# Séries de Tempo

## Aula 8 - Modelos vetoriais autorregressivos

Regis A. Ely

Departamento de Economia
Universidade Federal de Pelotas

04 de setembro de 2020

# Conteúdo

# Modelo VAR

- Modelos do tipo ADL ou ECM apenas consideram relações unidirecionais entre as variáveis e dependem da hipótese de exogeneidade estrita
- Modelos VAR são modelos de equações simultâneas em que é possível uma relação de *feedback* entre as variáveis endógenas, sendo todas tratadas de maneira simétrica[1]
- Um modelo VAR(p) bivariado é descrito como:

$$Y_t = c_1 + \phi_{11} Y_{t-1} + \ldots + \phi_{1p} Y_{t-p} + \gamma_{11} X_{t-1} + \ldots + \gamma_{1p} X_{t-p} + \varepsilon_{1t}$$
$$X_t = c_2 + \phi_{21} Y_{t-1} + \ldots + \phi_{2p} Y_{t-p} + \gamma_{21} X_{t-1} + \ldots + \gamma_{2p} X_{t-p} + \varepsilon_{2t}$$

[1] Também é possível incluir variáveis exógenas em um modelo VAR.

# Modelo VAR

- No modelo VAR os resíduos das equações $\varepsilon_{1t}$ e $\varepsilon_{2t}$ são ruídos brancos que podem ser correlacionados contemporaneamente
- Se as séries forem todas estacionárias, podemos estimar um modelo VAR em nível
- Se as séries forem não-estacionárias, podemos tirar diferenças até torná-las estacionárias e então estimar o modelo VAR[2]

[2] Na próxima aula veremos o caso em que as séries são não-estacionárias mas cointegradas.

# Modelo VAR

- Para construção dos modelos VAR utilizamos o mesmo ciclo de identificação, estimação, diagnóstico e previsão
- Entretanto, inicialmente devemos definir quais variáveis do modelo serão endógenas e fracamente exógenas, normalmente com base na teoria ou literatura
- A identificação da ordem $p$ a ser estimada requer a utilização de critérios como Akaike, BIC e Hannan-Quin
- A estimação pode ser realizada por máxima verossimilhança condicional, que é equivalente ao estimador de OLS

# Modelo VAR

- O diagnóstico dos modelos VAR requer a utilização de uma versão multivariada do teste de Ljung-Box, uma vez que os resíduos de todas as equações devem ser conjuntamente não autocorrelacionados
- Após estimado o modelo e realizado o diagnóstico, pode-se proceder a análise dos coeficientes e exercícios de simulação e previsão
- A previsão dos valores para cada equação é construída de maneira recursiva, semelhante a estrutura dos modelos AR
- Adicionalmente, pode-se fazer uma análise de simulação chamada de impulso-resposta, onde avaliamos como um choque em uma das variáveis do modelo se difunde ao longo do tempo em outras variáveis

# Causalidade de Granger

- Granger definiu causalidade em séries de tempo em termos de previsibilidade
- A variável $X$ causa Granger a variável $Y$, com respeito a um dado conjunto de informação, se o presente de $Y$ pode ser previsto mais eficientemente usando valores passados de $X$ do que não incluindo estes valores
- A eficiência na previsão é medida através do do erro quatrático médio de previsão (EQM)
- Note que para verificar se $X_t$ causa Granger $Y_t$ devemos testar a significância conjunta de todos os coeficientes passados[3] de $X_t$ na equação de $Y_t$

[3] Se tivermos mais do que duas equações no modelo VAR será necessário testar mais coeficientes conjuntamente para garantir causalidade.

# Causalidade de Granger

Defina $\{A_t, t = 0, \pm 1, \pm 2, \ldots\}$ o conjunto de toda a informação disponível até o instante $t$, e considere $\bar{A}_t = \{A_s : s < t\}$, $\bar{\bar{A}}_t = \{A_s : s \leq t\}$ e definições análogas para $\bar{X}_t$ e $\bar{Y}_t$, então:

1. $X_t$ causa Granger $Y_t$ $(X_t \rightarrow Y_t)$ se $EQM(Y_t|\bar{A}_t) < EQM(Y_t|\bar{A}_t - \bar{X}_t)$, ou seja, $Y_t$ pode ser melhor previsto usando o passado de $X_t$
2. $X_t$ causa Granger $Y_t$ instantaneamente $(X_t \Rightarrow Y_t)$[4] se $EQM(Y_t|\bar{A}_t, \bar{\bar{X}}_t) < EQM(Y_t|\bar{A}_t)$, ou seja, $Y_t$ pode ser melhor previsto utilizando o valor presente de $X_t$

Dizemos que há *feedback* se tanto $X_t$ causa Granger $Y_t$ quanto $Y_t$ causa Granger $X_t$

[4] Se $(X_t \Rightarrow Y_t)$ então automaticamente $(Y_t \Rightarrow X_t)$.

# Exemplo no R

Para estimação de modelos VAR vamos utilizar o pacote `vars` e a base de dados `uschange` do pacote `fpp2`, que contém dados macroeconômicos trimestrais dos Estados Unidos durante os anos de 1960 a 2016:

```
library(vars)
library(fpp2)
base <- uschange[, 1:2]
```

Iremos trabalhar apenas com as duas primeiras variáveis da base, a taxa de crescimento percentual do consumo agregado e da renda disponível

# Exemplo no R

Vamos inicialmente visualizar as observações iniciais e o gráfico das duas séries:

```r
head(base)
```

```
##         Consumption     Income
## 1970 Q1   0.6159862  0.9722610
## 1970 Q2   0.4603757  1.1690847
## 1970 Q3   0.8767914  1.5532705
## 1970 Q4  -0.2742451 -0.2552724
## 1971 Q1   1.8973708  1.9871536
## 1971 Q2   0.9119929  1.4473342
```

```r
autoplot(base)
```

# Exemplo no R

2.5
0.0
−2.5
base
1970
1980
1990
2000
2010
Time
series
Consumption
Income

# Estacionariedade

A primeira etapa antes de estimar as relações dinâmicas entre consumo e renda através do modelo VAR é checar a estacionariedade das séries. Para isso, utilizamos o teste `ADF` através da função `ur.df` em cada uma das séries:

```r
adf_cons <- ur.df(
  base[,1], type = "drift", lags = 5,
  selectlags = "AIC"
)
summary(adf_cons)
```

# Estacionariedade

```
...
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.29125    0.07899   3.687 0.000302 ***
## z.lag.1     -0.39220    0.08796  -4.459 1.46e-05 ***
## z.diff.lag1 -0.35360    0.08814  -4.012 8.88e-05 ***
## z.diff.lag2 -0.20654    0.07190  -2.873 0.004569 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.586 on 177 degrees of freedom
## Multiple R-squared:  0.3709, Adjusted R-squared:  0.3603
## F-statistic: 34.79 on 3 and 177 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -4.4591 9.9434
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
...
```

# Estacionariedade

```
...
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.70996    0.10458   6.788 1.63e-10 ***
## z.lag.1     -1.01010    0.11028  -9.159  < 2e-16 ***
## z.diff.lag  -0.08954    0.07408  -1.209    0.228
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9305 on 178 degrees of freedom
## Multiple R-squared:  0.5589, Adjusted R-squared:  0.554
## F-statistic: 112.8 on 2 and 178 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -9.1592 41.9531
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
...
```

# Estacionariedade

- Para as duas séries podemos rejeitar a hípotese nula de raíz unitária do teste ADF
- Assim, temos duas séries estacionárias e podemos proceder a estimação de um modelo VAR(p) bivariado
- Para a identificação da ordem do coeficiente $p$ podemos utilizar os critérios de Akaike, Schwarz, Hannan-Quin, etc.

# Identificação

A função `VARselect` indica o número de defasagens que minimiza os critérios de identificação:

```
VARselect(base, lag.max = 5)
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      5      1      1      5
##
## $criteria
##                 1          2         3          4          5
## AIC(n) -1.3950537 -1.3933323 -1.430640 -1.4416669 -1.4505174
## HQ(n)  -1.3522342 -1.3219665 -1.330728 -1.3132084 -1.2935126
## SC(n)  -1.2894271 -1.2172880 -1.184178 -1.1247871 -1.0632198
## FPE(n)  0.2478212  0.2482536  0.239174  0.2365714  0.2345181
```

# Estimação

Com o objetivo de parcimoniosidade neste exercício, vamos estimar um modelo VAR(1) bivariado, conforme o indicado pelo critério bayesiano de Schwarz

```
var_model <- vars::VAR(base, p = 1, type = "const")
summary(var_model)
```

A função `VAR` do pacote `vars` estima o modelo considerando todas as variáveis da base como endógenas e no argumento `type` é possível incluir constantes ou tendências

# Estimação

```
...
## Estimation results for equation Consumption:
## ============================================
## Consumption = Consumption.l1 + Income.l1 + const
##
##                Estimate Std. Error t value Pr(>|t|)
## Consumption.l1  0.29646    0.07485   3.961 0.000107 ***
## Income.l1       0.09434    0.05267   1.791 0.074908 .
## const           0.45811    0.06993   6.551 5.63e-10 ***
## ---
##
## Estimation results for equation Income:
## =======================================
## Income = Consumption.l1 + Income.l1 + const
##
##                Estimate Std. Error t value Pr(>|t|)
## Consumption.l1  0.51509    0.10728   4.801 3.27e-06 ***
## Income.l1      -0.25474    0.07549  -3.375 0.000903 ***
## const           0.51462    0.10024   5.134 7.21e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
...
```

# Estimação

- O modelo indica que a renda passada tem algum potencial para prever o consumo, sendo significante ao nível de 10%
- Já o consumo passado prevê fortemente a renda futura, devido ao fato que o aumento das vendas contribui para aumentar a renda da economia
- O consumo tem um componente autorregressivo significativo, indicando persistência nos hábitos de consumo
- Já a renda tem um coeficiente autorregressivo negativo, indicando uma tendência de reversão a média

# Diagnóstico

- Para avaliar a qualidade da especificação do modelo devemos analisar os resíduos através de um teste de autocorrelação multivariado
- Este teste pode ser realizado através da função `serial.test` no `R`, sendo que a rejeição da hipótese nula indica autocorrelação residual
- Assim, aplicaremos o teste as defasagens 10 e 15

# Diagnóstico

```r
serial.test(var_model, lags.pt=10, type="PT.asymptotic")
```

```
##
##  Portmanteau Test (asymptotic)
##
## data:  Residuals of VAR object var_model
## Chi-squared = 49.102, df = 36, p-value = 0.07144
```

```r
serial.test(var_model, lags.pt=15, type="PT.asymptotic")
```

```
##
##  Portmanteau Test (asymptotic)
##
## data:  Residuals of VAR object var_model
## Chi-squared = 56.858, df = 56, p-value = 0.4429
```

# Diagnóstico

- Em geral, quando utilizamos um número maior de defasagens no teste os resíduos parecem não estar correlacionados
- Entretanto, o teste com apenas 10 defasagens rejeita a hipótese nula de não autocorrelção ao nível de 10%
- Para resolver essa possível autocorrelação residual em defasagens mais baixas podemos acrescentar mais defasagens no nosso modelo VAR
- Neste exemplo, ficaremos com o modelo VAR(1) com o objetivo de manter a parcimoniosidade

# Causalidade de Granger

- Em um modelo VAR(1), o teste de causalidade de Granger é equivalente ao p-valor dos coeficientes das regressões do modelo VAR
- Entretanto, quando temos um número maior de defasagens, para avaliar a causalidade de Granger devemos realizar um teste conjunto de significância de todos os coeficientes
- Podemos realizar este teste através da função `causality` no 'R'
- Como temos duas variáveis endógenas no nosso modelo, devemos realizar dois teste, o primeiro com o consumo como variável causal e o segundo com a renda como variável causal

# Causalidade de Granger

```r
causality(var_model, cause = "Consumption")
```

```
## $Granger
##
##  Granger causality H0: Consumption do not Granger-cause Income
##
## data:  VAR object var_model
## F-Test = 23.053, df1 = 1, df2 = 366, p-value = 2.301e-06
##
##
## $Instant
##
##  H0: No instantaneous causality between: Consumption and Income
##
## data:  VAR object var_model
## Chi-squared = 24.227, df = 1, p-value = 8.561e-07
```

# Causalidade de Granger

```r
causality(var_model, cause = "Income")
```

```
## $Granger
##
##  Granger causality H0: Income do not Granger-cause Consumption
##
## data:  VAR object var_model
## F-Test = 3.2085, df1 = 1, df2 = 366, p-value = 0.07408
##
##
## $Instant
##
##  H0: No instantaneous causality between: Income and Consumption
##
## data:  VAR object var_model
## Chi-squared = 24.227, df = 1, p-value = 8.561e-07
```

# Causalidade de Granger

- Assim como identificado nos resultados das regressões do modelo VAR, o consumo granger causa a renda com significância de 1%
- Entretanto, a renda granger causa o consumo apenas com significância de 10%
- Para visualizar a dinâmica deste efeitos causais entre as duas variáveis podemos plotar a função impulso-resposta
- Esta função simula um choque em uma das variáveis e calcula a resposta ao longo do tempo que este choque causa em outra variável

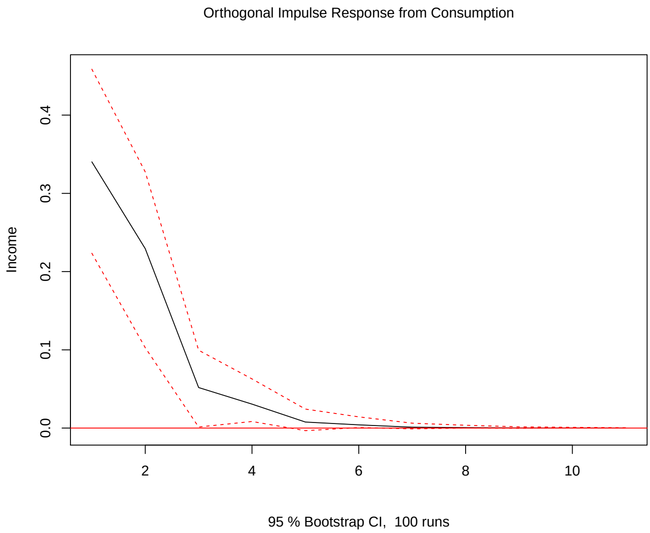
# Funções impulso-resposta

No `R` podemos estimar a função de impulso-resposta através do comando `irf`, utilizando como argumento o modelo estimado, a variável impulso e a variável resposta:

```
impulse_cons <- irf(
  var_model,
  impulse = "Consumption",
  response = "Income"
)
plot(impulse_cons)
```

Após estimar a função impulso-resposta do consumo na renda, podemos plotá-la utilizando a comando `plot`

# Funções impulso-resposta

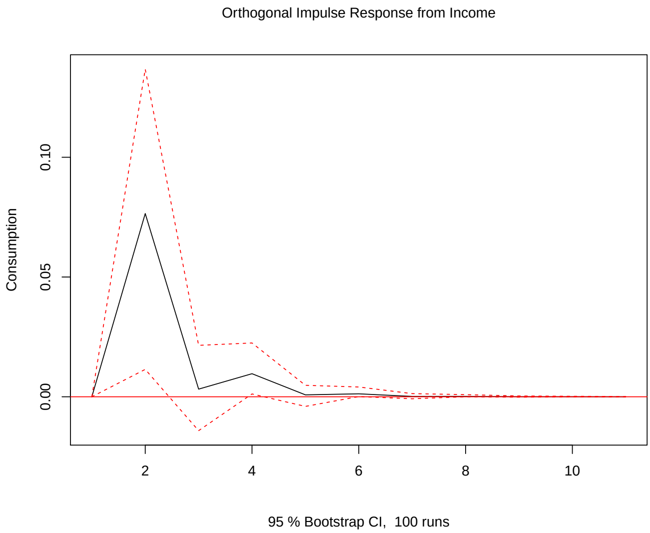


# Funções impulso-resposta

De maneira semelhante, vamos calcular a função impulso-resposta da renda no consumo:

```r
impulse_inc <- irf(
  var_model,
  impulse = "Income",
  response = "Consumption"
)
plot(impulse_inc)
```

# Funções impulso-resposta



# Previsão

Por fim, podemos fazer a previsão das duas variáveis endógenas utilizando o modelo VAR estimado através da função forecast:

```
var_fc <- forecast(var_model, h = 4)
var_fc
```

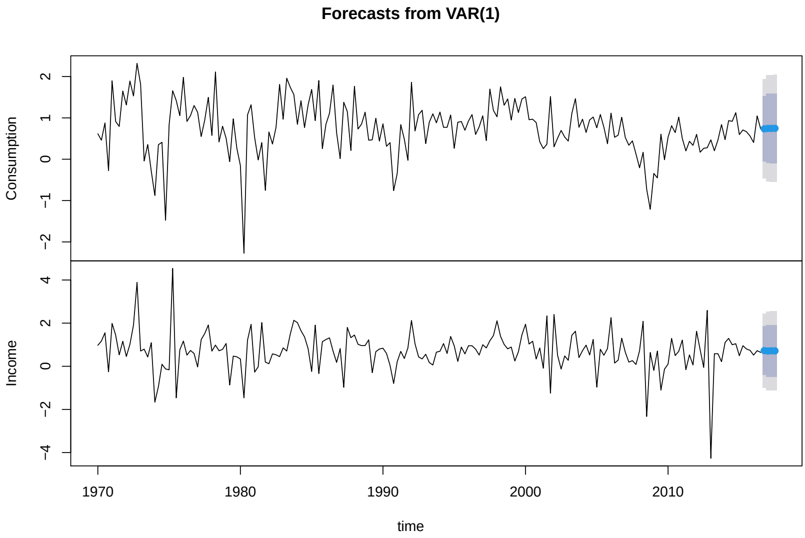
E plotar o resultado final:

```
plot(var_fc)
```

# Previsão

```
## Consumption
##         Point Forecast       Lo 80    Hi 80      Lo 95    Hi 95
## 2016 Q4      0.7352263 -0.05100310 1.521456 -0.4672078 1.937660
## 2017 Q1      0.7445835 -0.09384705 1.583014 -0.5376854 2.026852
## 2017 Q2      0.7456724 -0.09982648 1.591171 -0.5474065 2.038751
## 2017 Q3      0.7468792 -0.09958950 1.593348 -0.5476829 2.041441
##
## Income
##         Point Forecast      Lo 80    Hi 80      Lo 95    Hi 95
## 2016 Q4      0.7261971 -0.4006792 1.853073 -0.9972115 2.449606
## 2017 Q1      0.7083358 -0.4859362 1.902608 -1.1181456 2.534817
## 2017 Q2      0.7177056 -0.4842089 1.919620 -1.1204640 2.555875
## 2017 Q3      0.7158796 -0.4869998 1.918759 -1.1237657 2.555525
```

# Previsão



Forecasts from VAR(1)
Consumption
Income
time
2
1
0
−1
−2
4
2
0
−2
−4
1970
1980
1990
2000
2010