

---

## L'analyse en composantes principales de variables non stationnaires

*Principal components analysis of non stationary variables*

Philippe Casin, Christine Stachowiak et François Marque

---



### Édition électronique

URL : <http://journals.openedition.org/msh/12117>

DOI : 10.4000/msh.12117

ISSN : 1950-6821

### Éditeur

Centre d'analyse et de mathématique sociales de l'EHESS

### Édition imprimée

Date de publication : 30 décembre 2011

Pagination : 27-40

ISSN : 0987-6936

### Référence électronique

Philippe Casin, Christine Stachowiak et François Marque, « L'analyse en composantes principales de variables non stationnaires », *Mathématiques et sciences humaines* [En ligne], 196 | hiver 2011, document 1459, mis en ligne le 15 avril 2012, consulté le 01 mai 2019. URL : <http://journals.openedition.org/msh/12117> ; DOI : 10.4000/msh.12117

---

## L'ANALYSE EN COMPOSANTES PRINCIPALES DE VARIABLES NON STATIONNAIRES

Philippe CASIN<sup>1</sup>, Christine STACHOWIAK<sup>1</sup>, François MARQUE<sup>1</sup>

**RÉSUMÉ** – *Les séries chronologiques dont on dispose en économie sont souvent non stationnaires, et il n'est donc pas possible de les décrire à partir d'une analyse en composantes principales : en effet, l'analyse en composantes principales est basée sur l'analyse de la matrice des corrélations entre les variables, et ces corrélations sont fallacieuses. Cet article présente une technique pour décrire ce type de variables, en filtrant les données initiales par une suite d'analyse en composantes principales pour obtenir des séries résiduelles stationnaires ; l'analyse en composantes principales de ces séries est alors interprétable. Deux applications, l'une basée sur des données simulées et l'autre sur des données réelles, sont fournies.*

**MOTS-CLÉS** – Analyse en composantes principales, Cointégration, Variables non stationnaires

**SUMMARY** – Principal components analysis of non stationary variables

*Economical time series are often non stationary and then it is not possible to practice principal components analysis: this technique is based on analysis of the correlation matrix and correlations between original variables are spurious. The aim of this paper is to introduce a new technique which describes sets of non stationary variables. This technique is based on successive analyses of residuals which provide stationary variables. Two applications are given, the first one uses simulated data, the second one Nelson and Plosser's data.*

**KEYWORDS** – Cointegration, Non-stationary variables, Principal components analysis

### 1. INTRODUCTION

Effectuer un test de Student sur le coefficient de corrélation entre deux variables chronologiques peut conduire à des conclusions erronées concernant l'existence d'une relation linéaire entre ces variables : c'est le problème des régressions « fallacieuses ».

*A fortiori*, les résultats d'une analyse en composantes principales (ACP) étant basés sur la diagonalisation d'une matrice de corrélations ou d'une matrice de covariances sont ininterprétables lorsque certaines des variables sont non stationnaires, ce qui est souvent le cas en Économie.

L'objet de cet article est de proposer un algorithme permettant d'effectuer une ACP de variables dont certaines sont non stationnaires ; pour cela, nous utilisons des résultats établis en économétrie des séries temporelles et inexploités en analyse de données. Ces résultats concernent l'existence de relations de cointégration entre les

---

<sup>1</sup> Centre Européen de Recherche en Économie Financière et Gestion des Entreprises (CEREFIGE), UFR Droit, Économie, Administration, Université de Lorraine, Ile du Saulcy, 57005 Metz cedex, philippe.casin@wanadoo.fr ; stachowiak@univ-metz.fr ; marque@univ-metz.fr

variables non stationnaires. Ces résultats permettent de modéliser les relations de long terme qui existe entre les séries.

La suite de l'article se déroule comme suit. Dans la seconde section, nous exposons le problème avec notamment un exemple d'ACP fallacieuse. Dans la troisième section, nous rappelons comment l'ACP traite les séries non stationnaires, sans parvenir à mettre en évidence les éventuelles relations de cointégration entre celles-ci. La quatrième section présente la méthode d'ACP sur des séries non stationnaires. La cinquième section est consacrée à l'application de cet algorithme à deux exemples, l'un à partir de données simulées, l'autre à partir de données réelles, caractérisant l'économie des États-Unis de 1909 à 1970. La conclusion de ce papier fait l'objet de la sixième section.

## 2. LE PROBLÈME

### 2.1 LA RÉGRESSION FALLACIEUSE (“SPURIOUS REGRESSION”)

Considérons deux séries, notées  $X_1$  et  $X_2$  dont les valeurs à l'instant  $t$  sont notées respectivement  $X_{1,t}$  et  $X_{2,t}$ .  $X_{1,t}$  et  $X_{2,t}$  sont définies de la manière suivante :  $X_{j,t} = X_{j,t-1} + u_{j,t}$  pour  $j = 1, 2$ ,  $u_1$  et  $u_2$  étant deux bruits blancs indépendants.

Pour une simulation générant 10 000 valeurs de ces deux marches aléatoires calculées pour 5000 instants, la valeur de  $R^2$  varie de 0 à 0.936 avec une moyenne de 0.241 et un écart-type de 0.224.

Le coefficient de régression entre  $X_1$  et  $X_2$  n'est pas donc pas égal à 0 (cf. aussi [Gouriéroux et Monfort, 1995], [Lardic et Mignon, 2002], [Casin, 2009] par exemple), bien que les deux variables soient déterminées indépendamment l'une de l'autre ; les tests habituels concernant ce coefficient de régression ne sont pas utilisables, car ils conduisent dans le cas des variables non stationnaires à des conclusions erronées, c'est à dire à conclure à l'existence d'une relation linéaire entre les variables qui, en fait, est « fallacieuse ».

### 2.2. L'ACP FALLACIEUSE

#### 2.2.1. Le problème

Les résultats de l'analyse en composantes principales d'un ensemble de variables sont obtenus en diagonalisant la matrice de corrélations (ou la matrice de variances-covariances, s'il s'agit de variables non normées) entre les variables (cf [Bouroche, Saporta, 1980], [Casin, 1999], par exemple). Dans le cas de variables non stationnaires, le coefficient de corrélation est « fallacieux » et l'ACP effectuée donne elle aussi des résultats « fallacieux ».

#### 2.2.2. Un exemple d'ACP fallacieuse

Lorsque l'on considère des séries linéairement indépendantes et stationnaires, leur coefficient de corrélation est nul, l'ACP conduit donc à diagonaliser la matrice identité et on obtient des valeurs propres égales à 1.

Considérons ici 10 marches aléatoires  $X_j, j = 1, \dots, 10$  telles que  $X_{j,t} = X_{j,t-1} + u_{j,t}$ , les variables  $u_j$  désignant des bruits blancs deux à deux indépendants. On calcule les 1000 premières valeurs de chacune des 10 variables  $X_j$ , et on effectue 5000 fois l'ACP

du tableau à 1000 lignes et 10 colonnes obtenues, chacune des 5000 simulations étant obtenue pour des valeurs numériques des bruits blancs différentes. Les caractéristiques des valeurs propres sont alors les suivantes :

	Minimum	Maximum	Moyenne	Écart-type
Étape 1	0.270	0.793	0.490	0.084
Étape 2	0.078	0.377	0.212	0.047
Étape 3	0.039	0.217	0.116	0.028
Étape 4	0.019	0.150	0.069	0.019
Étape 5	0.010	0.102	0.043	0.013
Étape 6	0.007	0.070	0.028	0.009
Étape 7	0.005	0.049	0.018	0.005
Étape 8	0.003	0.042	0.012	0.004
Étape 9	0.001	0.024	0.008	0.002
Étape 10	< 0.001	0.016	0.004	0.002

TABLEAU 1. Caractéristiques des valeurs propres

Si l'on considère l'ACP « moyenne », l'examen de la colonne correspondante suggère de considérer 2 ou 3 étapes, celles-ci expliquant l'essentiel de la variance du nuage de points ; de plus, les dernières valeurs propres ayant des valeurs proches de 0, certaines des 10 variables semblent liées entre elles par des relations linéaires. En réalité, ces conclusions n'ont aucun sens, les séries de départ ne faisant que décrire des sommes de bruits blancs indépendants.

### 3. METTRE EN ÉVIDENCE LES RELATIONS DE LONG TERME

#### 3.2. NOTATION ET DÉFINITIONS

On note  $I(0)$  une variable stationnaire et de façon plus générale  $I(d)$  une variable dont la différence d'ordre  $d$  est stationnaire ; ainsi la variable  $X_t$  est  $I(1)$  si la série  $X_t - X_{t-1}$  est stationnaire. La non stationnarité d'une série peut être soit d'origine déterministe du fait de l'existence d'un trend linéaire (du type  $at + b$ , avec  $a$  différent de 0), soit d'origine stochastique comme dans le cas des marches aléatoires. Dans cet article, nous considérons uniquement des séries  $I(1)$  dont la non stationnarité est due à des causes stochastiques ; éventuellement, les séries initiales pourront être régressées par rapport à  $t$  pour se ramener à ce cas de figure en leur ôtant leur tendance temporelle.

Les  $p$  colonnes du tableau  $X$  sont donc constituées de  $p$  variables  $X_j$ ,  $j = 1, \dots, p$  dont les valeurs aux instants  $t = 1, \dots, T$  sont notées  $X_{j,t}$  ; ces variables sont soit  $I(1)$  et sans trend linéaire, ou  $I(0)$ . Les  $T$  lignes de ce tableau sont les  $p$  observations aux  $T$  différents instants.

#### 3.2. COINTÉGRATION ET RELATIONS DE LONG TERME

Une combinaison linéaire de variables  $I(1)$  peut être  $I(1)$ , mais peut être aussi  $I(0)$ . On dit, dans ce dernier cas, qu'il existe une relation de cointégration entre les variables. La

mise en évidence de relations de cointégration entre des séries non stationnaires est un chapitre important de l'économétrie des séries temporelles [Engle et Granger, 1987] dans le cas univarié – mise en évidence d'une relation de cointégration entre des variables – [Johansen, 1988], dans le cas multivarié – mise en évidence de plusieurs relations de cointégration linéairement indépendantes en utilisant l'analyse canonique. Ces relations de cointégration, qui lient entre elles des variables au cours du temps, sont appelées aussi relations de long terme, car elles lient entre elles les variables considérées en niveau absolu. À l'inverse, les relations de court-terme sont celles qui décrivent des relations entre les variations des variables, autrement dit les variables prises en différence qui sont stationnaires.

Ici, nous supposons qu'il existe  $r$  ( $r < p$ ) relations de cointégration entre les variables  $X_j, j = 1, \dots, p$ .

### 3.3. LE TRAITEMENT CLASSIQUE DES SÉRIES NON STATIONNAIRES EN ACP

Comme cela a été montré dans le paragraphe 2.2.1, effectuer l'ACP de séries non stationnaires conduit à des résultats fallacieux, donc sans intérêt.

Une possibilité consiste à construire à partir des séries non stationnaires des séries stationnaires et à traiter ensuite ces séries, obtenues à partir des séries initiales en calculant leurs accroissements relatifs ou absolus.

Mais en procédant ainsi, seules des relations de court terme sont mises en évidence entre les variables et non des relations de long terme. Ainsi, considérons les deux séries suivantes :  $X_{1,t} = X_{1,t-1} + u_t$  et  $X_{2,t} = 10 X_{1,t-1}$ . Ces deux séries ne sont pas stationnaires ; si l'on considère leurs différences premières, il s'agit de bruits blancs. Pourtant, il existe une relation vérifiée approximativement entre les variables, puisque  $X_{1,t} = 0.1 X_{2,t} + u_t$ , relation de long terme qu'il est intéressant de mettre en évidence. La relation de court terme correspondante est  $\Delta X_{1,t} = 0.1 \Delta X_{2,t} + \Delta u_t$ , qui ne sera pas forcément mise en évidence par une ACP si  $0.1 \Delta X_{2,t}$  est petit par rapport à  $\Delta u_t$ . A tout le moins, le coefficient de corrélation entre  $X_1$  et  $X_2$  est différent du coefficient de corrélation entre  $\Delta X_1$  et  $\Delta X_2$ , et donc les ACP correspondantes donnent des résultats très différents.

Filtrer les séries de départ pour ne considérer que leurs différences premières, c'est-à-dire pour ne s'occuper que de séries stationnaires, occulte donc les relations existant entre les niveaux absolus des variables. L'objet de cet article est de proposer une technique permettant d'effectuer l'ACP des données de départ et de mettre en évidence les relations de long terme, c'est à dire les relations linéaires non fallacieuses entre les variables.

## 4. L'ACP DE VARIABLES NON STATIONNAIRES

### 4.1. UNE PRÉSENTATION GÉOMÉTRIQUE DU PROBLÈME

Lorsqu'une variable est  $I(1)$ , alors sa variance est, en utilisant la notation de Landau,  $O_p(T)$ , c'est à dire est un infiniment grand équivalent à  $T$ , lorsque  $T$  tend vers l'infini, l'équivalence se faisant au sens d'une convergence en probabilité ; d'autre part, lorsqu'une variable est  $I(0)$ , alors sa variance est  $O_p(1)$ . ([Harris, 1997 ; Snell, 1999 ; Chigira, 2005], par exemple).

Les combinaisons linéaires non cointégrées, c'est à dire  $I(1)$ , ont une variance qui a pour ordre de grandeur  $T$  tandis que les combinaisons linéaires cointégrées, c'est-à-dire  $I(0)$ , sont  $O_p(1)$ . On en déduit [Engle et Granger, 1987] que s'il existe  $r$  ( $r < p$ ) relations de cointégration entre les variables  $X_j, j = 1, \dots, p$ , alors lorsque  $T$  tend vers l'infini, la matrice de covariances entre les variables admet pour limite une matrice finie singulière, dont le noyau est constitué par l'espace des  $r$  dernières composantes principales.

Dès lors, l'espace des  $p$  variables  $X_j$  peut être décomposé en deux sous-espaces orthogonaux :

- l'un, noté  $W_0$ , de dimension  $r$  est engendré par les dernières composantes principales et se caractérise par des variables  $O_p(1)$ , c'est-à-dire  $I(0)$  c'est-à-dire stationnaires
- l'autre, noté  $W_1$  est généré par des variables  $O_p(T)$ , c'est-à-dire  $I(1)$ , est de dimension  $p - r$ .

Aussi, si on effectue l'ACP du tableau des  $p$  variables, les  $p - r$  premières composantes principales constituent une base de l'espace  $W_1$  tandis que les  $r$  suivantes engendrent l'espace  $W_0$ .

Ce résultat, présenté ici de façon intuitive a été exploité par [Harris,1997 ; Snell, 1999 ; Chigira, 2005] pour mettre en évidence des relations de cointégration entre des séries  $I(1)$ , c'est-à-dire pour proposer une alternative à la méthode d'estimation de [Johansen, 1988], basée sur l'analyse canonique.

Dans ce contexte, les relations les plus intéressantes sont celles mises en évidence par les toutes dernières composantes principales, puisqu'elles décrivent les relations de cointégration les plus fortes entre les variables, leur variance résiduelle étant la plus faible possible. La dernière composante principale correspondant à une régression orthogonale ([Malinvaud, 1978], par exemple), la technique proposée revient donc à estimer par la régression orthogonale ce qui ne peut l'être par la régression des moindres carrés.

Ainsi, Harris [1997] teste la stationnarité des composantes principales, les variables étant les logarithmes de la consommation, de l'investissement et de la production par tête, exprimés en termes réels. Les données sont trimestrielles et concernent l'économie australienne de juin 71 à septembre 1994. Le test met en évidence deux relations de cointégration et de ces deux relations de cointégration se déduit alors un système permettant d'expliquer deux des trois variables.

Les résultats concernant la stationnarité des composantes principales n'ont pas été exploités en analyse des données, c'est à dire dans le cadre usuel de l'ACP comme technique descriptive d'un grand ensemble de données. Dans ce cadre là, les composantes principales les plus intéressantes ne sont pas les dernières, mais celles des composantes  $I(0)$  qui ont une variance maximale, autrement dit les premières composantes principales stationnaires.

La méthode que nous proposons ici – l'ACP de variables non stationnaires – est pour le modèle à correction d'erreur ce qu'est l'ACP usuelle à la régression multiple ; l'objet de l'ACP (sur des variables non stationnaires, comme ici, ou sur d'autres variables, pour l'ACP usuelle) est de décrire un grand ensemble de données à l'aide d'un petit nombre de variables synthétiques, alors que l'objet de la régression ou du modèle à correction d'erreur est de construire un modèle explicatif, à partir d'une variable endogène à expliquer et de variables exogènes explicatives.

#### 4.2. L'ALGORITHME DE CALCUL

Si on effectue une ACP sur des variables  $I(1)$ , les  $p - r$  premières composantes principales vont être  $I(1)$  et donc fallacieuses, les suivantes vont être  $I(0)$ , et un test de stationnarité (cf. paragraphe 4.3) permet de déterminer la valeur de  $r$ .

Mais en procédant ainsi, on se sert de résultats asymptotiques. Considérons une variable de départ  $I(0)$  et effectuons l'ACP sur l'ensemble des variables mixtes  $I(1)$  ou  $I(0)$  ; asymptotiquement, cette variable  $I(0)$  est orthogonale à la composante principale d'ordre 1, puisque celle-ci est  $I(1)$ . En pratique, l'échantillon ayant une taille finie, elle ne l'est sans doute pas exactement.

La deuxième étape de l'ACP est équivalente à une ACP des résidus des variables du tableau d'origine par rapport à la première composante principale (il s'agit d'un résultat classique). Le résidu d'une variable  $I(0)$  par la première composante principale est donc une combinaison linéaire d'une variable  $I(0)$  à expliquer par une variable  $I(1)$  explicative et peut donc être  $I(1)$ . Plus généralement, la composante principale d'ordre  $r$  du tableau de départ est la première composante principale du tableau des résidus des variables de départ par les  $r - 1$  premières composantes principales ; les résidus d'une variable  $I(0)$  par les  $r - 1$  premières composantes principales peuvent être  $I(1)$ .

En effectuant une ACP du tableau de départ, on peut ainsi transformer des variables  $I(0)$  en variables  $I(1)$ , ce qui est le contraire de l'objectif de l'analyse. Il est donc préférable de procéder ainsi, c'est-à-dire préférable d'effectuer des ACP emboîtées :

- on teste le caractère  $I(0)$  ou  $I(1)$  de chacune des variables, et on effectue l'ACP des seules variables  $I(1)$ , les composantes principales fallacieuses étant dans l'espace engendré par ces variables  $I(1)$  ;
- on considère les résidus des variables  $I(1)$  par la première composante principale de cette ACP ;
- on effectue l'ACP des seuls résidus qui sont  $I(1)$  ;
- etc., jusqu'au moment où tous les résidus sont  $I(0)$  ; les résultats de l'ACP sont alors tous  $I(0)$  et interprétables ;
- il ne reste alors plus qu'à effectuer une ACP (normée ou non) de ces résidus stationnaires.

#### 4.3. LES TESTS PRATIQUES

Les tests pratiqués par [Harris, 1998 ; Snell, 1999 ; ou Chigira, 2005] ont pour objet de déterminer le nombre  $r$  de relations de cointégration indépendantes existant entre les variables et testent donc le caractère  $I(0)$  ou  $I(1)$  de la composante principale d'ordre  $p - r$  ; c'est à dire que les tests débutent par la composante principale de variance la moins élevée pour se terminer lorsque le test conclut que la composante principale faisant l'objet du test est  $I(0)$ .

### 5. DEUX EXEMPLES D'APPLICATION

Dans ce paragraphe, nous allons illustrer la technique proposée par deux applications, la première sur des données simulées, la seconde sur des données réelles.

## 5.1. UN EXEMPLE ILLUSTRATIF À PARTIR DE DONNÉES SIMULÉES

Considérons les séries chronologiques  $x_j$  pour  $j = 1, \dots, 6$  ; ces 6 séries sont liées entre elles par les 2 relations de cointégration suivantes :

$$x_{1,t} = x_{1,t-1} + u_{1,t}$$

$$x_{2,t} = x_{2,t-1} + u_{2,t}$$

$$x_{3,t} = x_{1,t} + x_{2,t}$$

$$x_{4,t} = u_{4,t}$$

$$x_{5,t} = x_{4,t} + 0.2 x_{1,t}$$

$$x_{6,t} = 2x_{4,t} + u_{6,t}$$

Les variables  $x_1, x_2, x_3$  et  $x_5$  sont  $I(1)$ , tandis que  $x_4$  et  $x_6$  sont  $I(0)$ . Plus  $T$ , le nombre d'observations, est élevé, plus les 4 variables  $I(1)$  ont des normes élevées alors que les normes des variables  $x_4$  et  $x_6$  sont  $O_p(1)$ .  $x_3$  est, par construction dans le plan engendré par  $x_1$  et  $x_2$ . Dès lors, l'ACP du tableau décrivant les 6 variables admettra, pour  $T$  infini, comme deux premières composantes principales une base de l'espace engendré par  $x_1$  et  $x_2$ . Les 4 composantes principales suivantes décriront les résidus des variables  $x_3, x_4, x_5$  et  $x_6$  par les variables  $x_1, x_2$ . Autrement dit, les composantes d'ordre supérieur ou égal à 3 décrivent des relations non fallacieuses entre la partie non stochastique des variables de départ.

Pour tester la stationnarité des séries, nous optons dans ce paragraphe pour le test de [Dickey-Fuller, 1979,1981] qui est le test le plus couramment utilisé en économétrie.

Pour une simulation particulière pour laquelle  $T = 50\,000$  :

- 1) on teste la stationnarité des séries au seuil  $\alpha = 5\%$  par un test de Dickey-Fuller, on écarte  $x_4$  et  $x_6$  qui sont stationnaires ;
- 2) on effectue l'ACP non normée des séries  $x_1, x_2, x_3$  et  $x_5$ , les valeurs propres sont égales à 0.00, 0.97, 4050.47 et 53247.66 ;
- 3) on teste la stationnarité de la première composante principale ; celle-ci étant non stationnaire, on continue les calculs ;
- 4) on teste la stationnarité des résidus des 4 variables  $x_1, x_2, x_3$  et  $x_5$  par la première composante principale ; ceux-ci étant non stationnaires, on effectue l'ACP des résidus non stationnaires, les valeurs propres sont 0.00, 0.00, 0.97 et 4050.47 ;
- 5) on teste la stationnarité de la première composante principale de cette ACP, qui est donc la deuxième composante de l'ACP des séries  $x_1, x_2, x_3$  et  $x_5$  ; celle-ci étant non stationnaire, on continue les calculs ;
- 6) on teste la stationnarité des résidus des 4 variables  $x_1, x_2, x_3$  et  $x_5$  par cette composante principale ; ceux-ci sont stationnaires. Les variances de ces résidus sont égales respectivement à 0.014, 0.04, 0.04 et 0.945. On récupère donc bien la partie stationnaire de la série  $x_5$  ; pour  $x_1, x_2, x_3$ , les variances résiduelles sont quasi-nulles ;
- 7) on effectue alors l'ACP de  $x_4$  et  $x_6$  et des résidus des régressions de  $x_1, x_2, x_3$  et  $x_5$  par les deux composantes principales calculées en 3) et 5) ; les valeurs propres sont égales à 0.00, 0.00, 0.00, 0.00, 0.299 et 6.668. Les corrélations entre les 2 premières composantes principales et les variables sont alors :



	Composante 1	Composante 2
$x_6$	0.99	0.01
Résidu de $x_5$	0.94	0.33
$x_4$	0.94	0.33

TABLEAU 2. Corrélation avec les composantes principales

L'essentiel des relations entre les trois variables stationnaires est donc décrit par la première composante principale.

## 5.2. LES DONNÉES DE NELSON ET PLOSSER

La précédente méthode est maintenant appliquée (via le logiciel EViews) sur les données originales de Nelson et Plosser [1982]. Ces données concernent les USA pour la période 1909-1970, la fréquence est annuelle. Les séries présentant une tendance déterministe ont été mises de côté, conformément à la méthode appliquée. Nous avons de ce fait au total 13 variables : taux de chômage, emploi, vitesse de circulation de la monnaie, PNB nominal, PNB réel, PNB réel par tête, production industrielle, stock de monnaie (mesuré par M2), indice du prix du PNB, salaire nominal, salaire réel, l'indice SP500, indice des prix. Ces 13 séries sont centrées.

Dans ce paragraphe, nous utilisons le test de Phillips et Perron [1988] qui est mieux adapté à des variables macroéconomiques que le test de Dickey et Fuller et permet de tenir compte d'une éventuelle autocorrélation et/ou hétéroscédasticité des résidus ; les résultats de ce test sont donc plus robustes. Pour ce test, il est nécessaire de choisir un paramètre de troncature. En faisant varier ce paramètre, les résultats restent stables, et de ce fait nous avons choisi de retenir la valeur  $T^{1/4}$ , T étant le nombre d'observations.

**1<sup>ère</sup> étape.** Un test de Phillips-Perron est mis en place sur chacune d'entre elles. Ce test montre que toutes les séries sont non stationnaires. Afin de déterminer l'ordre d'intégration des séries initiales, le test de racine unitaire est également appliqué sur les séries en différence première et conclut à la stationnarité de celles-ci. De ce fait, nous pouvons affirmer que l'ensemble des variables est  $I(1)$ .

On effectue l'ACP non normée de cet ensemble de données. Pour ce faire, nous déterminons la matrice de variance-covariance associée aux données. Les 13 valeurs propres de cette matrice figurent dans le tableau suivant.

Le vecteur propre associé à la plus grande valeur propre est isolé. On teste la stationnarité de la première composante principale par le test de Phillips Perron. La valeur calculée de la statistique est de 2,03 ce qui nous permet de conclure en la non stationnarité de cette composante (la valeur critique au seuil de 5 % est de -1,95). De ce fait, nous régressons à présent nos séries initiales sur la première composante principale et nous nous intéressons à la stationnarité des résidus de ces régressions.

1	0,001
2	0,193
3	1,024
4	1,444
5	2,752
6	10,627
7	23,654
8	85,011
9	152,971
10	4 614,575
11	51 301,970
12	25 421 131,216
13	62 812 041 219,752

TABLEAU 3. Valeurs propres

**2<sup>nd</sup>e étape.** Un test de Phillips Perron est appliqué sur ces 13 résidus (nommés res1, res2, ..., res13). Les résultats figurent dans le tableau suivant :

Variable	Stat PP
Res1	-1.12
Res2	-1.51
Res3	-1.58
Res4	-1.82
Res5	-1.52
Res6	-2.27*
Res7	-1.89
Res8	-2.05*
Res9	-1.29
Res10	-0.927
Res11	-1.53
Res12	-2.15*
Res13	-1.75

TABLEAU 4. Test de Phillips Perron

Stat PP est la valeur de la statistique de Phillips Perron, \* signifie que l'hypothèse nulle de non stationnarité est rejetée au seuil de 5 %.

Nous remarquons que les résidus 6, 8 et 12 sont stationnaires, ces derniers sont mis de côté et on effectue à présent l'ACP des résidus non stationnaires.

**3<sup>ème</sup> étape.** l'ACP est effectuée sur les 10 résidus non stationnaires. Les valeurs propres associées figurent dans le tableau suivant :

1	1.46 <sup>E</sup> -8
2	0.38
3	1.58
4	6.07
5	19.90
6	81.41
7	132.25
8	4605.84
9	51275.86
10	25421117

TABLEAU 5. Valeurs propres

Nous étudions la stationnarité de la première composante principale, la statistique de Phillips et Perron est égale à -1,52. Nous pouvons donc en conclure que cette composante principale est non stationnaire au seuil de 5 %.

Les 10 précédents résidus sont régressés sur cette composante principale et nous étudions la stationnarité des résidus (nommés res14, ..., res23) de ces régressions. Le tableau 6 reprend l'ensemble des résultats.

Variable	Stat pp
Res14	-2.08*
Res15	-1.88
Res16	-3.03*
Res17	-3.25*
Res18	-1.55
Res19	-2.29*
Res20	-1.85
Res21	-1.77
Res22	-3.20*
Res23	-1.65

TABLEAU 6. Test de Phillips Perron

Les résidus stationnaires sont mis de côté. Nous effectuons l'ACP sur les résidus non stationnaires.

**4<sup>ème</sup> étape.** L'ACP est mise en place sur les 5 précédents résidus non stationnaires. Les valeurs propres figurent dans le tableau suivant :

1	9.89 <sup>E</sup> -6
2	2.52
3	8.64
4	42.30
5	49709.35

TABLEAU 7. Valeurs propres

Nous testons la stationnarité de la première composante principale. La statistique de Phillips Perron est égale à -1,85 ce qui signifie que la première composante principale est non stationnaire. Nous régressons de ce fait les 5 précédents résidus sur cette dernière et testons la stationnarité des résidus (nommés res24, ..., res28) de ces régressions. Les résultats figurent dans le Tableau 8.

Variable	Stat PP
Res24	-2.60*
Res25	-2.64*
Res26	-2.33*
Res27	-2.07*
Res28	-2.36*

TABLEAU 8. Test de Phillips-Perron

Nous remarquons que ces résidus sont tous stationnaires.

Ayant obtenu pour chacune des variables des résidus stationnaires, on calcule la variance de ces résidus :

Variable	variance
Taux de chômage	29,61
Indice des prix	56,66
Vitesse de circulation monnaie	0,075
PNB réel	139,26
PNB réel par tête	6184,59
Indice du prix du PNB	6,00
Production industrielle	46,47
Stock de monnaie	88,40
PNB nominal	0,00
Emploi	9,20
Salaire nominal	0,01
Salaire réel	3,53
Indice SP500	41,58

TABLEAU 9. Variance des résidus

puis les valeurs propres de l'ACP :

1	5.79 <sup>E-15</sup>
2	5 <sup>E-9</sup>
3	0.003
4	0.18
5	0.60
6	1.49
7	2.43
8	8.27
9	22.85
10	51.48
11	85.97
12	154.97
13	6170.61

TABLEAU 10. Valeurs propres

Et enfin les corrélations entre les résidus et les composantes principales et les corrélations entre les résidus. On notera  $C_i$  la composante n°  $i$ .

	C 1	C 2	C 3	C 4	C 5	C 6
Taux de chômage	.48	.69	.33	-.15	.31	-.11
Indice des prix	.54	-.49	-.14	.63	.04	.23
Vitesse de circulation monnaie	-.44	-.27	-.58	-.17	-.03	-.57
PNB réel	-.50	.81	.12	.26	.09	.03
PNB réel par tête	-.99	-.01	.00	.00	.00	.00
Indice du prix du PNB	-.58	.56	-.08	-.09	.05	.06
Production industrielle	.56	-.48	-.08	.58	.10	-.31
Stock de monnaie	.25	-.22	.93	.09	-.15	.00
PNB nominal	.87	-.05	-.02	-.42	-.07	.16
Emploi	.88	.01	.05	-.39	-.09	.16
Salaire nominal	.80	-.33	.14	-.40	.13	.16
Salaire réel	-.21	.88	.14	.07	.13	-.01
Indice SP500	-.18	.64	-.33	.21	-.63	-.07

TABLEAU 11. Corrélations entre composantes principales et résidus des variables

Les variances des résidus (cf. Tableau 9) sont très différentes, d'où des valeurs propres très différentes (cf. Tableau 10).

Les étapes successives de l'ACP décrivent des phénomènes linéairement indépendants pour la période considérée entre les résidus des variables :

- le premier axe est très fortement lié au PNB réel par tête, ce qui n'est pas étonnant car cette variable a une variance très forte ; sur cet axe, le PNB réel s'oppose au PNB

nominal, à l'emploi et au salaire nominal. Cette opposition est confirmée par l'examen des corrélations entre les résidus (cf. Tableau 12) ;

- le deuxième axe met l'accent sur la relation entre salaire réel et PNB réel ; cet axe est aussi lié positivement, dans une moindre mesure, au taux de chômage et à l'indice SP 500 ;
- le troisième axe oppose stock de monnaie et vitesse de circulation de la monnaie, mais la corrélation entre ces 2 variables n'est pas très élevée.

À partir de l'axe 4, les corrélations entre les variables et les axes deviennent trop faibles pour que l'on puisse mettre en évidence des relations entre les variables.

	Taux de chômage	Indice des prix	Vitesse cir. monnaie	PNB réel	PNB r. / tête	Indice du prix du PNB	Production industrielle	St. monnaie	PNB nominal	Emploi	Salaire nominal	Salaire réel	Indice SP 500
Taux de chômage	1.00												
Indice des prix	-.22	1.00											
Vitesse de circulation monnaie	-.52	-.26	1.00										
PNB réel	.34	-.51	-.13	1.00									
PNB réel par tête	-.49	-.53	.44	.49	1.00								
Indice du prix du PNB	.07	-.63	.14	.72	.57	1.00							
Production industrielle	-.11	.85	.00	-.52	-.55	-.64	1.00						
Stock de monnaie	.22	.16	-.60	-.18	-.25	-.36	.21	1.00					
PNB nominal	.40	.27	-.36	-.59	-.87	-.45	.22	.19	1.00				
Emploi	.47	.25	-.43	-.53	-.88	-.42	.20	.25	.99	1.00			
Salaire nominal	.29	.36	-.35	-.73	-.79	-.57	.33	.35	.92	.91	1.00		
Salaire réel	.57	-.52	-.27	.86	.20	.61	-.50	-.13	-.31	-.24	-.50	1.00	
Indice SP500	.04	-.28	.12	.57	.18	.43	-.31	-.38	-.24	-.21	-.58	.48	1.00

TABLEAU 12. Corrélations entre les résidus

## 6. CONCLUSION

Pour déterminer des relations entre des séries non stationnaires, très courantes en macro-économie, les économistes procèdent souvent à des transformations de ces séries pour les rendre stationnaires, notamment par différenciation.

Une autre manière de procéder est de rechercher des relations de cointégration, c'est-à-dire des combinaisons linéaires stationnaires de variable non stationnaires. La méthode proposée ici est l'application à l'ACP de cette recherche de relations de cointégration, les séries étant filtrées par leurs premières composantes principales de telle manière que les séries résiduelles sont alors stationnaires. Cette stationnarité est obtenue à partir des séries brutes, donc sans qu'il soit nécessaire de transformer les séries préalablement et fournit donc des informations sur les relations entre ces séries, considérées en niveau absolu. L'interprétation des résultats se fait alors comme classiquement en ACP.

#### BIBLIOGRAPHIE

- BOUROCHE J.-M., SAPORTA G. (1980), *L'analyse des données*, Paris, Presses Universitaires de France, Coll. Que sais-je ?
- CASIN PH. (1999), *Analyse des données et des panels de données*, Bruxelles et Paris, De Boeck.
- CASIN PH. (2009), *Économétrie, méthodes et applications avec EViews*, Paris, Éditions Technip.
- CHIGIRA H. (2006), *A test of cointegration rank based on principal component analysis*, Tokyo, Hitotsubashi University, Research Unit for Statistical Analysis in Social Sciences.
- DICKEY D.A., FULLER W.A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association* 74, p. 427-431.
- DICKEY D.A. FULLER W.A. (1981), "Likelihood ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica* 49, p. 1057-1072.
- ENGLE R., GRANGER C.W.J. (1987), "Cointegration and error correction: representation, estimation and testing", *Econometrica* 55, p. 251-276.
- GOURIEROUX CH., MONTFORT A. (1995), « Séries temporelles et modèles dynamiques », 2<sup>e</sup> éd., *Economica*.
- HARRIS Y. (1997), "Principal components analysis of cointegrated time series", *Econometric Theory* 13, p. 529-557.
- JOHANSEN S. (1988), "Statistical analysis of cointegration vectors", *Journal of economic dynamics and control* 12, p. 231-254.
- KWIATKOWSKI D., PHILLIPS P.C.B., SCHMIDT P., SHIN Y. (1992), "Testing the null hypothesis of stationarity against the alternative of a unit root", *Journal of Econometrics* 54, p. 159-178.
- MALINVAUD E. (1978), *Méthodes statistiques de l'économétrie*, 3<sup>e</sup> éd., Paris, Dunod.
- LARDIC S., MIGNON V. (2002), *Économétrie des séries temporelles macroéconomiques et financières*, Paris, Economica.
- NELSON C.R., PLOSSER C.I. (1982), "Trends and random walks in macroeconomic time series: some evidence and implications", *Journal of Monetary Economics* 10, p. 129-162.
- PHILLIPS P.C.B, PERRON P. (1988), "Testing for a unit root in time series regression", *Biometrika* 75, p. 335-346.
- SNELL A. (1999), "Testing for r versus r-1 cointegrating vectors", *Journal of Econometrics* 88, p. 151-191.