

# Modèle VAR stationnaire : revenu disponible et consommation des ménages aux Etats-Unis

Yannick Le Pen

2023-24

On utilise un modèle VAR pour étudier les relations entre le taux de croissance du revenu disponible et le taux de croissance de la consommation des ménages aux Etats-Unis. On importe des données réelles, désaisonnalisées et trimestrielles à partir du site DB.nomics.

On verra successivement :

1. l'estimation d'un modèle VAR
2. les tests de causalité au sens de Granger
3. l'analyse impulsion réponse
4. la prévision

De plus, on étudiera l'impact de l'inclusion de la période covid sur les estimations et comment essayer de prendre en compte cette période particulière.

## Importation des données depuis DB.nomics

```
rm(list=ls())
library(data.table)
library(rdbnomics)

# données sur le revenu disponible réel
data_inc<-rdb(ids = "BEA/NIPA-T20100/A067RX-Q")

# données sur les dépenses de consommation personnelles en volume
data_cons<-rdb(ids = "BEA/NIPA-T10106/DPCERX-Q")

# données sur le revenu disponible réel excluant les transferts sociaux
data_inc_wt<-rdb(ids = "BEA/NIPA-T20100/W875RX-Q")
```

On sélectionne les dates et les séries pour le revenu et la consommation qui sont fusionnées dans un seul dataframe. On conserve les données à partir de "1990-01-01".

```
df_inc<-data_inc[,c("period","value")]
colnames(df_inc)<-c("date","income")
df_cons<-data_cons[,c("period","value")]
colnames(df_cons)<-c("date","consumption")
df<-merge(df_inc,df_cons,by="date")
df<-df[["1980-01-01"<=df$date,]
```

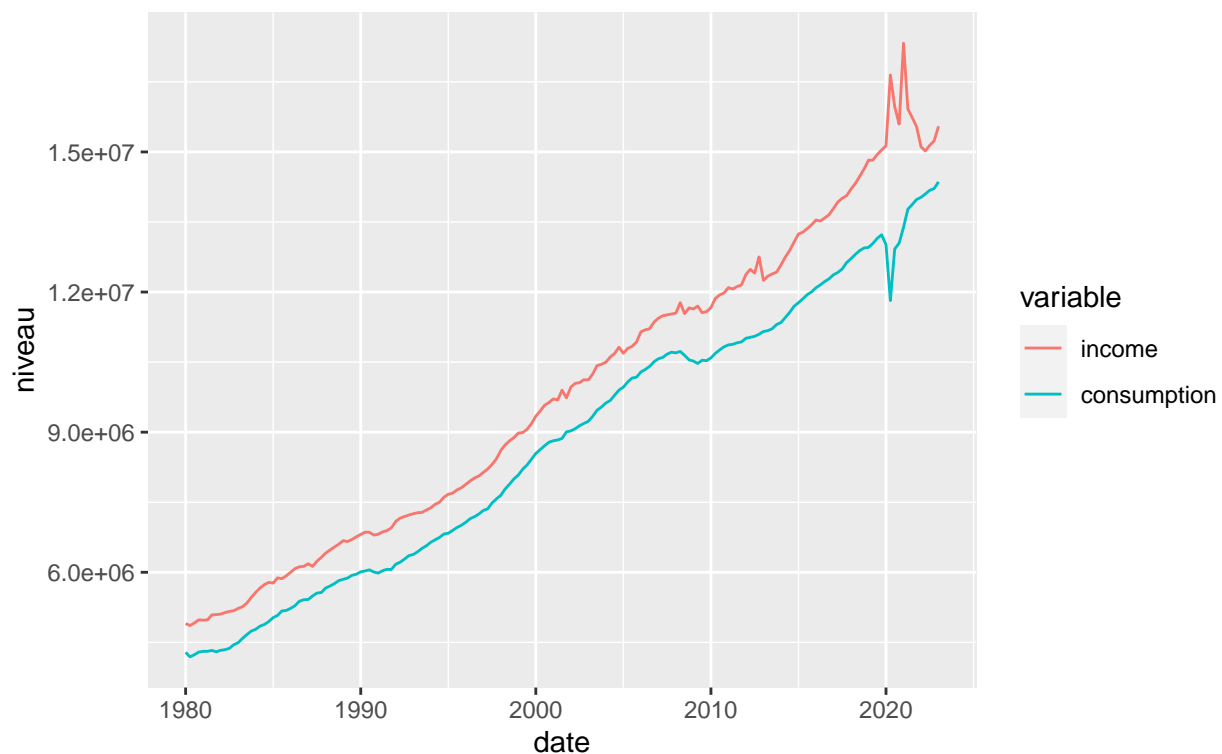
Les deux séries sont représentées sur le graphique ci-dessous. Les deux séries augmentent à un rythme régulier et en parallèle de 1990 à 2008. Les années 2008 et 2009 sont marquées par une baisse des deux séries, plus grande pour la consommation. Les deux séries connaissent ensuite une hausse régulière, à un rythme moins

élevé qu'avant 2008, jusqu'au déclenchement de l'épidémie de covid au début de l'année 2020. On observe alors une forte baisse temporaire de la consommation des ménages, qui retrouve ensuite son rythme de croissance. En raison des transferts sociaux aux ménages, le revenu disponible augmente fortement à deux reprises.

```
library(ggplot2)
library(reshape2)
data_melt=melt(as.data.frame(df),id.vars = 1)

p_us_T<-ggplot(data=data_melt,aes(x=date))+geom_line(aes(x=date,y=value,color=variable))
p_us_T+labs(y="niveau",title = "Revenu et consommation des ménages aux Etats-Unis",
            subtitle = "1990:T1 à 2023:T1")
```

Revenu et consommation des ménages aux Etats-Unis  
1990:T1 à 2023:T1



## Etude sur la période avant covid : 1990:T1 à 2019:T4

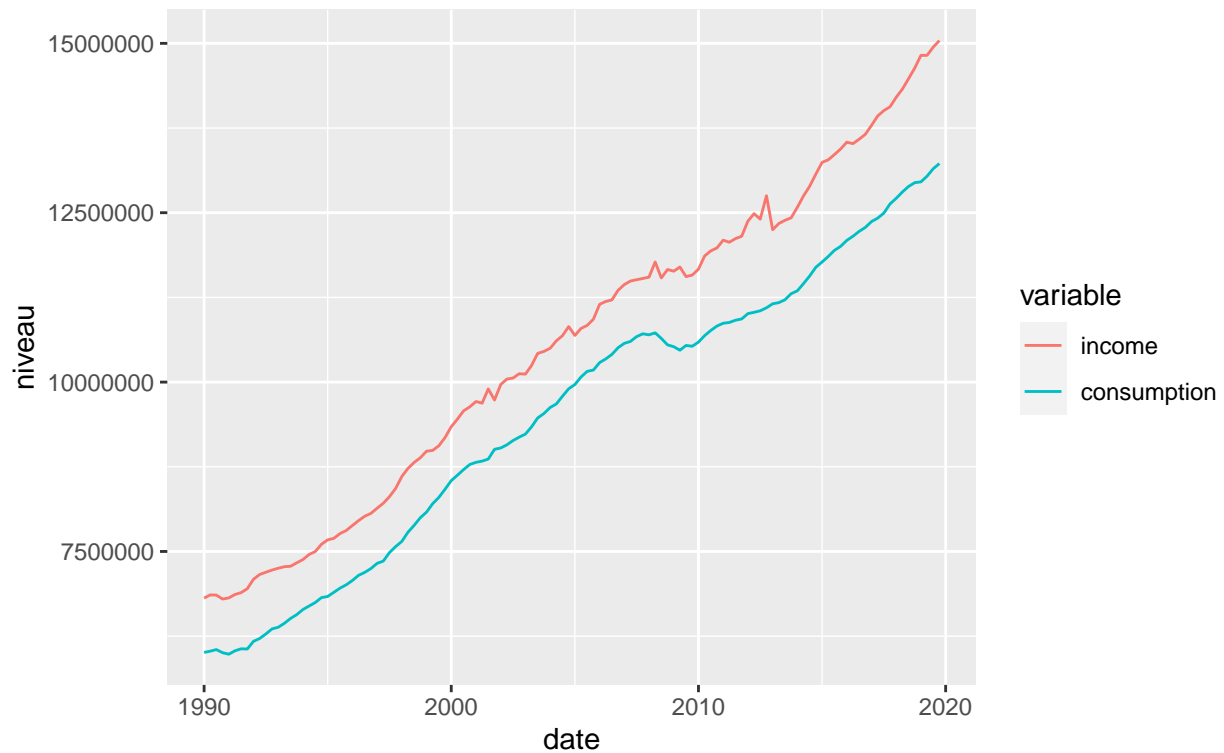
En raison de la spécificité de la période du covid, on se limite dans un premier temps à la période 1990:T1 à 2020:T1.

```
df_short<-df[("1990-01-01"<=df$date & df$date<"2020-01-01"),]

data_melt=melt(as.data.frame(df_short),id.vars = 1)

p_us<-ggplot(data=data_melt,aes(x=date))+geom_line(aes(x=date,y=value,color=variable))
p_us+labs(y="niveau",title = "Revenu et consommation des ménages aux Etats-Unis",
          subtitle = "1990:T1 à 2019:T4")
```

## Revenu et consommation des ménages aux Etats-Unis 1990:T1 à 2019:T4



### Taux de croissance du revenu disponible et de la consommation

On calcule les taux de croissance comme la différence première de variables en logarithme.

```
df_g_short<-data.frame(df_short$date[-1],diff(log(df_short$income)),diff(log(df_short$consumption)))  
# on perd une observation en calculant le taux de croissance  
colnames(df_g_short)<-c("date","g_inc","g_cons")  
  
data_melt=melt(as.data.frame(df_g_short),id.vars = 1)  
  
p_g_us<-ggplot(data=data_melt,aes(x=date))+geom_line(aes(x=date,y=value,color= variable))  
p_g_us+labs(y="taux",title = "Variation du revenu et de la consommation des ménages aux Etats-Unis",  
            subtitle = "1990:T1 à 2019:T4")
```

## Variation du revenu et de la consommation des ménages aux Etats-Unis 1990:T1 à 2019:T4



```
summary(df_g_short[c("g_inc", "g_cons")])
```

### Statistiques descriptives

```
##      g_inc      g_cons
## Min.   :-0.040186  Min.   :-0.009040
## 1st Qu.: 0.003696  1st Qu.: 0.003943
## Median : 0.007388  Median : 0.006702
## Mean   : 0.006658  Mean   : 0.006628
## 3rd Qu.: 0.010665  3rd Qu.: 0.009515
## Max.   : 0.027518  Max.   : 0.018550
```

```
sapply(df_g_short[c("g_inc", "g_cons")], sd)
```

```
##      g_inc      g_cons
## 0.008440164 0.004900454
```

Les statistiques descriptives montrent que les moyennes des deux taux de croissance sont quasiment égales. Le taux de croissance du revenu est plus variable que celui de la consommation, comme le montrent les écarts types, les maxima et minima.

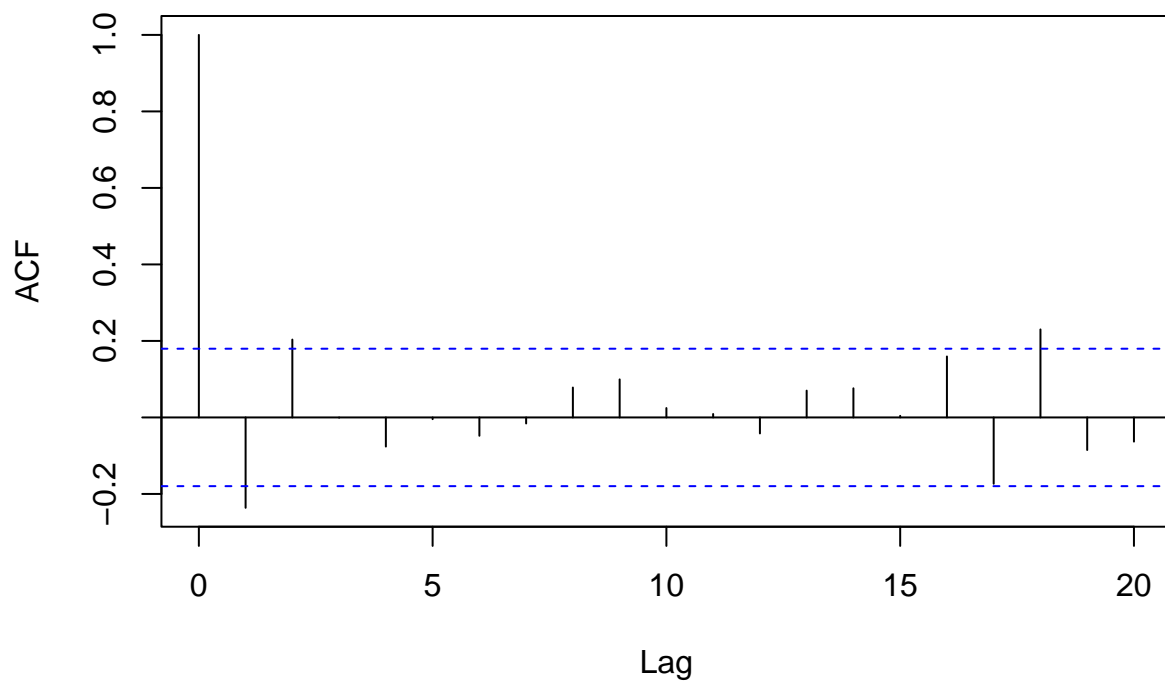
La corrélation entre les deux taux de croissance est égale à 0.28 et significative.

```
cor.test(df_g_short$g_inc, df_g_short$g_cons)
```

```
##
## Pearson's product-moment correlation
##
```

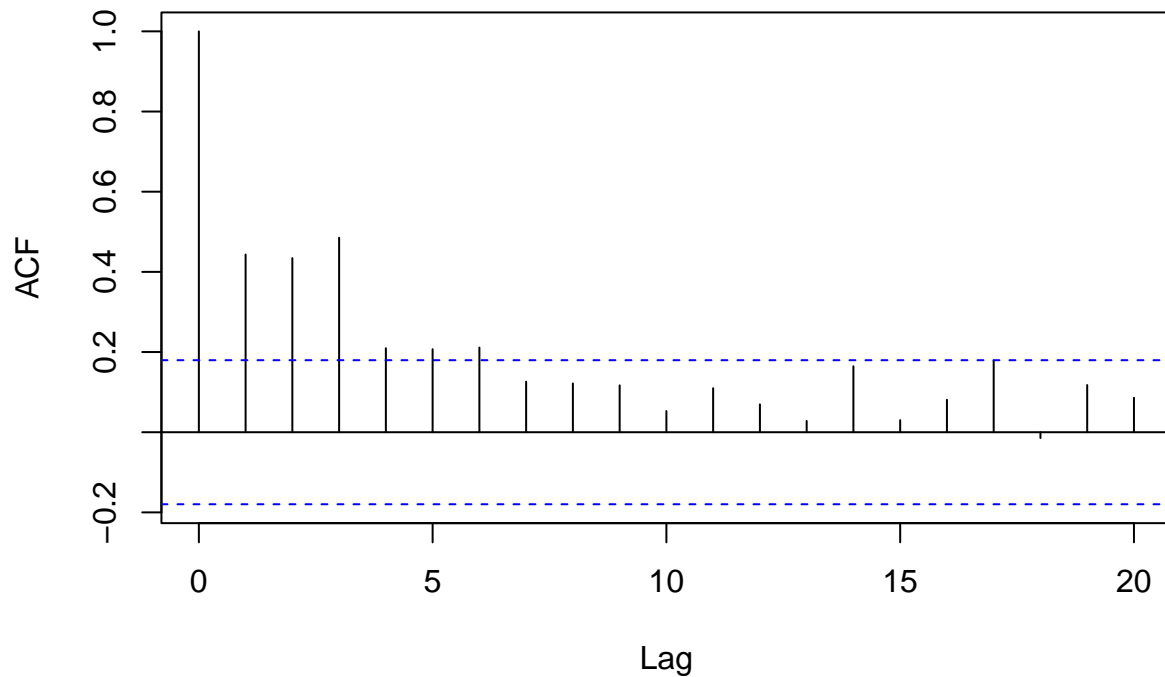
```
## data: df_g_short$g_inc and df_g_short$g_cons
## t = 3.1633, df = 117, p-value = 0.001988
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.1060522 0.4385294
## sample estimates:
## cor
## 0.2806897
acf(df_g_short$g_inc,main="Autocorrelogramme du taux de croissance du revenu disponible")
```

## Autocorrelogramme du taux de croissance du revenu disponible



```
acf(df_g_short$g_cons,main="Autocorrélogramme du taux de croissance de la consommation")
```

## Autocorrélogramme du taux de croissance de la consommation



## Modélisation VAR

On utilise la librairie “vars” pour estimer le modèle VAR.

### Sélection du nombre de retards optimal

On commence par déterminer le nombre de retards par l’application de critères d’information multivariés. On doit spécifier :

- lag.max = nombre maximal de retards
- type = forme de la composante déterministe “const”, “trend”, “both” ou “none” (1 seule spécification possible)

On choisit un nombre de retards maximum égal à 8 (correspond à deux années de recul pour des données trimestrielles) et l’on estime un modèle VAR avec une constante (la moyenne des taux de croissance n’est pas nulle.)

```
library(vars)
pselect<-VARselect(y=df_g_short[,c(2,3)],lag.max=8,type=c("const"))
pselect$criteria # Affichage des valeurs des critère de sélection
```

##	1	2	3	4	5
## AIC(n)	-2.078390e+01	-2.084549e+01	-2.087726e+01	-2.090670e+01	-2.086971e+01
## HQ(n)	-2.072448e+01	-2.074646e+01	-2.073862e+01	-2.072846e+01	-2.065186e+01
## SC(n)	-2.063744e+01	-2.060139e+01	-2.053552e+01	-2.046732e+01	-2.033269e+01
## FPE(n)	9.411924e-10	8.850601e-10	8.575651e-10	8.329976e-10	8.649020e-10
##	6	7	8		
## AIC(n)	-2.081628e+01	-2.078384e+01	-2.073989e+01		

```
## HQ(n) -2.055882e+01 -2.048677e+01 -2.040320e+01
## SC(n) -2.018162e+01 -2.005154e+01 -1.990994e+01
## FPE(n) 9.131489e-10 9.443564e-10 9.883045e-10
```

```
pselect$selection # nombre de retards optimal selon les critères de sélection
```

```
## AIC(n) HQ(n) SC(n) FPE(n)
##      4      2      1      4
```

Le nombre de retard optimal varie selon le critère :

- le critère AIC préconise 4 retards,
- le critère HQ préconise 2 retards,
- le critère SC préconise 1 retard,
- le nombre de retards optimaux sont néanmoins inférieurs à lag.max

Le critère AIC est connu pour surestimer le nombre de retards. On décide d'estimer le modèle VAR avec 1 retard.

### Estimation d'un modèle VAR(1)

```
var.1lag <-VAR(y = df_g_short[,c("g_inc","g_cons")],type="const",lag.max = 1)
summary(var.1lag) #Affichage des résultats avec summary
```

```
##
## VAR Estimation Results:
## =====
## Endogenous variables: g_inc, g_cons
## Deterministic variables: const
## Sample size: 118
## Log Likelihood: 888.19
## Roots of the characteristic polynomial:
## 0.4705 0.4153
## Call:
## VAR(y = df_g_short[, c("g_inc", "g_cons")], type = "const", lag.max = 1)
##
##
## Estimation results for equation g_inc:
## =====
## g_inc = g_inc.l1 + g_cons.l1 + const
##
##              Estimate Std. Error t value Pr(>|t|)
## g_inc.l1 -0.326877    0.089475  -3.653 0.000391 ***
## g_cons.l1  0.556250    0.154122   3.609 0.000456 ***
## const      0.005141    0.001259   4.085 8.19e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.007874 on 115 degrees of freedom
## Multiple R-Squared: 0.1519, Adjusted R-squared: 0.1371
## F-statistic: 10.3 on 2 and 115 DF, p-value: 7.7e-05
##
##
## Estimation results for equation g_cons:
## =====
```

```
## g_cons = g_inc.l1 + g_cons.l1 + const
##
##           Estimate Std. Error t value Pr(>|t|)
## g_inc.l1  0.1266923  0.0490151   2.585   0.011 *
## g_cons.l1 0.3821523  0.0844294   4.526 1.47e-05 ***
## const     0.0032794  0.0006895   4.756 5.77e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.004313 on 115 degrees of freedom
## Multiple R-Squared: 0.2415, Adjusted R-squared: 0.2283
## F-statistic: 18.31 on 2 and 115 DF, p-value: 1.248e-07
##
##
## Covariance matrix of residuals:
##           g_inc    g_cons
## g_inc    6.199e-05 1.036e-05
## g_cons    1.036e-05 1.860e-05
##
## Correlation matrix of residuals:
##           g_inc g_cons
## g_inc     1.0000 0.3051
## g_cons     0.3051 1.0000
serial.test(var.1lag, lags.pt = 10, type = "PT.adjusted")
##
## Portmanteau Test (adjusted)
##
## data:  Residuals of VAR object var.1lag
## Chi-squared = 45.002, df = 36, p-value = 0.1444
```

### Commentaires des résultats

$g\_inc.li$  et  $g\_cons.li$  représentent respectivement  $g\_inc$  et  $g\_cons$  avec un retard de  $i$  périodes.

- Equation du taux de croissance  $g\_inc$ 
  - le coefficient de  $g\_inc.l1$  est significatif et négatif pour un risque de première espèce de 1% : on a tendance à observer une alternance d'accélération et de ralentissement du taux de croissance trimestriel du revenu disponible.
  - le coefficient de  $g\_cons.l1$  est positif et significatif pour un risque de première espèce 1%,
  - la constante est significative pour un risque de première espèce de 1%.
  - le  $R^2$  est égal à 15%.
- Equation du taux de croissance  $g\_cons$ 
  - le coefficient de  $g\_inc.l1$  est positif et significatif pour un risque de première espèce de 1% ; une accélération de la croissance du revenu disponible a pour effet d'accélérer la croissance de la consommation
  - le coefficient de  $g\_cons.l1$  est positif et significatif pour un risque de première espèce 1% : le taux de croissance de la consommation est positivement corrélé à sa valeur passée
  - la constante est significative pour un risque de première espèce de 1%,
  - le  $R^2$  est égal à 24% : le pouvoir explicatif de la régression est plus important que pour  $g\_inc$

On peut afficher les graphiques des valeurs ajustées et des résidus pour chaque variables du VAR.



```
#par(mar=c(1, 1, 1, 1))
#plot(var.1lag)
```

**Affichage des racines du polynôme caractéristique** On étudie la stationnarité du VAR. La fonction `roots` renvoie les module des inverses des racines du polynôme caractéristique.

```
roots(var.1lag)
```

```
## [1] 0.4705295 0.4152543
```

Les inverses des racines sont de module strictement inférieur à 1 : la condition de stationnarité est donc satisfaite.

**Affichage de résultats** l'objet `var.1lag` contient les résultats du VAR auxquels on peut avoir accès en spécifiant leur nom.

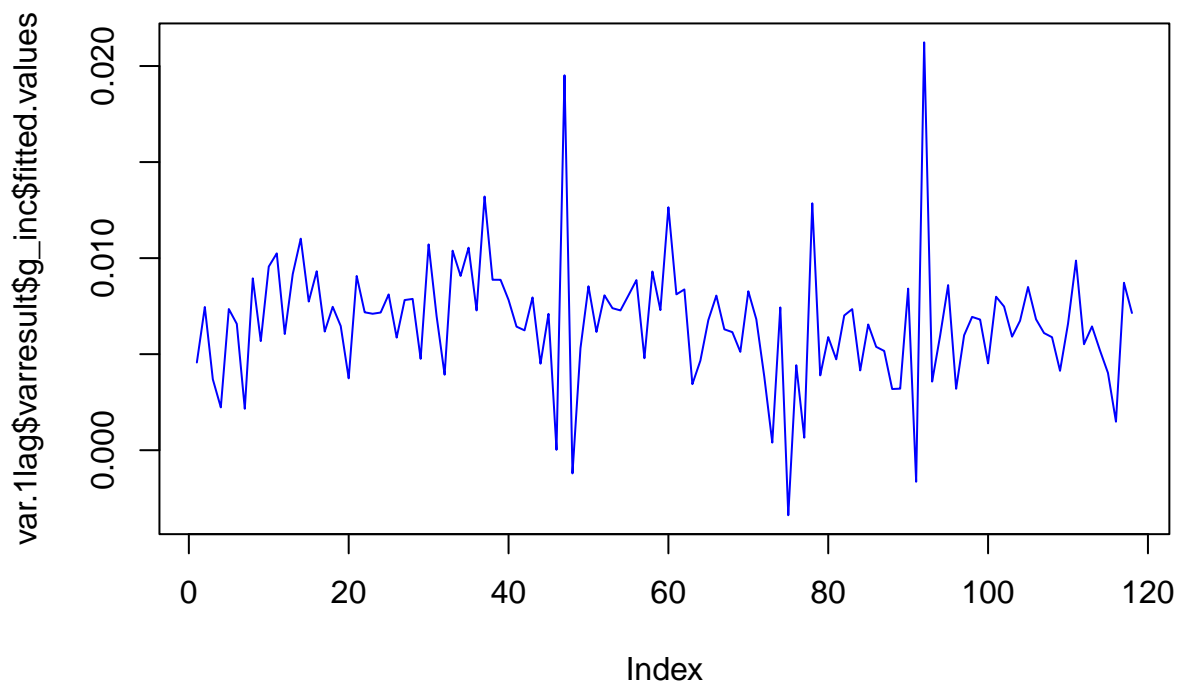
- **Exemple 1 : Liste des matrices des coefficients estimés du VAR**

```
A<-Acoef(var.1lag)
A
```

```
## [[1]]
##      g_inc.l1 g_cons.l1
## g_inc -0.3268771 0.5562500
## g_cons 0.1266923 0.3821523
```

- **Exemple 2 : Variable ajustée du taux de croissance du revenu `g_inc`**

```
plot(var.1lag$varresult$g_inc$fitted.values,type='line',col='blue')
```



## Test de causalité au sens de Granger

**Causalité de la consommation sur le revenu** On vérifie la significativité du coefficient de  $g\_conso_{t-1}$  dans l'équation de  $g\_income_t$ . Les hypothèses du test sont :

$$\begin{cases} H_0 : a_{12}^1 = 0 \Rightarrow g\_conso \text{ ne cause pas } g\_income \\ H_a : a_{12}^1 \neq 0 \Rightarrow g\_conso \text{ cause } g\_income \end{cases}$$

où  $a_{12}^1$  est le coefficient de la ligne 1 et colonne 2 de  $A$ . Les résultats du test figurent ci-dessous :

```
causality(var.1lag,cause=c('g_cons'))
```

```
## $Granger
##
##   Granger causality H0: g_cons do not Granger-cause g_inc
##
## data:  VAR object var.1lag
## F-Test = 13.026, df1 = 1, df2 = 230, p-value = 0.000377
##
##
## $Instant
##
##   H0: No instantaneous causality between: g_cons and g_inc
##
## data:  VAR object var.1lag
## Chi-squared = 10.046, df = 1, p-value = 0.001527
```

Les seuils critiques sont tirés d'une loi  $F(1, 224)$  :

$\alpha$	5%	10%
$f_{1-\alpha}(1, +\infty)$	2.99	4.60

La statistique de test égale à 5.93 est supérieure au seuil critique pour  $\alpha = 1\%$ . On rejette donc l'hypothèse nulle d'absence de causalité du taux de croissance de la consommation sur le taux de croissance du revenu. La probabilité critique du test est égale à 0,00% ce qui amène aux mêmes conclusions.

**Causalité du revenu sur la consommation** On teste si le taux de croissance du revenu "cause" le taux de croissance de la consommation au sens de Granger.

On vérifie la significativité du coefficient de  $g\_income_{t-1}$  dans l'équation de  $g\_consumption_t$ . Les hypothèses du test sont :

$$\begin{cases} H_0 : a_{21}^1 = 0 \Rightarrow g\_income \text{ ne cause pas } g\_consumption \\ H_a : a_{21}^1 \neq 0 \Rightarrow g\_income \text{ cause } g\_consumption \end{cases}$$

où  $a_{21}^1$  est le coefficient de la ligne 2 et colonne 1 de  $A_1$ . Les résultats du test figurent ci dessous :

```
causality(var.1lag,cause=c('g_inc'))
```

```
## $Granger
##
##   Granger causality H0: g_inc do not Granger-cause g_cons
##
## data:  VAR object var.1lag
## F-Test = 6.681, df1 = 1, df2 = 230, p-value = 0.01036
##
##
## $Instant
##
```

```
## H0: No instantaneous causality between: g_inc and g_cons
##
## data: VAR object var.1lag
## Chi-squared = 10.046, df = 1, p-value = 0.001527
```

La statistique de test égale à 2.58 est supérieure aux seuils critiques. On rejette donc l'hypothèse nulle d'absence de causalité du taux de croissance du revenu sur le taux de croissance de la consommation pour un risque de première espèce de 5%. La probabilité critique du test est égale à 0.020 ce qui amène aux mêmes conclusions.

**Conclusion** On observe que le taux de croissance de la consommation le taux de croissance du revenu au sens de Granger et réciproquement. La causalité s'exerce dans les deux sens et l'on parle alors d'**effet feedback**.

## Fonctions impulsion-réponse

Comme chacune des deux variables du VAR cause l'autre, les tests de causalité au sens de Granger ne permettent pas de dégager un ordre de classement des variables. Néanmoins on peut raisonnablement penser que les chocs sur le taux de croissance du revenu impactent instantanément le taux de croissance de la consommation plutôt que l'inverse. On décide de "placer" le taux de croissance du revenu `g_income` en première ligne et le taux de croissance de la consommation `g_conso` en seconde position.

On calcule les fonctions impulsion réponse :

- `var.1lag` : objet
- `n.ahead` : nombre de périodes futures pour lesquelles on calcule les impulsion réponse
- `ortho = TRUE` : on applique la décomposition de Cholesky
- `ci = 0.95` : on représente les bornes de l'intervalle de confiance à 95 % pour déterminer la significativité des IRF.

Les chocs sont égaux à 1 écart type du choc structurel.

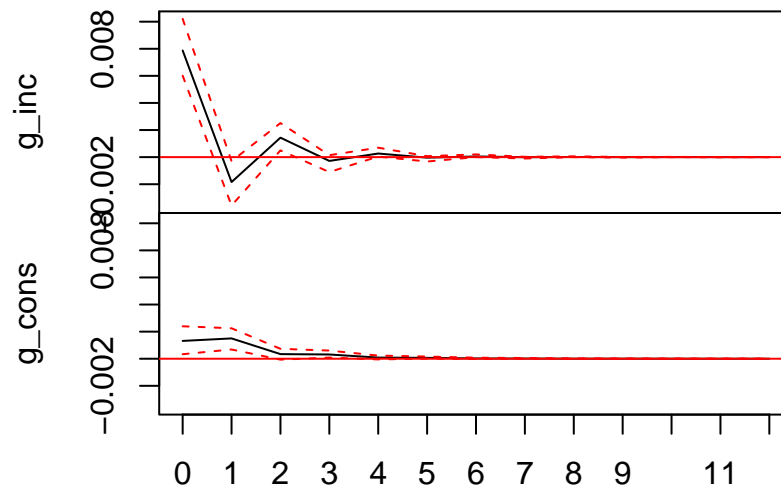
### Important

- En spécifiant les variables du VAR(1) on a écrit `var.1lag <- VAR(y = df_g_short[,c("g_inc", "g_cons")], type="const", lag.max = 1)` : ceci implique que la variable située en première ligne est `g_inc` et celle en seconde position `g_cons`.
- Si l'on souhaitait mettre `g_cons` en première ligne, il faudrait écrire

```
var.1lag <- VAR(y = df_g_short[,c("g_cons", "g_inc")], type="const", lag.max = 1)
```

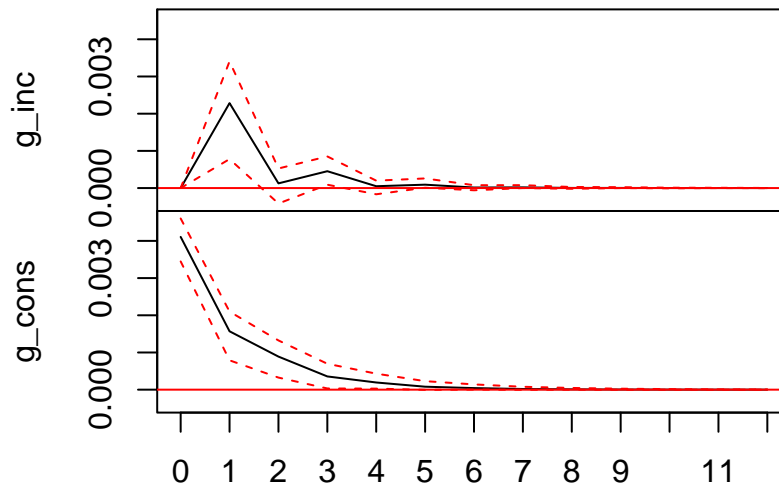
```
irf<-irf(var.1lag,n.ahead = 12,ortho = TRUE,ci=0.95)
plot(irf)
```

Orthogonal Impulse Response from g\_inc



95 % Bootstrap CI, 100 runs

### Orthogonal Impulse Response from g\_cons



95 % Bootstrap CI, 100 runs

- Effet d'un choc sur  $g\_inc$  : \*\* un effet immédiat, positif et significatif sur le taux de croissance de la consommation jusqu'au second trimestre.
- Effet d'un choc sur  $g\_cons$  : \*\* pas d'impact immédiat sur le taux de croissance du revenu disponible, \*\* un effet positif sur le revenu pour le trimestre suivant (à l'horizon  $h=1$ )

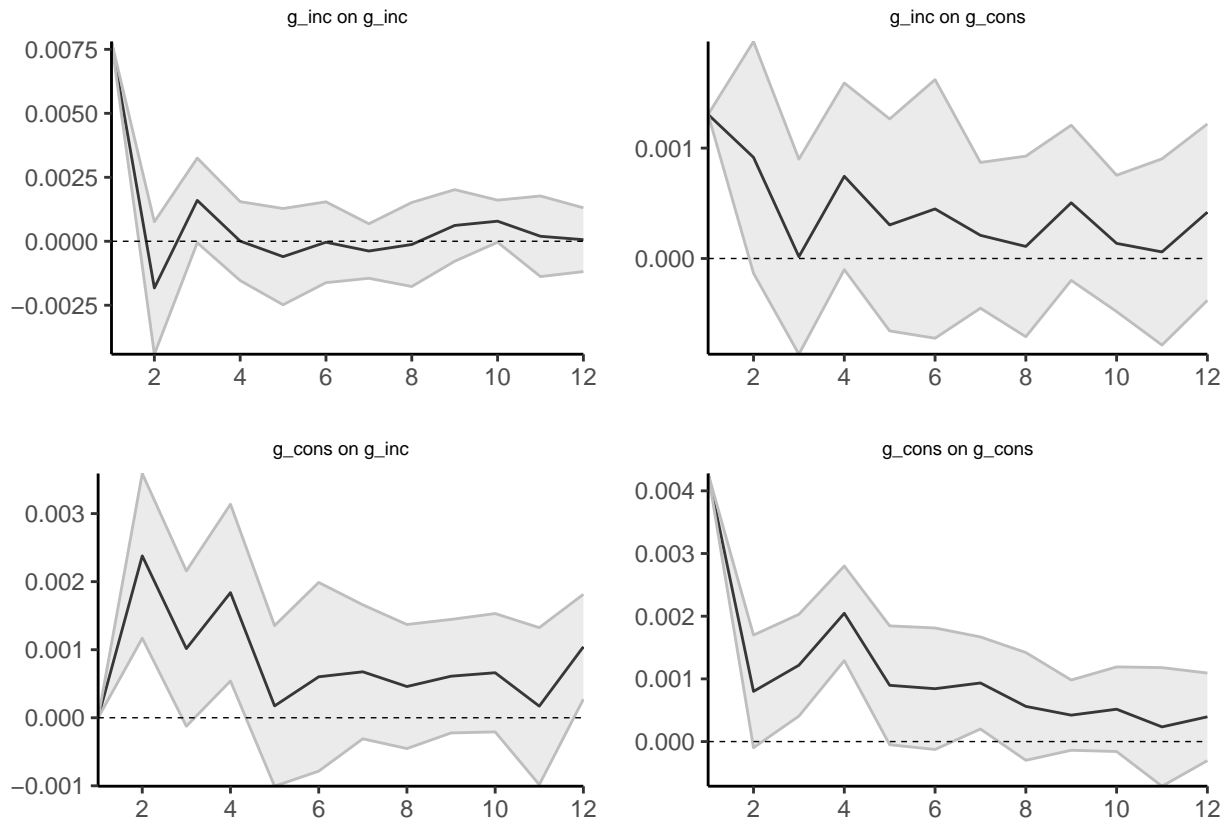
### Estimation des IRF par les projections locales

On estime les IRF par la méthode des projections locales. On spécifie les paramètres suivants :

- le nombre d'horizon est égal à 12
- Pour chaque horizon  $h = 1, \dots, 12$ , le nombre de retards dans la régression de la projection locale sera au plus égal à 8 ( $max\_lags = 8$ ) et choisi par la minimisation du critère BIC
- on applique des chocs égaux à 1 écart type du choc structurel :  $shock\_type = 0$ . ( $shock\_type = 1$  pour des chocs)
- Intervalle de confiance à 95% ( $confit=1.96$ ) ( $confit = 1.65$  pour un intervalle de confiance à 90% et  $confit=1$  pour un intervalle de confiance à 68%)

```
library(lpirfs)
irf_LP <- lp_lin(endog_data = df_g_short[, c(2, 3)],
                 lags_endog_lin = NaN,
                 lags_criterion = 'BIC',
                 max_lags = 8,
                 trend = 0,
                 shock_type = 0,
                 confint=1.96,
                 hor=12,
                 adjust_se=TRUE)
```

```
plot(irf_LP)
```



**\*\* Commentaires\*\***

- impact d'un choc sur g\_inc : peu de différence avec les fonctions obtenues à partir du VAR(1) (graphiques du haut)
- impact d'un choc sur g\_cons :
  - l'impact du choc est plus durable sur g\_inc
  - l'impact du choc est aussi plus durable sur g\_cons
  - Explication possible : les fonctions impulsion réponse sont estimées en déterminant de façon endogène le nombre de retards pour chaque horizon.

On peut afficher les valeurs de fonctions impulsion-réponse pour les chocs sur g\_inc (variable 1 du dataframe)

```
irf_LP$irf_lin_mean[, ,1]
```

```
##           [,1]           [,2]           [,3]           [,4]           [,5]
## [1,] 0.007805978 -0.0018259760 1.597588e-03 6.065962e-06 -0.0006009981
## [2,] 0.001304485 0.0009158394 1.655681e-05 7.445147e-04 0.0003039722
##           [,6]           [,7]           [,8]           [,9]          [,10]
## [1,] -3.764449e-05 -0.0003814607 -0.0001256139 0.0006180958 0.0007847352
## [2,] 4.490305e-04 0.0002099041 0.0001092348 0.0005044560 0.0001368489
##           [,11]          [,12]          [,13]
## [1,] 1.946615e-04 0.0000613235 -0.0003254562
## [2,] 5.939004e-05 0.0004195909 -0.0002422048
```

On peut afficher les valeurs de fonctions impulsion-réponse pour les chocs sur g\_cons (variable 2 du dataframe)

```
irf_LP$irf_lin_mean[,2]
```

```
##           [,1]           [,2]           [,3]           [,4]           [,5]           [,6]
## [1,] 0.000000000 0.0023786263 0.001016169 0.001838470 0.0001739160 0.0006017788
## [2,] 0.004276182 0.0008001524 0.001218040 0.002046363 0.0008972719 0.0008430222
##           [,7]           [,8]           [,9]           [,10]          [,11]
## [1,] 0.0006749255 0.0004587258 0.0006105987 0.0006615307 0.0001703716
## [2,] 0.0009345842 0.0005609018 0.0004214367 0.0005158122 0.0002336553
##           [,12]          [,13]
## [1,] 0.001042507 -0.0004170949
## [2,] 0.000394483  0.0004069005
```

L’affichage ci-dessous montre que le nombre de retards optimal inclus dans la projection locale varie selon la variable l’horizon, la variable qui subit un choc et celle dont on étudie la dynamique.

```
irf_LP$specs$chosen_lags
```

```
## $`Shock: g_inc`
## $`Shock: g_inc`[[1]]
##      [,1]
## [1,]    1
## [2,]    3
##
## $`Shock: g_inc`[[2]]
##      [,1]
## [1,]    1
## [2,]    2
##
## $`Shock: g_inc`[[3]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $`Shock: g_inc`[[4]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $`Shock: g_inc`[[5]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $`Shock: g_inc`[[6]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $`Shock: g_inc`[[7]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $`Shock: g_inc`[[8]]
##      [,1]
```

```

## [1,]    1
## [2,]    1
##
## $\Shock: g_inc`[[9]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $\Shock: g_inc`[[10]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $\Shock: g_inc`[[11]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $\Shock: g_inc`[[12]]
##      [,1]
## [1,]    1
## [2,]    1
##
##
## $\Shock: g_cons`
## $\Shock: g_cons`[[1]]
##      [,1]
## [1,]    1
## [2,]    3
##
## $\Shock: g_cons`[[2]]
##      [,1]
## [1,]    1
## [2,]    2
##
## $\Shock: g_cons`[[3]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $\Shock: g_cons`[[4]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $\Shock: g_cons`[[5]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $\Shock: g_cons`[[6]]
##      [,1]
## [1,]    1
## [2,]    1

```

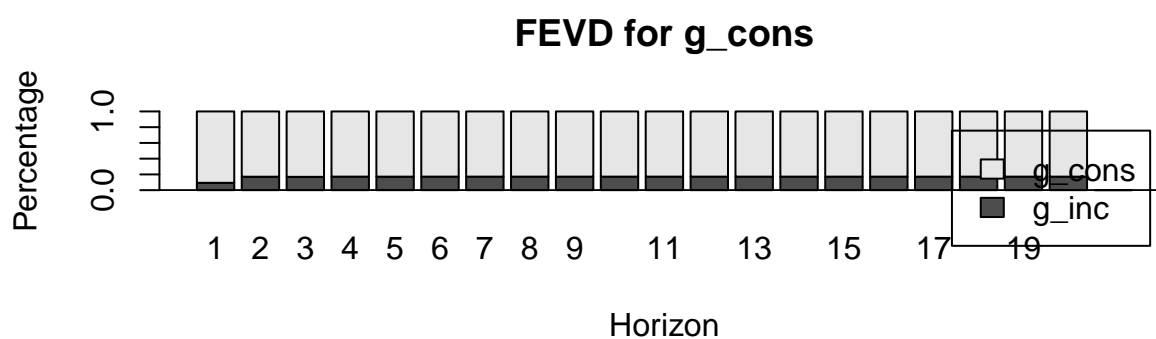
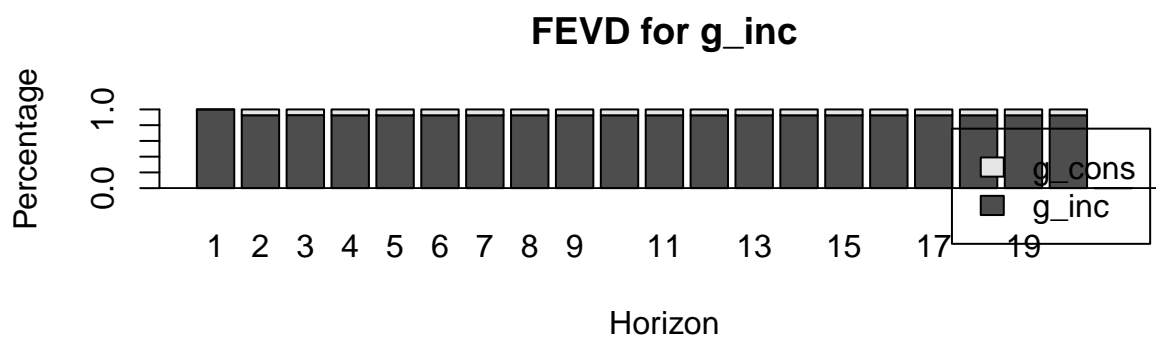


```
##
## $\`Shock: g_cons`[[7]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $\`Shock: g_cons`[[8]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $\`Shock: g_cons`[[9]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $\`Shock: g_cons`[[10]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $\`Shock: g_cons`[[11]]
##      [,1]
## [1,]    1
## [2,]    1
##
## $\`Shock: g_cons`[[12]]
##      [,1]
## [1,]    1
## [2,]    1
```

### Décomposition de la variance

```
dec_var <- fevd(var.1lag,n.ahead=20)

plot(dec_var)
```



## Calcul des prévisions

On calcule les prévisions pour les 10 observations suivantes à partir du VAR(1)

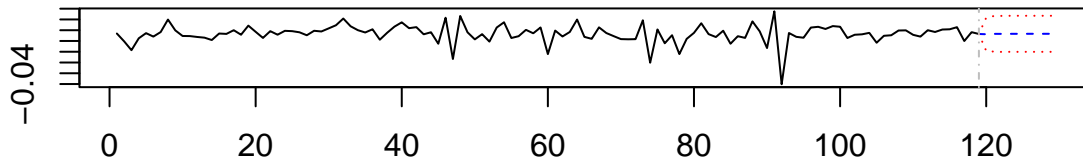
```
forecast<-predict(var.1lag,n.ahead = 10, ci=0.95)
```

```
summary(forecast)
```

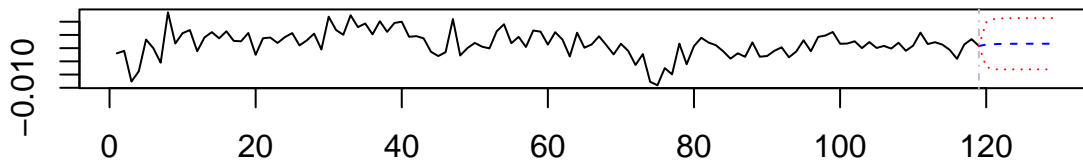
```
##          Length Class  Mode
## fcst         2   -none- list
## endog       238   -none- numeric
## model        10  varest list
## exo.fcst     0   -none- NULL
```

```
plot(forecast)
```

### Forecast of series g\_inc



### Forecast of series g\_cons



### Ajout d'une variable indicatrice en 2013:T1

Sur le graphique des taux de croissance, on observe une valeur atypique pour le premier trimestre de 2013. Cette baisse s'explique par des causes particulières liées à la distribution des salaires et des dividendes en décembre 2012 et aux contributions des employés aux assurances sociales cf BEA. On introduit une variable indicatrice pour prendre en compte cette observation particulière.

```
library(vars)
df_g_short$dummy_2013T1<-as.numeric(df_g_short$date=="2013-01-01")
pselect<-VARselect(y=df_g_short[,c(2,3)],lag.max=8,type=c("const"),exogen = df_g_short$dummy_2013T1)
pselect$criteria # Affichage des valeurs des critères de sélection
```

```
##           1           2           3           4           5
## AIC(n) -2.102441e+01 -2.107740e+01 -2.110390e+01 -2.110489e+01 -2.106510e+01
## HQ(n)  -2.094519e+01 -2.095857e+01 -2.094547e+01 -2.090684e+01 -2.082744e+01
## SC(n)  -2.082913e+01 -2.078447e+01 -2.071334e+01 -2.061668e+01 -2.047926e+01
## FPE(n)  7.400186e-10  7.019311e-10  6.837668e-10  6.834238e-10  7.116719e-10
##           6           7           8
## AIC(n) -2.101564e+01 -2.099071e+01 -2.094460e+01
## HQ(n)  -2.073837e+01 -2.067383e+01 -2.058811e+01
## SC(n)  -2.033215e+01 -2.020959e+01 -2.006583e+01
## FPE(n)  7.485083e-10  7.684329e-10  8.060949e-10
```

```
pselect$selection # nombre de retards optimal selon les critères de sélection
```

```
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      4      2      1      4
```

On obtient des résultats similaires à ceux du cas sans indicatrice.

```
var.1lag_wd <-VAR(y = df_g_short[,c(2,3)],type="const",lag.max = 1,exogen = df_g_short$dummy_2013T1)
summary(var.1lag_wd) #Affichage des résultats avec summary
```

### Estimation du VAR(1) avec l'indicatrice pour 2013:T1

```
##
## VAR Estimation Results:
## =====
## Endogenous variables: g_inc, g_cons
## Deterministic variables: const
## Sample size: 118
## Log Likelihood: 903.748
## Roots of the characteristic polynomial:
## 0.4672 0.3003
## Call:
## VAR(y = df_g_short[, c(2, 3)], type = "const", exogen = df_g_short$dummy_2013T1,
##     lag.max = 1)
##
##
## Estimation results for equation g_inc:
## =====
## g_inc = g_inc.l1 + g_cons.l1 + const + exo1
##
##           Estimate Std. Error t value Pr(>|t|)
## g_inc.l1 -0.207802   0.081770  -2.541  0.01239 *
## g_cons.l1  0.459550   0.137298   3.347  0.00111 **
## const     0.005343   0.001113   4.799 4.88e-06 ***
## exo1      -0.041636   0.007235  -5.755 7.44e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.006961 on 114 degrees of freedom
## Multiple R-Squared: 0.3428, Adjusted R-squared: 0.3255
## F-statistic: 19.82 on 3 and 114 DF, p-value: 2.076e-10
##
##
## Estimation results for equation g_cons:
## =====
## g_cons = g_inc.l1 + g_cons.l1 + const + exo1
##
##           Estimate Std. Error t value Pr(>|t|)
## g_inc.l1  0.1358698  0.0507723   2.676  0.00855 **
## g_cons.l1  0.3746992  0.0852507   4.395 2.49e-05 ***
## const     0.0032950  0.0006913   4.766 5.58e-06 ***
## exo1      -0.0032091  0.0044926  -0.714  0.47650
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.004322 on 114 degrees of freedom
## Multiple R-Squared: 0.2449, Adjusted R-squared: 0.225
```

```
## F-statistic: 12.33 on 3 and 114 DF,  p-value: 4.838e-07
```

```
##
```

```
##
```

```
##
```

```
## Covariance matrix of residuals:
```

```
##           g_inc    g_cons
```

```
## g_inc  4.846e-05 9.366e-06
```

```
## g_cons 9.366e-06 1.868e-05
```

```
##
```

```
## Correlation matrix of residuals:
```

```
##           g_inc g_cons
```

```
## g_inc  1.0000 0.3113
```

```
## g_cons 0.3113 1.0000
```

- La variable indicatrice est négative et significative dans l'équation de `g_inc`. Elle n'est pas significative dans l'équation de `g_conso`, ce qui était attendu
- Les valeurs des coefficients de l'équation de `g_inc` sont un peu modifiés mais ne changent pas de signe
- le  $R^2$  de l'équation de `g_inc` augmente et passe à 32%

**Tests de causalité et fonctions impulsion-réponse** Il n'y a pas de modification des résultats des tests de causalité ni des fonctions impulsion réponse.

```
causality(var.1lag_wd,cause=c('g_inc'))
```

```
## $Granger
```

```
##
```

```
## Granger causality H0: g_inc do not Granger-cause g_cons
```

```
##
```

```
## data: VAR object var.1lag_wd
```

```
## F-Test = 7.1613, df1 = 1, df2 = 228, p-value = 0.00799
```

```
##
```

```
##
```

```
## $Instant
```

```
##
```

```
## H0: No instantaneous causality between: g_inc and g_cons
```

```
##
```

```
## data: VAR object var.1lag_wd
```

```
## Chi-squared = 10.423, df = 1, p-value = 0.001245
```

```
causality(var.1lag_wd,cause=c('g_cons'))
```

```
## $Granger
```

```
##
```

```
## Granger causality H0: g_cons do not Granger-cause g_inc
```

```
##
```

```
## data: VAR object var.1lag_wd
```

```
## F-Test = 11.203, df1 = 1, df2 = 228, p-value = 0.0009551
```

```
##
```

```
##
```

```
## $Instant
```

```
##
```

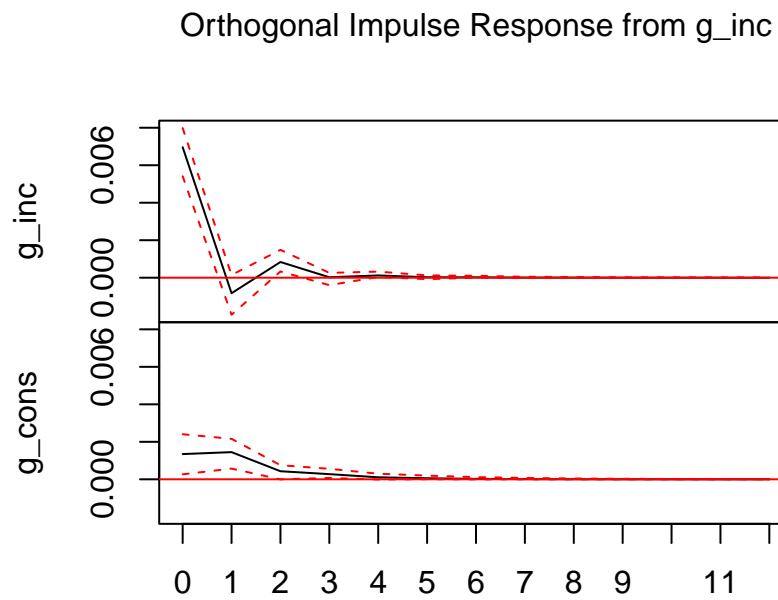
```
## H0: No instantaneous causality between: g_cons and g_inc
```

```
##
```

```
## data: VAR object var.1lag_wd
```

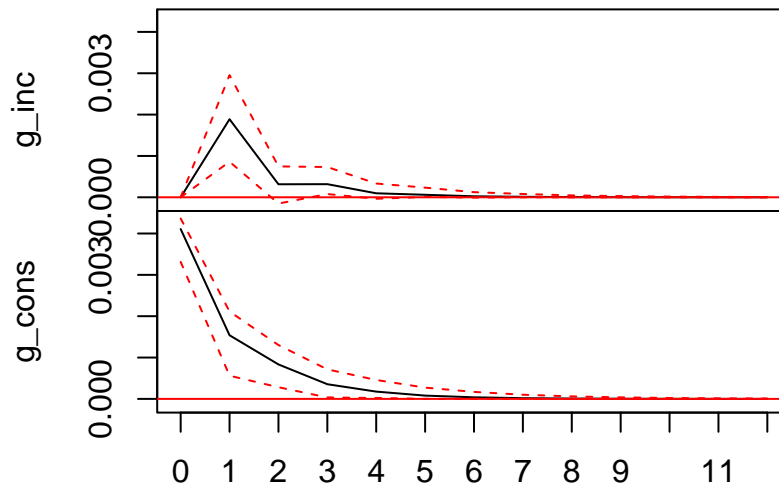
```
## Chi-squared = 10.423, df = 1, p-value = 0.001245
```

```
irf2<-irf(var.1lag_wd,n.ahead = 12,ortho = TRUE,ci=0.95)  
plot(irf2)
```



95 % Bootstrap CI, 100 runs

### Orthogonal Impulse Response from g\_cons



95 % Bootstrap CI, 100 runs

#### Commentaires

Peu de différences avec les impulsions-réponses précédentes sans la variable indicatrice.

## Etude sur la période 1990:T1 à 2023:T1

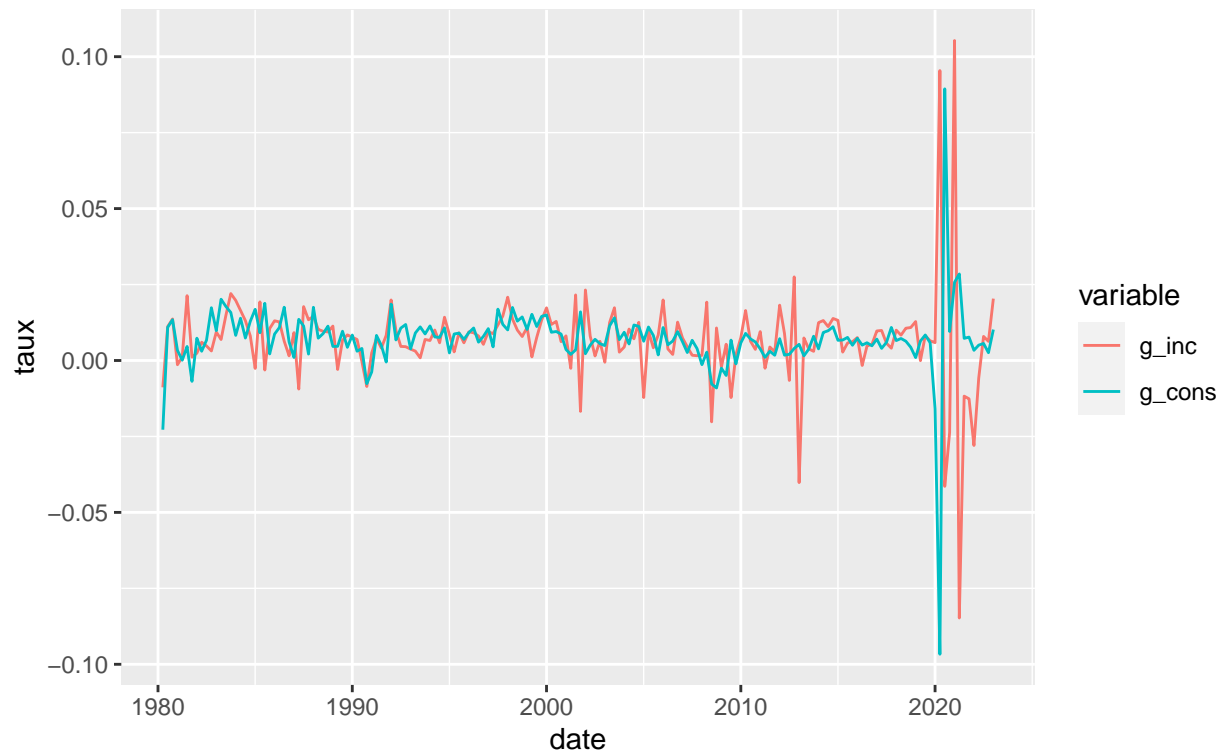
### Calcul et représentation des taux de croissance

```
df_g<-data.frame(df$date[-1],diff(log(df$income)),diff(log(df$consumption)))
colnames(df_g)<-c("date","g_inc","g_cons")
```

```
library(ggplot2)
library(reshape)
data_melt=melt(as.data.frame(df_g[c(1,2,3)]),id.vars = 1)
```

```
p_g_us<-ggplot(data=data_melt,aes(x=date))+geom_line(aes(x=date,y=value,color= variable))
p_g_us+labs(y="taux",title = "Variation du revenu et de la consommation des ménages aux Etats-Unis",sub
```

## Variation du revenu et de la consommation des ménages aux Etats-Unis 1990:T1 à 2023:T1



Le graphique des taux de croissance montre que les observations allant du premier trimestre de 2020 au second semestre de 2021 sont atypiques par le niveau des taux de croissance à la hausse comme à la baisse. On décide d'introduire des variables indicatrices pour prendre en compte ces observations.

```
tail(df_g,n=15)
```

```
##          date      g_inc      g_cons
## 158 2019-07-01  0.008170944  0.008401165
## 159 2019-10-01  0.006487573  0.005820191
## 160 2020-01-01  0.005869652 -0.015918036
## 161 2020-04-01  0.095428174 -0.096685377
## 162 2020-07-01 -0.041423594  0.089409157
## 163 2020-10-01 -0.023672366  0.009567867
## 164 2021-01-01  0.105316657  0.025737746
## 165 2021-04-01 -0.084749496  0.028493723
## 166 2021-07-01 -0.011731731  0.007285022
## 167 2021-10-01 -0.012603933  0.007690901
## 168 2022-01-01 -0.027970247  0.003346606
## 169 2022-04-01 -0.005813493  0.005053581
## 170 2022-07-01  0.007919452  0.005594890
## 171 2022-10-01  0.006286341  0.002557002
## 172 2023-01-01  0.020364516  0.010183317
```

### Estimation d'un modèle VAR

```
df_g$dummy_2013T1<-as.numeric(df_g$date=="2013-01-01")
df_g$dummy_2020T1<-as.numeric(df_g$date=="2020-01-01")
```



```
df_g$dummy_2020T2<-as.numeric(df_g$date=="2020-04-01")
df_g$dummy_2020T3<-as.numeric(df_g$date=="2020-07-01")
df_g$dummy_2020T4<-as.numeric(df_g$date=="2020-10-01")
df_g$dummy_2021T1<-as.numeric(df_g$date=="2021-01-01")
df_g$dummy_2021T2<-as.numeric(df_g$date=="2021-04-01")
```

On cherche le nombre de retards optimal en ajoutant ces variables indicatrices en prenant 8 retards au maximum

```
library(vars)
pselect<-VARselect(y=df_g[,c(2,3)],lag.max=8,type=c("const"),exogen = df_g[c(4,5,6,7,8,9,10)])
# Affichage des valeurs des critère de sélection
pselect$criteria
```

```
##              1              2              3              4              5
## AIC(n) -2.033654e+01 -2.039476e+01 -2.043853e+01 -2.051893e+01 -2.050991e+01
## HQ(n)  -2.018307e+01 -2.021060e+01 -2.022367e+01 -2.027338e+01 -2.023367e+01
## SC(n)  -1.995850e+01 -1.994112e+01 -1.990928e+01 -1.991408e+01 -1.982945e+01
## FPE(n)  1.472606e-09  1.389617e-09  1.330525e-09  1.228242e-09  1.240023e-09
##              6              7              8
## AIC(n) -2.051244e+01 -2.052883e+01 -2.050748e+01
## HQ(n)  -2.020551e+01 -2.019120e+01 -2.013916e+01
## SC(n)  -1.975638e+01 -1.969716e+01 -1.960020e+01
## FPE(n)  1.237713e-09  1.218583e-09  1.246098e-09
```

```
# nombre de retards optimal selon les critères de sélection
pselect$selection
```

```
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      7      4      1      7
```

On choisit d'estimer un modèle VAR(1).

```
var.4lag_T_wd <-VAR(y = df_g[,c(2,3)],type="const",lag.max = 1,exogen = df_g[c(4,5,6,7,8,9,10)])
summary(var.4lag_T_wd) #Affichage des résultats avec summary
```

```
##
## VAR Estimation Results:
## =====
## Endogenous variables: g_inc, g_cons
## Deterministic variables: const
## Sample size: 171
## Log Likelihood: 1264.565
## Roots of the characteristic polynomial:
## 0.3338 0.02519
## Call:
## VAR(y = df_g[, c(2, 3)], type = "const", exogen = df_g[c(4, 5,
##      6, 7, 8, 9, 10)], lag.max = 1)
##
##
## Estimation results for equation g_inc:
## =====
## g_inc = g_inc.l1 + g_cons.l1 + const + dummy_2013T1 + dummy_2020T1 + dummy_2020T2 + dummy_2020T3 + d
##
##              Estimate Std. Error t value Pr(>|t|)
## g_inc.l1      0.1016991  0.0553384   1.838 0.067940 .
```

```

## g_cons.l1      0.1857344  0.1045677   1.776 0.077587 .
## const         0.0050636  0.0010179   4.974 1.66e-06 ***
## dummy_2013T1 -0.0487859  0.0080783  -6.039 1.04e-08 ***
## dummy_2020T1 -0.0009348  0.0079812  -0.117 0.906909
## dummy_2020T2  0.0927241  0.0083373  11.122 < 2e-16 ***
## dummy_2020T3 -0.0382344  0.0144944  -2.638 0.009161 **
## dummy_2020T4 -0.0411296  0.0120864  -3.403 0.000841 ***
## dummy_2021T1  0.1008834  0.0081510  12.377 < 2e-16 ***
## dummy_2021T2 -0.1053041  0.0098477 -10.693 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.007956 on 161 degrees of freedom
## Multiple R-Squared:  0.7575, Adjusted R-squared:  0.744
## F-statistic: 55.89 on 9 and 161 DF, p-value: < 2.2e-16
##
##
## Estimation results for equation g_cons:
## =====
## g_cons = g_inc.l1 + g_cons.l1 + const + dummy_2013T1 + dummy_2020T1 + dummy_2020T2 + dummy_2020T3 + dummy_2020T4 + dummy_2021T1 + dummy_2021T2
##
##               Estimate Std. Error t value Pr(>|t|)
## g_inc.l1      0.0955982  0.0340998   2.803  0.00568 **
## g_cons.l1     0.2572476  0.0644353   3.992 9.92e-05 ***
## const         0.0048264  0.0006273   7.694 1.34e-12 ***
## dummy_2013T1 -0.0031657  0.0049779  -0.636  0.52571
## dummy_2020T1 -0.0228619  0.0049181  -4.649 6.92e-06 ***
## dummy_2020T2 -0.0979781  0.0051375 -19.071 < 2e-16 ***
## dummy_2020T3  0.1003320  0.0089315  11.233 < 2e-16 ***
## dummy_2020T4 -0.0142988  0.0074477  -1.920  0.05664 .
## dummy_2021T1  0.0207130  0.0050227   4.124 5.95e-05 ***
## dummy_2021T2  0.0069782  0.0060682   1.150  0.25186
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.004902 on 161 degrees of freedom
## Multiple R-Squared:  0.8343, Adjusted R-squared:  0.825
## F-statistic: 90.08 on 9 and 161 DF, p-value: < 2.2e-16
##
##
## Covariance matrix of residuals:
##           g_inc  g_cons
## g_inc  6.329e-05 7.876e-06
## g_cons 7.876e-06 2.403e-05
##
## Correlation matrix of residuals:
##           g_inc g_cons
## g_inc  1.0000 0.2019
## g_cons 0.2019 1.0000
##
## Commentaires **

```

- équation de `g_inc`
  - les coefficients de `g_inc.l1` et `g_cons.l1` ne sont significatifs que pour un risque de première espèce de 10 %. Les valeurs des coefficients sont très différentes de celles des estimations précédentes. L'inclusion de la période covid a un impact sur les estimations malgré l'inclusion des variables indicatrices.
  - toutes les indicatrices sauf `dummy_2020T1` sont significatives pour un risque de première espèce de 1%
  - le  $R^2$  est égal à 74,4%
- équation de `g_cons`
  - les coefficients de `g_inc.l1` et `g_cons.l1` ne sont significatifs que pour un risque de première espèce de 1 %. Les valeurs des coefficients sont assez proches de celles des estimations précédente
  - `dummy_2020T4` est significative pour un risque de première espèce de 10 %. Toutes les autres indicatrices sauf `dummy_2013T1` et `dummy_2021T2` sont significatives pour un risque de première espèce de 1%.
  - le  $R^2$  est égal à 83,4%

Nous observons que les indicatrices sont significatives pour au moins une équation, voire les deux.

### Tests de causalité au sens de Granger

```
causality(var.4lag_T_wd,cause=c('g_inc'))
```

```
## $Granger
##
## Granger causality H0: g_inc do not Granger-cause g_cons
##
## data:  VAR object var.4lag_T_wd
## F-Test = 7.8595, df1 = 1, df2 = 322, p-value = 0.005362
##
##
## $Instant
##
## H0: No instantaneous causality between: g_inc and g_cons
##
## data:  VAR object var.4lag_T_wd
## Chi-squared = 6.7005, df = 1, p-value = 0.009638
```

On rejette l'hypothèse nulle d'absence de causalité de `g_inc` sur `g_cons` pour un risque de première espèce de 1%.

```
causality(var.4lag_T_wd,cause=c('g_cons'))
```

```
## $Granger
##
## Granger causality H0: g_cons do not Granger-cause g_inc
##
## data:  VAR object var.4lag_T_wd
## F-Test = 3.1549, df1 = 1, df2 = 322, p-value = 0.07664
##
##
## $Instant
##
## H0: No instantaneous causality between: g_cons and g_inc
##
```

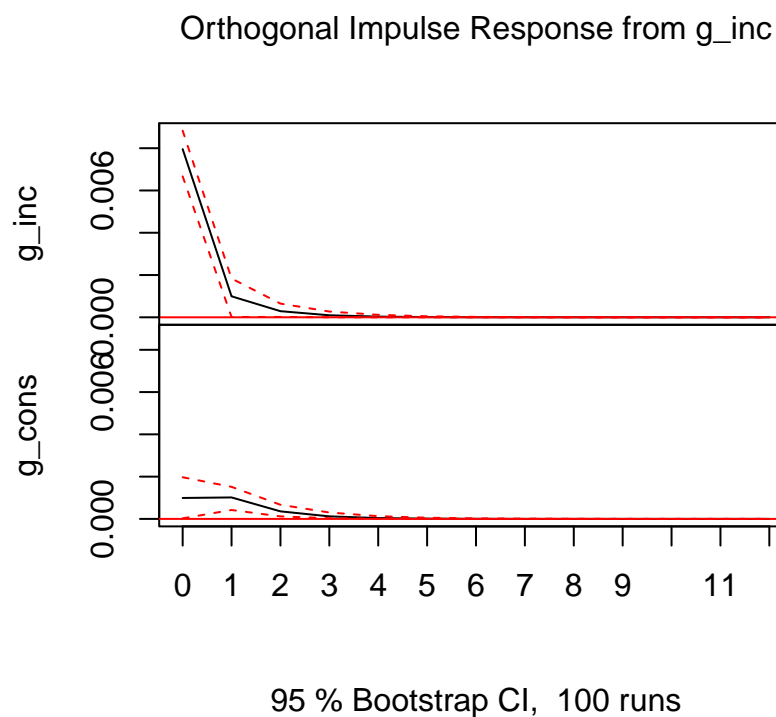
```
## data:  VAR object var.4lag_T_wd  
## Chi-squared = 6.7005, df = 1, p-value = 0.009638
```

On rejette l'hypothèse nulle d'absence de causalité de g\_cons sur g\_inc pour un risque de première espèce de 10% seulement.

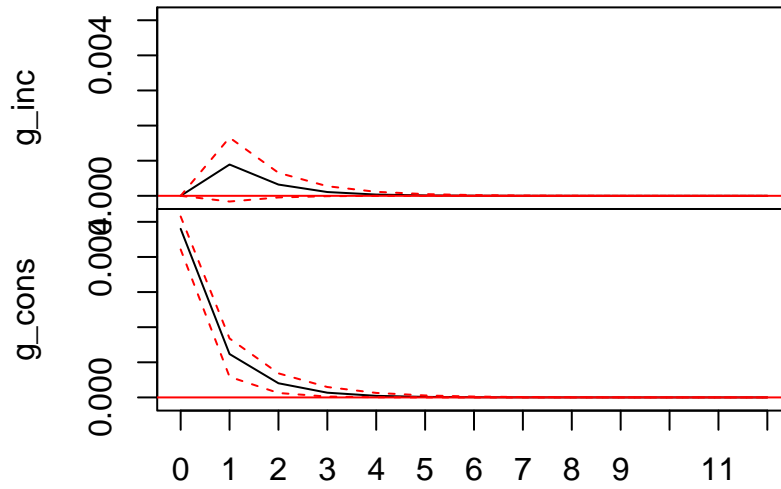
### Estimation des fonctions impulsion réponse

On garde le même ordre des variables pour le calcul des chocs structurels par la méthode de Cholesky.

```
irf3<-irf(var.4lag_T_wd,n.ahead = 12,ortho = TRUE,ci=0.95)  
plot(irf3)
```



Orthogonal Impulse Response from g\_cons



95 % Bootstrap CI, 100 runs

Un choc sur  $g\_cons$  n'a plus d'effet significatif sur  $g\_inc$ .