VECM modeli u teoriji signala

Brnjić, Marin

SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Preddiplomski sveučilišni studij elektrotehnike

Završni rad

VECM MODELI U TEORIJI SIGNALA

Rijeka, srpanj 2022.

Marin Brnjić
0069087163

SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Preddiplomski sveučilišni studij elektrotehnike

Završni rad

VECM MODELI U TEORIJI SIGNALA

Mentor: doc. dr. sc. Ivan Dražić

Rijeka, srpanj 2022.

Marin Brnjić
0069087163

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
POVJERENSTVO ZA ZAVRŠNE ISPITE

Rijeka, 18. ožujka 2022.

Zavod: **Zavod za matematiku, fiziku i strane jezike**
Predmet: **Inženjerska matematika ET**
Grana: **1.01.07 primijenjena matematika i matematičko modeliranje**

ZADATAK ZA ZAVRŠNI RAD

Pristupnik: **Marin Brnjić (0069087163)**
Studij: Preddiplomski sveučilišni studij elektrotehnike

Zadatak: **VECM modeli u teoriji signala // VECM models in signal theory**

Opis zadatka:

U radu je potrebno precizno definirati VECM modele u kontekstu slučajnih procesa. Za dane modele potrebno je opisati matematička i statistička svojstva te se osvrnuti na njihove prednosti i mane. Obrađene modele potrebno je staviti u kontekst primjene, a posebno se treba osvrnuti na primjere njihove implementacije u obradi signala i elektrotehnici.

Rad mora biti napisan prema Uputama za pisanje diplomskih / završnih radova koje su objavljene na mrežnim stranicama studija.

Brnjić

Zadatak uručen pristupniku: 21. ožujka 2022.

Mentor:

Doc. dr. sc. Ivan Dražić

Predsjednik povjerenstva za završni ispit:

Prof. dr. sc. Viktor Sučić

IZJAVA

Sukladno članku 8. Pravilnika o završnom radu, završnom ispitu i završetku prediplomskih sveučilišnih studija/stručnih studija Tehničkog fakulteta Sveučilišta u Rijeci od 1. veljače 2020., izjavljujem da sam samostalno izradio završni rad prema zadatku preuzetom dana 21.03.2022.

Rijeka, 08..07.2022.

Brnjić

Marin Brnjić

Zahvaljujem se mentoru doc. dr. sc. Ivanu Dražiću, na velikoj podršci i motivaciji prilikom pisanja završnoga rada. Također, veliko hvala mojoj obitelji i djevojci koji su mi bili velika podrška tijekom studija.

Sadržaj

1. Uvod **3**

2. Uvod u analizu vremenskih nizova **5**

2.1. Osnovni elementi teorije vjerojatnosti 5

2.2. Osnovni pojmovi iz teorije vremenskih nizova 8

2.3. Matematički model vremenskog niza 9

2.4. Stacionarni procesi 10

2.5. Primjeri slučajnih procesa 12

2.5.1. Bijeli šum 12

2.5.2. Gaussov bijeli šum 12

2.6. Autokovarijanca i autokorelacija 13

2.7. Autoregresijski modeli 14

2.7.1. Problem egzistencije AR procesa 15

2.8. Vektorski autoregresijski modeli 15

3. Kointegracija i korekcija pogreške **17**

3.1. Definicija kointegracije 17

3.2. Pojam ispravljanja pogreške 17

4. VECM model **19**

4.1. Multivarijantna VECM specifikacija 19

4.2. VECM model prema Johansenu 19

5. VECM u programskom paketu Stata **21**

5.1. Unos podataka 21

5.2. Deklariranje skupa podataka kao vremenskog niza 22

5.3. Određivanje optimalnog broja lagova 22

5.4. Model korekcije vektora pogreške 23

5.5. Generiranje predikcije 23

6. Empirijsko istraživanje **25**

6.1. Predikcija cijene Bitcoin-a 25

6.1.1. Uvodno o Bitcoin-u 25

6.1.2. Određivanje optimalnog broja lagova 26
6.1.3. Ponašanje za vrijeme padajućeg trenda 27
6.1.4. Ponašanje za vrijeme stacionarnog trenda 27
6.1.5. Ponašanje za vrijeme rastućeg trenda 28
6.1.6. Ponašanje tijekom dužeg vremenskog perioda 30
6.2. Predikcija proizvodnje, uvoza i izvoza prirodnog plina 30
6.2.1. Uvod 31
6.2.2. Određivanje optimalnog broja lagova 32
6.2.3. Proizvodnja prirodnog plina u Hrvatskoj 32
6.2.4. Uvoz prirodnog plina u Hrvatskoj 34
6.2.5. Izvoz prirodnog plina u Hrvatskoj 35

7. Zaključak **36**

Literatura **38**

Sažetak i ključne riječi **39**

Summary and key words **40**

1. Uvod

U gotovo svakom suvremenom inženjerskom poslu teži se analizom baza podataka dobiti što točnija buduća stanja. Primjerice, svakom poduzeću je predekcija budućih potreba neophodna, kako bi unaprijed mogle odrediti potrebne resurse za narednu godinu. Tako se npr. 2021. godine dogodila nestašica čipova u autoindustriji zbog loše analize tržišta tj. nije se očekivala velika potražnja za vrijeme globalne pandemije što se moglo izbjeći boljom analizom budućih stanja potražnje. Poznavati buduće stanje je nedvojbeno korisna informacija u bilo kojoj grani industrije, no postavlja se pitanje je li isto moguće postići uz visoki postotak točnosti.

Temeljna svrha ovog rada pod nazivom VECM modeli u teoriji signala jest empirijskim istraživanjima utvrditi pouzdanost i točnost VECM modela pri analizi budućih stanja signala. Točnije, bavimo se ispitivanjem odnosa između zadanih varijabli unutar empirijskih istraživanja kojima se u konačnici predviđa buduće stanje.

Signal se matematički modelira kao preslikavanje koje nekom vremenskom trenutku pridružuje vrijednost, tj. kao funkcija $t \mapsto x(t)$. Kako je signal matematički kocept, na njega se u matematičkom smislu stoga može gledati iz više kutova, a u ovom radu ćemo na signal gledati kao na vremenski niz, odnosno pretpostaviti ćemo da se radi o signalima u diskretnom vremenu. Također, kako nam je naglasak na predikcijama budućih stanja na njih možemo gledati samo stohastičkim, a ne determinističkim pristupom. Drugim riječima, na signal ćemo u ovom radu pretežito gledati kao na vjerojatnosno-statistički koncept.

VECM model pretpostavlja višedimenzionalni signal, tj. omogućava paralelno praćenje međusobno povezanih komponenti što ga čini prilično moćnim alatom.

Za početak kako bi uopće mogli razmatrati predikcije budućih stanja bitno je uvesti i objasniti teorijske pojmove nužne za definiranje i analizu VECM modela. Stoga su drugo i treće poglavlje posvećeni pojašnjenju nužnih pojmova za shvaćanje samog VECM modela poput definicija vjerojatnosti, vremenskog niza, slučajnih procesa, autokovarijance, autokorelacije, autoregresije, kointegracije, itd.

Nadalje, sljedeće poglavlje teoretski i matematički pokriva sam VECM model gdje se upoznajemo sa njegovim parametrima, specifikacijama i trendovima.

U petome poglavlju dane su detaljne upute za generiranje predikcija pomoću VECM modela unutar statističkog softverskog paketa STATA. Te u konačnici dolazimo do empirijskih istraživanja u šestome poglavlju gdje smo za dva primjera, unutar statističkog softverskog paketa STAT pomoću VECM modela generirali predikcije te analizirali dobivene rezultate.

U prvom primjeru predviđa se cijena kripto valute Bitcoin. (Bitcoin je decentralizirana, distribuirana i anonimna platna mreža koja funkcionira pomoću kompleksnog algoritma kojeg je izumio čovjek ili skupina ljudi pod pseudonimom Satoshi Nakamoto.) Varijable koje su uključene u empirijsko istraživanje cijene Bitcoin-a su Cijena zatvaranja, Dnevni maksimum, minimum i Volumen. Podaci korišteni u istraživanju odnose se na vremensku seriju od gotovo 10 mjeseci točnije od veljače 2021. godine do prosinca 2021. godine.

U drugom primjeru predviđa se proizvodnja, uvoz i izvoz prirodnog plina u Hrvatskoj. Varijable koje su uključene u empirijsko istraživanje su Proizvodnja, Uvoz i Izvoz. Podaci su preuzeti sa stranica Hrvatskog zavoda za statistiku, a odnose se na vremensku seriju od siječnja 2016. godine do ožujka 2022. godine.

Na samom kraju rada prikazani su rezultati analize modela te donesena zaključna razmatranja.

2. Uvod u analizu vremenskih nizova

U sklopu ovog poglavlja objasnit ćemo sve teorijske pojmove nužne za definiranje i analizu VECM modela. Ovo je poglavlje obrađeno koristeći reference [1, 2, 3, 4].

2.1. Osnovni elementi teorije vjerojatnosti

Teorija vjerojatnosti relativno je nova matematička disciplina koja se počela razvijati u 17. stoljeću. Na vjerojatnost možemo gledati kao na funkciju koja kvantificira šanse da se neki određeni događaj može dogoditi, a formalna matematička definicija dana je u nastavku.

Definicija 2.1. *Neka je Ω skup ishoda nekog eksperimenta nad kojim je definiran prostor događaja F. Funkciju $P : F \to \mathbb{R}$ zovemo **vjerojatnost** ako vrijedi:*

1. $P(A) \geqslant 0,\ \forall A \in F$,
2. $P(\Omega) = 1$,
3. $P\left(\bigsqcup_i A_i\right) = \sum_i P(A_i)$,

*pri čemu $\bigsqcup$ označava uniju disjunktnih događaja. Trojku (Ω, F, P) zovemo **prostor vjerojatnosti**.*

U sljedećem primjeru objasnit ćemo pojmove iz prethodne definicije.

Primjer 2.1. *Promatramo eksperiment bacanja igraće kocke. Tada je*

$$\Omega = \{1, 2, \cdots, 6\} \tag{2.1}$$

skup ishoda, a

$$F = \{0, \{1\}, \{2\}, \cdots, \{2, 3\}, \cdots, \Omega\} \tag{2.2}$$

prostor događaja. Tada primjerice vijedi:

$$P(\{2\}) = \frac{1}{6}, \qquad P(\{2, 3\}) = \frac{2}{6}. \tag{2.3}$$

Općenito vrijedi

$$P(A) = \frac{|A|}{|\Omega|}, \tag{2.4}$$

gdje $|\cdot|$ označava broj elemenata u skupu A. Ovakva definicija vjerojatnosti na skupu jednako vjerojatnih ishoda često se naziva klasičnom definicijom vjerojatnosti.

Sada ćemo definirati nešto apstraktniji pojam slučajne varijable.

Definicija 2.2. *Neka je zadan prostor vjerojatnosti (Ω, F, P). Funkciju koja svakom ishodu pridružuje realan broj zove se **slučajna varijabla**.*

Neformalno možemo reći, da ako kod klasične varijable znamo i vjerojatnost realizacije svake njene vrijednosti ta varijabla postaje slučajna varijabla.

Ako je skup mogućih vrijednosti slučajne varijable konačan ili diskretan, zovemo je diskretnom. S druge strane ako je skup vrijednosti kontinuiran zovemo je neprekidnom. Tako je primjerice broj igrača u nekoj igri na sreću je diskretna slučajna varijabla, dok bi zarada bila neprekidna slučajna varijabla.

Diskretna slučajna varijabla uglavnom se zadaje matrično na sljedeći način:

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix} \tag{2.5}$$

gdje x_i označava vrijednost, a p_i pripadne vjerojatnosti, odnosno vrijedi

$$P(X = x_i) = p_i. \tag{2.6}$$

Neprekidna slučajna varijabla zadaje se ili funkcijom gustoće $f(x)$ ili funkcijom distribucije $F(x)$. Funkcija distribucije definira se izrazom:

$$F(x) = P(X < x), \tag{2.7}$$

dok se funkcija gustoće definira s:

$$f(x) = F'(x). \tag{2.8}$$

Tako je primjerice uniformno distribuirana slučajna varijabla, tj. varijabla kod koje su sve vrijednosti jednako vjerojatne zadana sljedećom funkcijom gustoće:

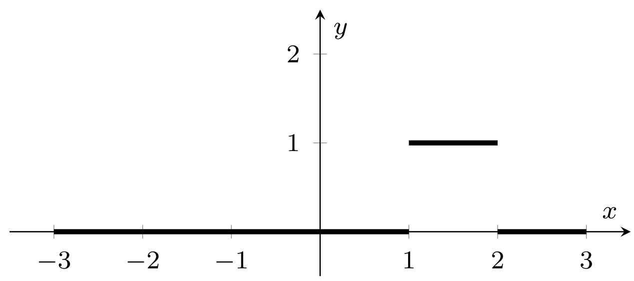
$$f(x) = \begin{cases} c, & x \in [a, b] \\ 0, & \text{inače}, \end{cases} \tag{2.9}$$

pri čemu je $c = \dfrac{1}{b-a}$. Na slici 2.1 možemo vidjeti grafički prikaz opisane slučajne varijable za $a = 1$ i $b = 2$.

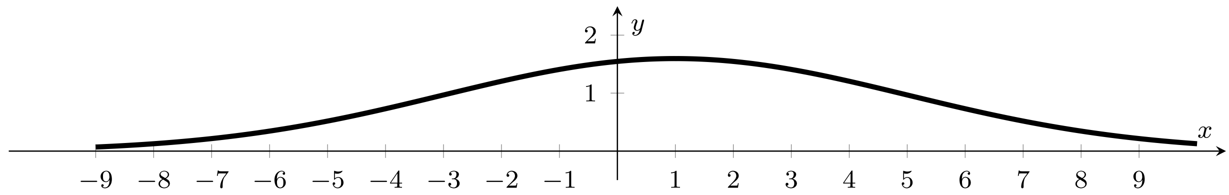
Često se koristi normalno distribuirana slučajna varijabla čija je funkcija gustoće dana s:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}, \tag{2.10}$$

pri čemu su μ i σ slobodni parametri. Njen grafički prikaz za $\mu = 1$ i $\sigma = 0.25$ prikazan je na slici 2.2. Napomenimo da se normalno distribuirana varijabla često naziva Gaussovom slučajnom varijablom.



Slika 2.1. Funkcija gustoće uniformno distribuirane slučajne varijable



Slika 2.2. Funkcija gustoće normalno distribuirane slučajne varijable

Broj oko kojeg se grupiraju vrijednosti slučajne varijable zove se matematičko očekivanje i označava se s $E(X)$. Broj koji mjeri rasap oko očekivane vrijednosti slučajne varijable zove se varijanca i označava se sa $V(x)$.

Matematičko očekivanje definira se s:

$$E(X) = \sum_i x_i \cdot p_i, \tag{2.11}$$

u diskretnom slučaju te

$$E(X) = \int_{\mathbb{R}} x \cdot f(x)dx, \tag{2.12}$$

u neprekidnom slučaju.

Varijanca je definirana izrazom

$$V(X) = E(X^2) - (E(X))^2 . \tag{2.13}$$

Slobodni parametri μ i σ kod normalno distribuirane slučajne varijable povezani su s matematičkim očekivanjem i varijancom, odnosno vrijedi da je kod normalno distribuirane slučajne varijable $E(X) = \mu$, a $V(X) = \sigma^2$.

Vrlo često, posebice u analizi vremenskih nizova, promatrat ćemo familije ili nizove slučajnih varijabli koje ćemo označavati s X_t pri čemu je t iz nekog diskretnog ili kontinuiranog skupa indeksa. Tako se primjerice stanje nekog sustava u vremenskom trenutku t može opisati familijom slučajnih varijabli X_t.

Za analizu vremenskih nizova od posebne su važnosti familije slučajnih varijabli koje se sastoje od jednako distribuiranih slučajnih varijabli čije realizacije ne ovise jedna o drugoj. Pritom se često koristi oznaka

$$X_t \sim IID\left(\mu, \sigma^2\right) \tag{2.14}$$

koja označava da sve slučajne varijable u familiji nezavisne, imaju jednaku distribuciju te da su njihova očekivanja jednaka μ, a varijance σ^2. Oznaka IID dolazi od engleskog naziva *identically and independently distributed*.

Spomenimo još jednu slučajnu varijablu koja se često koristi u analizi vremenskih nizova, a to je slučajna varijabla s tzv. umotanom normalnom razdiobom. Ona se definira sljedećom funkcijom gustoće:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} e^{-\dfrac{(x-\mu+2k\pi)^2}{2\sigma^2}}. \tag{2.15}$$

Iz ovog izraza može se zaključiti da je ova slučajna varijabla generirana klasičnom normalno distribuiranom slučajnom varijablom s parametrima μ i σ.

2.2. Osnovni pojmovi iz teorije vremenskih nizova

Vremenski niz je niz podataka $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$, prikupljenih u uzastopnim vremenskim trenucima iz skupa $T_0 = \{t_1, \ldots, t_n\}$, pri čemu je $t_1 < t_2 < \cdots < t_n$ i $x_{t_i} \in \mathbb{R}$, $i = 1, \ldots, n$. U ovom slučaju, kako je skup vremenskih trenutaka T_0 konačan govorimo o vremenskom nizu u diskretnom vremenu. Ovakav se niz često označava s $\{x_{\mathrm{t}}, t \in T_0\}$.

Moguće je da vrijednost nekog procesa poznajemo u svakom trenutku pa u tom slučaju govorimo o vremenskom nizu u neprekidnom vremenu (npr. broj ljudi u trgovini tijekom radnog dana može biti poznat u svakom trenutku). Skup T_0 može označavati npr. dane, mjesece, godine, sate itd.

Vremenski trenuci mogu biti raspoređeni:

1. ekvidistantno, kada je $t_2 - t_1 = \cdots = t_n - t_{n-1}$, te
2. neekvidistantno.

U ekvidistantnom slučaju možemo bez smanjenja općenitosti podatke označiti kao $x_1, \ldots, x_n$.

Problem analize niza podataka $x_1, \ldots, x_n$ može se promatrati i sa statističkog gledišta pri čemu se pretpostavlja da je niz $x_1, \ldots, x_n$ realizacija jednostavnog slučajnog uzorka $(X_1, \ldots, X_n)$ nezavisnog i jednako distribuiranog niza slučajnih varijabli. Ako u tom zahtjevu oslabimo pretpostavku jednake distribuiranosti povećava se broj statističkih metoda koje se na promatrani niz mogu primijeniti, pa prikladne postaju i regresijske metode. Svakako treba naglasiti da je u praksi vrlo često narušena pretpostavka nezavisnosti, a često i jednake distribuiranost.

Svrha analiza vremenskog niza je:

1. Modelirati i razumjeti stohastički mehanizam koji dovodi do realizacije kojom raspolažemo,
2. Predviđati buduće vrijednosti na osnovu poznatih podataka, što ćemo raditi u nadolazećim poglavljima.

2.3. Matematički model vremenskog niza

U ovom poglavlju dati ćemo matematički model kojim bi se mogao adekvatno opisati vremenski niz u diskretnom vremenu $\{x_{\mathrm{t}}, t \in T_0\}$. Iz navedenih primjera jasno je da nećemo tražiti deterministički model kod kojeg bi vrlo precizno mogli odrediti vrijednost promatrane pojave u svakom trenutku, već smo u potrazi sa stohastičkim modelom kojim je moguće procijeniti vjerojatnost da buduća vrijednost bude unutar nekih granica.

Stohastički model kojim se opisuje vremenski niz naziva se slučajni proces, a precizno je definiran u sljedećoj definiciji.

Definicija 2.3. *Slučajni proces* $\{X_t, t \in T\}, T \subseteq \mathbb{R}$ *familija je slučajnih varijabli definiranih na istom vjerojatnosnom prostoru* (Ω, F, P).

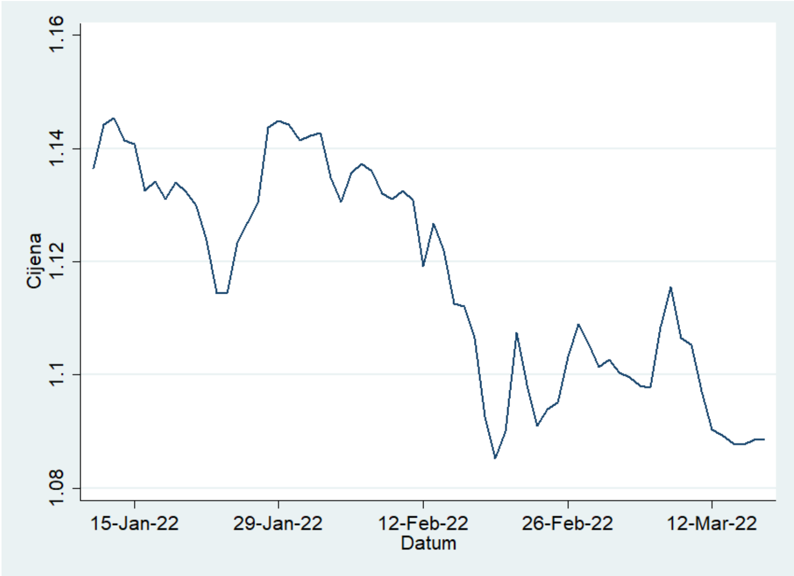
Svaki podatak vremenskog niza $x_{ti}, i = 1, \ldots, n$, smatrat ćemo realizacijom jedne slučajne varijable X_{t_i}, odnosno vremenski niz $\{x_t, t \in T_0\}$ smatramo realizacijom (ili dijelom realizacije) slučajnog procesa $\{X_t, t \in T\}, T_0 \subseteq T$. Slučajni proces $\{X_t, t \in T\}$ nazivamo model vremenskog niza. Često ćemo i za njega reći samo vremenski niz. Ovisno o vrsti vremenskog niza skup T može biti diskretan ili primjerice $[0, \infty)$.

Modelirati vremenski niz znači odrediti slučajni proces za kojeg vjerujemo da mu je vremenski niz jedna realizacija ili njezin dio, pri čemu se pod određivanjem slučajnog procesa misli na određivanje distribucije pripadnih slučajnih varijabli. Odrediti slučajni proces razmatramo u distribucijskom smislu.

Uočimo da je modeliranje vrlo težak problem, u smislu da na osnovu manjeg djela realizacija pokušavamo odrediti čitav slučajni proces. To će biti nemoguće bez pretpostavki na strukturu procesa.

Primjer 2.2. *U klasičnoj statistici niz* $(x_1, \ldots, x_n)$ *se modelira jednostavnim slučajnim uzorkom* $(X_1, \ldots, X_n)$*, odnosno nizom nezavisnih i jednako distribuiranih slučajnih varijabli. U tom slučaju možemo dosta toga reći o distribuciji iz koje dolazi uzorak jer zapravo imamo* n *realizacija iz iste distribucije. Statističke metode za ovakav problem su dobro razvijene. Zavisnost, iako otežava analizu, može dati bolju predikciju.*

Primjerice, vremenski niz dobijemo promatranjem napona baterije kroz svaki sat što nam predstavlja vremenski niz. Također, ukoliko pratimo cijenu kretanja dolara svaki dan dobijemo vremenski niz. Ovaj je vremenski niz grafički prikazan na sljedećoj slici.



Slika 2.3. Primjer vremenskog niza - kretanje cijene dolara

Vremenski niz može imati i više dimenzija, primjerice možemo analizirati uređeni par napona i struje $(u(t), i(t))$ gdje nam je napon $u(t)$, a struja $i(t)$. Također, ukoliko imamo dvije valute gdje je primjerice $x(t)$ cijena dolara, a $y(t)$ cijena eura također dobivamo dvodimenzionalni vremenski niz. Isto dobivamo primjerice i promatranjem zarade dva radnika. Jasno je da možemo navesti veliki broj primjera višedimenzionalnih vremenskih nizova, a bitno je naglasiti da ih promatramo kao uređene parove ili n-torke jer svaka komponenta, tj. dimenzija takvog vremenskog niza može ovisiti jedna o drugoj.

Takav se niz zapisuje kao $\{\boldsymbol{x}_t, t \in T_0\}$, gdje je

$$\boldsymbol{x}_t = \left(x_t^{(1)}, \dots, x_t^{(k)}\right) \in \mathbb{R}^k. \tag{2.16}$$

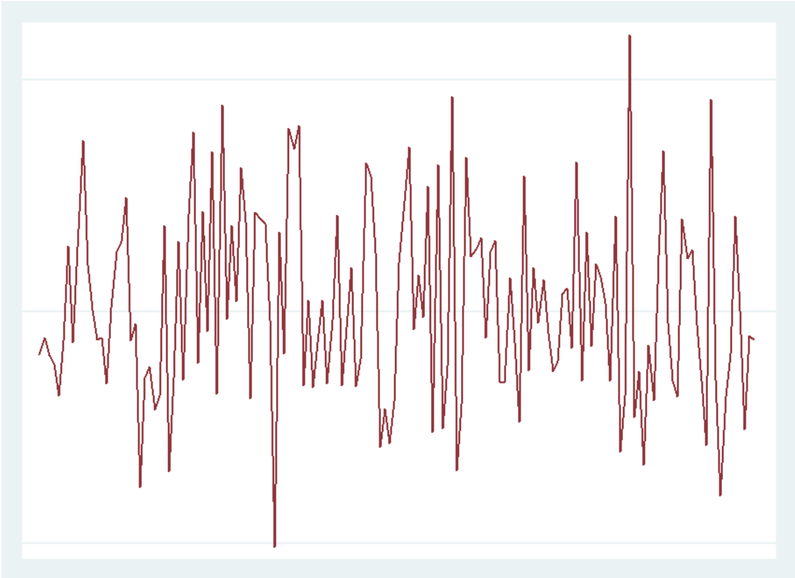
Model za ovakve podatke je vektorski slučajni proces $\{\boldsymbol{X}_t, t \in T\}$ s pripadnom familijom slučajnih vektora

$$\boldsymbol{X}_t = \left(X_t^{(1)}, \dots, X_t^{(k)}\right). \tag{2.17}$$

Model biramo tako da proces ima iste karakteristike koje uočavamo u podacima. To se prvenstveno odnosi na zavisnosti i ponašanje kroz vrijeme.

2.4. Stacionarni procesi

Stacionarnost znači da se statistička svojstva procesa koji generira vremenski niz ne mijenjaju tijekom vremena. To ne znači da se niz ne mijenja tijekom vremena, samo da se način na koji se mijenja ne mijenja s vremenom. Algebarski ekvivalent je stoga linearna funkcija, vrijednost linearne funkcije mijenja se kako x raste, ali način na koji se mijenja ostaje konstantan odnosno ima stalan nagib.

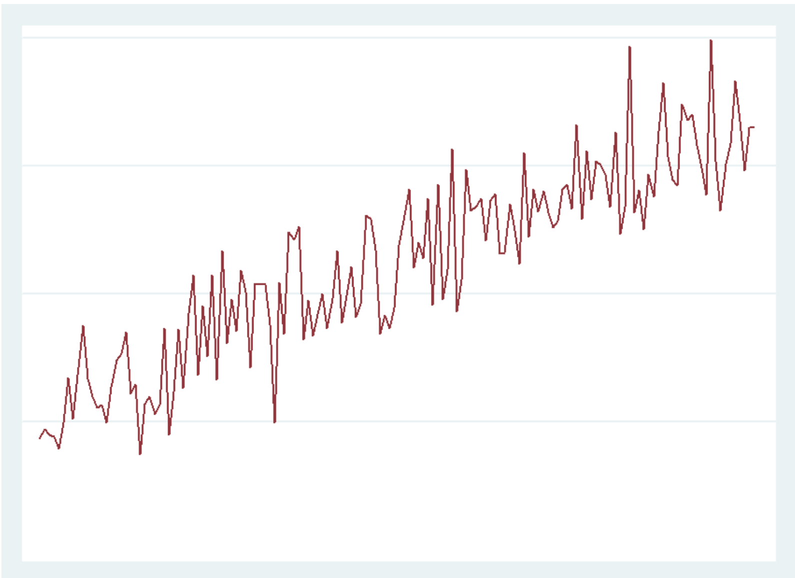


Slika 2.4. Primjer stacionarnog vremenskog niza

Na sljedećoj slici prikazan je jedan primjer stacionarnog vremenskog niza

Objasnimo zašto je stacionarnost važna. Naime, matematička analiza stacionarnih procesa daleko je jednostavnija od analize nestacionarnih, a u mnogim slučajevima jednostavni modeli poput stacionarnih procesa mogu biti iznenađujuće korisni, bilo kao građevni blokovi u izgradnji složenijih, ili kao korisne aproksimacije složenim fenomenima. Zbog ovih svojstava, stacionarnost je postala uobičajena pretpostavka za mnoge prakse i alate u analizi vremenskih nizova. To uključuje procjenu trenda, predviđanje i uzročno-posljedično zaključivanje.

Kako bi bolje ilustrirali pojam stacionarnosti, na sljedećoj slici dajemo primjer jednog nestacionarnog vremenskog niza.



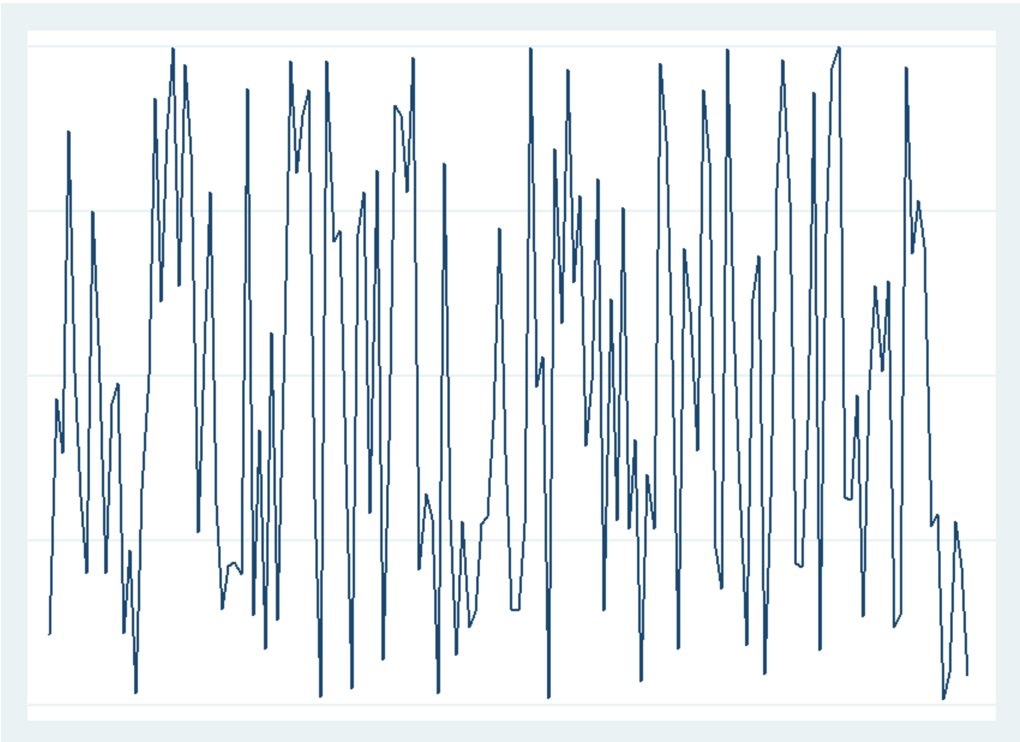
Slika 2.5. Primjer nestacionarnog vremenskog niza

2.5. Primjeri slučajnih procesa

U ovom poglavlju navodimo neke važne primjere slučajnih procesa koji su od velike teorijske važnosti u daljnoj analizi vremenskih nizova.

2.5.1. Bijeli šum

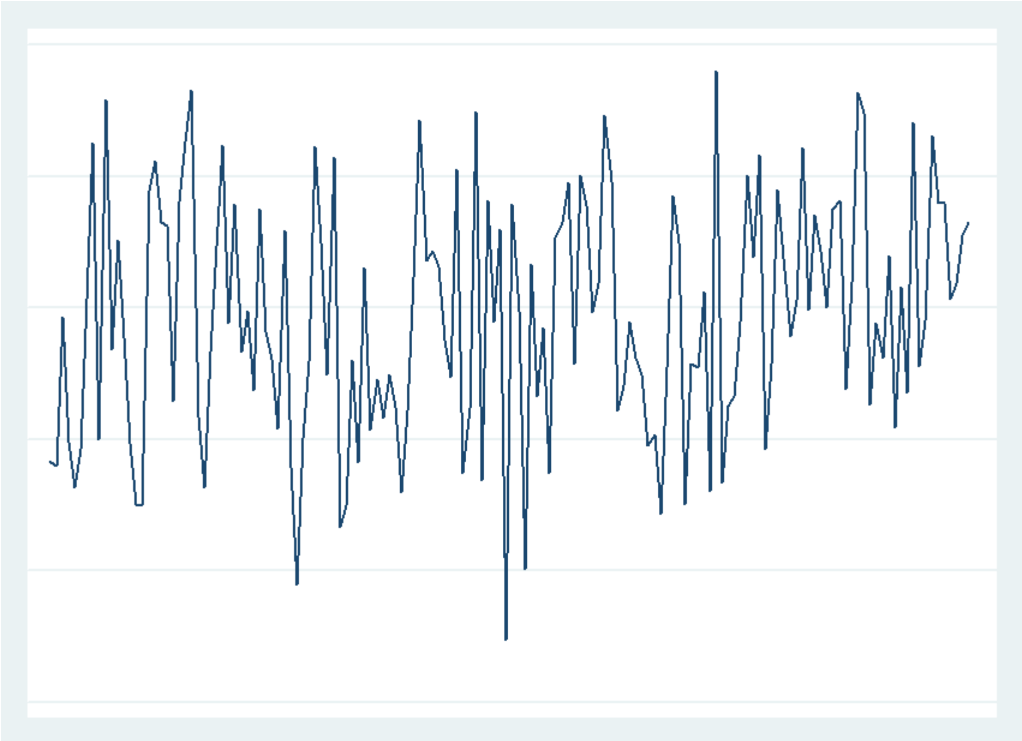
Već smo prije analizirali familiju slučajnih varijabli koju smo označili s $IID\left(\mu,\sigma^{2}\right)$. Ako kod takvu familiju promatramo kao slučajan proces, dobivamo slučajan proces koji se naziva nezavisno i jednoliko distribuiran šum (n.j.d. šum), a često i **bijeli šum**. Jedan primjer takvog procesa u kojem su sve slučajne varijable iz familije uniformno distribuirane prikazan je na sljedećoj slici.



Slika 2.6. Grafički primjer n.j.d. šuma s uniformnom distribucijom

2.5.2. Gaussov bijeli šum

Ako kod već opisanog bijelog šuma u pozadini stoji Gaussova ili normalna razdioba, govorimo o Gaussovom bijelom šumu. U dijelu literature pojam Gaussova bijelog šuma i općenitog bijelog šuma se izjednačava i takav se proces označava s $WN\left(\mu,\sigma^{2}\right)$. Jedan primjer Gaussova bijelog šuma grafički je prikazan na sljedećoj slici.



Slika 2.7. Grafički primjer Gaussova bijelog šuma

Sa slike se može vidjeti da je Gaussov bijeli šum stacionaran proces.

2.6. Autokovarijanca i autokorelacija

Pojam korelacije odnosi se na povezanost dviju slučajnih varijabli. Drugim riječima, dvije su slučajne varijable korelirane ako između njih postoji određena funkcijska povezanost. Pritom se uglavnom misli na linearnu povezanost, pa koreliranost znači da porastom vrijednosti jedne varijable, linearno rastu (ili padaju) vrijednosti druge varijable.

Kod vremenskih nizova prirodno je promatrati povezanost između slučajnih varijabli koje odgovaraju dvama različitim vremenskim trenutcima. Ako takva korelacija postoji, tada kažemo da je vremenski niz autokoreliran.

Analiza autokorelacije matematički je alat za pronalaženje ponavljajućih obrazaca, poput prisutnosti periodičnog signala zaklonjenog šumom ili identificiranja nedostajuće osnovne frekvencije u signalu koji impliciraju njegove harmoničke frekvencije. Autokorelaciju u vremenskom nizu koja mjeri povezanost trenutnog stanja sa stanjem nakon k trenutaka često obilježavamo s

$$R_X[m,k] = E[X_m X_{m+k}]. \tag{2.18}$$

Uz pojam korelacije usko je povezan i pojam kovarijance. Naime, dok je varijanca mjera varijabilnosti jedne slučajne varijable, odnosno njenog odstupanja od očekivane vrijednosti, kovarijanca bi predstavljala mjeru zajedničke varijabilnosti dviju slučajnih varijabli. Pomoću kovarijance možemo ustvrditi postojanje i smjer povezanosti dviju varijabli, dok pomoću korelacije tu informaciju proširujemo i snagom povezanosti. Ako se radi o vremenskom nizu, tada se kovarijanca prirodno naziva autokovarijanca, kao što je i korelacija postala autokorelacija.

Kovarijanca slučajnih varijabli oznčava se s $\mathrm{Cov}[X, Y]$, a kod vremenskih nizova i stohastički procesa često koristimo oznake

$$C_X(t, \tau) = \mathrm{Cov}[X(t), X(t + \tau)], \tag{2.19}$$

za stohastički proces i

$$C_X[m, k] = \mathrm{Cov}[X_m, X_{m+k}], \tag{2.20}$$

za vremenski niz.

2.7. Autoregresijski modeli

Autoregresijskim se modelima opisuju stohastički procesi koji generiraju vremenske nizove čije su vrijednosti autokorelirane. Autoregresijski model reda p je oblika:

$$X_t = \theta_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t. \tag{2.21}$$

Član procesa X_t u razdoblju t ovisi o članovima procesa u p prethodnih razdoblja i o slučajnoj varijabli Z_t, pri čemu Z_t predstavlja pogrešku. Koeficijenti ϕ_i predstavljaju utjecaje stanja X_{t-i} na stanje X_t.

Kako navedeni izraz predstavlja proces regresije u kojem se član X_t povezuje s p prethodnih članova procesa, $X_{t-1},\ X_{t-2}, \ldots, X_{t-p}$, model se stoga i naziva *autoregresijski model*. U tom kontekstu koeficijenti ϕ_i nazivaju se autoregresijski parametri, a θ_0 je konstantan član koji se najčešće zanemaruje.

Stohastički proces, odnosno vremenski niz koji se može prikazati autoregresijskim modelom često se naziva AR procesom. Ideja AR procesa je trenutnu vrijednost procesa prikazati kao linearnu kombinaciju nekoliko prošlih vrijednosti uz određenu grešku, kao što bi kod običnog regresijskog procesa povezivali zavisnu varijablu sa nizom nezavisnih varijabli.

Sada možemo i formalno definirati autoregresijski proces.

Definicija 2.4. *Slučajan proces $\{X_t, t \in \mathbb{Z}\}$ je autoregresijski proces reda $p \in \mathbb{N}$ ako je stacionaran i ako je:*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t, \quad t \in \mathbb{Z} \tag{2.22}$$

gdje je $\{Z_t, t \in \mathbb{Z}\} \sim WN\,(0, \sigma^2)$ i $\phi_1, \ldots, \phi_p \in \mathbb{R}, \phi_p \neq 0$.

Autoregresijki proces reda p označavamo s $AR(p)$.

Koristeći operator pomaka unazad, jednadžbu (2.22) možemo zapisati kao:

$$\phi(B)X_t = Z_t, \tag{2.23}$$

gdje je

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p, \quad z \in \mathbb{C}, \tag{2.24}$$

karakteristični polinom AR procesa dok je operator pomaka B definiran s

$$BX_t = X_{t-1}, B^2 X_t = X_{t-2}, B^3 X_t = X_{t-3}, \ldots \tag{2.25}$$

2.7.1. Problem egzistencije AR procesa

Može se postaviti pitanje uz koje uvjete za dane $\phi_1, \ldots, \phi_p \in \mathbb{R}$ postoji proces koji zadovoljava jednadžbu (2.22), kao i pitanje hoće li pripadni proces biti stacionaran. Razmotrimo da pitanja na primjeru $AR(1)$ procesa.

Promotrimo jednadžbu $AR(1)$ procesa:

$$X_t = \phi X_{t-1} + Z_t \tag{2.26}$$

Iteriranjem dobivamo

$$X_t = Z_t + \phi X_{t-1}, \tag{2.27}$$

$$= Z_t + \phi Z_{t-1} + \phi^2 X_{t-2}, \tag{2.28}$$

$$\vdots \tag{2.29}$$

$$= Z_t + \phi Z_{t-1} + \cdots + \phi^k Z_{t-k} + \phi^{k+1} X_{t-k-1}. \tag{2.30}$$

Sada je jasno da bi kao rješenje imalo smisla razmotriti proces

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{l-j}, \tag{2.31}$$

za koji se može pokazati da je dobro definiran ako je $|\phi| < 1$. Naime, vrijedi

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j} = \phi \sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j} + Z_t = \phi \sum_{j=0}^{\infty} \phi^j Z_{t-(j+1)} + Z_t = \phi X_{t-1} + Z_t. \tag{2.32}$$

Ako bi primjerice bilo $|\phi| > 1$ dobili bi neprirodno rješenje jer bi sadašnje vrijednosti ovisile o budućim vrijednostima šuma, što je za praktične primjene neprikladno (pogotovo kod predviđanja).

2.8. Vektorski autoregresijski modeli

Vektorski autoregresijski model (VAR) opisuje evoluciju skupa k varijabli, nazvanih endogene varijable, tijekom vremena. Endogena varijabla je ona čija je vrijednost određena odnosima uspostavljenim unutar modela u koji je uključena, a svako vremensko razdoblje je numerirano s $t = 1, \ldots, T$.

Varijable se skupljaju u vektoru, $\mathbf{y}_t$, koji je duljine k, odnosno vrijedi:

$$\mathbf{y}_t = \begin{bmatrix} y_t^1 \\ \vdots \\ y_t^k \end{bmatrix}, \tag{2.33}$$

pri čemu vektorska komponenta y_t^i označava vrijednost varijable Y^i u trenutku t.

VAR modele karakterizira njihov vremenski redoslijed, koji se odnosi na broj ranijih vremenskih razdoblja koje će model koristiti. Zaostajanje (lag) je vrijednost varijable u prethodnom vremenskom razdoblju. Tako je VAR s p zaostajanja model koji uključuje prethodnih p vremenskih razdoblja, te se zapisuje kao:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{e}_t, \tag{2.34}$$

gdje $\mathbf{y}_{t-r}$ označava vrijednost vektora $\mathbf{y}_t$ u trenutku $t-r$, $\mathbf{c}$ je vektor konstanti, dok je $\mathbf{A}_i$ vremenski nepromjenjiva $k \times k$-matrica, a $\mathbf{e}_t$ vektor pogreške.

Vektor pogreške mora zadovoljiti tri uvjeta:

1. $\mathrm{E}\left(e_t^i\right) = 0$, tj. očekivanje svake komponente vektora pogreške jednako je nuli,
2. $\mathrm{E}\left(\mathbf{e}_t \cdot \mathbf{e}_t'\right) = \Omega$, pri čemu je Ω pozitivno semidefinitna[1] matrica kovarijance,
3. $\mathrm{E}\left(\mathbf{e}_t \cdot \mathbf{e}_{t-k}'\right) = 0$, za sve $k \neq 0$, što znači da nema korelacije kroz vrijeme.

[1]Za realnu matricu $\mathbf{A}$ kažemo da je pozitivno semidefinitna ako je $\mathbf{v} \cdot \mathbf{A} \cdot \mathbf{v}' \geq 0$ za svaki stupčasti vektor $\mathbf{v}$.

3. Kointegracija i korekcija pogreške

U sklopu ovog poglavlja objasnit ćemo pojam kointegracije i pojam ispravljanja pogreške nužne za definiranje i analizu VECM modela. Ovo je poglavlje obrađeno koristeći reference [5, 6].

3.1. Definicija kointegracije

Kointegracija je važno statističko svojstvo vektora $\left(X_t^1, X_t^2, \ldots, X_t^k\right)$ varijabli vremenskog niza, a da bismo ga mogli objasniti moramo najprije objasniti pojam integracije. Naime, već smo spomenuli da je stacionarnost vremenskog niza poželjno svojstvo. Ako niz nije stacionaran, tada se stacionarnost može postići diferenciranjem, tj. formiranjem novog vremenskog niza razlika na sljedeći način:

$$\Delta X_t = X_t - X_{t-1}, \tag{3.1}$$

gdje je X_t proizvoljni vremenski niz. Ponekada za dobivanje stacionarnosti neće biti dovoljno jedino diferenciranje. U tom slučaju, ako smo za dobivanje stacionarnog niza trebali barem d uzastopnih diferenciranja, kažemo da je vremenski niz integriran reda d.

Sada možemo objasniti i pojam kointgracije. Ako promatramo više vremenskih nizova koji su integrirani reda d, a njihova linearna kombinacija ima red integracije manji od d, tada kažemo da su ti vremenski nizovi kointegrirani. Formalno, ako promatramo primjerice vremenske nizove X_t, Y_t i Z_t koji su integrirani reda d, a postoje koeficijenti a, b i c takvi da je $aX_t + bY_t + cZ_t$ integriran reda manjeg od d, tada su X_t, Y_t i Z_t kointegrirani.

Da bi za vektor $\left(X_t^1, X_t^2, \ldots, X_t^k\right)$ rekli da je kointegriran, sve njegove komponente moraju biti istog reda integracije, primjerice d te mora postojati linearna kombinacija komponenti čiji je red integracije strogo manji od d. Ta linearna kombinacija se obično naziva relacija ravnoteže. Može se pokazati da se komponente kointegriranog vektora previše ne odmiču jednad od druge, dok se kod vektora koji nije kointegriran mogu dogoditi značajna odstupanja kroz vrijeme.

3.2. Pojam ispravljanja pogreške

U vremenskom nizu možemo razlikovati kratkoročna i dugoročna odstupanja, a za pretpostaviti je da će dugoročna odstupanja korigirati ona kratkoročna u cilju očuvanja ravnoteže. Opisani princip se naziva principom ispravljanja pogreške.

S matematičkog stanovišta, logično je pretpostaviti da bi distinkcija između kratkoričnih i dugoročnih odstupanja bila korisna, odnosno da bi trebalo napraviti model kojim će se moći razlučiti njihovi međuosbni utjecaji. Takav se model naziva model za ispravljanje pogrešaka (ECM). Drugim riječima, ECM model izravno procjenjuje brzinu kojom se zavisna varijabla vraća u ravnotežu nakon promjene u drugim varijablama.

U najjednostavnijem slučaju ECM model funkcionira na sljedeći način. Promatramo regresijski model

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_{t-1} + \alpha_3 \Delta x_t + \zeta_t, \tag{3.2}$$

gdje y označava zavisnu, a x_t nezavisnu varijablu. Zavisna i nezavisna varijabla mogu biti i komponente vektora vremenskog niza. Kako smo već rekli diferenciranjem se postiže stacionarnost i jednadžba koji bi uključivala diferencirane varijable prikazivala bi odnos u ravnoteži. Tako možemo zaključiti da se koeficijent α_3 odnosi upravo na taj ravnotežni odnos, dok se koeficijenti α_1 i α_2 odnose na kratkoročna odstupanja, odnosno oni bi se mogli pretvoriti u koeficijent koji bi se nazivao koeficijent ispravljanja pogreške.

4. VECM model

VECM (eng. Vector Error Correction Model) je vektorski autoregresijski model koji ima ugrađen član s koeficijentom za ispravljanje pogreške. Uključivanjem tog člana u model može se postići lakše mjerenje utjecaja odmicanja od ravnoteže, a time i kvalitetniji prediktivni model koji mjeri svaki odmak od dugotrajne ravnoteže. VECM modeli se koriste za modeliranje stacionarnih odnosa između komponenti u vektoru vremenskog niza. Ovo poglavlje obrađeno je prema [7].

4.1. Multivarijantna VECM specifikacija

Kako bi došli do VECM modela, krećemo od standardnog i već opisanog VAR modela:

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \epsilon_t, \tag{4.1}$$

koji izražavamo u tzv. VECM specifikaciji na sljedeći način:

$$\Delta\mathbf{y}_t = \mathbf{v} + \mathbf{\Pi}\mathbf{y}_{t-1} + \sum_{i=1}^{p-1}\mathbf{\Gamma}_i\Delta\mathbf{y}_{t-i} + \epsilon_t, \tag{4.2}$$

gdje je

$$\mathbf{\Pi} = \sum_{j=1}^{j=p}\mathbf{A}_j - \mathbf{I}_k \tag{4.3}$$

i

$$\mathbf{\Gamma}_i = -\sum_{j=i+1}^{j=p}\mathbf{A}_j, \tag{4.4}$$

dok su $\mathbf{v}$ i ϵ_t iz gornjih jednadžbi identični.

Odmah uočavamo i sličnosti s već opisanim ECM modelom te zaključujemo da se matricom $\mathbf{\Pi}$ modelira ispravljanje pogreške, odnosno kratkoročni utjecaji, dok je s koeficijentima $\mathbf{\Gamma}_i$ prikazano ravnotežno ponašanje ili dugoročni utjecaj.

4.2. VECM model prema Johansenu

Osim ove općenite specifikacije, često se koristi i još jedna VECM specifikacija koju je uveo Johansen. Ta specifikacija poprima sljedeći oblik:

$$\Delta\mathbf{y}_t = \alpha\left(\beta\mathbf{y}_{t-1} + \mu + \rho t\right) + \sum_{i=1}^{p-1}\Gamma_i\Delta\mathbf{y}_{t-i} + \gamma + \tau t + \epsilon_t, \tag{4.5}$$

gdje možemo uočiti nove parametre μ, ρ, γ i τ. Te parametre ćemo zvati parametrima trenda.

Postavljanje ograničenja na parametre trenda daje pet slučajeva:

1. Neograničeni trend
   Ako nisu postavljena ograničenja na parametre trenda, model implicira da postoje kvadratni trendovi u razinama varijabli i da su kointegrirajuće jednadžbe stacionarne oko vremenskih trendova.

2. $\tau = 0$
   Postavljanjem $\tau = 0$, pretpostavljamo da su trendovi u razinama podataka linearni, ali ne i kvadratni. Ova specifikacija omogućuje da kointegrirajuće jednadžbe budu stacionarne u trendu.

3. $\tau = 0$ i $\rho = 0$
   Postavljanjem $\tau = 0$ i $\rho = 0$ isključujemo mogućnost da razine podataka imaju kvadratne trendove, a kointegrirajuće jednadžbe ograničavamo da budu stacionarne oko konstante. Budući da γ nije ograničen na nulu, ova specifikacija i dalje stavlja linearni vremenski trend u razini podataka.

4. $\tau = 0$, $\rho = 0$ i $\gamma = 0$
   Dodavanjem ograničenja da je $\gamma = 0$, pretpostavljamo da nema linearnih vremenskih trendova u razinama podataka. Ova specifikacija omogućuje da kointegrirajuće jednadžbe budu stacionarne oko konstante, ali ne dopušta nikakve druge trendove.

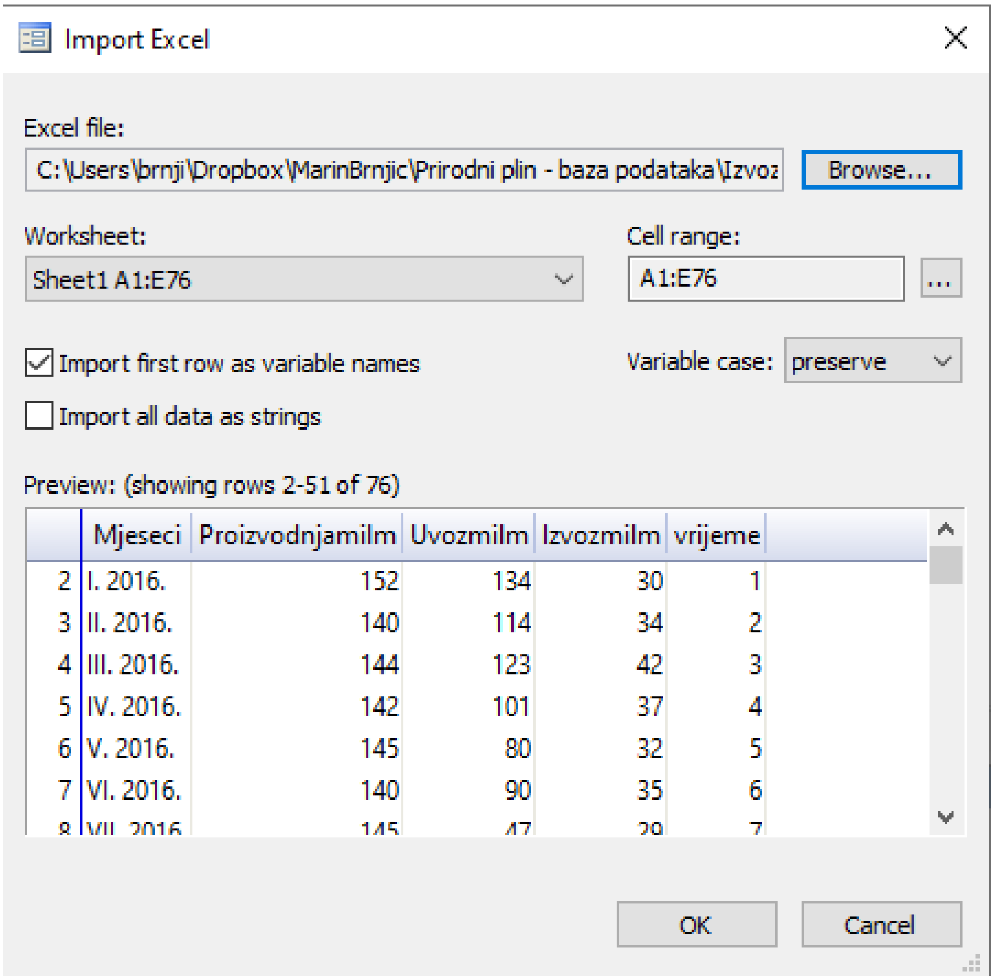
5. $\tau = 0$, $\rho = 0$, $\gamma = 0$ i $\mu = 0$
   Ova specifikacija pretpostavlja da ne postoje trendovi.

5. VECM u programskom paketu Stata

U ovome poglavlju spomenuti ćemo sve naredbe koje će biti korištene za empirijska istraživanja unutar statističkog softverskog paketa Stata.

5.1. Unos podataka

Za početak potrebno je unijeti našu bazu podataka koja je napisana u Excel-u, a to radimo na alatnoj traci klikom na File > Import > Excel spreadsheet.

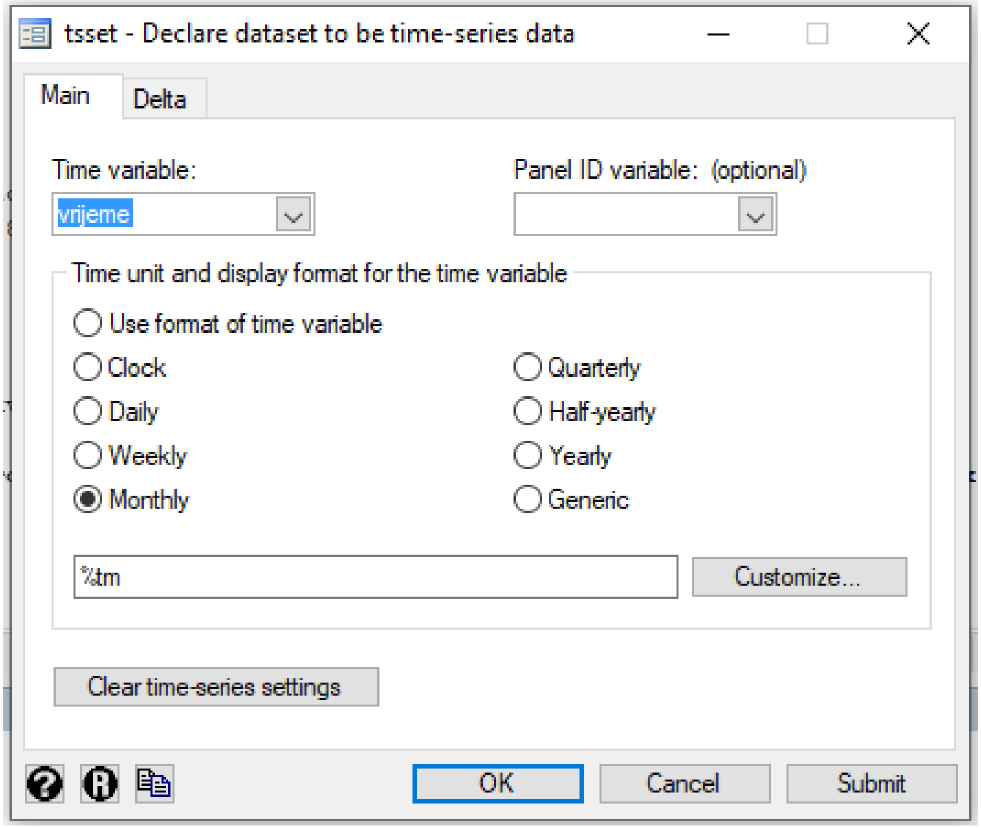


Slika 5.1. Prozor za unos podataka

Otvara se prozor za unos podataka unutar kojeg nakon što unesemo podatke moramo selektirati "Import first rows as variable names" kako bi nam se prvi redovi brojali kao imena varijabli.

5.2. Deklariranje skupa podataka kao vremenskog niza

Nakon što smo unijeli podatke potrebno je deklarirati vremenski niz. Na alatnoj traci kliknemo na Statistics > Survival analysis > Setup and utilities > Declare data to be survival-time data.



Slika 5.2. Prozor za deklariranje skupa podataka kao vremenske serije

Otvara se prozor za deklariranje skupa podataka kao vremenskog niza gdje je potrebno unijeti parametar "Time variable", odnosno varijablu koja nam varijabla predstavlja vrijeme u zadanome skupu podataka. Ovisno o podacima koje imamo potrebno je odabrati i format vremena. U ovome primjeru radimo sa mjesečnim podacima za proizvodnju, uvoz i izvor prirodnog plina u Hrvatskoj zbog čega odabiremo "Monthly".

Ove postupke moguće je skratiti koristeći naredbu:

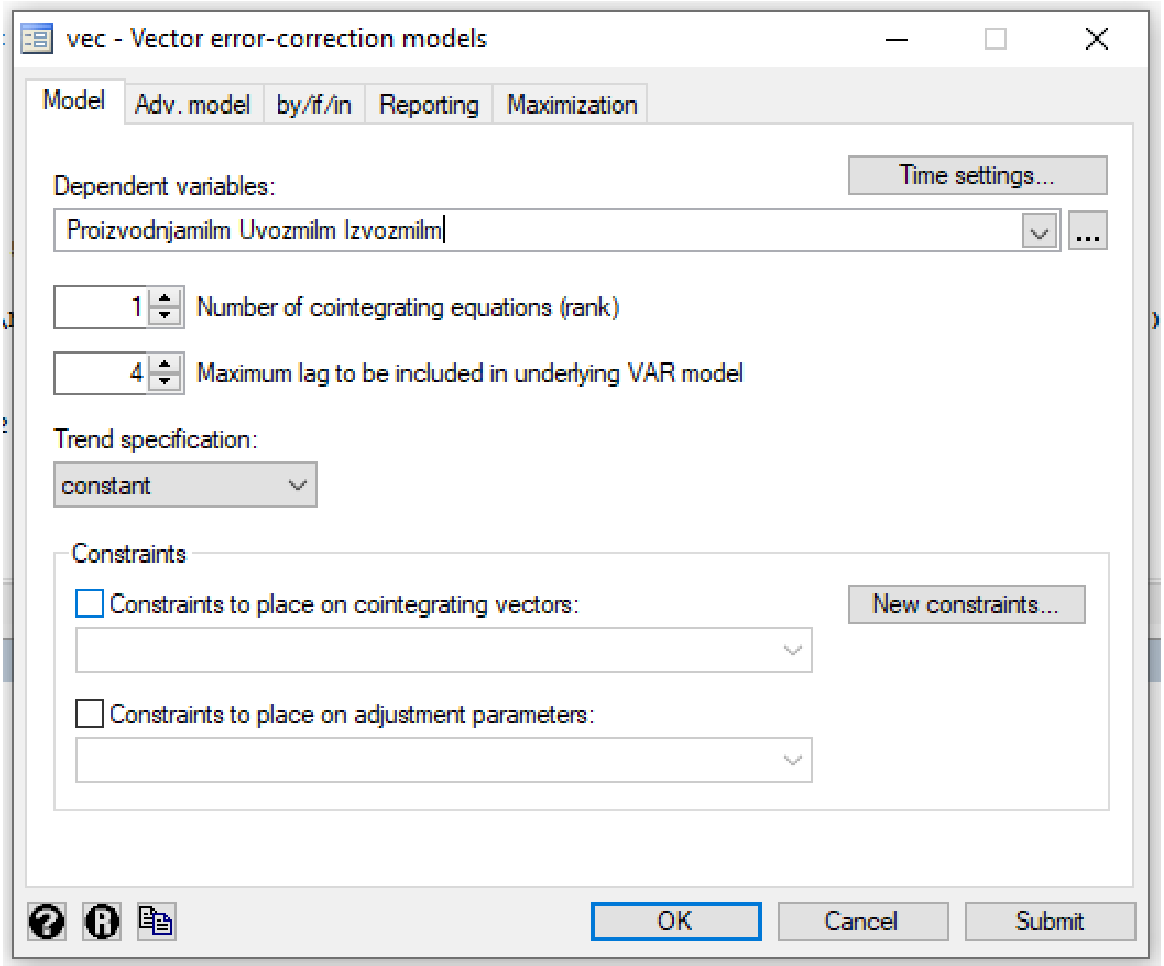
```
tsset vrijeme, monthly
```

5.3. Određivanje optimalnog broja lagova

Kako skupovi podataka nisu isti te se svaki skup podataka ponaša drukčije potrebno je odrediti optimalni broj lagova za uneseni skup podataka. Stata nam tu uskače u pomoć sa veoma korisnom naredbom varsoc koja nam ispisuje optimalan broj lagova za uneseni skup podataka. U sljedećem poglavlju bolje ćemo se upoznati sa ovom naredbom te ju i direktno koristiti.

5.4. Model korekcije vektora pogreške

Potrebno je podesiti parametre u modelu korekcije vektora pogreške (VECM) stoga na alatnoj traci kliknemo na Statistics > Multivariate time series > Vector error-correction model (VECM).



Slika 5.3. Prozor za podešavanje parametara modela korekcije vektora pogreške

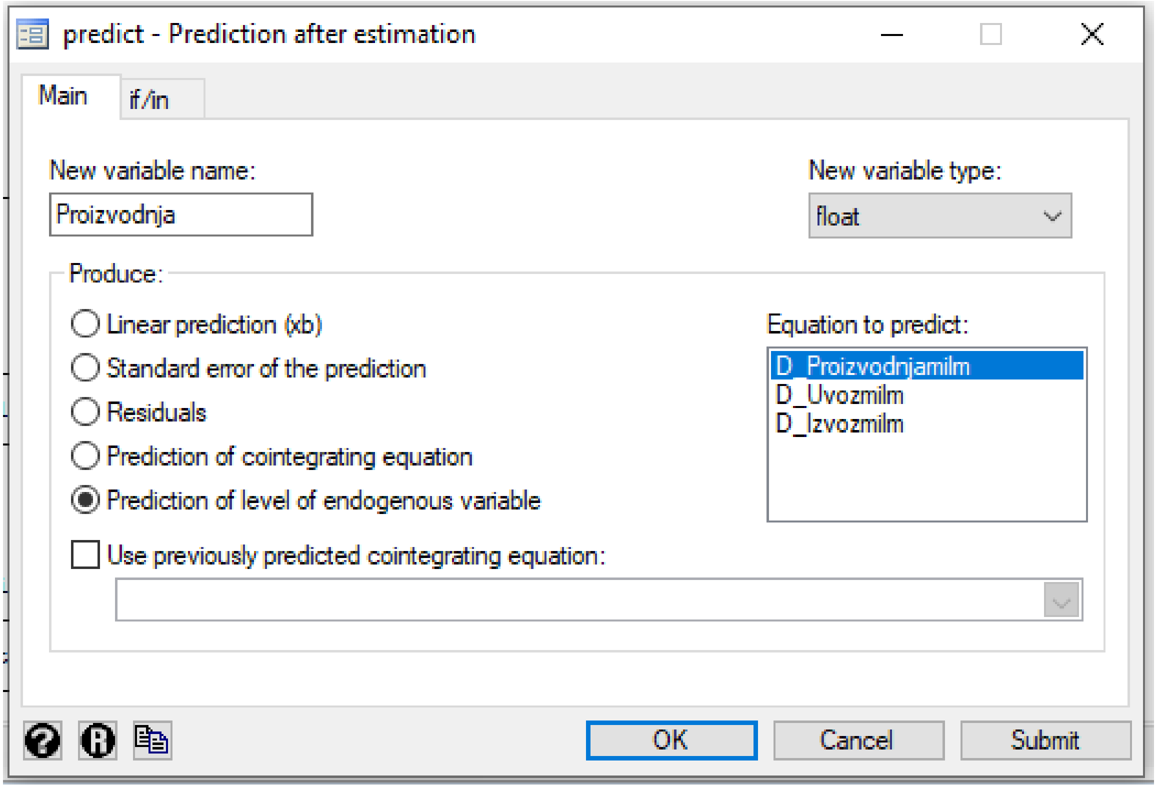
Otvara se prozor za podešavanje parametara modela korekcije vektora pogreške. Potrebno je pod "Dependent variables" unijeti varijable koje ovise jedna o drugoj. Te unijeti rank i broj lagova. Konkretno u ovom slučaju broj lagova je 4 (varsoc naredba).

Ovaj postupak može se skratiti unosom naredbe:

```
vec Proizvodnjamilm Uvozmilm Izvozmilm,
trend(constant) rank(1) lags(4)
```

5.5. Generiranje predikcije

U konačnici želimo dobiti predikciju za sljedeći mjesec. Predikcije generiramo tako da na alatnoj traci kliknemo Statistics > Postestimation > Predictions, residuals, etc.



Slika 5.4. Prozor za generiranje predikcija

Otvara se prozor za predviđanje nakon procjene unutar kojega je potrebno unijeti ime nove varijable koju generiramo i odabrati pod "Produce" opciju "prediction of level of edogenous variable" te ovisno koju varijablu želimo predvidjeti odaberemo ju u izborniku "Equation to predict".

Također je moguće skratiti cijeli postupak direktnim unosom naredbe za predikciju:

```
predict Proizvodnja, level
equation(D\textunderscore Proizvodnjamilm)
```

6. Empirijsko istraživanje

U ovome poglavlju baviti ćemo se predikcijima generiranih pomoću VECM modela unutar statističkog softverskog paketa Stata. Odnosno analizom baza podataka pokušati ćemo dobiti što točniju predikciju budućeg stanja. Konkretno u prvome dijelu istraživanja promatrati ćemo kripto valutu Bitcoin, analizirati njegovo ponašanje te pokušati odrediti njegovu vrijednost za sljedeći dan kroz sve vrste trendova i vremenskih perioda. U drugom dijelu istraživanja bavit ćemo problemom proizvodnje prirodnog plina u Hrvatskoj.

No, zašto uopće radimo predikcije? Poznavanje budućeg stanja nam u praksi može biti od jako velike koristi. Te tomu i služe predikcije kako bi smo otprilike znali buduće stanje našeg promatranog sustava. Naravno potrebna nam je što veća točnost predikcije budućeg stanja kako nam ta informacija ne bi bila totalno nasumična.

6.1. Predikcija cijene Bitcoin-a

U ovome podpoglavlju pomoću VECM modela unutar statističkog softverskog paketa Stata pokušati ćemo generirati što točnije predikcije cijene zatvaranja Bitcoin-a za jedan dan unaprijed kroz sve vrste trendova.

Za analizu koristimo sljedeće podatke: Cijena zatvaranja,
Dnevni maksimum,
Dnevni minimum,
Dnevni volumen.

Važno je napomenuti da analiziramo dnevne podatke odnosno podatke jednog dana.

Cilj je dobiti što točniju predikciju cijene zatvaranja za jedan dan unaprijed, no kako je to zbog velike volatilnosti Bitcoin-a skoro pa nemoguće zadovoljavajući rezultati bi bili ukoliko bi se predikcija kretala unutar raspona dnevnog maksimuma i minimuma.

6.1.1. Uvodno o Bitcoin-u

Bitcoin je decentralizirana, distribuirana, anonimna platna mreža, a ujedno i virtualna kriptovaluta koju ta platna mreža koristi. Bitcoin platna mreža funkcionira pomoću kompleksnog algoritma kojeg je izumio čovjek ili skupina ljudi pod pseudonimom Satoshi Nakamoto.

Plaćanja se provode na principu peer-to-peer mreže, bez središnje institucije ili jednog administratora. Novi bitcoin-i dodjeljuju se kao nagrada za obradu transakcija korisnicima koji nude svoju računalnu snagu za provjeru i spremanje uplata u javnoj knjizi.

Ovi pojedinci ili tvrtke uključuju se u aktivnosti održavanja Bitcoin mreže (tzv. "rudarenje") u zamjenu za transakcijske pristojbe i novonastale bitcoin-e. Osim rudarenjem, bitcoin-i se mogu kupiti novcem (u kunama, eurima, dolarima ili bilo kojoj drugoj valuti) ili se dobivaju u zamjenu za proizvode i usluge. Korisnici mogu slati i primati bitcoin-e elektroničkim putem za neznatnu naknadu pomoću računalnih programa (tzv. "novčanika"), koji se mogu nalaziti na osobnom računalu, mobilnom uređaju ili na internetu kao web aplikacija. Za razliku od kreditnih kartica, naknade za transakciju plaća kupac, a ne trgovac. Također je važno reći da su bitcoin transakcije nepovratne, jednom izvršenu transakciju nije moguće poništiti.

Peer to peer podrazumijeva:

> Koncept umrežavanja računala bez poslužitelja, gdje je svako računalo inteligentna radna stanica, koja pronalazi druga računala putem broadcast ethernet paketa, i komunicira s njima izravno, bez potrebe autorizacije na nekom centralnom poslužitelju. Primjer takve mreže su Microsoftove radne grupe (Workgroups), za razliku od domene (Domain) gdje se korisnici moraju prijaviti na centralni poslužitelj domene. [8]

6.1.2. Određivanje optimalnog broja lagova

Za početak potrebno je odrediti optimalan broj lagova za zadane varijable tj. baze podataka. Kod VECM modela koristimo sljedeće 4 varijable "Price, High, Low, Vol" to jest cijenu zatvaranja, dnevni maksimum i minimum te volumen.
Optimalni broj lagova određuje se naredbom:

```
varsoc (Price High Low Vol)
```

Naredba ispisuje tablicu koja nam daje optimalni broj lagova za naše podatke.

```
. varsoc (Price High Low Vol)

 Selection-order criteria
 Sample:  5-Jan-21 - 8-Dec-21                 Number of obs      =       338
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-13428.8				3.9e+29	79.4841	79.5022	79.5294
1	-12545.7	1766.2	16	0.000	2.3e+27	74.3532	74.4434	74.5794*
2	-12517.1	57.224	16	0.000	2.1e+27*	74.2786*	74.4409*	74.6858
3	-12513	8.1623	16	0.944	2.3e+27	74.3491	74.5835	74.9373
4	-12493.8	38.37*	16	0.001	2.2e+27	74.3303	74.6368	75.0994

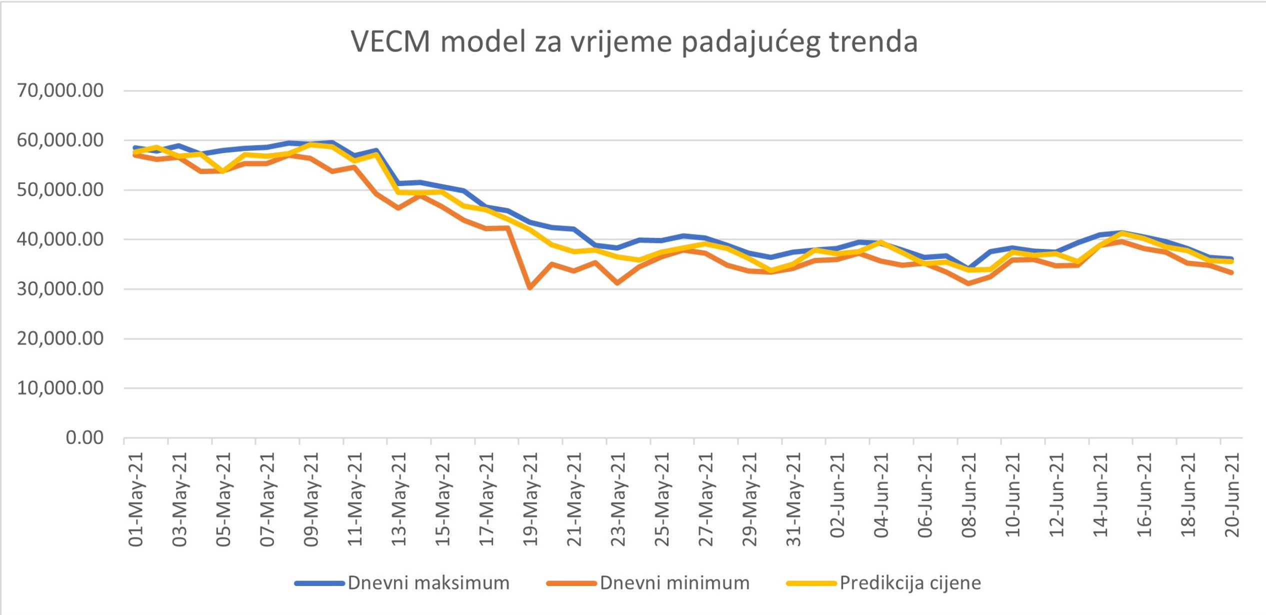
```
 Endogenous:  Price High Low Vol
  Exogenous:  _cons
```

Slika 6.1. Određivanje optimalnog broja lagova

Iz tablice očitavamo da je optimalan broj lagova za naše podatke 2.

6.1.3. Ponašanje za vrijeme padajućeg trenda

Analiziramo period koji traje 51 dan (01.05.2021. - 20.06.2021.) unutar kojega je cijena Bitcoin-a konstantno padala sa 57,000 $ sve do 29,310 $.



Slika 6.2. VECM model za vrijeme padajućeg trenda

Na slici 6.2. se može vidjeti kako su pogreške rijetke i da je žuta linija (predikcija cijene koju generira VECM model) gotovo stalno unutar granica dnevnog maksimuma i minimuma što nam daje odlične rezultate i postotke točnosti predikcija. Odstupanja su stvarno rijetka, predikcija je samo 6 puta bila izvan granica u periodu od 51 dana. 46 dana je predikcija bila točna što rezultira veoma velikim postotkom točnosti od 88% za vrijeme padajućeg trenda.

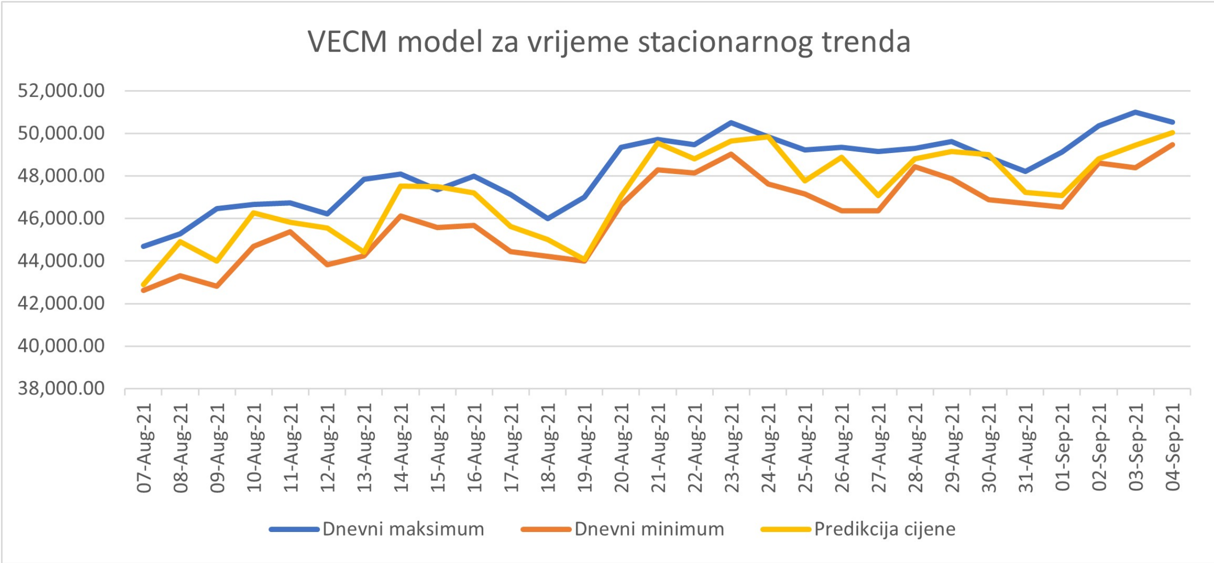
Zaključujemo da nam VECM model daje podosta točne predikcije za padajuće trendove s točnosti od čak 88%.

6.1.4. Ponašanje za vrijeme stacionarnog trenda

Analiziramo period od 29 dana (07.08.2021. - 04.09.2021.) unutar kojeg se cijena kretala gotovo stacionarno između dvije zone (44,000$ - 49,000$). (stacionarni trend je vrijeme dok cijena miruje ili se kreće unutar manjih zona)

Slika 6.3. na prvu izgleda kao rastući trend ali ako bolje pogledamo vidimo da se cijena Bitcoin-a kretala između 44,000$ - 48,000$ (07.08.2021. - 19.08.2021.) te se dogodio skok na druge dvije zone od 46,000$ - 50,000$ (20.08.2021. - 04.09.2021.), a cijena se nastavila kretati stacionarnim trendom.

Koristeći VECM model dobili smo točne predikcije 27 puta tj. 27 od 29 dana predikcija se nalazila unutar granica dnevnog maksimuma i minimuma te je samo 2 dana predikcija bila izvan



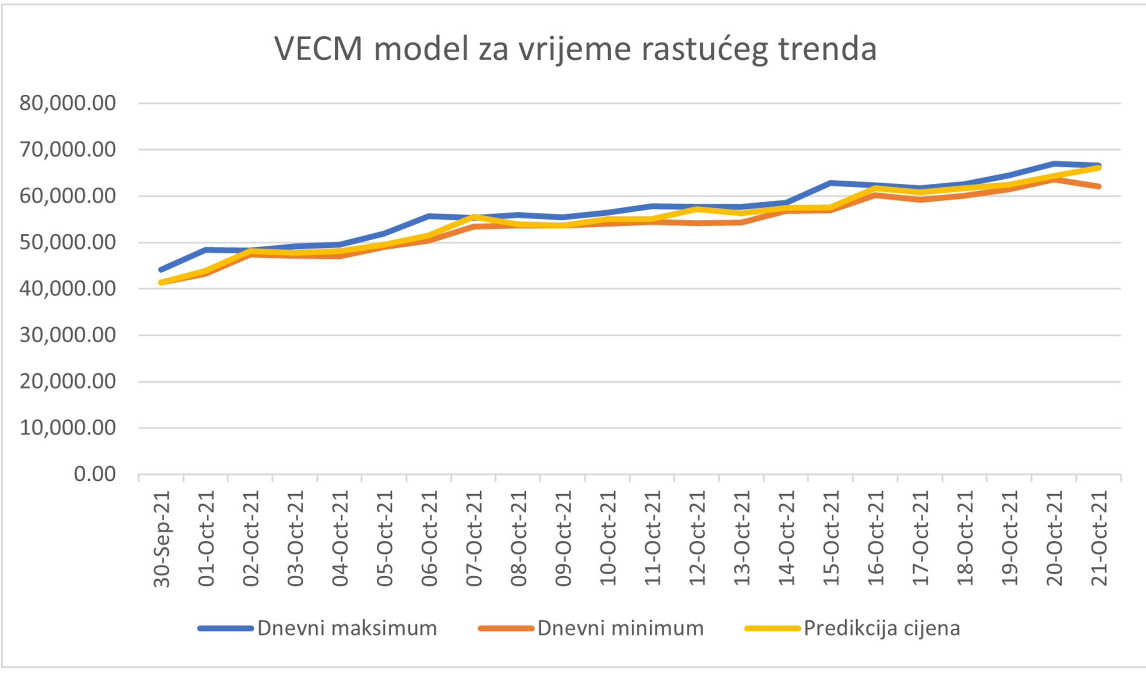
Slika 6.3. VECM model za vrijeme stacionarnog trenda

granica i to za samo malo više od 100$ što je pogreška od samo 0.3%.

Zaključujemo da nam VECM model daje jako točne predikcije i za stacionarni trend jer je postigao nevjerovatnu točnost od čak 93%.

6.1.5. Ponašanje za vrijeme rastućeg trenda

Ukoliko cijena konstantno raste bez značajnih padova možemo reći da smo ušli u rastući trend ili tako zvani "bull market". Analiziramo period od 22 dana (30.09.2021. - 21.10.2021.) unutar kojega je cijena rasla od 43,000$ do 66,000$.



Slika 6.4. VECM model za vrijeme rastućeg trenda

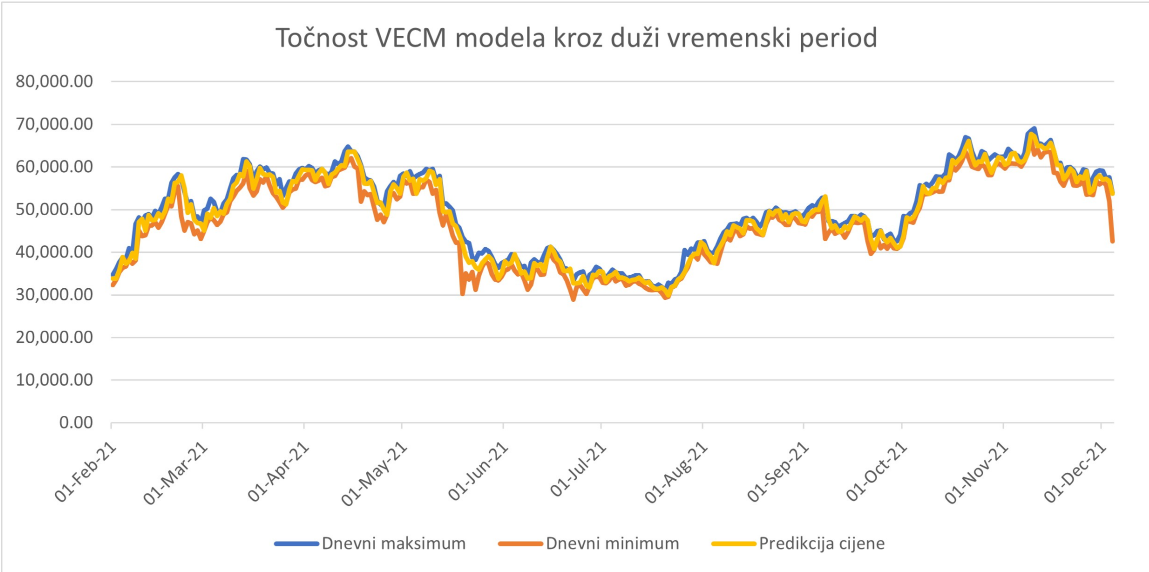
Možemo primjetiti da se predikcija (žuta linija) nalazi unutar granica dnevnog maksimuma i minimuma gotovo savršeno.

Postotak točnosti je veći za vrijeme rastućeg trenda nego za vrijeme padajućeg ali moramo uzeti u obzir da smo prilikom analize padajućeg trenda imali period od 51 dana dok za rastući trend koristimo period od 22 dana. Pa možemo reći da su im točnosti donekle iste.

Dakle, dobili smo točnu predikciju 20 puta odnosno 20 dana predikcija se nalazila unutar granica dnevnog maksimuma i minimuma, a izvan granica je bila samo 2 dana. Tako dobivamo jako veliki postotak točnosti od čak 91%. Zaključno VECM model daje također podosta točne predikcije i za rastući trend.

6.1.6. Ponašanje tijekom dužeg vremenskog perioda

Također ćemo promatrati točnost predikcija u periodu od 307 dana (otprilike 10 mjeseci). Unutar tih 10 mjeseci događala su se sva 3 trenda te je moguće dobiti dojam o točnosti VECM modela na duži vremenski period.



Slika 6.5. Točnost VECM modela kroz duži vremenski period

Zbog svoje dosta velike točnosti žuta linija (predikcija) na slici 5.5. izgleda kao da popunjava razmak između dnevnog maksimuma i minimuma.

Naravno postoji određeni broj pogreški ali postotci točnosti na duži vremenski period su također veoma zadovoljavajući.

Analizom podataka dobili smo zadovoljavajuće rezultate, predikcija je bila 261 puta točna odnosno dobili smo točnu predikciju (unutar granica) za 261 dan od ukupno 307 dana. Predikcija je bila izvan granica 46 dana što daje 15% pogreške.

Čak iako je bilo 46 pogreški od kojih je nekoliko skoro zanemarivo zbog greške od samo par desetaka dolara važno je uvidjeti da smo analizirali dosta veliki period od skoro godinu dana te da smo postigli jako veliku točnost od čak 85% na periodu od 307 dana.

Time zaključujemo da se VECM model može koristiti i za duže vremenske periode sa konstantnom i visokom točnošću.

6.2. Predikcija proizvodnje, uvoza i izvoza prirodnog plina

U ovome podpoglavlju pomoću VECM modela unutar statističkog softverskog paketa Stata pokušati ćemo dobiti što točnije predikcije proizvodnje, uvoza i izvoza prirodnog plina u Hrvatskoj za naredni mjesec. Podaci ovog istraživanja preuzeti su sa web stranica Državnog zavoda za statistiku te je važno napomenuti kako spaljene ili ispuhane količine plina nisu računate.

6.2.1. Uvod

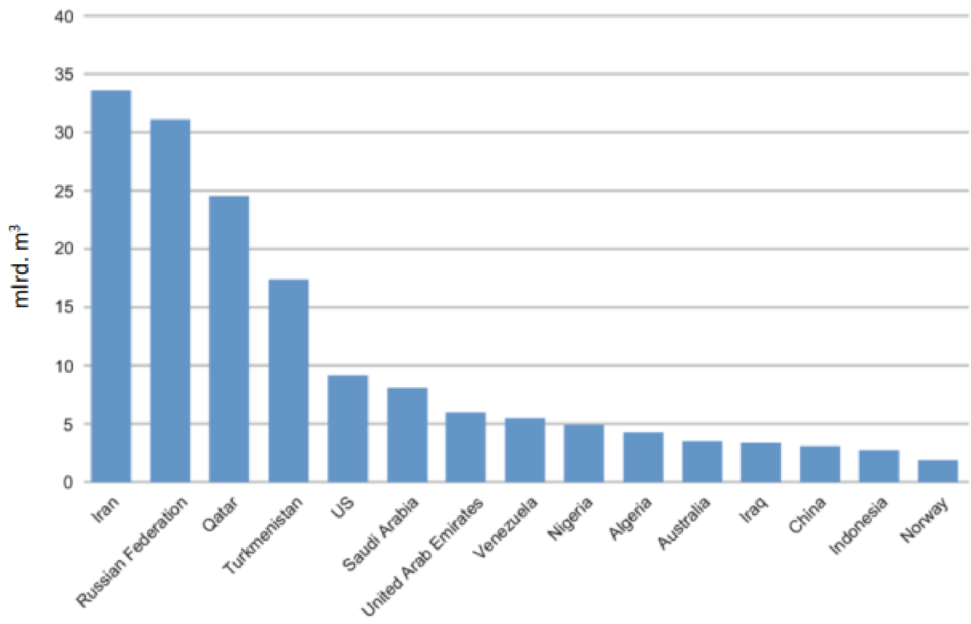
Prirodni plin smjesa je metana (molni udjel veći od 90%) s manjim udjelima etana, propana i viših ugljikovodika, a može sadržavati i nešto ugljikova dioksida, sumporovodika (takav se plin naziva kiselim), dušika, a katkad i helija i žive. S obzirom na udjel težih ugljikovodika razlikuju se: suhi plin, s neznatnim udjelom, i vlažni plin ili mokri plin, s povećanim i znatnim udjelom težih ugljikovodika iz plinskih i plinsko-kondenzatnih ležišta. Kao fosilno gorivo, prirodni plin ima ograničene zalihe. Procjene su da bi zalihe prirodnog plina, uz današnju razinu iskorištavanja, mogle potrajati još nekih sto godina. Najveći problemi s plinom leže u tome što se udio metana u njemu mijenja od države do države, pa tako na primjer udio metana u prirodnom plinu u Rusiji se kreće oko 98% dok je u Nizozemskoj taj udio od 80% do 85%. [9]

U 19. stoljeću, prirodni plin obično je dobivan kao usputni proizvod kod crpljenja nafte, pošto se oslobađaju određene količine plina kada tekućina prođe redukciju tlaka na putu iz podzemnih spremnika do površine, reakcijom sličnom poput otvaranja boce gaziranog pića. U takvim slučajevima prirodni plin koji nije imao potencijalno tržište u blizini crpilišta, postaje problem pošto ga je potrebno plinovodima dopremiti do krajnjeg korisnika. U 19. i početkom 20. stoljeću, takav neisplativi plin palio se na samim crpilištima nafte. Danas se pak takav plin upumpavanjem vraća natrag u podzemni spremnik od kuda je i došao, te se njegova distribucija odlaže za buduće potencijalno tržište ili zbog reguliranja tlaka u podzemnim spremnicima ne bi li se povećalo crpljenje preostale nafte. Ovisno o potražnji za prirodnim plinom ponekad se grade plinovodi s takvih crpilišta ako se utvrdi ekonomska isplativost istih. [9]

Druga mogućnost je da se prirodni plin otpremi kao tekućina ukapljivanjem GTL (engl. gas-to-liquids) tehnologijom u sitnetički benzin, dizel ili kerozin kroz Fischer-Tropschov postupak. Takav se plin lako distribuira putem tankera i konvencionalnih plinovoda. Smatra se da GTL plin gori čišće od naftnih goriva. Većina velikih naftnih korporacija distribuira GTL goriva. Ipak većina prirodnog plina komercijalno se vadi iz polja prirodnog plina. [9]

Najveći izvor zemnog plina u Republici Hrvatskoj se nalazi u Molvama gdje se proizvodi čak 70% plina za Republiku Hrvatsku. Tamo je i najmoderniji pogon za vađenje, prerađivanje i raspodjelu plina u ovom dijelu Europe.

S tvrtkom Gazprom, Rusija je trenutno najveći svjetski dobavljač prirodnog plina prema podacima iz 2020.-te godine. Procjenjue se da postoji oko 900 biljuna kubnih metara nekonvencionalnog plina poput plina iz škriljevca, od čega se smatra da je 180 bilijuna moguće iscrpiti. Mnoge znanstvene studije vide prirodni plin kao jedan od važnijih resursa za proizvodnju električne struje i toplana u budućnosti. Najveće svjetsko plinsko polje nalazi se u Qatarskom podmorju koje je procjenjeno na 25 bilijuna kubnih metara plina. Dovoljno da traje više od 420 godina uz optimalnu eksploataciju. Drugo najveće polje nalazi se pod Iranskim morem Perzijskog zaljeva i reda je veličine od 8-14 bilijuna kubnih metara prirodnog plina. [9]



Slika 6.6. 15 država s najvećim dokazanim rezervama prirodnog plina u milijardama kubičnih metara (2014.)

6.2.2. Određivanje optimalnog broja lagova

Za početak potrebno je odrediti broj lagova za zadane varijable. No, u ovome podpoglavlju radimo sa mjesečnim podacima, a ne sa dnevnim podacima kao u prethodnome podpoglavlju. Kod VECM modela koristimo varijable proizvodnje, uvoza i izvoza prirodnog plina te ćemo pokušati generirati što točnije predikcije za sve 3 varijable. Odnosno pokušati ćemo dobiti što točniju predikciju količine plina koju će Republika Hrvatska proizvesti, uvesti i izvesti za naredni mjesec (u mil. kubičnih metara).

Optimalni broj lagova određuje se naredbom:

```
varsoc (Proizvodnja Uvoz Izvoz)
```

Naredba ispisuje tablicu koja nam daje najoptimalniji broj lagova za naše podatke.

Iz tablice očitavamo da je optimalan broj lagova za naše podatke 4.

6.2.3. Proizvodnja prirodnog plina u Hrvatskoj

Analiziramo period od 2 godine i 3 mjeseca odnosno predikcije bilježimo od siječnja 2020. god. do ožujka 2022. god. s bazom podataka koja se kreće od siječnja 2016. godine. Cilj je dobiti što točniju predikciju proizvedene količine prirodnog plina u Hrvatskoj za naredni mjesec. Dozvoljena je pogreška predikcije od 8%, sve iznad toga je uzimano kao netočna odnosno nedovoljno točna predikcija. Nikakvi vanjski utjecaji nisu uzimani u obzir poput ekonomije, cijene plina, pro-

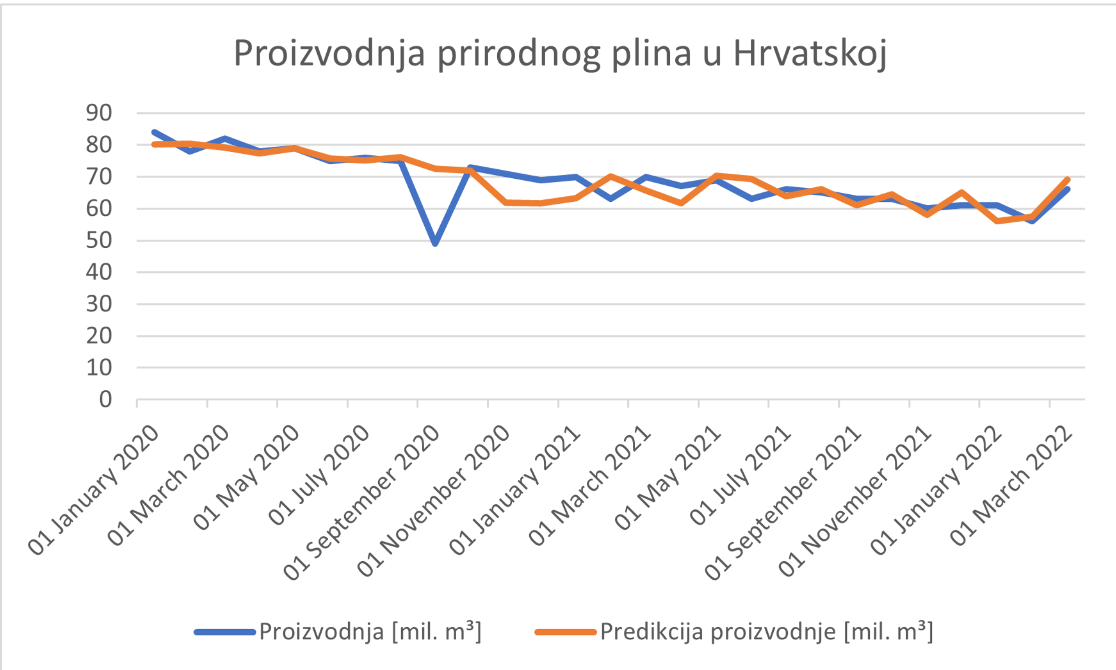
```
. varsoc Proizvodnjamilm Uvozmilm Izvozmilm

   Selection-order criteria
   Sample:  1960m6 - 1964m11                     Number of obs      =        54
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-663.483				1.1e+07	24.6846	24.7272	24.7951
1	-568.517	189.93	9	0.000	437059	21.5006	21.6711	21.9426
2	-541.191	54.652	9	0.000	222496	20.8219	21.1202	21.5954*
3	-528.907	24.569	9	0.003	198706	20.7002	21.1264	21.8052
4	-506.915	43.984*	9	0.000	124784*	20.2191*	20.7731*	21.6555

```
   Endogenous:  Proizvodnjamilm Uvozmilm Izvozmilm
    Exogenous:  _cons
```

Slika 6.7. Određivanje optimalnog broja lagova

blemi prilikom vađenja plina itd. već samo mjesečni podaci o proizvodnji, uvozu i izvozu plina u Hrvatskoj.



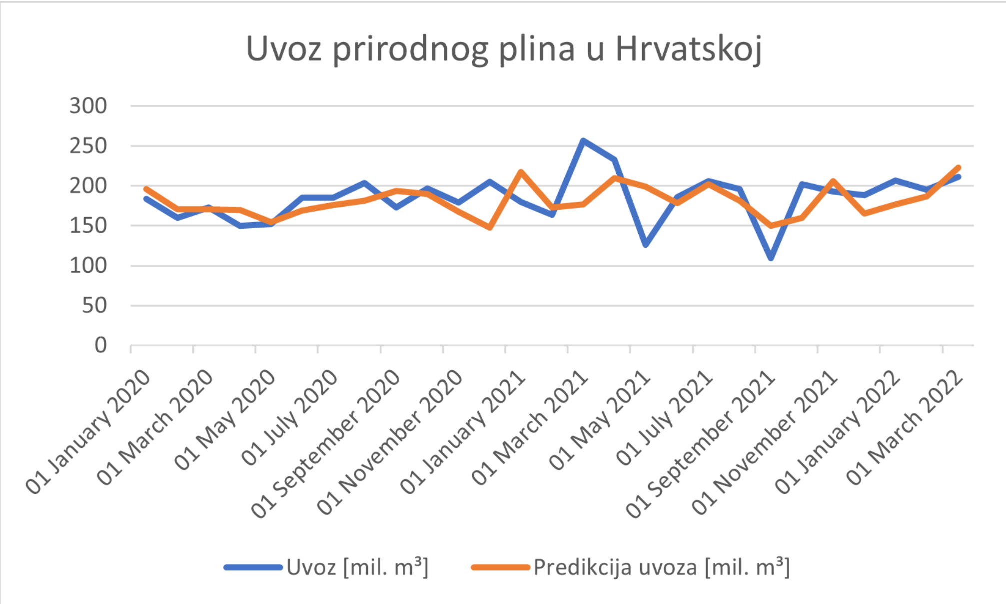
Slika 6.8. Usporedbna stvarnih i predviđenih količina proizvedenog prirodnog plina

Analizom dobivenih podataka sa slike 6.8. možemo uočiti da je predikcija bila točna 20 puta od sveukupno 27 promatranih mjeseci. Što daje točnost od 74.07% čime se VECM model pokazao čak i dobrim modelom s visokom točnosti pri analizi proizvodnje s time da uopće nisu uzimati nikakvi vanjski čimbenici već samo baza podataka koja datira od siječnja 2016.-te godine.

Dogodilo se 7 pogreški i to skoro sve za vrijeme naglog pada u proizvodnji (rujan 2020. godine) kada je Hrvatska proizvela samo 49 mil. kubičnih metara plina te kako koristimo broj lagova 4 narednih par mjeseci smo dobivali krivu predikciju jer nam model generira predikcije s težinom na zadnja 4 mjeseca. Zbog čega je došlo do 5 pogreški baš u tome periodu. Ova pogreška bi se mogla eliminirati dorađivanjem modela da mu se smanji broj lagova ili unese neka nova zavisna varijabla. Čime bi se točnost povećala na preko 85%.

6.2.4. Uvoz prirodnog plina u Hrvatskoj

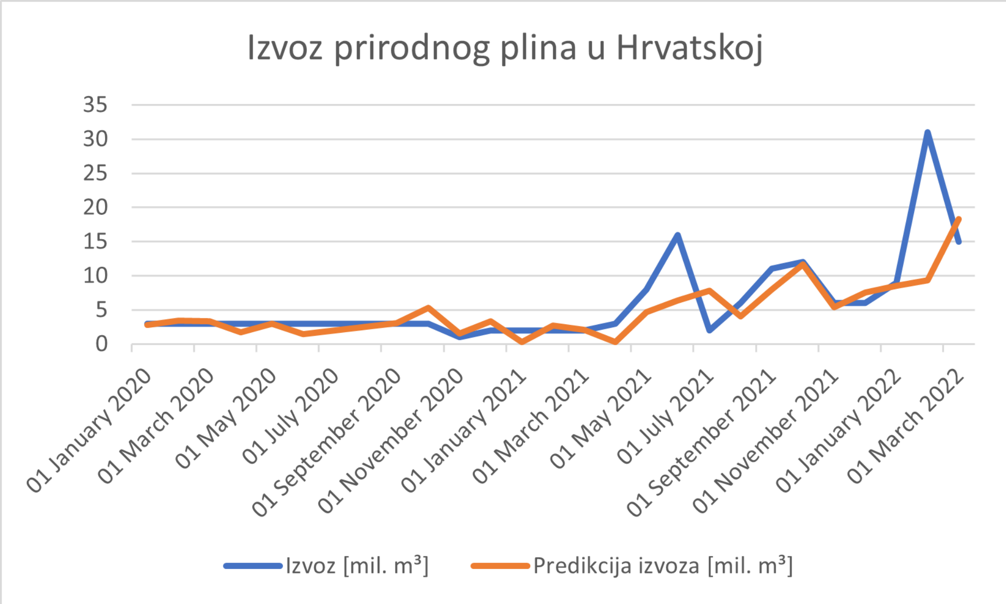
Također analiziramo period od 2 godine i 3 mjeseca s bazom podataka koja se kreće od siječnja 2016. godine. Cilj je dobiti što točniju predikciju uvezene količine prirodnog plina u Hrvatskoj za naredni mjesec. Također je dozvoljena pogreška predikcije od 8%, te je sve iznad toga je uzimano kao netočna odnosno nedovoljno točna predikcija. Također nikakvi vanjski utjecaji nisu uzimani u obzir što će na kraju utjecati na postotak točnosti predikcija jer uvoz ovisi o proizvedenoj količini plina ali također o politici i potrebama države.



Slika 6.9. Usporedbna stvarnih i predviđenih količina uvezenog prirodnog plina

Sa slike 6.9. analizom podataka možemo uočiti da je predikcija uvezene količine prirodnog plina bila točna 14 puta od 27 promatranih mjeseci. Što daje točnost od samo 51.85%.

Pogreške se konstantno javljaju jer je uvoz prirodnoga plina zasnovan na potrebi države i usko je povezan s politikom same države zbog čega ne dobivamo zadovoljavajuću točnost koristeći VECM model. Također zbog naglih promjena tzv. volatilnosti podataka model nije u stanju popratiti stvarne vrijednosti zbog visokog broja lagova. Bitcoin je također dosta volatilan ali smo upravo zbog njegove volatilnosti koristili manji broj lagova. Zaključno najbolje je ne koristiti VECM model kod predikcija uvoza zbog premale točnosti od 51.85% što je u rangu "bacanja novčica".



Slika 6.10. Usporedbna stvarnih i predviđenih količina izvezenog prirodnog plina

6.2.5. Izvoz prirodnog plina u Hrvatskoj

Također analiziramo period od 2 godine i 3 mjeseca s bazom podataka koja se kreće od siječnja 2016. godine. Cilj je dobiti što točniju predikciju izvezene količine prirodnog plina u Hrvatskoj za naredni mjesec. Također je dozvoljena pogreška predikcije od 8%, te je sve iznad toga je uzimano kao netočna odnosno nedovoljno točna predikcija. Nikakvi vanjski utjecaji nisu uzimani u obzir što će na kraju utjecati na postotak točnosti predikcija jer izvoz ovisi samo o politici države i višku proizvedenih količina što model jednostavno ne može uzeti u obzir.

Analizom podataka sa slike 6.10. možemo uočiti da nam je predikcija bila točna samo 6 puta od 27 promatranih mjeseci. Što je katastrofalan postotak točnosti od samo 22%.

Do ovako puno pogreški dolazi zbog toga što izvoz ovisi o politici države i eventualno o višku proizvedene količine ali izvoz Hrvatske je užasno malen, a uvoz gotovo troduplo veći nego proizvodnja. Zaključno promatranjem samo povijesnih podataka također nije moguće dobiti zadovoljavajuće predikcije za izvoz.

7. Zaključak

U ovome radu pojasnili smo pojmove vremenskih nizova, autokovarijance, autokorelacije, autoregresije, kointegracije i korekcije pogreške te pojmove nužne za shvaćanje istih. Nadalje, smisao ovog rada bio je uvesti matematički model za generiranje predikcija budućih stanja za dva prezentirana primjera. Prezentirani su primjeri iz različitih struka, odnosno iz grane ekonomije i energetike.

Vremenski niz je čisto statistički koncept koji omogućava promatranje evolucije neke veličine kroz vrijeme, drugim riječima na vremenski niz možemo gledati kao na signal. Naime, signal nije ništa drugo do funkcija koja vremenskom trenutku t pridružuje vrijednost $x(t)$. Upravo stoga možemo reći da smo u ovom radu objasnili jednu od metoda za prediktiranje budućih vrijednosti signala, pri čemu smo pretpostavljali da je vremenska varijabla diskretna.

Model koji se koristio kroz ovaj rad je VECM model. VECM model je zapravo vektorski autoregresijski (VAR) model s dodanim članom za ispravljanje pogrešaka. VECM model s matematičkog je stanovišta objašnjen u u četvrtome poglavlju, a navedena je i Johansenova specifikacija s različitim trendovima kod VECM modela.

Nadalje, u petom poglavlju, detaljno je objašnjen postupak kojim se generiraju predikcije unutar statističkog softverskog paketa Stata, dok su u šestom poglavlju opisana provedena empirijska istraživanja.

Zaključno možemo reći da je za predikciju cijene Bitcoin-a VECM model odlično rješenje i daje dosta visoku točnost predikcije za jedan dan unaprijed. Trend nije uopće bitan prilikom predikcije niti duljina analiziranog perioda jer je točnost za sva istraživanja bila iznad 85% tj. najlošiju točnost smo dobili za period od 307 dana gdje je točnost bila 85% što je i dalje dosta visoka i po našim kriterijima zadovoljavajuća točnost.

Međutim, iako je za predikciju cijene Bitcoin-a VECM model dobro rješenje u sljedećem empirijskom istraživanju o proizvodnji, uvozu i izvozu prirodnog plina u Hrvatskoj se baš i nije pokazao kao kvalitetno rješenje za sve varijable. Najveća točnost predikcija je postignuta za proizvodnju od 85% ukoliko je dozvoljeno 8% odstupanja od stvarne vrijednosti. No za uvoz i izvoz prirodnog plina model ne daje nimalo zadovoljavajuću točnost od 51.85% za uvoz te samo 22% za izvoz (dozvoljeno odstupanje od stvarne vrijednosti je također 8%). Na uvoz i izvoz utječe politika same države, ekonomija svijeta i drugi čimbenici koje VECM model jednostavno ne može razmatrati tj. uzeti u obzir.

Na kraju možemo reći da je VECM moćan alat u teoriji signala kojim se mogu predvidjeti buduća stanja signala u uvjetima kada je potrebno razmatrati višedimenzionalne signale, odnosno vremenske nizove koji se sastoje od više komponenti. Empirijski smo pokazali da u slučaju kad na promatrani vektor signala ne utječu previše vanjski faktori, što je bio slučaj kod Bitcoina-a VECM model daje praktički savršene predikcije, međutim u modelima kod kojih je za očekivati da vanjski faktori imaju značajan utjecaj, kao što je to kod uvoza i izvoza prirodnog plina, čisti VECM model neće biti od prevelike koristi.

Kao bitan faktor se pokazala i sama volatilnost promatranog vremenskog niza, odnosno za dobru predikciju model kroz čitavo promatrano razdoblje treba biti podjednake volatilnosti. Svako povećanje volatilnosti uzrokovalo je i narušavanje točnosti predikcije.

Literatura

[1] Grahovac, D.: "Analiza vremenskih nizova", dijelovi predavanja, s Interneta, https://www.mathos.unios.hr/index.php/30-homepage/odjel/nastava/kolegij/253-analiza-vremenskih-nizova, 24. travnja 2022.

[2] Hendry, David F.; Juselius, K.: "Explaining Cointegration Analysis: Part 1", The Energy Journal 21, 2000.

[3] Erjavec, N.; Bahovec, V.: "Uvod u ekonometrijsku analizu", Element, Zagreb, 2009.

[4] "Vector autoregression", s Interneta, https://en.wikipedia.org/wiki/Vector_autoregression, 12. travnja 2022.

[5] "Cointegration", s Interneta, https://hhr.wiki/detial/Cointegration, 12. travnja 2022.

[6] "Error correction model", s Interneta, https://en.wikipedia.org/wiki/Error_correction_model, 12. travnja 2022.

[7] "Introduction to vector error-correction models", s Interneta, https://www.stata.com/manuals/tsvecintro.pdf, 4. svibnja 2022.

[8] "Bitcoin", s Interneta, https://hr.wikipedia.org/wiki/Bitcoin, 14. svibnja 2022.

[9] "Prirodni plin", s Interneta, https://hr.wikipedia.org/wiki/Prirodni_plin, 25. ožujka 2022.

Sažetak i ključne riječi

Ovaj rad opisuje i definira teorijske pojmove poput vremenskog niza, autokovarijance, autokorelacije, autoregresije, kointegracije i korekcije pogreške nužne za definiranje i analizu VECM modela, koji je također detaljno objašnjen i matematički izveden. Zatim je pokazan postupak kojim se generiraju predikcije uz pomoć statističkog programa Stata. Na kraju ovog rada prezentirana su dva primjera iz različitih struka, kod kojih smo pokušali dobiti što točnije predikcije budućih stanja.

Ključne riječi: VECM model, vremenski niz, pojma korekcije pogreške, autoregresija, kointegracija, autokovarijanca, autokorelacija, slučajna varijabla, VAR, stacionarni procesi, vjerojatnost.

Summary and key words

This paper describes and defines theoretical concepts such as time series, autocovariance, autocorrelation, autoregression, cointegration and error correction necessary for defining and analyzing VECM models, which is also explained in detail and mathematically derived. Next, the procedure by which predictions are generated with the help of the Statistic program Stata is shown. At the end of this paper, two examples from different professions were presented, in which we tried to get as accurate predictions of future conditions as possible.

Keywords: VECM model, time series, error correction concept, autoregression, cointegration, autocovariance, autocorrelation, random variable, VAR, stationary processes, probability.