

Jihočeská univerzita v Českých Budějovicích

Ekonomická fakulta

Katedra aplikované matematiky a informatiky

Teze diplomové práce

Regresní studium ekonomických jevů a politické stability

Vypracoval: Bc. František Žiška

Vedoucí práce: RNDr. Tomáš Roskovec, Ph.D.

České Budějovice

2023

Klíčová slova: logistická regrese, ekonomické ukazatele, volby, predikce

Anotace:

Tato práce zkoumá vztahy mezi mnoha různými ekonomickými ukazateli a změnou vedoucí strany ve vládě. Tyto vztahy nebo vlivy mohou být malé a nevýznamné, nebo mohou být statisticky významné. Teoretická část nejprve vysvětluje, jak ekonomika ovlivňuje politiku a naopak. Druhá část teorie obsahuje všechny ekonomické ukazatele použité v analytické části této práce s několika základními vysvětleními ukazatelů a několika příklady vývoje uvedených ukazatelů. Další částí práce je metodika, kde je podrobně vysvětlen sběr dat a popsány všechny analytické nástroje a postupy. Poslední část, praktická, se zaměřuje na využití metod uvedených v metodice k nalezení statisticky významného vztahu mezi ekonomickým ukazatelem a změnou předsednictví a výsledkem je zjištění, které ukazatele nejvíce ovlivňují volby. Kromě toho byl vytvořen kód v programovacím jazyce Python, který byl použit k vytvoření všech analýz v této práci.

Úvod a cíl

„V praxi existuje silný vztah mezi ekonomikou a politikou, protože výkonnost ekonomiky je jedním z klíčových politických problémů.“ (Asiimwe, 2015)

Lidé ve státě, kteří mají právo volit, jsou často předem rozhodnutí, jakého kandidáta budou v následujících volbách volit. Toto rozhodnutí voliče může mít mnoho důvodů, ať už rozhodnutí z minulých zkušeností nebo například rozhodnutí, kterému napomohla velká volební kampaň, kterou kandidát zorganizoval. Ale co když volič činí toto rozhodnutí i z jiných důvodů, jako například vysoká míra inflace či zadlužení státu, které je v poslední době velice diskutované téma? Ekonomika státu se týká všech subjektů na daném území, ať už se jedná o nezaměstnanou osobu či o osobu se zaměstnáním. Proto zde nastává otázka: „Má ekonomika ve státě vliv na výsledky voleb?“

Cílem této práce je zjistit a které ekonomické ukazatele mají vliv na změnu současné vlády. Zda se voliči rozhodnou změnit současnou vládu, pokud je ekonomický stav nepříznivý, anebo jestli budou věřit ve stabilitu a zvolí stejnou vládu, která byla v čele jako doposud.

Ekonomické ukazatele, která tato práce bude zkoumat, jsou míra inflace, míra nezaměstnanosti, veřejný dluh, hrubý domácí produkt, ratingová známka a vývoz a dovoz. Vybraná oblast analýzy a sběru dat o těchto ekonomických ukazatelích jsou státy Evropské unie za posledních 31 let, od roku 1990 až do roku 2021.

V této práci je představena analýza vztahu mezi ekonomickými ukazateli a změnou vlády za pomoci historických dat. Analýza v této práci je z velké části zaměřena na logistickou regresi, která zkoumá závislost jedné binární proměnné (změna vlády) na jedné či více nezávislých proměnných.

Přínosem této práce je zjištění, zda a jak velký vliv mají jednotlivé ekonomické ukazatele na změnu vlády. Pro zkoumání tohoto vztahu jsem také připravil zdrojový kód v programovacím jazyce Python, který může být využit na analýzu dat o změně vlády a o ekonomických ukazatelích pro jinou sadu dat, například chování voličů v konkrétním státě či regionu.

Metodika

Cílem této práce je zjistit, pokud a v jaké míře mají vybrané makroekonomické ukazatele vliv na politickou stabilitu státu.

Sběr dat je nedílnou součástí každé statistické studie. Existuje mnoho různých metod sběru dat, které se používají, v závislosti na cíli a požadované kvalitě dat. Mezi nejčastější metody sběru dat patří dotazníky, průzkumy, rozhovory, pozorování a experimenty. V této práci se zaměříme na sběr historických dat o volbách ve vybraných státech a hodnot ekonomických ukazatelů v daných zemích v období voleb. Tato forma dat je jasně daná a většinou lépe popisují realitu než průzkumy a experimenty například formou dotazníků, které se obecně nemusí zcela shodovat s výsledky voleb a s daty, která zkoumáme.

Nejdříve je potřeba určit, pro které země by studium vlivu ekonomických ukazatelů mělo vysoký význam. V této práci jsem se rozhodl, že budu sledovat především státy Evropské unie. Důvod výběru této skupiny států je zjištění rozdílů ve vlivu ekonomických ukazatelů na změnu vlády v evropských zemích.

Dále je zapotřebí stavit časový úsek, od kterého a do kterého tato data sbírat. Po dlouhém rozhodování jsem stanovil rok 1990 jako počáteční rok sběru historických dat o volbách a ekonomických ukazatelů a jako konečný rok jsem stanovil 2021, protože data pro rok 2022, především ze strany ekonomických ukazatelů, nejsou prozatím přesná, upravená nebo jednoduše dohledatelná. Takže uvažujeme každý stát Evropské unie od roku 1990 do roku 2021, což znamená přes 30 let historických dat o volbách a ekonomických ukazatelích pro každou zemi EU.

V tuto chvíli máme stanovené státy a časový úsek. Dalším problémem u sběru dat politických voleb je důležité určit, které volby jsou v daném státě „nejdůležitější“. S jistotou lze usoudit, že komunální volby nebudou těmi hlavními volbami. Takže nám zbývají volby prezidentské a volby do parlamentu, který má na starost sestavení vlády. Zde může být volba správných dat matoucí, ale záleží především na politickém systému daného státu. Příkladem států, pro které jsem vybral prezidentské volby jako hlavní volby, jsou z velké většiny poloprezidentské republiky, jako například Francie a Kypr. Ve většině států Evropské unie jsou hlavní volby do parlamentu, na základě kterých se skládá vláda. Zde jsem se rozhodl vybírat ve většině případů předsedu vlády a jeho volební stranu či koalici za vyhrávající stranu ve volbách, protože ve většině případů strana či koalice s největším počtem hlasů vyhrává post předsedy vlády. Samozřejmě, že to tak být nemusí, ale v absolutní většině případů se tato hypotéza potvrdila.

Z těchto dat je poté potřeba zjistit, kdy proběhly změny ve volbách. Zde máme z datového setu, který jsem vytvořil dvě možnosti, jak stanovit změnu ve vládě, a to:

- Změna předsedy vlády se počítá jako změna ve vládě, anebo
- Změna výherní politické strany či koalice se počítá jako změna ve vládě.

Dle mého zvážení dává větší smysl změna výherní politické strany, protože pokud se změní předseda vlády, tak pořád je zde možnost, že patří ke stejné politické straně jako předešlý předseda. Z druhé strany pohledu je velice nepravděpodobné, že stejný předseda vlády bude znovu zvolen, ale výherní strana bude jiná. Samozřejmě, že je zde možnost, že stejný předseda vlády je součástí jiné volební strany, ale v datech z Evropské unie jsem se setkal pouze s tím, že se daná strana pouze přejmenovala či byla součástí jiné koalice, ale hlavní zaměření na politickém spektru dané strany bylo stejné.

V teoretické části práce byly popsány ekonomické ukazatele použité v praktické části. Mezi ně patří inflace, nezaměstnanost, hrubý domácí produkt, veřejný dluh a obchodní bilance. Po zvážení bylo nutné upravit některé ekonomické ukazatele na jiný tvar, ale ve své podstatě jsou to ukazatele blízko spojené ke své původní podobě.

Nejdříve je zapotřebí určit, v jakém časovém pásmu data budeme sbírat. Výběr časového úseku pro ekonomické ukazatele v této práci byl celý rok. Důvod výběru celého roku oproti čtvrtletí či měsíce byla především reakce lidí na ekonomické ukazatele. Pokud se volby konají na začátku roku, tak pokud nová vláda dokáže změnit za celý rok průměrné ukazatele, tak to je jenom dobře pro tuto práci, což ukazuje, že zlepšení situace od minulé vlády vede ke změně vlády. Pokud se volby konají na konci roku, tak roční ukazatele také dávají smysl, protože pokud v daný rok tamní obyvatelstvo zažívá ekonomickou krizi a ekonomické ukazatele, jako například inflace, skokově vzroste, tak je velice pravděpodobné, že nespokojené obyvatelstvo bude chtít současnou vládu změnit.

Po nasbírání všech ekonomických dat a připojení k datům politickým, tak celkový počet pro výpočty je přes 30 souborů s daty o státech. Je třeba brát v potaz, že krom informací o jednotlivých státech, byl vytvořen i datový zdroj, který bere celou Evropskou unii jako jeden velký datový zdroj. Tento zdroj je využit jako hlavní zdroj dat a využit v praktické části na ukázky.

Korelační analýza

Pro vytváření statistických a analytických poznatků je nejdříve důležité zjistit, pokud a jak moc na sobě všechny pozorované proměnné. Rozsah korelace může být záporný i kladný a pohybuje se od -1 do 1. Čím více se korelace blíží -1 a 1, tím větší je závislost mezi danými proměnnými. Na druhou stranu čím více se korelace pohybuje kolem 0, tím jsou na sobě proměnné méně závislé.

Korelace je základní statistický a analytický nástroj, který dokáže velice jednoduše zobrazit souvislosti mezi daty. Obecně pro většinu modelů je výhodné mít data, kde nezávislé veličiny mají mezi sebou co nejmenší korelaci a zároveň vysokou korelaci se závislou proměnnou.

Likelihood Ratio Test (LLR)

Tento test je při logistické regresi potřeba provést, aby bylo jasné, zda model se všemi proměnnými (ekonomickými ukazateli) je statisticky signifikantní či nikoli.

LLR test můžeme rozdělit do dvou částí:

- Když se použijí hodnoty nezávislých proměnných pro predikci
- Když se hodnoty nezávislých proměnných nevyužijí pro predikci

Díky tomuto testu můžeme poznat, pokud daný model se všemi prediktory jako celek představuje statisticky signifikantní model pro predikci či nikoli. Pokud vyjde p-value menší než 5 %, tak můžeme usoudit, že daný model s nezávislými proměnnými má vliv na predikci závislé proměnné nebo přesněji můžeme zamítnout nulovou hypotézu. Na druhou stranu, pokud p-value vyjde větší než 5 %, tak nelze vyvrátit nulovou hypotézu. To znamená, že při použití modelu logistické regrese bez nezávislých proměnných lze predikovat závislou proměnnou a nezávislé proměnné nemají vliv na predikci.

Regresní analýza

Protože formát vstupních dat této práce je jedna závislá binární proměnná a více nezávislých spojitých proměnných, tak lze zde využít logistickou regresní analýzu a následní testování statistických hypotéz.

Nejdříve je důležité si stanovit nulovou hypotézu, což je tvrzení, které se snažíme vyvrátit. Dále je zapotřebí stanovit hypotézu alternativní, která se snaží vyvrátit nulovou hypotézu. Poslední

důležitým faktorem je hladina významnosti. Ta se ve většině případů stanovuje na 5 %, čímž budeme mít jistotu, že z 95 % jsme vybrali správně rozhodnutí.

V této práci je stanovena nulová hypotéza jako „Daná závislá proměnná není závislá na nezávislé proměnné“. Tato hypotéza nastane v případě, pokud test hladiny významnosti (p-value) bude vyšší než 0,05. Bude to znamenat, že nezávislá proměnná nemá statisticky významný vliv na závislou proměnnou.

Dále v této práci je využita alternativní hypotéza ve tvaru „Daná závislá proměnná je závislá na nezávislé proměnné“. Tato hypotéza vyvrací nulovou hypotézu a situace ve prospěch alternativy nastane ve chvíli, kdy daný ukazatel bude mít hladinu významnosti (p-value) menší než 0,05. Takový výsledek bude vypovídat o určité závislosti mezi závislou a nezávislou proměnnou.

Logistická regrese

Logistická regrese je jednou ze základních statistických modelů pro klasifikaci. Tento model je založen na predikci pravděpodobnosti příslušnosti pozorování k jedné z vybraných kategorií. V této práci jsou kategorie dvě a to, jestli nastala změna vlády (jednička) či nikoli (nula).

U logistické regrese se pro výpočet využívá metoda maximální věrohodnosti (Maximum likelihood).

„Nechť X označuje vektor prediktorů $\{X_1, X_2, \dots, X_k\}$, pro první pokus o modelování odezvy lze použít běžnou lineární regresí

$$E\{Y|X\} = X\beta$$

protože očekávání binární proměnné Y je $\text{Prob}\{Y = 1\}$.“ (Harrelll, 2015)

Z hlediska pravděpodobnosti, že $Y = 1$ při daném X , hodnoty prediktorů jsou následující:

$$\text{Prob}\{Y = 1|X\} = [1 + \exp(-X\beta)]^{-1}$$

„ $X\beta$ znamená $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$. Binární logistický regresní model byl vyvinut především Coxem a Walkerem a Duncanem. Regresní parametry β jsou odhadnuty metodou maximální věrohodnosti“ (Harrelll, 2015)

Samotná logistická funkce má takovéto vyjádření:

$$P = [1 + \exp(-x)]^{-1}$$

Výsledky a diskuse

Finální výsledek práce je zjištění, že záleží převážně na míře nezaměstnanosti s využitím modelu dat vycházejícího z metodologie.

Obrázek 1: Souhrn logit nezaměstnanosti

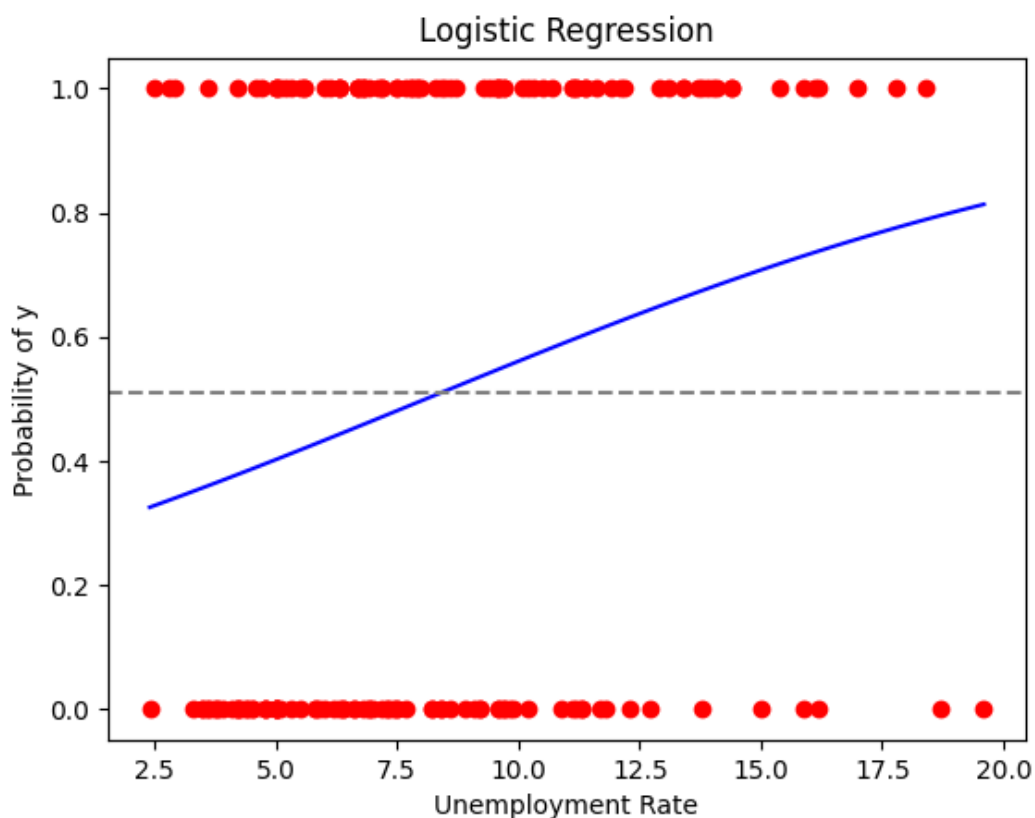
Logit Regression Results						
=====						
Dep. Variable:	Change?	No. Observations:	167			
Model:	Logit	Df Residuals:	165			
Method:	MLE	Df Model:	1			
Date:	Sat, 08 Apr 2023	Pseudo R-squ.:	0.03822			
Time:	11:26:49	Log-Likelihood:	-111.31			
converged:	True	LL-Null:	-115.73			
Covariance Type:	nonrobust	LLR p-value:	0.002936			
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	-1.0357	0.403	-2.570	0.010	-1.826	-0.246
Unemployment Rate	0.1280	0.045	2.858	0.004	0.040	0.216
=====						

Zdroj: Autor práce

Ze souhrnu na Obrázku 1 můžeme vidět, že nezaměstnanost má velký vliv na změnu vlády. Zamítáme nulovou hypotézu LLR testu, u kterého vyšlo p-value menší než 5 % a stejně tak zamítáme nulovou hypotézu u logistické regrese, protože p-value nezaměstnanosti vyšlo 0,004, což je méně než 5 %. Nyní můžeme usoudit, že proměnná podezřelá v základní korelační analýze se nám potvrdila, a že míra nezaměstnanosti má statisticky významný vliv na změnu vlády při volbách.

Obrázek 2: Grafické znázornění dat nezaměstnanosti



Zdroj: Autor práce

Na Obrázku 2 můžeme vidět, že čím vyšší je nezaměstnanost, tím vyšší je šance toho, že se změní vláda. Tento vliv je očekávaný, protože pokud je více lidí nezaměstnaných, tím může být více lidí, kteří jsou nespokojeni se současnou vládou a chtějí změnu.

Závěr

Cílem této diplomové práce bylo zjistit vlivy vybraných ekonomických ukazatelů na politickou stabilitu. Politická stabilita byla definována jako žádná změna ve vládě.

Praktická část nejdříve pojednává o práci s daty, kdy bylo zapotřebí data očistit od odlehlých hodnot. Po očištění těchto hodnot přišla na práci samotná analýza dat. Korelační analýza vyvodila několik předběžných výsledků, nejdůležitější byla nezanedbatelná pozitivní korelace mezi nezaměstnaností a závislou proměnnou o změnách vlády.

V hlavní části regresní analýzy pomocí logistické regrese bylo nejdříve zjištěno pomocí Likelihood Ratio Test (LLR), že při použití všech pěti nezávislých proměnných model logistické regrese není dostatečně věrohodný. Proto po testování byl odebrán jeden ekonomický ukazatel, který věrohodnost značně snižoval, a byl to dluh k HDP. Po stanovení dostatečné věrohodnosti modelu se čtyřmi ekonomickými ukazateli byla provedena regresní analýza pomocí regrese na všechny ekonomické ukazatele jednotlivě. Jediný ukazatel, který dokázal vyvrátit nulovou hypotézu o chybějícím vlivu na změnu vlády, byl míra nezaměstnanosti. Míra nezaměstnanosti se ukázala jako velmi vlivným ukazatelem při změně vlády.

V části testování a povahy dat je popsán průběh testování několika datových zdrojů u regresní analýzy a je zde popsáno, jak se jednotlivé datové zdroje od sebe lišily a jaké byly výsledky analýzy těchto dat. V poslední kapitole práce byl představen vedlejší produkt této práce. Je to skript pro predikci voleb, který se dokáže naučit na zadaných historických datech a dokáže předpovídat změnu ve vládě při zadané míře ekonomických ukazatelů.

Výstupem této práce je závěr, že ze zkoumaných ekonomických ukazatelů je statisticky významný pouze jeden z nich, míra nezaměstnanosti. Nezaměstnanost může mít velký vliv na rozhodování obyvatelstva, zda se vláda změní či nikoli. Dalším výstupem této práce jsou zdrojové kódy pro analýzu výpočetních technik uvedených v metodologii a posledním výstupem je zdrojový kód programu pro predikci voleb při uživatelem zadaných parametrech.

Mým plánem do budoucna s touto prací je pokračovat ve sběru dat o více ekonomických ukazatelích a porovnávat více zemí, rozdíly mezi danými zeměmi. Kromě toho chci také rozšířit tuto práci o porovnání různých kontinentů, a na jakých kontinentech je jaký ekonomický ukazatel důležitější, a který je zase méně významný. Příkladem mohu uvést, že při prvních pokusech o porovnání Evropské unie a Spojených států amerických (USA) jsem zjistil, že pro voliče v USA je důležitější růst HDP než pro voliče v Evropské unii.

Seznam literatury použité v tezi

- [1] Asiimwe, D. (2015). The Relationship between Politics, Economics and Development. Kampala: Kampala International University. Retrieved from:
https://www.academia.edu/13783141/The_Relationship_between_Politics_Economics_and_Development
- [2] Harrell, F. E. (2015). Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis (2nd ed.). Springer International Publishing.

Poznámka: Plný seznam použité literatury je uveden v textu diplomové práce

Další bibliografické údaje

Type of thesis: master

Author: František ŽIŠKA

Academic Year: 2022/2023

Department: Department of Applied Mathematics and Informatics, Faculty of Economics, University of South Bohemia in Ceske Budejovice, Czech Republic

Thesis supervisor: RNDr. Tomáš Roskovec, Ph.D.

Number of pages: 65

Language of thesis: CZ

Title of Thesis: Regression study of economic indicators and political stability

Annotation:

This thesis explores relationships among many different economic indicators and a change of leading party in a government. These relationships or influences may be small and insignificant, or they can be statistically significant. Firstly, theoretical part explains how economics influence politics and vice versa. Secondly, theoretical part includes all economic indicators used in the analytical part of this thesis with some basic explanations of the indicators and some examples of a development of the said indicators. Next part of the thesis is methodology, where the collection of data is thoroughly explained, and all the analytical tools and practices are described there too. The last part, the practical part, focuses on using the tools mentioned in methodology to find a statistically significant relationship between the economic indicator and the change of presidency and resulting in a statement, which indicators influence elections the most. Apart from that, there is a code in Python which was used to create all the analysis in this paper.

Key words: logistic regression, economic indicators, elections, Python, prediction