

# Projet d'économétrie

VOIRIN Tristan et SHMITT Rémi

M1 Economie et Finance

## Contents

<b>Introduction</b>	<b>2</b>
<b>Partie 1 : Modélisation univariée</b>	<b>2</b>
1.1 Présentation des séries . . . . .	2
1.2 Autocorrélogrammes simples et partiels . . . . .	3
1.3 Test de racine unitaire et stationnarisation . . . . .	5
1.3 Détermination du modèle ARMA optimal . . . . .	10
1.4 Prévision . . . . .	12
<b>Partie 2 : Modélisation multivariée</b>	<b>15</b>
2.1 Estimation du modèle VAR . . . . .	16
2.2 Test de causalité . . . . .	17
2.3 Analyse impulsion-repulsion . . . . .	18
2.4 Test de cointégration . . . . .	20
<b>Conclusion</b>	<b>20</b>

*Ce document contient une synthèse de résultats générés par du code R. L'intégralité des sorties brutes du code n'y figurent donc pas directement, mais la plupart sont présentées sous forme de tableaux afin de respecter la contrainte donnée de 20 pages maximum et permettre une meilleure lisibilité. Toutefois, toutes les cellules de code essentielles à la réalisation de ce document, ainsi que leurs sorties, sont incluses et commentées dans le fichier RMarkdown associé.*

# Introduction

Étant considéré comme l'un des indicateurs économiques les plus importants, le taux de chômage est, depuis maintenant plusieurs décennies, au coeur de nos actualités. Ses fluctuations sont suivies quotidiennement car elles ont des conséquences directes sur notre pays, touchant non seulement l'économie, mais aussi les sphères politique et sociale. Comprendre cet indicateur est donc crucial, en identifiant par exemple les autres indicateurs avec lesquels il peut être corrélé. Dans cette optique, de nombreuses études ont examiné la relation entre chômage et inflation, comme en témoigne la célèbre courbe de Phillips. Pour notre projet, nous avons choisi d'explorer une relation moins étudiée, mais néanmoins pertinente : le lien entre taux de chômage et taux de croissance des crédits accordés aux entreprises non financières, que nous suspectons être causale.

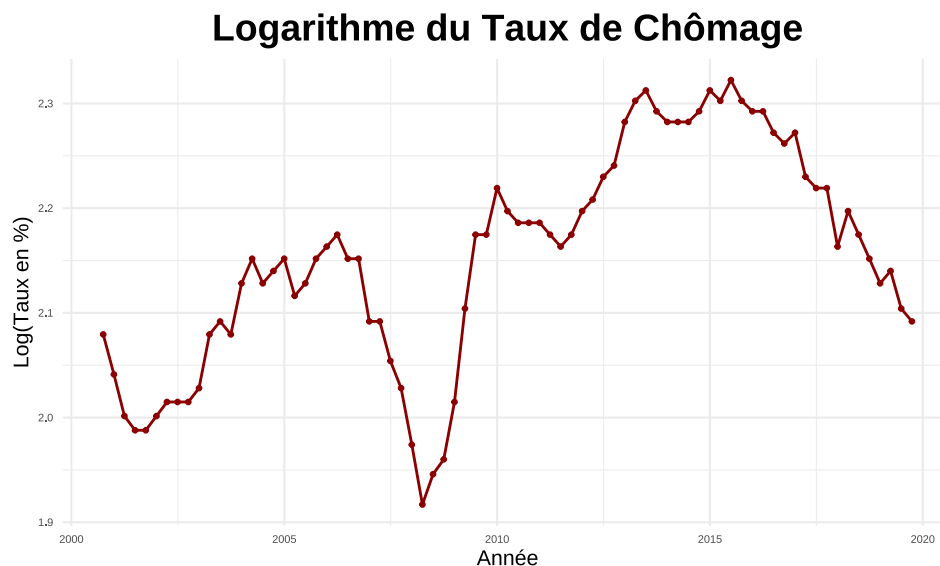
Les variations du taux de chômage exercent une influence directe sur la consommation des ménages, un des moteurs clés de la demande globale, et par extension, sur l'activité économique. De plus, les crédits accordés aux entreprises non financières jouent un rôle essentiel dans le financement de l'investissement, stimulant l'innovation et la croissance. Analyser les interactions entre ces deux indicateurs permet alors de mieux comprendre certains mécanismes macroéconomiques clés.

## Partie 1 : Modélisation univariée

### 1.1 Présentation des séries

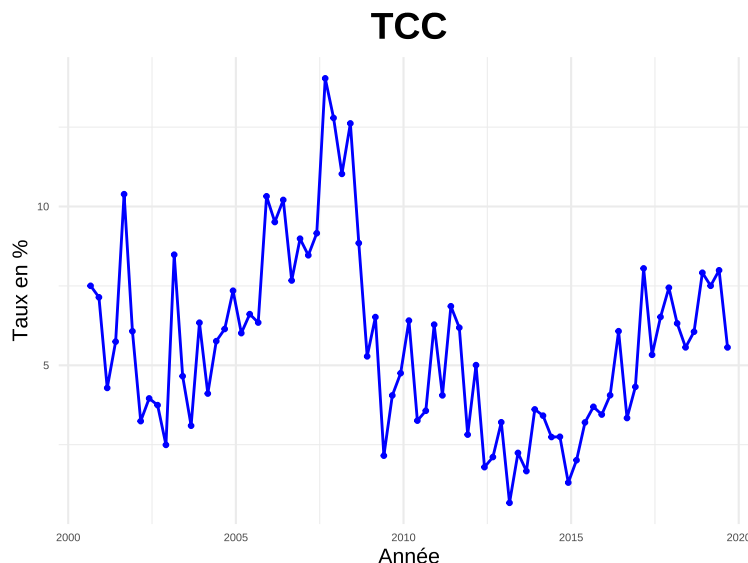
Pour étudier cette relation, nous avons collecté une série de données couvrant une période commune pour chaque variable. Nous avons utilisé les données de la section *webStat* du site de la Banque de France, comprenant deux séries sous format CVS, de septembre 2000 à septembre 2019, pour un total de 77 observations.

Le tx chômage, est observé mensuellement et exprimé en pourcentage de la population active. Cette série, couvrant la période de septembre 2000 à septembre 2019, présente une évolution notable, fluctuante entre 6 % et 12 %. Il convient de noter que cette série est déjà corrigée des variations saisonnières, ce qui évite la nécessité d'un ajustement supplémentaire.



Visuellement la série semble adopter une tendance stochastique et être de type DS, donc non stationnaire. En effet, si l'on ne considère pas la crise de 2008, on observe que l'évolution des valeurs de la série n'a cessé de varier.

Le taux de croissance des crédits accordés aux entreprises non financières, que nous abrègerons TCC, est observé mensuellement et mesuré en taux annualisés. De ce fait, il n'est **pas requis d'effectuer une transformation logarithmique**. Néanmoins, un léger filtre de la série sur *Excel* a été nécessaire en supprimant 3 observations sur 4 afin de la rendre trimestrielle et donc compatible avec la série sur le taux de chômage. Ici encore, la série est corrigée des variations saisonnières.

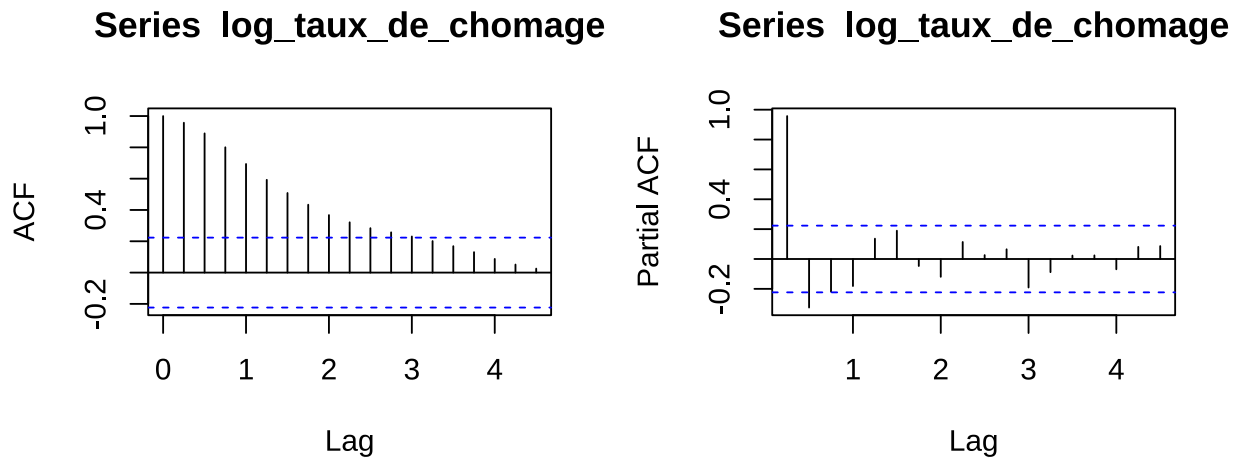


On remarque que le TCC a été positif sur toute la période considérée. Cependant, sa valeur a constamment fluctué avec notamment un pic d'augmentation atteignant 14% entre 2013 et 2015. On remarque que le taux de chômage a toujours été positif et a fortement fluctué entre septembre 2000 et septembre 2019. Le taux de chômage semble admettre une borne inférieure à 6% et une borne supérieure aux alentours de 12%.

## 1.2 Autocorrélogrammes simples et partiels

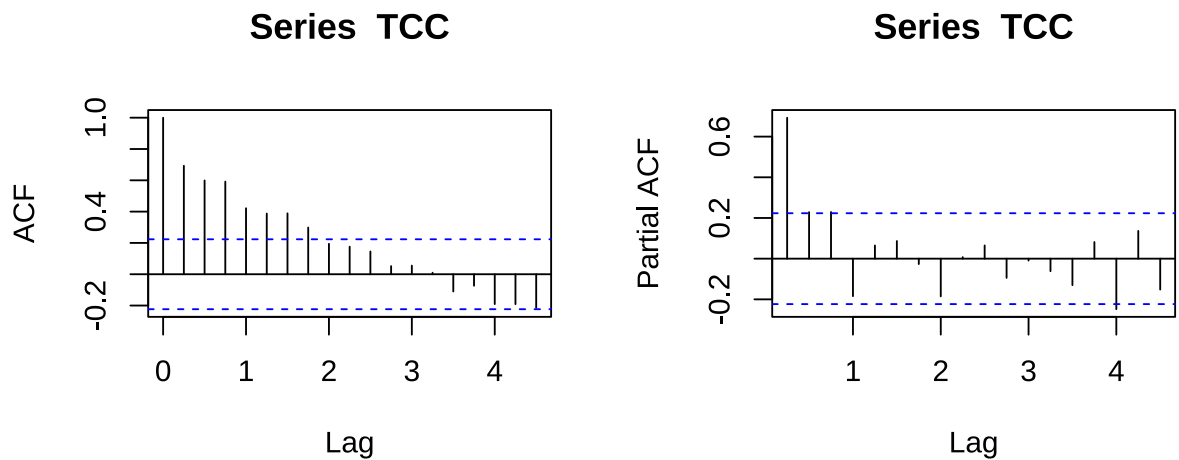
Une première approche permettant d'appréhender la stationnarité présente ou non dans une série est celle de l'analyse de son autocorrelation. L'autocorrelat simple d'ordre  $h$  d'une série est une mesure de sa corrélation avec une version décalée de  $h$  unités temporelles d'elle même. On l'exprime :  $\hat{\rho}(h) = \frac{\hat{\Gamma}(h)}{\hat{\Gamma}(0)}$  avec  $\hat{\Gamma}(h)$  la fonction d'autocovariance d'ordre  $h$ . On la différencie de l'autocorrelation **partielle** d'ordre  $h$  qui ne tient pas compte de l'influence des termes d'ordre inférieurs à  $h$ . Typiquement, une série non stationnaire verra son autocorrelation simple décroître très lentement avec  $h$ .

## Le taux de chômage



Les autocorrélogrammes du taux de chômage couplée à l'analyse graphique suggèrent que la série n'est pas stationnaire. En effet, l'ACF décroît lentement avec  $h$ , et seuls les 2 premiers termes de la PACF sont significativement différents de 0.

## Le TCC



Là encore, les autocorrélogrammes du TCC suggèrent que la série n'est pas stationnaire. En effet, l'ACF décroît lentement avec  $h$ , tandis que seuls les 2 premiers termes de la PACF sont significativement différents de 0.

### 1.3 Test de racine unitaire et stationnarisation

Les racines d'un processus stationnaire doivent être supérieures à 1 en module. Si ce n'est pas le cas, le processus est dit non stationnaire et présente donc une tendance déterministe ou stochastique. Le test de Dickey-Fuller augmenté couplé au test KPSS nous permettent alors de vérifier rigoureusement la stationnarité de nos séries.

Voici les modèles utilisés séquentiellement dans le test de Dickey-Fuller augmenté.

- M3:

$$\Delta X_t = c + bt + \rho X_{t-1} + \sum_{j=1}^p \delta_j \Delta X_{t-j} + \varepsilon_t$$

- M2:

$$\Delta X_t = c + \rho X_{t-1} + \sum_{j=1}^p \delta_j \Delta X_{t-j} + \varepsilon_t$$

- M1:

$$\Delta X_t = \rho X_{t-1} + \sum_{j=1}^p \delta_j \Delta X_{t-j} + \varepsilon_t$$

On teste ensuite sur ces modèles, un test de racine unitaire et un test de significativité des paramètres.

Le test de racine unitaire consiste à tester les hypothèses suivantes:

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho < 0 \end{cases}$$

On utilise alors la statistique de test:

$$t = \frac{\hat{\rho}}{\sigma_{\hat{\rho}}}$$

Le test de significativité des paramètres consiste à tester les hypothèses suivantes (différentes pour chaque modèle).

- M3:

$$\begin{cases} H_0 : b = 0 \\ H_1 : b \neq 0 \end{cases}$$

On utilise ensuite la statistique de test:

$$t = \frac{\hat{b}}{\sigma_{\hat{b}}}$$

- M2:

$$\begin{cases} H_0 : c = 0 \\ H_1 : c \neq 0 \end{cases}$$

Statistique de test:

$$t = \frac{\hat{c}}{\sigma_{\hat{c}}}$$

- M1: le test de racine unitaire est suffisant.

Ensuite, nous réaliserons un test KPSS consistant à tester les hypothèses suivantes:

$$\begin{cases} H_0 : \text{absence de racine unitaire} \\ H_1 : \text{présence de racine unitaire} \end{cases}$$

On utilise alors la statistique de test suivante:

$$LM = \frac{1}{s^2 T} \sum_{t=1}^T P T_t^2$$

### 1.3.1 Le taux de chômage

#### a. Test ADF

- **Modele avec constante et terme déterministe M3**

On commence donc par tester le modèle avec constante et terme déterministe.

Table 1: Resultats du test ADF sur M3

Statistique	Valeur
Statistique ADF	-3.596
Valeur critique tau3 (5%)	-3.450
$t$ stat du terme déterministe	-2.569
Valeur critique (5%) pour $tt$	1.960

On voit que la statistique ADF est inférieure à la valeur critique  $tau3$  au seuil de 5%. On rejette donc  $H_0$  au seuil de 5% et on conclut à l'absence d'une RU. Cependant, on doit donc lire la valeur critique à 5% (pour le test déterministe) dans la table de Student à  $77 - (4 + 3) = 70 > 30$ , les degrés de liberté étant supérieure à 30, on lit la valeur critique dans la table de la loi normale soit 1,96. Or on voit que la  $t$ -value du terme déterministe  $tt$  est supérieure en valeur absolue à 1.96. On rejette donc  $H_0$  au seuil de 5%. On en conclut que le taux de chômage suit un processus  $(X_t) \sim I(0) + C + T$ .

#### b. Test KPSS

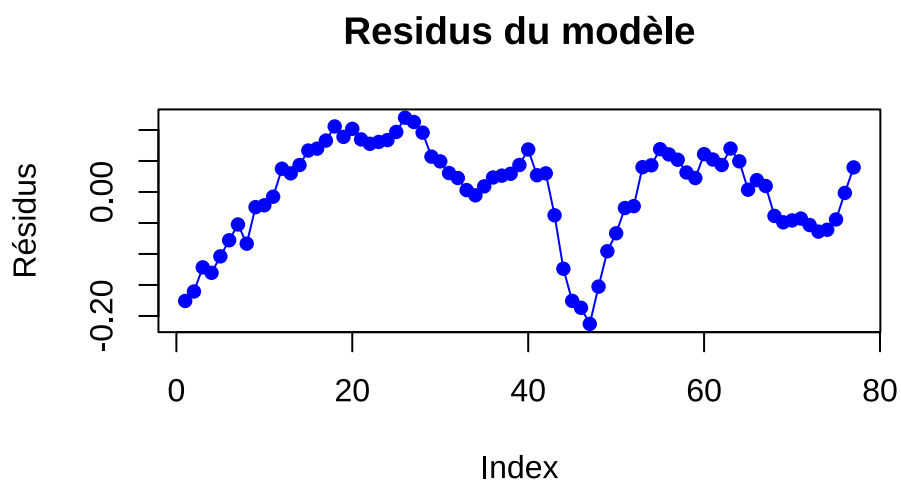
Table 2: Resultats du test KPSS

Statistique	Valeur
Statistique KPSS	0.153
Valeur critique 5%	0.146

La statistique KPSS étant supérieure à la valeur critique, on rejette  $H_0$  au seuil de 5%.

Les résultats des tests statistiques suggèrent que la série est stationnaire autour d'une tendance (TS). Le test ADF, réalisé avec un terme de tendance, rejette l'hypothèse nulle de racine unitaire au seuil de 5% (statistique de test : -3.5957, valeur critique à 5% : -3.45). De plus, le terme de tendance  $tt$  est significatif (p-value = 0.012), ce qui confirme qu'une composante déterministe influence la série. Par ailleurs, bien que le test KPSS rejette l'hypothèse nulle de stationnarité, cela est cohérent avec sa sensibilité aux tendances résiduelles et à l'autocorrélation. Ces résultats montrent que la série, bien qu'influencée par une tendance déterministe, est stationnaire une fois cette dernière retirée, ce qui caractérise une série TS.

#### c Stationnarisation



Suite au test de Dickey-Fuller augmentée sur le M3, on conclut à l'absence de racine unitaire mais à la non significativité du coefficient b. On passe donc au test sur M2.

Grâce au test de Dickey-Fuller augmentée sur le M2, on conclut à l'absence de racine unitaire et à la non significativité de la constante.

Table 3: Resultats du test ADF sur M1

Statistique	Valeur
Statistique ADF	-3.546
Valeur critique tau1 (5%)	-1.950

Suite au test de Dickey-Fuller simple sur le M1, on conclut à l'absence de racine unitaire. On a donc bien que la série des résidus est stationnaire car intégrée d'ordre 0. On vérifie notre résultat avec un test KPSS.

Table 4: Resultats du test KPSS

Statistique	Valeur
Statistique KPSS	0.153
Valeur critique 5%	0.463

La statistique KPSS étant inférieure à la valeur critique, on ne rejette pas  $H_0$ . La série *log\_taux\_de\_chomage* est donc de type TS, ce qui signifie qu'elle est stationnaire une fois que la tendance linéaire est retirée. Cela va dans le même sens que le test ADF, où la série ne présente pas de racine unitaire et semble stationnaire.

### 1.3.2 Le TCC

#### a. Test ADF

- Modèle avec constante et terme déterministe M3

On commence donc par tester le modèle avec constante et terme déterministe :

Table 5: Resultats du test ADF sur M3

Statistique	Valeur
Statistique ADF	-2.484
Valeur critique tau3 (5%)	-3.450
$t$ stat du terme déterministe	-2.484
Valeur critique (5%) pour $tt$	3.170

La  $t$  value du terme  $z.lag.1$  est supérieure à la valeur critique  $tau3$  ce qui nous conduit donc à ne pas rejeter  $H_0$  au seuil de 5% et ainsi conclure à la présence de RU. Cependant, la  $t$  value du terme déterministe  $tt$  est inférieure à la valeur critique 3.17 lu dans la table de *Dickey Fuller*. Le coefficient n'est donc pas significatif au seuil de 5% et nous devons alors considérer le modèle 2.

- **Modele avec constante M2**

Table 6: Resultats du test ADF sur M2

Statistique	Valeur
Statistique ADF	-2.241
Valeur critique tau2 (5%)	-2.890
$t$ stat de $c$	2.069
Valeur critique (5%) pour $c$	2.880

Ici encore, la  $t$  value du terme  $z.lag.1$  est supérieure à la valeur critique  $tau2$ . Le terme constant n'est pas significatif au seuil de 5% car sa  $t$  value est inférieure au quantile 2.88 de la table de *Dickey Fuller*. Nous devons alors considérer le modèle 1.

- **Modele simple M1**

Table 7: Resultats du test ADF sur M1

Statistique	Valeur
Statistique ADF	-0.762
Valeur critique tau1 (5%)	-1.950

Finalement, on est encore amené à rejeter  $H_0$  car la  $t$  value du terme  $z.lag.1$  est supérieure à la valeur critique  $tau1$  et nous pouvons donc conclure à la présence d'une racine unitaire.

La stratégie de test ADF nous amène donc à considérer un processus intégré d'ordre 1 de la forme  $(X_t) \sim I(1)$ .

## b. Test KPSS

Table 8: Resultats du test KPSS

Statistique	Valeur
Statistique KPSS	0.365
Valeur critique 5%	0.463

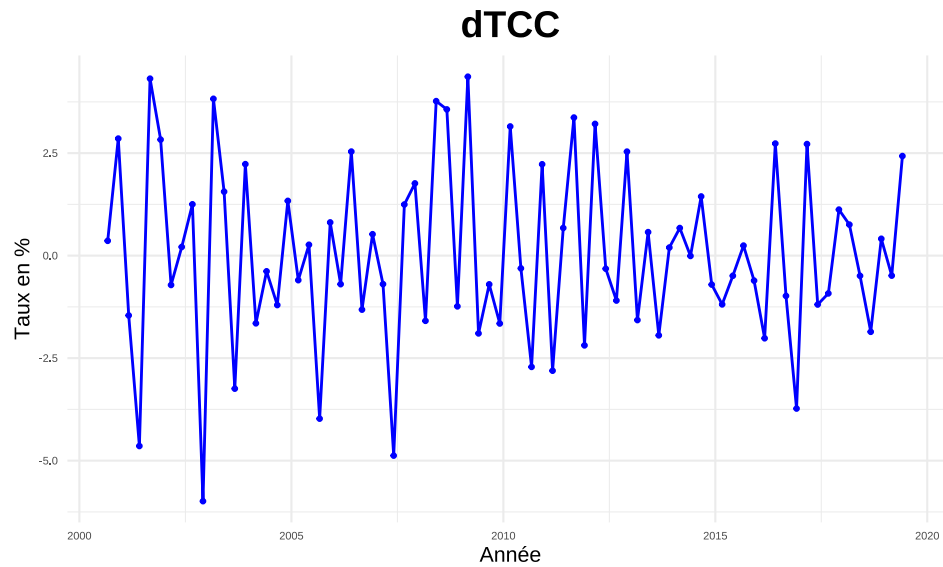


Statistique	Valeur
-------------	--------

La  $t$  value étant supérieure à la valeur critique 0.146, nous sommes amené à rejeter l'hypothèse nulle de stationnarité au seuil de 5%, ce qui est cohérent avec les résultats précédents.

### c Stationnarisation

Les résultats des tests ADF et KPSS montrent que la série suit un processus DS sans dérive. Il suffit donc de la différencier pour la stationnariser. On obtient alors une nouvelle série  $dTCC$ , dont l'interprétation est moins directe : il s'agit des variations d'un taux de croissance.



Visuellement, cette nouvelle série semble bien stationnaire en étant centrée autour de 0, soit un processus  $(X_t) \sim I(0)$ . Néanmoins, afin, de garantir la stationnarité de cette série et ainsi pouvoir déterminer le modèle ARMA adapté, il est nécessaire de réaliser à nouveau les tests ADF et KPSS.

Table 9: Resultats du test ADF sur M1 de  $dTCC$

Statistique	Valeur
Statistique ADF	-9.299
Valeur critique tau1 (5%)	-1.950

Après avoir rejeté les modèles M3 et M2 en raison de la non significativité de leur terme de tendance et de leur terme constant (*cf. fichier markdown*), nous testons finalement le modèle M1. Dans celui-ci, la statistique ADF est inférieure à la valeur critique  $\tau_{01}$  ce qui nous amène à rejeter  $H_0$ , impliquant donc l'absence de racine unitaire.

Table 10: Resultats du test KPSS sur  $dTCC$

Statistique	Valeur
Statistique KPSS	0.046
Valeur critique 5%	0.463

Finalement, la  $t$  value étant inférieure à la valeur critique 0.146, nous sommes amené à accepter l'hypothèse nulle de stationnarité au seuil de 5%, ce qui est cohérent avec les résultats du test ADF.

On peut donc affirmer, toujours au seuil  $\alpha = 0.05$ , que la série  $dTCC$  suit un processus stationnaire.

### 1.3 Détermination du modèle ARMA optimal

#### 1.3.1 Recherche des retards $p$ et $q$

La stationnarité de la série étant vérifiée, il est donc possible de la modéliser par un processus ARMA( $p,q$ ). Cependant, il faut déterminer le nombre optimal de retards  $p$  et  $q$  du modèle. Pour cela, on détermine en premier lieu le critère AIC de tous les modèles ARMA allant jusqu'à l'ordre 4 et on choisit celui ayant le critère le plus faible.

Table 11: Critère AIC d'un ARMA( $p,q$ )

	q=0	q=1	q=2	q=3	q=4
p=0	340.24	328.09	329.97	330.99	328.08
p=1	332.41	330.01	326.08	327.75	327.94
p=2	327.79	328.11	327.69	329.67	325.80
p=3	328.58	330.09	329.10	325.52	327.40
p=4	329.01	330.16	326.07	328.23	326.42

C'est donc le modèle ARMA(3,3) qui a le critère AIC le plus faible. Il faut désormais s'assurer de la significativité de ses coefficients.

Table 12: Estimation des paramètres de l'ARMA(3,3) avec  $c$

Coefficient	Estimate	Std_Error	z_value	Pr_z
ar1	0.254	0.0870	2.907	0.004
ar2	-0.476	0.0720	-6.568	5.1e-11
ar3	0.717	0.0850	8.398	2.2e-16
ma1	-0.676	0.0741	-9.117	2.2e-16
ma2	0.676	0.2080	3.242	0.001
ma3	-0.999	0.2440	-4.099	4.1e-05
intercept	0.033	0.0360	0.910	0.363

Le terme constant n'étant pas significatif, on considère un modèle sans celui-ci.

Table 13: Estimation des paramètres du ARMA(3,3) sans  $c$

Coefficient	Estimate	Std_Error	z_value	Pr_z
ar1	0.268	0.089	3.020	0.003
ar2	-0.470	0.084	-5.606	0.000
ar3	0.733	0.086	8.495	0.000
ma1	-0.675	0.411	-1.643	0.100
ma2	0.673	1.706	0.395	0.693
ma3	-0.998	2.104	-0.474	0.635

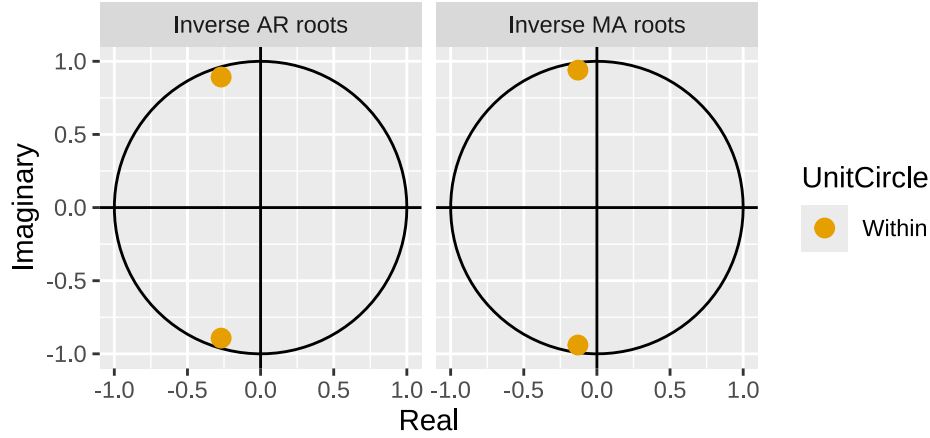
Une fois le terme constant retiré du modèle, aucun des coefficients  $MA$  n'est significatif. Pour pallier cela, 2 stratégies sont possibles :

- La première consiste simplement à retirer tous les termes *MA* soit à considérer un modèle ARMA(3,0). Cependant, en continuant cette stratégie, l'on résulte finalement avec un modèle ARMA(1,0) ayant un critère AIC de 332.41.
- La seconde consiste à retirer uniquement un seul terme *MA* afin d'analyser ensuite comment évolue la significativité des autres termes. En effectuant cela, on est amené à tester un modèle ARMA(3,2) et on constate qu'un seul coefficient *AR* n'est pas significatif. On considère donc finalement un modèle ARMA(2,2) dont tous les coefficients sont significatifs et ayant un critère AIC de 327.69.

On choisira donc le modèle ARMA(2,2) du fait de son critère AIC plus faible que celui de l'ARMA(1,0). Une fois le modèle estimé, on peut l'écrire :

$$X_t = -0.5396X_{t-1} - 0.8693X_{t-2} + \epsilon_t + 0.2594\epsilon_{t-1} + 0.9001\epsilon_{t-2}$$

On vérifie ensuite que les racines du polynôme  $\phi(L)$  satisfont bien les conditions de stationnarité en étant contenue dans le cercle unitaire.



L'analyse des racines inverses du modèle montre qu'elles sont bien à l'intérieur du cercle unité, ce qui confirme que le modèle satisfait les conditions de stationnarité et d'inversibilité. Cependant, leur proximité avec le cercle unité suggère une persistance des chocs, qui peut entraîner une autocorrélation prolongée dans les données. Bien que cette caractéristique ne remette pas en cause la stationnarité, elle pourrait affecter la précision des prévisions à long terme.

### 1.3.2 Analyse des résidus

Il est nécessaire de vérifier la qualité d'ajustement du modèle ARMA(2,2) en s'assurant que ses résidus se comportent comme un bruit blanc : non autocorrélés, de variance constante et suivant une loi normale.

#### a. Autocorrélation

Pour vérifier l'absence d'autocorrélation dans les résidus du modèle ARMA(2,2), nous avons choisi d'utiliser le test de Ljung-Box. La statistique de test  $Q^*$  est calculée ainsi :

$$Q^* = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}$$

où  $n$  est la taille de l'échantillon,  $h$  est le nombre de décalages, et  $\hat{\rho}_k$  l'autocorrélation des résidus au décalage  $k$ . Sous l'hypothèse nulle  $H_0$ ,  $Q^*$  suit une loi du chi-deux avec  $h - \text{df}$  degrés de liberté.

Pour notre modèle ARMA(2,2),  $Q^* = 4.908$ , avec  $\text{df} = 4$  et une  $p$ -value de 0.2969. Cette  $p$ -value étant supérieure à 0.05, nous ne rejetons pas  $H_0$ , ce qui indique que **les résidus sont bien non autocorrélés** et que le modèle est adapté.

### b. Homoscédasticité

Le test d'absence d'effet ARCH de *Engle* est utilisé pour détecter la présence d'hétéroscédasticité dans les résidus du modèle. La statistique de test utilisée est celle du multiplicateur de Lagrange et est calculée ainsi :

$$LM = T' \cdot R^2$$

où  $T' = T - p$  avec  $T$  le nombre d'observations et  $p$  le nombre de retard retenus pour le test.  $R^2$  est le coefficient de détermination de la régression des carrés des résidus sur leurs propres retards. Sous l'hypothèse nulle  $H_0$ , cette statistique suit une loi du chi-deux avec  $\text{df}$  degrés de liberté, correspondant au nombre de retards testés.

Pour notre modèle,  $LM = 16.099$  avec  $\text{df} = 20$  et une  $p$ -value de 0.7105. Comme cette  $p$ -value est bien supérieure à 0.05, nous ne rejetons pas  $H_0$ . Cela indique **l'absence d'effets ARCH significatifs dans les résidus**, ce qui confirme la stabilité de la variance du modèle.

### c. Distribution normale

Dans le cadre de l'évaluation de la normalité des résidus, le test de Jarque-Bera est couramment utilisé pour détecter toute déviation significative de l'asymétrie (skewness) et de l'aplatissement (kurtosis) par rapport à ceux d'une distribution normale. Sous l'hypothèse nulle ( $H_0$ ) de normalité, la statistique de test suit une distribution du chi-deux avec deux degrés de liberté.

$$JB = \frac{T}{6} \left( \hat{S}^2 + \frac{(\hat{K} - 3)^2}{4} \right)$$

où la statistique suit une loi du  $\chi^2$  à deux degrés de liberté sous l'hypothèse nulle de normalité des résidus.

Les termes sont définis comme suit :

- $T$  : nombre d'observations
- $k$  : nombre de paramètres estimés (0 pour une série brute)
- $\hat{S} = \frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}}$  : coefficient d'asymétrie (skewness)
- $\hat{K} = \frac{\hat{\mu}_4}{\hat{\mu}_2^2}$  : coefficient d'aplatissement (kurtosis)
- $\hat{\mu}_k = \frac{\sum_{t=1}^T (y_t - \bar{y})^k}{T}$  : moment empirique centré d'ordre  $k$

Dans notre cas,  $X^2 = 1.2533$ , avec une  $p$ -value de 0.5344. Cette  $p$ -value étant supérieure à 0.05, nous ne rejetons pas  $H_0$ , ce qui suggère que **les résidus suivent bien une distribution normale**.

## 1.4 Prévision

Nous savons désormais que le comportement des résidus du modèle est satisfaisant, il est donc possible d'établir des prévisions. Cependant, comme mentionné précédemment, ces prévisions seront limitées à un horizon assez court du à la proximité des racines du modèle avec le cercle unitaire

### 1.4.1 Hypothèses nécessaires

#### a. Espérance

Pour un horizon de prévision  $h$ , la prévision optimale de la valeur future  $Y_{T+h}$ , notée  $\hat{Y}_T(h)$ , est donnée par l'espérance conditionnelle de  $Y_{T+h}$  sachant l'ensemble des informations disponibles à l'instant  $T$ , noté  $I_T$  :

$$\hat{Y}_T(h) = E(Y_{T+h} \mid I_T)$$

avec  $I_T = (Y_1, \dots, Y_T)$

Les prévisions s'appuient sur les propriétés suivantes :

1. La prévision optimale de  $Y_{T+i}$  à partir de l'information disponible en  $T$ , notée  $\hat{Y}_T(i)$ , est donnée par :

$$E(Y_{T+i} \mid I_T) = \hat{Y}_T(i) \quad \text{pour } i > 0$$

2. Les innovations  $\epsilon_t$  sont indépendantes et centrées, ce qui implique :

$$E(\epsilon_{T+i} \mid I_T) = 0 \quad \text{pour } i > 0$$

Dans le contexte de la prévision en  $T$  : - Les valeurs passées  $Y_{T-i}$  (pour  $i \geq 0$ ) sont connues. - Les valeurs futures  $Y_{T+i}$  (pour  $i > 0$ ) sont inconnues et donc remplacées par leur prévision  $\hat{Y}_T(i)$ . - De même, les innovations passées  $\epsilon_{T-i}$  sont connues, tandis que les innovations futures  $\epsilon_{T+i}$  sont supposées de moyenne nulle.

#### b. Intervalle de précision

On définit l'erreur de prévision de  $Y_{T+h}$  faite en  $T$  :

$$\hat{\epsilon}_T(h) = Y_{T+h} - \hat{Y}_T(h)$$

Sous l'hypothèse de normalité des chocs, vérifiée par le test de *Jarque Bera*, l'erreur de prévision  $\hat{\epsilon}_T(h)$  suit une loi normale  $\mathcal{N}(0, V)$ , avec  $V = V(\hat{\epsilon}_T(h))$ .

La statistique normalisée associée est définie par :

$$\frac{Y_{T+h} - \hat{Y}_T(h)}{\sqrt{\hat{V}}} \sim \text{St}(T - K)$$

où  $\text{St}(T - K)$  représente la loi de Student si  $T - K < 30$ , et une loi normale  $\mathcal{N}(0, 1)$  pour  $T - K \geq 30$ .

L'intervalle de prévision au niveau de confiance  $1 - \alpha$  est alors donné par :

$$\text{IP} = \left[ \hat{Y}_T(h) - t_{1-\frac{\alpha}{2}} \sqrt{\hat{V}}, \hat{Y}_T(h) + t_{1-\frac{\alpha}{2}} \sqrt{\hat{V}} \right]$$

où  $t_{1-\frac{\alpha}{2}}$  est le quantile d'ordre  $1 - \frac{\alpha}{2}$  de la loi de Student ou de la loi normale.

### 1.4.2 Calcul

```
# Coefficients du modèle
coefficients <- coef(model_d)
phi1 <- coefficients["ar1"] # Coefficient AR(1)
phi2 <- coefficients["ar2"] # Coefficient AR(2)
theta1 <- coefficients["ma1"] # Coefficient MA(1)
```

```
theta2 <- coefficients["ma2"] # Coefficient MA(2)
```

```
# Dernières valeurs des séries et des résidus
```

```
Y_T <- tail(model_d$x, 1)
```

```
Y_T_1 <- tail(model_d$x, 2)[1]
```

```
eps_T <- tail(residuals(model_d), 1)
```

```
eps_T_1 <- tail(residuals(model_d), 2)[1]
```

- Pour  $h = 1$  :

```
Y_t1 <- phi1 * Y_T + phi2 * Y_T_1 + theta1 * eps_T + theta2 * eps_T_1
print(Y_t1)
```

```
##           Qtr1
## 2021 -1.117214
```

- Pour  $h = 2$  :

```
eps_T1 <- 0 # Résidu futur supposé nul
```

```
Y_t2 <- phi1 * Y_t1 + phi2 * Y_T + theta1 * eps_T1 + theta2 * eps_T
```

```
print(Y_t2)
```

```
##           Qtr1
## 2021 2.161302
```

- Pour  $h = 3$  :

```
eps_T2 <- 0 # Résidu futur supposé nul
```

```
Y_t3 <- phi1 * Y_t2 + phi2 * Y_t1 + theta1 * eps_T2 + theta2 * eps_T1
```

```
print(Y_t3)
```

```
##           Qtr1
## 2021 -0.1950931
```

À partir du package *forecast* de *R*, on vérifie ensuite facilement la validité de ces prévisions et leur intervalle :

	Point Forecast	Lo 95	Hi 95
2021 Q2	-1.117	-4.956	2.722
2021 Q3	2.161	-1.825	6.148
2021 Q4	-0.195	-4.242	3.852
2022 Q1	-1.773	-5.859	2.312



On peut désormais établir les prévisions de la série non filtrée  $TCC$ . En effet, nous avons établi que la série suivait un processus de type  $DS$  de la forme :

$$(1 - L)X_t = Y_t$$

$$\iff X_t = Y_t + X_{t-1}$$

On peut donc déduire les prévisions de  $X_t$  à des horizons de 1, 2 et 3 de manière itérative avec  $X_t = 5.563$ .

- Pour  $h = 1$  :

$$\begin{aligned} \hat{X}_T(1) &= E(Y_{T+1} + X_T \mid \mathcal{I}_T) = \hat{Y}_T(1) + X_T \\ \iff \hat{X}_T(1) &= -1.117 + 5.563 \\ \iff \hat{X}_T(1) &= 4.446 \end{aligned} \tag{1}$$

- Pour  $h = 2$  :

$$\hat{X}_T(2) = 2.161 + 5.563 = 7.724$$

- Pour  $h = 3$  :

$$\hat{X}_T(3) = -0.195 + 5.563 = 5.368$$

## Partie 2 : Modélisation multivariée

Dans cette section, nous élargissons notre analyse en adoptant une approche multivariée afin d'explorer les relations dynamiques entre le taux de chômage et le taux de croissance des crédits accordés aux entreprises non financières. Cette perspective permet de modéliser les interactions entre ces deux variables économiques clés et d'évaluer dans quelle mesure leurs évolutions conjointes peuvent améliorer la qualité des prévisions. Nous nous appuyons sur des modèles VAR (Vector AutoRegressive) pour capturer la structure temporelle et les interdépendances entre ces séries, tout en examinant les mécanismes de transmission potentiels au sein du système économique.

## 2.1 Estimation du modèle VAR

Pour procéder à l'estimation du modèle, il est nécessaire de déterminer le nombre de retards optimal de celui-ci. Pour cela, on va estimer différents modèles  $\text{VAR}(p)$  (avec  $p$  compris entre 0 et 6) et choisir celui qui minimise le critère  $AIC$ .

Table 15: Critères AIC par lag

Criteria	Lag 1	Lag 2	Lag 3	Lag 4
AIC	-5.64	-5.87	-5.99	-6.04

Le critère d'information  $AIC$  étant minimal pour  $p = 4$ , nous pouvons maintenant procéder à l'estimation d'un  $\text{VAR}(4)$ . Cependant, il est nécessaire de faire un compromis concernant la tendance déterministe du VAR : la série  $dTCC$  a été différenciée mais la série  $\log\_taux\_de\_chomage$  est de type TS. Nous avons donc choisi de n'inclure aucune composante déterministe ou constante dans le VAR car la série  $dTCC$  est déjà stationnaire après différenciation, et la tendance de la série  $\log\_taux\_de\_chomage$  a été correctement capturée dans sa transformation. Par conséquent, un modèle VAR sans constante ni tendance est approprié pour nos données.

Table 16: Résultats de l'estimation du VAR(5)

	Equation	Variable	Estimate	StdError	TValue	PValue
dTCC.dTCC.l1	dTCC	dTCC.l1	-0.59	0.12	-4.85	0.0000
dTCC.z.l1	dTCC	z.l1	-6.93	8.68	-0.80	0.4276
dTCC.dTCC.l2	dTCC	dTCC.l2	-0.54	0.13	-4.08	0.0001
dTCC.z.l2	dTCC	z.l2	-30.84	12.81	-2.41	0.0190
dTCC.dTCC.l3	dTCC	dTCC.l3	-0.19	0.13	-1.41	0.1633
dTCC.z.l3	dTCC	z.l3	22.53	12.82	1.76	0.0837
dTCC.dTCC.l4	dTCC	dTCC.l4	-0.28	0.12	-2.39	0.0200
dTCC.z.l4	dTCC	z.l4	9.30	9.00	1.03	0.3057
z.dTCC.l1	z	dTCC.l1	0.00	0.00	-0.41	0.6796
z.z.l1	z	z.l1	1.09	0.12	8.84	0.0000
z.dTCC.l2	z	dTCC.l2	0.00	0.00	0.01	0.9953
z.z.l2	z	z.l2	0.04	0.18	0.24	0.8137
z.dTCC.l3	z	dTCC.l3	0.00	0.00	1.09	0.2783
z.z.l3	z	z.l3	-0.04	0.18	-0.24	0.8142
z.dTCC.l4	z	dTCC.l4	0.00	0.00	1.51	0.1358
z.z.l4	z	z.l4	-0.22	0.13	-1.70	0.0932

Après estimation du  $\text{VAR}(4)$ , on remarque que la plupart des coefficients ne sont pas significatifs. Ceci peut être causé par deux raisons :

- une forte **multicolinéarité** : cela peut se produire lorsque les variables explicatives dans le modèle sont fortement corrélées entre elles, ce qui rend difficile l'estimation précise des coefficients. La multicolinéarité peut entraîner des intervalles de confiance larges pour les coefficients et rendre les tests de significativité moins fiables.
- un **manque de corrélation** : si les deux séries ne sont pas suffisamment corrélées, le modèle VAR peut ne pas capter une relation significative entre elles. Dans ce cas, les coefficients associés aux variables non corrélées peuvent être statistiquement non significatifs.



Pour autant, les racines du modèle en module sont bien inférieures à 1, garantissant sa stationnarité.

Table 17: Racines du modèle VAR(5)

X1	X2	X3	X4	X5	X6	X7	X8
0.87	0.87	0.78	0.78	0.58	0.58	0.5	0.5

De plus, on vérifie bien que les résidus du modèle ne sont pas autocorrélés :

Table 18: Portmanteau Test (adjusted)

Statistique	Valeur
Chi-squared	18.87
Degrés de liberté (df)	20.00
p-value	0.53

La *pvalue* étant supérieure à 5%, on ne rejette pas  $H_0$  et pouvons conclure que les résidus ne sont pas autocorrélés.

## 2.2 Test de causalité

Avant de procéder à l'analyse de la causalité, il est essentiel d'examiner si l'une des variables permet de prédire les variations de l'autre. Pour ce faire, nous utilisons le test de causalité au sens de Granger, qui évalue si les informations passées d'une série améliorent la prévision d'une autre.

Table 19: Résultat du test de Granger

Hypothèse	F_Test	df1	df2	p_value
dTCC ne cause pas z	0.804	5	128	0.52
z ne cause pas dTCC	6.239	5	128	0.00

Les résultats du test sont particulièrement intéressants. Ils **contredisent** notre intuition initiale selon laquelle le taux de croissance des crédits accordés aux entreprises influencerait le taux de chômage. En effet, la *p-value* de l'hypothèse selon laquelle il n'y a pas de causalité entre la différentielle du taux de croissance des crédits (*dTCC*) et le taux de chômage (*z*) est de 0.7, ce qui indique que nous ne rejetons pas l'hypothèse nulle. Autrement dit, il n'y a pas de preuve significative de causalité de *dTCC* vers *z*.

En revanche, la *p-value* de l'hypothèse selon laquelle le taux de chômage (*z*) ne cause pas la différentielle du taux de croissance des crédits (*dTCC*) est très faible ( $3.06e - 05$ ), ce qui nous pousse à rejeter l'hypothèse nulle. Cela suggère que le taux de chômage a une influence significative sur la différentielle du taux de croissance des crédits accordés aux entreprises.

Ces résultats peuvent refléter l'impact des conditions du marché du travail sur les anticipations économiques des entreprises. Un chômage élevé pourrait inciter à une prudence accrue dans l'octroi ou la demande de crédits, notamment en raison d'une consommation atone et d'une confiance affaiblie dans la capacité de remboursement. À l'inverse, un marché du travail en bonne santé favorise les perspectives de croissance, stimulant ainsi la dynamique de crédit.

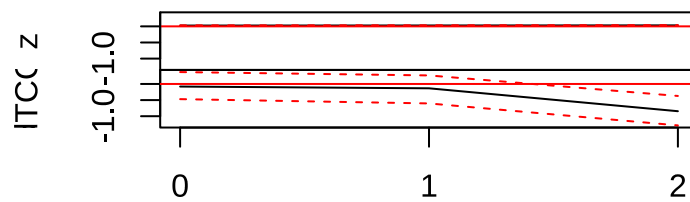
## 2.3 Analyse impulsion-repulsion

Nous prolongeons maintenant l'analyse des relations dynamiques entre les variables en recourant à l'approche impulsion-répulsion. Cette méthodologie permettant d'examiner la réponse des variables du système à un choc exogène affectant une variable spécifique. Pour cela, nous passerons par deux méthodes : celle des VAR (a) et celle des projections locales (b).

### a. Méthode des VAR

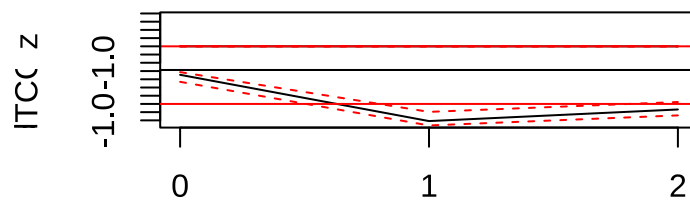
Dans le cadre de l'analyse VAR, nous utilisons une décomposition de Choleski pour identifier les chocs exogènes, en plaçant les variables selon leur causalité. Le test de Granger a montré que  $z$  (résidus du log taux de chômage) cause  $dTCC$  (différentiel de croissance des crédits), justifiant ainsi de placer  $z$  en première position. La décomposition de Choleski suppose que les chocs sont orthogonaux, ce qui permet d'isoler les effets spécifiques de chaque variable. Cet ordre est crucial, car une mauvaise spécification peut biaiser les résultats.

Orthogonal Impulse Response from  $z$



95 % Bootstrap CI, 100 runs

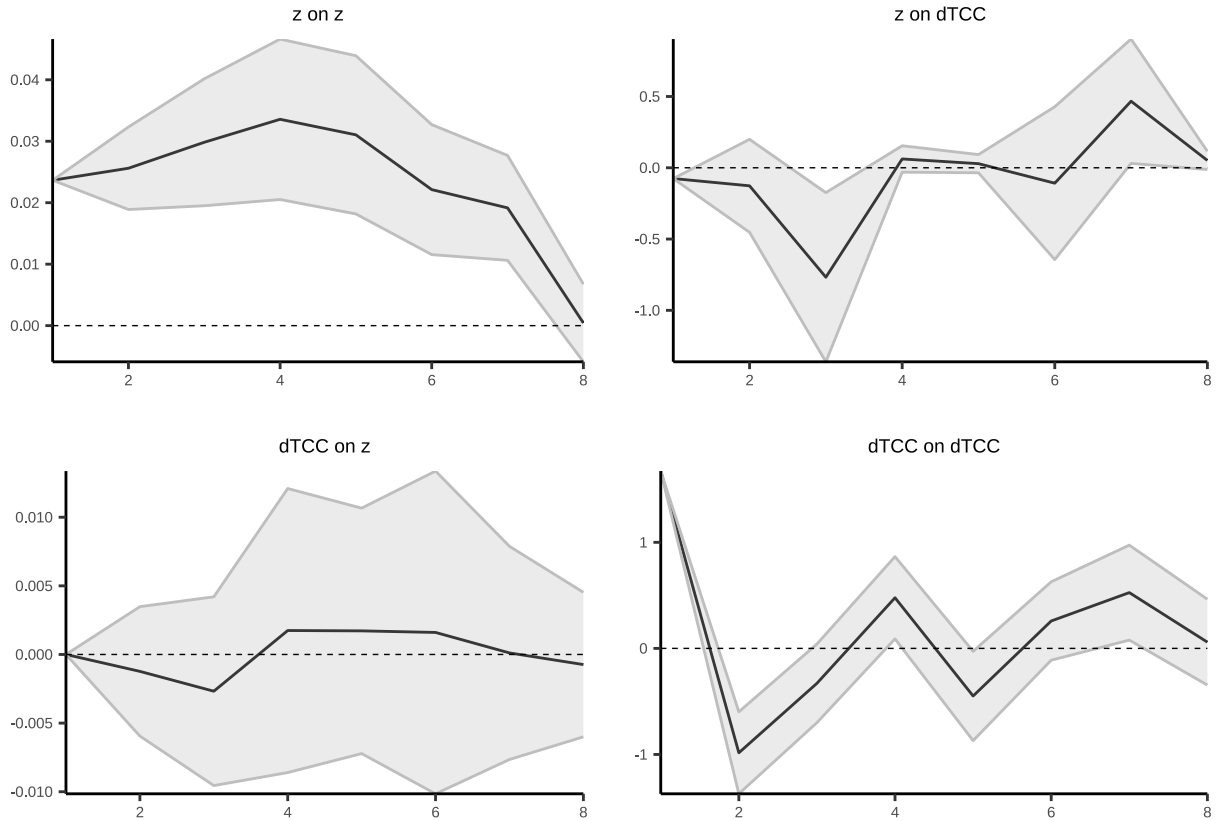
Orthogonal Impulse Response from  $dTCC$



95 % Bootstrap CI, 100 runs

Les réponses impulsionnelles révèlent des interactions dynamiques asymétriques entre  $z$  et  $dTCC$ . Un choc orthogonal sur  $z$  provoque une réponse négative transitoire de  $dTCC$ , avec un effet maximal au premier horizon et un retour progressif à l'équilibre. Inversement, un choc sur  $dTCC$  engendre une réaction initialement négative de  $z$ , dont l'amplitude est plus modérée et s'estompe rapidement. Les intervalles de confiance à 95 % montrent que ces effets, bien que transitoires, sont significatifs dans les premiers horizons.

## b. Méthode des projections locales



- **Réponse de  $z$  à un choc orthogonal sur  $z$  :** le graphique en haut à gauche montre une persistance positive des effets de  $z$  sur lui-même, avec une tendance décroissante au fil des horizons temporels. Les intervalles de confiance encadrent ces effets, confirmant leur significativité initiale.
- **Réponse de  $dTCC$  à un choc orthogonal sur  $z$  :** en haut à droite, la réponse initiale de  $dTCC$  à un choc sur  $z$  est négative avant de fluctuer autour de zéro. Ces oscillations reflètent une dynamique complexe, mais leur significativité est limitée aux premiers horizons.
- **Réponse de  $z$  à un choc orthogonal sur  $dTCC$  :** en bas à gauche, les effets sur  $z$  d'un choc sur  $dTCC$  sont très faibles et proches de zéro, indiquant une absence de relation significative entre les deux variables pour cette direction de causalité.
- **Réponse de  $dTCC$  à un choc orthogonal sur  $dTCC$  :** en bas à droite,  $dTCC$  répond à ses propres chocs de manière oscillatoire avec une forte réponse initiale, suivie d'une dissipation progressive. Les intervalles de confiance suggèrent une significativité notable sur les premiers horizons.

## 2.4 Test de cointégration

Pour terminer, nous allons tester si les séries initiales (non différenciées) présentent une relation de cointégration à l'aide du test de Johansen. Basé sur un VAR, ce test identifie le nombre de relations de cointégration via la statistiques de trace. La présence de cointégration justifie l'estimation d'un modèle vectoriel à correction d'erreur (VECM) pour intégrer les dynamiques de court et long terme. Cette section applique et interprète ce test afin de valider ou non le modèle VAR initial.

Le test de cointégration de Johansen permet de tester les hypothèses suivantes :

- $H_0 : r = 0$  (absence de relation de cointégration entre les variables)
- $H_1 : r > 0$  (présence d'au moins une relation de cointégration entre les variables)

La statistique de test pour  $r = 0$  est de 14.16, inférieure à la valeur critique au seuil de 5% (19.96). Nous ne rejetons donc pas  $H_0$  ce qui indique l'absence de relation de cointégration. Il est alors inutile d'estimer un VECM.

On peut donc déduire que le VAR en niveau reste approprié pour modéliser les dynamiques entre les variables

## Conclusion

Cette étude économétrique a permis de mieux comprendre les relations dynamiques entre le taux de chômage et le taux de croissance des crédits accordés aux entreprises non financières. L'approche univariée a d'abord permis de stationnariser les séries pour modéliser leurs dynamiques individuelles à l'aide de processus ARMA. Les prévisions réalisées, bien que limitées par la proximité des racines du modèle avec le cercle unitaire, se sont révélées robustes à court terme.

Dans un second temps, l'analyse multivariée a mis en évidence des relations intéressantes entre les deux séries. Le modèle VAR a révélé une causalité significative du taux de chômage vers la croissance des crédits, mais pas l'inverse, ce qui a été confirmé par les tests de causalité de Granger. Par ailleurs, l'analyse impulsion-répulsion a montré des interactions dynamiques asymétriques, et le test de cointégration a mis en lumière une relation de long terme entre ces deux variables.