Package MTS

Jules CORBEL & Paul GUILLOTTE

05/02/2019

Nous nous intéresserons dans ce document à la mise en place de modèles VAR à l'aide du package **MTS** afin de prédire la masse salariale trimestrielle. Un modèle VAR, pour Vecteur AutoRégressif, a pour objectif de capturer les interdépendances entre les différentes séries temporelles à notre disposition. Ainsi, chaque variable est expliquée par ses propres valeurs passées ainsi que par les valeurs passées des autres variables du modèle.

Visualisation des séries

Même contenu que pour le package vars

Transformation des séries

Même contenu que pour le package vars

Calcul de l'ordre p

Afin de mettre en place une modélisation VAR, nous devons dans un premier temps nous intéresser à l'ordre p du modèle VAR. L'ordre p correspond à l'ordre de l'opérateur de retard, c'est-à-dire le nombre de valeurs du passé qui ont un impact sur la valeur à un instant défini. Dans le package **MTS**, la fonction utilisée est VARorder, qui comme VARselect utilise les critères d'AIC, BIC et HQ afin de déterminer l'ordre du processus. Toutefois, le critère FPE n'est pas présent, ce qui nous conforte dans notre idée qu'il n'est pas très utile à notre étude.

Cependant, nous avons également en notre possession un autre critère, la statistique du test de Tiao-Box ainsi que sa p-value. Ce test évalue pour un ordre i la significativité des coefficients de la matrice A_i. Ce test permet donc de déterminer si le modèle d'ordre i est meilleur que celui d'ordre *i-1*. La statistique de test est la suivante : $M(i) = -(T - K - i - \frac{3}{2})\ln(\frac{det(\hat{\Sigma}_1)}{det(\hat{\Sigma}_{i-1})})$, qui suit une loi du $\chi^2_{k^2}$. Dans l'exemple que nous prenons, la première p-value supérieure à 0.05 est celle pour la matrice A_5. L'ordre que nous devons retenir par rapport à ce test est donc 4.

```r
selec <- VARorder(cbind(MSEStaTrain, PIBStaTrain, SMICStaTrain, TCHOFStaTrain), maxp=8)
```

```
## selected order: aic =  4
## selected order: bic =  0
## selected order: hq =  3
## Summary table:
##       p      AIC      BIC       HQ    M(p) p-value
##  [1,] 0 -31.0540 -31.0540 -31.0540  0.0000  0.0000
##  [2,] 1 -30.9694 -30.5576 -30.8026 20.2744  0.2081
##  [3,] 2 -31.5829 -30.7594 -31.2494 78.3535  0.0000
##  [4,] 3 -31.9078 -30.6725 -31.4076 51.4059  0.0000
##  [5,] 4 -32.0011 -30.3540 -31.3341 31.1393  0.0129
##  [6,] 5 -31.9455 -29.8867 -31.1118 18.7133  0.2838
```

```
##  [7,] 6 -31.9427 -29.4721 -30.9423 21.2987  0.1673
##  [8,] 7 -31.9194 -29.0371 -30.7522 18.7326  0.2828
##  [9,] 8 -31.8698 -28.5757 -30.5359 15.9828  0.4542
```

Comme pour le package VARS, l'AIC associé aux différents modèles diminue en même temps que l'ordre augmente. Si l'on conserve un ordre raisonnable, le meilleur AIC est également pour un modèle d'ordre 4. Le BIC nous donne lui un modèle d'ordre 2. Cela correspond aux mêmes ordres que ceux données par le package **vars**.

L'article de Ruey S. Tsay nous conseille d'utiliser en priorité le test de Tiao-Box pour déterminer l'ordre, nous retenons donc le modèle d'ordre 4.

Estimation du modèle

Dans le package **MTS**, la fonction utilisée pour construire des modèles VAR est *VAR*, qui prend en entrée la série temporelle multivariée et l'ordre du processus. On affiche ci-dessous les résultats renvoyés par la fonction sur le modèle d'ordre 4

```
modele<-VAR(cbind(MSEStaTrain, PIBStaTrain, SMICStaTrain, TCHOFStaTrain), p=4, output=F)
print("Coefficients du modèle pour un retard d'ordre 1")
```

```
## [1] "Coefficients du modèle pour un retard d'ordre 1"
```

```
coeff<-modele$coef[2:5,]
colnames(coeff)<-rownames(coeff)<-c("MSE", "PIB", "SMIC", "TCHOF")
coeff
```

```
##                MSE           PIB        SMIC      TCHOF
## MSE   -0.49912624 -0.0308097019  0.05328488 -0.8445102
## PIB    0.92777848  0.1257671961  1.21484663 -5.9314489
## SMIC   0.06599595 -0.0051246390 -0.49802200  0.6626694
## TCHOF -0.01352986  0.0006699293 -0.01008137 -0.3235506
```

```
print("Erreurs standard associées aux coefficients")
```

```
## [1] "Erreurs standard associées aux coefficients"
```

```
se<-modele$secoef[2:5,]
colnames(se)<-rownames(se)<-c("MSE", "PIB", "SMIC", "TCHOF")
se
```

```
##              MSE         PIB       SMIC     TCHOF
## MSE   0.10663462 0.013939764 0.22438615 0.8210570
## PIB   0.80820760 0.105652590 1.70067285 6.2229750
## SMIC  0.05296213 0.006923452 0.11144569 0.4077938
## TCHOF 0.01327681 0.001735605 0.02793775 0.1022277
```

```
matrix(data=c(modele$aic, modele$bic, modele$hq), nrow=1, byrow=T, dimnames=list(NULL, c("AIC", "BIC",
```

```
##            AIC       BIC        HQ
## [1,] -31.91369 -30.26664 -31.24674
```

Vérification des hypothèses

Vérification de la stabilité

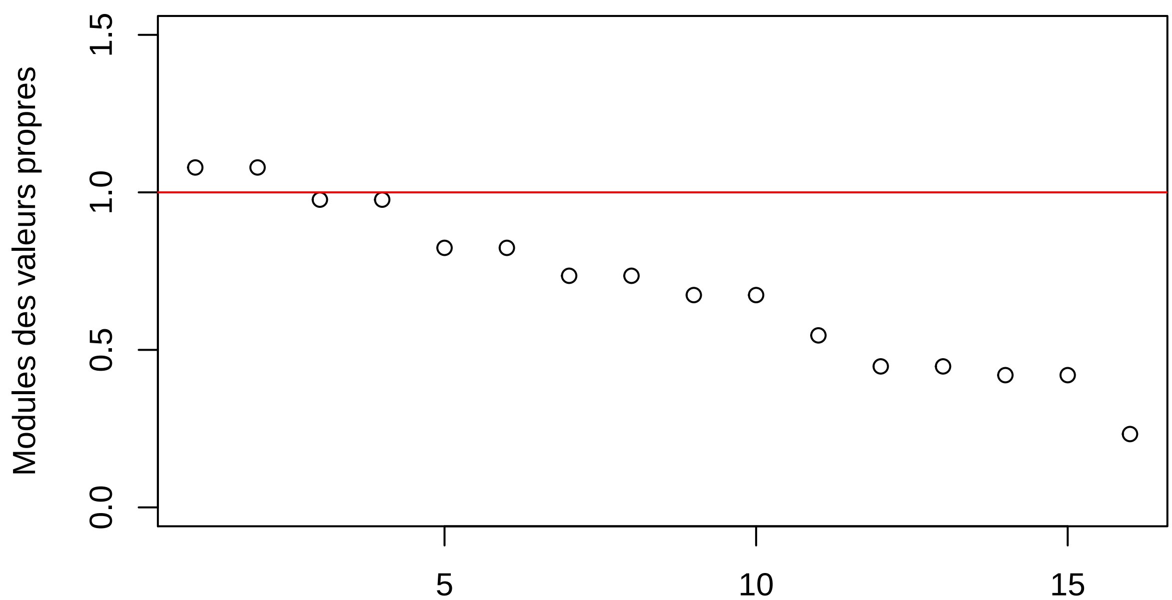
Comme pour le package **vars**, nous vérifions que les modules des valeurs propres de la matrice A sont tous inférieurs à 1 . R. Tsay définit cependant la matrice A différemment de B. Pfaff, soit en la retournant

$$A = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & I \\ A_p & A_{p-1} & A_{p-2} & \cdots & A_1 \end{bmatrix}$$

```
A <- matrix(0,nrow=16, ncol=16)
    A[1:4,5:8] = diag(4)
    A[5:8,9:12] = diag(4)
    A[9:12,13:16] = diag(4)
    A[13:16,1:4] = modele$coef[2:5,]
    A[13:16,5:8] = modele$coef[6:9,]
    A[13:16,9:12] = modele$coef[10:13,]
    A[13:16,13:16] = modele$coef[14:17,]
    vp<-eigen(A)$values
    Mod(vp)
```

```
##  [1] 1.0790210 1.0790210 0.9768390 0.9768390 0.8238857 0.8238857 0.7351693
##  [8] 0.7351693 0.6739761 0.6739761 0.5460153 0.4475344 0.4475344 0.4198056
## [15] 0.4198056 0.2327559
```

```
plot(seq(1,16), Mod(vp), xlab="",
     ylab="Modules des valeurs propres", ylim=c(0,1.5))
abline(h=1, col="red")
```

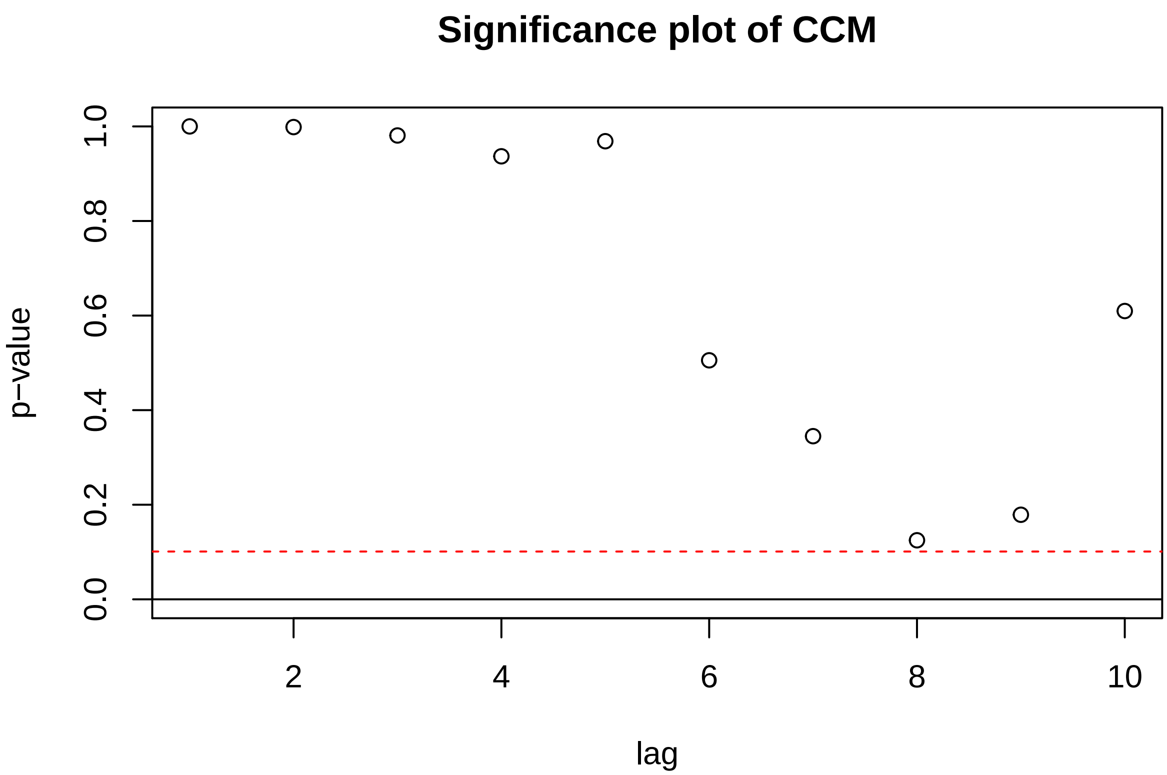


Pas stable ????????????????????????????

Corrélation croisée

Nous vérifions ensuite que les résidus ne comportent ni d'autocorrélation, ni de corrélation croisée. Pour cela, on utilise la fonction *ccm*, qui vérifie les matrices de corrélation croisée pour un lag donné. On définit la matrice de variance-covariance pour un lag p comme $\hat{\Gamma}_p = \frac{1}{T}\sum_{i=p+1}^{T}(r_t - \bar{r})(r_{t-p} - \bar{r})$. La matrice de corrélation associée vaut donc $\rho_p = \hat{D}^{-1}\hat{\Gamma}_p\hat{D}^{-1}$.

```
crossCorr<-ccm(modele$residuals, lag=10, output=F)
plot(crossCorr$pvalue, xlab = "lag", ylab = "p-value", ylim = c(0,1), main="Significance plot of CCM")
abline(h = 0)
crit = 2/sqrt(length(modele$residuals))
abline(h = crit, lty = 2, col="red")
```



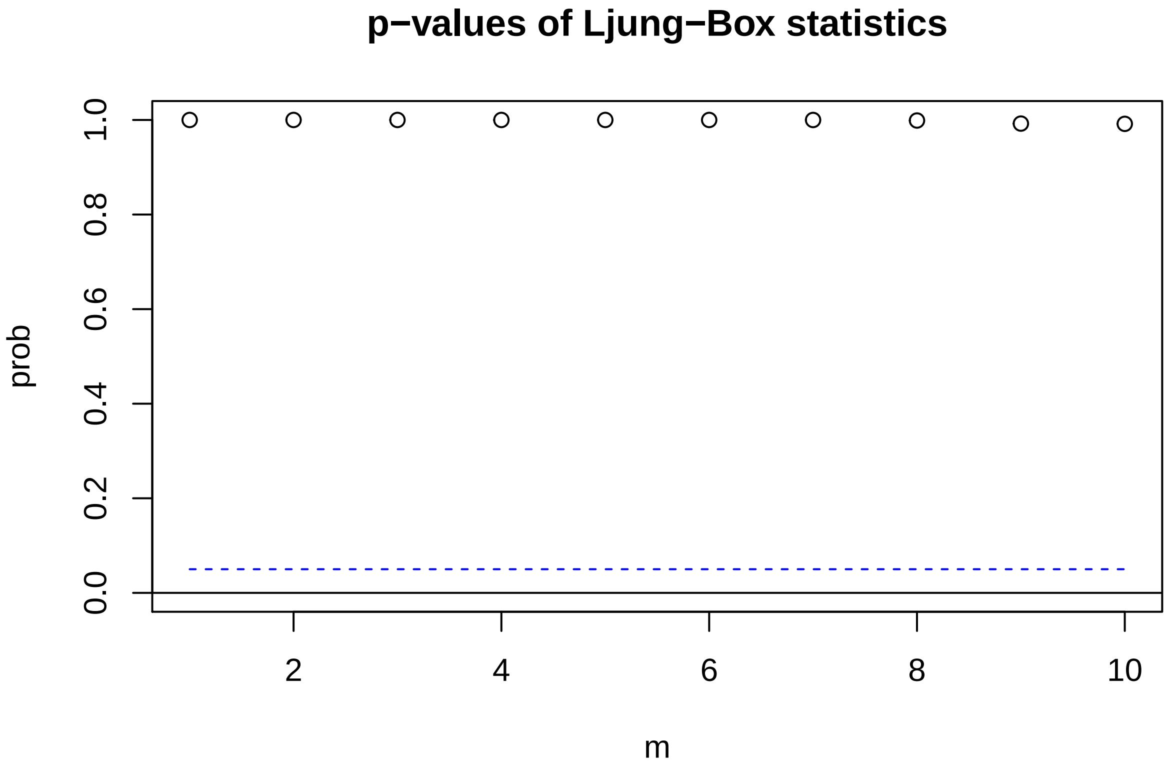
Le graphique ci-dessus représente le résultat du test d'égalité de la matrice $\hat{\Gamma}_p$ à 0. On ne rejette pas l'hypothèse nulle peu importe le lag, ce qui nous indique que les résidus ne comporte pas de corrélation croisée.

Nous effectuns ensuite le test de Ljung-Box multivariée, qui teste à la fois l'autocorrélation et la corrélation croisée.

```
mq(modele$residuals, lag=10)
```

```
## Ljung-Box Statistics:
##            m       Q(m)     df    p-value
##   [1,]    1.00      1.13   16.00     1.00
##   [2,]    2.00      5.38   32.00     1.00
##   [3,]    3.00     12.09   48.00     1.00
##   [4,]    4.00     20.65   64.00     1.00
##   [5,]    5.00     28.03   80.00     1.00
##   [6,]    6.00     43.61   96.00     1.00
##   [7,]    7.00     61.61  112.00     1.00
```

```
## [8,]   8.00    84.69  128.00     1.00
## [9,]   9.00   106.12  144.00     0.99
## [10,]  10.00  120.26  160.00     0.99
```



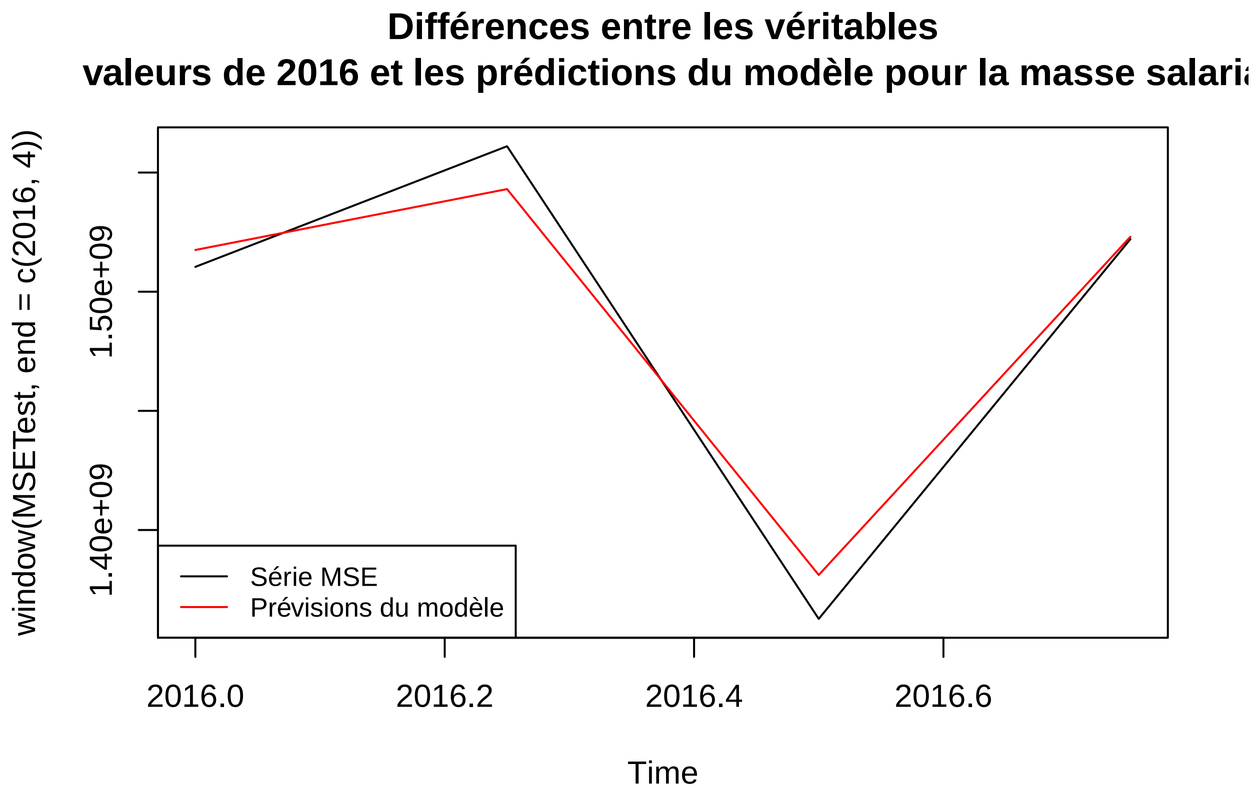
Prévisions

Maintenant que nous avons estimé l'ordre des différents modèle VAR, et que nous avons explicité l'estimation des modèles, nous cherchons désormais à trouver celui dont les prédictions sont les plus proches de la réalité.

Après avoir comparé tous les modèles possibles (7 : 3 modèles avec deux variables, 3 modèles avec trois variables et un modèle avec les quatre variables), nous nous apercevons que le meilleur en terme de prédictions est le modèle prenant en compte le SMIC et le PIB, avec un ordre égal à 4.

```r
#SMIC
VARorder(cbind(MSEStaTrain, SMICStaTrain, PIBStaTrain), maxp=10)
modele<-VAR(cbind(MSEStaTrain, SMICStaTrain, PIBStaTrain), p=4)
pred <- VARpred(modele, 4)
```

```r
plot(window(MSETest, end=c(2016,4)), main="Différences entre les véritables
     valeurs de 2016 et les prédictions du modèle pour la masse salariale")
#Reconstruction de la variable stationnaire
recon <- ts(pred$pred[,1], start=2016, frequency=4) * MSETrendTest * MSESeasonalTest
lines(recon, col = "red")
legend('bottomleft', legend = c('Série MSE', 'Prévisions du modèle'),
       col=c('black', 'red'), lty=1, cex=0.8)
```



Nous nous intéressons donc à l'erreur quadratique moyenne de cette prévision, qui est inférieure à celle obtenue pour le meilleur modèle effectué avec le package **vars**.

```
EQM(MSETest, recon)
```

```
## [1] 1.188978e+14
```