VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

FINANSŲ IR DRAUDIMO MATEMATIKOS BAKALAURO STUDIJŲ PROGRAMA

Bakalauro baigiamasis darbas

COVID-19 įtaka Europos bendrojo vidaus produkto prognozavimui

COVID-19 Impact on European Gross Domestic Product Forecasting

Roberta Vaštakaitė

Darbo vadovas Dr. doc. Martynas Manstavičius

VILNIUS 2021

Santrauka

COVID-19 įtaka Europos bendrojo vidaus produkto prognozavimui

Šiame darbe naudojantis ARIMA modeliu, prognozuojamas Europos bendrojo vidaus produkto rodiklis ir analizuota ar Covid-19 pandemija darė įtaką šio finansinio rodiklio prognozėms. Nagrinėjama statistikos ir laiko eilučių teorija, ARIMA modelio sudarymas bei papildomi prognozavimui reikalingi metodai.

Raktiniai žodžiai : Bendrasis vidaus produktas, prognozavimas, ARIMA modeliai.

Abstract

COVID-19 Impact on European Gross Domestic Product Forecasting

In this work European Gross Domestic Product was forecasted using ARIMA models and analysed if Covid-19 pandemic did impact this financial measure forecasts. Also analysed statistics and time series theory, ARIMA model composition and additional methods for time series forecasting.

Key words : Gross Domestic Product, forecasting, ARIMA models.

Turinys

1 Įvadas

Covid-19 yra infekcinė užkrečiamoji liga, kuri pasireiškia tokiais simptomais kaip kosulys, karščiavimas ar kvėpavimo sutrikimai. 2019 m. pabaigoje ši liga pradėjo plisti Kinijoje, tačiau neilgai trukus ji paplito visame pasaulyje. 2020 m. kovo 11 d. Pasaulio sveikatos organizacija [13] šį ligos protrūkį paskelbė pandemija. Kadangi šiai ligai būdingas lengvas užkrėčiamumas, daugelis pasaulio valstybių ėmėsi priemo-nių užkirsti kelią ligos plitimą. Siekant suvaldyti Covid-19 užkrečiamumą, valstybės nusprendė uždaryti šalies sienas, riboti parduotuvių, kavinių ir restoranų darbo laiką, masiniai susirinkimai buvo uždrausti. Tokios industrijos kaip turizmas, maitinimas ar renginių veikla visiškai sustojo. Verslams sustabdžius ar apribojus darbą, ekonominiai padariniai neišvengiami. BVP rodiklis buvo stipriai paveiktas tokių ekonominių pokyčių. Eurostat duomenimis 2020 m. Europos Sąjungos bendras vidaus produktas susitraukė 6,1 procentinio punkto. Toks netikėtas ekonominis šokas, nebuvo numatytas jokiose ekonominėse prognozėse.

Šiuo metu pandemija dar vis dar tęsiasi, todėl ekonominė žala jaučiama ir bus jaučiama ateityje. Pandemijos nuostoliai kiekvieną dieną iš naujo perskaičiuojami ir bandoma nuspėti kokie ekonominiai padariniai dar laukia ir kiek laiko truks ekonomikai atsistyti po šios krizės.

Šio darbo tikslas naudojantis EU BVP duomenimis ir matematiniais modeliais sužinoti kaip Covid-19 pandemija gali paveikti Europos ekonomiką ateinančius 2 metus. Šio darbo tyrimas – naudojantis laiko eilučių teorija, pritaikyti ARIMA (angl. *Auto Regressive Integrated Moving Average*) modelį, kuris pagal sukauptus praeities duomenis, bando nuspėti būsimas rodiklio reikšmes.

Šį darbą sudaro teorinė ir tyriamoji dalys. 2 skyriuje apžvelgiama statistikos teorija, 3 skyriuje – laiko eilučių teorija, 4 skyriuje plačiau išanalizuotas ARIMA modelio sudarymas, 5 skyriuje – keletas papildomų laiko eilučių analizės metodų, kurie reikalingi tyrimui, 6 skyriuje apžvelgiama bendrojo vidaus produkto sąvoka ir šio rodiklio reikšmė, 7 skyriuje atliekamas tyrimas: duomenų pavaizdavimas, jų stacionarumo tikrinimas, ieškoma geriausio ARIMA modelio ir duomenų prognozavimas, 8 skyriuje apžvelgiami gauti rezultai ir išvados.

2 Statistikos teorija

Tyrimo atlikimui, reikalingi tokie statistikos apibrėžimai kaip: vidurkis, dispersija, kovaricija, hipotezė ir kiti.

Aprašomoji statistika

Vidurkis

Vidurkis - tai taškas, kuris vidutiniškai artimiausias viesiem statistinės eilutės elementams. Tarkime, kad M koks nors skaičius. Atstumą tarp M ir statistinės eilutės elementų matuojame taip:

$$f(M) = (x_1 - M)^2 + (x_2 - M)^2 + \cdots + (x_n - M)^2.$$

Funkcija $f(M)$ pasiekia miniminumą taške $(x_1 + x_2 + \ldots + x_n)/n$. Pastarąjį skaičių vaidname imties vidurkiu (vidurkiu, aritmetiniu vidurkiu, empiriniu vidurkiu) ir žymime $\bar{x}$. Taigi vidurkis yra ne kas kita kaip visų statistinės eilutės elementų suma, padalyta iš jų skaičiaus [5]. Vidurkis yra užrašomas:

$$\textit{Imties vidurkis} \quad \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j,$$

$$\textit{Populiacijos vidurkis} \quad \mu = \frac{1}{N}\sum_{j=1}^{N} x_j.$$

Atsitikrinio dydžio vidurkis

Tolydaus atsitiktinio dyždio turinčio vidurkis [5]:

$$\mathbb{E}X = \int_{-\infty}^{\infty} xg(x)dx,$$

kur $g(x)$ yra X tankis, o X absoliučiai tolydus atsitiktinis dydis.

Dispersija

Imties dispersija parodo duomenų sklaidą apie vidurkį [5]:

$$\textit{Imties dispersija} \quad s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2,$$

$$\textit{Populiacijos dispersija} \quad \sigma^2 = \frac{1}{N}\sum_{j=1}^{n}(x_j - \mu)^2.$$

Atsitikrinio dydžio disperija

Atsitiktinio dydžio dispersija yra žymima [5]:

$$\mathbb{D}X = \mathbb{E}(X - \mathbb{E}X)^2.$$

Standartinis nuokrypis

Kvadratinė šaknis iš dispersijos vadinama teoriniu standartiniu nuokrypiu [5]:

$$\textit{Standartinis nuokrypis} \quad \sigma = \sqrt{\mathbb{D}X}.$$

Kovariacija

Kovariacija - tai skaitinė charakteristika, įvertinanti dviejų atsitiktinių dydžių tiesinę priklausomybę [5]:

$$X \textit{ ir } Y \textit{ kovariacija} \quad cov(X,Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y).$$

Koreliacija

Atsitiktinių dydžių X ir Y koreliacijos koeficientas yra apibrėžiamas taip [5] :

$$\varrho(X,Y) = \frac{cov(X,Y)}{\sqrt{\mathbb{D}X\mathbb{D}Y}} = \frac{\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y}{\sqrt{\mathbb{D}X\mathbb{D}Y}}.$$

Liekamųjų paklaidų kvadratų suma (SSE)

Liekamųjų paklaidų kvadratų suma (angl. *Error Sum of Squares*) parodo, kiek faktiškos stebėjimų reikšmės nukrypsta nuo apskaičiuotųjų pagal regresijos modelį [9]. SSE skaičiuojama :

$$SSE = \sum_{j=1}^{n} (X_j - \hat{X}_j)^2.$$

čia $\hat{X}_j$ yra regresijos tiesės reikšmė.

Maksimalus tikimybės įvertis

Maksimalus tikimybės įvertis skaičiuojamas naudojantis formule [11] :

$$\hat{\sigma}_k^2 = \frac{SSE(j)}{n},$$

kur $SSE(k)$ yra liekamųjų paklaidų kvadratų sumų modeliui su j regresijos koeficientu.

Statistinės išvados

Hipotezė

Bet koks teiginys apie populiacijos parametrų reikšmes vadinamas parametrine hipoteze. Statistinę parametrinę hipotezę sudaro du alternatyvūs teiginiai apie galimas parametro reikšmes. Problema formuluojama kaip spėjimas apie galimas parametro θ reikšmes, priklausančias Θ_0, pateikiant alternatyvą, kad θ priklauso Θ_1 [5].

$$\begin{cases} H_0 : \theta \in \Theta_0, \\ H_1 : \theta \in \Theta_1. \end{cases}$$

Statistika

Atsitiktinės imties mati funkcija $f(X_1, X_2, \ldots, X_n)$ vadinama statistika.

Statistinėms išvadoms naudojama kokia nors imties duomenų funkcija - statistika. Statistika yra atsitiktinių dyždių funkcija, todėl ji taip pat yra atsitiktinis dydis, jei funkcija yra mati [5].

Kriterijaus reikšmingumo lygmuo

Kriterijaus reikšmingumo lygmuo yra išreiškiamas [5]:

$$\alpha = \mathbb{P}(H_0 \ atmetame | H_0 \ teisinga).$$

P-reikšmė

Tarkime, tikriname hipotezę su vienpuse alternatyva :

$$\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta > \theta_0. \end{cases}$$

Tegul sprendimui naudojama statistika $T(X_1, X_2, \ldots, X_n)$ yra tolydi. Tarkime, $t*$ yra statistikos T realizacija.

Tikimybė, kad kriterijaus statistika T (tuo atveju, kai H_0 teisinga) ne mažesnė už stebimą realizaciją $t*$, vadinama p-reikšme [5].

$$p = \mathbb{P}(T \geq t*),$$

kai H_0 teisinga.

Tegul α yra reikšmingumo lygmuo, o p yra p-reikšmė [5].

Jeigu $p < \alpha$, tai hipotezė H_0 atmetama.

Jeigu $p \geq \alpha$, tai hipotezė H_0 neatmetama.

Pasikliautinieji intervalai

Parametrų įvertiniai yra atsitiktiniai dydžiai. Jų realizacijos yra išsibarsčiusios apie tikrąją parametro reikšmę. Taikymams svarbu žinoti intervalą, kuriam gali priklausyti nežinomas parametras. Tarkime, stebime atsitiktinį dydį, turintį skirstinį P_θ su nežinomuoju $\theta \in \Theta$. Pasirenkame skaičių $0 < Q < 1$. Tegul $\hat{\theta_1}, \hat{\theta_2}$ dvi tokios statistikos, kad $P(\hat{\theta_1} \leq \theta \leq \hat{\theta_2}) = Q$, intervalas $[\hat{\theta_1}, \hat{\theta_2}]$ vadinamas parmetro θ pasikliautinuoju intervalu [5].

3 Laiko eilučių teorija

Laiko eilutė

Atsitiktinio dydžio reikšmes tam tikrais laiko momentais galima užrašyti šių reikšmių seka - laiko eilute. Tokie stebėjimai turėtų būtų atlikti vienodais laiko intervalais. Tiriant laiko eilutes, galima nuspėti kokias reikšmes galima gauti ateityje arba kitaip tariant, galimas atsitiktinio dydžio reikšmių prognozavimas.

Laiko eilutė yra užrašoma:

$$X_t = m_t + s_t + Y_t, \quad t = 1, \ldots, n, \quad t \in \mathbb{T},$$

čia m_t yra lėtai kintanti funkcija, dar vadinama trendu, s_t yra funkcija su žinomu periodu d, vadinama sezoniškumo komponente ir Y_t yra atsitiktinis klaidžiojimas [3].

Trendas

Trendas atspindi pagrindines ir ilgalaikes laiko eilutės tendencijas, esminius eilutės bruožus. Trendas gali būti tiesinis arba netiesinis (logaritminis, eksponentinis ir kiti).

Sezoniškumas

Sezoniškumas - tai reguliarus stebimų reikšmių didėjimas ar mažėjimas apibrėžtais laiko periodais. Sezoniškumas yra ciklinis procesas, pavyzdžiui : oro temperatūra. Oro temperatūra vasarą ir žiemą drastiškai skiriasi, bet žiūrit į keletos ar kelesdešimties metų duomenų eilutę, sezoniškumas yra matomas, vasarą temperatūra yra aukšta, o žiemą - žema. Dažnai iš grafinio duomenų atvaizdavimo galima nustatyti, jog sezoniškumas egzistuoja.

Autokovariacinė funkcija

Tegu $\{X_t, t \in \mathbb{T}\}$ yra procesas, turintis baigtinę dispersiją $\mathbb{D}(X_t) < \infty, t \in \mathbb{T}$, tada proceso X_t autokovariacinė funkcija yra apibrėžiama [6] :

$$\gamma X(r, s) = Cov(X_r, X_s) = \mathbb{E}(X_r - \mathbb{E}X_r)(X_s - \mathbb{E}X_s).$$

Stacionarumas

Laiko eilutė $\{X_t, t \in \mathbb{Z}\}$ vadinama stacionaria ($\mathbb{Z}$ yra sveikųjų skaičių aibė) [6], jei

- $\mathbb{E}X_t = \mu, t \in \mathbb{Z}$,
- $\mathbb{E}|X_t|^2 < \infty, \forall t$,
- $\gamma X(r, s) = \gamma X(r + t, s + t), \quad r, s, t \in \mathbb{Z}$.

Nuo laiko eilutės stacionarumo priklauso vidurkio funkcijos pavidalas. Stacionarus procesas proceso laiko eilutės reikšmės kinta atsitiktinai kiekvienu momentu, tačiau vidurkis gali ilgą laiką nekisti. Nestacionarių procesų vidurkis nėra pastovus ir ilguoju laikotarpiu kinta. Todėl svarbu, kad laiko eilutė būtų stacionari, jog ji būtų nuspėjama [12].

Autokoreliacinė funkcija (ACF)

Autokoreliacinė funkcija matuoja tiesinę laiko eilutės priklausomybę laiku t. ACF priklauso tik nuo laiko eilutės elementų, kurie skiriasi per k intervalų. Ši funkcija apibrėžiama [3] :

$$\rho(k) = \frac{\sum_{t=k+1}^{n}(X_t - \overline{X}_t, X_{t-k} - \overline{X}_t)}{\sum_{t=1}^{T}(X_t - \overline{X}_t)}.$$

Dalinė autokoreliacinė funkcija (PACF)

Dalinė autokoreliacinė funkcija apibrėžiama [3]:

$$\alpha(k) = Corr(X_{t+k} - \hat{X}_{t+k}, X_t - \hat{X}_t),$$

čia $\hat{X}_t$ yra nario X_t prognozė. $\alpha(k)$ galime suprasti kaip koreliaciją tarp X_{t+k} ir X_t, atsižvelgiant į tarpines reikšmes arba kaip sąlyginę koreliaciją.

Baltas triukšmas

Stacionarus procecas $\{w_t, t \in \mathbb{Z}\}$ vadinamas balto triukšmo procesu kiekvienam t jeigu [3]:

- $\mathbb{E}w_t = 0,$
- $\gamma(h) = \begin{cases} \sigma^2, & h = 0, \\ 0, & h \neq 0. \end{cases}$

Baltas triukšmas taip gali būti žymimas $w_t \sim WN(0, \sigma^2)$.

Poslinkio operatorius

Poslinkio operatorius B (angl. *Backshift, lag*) kiekvienam t yra apibrėžiamas [6]:

$$BX_t = X_{t-1},$$

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t.$$

.

4 ARIMA modelis

ARIMA modelis yra laiko eilučių modelis, kuris leidžia X_t priklausomiems kintamamiesiems būti nuspėtiems pagal praeities reikšmes. ARMA sudarytas iš 2 kitų skirtingų modelių: AR (angl. *autoregressive*) ir MA (angl. *moving average*) modelių [11].

4.1 AR modelis

Autoregresiniai modeliai yra grįsti tuo, jog dabartinė reikšmė X_t, gali būti paaiškinta kaip $X_{t-1},\ X_{t-2}, \dots,\ X_{t-p}$ funkcija, kur parametras p yra skaičius žingsnių į praeities reikšmes, kurių reikia prognozuoti reikšmes į priekį [11].

Autoregresinis modelis AR(p) :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + w_t,$$

kur X_t yra stacionarus, w_t yra baltas triukšmas, o $\phi_1, \dots, \phi_p$ yra parametrai ir $\phi_p \neq 0$.

Autoregresinis polinomas yra užrašomas:

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p, \quad \phi_q \neq 0.$$

4.2 MA modelis

Slenkančio vidurkio modelis yra alternatyva autoregresiniam modeliui, kur X_t kairioje pusėje nelygybė yra manoma jog bus tiesinė, o slenkančio vidurkio modelis su parametru q ir baltu triukšmu w_t yra dešinoje lygybės pusėje nurodo lygybę kuri yra derinama tiesiškai iš surinktų duomenų.
Slenkančio vidurkio MA (q) modelis yra užrašomas [11]:

$$X_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q},$$

kur w_t yra baltas triukšmas, o $\theta_1, \dots, \theta_q$ yra parametrai.

Slenkančio vidurkio polinomas :

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q, \quad \theta_q \neq 0.$$

4.3 ARMA modelis

Procesas $\{X_t, t \in \mathbb{Z}\}$ yra vadinamas ARMA(p,q) procesu, jeigu $\{X_t\}$ yra stacionarus ir $\forall t$ [11]:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q},$$

kur $\phi_p \neq 0$, $\theta_q \neq 0$ ir $\sigma_w^2 > 0$. Parametrai p ir q yra vadinami autoregresiniu ir slenkančio vidurkio parametrais atitinkamai.

ARMA(p, q) procesą dar galima užrašyti :

$$\phi(B) X_t = \theta(B) w_t, \quad t \in \mathbb{Z}.$$

Kauzalumas

Procesas ARMA(p,q) yra vadinamas kauzaliu, jeigu egzistuoja konstantų seka $\{\psi_j\}$ tokia, kad [6]:

$$\sum_{j=0}^{\infty} |\psi_j| < \infty,$$

$$X_t = \sum_{j=0}^{\infty} \psi_j z_{t-j}, \quad j \in \mathbb{Z}.$$

4.4 ARIMA modelis

Bendresnis ARMA modelio atvejis, kuris apima ir nestacionarius procesus, yra ARIMA modelis.

ARIMA(p,d,q) - autoregresinis integruotas slenkančio vidurkio procesas (angl. *autoregressive integrated moving average*). Parametras p yra autoregresinis, d - integruotas, q - slenkančio vidurkio parametras [11].

Tarkime d yra neneigiamas sveikasis skaičius, tada procesas $\{X_t\}$ yra vadinamas ARIMA(p,d,q) procesu, jeigu procesas $Y_t = (1-B)^d X_t$ yra kauzalus ARMA (p,q) procesas. Vadinasi, procesas $\{X_t\}$ tenkina lygtį

$$\phi(B)(1-B)^d X_t = \theta(B) w_t$$

Stacionariam procesui $d = 0$ modelis supaprastinamasiki iki ARMA(p,q), t.y. ARIMA$(p,0,q)$ = ARMA(p,q).

5 Papildomi metodai

5.1 Dickey – Fuller testas

Stacionarumas gali būti patikrintas Dickey-Fuller testu. Šis testas ieško vienetinių šaknų stacionarumui. Bendros šaknys gali reikšti nenuspėjamus rezultatus laiko eilučių analizėje. Toks stacionarumo tikrinimas, yra naudojamas statistikos testo principu, kai gauname reikšmę, kuri nurodo ar galima atmesti hipotezę, kad laiko eilutė yra nestacionari. Tikrinant stacionarumą, kritinės reikšmės - Stjudento skirstinio α lygmens kritinės reikšmės [4].

Tarkime, kad:

$$X_t = \phi X_{t-1} + w_t,$$

kur w_t yra baltas triukšmas.

Testo hipotezė:

- $H_0 : \phi = 1$ (egzistuoja vienetinė šaknis ir laiko eilutė nėra stacionari),
- $H_1 : \phi < |1|$ (neegzistuoja vienetinė šaknis laiko eilutė yra stacionari).

Naudojantis pateiktais teiginiais, sudarome statistiką:

$$\tau = \frac{\hat{\phi} - 1}{SSE(\hat{\phi})}.$$

Jei statistikos τ reikšmė yra mažesnė nei $-3,45$, galime atmesti nulinę hipotezę ir teigti jog laiko eilutė yra stacionari. Kitu atveju, laiko eilutę turi vienetinę šaknį ir nulinės hipotezės atmesti negalime.

5.2 ACF ir PACF ARIMA modelio nustatymui

Jei laiko eilutė yra stacionari, kitas žingsnis būtų nustatyti reikiamą ARIMA modelį pateiktiems duomenims. Nubraižius autokoreliacijos ir dalinės autokoreliacijos grafikus galime identifikuoti AR(p) ir MA(q) parametrus. ACF grafikas atvaizduoja koreliacijos koeficientą tarp laiko eilutės duomenų ir jos poslinkio. PACF grafikas rodo dalinės korealiacijos koeficientą tarp laiko eilutės ir jos poslinkio.

Jeigu PACF grafikas rodo žymų reikšmių kritimą, kur poslinkio opearatorius yra teigiamas, tai paskutinė reikšmė nurodo AR modelio parametrą p. Poslinkis, kur PACF pastebimai sumažėja yra yra AR parametro p reikšmė. Analogiškai ACF jei grafikas rodo reikšmių kritimą ir poslinkio operatorius yra neigiamas. Poslinkis, kur ACF pastebimai sumažėja yra MA parametro q reikšmė [10].

5.3 Kriterijai modelio tinkamumui

Žemiau pateikti kriterijai naudojami atrinkti geriausią modelį iš keletos konkuruojančių modelių. Atsižvelgiant į duomenų reikšmes, kriterijai įvertina modelio kokybę. Visų kriterijų savybė tokia, kad kuo kriterijaus statistikos reikšmė mažesnė, tuo modelis geresnis.

Akaike informacinis kriterijus (AIC)

$$AIC = \log \hat{\sigma}_k^2 + \frac{n+2k}{n},$$

kur $\hat{\sigma}_k^2$ yra maksimalus tikimybės įvertis, k yra parametrų skaičius modelyje ir n stebėjimų skaičius [11].

Akaike - Bajeso informacinis kriterijus (AICc)

Šis kriterijus yra naudojamas mažesniam duomenų kiekiui nei AIC atveju.

$$AICc = \log \hat{\sigma}_k^2 + \frac{n+k}{n-k-2},$$

kur $\hat{\sigma}_k^2$ yra maksimalus tikimybės įvertis, k yra parametrų skaičius modelyje ir n stebėjimų skaičius [11].

Bajeso informacinis kriterijus (BIC)

$$BIC = \log \hat{\sigma}_k^2 + \frac{k \log n}{n},$$

kur $\hat{\sigma}_k^2$ yra maksimalus tikimybės įvertis, k yra parametrų skaičius modelyje ir n stebėjimų skaičius [11].

6 Bendrasis vidaus produktas

Bendrasis vidaus produktas (BVP) – rodiklis, apibrėžiamas kaip prekių ir paslaugų sukurtų šalyje vertė tam tikru laiko periodu [2].

$$BVP = C + I + G + (X - M),$$

kur C - vartojimas (*consumption*), I - investicijos (*investment*), G - valstybės išlaidos (*goverment spending*), X - eksportas (*exports*), M - importas (*imports*).

Bendras vidaus produktas yra vienas svarbiausių rodiklių nurodant valstybės ekonominį lygį. Šiuolaikiniame pasaulyje šis rodiklis plačiai naudojamas stebėti ekonominį aktyvumą šalyse, žemyne ar pasaulyje. Taip pat BVP yra geras rodiklis lyginti šalių ekonomiką ir jų išsivystymo lygį.

BVP prognozavimas

Naudojantis Box-Jenkins metodologija bendrojo vidaus produkto prognozavimas yra naudojamas jau daugelį metų. Autoregresinis integruotas slenkančio vidurkio modelis yra vienas iš plačiausiai naudojamų BVP rodiklio prognozavimo metodų. Naudojantis praėjusių metų (pusmečio ar ketvirčio) duomenimis, juos išanalizavus ir pritaikius tinkamą ARIMA modelį, gaunamos reikšmės, kurios yra BVP rodiklio prognozės.

Bendrojo vidaus produkto prognozėms, buvo pritaikyti tokie modeliai - ARIMA (1,1,4) - Indijos BVP, paveiktam Covid-19 krizės [8], ARIMA (1,2,1) Egipto BVP prognozėms [1], ARIMA (1,1,1) Sudano BVP prognozėms [7].

BVP ir COVID-19

Bet kuriai bendrojo vidaus produkto komponentei sumažėjus, BVP tiesiškai sumažėja. Galima teigti, jog visos bendrojo vidaus produkto sudedemosios komponentės sumažėjo dėl Covid-19 pandemijos atsargumo priemonių. Žmonės mažiau vartojo, nes daugelis veiklos rūšių nevykdė veiklos. Investicijos taip pat galėjo mažėti dėl situacijos nenuspėjamumo. Importas ir eksportas galimai sumažėjo, dėl neveikiančių įmonių Europoje ir už jos ribų. Covid-19 pasaulyje pradėjo sklisti dar 2020 metų pirmame ketvirtyje, nors didžiausias smukimas matomas 2 ketvirtyje. Eurostat duomenimis bendras EU BVP per 2020 metus susitraukė per 6,1 %. Labiausiai BVP smuko Ispanijoje (-10,8%), Italijoje (-8,9%) ir Prancūzijoje (-8,1%), kur Covid-19 susirgimų buvo registruota daugiausiai [Eurostat duomenys, Priedas Nr. 1].

7 Tyrimas

Pasiruošimas ARIMA modelio pritaikymui

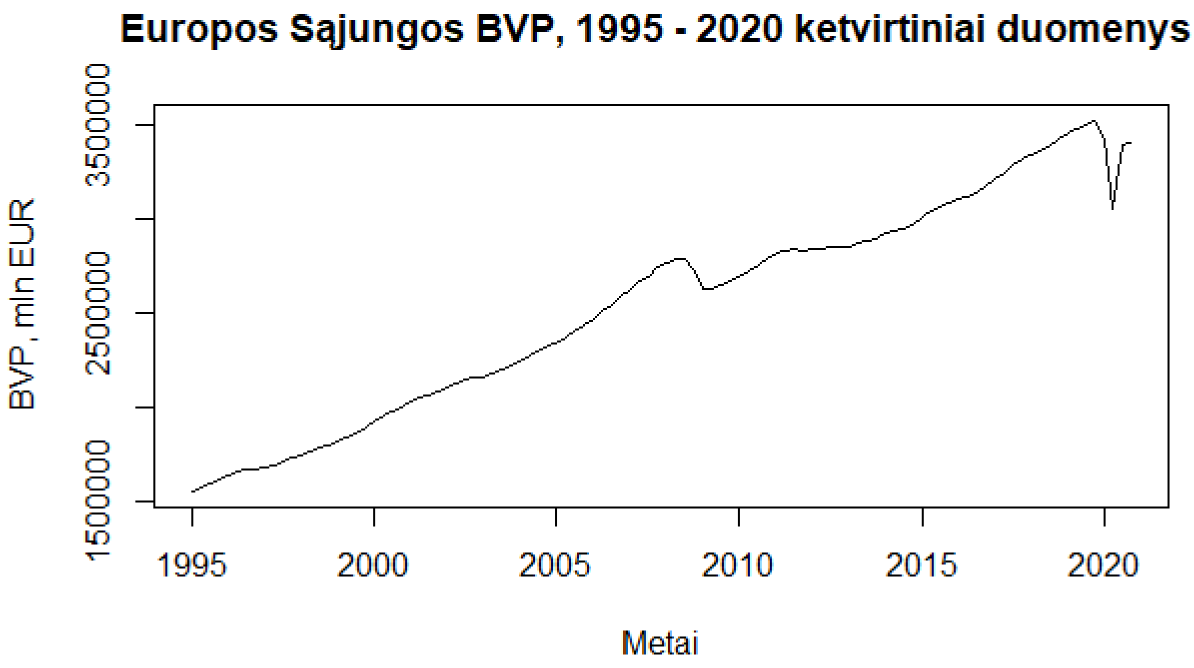
Prieš pradedant naudoti ARIMA modelį, reikia pasiruošimo. Mokslininkai G. Box ir J. Jenkins išskyrė keletą žingsnių, kuriais patariama vadovautis prognozuojant ARIMA modeliu:

- Duomenų pavaizdavimas,
- Stacionarumo patikrinimas,
- Duomenų transformavimas (priklausomai nuo situacijos),
- Modelio pasirinkimas,
- Prognozavimas.

7.1 Duomenų analizė

Duomenų pavaizdavimas

Tyrimui pradėti, pasirinkti yra Europos Sąjungos BVP ketvirtiniai duomenys. Duomenys pasirinkti kas ketvirtį, tokiu būdu stebėjimų skaičius žymiai didesnis, todėl galima tikėtis tikslesnių prognozių. *Eurostat* kas ketvirtį pateikia Europos sąjungos ir visų jos šalių (taip pat Europos šalių kurios ir nepriklauso ES) bendrojo vidaus produkto rodiklius. Šie parinkti duomenys jau yra pakoreguoti mažinant sezoniškumo įtaką. Todėl galima laikyti, jog šie duomenys neturi sezoniškumo komponentės (Priedas 1). Šie duomenys pateikti žemiau esamčiame grafike.



1 pav.: Europos Sąjungos bendrasis vidaus produktas, 1995 - 2020 ketvirtiniai duomenys

Stacionarumas

Prieš pradedant duomenų tyrimą, svarbu patikrinti ar laiko eilutė yra stacionari. Naudojantis programa R, stacionarumui galima naudoti Papildytą Dickey-Fuller (ADF) testą (*adf.test*).

Statistikos reikšmė	P-reikšmė
-2,3585	0,428

1 lentelė: ADF testas realiems duomenims

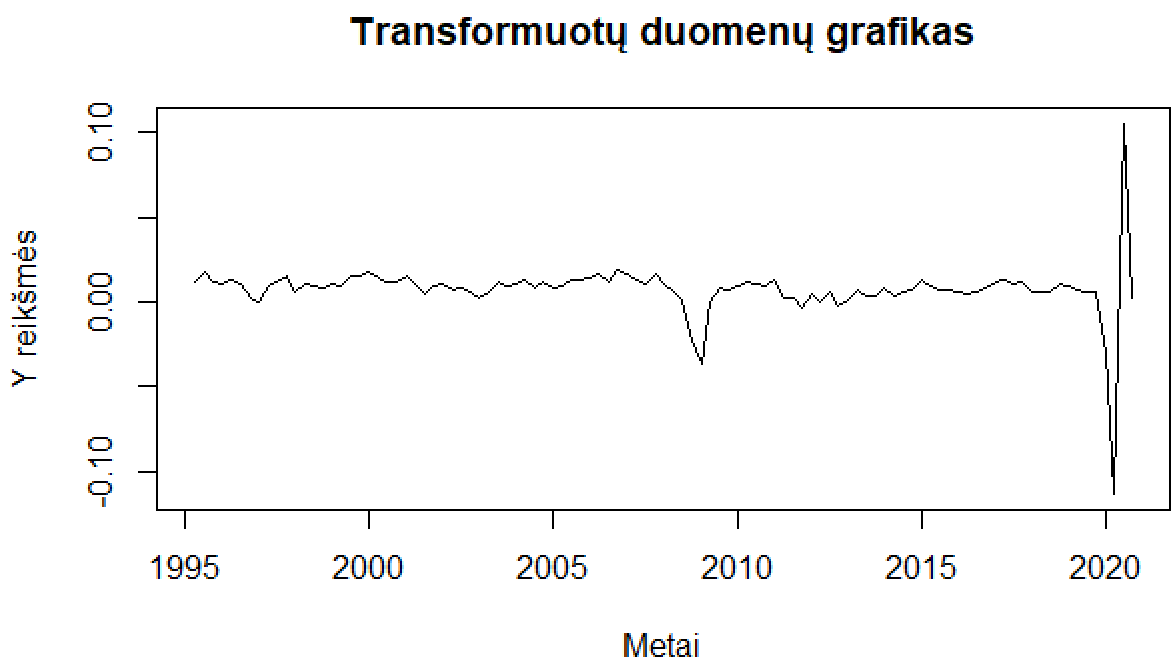
Pasirinkame, jog reikšmingumo lygmuo $\alpha = 0,05$. Atlikus šį testą, testo reikšmė lygi -2,3585, p-reikšmė = 0,428. Todėl nulinės hipotezės negalime atmesti, tai reiškia, kad duomenys yra nestacionarūs.

Duomenų transformavimas

Kadangi pradiniai duomenys nėra stacionarūs, išgauti stacionarią eilutė, gautus duomenis galime naudojant tokią formulę:

$$Y_t = \triangle \log X_t = \log X_t - \log X_{t-1}$$

Naudojantis šia formule, sukuriame naują eilutę, kuri stabilizuoja dispersiją, kuri pradiniu atveju yra nepastovi skirtingu laiko momentu. Transformuoti duomenys pateikti žemiau esančiame grafike.



2 pav.: Tranformuoti Europos Sąjungos BVP duomenys (logaritmų skirtumas)

Matome, jog Y reikšmės yra išsibarsčiusios apie nulį. 2020 metų BVP pokytis, vis dar matomas ir daro įtaką tolesniems skaičiavimams.

Naujiems transformuotiems duomenims atliekame ADF testą, t.y. tikriname jų stacionarumą.

Statistikos reikšmė	P-reikšmė
-4,1816	0,01

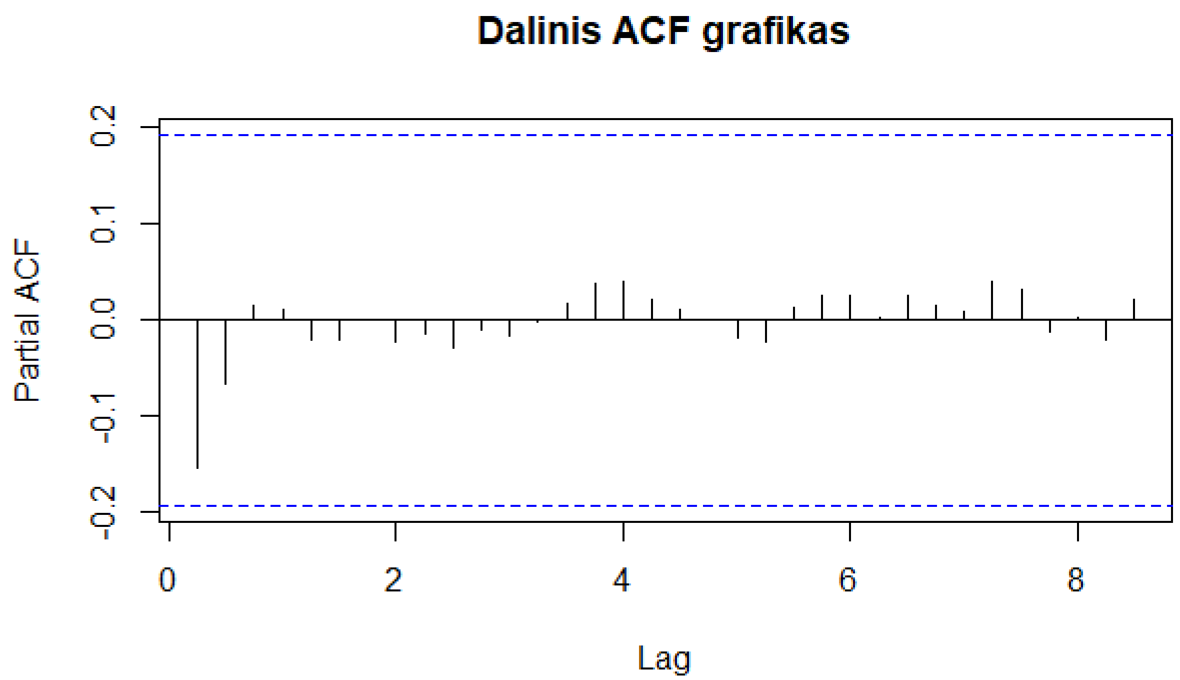
2 lentelė: ADF testas transformuotiems duomenims

Pasirinkame, jog reikšmingumo lygmuo $\alpha = 0,05$. Dabar atlikus ADF testą, gavome, jog, testo reikšmė yra -4,1816, o p-reikšmė lygi 0,01. O tai yra mažiau už reikšmingumo lygmenį α. Todėl galime atmesti H_0 hipotezę, ir teigti kad duomenys yra stacionarūs. Dabar laiko eilutė yra stacionari, todėl galima ieškoti geriausio ARIMA modelio varianto prognozavimui.

7.2 Modelio parinkimas

Modeliui pasirinkti, galima paskaičiuoti autokoreliacijos ir dalinės autokoreliacijos funkcijas. Iš šių funkcijų grafikų galime nuspėti, koks ARIMA modelis yra tinkamiausias turimiems duomenims. Atliekame ACF ir PACF testus, kad galėtume identifikuoti tinkamą modelį.

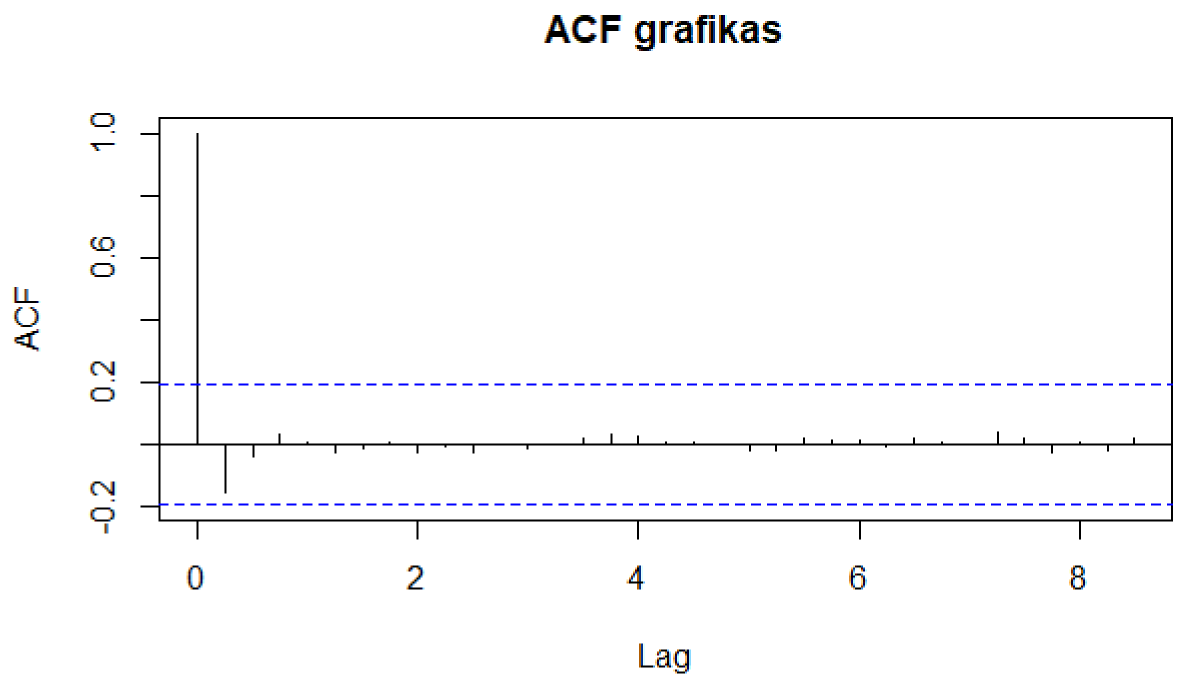
Panaudoję funkciją *adf* ir pasirinkę parametrą *partial*, gauname dalinį autokoreliacijos grafiką.



3 pav.: Dalinė autokoreliacinė funkcija tranformuotiems Europos Sąjungos BVP duomenims (PACF)

Pažvelgus į šį grafiką, įdidžiausios reikšmės yra ties 1 ir 2 reikšme, tačiau jos yra gana mažos ir panašios į kitas reikšmes, todėl iš šio grafiko galime nuspėti jog AR parametro reikšmė yra lygi 0.

Išsiaiškinti koks galėtų būti MA parametras, brėžiame autokoreliacijos grafiką.



4 pav.: Autokoreliacinė funkcija tranformuotiems Europos Sąjungos BVP duomenims (ACF)

Pažvelgus į šitą grafiką matome, kad pirmoji reikšmė yra didesnė, nei kitos reikšmės. Kadangi matomasis didelis skirtumas tarp pirmos ir likusių reikšmių, galima daryti prielaidą jog MA parametro reikšmė galėtų būti lygi 1.

Iš šių dviejų grafikų reikšmių, galime daryti išvadą, jog geriausias modelis šiai laiko eilutei yra MA(1).

Naudojantis programa R, taip pat galima panaudoti funkcija auto.arima, kuri naudojantis vienetinių šaknų testo algoritmais ir ankščiau įvardytais kriterijais, gali pasiūlyti geriausią ARIMA modelio variantą.

Modelis	Standartinis nuokrypis	AIC	AICc	BIC
ARIMA (0,0,1)	0,1001	-540,55	-540,31	-532,65

3 lentelė: ARIMA modelio kriterijai

Ši funkcija taip pat siūlo naudoti ARIMA (0,0,1) ar kitaip MA(1) modelį. Visų kriterijų reikšmės gana žemos, todėl galima manyti, jog modelis yra tinkamas šiems duomenims.

MA (1) modelis :

$$X_t = w_t + \theta_1 w_{t-1}, \quad -\infty < \theta < \infty.$$

7.3 Modelio pritaikymas

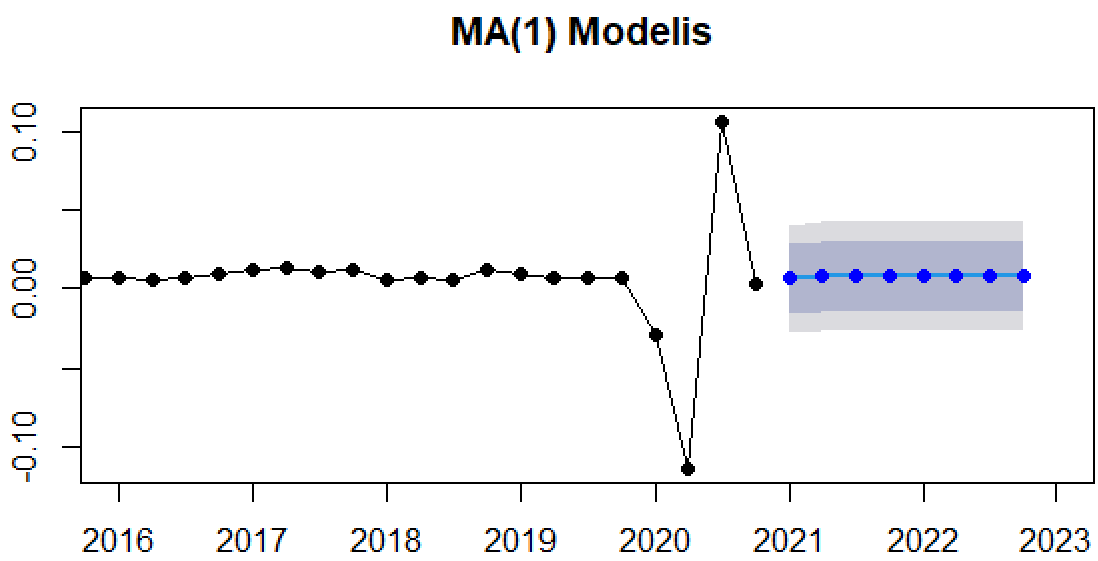
Duomenis prognizuosime 2 metus arba 8 ketvirčius į ateitį. Pasirinktas neilgas laitarpis, nes duomenų imtis nebuvo didelė, todėl vengiant didelių netikslumų, prognozuojame tik 8 laiko intervalus į priekį.

Dabar pasirinkus šį prognozavimo metodą, pasinaudojus funkcijomis *arima* ir *forecast*, galime prognozuoti laiko eilutės reikšmes į priekį. ARIMA modelio MA(1) reikšmės :

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021 Q1	0.006325478	-0.01572609	0.02837704	-0.02739948	0.04005044
2021 Q2	0.007617646	-0.01473894	0.02997423	-0.02657381	0.04180910
2021 Q3	0.007617646	-0.01473894	0.02997423	-0.02657381	0.04180910
2021 Q4	0.007617646	-0.01473894	0.02997423	-0.02657381	0.04180910
2022 Q1	0.007617646	-0.01473894	0.02997423	-0.02657381	0.04180910
2022 Q2	0.007617646	-0.01473894	0.02997423	-0.02657381	0.04180910
2022 Q3	0.007617646	-0.01473894	0.02997423	-0.02657381	0.04180910
2022 Q4	0.007617646	-0.01473894	0.02997423	-0.02657381	0.04180910

4 lentelė: MA (1) prognozių reikšmės

Iš gautų duomenų matome, jog nuo antrojo ketvirčio iki pat aštuntojo, reikšmės išlieka vienodos. Brėžiame transformuotų duomenų grafiką kartu su naujomis, prognozių reikšmėmis.



5 pav.: MA(1) modelio prognozės Europos Sąjungos BVP 2021 -2022 metams (transformuoti duomenys)

Iš šio grafiko galime matyti, jog 2020 metų duomenys, reikšmingai nepakeitė prognozės, reikšmės liko artimos nuliui ir praėjusių metų duomenys, reikšmingos įtakos prognozei nepadarė. Pasikliautinojo intervalo reikšmės, taip pat didelių nukrypimų nenurodo.

8 Išvados

Atlikus Europos Sąjungos bendrojo vidaus produkto rodiklio analizę, gauta jog duomenys nėra stacionarūs ir jiems pritaikyti modelį būtų gana netikslu. Šiuos duomenis transformavus, reikšmės išbarstė aplink nulį ir atlikus stacionarumo testą, gauta jog laiko eilutės reikšmės yra stacionarios. Nubraižius autokoreliacijos ir dalinės autokoreliacijos grafikus, rasta jog geriausias modelis duomenų prognozavimui yra MA(1). Taip pat atlikus Dickey – Fuller testą buvo pasiūlytas MA(1) modelis. Naudojantis programa R, gautos prognozės reikšmės 2021 - 2022 metams. Šios reikšmės taip pat liko išsibarsčiusios aplink nulį. Todėl yra matoma, jog naudojantis MA(1) modeliu, 2020 metų BVP kritimas dėl Covid-19 pandemijos, prognozėms reikšmingos įtakos nepadarė.

9 Literatūra

[1] M. R. Abonazel, A. I. Abd-Elftah. *Forecasting Egyptian GDP Using ARIMA Models*. Cairo University, 2019.

[2] P. Bondarenko. *Gross domestic product*. Encyclopædia Britannica, 2016.

[3] R.J. Brockwell, R.A. Davis. *Time Series: Theory and Methods.* Springer, 2006.

[4] A. Buteikis. *Time Series with Unit Root*.
http://web.vu.lt/mif/a.buteikis/wp-content/uploads/2019/03/Lecture_04.pdf

[5] V. Čekanavičius, G. Murauskas. *Statistika ir jos taikymai*. Vilnius: TEV, 2001.

[6] E. Gutauskaitė. *Laiko eilučių paskaitų konspektas*. Vilniaus universitetas, 2020.

[7] H. M. Hassan Ali, A. M. Ali Haleeb. *Modelling GDP for Sudan using ARIMA*. University of Khartoum, 2020.

[8] Dr. Imtinungsang Jamir. *Forecasting Potential Impact of COVID-19 Outbreak on India's GDP using ARIMA model*. The ICFAI University, 2020.

[9] V. Karpuškienė, A. Davidovič, O. Davidovič, K. Majevskaja, J. Mečkovski, S. Meškelytė, L. Mociūnaitė, G. Rupeika, N. Šikšniūtė. *Ekonometrijos Virtuvė*. Vilniaus universitetas, 2017.

[10] E. Paparoditis, D. N. Politis. *The asymptotic size and power of the augmented Dickey–Fuller test for a unit root*. Econometric Reviews, 2016.

[11] R.H. Shumway, D.S. Stoffer. *Time Series Analysis and Its Applications: with R examples*. Springer, 2017.

[12] L. Stabingienė. *Ekonometrika, Elektroninė mokomoji knyga*. Klaipėdos universitetas, 2014.
http://www.ilab.lt/stabingiene/sk8 _2.html

[13] World Health Organization. *Rolling updates on coronavirus desease (COVID-19)*.
https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen

A Priedai

1. Europos Sąjungos bendrasis vidaus produktas, 1995-2020 metai

Data	BVP, mln EUR	Data	BVP, mln EUR
1995-01-01	1551975,4	2008-01-01	2770308,3
1995-04-01	1570377,2	2008-04-01	2787739,2
1995-07-01	1598012,4	2008-07-01	2790707,9
1995-10-01	1617255,2	2008-10-01	2728569,5
1996-01-01	1634056,2	2009-01-01	2630375,3
1996-04-01	1655365,0	2009-04-01	2629771,3
1996-07-01	1672648,4	2009-07-01	2650798,4
1996-10-01	1676121,4	2009-10-01	2671489,3
1997-01-01	1676538,3	2010-01-01	2695731,3
1997-04-01	1693651,5	2010-04-01	2728478,3
1997-07-01	1713218,4	2010-07-01	2757472,5
1997-10-01	1739687,3	2010-10-01	2784941,9
1998-01-01	1749938,3	2011-01-01	2821578,5
1998-04-01	1769839,8	2011-04-01	2830061,2
1998-07-01	1786386,5	2011-07-01	2837408,1
1998-10-01	1802322,3	2011-10-01	2827127,2
1999-01-01	1820945,5	2012-01-01	2840018,1
1999-04-01	1838126,9	2012-04-01	2840944,3
1999-07-01	1867022,6	2012-07-01	2857190,4
1999-10-01	1896156,7	2012-10-01	2851860,0
2000-01-01	1929429,4	2013-01-01	2856480,7
2000-04-01	1957793,9	2013-04-01	2876700,1
2000-07-01	1980397,8	2013-07-01	2888962,0
2000-10-01	2003205,2	2013-10-01	2900197,9
2001-01-01	2034374,0	2014-01-01	2924099,6
2001-04-01	2055705,5	2014-04-01	2935979,1
2001-07-01	2066863,5	2014-07-01	2954380,0
2001-10-01	2087657,4	2014-10-01	2974780,5
2002-01-01	2109940,2	2015-01-01	3013453,1
2002-04-01	2126536,9	2015-04-01	3042021,8
2002-07-01	2145591,9	2015-07-01	3065077,1
2002-10-01	2159785,7	2015-10-01	3086174,4
2003-01-01	2164755,7	2016-01-01	3106001,1
2003-04-01	2178263,7	2016-04-01	3122218,9
2003-07-01	2205163,1	2016-07-01	3143000,7

2003-10-01	2225220,9	2016-10-01	3171504,7
2004-01-01	2248712,4	2017-01-01	3210541,2
2004-04-01	2279327,4	2017-04-01	3251980,1
2004-07-01	2298729,6	2017-07-01	3287969,8
2004-10-01	2326769,1	2017-10-01	3326938,3
2005-01-01	2346870,3	2018-01-01	3346106,4
2005-04-01	2370571,1	2018-04-01	3368309,1
2005-07-01	2401646,5	2018-07-01	3386871,4
2005-10-01	2434858,7	2018-10-01	3425651,8
2006-01-01	2469570,7	2019-01-01	3459071,6
2006-04-01	2512374,9	2019-04-01	3483170,7
2006-07-01	2542654,2	2019-07-01	3505729,0
2006-10-01	2590574,3	2019-10-01	3528283,3
2007-01-01	2634253,0	2020-01-01	3426575,7
2007-04-01	2668263,4	2020-04-01	3059145,2
2007-07-01	2697848,0	2020-07-01	3400318,0
2007-10-01	2741882,2	2020-10-01	3496240,9

2. R kodas

```
pradiniai <- read.csv("C:~.csv", header = TRUE, stringsAsFactors = FALSE)
#Duomenų užkėlimas
new.date1 = as.yearqtr(pradiniai$TIME_PERIOD, "%Y-Q%d") #Datos formatas
eilute = ts(data = pradiniai$OBS_VALUE, frequency = 4, start=c(1995),end = c(2020,4))
#Laiko eilute
plot.ts(eilute, main = "Europos Sąjungos BVP, 1995-2020 ketvirčio duomenys",
xlab = "Metai", ylab = "BVP, mln EUR") #Pradinių duomenų grafikas
adf.test(eilute) #Stacionarumo testas
dlrGDP <- diff(log(eilute)) #Transformavimas
plot.ts(dlrGDP, main = "Transformuotų duomenų grafikas", xlab = "Metai",
ylab = "Y reikšmės") #Transformuotų duomenų laiko eilutė
adf.test(dlrGDP) #Stacionarumo testas
acf(dlrGDP, type = "correlation", lag = 34, main = "ACF grafikas")
#Autokoreliacijos grafikas
acf(dlrGDP, type = "partial", lag = 34, main = "Dalinis ACF grafikas")
#Dalinės autokoreliacijos grafikas
a2 <- auto.arima(dlrGDP) #Modelio parinkimas
a2
ma1 <- Arima(dlrGDP, order=c(0, 0, 1)) #Ma(1) modelis ir prognozės
ma1.f <- forecast(ma1, 8)
plot(ma1.f, type="o", pch=16, xlim=c(2016, 2022), main="MA(1) Modelis ")
lines(ma1.f$mean, type="p", pch=16, lty="dashed", col="blue") #Prognozių grafikas
```