	6,508	6,508	5,033	5,205	5,003	4,329	4,895	5,379
Hry	4,655	4,655	4,613	4,622	4,490	4,345	4,464	4,612
Knihy a literatura	7,917	7,917	7,862	6,726	7,014	0,167	5,679	7,862
Nemovitosti	6,137	6,137	5,401	5,402	5,404	0,009	4,513	5,401
Počítače a elektronika	4,977	4,977	4,942	4,942	4,913	0,155	3,835	4,942

Zdroj: vlastní zpracování

Z výsledků na tréninkových datech je patrné, že pro všechny kategorie je nejvhodnější použití metody na bázi umělých neuronových sítí Nnetar. Jak už ale z předchozích výzkumů vyplynulo, v souladu s odbornou literaturou a jak je uvedeno v předchozích kapitolách, na tréninkových datech je metoda umělých neuronových sítí téměř vždy velmi úspěšná. Tento úspěch je dán tím, že model hledá ve shodě s daty a přímo na míru tréninkovým datům. Pokud takto vytvořený model necháme predikovat časový horizont v délce testovacích dat a pak s testovacími daty srovnáme, obvykle výsledky už tak výrazně lepší než u ostatních modelů nejsou.

Výsledky míry přesnosti RMSE na tréninkových datech vypovídají pouze o schopnosti modelu tato data popsat, což pro praktické využití nemá význam. Proto je daleko významnější hodnotit výsledky RMSE na datech testovacích. Výsledky výpočtu měr přesnosti RMSE vypočítané na testovacích datech jsou uvedeny v tabulce 7-2.

Tabulka 7-2 RMSE modelů pro GT data vyhledávání kategorií na testovacích datech.

	Naive	Seasonal	ETS	ARIMA	BATS	NNETAR	hybrid Model	Thetam
Automobily	9,34	9,34	8,42	7,97	8,42	6,86	8,27	11,09
Cestování	14,18	14,18	14,18	23,51	27,73	18,26	18,64	15,52
Domácí mazlíčci	10,32	10,32	11,14	11,20	11,24	13,19	11,34	10,38
Domov a zahrada	8,90	8,90	7,73	9,40	7,74	12,58	8,14	8,75
Hobby a volný čas	6,90	6,90	7,49	8,23	7,80	9,79	7,83	7,50
Hry	10,50	10,50	9,99	10,50	7,12	7,41	8,47	8,12
Knihy a literatura	12,70	12,70	12,70	12,62	12,40	14,87	11,69	13,15
Nemovitosti	11,41	11,41	11,42	11,42	11,42	10,64	10,93	11,04
Počítače a elektronika	11,31	11,31	11,31	11,31	11,48	8,37	9,50	9,89

Zdroj: vlastní zpracování

Dle výsledků na testovacích datech už se vhodnost jednotlivých modelů pro daná data liší. Pro data kategorie automobily je nejvhodnější model Nnetar. Pro kategorii cestování model Thetam. Pro kategorii domácí mazlíčci je dokonce nejvhodnější model naive, respektive seasonal naive. Pro kategorii domov a zahrada model Bats. Pro hobby a volný čas opět model naive, respektive seasonal naive. Pro kategorii hry je nejpřesnější model Bats. Pro kategorii knihy a literatura je nejpřesnější HybridModel. Kategorie nemovitosti má největší přesnost při použití Nnetar a kategorie počítače a elektronika je také nejpřesněji predikovatelná pomocí Nnetar. Pro výslednou predikci jsou samozřejmě použité modely nejpřesnější na testovacích datech. Výsledné modely jsou pro přehlednost uvedeny v tabulce 7-3.

Tabulka 7-3 Nejpřesnější modely pro jednotlivé kategorie vyhledávání

	Nejpřesnější model
Automobily	Nnetar
Cestování	Thetam
Domácí mazlíčci	naive, respektive seasonal naive
Domov a zahrada	Bats
Hobby a volný čas	naive, respektive seasonal naive
Hry	Bats
Knihy a literatura	HybridModel
Nemovitosti	Nnetar
Počítače a elektronika	Nnetar

Zdroj: vlastní zpracování

7.2 VÝSLEDKY MODELŮ

Nejčastěji vykazoval nejvyšší přesnost opět model Nnetar založený na umělých neuronových sítích. Trochu překvapivé je, že u dvou kategorií ani jeden model nepřekročil přesnost jednoduchého modelu naive, respektive seasonal naive.

Pro tři kategorie vyhledávání, a to automobily, nemovitosti a počítače a elektronika byl vhodný model Nnetar s parametry NNAR(13,7). Model je průměrem 20 samostatných neuronových sítí. Použití více sítí a jejich průměrování je technika zlepšující stabilitu a přesnost předpovědí. Každá z těchto sítí má strukturu:

- 13 vstupních neuronů (pravděpodobně odpovídajících počtu zpožděných hodnot časové řady, které jsou použity k předpovědi další hodnoty)
- 7 skrytých neuronů ve skryté vrstvě
- 1 výstupní neuron, který generuje předpověď
- 106 vah celkem v každé síti, které jsou parametry modelu upravované během tréninku, aby se minimalizovala chyba předpovědi.

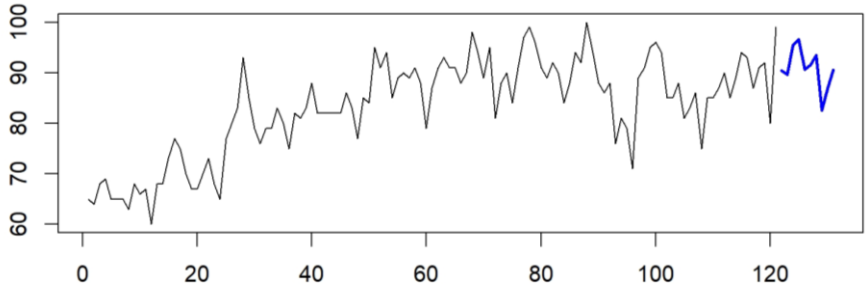
Všechny tři modely používají lineární aktivační funkci ve výstupní vrstvě, což znamená, že výstupní hodnoty jsou lineární kombinací vstupů. Rozptyl reziduí měření sigmou ^2 a samotné výsledné hodnoty předpovědí jsou samozřejmě již jiné pro každý ze tří odhadovaných modelů. Výsledky těchto tří modelů jsou v tabulce 7-4.

Tabulka 7-4 Předpovědi na 10 měsíců modelů s nejpřesnější metodou Nnetar

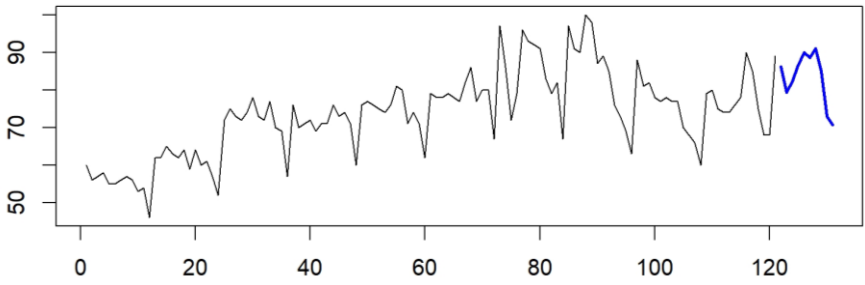
	Hodnota předpovědi vyhledávání kategorie automobily	Hodnota předpovědi vyhledávání kategorie nemovitosti	Hodnota předpovědi vyhledávání kategorie počítače a elektronika
sigma ^2	0,03242	0,04016	0,08461
01.01.2024	92,51641	84,06985	79,52066
01.02.2024	89,97900	77,44700	83,20610
01.03.2024	96,88161	79,86832	76,57105
01.04.2024	94,24411	84,36523	72,89085
01.05.2024	88,82445	81,32400	73,93731
01.06.2024	92,58345	83,88813	70,14573
01.07.2024	91,03155	83,83169	77,14800
01.08.2024	82,35544	79,66038	79,90827
01.09.2024	88,24738	73,13820	88,53621
01.10.2024	90,96522	72,78992	85,42924

Zdroj: vlastní zpracování

Z výsledků je patrné, že nejpřesněji Nnetar odhadl predikci vyhledávání kategorie automobily. Všechny kategorie jsou prognózovány s větším než polovičním zájmem o vyhledávání. Prognózy pro kategorie automobily a nemovitosti neobsahuje jednoznačný trend po celé prognózované období. To lze vidět i na obrázcích 7-1 a 7-2. Měsíce růstu jsou střídány měsíci poklesu, podobně jako tomu je u původní časové řady. To znamená, že v rámci tržního prostředí s žádnou z daných kategorií nemůžeme očekávat enormní vzestup poptávky, ale také ani enormní pokles. Výrobky a služby nabízené v rámci daných kategorií budou mít i v následujícím období pravděpodobně stagnující trend. Pro kategorii počítače a elektronika lze očekávat růst.

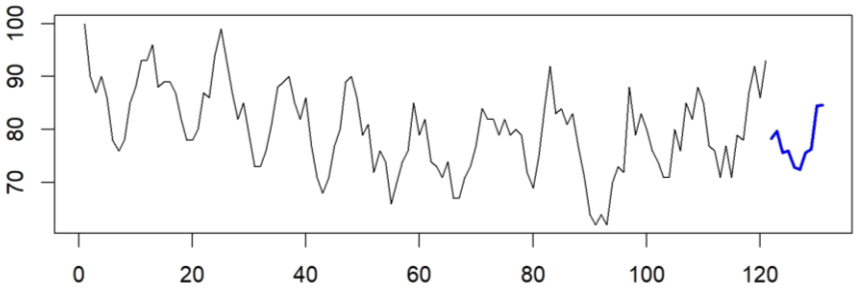


Obrázek 7-1 Prognóza vyhledávání kategorie automobily
Zdroj: vlastní zpracování



Obrázek 7-2 Prognóza vyhledávání kategorie nemovitosti
Zdroj: vlastní zpracování

Kategorie automobily a nemovitosti na konci prognózovaného období nedosahují ani hodnoty z počátku prognózy. Naopak kategorie počítače a elektronika na konci období vykazuje vyšší hodnoty než na svém počátku, lze tedy očekávat jejich růst. Toto je patrné i na obrázku 7-3, kde ovšem z historického vývoje vidíme, že v minulosti mělo vyhledávání této kategorie i vyšší hodnoty.



Obrázek 7-3 Prognóza vyhledávání kategorie počítače a elektronika
Zdroj: vlastní zpracování

Z obrázků je patrné, že model Nnetar z balíčku *forecast* nevypočítává 80% a 95% hranice prognózy. Ty by bylo nutno dopočítat ručně, což je neefektivní a v praxi těžko použitelné, proto i v této publikaci je model Nnetar vykreslován je formou linie předpovědi.

Pro GT data vyhledávání v kategorii domov a zahrada a kategorii hry je nejpřesnější metoda BATS s parametry uvedenými v tabulce 7-5.

Tabulka 7-5 Parametry modelu BATS

Parametr	Kategorie domov a zahrada	Kategorie hry
Model	BATS(0, {0,0}, -, -)	BATS(0,011, {0,0}, 0,874, -)
Lambda	0,000287	0,011369
Alfa	0,4886646	0,963195
Beta	-	-0,1496907
Damping parametr	-	0,874193
Sigma	0,06682066	0,06195012
AIC	957,0086	963,2315

Zdroj: vlastní zpracování

Pro kategorii domov a zahrada tedy můžeme konstatovat, že model BATS použil pro časovou řadu jednoduchý lineární přístup bez sezónních či cyklických komponentů. Model dosahuje mírného vyhlazení (alfa = 0,4886646) a má relativně nízkou odchylku reziduí (sigma = 0,06682066).

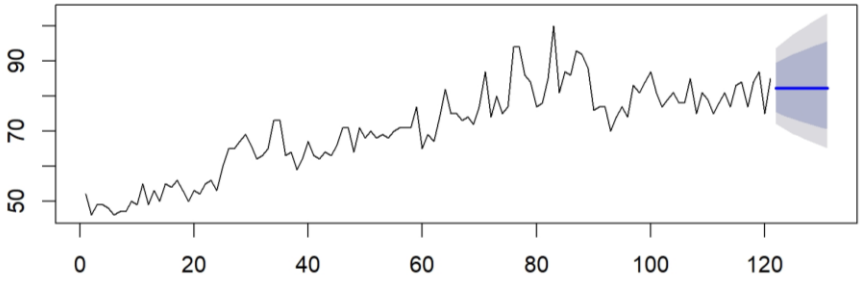
Model BATS pro časovou řadu GT dat kategorie hry použil mírnou Box-Cox transformaci (lambda = 0,011), která je téměř lineární. Model má vysokou hodnotu parametru alfa (0,963195), což naznačuje, že silně reaguje na nové údaje. Trend je klesající (beta = -0.1496907) a damping parametr pro sezónní složku je 0,874193. Standardní odchylka reziduí (sigma = 0,06195012) je relativně nízká, což znamená, že model dobře odpovídá datům.

Tabulka 7-6 Předpovědi na 10 měsíců modelů s nejpřesnější metodou BATS

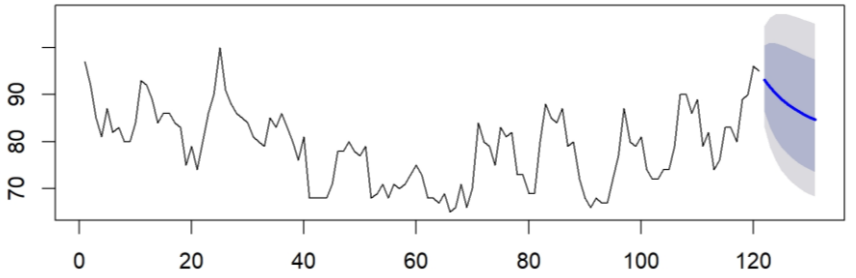
	Hodnota předpovědi vyhledávání kategorie domov a zahrada	Hodnota předpovědi vyhledávání kategorie hry
01.01.2024	82,19489	93,13172
01.02.2024	82,19489	91,56556
01.03.2024	82,19489	90,21779
01.04.2024	82,19489	89,05566
01.05.2024	82,19489	88,05186
01.06.2024	82,19489	87,18352
01.07.2024	82,19489	86,43136
01.08.2024	82,19489	85,77909
01.09.2024	82,19489	85,21286
01.10.2024	82,19489	84,72090

Zdroj: vlastní zpracování

Jak vidíme z tabulky 7-6 i z obrázku 7-4, je trend pro vyhledávání kategorie domov a zahrada naprosto stagnující. Hodnoty jsou lineární a očekává se, že se nebudou po celé prognózované období měnit. To je i trendem v posledních letech. Je tedy zřejmé, že po předchozích letech trvalého růstu už tato kategorie vyčerpala svůj potenciál poptávky a v budoucnosti to bude kategorie spíše s poptávkou stagnující.



Obrázek 7-4 Prognóza vyhledávání kategorie domov a zahrada
Zdroj: vlastní zpracování



Obrázek 7-5 Prognóza vyhledávání kategorie hry
Zdroj: vlastní zpracování

Kategorie hry je na tom poněkud hůře a podle obrázku 7-5 lze očekávat klesající poptávku po herních produktech. V této kategorii ovšem není rozlišeno, zda se jedná o hry pouze deskové, nebo i počítačové. Jelikož je to tedy dost široký pojem, pro prognózu tržní poptávky v tomto odvětví je potřeba dotazování zpřesnit. Pak už by šlo i vyhledávání pojmu, nikoli kategorie. Z prognózy můžeme pouze obecně konstatovat, že u těchto produktů lze očekávat pokles.

Pro kategorii cestování je na základě míry přesnosti RMSE nejpřesnější model Thetam z balíčku forecTheta. Tento model má parametry uvedené v tabulce 7-7.

Tabulka 7-7 Parametry modelu Thetam

Parametr	Kategorie cestování
alpha	0,9999
Initial states	54,9972
sigma	8,7181
AIC	1108,301

Zdroj: vlastní zpracování

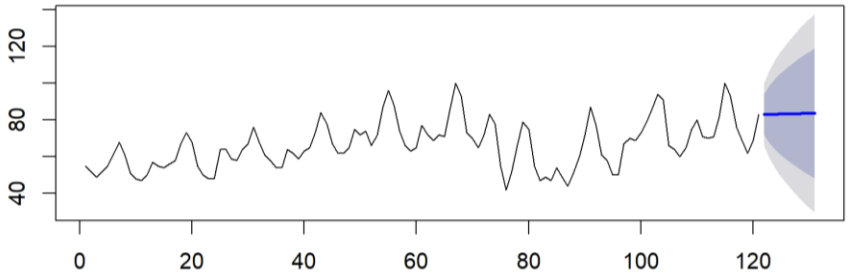
Model Thetam pro časovou řadu kategorie cestování využívá jednoduchý ETS model bez trendu a sezónnosti s velmi vysokým hladicím parametrem alpha, který má hodnotu 0,9999. Počáteční úroveň je přibližně 55 a model předpokládá vysokou variabilitu chyb (sigma 8.7181).

Tabulka 7-8 Předpovědi na 10 měsíců modelů s nejpřesnější metodou Thetam

	Hodnota předpovědi vyhledávání kategorie cestování
01.01.2024	83,06937
01.02.2024	83,14014
01.03.2024	83,21091
01.04.2024	83,28167
01.05.2024	83,35244
01.06.2024	83,42320
01.07.2024	83,49397
01.08.2024	83,56474
01.09.2024	83,63550
01.10.2024	83,70627

Zdroj: vlastní zpracování

Prognóza uvedená v tabulce 7-8 poskytuje bodové odhady a intervaly spolehlivosti pro několik období do budoucna, přičemž intervaly naznačují nejistotu v předpovědích. Na obrázku 7-6 je tato metoda vyobrazena graficky.



Obrázek 7-6 Prognóza vyhledávání kategorie cestování
Zdroj: vlastní zpracování

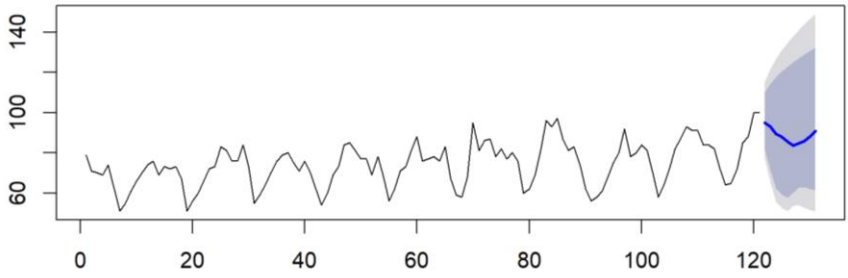
Další model, který byl určen jako nepřesnější pro časovou řadu vyhledávání kategorie knihy a literatura, je hybridModel. Tento model spojuje modely auto.arima, ets, Thetam, Nnetar, Bats. Zahrnuje je do hybridního modelu s váhou 0,2 pro každý z modelů. Hodnota předpovědí tohoto hybridního modelu je uvedena v tabulce 7-9.

Tabulka 7-9 Předpovědi na 10 měsíců modelů s nejpřesnější metodou hybridModel

	Hodnota předpovědi vyhledávání kategorie knihy a literatura
01.01.2024	94,75199
01.02.2024	92,69008
01.03.2024	89,74901
01.04.2024	87,19388
01.05.2024	84,12845
01.06.2024	83,16918
01.07.2024	84,41350
01.08.2024	85,31306
01.09.2024	87,35745
01.10.2024	89,15212

Zdroj: vlastní zpracování

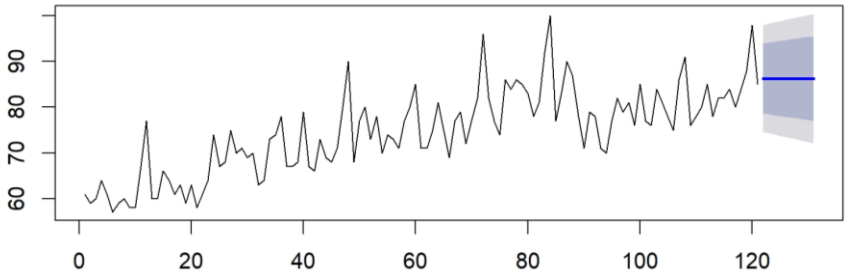
Tento model má kolísavou tendenci, nicméně celkový trend můžeme označit jako klesající. Je možné, že o několik měsíců později by hodnoty dosáhly hodnot počátečních, ale to už není předmětem naší prognózy. Z obrázku 7-7 také nevyčteme jednoznačnou odpověď na to, zda poptávka v tomto odvětví má potenciál růst.



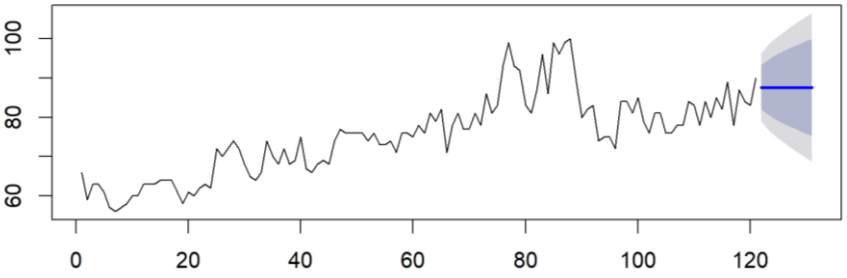
Obrázek 7-7 Prognóza vyhledávání kategorie knihy a literatura
Zdroj: vlastní zpracování

Poslední dvě časové řady vyhledávání, kategorie domácí mazlíčci a hobby a volný čas, neměly žádný model, který by překonal jednoduchou metodu sezónní naivní. Pro ni platí jednoduchá logika, a nemá tudíž žádné další složitější parametry. Pro kategorii domácí mazlíčci je prognóza lineární, a tedy hodnota stagnující na úrovni 87,62178. Ta je predikována po celých 10 měsíců.

Pro kategorii hobby a volný čas je tomu velmi podobně. Po celý predikovaný interval metoda předpovídá hodnotu 86,30482.



Obrázek 7-8 Prognóza vyhledávání kategorie domácí mazlíčci
Zdroj: vlastní zpracování



Obrázek 7-9 Prognóza vyhledávání kategorie hobby a volný čas
Zdroj: vlastní zpracování

Obrázky 7-8 a 7-9 jen potvrzují, že prognóza je velmi jednoduchá a o poptávce po produktech z těchto kategorií nám mnoho neřekne. I to mohou být reálné výstupy z prognostické analytiky. Není to vždy jen jednoznačná odpověď na naše otázky o budoucí poptávce po námi zvolených produktech. Výsledky prognostické analytiky vždy budou odpovídat jen datům, která analyzují, a pokud v nich není jasný trend, nenaleznou jej.

7.3 SHRNUTÍ PREDIKCE TRŽNÍ POPTÁVKY NA ZÁKLADĚ GT DAT

Nakonec je vhodné shrnout predikce jednotlivých kategorií. Výsledky prognóz jsou slovně shrnuty v tabulce 7-10.

Tabulka 7-10 Výsledky predikce pro jednotlivé kategorie vyhledávání

	Detailní popis trendů na 10 měsíců	Predikce na konci intervalu
Automobily	Kolísavý trend s dvěma vrcholy, daný cykličností poptávky	Pokles
Cestování	Mírný nárůst po celé predikované období	Růst
Domácí mazlíčci	Stagnace	Stagnace
Domov a zahrada	Stagnace	Stagnace
Hobby a volný čas	Stagnace	Stagnace
Hry	Pokles po celé období	Pokles
Knihy a literatura	Pokles 6 měsíců, nižší nárůst 4 měsíce	Pokles
Nemovitosti	Kolísavý trend s dvěma vrcholy, daný cykličností poptávky	Pokles
Počítače a elektronika	Růst po celou dobu predikce	Růst

Zdroj: vlastní zpracování

Z tabulky vyplývá fakt, že čtyři kategorie vykazují pokles, tři kategorie vykazují stagnaci a pouze u dvou kategorií byl zaznamenán růst. Jedná se o vyhledávání kategorie cestování a kategorie počítače a elektronika. Oblasti cestování a počítače a elektronika mají tedy velký potenciál růstu poptávky v roce 2024. U kategorie počítače a elektronika je to logický důsledek čtvrté průmyslové revoluce a prudkého rozvoje počítačových softwarů, umělé inteligence i sociálních sítí. Lidé zřejmě musejí i svůj hardware přizpůsobit těmto dynamickým změnám. Růst poptávky po cestování zase může být důsledkem ukončení pandemie a následného znovuobnovení cestovního ruchu. Lidé cestování během pandemie odkládali a nyní opět dochází k růstu zájmu. Je zřejmé, že zahájení podnikání v těchto dvou oblastech může mít pozitivní výsledky. Naopak zahájit podnikání v oblasti her, knih a literatury, případně nemovitostí a automobilů může být problematické. I když je možné, že v oblasti automobilů a nemovitostí je výsledek dán jen tím, že prognózované období skončilo před vrcholem cyklu. Tyto kategorie se vyznačují značnou variabilitou a jednoznačně definovat trend na 10 měsíců je obtížné.

Kapitola 8

Využití GT dat při predikci poptávky u konkurence

V této kapitole budou vyhodnoceny značky automobilů, zvolených dle motivace popsané v kapitole Příprava dat pro prognózování. Cílem kapitoly je popsat, jak by mohla data Google Trends napomoci při odhadu potenciálu poptávky našich konkurentů. Jak už bylo řečeno, získat data konkurence je téměř nemožné. Ale data vyhledávání produktů konkurence nám mohou být významným pomocníkem k analýze naší konkurence. Získáme je lehce a zdarma. Přesto nám poskytnou významné informace o tom, zda potenciální poptávka konkurence poroste či nikoli. Vyhledávání pojmů nebo kategorií obvykle předchází samotným nákupům, proto tuto informaci o naší konkurenci získáme s předstihem a můžeme tomu přizpůsobit svá podnikatelská rozhodnutí.

Na základě vyhledávání jednotlivých značek vozidel je zajímavé odhadnout, zda budou mít potenciální poptávku rostoucí či klesající. Kapitola je opět rozdělena na dvě části, a to na výběr nejvhodnějšího modelu a samotnou predikci.

8.1 VYHLEDÁVÁNÍ NEJPŘESNĚJŠÍCH MODELŮ

Výsledky nejprve na tréninkových a pak na testovacích datech jsou v tabulkách 8-1 a 8-2. Data byla opět rozdělena v poměru 80:20.

Tabulka 8-1 RMSE modelů pro GT data vyhledávání typů vozidel na tréninkových datech

	Naive	Seasonal	ETS	ARIMA	BATS	NNETAR	hybrid Model	Thetam
Toyota	5,080	5,080	5,049	5,049	5,045	4,967	5,036	5,049
Ford	10,515	10,515	10,451	9,744	9,959	7,329	9,174	10,451
Tesla	0,950	0,950	0,843	0,843	0,861	0,639	0,789	0,843
Nissan	3,841	3,841	3,817	3,716	3,845	3,364	3,644	3,817
Honda	7,914	7,914	7,867	7,424	7,479	6,381	7,225	7,867

Zdroj: vlastní zpracování

Přesnost byla hodnocena na základě kritéria RMSE. Dle tohoto kritéria jsou stejně jako v předchozích kapitolách v tréninkových datech nejpřesnější modely založené na umělých neuronových sítích Nnetar. To ovšem, jak už bylo řečeno, nemá velkou vypovídací schopnost, protože až na datech testovacích zjistíme, zda model dokáže predikovat přesně i do budoucna.

Tabulka 8-2 RMSE modelů pro GT data vyhledávání typů vozidel na testovacích datech.

	Naive	Seasonal	ETS	ARIMA	BATS	NNETAR	hybrid Model	Thetam
Toyota	5,715	5,715	5,715	5,715	5,811	9,274	6,157	5,013
Ford	11,715	11,715	11,715	8,827	12,723	11,920	11,225	11,787
Tesla	4,098	4,098	2,876	2,895	2,694	1,615	2,654	3,175
Nissan	2,700	2,700	2,700	3,515	2,700	2,714	2,789	2,814
Honda	8,319	8,319	8,319	8,757	8,758	11,267	6,885	9,739

Zdroj: vlastní zpracování

U zhodnocení RMSE na datech testovacích jsou výsledky již naprosto odlišné. Pro značku Toyota je nejvhodnější prognostický model Thetam, pro značku Ford hybridModel, pro značku Tesla Nnetar, pro značku Nissan je to model Bats a pro značku Honda je nejvhodnější opět model hybridModel.

8.2 VÝSLEDKY MODELŮ

Pro značku Toyota, jak jsme ověřili v předchozí kapitole, je nejvhodnější, respektive nejpřesnější model Thetam. Tento model má parametry uvedené v tabulce 8-3. Hodnota parametru alpha je 0,875, což je hladicí koeficient v modelu exponenciálního vyrovnávání. Vysoká hodnota alpha (blízko 1) naznačuje, že model klade velký důraz na poslední pozorování a méně na starší data. To znamená, že model rychle reaguje na jejich změny.

Počáteční stav (Initial states) je 27,4917. Tato hodnota reprezentuje počáteční odhad stavu úrovně (level) časové řady. V modelu exponenciálního vyrovnávání (ETS) je úroveň časové řady jedním z hlavních komponentů a tato hodnota určuje počáteční odhad úrovně. Nižší hodnota sigma naznačuje menší variabilitu nevysvětlených změn v datech.

K porovnání různých modelů se používá AIC (Akaikeho informační kritérium). Nižší hodnota AIC indikuje lepší přizpůsobení modelu datům. AIC je založen na ztrátě informace, kde menší hodnota značí, že model dobře vysvětluje data s minimálními ztrátami informace. Zde je však hodnota poměrně vysoká.

Tabulka 8-3 Parametry modelu Thetam

Parametr	Toyota
alpha	0,875
Initial states	27,4917
sigma	5,1687
AIC	1091,336

Zdroj: vlastní zpracování

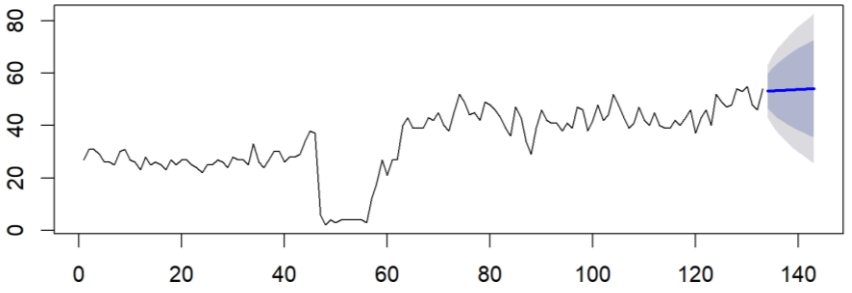
Předpovědi na následujících 10 měsíců jsou uvedeny v tabulce 8-4. Z těchto hodnot jednoznačně vyplývá, že vyhledávání automobilu Toyota bude mít v nejbližší době skoro stagnující tendenci. Hodnoty jsou takřka stejné a liší se pouze desetinnými místy. Můžeme ale konstatovat, že se jedná o mírný nárůst.

Tabulka 8-4 Předpovědi na 10 měsíců vyhledávání automobilu Toyota

	Hodnota předpovědi vyhledávání automobilu Toyota
01.01.2024	53,16854
01.02.2024	53,27703
01.03.2024	53,38553
01.04.2024	53,49403
01.05.2024	53,60252
01.06.2024	53,71102
01.07.2024	53,81952
01.08.2024	53,92802
01.09.2024	54,03651
01.10.2024	54,14501

Zdroj: vlastní zpracování

Při pohledu na graf v obrázku 8-1 lze také spatřit takřka stagnaci s mírným nárůstem. Potenciální poptávka, měřená vyhledáváním značky, tak bude velmi mírně růst.



Obrázek 8-1 Prognóza vyhledávání značky Toyota
Zdroj: vlastní zpracování

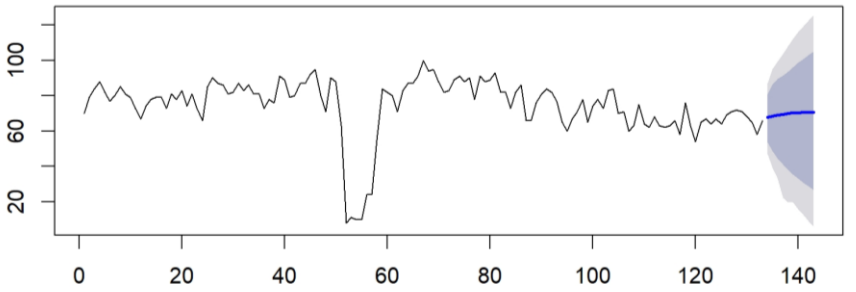
Vyhledávání značky Ford a Honda je realizováno pomocí hybridModelu, který byl v předchozí kapitole označen za nejpřesnější. Model slučuje auto.arima, ETS, Thetam, Nnetar a Bats. Všechny s podíly 0,2.

Tabulka 8-5 Předpovědi na 10 měsíců modelů s nejpřesnější metodou hybridModel

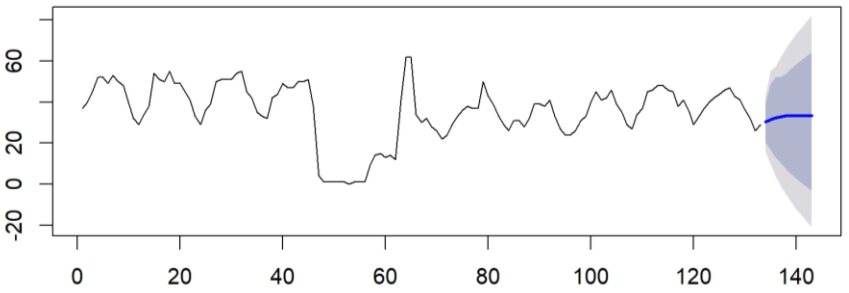
	Hodnota předpovědi vyhledávání automobilů Ford	Hodnota předpovědi vyhledávání automobilů Honda
01.01.2024	67,86419	30,47183
01.02.2024	68,62045	31,46134
01.03.2024	69,14394	32,25502
01.04.2024	69,60839	32,95398
01.05.2024	70,02986	33,33168
01.06.2024	70,36542	33,46277
01.07.2024	70,61776	33,44485
01.08.2024	70,80356	33,35791
01.09.2024	70,93964	33,25244
01.10.2024	71,03896	33,15480

Zdroj: vlastní zpracování

Výsledky predikce shrnuje tabulka 8-5. Z ní vyplývá, že pro automobil značky Ford můžeme očekávat zvýšení zájmu. Totéž platí pro automobil Honda, pro něj jsou hodnoty potenciálního zájmu o vyhledávání na nižší úrovni. Obrázek 8-2 vykresluje situaci prognózy vyhledávání značky Ford. Obrázek 8-3 pak prognózy vyhledávání značky Honda. U obou je vidět mírný nárůst, jak už bylo řečeno.



Obrázek 8-2 Prognóza vyhledávání značky Ford
Zdroj: vlastní zpracování



Obrázek 8-3 Prognóza vyhledávání značky Honda
Zdroj: vlastní zpracování

Pro vyhledávání značky Tesla byl nalezen nejpřesnější model NNAR(3,2). To znamená, že se jedná o autoregresivní neuronovou síť s těmito parametry:

- 3: Počet zpožděných vstupů (lagů) časové řady použitých jako vstupy do sítě.
- 2: Počet neuronů ve skryté vrstvě.

Výsledný model je průměrem 20 neuronových sítí, každá z nich má strukturu 3-2-1. To znamená tři vstupní neurony (odpovídající třem zpožděným hodnotám), dva skryté neurony a jeden výstupní neuron. Celkový počet váhových parametrů v každé síti je 11. Neuronová síť použila lineární výstupní jednotky. To znamená, že výstupní neuron má lineární aktivační funkci. Předpovědi jsou shrnuty v tabulce 8-6.

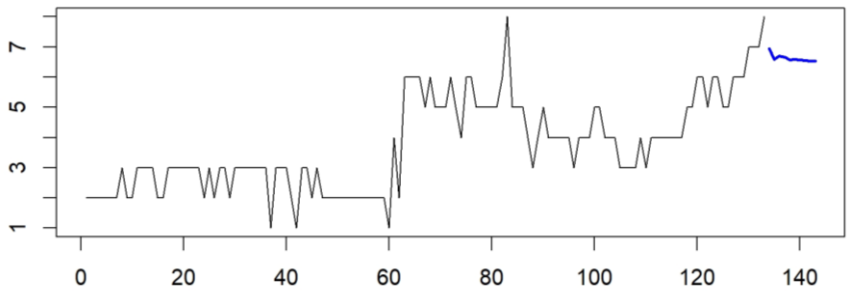
Tabulka 8-6 Předpovědi na 10 měsíců vyhledávání automobilu Tesla

	Hodnota předpovědi vyhledávání automobilu Tesla
01.01.2024	6,856561
01.02.2024	6,484022
01.03.2024	6,579160
01.04.2024	6,560648
01.05.2024	6,471893
01.06.2024	6,471332
01.07.2024	6,471044
01.08.2024	6,447602
01.09.2024	6,439307
01.10.2024	6,437869

Zdroj: vlastní zpracování

Hodnota sigma^2 je odhad rozptylu chyb modelu. V tomto případě je odhadnutý rozptyl chyb 0,4591. Ten lze interpretovat jako průměrný rozptyl (nebo

variabilitu) mezi předpovězenými hodnotami a skutečnými hodnotami v trénovacích datech.



Obrázek 8-4 Prognóza vyhledávání značky Tesla
Zdroj: vlastní zpracování

Jak už bylo uvedeno i v předchozí kapitole, model Nnetar má nastavení 85% a 90% hranice, jak vidíme na obrázku 8-4. Ruční dopočítávání je v praxi nereálné, a tak i zde je uváděn v tomto tvaru, pro praxi by pak stálo na zvážení doplnění modelu pomocí dalších kódů, případně využití jiných modelů na bázi umělých neuronových sítí.

Model BATS (s transformačními Box-Cox, ARMA chybami, trendem a sezónností) je jedním z pokročilých modelů pro časové řady, který je vhodný pro složité a sezónní časové řady. Jeho parametry jsou uvedeny v tabulce 8-7. Model BATS (1, {0,0}, -, -) naznačuje, že byl použit model BATS. Box-Cox transformace byla aplikována (lambda není rovna 0). {0,0} indikuje, že na chyby nebyl aplikován žádný ARMA model. Žádný trend (přirozený nebo dampovaný) nebyl identifikován, a nakonec nebyla identifikována ani žádná sezónnost.

Tabulka 8-7 Parametry modelu Bats

Parametr	Nissan
Model	(1, {0,0}, -, -)
Alfa	0,8818625
Sigma	3,384382
AIC	978,716

Zdroj: vlastní zpracování

Hodnota hladicího parametru alfa určuje míru vyhlazení úrovně časové řady. Vysoká hodnota (blízká 1) naznačuje, že model klade větší váhu na nedávná data při předpovědi úrovně. Počáteční odhad stavu úrovně časové řady (23,40763) odhaduje výchozí bod pro vyhlazovací algoritmus modelu. Odhad standardní odchylky chybového termínu modelu je 3.384382. Toto číslo udává míru rozptylu mezi skutečnými hodnotami a hodnotami předpovězenými modelem. Nižší sigma znamená menší rozptyl a lepší předpověď.

Pro model je hodnota AIC 978,716. AIC se používá k porovnání různých modelů a hodnotí, jak dobře model vysvětluje data s ohledem na počet parametrů.

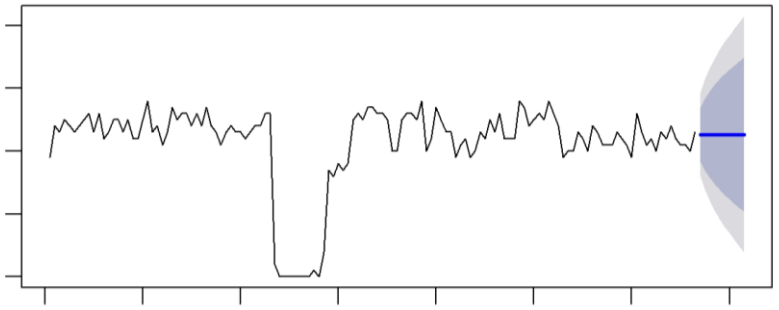
Nižší hodnota AIC znamená lepší model. Ten je preferován, protože poskytuje lepší rovnováhu mezi přesností modelu a jeho složitostí.

Tabulka 8-8 Předpovědi na 10 měsíců vyhledávání automobilu Nissan

	Hodnota předpovědi vyhledávání automobilu Nissan
01.01.2024	22,65978
01.02.2024	22,65978
01.03.2024	22,65978
01.04.2024	22,65978
01.05.2024	22,65978
01.06.2024	22,65978
01.07.2024	22,65978
01.08.2024	22,65978
01.09.2024	22,65978
01.10.2024	22,65978

Zdroj: vlastní zpracování

Hodnota předpovědí je uvedena v tabulce 8-8 a opět zde můžeme vidět stagnující trend. Jednotlivé předpovědi pro všechny měsíce jsou absolutně stejné. Grafické znázornění této stagnace je uvedeno na obrázku 8-5.



Obrázek 8-5 Prognóza vyhledávání značky Nissan
Zdroj: vlastní zpracování

8.3 SHRNUTÍ ANALÝZY KONKURENCE POMOCÍ PREDIKCE NA ZÁKLADĚ GT DAT

Nakonec provedeme opět srovnání jednotlivých modelů a vyhodnocení v tabulce 8-9. U žádného z modelů nebyl predikován pokles. Takže nelze s jistotou tvrdit, že o některý z typů automobilů bude v roce 2024 menší zájem. Automobil Tesla a Nissan zaznamenaly stagnaci. U automobilu Nissan se jedná o stagnaci po celou dobu sledovaného prognózovaného období. U automobilu Tesla se projevovalo mírné kolísání v řádu desetinných hodnot, to ovšem lze také označit za stagnaci.

Růst zájmu o vyhledávání značek automobilů zaznamenaly automobilky Toyota, Ford a Honda. Lze tedy očekávat, že když subjekty více vyhledávají tyto značky, může to znamenat potenciální růst poptávky. Pro potenciální analýzu konkurence to může být klíčová informace.

Tabulka 8-9 Výsledky predikce pro jednotlivé vyhledávání automobilů

	Detailní popis trendu na 10 měsíců	Predikce na konci intervalu
Toyota	Mírný růst po celou predikci	Růst
Ford	Mírný růst po celou predikci	Růst
Tesla	Stagnace s mírným kolísáním	Stagnace
Nissan	Stagnace	Stagnace
Honda	Mírný růst po celou predikci	Růst

Zdroj: vlastní zpracování

Tyto údaje si může spočítat kdokoli bez nutnosti získání interních informací automobilek. Je to tedy jednoduchý nástroj analýzy konkurence. S využitím adekvátní výpočetní techniky jsou údaje dostupné v řádu několika hodin (případně jen minut) a podnik tak má okamžitou informaci, jaký je zájem o vyhledávání všech jeho konkurentů. V Google Trends datech samozřejmě lze zadat i vyhledávání konkrétního pojmu či konkrétního podniku a v tomto smyslu to může být zajímavé i pro analýzu konkurence malých a středních podniků.

Z analýzy také vyplynulo, že všechny prognózy vykazovaly nejvyšší přesnost při použití modernějších metod prognózování. Základní statistické metody, jako ETS nebo arima, nebyly uplatněny vůbec. To klade zase vyšší nároky na výpočetní techniku, což může znamenat pro malé a střední podniky jisté omezení. Nicméně dnes již i modely založené na umělé inteligenci lze vypočítat na běžném počítači bez extrémních nároků na výpočetní výkon. Také znalosti programovaní a využívání moderních statistických jazyků jsou u nastupující generace podnikatelů poměrně vysoké. Informatiku se dnes již učí děti na prvním stupni základních škol a základy programování jsou i ve výukových plánech druhých stupňů základních škol. Z toho důvodu lze předpokládat, že náročnost modelů a nároky na výpočetní znalosti již pro nastupující generaci malých a středních podnikatelů nebudou omezením.

Kapitola 9

Diskuze výsledků výzkumu a návrhy k další vědecké práci

Prognózování je individuální proces a pro každý podnik i pro každý produkt je nutné stále znovu vybírat vhodný model. Proces aplikovaný dle obrázku 2-6 v této monografii není jednorázový akt, ale stále se opakující proces. V podnicích je proto nutné ho co nejvíce automatizovat, což výpočty v této monografii umožňují. Dále je také potřebné neustále reflektovat vývoj v ekonomice i ve společnosti a v souladu s ním aktualizovat používané metody, které se neustále vyvíjejí. Ty, které ještě v minulých desetiletích poskytovaly poměrně přesné prognózy, v desetiletích následujících již mohou být zastaralé a nepřesné.

Pro podniky je prognózování poptávky samozřejmě nutnou záležitostí. Ukazuje se však, že sofistikovanější modely predikce poptávky zůstávají stále doménou velkých podniků. Malé a střední podniky častěji zůstávají u kvalitativních metod a u odborných odhadů svých manažerů či majitelů. Nutnost realizace predikcí ovšem zůstává na bedrech finančních institucí. Zde si lze jakékoli obchodování bez predikcí těžko představit. V těchto institucích je také stále zřetelný tlak na výzkum a vývoj v oblasti prognózování. Modely je nutné realizovat s vysokou přesností a v adekvátním čase, takže jsou náročné i na výpočetní čas. Vědecký vývoj prognózování a pro ně využívání GT je ve financích stále zřetelnější a v této oblasti je v současnosti více a více otevřených vědeckých otázek.

Samozřejmě i nadále budou probíhat diskuze o tom, jak jsou výsledky prognózování ovlivnitelné šoky. Ty lze předpovídat velmi obtížně, klasickými statistickými metodami je předpovídat prakticky nejde. Šoky také mohou být nejen ekonomické, ale také politické, zdravotní či sociální. Například během pandemie covid se ukázalo, že běžné způsoby prognózování akciových a jiných finančních trhů často selhávají. A byly to prognózy využívající právě GT data, která se jevily jako vhodná alternativa ke klasickému použití historických dat, což potvrzují i Maddodi a Kunte (2024). Ve své studii autoři předpovídali výkyvy akciového trhu na indickém finančním trhu. K tomu využili GT data a vybudovali

prediktivní model s maximální přesností 98,95 % právě při předvídání v krizových situacích a v situacích ekonomických šoků. Model je zaměřen krátkodobě. Při prognózování poptávky v době pandemie covid byla GT data také vhodná, což potvrzuje i Kolková a Rozehnal (2022).

Důležitost prognózování hlavně ve finančních institucích je dnes podpořeno již řadou výzkumů. Častým předmětem výzkumu je prognózování směnných kurzů na základě GT dat. Prognózování měnových kurzů je jinak poměrně složité a často málo přesné. GT data mohou pomoci tyto prognózy učinit přesnější, a tudíž využitelnější. Odhadem směnných kurzů na základě GT dat se zabývá například Bulut (2018). Ve svém článku autor prognózuje směnné kurzy na základě širokého přehledu literatury, který spojuje pohyby směnných kurzů s makroekonomickými fundamenty. Autor zkoumal 11 směnných kurzů zemí OECD v období 10 let. Zjistil, že předpovědi založené na Trendech Google lépe určují směr změn měsíčních nominálních směnných kurzů po éře velké recese (2008–2009) a předpovědi založené na GT datech poskytují přesnější výsledky než modely parity kupní síly. Nicméně, jak sám uvádí, je nutný ještě další výzkum v této oblasti, aby byla prozkoumána prediktivní síla informací.

Dle Ita et al. (2021) jsou předpovídány sazby USD/JPY pomocí tří strukturálních a dvou autoregresních modelů. Autoři pak zkoumají, zda jejich index sentimentu může zlepšit prediktivní sílu těchto modelů. Výsledky jednoznačně potvrzují, že přidání indexu sentimentu založeného na GT datech zpřesnilo prognózu směnného kurzu.

Jiný dokument (Herzog & dos Santos, 2021) pak studuje sílu metrik intenzity online vyhledávání, měřených společností Google, pro předpovídání směnných kurzů. Autoři používají panelová data ze čtvrtletních časových řad od roku 2004 do roku 2018 a deseti mezinárodních zemí s nejvyšším objemem obchodů s měnami. Do panelových dat zařazují také metriky intenzity vyhledávání Google. Z článku vyplývá závěr, že online vyhledávání zlepšuje celkově ekonometrické modely.

Směnné kurzy ovšem nejsou jedinou veličinou, která je ve financích pro prognózování pomocí GT dat vhodná. Ačkoli právě směnné kurzy by mohly být lépe vysvětleny institucionálními, behaviorálními a jinými determinanty a očekáváním trhu než ekonomickými fundamenty a historickým vývojem, jak uvádí už v roce 2016 Kapounek a Kučerová. Z toho důvodu má možná využití GT dat pro jejich prognózování největší vědecký potenciál. Nicméně pomocí GT dat můžeme dopředu odhadnout turbulence na finančních trzích vůbec. To analyzovali Petropoulos et al. (2022). Aplikovali metodiku těžby textu na sadu 10.000 projevů centrální banky a vytvořili finanční slovník, na jehož základě používají indexy Google Trends k měření zájmu lidí o finanční zprávy. Výsledkem bylo zjištění, že dotazy Google zprostředkovávají informace schopné předvídat budoucí turbulence na trhu v krátkém časovém období (jeden měsíc) a že jasně překonávají algoritmy Deep Learning oproti referenčním technikám.

GT data jsou také předmětem výzkumu technik optimalizace portfolia, například Kupfer a Zorn (2019). V tomto vědeckém článku na základě informací o relativním nákupním zájmu na základě GT dat předpovídají budoucí prodeje, a to začleňují do technik optimalizace portfolia. A to na základě jednoduché logiky, že čím vyšší objem vyhledávání firmy, tím vyšší pozici má firma v procesu optimalizace. U vzorku firem z módního průmyslu se to ukazuje jako vhodný nástroj optimalizace.

Dle Deba (2023) taktéž ve financích lze GT data využít pro analýzu volatility akciových kurzů. Autor zde analyzoval pohyby cen akcií tří hlavních leteckých společností na základě míry zájmu o určitá témata na internetu. Použil nejen GT data, ale i data z Twitteru. Na základě nich pak vytvořil sadu prediktorů, které vhodně upravil v modelu GARCH. Pro krátkodobé prognózy se tento postup ukázal přesnější než jiné porovnávané modely. Ve studii Zhong a Raghib (2019) využili i novou službu Google Correlate a zautomatizovali výběr nové sady vyhledávacích výrazů pro každé obchodní rozhodnutí. Zde je dokonce předpoklad, že tento adaptivní rozhodovací rámec může mít hodnotu nejen pro finanční aplikace, ale také v jiných oblastech výpočetní sociální vědy, kde lze z dat digitální stopy odvodit vazby mezi aspekty kolektivního lidského chování a online vyhledáváním. Ve studii Phama et al. (2024) si autoři kladou za cíl prozkoumat vztah mezi vyhledáváním fintech a výnosy bankovních akcií, což jsou měřítka fintech a bank. Údaje o časových řadách pro fintech a výnosy z bankovních akcií byly získány od společností Google Trends a Vietstock. Klíčovým zjištěním této studie je přítomnost souběžné negativní změny a obousměrné kauzality mezi výnosy bankovních akcií a fintech úvěrů.

Jako jeden z parametrů pro predikci ceny bitcoinů jsou GT data použitá i v článku Fakharchiana (2023). Zde autor navrhl prognostického asistenta zahrnujícího různé významné faktory pro předpovídání ceny bitcoinu s nejvyšší přesností. Zahrnutím i GT dat se mu podařilo získat přesnost i u takto kolísavého aktiva s hodnotou MSE 0,001. Ve studii Apergise et al. (2023) byl zkoumán dynamický přenosový mechanismus mezi zpravodajským sentimentem COVID-19 (Google Trends Index) a indexem S&P100, indexy volatility ropy a zlata. Výsledky jednoznačně potvrdily, že GT data mohou být skutečně důležitým prediktorem ekonomických šoků.

Je tedy zřejmé, že v prognózování poptávky v oblasti financí otevírají GT data velkou možnost dalšího vědeckého výzkumu a tvoří obrovský vědecký potenciál. Prognózování poptávky po jednotlivých produktech, jak je vyčísleno v této monografii, je stále předmětem výzkumu, ale v oblasti financí jsou možnosti výzkumu ještě daleko širší. Zde je také možno zkoumat poptávku po finančních produktech, například životním či neživotním pojištění, a spojit tak logistické prognózování poptávky a finančních veličin.

Nicméně GT data nejsou jen moderním, přesným a spolehlivým prostředkem k prognózování, ale je potřeba podrobit výzkumu také jejich nevýhody. Ty jsou

také již popisovány a přesnost GT dat je někdy i zpochybňována. Ve studii Cebriána a Domenecha (2023) autoři došli k závěru, že nepřesnost údajů GT není zanedbatelná. Ačkoli jeden vědecký článek zpochybňující přesnost GT dat je nezničí jako zdroj dat pro sociální a ekonomické analýzy, je zřejmé, že o rozsahu a determinantech nepřesností je stále jen málo studií a informací. Budoucí výzkumné práce by tedy měly prozkoumat a změřit tyto problémy v široké škále kontextů tak, aby v budoucnu bylo možno s nepřesnostmi adekvátně počítat. Cebrián a Domenech (2023) předpokládají, že nepřesnosti pravděpodobně souvisí s interním procesem, který Google používá k výpočtu, nicméně další relevantní studie na toto téma chybí.

Kapitola 10

Závěr

Prognózování je velmi významným úkolem výzkumu v logistice i v celé podnikové ekonomice. Tento úkol je již řešen množstvím zajímavých publikací a je řada autorů, kteří této oblasti věnují celý svůj vědecký život. Často se ale potýkají s nedostatkem dat, která by mohli pro jakoukoli datovou analýzu využít. Získat kvalitní, a to úplná a relevantní data z podniku je často práce spíše detektivní než vědecká. Vědci přitom často ani nepotřebují informace, které z výsledků vyplynou, ale data na ověření nebo porovnání vyvíjených metodik. Z tohoto pohledu se mohou GT data jevit jako velice vhodná alternativa.

Monografie v první kapitole uvádí prognózování poptávky do souvislosti s logistikou a definuje základní pojmy v ní používané. Kapitola druhá definuje členění metod prognózování poptávky. Monografie se dále zabývá již jen kvantitativními metodami. V této kapitole je také definován postup aplikace metod prognózování.

Kapitola třetí je aplikací tohoto postupu a po definici GT dat už je zde vyčíslena statistická analýza dat a provedena jejich analýza grafická. Je tak metodologickým východiskem pro výpočty v následujících kapitolách.

GT data se dají využít nejen pro vědecké účely, ale i pro individuální prognózování poptávky. Vyhledávání na Google většinou předchází skutečným obchodům. Z tohoto pohledu by informace z prognózování poptávky na základě GT dat mohly být velice cenné. To je zkoumáno v kapitole 6 na datech vyhledávání automobilu Toyota Corolla. Prognóza predikuje mírný nárůst poptávky po celé prognózované období. Tato informace je pro podnik jistě pozitivní.

Další zajímavou možností, jak využít GT data je prognóza kategorií vyhledávání. Tyto kategorie můžeme s trochou zjednodušení považovat za podklad pro vyjádření tržní poptávky po produktu nebo službě. Tato informace může být zajímavá pro start-up firmy či subjekty, které o zahájení podnikání v určité oblasti teprve uvažují. Tomuto tématu je věnována kapitola 7. V ní bylo vyčísleno, že potenciál růstu tržní poptávky mají počítače a elektronika a

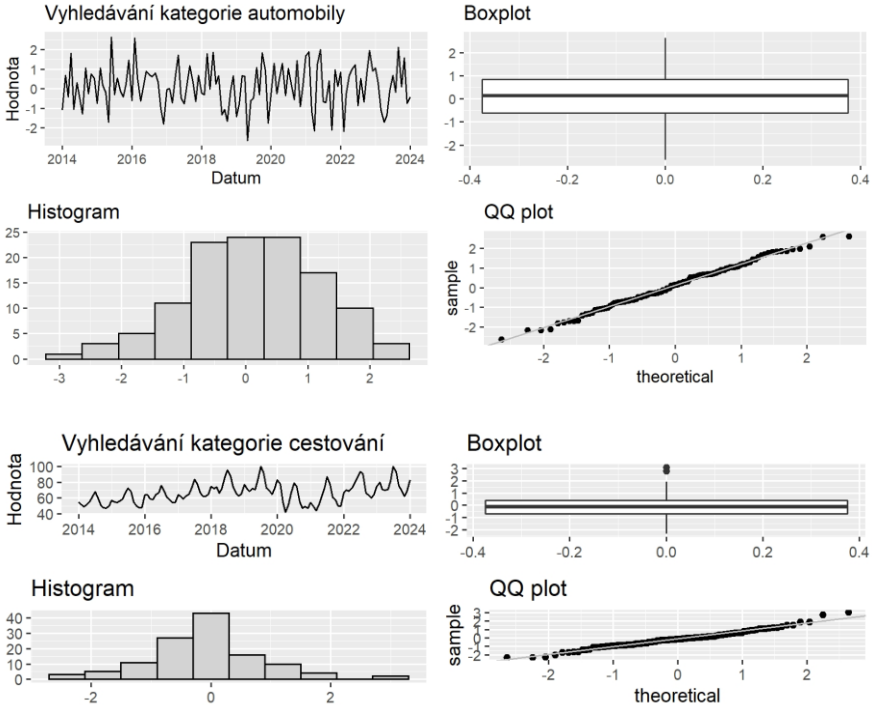
podnikání v cestování. Ostatní kategorie (respektive tržní poptávka) vykazují stagnaci nebo pokles.

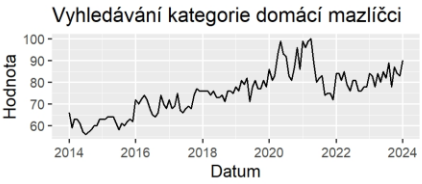
Neméně zajímavá je kapitola 8, která se zabývá analýzou konkurence pomocí prognózy GT dat. Získat konkrétní data konkurence za účelem analýzy je už nadlidský úkol, protože vcelku oprávněně firmy svým konkurentům údaje o svých obchodech běžně neposkytují. GT data by tak mohla tvořit jeden z mála, ne-li jediný způsob, jak potenciální poptávku po konkurenci vyčíslit.

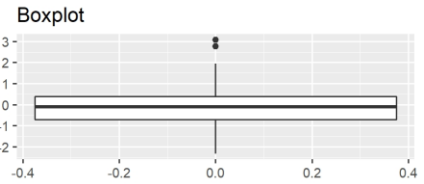
Touto monografií samozřejmě vědecký výzkum autorky nekončí, ale je spíše jejím začátkem. Stále vznikají nové, složitější a přesnější metody. A vědci na celém světě bádají nad vytvořením nejpřesnější metody univerzálně vhodné pro všechny typy poptávky. Nicméně je také nutno řešit problematiku využití takto sofistikovaných metod pro podnikovou praxi. Ani sebedokonalejší a sebepřesnější metoda, pokud není schopna aplikace v běžných provozech a při běžných statistických znalostech podnikatelů a manažerů, nemůže obstát a zůstane jen akademickou záležitostí. V dalším výzkumu je tedy nutno hledat řešení sporu mezi přesností metod a jejich aplikovatelností v praxi.

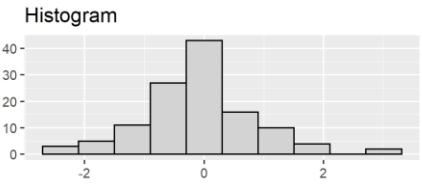
Přílohy

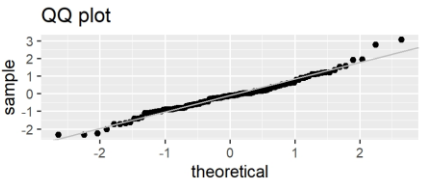
Příloha 1 Popisné grafy GT dat vyhledávání vybraných kategorií

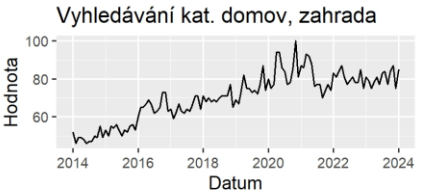


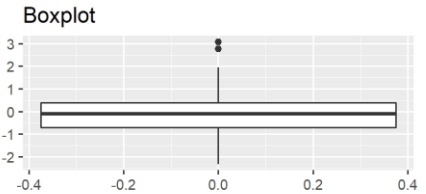

Vyhledávání kategorie domácí mazlíčci
Hodnota
Datum

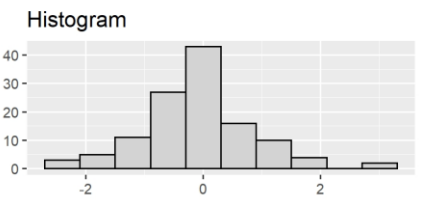

Boxplot

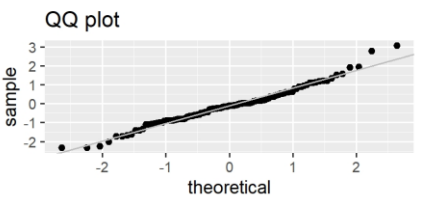

Histogram

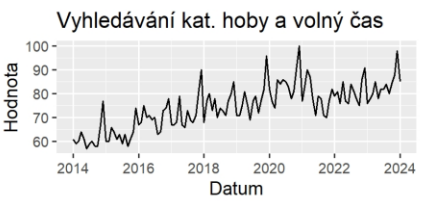

QQ plot
sample
theoretical

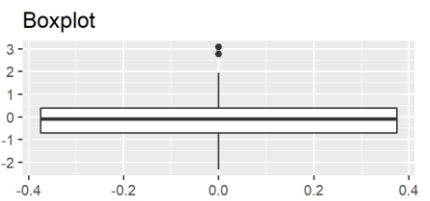

Vyhledávání kat. domov, zahrada
Hodnota
Datum

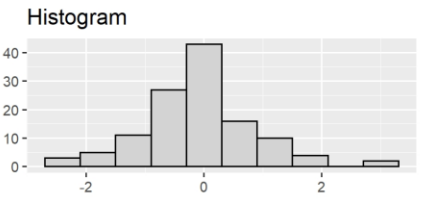

Boxplot

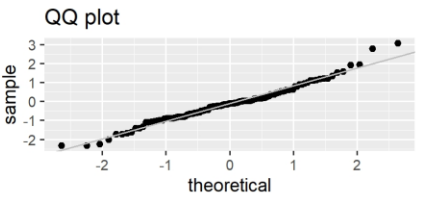

Histogram

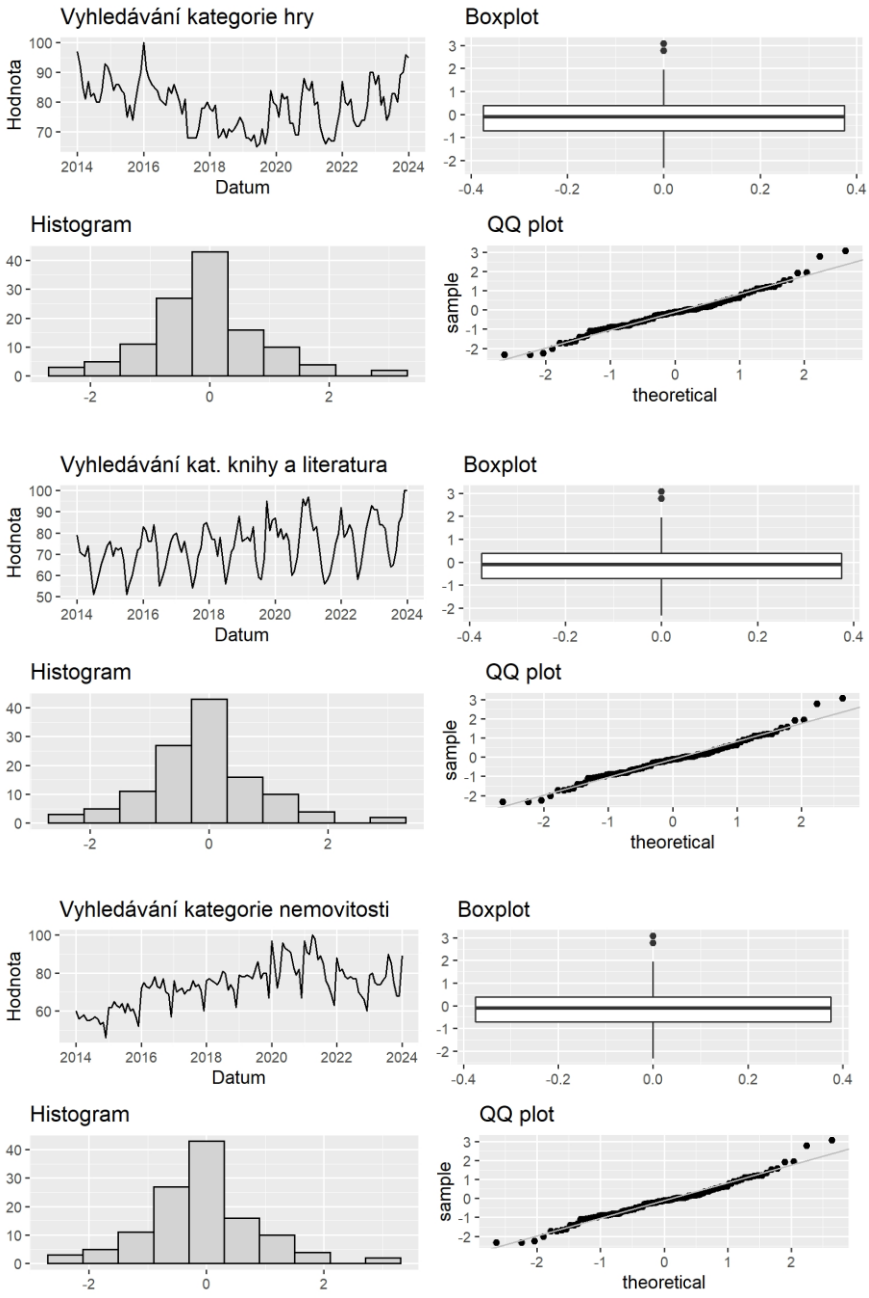

QQ plot
sample
theoretical

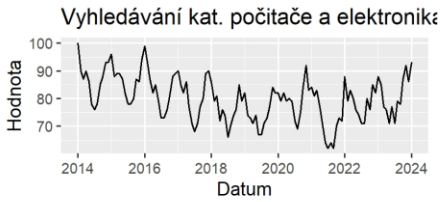

Vyhledávání kat. hoby a volný čas
Hodnota
Datum

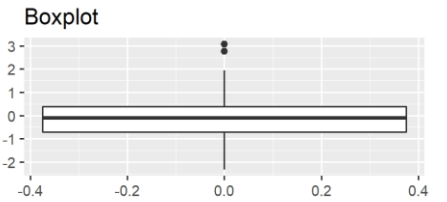

Boxplot

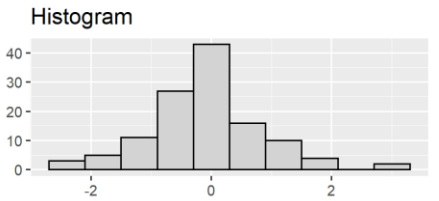

Histogram

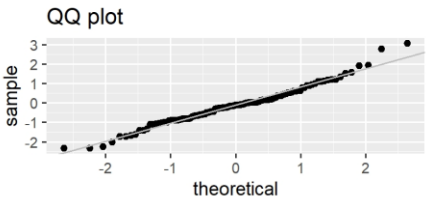

QQ plot
sample
theoretical


Vyhledávání kategorie hry
Hodnota
Datum
Boxplot
Histogram
QQ plot
sample
theoretical
Vyhledávání kat. knihy a literatura
Hodnota
Datum
Boxplot
Histogram
QQ plot
sample
theoretical
Vyhledávání kategorie nemovitosti
Hodnota
Datum
Boxplot
Histogram
QQ plot
sample
theoretical

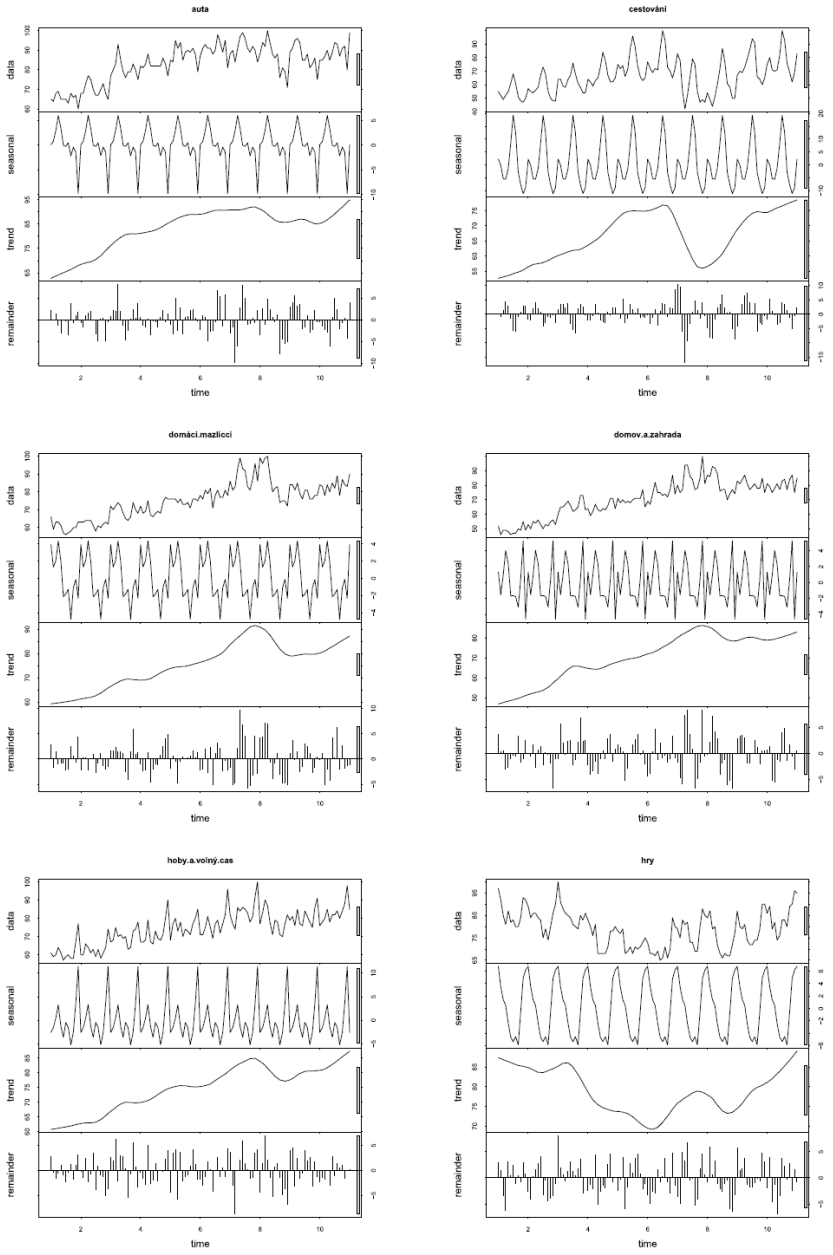

Vyhledávání kat. počitače a elektronika
Hodnota
100
90
80
70
2014
2016
2018
2020
2022
2024
Datum

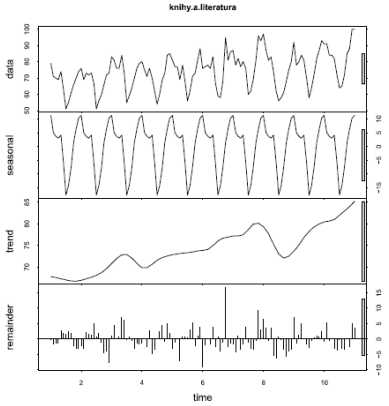

Boxplot
3
2
1
0
-1
-2
-0.4
-0.2
0.0
0.2
0.4

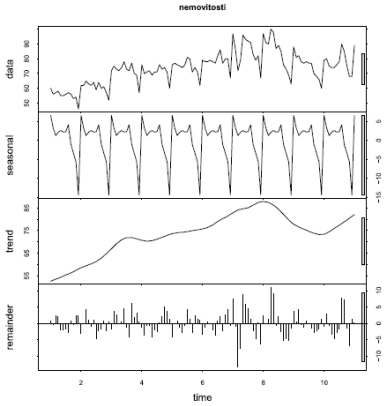

Histogram
40
30
20
10
0
-2
0
2

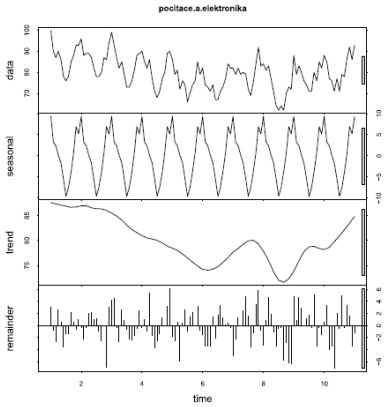

QQ plot
sample
3
2
1
0
-1
-2
-2
-1
0
1
2
theoretical

Příloha 2 Dekompozice GT dat vyhledávání jednotlivých kategorií

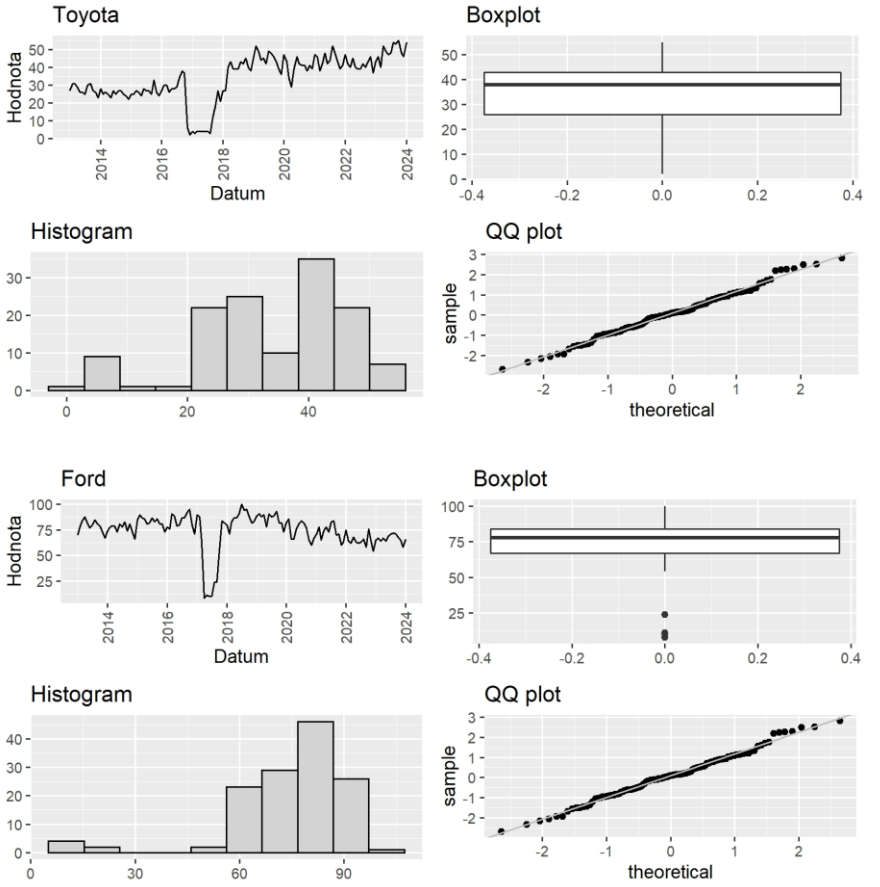


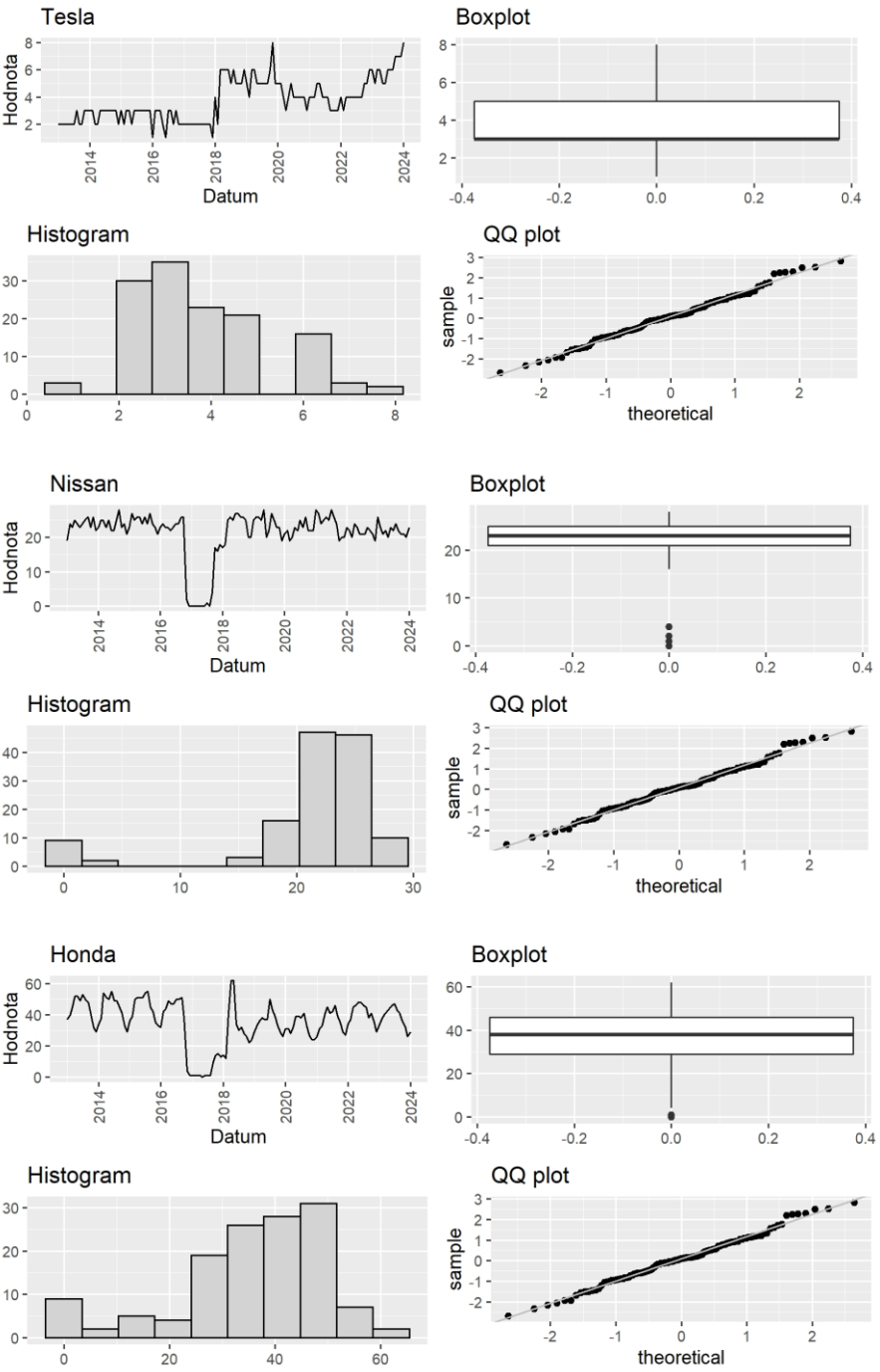

knihy.a.literatura
data
seasonal
trend
remainder
time


nemovitosti
data
seasonal
trend
remainder
time

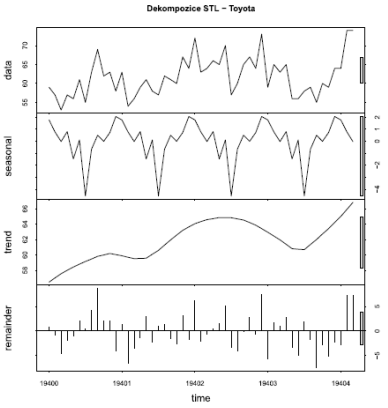

pocitace.a.elektronika
data
seasonal
trend
remainder
time

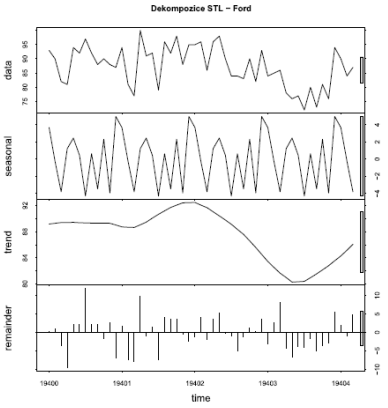
Příloha 3 Popisné grafy GT dat vyhledávání vybraných značek automobilů

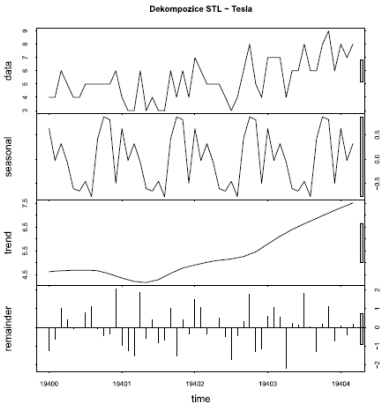


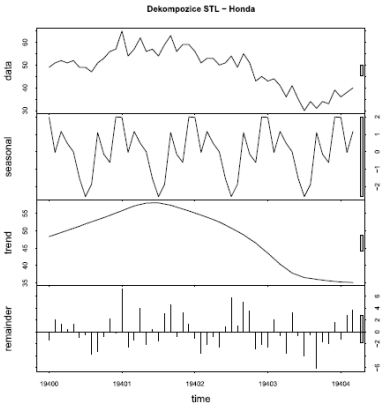

Tesla
Hodnota
Datum
Boxplot
Histogram
QQ plot
sample
theoretical
Nissan
Hodnota
Datum
Boxplot
Histogram
QQ plot
sample
theoretical
Honda
Hodnota
Datum
Boxplot
Histogram
QQ plot
sample
theoretical

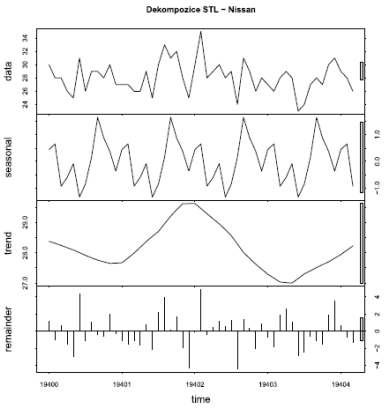
Příloha 4 Dekompozice GT dat vyhledávání jednotlivých značek automobilů









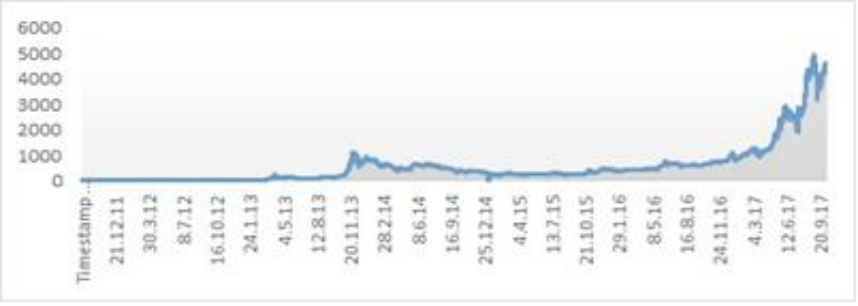


Příloha 5 Vizualizace dat v programu Excel a SPSS

Aplikace moderních přístupů k vizualizaci a deskripci dat v podnikové ekonomice je provedena na datech Airbnb v New York City. Zkoumaný datový soubor obsahuje aktivity v rámci této podnikatelské platformy za rok 2019. K dispozici jsou informace o názvu ubytování a křestním jménu ubytovaného, oblast ubytování, zeměpisná délka a šířka, typ pokoje, jeho cena v dolarech, minimální počet nocí zde strávených, počet recenzí daného ubytovacího zařízení, počty dní v roce, po kterých je ubytování k dispozici. Data jsou získána z portálu kaggle (2019). Celkem je analyzováno 47 905 ubytovacích prostor, 44 % z nich se nachází na Manhattan, 41 % v Brooklynu, 15 % v jiných částech New York City. Jedná se o pronájem celého domu či apartmánu (52 %), soukromého pokoje (46 %) nebo ubytování ve sdíleném pokoji (2 %).

Grafická analýza dat spočívá ve vizualizaci dat pomocí nejrůznějších typů grafů. Některé jsou pro podnikovou ekonomiku vhodnější, jiné nabízené v tabulkových procesorech nebo statistických programech se téměř nepoužívají. V této publikaci jsou použity nejdůležitější typy grafů aplikovaných v podnikové ekonomice.

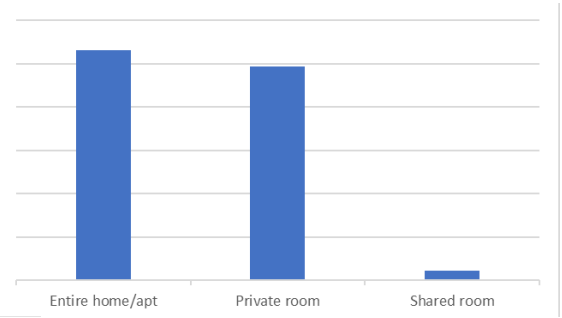
Grafická analýza časových řad obvykle začíná **spojnicovým grafem**. Jeho princip spočívá v jednoduchém zakreslení hodnot časové řady do souřadnicového systému, kde na ose x obvykle bývá proměnná vyjadřující čas. Do tohoto grafu lze vykreslit i jiné proměnné a vizuálně je tak porovnat; v případě, že se dvě časové řady liší měřítkem, je možné použít i pravou vertikální osu, pak jsou však možnosti porovnávání velmi omezené. Vybraná data Airbnb v New York City netvoří časovou řadu. Ukázka spojnicového grafu (obrázek 1) je aplikována na datech hodnoty bitcoinu.



Obrázek 1 Ukázka spojnicového grafu na hodnotě Bitcoinu

zdroj: Kolková, 2018c

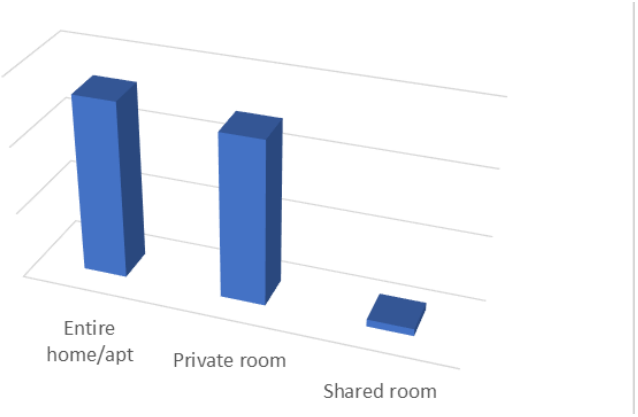
Dalším používaným grafem je **sloupcový graf**. Opět i zde je možno volit celou řadu různých grafických úprav. Často, zejména v populárně vědeckých článcích je volen i graf trojrozměrný. Ten je ale nutno používat velmi opatrně, protože způsobuje celou řadu zkreslení.



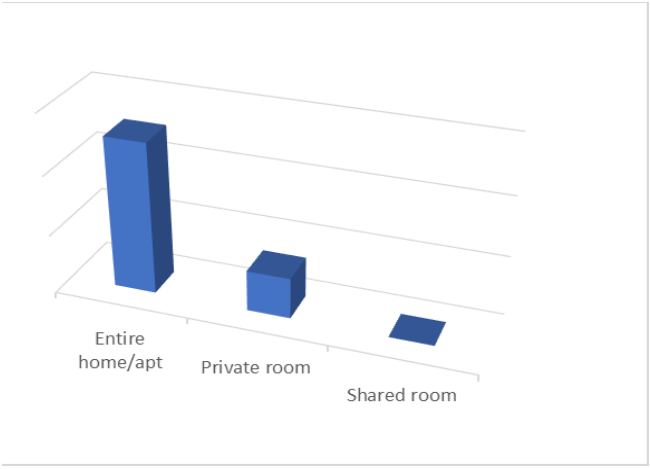
Obrázek 2 Sloupcový graf počtu ubytování v Airbnb New York city dle typu v Excelu

Jako nejvhodnější ze sloupcových grafů se v podnikové ekonomice jeví zřejmě graf sloupcový neprostorový a bez efektů. Obrázek 2 znázorňuje graf počtu různých typů ubytování. Je patrné, že nejvíce si lidé prostřednictvím AIRBN v New York City pronajímali celé domy nebo celé apartmány, soukromé pokoje jsou podle grafu využívány také poměrně často, a to jen s malým rozdílem oproti celým domům a apartmánům. Konečně, sdílené pokoje, kde je možno ubytovat se i s cizími lidmi dohromady, byly zastoupeny výrazně méně.

Na obrázku 3 jsou tatáž data ukázána v prostoru, tj. 3D. Trochu zkosený pohled vyvolá dojem, že celých domů a apartmánů bylo pronajato téměř stejné množství. Jak je na první pohled vidět, vizualizace při použití graficky sice půvabných tvarů může vést ke zkreslujícím závěrům.

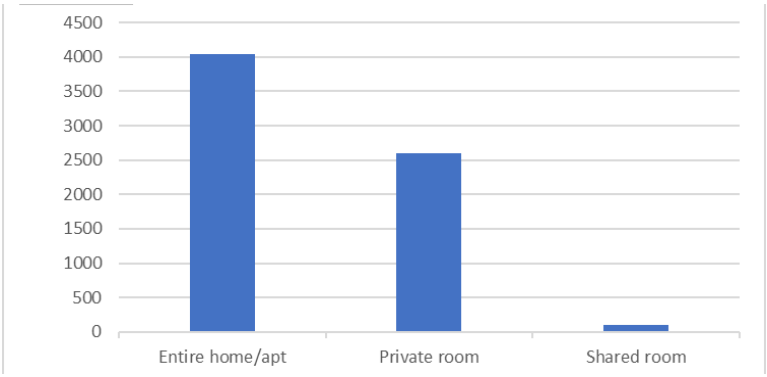


Obrázek 3 Prostorový sloupcový graf počtu ubytování v Airbnb New York city dle typu v Excelu



Obrázek 4 Prostorový sloupcový graf počtu ubytování v Airbnb New York city dle typu v jiném měřítku v Excelu

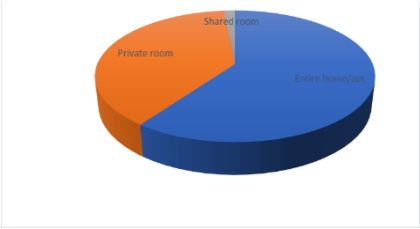
V předchozích grafech také chybí velmi významný prvek, a to hodnoty na ose. Tento záměr lze dokreslit v grafu na obrázku 4, kde počátek osy nebude v 0, ale ve vyšších hodnotách. Je zcela jasné, že bez těchto údajů čtenář z grafu vyčte, že celých domů a apartmánů se v New York City pronajímá výrazně větší množství než jakéhokoli jiného typu ubytování.



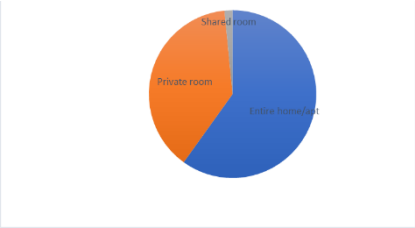
Obrázek 5 Sloupcový graf počtu ubytování v Airbnb New York City dle typu se správným měřítkem v Excelu

Zcela jednoznačné zobrazení poskytuje graf bez prostorových efektů a se správným měřítkem – viz obrázek 5.

Podobného zkreslení lze dosáhnout i využitím **grafů koláčových**. Pomocí zkoseného grafu navodíme pocit, že celých domů a apartmánů je využito výrazně více, jak lze vidět v kontrastu obrázků 6 a 7.

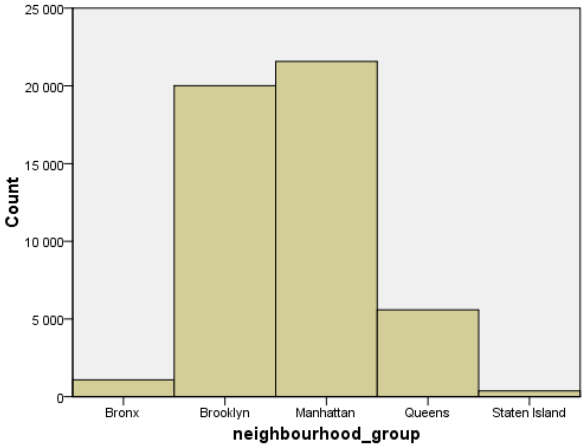


Obrázek 6 Zkosený koláčový graf počtu ubytování v Airbnb New York City dle typu



Obrázek 7 Koláčový graf počtu ubytování v Airbnb New York City dle typu

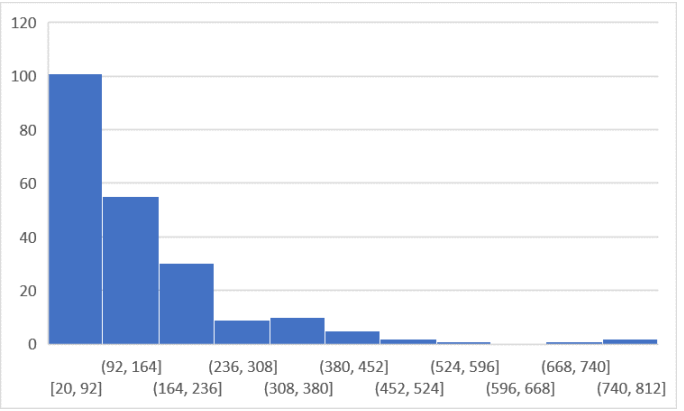
Tyto efekty lze samozřejmě využít k manipulaci s veřejností a míněním zájmových skupin, nicméně v solidním výzkumu tyto manipulace nemají své místo. Ve vědeckém výzkumu jsou zcela neetické.



Obrázek 8 Sloupcový graf umístění ubytování v Airbnb New York City v SPSS

Statistické programy jako SPSS už samy doporučují grafy dle nejvhodnějšího vizuálního stylu, který neumožňuje vizuální zkreslení. Na obrázku 8 je sloupcový graf ubytování v Airbnb New York City dle čtvrti, ve které je ubytování nabízeno. Z grafu jednoznačně vyplývá, že lokality Manhattan a Brooklyn jsou nejžádanější destinace v Airbnb v New York City. Ubytování v Bronxu a Staten Island jsou spíše okrajovou záležitostí.

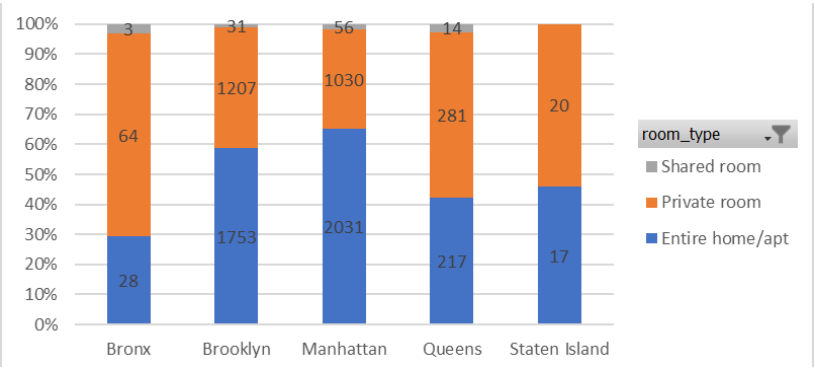
Zejména při zpracování dat pro statistické účely je také často volen **histogram**. Je to grafické znázornění četností pomocí sloupcového grafu u souboru hodnot rozděleného do tříd. V následujícím histogramu na obrázku 9 vidíme opět data Airbnb New York City, tentokrát údaje o ceně; hodnoty jsou v USD za noc.



Obrázek 9 Histogram cen ubytování u Airbnb New York City

Z histogramu na obrázku 9 můžeme vyčíst, že ceny ubytování u Airbnb New York City se nejčastěji pohybují v rozmezí od 20–92 USD. Graf lze také graficky vyjádřit různě a sloupce tak mohou mít nejen tvar obdélníků, ale i hranolů, kuželů atd., a to v nejrůznějších barvách. Opět lze doporučit výraznou obezřetnost při volbě grafického zpracování.

Specifickou možností sloupcových grafů, která nám poskytne spoustu informací, je skládaný sloupcový graf. V následujícím obrázku 10 je znázorněn typ ubytování Airbnb v jednotlivých částech New York City. Tento graf je znázorněn včetně počtu ubytovaných. Lze z něj vyčíst podíl jednotlivých typů ubytování v různých destinacích New Yorku. Například v Bronxu převládá pronajímání pokojů, zatímco na Manhattanu si lidé pronajímají většinou celé domy či apartmány. Nejvíce sdílených pokojů je na Manhattanu, ale v poměru k ostatním ubytovacím možnostem je na Manhattanu tato forma nejméně častá. Také lze vidět, že v Bronxu je pronajímání celých pokojů nebo apartmánů výrazně méně častou variantou než pronajímání soukromých pokojů, což v ostatních částech New Yorku není.

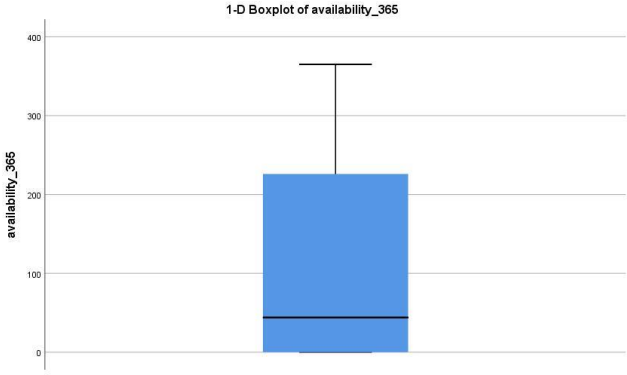


Obrázek 10 Stoprocentní skládaný graf typu ubytování a destinace u AIRBN New York City

Posledním velmi užívaným grafem v moderní grafické analýze podnikových dat je jistě **krabicový graf**. Poprvé byl využit v roce 1977 statistikem J. W. Turkeyem (Lishmanová, 2012). Dnes je možno setkat se s různým označením tohoto grafu: krabička s vousy, vousatá krabička, ve statistických programech se vžil název Boxplot. Konstrukce je založena na pěti základních statistických charakteristikách, a to obvykle mediánu (případně střední hodnotě), horním a dolním kvartilu, minimální a maximální hodnotě souboru dat a bývá doplněna o odlehlá či vzdálená pozorování.

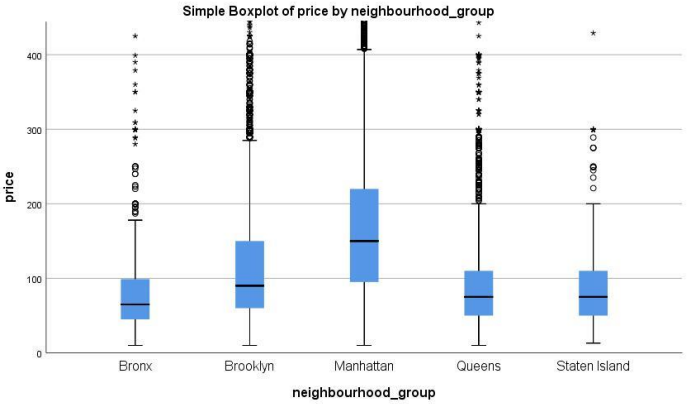
Tento graf je již i součástí programu Excel (od verze Office 2016), nicméně vzhledem k tomu, že jeho praktické využití v tomto programu je prozatím pouze okrajové, v dalším textu je dána přednost statistickým programům SPSS, případně R.

Na obrázku 11 je krabicový graf, na kterém opět analyzujeme data Airbnb New York City. A to tentokrát dostupnost jednotlivých ubytování v průběhu roku. Lze vidět, že zdaleka ne všechna ubytování jsou k dispozici 365 dní v roce, naopak medián je dokonce nižší než 50 dní. Horní kvartil je vyšší než hodnota 200, to znamená, že 75 % typů ubytování jsou dostupná více než 200 dní v roce.



Obrázek 11 Krabicový graf dostupnosti ubytování na Airbnb New York City

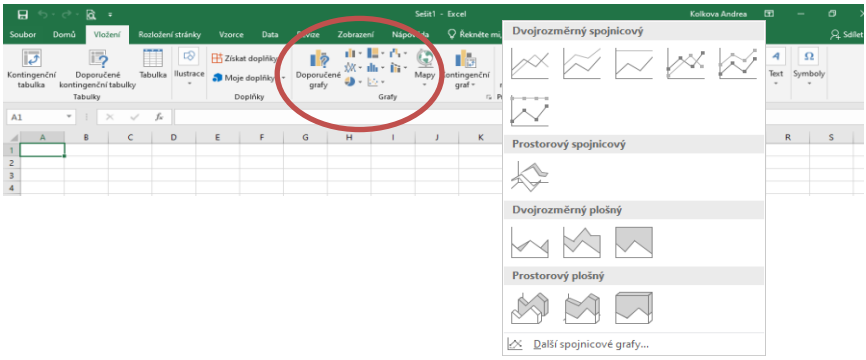
Krabicové grafy můžeme i porovnávat a získat tak další informace ke srovnání. V následujícím grafu na obrázku 12 lze porovnat cenu ubytování Airbnb New York City v jednotlivých částech města na základě krabicového diagramu. Lze vidět, že ceny na Manhattanu daleko převyšují ostatní, nicméně v ostatních destinacích je patrná řada odlehlých pozorování, které svědčí o tom, že i zde je možno narazit na luxusní typy ubytování. Také vidíme, že rozptýlení cen na Manhattanu je nejvyšší ze všech čtvrtí.



Obrázek 12 Krabicový graf cen ubytování na Airbnb New York City dle čtvrtí New Yorku

Příloha 6 Postup při grafické analýze v Excelu a SPSS

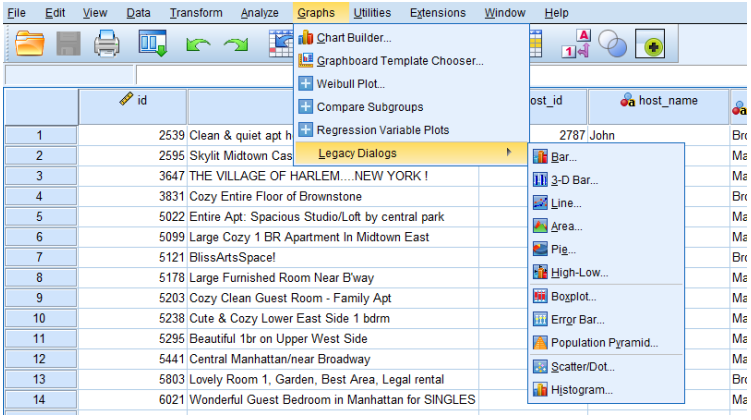
V Excelu je vytvoření grafů velmi intuitivní – v záložce VLOŽENÍ vybereme typ grafu spojnicového dle obrázku 1. Ze stejné nabídky se pak vybírají i další možnosti grafů. Jejich různorodost je dána spíše grafickým cítěním analytika, jejich vypovídací schopnost se změnou příliš nemění.



Obrázek 11 Grafická analýza v Excelu

V SPSS je grafická analýza zřejmě jednodušší záležitostí. V zásadě máme čtyři možnosti, jak vytvořit graf. Všechny nalezneme v základním příkazu Graphs, jak je vidět na obrázku 2:

- Legacy Dialog,
- Chart Builder,
- Craphboard Template Chooser,
- Graf z pivotní tabulky.

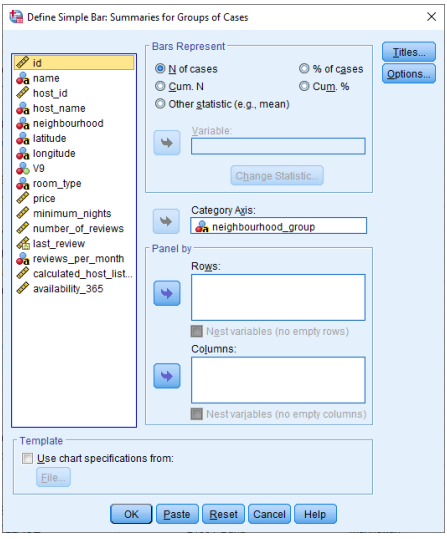


Obrázek 12 Dialogové okno pro výběr možností tvorby grafu v SPSS

Legacy Dialog je příkaz pro zadávání pomocí standardních dialogů, což je asi nejjednodušší způsob umožňující výběr různých typů grafů. Jedná se o standardní součást SPSS. Možnost vytvořit graf tímto způsobem lze zahájit

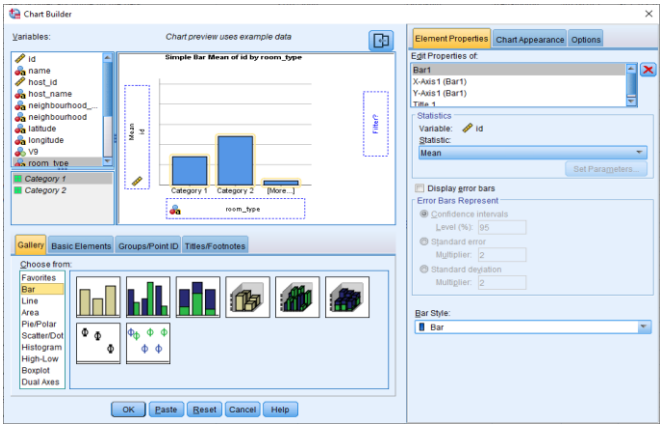
kliknutím na příslušný graf a následně v úvodním dialogovém okně určíme objekty, které bude zobrazovat.

Zde jsou na výběr **Summaries for groups of cases**, kde lze graficky porovnat skupiny, které jsou určené hodnotami kategorizované proměnné, což lze využít zejména pro sloupcový nebo koláčový graf četností. Další volbou je **Summaries of separate variables**, kde graficky vyjadřujeme hodnoty několika proměnných. Poslední jsou **Values of individual cases**, kdy zobrazujeme jednotlivé případy z datové matice. Následně už jen vybereme proměnné, které v grafu chceme mít v jednotlivých kategoriích. Na obrázku 3 je znázorněno vytvoření sloupcového grafu přes Summaries for groups of cases s výběrem kategorie osy polohy ubytování. Další možnosti tvorby grafů tímto způsobem jsou velmi podobné a lze je tvořit intuitivně, uživatelsky velmi přívětivým způsobem.



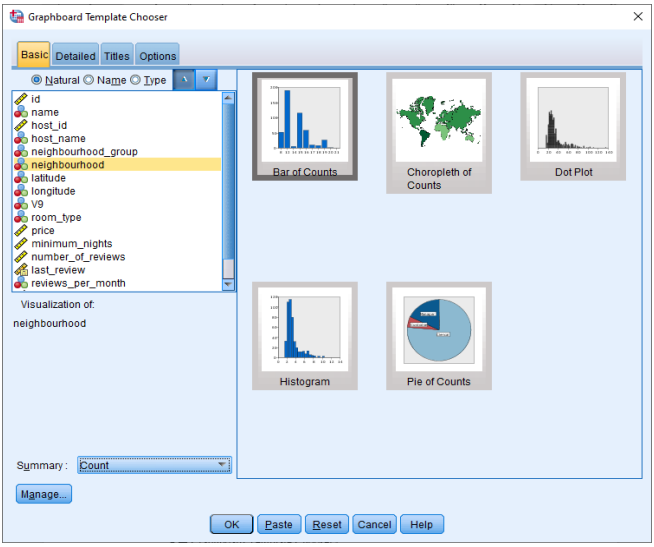
Obrázek 13 Dialogové okno pro vytvoření sloupcového grafu pro Summaries of separate variables pro polohu ubytování v Airbnb New York City

Chart Builder je druhý způsob, jak vytvořit vhodný graf k analýze. Tento nástroj představuje jakési interaktivní rozhraní pro tvorbu grafů. Uprostřed dialogového okna lze vidět náhled budoucího grafu a tažením myší nanášíme na jednotlivé osy, případně do filtru vhodné proměnné. Na obrázku 4 je dialogové okno pro sloupcový graf typu ubytování nabízených v Airbnb New York City. Pomocí tažení myší z nabídek Gallery, Basic Elements můžeme do grafu vložit další prvky jako proměnné, nadpisy, poznámky i typ grafu. V pravé části dialogového okna na záložce Options dokonce můžeme zvolit, jak bude graf reagovat na chybějící hodnoty, které budou dále diskutovány v následujících podkapitolách.



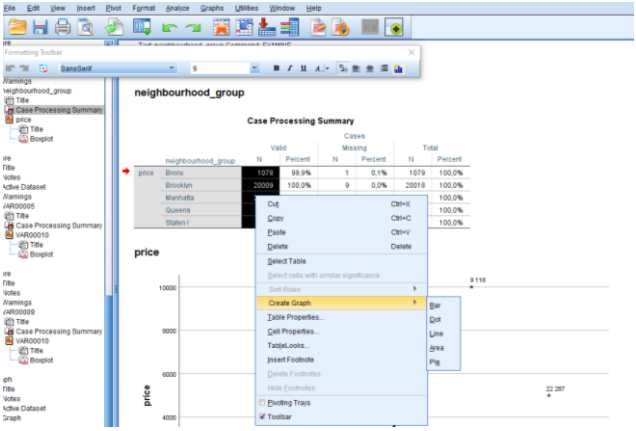
Obrázek 14 Dialogové okno Chart Builder pro vytvoření sloupcového grafu typu ubytování v Airbnb New York City

Graphboard Template Chooser umožňuje tvořit grafy s využitím přednastavených šablon, případně podle vlastních šablon vytvořených v samostatném programu IBM SPSS Visualization Designer. Dialogové okno zde má čtyři části, jak vidíme na obrázku 5, a to Basic, Detailed, Titles a Options V záložce Basic zvolíme proměnné a následně můžeme vybrat šablonu grafů. Záložka Detailed pak slouží k podrobné specifikaci grafu, v Titles můžeme vložit vlastní nadpis a Options použijeme k definici výstupního objektu, změně grafického vzhledu či specifikaci způsobu zacházení s chybějícími hodnotami. Za zmínku stojí také tlačítko Manage, nacházející se na každé záložce, díky kterému můžeme spravovat dostupné šablony a načítat nové.



Obrázek 15 Dialogové okno Graphboard Template Chooser

Graf můžeme vytvářet také ve výstupovém okně přímo z **pivotní tabulky**. Tvorba je opět poměrně jednoduchá, poklepáním na tabulku se dostaneme do editačního modu a po označení hodnot, které chceme graficky zobrazovat, se prostřednictvím pravého tlačítka myši můžeme dostat do Create Graph. Následně už vybíráme graf dle našich potřeb. Ukázka vytvoření grafu z pivotní tabulky pro polohu ubytování v Airbnb New York City nalezneme v obrázku 6.



Obrázek 16 Dialogové okno pro vytvoření grafu z pivotní tabulky pro polohu ubytování v Airbnb New York City

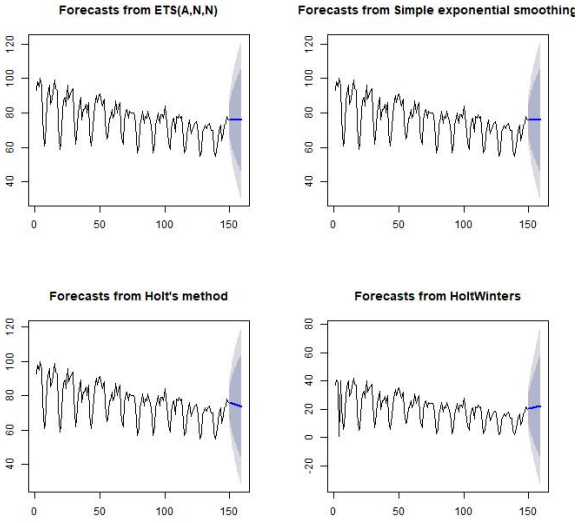
Příloha 7 Grafická analýza v jazyce R

Grafika v R je jednou z velmi propracovaných součástí programu. Mnohé grafické funkce jsou již součástí základní „výbavy“ jazyka R. Nicméně existuje i řada balíčků, které tyto grafické funkce ještě doplňují. A dělají tak z R nástroj pro plnohodnotné grafické vyjádření všech statistických funkcí. Nejpopulárnější balíček pro vizualizaci dat je *ggplot2*. Byl vytvořen v roce 2005 Hadleyem Wickhamem (Wickham, 2009).

Pro základní práci s grafy v R jsou základní funkce mnohdy dostačující. Pro grafické vyjádření prognóz v této publikaci jsou použity kódy *plot*. Velice užitečnou je také funkce *split.screen*. Ta umožňuje v jednom grafickém prvku porovnat i více grafů. Ukázka použití kódu rozdělení grafu na čtyři části je v následujícím boxu:

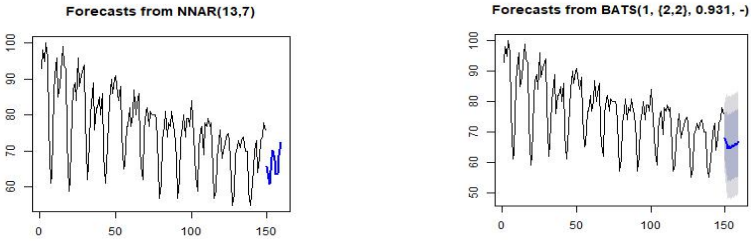
```
split.screen(c(2,2))
screen(1)
plot(forecast(fit1))
screen(2)
plot(forecast(fit2))
screen(3)
plot(forecast(fit3))
screen(4)
plot(forecast(fit4))
```

Výsledkem takového kódu je pak skupina čtyř grafů jednotlivých prognóz. Ukázka je v grafu na obrázku 1. Na tomto grafu je vyjádřena prognóza poptávky po klasické literatuře tak, jak byla prognózována na tréninkových datech, a to datech celosvětových.



Obrázek 1 Ukázka grafického vyjádření prognóz poptávky dle vybraných metod exponenciálního vyrovnání

Pokud bychom chtěli využít skupinu pouze dvou grafů, změnilo by se rozdělení obrazovky *na split.screen(c(1,2))*, případně *split.screen(c(2,1)).* Dle toho bychom mohli grafy seřadit do řádků nebo do sloupců. Obrázek 2 ukazuje první možnost, a to seřazení grafů do jednoho řádku.



Obrázek 2 Ukázka grafického vyjádření prognózy poptávky

Velice užitečnou funkcí pro znázorňování prognóz je pak třeba vložení linie historických nákupů při hodnocení prognózy na tréninkových datech, a to použitím kódu *lines*. Pokud tuto linii chceme například čárkovanou, je možno doplnit příkazem *lty="dashed"*. Ukázka kódu je opět v následujícím boxu:

```
plot(fit1)
lines(nakupy.cr[1:150,1])
lines(nakupy.cr[1:150,1],lty="dashed")
```

To samozřejmě ani zdaleka není celý výčet grafických funkcí, které R nabízí. Pokud bychom chtěli kompletní popis a praktické ukázky všech grafických funkcí R, vydalo by to jistě na další knihu, možná i několik knih. To ovšem není předmětem této publikace.

Příloha 8 Explorativní analýza dat

Funkce používané pro výpočet ukazatelů EDA jsou shrnuty v kategorii statistické, ta však obsahuje i řadu dalších statistických funkcí. Následující tabulka 1 shrnuje funkce v Excelu sloužící nejčastěji k výpočtu základních ukazatelů EDA.

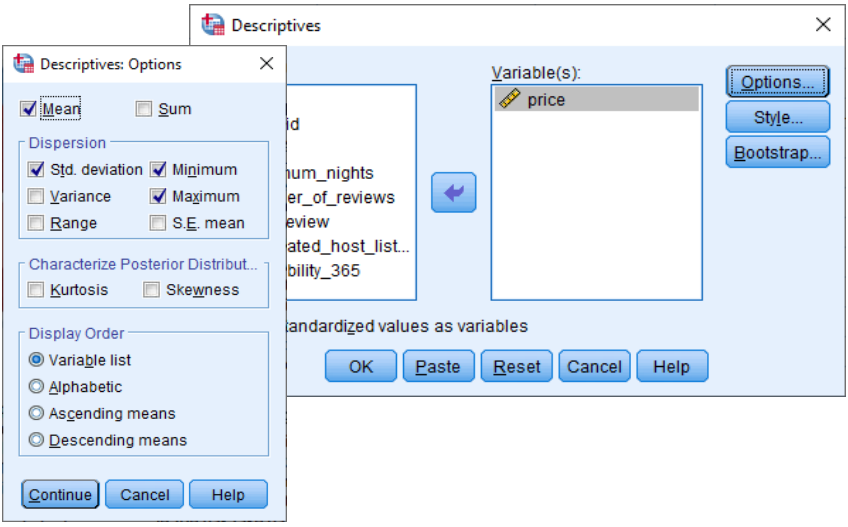
Tabulka 1 Nejčastěji používané funkce pro výpočet EDA v Excelu

Statistická charakteristika		Funkce v Excelu
Míry polohy	střední hodnota	AVERAGEA
	minimum	MIN
	maximum	MAX
	medián	MEDIAN
	modus	MODE.SNGL
	kvartily	QUARTIL (minimální hodnota=0, první kvartil=1, medián=2, třetí kvartil=3, maximální hodnota=4)
Míry variability	rozptyl	VAR.VÝBĚR
	směrodatná odchylka	SMODCH.VÝBĚR
	variační koeficient	SMODCH.VÝBĚR/AVERAGEA
Míry tvaru	šikmost	SKEW
	špičatost	KURT

Následně je nutné údaje vložit do textu či tabulky a vytvořit uživatelsky přívětivý výstup. Pro časovou náročnost tohoto postupu a jeho dnes již velmi malé využití v praxi, zde nebude demonstrován.

Při využití statistického programu SPSS je cesta k EDA poněkud kratší a základní charakteristiky získáme najednou použitím nabídky Analyse – Descriptive statistic – Descriptives. Po výběru proměnné je možno na tlačítku Options vybrat, jaké charakteristiky chceme vypočítat. V následujícím obrázku 1 je zvolen výpočet všech charakteristik.

Z anglických názvů je již patrné, jaké výsledky nám může SPSS poskytnout. Mean je průměr, Sum označuje celkový součet, Std. Deviation je směrodatná odchylka, Variance rozptyl, Range představuje rozpětí, S.E.mean je standardní chyba průměru Kurtosis špičatost a Skewness šikmost. Sekce Display Order je pro určení pořadí, v jakém budou proměnné zobrazeny ve výstupní tabulce.



Obrázek 1 Volba popisných charakteristik EDA v programu SPSS

EDA v programu SPSS byla provedena opět na data Airbnb v New York City, tentokrát jen pro proměnnou *Price* (cena sdíleného ubytování). Výstupem programu SPSS, tak jak byl zadán v obrázku 1, je tabulka 2. Z ní lze vyčíst, že průměrná cena ubytování je 152,79 USD se standardní chybou průměru 1,083. Směrodatná odchylka je 238,831. Z hodnot šikmosti, která dosahuje 19,151, je patrné, že u ceny převažují hodnoty menší než průměr a jsou pozitivně zešikmeny. Z hodnot špičatosti lze vyčíst, že statistické rozložení proměnné cena je velmi zešikmené.

Tabulka 2 Popisná statistika EDA pro Airbnb v New York City

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
price	48618	9990	10	10000	152,79	1,083	238,831	57040,302	19,151	,011	591,879	,022
Valid N (listwise)	48618											

EDA v programu R opět umožňuje vyčíslit všechny charakteristiky najednou. Stačí jednoduchý příkaz *Summary* a základní charakteristiky jsou vyčísleny:

```
summary(data_airbnb$VAR00010)
     price
 Min.   :   10.0
 1st Qu.:   69.0
 Median :  107.0
 Mean   :  152.8
 3rd Qu.:  175.0
 Max.   :10000.0
 NA's   :11
```

Také můžeme vypočítat směrodatnou odchylku, rozptyl nebo jakýkoli kvantil.

```
sd(data_airbnb$VAR00010,na.rm = T)
var(data_airbnb$VAR00010,na.rm = T)
quantile(data_airbnb$VAR00010,prob = 0.3,na.rm = T)
```

Nebo jednoduše proměnnou pojmenujeme price *price<-data_airbnb$VAR00010* a dále už jen zadáváme příkazy s touto proměnnou.

```
summary(price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   10.0    80.0   120.0   168.9   189.0  9999.0
 sd(price)
[1] 286.0369
 var(price)
[1] 81817.1
quantile(price, probs=0.3)
30%
 90
```

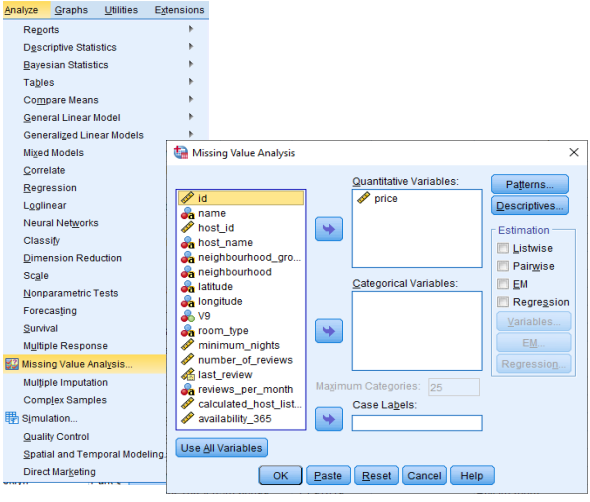
Pro další statistické veličiny, jako například šikmost nebo špičatost, je nutno už využít další balíček pro statistické funkce. nejznámější je *dplyr*. Funkce šikmosti a špičatosti je také součástí balíčku *moments.*

```
skewness(price)
[1] 16.87725
kurtosis(price)
[1] 416.1551
```

Příloha 9 Analýza chybějících a odlehlých hodnot

Vyhledávání a řešení chybějících a odlehlých hodnot v Excelu je již časově velmi náročná záležitost spočívající ve filtrování dat a jejím následném ručním nahrazování. Z toho důvodu to nelze jako efektivní způsob doporučit a ani zde nebude aplikována. Jistým umožněním využití Excelu je snad jen grafická analýza pomocí krabicového diagramu, který umožňuje odlehlá pozorování vynechat nebo ne. Odlehlá pozorování můžeme vidět zaškrtnutím políčka zobrazit vnější body ve formátu datové řady. Možnosti Excelu jsou vyzkoušeny opět na datech ceny sdíleného ubytování. Na obrázku 1 je vidět porovnání krabicového diagramu, když vnější body nebyly zobrazeny, a druhý graf stejných hodnot při zobrazení vnějšího hodnot, samozřejmě je nutno si povšimnout nutnosti změnit měřítko v druhém grafu. Z obou grafů můžeme vyčíst kvartilové rozpětí a jednoznačně tak říct, kde se nachází 50 % všech dat. Také, že 75 % cen ubytování je nižších než 170 USD a naopak 75 % z nich je vyšších než 58,75 USD. Od ceny 20 USD do 325 USD je rozpětí všech cen ubytování s výjimkou několika odlehlých hodnot, tedy individuálních cen přesahujících tuto hodnotu.

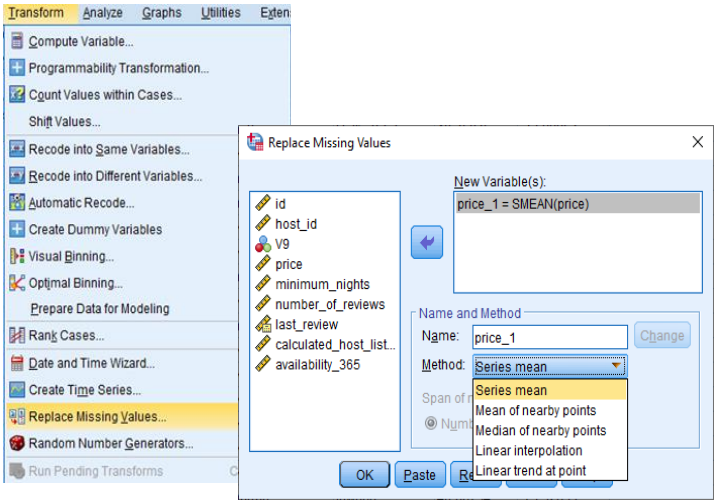
Obrázek 1 Krabicový diagram cen sdíleného ubytování Airbnb v New York City



Obrázek 2 Dialogové okno SPSS pro hledání odlehlých a chybějících hodnot

Práce ve statistickém programu SPSS se zdá být již mnohem efektivnější. Hledání odlehlých a chybějících hodnot je opět umožněno jedním dialogovým oknem, jak deklaruje obrázek 2.

Pokud následně chceme chybějící hodnoty i vyřešit, je postup opět automatizován a na výzkumníkovi je pouze, jaký zvolí způsob nahrazení chybějících dat. Obrázek 3 definuje možnosti, které při řešení máme, a to střední hodnotu řady (series mean), průměr z nejbližších bodů (mean of nearby points), medián z nejbližších bodů (median of nearby points), lineární interpolaci (linear interpolation) a lineární trend v bodu (linear trend at point).



Obrázek 3 Dialogové okno SPSS pro řešení odlehlých a chybějících hodnot

Při vyhledávání chybějících a odlehlých hodnot v SPSS jsou opět použity hodnoty ceny sdíleného ubytování. Výsledky uvádí tabulka 1 a je patrné, že chybějící hodnota se v daných datech vyskytla 11krát, nicméně extrémně vysokých hodnot dosahuje 765krát.

Tabulka 1 Vyhledání odlehlých a chybějících hodnot Airbnb v New York City

Univariate Statistics

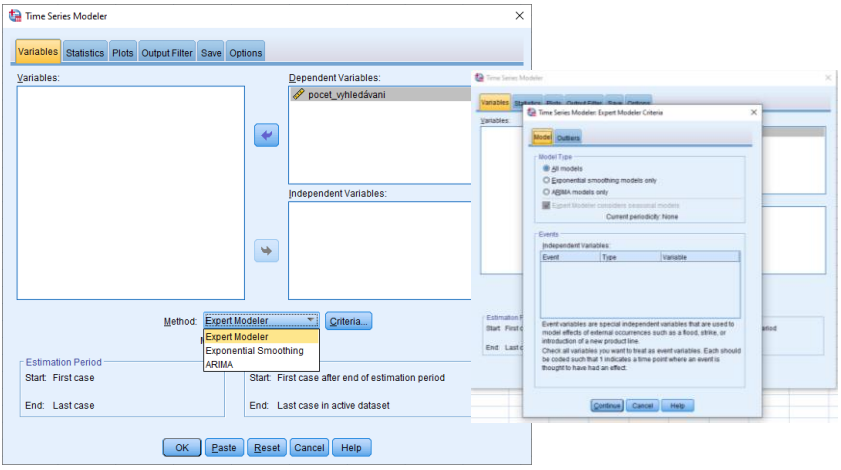
	N	Mean	Std. Deviation	Missing		No. of Extremes[a]	
				Count	Percent	Low	High
price	48618	152,79	238,831	11	,0	0	765

a. Number of cases outside the range (Mean - 2*SD, Mean + 2*SD).

Při detekování chybějících hodnot v jazyce R lze použít jednoduchý příkaz *is.na(price)*, a pokud chybějící hodnoty chceme vynechat, stačí zadat *na.omit(price)*.

Příloha 10 Prognózování v programu SPSS

Nejjednodušším způsobem využití SPSS pro prognózování je v nabídce *Analyse* vybrat *Forecasting* a zvolit *Create Traditional Models*. V této nabídce si můžeme vybrat nástroj *Expert Modeler*. Je velmi výhodný, protože z modelů vybere ten nejlepší a pak jej aplikuje na výslednou prognózu. Samozřejmě můžeme i volit *Exponential Smoothing* a pak konkrétní ETS model nebo ARIMA. Je to využívané zejména, když chceme modely porovnávat a potřebujeme všechny vyčíslit. Pro využití pouze pro prognózování je *Expert modeler* jednodušší a rychlejší volbou. Dále samozřejmě zvolíme proměnnou. Na obrázku 1 je volba proměnné podle počtu vyhledávání pomocí *Expert modeler*.



Obrázek 1 Dialogové okno pro volbu modelu prognózy v SPSS

Výstupem z prognóz je celá řada dat. Prvním z nich je samozřejmě automatické vyhledání nejvhodnějšího modelu z nabízených. V případě našich dat vyhledávání na Google, je to model ARIMA s následujícími parametry v tabulce 1.

Tabulka 1 Popis modelu ARIMA pro prognózování počtu vyhledávání pojmu "klasická literatura"

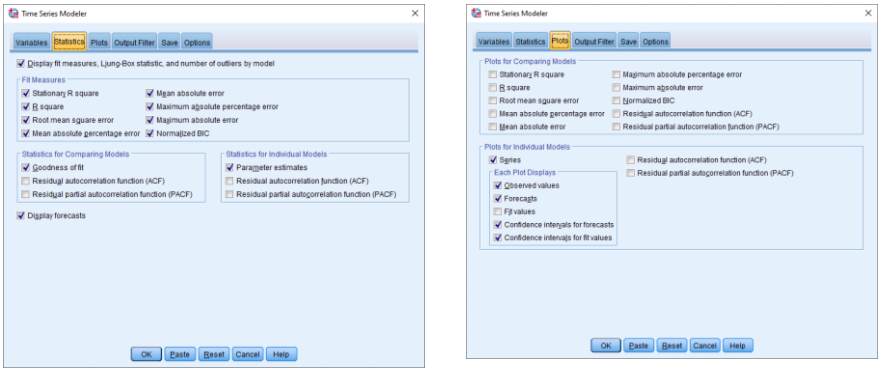
Model Description

	Model Type
pocet_vyhledávani	ARIMA(0,0,12)

ARIMA Model Parameters

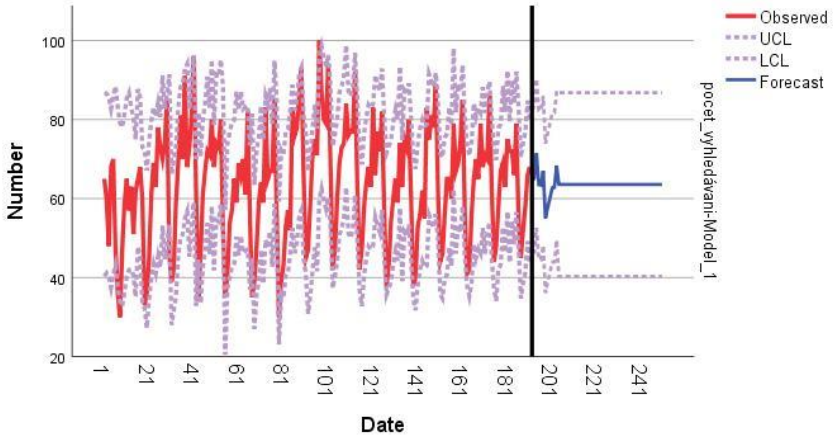
			Estimate	SE	t	Sig.
pocet_vyhledávani	Constant		63,553	1,441	44,090	,000
	MA	Lag 1	-,314	,060	-5,192	,000
		Lag 2	-,188	,064	-2,925	,004
		Lag 12	-,779	,094	-8,333	,000

Samozřejmě dalším požadavkem je zjistit míru přesnosti tohoto modelu. SPSS opět nabízí řadu ukazatelů. Ty budou diskutovány v následujících kapitolách. Stejně lze i vybrat grafy, které chceme automaticky zobrazit. Pro názornost byly požadovány výpočty velkého množství ukazatelů měr přesnosti a grafy, jak ukazuje obrázek 2.



Obrázek 2 Dialogové okno výběru ukazatelů měr přesnosti a grafů v SPSS

SPSS nabízí samozřejmě ještě řadu dalších výstupů, jako jsou například predikované hodnoty. I v grafických výstupech nabízí celou řadu možností. Na obrázku 3 je prognóza vývoje počtu vyhledávání na následujících 50 měsíců (což je záměrně přehnaná délka horizontu, zvolená pouze z důvodu lepší vizuální orientace).



Obrázek 3 Graf prognózy vývoje vyhledávání klíčového slova "klasická literatura" v SPSS.

Příloha 11 Prognózování v jazyce R

V této příloze je proveden příklad výpočtu v jazyce R modelu ARIMA a ETS. Aplikace bude provedena na stejných datech „literature“. Nejprve je nutno naimportovat si datový soubor a pojmenovat proměnnou, v našem případě je pojmenovaná „literature“. Tyto procesy jsou založeny na základních znalostech jazyka R, proto zde nebudou vypisována. Samotné prognózování se zjednodušuje na jeden jediný příkaz, a to u výpočtu metod na základě exponenciálního vyrovnání:

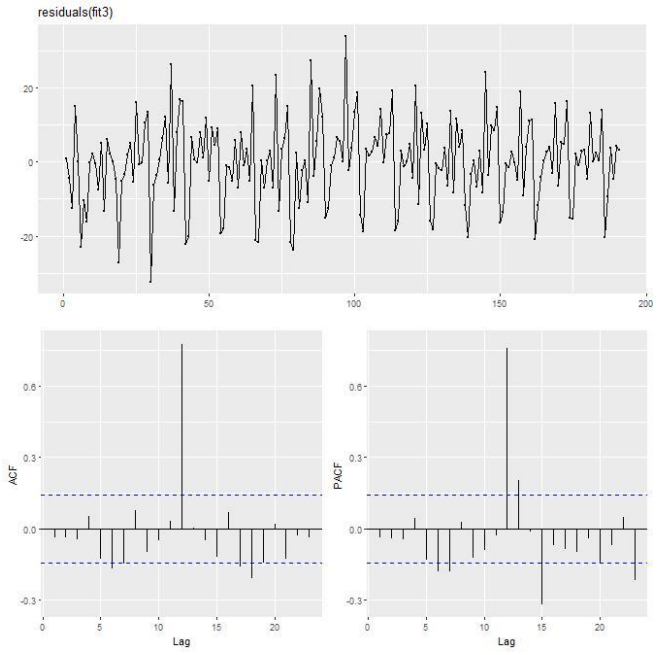
```
fit<-ets(literature)
```

Samozřejmě je žádoucí vidět výsledky, a tak jednoduchým příkazem *summary(fit)* si je můžeme prohlédnout. Výsledek je pak v R viditelný v tomto formátu:

```
ETS(M,A,N)

  Smoothing parameters:
    alpha = 0,9999
    beta  = 0,0023

  Initial states:
    l = 64,9757
    b = 3,3455

  sigma:  0,1993

     AIC     AICc      BIC
1989.593 1989.917 2005.854


Training set error measures:
                 ME     RMSE      MAE       MPE     MAPE      MASE
Training set -2,68074 13,80168 10,44897 -7,153848 18,35657
0.9632726
                    ACF1
Training set -0,04417568
```

Pojmy jako ME, RMSE atd. jsou výsledkem funkce *summary*, ale jejich vysvětlení je detailně diskutováno v kapitole 5, míry přesnosti zde jsou uvedeny jen pro úplnost. Podle potřeby je samozřejmě vhodné dané údaje upravit a vložit například do tabulky. Pomocí funkce *autoplot(forecast(fit))* pak můžeme vykreslit graf prognózy počtu vyhledávání pojmu klasická literatura na vyhledávačích Google.

Při prognóze pomocí metod ARIMA v jazyce R opět využijeme balíček *forecast*. Nabízí v zásadě dvě cesty k prognóze ARIMA modelů, a to využití automatizovaného algoritmu pro výběr nejlepšího ARIMA modelu, a to funkce *auto.arima*, nebo výběr modelu samostatně na základě výpočtů ACF/PACF k výběru možných modelů a následně výpočtu AIS_c pro výběr toho nejlepšího; graf takto vypočítaných ukazatelů je vidět na obrázku 1.



Obrázek 1 Výpočty ACF/PACF k možnému výběru nejvhodnějších modelů.

Pro podnikové účely je samozřejmě žádoucí co největší automatizace výpočtů, jelikož podniku obvykle nejde o vyčíslení statistických veličin všech možných modelů, ale o praktické využití pro prognózu. Proto i v této publikaci budeme postupovat s co největší mírou automatizace. Výpočet prognózy na základě automatického výběru nejvhodnějšího ARIMA modelu je zjednodušen na volání:

```
fit<-auto.arima(literature)
```

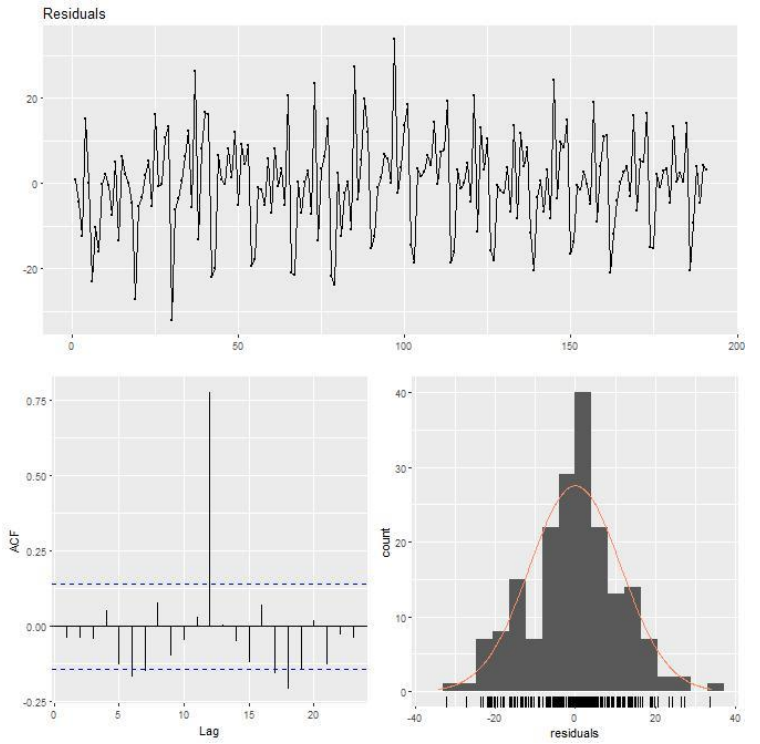
Následně si výsledky opět zviditelníme pomocí *summary(fit)* a výsledkem jsou následující veličiny:

```
ARIMA(0,0,2) with non-zero mean
Coefficients:
         ma1     ma2     mean
      0,7531  0,4228  63,7474
s.e.  0,0700  0,0682   1,7875

sigma^2 estimated as 131,9:  log likelihood=-736,14
AIC=1480,28   AICc=1480,49   BIC=1493,28

Training set error measures:
                      ME     RMSE     MAE       MPE     MAPE
MASE
Training set -0,01317267 11,39636 8,68512 -3,53339 14,69985
0.800666
                    ACF1
Training set -0,03838939
```

Přestože již víme, že je tento model pro praktické využití nejvhodnější, vytvoříme graf reziduí a pomocí jazyka R vypočítáme i příslušné testy. Což se opět zjednoduší na jeden příkaz:

```
checkresiduals(fit)
```

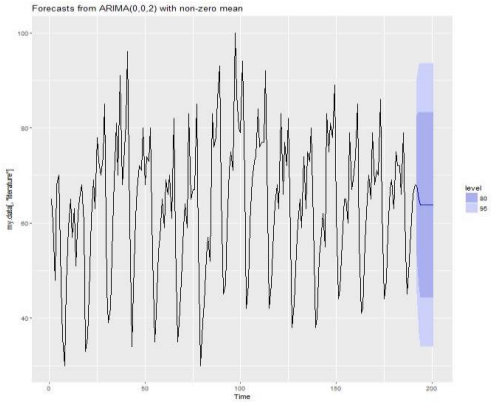
Výsledky tohoto příkazu jsou v grafu na obrázku 2 a následujícím textu.



Obrázek 2 Graf reziduí při prognóze vyhledávání pojmu klasická literatura pomocí jazyka R

```
Results of Hypothesis Test
--------------------------
Alternative Hypothesis:

Test Name:                          Ljung-Box test

Data:                               residuals

Test Statistic:                     Q* = 18,32377

Test Statistic Parameter:           df = 7

P-value:                            0,01059156

Model df: 3.   Total lags used: 10
```

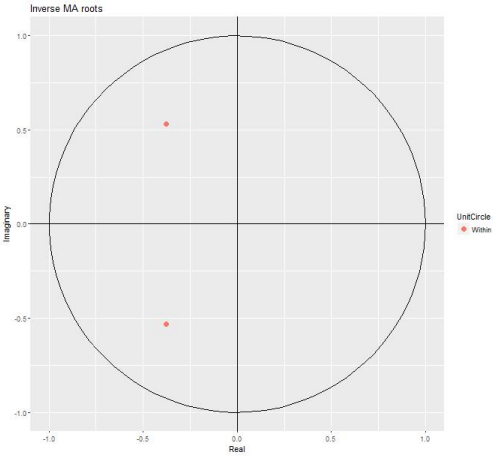
Nyní můžeme přistoupit k samotnému grafickému znázornění prognózy pomocí ARIMA modelu. Příkaz je stejný jako u ETS modelů, a to *autoplot(forecast(fit))*, výsledek je vidět na obrázku 3.



Obrázek 3 Grafické znázornění prognózy ARIMA modelu v jazyce R

Jazyk R také umožňuje vykreslení charakteristických kořenů modelů pomocí *autoplot(fit).* V našem případě jsou všechny kořeny uvnitř kruhu jednotky, jak ukazuje graf na obrázku 4. R již zajišťuje, že vybraný model je jak stacionární, tak invertní.

Funkce *Arima* nikdy nevrátí model s inverzními kořeny mimo kruh jednotky. Funkce *auto.arima* je ještě přísnější a nevybere ani model s kořeny v blízkosti kruhu jednotky. To je ale již spíše součástí pokročilé statistiky a pro podnikové účely postačí vědomí, že balíček *forecast* je naprogramován správně, bez nutnosti to ověřovat.



Obrázek 4 Vykreslení charakteristických kořenů modelů, a to pomocí autoplot(fit) v jazyce R

Příloha 12 Míry přesnosti v SPSS

Pokud budeme v programu SPSS hodnotit model predikce poptávky knih na základě vyhledávání pojmu klasická literatura, který byl využit v kapitole dekompozice, lze použít celou řadu měr přesnosti. Zde je aplikován výpočet měr přesnosti na model v ex-post predikci, respektive na tréninková data. SPSS umožňuje ve standardním nastavení výpočet statistik:

- Stacionární R^2,
- R^2,
- RMSE,
- MAPE,
- MAE,
- MaxAPE,
- MaxAE,
- Normalizované BIC,
- Ljung–Box statistiky.

Tabulka 1 udává výsledky při použití všech dostupných měr přesnosti. Tyto tabulky znázorňují míru přesnosti prognózy poptávky fiktivních dat za použití modelu ARIMA(0,0,12), a to pouze na tréninkových datech.

Tabulka 1 Výsledky měr přesnosti v programu SPSS

Model Statistics

			Model Fit statistics							Ljung-Box Q(18)			
Model	Number of Predictors	Stationary R-squared	R-squared	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normalized BIC	Statistics	DF	Sig.	Number of Outliers
pocet_vyhledávani-Model_1	0	,584	,584	9,490	12,811	7,138	88,175	29,175	4,610	122,967	15	,000	0

Výsledky však můžeme ještě zpřesnit přidáním nejen středních hodnot měr přesnosti, ale dalších statistik, jako jsou minimum, maximum a percentilové charakteristiky uvedené v tabulce 2.

Tabulka 2 Výsledky měr přesnosti v programu SPSS – detailní vyjádření

Model Fit

					Percentile						
Fit Statistic	Mean	SE	Minimum	Maximum	5	10	25	50	75	90	95
Stationary R-squared	,584	.	,584	,584	,584	,584	,584	,584	,584	,584	,584
R-squared	,584	.	,584	,584	,584	,584	,584	,584	,584	,584	,584
RMSE	9,490	.	9,490	9,490	9,490	9,490	9,490	9,490	9,490	9,490	9,490
MAPE	12,811	.	12,811	12,811	12,811	12,811	12,811	12,811	12,811	12,811	12,811
MaxAPE	88,175	.	88,175	88,175	88,175	88,175	88,175	88,175	88,175	88,175	88,175
MAE	7,138	.	7,138	7,138	7,138	7,138	7,138	7,138	7,138	7,138	7,138
MaxAE	29,175	.	29,175	29,175	29,175	29,175	29,175	29,175	29,175	29,175	29,175
Normalized BIC	4,610	.	4,610	4,610	4,610	4,610	4,610	4,610	4,610	4,610	4,610

Příloha 13 Míry přesnosti v jazyce R

V jazyce R je vhodné opět využít balíčku f*orecast*, i když míry přesnosti nabízejí i jiné. Samotné vyčíslení měr přesnosti je ještě jednodušší a jednoduchým příkazem *accuracy* vyčíslíme míry přesnosti, které tento balíček v základu nabízí. A to jsou:

- ME,
- RMSE,
- MAE,
- MPE,
- MAPE,
- MASE.

Dále také statistiky:

- AIC, Akaikeova informačního kritéria
- AICc,
- BIC,
- ACF1.

Příklad kódu pro výpočet měr přesnosti modelu ETS (exponenciální vyrovnání – viz předchozí kapitoly) je v následujícím boxu. Je zde opět využito fiktivních dat „literature“, tedy data použitá i v předchozí kapitole.

```
fit<-ets(literature[1:200,1], h=100)
accuracy(fit)
accuracy(fit,literature[201:300,1])
```

Výsledkem pak je jednak míra přesnosti jak v tréninkových datech, tak v testovacích. Výsledky mohou být různé, pro tréninková data samozřejmě vyjde přesnost vyšší, naopak u dat testovacích bude přesnost nižší, pak ji porovnáváme s jinými modely k určení míry přesnosti daného prognostického modelu.

V následující tabulce 1 vidíme rozdíly v přesnosti dat testovacích a tréninkových na ilustračním příkladu výpočtu měr přesnosti při výpočtu prognózy na akciovém trhu. Je zde patrné, že na testovacích datech prognóza vykazuje vyšší chybu než na datech tréninkových.

Tabulka 1 Míry přesnosti pro data tréninková a testovací

	ME	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**
Tréninková data	0,4393	15,8461	8,9596	0,0213	0,5534	1,0000
Testovací data	0,8900	78,1811	63,3112	-0,1676	3,7897	7,0663

Příloha 14 Moderní statistické programy využívané při prognózování a řízení v podnikové ekonomice

Prognózování je možno výrazně zjednodušit pomocí moderních statistických programů. Základní metody prognózování umožňují i jednodušší programy jako Excel nebo SPSS. Zde ovšem narážíme na omezení možných metod výpočtu. Ty se zaměřují pouze na jednoduché statistické modely, které je již dnes obvyklé doplňovat a kombinovat s jinými. V podnicích, kde prognostické metody používají v praxi, obvykle Excel není plně dostačujícím programem, který by zajistil vhodnou prognózu v přijatelném čase a s přijatelnou náročností. Nyní je na trhu k dispozici řada softwarů, které je možno označit za statistické.

Existující systémy pro prediktivní analýzu jsou dodávány například společnostmi SAS (SAS Enterprise Miner), SPSS (Insightful Miner), KXEN (KXEN Analytic Framework), IBM (DB2 Intelligent Miner for Data), Dynamic Future (PREWIT), Economic Wizard (expertní systém Forecast Wizard). Dle Cipra (2013) dalšími může být i EViews (od dodavatele QMS Software, Inc.), GAUSS (Aptech Systems, Inc.), LIMDEP (Econometric Software), Mathematika (Wolfram Research, Inc.), MatLab (MathWorks, Inc.), RATS (Estima), SHAZAM (Dept.of Economics, Univesity of British Columbia) a TSP (TSP International), R (Free Software Foundation, Inc.), Stata (StataCorp).

V praxi jsou nabízeny i mnohé firemní produkty usnadňující predikce firmám, tyto modely a firmy byly již definovány v kapitole 6. Často jsou však placené a pro základní výzkum nevhodné zejména pro nezveřejňování postupů s kódy použitými v modelech. Samozřejmě, v množství programů, které trh nabízí, jsou určití lídři, kteří tvoří základní rámec používaných statistických programů. Také technologie použité v modelech procházejí vývojem. Z historického hlediska jsou nejpoužívanější technologie zhruba řazeny:

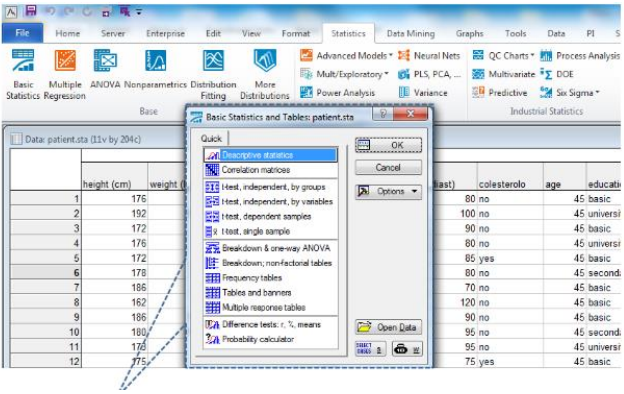
- 1995 SPSS,
- 2003 SPSS Modeler,
- 2013 R,
- 2016 Python,
- 2018 Deep Learning a nástroje umožňující jejich aplikaci,
- Od roku 2020 se pak do popředí dostává systémy AutoML, kdy už jde o řízenou umělou inteligenci.

V následujících kapitolách jsou popsány nejčastěji využívané anebo nejzajímavější a nejnovější statistické softwary. Samozřejmě se nejedná o dokonalý celkový přehled statistických produktů, které jsou dnes nabízeny.

Statistica

STATISTICA je komplexní systém obsahující prostředky pro správu dat, jejich analýzu, vizualizaci a vývoj uživatelských aplikací. Poskytuje široký výběr základních i pokročilých technik speciálně vyvinutých pro podnikání, vytěžování dat, vědu a inženýrské aplikace. Statistica je jedním z nejvýznamnějších konkurentů SPSS.

Její velkou výhodou je uživatelská přívětivost pro uživatele MS Office, protože z velké části kopíruje jeho panely nástrojů. Také je mírně levnější než SPSS. Samozřejmě díky menšímu rozšíření v ČR je jeho hlavní nevýhodou nižší možnost sdílet práci. Ukázku uživatelského rozhraní lze vidět na obrázku 1.



Obrázek 1 Ukázka uživatelského prostředí programu Statistica

Zdroj: statsoft.cz

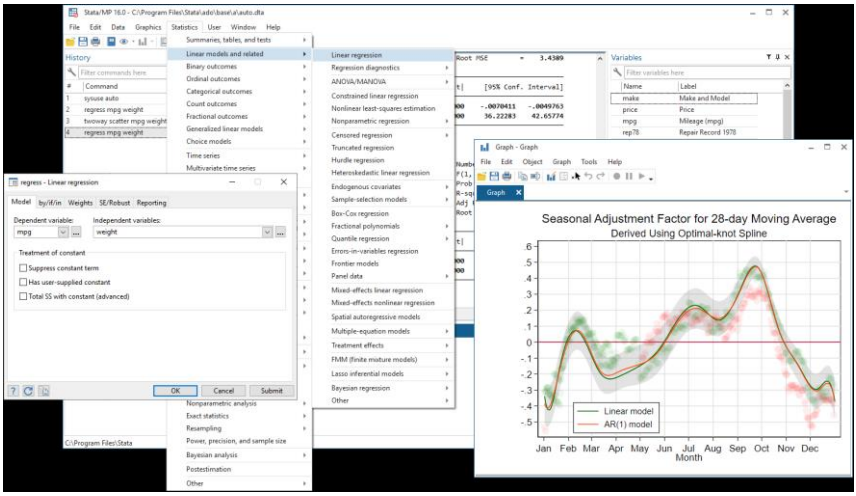
SPSS

SPSS je statistický a analytický nástroj pro univerzální použití. Název původně znamenal *Statistical Package for the Social Sciences* a jeho první uvedení na trh je datováno už v roce 1968. Je to název americké softwarové firmy sídlící v Chicagu. Od roku 2009 ji vlastní společnost IBM, proto je dnes software označovaný jako IBM SPSS**.** V letech 2009 až 2010 byl znám pod názvem PASW (*Predictive Analytics SoftWare*), před rokem 2009 jen SPSS. Tento produkt můžeme označit jako vládce trhu se statistickým softwarem a využívají ho jak téměř všechny vysoké školy, tak i řada podniků.

Software je určen jak pro zkušenější, tak pro začínající statistiky, jeho ovládání je poměrně intuitivní a je založeno na modulárním řešení. Uživatel si tak může vytvořit sestavu nejvhodnější pro řešení svých úloh. Jeho hlavní nevýhodou je samozřejmě poněkud vyšší cena a také velké nároky na hardware. Uživatelské rozhraní je představeno v mnoha ukázkách v přílohách této publikace.

Stata

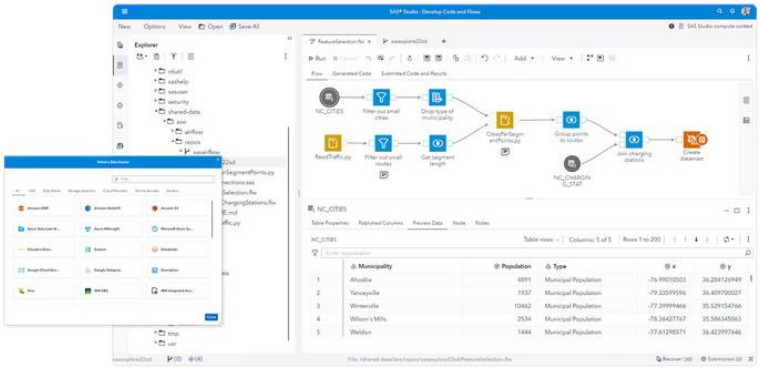
Stata je statistický softwarový balíček pro obecné použití. Uživatelské rozhraní je znázorněno na obrázku 2. Tento software byl vytvořený v roce 1985 společností StataCorp. Většina jeho uživatelů jsou výzkumní pracovníci zejména v oblasti ekonomiky, sociologie, politických věd, biomedicíny a epidemiologie. Funkce Stata zahrnují správu dat, statistickou analýzu, grafiku, simulaci, regresi a vlastní programování. Má také systém šíření uživatelsky psaných programů, který umožňuje nepřetržitý růst. Název Stata vychází ze zkratky slov „statistika“ a „data“.



Obrázek 2 Ukázka uživatelského prostředí programu STATA
Zdroj: stata.com

SAS Systém

SAS System nebo jen SAS (původně *Statistical Analysis System*) je integrovaný systém softwarových produktů, vyráběný firmou SAS Institute. Slouží jednak ve firmách jako databázový systém, jednak jako nástroj pro analýzu a obchodní využití dat a jednak pro statistickou analýzu dat ve vědě a technologii. Jedná se o modulární software, takže zákazník může využít jen ty části, které se hodí jemu. SAS obsahuje vlastní programovací jazyk, označovaný rovněž jako SAS. Obrázek 3 znázorňuje uživatelské rozhraní systému SAS.



Obrázek 3 Ukázka uživatelského rozhraní programu SAS systém

Zdroj: support.sas.com

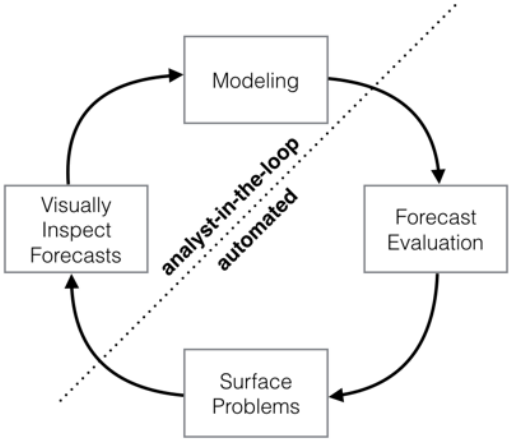
Jazyk R

R je označován spíše jako výpočetní prostředí než statistický software a díky programovacím možnostem s ním lze vypočítat i to, na co běžné funkce klasických statistických software nestačí. Hlavní výhodou je samozřejmě jeho nulová cena. Jazyk R využívá balíčků, které k tomuto programu poskytují vědci i odborníci z praxe zdarma v centrálním repositáři CRAN. Možnosti využití tohoto programu tak neustále rostou. Pro verzi 3.6.2 je těchto balíčků, dle údajů z dubna roku 2023, k dispozici 19 352.

Balíčků, kterými je možno podpořit prognózování v moderních statistických programech jako R, je také několik. Jedním z nich je například známý *forecast* od Hyndmana a Khandakara (2008). Dalšími mohou být *bfast*, který vznikl v roce 2014 na základě výzkumu Verbesselta et al. (Verbesselt et al., 2014). Balíček *hts*, který obsahuje metody pro vizualizaci, analýzu a prognózování hierarchických časových řad a jeho interpretace, je obsažen v Hyndman et al. (2015). Dalším může být *MEFM*, sada nástrojů pro implementaci Monash Electricity, prognostický model, založený na článku Hyndmana a Fana (2010). Balíček *robets*

realizuje prognózy časových řad s robustním exponenciálním vyrovnáním. Je postaven na práci Hyndmana a Khandakara (2008) a je popsán v textu Crevitse a Crouxe (2017). Pomocí *stR* lze vyčíslit dekompozice časových řad, metody jsou popsány Dokumentovem a Hyndmanem (2016). *Thief* je temporální hierarchický prognostický model – tato metoda byla popsána ve "Forecasting with temporal hierarchies" spoluautorů Athanasopoulose et al. (2017). *Tsfeatures* umožňuje extrakci funkce časové řady, dle Hyndmana et al. (2015). Dalším z nových balíčků může být například *Prophet*, který vytvořili odborníci z Facebooku (Taylor & Letham, 2018). Byl využit i v dalších článcích (Navrátil & Kolková, 2019) a (Kolková & Ključnikov, 2022).

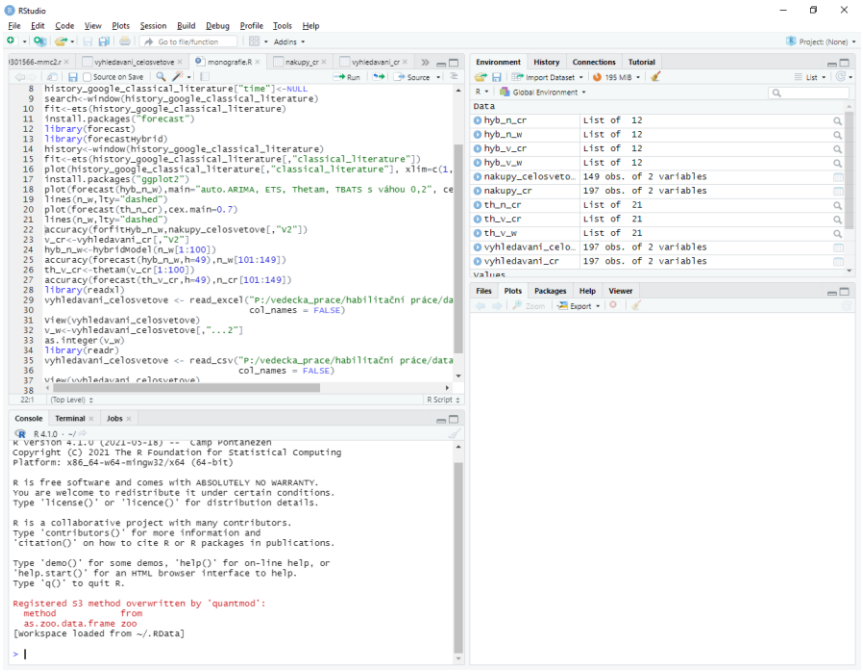
V samotném prognózování a dekompozici je jistě stále důležitá role analytika, nicméně řadu otázek lze již automatizovat. Jenom tato automatizace umožní využití modelu v širším měřítku, a to i pro manažery a podnikatele bez širokých statistických a programovacích znalostí. Modely určené pro podnikovou sféru musí splňovat požadavky určité jednoduchosti a snadné interpretovatelnosti. Taylor a Letham (2018) definovali nejlepší využití lidských a automatizovaných úkolů. Schematický pohled na tuto automatizaci je na obrázku 4.



Obrázek 4 Schéma automatizace úkolů při prognózování
Zdroj: Taylor a Letham, 2018

Taylor a Letham (2018) začínají své modelování pomocí analytika, který definuje dostatečně flexibilní zadání se subjektivním výkladem parametrů. Následuje již automatizovaný proces, kterým se simulují prognózy a vyčíslují míry jejich přesnosti. Pokud nastane situace, kdy míra přesnosti nedosahuje požadovaných hodnot nebo nastanou problémy s jinými aspekty, model označí tyto problémy a předá lidskému analytikovi, obvykle i s pořadím priority. Analytik pak může model vizuálně posoudit a následně adekvátně upravit. Zde už lze spatřovat počátek AutoML systému, o kterém už byla zmínka na začátku kapitoly.

Jazyk R lze využít v různých prostředích, nejznámější je R-studio, jejichž uživatelské rozhraní je na obrázku 5. Velmi zajímavé je i prostředí Rkward (Roediger et al., 2012), které upravili i odborníci z VŠB–TU Ostrava (Janurová et al., 2016).



Obrázek 5 Ukázka RStudia s programovacím jazykem R

Zdroj: vlastní zpracování v RStudiu

Python

Dalším programovacím jazykem, který lze využít pro prognózování v podnikové ekonomice, je Python. Funguje na podobném principu jako jazyk R. Rovněž je vyvíjen jako open source projekt s celou řadou dalších balíčků. Python stejně jako R nebo i Matlab patří mezi programovací jazyky, které se lze poměrně snadno a rychle naučit. Mezi vědeckou veřejností jsou příznivci jak jazyka R, tak Pythonu. Příznivci Pythonu argumentují jednoduššími kódy a prakticky stejnými možnostmi využití balíčku jako R a nejspíše lze očekávat v budoucnu větší příklon směrem od R k Pythonu.

Python navrhl Guido van Rossum už v roce 1991 v Matematickém centru Stiching v Nizozemsku. A jeho jméno opravdu pochází z názvu zábavného pořadu Monty Python létající cirkus. V současnosti se aktuálně používá verze Python 3.0.

Autorka v Pythonu realizovala několik vědeckých článků (Kolková & Ključnikov, 2022) nebo (Kolková & Navrátil, 2021) a dospěla k závěru, že Python má ve svém použití mnoho výhod, a i když se jedná o software s otevřeným zdrojovým kódem, umožňuje profesionální a sofistikovanou práci v oblasti správy dat a statistických výpočtů.

Python na poli prognózování poptávky již několik let plní nezastupitelnou úlohu. Poptávku po biopalivech prognózují Paula aj. (Paula et al., 2022), kde využili modely Arima a modely založené na umělé inteligenci. K výpočtům použili balíčky *Scikit-Learn*, *Gradient Boosting* a *forecast*. V článku (Khokhlova & Khokhlova, 2022) analyzovali technologické trendy pomocí rozsáhlých open source dat o patentech. Na základě této analýzy se pokoušejí předvídat budoucí poptávku po dovednostech na trhu práce. Python dokonce et al. používají pro predikci poptávky po nabíjení elektrických vozidel (Zhang et al., 2021). Zde na základě dat z městských bodů a časoprostorových charakteristik obyvatel autoři pomocí Pythonu vytvářejí úplný model předpovědi poptávky po nabíjení. Optimalizaci problémů s plánováním agregované výroby s nebo bez ztráty produktivity řešili Rehman et al. (2021). Jejich výsledky naznačují, že ztráta produktivity má přímý dopad na poptávku po pracovních silách. Pro výpočet využili PuLP, což je modelovací rámec opět v režimu open source a řeší výpočty lineárního celočíselného programování.

V podnikání se Python používá i v jiných oblastech než jen prognózování poptávky a jeho využití je velmi široké. Pomocí Pythonu výzkumníci hledají vhodné umístnění turisticky atraktivních cílů (Michalkó et al., 2022). Rozhodovací analýzu multifaktorového úvěrového rizika provádějí Tang et al. (2022), kteří pomocí Pythonu řeší úlohy logistické regrese i neuronové sítě pro malé a střední podniky. Celá řada studií se také zabývá problematikou finančního trhu, například Ding et al. (2022) prognózuje na základě hlubokého učení vývoj Shanhai Stock Exchange 50 Index. A mohli bychom pokračovat jistě celou řadou dalších studií. Python je dnes již nástroj, který výzkumníci z ekonomických oborů celého světa používají zcela běžně. Lze předpokládat, že nebude dlouho trvat a dostane se i do běžné výuky na vysokých školách ekonomických a bude postupně nahrazovat programy jako Matemathica, Statistica nebo SPSS.

Nyní se mnoho datových vědců zaměřuje hlavně na studium programovacího jazyka, kterému dnes opravdu dominuje Python. Často ale nedisponují dostatečnou znalostí statistiky či jiných aplikovaných věd. Mezi nejlepší knihovny, pomocí kterých se lze učit statistiku, patří Scipy.Stats, Pingouin a Statsmodels. Všechny tyto knihovny lze získat v režimu open source, takže jsou k dispozici každému. Vyjmenované knihovny usnadňují práci datovým vědcům, pro něž není statistika hlavním oborem vzdělání.

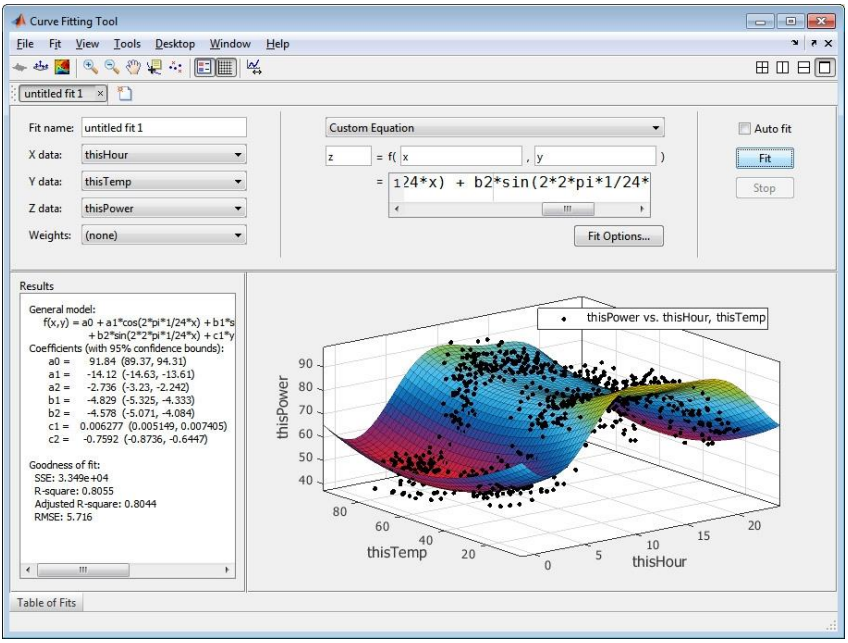
Stejně jako R i Python lze použít v různých prostředích. Nejčastěji používané v akademické sféře je ale i zde volně dostupné integrované vývojové prostředí – Visual Studio, ať již ve verzi Code nebo 2022 či jiné. Python lze ale aplikovat i v R-studiu a kombinovat jeho využití s jazykem R. Samozřejmě existují i další prostředí, jako PyCharm.

Matlab

Matlab je integrovaný systém zahrnující nástroje pro symbolické a numerické výpočty, analýzu a vizualizaci dat, modelování a simulace dějů. Název MATLAB vznikl zkrácením slov MATrix LABoratory, což se dá přeložit jako „maticová laboratoř“. Odpovídá to skutečnosti, že klíčovou datovou strukturou při výpočtech v MATLABu jsou matice.

Vytvořil ho Cleve Moler z univerzity v Novém Mexiku na konci 70. let, aby usnadnil práci svým studentům, kteří se tak nemuseli učit programovací jazyk Fortran. Programování ve Fortranu bylo složité a skrývalo mnoho nástrah při řešení matematických výpočtů. MATLAB se velice rychle rozšířil i na další univerzity. V roce 1983 se o MATLAB začal zajímat Jack Little na Stanfordově univerzitě. Do té doby byl MATLAB zdarma, Jack Little v něm však viděl i možnost ekonomického zhodnocení. MATLAB přepsal do jazyka C a v roce 1984 Little, Moler a Steve Bergert společně založili společnost MathWorks, která nabídla produkt na trh.

Hlavní oblastí využití jsou technické obory a ekonomie. Někteří odborníci nepovažují MATLAB za programovací jazyk, jiní o něm zase říkají, že je velice cenným a užitečným programovacím jazykem. Hlavním konkurentem je samozřejmě jazyk R a Python, jejichž hlavní výhodou je nulová cena. Naproti tomu je MATLAB označován za rychlejší. Ukázku uživatelského rozhraní v softwaru MATLAB znázorňuje obrázek 6.



Obrázek 6 Ukázka možností softwaru MATLAB

Zdroj: mathworks.com

AutoML systémy

Automatizované Machine Learning, označovaný také jako automatizovaná ML nebo AutoML, je proces automatizace časově náročného zpracování iterativních úloh vývoje modelů strojového učení.

AutoML systémy jsou novinkou a v posledních deseti letech se o nich velmi diskutuje, nicméně praktický rozvoj je datován kolem roku 2020. Jejich hlavní podstatou je vytvořit systém, který bude realizovat celou predikční analytiku sám. Jinými slovy, můžeme ho pojmenovat jako umělou inteligenci, která řídí umělou inteligenci. Měl by být schopen samostatné optimalizace i samostatné tvorby uživatelsky příznivých výstupů. Takto by byl vhodný vlastně pro každého i bez znalostí informatiky či prognostických modelů. Co samozřejmě modely nedokážou, je definice samotných úloh a obchodních cílů, ke kterým má výstup systému sloužit.

V současné době jsou tyto systémy nabízeny několika firmami. Za zmínku stojí i open source možnosti, jež nabízí například AutoSklearn, Auto Keras, TPOT. Další skupinu poskytovatelů těchto systémů tvoří specializované start-upy, které vznikly právě za účelem takové nabídky. Mohou to být například Data Robot nebo H2O a řada dalších. Nicméně ani standardní softwarové firmy v tomto vývoji nezaostávají, a tak své AutoML systémy již nabízí i IBM SPSS, Google, Amazon, MS Azure a jiné, definované i v předchozích kapitolách.

Seznam příloh

Seznam obrázků

Seznam tabulek

Literatura

ABAY, K., TAFERE, K., WOLDEMICHAEL, A. (2020). Winners and Losers from COVID-19: Global Evidence from Google Search. *World Bank Policy Research Working Paper*, No. 9268. Washington D.C.: World Bank. https://papers.ssrn.com/abstract=3617347.

ABBASIMEHR, H., SHABANI, M., YOUSEFI, M. (2020). An optimized model using LSTM network for demand forecasting. *Computers & Industrial Engineering* 143(1): 345-366. https://doi.org/10.1016/j.cie.2020.106435.

ALON, M. Q., SADOWSKI, R. J. (2001). Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. *J. Retail. Consumer Service* 8(3): 147-156.

ANUPAM, A., LAWAL, I. (2023). Forecasting air passenger travel: A case study of Norwegian aviation industry. *Journal of Forecasting*. https://doi.org/10.1002/for.305.

APERGIS, N., CHATZIANTONIOU, I., GABAUER, D. (2023). Dynamic connectedness between COVID-19 news sentiment, capital and commodity markets. *Applied Economics 55*(24): 2740-2754. https://doi.org/10.1080/00036846.2022.2104804.

ARAZMJOO, H. R. (2019). A Multi-Dimensional Meta-Heuristic Model for Managing Organizational Change. *Management Decision* 58(3): 526-543. http://doi: 10.1108/MD-07-2018-0795.

ARMSTRONG, J. S. (2001). *Principles of Forecasting: a Handbook for Researchers and Practitioners*. New york: Kluwer Academic Publishers.

ARMSTRONG, J., GREEN, K. (2014). *Methodology of Tree of Forecasting.* Dostupné na: http://forecastingprinciples.com/index.php/methodology-tree.

ASSIMAKOPOULOS, V., NIKOLOPOULOS, K. (2000). The theta model: a decomposition approach to forecasting. *International Journal of Forecasting* 16(4): 521-530. https://doi.org/10.1016/S0169-2070(00)00066-2.

BAKER, S., FRADKIN, A. (2017). The impact of unemployment insurance on job search: Evidence from google search data. *Review of Economics and Statistics* 99(5): 756-768. http://dx.doi.org/10.1162/REST_a_00674.

BALAJI PRABHU, B., DAKSHAYINI, M. (2020). Computational Performance Analysis of Neural Network and Regression Models in Forecasting the Societal Demand for Agricultural Food Harvests. *International Journal of Grid and High Performance Computing 12*(4): 35-47. https://doi.org/10.4018/IJGHPC.2020100103.

BALL, W., COXETER, H. (1987). *Mathematical Recreations and Essays*. 13th Edition, Dover, New York.

BANDARA, K., BERGMEIR, C., SMYL, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Systems with Applications 140*. https://doi.org/10.1016/j.eswa.2019.112896.

BASAK, A., RAHMAN, A., DAS, J., HOSONO, T., KISI, O. (2022). Drought forecasting using the Prophet model in a semi-arid climate region of western India. *Hydrological Sciences Journal 2022*: 1-21. https://doi.org/10.1080/02626667.2022.2082876.

BELL, E., BRYMAN, A., HARLEY, B. (2018). *Business Research Methods* (1st ed.). Oxford: Oxford University Press.

BELL, F. (2018). *Introduction to Forecasting*. Dostupné na: Qcon.ai: https://qcon.ai/qconai-2018-presentation-videos.

BERGMEIR, CH., HYNDMAN, R. J., BENÍTEZ, J. M. (2016). Bagging Exponential Smoothing Methods Using STL Decomposition and Box–Cox Transformation. *International Journal of Forecasting* 32: 303–312. http://doi:10.1016/j.ijforecast.2015.07.002.

BHATTACHERJEE, A. (2019). *Social Science Research: Principles, Methods and Practices* (., 2nd ed.). University of South Florida, Florida. Dostupné z: https://usq.pressbooks.pub/socialscienceresearch/.

BROWN, R. G. (1959). Statistical Forecasting for Inventory Control. McGraw/Hill.

Bible, překlad 21 století. (2009). Praha: Biblion, o.s.

BOROVCOVÁ, M., ŠPAČKOVÁ, A. (2018). Determination and Verification of the Key Assessment Indicators for the Insurance Market by Applying the Decomposition Multi- attribute Methods and Regression Analysis. In: *EUROPEAN FINANCIAL SYSTEMS 2018: Proceedings of the 15th International Scientific Conference*.pp. 44-52.

BUI, D., LE, P., CAO, M., PHAM, T., PHAM, D. (2020). Accuracy improvement of various short-term load forecasting models by a novel and unified statistical data-filtering method. *International Journal of Green Energy 17*(7): 382-406. https://doi.org/10.1080/15435075.2020.1761810.

BULUT, L. (2018). Google Trends and the forecasting performance of exchange rate models. *Journal of Forecasting* 37(3), 303-315.
https://doi.org/10.1002/for.2500

BUREAU, U. C. (2011). X-12-ARIMA Reference Manual, Statistical Research Division -Time Series Research Staff. Dostupné na: http://www.census.gov/srd/www/x12a/

CASTI, J. (2002). I Know what You'll Do Next Summer. *New Scientist* 175(2358): 29.

CEBRIÁN, E., DOMENECH, J. (2023). Is Google Trends a quality data source?. *Applied Economics Letters 30*(6): 811-815. https://doi.org/10.1080/13504851.2021.2023088.

CIPRA, T. (2013). *Finanční ekonometrie.* Praha: Ekopress.

CLEVELAND, R. B., CLEVELAND, W. S., MCRAE, J., a TERPENNING, I. (1990). STL: a seasonal- trend decomposition procedure based on loess. *J. O. Statt* 6: 3-73.

COURNÈDE, B., V. ZIEMANN, F. DE PACE (2020). Housing amid Covid-19: Policy responses and challenges. OECD, *OECD Policy Responses to Coronavirus* (COVID-19).

COXETER, H. S. M. (1969). The Golden Section and Phyllotaxis. *Introduction to Geometry*, 2nd ed. New York: Wiley. ISBN 0-471-50458-0.

CREVITS, R., CROUX, C. (2017). *Forecasting Using Robust Exponential Smoothing with Damped Trend and Seasonal Components*. Dostupné na: https://ssrn.com/abstract=3068634 nebo http://dx.doi.org/10.2139/ssrn.3068634.

CRONE, S., HIBON, M., NIKOLOPOULOS, K. (2011). Advances in forecasting with Neutral Networks? Empirical Evidence from The NN3 Competetion on Time Series Prediction. *International Journal of Forecasting* 27: 635-660. http://doi:10.1016/j.ijforecast.2014.03.011.

CUHADAR, M. (2020). Modelling and Forecasting Inbound Tourism Demand to Croatia using Artificial Neural Networks: A Comparative Study. *Journal of Tourism and Services 11*(21): 55-70. https://doi.org/10.29036/jots.v11i21.171.

DAGUM, E., BIANCONCINI, S. (2016). *Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation* (Statistics for Social and Behavioral Sciences) 1st ed. New York: Springer.

DAIM, T., RUEDA, G., MARTIN, H., GERDSRI, P. (2006). Forecasting Emerging Technologies: Use of Bibliometric and Patent Analysis. *Technological Forecasting and Social Change* 73 (8): 981–1012. http://doi:10.1016/j.techfore.2006.04.004.

DEB, S. (2023). Analyzing airlines stock price volatility during COVID -19 pandemic through internet search data. *International Journal of Finance & Economics*, *28*(2), 1497-1513. https://doi.org/10.1002/ijfe.2490.

DE LIVERA, A.M., HYNDMAN, R. J., SNYDER R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association* 106(496): 1513-1527. http://doi:10.1198/jasa.2011.tm09771.

DING, Y., SUN, N., XU, J., LI, P., WU, J., TANG, S., JAN, N. (2022). Research on Shanghai Stock Exchange 50 Index Forecast Based on Deep Learning. *Mathematical Problems in Engineering 2022*: 1-9. https://doi.org/10.1155/2022/1367920.

DOERR, S., GAMBACORTA, L. (2020). Identifying regions at risk with Google Trends: the impact of Covid-19 on US labor markets. *BIS Bulletins.* https://ideas.repec.org/p/bis/bisblt/8.html.

DOKUMENTOV, A., HYNDMAN, R.J. (2016). Long-term forecasts of age-specific participation rates with functional data models. *Working Paper*. http://business.monash.edu/ econometrics-and-business-statistics/research/publications

DOSTÁL, P. R. (2005). *Pokročilé metody manažerského rozhodování: konkrétní příklady využití metod v praxi*. Praha: Grada.

DOSTÁL, P., RAIS, K., SOJKA, Z. (2005). *Pokročilé metody manažerského rozhodování: pro manažery, specialisty, podnikatele a studenty, konkrétní příklady využití metod v praxi*. Praha: Grada Publishing.

DUDEK, G., PELKA, P., SMYL, S. (2022). A Hybrid Residual Dilated LSTM and Exponential Smoothing Model for Midterm Electric Load Forecasting. *IEEE Transactions on Neural Networks and Learning Systems 33*(7): 2879-2891. https://doi.org/10.1109/TNNLS.2020.3046629.

EIKELAND F., O., BIANCHI, F., APOSTOLERIS, H., HANSEN, M., CHIOU, Y., CHIESA, M. (2021). Predicting Energy Demand in Semi-Remote Arctic Locations. *Energies 14*(4). https://doi.org/10.3390/en14040798.

ELLIOTT, R.N., COLLINS, CH.J. (1938). *The Wave Principle.*

ELLIOTT, R.N. (1946). *Nature's Law: The Secret of the Universe.* New York: Snowballpublishing, (reprodukováno ze strojopisné kopie, 2011).

ELSENBROICH, C., PAYETTE, N. (2020). Choosing to Cooperate: Modelling Public Goods Games with Team Reasoning. *Journal of Choice modelling* 34. http://doi:10.1016/j.jocm.2020.100203.

ESMAELIAN, M., TAVANA, M., CAPRIO, D., ANSARI, R. (2017). A Multiple Correspondence Analysis Model for Emaluating Technology Foresight Methods. *Technological Forecasting a Social Cahnge* 125: 188-205. http://doi:10.1016/j.techfore.2017.07.022.

FILDES, R., MA, S., KOLASSA, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting* 38(4): 1283-1318. https://doi.org/10.1016/j.ijforecast.2019.06.004.

Forecast Pro. (2022). https://www.forecastpro.com/solutions/forecast-pro/?gclid=EAIaIQobChMIluHb7bWs-AIVmvdRCh2b4A9iEAAYASACEgKMxfD_BwE. Retrieved 2022-06-28, from https://www.forecastpro.com/solutions/forecast-pro/?gclid=EAIaIQobChMIluHb7bWs-AIVmvdRCh2b4A9iEAAYASACEgKMxfD_BwE.

FRANEK, J. (2016). Application of selecting measures in data envelopment analysis for company performance rating. In: 3*4th International Conference Mathematical Methods in Economics*, pp. 219-224.

FAKHARCHIAN, S. (2023). Designing a forecasting assistant of the Bitcoin price based on deep learning using market sentiment analysis and multiple feature extraction. *Soft Computing 27*(24): 18803-18827. https://doi.org/10.1007/s00500-023-09028-5.

GÁL, F. (1999). *Možnosť a skutočnosť.* Bratislava: Obzor.

GANGULY, D. D. (2017). Optimalized Hydel-Thermic Operative Planning Using IRECGA. In: *3rd International Conference on Research in Computational Intelligence*. Kalkata: India.

GAŠPARÍKOVÁ, J. (2007). Is New Economic Research More Qualitative? (Treatment on Forecasting Methods). *Ekonomický časopis/Journal of Economics* 55(3): 287.

GILLILAND, M. (2020). The value added by machine learning approaches in forecasting. *International Journal of Forecasting 36*(1): 161-166. https://doi.org/10.1016/j.ijforecast.2019.04.016.

GIMENEZ, E., NOVO, J. (2020). A Theory of Succession in Family Firms. *Journal of Family and Economic Issues* 41(1): 96-120. http://doi:1 0.1007/s10834-019-09646-y.

GOMÉZ, V., MARAVALL, A. (1997). *Programs TRAMO (Time Series Regression with ARIMA Noise, Missing Observations, and Ooutliers) and SEATS (Signal Extraction) in ARIMA Time Series*. Dostupné na: http://www.bde.es/servicio/software/econome.htm.

GONZALES, F., A. JAAX, A. MOUROUGANE (2020). Nowcasting aggregate services trade. *A pilot approach to providing insights into monthly balance of payments data*. Paris: OECD.

GREEN, K. C., ARMSTRONG, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research 68*(8): 1678-1685. https://doi.org/10.1016/j.jbusres.2015.03.026.

GUO, Y., ZHANG, Y., BOULAKSIL, Y., TIAN, N. (2021). Multi-Dimensional Spatiotemporal Demand Forecasting and Service Vehicle Dispatching for Online Car-Hailing Platforms. *International Journal of Production Research*. http://doi:10.1080/00207543.2021.1871675.

GYA, R. (2020). *Fast forward: Rethinking supply chain resilience for a post-COVID-19 world*. Capgemini Research Institute, 44. dostupné z https://www.capgemini.com/wp-content/uploads/2020/11/Fast-forward_Report.pdf.

HAYKIN, S. (1994). *Neural Networks: A Comprehensive Foundation*. New York: Macmillan Coll Div. ISBN-10 : 0023527617.

HERZOG, B., DOS SANTOS, L. (2021). Google Search in Exchange Rate Models: Hype or Hope? *Journal of Risk and Financial Management* 14(11). https://doi.org/10.3390/jrfm14110512.

HLADÍK, T. (2012). *Forecasting, demand planning a řízení zásob: Skrytý potenciál*. Dostupné na: https://docplayer.cz/1875852-Forecasting-demand-planning-a-rizeni-zasob-skryty-potencial-tomas-hladik-logio.html.

HOLCR, K. (1981). *Vojenská prognostika.* Praha: Naše vojsko.

HOLT, C. E. (1957). *Forecasting seasonal and trends by exponentially weighted averages*. Pittsburgh: Carnegie Institute of Technology, DOI: 10.1016/j.ijforecast.2003.09.015.

HURST, H. E. (1951). Long-term storage capacity of reservoirs. T*ransactions of the American Society of Civil Engineers* (116): 770–808.

HYNDMAN, R., ATHANASOPOULOS, G. (2018). *Forecasting, Principles and Practise*. Middletown: Otext.

HYNDMAN, R. J., ATHANASOPOULOS, G. (2021). *Forecasting: principles and practice (3rd edition)*. OTexts: Melbourne, Australia: OTexts.com/fpp3.

HYNDMAN, R., FAN, S. (2010). Density Forecasting for Long-Term Peak Electricity Demand. *IEEE Transactions on Power Systems* 25(2): 1142-1153. http://doi:10.1109/TPWRS.2009.2036017.

HYNDMAN, R., KHANDAKAR, Y. (2008). Automatic Time Series Forecasting: the Forecast Package for R. *Journal of Statistical Software* 27(3): 1-22. http://doi:10.18637/jss.v027.i03.

HYNDMAN, R. J., KOEHLER, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting* 22: 679–688. http://doi: 10.1016/j.ijforecast.2006.03.001.

HYNDMAN, R., ATHANASOPOULOS, G., SHANG, H. (2015). *hts: AnRPackage for Forecasting Hierarchical orGrouped Time Series*. Dodtupné na: http://cran. unej. ac. id/web/packages/hts/vignettes/hts. pdf.

HYNDMAN, R., WANG, E., LAPTEV, N. (2015). Large-Scale Unusual Time Series Detection. In: *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. Atlantic City: IEEE, pp. 1616-1619.

CHALMERS, A. (2013). *What is This Thing Called Science?* (UQ Press). St. Lucia.

CHEN, C., TWYCROSS, J., GARIBALDI, J. M. (2017). A new accuracy measure based on bounded relative error for time series forecasting. *PLoS ONE* 12(3). http://doi.org/10.1371/journal.pone.0174202.

CHU, C.W., ZHANG, G.P. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *Int. J. Prod. Econ*omy 86: 217-231.

IBM Cognos Analytics. (2021). Citováno 2022-06-28, dostupné z: https://www.ibm.com/docs/cs/cognos-analytics/11.1.0?topic=stories-get-started-dashboards.

ITO, T., MASUDA, M., NAITO, A., TAKEDA, F. (2021). Application of Google Trends-based sentiment index in exchange rate prediction. *Journal of Forecasting* 40(7): 1154-1178. https://doi.org/10.1002/for.2762.

JAGANATHAN, S., PRAKASH, P. (2020). A combination-based forecasting method for the M4-competition. *International Journal of Forecasting 36*(1): 98-104. https://doi.org/10.1016/j.ijforecast.2019.03.030.

JANA, R., GHOSH, I., WALLIN, M. (2022). Taming energy and electronic waste generation in bitcoin mining: Insights from Facebook prophet and deep neural network. *Technological Forecasting and Social Change 178*. https://doi.org/10.1016/j.techfore.2022.121584.

JANUROVA, K., LITSCHMANNOVA, M., SKOPAL, R., KURANOVA, P., BELOCH, M.. (2016). Supporting Freeware For Statistical Lectures - RKWard. In: *10TH International Days of Statistics And Economics*. Praha: MELANDRIUM, 711-722.

KAPOUNEK, S. KUČEROVÁ, Z. (2016). Google Trends and Exchange Rate Movements: Evidence from Wavelet Analysis. 34th International Conference Mathematical Methods in Economics. Liberec: Technická univerzita Liberec, 377-382.

KATAEV, M. B. (2020). Fuzzy Model Estimation of the Risk Factors Impact on the Target of Promotion of the Software Product. *Enterprise Information Systems*. http://doi:1 0.1080/17517575.2020.1713407.

KESTEN, C., ARMSTRONG, J. (2015). Simple Versus Complex Forecasting: the Evidence. *Journal of Business Research* 68: 1678-1685. http://doi:10.1016/j.jbusres.2015.03.026.

KHOKHLOVA, O. A., KHOKHLOVA, A. N. (2022). Analysis of technological trends to identify skills that will be in demand in the labor market with open-source data using machine learning methods. *Izvestiya Of Saratov University. New Series. Series: Mathematics. Mechanics. Informatics 22*(1): 123-129. https://doi.org/10.18500/1816-9791-2022-22-1-123-129.

KOLASSA, S. (2022). Commentary on the M5 forecasting competition. *International Journal of Forecasting 38*(4): 1562-1568. https://doi.org/10.1016/j.ijforecast.2021.08.006.

KOLKOVÁ, A. (2017). Application of Fibonacci Numbers on Technical Analysis of EUR/USD Currency Pair. In: *International Day o Science 2017*. Olomouc: Moravian university college Olomouc, 99-107.

KOLKOVÁ, A. (2018a). Indicators of Technical Analysis on the Basis of Moving Averages as Prognostic Methods in the Food Industry. *Journal of competetivess* 10(4): 102–119. http://doi:1 0.7441/joc.2018.04.07.

KOLKOVÁ, A. (2018b). Measuring the Accuracy of Quantitative Prognostic Methods and Methods Based on Technical Indicators in the Field of Tourism. *Journal Acta Oeconomica Universitatis Selye* 7(1): 58-70.

KOLKOVÁ, A. (2018c). The use of Cryptocurency in Entreprises. In: *6th International ConferenceInnovation Management, Entrepreneurship and Sustainability*. Praha: VŠE, 541-553.

KOLKOVÁ, A. (2019). Aplication of Artificial Neural Networks for Forecasting in Business. In: *7th International Conference on Innovation Management, Entrepreneurship and Sustainability.* Praha: VŠE Praha, 359-368.

KOLKOVÁ, A. (2020). The Application of Forecasting Sales In Services To Incerase Business Competetiveness. *Journal of Competetiveness* 12(2): 90-105. http://doi:1 0.7441/joc.2020.02.06.

KOLKOVÁ, A. (2020b). Exploratory data analysis as a tool for risk management of accommodation services Airbnb New York City. In *10th International Scientific Conference on Managing and Modelling of Financial Risks*. Ostrava: VŠB-TU of Ostrava, Faculty of Economics, Department of Finance, pp. 98-106.

KOLKOVÁ, A., KLJUČNIKOV, A. (2021). Demand forecasting: an alternative approach based on technical indicator Pbands. *Oeconomia Copernicana 12*(4): 1063-1094. https://doi.org/10.24136/oc.2021.035.

KOLKOVÁ, A., KLJUČNIKOV, A. (2022). Demand forecasting: AI-based, statistical and hybrid models vs practice-based models - the case of SMEs and large enterprises. *Economics And Sociology*, *15*(4): 39-62. https:/ /doi.org/10.14254/2071-789X.2022/15-4/2.

KOLKOVÁ, A., NAVRÁTIL, M. (2021). Demand Forecasting in Python: Deep Learning Model Based on LSTM Architecture versus Statistical Models. Acta Polytechnica Hungarica 18(8): 123-141. https://doi.org/10.12700/APH.18.8.2021.8.7.

KOLKOVA, A., ROZEHNAL, P. (2022). Hybrid demand forecasting models: pre-pandemic and pandemic use studies. *Equilibrium 17*(3): 699-725. https://doi.org/10.24136/eq.2022.024.

KOLKOVÁ, A., ROZEHNAL, P., GAŽI, F., FAJMON, L. (2022). The Use of Quantitative Methods in Business Practice: Study of Czech Republic. *International Journal of Entrepreneurial Knowledge 10*(1): 80-99. https://doi.org/10.37335/ijek.v10i1.159.

KOLKOVÁ, A., MACUROVÁ, P. (2022). Cyclo-Mobility as a Contribution to Sustainable Development: Analysis and Forecast of Demand for Bicycles and their Components Before and During the Pandemic Time. In *ARDIELLI, E. (ed.). Proceedings of the 5th International Scientific Conference Development and Administration of Border Areas of the Czech Republic and Poland – Support for Sustainable Development, RASPO 2022*. Ostrava: VSB - Technical University of Ostrava, pp. 137-147.

KOSTIN, K. (2019). Investment Appeal Assessment of International Corporations. *Strategic management* 24(1): 3-9. http://doi:1 0.5937/StraMan1901003K.

KUPFER, A., ZORN, J. (2019). Valuable information in early sales proxies: The use of Google search ranks in portfolio optimization. *Journal of Forecasting 38*(1): 1-10. https://doi.org/10.1002/for.2547.

LI, X., LAW, R., XIE, G., WANG, S. (2021). Review of Tourism Forecasting Research with Internet Data. *Tourism Management,* 83. http://doi:10.1016/j.tourman.2020.104245.

LISHMANOVÁ, M. (2012). *Úvod do statistiky.* Ostrava: VŠB- TU Ostrava.

MACUROVÁ, P., KLABUSAJOVÁ, N., TVRDOŇ, L. (2018). *Logistika, 2. vyd.* Ostrava: VŠB-TU Ostrava.

MADDODI, S., KUNTE, S. R. (2024). Precision forecasting in perilous times: stock market predictions leveraging google trends and momentum indicators during COVID-19. *Managerial Finance 50*(10): 1747-1772. https://doi.org/10.1108/MF-02-2024-0128.

MAKRIDAKIS, S., HIBON, G. (2000). The M3-Competition: Results, Conclusions and Implications. *International Journal of Forecasting* 16(4): 451-476. http://doi:10.1016/S0169-2070(00)00057-1.

MAKRIDAKIS, S., HIBON, M. (1979). Accuracy of Forecasting: An Empirical Investigation (With Discussion). *Journal of the Royala Statistical Society* 142: 97-145.

MAKRIDAKIS, S., HIBON, M., LAWRENCE, M., MILLS, T., ORD, K., SIMMONS, L. (1993). The M2-Competition: a Real-Time Judgmentally Based Forecasting Study. *International Journal of Forecasting* 9(1): 5-22. http://doi:10.1016/0169-2070(93)90044-N.

MAKRIDAKIS, S., SPILIOTIS, E., ASSIMAKOPOULOS, V. (2018). The M4 Competition: Results, Findings, Conclusion and Way Forward. *International Journal of Forecasting* 34: 802–808. http://doi:1 0.1016/j.ijforecast.2018.06.001.

MAKRIDAKIS, S., PETROPOULOS, F., SPILIOTIS, E. (2022). The M5 competition: Conclusions. *International Journal of Forecasting 38*(4):1576-1582. https://doi.org/10.1016/j.ijforecast.2022.04.006.

MAKRIDAKIS, S., SPILIOTIS, E., ASSIMAKOPOULOS, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, article in press. https://doi.org/10.1016/j.ijforecast.2021.11.013

MANSOOR, M., GRIMACCIA, F., LEVA, S., MUSSETTA, M. (2021). Comparison of Echo State Network and Feed-Forward Neural Networks in Electrical Load Forecasting for Demand Response Programs. *Mathematics and Computers in Simulation* 184: 282-293. http://doi:10.1016/j.matcom.2020.07.011

MARANON, M. K. (2018). Exploring the Elliott Wave Principle to Interpret Metal Commodity Price Cycles. *Environmental studies* (59): 125-138. http://doi:10.1016/j.resourpol.2018.06.010

MARČEK, D. (2013). *Pravděpodobnostné modelovanie a soft computing v ekonomike*. Ostrava: VŠB - TU Ostrava.

MARČEK, D. (2016). *Supervizované a nesupervizované učení z dat: statistický a soft přístup*. SAEI, 45. Ostrava: VSB – TU Ostrava.

MARČEK, D. (2019). Comparison of Predictive Statistical Learning Accuracy with Computational Intelligence Methods. In *IEEE 15th International Scientific Conference on Informatics*. IEEE, pp. 317-322.

MARTINO, J. (2003). A Review of Selected Recent Advances in Technological Forecasting. *Technological Forecasting and Social Change* 70 (8): 719-733. http://doi:10.1016/S0040-1625(02)00375-X

MCGREGOR, J. (2013). *Predictive Analysis with SAP®.* SAP Press.

MCCARTHY, T. , DAVIS, D. , GOLICIC, S., MENTZER, J. (2006). The evolution of sales forecasting management: a 20-year longitudinal study of forecasting practices. *Journal of Forecasting* 25(5), 303-324. https://doi.org/10.1002/for.989

MICHALKÓ, G., BAL, D., ERDÉLYI, É. (2022). Repositioning Budapest's tourism milieu for a post-Covid-19 period: Visual content analysis. *European Journal of Tourism Research*, 32. https://doi.org/10.54055/ejtr.v32i.2602

MIKUŠOVÁ, M., ČOPÍKOVÁ, A. (2016). What Business Owners Expect from a Crisis Manager? A Competency Model: Survey Results from Czech Businesses. *Journal of Contingencies and Crisis Management* 24(3): 162-180. https://doi.org/10.1111/1468-5973.12111

MILLER, P., SWINEHART, K. (2010). Technological Forecasting: a Strategic Imperative. *JGBM* 6 (2): 1-5.

MORGAVI, H. (2020). A GARCH model to now-cast private consumption using Google trends data. *OECD Economics Department Working Paper*. Paris: OECD.

MORO, S., CORTEZ, P. (2015). Business Intelligence in Banking: a Literature Analysis from 2002 To 2013 Using Text Mining and Latent Dirichlet Allocation. *Expert Syst. Application* 42 (3): 1314-1324. http://doi:10.1016/j.eswa.2014.09.024

MONTERO-MANSO, P., ATHANASOPOULOS, G., HYNDMAN, R.J., TALAGALA, T.S. (2020). FFORMA: Feature-based forecast model averaging. *International Journal of Forecasting* 36(1): 86-92. ISSN 01692070. Dostupné z: doi:10.1016/j.ijforecast.2019.02.011

NASH, J. (1996). *Essays on Game Theory.* Cheltenham: Edward Elgar.

NAVRÁTIL, M., KOLKOVÁ, A. (2019). Decomposition and Forecasting Time Series in the Business Economy Using Prophet Forecasting Model. *Central European Business Review 2019* 8(4): 26-39. http://doi:10.18267/j.cebr.221.

NIKOLOPOULOS, K., PUNIA, S., SCHAFERS, A., TSINOPOULOS, C., VASILAKIS, C. (2020). Forecasting and Planning during a Pandemic: COVID-19 Growth Rates, Supply Chain Disruptions, and Governmental Decisions. *European Journal of Operational Research* 290(1): 99-115. http://doi:10.1016/j.ejor.2020.08.001.

NING, Y., KAZEMI, H., TAHMASEBI, P. (2022). A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. *Computers & Geosciences*, 164. https://doi.org/10.1016/j.cageo.2022.105126.

PANDEY, P., BOKDE, N., DONGRE, S., GUPTA, R. (2021). Hybrid Models for Water Demand Forecasting. *Journal of Water Resources Planning and Management* 147(2). http://doi:10.1061/(ASCE)WR.1943-5452.0001331.

PAPPAS, T. (1989). *The Joy of Mathematics*. San Carlos, CA: Tetra. ISBN 0-933174-65-9.

PAULA, J. DE S., TEIXEIRA, L. L., RODRIGUES, S. B., HICKMANN, T., CORREA, J. M., RIBEIRO, L. DA S. (2022). Aplicação de técnicas de aprendizado de máquina e estatística na previsão da demanda de biocombustíveis. *Revista De Gestão E Secretariado 13*(4): 2559-2572. https://doi.org/10.7769/gesec.v13i4.1488.

PAVELKOVÁ, D., HOMOLKA, L. (2018). Passenger Car Sales Projections: Measuring the Accuracy of a Sales Forecasting Model. *Ekonomický časopis/Journal of Economics* 66(3): 227.

PAWLIKOWSKI, M., CHOROWSKA, A. (2020). Weighted ensemble of statistical models. *International Journal of Forecasting 36*(1): 93-97. https://doi.org/10.1016/j.ijforecast.2019.03.019.

PENG B., HAIYAN S., GEOFFREY I. C. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management* 45: 181-193. http://doi:10.1016/j.tourman.2014.04.005.

PEREIRA, L. N., CERQUEIRA, V. (2021). Forecasting hotel demand for revenue management using machine learning regression methods. *Current Issues in Tourism*, 1-18. https://doi.org/10.1080/13683500.2021.1999397.

PEREIRA DA VEIGA, C., PEREIRA DA VEIGA, C.R., PUCHALSKI, W., COELHO, L. S., TORTATO, U. (2016). Demand forecasting based on natural computing approaches applied to the foodstuff retail segment. *Journal of Retailing and Consumer Services* 31: 174-181. http://doi:10.1016/j.jretconser.2016.03.008.

PETERKOVÁ, J., FRANEK, J. (2018). Decision Making Support for Managers In Innovation Management: a PROMETHEE approach. *International Journal of Innovation* 6(3): 256-274. https://doi.org/10.5585/iji.v6i3.236.

PETERSON, S. P. (1993). Forecasting dynamics and convergence to market fundamentals. *Journal of Economic Behavior & Organization*. 22(3): 269-284. https://doi.org/10.1016/0167-2681(93)90002-7.

PETROPOULOS, A., SIAKOULIS, V., STAVROULAKIS, E., LAZARIS, P., VLACHOGIANNAKIS, N. (2022). Employing Google Trends and Deep Learning in Forecasting Financial Market Turbulence. *Journal of Behavioral Finance 23*(3): 353-365. https://doi.org/10.1080/15427560.2021.1913160.

PHAM, T. P., PAVELKOVA, D., POPESKO, B., HOANG, S. D., HUYNH, H. T. (2024). Relationship between fintech by Google search and bank stock return: a case study of Vietnam. *Financial Innovation* 10(1). https://doi.org/10.1186/s40854-023-00576-1.

PISU, M., COSTA, H., HWANG, H. (2020). Digital platforms and the COVID-19 crisis. *OECD Economics Department Working Paper*. Paris: OECD.

PTAK, C. (2018). The Demand Driven Adaptive Enterprise Model. Dynasys. Dostupné z https://blog.dys.com/ddae-model/.

RADVANSKÝ, M., PÁLENÍK, V., SLOBODNÍKOVÁ, S. (2010). Midterm Forecast of Slovak Economy for the Period 2010 – 2013 with Outlook to 2015. *Ekonomický časopis/Journal of Economics* 58(6): 614.

REHMAN, H. U., AHMAD, A., ALI, Z., BAIG, S. A., MANZOOR, U. (2021). Optimization of Aggregate Production Planning Problems with and without Productivity Loss using Python Pulp Package. *Management And Production Engineering Review 12*(4): 38-44. https://doi.org/10.24425/mper.2021.139993.

REID, D. (1969). A *comparative Study of Time Series Prediction Techniques on Economic Data*. Nottingham: University of Nottingham.

RIBEIRO, A. C. (2019). Elliott's Wave Theory in The Field of Econophysics and its Application to the PSI20 in the Context Of Crisis. *Estudios de Economia Aplicada* 37(2): 41-53.

ROEDIGER, S., FRIEDRICHSMEIER, T., KAPAT, P., MICHALKE, M. (2012). RKWard: A Comprehensive Graphical User Interface and Integrated Development Environment for Statistical Analysis with R. *Journal of Statistical Software* 49(9): 1-34.

SHAUB, D. (2020). Fast and accurate yearly time series forecasting with forecast combinations. *International Journal of Forecasting 36*(1): 116-120. https://doi.org/10.1016/j.ijforecast.2019.03.032.

SHAUB, D., ELLIS, P. (2018). *Package 'forecastHybrid'.* Retrieved from https://gitlab.com/dashaub/forecastHybrid.

SARITAS, O., PACE, L., STALPERS, S. (2013). Stakeholder Participation and Dialogue in Foresight. In: *Participation and Interaction Foresight. Dialogue, Dissemination and Visions*. Cheltenham: Edward Elgar Publishing Limited, 37-70.

SIEGEL, E. (2015). P*redictive Analytics: The Power to Predict who Will Click, Buy, Lie, or Die*. Hoboken: John Wiley a Sons, Inc.

SMYL, S., BERGMEIR, C., WIBOWO, E., A NG, T. W. (2019). *Rlgt: Bayesian exponential smoothing models with trend modifications*. R package version 01-2.

SMYL, S., KUBER, K. (2016). Data preprocessing and augmentation for multiple short time series forecasting with recurrent neural networks. In *36th international symposium on forecasting*.

SMYL, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting* 36(1): 75-85. https://doi.org/10.1016/j.ijforecast.2019.03.017.

SMYL, S., ZHANG, Q. (2015). Fitting and extending exponential smoothing models with Stan. In *35th international symposium on forecasting.*

SOYER, E., HOGARTH, R. (2012). Illusion of Predictability: How Regressions Statistics Mislead Experts. *International Journal od Forecasting* 28: 695-711. http://doi:10.1016/j.ijforecast.2012.02.002.

SPILIOTIS, E., MAKRIDAKIS, S., SEMENOGLOU, A., ASSIMAKOPOULOS, V. Comparison of statistical and machine learning methods for daily SKU demand forecasting. *Operational Research*. https://doi.org/10.1007/s12351-020-00605-2.

STEHEL, V. (2019). *Využití teorie her při řízení podniku.* Plzeň: Aleš Čeněk, s.r.o.

SWAMINATHAN, K., VENKITASUBRAMONY, R. (2024). Demand forecasting for fashion products: A systematic review. *International Journal of Forecasting* 40(1): 247-267. https://doi.org/10.1016/j.ijforecast.2023.02.005.

ŠIMEČEK, P. (2019). *Statistical vs. Deep Learning Methods for Time Series Forecasting*. Dostupné na: http://www.mlmu.cz/archiv/.

ŠTĚDROŇ, B., PALÍŠKOVÁ, M., SOUČEK, Z., DVOŘÁK, A., TILINGER, P., A KOL. (2019). *Prognostika.* Praha: C.H.Beck.

TANG, Q., SHI, W. Y., NADEEM, M. F. (2022). Decision Analysis of Multifactor Credit Risk Based on Logistic Regression and BP Neural Network. *Mathematical Problems in Engineering* 2022: 1-6. https://doi.org/10.1155/2022/5194443.

TASSONE, E., ROHANI, F. (2017, April 17). *Our Quest For Robust Time Series Forecasting at Scale*. Dostupné na: http://www.unofficialgoogledatascience.com/2017/04/our-quest-for-robust-time-series.html.

TAYLOR, J. (2003). Exponential Smoothing with a Damped Multiplicative Trend. *International Journal of Forecasting* 19: 715–725. http://doi:10.1016/S0169-2070(03)00003-7.

TAYLOR, S., LETHAM, B. (2018). Forecasting at Scale. *The American Statistician* 72: 37-45. http://doi:10.1080/00031305.2017.1380080

TURKEY, J. W. (1977). *Exploratory Data Analysis.* California: Addison-Wesley Publishing Company.

VENABLES, W., SMITH, D. (2020). *An Introduction to R.* Dostupné na: https://cran.r-project.org/.

VERBESSELT, J., ZEILEIS, A., HYNDMAN, R. (2014). *bfast: Breaks for Additive Season and Trend (BFAST).* Dostupné na: https://CRAN.R-project.org/package=bfast.

VINCUR, P., ZAJAC, Š. (2007). *Úvod do prognostiky.* Bratislava: Sprint.

VISHNEVSKIY, K., KARASEV, O., A MEISSNER, D. (2015). Integrated Roadmaps And Corporate Foresight as Tools Of Innovation Management: The Case of Russian Companies. *Technol. Forecasting Social. Change* 90: 433-443. http://doi:10.1016/j.techfore.2014.04.011.

WANG, J., DU, X., QI, X. (2022). Strain prediction for historical timber buildings with a hybrid Prophet-XGBoost model. *Mechanical Systems and Signal Processing* 179. https://doi.org/10.1016/j.ymssp.2022.109316.

WANG, L., CHEN, H. (2022). Tourism Demand Forecast Based on Adaptive Neural Network Technology in Business Intelligence. *Computational Intelligence and Neuroscience* 2022: 1-14. https://doi.org/10.1155/2022/3376296.

WANG, J., LUO Y., TANG L., GE P. (2018). Modelling a combined forecast algorithm based on sequence patterns and near characteristics: An application for tourism demand forecasting. *Chaos, Solutions & Fractals* 108: 136-147. http://doi: 10.1016/j.chaos. 2018.01.028.

WICKHAM, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, USA: SPRINGER. http://doi:10.1007/978-0-387-98141-3.

WICKHAM, H., CHANG, W., HENRY, L., PEDERSEN, T., TAKAHASH, K., WILKE, C., DUNNINGTON, D. (2020). *Create Elegant Data Visualisations Using the Grammar of Graphics.* Dostupné na: https://cran.r-project.org/.

WICHER, P. Z., ZAPLETAL, F., LENORT, R. (2019). Sustainability Performance Assessment of Industrial Corporation Using Fuzzy Analytic Network Process. *Journal of Cleaner Production* 241. http://doi:1 0.1016/j.jclepro.2019.118132.

WISNIEWSKI, M. (1996). *Quantitative Methods for Decision Makers.* London: Pitman Publishing.

www.kaggle.com. (2019). Dostupné na: https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data.

ZELLNER, A. (2001). S*implicity, Inference and Modelling: Keep it Sophisticatedly Simple*. Cambridge: Cambridge University Press.

ZHANG, Q., YANG, S., ZENG, Q., YU, T. (2020). Storage Pricing Model of Container Yards Under Fluctuating Demand. *Applied Economics*. http://doi:10.1080/00036846.2020.1733476.

ZHANG, K. Y., TIAN, Y. J., SHI, S. S., SU, Y., XU, L. C., ZHANG, M. X. (2021). Electric Vehicle Charging Demand Forecasting Based on City Grid Attribute Classification. In *11th International Conference on Power and Energy Systems (ICPES)*. ELECTR NETWORK: IEEE Power & Energy Soc., pp. 592-597.

ZHONG, X., RAGHIB, M. (2019). Revisiting the use of web search data for stock market movements. *Scientific Reports* 9(1). https://doi.org/10.1038/s41598-019-50131-1.

ZMEŠKAL, Z. (2014). Two-person bi-matrix and N-person oligopolistic non-cooperative game models with fuzzy payoff. In: *7th International Scientific Conference on Managing and Modelling of Financial Risks*. Ostrava: VŠB - TU Ostrava, pp.897-903.

ZMEŠKAL, Z., TICHÝ, T., DLUHOŠOVÁ, D. (2013). *Finanční modely: koncepty, metody, aplikace.* Praha: Ekopress.

Zpráva o vývoji podnikatelského prostředí v České republice v roce 2019 (2019). Ministerstvo průmyslu a obchodu ČR, [cit. 2022-6-17]. Dostupné z: https://www.mpo.cz.

Rejstřík

Demand forecasting using Google Trends data

Andrea Kolková

Summary

Forecasting in the business economy is one of the key research questions today. The publication focuses on quantitative methods of demand prediction. Contains the results of several years of research as well as completely new findings. However, what researchers often have a problem with is obtaining data. Companies very often do not want to provide their data, especially for fear of revealing some business secrets or tactics. From this point of view, the starting point can be the use of Google Trends data. The goal is then to verify the possibilities of using GT data in demand forecasting.

The monograph is divided into 10 parts. The first part is devoted to basic methodological and theoretical starting points. It also puts the term demand into the context of logistics and logistics management. The next chapter forms the basic framework of current scientific research in quantitative demand forecasting methods. It also defines the thought process in forecasting and its evaluation. The third chapter concerns the definition of GT data and the method of their creation, as well as the description and decomposition of the data. The fourth chapter introduces the reader to quantitative forecasting methods, divided into statistical methods, methods inspired by nature, including methods of artificial neural networks, combined and hybrid methods, and methods developed in current practice.

The key chapters of the publication are chapters five to seven. These already apply the possibilities of using GT to a specific forecast of demand, first of all individual demand, then also demand for selected search categories (respectively market demand). The seventh chapter then describes how forecasting based on GT data can be used for competitor analysis.

The publication can be recommended above all to scientific workers who continue the scientific questions raised in this monograph. Also to researchers who are working on the development of new methodologies in the field of business economics and face the problems of lack of data. The publication may also be of

interest to practitioners dealing with forecasting. Last but not least, it can be used by students of subsequent master's or doctoral study programs as background material for their diploma or dissertation theses on this topic.

About the author

Ing. Andrea Kolková, Ph.D. (1978)

Andrea Kolková is an assistant professor at the Department of Business Administration, Faculty of Economics, VSB Technical University Ostrava. In 2006, she received a PhD degree in Finance. From 2004 to 2017, she worked at the University of Entrepreneurship and Law, Plc. as an assistant professor and in the positions of manager for master's degree, secretary of the department and subsequently deputy head. In the years 2006–2009 and 2010–2015 only as a part-time assistant professor. In the years 2001–2006, she also worked part-time at the Higher Vocational School Ahol, Plc. She is a volunteer financial manager in Junak - Czech Scouting (Scout Center 816.11).

Between 2001 and 2006, she published 15 scientific articles at conferences and in scientific journals. She published these mainly in the area of financial markets and forecasting financial variables in a short time. She dealt with the issue of stock exchange indices and stock market option contracts, including compound option positions. In the years 2006–2015, she interrupted scientific and publishing activities due to maternity duties (four children). During these years, she devoted herself to pedagogical work, supervision of diploma and bachelor's theses and self-education, especially in the field of business and business economics. She also wrote several study textbooks. After 2016, she resumed her scientific and publishing activities in the field of business economics. Now she deals with logistical demand forecasting, forecasting of other business variables, as well as the area of small and medium-sized enterprises. After 2016, she published another 30 articles in scientific journals and conferences. Of these articles, there are seven impacted journals, three with a Q1 rating, two with a Q2 rating and two articles with a Q3 rating. In all these articles, the author has the majority of the authorship.

She is the author and co-author of eight university textbooks, published both paperback and electronically. She has participated in several scientific and pedagogical projects (Grant Agency of the Czech Republic, Grant Agency of the Academic Alliance, Student Grant Competition, Innovation and Development Project).

Prognózování poptávky s využitím Google Trends dat
Andrea Kolková
Vydala VŠB-TUO
1. vydání 2025
Tisk VŠB-TUO
Náklad 100 ks
Počet stran 204

Volume 1
Prokop, L., Medvec, Z., Zmeškal, Z. (2009). *Problematika oceňování nedodané energie v průmyslu,* SAEI, vol. 1. Ostrava: VSB – TU Ostrava.
Volume 2
Hrubý, P. et al. (2010). *Víceúrovňové modelování podnikových procesů (systém REA)*, SAEI, vol. 2. Ostrava: VSB – TU Ostrava.
Volume 3
Kutscherauer, A. et al. (2010). *Regionální disparity. Disparity v regionálním rozvoji země, jejich pojetí, identifikace a hodnocení*, SAEI, vol. 3. Ostrava: VSB – TU Ostrava.
Volume 4
Hančlová, J. et al. (2010). *Makroekonometrické modelování české ekonomiky a vybraných ekonomik EU*, SAEI, vol. 4. Ostrava: VSB – TU Ostrava.
Volume 5
Martiník, I. (2015). *Modelování objektově orientovaných programových systémů Petriho sítěmi*, SAEI, vol. 5. Ostrava: VSB – TU Ostrava.
Volume 6
Tichý, T. (2010). *Simulace Monte Carlo ve financích: Aplikace při ocenění jednoduchých opcí*, SAEI, vol. 6. Ostrava: VSB – TU Ostrava.
Volume 7
Doleželová, H., Halásek, D. (2011). *Služby v obecném hospodářském zájmu v EU. Komparace České republiky a Německa*, SAEI, vol. 7. Ostrava: VSB – TU Ostrava.
Volume 8
Machová, Z., Kaštan, M., Janíčková, L., Kotlán, I. (2011). *Tax Burden in OECD Countries: WTI Application*, SAEI, vol. 8. Ostrava: VSB – TU Ostrava.
Volume 9
Tichý, T. (2011). *Lévy Processes in Finance: Selected Applications with Theoretical Background,* SAEI, vol. 9. Ostrava: VSB – TU Ostrava.
Volume 10
Macurová, P. et al. (2011). *Řízení rizik v logistice*, SAEI, vol. 10. Ostrava: VSB – TU Ostrava.
Volume 11
Dvoroková, K. et al. (2012). *Ekonometrické modelování konvergence ekonomické a cenové úrovně. Analýza průřezových a panelových dat*, SAEI, vol. 11. Ostrava: VSB – TU Ostrava.
Volume 12
Melecký, M., Melecký, A. (2012). *Aplikované modelování pro potřeby hospodářské politiky*, SAEI, vol. 12. Ostrava: VSB – TU Ostrava.
Volume 13
Halásková, M. (2012). *Veřejná správa a veřejné služby v zemích Evropské unie*, SAEI, vol. 13. Ostrava: VSB – TU Ostrava.
Volume 14

Šebestíková, V. et al. (2012). *Daňová a sociální optimalizace ve vztahu k nezaměstnanosti v České republice*, SAEI, vol. 14. Ostrava: VSB – TU Ostrava.
Volume 15
Pytlíková, M. et al. (2012). *Gender Wage Gap and Discrimination in the Czech Republic*, SAEI, vol. 15. Ostrava: VSB – TU Ostrava.
Volume 16
Hučka, M. et al. (2012). *Správa společností v zemích střední a východní Evropy*, SAEI, vol. 16. Ostrava: VSB – TU Ostrava.
Volume 17
Vrabková, I. (2012). *Perspektivy řízení kvality ve veřejné správě*, SAEI, vol. 17. Ostrava: VSB – TU Ostrava.
Volume 18
Marček, D. (2013). *Pravděpodobnostné modelovanie a soft computing v ekonomike*, SAEI, vol. 18. Ostrava: VSB – TU Ostrava.
Volume 19
Čulík, M. (2013). *Aplikace reálných opcí v investičním rozhodování firmy*, SAEI, vol. 19. Ostrava: VSB – TU Ostrava.
Volume 20
Šnapka, P. et al. (2013). *Rozhodování a rozhodovací procesy v organizaci (vybrané problémy)*, SAEI, vol. 20. Ostrava: VSB – TU Ostrava.
Volume 21
Šimíčková, M., Drastichová, M. (2013). *Ekonomie udržitelnosti – alternativní přístupy a perspektivy*, SAEI, vol. 21. Ostrava: VSB – TU Ostrava.
Volume 22
Kutscherauer, A., Šotkovský, I., Adamovský, J., Ivan, I. (2013). *Socioekonomická geografie a regionální rozvoj. Regionální analýzy v přístupech socioekonomické geografie k regionálnímu rozvoji,* SAEI, vol. 22. Ostrava: VSB – TU Ostrava.
Volume 23
Kutscherauer, A., Václavková, R., Malinovský, J. et al. (2013). *Komplementární přístupy k podpoře regionálního a municipálního rozvoje,* SAEI, vol. 23. Ostrava: VSB – TU Ostrava.
Volume 24
Komárková, Z., Frait, J., Komárek, L. (2013). *Macroprudential Policy in a Small Economy*, SAEI, vol. 24. Ostrava: VSB – TU Ostrava.
Volume 25
Komárek, L. et al. (2013). *Money, Pricing and Bubbles*, SAEI, vol. 25. Ostrava: VSB – TU Ostrava.
Volume 26
Tichý, T. (2013). *Backtesting of VaR Based Models: Methodological Review and Selected Applications*, SAEI, vol. 26. Ostrava: VSB – TU Ostrava.

Volume 27
Fojtíková, L. et al. (2014). *Postavení Evropské unie v podmínkách globalizované světové ekonomiky*, SAEI, vol. 27. Ostrava: VSB – TU Ostrava.
Volume 28
Dluhošová, D. et al. (2014). *Financial Management and Decision-making of a Company. Analysis, Investing, Valuation, Sensitivity, Risk, Flexibility*, SAEI, vol. 28. Ostrava: VSB – TU Ostrava.
Volume 29
Zmeškal, Z. et al. (2014). *Aplikace vícekriteriálních dekompozičních metod rozhodování v oblasti podnikové ekonomiky a managementu*, SAEI, vol. 29. Ostrava: VSB – TU Ostrava.
Volume 30
Toloo, M. (2014). *Introduction to Data Envelopment Analysis with Basic Models and Applications*, SAEI, vol. 30. Ostrava: VSB – TU Ostrava.
Volume 31
Janasová, E., Slavata, D., Ardielli, J. (2014). *Specifikace míry kapitalizace vybraného segmentu realitního trhu. Analyticko-statistická studie trhu s byty v Ostravě*, SAEI, vol. 31. Ostrava: VSB – TU Ostrava.
Volume 32
Tichý, T. (2015). *Simulation Monte Carlo In Finance*. *Pricing of (Exotic) Options*, SAEI, vol. 32. Ostrava: VSB – TU Ostrava.
Volume 33
Kresta, A. (2015). *Financial Engineering in Matlab: Selected Approaches and Algorithms*, SAEI, vol. 33. Ostrava: VSB – TU Ostrava.
Volume 34
Macháček, M. (2015). *Soumrak ekonomie? K problému formalizace a krize smyslu společenské vědy*, SAEI, vol. 34. Ostrava: VSB – TU Ostrava.
Volume 35
Macek, R. et al. (2015). *Effective Corporate Taxation in the European Union and the Czech Republic*, SAEI, vol. 35. Ostrava: VSB – TU Ostrava.
Volume 36
Vrabková, I., Vaňková, I. (2015). *Evaluation Models of Efficiency and Quality of Bed Care in Hospitals*, SAEI, vol. 36. Ostrava: VSB – TU Ostrava.
Volume 37
Bazsová, B., Křížová, A., Řeháček, P. (2015). *Teorie organizace. Přístupy, metody, nástroje, softwarová podpora*, SAEI, vol. 37. Ostrava: VSB – TU Ostrava.
Volume 38
Šotkovský, I. (2015*). Socioekonomické struktury v rozvoji regionu: Postavení obyvatelstva Moravskoslezska v prostorové struktuře regionů soudržnosti ČR*, SAEI, vol. 38. Ostrava: VSB – TU Ostrava.
Volume 39

Fojtíková, L., Vahalík, B. (2015). *Zahraniční obchod České republiky pod vlivem globalizačních jevů a procesů*, SAEI, vol. 39. Ostrava: VSB – TU Ostrava.
Volume 40
Melecký, A., Melecký, M. (2015). *Řízení vládního dluhu: strategie, riziko, alokace*, SAEI, vol. 40. Ostrava: VSB – TU Ostrava.
Volume 41
Hančlová, J. et al. (2015). *Optimization Problems in Economics and Finance*, SAEI, vol. 41. Ostrava: VSB – TU Ostrava.
Volume 42
Peterková, J., Ludvík, L. (2015). *Řízení inovací v průmyslovém podniku*, SAEI, vol. 42. Ostrava: VSB – TU Ostrava.
Volume 43
Kresta, A. (2016). *Finanční modelování s aplikacemi v prostředí MATLAB*, SAEI, vol. 43. Ostrava: VSB – TU Ostrava.
Volume 44
Melecký, L., Staníčková, M. (2015). *Soudržnost a konkurenceschopnost vybraných zemí a regionů Evropské unie*, SAEI, vol. 44. Ostrava: VSB – TU Ostrava.
Volume 45
Marček, D. (2016). *Supervizované a nesupervizované učení z dat: Statistický a soft přístup*, SAEI, vol. 45. Ostrava: VSB – TU Ostrava.
Volume 46
Friedrich, V. (2017). *Postojové a hodnoticí škály v marketing a managementu. Vybrané statistické metody a aplikace*, SAEI, vol. 46. Ostrava: VSB – TU Ostrava.
Volume 47
Vrabková, I., Vaňková, I., Bečica, J., Kryšková, Š. (2017). *Příspěvkové organizace. Postavení, úkoly a technická efektivnost*, SAEI, vol. 47. Ostrava: VSB – TU Ostrava.
Volume 48
Czudek, D., Dvořáková, P., Kozieł, M., Krügerová, M., Kubica, J., Závada, V. (2014). *Vybrané právní a ekonomické aspekty islámského finančnictví*, SAEI, vol. 48. Ostrava: VSB – TU Ostrava.
Volume 49
Peterková, J. (2018). *Využití konceptů inovací v průmyslovém podniku*, SAEI, vol. 49. Ostrava: VSB – TU Ostrava.
Volume 50
Velčovská, Š. (2018). *Zhodnocení mezigeneračních postojů českých spotřebitelů ke značkám kvality potravin*, SAEI, vol. 50. Ostrava: VSB – TU Ostrava.
Volume 51
Staníčková, M. (2018). *EU Competitiveness and Resilience: Evidenc-based on Regional Leve*l, SAEI, vol. 51. Ostrava: VSB – TU Ostrava.
Volume 52

Drastichová, M. (2018). *The Theory and Measurement of Sustainable Development*, SAEI, vol. 52. Ostrava: VSB – TU Ostrava.
Volume 53
Zapletal, F. (2018). *Optimization Models for Emissions Management*, SAEI, vol. 53. Ostrava: VSB –TU Ostrava.
Volume 54
Jenčová, S. (2018). *Aplikácia pokročilých metód vo finančno-ekonomickej analýze elektrotechnického odvetvia Slovenskej republiky,* SAEI, vol. 54. Ostrava: VSB – TU Ostrava.
Volume 55
Balcar, J. (2018). *Soft Skills on the Czech Labour Market*, SAEI, vol. 55. Ostrava: VSB – TU Ostrava.
Volume 56
Vitali, S. (2018). *Pension Fund Management in a Stochastic Optimization Framework*, SAEI, vol. 56. Ostrava: VSB – TU Ostrava.
Volume 57
Hozman, J. et al. (2018). *Robust Numerical Schemes for Pricing of Selected Options*, SAEI vol. 57. Ostrava: VSB – TU Ostrava.
Volume 58
Martiník, I. (2018). *Petriho sítě s časovými známkami*, SAEI vol. 58. Ostrava: VSB – TU Ostrava.
Volume 59
Kouaissah N., Ortobelli S., Tichý T. (2018). *Portfolio Theory and Conditional Expectations: Selected Models and Applications*, SAEI, Vol. 59. Ostrava: VŠB – TU Ostrava.
Volume 60
Hodula, M. (2019). *Stínové bankovnictví v Evropě: pojetí, regulace a faktory růstu*, SAEI, vol. 60. Ostrava: VŠB – TU Ostrava.
Volume 61
Cassader, M., Tichý, T., Vitali, S. (2019). *Benchmark Tracking Portfolio Problems with Stochastic Ordering Constraints*, SAEI, vol. 61. Ostrava: VSB – TU Ostrava.
Volume 62
Kovářová, E. et al. (2019). *Společensko-ekonomické předpoklady sociálního vyloučení dětí a mládeže v České republice,* SAEI, vol. 62. Ostrava: VSB – TU Ostrava.
Volume 63
Komárek, L. et al. (2019). *Stabilization Policies: Selected Challenges in the International Context,* SAEI, vol. 63. Ostrava: VSB – TU Ostrava.
Volume 64
Krajňák, M. (2020). *Osobní důchodová daň v České republice se zaměřením na příjmy ze závislé činnosti,* SAEI, vol. 64. Ostrava: VŠB – TU Ostrava.

Volume 65
Seďa, P. (2020). *Teoretické aspekty hypotézy efektivního trhu a empirie na akciových trzích v zemích střední Evropy,* SAEI, vol. 65. Ostrava: VŠB – TU Ostrava.
Volume 66
Janků, J. (2020). *Ekonomické důsledky neinformovanosti voliče: Politicko-rozpočtový cyklus,* SAEI, vol. 66. Ostrava: VŠB-TUO.
Volume 67
Tománek, P. (2022). *Rozpočtové určení daní v ČR,* SAEI, vol. 67. Ostrava: VŠB-TUO.
Volume 68
Wu, X. (2023). *The impact of selected moderators on the relationship between CSR and profitability: Evidence form China*, SAEI, vol. 68. Ostrava: VŠB-TUO
Volume 69
Novotná, M. (2024). *Micro-Modelling Approaches for Credit Rating and Corporate Survival*, SAEI, vol. 69. Ostrava: VŠB-TUO