



Premesse

Progetto

Real Gross Domestic Product

GDP in funzione di altre misure macroeconomiche

Premesse

Appendice

Il movente

Lo studio nasce dalla curiosità di una studentessa di toccare con mano alcune delle misure macroeconomiche italiane più importanti, spesso menzionate da economisti e media.

I dati

Tutti i dati analizzati nel seguente studio sono stati scaricati dal portale FRED (<https://fred.stlouisfed.org>), "Federal Reserve Economic Data", un vasto archivio di dati gestito dalla Federal Reserve Bank of St. Louis. I dati scelti sono tutti rappresentati da serie storiche, tutte a frequenza trimestrale e non stagionalmente adattate.

La finestra temporale prescelta è quella che va dal primo trimestre del 1996 al quarto trimestre del 2022 e i dati sono tutti riferiti al contesto economico italiano.

Gli strumenti

Gli strumenti statistici utilizzati sono quelli appresi durante il corso di *Laboratorio di Statistica*, incentrato sull'analisi delle serie storiche e sull'interpretazione dei relativi risultati.

I packages R necessari per l'analisi che si intende condurre sono i seguenti.

```
library(forecast)
library(quantmod)
library(GGally)
library(tidyverse)
library(dygraphs)
```

Real Gross Domestic Product

Il Gross Domestic Product (GDP), che in italiano traduciamo con Prodotto Interno Lordo (PIL), è una misura del valore di tutti i beni e servizi finali prodotti nel Paese in un determinato periodo di tempo. La misura *Real GDP* tiene conto dell'effetto dell'inflazione.

Il Real GDP riesce a fornire una misura accurata della crescita economica dell'Italia e può essere utilizzato per confrontare l'andamento dell'economia italiana con quello di altri Paesi o con il passato.

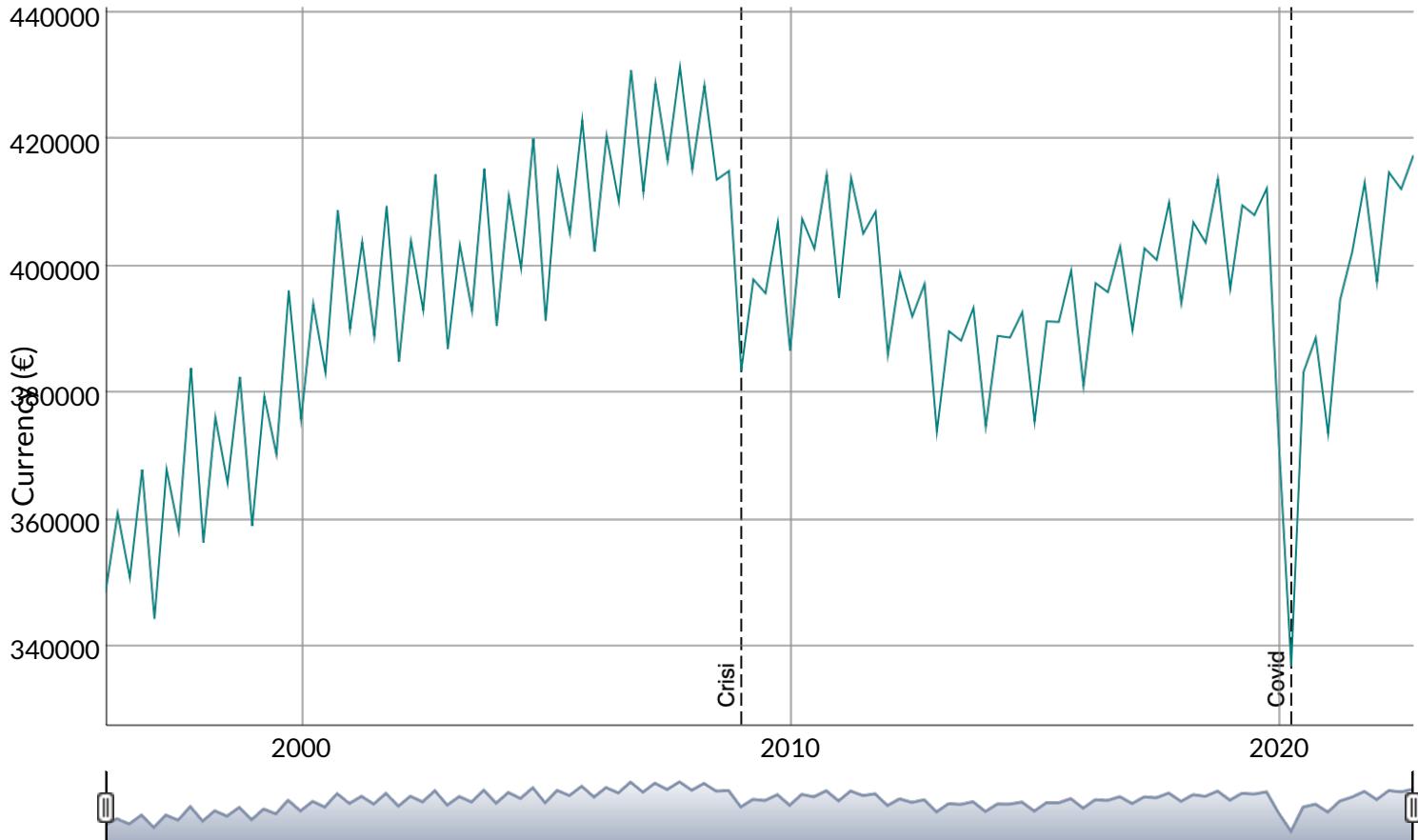
Nota: Ci si riferirà a questa serie storica sempre con l'acronimo GDP.

```
gdp <- ts(getSymbols("NGDPRNSAXDCITQ", src="FRED", auto.assign = FALSE), start=1996, frequency=4)
```

Fonte: International Monetary Fund, Real Gross Domestic Product for Italy [NGDPRNSAXDCITQ], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/NGDPRNSAXDCITQ> (<https://fred.stlouisfed.org/series/NGDPRNSAXDCITQ>), May 11, 2023.

```
dyRangeSelector(dygraph(gdp,
                         main="Real Gross Domestic Product",
                         ylab="Currency (€)") %>%
  dyEvent("2020-04-01", "Covid", labelLoc = "bottom") %>%
  dyEvent("2009-01-01", "Crisi", labelLoc = "bottom")
)
```

Real Gross Domestic Product



La prima informazione che è possibile ricavare dai dati e dalla loro visualizzazione riguarda l'andamento medio di lungo periodo della serie: il GDP Italiano ha avuto un andamento medio crescente fino alla metà del 2008, per poi mostrare un rapido calo fino agli inizi del 2009. Questa fase coincide con la crisi finanziaria (https://it.wikipedia.org/wiki/Crisi_finanziaria_del_2007-2008) del 2007-2008. Dopo una fase di ripresa, durata fino alla metà del 2011, il trend è tornato ad essere decrescente fino agli inizi del 2014, per poi riprendere in salita fino alla fine del 2019. Nel 2020 l'emergenza Covid-19 (<https://it.wikipedia.org/wiki/COVID-19>) ha avuto effetti drammatici sull'economia di ogni Paese e il GDP italiano ha toccato rapidamente, nella metà del 2020, i minimi storici. Tuttavia, la ripresa è stata rapidissima, mostrando un trend crescente ad alta pendenza.

Pulizia dei dati

Valori mancanti (missing values) o anomali (outliers) potrebbero avere effetti sui modelli che si andranno a costruire e sui relativi risultati. Risulta pertanto indispensabile verificare l'eventuale presenza di tali valori.

Missing values

```
anyNA(gdp)
```

```
## [1] FALSE
```

Outliers

```
tsoutliers(gdp)
```

```
## $index
## [1] 95 96 98
##
## $replacements
## [1] 399055.6 397351.1 386198.6
```

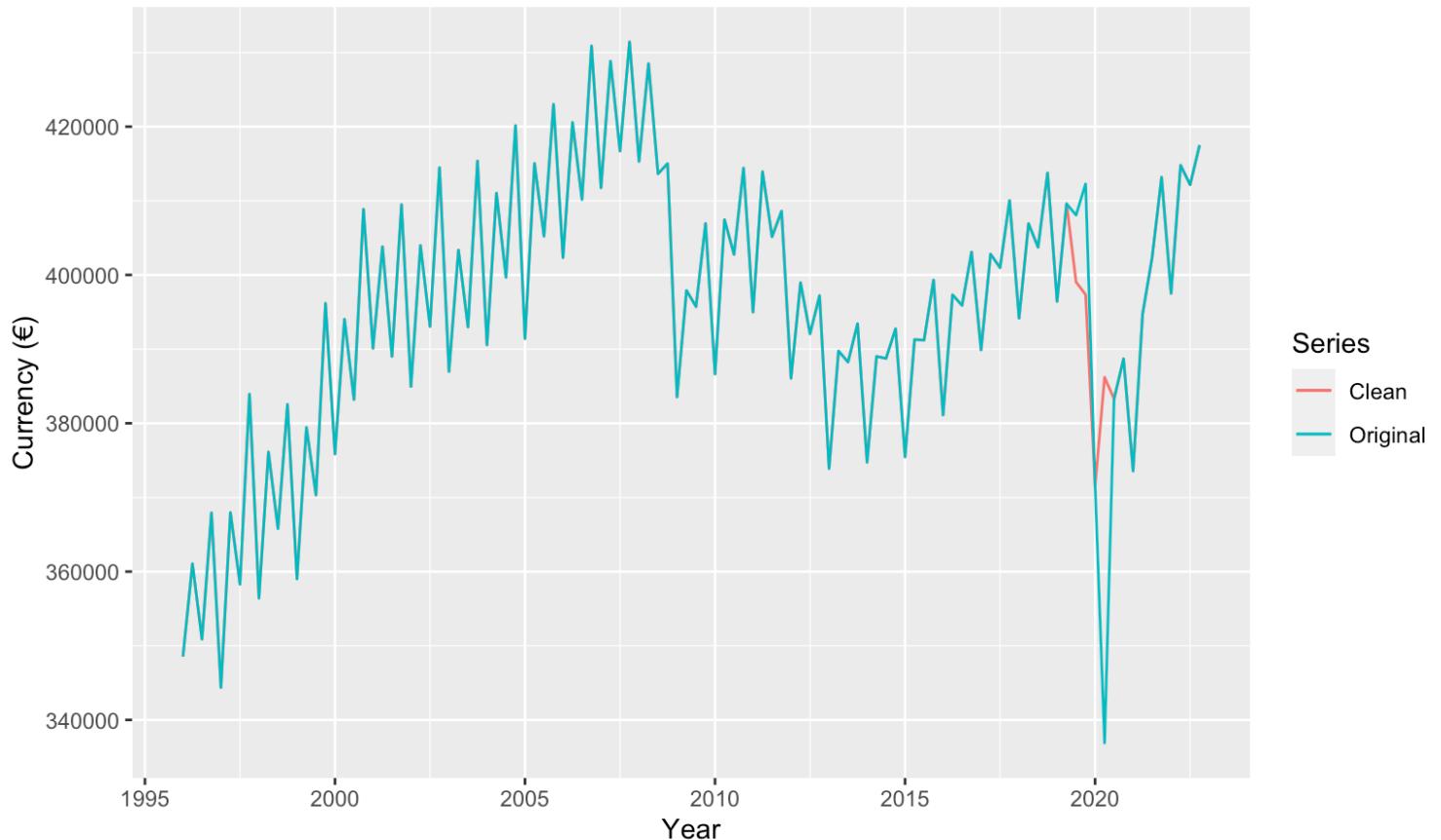
La serie storica oggetto di studio non presenta missing values, pertanto non sarà necessario gestirli con metodi di imputazione; tuttavia, presenta outliers. I valori anomali presenti nella serie sono frutto degli effetti economici dell'emergenza Covid-19 (<https://it.wikipedia.org/wiki/COVID-19>).

La funzione `tsclean()` utilizza la decomposizione robusta STL per la stima dei missing values e la sostituzione degli outliers. Possiamo utilizzarla per creare una serie storica in cui viene applicata la sostituzione degli outliers, in modo da poter confrontare i risultati dei metodi che verranno applicati sulle due serie.

```
gdp_clean <- tsapply(gdp)
```

```
autoplot(gdp_clean, series="Clean") +
  autolayer(gdp, series="Original") +
  ggtitle("Original series VS Clean series") +
  xlab("Year") +
  ylab("Currency (€)") +
  guides(colour=guide_legend(title="Series"))
```

Original series VS Clean series

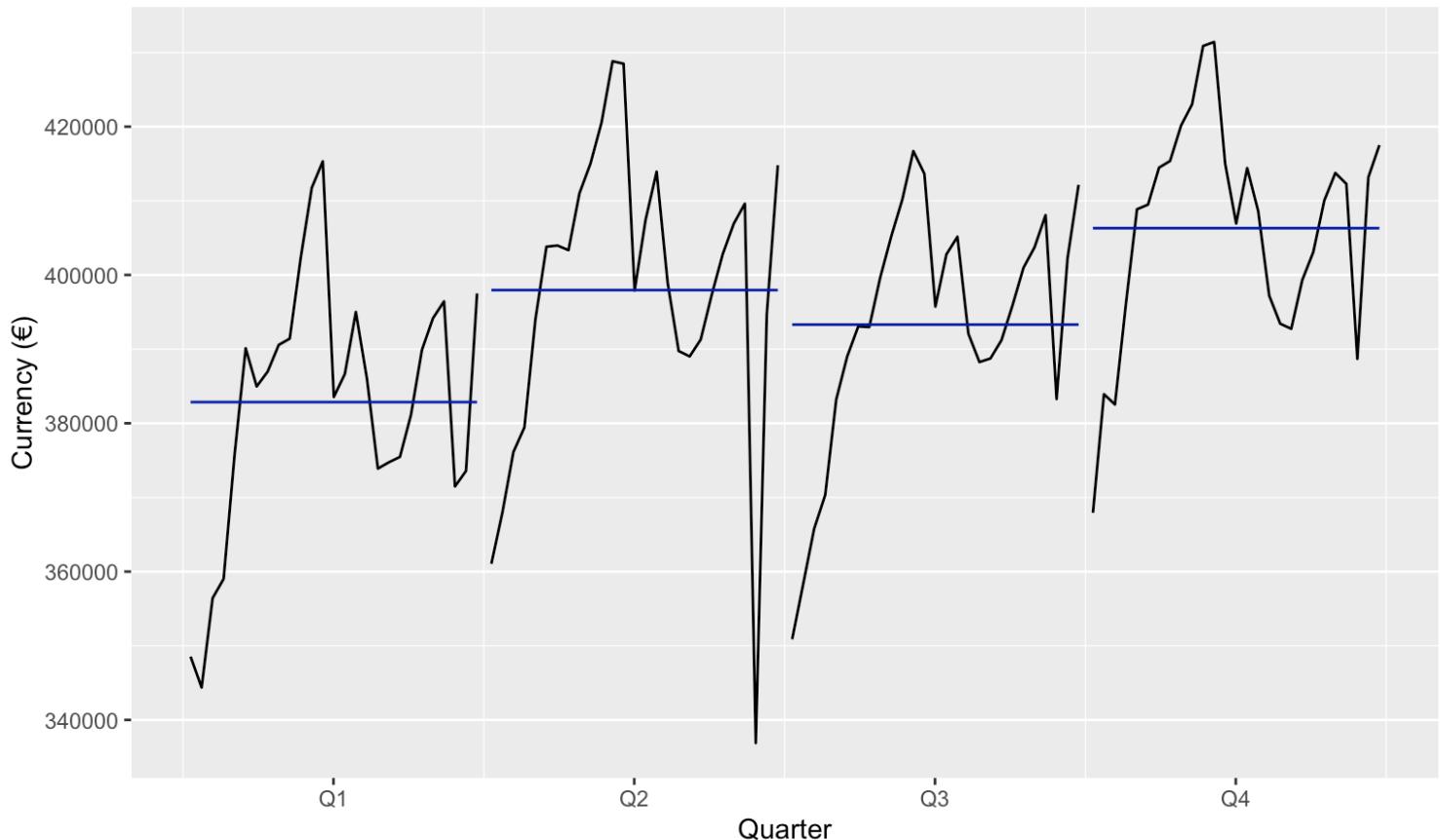


Stagionalità

Si prosegue con la ricerca di un'eventuale stagionalità all'interno della serie storica del GDP.

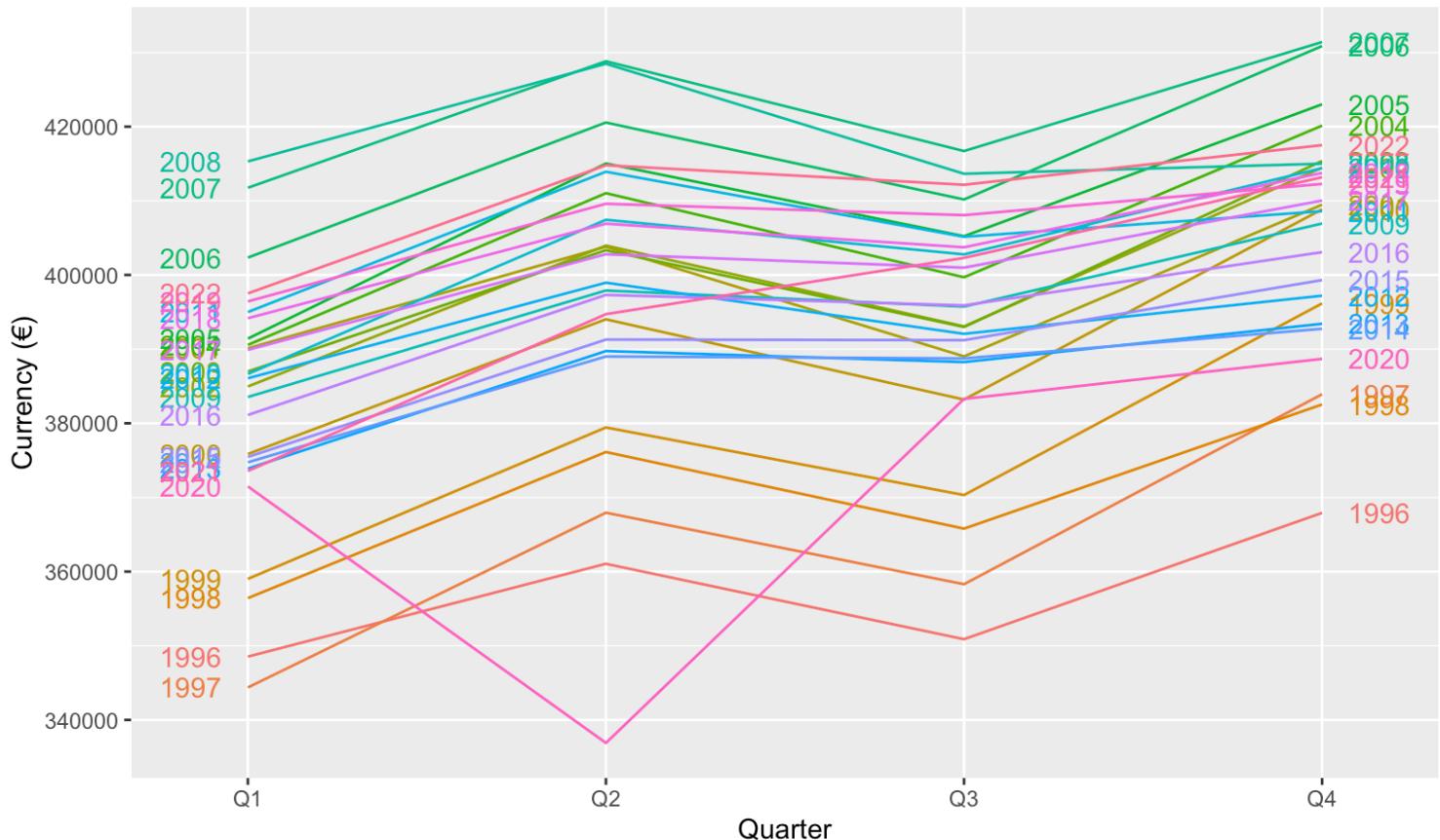
```
ggsubseriesplot(gdp) +  
  ylab("Currency (€)") +  
  ggtitle("Seasonal subseries plot: Real Gross Domestic Product")
```

Seasonal subseries plot: Real Gross Domestic Product



```
ggseasonplot(gdp, year.labels = TRUE, year.labels.left = TRUE) +
  ylab("Currency (€)") +
  xlab("Quarter")
```

Seasonal plot: gdp



```
ggtitle("Seasonal plot: Real Gross Domestic")
```

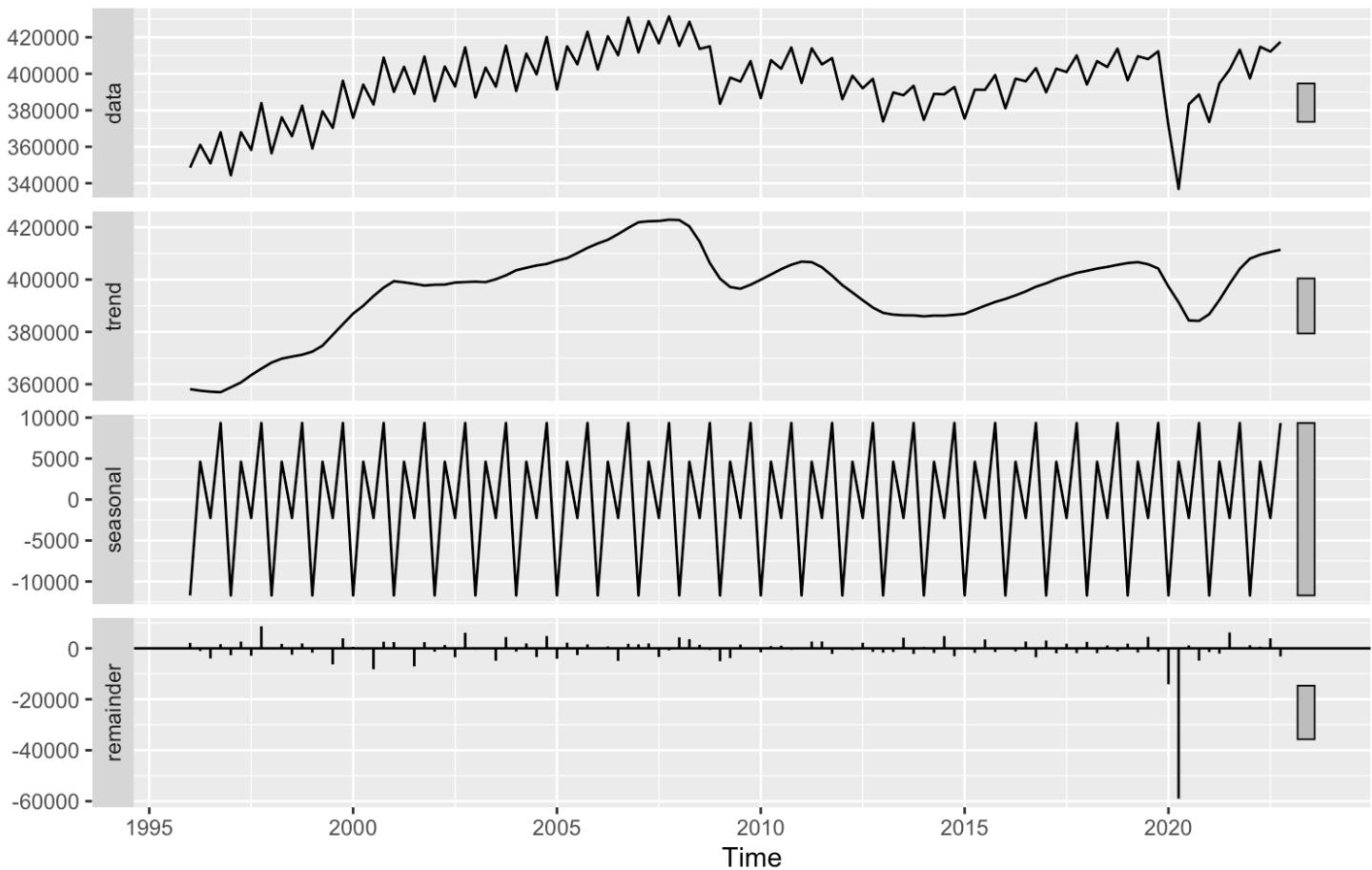
```
## $title
## [1] "Seasonal plot: Real Gross Domestic"
##
## attr(,"class")
## [1] "labels"
```

Come si poteva prevedere, il GDP può essere influenzato dalla stagionalità in quanto le attività economiche possono variare in modo regolare e prevedibile a seconda delle stagioni: il GDP mostra un andamento crescente il primo e il terzo trimestre, al contrario di ciò che accade il secondo e il quarto. Tuttavia, sembra che negli ultimi anni (orientativamente dal 2014) questo andamento sia stato meno netto e gli anni 2020 e 2021 hanno avuto andamenti anomali per le stesse considerazioni sopracitate legate all'emergenza Covid-19.

Decomposizione

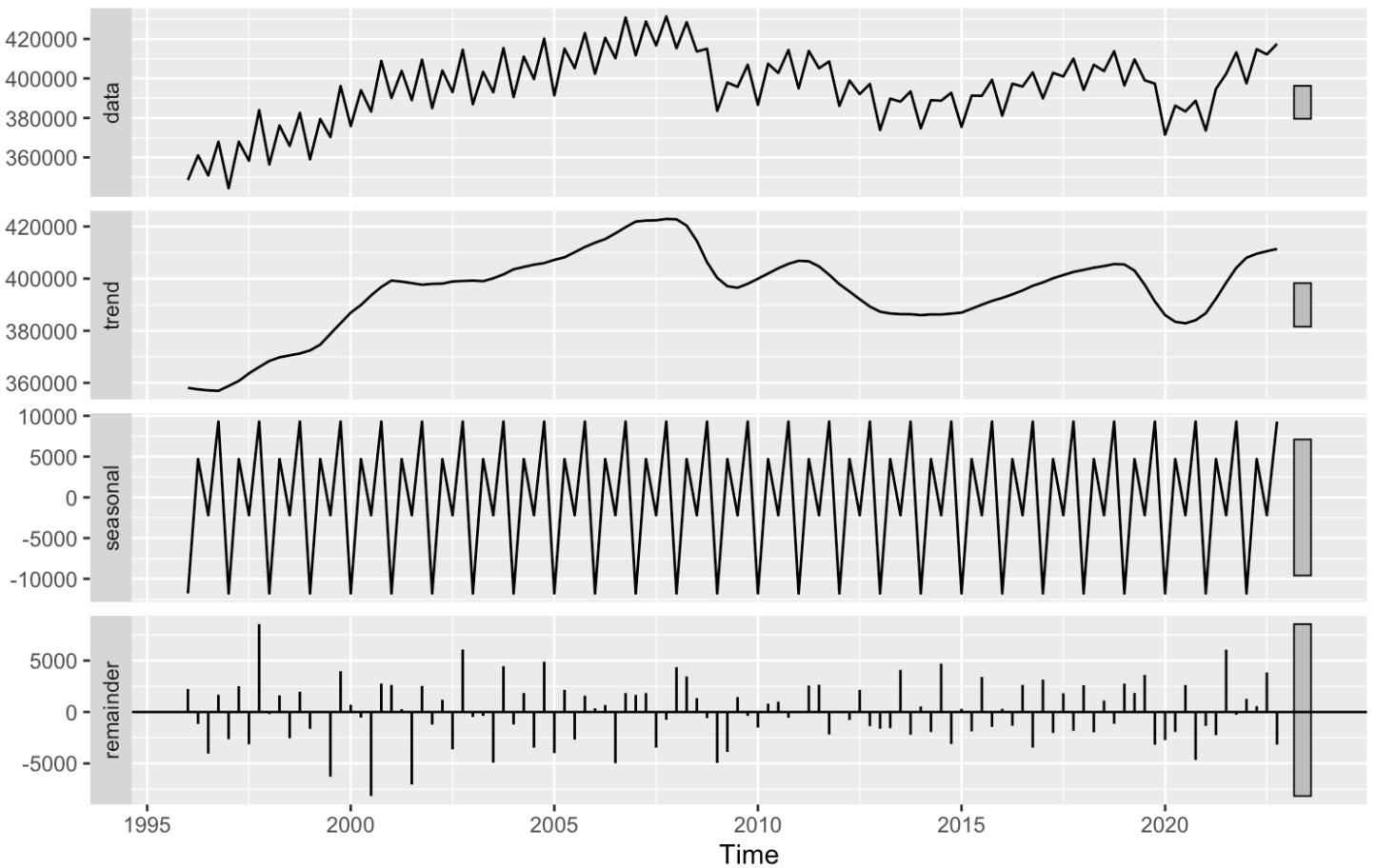
È possibile decomporre la serie storica GDP isolando le diverse sue componenti: trend-ciclo, stagionalità e residuale. Si procede con la decomposizione STL (Seasonal and Trend decomposition using Loess), in quanto è la più robusta ai valori anomali.

```
stl_fit <- stl(ts(as.vector(gdp), start=1996, frequency = 4), s.window = "periodic", robust=TRUE)
autoplot(stl_fit)
```



La decomposizione applicata ai dati originali appare fortemente influenzata dai valori anomali, nonostante la decomposizione STL sia la più robusta. Pertanto, si applica la decomposizione alla serie `gdp_clean`.

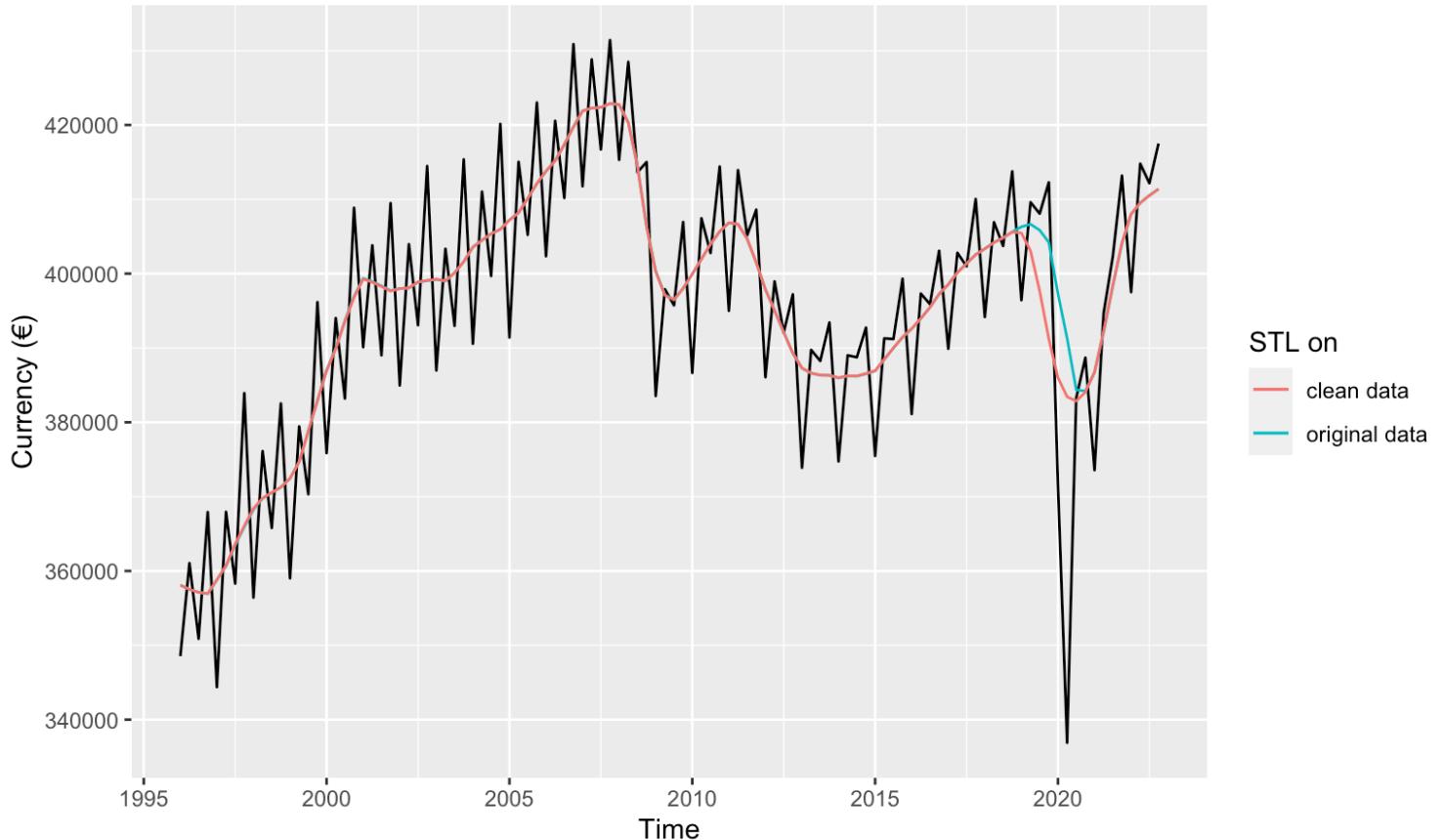
```
stl_fit2 <- stl(ts(as.vector(gdp_clean), start=1996, frequency = 4), s.window = "periodic", robust)
autoplot(stl_fit2)
```



La componente principale risulta essere quella di trend.

```
autoplot(gdp) +
  autolayer(stl_fit$time.series[,2], series="original data") +
  autolayer(stl_fit2$time.series[,2], series="clean data") +
  ylab("Currency (€)") +
  ggtitle("STL decomposition") +
  guides(colour=guide_legend(title="STL on"))
```

STL decomposition



Possiamo, ora che sono state separate le componenti di trend e stagionalità, calcolarne la forza.

```
var_R <- var(stl_fit2$time.series[,3])
var_RT <- var(stl_fit2$time.series[,3] + stl_fit2$time.series[,2])
var_RS <- var(stl_fit2$time.series[,3] + stl_fit2$time.series[,1])
```

Forza del trend:

$$F_t = \max(0; 1 - \frac{Var(R_t)}{Var(R_t + T_t)})$$

```
max(0,1-(var_R/var_RT))
```

```
## [1] 0.9661618
```

Forza della stagionalità:

$$F_s = \max(0; 1 - \frac{Var(R_t)}{Var(R_t + S_t)})$$

```
max(0,1-(var_R/var_RS))
```

```
## [1] 0.8836213
```

Entrambi gli indici sono compresi tra 0 e 1. Essendo entrambi prossimi a 1, $F_t = 0.97$ e $F_s = 0.88$, entrambe le componenti sono forti, ma il trend risulta essere la componente principale.

Autocorrelazione

Per misurare la forza del legame lineare tra i valori della serie storica su due istanti di tempo diversi, t e $t-h$, dove h è detto ritardo o lag, si procede con lo studio dell'autocorrelazione della serie del GDP italiano.

L'autocorrelazione è data da

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

dove

$$\gamma(h) = \frac{1}{T} \sum_{t=1}^{T-h} (x_t - \bar{x})(x_{t+h} - \bar{x})$$

e $\gamma(0)$ si riduce a

$$\gamma(0) = \frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})^2$$

ovvero alla nozione di varianza.

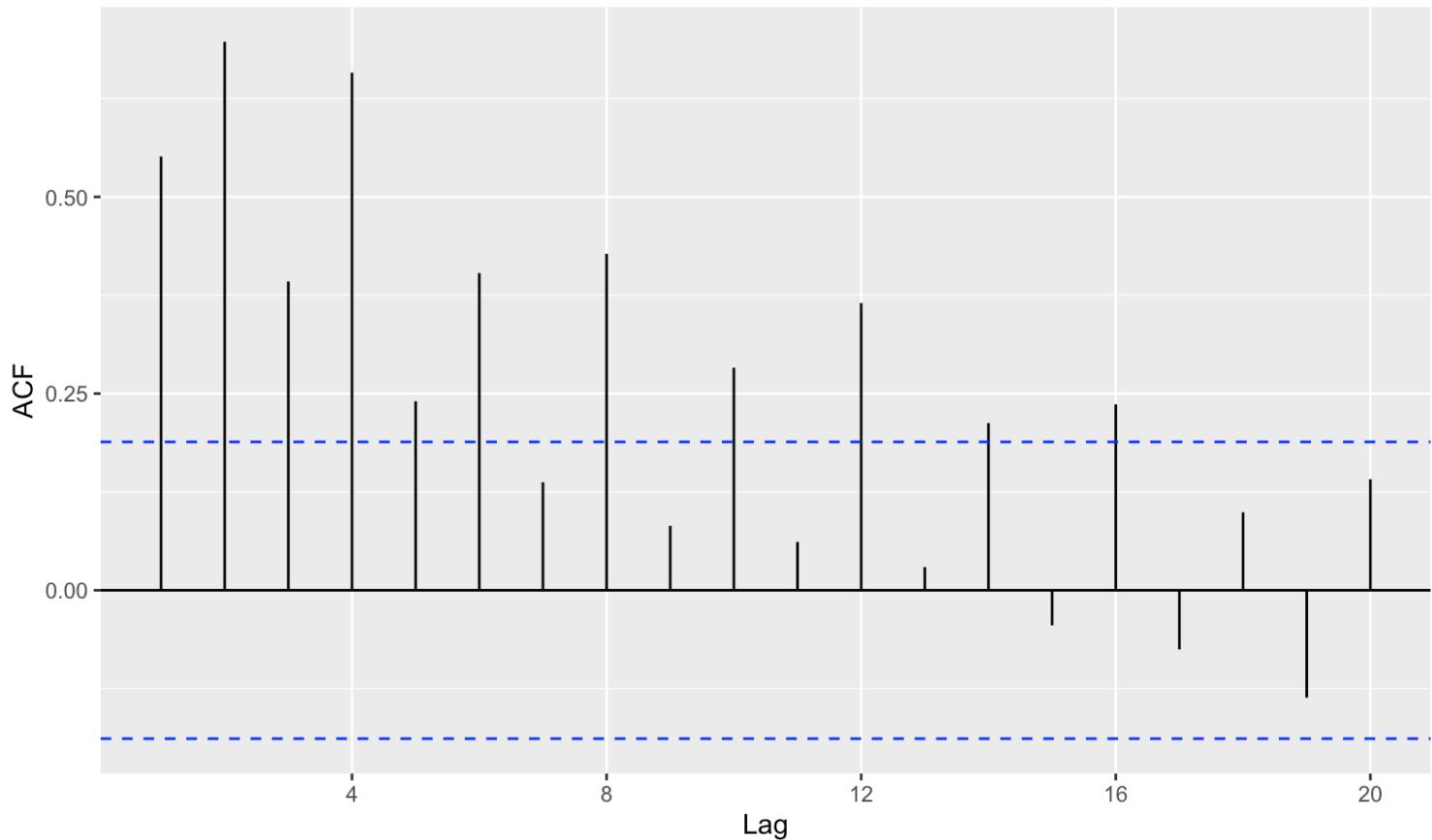
L'insieme di tutti i valori di $\rho(h)$ è detto funzione di autocorrelazione. Si tratta di una funzione simmetrica intorno allo zero.

```
ggAcf(gdp, plot=FALSE)
```

```
##  
## Autocorrelations of series 'gdp', by lag  
##  
##      0      1      2      3      4      5      6      7      8      9      10  
##  1.000  0.551  0.697  0.393  0.658  0.240  0.403  0.137  0.428  0.082  0.283  
##      11     12     13     14     15     16     17     18     19     20  
##  0.061  0.365  0.030  0.213 -0.045  0.236 -0.075  0.099 -0.137  0.141
```

```
ggAcf(gdp)
```

Series: gdp



Dai valori delle autocorrelazioni, calcolati per diversi ritardi, risulta evidente che le autocorrelazioni di ritardo $h = 2$ sono quelle dal valore maggiore, coerentemente con le considerazioni fatte precedentemente circa la stagionalità della serie. Ciò si riflette anche nel grafico che mostra regolarità nel comportamento delle autocorrelazioni.

I valori delle autocorrelazioni permettono di rilevare la presenza di trend, in quanto inizialmente sono tutte positive e decadono a zero lentamente, e del relativo cambio di pendenza, in quanto c'è un cambio di direzione per i valori delle autocorrelazioni.

Nonostante i valori delle autocorrelazioni nella prima parte del grafico ricadano al di fuori dell'intervallo di confidenza (fissato al 95%), mostrando delle autocorrelazioni statisticamente significative, queste diventano statisticamente nulle nella parte finale, per valori elevati di h .

Previsioni

I dati di una serie storica, in questo caso del GDP italiano, possono essere utilizzati per fare previsione sul valore futuro che questa variabile assumerà. Chiaramente l'intervallo di confidenza diventa via via più ampio all'allontanarsi dall'ultima osservazione a disposizione, rendendo le previsioni a lungo termine più incerte.

La presenza di valori anomali influisce notevolmente sulle capacità predittive dei modelli applicati. I modelli verranno addestrati sulle osservazioni dal 1996 al 2018 e testati sulle osservazioni dal 2019 al 2022 della serie storica trasformata `gdp_clean` per ridurre l'effetto dell'emergenza Covid sui dati.

Modelli benchmark

Di seguito verranno implementati i più semplici metodi di previsione, seppur con poche pretese conoscendo le caratteristiche della serie:

- Average method

$$y_{T+h|T} = \frac{\sum_{t=1}^T y_t}{T}$$

- Naïve method

$$y_{T+h|T} = y_T$$

- Seasonal naïve method

$$y_{T+h|T} = y_{T+h-m(k+1)}$$

- Drift method

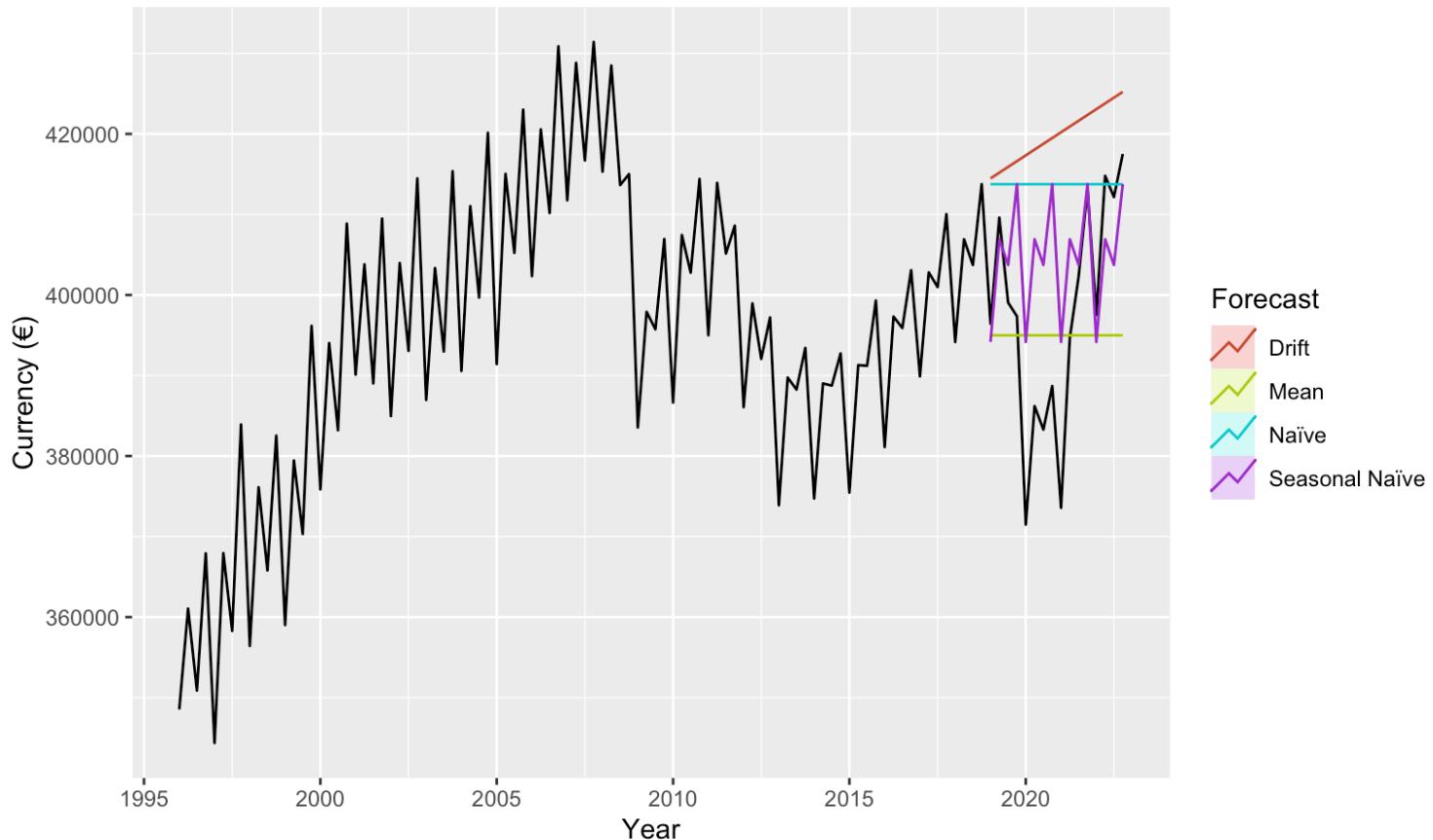
$$y_{T+h|T} = y_T + h \frac{y_T - y_1}{T - 1}$$

```
train <- ts(gdp_clean[1:92], start=1996, frequency=4)
test <- window(gdp_clean, start=2019, end=c(2022,4))
```

```
mean_fit <- meanf(train, 16)
naive_fit <- naive(train, 16)
snaive_fit <- snaive(train, h=16)
drift_fit <- rwf(train, 16, drift=TRUE)
```

```
autoplot(gdp_clean) +
  autolayer(mean_fit, series="Mean", PI=FALSE) +
  autolayer(naive_fit, series="Naïve", PI=FALSE) +
  autolayer(snaive_fit, series="Seasonal Naïve", PI=FALSE) +
  autolayer(drift_fit, series="Drift", PI=FALSE) +
  xlab("Year") +
  ylab("Currency (€)") +
  ggtitle("Real Gross Domestic Product") +
  guides(colour=guide_legend(title="Forecast"))
```

Real Gross Domestic Product



```
rbind(
  "Average" = accuracy(mean_fit, test)[2,c(2,3,5,6)],
  "Naïve" = accuracy(naive_fit, test)[2,c(2,3,5,6)],
  "Seasonal Naïve" = accuracy(snaive_fit, test)[2,c(2,3,5,6)],
  "Drift" = accuracy(drift_fit, test)[2,c(2,3,5,6)])
```

	RMSE	MAE	MAPE	MASE
## Average	13891.99	11374.26	2.863320	1.672807
## Naïve	21352.64	16987.33	4.393137	2.498318
## Seasonal Naïve	13754.56	10821.43	2.785163	1.591503
## Drift	25741.42	22480.21	5.771592	3.306153

Chiaramente nessuno di questi metodi di previsione risulta efficace, ma tra tutti il seasonal naïve method è l'unico che riesce a cogliere la struttura stagionale della serie, e mostra il più basso tasso di errore. Il seasonal naïve method è incapace di cogliere la componente di trend presente nei dati, componente colta dal drift method. Il metodo che garantisce un'accuratezza previsiva prossima a quella garantita dal seasonal naïve method, è l'average method, pur essendo il modello più semplice.

No Covid Data

Nonostante i modelli siano stati testati sulla serie `gdp_clean`, nessun metodo di previsione addestrato sulla serie fino al 2019 sarebbe stato in grado di prevedere quanto accaduto nel 2020 al GDP italiano. Può essere utile valutare l'accuratezza di questi metodi addestrandoli sulle osservazioni dal 1996 al 2016 e testandoli sulle osservazioni dal 2017 al 2019 della serie storica originale `gdp`.

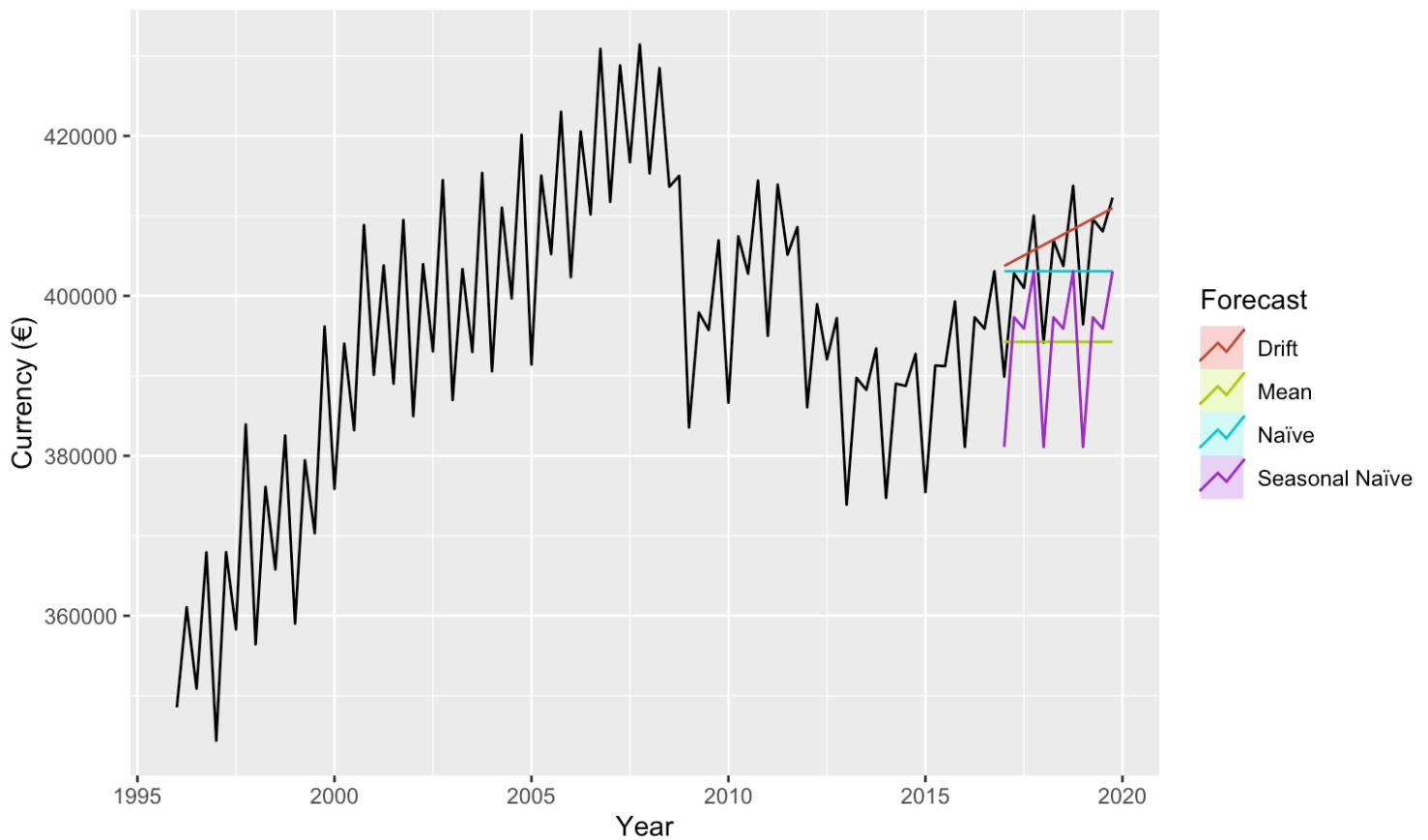
```
gdp_nc <- ts(gdp[1:96], start=1996, frequency=4)
```

```
train_nc <- ts(gdp_nc[1:84], start=1996, frequency=4)
test_nc <- window(gdp_nc, start=2017)
```

```
mean_fit_nc <- meanf(train_nc, 12)
naive_fit_nc <- naive(train_nc, 12)
snaive_fit_nc <- snaive(train_nc, h=12)
drift_fit_nc <- r wf(train_nc, 12, drift=TRUE)
```

```
autoplot(gdp_nc) +
  autolayer(mean_fit_nc, series="Mean", PI=FALSE) +
  autolayer(naive_fit_nc, series="Naïve", PI=FALSE) +
  autolayer(snaive_fit_nc, series="Seasonal Naïve", PI=FALSE) +
  autolayer(drift_fit_nc, series="Drift", PI=FALSE) +
  xlab("Year") +
  ylab("Currency (€)") +
  ggtitle("Real Gross Domestic Product") +
  guides(colour=guide_legend(title="Forecast"))
```

Real Gross Domestic Product



```
rbind(  
  "Average" = accuracy(mean_fit_nc, test_nc)[2,c(2,3,5,6)],  
  "Naïve" = accuracy(naive_fit_nc, test_nc)[2,c(2,3,5,6)],  
  "Seasonal Naïve" = accuracy(snaive_fit_nc, test_nc)[2,c(2,3,5,6)],  
  "Drift" = accuracy(drift_fit_nc, test_nc)[2,c(2,3,5,6)])
```

```
##                                     RMSE       MAE      MAPE      MASE  
## Average      12157.667 10544.807 2.585153 1.5139682  
## Naïve        7254.374  6161.831 1.528049 0.8846834  
## Seasonal Naïve 10155.249  9703.442 2.403862 1.3931695  
## Drift         6999.073  5140.116 1.289655 0.7379909
```

Il seasonal naïve method, pur essendo l'unico in grado di cogliere la struttura stagionale della serie, presenta il secondo più elevato tasso di errore. Il più basso tasso di errore è fornito dal drift method, l'unico in grado di cogliere la componente di trend presente nei dati.

Natural cubic smoothing splines

Si può provare ad utilizzare un metodo di previsione più complesso e verificare se ciò porta ad un'accuratezza previsiva maggiore. Di seguito viene implementata la cubic smoothing spline, una forma di smoothing spline che utilizza funzioni cubiche a tratti per modellare la curva di regressione. Ogni osservazione rappresenta un nodo e, per evitare overfitting, vengono imposti vincoli sui coefficienti.

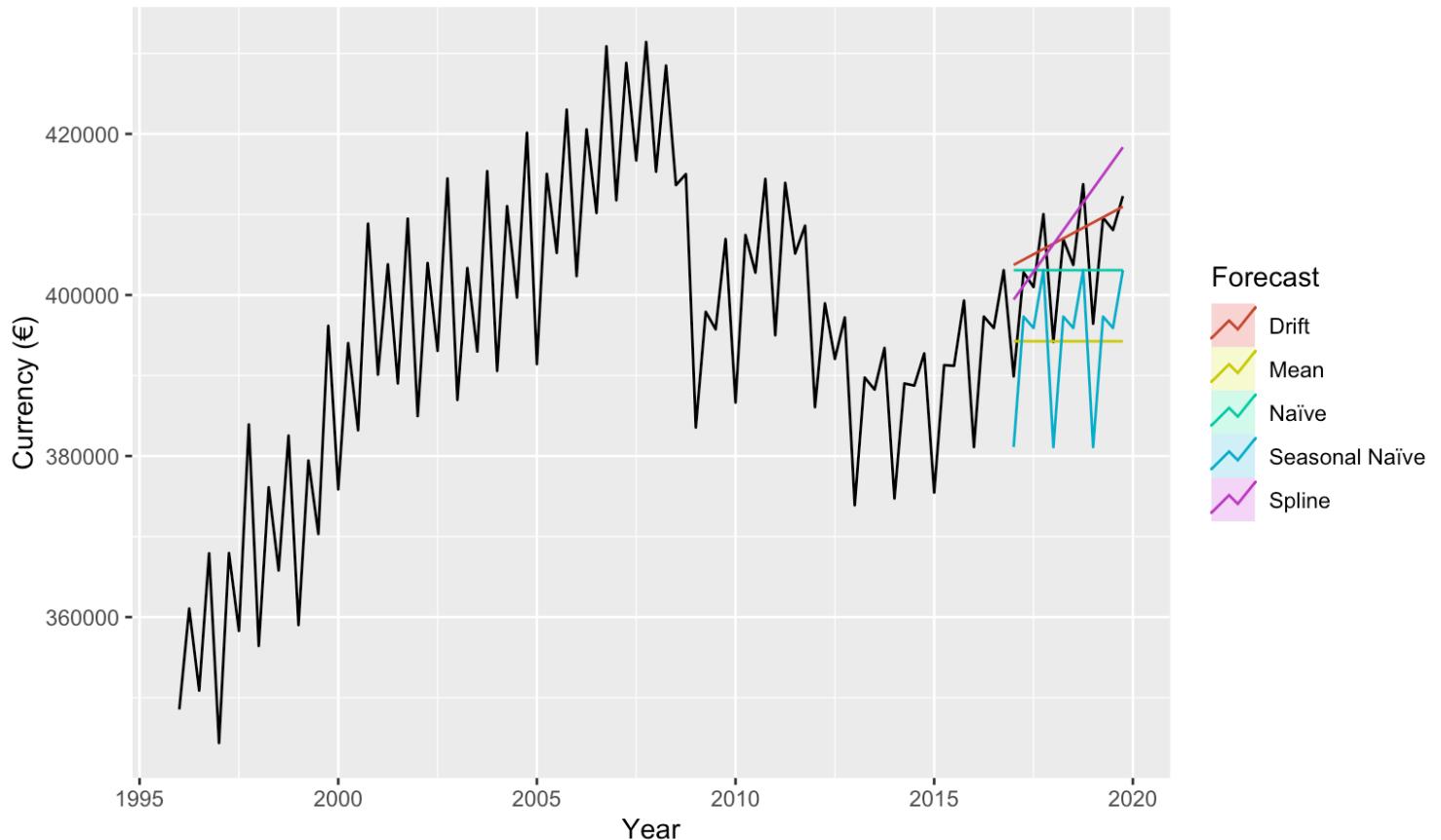
```
spline_fit_nc <- splinef(train_nc, 12)
```

```
autoplot(spline_fit_nc)
```



```
autoplot(gdp_nc) +  
  autolayer(mean_fit_nc, series="Mean", PI=FALSE) +  
  autolayer(naive_fit_nc, series="Naïve", PI=FALSE) +  
  autolayer(snaive_fit_nc, series="Seasonal Naïve", PI=FALSE) +  
  autolayer(drift_fit_nc, series="Drift", PI=FALSE) +  
  autolayer(spline_fit_nc, series="Spline", PI=FALSE) +  
  xlab("Year") +  
  ylab("Currency (€)") +  
  ggtitle("Real Gross Domestic Product") +  
  guides(colour=guide_legend(title="Forecast"))
```

Real Gross Domestic Product



```
rbind(
"Average" = accuracy(mean_fit_nc, test_nc)[2,c(2,3,5,6)],
"Naïve" = accuracy(naive_fit_nc, test_nc)[2,c(2,3,5,6)],
"Seasonal Naïve" = accuracy(snaive_fit_nc, test_nc)[2,c(2,3,5,6)],
"Drift" = accuracy(drift_fit_nc, test_nc)[2,c(2,3,5,6)],
"Spline" = accuracy(spline_fit_nc, test_nc)[2,c(2,3,5,6)])
```

	RMSE	MAE	MAPE	MASE
## Average	12157.667	10544.807	2.585153	1.5139682
## Naïve	7254.374	6161.831	1.528049	0.8846834
## Seasonal Naïve	10155.249	9703.442	2.403862	1.3931695
## Drift	6999.073	5140.116	1.289655	0.7379909
## Spline	7830.295	6396.832	1.595106	0.9184236

Nonostante la complessità del modello ora utilizzato sia maggiore, le sue performance previsive non superano quelle di modelli semplici come il drift method o il naïve method.

Analisi dei residui

Per verificare se un modello ha catturato in modo adeguato le informazioni contenute nei dati, è necessario effettuare l'analisi dei residui, che dovrebbero avere le due seguenti proprietà: essere incorrelati, per mostrare che tutte le informazioni presenti nei dati sono state catturate dal modello, e avere media zero, per mostrare che le previsioni sono non distorte.

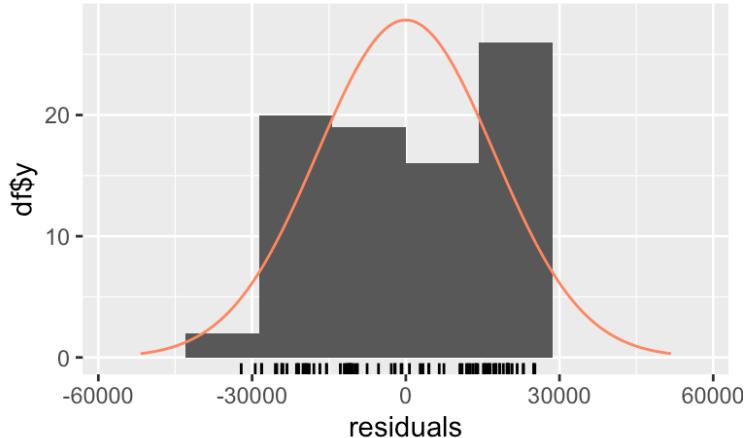
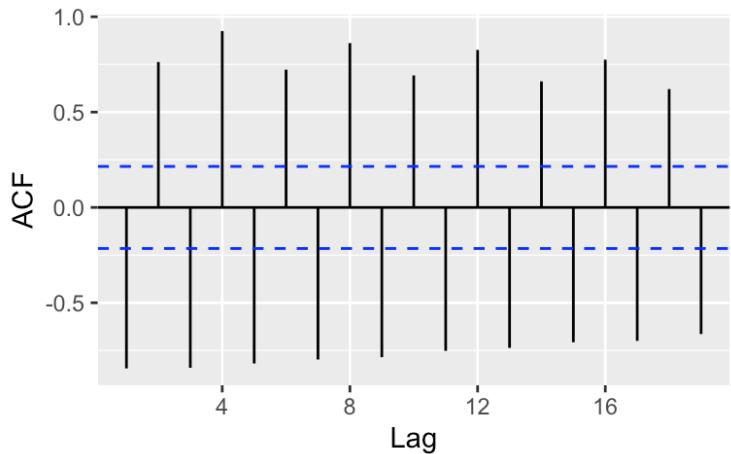
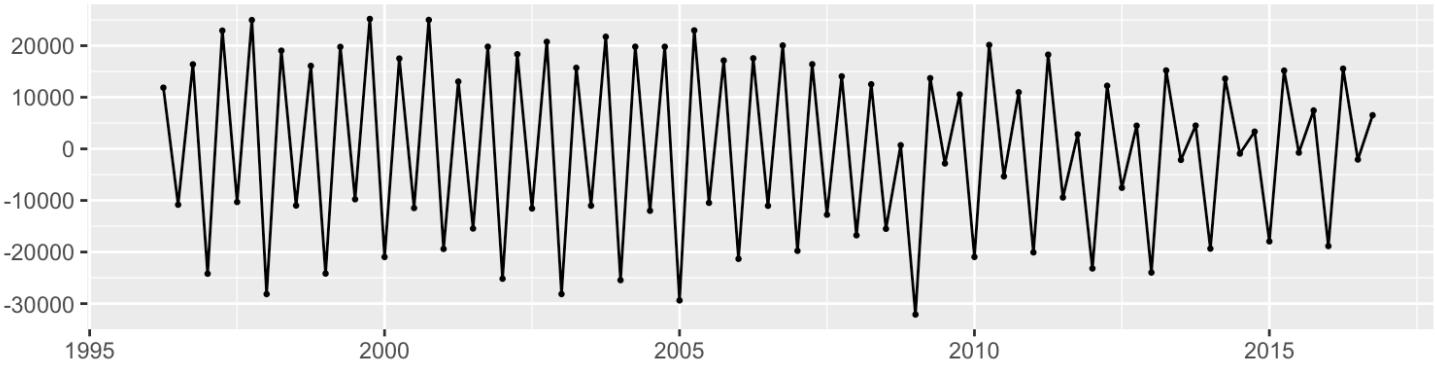
Inoltre, per la costruzione degli intervalli di confidenza, è utile che i residui abbiano varianza costante e siano normalmente distribuiti.

In sostanza, un buon modello è quello in cui i residui sono white noise.

Drift method

```
checkresiduals(drift_fit_nc)
```

Residuals from Random walk with drift



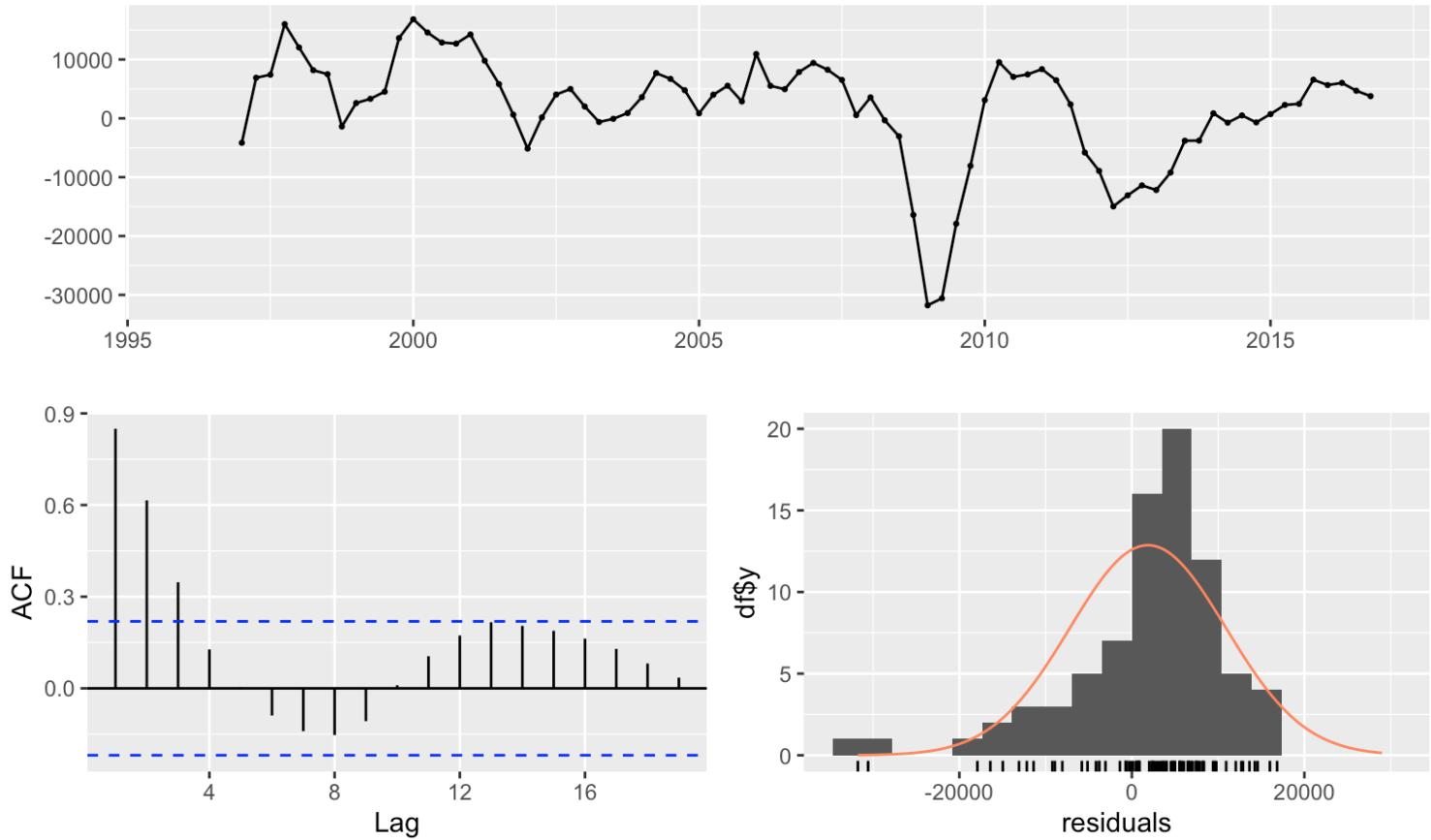
```
##  
## Ljung-Box test  
##  
## data: Residuals from Random walk with drift  
## Q* = 488.21, df = 8, p-value < 2.2e-16  
##  
## Model df: 0. Total lags used: 8
```

Nonostante il metodo drift sia quello con una maggiore accuratezza previsiva, questo presenta residui altamente correlati e non è in grado di cogliere la stagionalità presente nei dati, visibile qui all'interno dei residui e del loro correlogramma. Come si poteva intuire, il test di Ljung-Box per l'incorrelazione dei coefficienti ha un p-value minore di $\alpha = 0.05$, quindi rifiutiamo l'ipotesi nulla di incorrelazione dei residui. Inoltre, la distribuzione non è normale.

Seasonal Naïve Method

```
checkresiduals(snaive_fit_nc)
```

Residuals from Seasonal naive method



```
##  
## Ljung-Box test  
##  
## data: Residuals from Seasonal naive method  
## Q* = 108.08, df = 8, p-value < 2.2e-16  
##  
## Model df: 0. Total lags used: 8
```

Il metodo seasonal naïve, nonostante abbia una bassa accuratezza previsiva, presenta residui correlati solo per ritardi $h = 1, 2$ ed è in grado di cogliere la stagionalità presente nei dati. Inoltre, la sua distribuzione si può approssimare ad una normale. È presente asimmetria negativa a causa di due valori anomali. Infatti, anche dalla serie dei residui è evidente come il modello soffra della presenza degli outliers, legati alla crisi finanziaria del 2008. Anche in questo caso il test di Ljung-Box per l'incorrelazione dei coefficienti ha un p-value minore di $\alpha = 0.05$, quindi rifiutiamo l'ipotesi nulla di incorrelazione dei residui.

GDP in funzione di altre misure macroeconomiche

Per la costruzione di un modello per la previsione del GDP verranno di seguito prese in considerazione altre serie storiche macroeconomiche. Tuttavia, in questa sede non si procede con l'analisi dettagliata delle suddette serie (su cui ci sarebbe molto da dire), ma solo delle caratteristiche necessarie agli scopi dello studio.

Per ciascuna delle seguenti serie storiche, i dati estratti sono sullo stesso intervallo temporale considerato per la serie del GDP.

Real Exports of Goods and Services

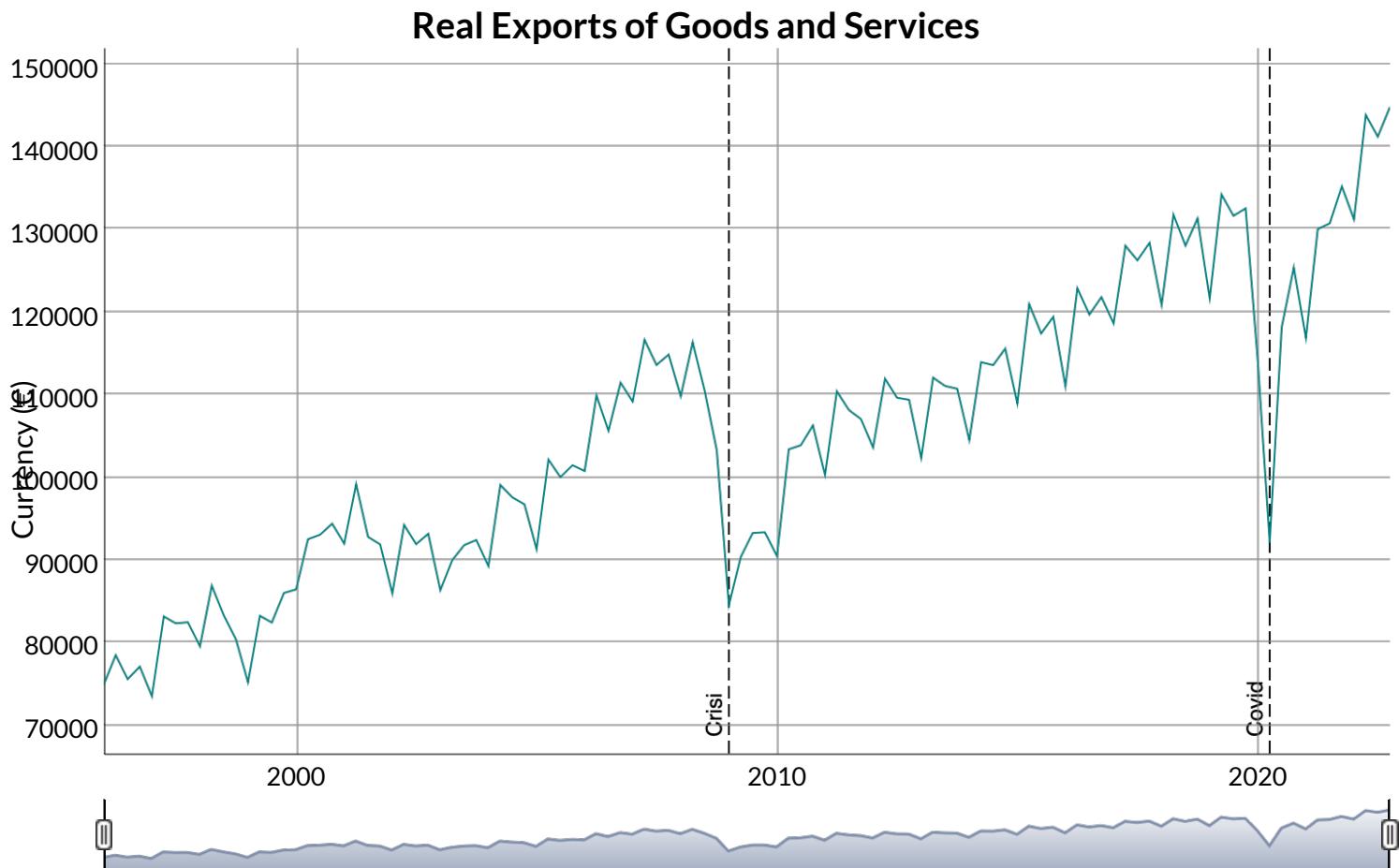
Il Real Exports of Goods and Services si riferisce al valore totale di beni e servizi che l'Italia esporta verso altri Paesi, adeguato all'inflazione. Si tratta di un indicatore della competitività internazionale di un Paese. Per l'Italia, le esportazioni sono una parte significativa dell'economia, poiché il Paese è noto per le sue esportazioni di prodotti di alta qualità.

Nota: Ci si riferirà a questa serie storica sempre con la sigla EXP.

Fonte: International Monetary Fund, Real Exports of Goods and Services for Italy [NXRNSAXDCITQ], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/NXRNSAXDCITQ> (<https://fred.stlouisfed.org/series/NXRNSAXDCITQ>), May 11, 2023.

```
exp <- ts(getSymbols("NXRNSAXDCITQ", src="FRED", auto.assign = FALSE), start=1996, frequency=4)
```

```
dyRangeSelector(dygraph(exp,
  main="Real Exports of Goods and Services",
  ylab="Currency (€)") %>%
  dyEvent("2020-04-01", "Covid", labelLoc = "bottom") %>%
  dyEvent("2009-01-01", "Crisi", labelLoc = "bottom")
)
```



Real Imports of Goods and Services for Italy

Il Real Imports of Goods and Services si riferisce al valore totale di beni e servizi che l'Italia importa da altri Paesi, adeguato all'inflazione. Si tratta di un indicatore economico della dipendenza dell'Italia dalle importazioni.

Come molte altre economie moderne, l'Italia importa una vasta gamma di beni e servizi.

Nota: Ci si riferirà a questa serie storica sempre con la sigla IMP.

Fonte: International Monetary Fund, Real Imports of Goods and Services for Italy [NMRNSAXDCITQ], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/NMRNSAXDCITQ> (<https://fred.stlouisfed.org/series/NMRNSAXDCITQ>), May 11, 2023.

```
imp <- ts(getSymbols("NMRNSAXDCITQ", src="FRED", auto.assign = FALSE), start=1996, frequency=4)
```

```
dyRangeSelector(dygraph(imp,
  main="Real Imports of Goods and Services for Italy",
  ylab="Currency (€)") %>%
  dyEvent("2020-04-01", "Covid", labelLoc = "bottom") %>%
  dyEvent("2009-01-01", "Crisi", labelLoc = "bottom")
)
```



Early Estimate of Quarterly ULC Indicators: Total Labor Productivity

L'ULC, acronimo di Unit Labor Cost, è un indicatore macroeconomico che misura il costo del lavoro necessario per produrre un'unità di prodotto. L'ULC si calcola come il rapporto tra il costo del lavoro e la produttività del lavoro.

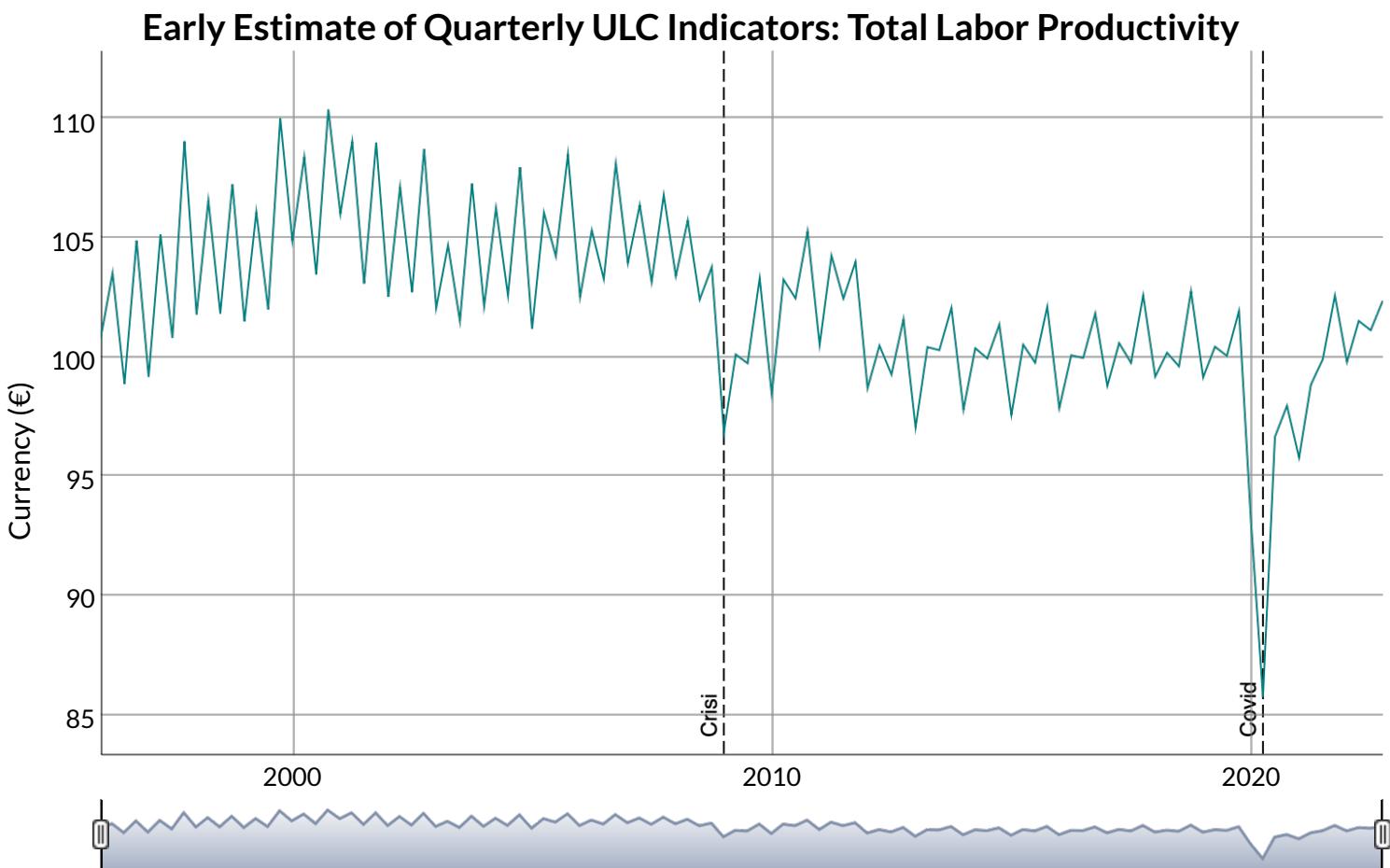
Questo indicatore può fornire informazioni importanti sulla capacità produttiva dell'economia italiana e sulla sua competitività internazionale.

Nota: Ci si riferirà a questa serie storica sempre con la sigla LAB.

Fonte: Organization for Economic Co-operation and Development, Early Estimate of Quarterly ULC Indicators: Total Labor Productivity for Italy [ULQELP01ITQ661N], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/ULQELP01ITQ661N> (<https://fred.stlouisfed.org/series/ULQELP01ITQ661N>), May 11, 2023.

```
lab <- ts(getSymbols("ULQELP01ITQ661N", src="FRED", auto.assign = FALSE) %>%  
  window(start="1996-01-01"), start=1996, frequency=4)
```

```
dyRangeSelector(dygraph(lab,  
  main="Early Estimate of Quarterly ULC Indicators: Total Labor Productivity",  
  ylab="Currency (€)") %>%  
  dyEvent("2020-04-01", "Covid", labelLoc = "bottom") %>%  
  dyEvent("2009-01-01", "Crisi", labelLoc = "bottom")  
)
```



Real Final Consumption Expenditure for Italy

Con *Real Final Consumption Expenditure* ci si riferisce al valore totale della spesa finale effettuata dai residenti italiani per l'acquisto di beni e servizi, adeguato all'inflazione.

La spesa finale delle famiglie è un importante indicatore della domanda interna: una spesa finale elevata può indicare un aumento del reddito disponibile e una maggiore fiducia dei consumatori nell'economia. Tuttavia,

una spesa finale eccessivamente elevata può anche indicare una maggiore inflazione o un aumento del debito delle famiglie.

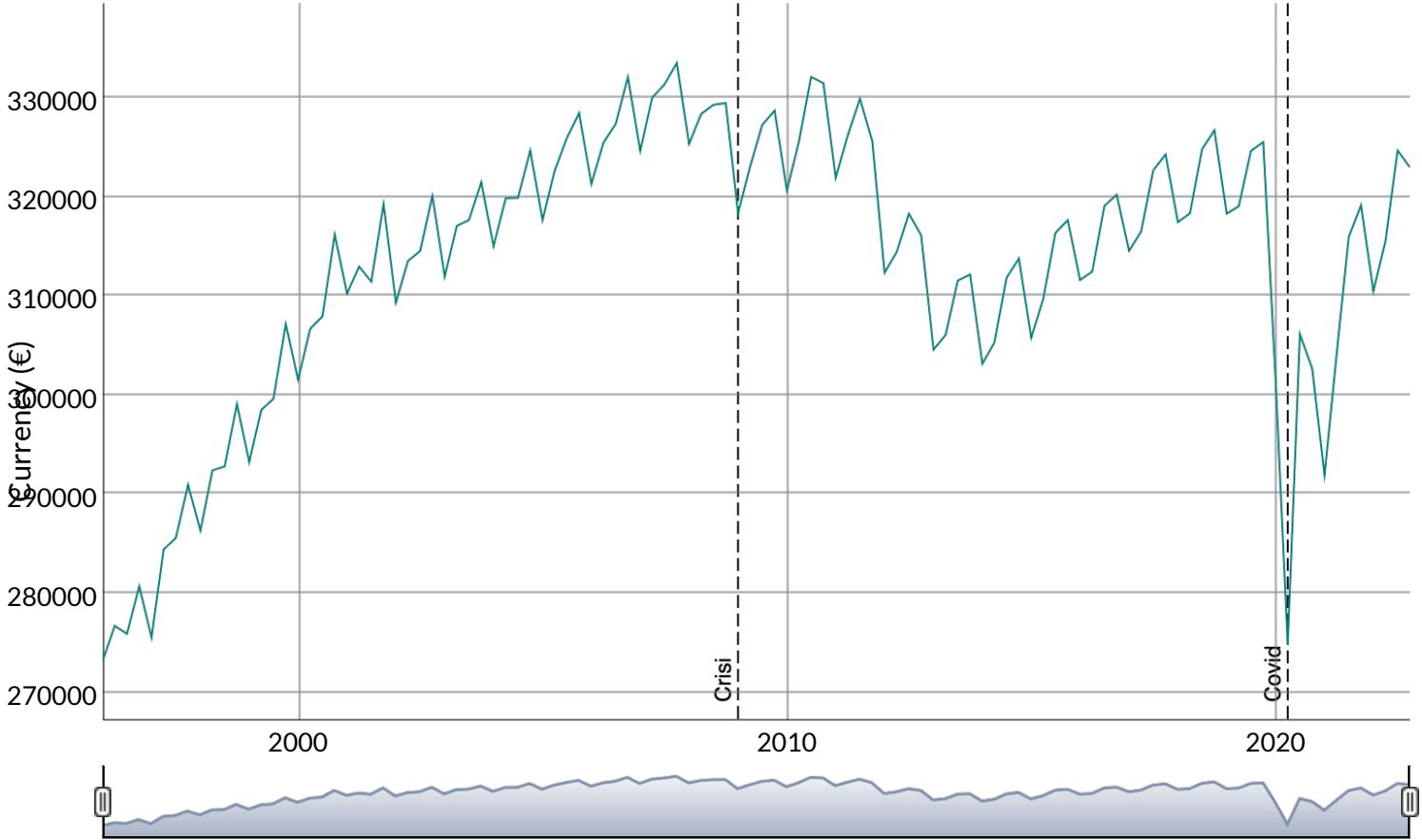
Nota: Ci si riferirà a questa serie storica sempre con la sigla CONS.

Fonte: International Monetary Fund, Real Final Consumption Expenditure for Italy [NCRNSAXDCITQ], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/NCRNSAXDCITQ> (<https://fred.stlouisfed.org/series/NCRNSAXDCITQ>), May 11, 2023.

```
cons <- ts(getSymbols("NCRNSAXDCITQ", src="FRED", auto.assign = FALSE), start=1996, frequency=4)
```

```
dyRangeSelector(dygraph(cons,
  main="Real Final Consumption Expenditure for Italy",
  ylab="Currency (€)") %>%
  dyEvent("2020-04-01", "Covid", labelLoc = "bottom") %>%
  dyEvent("2009-01-01", "Crisi", labelLoc = "bottom")
)
```

Real Final Consumption Expenditure for Italy



Real General Government Final Consumption Expenditure for Italy

Il Real General Government Final Consumption Expenditure si riferisce al valore totale della spesa finale effettuata dal governo per l'acquisto di beni e servizi (ad esempio le spese per la difesa, l'istruzione, la salute, la cultura, ecc.), adeguato all'inflazione.

La spesa pubblica può influire sull'economia nazionale attraverso il finanziamento di progetti infrastrutturali e di sviluppo, la creazione di posti di lavoro e l'aumento della domanda di beni e servizi. Tuttavia, una spesa

pubblica eccessivamente elevata può anche portare a un aumento del debito pubblico e ad un aumento dei tassi di interesse.

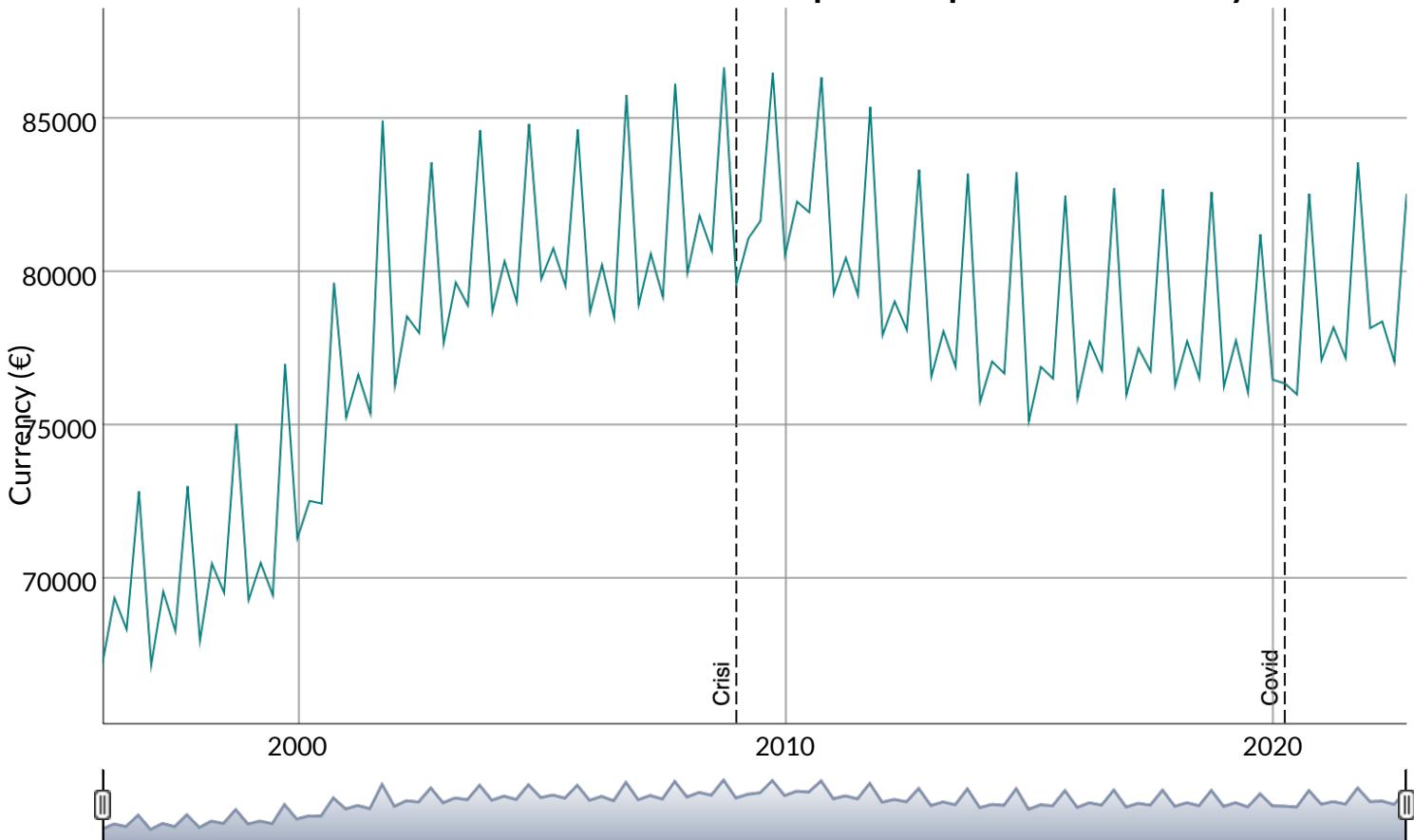
Nota: Ci si riferirà a questa serie storica sempre con la sigla CONSGG.

Fonte: International Monetary Fund, Real General Government Final Consumption Expenditure for Italy [NCGGRNSAXDCITQ], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/NCGGRNSAXDCITQ> (<https://fred.stlouisfed.org/series/NCGGRNSAXDCITQ>), May 11, 2023.

```
consGG <- ts(getSymbols("NCGGRNSAXDCITQ", src="FRED", auto.assign = FALSE), start=1996, frequency=4)
```

```
dyRangeSelector(dygraph(consGG,
  main="Real General Government Final Consumption Expenditure for Italy",
  ylab="Currency (€)") %>%
  dyEvent("2020-04-01", "Covid", labelLoc = "bottom") %>%
  dyEvent("2009-01-01", "Crisi", labelLoc = "bottom")
)
```

Real General Government Final Consumption Expenditure for Italy



Consumer Price Index: Total All Items for Italy

Il Consumer Price Index è un indicatore economico che misura la variazione dei prezzi di un paniere di beni e servizi consumati dalle famiglie. In pratica, l'indice dei prezzi al consumo fornisce una misura dell'inflazione in Italia.

L'aumento dei prezzi al consumo può avere un impatto significativo sulla capacità di acquisto delle famiglie e sull'andamento dell'economia.

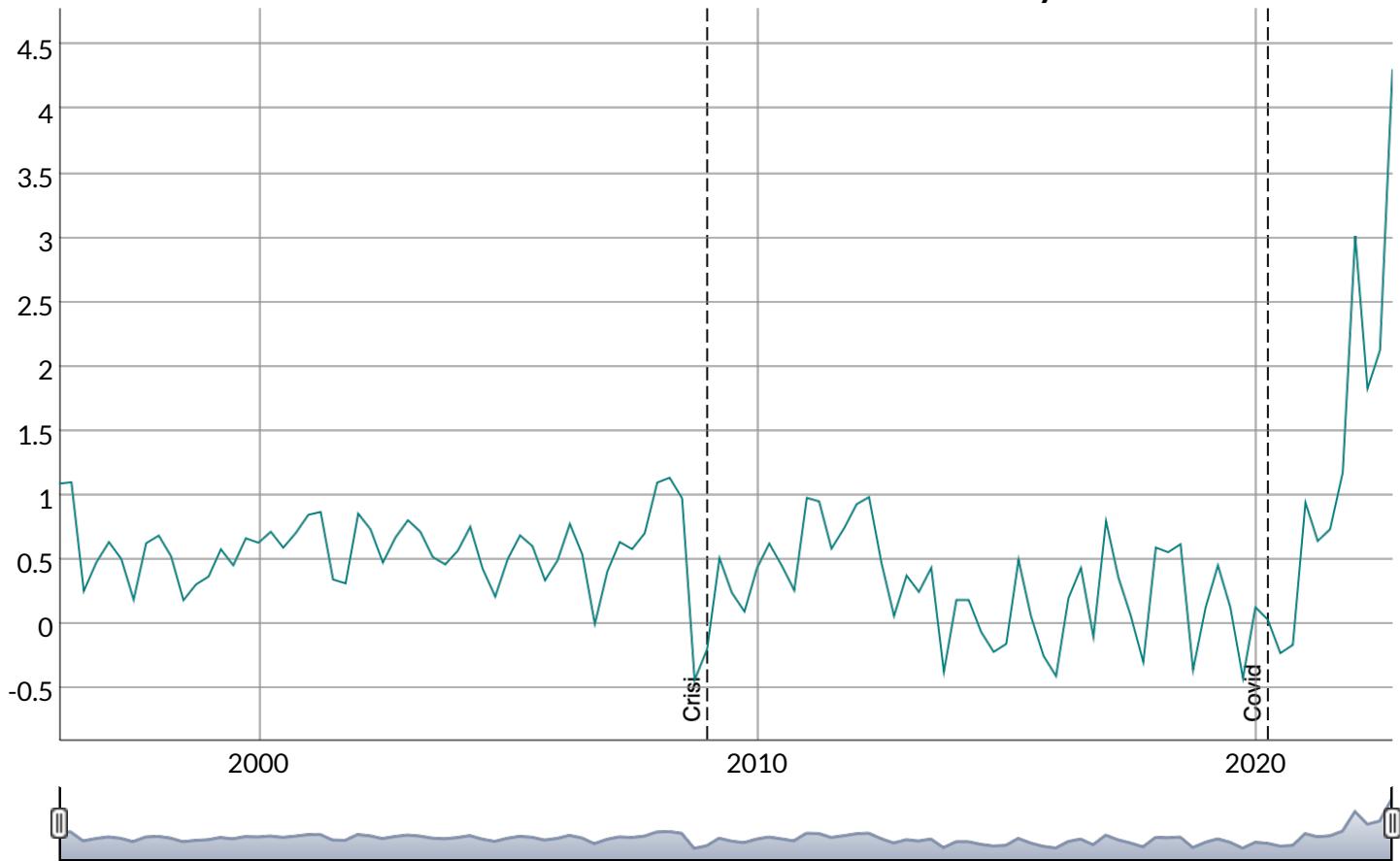
Nota: Ci si riferirà a questa serie storica sempre con la sigla CPI.

Fonte: Organization for Economic Co-operation and Development, Consumer Price Index: Total All Items for Italy [CPALTT01ITQ657N], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/CPALTT01ITQ657N> (<https://fred.stlouisfed.org/series/CPALTT01ITQ657N>), May 10, 2023.

```
cpi <- ts(getSymbols("CPALTT01ITQ657N", src="FRED", auto.assign = FALSE) %>%  
  window(start="1996-01-01", end="2022-10-01"), start=1996, frequency=4)
```

```
dyRangeSelector(dygraph(cpi,  
  main="Consumer Price Index: Total All Items for Italy") %>%  
  dyEvent("2020-04-01", "Covid", labelLoc = "bottom") %>%  
  dyEvent("2009-01-01", "Crisi", labelLoc = "bottom")  
)
```

Consumer Price Index: Total All Items for Italy



Real Gross Capital Formation for Italy

Con Real Gross Capital Formation ci si riferisce al valore totale degli investimenti effettuati dalle imprese, sia pubbliche che private, in beni di capitali al netto della deprezzamento e adeguato all'inflazione.

L'investimento in beni di capitale è una componente importante della spesa nazionale, poiché consente di

aumentare la produttività e la capacità produttiva dell'economia italiana.

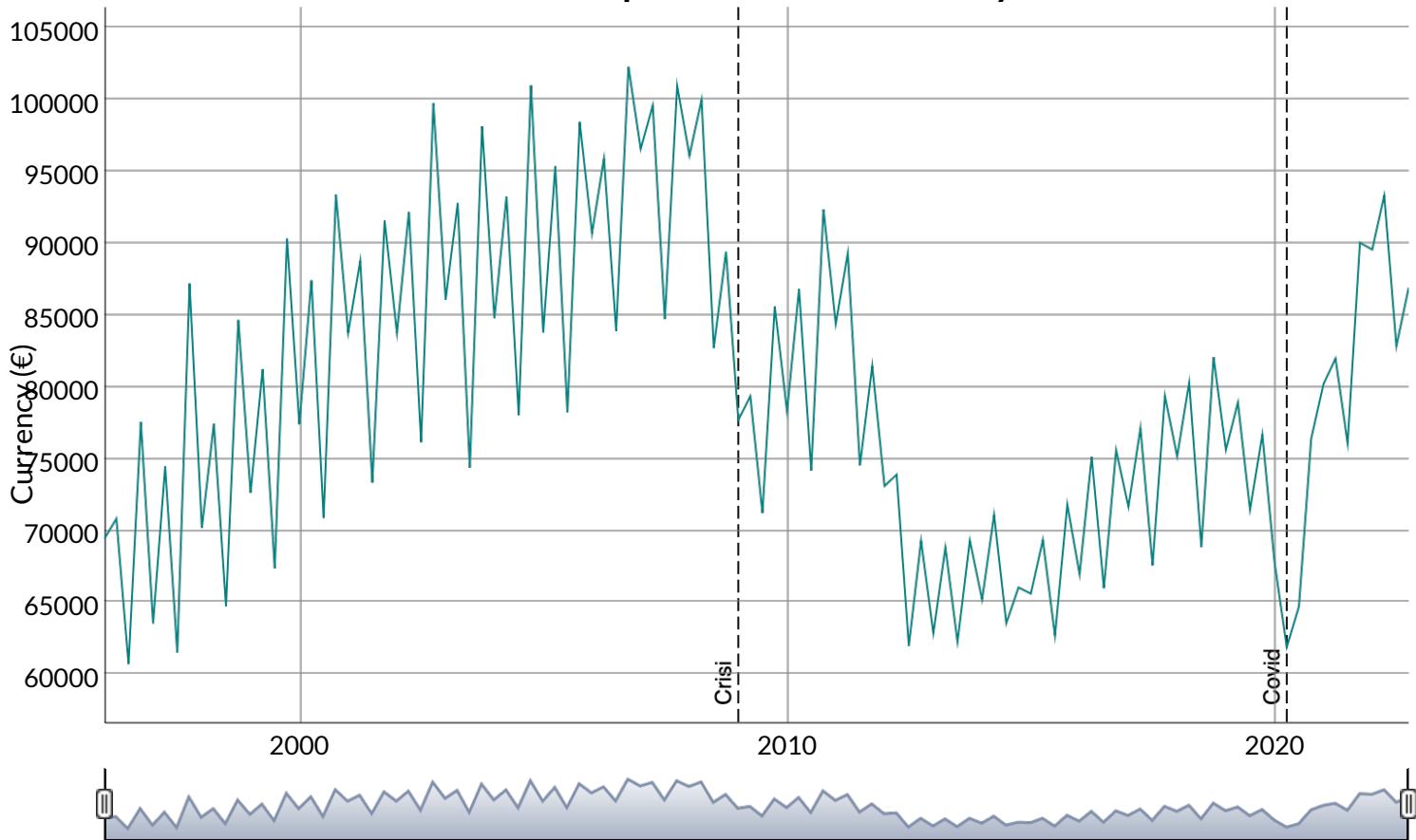
Nota: Ci si riferirà a questa serie storica sempre con la sigla CAP.

Fonte: International Monetary Fund, Real Gross Capital Formation for Italy [NIRNSAXDCITQ], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/NIRNSAXDCITQ> (<https://fred.stlouisfed.org/series/NIRNSAXDCITQ>), May 11, 2023.

```
cap <- ts(getSymbols("NIRNSAXDCITQ", src="FRED", auto.assign = FALSE), start=1996, frequency=4)
```

```
dyRangeSelector(dygraph(cap,
  main="Real Gross Capital Formation for Italy",
  ylab="Currency (€)") %>%
  dyEvent("2020-04-01", "Covid", labellLoc = "bottom") %>%
  dyEvent("2009-01-01", "Crisi", labellLoc = "bottom")
)
```

Real Gross Capital Formation for Italy

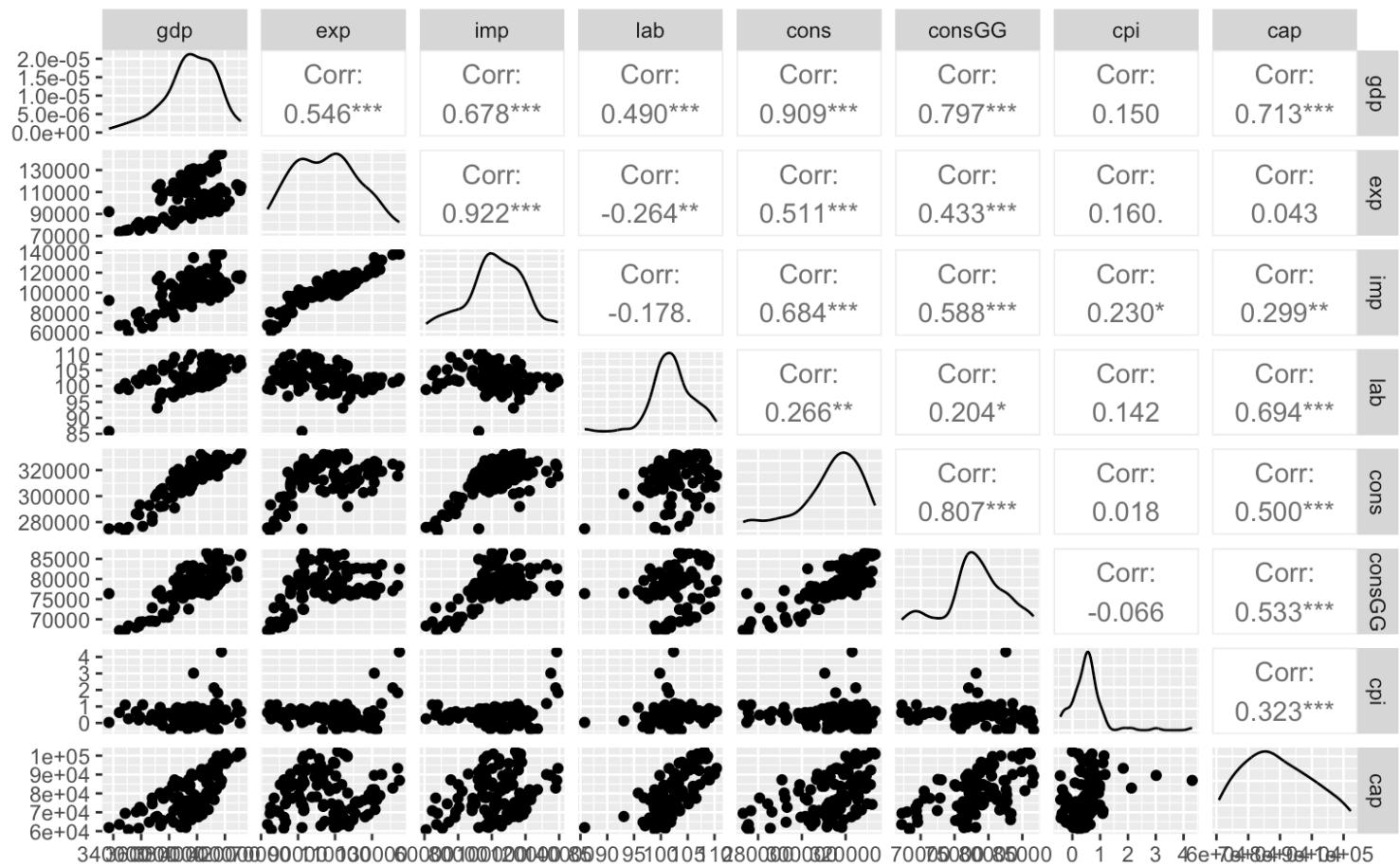


Costruzione del dataset

```
data <- cbind(gdp, exp, imp, lab, cons, consGG, cpi, cap)
colnames(data) <- c("gdp", "exp", "imp", "lab", "cons", "consGG", "cpi", "cap")
```

Per evitare i problemi legati alla multicollinearità, bisogna assicurarsi che le variabili non siano altamente correlate tra loro.

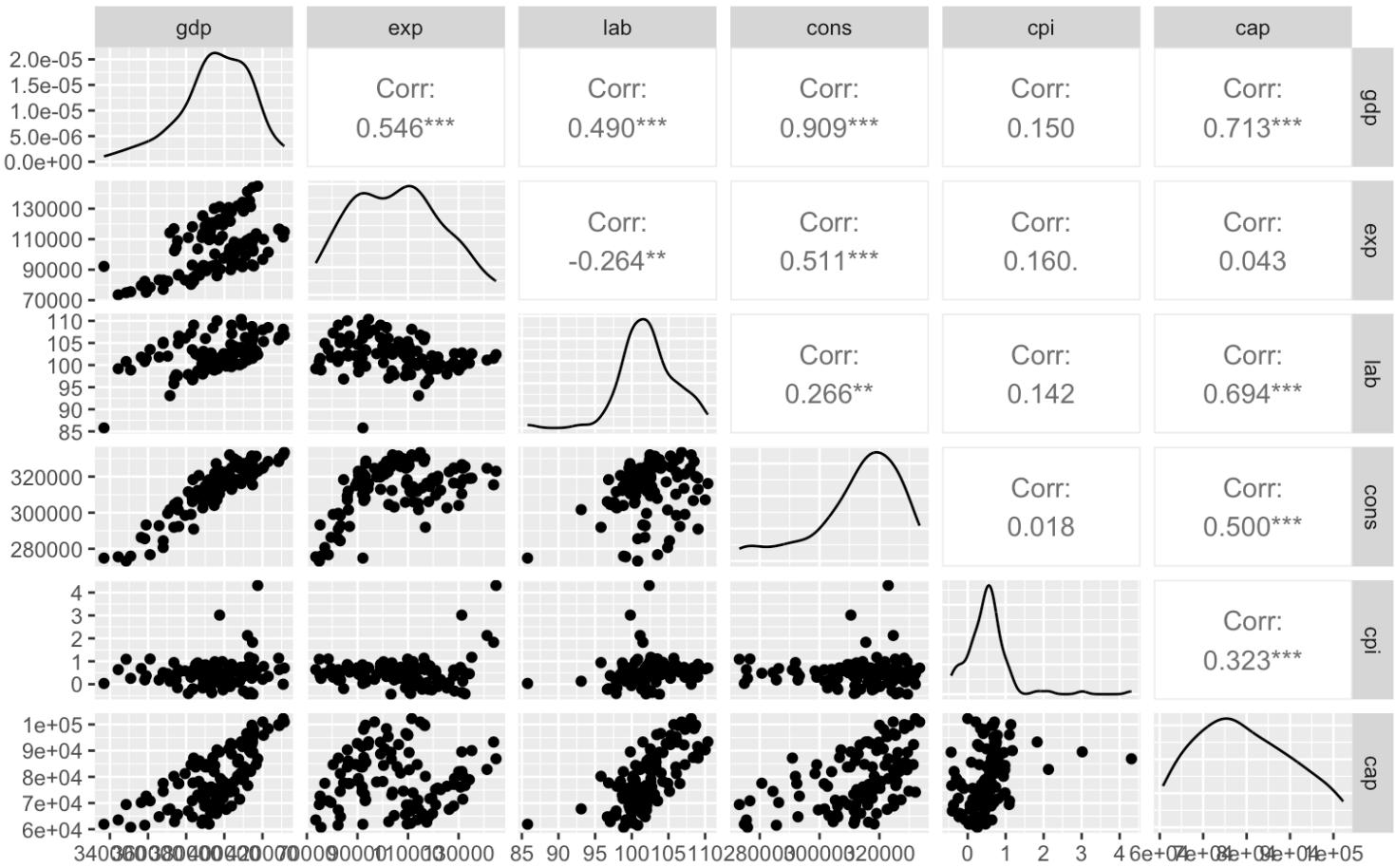
```
ggpairs(as.data.frame(data))
```



Risulta opportuno escludere le variabili `imp` e `consGG`, rispettivamente altamente correlate alle variabili `exp` e `cons`.

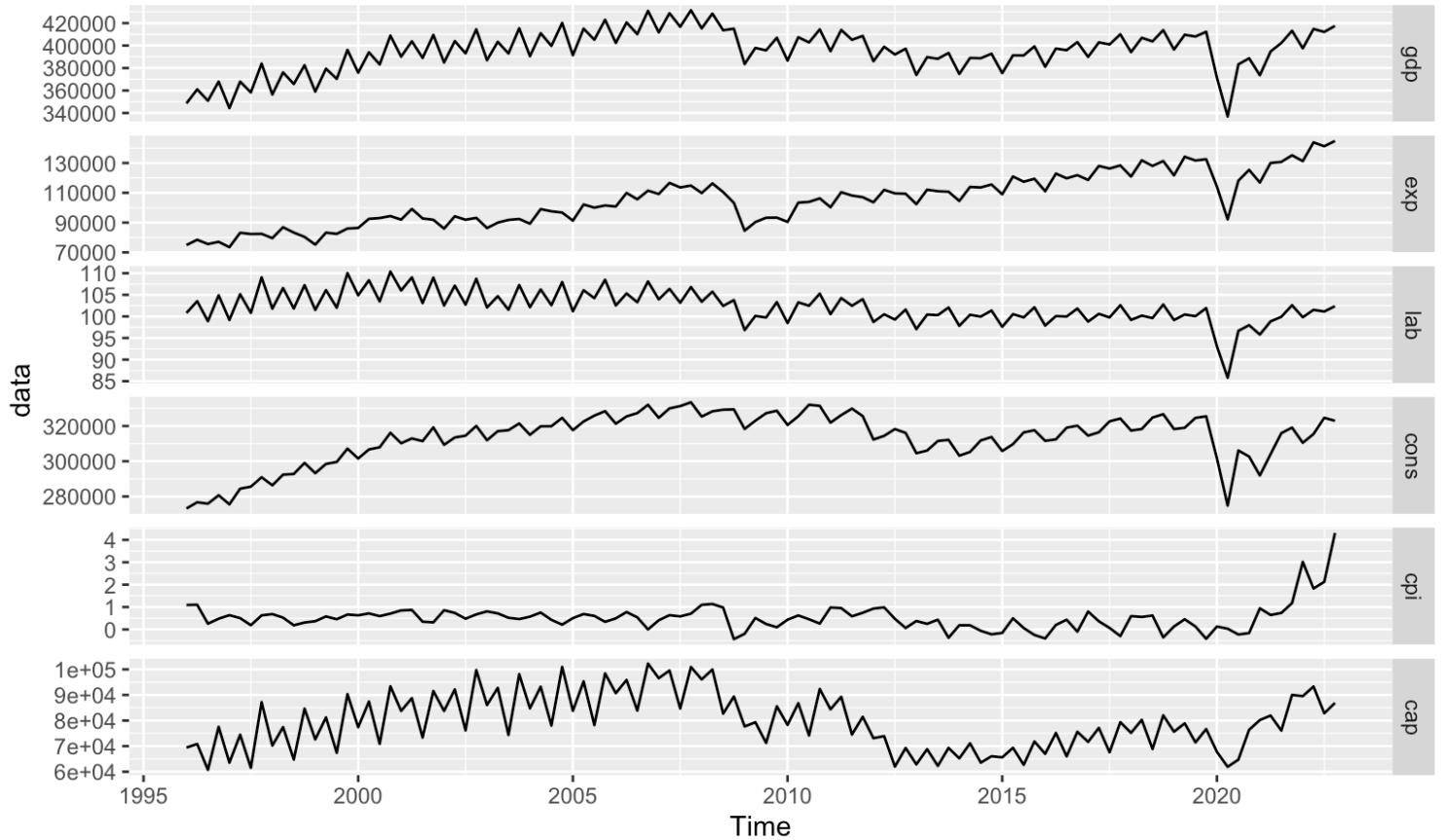
```
data <- cbind(gdp, exp, lab, cons, cpi, cap)
colnames(data) <- c("gdp", "exp", "lab", "cons", "cpi", "cap")
```

```
ggpairs(as.data.frame(data))
```



Dagli scatterplot risulta evidente l'esistenza di una relazione lineare tra la variabile GDP e le covariate prese in considerazione, suggerendo di poter costruire un modello di regressione lineare.

```
autoplot(data, facets = TRUE)
```



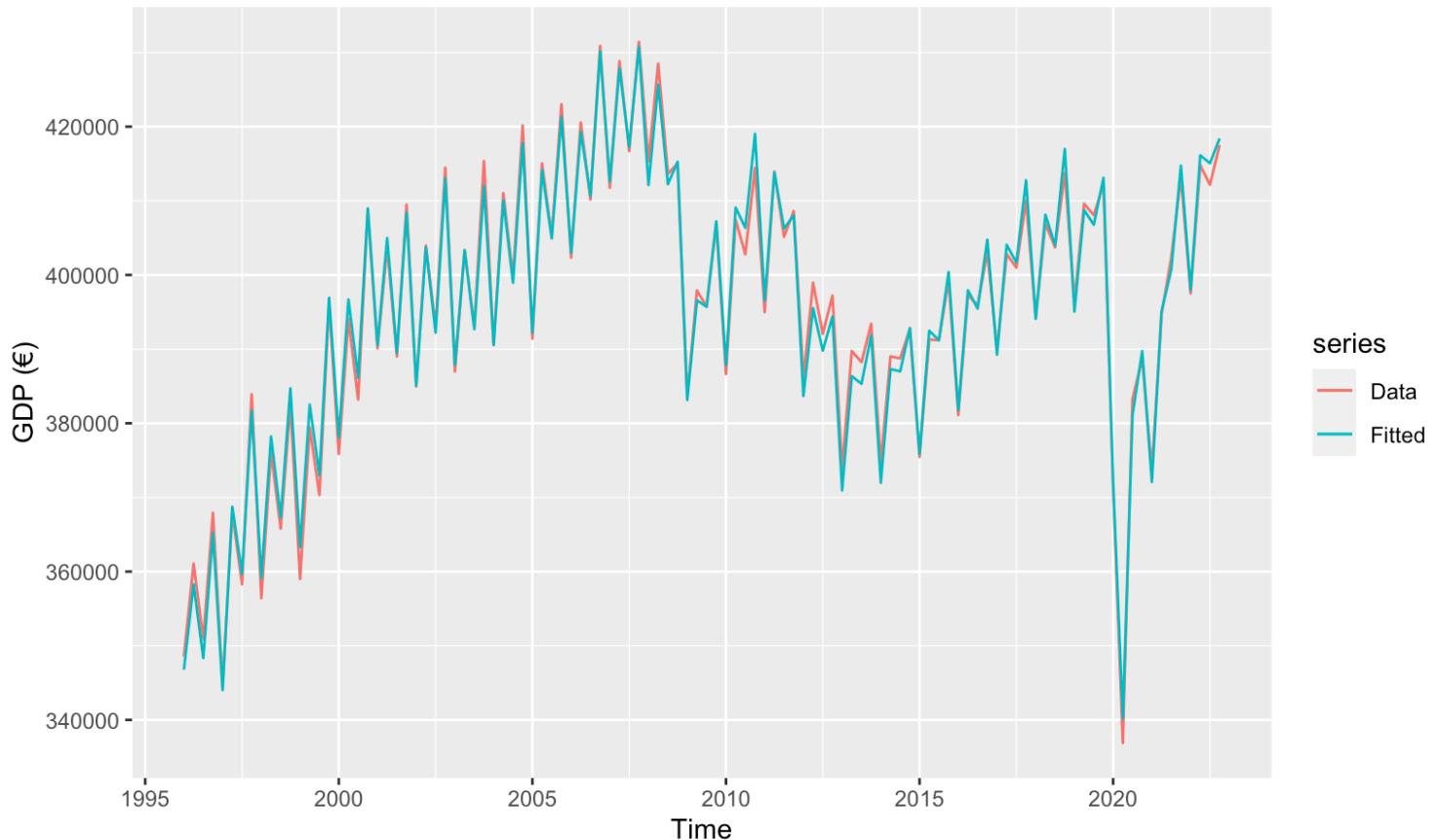
Modello di regressione lineare multiplo

Il modello più semplice che possiamo considerare è il modello di regressione lineare, che assume la presenza di un legame lineare tra la variabile `gdp` e le covariate sopra presentate. Inoltre, si aggiungono le variabili `trend` e `season`, legate alla struttura della serie storica dipendente.

$$gdp = \beta_0 + \beta_1 exp + \beta_2 lab + \beta_3 cons + \beta_4 cpi + \beta_5 cap + \beta_6 trend + \beta_7 season + \epsilon$$

```
mod <- tslm(gdp ~ exp + lab + cons + cpi + cap + trend + season, data=data)
```

```
autoplot(gdp, series="Data") +
  autolayer(fitted(mod), series="Fitted") +
  ylab("GDP (€)")
```



Il modello sembra adattarsi molto bene ai dati, ma è necessario analizzare le sue statistiche di sintesi.

```
summary(mod)
```

```

## 
## Call:
## tslm(formula = gdp ~ exp + lab + cons + cpi + cap + trend + season,
##       data = data)
## 
## Residuals:
##      Min      1Q Median      3Q     Max 
## -4612.2 -1128.3   -94.5  1330.7  3438.6 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 2.443e+04  1.187e+04   2.058  0.042259 *  
## exp         4.583e-01  4.093e-02   11.196 < 2e-16 *** 
## lab          4.127e+02  1.559e+02   2.647  0.009452 **  
## cons         7.741e-01  2.479e-02   31.222 < 2e-16 *** 
## cpi          -4.871e+02  3.709e+02  -1.313  0.192145    
## cap          5.126e-01  3.528e-02   14.532 < 2e-16 *** 
## trend        -1.196e+02  3.050e+01  -3.922  0.000163 *** 
## season2      5.416e+03  5.898e+02   9.182 7.21e-15 *** 
## season3      4.230e+03  6.477e+02   6.531 2.91e-09 *** 
## season4      5.847e+03  7.474e+02   7.822 6.07e-12 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 1891 on 98 degrees of freedom 
## Multiple R-squared:  0.991, Adjusted R-squared:  0.9902 
## F-statistic: 1202 on 9 and 98 DF,  p-value: < 2.2e-16

```

Un dato particolarmente rivelante è che tutte le variabili, ad eccezione del `cpi` (indice dei prezzi al consumo) sono statisticamente significative. Ciò è confermato anche dal fatto che il p-value della F-statistic ci porta a rifiutare l'ipotesi nulla H_0 di nullità di tutti i predittori e ad accettare l'ipotesi alternativa H_1 di almeno un predittore non nullo.

Il modello così costruito presenta un elevatissimo indice R^2 corretto, una misura della bontà di adattamento. L'indice R^2 indica la percentuale di varianza nella variabile dipendente spiegata dalle variabili indipendenti, ma non tiene conto del numero di variabili indipendenti utilizzate nel modello, il che può portare a una sovrastima dell'efficacia del modello quando vengono utilizzate molte variabili indipendenti. L'indice R^2 corretto aumenta solo se l'aggiunta di una variabile indipendente migliora effettivamente la capacità del modello di spiegare la variazione nella variabile dipendente.

L'indice R^2 corretto, così come quello standard, ha un valore compreso tra 0 e 1: quanto più il valore è prossimo a 1, tanto migliore è l'adattamento del modello ai dati. In questo caso l'indice corretto $R^2 = 0.99$ mostra un ottimo adattamento del modello ai dati.

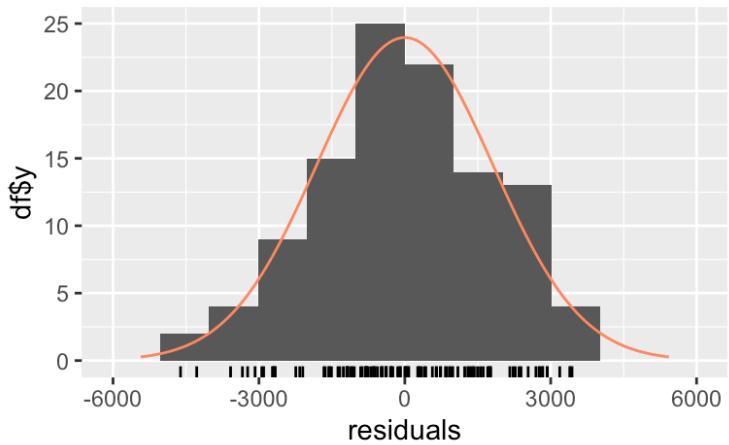
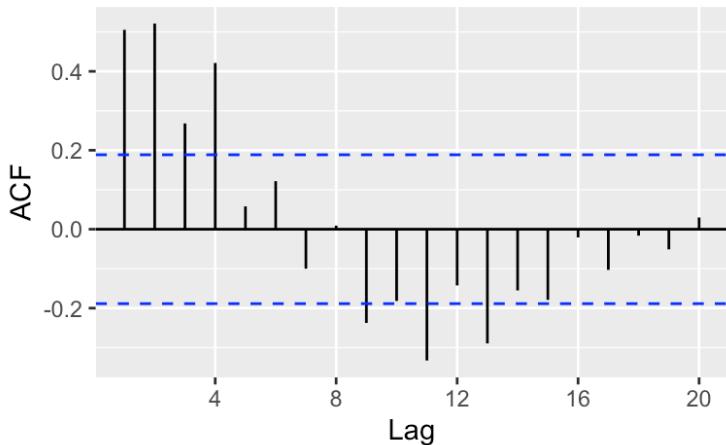
Tuttavia, è indispensabile effettuare un'analisi dei residui. Dalle statistiche sopra riportate possiamo dedurre che la distribuzione dei residui è leggermente asimmetrica positiva e presenta uno standard error, misura della varianza dei residui, molto elevato. Si deduce che il modello di regressione non è in grado di spiegare tutta la variabilità presente nella variabile dipendente.

Analisi dei residui

Per i motivi sopra menzionati, è importante effettuare un'analisi dei residui per valutare il modello di regressione.

```
checkresiduals(mod)
```

Residuals from Linear regression model



```
##  
## Breusch-Godfrey test for serial correlation of order up to 13  
##  
## data: Residuals from Linear regression model  
## LM test = 58.143, df = 13, p-value = 1.126e-07
```

I residui mostrano eteroschedasticità, appaiono correlati e il p-value del test di Breusch-Godfrey è molto minore di 0.05, quindi non c'è evidenza empirica per poter affermare che i residui sono in linea con un processo white noise. Tuttavia, la correlazione sembra essere forte solo per i ritardi $h = 1, 2$ e la distribuzione si può approssimare ad una normale.

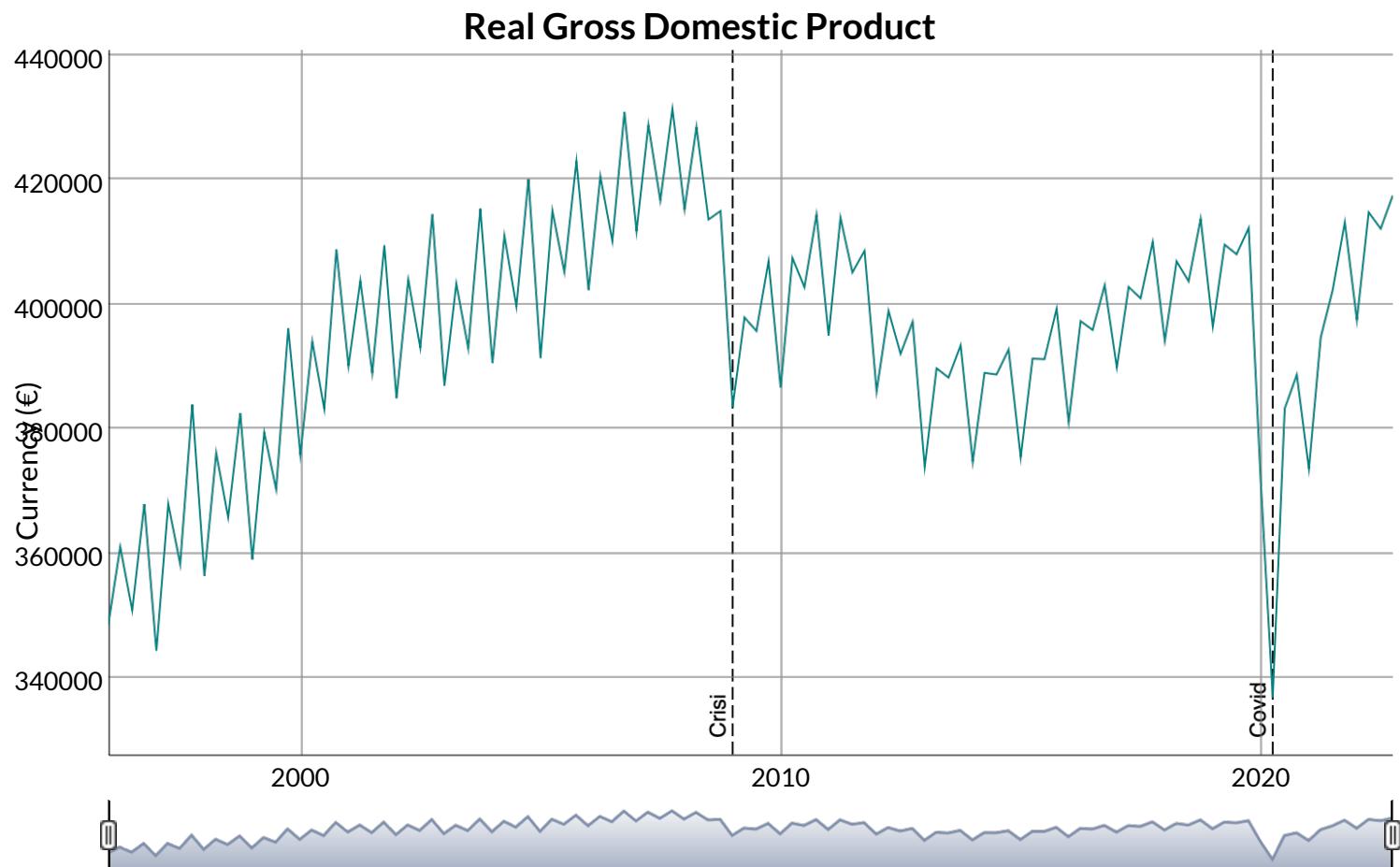
Modello di regressione lineare a tratti

Si potrebbe ipotizzare che questo problema sia legato ai cambi di pendenza nella serie storica del GDP, dunque si procede con la definizione di un modello lineare a tratti.

```

dyRangeSelector(dygraph(gdp,
  main="Real Gross Domestic Product",
  ylab="Currency (€)") %>%
  dyEvent("2020-04-01", "Covid", labelLoc = "bottom") %>%
  dyEvent("2009-01-01", "Crisi", labelLoc = "bottom")
)

```



Si procede con l'utilizzo di due nodi, uno in corrispondenza del quarto trimestre del 2007 e l'altro del primo trimestre del 2014.

```

t <- time(gdp)
tau.2007 <- c(2007,4)
tau.2014 <- c(2014,1)
tau1 <- ts(pmax(0, t - tau.2007), start = 1996)
tau2 <- ts(pmax(0, t - tau.2014), start = 1996)
mod_pw <- tslm(gdp ~ t + tau1 + tau2 + exp + lab + cons + cpi + cap + season)
summary(mod_pw)

```

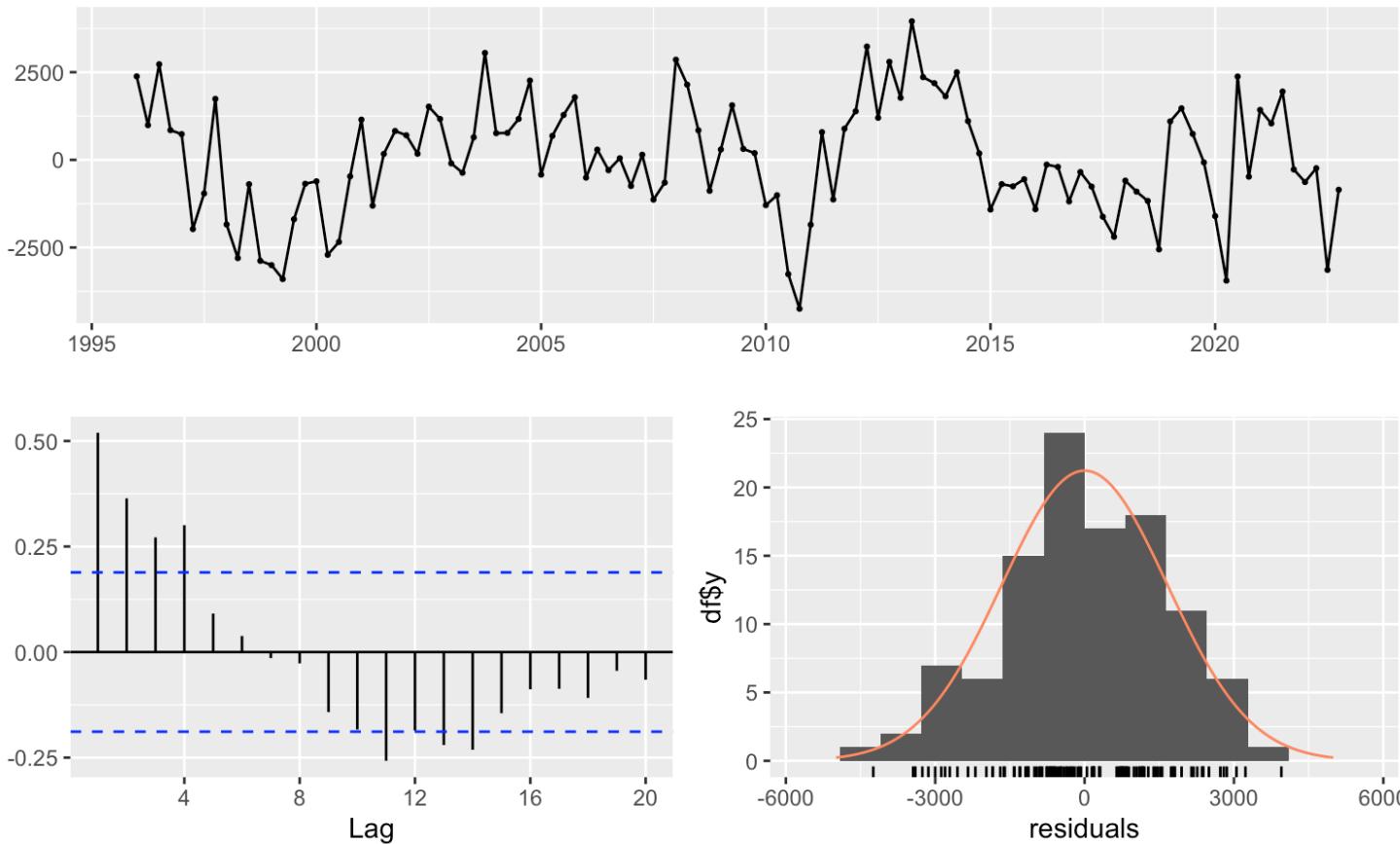
```

## 
## Call:
## tslm(formula = gdp ~ t + tau1 + tau2 + exp + lab + cons + cpi +
##       cap + season)
## 
## Residuals:
##      Min      1Q Median      3Q     Max 
## -4244.5 -970.6 -86.9 1152.4 3952.4 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1.475e+06 3.712e+05  3.973 0.000137 ***
## t          -7.121e+02 1.826e+02 -3.901 0.000178 *** 
## tau1        1.836e+02 2.393e+02  0.768 0.444650  
## tau2        -3.777e+02 1.476e+02 -2.558 0.012090 *  
## exp         5.526e-01 4.418e-02 12.509 < 2e-16 ***
## lab         -1.806e+01 1.806e+02 -0.100 0.920590  
## cons         7.951e-01 2.621e-02 30.332 < 2e-16 *** 
## cpi         -2.896e+02 3.477e+02 -0.833 0.406889  
## cap          5.238e-01 3.361e-02 15.582 < 2e-16 *** 
## season2     3.963e+05 2.535e+05  1.563 0.121256  
## season3     4.090e+03 6.027e+02  6.786 9.48e-10 *** 
## season4     3.977e+05 2.536e+05  1.568 0.120129  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 1756 on 96 degrees of freedom
## Multiple R-squared:  0.9924, Adjusted R-squared:  0.9915 
## F-statistic: 1142 on 11 and 96 DF,  p-value: < 2.2e-16

```

```
checkresiduals(mod_pw)
```

Residuals from Linear regression model

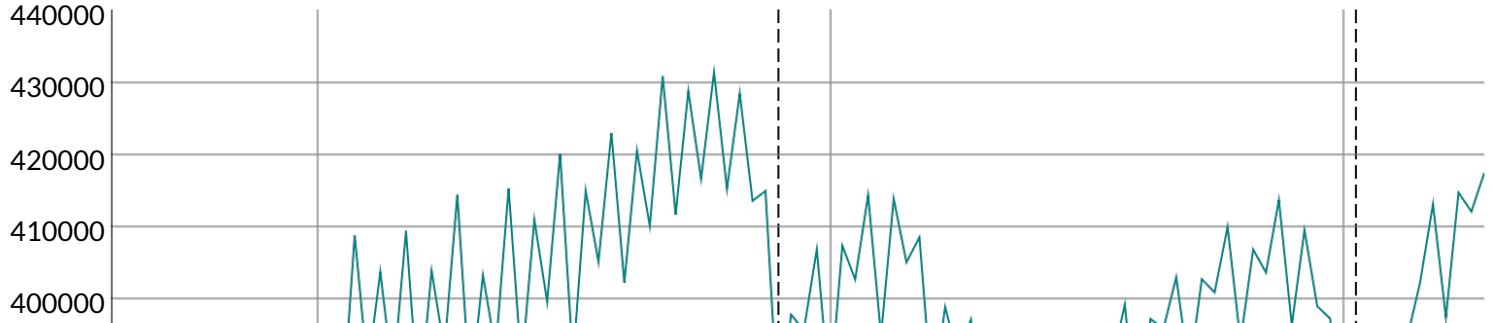


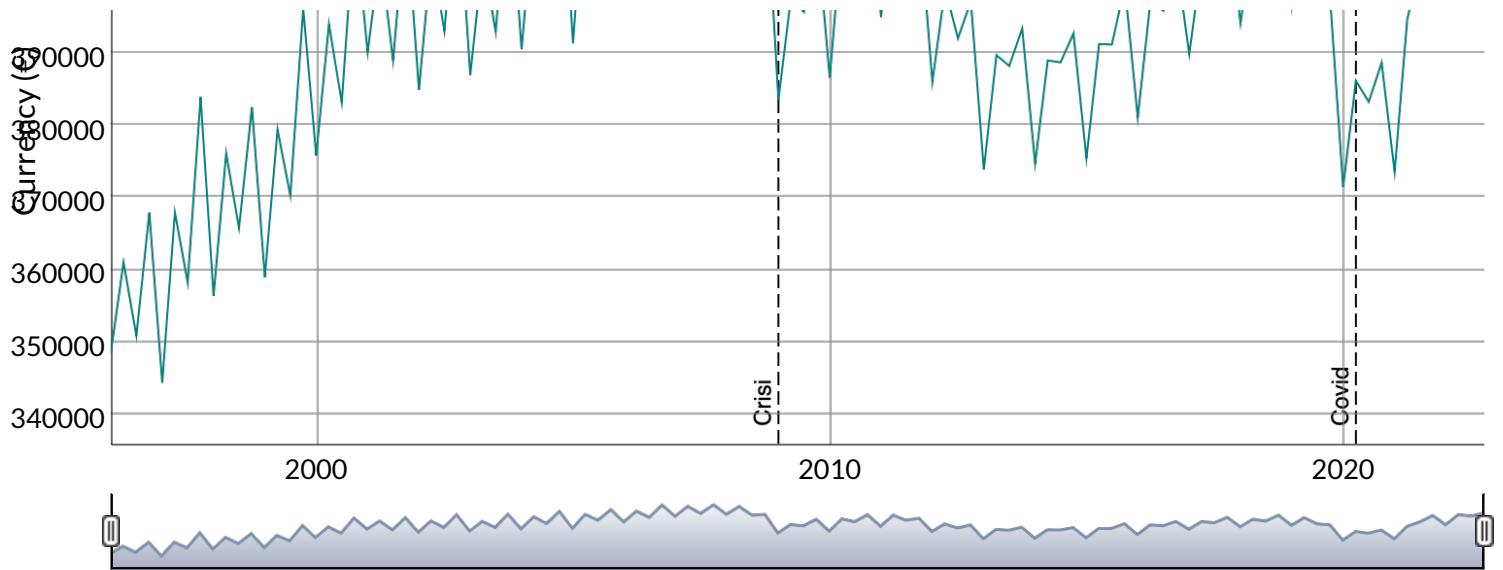
```
## 
## Breusch-Godfrey test for serial correlation of order up to 15
## 
## data: Residuals from Linear regression model
## LM test = 43.828, df = 15, p-value = 0.000117
```

Il modello presenta ancora i residui correlati, ma il p-value del test di Breusch-Godfrey è aumentato, pur restando sotto il livello $\alpha = 0.05$. Si potrebbe, pertanto, pensare di aggiungere altri due noti, consapevoli del rischio di overfitting.

```
dyRangeSelector(dygraph(gdp_clean,
                         main="Real Gross Domestic Product",
                         ylab="Currency (€)") %>%
  dyEvent("2020-04-01", "Covid", labelLoc = "bottom") %>%
  dyEvent("2009-01-01", "Crisi", labelLoc = "bottom")
)
```

Real Gross Domestic Product





```
t <- time(gdp)
tau.2007 <- c(2007,4)
tau.2014 <- c(2014,1)
tau.2018 <- c(2018,4)
tau.2020 <- c(2020,3)
tau1 <- ts(pmax(0, t - tau.2007), start = 1996)
tau2 <- ts(pmax(0, t - tau.2014), start = 1996)
tau3 <- ts(pmax(0, t - tau.2018), start = 1996)
tau4 <- ts(pmax(0, t - tau.2020), start = 1996)
mod_pw2 <- tslm(gdp ~ t + tau1 + tau2 + tau3 + tau4 + exp + lab + cons + cpi + cap + season)
summary(mod_pw2)
```

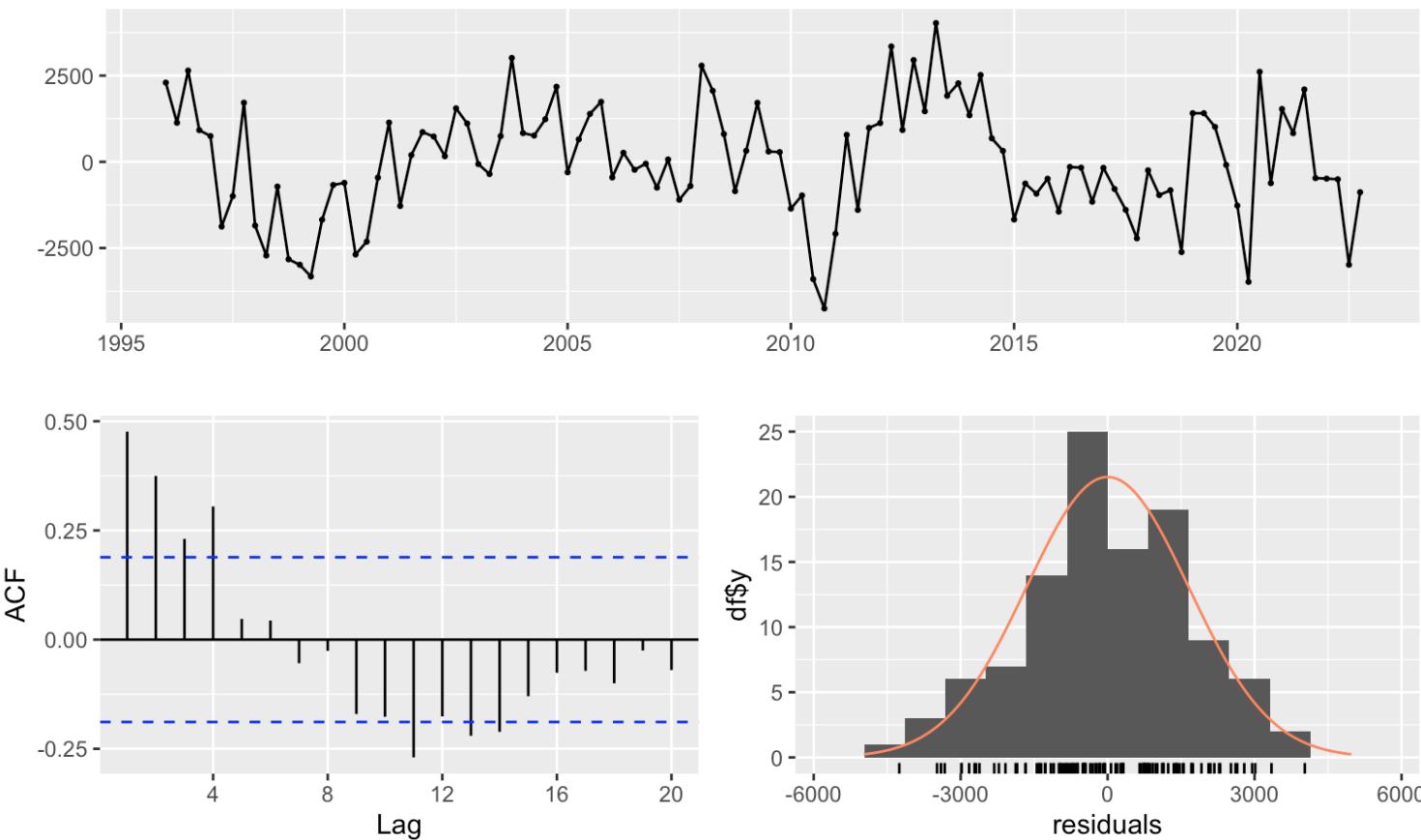
```

## 
## Call:
## tslm(formula = gdp ~ t + tau1 + tau2 + tau3 + tau2 + exp + lab +
##       cons + cpi + cap + season)
## 
## Residuals:
##      Min      1Q Median      3Q     Max 
## -4252.4 -966.2 -77.6 1126.5 4025.0 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1.498e+06 3.727e+05  4.020 0.000116 ***
## t          -7.238e+02 1.833e+02 -3.949 0.000151 *** 
## tau1        3.008e+02 2.754e+02  1.092 0.277495  
## tau2        -6.924e+02 3.936e+02 -1.759 0.081773 .  
## tau3        2.251e+02 2.609e+02  0.863 0.390465  
## exp         5.533e-01 4.425e-02 12.505 < 2e-16 ***
## lab         -9.249e+00 1.812e+02 -0.051 0.959391  
## cons        7.882e-01 2.746e-02 28.707 < 2e-16 *** 
## cpi         -3.299e+02 3.513e+02 -0.939 0.350055  
## cap         5.359e-01 3.646e-02 14.696 < 2e-16 *** 
## season2    3.421e+05 2.615e+05  1.308 0.193975  
## season3    4.200e+03 6.169e+02  6.809 8.84e-10 *** 
## season4    3.435e+05 2.616e+05  1.313 0.192409  
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1759 on 95 degrees of freedom
## Multiple R-squared:  0.9925, Adjusted R-squared:  0.9915 
## F-statistic: 1044 on 12 and 95 DF,  p-value: < 2.2e-16

```

```
checkresiduals(mod_pw2)
```

Residuals from Linear regression model



```
##  
## Breusch-Godfrey test for serial correlation of order up to 16  
##  
## data: Residuals from Linear regression model  
## LM test = 43.329, df = 16, p-value = 0.0002492
```

Il modello presenta ancora i residui correlati, ma il p-value del test di Breusch-Godfrey è aumentato, pur restando sotto il livello $\alpha = 0.05$.

Bisogna fare un'importante precisazione: la correlazione nei residui non rende le previsioni distorte, quindi queste non sono sbagliate, rende solo gli intervalli di previsione più ampi e quindi la stima più incerta.

```
rbind("lineare multiplo" = CV(mod),  
      "lineare a tratti v.1" = CV(mod_pw),  
      "lineare a tratti v.2" = CV(mod_pw2))
```

```
##  
## lineare multiplo      CV      AIC     AICc      BIC     AdjR2  
## lineare multiplo    4203785 1641.201 1643.951 1670.705 0.9901983  
## lineare a tratti v.1 3757857 1627.032 1630.904 1661.899 0.9915435  
## lineare a tratti v.2 3799479 1628.189 1632.705 1665.739 0.9915209
```

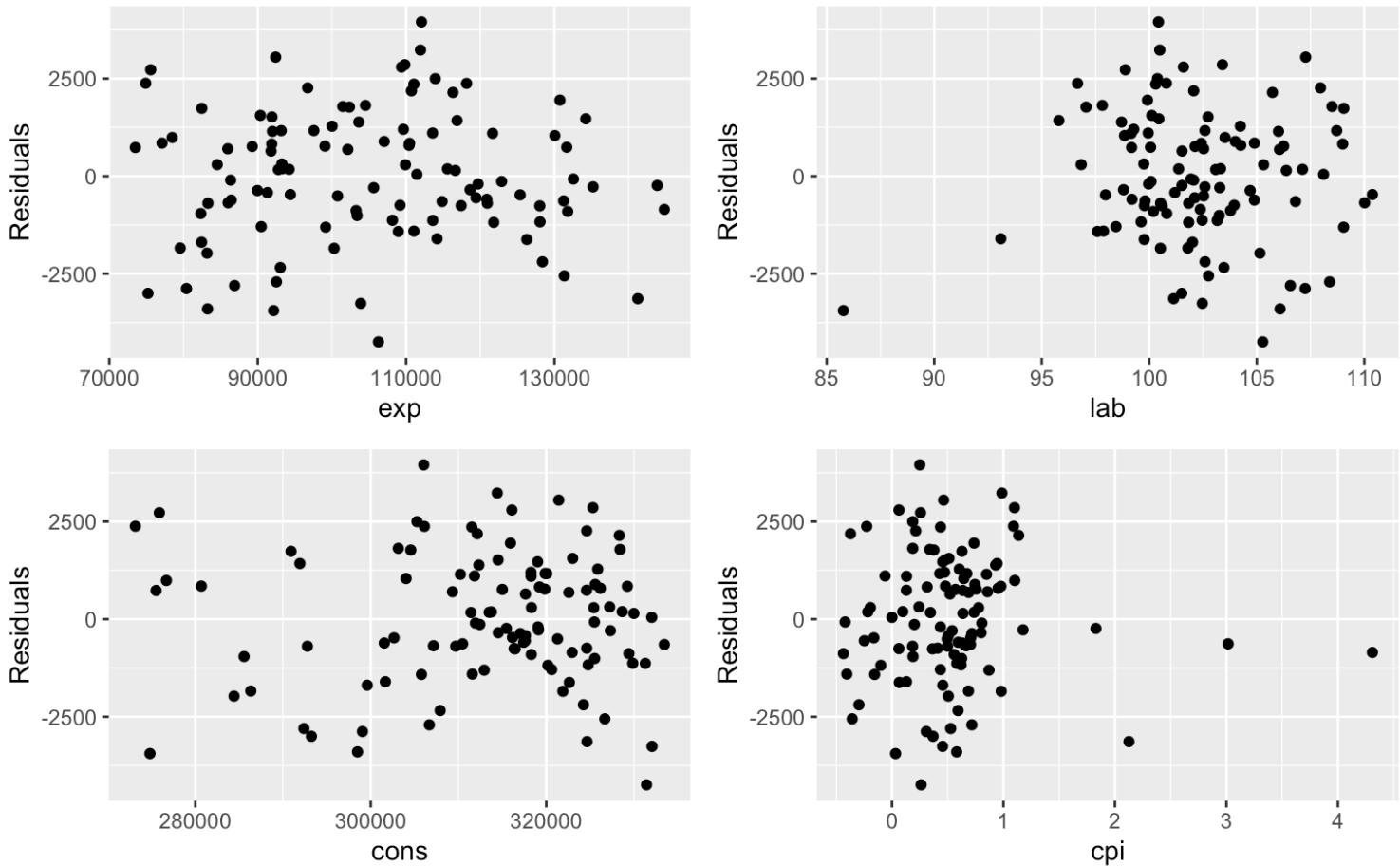
Siccome in termini di accuratezza previsiva tutti e tre i modelli presentati offrono ottime performance, il modello prescelto è il più semplice tra i due modelli lineari a tratti.

Plot dei residui

Per assicurarsi che il modello lineare sia adatto, si può verificare che lo scatter dei residui non mostri alcun pattern.

```
df <- as.data.frame(data)
```

```
df[, "Residuals"] <- as.numeric(residuals(mod_pw))
p1 <- ggplot(df, aes(x=exp, y=Residuals)) +
  geom_point()
p2 <- ggplot(df, aes(x=lab, y=Residuals)) +
  geom_point()
p3 <- ggplot(df, aes(x=cons, y=Residuals)) +
  geom_point()
p4 <- ggplot(df, aes(x=cpi, y=Residuals)) +
  geom_point()
gridExtra::grid.arrange(p1, p2, p3, p4, nrow=2)
```



Al netto di alcuni valori anomali, i residui sembrano essere randomly scattered.

Previsioni

Su test set

È possibile utilizzare il modello di regressione lineare multiplo costruito `mod` per fare previsioni e valutare la bontà previsiva del modello. Per i motivi di cui sopra, il modello viene addestrato sulle osservazioni dal 1996 al 2016 e testato sulle osservazioni dal 2017 al 2019 della serie storica originale `gdp`.

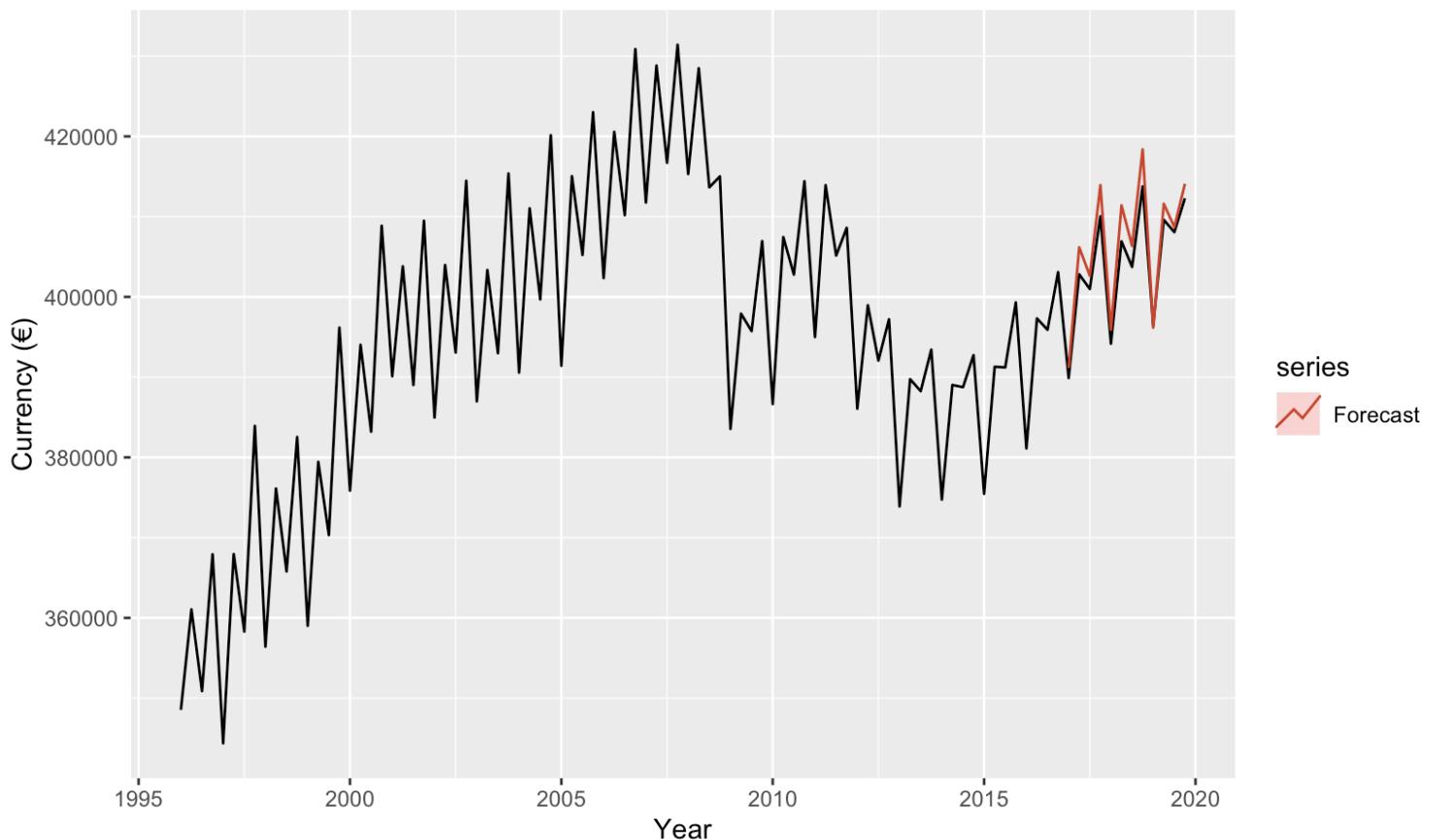
```
nc_train <- ts(data[1:84,], start=1996, frequency = 4)
nc_test <- as.data.frame(ts(data[85:96,], start=2017, frequency = 4))
```

```
mod_nc <- tslm(gdp ~ exp + lab + cons + cpi + cap + trend + season, data=nc_train)
```

```
fcast_mod <- forecast(mod_nc, nc_test)
```

```
autoplot(gdp_nc) +
  autolayer(fcast_mod, series="Forecast", PI=FALSE) +
  xlab("Year") +
  ylab("Currency (€)") +
  ggtitle("Real Gross Domestic Product")
```

Real Gross Domestic Product



```
accuracy(fcast_mod, test_nc)[2,c(2,3,5,6)]
```

##	RMSE	MAE	MAPE	MASE
##	2736.1123039	2360.1509437	0.5810829	0.3388581

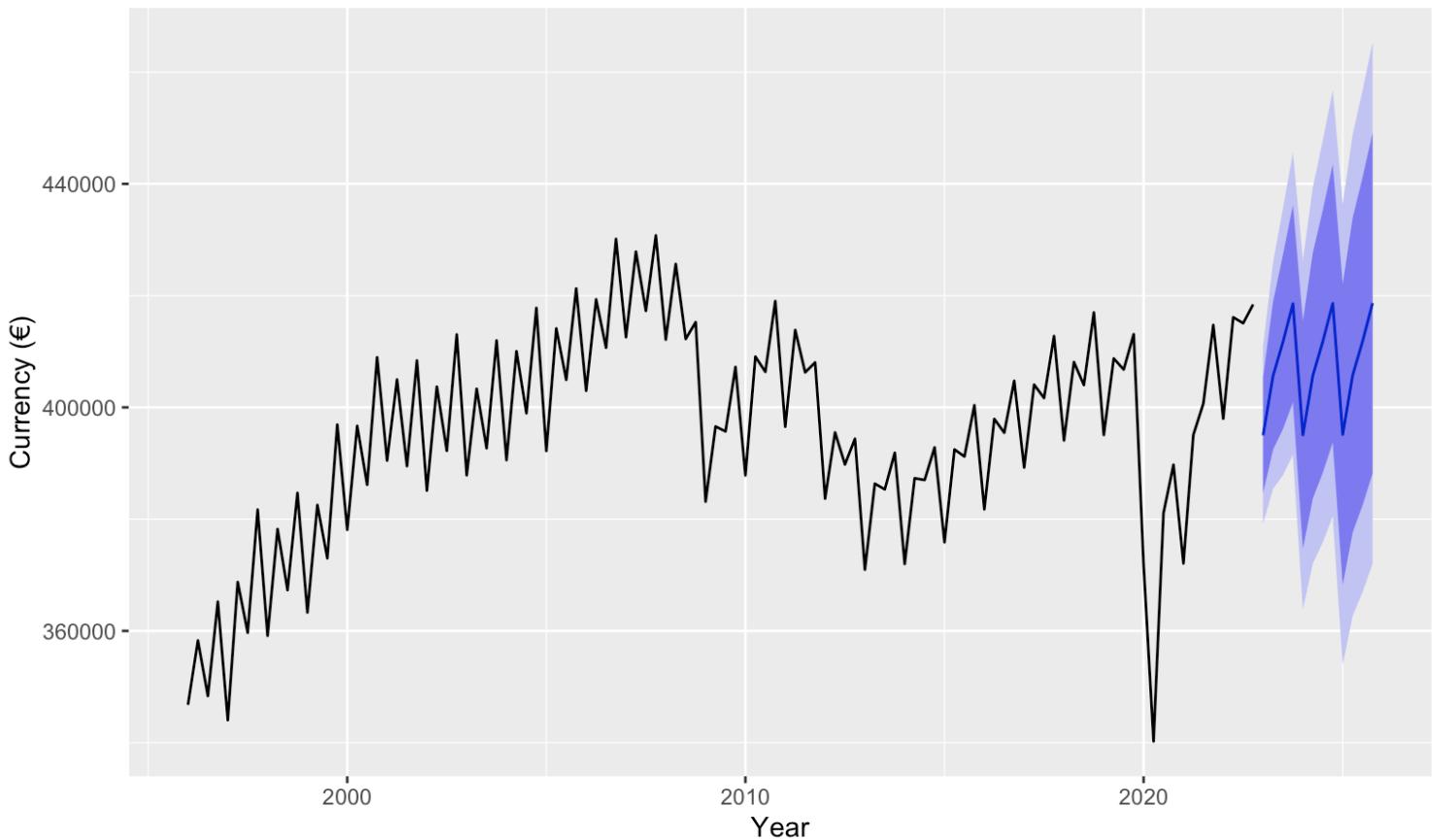
Sul futuro

Possiamo, inoltre, utilizzare il modello così costruito per fare previsione sui valori futuri di GDP, previsioni che non possiamo testare.

```
fcast_mod_new <- forecast(mod$fitted.values, h=12)
```

```
autoplot(fcast_mod_new) +  
  ggtitle("Forecasting") +  
  xlab("Year") +  
  ylab("Currency (€)")
```

Forecasting



Grazie!

Appendice

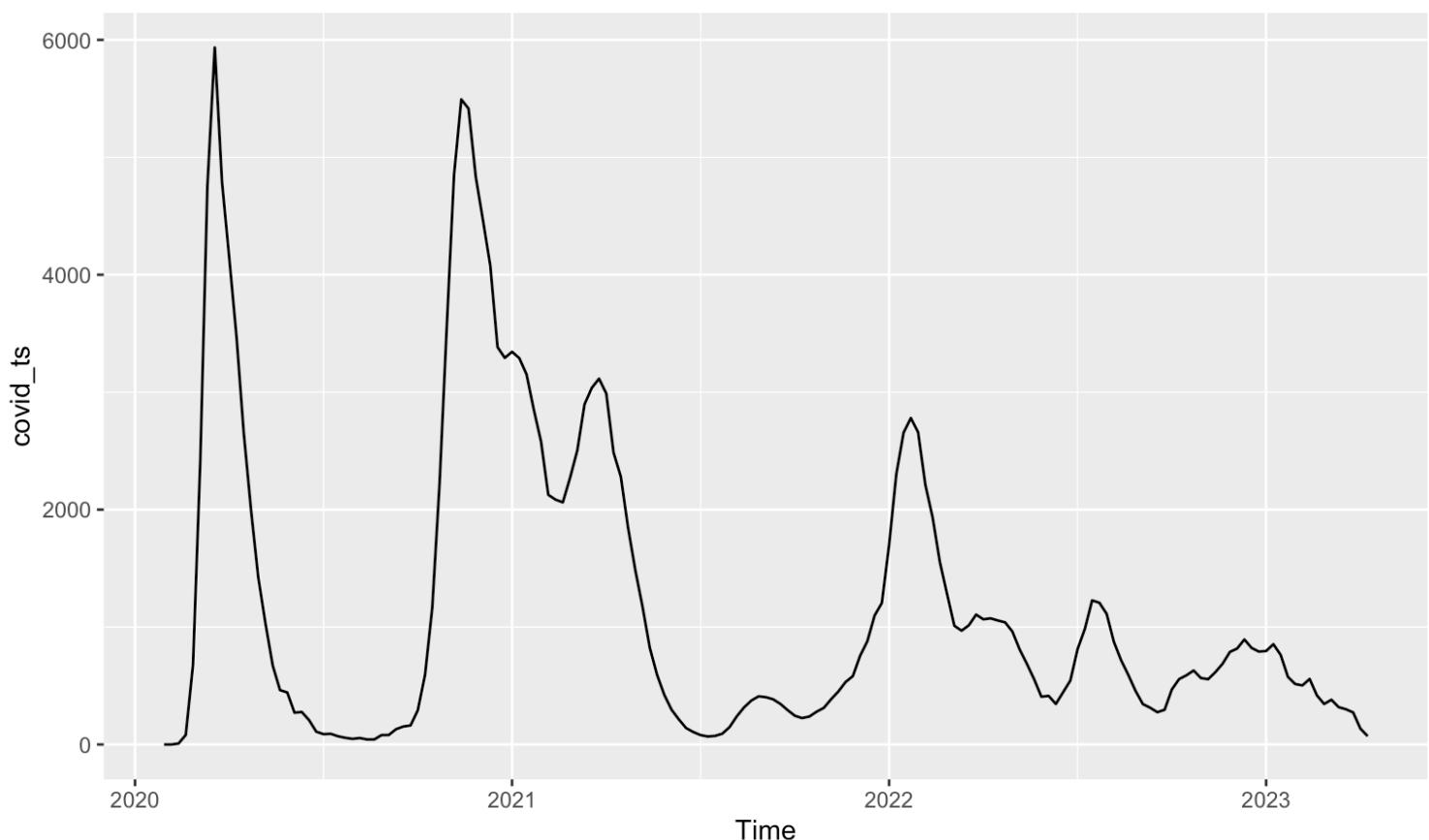
Di seguito vengono riportati alcuni dati relativi ai decessi causati dal Covid-19.

I dati sono solo stati solo visulizzati a scopo esplorativo al fine di avere una visione migliore su come il Covid abbia potuto influenzare il GDP.

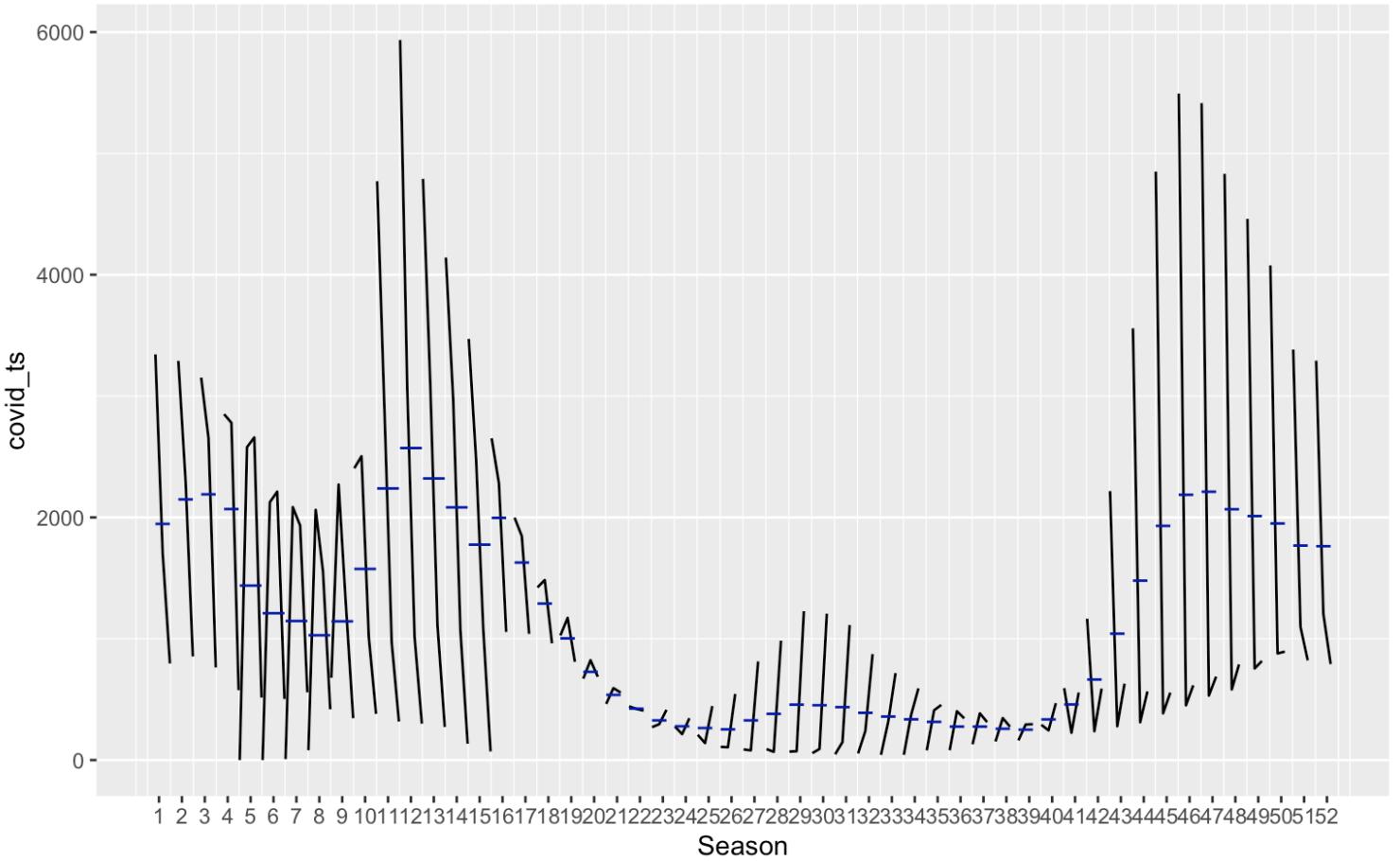
Non è stata fatta modellazione con questi dati in quanto aggregando la serie da settimanale a trimestrale, come la serie GDP oggetto di studio, le osservazioni erano così poche da non permettere di fare inferenza.

```
covid <- read.csv("/Users/mariapiatedesco/Downloads/HEALTH_MORTALITY_11052023211912210.csv") %>%  
  select(c(WEEK, YEAR, Value)) %>%  
  arrange(YEAR, WEEK)
```

```
covid_ts <- ts(covid[,3], start=c(2020,5), frequency = 52)  
autoplot(covid_ts)
```



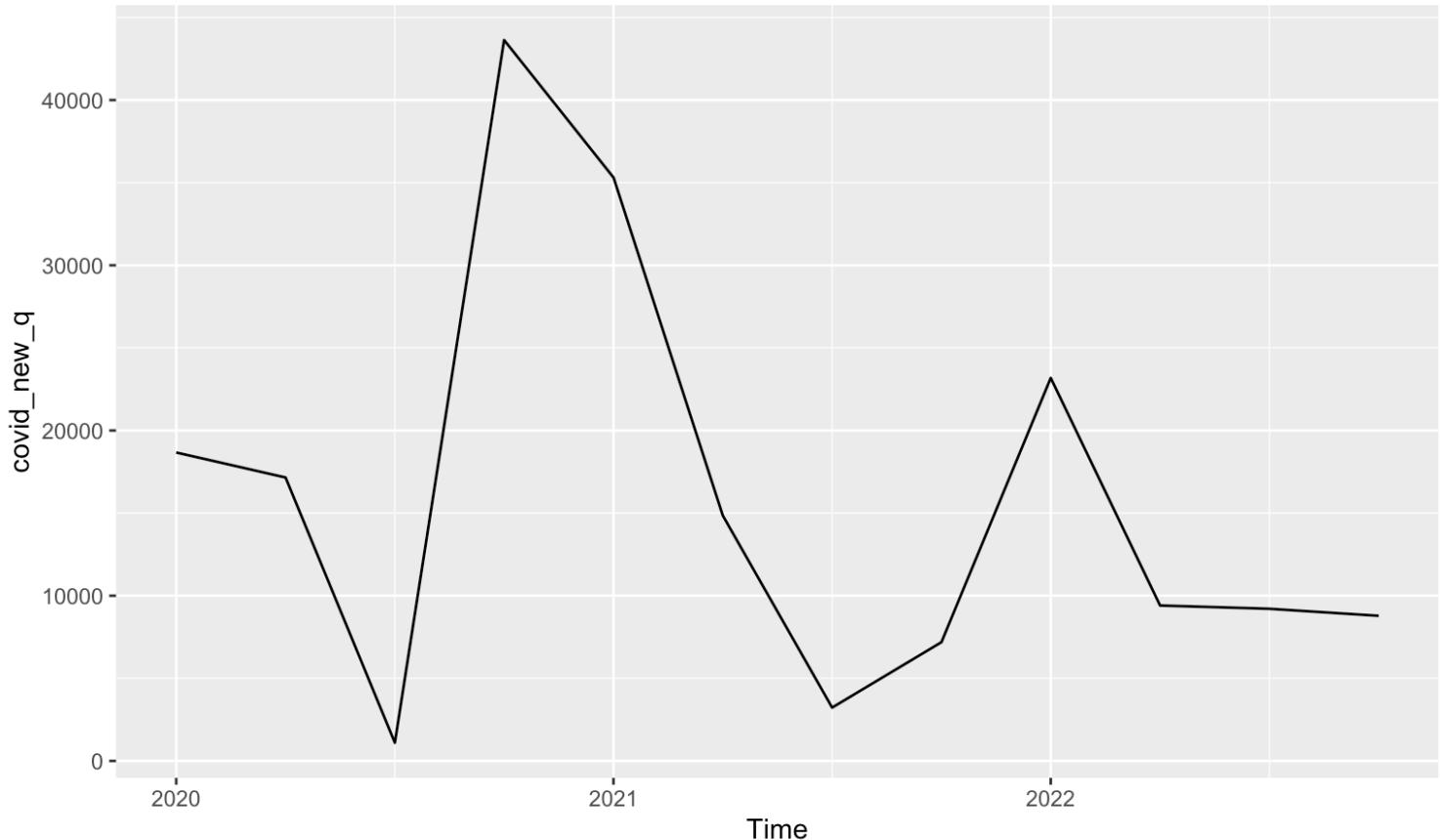
```
ggsubseriesplot(covid_ts)
```



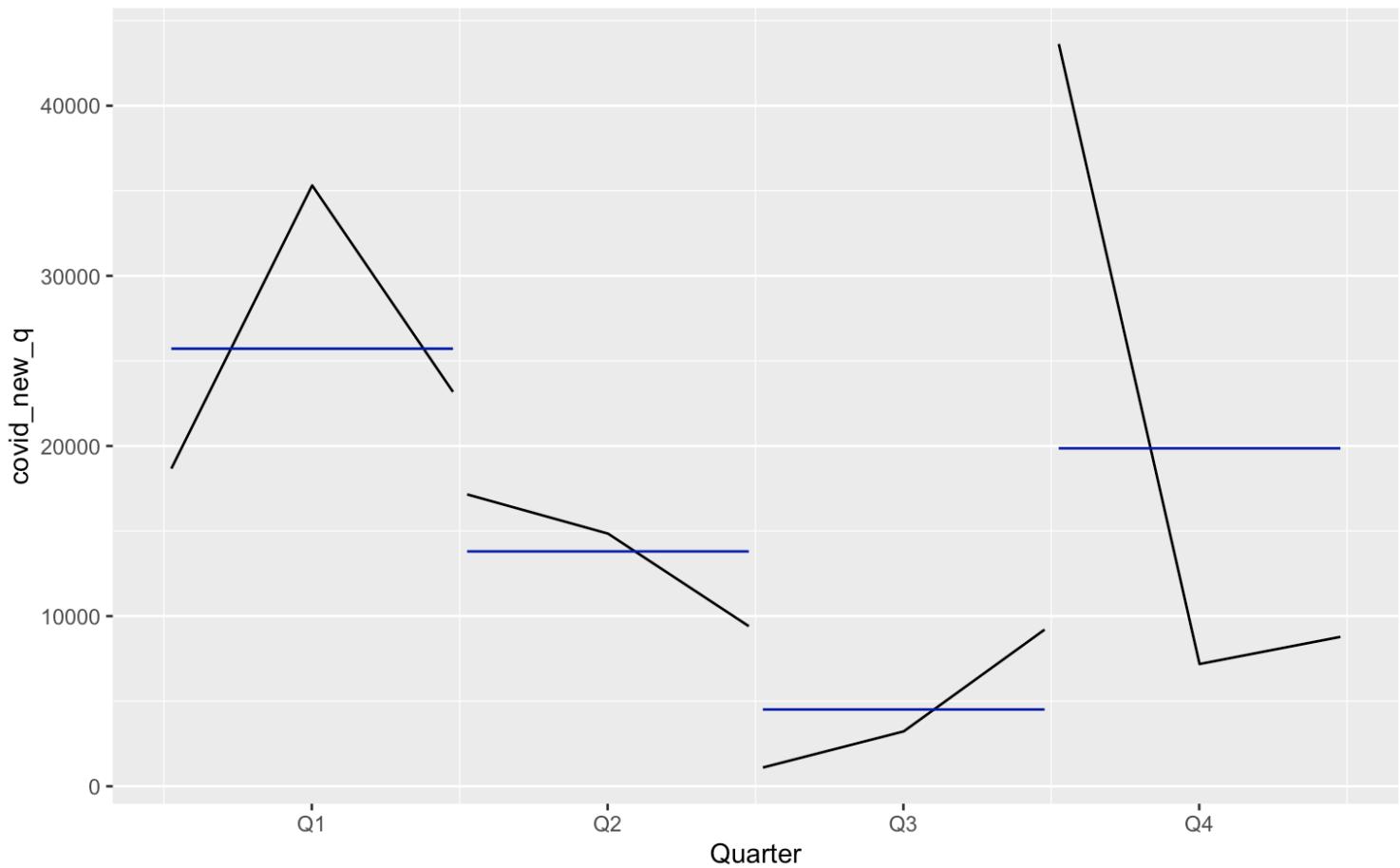
```
covid_new <- covid %>%
  add_row(WEEK=1, YEAR=2020, Value=0) %>%
  add_row(WEEK=2, YEAR=2020, Value=0) %>%
  add_row(WEEK=3, YEAR=2020, Value=0) %>%
  add_row(WEEK=4, YEAR=2020, Value=0) %>%
  filter(YEAR<2023) %>%
  arrange(YEAR, WEEK) %>%
  select(Value)
```

```
covid_new_ts <- ts(covid_new, start=c(2020,1), frequency = 52)
covid_new_q <- aggregate(covid_new_ts, nfrequency = 4)
```

```
autoplot(covid_new_q)
```



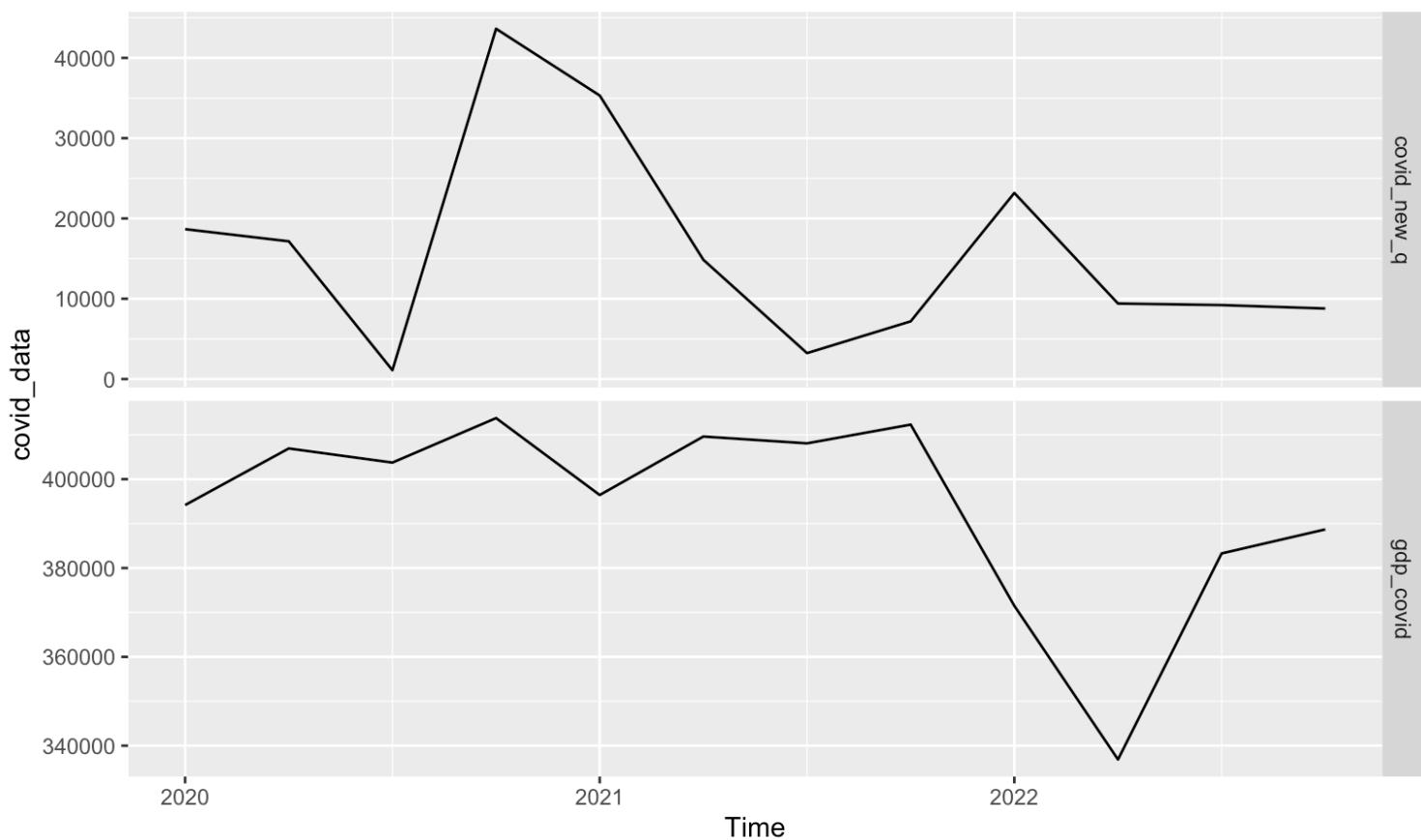
```
ggsubseriesplot(covid_new_q)
```



```
gdp_covid <- ts(gdp[89:100], start=c(2020,1), frequency = 4)
```

```
covid_data <- cbind(covid_new_q, gdp_covid)
```

```
autoplot(covid_data, facets=TRUE)
```



```
cor.test(gdp_covid, covid_new_q)
```

```
##  
## Pearson's product-moment correlation  
##  
## data: gdp_covid and covid_new_q  
## t = 0.44283, df = 10, p-value = 0.6673  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.4728543 0.6600501  
## sample estimates:  
## cor  
## 0.1386816
```