

# Modelování ekonomických časových řad

Semestrální práce do předmětu  
Modely ekonomických časových řad

## Obsah

Úvod .....	3
Analýza jednorozměrných nestacionárních nesezónních časových řad .....	3
Analýza dvojrozměrné časové řady .....	11

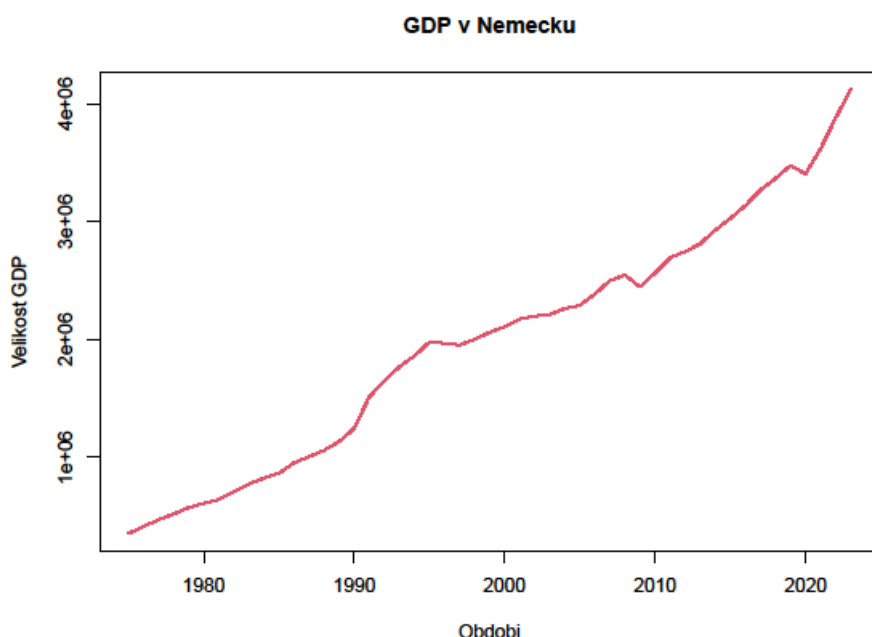
# Úvod

Budeme modelovat ekonomickou časovou řadu HDP v Německu. Data ve formátu Excel a skript v jazyce R budou přiloženy v příloze. Pracujeme s knihovnami v jazyce R jako jsou *readxl* pro načtení excelového souboru, *tseries* pro analýzu časových řad, *forecast* pro předpovídání a *urca* pro jednorozměrné testy časových řad.

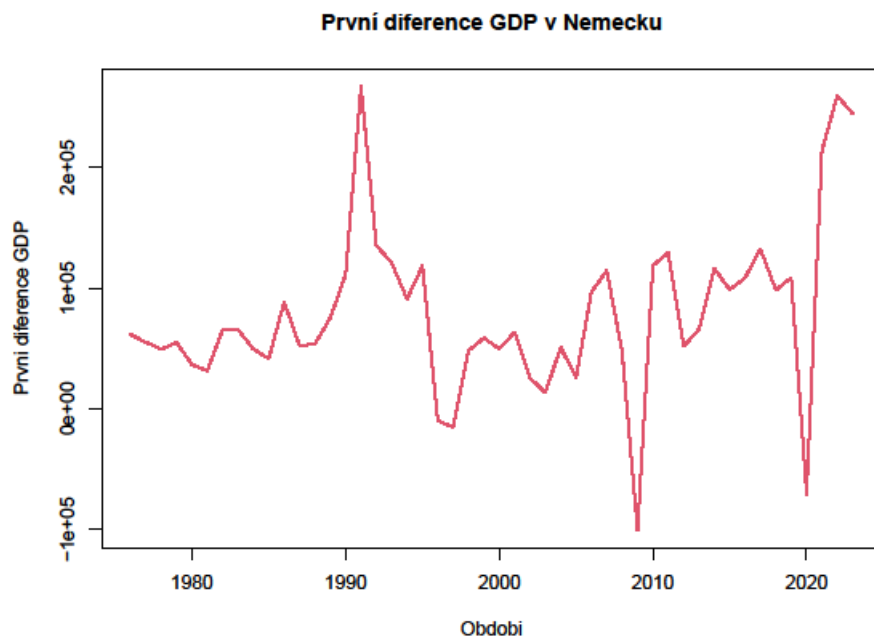
V druhé části se budeme věnovat časové řadě o dvou rozměrech: HDP v ČR a Mzdy a platy v ČR. Načítáme knihovny *tseries*, *forecast*, *readxl*, *urca*, *vars*.

## Analýza jednorozměrných nestacionárních nesezónních časových řad

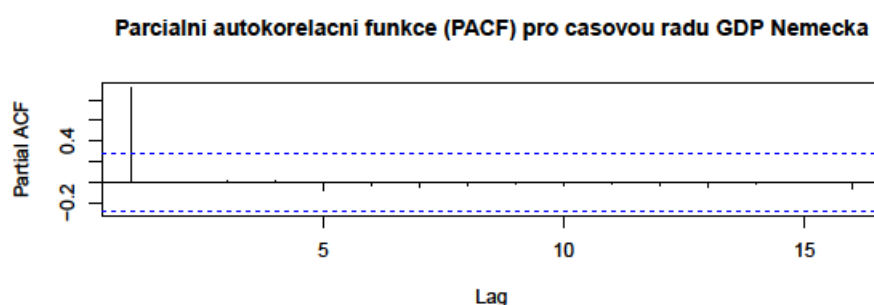
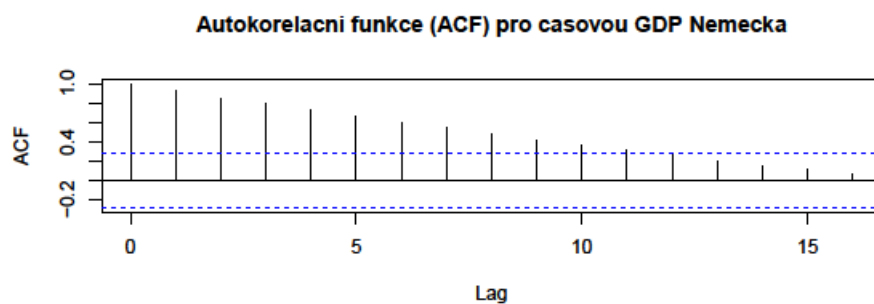
Data načítáme ze souboru *project\_data.xlsx* a vybíráme list *Sheet1*. Odstraňujeme řádky obsahující chybějící hodnoty. Převedeme sloupec *GDP* na numerický formát, aby bylo možné s ním pracovat jako s číselnými hodnotami. Vytváříme časovou řadu GDP z dat s počátečním bodem určeným minimálním časovým indexem (rok 1975) a frekvencí 1. Řadu si vizualizujeme:



Dále si zobrazíme první diferenci řady. První diference se často používají k transformaci nestacionárních řad na stacionární formu, což usnadňuje analýzu a predikci.



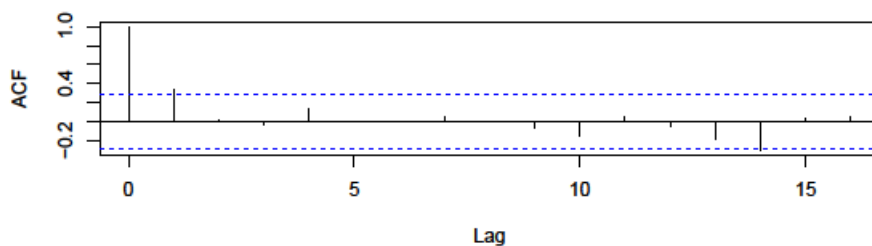
Dále porovnáváme autokorelační funkci (ACF) a parciální autokorelační funkci (PACF).



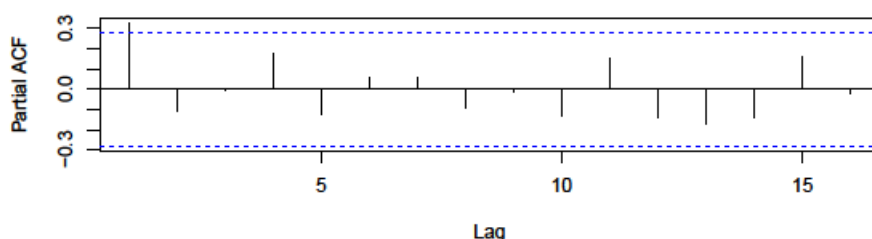
Z vizualizace lze vyčíst, že se jedná o typický průběh ACF a PACF pro autoregresní proces prvního řádu  $AR(1)$  ve tvaru  $Y_t = \phi_1 Y_{t-1} + a_t$ , kde  $\phi \in (0,1)$ .

Zobrazení ACF a PACF pro první diference časové řady GDP může poskytnout informace o autokorelaci či stacionaritě v upravených datech.

**Autokorelační funkce (ACF) pro časovou řadu prvních diferencí GDP Německa**



**Parcialní autokorelační funkce (PACF) pro časovou řadu prvních diferencí GDP Německa**



Vidíme, že jsou první difference časové řady stacionární, což může naznačovat, že byly úspěšně odstraněny trendy nebo sezónní vzory.

Vbudovanými testy v jazyce R na jednotkový kořen, jako jsou Augmented Dickey-Fuller, Phillips-Perron a KPSS (*funkcemi* `adf.test()`, `pp.test()`, `kpss.test()` v R) si numericky ověříme stacionaritu řad. Výsledek ADF testu pro neupravenou časovou řadu GDP ukázal Dickey-Fuller statistiku rovnu -0.98619 s p-hodnotou 0.9311. Vzhledem k tomu, že p-hodnota je vyšší než běžně používaná hladina významnosti 0.05, nemůžeme zamítnout nulovou hypotézu, a tedy nemáme dostatek důkazů pro to, abychom tvrdili, že neupravená časová řada je stacionární. Výsledek PP testu pro neupravenou časovou řadu GDP ukázal Dickey-Fullerovu statistiku -3.7118 s p-hodnotou 0.8996. Opět, pokud bychom zvolili hladinu významnosti 0.05, nedostáváme se do oblasti zamítnutí nulové hypotézy o přítomnosti jednotkového kořene. Výsledek KPSS testu pro neupravenou časovou řadu GDP ukázal testovací statistiku 1.3106 s p-hodnotou 0.01. Vzhledem k tomu, že p-hodnota je menší než běžně používaná hladina významnosti 0.05, zamítáme nulovou hypotézu o stacionaritě, což naznačuje, že neupravená časová řada není stacionární. Třemi testy jsme si tedy utvrdili, že řada je nestacionární.

Provedeme testy i na diferencovaných datech, abychom se ujistili, že jsme dosáhli požadované stacionarity po diferencování. Provedli jsme ADF test na prvních diferencích časové řady. Výsledek ukázal Dickey-Fuller statistiku -1.9401 s p-hodnotou 0.598. P-hodnota je vysoká, což znamená, že nemůžeme zamítnout nulovou hypotézu o přítomnosti jednotkového kořene, ale na druhou stranu to naznačuje, že první difference jsou stacionární. Dále jsme provedli PP test na prvních diferencích časové řady. Výsledek ukázal Dickey-Fuller statistiku -29.985 s p-hodnotou 0.01. Opět, i když p-hodnota je nízká, což naznačuje stacionaritu, nemůžeme jednoznačně zamítnout nulovou hypotézu. Po provedení KPSS testu výsledek ukázal testovací statistiku 0.28237 s p-hodnotou 0.1. Tato p-hodnota je vyšší než obvyklá hladina významnosti 0.05, což naznačuje, že neexistují dostatečné

důkazy pro zamítnutí nulové hypotézy o stacionaritě. Všemi testy jsme si utvrdili, že diferencovaná řada je stacionární.

Po provedení testů na diferencovaných datech je běžným postupem navrhnout vhodný model ARIMA pomocí funkce *auto.arima()*. Tato funkce automaticky vybírá model ARIMA na základě informací z analýzy časové řady.

Model pro neupravenou časovou řadu (GDP):

```
## Series: GDP
## ARIMA(1,1,0) with drift
##
## Coefficients:
##          ar1      drift
##      0.3668  80250.23
## s.e.  0.1420  14647.07
##
## sigma^2 = 4.389e+09: log likelihood = -600.02
## AIC=1206.04  AICc=1206.58  BIC=1211.65
```

Model pro první diferenci časové řady (GDP):

```
## Series: diff(GDP)
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1      mean
##      0.3668  80250.23
## s.e.  0.1420  14647.38
##
## sigma^2 = 4.389e+09: log likelihood = -600.02
## AIC=1206.04  AICc=1206.58  BIC=1211.65
```

Oba navržené modely mají stejné hodnoty koeficientů AR(1) a driftu, což je očekávatelné, protože jsme pracovali s prvními diferencemi nebo s neupravenými daty. Tato shoda naznačuje, že první diference dosáhly stacionarity a můžeme očekávat, že modely budou mít podobné výsledky při predikci.

Tím, že je první diference stacionární, provedeme odhad ARIMA modelu naší časové řady GDP s pomocí funkce *Arima()*. Model jsme specifikovali s parametry `order=c(1,1,0)`, což znamená, že se jedná o autoregresní model prvního řádu s jedním diferencováním.

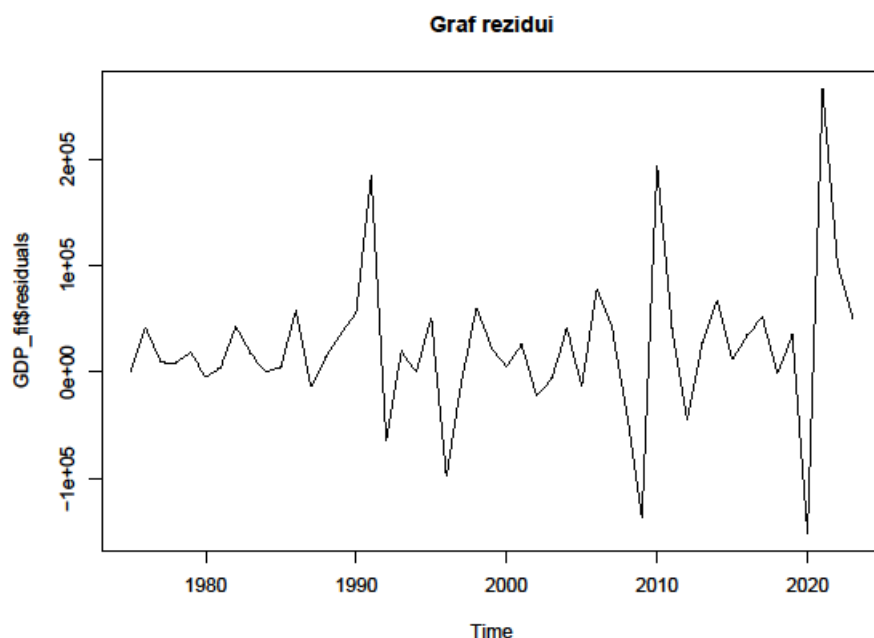
```
GDP_fit <- Arima(GDP, order=c(1,1,0))
summary(GDP_fit)
```

```
## Series: GDP
## ARIMA(1,1,0)
##
## Coefficients:
##          ar1
##      0.7455
## s.e.  0.1027
##
```

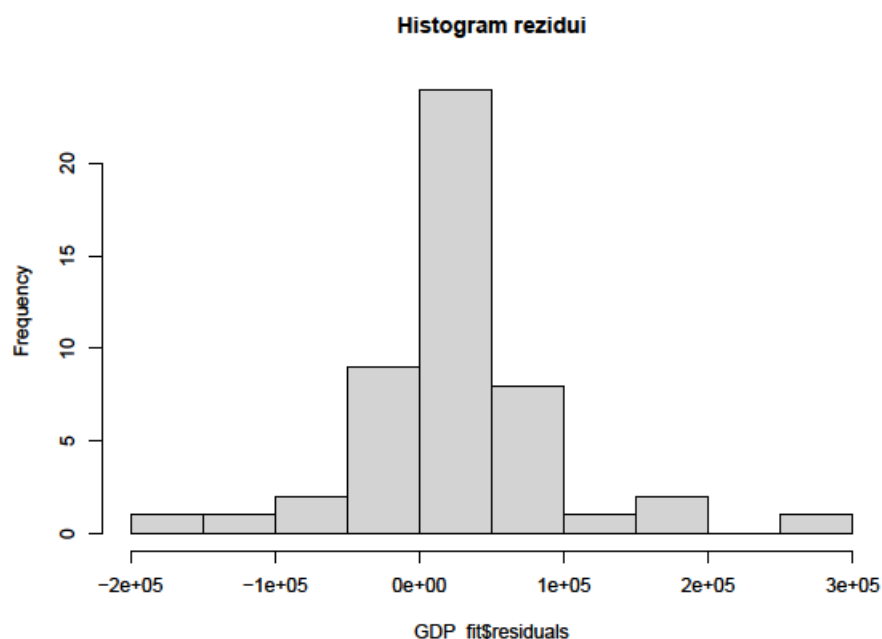
```
## sigma^2 = 5.441e+09: log likelihood = -606.02
## AIC=1216.04 AICc=1216.31 BIC=1219.79
##
## Training set error measures:
##           ME      RMSE      MAE      MPE      MAPE      MASE      A
CF1
## Training set 22875.48 72240.91 47256.26 1.441206 2.510436 0.545264 -0.23
217
```

Následně jsme vypsal shrnutí tohoto modelu pomocí funkce *summary()*. Z tohoto shrnutí můžeme vyčíst například, že se skutečně jedná o autoregresní model prvního řádu s jedním diferencováním. Odhadovaný koeficient autoregrese (*ar1*) je 0.7455 se standardní chybou 0.1027. Hodnoty AIC a BIC jsou docela nízké, což naznačuje, že model by měl být docela vhodný, avšak nemáme jiné modely, s jakými jej porovnat.

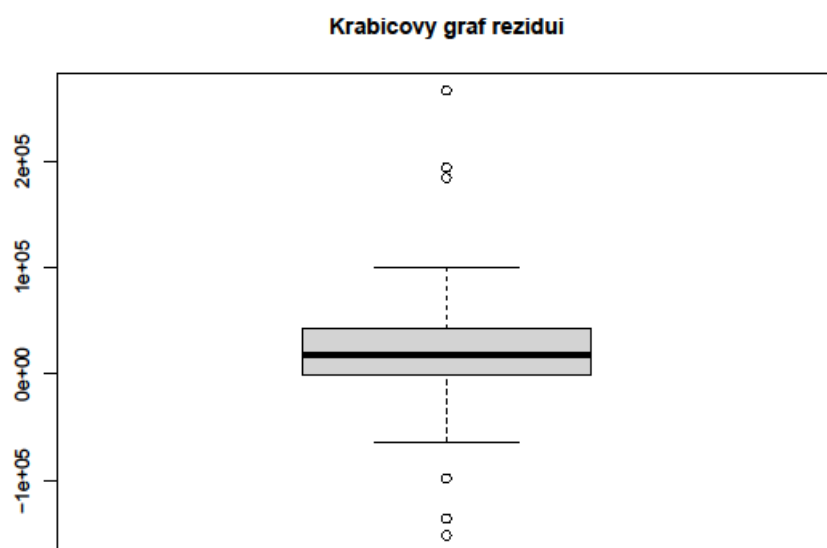
Nyní provedeme verifikaci modelu, prvně za pomoci grafů. Vizualizujeme si rezidua modelu:



Rezidua sice vykazují nějaké výkyvy, avšak ty nejsou pravidelné. Model tedy bere v úvahu sezónnost časové řady. Necháme si vykreslit histogram reziduí.

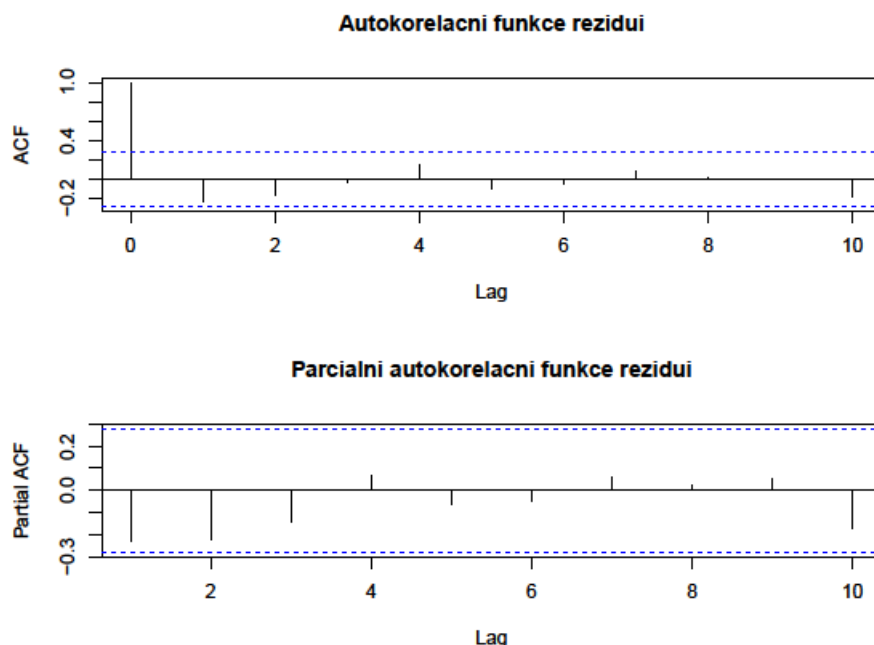


A zjišťujeme, že se rezidua řídí normálním rozdělením. Necháváme si vykreslit boxplot.



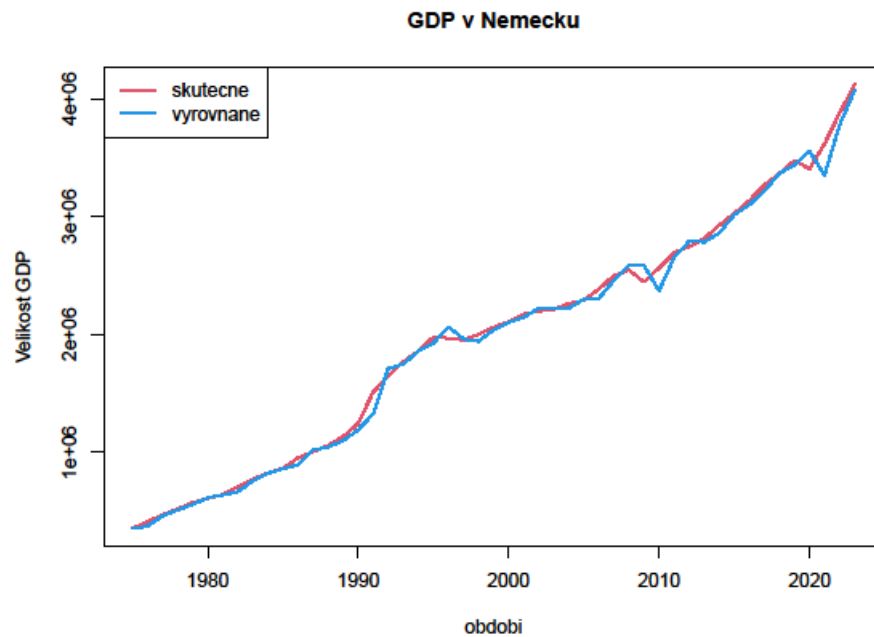
Medián reziduí se pohybuje okolo nuly. Graf je celkově symetrický nevyjímaje horního a dolního kvartilů. Nicméně můžeme pozorovat pár odlehlých pozorování. Provedeme vizualizaci autokorelační funkce (ACF) a parciální autokorelační funkce (PACF) reziduí z odhadnutého ARIMA modelu.



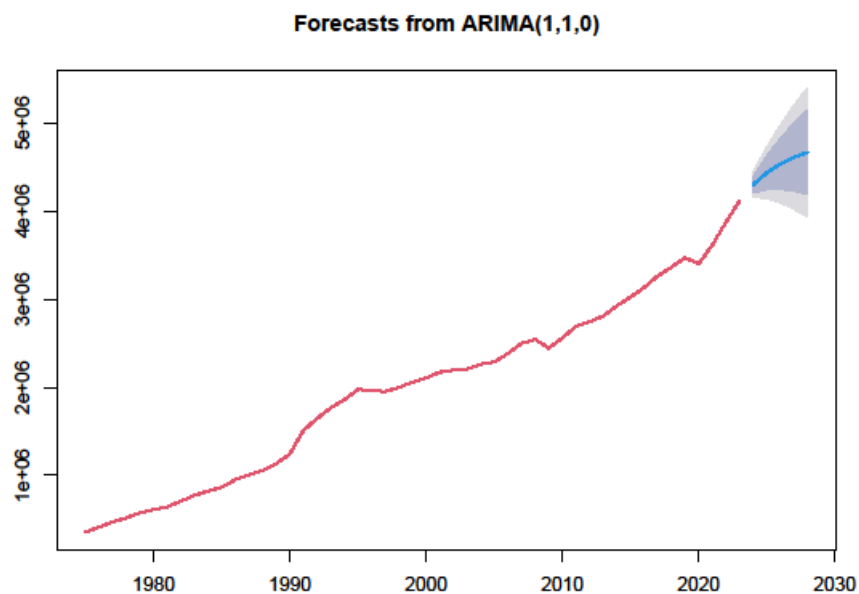


Vidíme, že je to bílý šum. Nyní provedeme na reziduích 3 testy; Box-Ljungův test, který se používá k testování nezávislosti reziduí. V našem případě je p-hodnota 0.5605, což naznačuje, že nemáme dostatek důkazů k zamítnutí nulové hypotézy nezávislosti reziduí na hladině významnosti 0.05. Shapiro-Wilkův test, který se používá k testování normality dat. V našem případě je p-hodnota  $8.814 \times 10^{-5}$ , což je méně než 0.05, takže máme dostatek důkazů na zamítnutí nulové hypotézy o normalitě dat. Jarque-Bera test testuje normalitu dat, ale je citlivější na odchylky od normálního rozdělení ve tvaru a šikmosti. V našem případě je p-hodnota  $5.261 \times 10^{-7}$ , což je méně než 0.05, takže máme dostatek důkazů na zamítnutí nulové hypotézy o normalitě dat.

Dále si zobrazíme samotný model v porovnání s našimi daty.



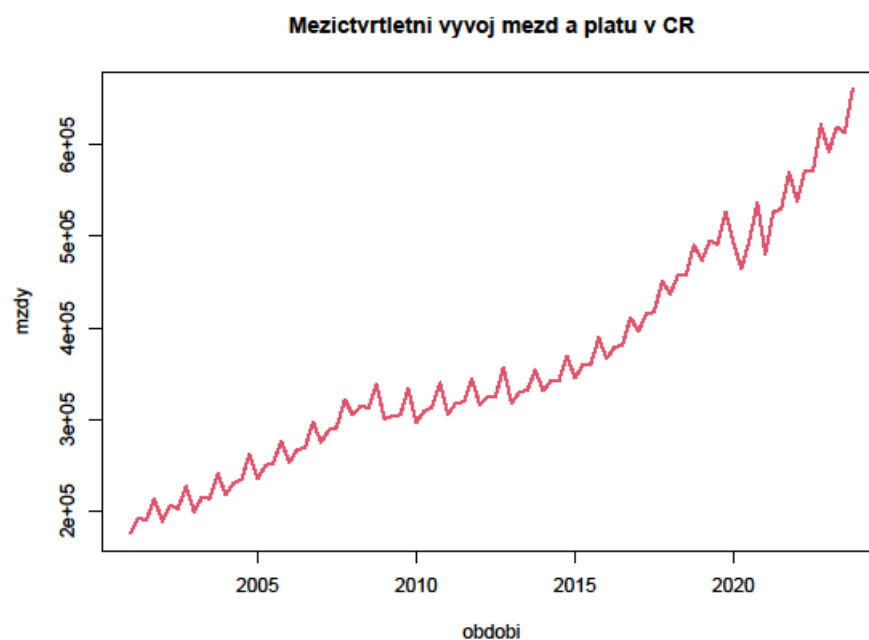
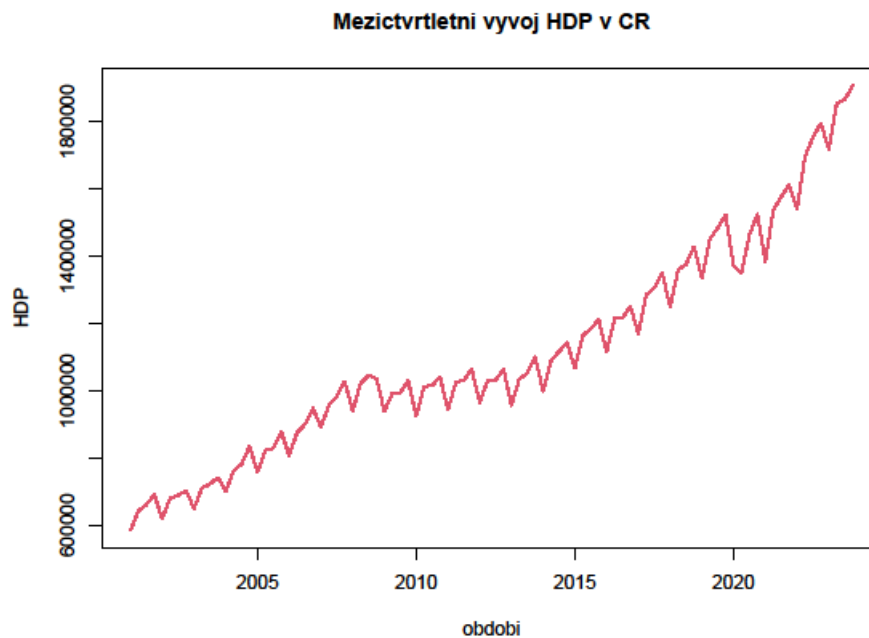
Vidíme, že na pohled model velmi dobře fituje naše data. A nakonec provedeme predikci na základě našeho modelu do následujících 5 let.



Predikce nám vyšla pozvolna rostoucí pro HDP v Německu pro všechny body v 95% intervalu spolehlivosti predikce.

# Analýza dvojrozměrné časové řady

Jak už bylo zmíněno, jedná se o dvě časové řady: HDP v ČR a Mzdy a platy v ČR. Obě řady mají počátek v roce 2001 a konec v roce 2023. Záznamy jsou kvartální (tedy máme 92 záznamů).



Již po vizualizaci je možné si všimnout, že obě řady jsou nestacionární a bude potřeba je diferencovat. Ověříme si to adf testem.

```
adf.test(HDP)
```

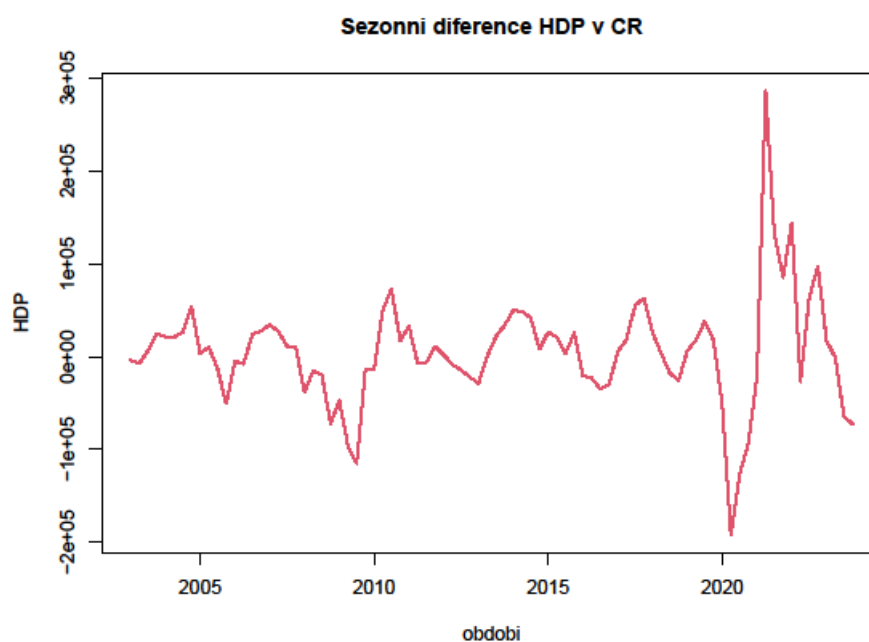
```
##
## Augmented Dickey-Fuller Test
##
## data: HDP
## Dickey-Fuller = -0.42144, Lag order = 4, p-value = 0.9834
## alternative hypothesis: stationary

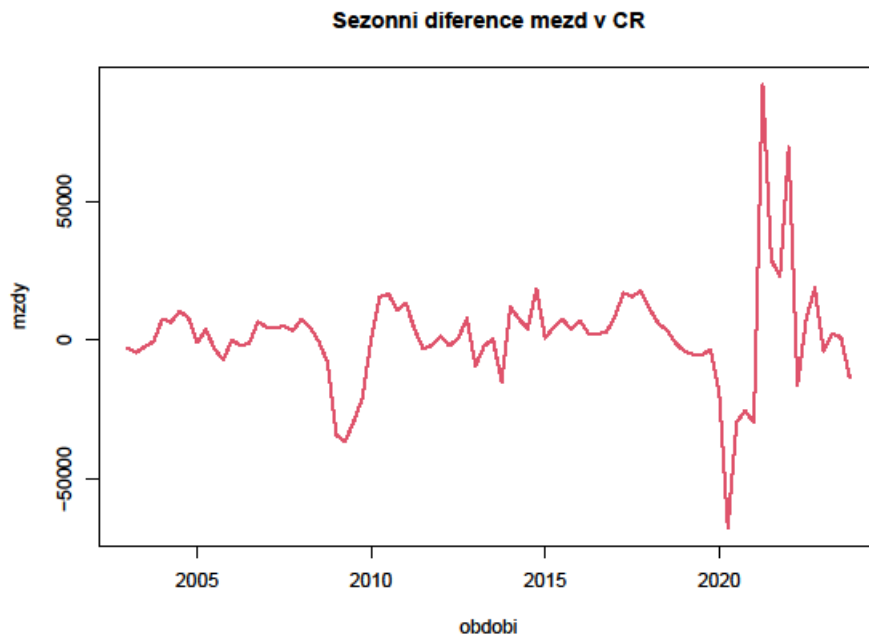
adf.test(mzdy)

## Warning in adf.test(mzdy): p-value greater than printed p-value

##
## Augmented Dickey-Fuller Test
##
## data: mzdy
## Dickey-Fuller = 0.16877, Lag order = 4, p-value = 0.99
## alternative hypothesis: stationary
```

V obou případech s p-hodnotami mnohem vyššími než hladina významnosti 0,05; není dostatek důkazů pro zamítnutí nulových hypotéz. Nulové hypotézy tedy nezamítáme, což naznačuje, že řady jsou nestacionární. Řady tedy sezónně diferencujeme.





Ověříme Phillips-Perron testem stacionaritu.

```
pp.test(HDP_diff) # Phillips-Perron test

##
##  Phillips-Perron Unit Root Test
##
## data:  HDP_diff
## Dickey-Fuller Z(alpha) = -23.941, Truncation lag parameter = 3, p-value
## = 0.02233
## alternative hypothesis: stationary

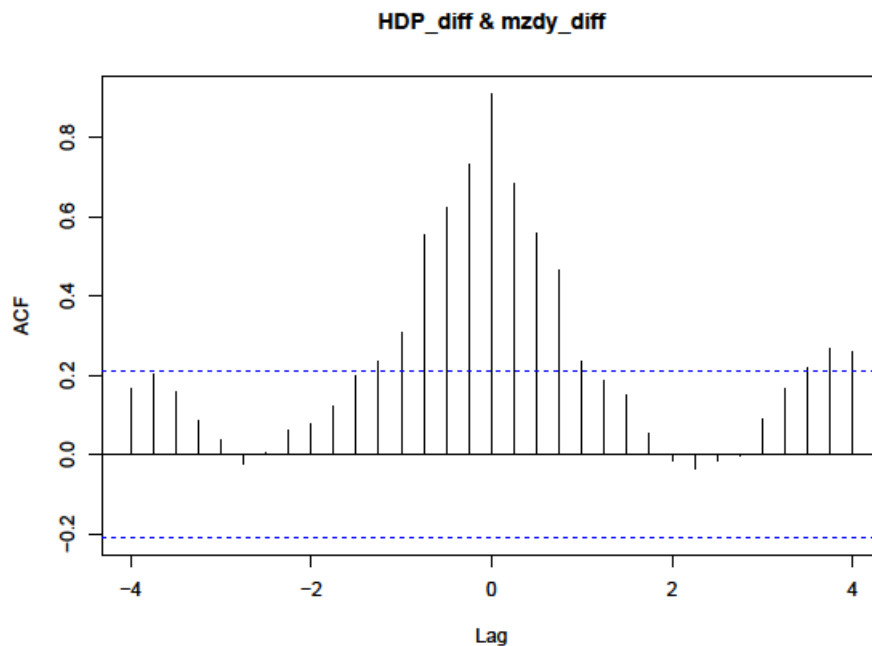
pp.test(mzdy_diff) # Phillips-Perron test

## Warning in pp.test(mzdy_diff): p-value smaller than printed p-value

##
##  Phillips-Perron Unit Root Test
##
## data:  mzdy_diff
## Dickey-Fuller Z(alpha) = -32.7, Truncation lag parameter = 3, p-value =
## 0.01
## alternative hypothesis: stationary
```

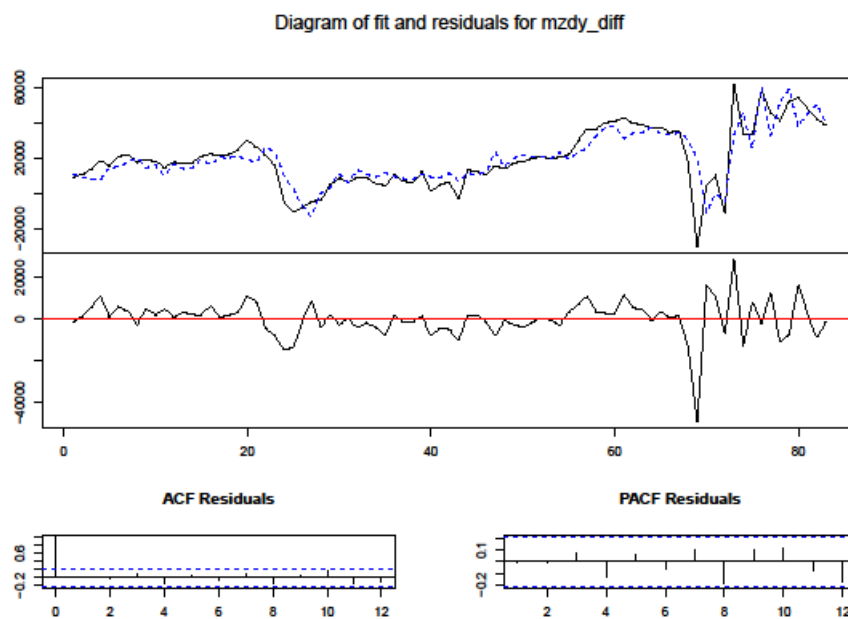
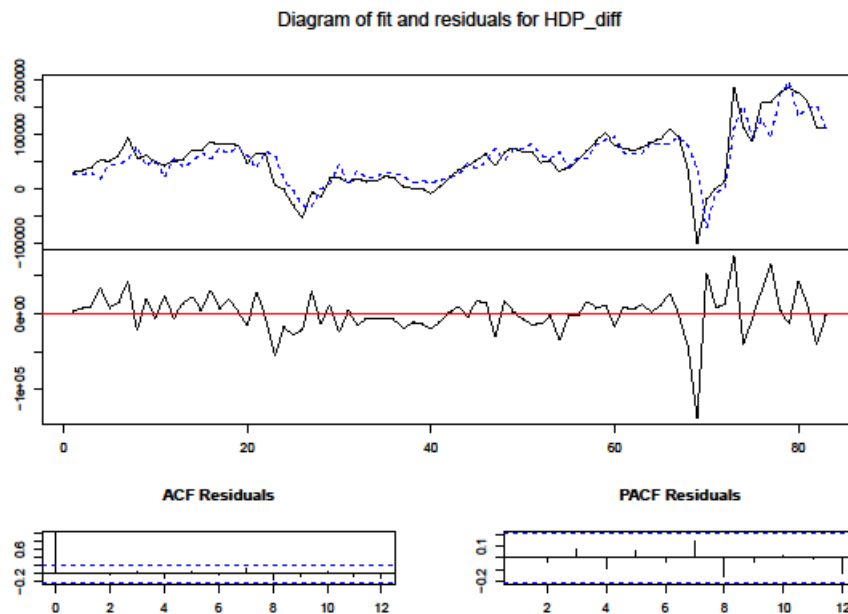
Pro obě proměnné zamítáme nulovou hypotézu a nestacionaritě a nezamítáme alternativní hypotézu na hladině významnosti 0,05.

Necháme si vykreslit křížový korelogram.



Vysoká hodnota křížové korelace při zpoždění (lag) 0, blízko k 0.8, naznačuje, že existuje silná lineární pozitivní závislost mezi změnami mezd a změnami hrubého domácího produktu v tom samém časovém okamžiku. Pokles hodnoty křížové korelace se zvětšujícím se zpožděním naznačuje, že změny v jedné proměnné mají tendenci se projevovat v druhé proměnné s určitým časovým zpožděním.

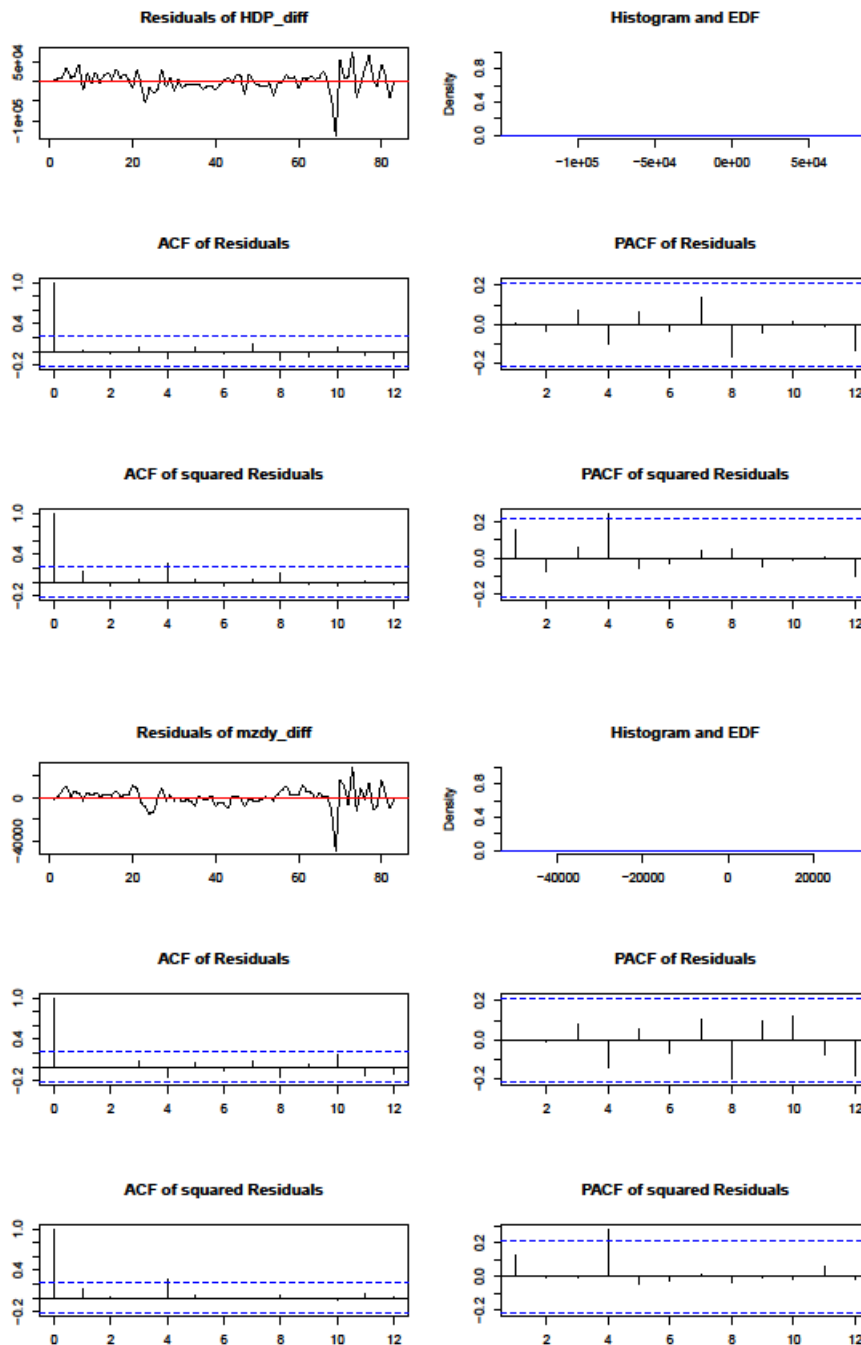
Nyní k výběru vhodného modelu pomocí funkce *VARselect* pro výběr modelu vektorové autoregrese (VAR) s použitím různých specifikací (konstantní, trend, obě, žádná) a jejich vyhodnocení na základě různých informačních kritérií (AIC, HQ, SC, FPE). Všechny specifikace vyhodnotily model řádu 5. Za pomoci funkce *VAR* a diagnostik modelů (Portmanteauva testu, analýzy reziduí, ...) se rozhodujeme finálně pro model s trendem. Model s trendem a jeho diagnostiky si vypíšeme/vykreslíme



Provedeme Portmanteau test

```
## Portmanteau Test (asymptotic)
##
## data: Residuals of VAR object hdp mzdy.odhad4
## Chi-squared = 41.424, df = 44, p-value = 0.5826
```

Protože p-hodnota (0,1824) je vyšší než běžná hladina významnosti 0,05, neexistuje dostatečný důkaz pro zamítnutí nulové hypotézy o neexistenci autokorelace v reziduích. V reziduích modelu VAR nebyla zjištěna žádná významná autokorelace.



Provedeme ARCH test (Autoregressive Conditional Heteroskedasticity), který pro vícerozměrná data posuzuje, zda v reziduích vícerozměrného modelu časové řady důkaz podmíněná heteroskedasticita.

```
## ARCH (multivariate)
##
## data: Residuals of VAR object hdpmzdy.odhad4
## Chi-squared = 39.262, df = 45, p-value = 0.7128
```

Protože p-hodnota (0,7128) je vyšší než běžná hladina významnosti 0,05, neexistuje dostatečný důkaz pro zamítnutí nulové hypotézy o neexistenci ARCH efektů v



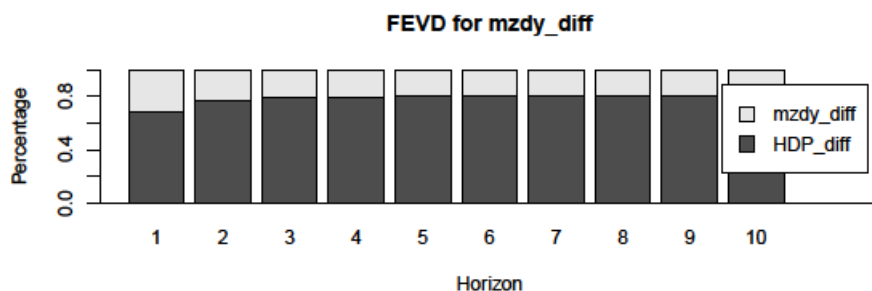
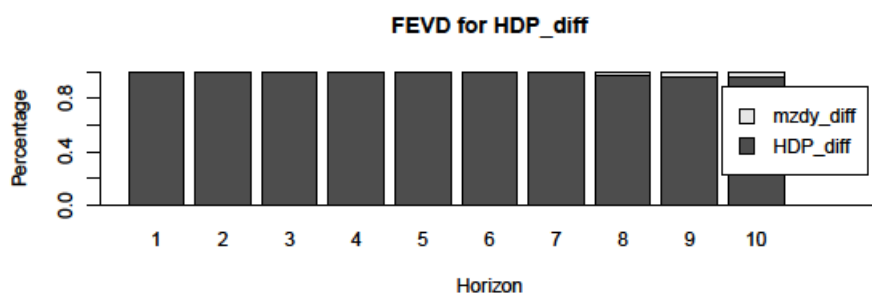
reziduích. Jinými slovy, v reziduích modelu VAR neexistuje žádný významný důkaz podmíněné heteroskedasticity.

Dále testujeme normalitu, šikmost a špičatost reziduí.

```
## $JB
##
## JB-Test (multivariate)
##
## data: Residuals of VAR object hdp mzdy.odhad4
## Chi-squared = 176.07, df = 4, p-value < 2.2e-16
##
## $Skewness
##
## Skewness only (multivariate)
##
## data: Residuals of VAR object hdp mzdy.odhad4
## Chi-squared = 28.644, df = 2, p-value = 6.025e-07
##
## $Kurtosis
##
## Kurtosis only (multivariate)
##
## data: Residuals of VAR object hdp mzdy.odhad4
## Chi-squared = 147.43, df = 2, p-value < 2.2e-16
```

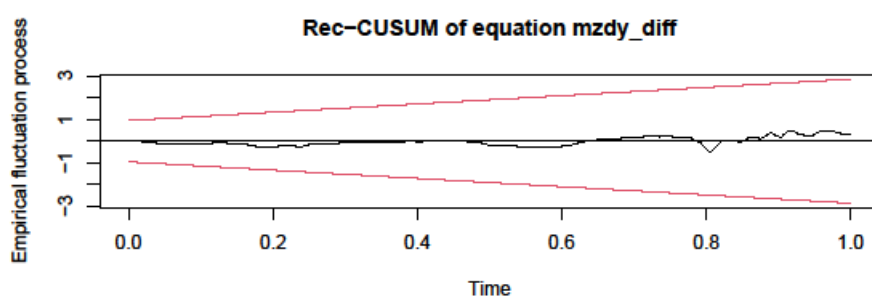
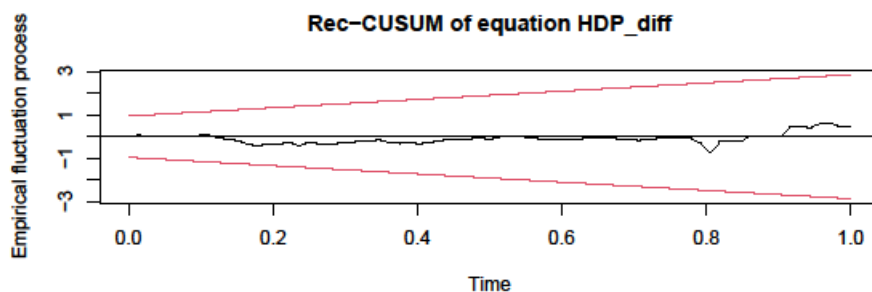
Nulové hypotézy zamítáme ve všech třech případech, tedy: rezidua nejspíše nejsou normálně rozdělena, mají nadměrnou šikmost i špičatost.

Nyní uděláme rozklad chyby předpovědi pro každou proměnnou ve VAR modelu při predikci.

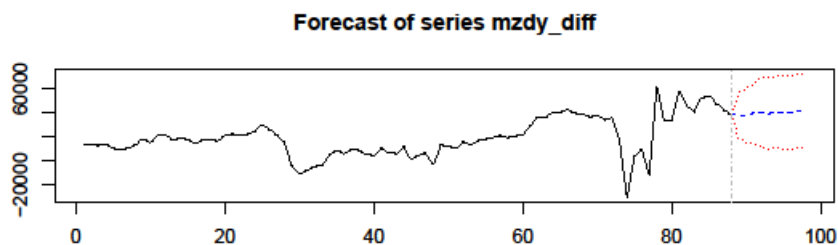
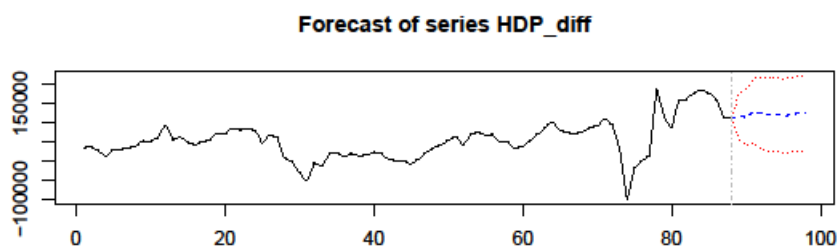


Proměnná HDP není skoro vůbec ovlivněna proměnnou mzdy. Naopak proměnná mzdy je ovlivněna z 68 % proměnnou HDP.

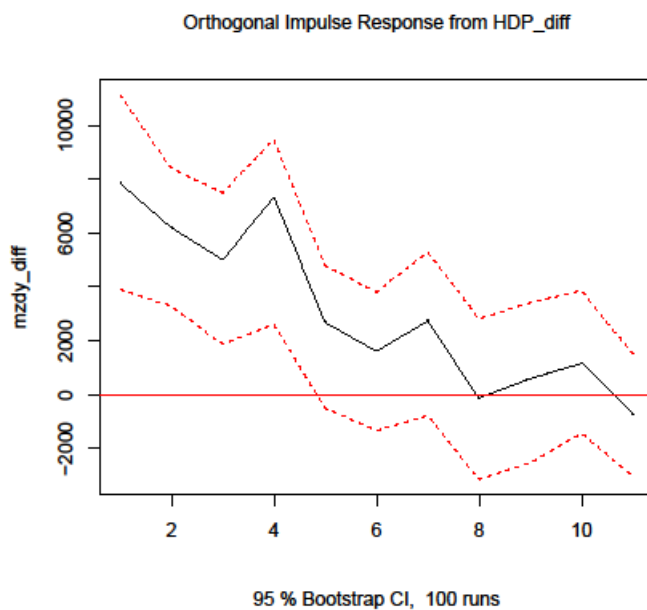
Dále si vykreslíme graf znázorňující stabilitu VAR modelu



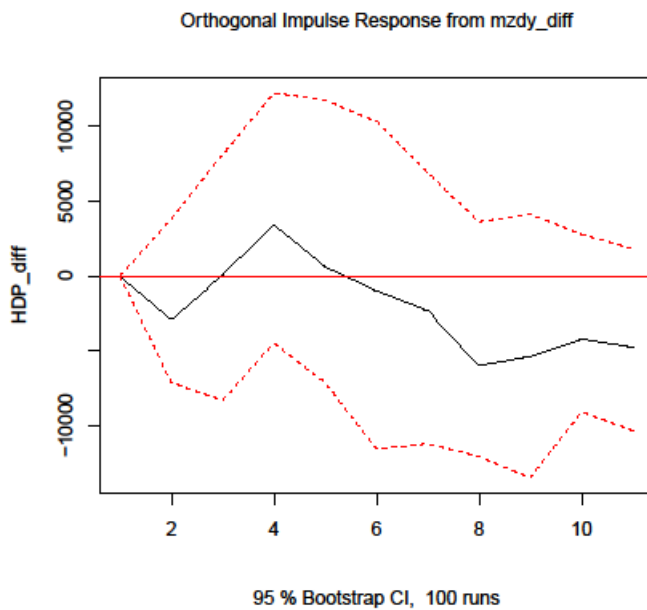
Statistiky se pohybují v rámci určitého rozmezí, model je tedy stabilní. Dále jsme si vykreslili predikci obou proměnných s 95% intervaly spolehlivosti



Nyní se přesuňme na impulse-response analýzu. Prvně si zobrazíme impulzní odezvu ve směru, kdy HDP je impuls a mzdy jsou odezva.



Celkově lze říci, že HDP má negativní vliv na mzdy a platy. Dále si zobrazíme impulzní odezvu ve směru, kdy mzdy jsou impuls a HDP je odezva.



V prvních čtyřech obdobích mají mzdy pozitivní dopad na HDP, naopak dále negativní. Avšak vidíme, že analýza je statisticky nevýznamná.

Nakonec provedeme testy na Grangerovu kauzalitu

```
causality(hdpmzdy.odhad4, cause="HDP_diff")
## $Granger
##
## Granger causality H0: HDP_diff do not Granger-cause mzdy_diff
```

```

##
## data: VAR object hdp mzdy.odhad4
## F-Test = 1.493, df1 = 5, df2 = 144, p-value = 0.1957
##
##
## $Instant
##
## H0: No instantaneous causality between: HDP_diff and mzdy_diff
##
## data: VAR object hdp mzdy.odhad4
## Chi-squared = 33.551, df = 1, p-value = 6.942e-09

causality(hdp mzdy.odhad4, cause="mzdy_diff")

## $Granger
##
## Granger causality H0: mzdy_diff do not Granger-cause HDP_diff
##
## data: VAR object hdp mzdy.odhad4
## F-Test = 0.89088, df1 = 5, df2 = 144, p-value = 0.489
##
##
## $Instant
##
## H0: No instantaneous causality between: mzdy_diff and HDP_diff
##
## data: VAR object hdp mzdy.odhad4
## Chi-squared = 33.551, df = 1, p-value = 6.942e-09

```

V obou případech nezamítáme nulovou hypotézu, tedy nelze říct, že by existovala Grangerova kauzalita mezi mzdami a HDP nebo naopak.

Naopak v obou případech Chí-kvadrát test ukazuje velkou statistickou signifikanci testu. To znamená, že máme dostatečné důkazy na to, abychom zamítli nulovou hypotézu a mohli přijmout alternativní hypotézu o existenci okamžité příčinnosti mezi mzdami a HDP a naopak. To znamená, že změny v jedné proměnné mají okamžitý vliv na změny v druhé proměnné.