Lecture 7, October 20, 2025

Forecasting VAR models

Marta Boczon

Department of Economics

Copnehagen Business School

mbo.eco@cbs.dk

In [23]:
```
#install.packages("quantmod")
#install.packages("fredr")
#install.packages("ggfortify")
#install.packages('urca')
#install.packages("dynlm")
#install.packages("stargazer")
#install.packages("pracma")
#install.packages("dLagM")
#install.packages("gets")
#install.packages("car")
#install.packages("lmtest")
#install.packages("vars")
options(warn=-1)
```

Roadmap

Today's lecture will follow an **exercise-based format** designed to combine theory, coding, and interpretation in a hands-on way. We'll complete **five structured exercises**, each taking approximately **20 minutes**, followed by a short **discussion and debrief**.

Goal for today

- Revisit and reinforce what we did **last week** (lag selection, identification, interpretation).

Expected outcomes

By the end of the session, you should be able to:

- Prepare macroeconomic time series for VAR analysis.
- Estimate both reduced-form VARs and identified SVARs.
- Generate and interpret **impulse response functions (IRFs)**.

In [2]:
```
# Import the dataset from a CSV file downloaded from FRED
cpi = read.csv("CPIAUCSL.csv")
urate = read.csv("UNRATE.csv")

# Display the first few observations to data check stzructure
head(cpi)
head(urate)

# Display the last few observations to data check structure
tail(cpi)
tail(urate)

# Convert the second column of the dataset into a time series object
# Exclude the last observation because the most recent value is "Pending" (NA after import)
tscpi = ts(cpi[1:(nrow(cpi)-1), 2], frequency = 4, start = c(1947, 1))
tsurate = ts(urate[1:(nrow(urate)-1), 2], frequency = 4, start = c(1948, 1))

# Define inflation
tsinf = diff(log(tscpi)) * 100

# Inflation is stationary in levels
dinf = tsinf

# Unemployment rate is stationary in levels
durate = tsurate
```

A data.frame: 6 × 2

	observation_date	CPIAUCSL
	<chr>	<dbl>
1	1947-01-01	21.700
2	1947-04-01	22.010
3	1947-07-01	22.490
4	1947-10-01	23.127
5	1948-01-01	23.617
6	1948-04-01	23.993

A data.frame: 6 × 2

	observation_date	UNRATE
	<chr>	<dbl>
1	1948-01-01	3.7
2	1948-04-01	3.7
3	1948-07-01	3.8
4	1948-10-01	3.8
5	1949-01-01	4.7
6	1949-04-01	5.9

A data.frame: 6 × 2

	observation_date	CPIAUCSL
	<chr>	<dbl>
310	2024-04-01	313.096
311	2024-07-01	314.183
312	2024-10-01	316.539
313	2025-01-01	319.492
314	2025-04-01	320.800
315	2025-07-01	NA

A data.frame: 6 × 2

	observation_date	UNRATE
	<chr>	<dbl>
306	2024-04-01	4.0
307	2024-07-01	4.2
308	2024-10-01	4.1
309	2025-01-01	4.1
310	2025-04-01	4.2
311	2025-07-01	NA

Exercise 1 (15')

For US real GDP:

1. Import the data.
2. Test for a unit root.
3. If necessary, transform it into a stationary series.

Solution to Exercise 1

In [3]:

```
# Import the dataset from a CSV file downloaded from FRED
gdp = read.csv("GDPC1.csv")

# Display the first few observations to data check stzructure
head(gdp)

# Display the last few observations to data check structure
tail(gdp)

# Convert the second column of the dataset into a time series object
tsgdp = ts(gdp[, 2], frequency = 4, start = c(1947, 1))
```

A data.frame: 6 × 2

	observation_date	GDPC1
	<chr>	<dbl>
1	1947-01-01	2182.681
2	1947-04-01	2176.892
3	1947-07-01	2172.432
4	1947-10-01	2206.452
5	1948-01-01	2239.682
6	1948-04-01	2276.690

A data.frame: 6 × 2

	observation_date	GDPC1
	<chr>	<dbl>
309	2024-01-01	23082.12
310	2024-04-01	23286.51
311	2024-07-01	23478.57
312	2024-10-01	23586.54
313	2025-01-01	23548.21
314	2025-04-01	23770.98

In [4]:

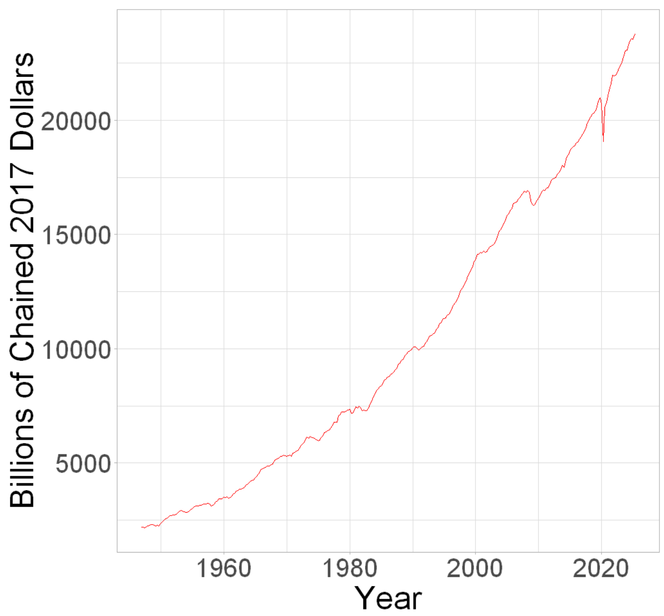
```
# Load required libraries
library(ggplot2)   # for plotting
library(ggfortify) # for autoplot of time series objects

# Set default figure size (width = 24, height = 8)
options(repr.plot.width = 24, repr.plot.height = 8)

# Create an initial time series plot of tspce in red
fig = autoplot(tsgdp, colour = 'red')

# Customize the plot
fig = fig +
  theme(aspect.ratio = 1) +                        # make plot square-shaped
  theme_light() +                                  # use a light theme
  theme(aspect.ratio = 1) +                        # reinforce square aspect ratio
  theme(plot.margin = ggplot2::margin(0.2, 0.2, 0.2, 0.2, "cm")) + # adjust margins
  theme(text = element_text(size = 30)) +          # set large font size for readability
  labs(x = "Year") +                               # label x-axis
  labs(y = "Billions of Chained 2017 Dollars")    # label y-axis

# Display the plot
fig
```



In [5]:

```
library(urca)
summary(ur.df(tsgdp, type='trend', lags=8, selectlags="AIC"))
```

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-1797.63   -32.40     3.44    46.24  1107.13

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.194420  18.513223   0.875   0.3824
z.lag.1     -0.004476   0.007428  -0.603   0.5472
tt           0.690343   0.521043   1.325   0.1862
z.diff.lag  -0.147127   0.057156  -2.574   0.0105 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 148.1 on 301 degrees of freedom
Multiple R-squared:  0.05995,   Adjusted R-squared:  0.05058
F-statistic: 6.399 on 3 and 301 DF,  p-value: 0.0003245


Value of test-statistic is: -0.6026 28.452 7.7571

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.98 -3.42 -3.13
phi2  6.15  4.71  4.05
phi3  8.34  6.30  5.36
```

ADF Test Interpretation

Test statistics:

$\tau_3 = -0.6026$, $\varphi_2 = 28.452$, $\varphi_3 = 7.7571$

Critical values:

Significance	1%	5%	10%
τ_3	−3.98	−3.42	−3.13
φ_2	6.15	4.71	4.05
φ_3	8.34	6.30	5.36

Interpretation:

- We **reject the null hypothesis** that both the **constant and deterministic trend are zero**, because $\varphi_2 = 28.452 > 6.15$.
- However, at the **1% significance level**, we **cannot reject** that the **deterministic trend is zero** since $\varphi_3 = 7.7571 < 8.34$.
- Therefore, we **retest with type = "drift"**, assuming the series has a constant but no trend.

In [6]:
```
library(urca)
summary(ur.df(tsgdp, type='drift', lags=8, selectlags="AIC"))
```

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-1807.16   -34.67     5.05    46.78  1110.02

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 27.003256  16.639597   1.623 0.105669
z.lag.1      0.005190   0.001401   3.705 0.000252 ***
z.diff.lag  -0.150104   0.057183  -2.625 0.009107 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 148.3 on 302 degrees of freedom
Multiple R-squared:  0.05447,	Adjusted R-squared:  0.04821
F-statistic: 8.699 on 2 and 302 DF,  p-value: 0.0002123


Value of test-statistic is: 3.7046 41.696

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.44 -2.87 -2.57
phi1  6.47  4.61  3.79
```

ADF Test Interpretation (drift case)

Test statistics:

$\tau_2 = 3.7046$, $\varphi_1 = 41.696$

Critical values:

Significance	1%	5%	10%
τ_2	−3.44	−2.87	−2.57
φ_1	6.47	4.61	3.79

Interpretation:

- **Deterministic term:** Since $\varphi_1 = 41.696$ exceeds its critical values, we **reject** the null that the **constant (drift) is zero** → the model **should include a constant** (specification with drift is correct).
- **Unit root:** The ADF statistic $\tau_2 = 3.7046$ is **not more negative** than the critical values, so we **cannot reject** the **unit root** null at conventional levels.

In [7]:
```
# Take the first difference of Log(GDP) to obtain the growth rate (stationary series)
dgdp = diff(log(tsgdp))
```

In [8]:
```
library(urca)
summary(ur.df(dgdp, type='drift', lags=8, selectlags="AIC"))
```

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.087406 -0.004101 -0.000198  0.004367  0.078958 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.0061336  0.0008576   7.152 6.52e-12 ***
z.lag.1     -0.7897137  0.0754362 -10.469  < 2e-16 ***
z.diff.lag  -0.1005670  0.0570089  -1.764   0.0787 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01101 on 301 degrees of freedom
Multiple R-squared:  0.4452,	Adjusted R-squared:  0.4415 
F-statistic: 120.8 on 2 and 301 DF,  p-value: < 2.2e-16


Value of test-statistic is: -10.4686 54.801 

Critical values for test statistics: 
      1pct  5pct 10pct
tau2 -3.44 -2.87 -2.57
phi1  6.47  4.61  3.79
```

ADF Test Interpretation (first difference)

Test statistics:
$\tau_2 = -10.4686$, $\varphi_1 = 54.801$

Critical values:

Significance	1%	5%	10%
τ_2	−3.44	−2.87	−2.57
φ_1	6.47	4.61	3.79

Interpretation:

- **Deterministic term:** Since $\varphi_1 = 54.801$ is much larger than its critical values, we **reject** the null that the **constant (drift) is zero** → the drift specification remains appropriate.
- **Unit root:** The ADF statistic $\tau_2 = -10.4686$ is far **below** the critical values, so we **reject** the null hypothesis of a **unit root**.
- Therefore, the series is **stationary after first differencing**, indicating there is **no second unit root**.

Exercise 2 (15')

Consider two reduced-form VAR models:

i. a two-variable system of real GDP and the unemployment rate in the US;
ii. a three-variable system of real GDP, inflation, and the unemployment rate in the US.

For each system:

1. Create a joint time-series object and report the number of observations.
2. Use AIC to determine the optimal number of lags.
3. Write out the full system of equations assuming 2 lags for 2 variable system, and 3 lags for three variable system.
4. Rerwrite the system using matrices.
5. Compute the total number of parameters to be estimated.

Solution to Exercise 2

In [9]:
```
# Align series on common dates (keep only overlapping periods)
data2 = ts.intersect(dgdp, durate)          # 2-variable system
data3 = ts.intersect(dgdp, durate, dinf)     # 3-variable system

# Number of overlapping observations in each system
```

```
nrow(data2)
nrow(data3)
```

310
310

In [10]:
```
head(data2)
```

A matrix: 6 × 2 of type dbl

dgdp	durate
0.014948096	3.7
0.016388742	3.7
0.005728742	3.8
0.001132224	3.8
-0.013861770	4.7
-0.003402356	5.9

In [11]:
```
head(data3)
```

A matrix: 6 × 3 of type dbl

dgdp	durate	dinf
0.014948096	3.7	2.0966026
0.016388742	3.7	1.5795329
0.005728742	3.8	1.6698053
0.001132224	3.8	-0.9223866
-0.013861770	4.7	-0.9560302
-0.003402356	5.9	-0.1086502

Each system contains 310 observations.

In [24]:
```
library(vars)

out = VARselect(data2, lag.max = 8, type = "const")
out$selection
```

AIC(n): 2 **HQ(n):** 2 **SC(n):** 2 **FPE(n):** 2

According to AIC, the two-variable system should be estimated on **two** lags since AIC(n):2

In [13]:
```
library(vars)

out = VARselect(data3, lag.max = 8, type = "const")
out$selection
```

AIC(n): 3 **HQ(n):** 2 **SC(n):** 1 **FPE(n):** 3

According to AIC, the three-variable system should be estimated on **three** lags since AIC(n):3

VAR(2) with two variables (GDP growth and unemployment)

Equation-by-equation form (2 lags):

$$\begin{aligned} dgdp_t &= c_1 + \phi_{11,1}\, dgdp_{t-1} + \phi_{12,1}\, durate_{t-1} + \phi_{11,2}\, dgdp_{t-2} + \phi_{12,2}\, durate_{t-2} + u_{1t}, \\ durate_t &= c_2 + \phi_{21,1}\, dgdp_{t-1} + \phi_{22,1}\, durate_{t-1} + \phi_{21,2}\, dgdp_{t-2} + \phi_{22,2}\, durate_{t-2} + u_{2t}. \end{aligned}$$

Matrix form:

$$\begin{bmatrix} dgdp_t \\ durate_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + A_1 \begin{bmatrix} dgdp_{t-1} \\ durate_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} dgdp_{t-2} \\ durate_{t-2} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \quad A_j = \begin{bmatrix} \phi_{11,j} & \phi_{12,j} \\ \phi_{21,j} & \phi_{22,j} \end{bmatrix}, \ j = 1, 2.$$

Parameter count: Per equation: $kp + 1 = 2 \times 2 + 1 = 5$ → Total across 2 equations: **10** plus **3** free parameters in Σ_u, where $\Sigma_u = \mathrm{Var}(u_t) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$, so **13** parameters.

VAR(3) with three variables (GDP growth,unemployment, inflation)

Equation-by-equation form (3 lags):

$$\begin{aligned}
dgdp_t &= c_1 + \phi_{11,1}\, dgdp_{t-1} + \phi_{12,1}\, durate_{t-1} + \phi_{13,1}\, dinf_{t-1} \\
&\quad + \phi_{11,2}\, dgdp_{t-2} + \phi_{12,2}\, durate_{t-2} + \phi_{13,2}\, dinf_{t-2} \\
&\quad + \phi_{11,3}\, dgdp_{t-3} + \phi_{12,3}\, durate_{t-3} + \phi_{13,3}\, dinf_{t-3} + u_{1t}, \\
durate_t &= c_2 + \phi_{21,1}\, dgdp_{t-1} + \phi_{22,1}\, durate_{t-1} + \phi_{23,1}\, dinf_{t-1} \\
&\quad + \phi_{21,2}\, dgdp_{t-2} + \phi_{22,2}\, durate_{t-2} + \phi_{23,2}\, dinf_{t-2} \\
&\quad + \phi_{21,3}\, dgdp_{t-3} + \phi_{22,3}\, durate_{t-3} + \phi_{23,3}\, dinf_{t-3} + u_{2t}, \\
dinf_t &= c_3 + \phi_{31,1}\, dgdp_{t-1} + \phi_{32,1}\, durate_{t-1} + \phi_{33,1}\, dinf_{t-1} \\
&\quad + \phi_{31,2}\, dgdp_{t-2} + \phi_{32,2}\, durate_{t-2} + \phi_{33,2}\, dinf_{t-2} \\
&\quad + \phi_{31,3}\, dgdp_{t-3} + \phi_{32,3}\, durate_{t-3} + \phi_{33,3}\, dinf_{t-3} + u_{3t}.
\end{aligned}$$

Matrix form:

$$\begin{bmatrix} dgdp_t \\ durate_t \\ dinf_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + A_1 \begin{bmatrix} dgdp_{t-1} \\ durate_{t-1} \\ dinf_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} dgdp_{t-2} \\ durate_{t-2} \\ dinf_{t-2} \end{bmatrix} + A_3 \begin{bmatrix} dgdp_{t-3} \\ durate_{t-3} \\ dinf_{t-3} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}, \quad A_j = \begin{bmatrix} \phi_{11,j} & \phi_{12,j} & \phi_{13,j} \\ \phi_{21,j} & \phi_{22,j} & \phi_{23,j} \\ \phi_{31,j} & \phi_{32,j} & \phi_{33,j} \end{bmatrix}, \; j = 1, 2, 3.$$

Parameter count: Per equation: $kp + 1 = 3 \times 3 + 1 = 10$. Across 3 equations: **30** regression parameters, plus **6** free parameters in the 3×3 covariance matrix Σ_u, where $\Sigma_u = \mathrm{Var}(u_t) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}$, so **36** parameters.

Exercise 3 (15')

Building on Exercise 2, in which you determined the optimal lag length for a reduced-form VAR, do the following for each system:

1. Rewrite the model in structural form as a system of equations.
2. Rewrite the model in matrix form.
3. Count the total number of parameters in the structural form (intercepts, contemporaneous matrix, and lag matrices).

Solution to Exercise 3

SVAR(2) with two variables (GDP growth and unemployment)

Equation-by-equation form (2 lags):

$$a_{11}\, dgdp_t + a_{12}\, durate_t = \alpha_1 + \beta_{11,1}\, dgdp_{t-1} + \beta_{12,1}\, durate_{t-1} + \beta_{11,2}\, dgdp_{t-2} + \beta_{12,2}\, durate_{t-2} + b_{11}\, \varepsilon_{1t} + b_{12}\, \varepsilon_{2t}.$$

$$a_{21}\, dgdp_t + a_{22}\, durate_t = \alpha_2 + \beta_{21,1}\, dgdp_{t-1} + \beta_{22,1}\, durate_{t-1} + \beta_{21,2}\, dgdp_{t-2} + \beta_{22,2}\, durate_{t-2} + b_{21}\, \varepsilon_{1t} + b_{22}\, \varepsilon_{2t}.$$

Matrix form:

$$A \begin{bmatrix} dgdp_t \\ durate_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + A_1 \begin{bmatrix} dgdp_{t-1} \\ durate_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} dgdp_{t-2} \\ durate_{t-2} \end{bmatrix} + B \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad A_j = \begin{bmatrix} \beta_{11,j} & \beta_{12,j} \\ \beta_{21,j} & \beta_{22,j} \end{bmatrix}, \; j = 1, 2, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

Number of strucutral unknowns: **18**.

SVAR(3) with three variables (GDP growth, unemployment, and inflation)

Equation-by-equation form (3 lags):

$$\begin{aligned}
a_{11}\, dgdp_t + a_{12}\, durate_t + a_{13}\, dinf_t &= \alpha_1 + \beta_{11,1}\, dgdp_{t-1} + \beta_{12,1}\, durate_{t-1} + \beta_{13,1}\, dinf_{t-1} \\
&\quad + \beta_{11,2}\, dgdp_{t-2} + \beta_{12,2}\, durate_{t-2} + \beta_{13,2}\, dinf_{t-2} \\
&\quad + \beta_{11,3}\, dgdp_{t-3} + \beta_{12,3}\, durate_{t-3} + \beta_{13,3}\, dinf_{t-3} \\
&\quad + b_{11}\, \varepsilon_{1t} + b_{12}\, \varepsilon_{2t} + b_{13}\, \varepsilon_{3t}. \\
a_{21}\, dgdp_t + a_{22}\, durate_t + a_{23}\, dinf_t &= \alpha_2 + \beta_{21,1}\, dgdp_{t-1} + \beta_{22,1}\, durate_{t-1} + \beta_{23,1}\, dinf_{t-1} \\
&\quad + \beta_{21,2}\, dgdp_{t-2} + \beta_{22,2}\, durate_{t-2} + \beta_{23,2}\, dinf_{t-2} \\
&\quad + \beta_{21,3}\, dgdp_{t-3} + \beta_{22,3}\, durate_{t-3} + \beta_{23,3}\, dinf_{t-3} \\
&\quad + b_{21}\, \varepsilon_{1t} + b_{22}\, \varepsilon_{2t} + b_{23}\, \varepsilon_{3t}.
\end{aligned}$$

$$\begin{aligned} a_{31}\, dgdp_t + a_{32}\, durate_t + a_{33}\, dinf_t = \alpha_3 &+ \beta_{31,1}\, dgdp_{t-1} + \beta_{32,1}\, durate_{t-1} + \beta_{33,1}\, dinf_{t-1} \\ &+ \beta_{31,2}\, dgdp_{t-2} + \beta_{32,2}\, durate_{t-2} + \beta_{33,2}\, dinf_{t-2} \\ &+ \beta_{31,3}\, dgdp_{t-3} + \beta_{32,3}\, durate_{t-3} + \beta_{33,3}\, dinf_{t-3} \\ &+ b_{31}\, \varepsilon_{1t} + b_{32}\, \varepsilon_{2t} + b_{33}\, \varepsilon_{3t}. \end{aligned}$$

Matrix form:

$$A \begin{bmatrix} dgdp_t \\ durate_t \\ dinf_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + A_1 \begin{bmatrix} dgdp_{t-1} \\ durate_{t-1} \\ dinf_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} dgdp_{t-2} \\ durate_{t-2} \\ dinf_{t-2} \end{bmatrix} + A_3 \begin{bmatrix} dgdp_{t-3} \\ durate_{t-3} \\ dinf_{t-3} \end{bmatrix} + B \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad A_j = \begin{bmatrix} \beta_{11,j} & \beta_{12,j} & \beta_{13,j} \\ \beta_{21,j} & \beta_{22,j} & \beta_{23,j} \\ \beta_{31,j} & \beta_{32,j} & \beta_{33,j} \end{bmatrix}, \; j = 1, 2, 3, \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}.$$

Number of strucutral unknowns: **48**.

Exercise 4 (15')

Building on Exercise 3, do the following for each system:

1. Rewrite the matrix system after imposing restrictions to identify the structural parameters via Cholesky identification. Count the number of restrictions. In the two-variable system, assume the unemployment rate does not have a contemporaneous effect on GDP. In the three-variable system, assume GDP can have contemporaneous effects on the other two variables, but neither inflation nor unemployment has a contemporaneous effect on GDP. Moreover, assume inflation does not affect unemployment contemporaneously.
2. Estimate the reduced-form VAR.
3. Use the reduced-form estimates to identify the structural VAR.

Solution to Exercise 4

SVAR(2) with two variables (GDP growth and unemployment)

Cholesky identification:

$$A \begin{bmatrix} dgdp_t \\ durate_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + A_1 \begin{bmatrix} dgdp_{t-1} \\ durate_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} dgdp_{t-2} \\ durate_{t-2} \end{bmatrix} + B \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

where

$$A = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix}, \quad A_j = \begin{bmatrix} \beta_{11,j} & \beta_{12,j} \\ \beta_{21,j} & \beta_{22,j} \end{bmatrix}, \; j = 1, 2, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Number of restirctions: **5**.

SVAR(3) with three variables (GDP growth, unemployment, and inflation)

Cholesky identification:

$$A \begin{bmatrix} dgdp_t \\ durate_t \\ dinf_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + A_1 \begin{bmatrix} dgdp_{t-1} \\ durate_{t-1} \\ dinf_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} dgdp_{t-2} \\ durate_{t-2} \\ dinf_{t-2} \end{bmatrix} + A_3 \begin{bmatrix} dgdp_{t-3} \\ durate_{t-3} \\ dinf_{t-3} \end{bmatrix} + B \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}$$

where

$$A = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad A_j = \begin{bmatrix} \beta_{11,j} & \beta_{12,j} & \beta_{13,j} \\ \beta_{21,j} & \beta_{22,j} & \beta_{23,j} \\ \beta_{31,j} & \beta_{32,j} & \beta_{33,j} \end{bmatrix}, \; j = 1, 2, 3, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Number of restrictions: **12**.

In [14]:

```
library(vars)

# Estimate a reduced-form VAR(2)
sys2.VAR2 = VAR(data2, p = 2, type = "const")
```

```
# Lower-triangular A-pattern for recursive (Cholesky) SVAR
amat2 = matrix(NA, 2, 2)   # NA = to be estimated
amat2[1, 2] = 0            # a12 = 0: unemployment rate ⟶ GDP (no contemporaneous effect)

# Identify a structural VAR
sys2.SVAR2 = SVAR(sys2.VAR2, Amat = amat2, estmethod = "scoring", lrtest = FALSE)

summary(sys2.SVAR2)
```

```
SVAR Estimation Results:
========================

Call:
SVAR(x = sys2.VAR2, estmethod = "scoring", Amat = amat2, lrtest = FALSE)

Type: A-model
Sample size: 308
Log Likelihood: 696.435
Method: scoring
Number of iterations: 99

Estimated A matrix:
        dgdp durate
dgdp   82.70  0.000
durate 63.85  1.981

Estimated standard errors for A matrix:
        dgdp  durate
dgdp   3.332 0.00000
durate 5.369 0.07983

Estimated B matrix:
       dgdp durate
dgdp      1      0
durate    0      1

Covariance matrix of reduced form residuals (*100):
           dgdp  durate
dgdp    0.01462 -0.4711
durate -0.47115 40.6576
```

In [15]:

```
library(vars)

# Estimate a reduced-form VAR(3)
sys3.VAR3 = VAR(data3, p = 3, type = "const")

# Lower-triangular A-pattern for recursive (Cholesky) SVAR
amat3 = matrix(NA, 3, 3)   # NA = to be estimated
amat3[1, 2] = 0            # a12 = 0: unemploymenr rate ⟶ GDP (no contemporaneous effect)
amat3[1, 3] = 0            # a13 = 0: inflation ⟶ GDP (no contemporaneous effect)
amat3[2, 3] = 0            # a23 = 0: inflation ⟶ unemployment rate (no contemporaneous effect)

# Identify a structural VAR
sys3.SVAR3 = SVAR(sys3.VAR3, Amat = amat3, estmethod = "scoring", lrtest = FALSE)

summary(sys3.SVAR3)
```

```
SVAR Estimation Results:
========================

Call:
SVAR(x = sys3.VAR3, estmethod = "scoring", Amat = amat3, lrtest = FALSE)

Type: A-model
Sample size: 307
Log Likelihood: 462.953
Method: scoring
Number of iterations: 99

Estimated A matrix:
          dgdp durate  dinf
dgdp   83.0709  0.000 0.000
durate 63.5198  1.973 0.000
dinf   -0.2772  0.334 1.946

Estimated standard errors for A matrix:
        dgdp  durate    dinf
dgdp   3.352 0.00000 0.00000
durate 5.390 0.07962 0.00000
dinf   5.968 0.11340 0.07852

Estimated B matrix:
       dgdp durate dinf
dgdp      1      0    0
durate    0      1    0
dinf      0      0    1

Covariance matrix of reduced form residuals (*100):
           dgdp  durate     dinf
dgdp    0.01449 -0.4666  0.08216
durate -0.46658 40.7160 -7.05594
dinf    0.08216 -7.0559 27.63670
```

Exercise 5 (15')

Building on Exercise 4, do the following for each system:

1. Analyze the response of unemployment to a one–standard-deviation structural shock to GDP for the next 4 years.
2. Interpret the impulse responses.

Solution to Exercise 5

In [16]:

```
# SVAR IRF: structural identification (A/B restrictions in sys2.SVAR2)
irf.sys2.SVAR2 = irf(
  sys2.SVAR2,
  impulse  = "dgdp",     # shock variable
  response = "durate",   # variable whose response we trace
  n.ahead  = 16,         # horizons
  boot     = TRUE,       # bootstrap CIs (irf/Lower/Upper)
  seed     = 123         # reproducible bands (used only when boot=TRUE)
)
head(irf.sys2.SVAR2$irf$dgdp)   # first horizons of durate to a structural dgdp shock


# VAR IRF: reduced-form with Cholesky orthogonalization
irf.sys2.VAR2 = irf(
  sys2.VAR2,
  impulse  = "dgdp",
  response = "durate",
  n.ahead  = 16,
  ortho    = TRUE,       # Cholesky; ordering in sys2.VAR2 matters
  boot     = FALSE,      # no CIs; 'seed' has no effect
  seed     = 123
)
head(irf.sys2.VAR2$irf$dgdp)    # first horizons under Cholesky ID
```

A matrix: 6 × 1 of type dbl

durate
-0.3896619
-0.6090296
-0.6710500
-0.6482565
-0.5979143
-0.5372704

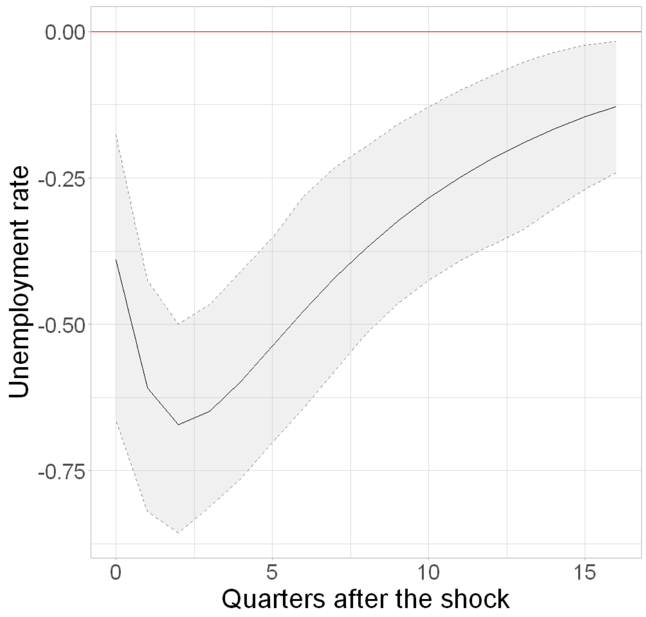
A matrix: 6 × 1 of type dbl

durate
-0.4585893
-0.5824296
-0.6156748
-0.5825826
-0.5328237
-0.4764961

In [25]:

```
library(devtools)
source_url("https://raw.githubusercontent.com/anguyen1210/var-tools/master/R/extract_varirf.R");
library(dplyr)
temp = extract_varirf(irf.sys2.SVAR2)
temp %>%
  ggplot(aes(x=period, y=irf_dgdp_durate, ymin=lower_dgdp_durate, ymax=upper_dgdp_durate)) +
  geom_hline(yintercept = 0, color="red") +
  geom_ribbon(fill="grey", alpha=.2, color="grey50", linetype="dashed") +
  geom_line() +
  theme_light() +
  theme(aspect.ratio=1) +
  theme(plot.margin = ggplot2::margin(0.2, 0.2, 0.2, 0.2, "cm")) +
  theme(text=element_text(size=25)) +
  labs(x = "Quarters after the shock") +
  labs(y = "Unemployment rate")
```

```
i SHA-1 hash of file is "cb23ce6a462d73b67920ae56dc83c828d18fe7e4"
```



In [18]:

```
# SVAR IRF: structural identification (A/B restrictions in sys3.SVAR3)
irf.sys3.SVAR3 = irf(
  sys3.SVAR3,
  impulse  = "dgdp",     # shock variable
  response = "durate",   # variable whose response we trace
  n.ahead  = 16,         # horizons
  boot     = TRUE,       # bootstrap CIs (irf/Lower/Upper)
  seed     = 123         # reproducible bands (used only when boot=TRUE)
)
head(irf.sys3.SVAR3$irf$dgdp)   # first horizons of durate to a structural dgdp shock


# VAR IRF: reduced-form with Cholesky orthogonalization
irf.sys3.VAR3 = irf(
  sys3.VAR3,
  impulse  = "dgdp",
  response = "durate",
  n.ahead  = 16,
  ortho    = TRUE,       # Cholesky; ordering in sys3.VAR3 matters
  boot     = FALSE,      # no CIs; 'seed' has no effect
  seed     = 123
)
head(irf.sys3.VAR3$irf$dgdp)    # first horizons under Cholesky ID
```

A matrix: 6 × 1 of type dbl

durate
-0.3875892
-0.5754136
-0.6402268
-0.6442955
-0.5833409
-0.5131715

A matrix: 6 × 1 of type dbl

durate
-0.4567854
-0.5532078
-0.5764883
-0.5663070
-0.5048055
-0.4412406

In [19]:

```
library(devtools)
source_url("https://raw.githubusercontent.com/anguyen1210/var-tools/master/R/extract_varirf.R");
library(dplyr)
temp = extract_varirf(irf.sys3.SVAR3)
temp %>%
  ggplot(aes(x=period, y=irf_dgdp_durate, ymin=lower_dgdp_durate, ymax=upper_dgdp_durate)) +
  geom_hline(yintercept = 0, color="red") +
  geom_ribbon(fill="grey", alpha=.2, color="grey50", linetype="dashed") +
  geom_line() +
  theme_light() +
  theme(aspect.ratio=1) +
  theme(plot.margin = ggplot2::margin(0.2, 0.2, 0.2, 0.2, "cm")) +
  theme(text=element_text(size=25)) +
  labs(x = "Quarters after the shock") +
  labs(y = "Unemployment rate")
```

```
ℹ SHA-1 hash of file is "cb23ce6a462d73b67920ae56dc83c828d18fe7e4"
```

Comparison: Unemployment response in 2-variable vs. 3-variable SVAR

Overall pattern
Both figures show that the unemployment rate **declines following a positive structural GDP shock** and gradually returns to zero.
The negative short-run effect and eventual mean reversion are consistent with **Okun's law** and standard macroeconomic dynamics.

Similarities

- **Direction of effect:**
  In both SVARs, a positive GDP-related structural shock leads to a **temporary drop in unemployment**, as higher output reduces joblessness.
- **Timing:**
  The strongest effect occurs within the first few quarters, and the response **returns toward zero over roughly 12–16 quarters**.
- **Identification:**
  Both models use the **same Cholesky decomposition** for structural identification, ensuring the shocks are comparable across specifications.

Differences

- **Magnitude:**
  The **two-variable SVAR** produces a slightly **larger initial decline** in unemployment (≈ −0.6 to −0.8 p.p.), while the **three-variable SVAR** shows a more moderate effect (≈ −0.5 p.p.).
- **Persistence and statistical significance:**
  The **three-variable SVAR response becomes insignificant sooner** — the confidence band includes zero already around **6–8 quarters** — whereas in the **two-variable SVAR**, the response remains **significant for longer** (roughly up to 10–12 quarters). This suggests that including the third variable absorbs some of the variation previously attributed to GDP shocks, reducing the estimated persistence.
- **Model dimensionality and precision:**
  Adding a third variable increases model complexity, leading to **wider confidence bands** and potentially less precise estimates for each impulse response. The smaller model, by contrast, concentrates the dynamics in fewer equations, which can yield **sharper and longer-lasting responses**.

Interpretation

Both models tell a consistent story:
a positive structural GDP shock temporarily lowers unemployment.

However, the **two-variable SVAR** shows a **stronger and more persistent** response, while the **three-variable SVAR** yields a **smaller and less persistent** effect that becomes **insignificant sooner**.
This pattern is typical when expanding a VAR system — additional variables can improve realism but may also **dilute and statistically weaken individual impulse responses**.

Forecasting VAR models

In [20]:
```
# Generate forecasts from the VAR model
prd = predict(
  sys2.VAR2,   # the estimated VAR object
  n.ahead = 10, # number of steps ahead to forecast (here: 4 periods)
  ci = 0.95    # confidence interval level (95%)
)
```

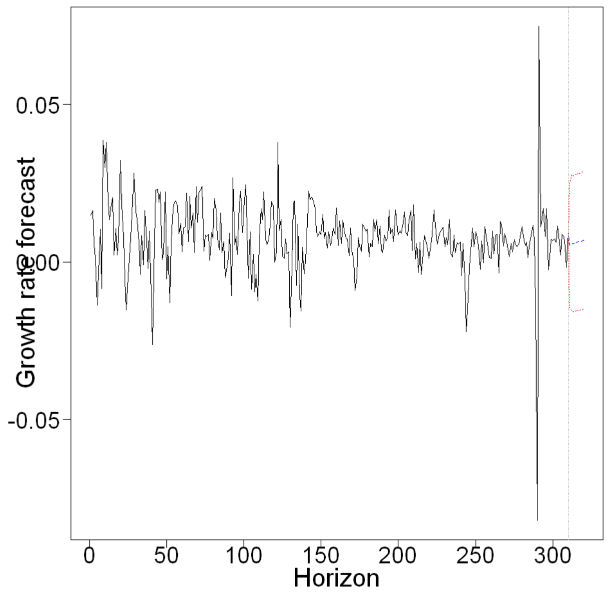
```
# Display the forecast results
print(prd)
```

```
$dgdp
            fcst       lower      upper         CI
 [1,] 0.005269718 -0.01480743 0.02534686 0.02007715
 [2,] 0.006045592 -0.01543808 0.02752926 0.02148367
 [3,] 0.005728474 -0.01584857 0.02730552 0.02157705
 [4,] 0.005961434 -0.01568938 0.02761225 0.02165082
 [5,] 0.006090112 -0.01559972 0.02777995 0.02168984
 [6,] 0.006270006 -0.01547303 0.02801304 0.02174303
 [7,] 0.006426084 -0.01536283 0.02821500 0.02178892
 [8,] 0.006572835 -0.01525568 0.02840135 0.02182851
 [9,] 0.006702586 -0.01515763 0.02856280 0.02186021
[10,] 0.006817891 -0.01506734 0.02870312 0.02188523

$durate
         fcst    lower    upper       CI
 [1,] 4.340476 3.090252 5.590701 1.250225
 [2,] 4.477928 2.726271 6.229584 1.751656
 [3,] 4.620004 2.469140 6.770867 2.150863
 [4,] 4.749464 2.303854 7.195074 2.445610
 [5,] 4.866905 2.201724 7.532086 2.665181
 [6,] 4.971091 2.143337 7.798845 2.827754
 [7,] 5.063110 2.114350 8.011870 2.948760
 [8,] 5.143974 2.104809 8.183139 3.039165
 [9,] 5.214912 2.107884 8.321940 3.107028
[10,] 5.277062 2.118894 8.435229 3.158168
```

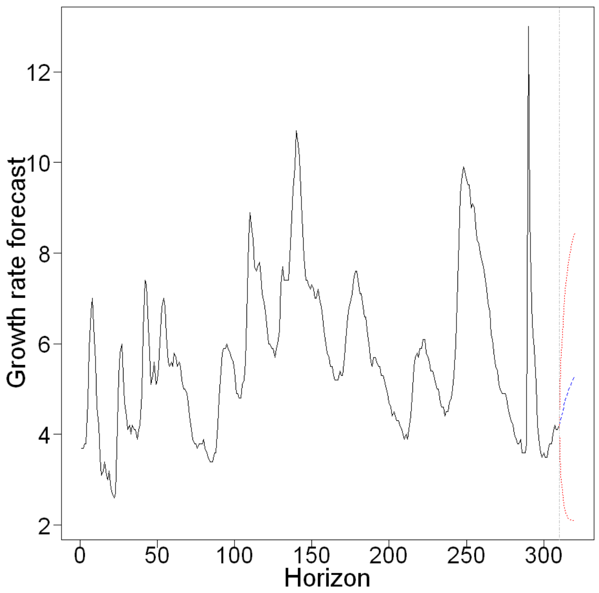
In [21]:
```
par(
  pty = "s",                  # square plotting region (aspect ratio ~ 1)
  cex.axis = 1.8,             # tick label size (1 = default)
  cex.lab  = 2.0,             # axis title size
  mgp = c(2.2, 0.7, 0),       # axis title, labels, line spacing
  las = 1,                    # horizontal y-axis labels
  mar = c(4, 5, 1.5, 1)      # margins: bottom, left, top, right
)

# Plot only the dgdp forecast, single panel, with your own labels
plot(
  prd,
  names     = "dgdp",         # only this variable
  plot.type = "single",       # single panel
  xlab      = "Horizon",
  ylab      = "Growth rate forecast",
  main      = ""              # remove default title
)
```



In [22]:
```
par(
  pty = "s",                  # square plotting region (aspect ratio ~ 1)
  cex.axis = 1.8,             # tick label size (1 = default)
  cex.lab  = 2.0,             # axis title size
  mgp = c(2.2, 0.7, 0),       # axis title, labels, line spacing
  las = 1,                    # horizontal y-axis labels
  mar = c(4, 5, 1.5, 1)       # margins: bottom, left, top, right
)

# Plot only the dgdp forecast, single panel, with your own labels
plot(
  prd,
  names     = "durate",       # only this variable
  plot.type = "single",       # single panel
  xlab      = "Horizon",
  ylab      = "Growth rate forecast",
  main      = ""              # remove default title
)
```


Growth rate forecast
Horizon
12
10
8
6
4
2
0
50
100
150
200
250
300

In [ ]: