

Predikce ceny akcie podniku z Manahry

Václav Jež¹

12. ledna 2025

Abstrakt: Tato práce cílí na průzkum přesnosti predikcí ceny akcie v jednoduchém uzavřeném prostředí. Konkrétně se jedná o manažerskou hru, kde figurují automobilové podniky a jejich akcie. Práce využívá přímočarý přístup s využitím lineárně regresního modelu aplikovaného na časové řady získané z této hry. Soustředí se na jeden podnik a na základě jeho klíčových ukazatelů predikuje hodnotu jeho akcie. Následně pomocí metriky RMSE a vizualizace, je ověřována hypotéza, zdali lineárně regresní model dokáže spolehlivě predikovat hodnotu akcie na datech z tohoto prostředí. Práce nachází potvrzení této hypotézy v relativně nízkém RMSE a přesností predikovaných hodnot na grafu. Závěrem je cíl práce vyhodnocen jako splněný s potvrzenou hypotézou a je doporučen bližší výzkum na kointegraci.

Klíčová slova: Predikce; Akcie; Lineární regrese; RMSE; Stacionarita

JEL klasifikace: G0, G1

1 Úvod

Svět akcií a podobných volatilních aktiv může být poměrně složitý a nemusí být jednoduché se v něm vyznat. K lepší představě o tom co nám, která akcie může v budoucnu přinést mohou pomoci predikce. Avšak predikce cen akcií nebo podobných volatilních aktiv může být v reálném světě složitá a poměrně nespolehlivá. Proto si tato práce klade za cíl vyzkoušet, jak přesné mohou být právě takové predikce v menším simulovaném prostředí. Spoustu úvah a postupů vychází z učebnice Základů ekonometrie (Němec, 2019).

Moje motivace pro výběr tohoto tématu byla, že v rámci manažerské hry (Manahry) v předmětu Management byla simulována burza a my jsme mohli zhodnocovat svoje prostředky. Chtěl jsem zkusit využít nabyté znalosti ze Základů ekonometrie a zkusit predikovat hodnotu akcií, tak, abych měl na burze nějaký přehled o budoucím vývoji, a tak jistou výhodu. Průběžně jsem si obstarával data, ale nakonec jsem v průběhu semestru model vůbec nevytvořil. Napadlo mě tedy využít tato data pro tento projekt a zkusit prozkoumat, jestli predikce v takovém prostředí s využitím jednodušších metod budou fungovat relativně dobře. S takovými panelovými daty se nabízí spousta zajímavých způsobů, jak s nimi nakládat. Nejrozumnější a nejzajímavější mi ale přijde, zůstat u otázky predikcí cen akcií.

Tato práce dále seznámí čtenáře se specifickým prostředím využitých dat, teoretickým modelem vyplývajícím z tohoto prostředí a s cílem práce. Dále budou popsána využitá data pomocí tabulek a grafů, budou řešeny problémy týkající se předpokladů modelu a následně bude model odhadnut a testován na testovacích datech.

¹ Masarykova univerzita, Ekonomicko–správní fakulta, obor: Ekonomie, 535432@mail.muni.cz

2 Prostředí, teoretický model a cíl

Manahra se odehrává na trhu s automobily, kde figurují jednotlivé seminární skupiny reprezentující výrobní firmy. Fungování podniku je poměrně komplikovaná záležitost, ale tato práce se zaměřuje na burzu s akcemi. Zmíněná burza zde funguje tak, že se jí každý student Managementu může účastnit buď jako jednotlivec, nebo jako celý podnik. Na burze také obchoduje simulovaná veřejnost. Hra, jako celek se odehrává po kolech a ta probíhají 1x týdně. Zde probíhají všechny hlavní události (trh s auty, zveřejnění finančních výkazů). Avšak jednotlivá zobchodování burzy probíhá po kolech, a to vždy 7x za týden, celkem 49x + 1 poslední kolo otevřené jen pro simulovanou veřejnost. Fungování burzy bylo dobře popsáno v pravidlech (Masarykova univerzita, 2024) a dalo by se shrnout tak, že hodnota akcie se odvíjí hlavně od ukazatele „čisté obchodní jmění na akcii“, počítáno jako podíl vlastního kapitálu podniku a fixního počtu vydaných akcií. Obchodování simulované veřejnosti se odvíjí také od očekávané změny hodnoty akcie v příštím kole postavené na základě analýzy finančních výkazů a výsledků na trhu s auty. Nelze bohužel říct jakou váhu měla tato veřejnost na celkových změnách cen, protože samozřejmě s cenou hýbe i chování reálných hráčů a podniků. Toto prostředí je specifické svojí relativní jednoduchostí oproti reálné burze a také uzavřeností, kde na trhu je omezen počet účastníků, celkových hráčů a podniků.

Na základě zmíněných pravidel a vlastního uvážení je odvozena prvotní „teoretická“ specifikace modelu 1.

Model 1:

$$cena_t = obch_jmeni_akcie_{t-1} + objem_burza_{t-1} + neprodana_auta_{t-1} + HV_{t-1}$$

Cena_t označuje hodnotu akcie v Kč, obchodní jmění akcie označuje vlastní kapitál na počet emitovaných akcií, objem burzy je počet zobchodovaných akcií, neprodaná auta je počet neprodaných aut podniku na jejich trhu, a HV označuje výsledek hospodaření podniku. Za poznámku stojí, že všechny nezávislé proměnné v tomto modelu jsou uvedeny v čase t-1, protože jejich data v čase t ještě nejsou dostupná, a tak výsledná se cena v čase t odvíjí od rozhodnutí udělaných na posledně známých hodnotách. Důvody zahrnutí obchodního jmění a HV jsou popsány v pravidlech hry a proměnné neprodaná auta a obchodní objem do modelu zahrnují trh s automobily a reálné hráče a podniky.

Tato práce si klade jako cíl zodpovědět hypotézu, zdali lineární regresní model dokáže spolehlivě predikovat hodnotu akcie podniku v jednodušším omezeném prostředí Manahry. K odpovědi na tuto otázku vede bližší specifikace a následný odhad ekonometrického modelu popsané v další kapitole.

3 Data

Výzkum používá data, která byla webscrapnutá přímo ze stránek Manahry (Masarykova univerzita, 2024). Konkrétně jde o data proměnných popsaných v kapitole 2 a týkají se pouze jednoho podniku, 17, kterého jsem byl členem. Je důležité zmínit, že Manahra měla celkově 7 týdnů + „0“ výchozí kolo. Toto „0“ kolo není zahrnuto do dat vstupujících do modelu, z důvodu absence dat vysvětlujících proměnných kvůli jejich posunutí na čas t-1. Druhý důvod vynechání je stejný i pro kolo 50 a to ten, že reální hráči a podniky v tomto kole nemohli obchodovat. Kvůli tomuto externímu omezení by tato pozorování mohla zkreslovat vztahy mezi proměnnými a nepřinášet reálnou informaci do modelu, a tak jsou zahrnuta jen pozorování 1-49. Také je zde problém, že proměnné mají různou frekvenci pozorování. Jak

bylo zmíněno v kapitole 2, obchodování na burze probíhá 7x týdně. Naproti tomu zveřejnění výkazů a výsledků trhu s auty probíhá pouze 1x týdně. Tyto události jsou však zdrojem dat pro proměnné `obch_jmeni_akcie`, `neprodana_auta` a `HV`. Velikost vzorku je už tak poměrně malá, a tak agregace více frekvenčních dat (`cena`, `objem_burza`) nedává smysl. Problém je tedy řešen tak, že vždy zbylých 6 jednotek času před dalším pozorováním je doplněno poslední známou hodnotou. Jelikož se v rámci hry v těchto obdobích opravdu neměnili, dává to smysl, ale zároveň data nejsou příliš variabilní. Tudíž, všechny tyto proměnné jsou vždy pro každý po sobě jdoucí blok 7 obchodních dní konstantní.

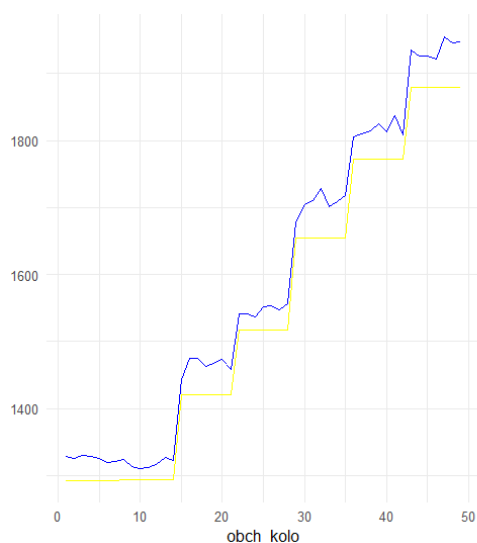
Tabulka 1: popisné statistiky zájmových proměnných

<code>cena</code>	<code>objem_burza_t-1</code>	<code>obch_jmeni_akcie_t-1</code>	<code>HV_t-1</code>	<code>neprodana_auta_t-1</code>
Min.: 1311	Min.: 1000000	Min.: 1292	Min.: 64352	Min.: 0
1st Qu.: 1329	1st Qu.: 2792534	1st Qu.: 1294	1st Qu.: 1619415	1st Qu.: 0
Median: 1548	Median: 2799965	Median: 1517	Median: 2999203	Median: 10073
Mean: 1588	Mean: 2737371	Mean: 1547	Mean: 2581454	Mean: 15698
3rd Qu.: 1809	3rd Qu.: 2809203	3rd Qu.: 1772	3rd Qu.: 3526946	3rd Qu.: 28310
Max.: 1954	Max.: 2831278	Max.: 1879	Max.: 3841500	Max.: 46134

Zdroj: vlastní zpracování v R

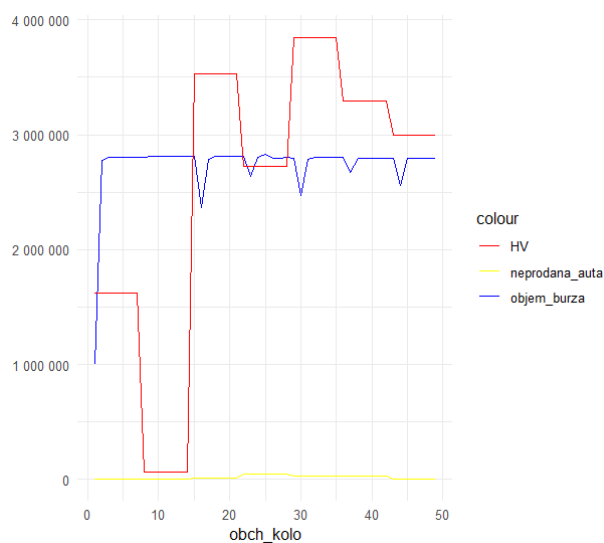
Na tabulce 1, lze vidět popisné statistiky všech proměnných. U proměnné `neprodaná auta` je vidět, že v některých obdobích se auta prodala všechna, ale v její maximální hodnotě jejich počet byl vysoký. Velká variabilita je vidět také u objemu obchodů, nebo `HV`, kde je velký rozdíl mezi maximální a minimální hodnotou. Naproti tomu proměnná `cena` je poměrně stabilní.

Graf 1: Malé měřítko



Zdroj: vlastní zpracování v R

Graf 2: Velké měřítko



Zdroj: vlastní zpracování v R

Při vizualizaci jsou proměnné rozděleny podle jejich měřítka, protože jsou ve velmi rozdílných jednotkách. Proměnnou neprodaná auta bohužel nelze přiřadit ani do jednoho, aby se dala hezky vykreslit. Na grafu 1, je vidět opravdu zajímavý, výrazný, ale očekávaný vztah mezi těmito dvěma proměnnými, jak byl popsán z pravidel v kapitole 2. Na grafu 1 lze také pozorovat rostoucí trend obou řad. Tento graf také může poukazovat na sezonnost. U ceny by mohla být přítomna kvůli odehrání hlavního kola hry každých 7 dní, které, zdá se, má pokaždé výrazný efekt.

4 Ekonometrický model

Jak je z vlastního cíle práce zřejmé, k empirickým výsledkům je použit lineární regresní model. Model je odhadnut pomocí OLS estimátoru. Model je prvně odhadnut na trénovací sadě dat (85 % celkových) a následně jsou na testovacích datech (zbytku celkových dat) odhadnuty predikce. K hodnocení přesnosti predikcí je dále využita metrika Root Mean Squared Error (RMSE) a graf jako vizuální podpora. Tato kapitola dále testuje nutné předpoklady pro model 1 a jako závěr ukáže nutné úpravy, aby předpoklady byly splněny.

1. Stacionarita

Jak bylo avizováno v předchozí kapitole u proměnné cena a obchodní jmění akcie by tento významný předpoklad mohl být porušen přítomným trendem. Při pohledu na ACF proměnné ceny (příloha 1) je opět vidět, že řada cena nejspíš nebude stacionární, protože má vysokou autokorelaci, která klesá velmi pomalu. Rozhodující informace jsou však získány z formálních testů stacionarity.

Tabulka 2: Testy stacionarity

Variable	KPSS_p_value	ADF_p_value	PP_p_value	Recommended_Diffs
cena	0,01	0,05770292	0,222970543	1
obch_jmeni_akcie_t-1	0,01	0,05111343	0,112375963	1
HV_t-1	0,013222617	0,429595057	0,473199124	1
neprodana_auta_t-1	0,089875945	0,860146696	0,83673945	1
objem_burza_t-1	0,1	0,01	0,01	0

Zdroj: vlastní zpracování v R

V tabulce je také uvedený doporučený řád difference, který je vzatý z funkce `ndiffs` (balík `forecast`). Tato funkce sama na základě testů stacionarity vrátí počet diferencí nutných k jejímu dosažení. Z tabulky lze vidět, že různé testy mají různé závěry. Vesměs se ale shodnou, že první 4 proměnné vykazují nestacionaritu. Zajímavé by dále mohlo být, podívat se na test kointegrace mezi těmito 4 nestacionárními proměnnými. Je tedy proveden Engle-Granger test kointegrace, který opravdu potvrzuje kointegraci. Nabízí se zde tedy zcela jiný model, který kointegraci může využít. Tato práce se však tímto přístupem nebude dále zabývat a bude se dále držet poměrně přímočařejší lineární regrese. Tudíž na všechny nestacionární proměnné je aplikována difference. Když je stacionarita stejným způsobem testována na diferencovaných řadách, všechny řady podmínku stacionarity splňují. Hodí se také poznamenat, že indikace sezonnosti v grafu 1 se nepotvrzuje, protože řady jsou stacionární po jedné diferenci mezi x_t a x_{t-1} .

2. Nepřítomná autokorelace reziduí

Další velmi významný předpoklad je nezávislost reziduí na sobě v čase. Tento předpoklad je otestován pomocí Ljung-Boxova testu, který vychází s p-hodnotou 0.005778 a tudíž, předpoklad vyvrací. Když je ale přidán autoregresní prvek ($cena_{t-1}$) mezi vysvětlující proměnné, p-hodnota tohoto testu vychází 0.08709. Tato hodnota indikuje nezamítnutí H_0 o nezávislosti reziduí a tudíž předpoklad autokorelace je také splněn.

3. Homoskedasticita

White's test (p-hodnota 1.00) i Breusch-Pagan test (p-hodnota 0.716) nezamítají H_0 o homoskedasticitě, a tak je tento předpoklad splněn.

4. Normalita reziduí

Shapiro-wilk test s p-hodnotou 0.9173 nezamítá H_0 o normalitě, předpoklad je tedy splněn.

5. Multikolinearita

Hodnoty VIF vychází pro všechny proměnné menší než 10, a tak je tento předpoklad také splněn.

Je také vhodně zahrnout proměnnou objem burza a HV v logaritmu, protože mají naprosto odlišné jednotky od ostatních proměnných, a takto je snadnější je interpretovat. Na žádný závěr testu toto nemělo vliv. Je na místě úpravy, udělané v této kapitole, shrnout do finální upravené specifikace modelu, která bude použita na predikce. Tato specifikace byla znovu testována na všechny testy a jejich závěry byly stejné.

Model 2: finální model

$$\Delta cena_t = \Delta cena_{t-1} + \Delta obch_jmeni_{t-1} + \log(objem_burza_{t-1}) + \Delta neprodana_auta_{t-1} + \Delta \log(HV_{t-1})$$

5 Odhadnutý model

Tabulka 3: odhad modelu 2

Dependent variable: Δ cena			
part	term	statistic	(1)
estimates	(Intercept)	estimate	5750.012***
estimates	(Intercept)	std.error	(1653.242)
estimates	Δ cena_t-1	estimate	-0.256**
estimates	Δ cena_t-1	std.error	(0.117)
estimates	Δ obch_jmeni_t-1	estimate	0.840***
estimates	Δ obch_jmeni_t-1	std.error	(0.049)
estimates	log(objem_burza_t-1)	estimate	-387.173***
estimates	log(objem_burza_t-1)	std.error	(111.345)
estimates	Δ neprodana_auta_t-1	estimate	0.000
estimates	Δ neprodana_auta_t-1	std.error	(0.000)
estimates	Δ log(HV_t-1)	estimate	2.222
estimates	Δ log(HV_t-1)	std.error	(2.108)
Num.Obs.			39
R2			0.924
R2 Adj.			0.913
F			80.449
RMSE			9.08

V závorkách jsou uvedeny směrodatné chyby, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, Zdroj: vlastní zpracování v R

Na tabulce odhadnutého modelu můžeme vidět, že model je poměrně vysoce přesný z hlediska vysvětlované variability proměnné R2, která je 92 %. Relativně nízký se zdá indikátor RMSE na trénovacích datech, kdy model se v průměru spletl jen o 9 Kč. V modelu sice figurují statisticky nevýznamné proměnné (neprodana_auta, HV), ale rozhodl jsem se je v modelu nechat i na základě p-hodnoty RESET testu 0.334, což indikuje nezamítnutí H0: správné specifikace modelu. Model je tedy připravený na predikce. Bohužel z důvodu málo pozorování je i testovacích dat málo (8).

6 Predikce modelu

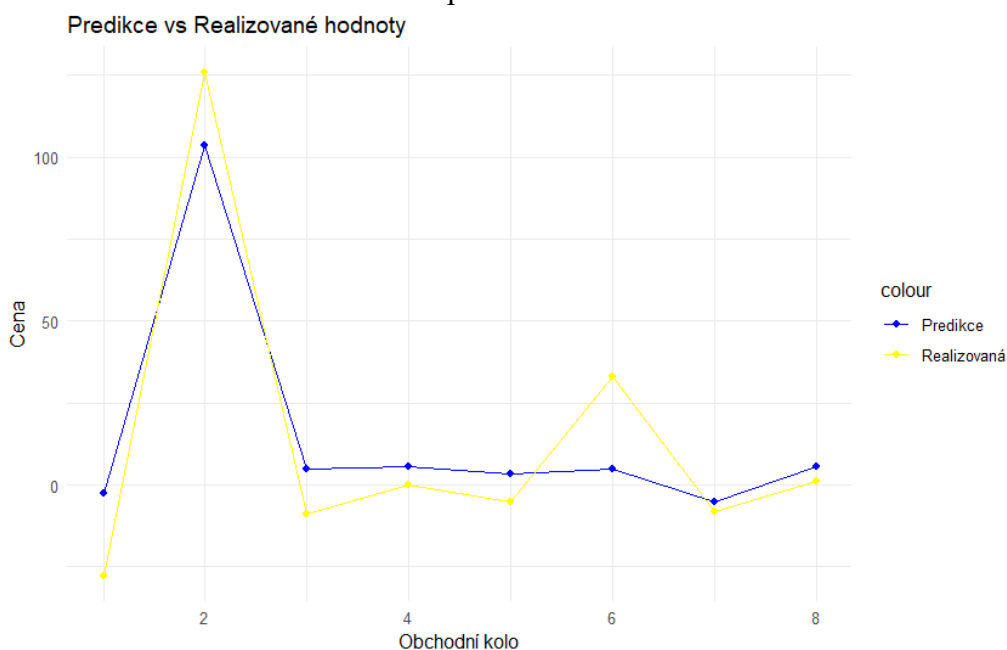
Z důvodu transformace závislé proměnné diferencí, model samozřejmě predikuje změnu ceny v budoucím čase. MSE predikcí je: 283.7 a RMSE: 16.8.

Tabulka 4: predikce modelu 2

predikce	realizovaná
2,733426512	-28
103,4339965	126
4,826596002	-9
5,721780813	0
3,347393568	-5
4,71535997	33
-	-8
5,070246348	
5,481267352	1

Zdroj: vlastní zpracování v R

Graf 3: predikce modelu 2



Zdroj: vlastní zpracování v R

7 Závěr

Na výsledcích je vidět, že model má průměrnou chybu okolo 16,8 Kč. Je poněkud vyšší než na trénovacích datech, což může značit, že model neumí tak dobře generalizovat vztahy z těchto trénovacích dat na nová data. Pořád se ale jedná o nízkou hodnotu, když se připomene,

že směrodatná odchylka změny ceny je 35.5. Na grafu je vizuálně vidět, že model dokáže predikovat dobře. Bylo očekáváno, že model by mohl predikovat hodnoty akcie dobře, díky popsaným vztahům v kapitole 2. Graf vypadá ale opravdu dobře a předčil očekávání. I metrika průměrné chyby v predikcích, která je poloviční než směrodatná odchylka značí úspěch.

Cíl práce je tedy označen jako splněný. Odpověď na hypotézu, zdali dokáže lineární regrese spolehlivě predikovat ceny akcie podniku v tomto omezeném prostředí je její potvrzení. Přínos práce je hlavně v poměrně přímočarém a jednoduše interpretovatelném postupu, který se ukázal jako efektivní. Téma práce je velice zajímavé a určitě si zaslouží bližší pozornost. Doporučení pro bližší prozkoumání se týká využití ECM nebo VECM modelu na bližší analýzu nalezené kointegrace, kterou tato práce pouze zmiňuje.

Reference

- [1] Němec, D. 2019. *Základy ekonometrie*. Brno: Ekonomicko-správní fakulta Masarykovy univerzity
- [2] Masarykova univerzita. 2024. *Manahra*. [online]. Dostupné z: <http://manahra.cz/>

A Přílohy

Všechny testy, přechodné modely, a data jsou k nalezení v R scriptu v archivu odevzdaném s tímto souborem.

Příloha 1: ACF řady cena

