

Direction de la Diffusion et de l'Action régionale

H 2018/01

Structurel, résiduel, géographique :

**Principe et mise en œuvre des approches comptable et
économétrique**

Claire Kubrak

Document de travail



Institut National de la Statistique et des Études Économiques

Institut National de la Statistique et des Études Économiques

*Série des documents de travail
de la Direction de la Diffusion et de l'Action Régionale*

Structurel, résiduel, géographique :

Principe et mise en œuvre des approches comptable et
économétrique

Claire Kubrak

Insee, PSAR Études économiques régionales

Janvier 2018

*Ces documents de travail ne reflètent pas la position de l'INSEE et n'engagent que
leurs auteurs.*

Working papers do not reflect the position of INSEE but only their author's views.

Résumé

Ce document présente des méthodes permettant d'analyser les différences d'évolution entre territoires. Ces différences peuvent tenir à des structures sectorielles qu'il s'agira d'isoler pour faire apparaître des effets pouvant être attribués à une plus ou moins grande compétitivité locale, à des capacités d'innovation différentes, par exemple.

Des méthodes d'analyse très utilisées se sont développées à partir de la modélisation « shift-share », qui permet une analyse que l'on qualifiera ici de structurelle-résiduelle. Des approches fondées sur l'analyse de la variance ont été développées par la suite, qui visent à faire intervenir de façon plus explicite la dimension territoriale. On regroupe souvent ces modèles sous le terme de structurels géographiques, malgré quelques réserves sur l'utilisation du terme géographique.

La première partie du document présente les deux approches : décomposition « comptable » des évolutions et analyse économétrique. Dans une deuxième partie, les deux approches sont mises en oeuvre sur un même exemple : l'analyse de la croissance de l'emploi entre 2000 et 2014 dans les régions françaises. Les résultats sont détaillés et commentés avec les principaux éléments d'interprétation propres à chaque méthode.

Mots clés : modèle, structurel, comptable, analyse de la variance, autocorrélation spatiale

Abstract

This document presents methods for analysing the differences in developments between territories. These differences may be due to sectoral structures which will need to be isolated to show effects that can be attributed to, for example, a greater or lesser degree of local competitiveness, different capacities for innovation.

Frequently used analytical methods have developed from the "shift-share" modeling, which allows an analysis that can be described here as structural-residual. Approaches based on variance analysis were subsequently developed, which aim at more explicitly involving the territorial dimension. These models are often grouped under the term geographic structures, despite some reservations about the use of the term geographic..

The first part of the document presents the two approaches: accounting decomposition and econometric analysis. In a second part, the two approaches are implemented on the same example: the analysis of employment growth between 2000 and 2014 in French regions. The results are detailed and commented on with the main elements of interpretation specific to each method.

Keywords: model, shift-share, accounting, variance analysis, spatial autocorrelation

Table des matières

INTRODUCTION.....	7
PARTIE A : DEUX MÉTHODES D'ANALYSE.....	9
I. L'approche « descriptive » ou « comptable ».....	9
I.1 Le modèle classique.....	9
I.2 Variante Esteban-Marquillas.....	10
I.3 Variantes spatialisées des modèles classique et Esteban-Marquillas.....	11
II. L'approche économétrique.....	14
II.1. Le modèle structurel-résiduel.....	14
II.2 Le modèle structurel-géographique.....	15
II.3 Estimation du modèle structurel-géographique.....	16
II.4. Raisonner « en écart à la moyenne ».....	17
PARTIE B : MISE EN ŒUVRE PRATIQUE.....	20
I. Remarques générales sur l'interprétation des résultats.....	20
II. Mise en œuvre de la méthode « descriptive ».....	21
II.1. Décomposition structurelle-résiduelle.....	21
II.2 Des résultats détaillés selon la structure.....	22
II.3 Effets croisés, effets séparés.....	24
III. Mise en œuvre de la méthode économétrique.....	26
III.1 Des effets de structure significatifs ?.....	26
III.2 Résultats de l'estimation du modèle.....	27
III.3 Complément : exemple d'introduction d'un effet croisé.....	33
CONCLUSION.....	35
BIBLIOGRAPHIE.....	36

Introduction

Si l'on s'intéresse à la croissance de l'emploi dans une région sur quelques années, le résultat donne un premier niveau d'information : création ou destruction de l'emploi, ampleur de cette croissance ou de cette décroissance. La comparaison avec la croissance de l'emploi des autres régions, ou avec la croissance nationale permet ensuite de diagnostiquer si la région a connu une croissance plus ou moins forte que l'ensemble du territoire, ou que d'autres régions. De fait, la comparaison des indicateurs entre zones s'avère souvent pertinente et nécessaire, pour disposer d'un positionnement relatif du territoire étudié.

Cependant, des différences structurelles entre les zones peuvent expliquer les écarts constatés. Typiquement, en termes de croissance économique, les régions n'ont pas les mêmes spécialisations en termes d'activité, le même tissu productif. Si la région est fortement industrielle, les difficultés économiques de ce secteur pénalisent la croissance de l'emploi régional. Dans quelle mesure « le fait qu'elle soit plus industrielle » explique les écarts de croissance de cette région par rapport aux autres ?

L'analyse *shift-share*, datant de la fin des années 1950 a constitué la première réponse à cette question. Elle distinguait un effet lié à l'économie nationale (national share), un effet lié à la structure sectorielle (proportional shift) et un facteur lié aux différences locales de productivité (« differential shift »). Cette décomposition a été qualifiée de structurelle-résiduelle. L'effet structurel est mesuré en appliquant la tendance nationale à la structure sectorielle de la région. Il correspond à la croissance attendue pour la région si chaque secteur connaissait la même croissance qu'au niveau national. Il faut bien voir que le « structurel » de cette appellation renvoie à une structure sectorielle. (On pourrait aussi considérer le découpage du territoire comme une structure). L'effet résiduel correspond à la part de la croissance régionale non expliquée par la structure sectorielle. Dans la littérature francophone, un pas supplémentaire a souvent été franchi en qualifiant cet effet résiduel de géographique, parce qu'il serait associé à un territoire.

Les méthodes shift-share ont fait l'objet de nombreux développements dans la littérature anglo-saxonne notamment. La critique de ces modèles (Berzeg 1978 ; Knudsen 1991) a conduit à explorer d'autres modélisations, de nature économétrique. Dans un premier temps, on a construit des modèles d'analyse de la variance à un facteur explicatif (la répartition sectorielle), équivalant au modèle comptable, le résidu de la régression fournissant la composante résiduelle du modèle. Dans un second temps, les territoires d'intérêt sont introduits dans un modèle d'analyse à deux facteurs : le facteur sectoriel d'une part, le

facteur territorial d'autre part. Ce modèle permet de tester les effets fixes liés à chaque territoire. Le résidu de la régression ne correspond plus dans ce cas à la composante résiduelle des modèles comptables. On peut dans ce cas parler de « structurel-géographique », en ne perdant pas de vue que la dimension géographique se réduit à une partition du territoire en zones d'étude. La géographie au sens fort renvoie à des proximités et des distances, ce qui n'est pas le cas dans le modèle. On ne conservera la formulation « structurelle-géographique », que dans le cas de l'analyse de la variance à deux facteurs.

La première partie de ce document est consacrée à la présentation des deux familles de méthodes :

- méthodes basées sur des décompositions comptables à travers le modèle de base présenté par Dunn(1960), et deux variantes basées sur des travaux d'Esteban-Marquillas (1972) et de Mayor et Lopes(2005) ;

- méthodes économétriques à un ou deux facteurs explicatifs permettant de tester la significativité des effets.

Dans la deuxième partie du document, les deux approches sont mises en œuvre sur un même exemple : l'analyse de la croissance de l'emploi entre 2000 et 2014 dans les régions françaises. Les résultats sont détaillés et commentés avec les principaux éléments d'interprétation propres à chaque méthode.

Les méthodes présentées et leurs applications pratiques sont celles utilisées dans les travaux du PSAR Études économiques régionales de l'Insee. À ce titre, ce document est le produit d'un travail collectif, initié en 2004 par Benoît Mulkay puis par Camille Rousseau. Toutefois, certaines des interprétations sont personnelles et les erreurs qui apparaîtraient ne sont imputables qu'à l'auteur du document.

PARTIE A : DEUX MÉTHODES D'ANALYSE

I. L'approche « descriptive » ou « comptable »

I.1 Le modèle classique

a) Présentation

On dispose d'une variable d'intérêt, quantitative, qui est un taux d'évolution ou un ratio (X) et d'une variable de structure, qualitative, qui permet de répartir la population en classes. Les données de base sont les valeurs X_{ri} , soit le taux observé dans la zone r pour le groupe de caractéristique i . Le dénominateur du taux ou du ratio est l'effectif de la sous-population : Z_{ri} .

On peut calculer :

- le taux moyen observé dans la zone r : c'est la moyenne pondérée des taux de l'ensemble des catégories observées dans cette zone

$$X_{r.} = \sum_i \left(\frac{Z_{ri}}{Z_{r.}} X_{ri} \right), \text{ avec } Z_{r.} : \text{effectif de la zone } r$$

- le taux moyen pour la catégorie i : c'est la moyenne pondérée des taux observés dans l'ensemble des régions pour la catégorie i

$$X_{.i} = \sum_r \left(\frac{Z_{ri}}{Z_{.i}} X_{ri} \right), \text{ avec } Z_{.i} : \text{effectif de la catégorie } i$$

- le taux moyen sur l'ensemble du territoire :

$$X_{..} = \sum_{r,i} \left(\frac{Z_{ri}}{Z_{..}} X_{ri} \right), \text{ avec } Z_{..} : \text{effectif total de la variable de pondération.}$$

L'écart entre la valeur observée dans une zone r et la valeur observée pour l'ensemble du territoire se décompose en deux effets :

$$X_{r.} - X_{..} = \sum_i \frac{Z_{ri}}{Z_{r.}} (X_{ri} - X_{..}) = \sum_i \frac{Z_{ri}}{Z_{r.}} (X_{.i} - X_{..} + X_{ri} - X_{.i})$$

$$X_{r.} - X_{..} = \underbrace{\sum_i \frac{Z_{ri}}{Z_{r.}} (X_{.i} - X_{..})}_{\text{EFFET STRUCTUREL}} + \underbrace{\sum_i \frac{Z_{ri}}{Z_{r.}} (X_{ri} - X_{.i})}_{\text{EFFET RESIDUEL}} \quad (1)$$

L'effet structurel est la moyenne des écarts entre les taux observés pour chaque catégorie i sur l'ensemble du territoire et le taux moyen global, pondérée par le poids de chaque catégorie i dans la région r .

L'effet résiduel est une moyenne pondérée des écarts entre la valeur observée dans la région r et celle observée sur l'ensemble du territoire pour chacune des catégories i . Il correspond à l'écart à la moyenne qui reste une fois qu'on a pris en compte l'effet de la variable de structure.

b) Mise en œuvre

Cette analyse peut être généralisée au cas de plusieurs facteurs structurels, en considérant chaque croisement possible entre les différentes variables comme une modalité particulière d'une nouvelle variable synthétique.

c) Limites de la méthode

- [critique de Rosenfeld (1959)] : L'effet résiduel ne dépend pas uniquement du dynamisme spécifique de la zone r pour la caractéristique i . Il dépend également de la spécialisation de la zone dans cette caractéristique i , par le facteur $Z_{ri} / Z_{r.}$. Ainsi, une partie de l'effet structurel est affectée à l'effet résiduel. Une variante, proposée par Esteban-Marquillas, est décrite au § 1.2.
- Le découpage territorial ne prend pas en compte les interactions entre zones, ce qui a conduit certains auteurs à proposer des variantes spatialisées du modèle : Nazara and Hewings (2004), Mayor and Lopez (2005).
- L'approche descriptive ne permet pas de tester la significativité des effets structurels et résiduels. D'où le recours à des approches économétriques utilisant les modèles d'analyse de la variance (voir § II).

1.2 Variante Esteban-Marquillas

Pour répondre à la critique de Rosenfeld, Esteban-Marquillas (EM) propose en 1972 une variante du modèle faisant intervenir une nouvelle composante, dite homothétique.

On définit $Z'_{ri} = Z_{r.} \frac{Z_{.i}}{Z_{..}}$

Z'_{ri} correspond à l'effectif qu'on aurait pour la caractéristique i dans la zone r si la structure dans la zone r était la même qu'au niveau national.

L'égalité (1) s'écrit alors :

$$X_{r.} - X_{..} = \sum_i \frac{Z_{ri}}{Z_{r.}} (X_{.i} - X_{..}) + \sum_i \frac{Z'_{ri}}{Z_{r.}} (X_{ri} - X_{.i}) + \sum_i \left(\frac{Z_{ri}}{Z_{r.}} - \frac{Z'_{ri}}{Z_{r.}} \right) (X_{ri} - X_{.i}) \text{ ou encore}$$

$$X_{r.} - X_{..} = \sum_i \frac{Z_{ri}}{Z_{r.}} (X_{.i} - X_{..}) + \sum_i \frac{Z_{.i}}{Z_{..}} (X_{ri} - X_{.i}) + \sum_i \left(\frac{Z_{ri}}{Z_{r.}} - \frac{Z_{.i}}{Z_{..}} \right) (X_{ri} - X_{.i}) .$$

L'effet résiduel du modèle de Dunn(1960) se décompose en un **effet de compétitivité** qui correspond au terme $\sum_i \frac{Z_{.i}}{Z_{..}} (X_{ri} - X_{.i})$ indépendant de la structure de la zone r.

C'est une moyenne, pondérée par la structure nationale, des écarts entre la valeur observée dans la région r et celle observée sur l'ensemble du territoire pour chacune des catégories i.

Le terme additionnel $\sum_i \left(\frac{Z_{ri}}{Z_{r.}} - \frac{Z_{.i}}{Z_{..}} \right) (X_{ri} - X_{.i})$ est dit **effet d'allocation**.

L'effet d'allocation sera positif si la zone r est spécialisée dans les secteurs i pour lesquels elle présente des avantages comparatifs par rapport au niveau national :

$$\frac{Z_{ri}}{Z_{r.}} > \frac{Z_{.i}}{Z_{..}} \text{ et } X_{ri} > X_{.i}$$

Il sera également positif si les secteurs i pour lesquels la zone est peu compétitive sont sous-représentés :

$$\frac{Z_{ri}}{Z_{r.}} < \frac{Z_{.i}}{Z_{..}} \text{ et } X_{ri} < X_{.i}$$

Au contraire, l'effet d'allocation sera négatif si la région est spécialisée dans des secteurs pour lesquels elle n'est pas compétitive ($X_{ri} < X_{.i}$).

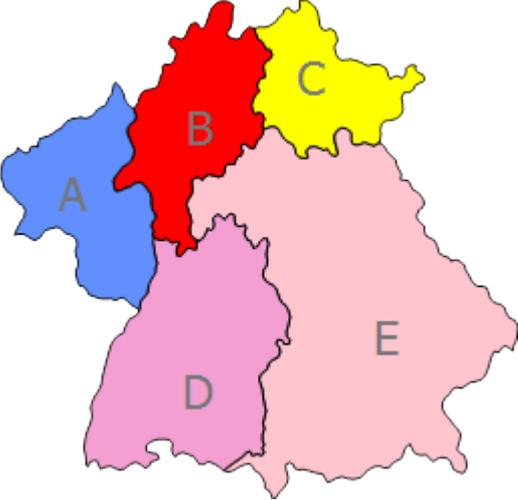
Dans le cas limite où la structure de la région serait la même qu'au niveau national, l'effet d'allocation serait nul.

I.3 Variantes spatialisées des modèles classique et Esteban-Marquillas

Les approches présentées en I.1 et I.2 ne sont pas complètement spatiales. Elles font certes intervenir un découpage, mais une permutation des zones ne changerait rien. C'est pourquoi différents auteurs ont proposé dans les années récentes des indicateurs prenant en compte les relations de proximité entre territoires.

Comme pour le calcul des indicateurs d'autocorrélation spatiale, la proximité entre territoires est appréhendée à l'aide d'une matrice de contiguïté, qui, dans sa variante la plus simple,

vaut 1 si les territoires ont une frontière, 0 sinon, une zone n'étant pas contiguë à elle-même par définition.

	<table border="1" data-bbox="906 230 1350 685"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>0</td> <td>1</td> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <th>B</th> <td>1</td> <td>0</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <th>C</th> <td>0</td> <td>1</td> <td>0</td> <td>0</td> <td>1</td> </tr> <tr> <th>D</th> <td>1</td> <td>1</td> <td>0</td> <td>0</td> <td>1</td> </tr> <tr> <th>E</th> <td>0</td> <td>1</td> <td>1</td> <td>1</td> <td>0</td> </tr> </tbody> </table>		A	B	C	D	E	A	0	1	0	1	0	B	1	0	1	1	1	C	0	1	0	0	1	D	1	1	0	0	1	E	0	1	1	1	0
	A	B	C	D	E																																
A	0	1	0	1	0																																
B	1	0	1	1	1																																
C	0	1	0	0	1																																
D	1	1	0	0	1																																
E	0	1	1	1	0																																
Territoires	Matrice des poids																																				

D'une certaine manière, la variante spatialisée du découpage structurel-résiduel n'est pas sans rappeler le modèle d'Esteban-Marquillas(1972). On part du modèle classique, puis on ajoute et retranche un terme. Dans Esteban-Marquillas, on faisait intervenir une composante qualifiée d'homothétique. Dans le modèle spatialisé, on va faire intervenir une composante de voisinage. $Z_{r,i}^v$ désigne l'effectif calculé dans le voisinage de i , w_{ij} désignant la matrice des poids.

$$Z_{r,i}^v = \sum_{k \in V} w_{ik} Z_{rk}$$

La variante spatiale la plus simple est alors :

$$X_{r.} - X_{..} = \sum_i \frac{Z_{ri}}{Z_{r.}} (X_{.i} - X_{..}) + \sum_i \frac{Z_{ri}^v}{Z_{r.}} (X_{ri} - X_{.i}) + \sum_i \left(\frac{Z_{ri}}{Z_{r.}} - \frac{Z_{ri}^v}{Z_{r.}} \right) (X_{ri} - X_{.i})$$

La difficulté de cette formulation vient de ce que : $\sum_{r,i} Z_{ri}^v \neq \sum_{r,i} Z_{ri}$

Une des façons de résoudre cette question (Mayor et Lopes 2005) est de reconstruire une composante homothétique à la Esteban-Marquillas, en prenant en compte les effets spatiaux. Elle peut s'écrire :

$$Z_{r,i}^{vw} = Z_{r.} \frac{Z_{.i}^v}{Z_{..}}$$

L'indicateur spatialisé est alors :

$$X_{r.} - X_{..} = \sum_i \frac{Z_{ri}}{Z_{r.}} (X_{.i} - X_{..}) + \sum_i \frac{Z_{ri}^{vv}}{Z_{r.}} (X_{ri} - X_{.i}) + \sum_i \left(\frac{Z_{ri}}{Z_{r.}} - \frac{Z_{ri}^{vv}}{Z_{r.}} \right) (X_{ri} - X_{.i})$$

Le deuxième terme est parfois qualifié d'effet compétitif spatial net, le troisième d'effet d'allocation spatialisé.

Une présentation plus détaillée de cette méthode, et d'une autre méthode reposant sur le filtrage spatial des données peut être trouvée dans Mayor et Lopez (2009).

On en donne une illustration avec l'évolution de l'emploi salarié dans une nomenclature en 6 postes. Le calcul a été effectué pour l'ensemble des zones d'emploi de France métropolitaine, mais publié sur les plus importantes dans le tableau ci-dessous.

Les évolutions sont en pourcentage, entre 2008 et 2011. L'exemple est différent de celui de la partie B de ce document. Le maillage en zone d'emploi se prête mieux à la prise en compte de la proximité spatiale que les régions, qui sont trop vastes.

On trouvera ci-dessous les résultats obtenus avec les différentes méthodes :

- modèle de Dunn(1960),
- modèle d'Esteban-Marquillas dans ses versions spatiales et non spatiales.

Par construction la somme des effets d'allocation et de compétitivité des modèles d'Esteban-Marquillas sont égaux au résidu du modèle de Dunn. Sur cet exemple, les effets d'allocation sont assez faibles relativement aux effets de compétitivité.

Zone d'emploi	Modèle Dunn				Modèle Esteban-Marquillas			
	Territoire	National	Structure	Residu	Non spatial		Spatial	
					Compétitivité	Allocation	Compétitivité	Allocation
Roissy - Sud Picardie	14,5	10,5	2,1	1,9	2,1	-0,2	2,4	-0,6
Saint-Étienne	2,2	10,5	-3,0	-5,4	-5,0	-0,4	-5,2	-0,2
Toulouse	29,6	10,5	1,2	17,9	18,5	-0,6	16,1	1,8
Paris	8,4	10,5	4,2	-6,3	-7,3	1,0	-7,6	1,3
Saclay	13,8	10,5	1,8	1,5	2,6	-1,2	1,0	0,5
Orly	13,2	10,5	5,0	-2,2	1,8	-4,1	-0,7	-1,5
Rouen	8,0	10,5	-0,7	-1,7	-1,9	0,1	-1,8	0,1
Tours	11,3	10,5	0,0	0,8	1,0	-0,2	0,5	0,3
Orléans	7,4	10,5	0,1	-3,2	-2,9	-0,2	-2,2	-0,9
Lille	15,6	10,5	1,9	3,2	2,8	0,4	1,1	2,2
Nancy	4,9	10,5	1,0	-6,6	-5,8	-0,9	-6,2	-0,4
Metz	3,2	10,5	0,4	-7,7	-6,8	-0,9	-6,9	-0,7
Strasbourg	4,8	10,5	1,2	-7,0	-6,3	-0,7	-8,5	1,5
Nantes	26,1	10,5	0,3	15,2	15,2	0,0	14,5	0,7
Rennes	22,0	10,5	0,3	11,2	11,9	-0,7	11,3	-0,2
Bordeaux	18,8	10,5	0,9	7,5	7,9	-0,4	7,3	0,1
Grenoble	12,7	10,5	-1,5	3,7	3,8	-0,1	3,9	-0,1
Lyon	16,4	10,5	0,2	5,8	6,1	-0,3	5,6	0,2
Clermont-Ferrand	7,0	10,5	-0,4	-3,1	-3,0	-0,1	-2,9	-0,2
Montpellier	31,5	10,5	3,1	17,9	18,7	-0,8	18,1	-0,2
Nice	11,9	10,5	3,5	-2,1	-1,1	-1,0	-1,7	-0,4
Marseille - Aubagne	15,8	10,5	2,5	2,8	3,6	-0,8	2,8	0,1
Toulon	20,8	10,5	2,7	7,6	9,8	-2,2	8,5	-0,8

Il a été réalisé avec R, l'écriture étant simple. Il existe aussi un package [REAT: Regional Economic Analysis Toolbox](#) permettant de réaliser des analyses.

II. L'approche économétrique

L'approche classique ou comptable est d'une grande simplicité et, à partir du moment où l'on dispose des données par zone et variable structurelle, peut être mise en œuvre avec une grande économie de moyens. Elle ne permet cependant pas de tester la pertinence des effets de structure pris en compte, ni de savoir si les effets structurels calculés sont significatifs.

L'approche économétrique vise à apporter des éléments de réponse sur ces points.

II.1. Le modèle structurel-résiduel

Le modèle structurel-résiduel s'écrit :

$$X_{r,i} = m + e_i + \varepsilon_{r,i}$$

$$\text{avec } \sum_i Z_{.i} e_i = 0$$

{ Avec les hypothèses suivantes : $E(\varepsilon_{r,i}) = 0$
 $Var(\varepsilon_{r,i}) = \sigma^2 / Z_{r,i}$

Ce modèle correspond à une analyse de la variance pondérée. On peut trouver des références aux méthodes d'analyse de la variance dans Confais et Leguen(2007), Dagnelie(1998) pour une présentation générale, Jayet(1993) pour une application aux modèles présentés dans ce document. Les travaux pionniers sont présentés dans Berzeg(1978) ou Knudsen et Barf(1991).

Dans cette modélisation, e_i s'interprète comme l'écart à la moyenne dû à la modalité i . L'estimateur \hat{m} de m est $X_{..}$, l'estimateur \hat{e}_i de e_i est $(X_{.i} - X_{..})$ et les résidus de l'estimation $\hat{\varepsilon}_{r,i}$ sont égaux à $(X_{r,i} - X_{.i})$. La contrainte $\sum_i Z_{.i} e_i = 0$ est nécessaire pour pouvoir estimer de manière unique les paramètres du modèle. Si on ne rajoutait pas cette information supplémentaire, il y aurait une colinéarité parfaite entre la constante et la somme des variables structurelles et le modèle ne serait pas estimable. La contrainte d'identification peut prendre plusieurs formes : pour le choix de la contrainte, voir la partie II.4. Raisonner « en écart à la moyenne ».

Les résultats obtenus pour les effets structurels et résiduels dans la zone r sont identiques à ceux de l'approche descriptive :

$$\text{Effet structurel} = \sum_i \left(\frac{Z_{ri}}{Z_{r.}} \right) \hat{e}_i$$

$$\text{Effet résiduel} = \sum_i \left(\frac{Z_{ri}}{Z_{r.}} \right) \hat{\epsilon}_{ri}$$

L'effet résiduel (relatif à un territoire r) est une moyenne pondérée des résidus calculés sur toutes les observations relatives à ce territoire (les $\hat{\epsilon}_{ri}$).

Les résultats identiques à ceux de l'approche descriptive en montrent les limites. Dans l'approche descriptive, il n'y a par définition aucune hypothèse sur la composante résiduelle. Dans l'analyse de variance à un facteur, les résidus $\hat{\epsilon}_{ri}$ sont d'espérance nulle. Par voie de conséquence, les effets résiduels du modèle, dont on souhaiterait qu'ils traduisent un effet propre à la région sont d'espérance nulle (Jayet 1993).

Une solution est d'introduire la variable de partition g_r (le zonage) comme facteur explicatif, pour essayer de mesurer explicitement l'effet géographique.

II.2 Le modèle structurel-géographique

a) Description

Lorsque l'on ajoute un facteur géographique, le modèle devient¹ :

$$X_{ri} = m + e_i + g_r + \mu_{ri} \quad \text{avec} \quad \sum_i Z_{.i} e_i = \sum_r Z_{r.} g_r = 0$$

On suivra ici Jayet (1993) ou Gagné et al.(2005) pour le qualifier de structurel-géographique, sous réserve des remarques formulées dans l'introduction.

C'est maintenant un modèle d'analyse de la variance à deux facteurs : une caractéristique structurelle et la zone géographique (remarque : on peut introduire d'autres variables structurelles). Une double contrainte est ici nécessaire ($\sum_i Z_{.i} e_i = \sum_r Z_{r.} g_r = 0$) pour que le modèle soit identifiable.

L'effet structurel est calculé comme la moyenne des paramètres e_i , pondérée par la structure de la zone étudiée. L'effet géographique est directement obtenu par l'estimation des paramètres g_r .

$$\text{Effet structurel} = \sum_i \left(\frac{Z_{ri}}{Z_{r.}} \right) \hat{\epsilon}_i$$

$$\text{Effet géographique} = \hat{g}_r$$

Grâce à la matrice de variance-covariance des paramètres estimés, on peut calculer un écart-type pour les effets structurels et géographiques et tester leur significativité. Dans ce modèle, les résidus s'interprètent de façon classique et traduisent les limites de la spécification (variables omises notamment).

b) Mise en œuvre

Le modèle est généralisable à plusieurs variables de structures. Dans le cas de deux variables structurelles, la formulation est :

¹ Cette spécification est proposée par Hubert JAYET [1993].

$$X_{r i j} = m + e1_i + e2_j + g_r + \mu_{r i j}$$

$$\text{avec } \sum_i Z_{.i.} e1_i = \sum_j Z_{..j} e2_j = \sum_r Z_{r..} g_r = 0$$

où $X_{r i j}$ est le taux observé sur la sous-population de la zone r qui prend respectivement les caractéristiques i et j pour les deux facteurs structurels.

- m est la constante, elle correspond à la valeur moyenne sur l'ensemble du territoire,
- $e1_i$ est l'effet de la i-ème modalité de la première variable structurelle,
- $e2_j$ est l'effet de la j-ème modalité de la deuxième variable structurelle,
- g_r est l'effet de la région r,
- $\mu_{r i j}$ est le terme d'erreur associé à l'estimation,
- $Z_{.i.}$ est le nombre d'individus appartenant à la catégorie i de la première variable structurelle,
- $Z_{..j}$ est l'effectif de la catégorie j de la deuxième variable structurelle,
- $Z_{r..}$ est l'effectif observé dans la zone r.

Ici les effets géographiques et structurels sont définis par :

$$\text{Effet géographique} = \hat{g}_r$$

$$\text{Effet structurel} = \sum_i \left(\frac{Z_{r i.}}{Z_{r..}} \right) \hat{e}1_i + \sum_j \left(\frac{Z_{r. j}}{Z_{r..}} \right) \hat{e}2_j$$

II.3 Estimation du modèle structurel-géographique

a) Estimation par les moindres carrés pondérés

Prenons le modèle avec un facteur structurel et un facteur géographique :

$$X_{r i} = m + e_i + g_r + \mu_{r i}$$

$$\text{avec } \sum_i Z_{.i} e_i = \sum_r Z_{r.} g_r = 0$$

Les hypothèses suivantes sont nécessaires pour réaliser les tests statistiques :

- les erreurs $\mu_{r i}$ sont des variables aléatoires d'espérance nulle et de variance constante,

$$E(\mu_{r i}) = 0$$

$$Var(\mu_{r i}) = \sigma^2 / Z_{r i}$$

- les erreurs $\mu_{r i}$ ne sont pas corrélées entre elles,
- les erreurs $\mu_{r i}$ sont normalement distribuées.

$$\mu_{r i} \approx N(0, \sigma^2)$$

Or en présence d'effets structurels et géographiques, l'hypothèse d'homoscédasticité n'est pas vérifiée ; la variance des erreurs n'est pas constante, elle dépend de la région r et de la catégorie i :

$$\text{Var}(\mu_{ri}) = \sigma^2_{ri}$$

On utilise dans ce cas une régression pondérée, en prenant comme pondération les effectifs associés à la région r et à la catégorie i (Z_{ri}). En pratique la régression pondérée, qui est un cas particulier de la méthode des moindres carrés généralisés (MCG), revient à faire l'hypothèse :

$$\mu_{ri} \approx N(0, \sigma^2_{ri})$$

$$\text{avec } \sigma^2_{ri} = \frac{\sigma^2}{Z_{ri}}$$

b) Procédure SAS utilisée

La mise en œuvre de ce modèle est effectuée avec le logiciel SAS. La procédure privilégiée pour l'estimation du modèle structurel-géographique est la PROC MIXED. Cette procédure est préférée à la PROC REG, car elle peut traiter les variables catégorielles. Elle est également préférée à la PROC GLM, car elle fournit dans une table SAS la matrice de variance-covariance des paramètres, qui permet d'effectuer les tests d'hypothèses souhaités.

Remarque : en organisant les données de façon adéquate, on peut utiliser facilement les fonctions de R.

II.4 Raisonner « en écart à la moyenne »

Pour estimer les paramètres d'un modèle d'analyse de la variance, la PROC MIXED (comme la PROC GLM) pose par défaut la contrainte identifiante suivante : pour chaque variable catégorielle, le coefficient de la dernière modalité est mis à zéro. Dans le cas d'un modèle d'analyse structurelle-géographique avec un effet structurel, cette contrainte s'écrit :

$$e_I = g_R = 0$$

avec I modalité de référence pour la variable structurelle et R région de référence. Avec cette contrainte, les différents paramètres s'interprètent par rapport à la modalité de référence.

On peut mesurer les effets géographiques et structurels **en écart à la moyenne, plutôt qu'en écart à une modalité de référence**. Pour cela, on transforme les coefficients fournis par la PROC MIXED. La transformation consiste à centrer les coefficients estimés. Dans le

cas d'un modèle avec un effet structurel et une variable géographique, cela revient à imposer la contrainte identifiante suivante :

$$\sum Z_{.i} e_i = \sum Z_{r.} g_r = 0$$

Nous présentons ci-dessous le calcul qui permet de passer des coefficients « en écart à la modalité de référence » aux coefficients « en écart à la moyenne ».

a) Calcul dans le cas du modèle à un facteur structurel :

Le modèle est le suivant :

$$\left\{ \begin{array}{l} X_{ri} = m + e_i + g_r + \mu_{ri} \\ \sum Z_{.i} e_i = \sum Z_{r.} g_r = 0 \end{array} \right.$$

où X_{ri} est le taux observé dans la zone r pour les individus de caractéristique i.

$Z_{.i}$ est l'effectif de la catégorie i.

$Z_{r.}$ est l'effectif de la zone r.

L'estimateur \hat{e}_i de e_i est fourni par défaut par la PROC MIXED². Il est égal à $(X_{.i} - X_{.I})$ et décrit l'effet de la modalité i par rapport à la modalité de référence I. On centre le coefficient \hat{e}_i , c'est-à-dire qu'on lui soustrait la moyenne pondérée des différents coefficients du même facteur :

$$\begin{aligned} \hat{e}_i - \sum_i \frac{Z_{.i}}{Z_{..}} \hat{e}_i &= (X_{.i} - X_{.I}) - \sum_i \frac{Z_{.i}}{Z_{..}} (X_{.i} - X_{.I}) \\ &= X_{.i} - X_{.I} - \sum_i \frac{Z_{.i}}{Z_{..}} X_{.i} + X_{.I} \sum_i \frac{Z_{.i}}{Z_{..}} \\ &= X_{.i} - X_{.I} - X_{..} + X_{.I} \\ &= X_{.i} - X_{..} \\ &= \hat{e}'_i \end{aligned}$$

On obtient le coefficient transformé \hat{e}'_i , qui correspond à l'effet de la modalité i par rapport à la moyenne.

b) Calcul dans le cas du modèle à deux facteurs structurels :

Le modèle s'écrit :

$$\begin{array}{l} X_{rij} = m + e1_i + e2_j + g_r + \mu_{rij} \\ \text{avec } \sum_i Z_{.i.} e1_i = \sum_j Z_{..j} e2_j = \sum_r Z_{r..} g_r = 0 \end{array}$$

² Comme par la PROC GLM.

Chaque paramètre $e1_i$, correspondant à l'effet de la i-ème modalité de la première variable de structure par rapport à la modalité de référence I , est estimé par $\hat{e}1_i = (X_{.i.} - X_{.I.})$. En soustrayant à $\hat{e}1_i$ sa moyenne, on obtient le paramètre transformé $\hat{e}'1_i$, qui est l'effet de la modalité i par rapport à la moyenne.

$$\begin{aligned} \hat{e}'1_i - \sum_i \frac{Z_{.i.}}{Z_{...}} \hat{e}1_i &= (X_{.i.} - X_{.I.}) - \sum_i \frac{Z_{.i.}}{Z_{...}} (X_{.i.} - X_{.I.}) \\ &= X_{.i.} - X_{.I.} - \sum_i \frac{Z_{.i.}}{Z_{...}} X_{.i.} + X_{.I.} \sum_i \frac{Z_{.i.}}{Z_{...}} \\ &= X_{.i.} - X_{.I.} - X_{...} + X_{.I.} \\ &= X_{.i.} - X_{...} \\ &= \hat{e}'1_i \end{aligned}$$

c) Calcul dans le cas du modèle avec interaction :

Le modèle avec une interaction entre deux variables structurelles s'écrit :

$$\begin{aligned} X_{rij} &= m + e1_i + e2_j + u_{ij} + g_r + \mu_{rij} \\ \sum_i Z_{.i.} e1_i &= \sum_j Z_{..j} e2_j = \sum_{ij} Z_{.ij} u_{ij} = \sum_r Z_{r..} g_r = 0 \end{aligned}$$

Le terme u_{ij} décrit l'effet combiné des caractéristiques i et j . Le coefficient \hat{u}_{ij} estimé par la PROC MIXED est égal à $(X_{.ij} - X_{.IJ})$, il s'interprète comme l'effet croisé des caractéristiques i et j par rapport à la modalité de référence $I \times J$. On soustrait à ce coefficient la moyenne des effets croisés :

$$\left\{ \begin{aligned} &\hat{u}_{ij} - \sum_{ij} \frac{Z_{.ij}}{Z_{...}} \hat{u}_{ij} \\ &= (X_{.ij} - X_{.IJ}) - \sum_{ij} \frac{Z_{.ij}}{Z_{...}} (X_{.ij} - X_{.IJ}) \\ &= X_{.ij} - X_{.IJ} - \sum_{ij} \frac{Z_{.ij}}{Z_{...}} X_{.ij} + X_{.IJ} \sum_{ij} \frac{Z_{.ij}}{Z_{...}} \\ &= X_{.ij} - X_{.IJ} - X_{...} + X_{.IJ} \\ &= X_{.ij} - X_{...} \\ &= \hat{u}'_{ij} \end{aligned} \right.$$

On obtient \hat{u}'_{ij} , l'effet conjoint des caractéristiques i et j par rapport à la moyenne.

PARTIE B : MISE EN ŒUVRE PRATIQUE

Dans cette partie, quelques-unes des méthodes présentées précédemment sont mises en œuvre sur un même jeu de données. L'objectif est double : il s'agit d'une part de proposer un exemple d'interprétation des différentes analyses en insistant sur les spécificités de chaque méthode et d'autre part de comparer les résultats produits par les différentes méthodes.

L'exemple porte sur l'analyse de la croissance de l'emploi dans les régions métropolitaines françaises. L'indicateur étudié est le taux d'évolution de l'emploi entre 2000 et 2014 :

$$\frac{\text{emploi 2014} - \text{emploi 2000}}{\text{emploi 2000}} * 100$$

. La variable géographique est la région, les variables structurelles sont le secteur d'activité (en cinq secteurs) et le statut de l'emploi (salarié ou non salarié). La source utilisée est Estel : Estimations d'emplois localisés.

I. Remarques générales sur l'interprétation des résultats

Les résultats obtenus, quelle que soit la méthode retenue, dépendent du détail des variables. On n'obtiendra pas les mêmes résultats si on utilise en maille géographique les régions, les départements, les zones d'emplois, etc. De même, le découpage en classes des variables de structure influera sur l'importance et / ou la significativité des effets de structure. Par exemple, les secteurs d'activité peuvent être plus ou moins agrégés (en 5, 21, 38 classes). En pratique, il est conseillé de mettre en œuvre l'analyse structurelle-géographique avec différents niveaux de nomenclature, ce qui permet d'enrichir l'analyse.

L'analyse est par nature relative aux variables utilisées. Il est le plus souvent impossible de prendre en compte tous les effets de structure. Par exemple, les facteurs théoriques influençant la croissance de l'emploi dépassent très largement l'activité et le statut de l'emploi. Une partie n'est d'ailleurs pas disponible dans les sources de données. L'étude doit s'efforcer de prendre en compte les variables les plus significatives. Dans toutes les méthodes exposées, les effets structurels non pris en compte sont de fait intégrés dans les effets géographiques. L'interprétation des résultats doit en tenir compte et ne pas analyser ces effets comme relevant uniquement du territoire.

II. Mise en œuvre de la méthode « descriptive »

II.1. Décomposition structurelle-résiduelle

Le taux de croissance moyen de l'emploi entre 2000 et 2014 est de 6,2 % pour la France métropolitaine.

Analyse « descriptive » structurelle-résiduelle

Région	Taux observé	Taux attendu	Effet structurel	Effet résiduel
ILE-DE-FRANCE	6,2	8,6	2,4	-2,4
CENTRE-VAL-DE-LOIRE	-0,3	4,3	-1,9	-4,6
BOURGOGNE-FRANCHE-COMTE	-2,5	3,6	-2,6	-6,1
NORMANDIE	-0,2	4,4	-1,8	-4,7
HAUTS-DE-FRANCE	0,8	4,7	-1,5	-3,9
GRAND EST	-2,8	4,0	-2,3	-6,8
PAYS DE LA LOIRE	10,0	3,9	-2,3	6,1
BRETAGNE	9,7	5,4	-0,8	4,3
NOUVELLE-AQUITAINE	7,8	6,1	-0,2	1,7
OCCITANIE	16,0	7,7	1,4	8,3
AUVERGNE-RHONE-ALPES	8,3	5,2	-1,1	3,1
PROVENCE-ALPES-COTE D'AZUR	15,0	9,7	3,5	5,3
CORSE	31,0	10,1	3,9	20,9

Ci-dessus, le **taux observé** est le taux de croissance de l'emploi entre 2000 et 2014 observé dans la région considérée.

Le **taux attendu** est le taux de croissance qu'aurait enregistré la région si la croissance de l'emploi dans chaque secteur et pour chaque statut d'emploi avait été la même qu'en moyenne en France métropolitaine. C'est le taux obtenu en appliquant les taux de croissance nationaux par secteur et statut de l'emploi à la structure de la région. Il correspond également à l'écart entre le taux observé et l'effet résiduel.

$$\text{Taux attendu} = \sum_i \frac{Z_{ri}}{Z_r} X_{.i}$$

L'**effet structurel** s'interprète comme la part de l'écart au taux national expliquée par la structure secteur*statut de la région. En théorie, il s'agit des écarts entre taux moyens nationaux par secteur et statut et taux moyen national global appliqués à la structure de la région. C'est également la différence entre le taux de croissance régional attendu et le taux de croissance moyen national.

$$\text{Effet structurel} = \sum_i \frac{Z_{ri}}{Z_{r.}} (X_{.i} - X_{..})$$

L'**effet résiduel** est l'écart entre les taux secteur*statut régional et national pondérés par la structure de la région. Il rend compte des évolutions spécifiques de l'emploi de la région, selon les différents secteurs et selon le statut de l'emploi. C'est également la différence entre le taux de croissance régional observé et le taux de croissance régional attendu.

$$\text{Effet résiduel} = \sum_i \frac{Z_{ri}}{Z_{r.}} (X_{ri} - X_{.i})$$

Les égalités suivantes sont vérifiées :

- taux France – taux région = effet structurel + effet résiduel
- effet structurel = taux attendu région – taux France
- effet résiduel = taux région – taux attendu région

Remarque : cette méthode ne permet pas de tester la significativité des effets structurels et résiduels.

II.2 Des résultats détaillés selon la structure

La méthode descriptive permet, pour chaque région, de décomposer les effets structurels et résiduels selon les différentes modalités des variables structurelles.

Exemple : Analyse structurelle-résiduelle du taux de croissance de l'emploi entre 2000 et 2014

Effets de la structure par secteur et statut de l'emploi dans la région Paca

Région PROVENCE-ALPES-COTE D AZUR :
Analyse descriptive structurelle-résiduelle

Taux France	Taux observé PACA	Écart	Taux attendu PACA	Effet structurel	Effet résiduel
6,2	15,0	8,8	9,7	3,5	5,3

En PACA, le taux de croissance de l'emploi est supérieur de 8,8 points au taux moyen national, dont 3,5 points correspondent à un effet de structure.

Le tableau ci-dessous donne pour chaque croisement des variables structurelles : la structure de la zone, la structure du territoire de référence, le taux de croissance de l'emploi dans la région, le taux de croissance de l'emploi dans le territoire de référence et la décomposition des effets structurel et résiduel.

Effets des modalités

secteur	statut	Poids ZONE	Poids REF	Tx ZONE	Tx REF	Effet structurel	Effet résiduel
agriculture	non salarié	1,2	2,2	-23,4	-24,2	0,30	0,01
agriculture	salarié	1,1	1,1	-29,0	-11,1	-0,01	-0,20
construction	non salarié	1,2	0,9	75,1	52,7	0,16	0,27
construction	salarié	5,2	5,0	7,4	6,7	0,00	0,03
industrie	non salarié	0,6	0,5	21,4	13,5	0,01	0,05
industrie	salarié	9,6	16,2	-10,1	-22,5	1,88	1,19
tertiaire marchand	non salarié	6,1	4,2	37,0	36,8	0,60	0,02
tertiaire marchand	salarié	42,3	40,7	15,0	10,8	0,07	1,77
tertiaire non marchand	non salarié	2,1	1,3	54,0	53,1	0,38	0,02
tertiaire non marchand	salarié	30,5	28,0	17,9	10,8	0,12	2,16

Pour un croisement secteur*statut donné, l'effet structurel pour la région r s'écrit :



$$\text{Effet structurel}_{r,\text{secteur}*\text{statut}} = \left(\frac{Z_{r,\text{secteur}*\text{statut}}}{Z_{r.}} - \frac{Z_{.,\text{secteur}*\text{statut}}}{Z_{..}} \right) (X_{.i} - X_{..})$$

poids ZONE

poids REF

taux REF

taux REF

secteur*statut

global

avec $\text{Effet structurel}_r = \sum_{\text{secteur}*\text{statut}} \text{effet_structurel}_{r,\text{secteur}*\text{statut}}$

et l'effet résiduel pour la région r s'écrit :



$$\text{Effet résiduel}_{r,\text{secteur}*\text{statut}} = \frac{Z_{r,\text{secteur}*\text{statut}}}{Z_{r.}} (X_{ri} - X_{.i})$$

poids ZONE

taux ZONE

taux REF

secteur*statut

secteur*statut

avec $\text{Effet résiduel}_r = \sum_{\text{secteur}*\text{statut}} \text{effet_résiduel}_{r,\text{secteur}*\text{statut}}$

En Paca, on observe un effet structurel de 3,5 %, donc assez important compte tenu de la valeur du taux de croissance. La décomposition de l'effet structurel selon les modalités des variables de structure montre que cet effet est dû en grande partie aux salariés de l'industrie.

Le nombre de salariés de l'industrie a beaucoup diminué entre 2000 et 2014 en France (-22,5 %), mais ces salariés de l'industrie sont moins nombreux dans la région qu'en moyenne nationale : 9,6 % contre 16,2 % en France. Cette catégorie joue donc dans le sens d'un taux de croissance plus élevé qu'en moyenne dans la région.

Les effets résiduels portent eux sur les taux de croissance spécifiques de la zone. Les effets les plus importants concernent les salariés du tertiaire marchand et non marchand. Dans les deux cas, le taux de croissance de la zone est bien supérieur à celui observé en moyenne. Ces catégories de salariés contribuent donc à augmenter le taux de croissance de la région.

II.3 Effets croisés, effets séparés

Dans l'analyse croisée des variables structurelles, présentée ci-dessus, les effets sont parfois difficiles à interpréter. On aimerait pouvoir distinguer, dans l'effet structurel, ce qui relève du secteur et ce qui relève du statut de l'emploi, pour hiérarchiser les effets propres des deux variables et mettre en évidence les atouts et faiblesses de la structure de l'emploi dans la région. Cela est réalisable avec l'approche économétrique (cf. partie A, chapitre II) mais pas avec l'approche descriptive. On peut toutefois obtenir des éléments de réponse en réalisant deux analyses à une seule variable structurelle : une pour chacun des deux facteurs secteur et statut de l'emploi.

Cependant, en théorie, la somme des effets dits « séparés », c'est-à-dire obtenus par deux analyses simples séparées, ne correspond pas à l'effet dit « croisé » et on ne peut pas décomposer l'effet structurel total en un effet secteur et un effet statut. Plus précisément, l'égalité n'est vérifiée que lorsque pour chaque croisement, le taux de croissance de l'emploi pour le croisement considéré est égal à la moyenne des taux de croissance des modalités séparées. Par exemple, $TxREF_{\text{agriculture}*\text{salarié}} = 0.5*(TxREF_{\text{agriculture}} + TxREF_{\text{salarié}})$. En pratique, la somme des effets « séparés » et l'effet « croisé » sont souvent de même sens et de niveaux proches. Dans les cas où l'on a quasi-égalité, on peut appréhender la part de chaque variable dans l'effet structurel via l'analyse des effets « séparés ».

En revanche, les effets résiduels obtenus dans les analyses séparées ne peuvent pas être comparés à celui obtenu dans l'analyse croisée. En effet, dans les analyses séparées, l'effet résiduel intègre les effets de structure non pris en compte dans l'effet structurel. La déclinaison de l'effet résiduel par modalité permet juste de mettre en évidence les secteurs et le statut d'emploi qui augmentent plus (ou moins) dans la zone qu'au niveau du territoire de référence.

Même exemple avec des analyses séparées :

Analyse structurelle-résiduelle

Structure=secteur

Structure	Taux attendu	Effet structurel	Effet résiduel
secteur	8,7	2,5	6,3

Structure=statut

Structure	Taux attendu	Effet structurel	Effet résiduel
statut	6,7	0,4	8,4

Ici, on retrouve bien le résultat selon lequel la structure par secteur dans la région est très favorable à la croissance de l'emploi, grâce surtout au faible poids de l'industrie. La structure entre salariés et non salariés est également favorable mais dans une moindre mesure. La somme des effets séparés fait 2,9 et est donc inférieure à l'effet structurel de l'analyse croisée (3,5). Ces résultats permettent toutefois de hiérarchiser les facteurs et d'identifier quelles modalités pèsent le plus fortement dans l'effet de la structure.

Effets des modalités

type=secteur

secteur	statut	Poids ZONE	Poids REF	Tx ZONE	Tx REF	Effet structurel	Effet résiduel
agriculture		2,3	3,3	-26,1	-19,9	0,25	-0,15
construction		6,4	5,9	20,2	13,5	0,04	0,42
industrie		10,3	16,7	-8,2	-21,4	1,77	1,35
tertiaire marchand		48,4	44,9	17,8	13,2	0,25	2,21
tertiaire non marchand		32,6	29,3	20,2	12,7	0,21	2,46

type=statut

secteur	statut	Poids ZONE	Poids REF	Tx ZONE	Tx REF	Effet structurel	Effet résiduel
	non salarié	11,2	9,0	37,0	24,5	0,41	1,40
	salarié	88,8	91,0	12,3	4,4	0,04	6,97

III. Mise en œuvre de la méthode économétrique

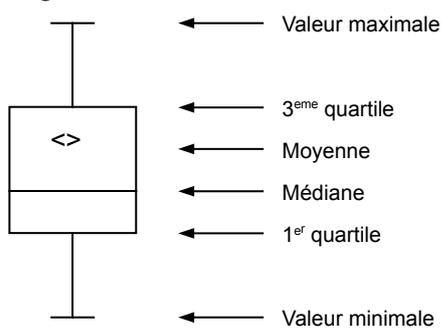
La même analyse a été produite en utilisant le modèle économétrique d'analyse structurelle-géographique. Les résultats globaux, à savoir effets structurel et géographique par région, sont sensiblement identiques, mais les éléments d'interprétation disponibles diffèrent.

III.1 Des effets de structure significatifs ?

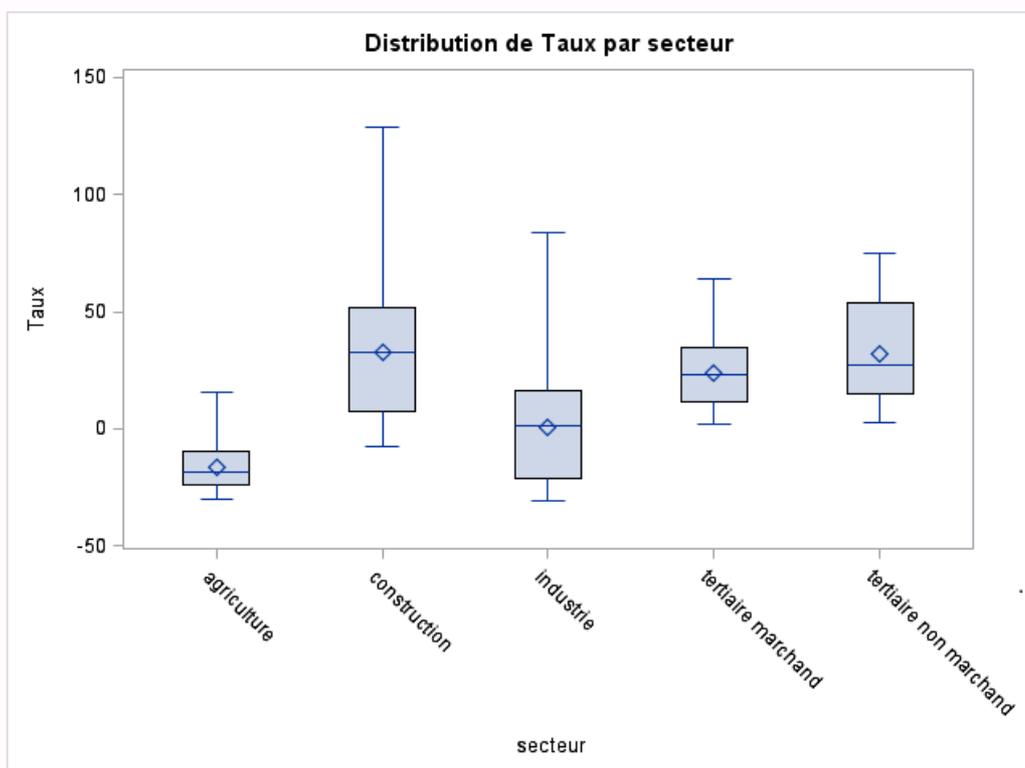
Un des principaux apports de l'approche économétrique est qu'elle permet de tester la significativité des effets mesurés, contrairement à l'approche comptable ou descriptive. L'observation graphique des taux de croissance montre des différences selon les secteurs d'activité et le statut de l'emploi mais sont-elles significatives ?

Pour chaque variable structurelle, les graphiques sous forme de « boîtes à moustaches » permettent d'illustrer la distribution des observations dans chacune des modalités et d'apprécier visuellement l'existence d'un effet de structure.

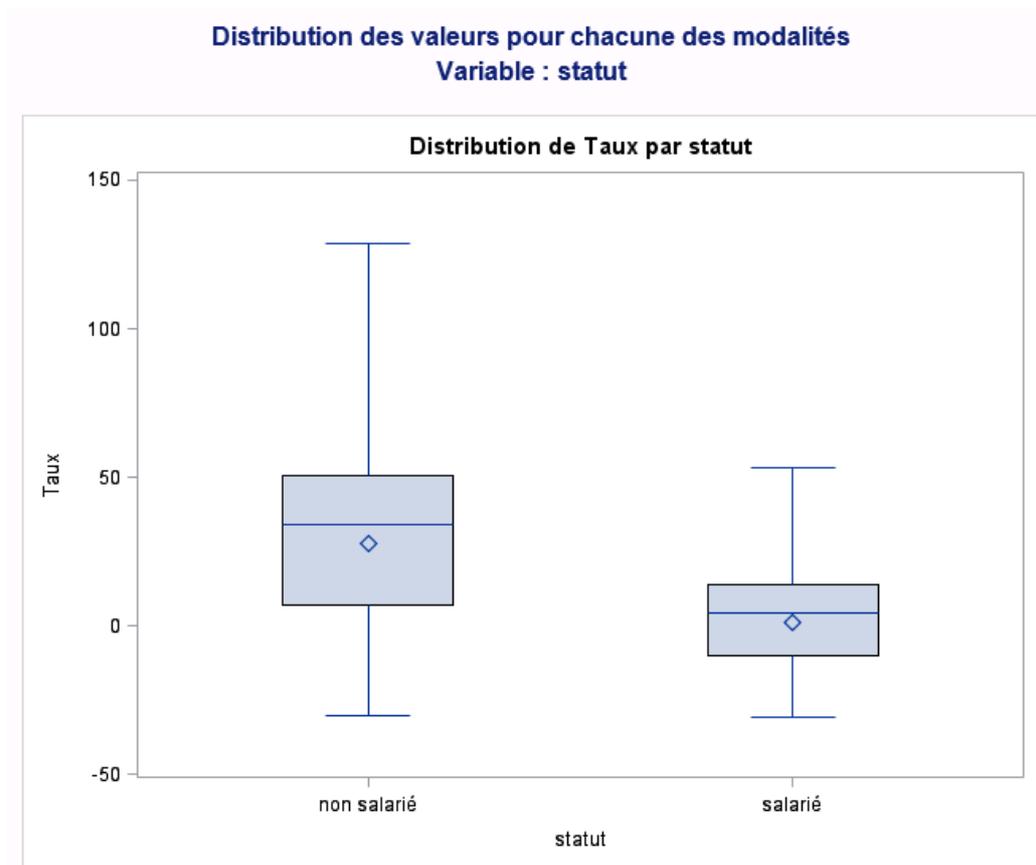
Légende des boîtes à moustache :



Distribution des valeurs pour chacune des modalités
Variable : secteur



Les taux de croissance régionaux de l'emploi dans l'agriculture et l'industrie sont en moyenne inférieurs à ceux des autres secteurs. On peut supposer l'existence d'un effet de structure. Cependant, les distributions se chevauchent, ce qui signifie que le taux de croissance de l'emploi n'est pas systématiquement plus bas dans ces deux secteurs. Cela peut varier d'une région à l'autre ou encore selon le statut de l'emploi.



Le statut d'emploi est clairement discriminant pour l'évolution de l'emploi. Le taux de croissance de l'emploi non salarié est presque systématiquement supérieur à celui de l'emploi salarié.

III.2 Résultats de l'estimation du modèle

a) Validité globale du modèle

ANALYSE STRUCTURELLE - GEOGRAPHIQUE

*La variable expliquée est : $100 * \text{evol0014} / \text{emploi2000}$*

La variable de pondération est : emploi2000

Les variables explicatives sont : reg secteur statut

La table SAS agrégée comprend 130 observations.

ANALYSE DE LA VARIANCE ET TEST DE VALIDITE GLOBALE DU MODELE

Source	DF	Somme_des_carres	Carres_moyens	Stat_F	Proba_critique	R2
Modèle	17	6629715979.10	389983292.89	52.73	0.0000	0.8889
Erreur	112	828365268.85	7396118.47	.	.	.
Total	129	7458081247.95

La statistique de Fisher (Stat_F) permet de tester la validité globale du modèle. Lorsque la probabilité critique associée est inférieure à 0,05, cela signifie que la probabilité d'accepter à tort la pertinence globale du modèle est inférieure à 5 %.

Dans notre exemple, la probabilité critique est inférieure à 0,0001 : on juge le modèle globalement significatif avec une probabilité de se tromper inférieure à 1/1000. Le R² mesure le pourcentage de variabilité des données expliqué par le modèle.

b) Les tests de significativité des variables explicatives

Chaque variable explicative est significative au seuil de 1/1000. Autrement dit, la localisation géographique, l'appartenance sectorielle et le statut d'emploi ont un effet sur le taux de croissance de l'emploi.

SIGNIFICATIVITE DES VARIABLES EXPLICATIVES

Effet	DF	Somme_des_carres_III	DF_num	Stat_F	Proba_critique
reg	12	606 636 588.68	112	6.84	0.0000
secteur	4	5 017 422 698.13	112	169.60	0.0000
statut	1	1 240 692 116.02	112	167.75	0.0000

Les valeurs des statistiques de Fisher font apparaître une hiérarchie entre les facteurs explicatifs : le statut d'emploi et le secteur d'activité sont beaucoup plus explicatifs que le lieu géographique.

c) L'effet des modalités des variables explicatives

On donne d'abord la moyenne de la variable étudiée, car les effets structurels et géographiques s'interprètent en écart à la moyenne. Ici, la croissance de l'emploi observée sur l'ensemble des régions entre 2000 et 2014 est de 6,22%.

MOYENNE DE LA VARIABLE EXPLIQUEE

moyenne : 6.2176

1- Les variables structurelles

Toutes les modalités des variables SECTEUR et STATUT ont un effet significatif au seuil de 5 %. Dans l'agriculture et l'industrie, il y a un déficit de croissance par rapport au taux de croissance moyen : les régions spécialisées dans ces secteurs ont structurellement une croissance plus faible. Inversement, les secteurs de la construction et des services, marchands et non marchands offrent des possibilités de croissance.

EFFETS DES MODALITES DES VARIABLES STRUCTURELLES

Variable : secteur

Secteur	Effet	Ecart_type	Stat_T	Proba_critique
Agriculture	-42.9647	3.1999	-13.43	0.0000
Construction	5.4405	2.1777	2.50	0.0139
Industrie	-25.5178	1.2290	-20.76	0.0000
Tertiaire marchand	6.9801	0.6099	11.44	0.0000
Tertiaire non marchand	7.5661	0.8501	8.90	0.0000

Remarque : l'effet structurel dépend de la finesse du découpage en classes de la variable de structure. On n'obtiendra pas les mêmes valeurs des effets de structure, ni la même significativité, en réalisant l'analyse à un autre niveau d'agrégation (A21 ou A38). En pratique il est conseillé de mettre en œuvre l'analyse structurelle-géographique avec différents niveaux de découpage, ce qui permet d'enrichir l'analyse.

Variable : statut

statut	Effet	Ecart_type	Stat_T	Proba_critique
Salarié	-2.4082	0.1859	-12.95	0.0000
Non salarié	24.2992	1.8761	12.95	0.0000

Pour ce qui est du statut d'emploi, la catégorie « non salarié » a un effet positif sur l'emploi. L'emploi non salarié a en effet fortement progressé au cours de la période 2000-2014.

2- La variable géographique

Le tableau ci-dessous donne la valeur de l'effet géographique dans chaque région. C'est l'effet propre de la région, qui n'est expliqué ni par la structure d'activité ni par la répartition entre emploi salarié et non salarié. L'effet géographique d'une région est significatif (significativement différent de zéro) au seuil de 5% lorsque la probabilité critique associée à la statistique T de Student est inférieure à 0,05.

Cinq régions ont un effet géographique favorable : Pays de la Loire, Occitanie, Auvergne-Rhône-Alpes, Paca et Corse. Au contraire, l'effet est significativement négatif en Île-de-France, Bourgogne-Franche-Comté, Normandie, Hauts-de-France et Grand Est.

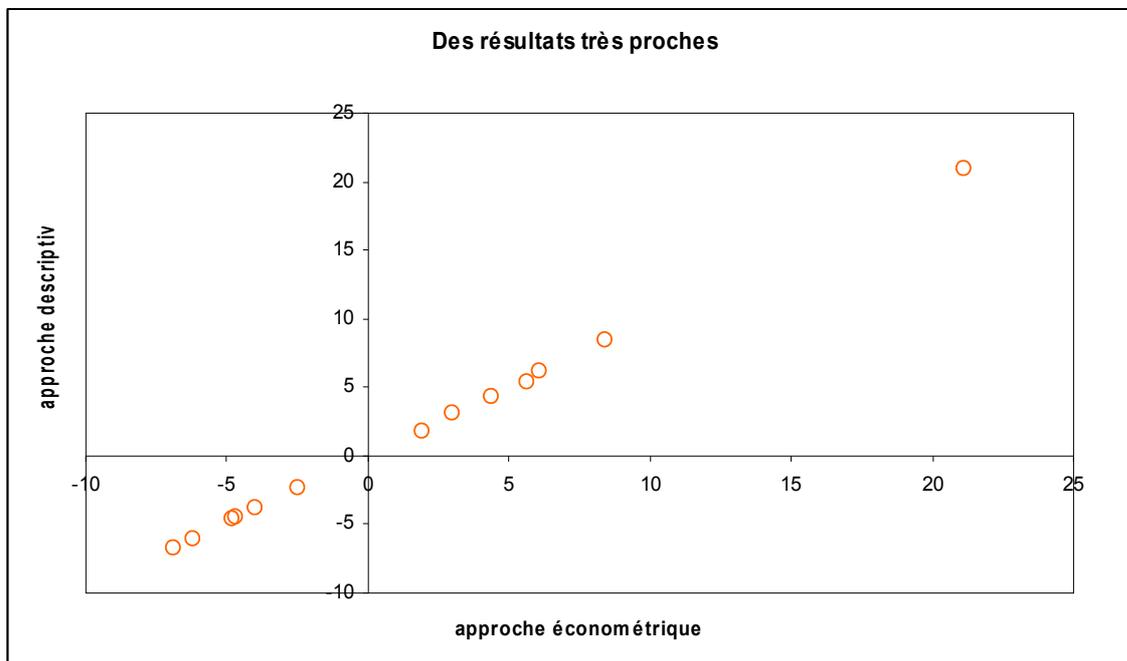
EFFETS DES MODALITES DE LA VARIABLE GEOGRAPHIQUE

reg	Effet	Ecart_type	Stat_T	Proba_critique
ILE-DE-FRANCE	-2.4697	1.0090	-2.45	0.0159
CENTRE-VAL-DE-LOIRE	-4.6392	2.6692	-1.74	0.0850
BOURGOGNE-FRANCHE-COMTE	-6.1588	2.5266	-2.44	0.0164
NORMANDIE	-4.7847	2.3379	-2.05	0.0430
HAUTS-DE-FRANCE	-3.9654	1.7691	-2.24	0.0270
GRAND EST	-6.8484	1.7617	-3.89	0.0002
PAYS DE LA LOIRE	6.1447	2.2466	2.74	0.0073
BRETAGNE	4.3944	2.4240	1.81	0.0725
NOUVELLE-AQUITAINE	1.9496	1.7688	1.10	0.2727
OCCITANIE	8.4226	1.8942	4.45	0.0000
AUVERGNE-RHONE-ALPES	3.0438	1.4618	2.08	0.0396
PROVENCE-ALPES-COTE D AZUR	5.6578	1.9736	2.87	0.0050
CORSE	21.1443	8.8514	2.39	0.0186

En jaune : les régions ayant un effet géographique positif et significatif au seuil de 5%.

En bleu : les régions ayant un effet géographique négatif et significatif au seuil de 5%.

Les effets géographiques sont presque identiques à ceux obtenus avec l'approche descriptive :



d) Décomposition entre effet structurel et effet géographique, et détail des effets structurels

L'écart entre le taux de croissance régional et le taux de croissance moyen national (colonne « Ecart_total ») se décompose entre l'effet géographique et l'effet structurel³ :

- L'effet géographique est obtenu directement par les résultats du modèle d'analyse de la variance (cf. tableau précédent : Effets des modalités de la variable géographique).
- L'effet structurel au niveau régional est calculé en pondérant les effets des secteurs et du statut d'emploi selon de la structure de la région.

Dans l'explication des écarts au taux de croissance moyen, on remarque que l'effet géographique joue un rôle souvent plus important que les facteurs structurels.

³ Il existe aussi un effet résiduel, qui correspond à la part non expliquée par le modèle. Mais lorsqu'il est agrégé au niveau des zones géographiques, l'effet résiduel est très proche de zéro.

DECOMPOSITION DE L'ECART A LA MOYENNE ENTRE EFFET STRUCTUREL ET EFFET GEOGRAPHIQUE

reg	tx_emp	Ecart_total	Effet_structuel	Effet_geographique
ILE-DE-FRANCE	6.22	0.00	2.47	-2.47
CENTRE-VAL-DE-LOIRE	-0.28	-6.50	-1.86	-4.64
BOURGOGNE-FRANCHE-COMTE	-2.49	-8.71	-2.55	-6.16
NORMANDIE	-0.21	-6.43	-1.65	-4.78
HAUTS-DE-FRANCE	0.76	-5.45	-1.49	-3.97
GRAND EST	-2.79	-9.01	-2.16	-6.85
PAYS DE LA LOIRE	10.04	3.82	-2.33	6.14
BRETAGNE	9.73	3.51	-0.89	4.39
NOUVELLE-AQUITAINE	7.76	1.54	-0.41	1.95
OCCITANIE	16.01	9.79	1.37	8.42
AUVERGNE-RHONE-ALPES	8.30	2.08	-0.96	3.04
PROVENCE-ALPES-COTE D AZUR	15.04	8.82	3.16	5.66
CORSE	31.01	24.80	3.65	21.14

Le tableau ci-dessous détaille les effets structurels. Il permet d'observer l'importance relative des facteurs. L'effet sectoriel est prépondérant dans toutes les régions sauf en Occitanie, où l'effet du statut de l'emploi explique la quasi-totalité de l'effet structurel. Dans cette région, le poids de l'emploi non salarié est nettement supérieur à la moyenne française (13,7 % contre 9,0 % en moyenne).

En Normandie, c'est la structure sectorielle qui explique l'effet de structure. Dans cette région, l'industrie est sur-représentée alors que c'est le secteur qui a perdu le plus d'emplois au niveau national sur la période étudiée (-21,4 %). Autre exemple, en Nouvelle-Aquitaine, l'effet sectoriel très négatif du fait du poids important de l'agriculture est compensé par un effet positif du statut de l'emploi grâce à une sur-représentation des non-salariés. L'effet de structure global est alors proche de 0 (-0,41).

DETAIL DES EFFETS STRUCTURELS

reg	Effet structurel	Effet secteur	Proba secteur	Effet statut	Proba statut
ILE-DE-FRANCE	2.47	3.50	0.0000	-1.03	0.0000
CENTRE-VAL-DE-LOIRE	-1.86	-1.93	0.0000	0.07	0.0000
BOURGOGNE-FRANCHE-COMTE	-2.55	-2.79	0.0000	0.23	0.0000
NORMANDIE	-1.65	-1.70	0.0000	0.05	0.0000
HAUTS-DE-FRANCE	-1.49	-0.94	0.0000	-0.55	0.0000
GRAND EST	-2.16	-1.74	0.0000	-0.42	0.0000
PAYS DE LA LOIRE	-2.33	-2.60	0.0000	0.27	0.0000
BRETAGNE	-0.89	-1.62	0.0000	0.73	0.0000
NOUVELLE-AQUITAINE	-0.41	-1.35	0.0000	0.94	0.0000
OCCITANIE	1.37	0.11	0.3053	1.26	0.0000
AUVERGNE-RHONE-ALPES	-0.96	-1.22	0.0000	0.26	0.0000
PROVENCE-ALPES-COTE D AZUR	3.16	2.57	0.0000	0.59	0.0000
CORSE	3.65	2.43	0.0000	1.22	0.0000

III.3 Complément : exemple d'introduction d'un effet croisé

Dans l'exemple du taux de croissance de l'emploi, on peut se demander si l'effet du statut d'emploi est toujours le même, quel que soit le secteur d'activité. Pour le vérifier, on introduit les variables structurelles non plus séparément, mais en interaction.

On obtient en résultat l'effet de chaque croisement entre les modalités des deux variables de structure.

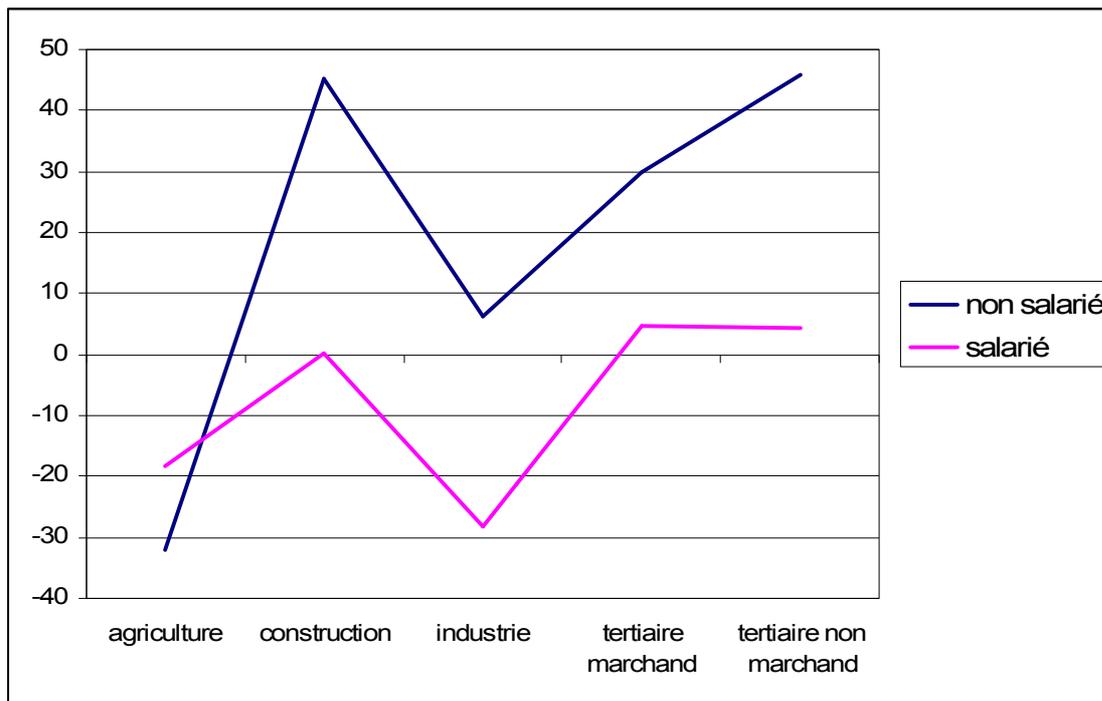
EFFETS DES MODALITES DES VARIABLES STRUCTURELLES

Interaction : secteur*statut

secteur	statut	Effet	Ecart_type	Stat_T	Proba_critique
Agriculture	Non salarié	-31.9368	2.5734	-12.41	0.0000
Agriculture	Salarié	-18.3887	3.6765	-5.00	0.0000
Construction	Non salarié	45.0849	4.0907	11.02	0.0000
Construction	Salarié	0.3675	1.6693	0.22	0.8261
Industrie	Non salarié	6.2900	5.3407	1.18	0.2415
Industrie	Salarié	-28.1782	0.8787	-32.07	0.0000
Tertiaire marchand	Non salarié	29.7995	1.8395	16.20	0.0000
Tertiaire marchand	Salarié	4.7243	0.4689	10.08	0.0000
Tertiaire non marchand	Non salarié	45.9464	3.3542	13.70	0.0000
Tertiaire non marchand	Salarié	4.5152	0.6146	7.35	0.0000

Pour visualiser l'existence ou non d'une interaction entre les variables, on peut réaliser à partir du tableau précédent le graphique ci-dessous. Les deux courbes représentent l'effet structurel de chaque statut d'emploi, en fonction des secteurs d'activité.

Graphique 3 : Effet estimé du statut de l'emploi selon le secteur d'activité (écart au taux de croissance moyen)



L'effet du statut de l'emploi varie selon les secteurs. Dans l'agriculture, le nombre de non salariés a diminué sur la période sous l'effet de la concentration de la production agricole. Au contraire, il a fortement augmenté dans tous les autres secteurs. Dans l'industrie, l'emploi salarié a diminué, pénalisé par la récession de 2008-2009. En contrepartie, le nombre de non salariés a beaucoup augmenté, la pression incitant certaines personnes en situation précaire à créer leur propre emploi. Le schéma est le même pour la construction mais avec un maintien de l'emploi salarié sur la période. Dans les services, l'emploi salarié a progressé, mais dans une proportion bien moindre que l'emploi non salarié.

Conclusion

Dans ses travaux, l'Insee a privilégié l'approche fondée sur l'analyse de la variance, qui permet d'introduire de façon explicite le territoire comme variable explicative, et de construire des tests permettant de mesurer la significativité des effets. Certains auteurs, sans contester l'intérêt de l'approche économétrique, en soulignent les limites (Fothergill & Gudgin;1979). La méthode shift-share initiale reste donc très utilisée dans la pratique (Carré & Levratto 2011).

Les méthodes shift-share ont fait l'objet de nombreux développements méthodologiques, dont certains ont été évoqués dans la première partie. D'autres méthodes comme le « spatial filtering » ont également été proposées (Major et Lopez ; 2009) sans faire l'objet d'utilisations pratiques.

Bibliographie

- ARCELUS, Francisco J. An Extension of Shift-Share Analysis. *Growth and change*, 1984, vol. 15, no 1, p. 3-8.
- ARTIGE, Lionel et NEUSS, Leif. A New Shift-Share Method. *Growth and Change*, 2014, vol. 45, no 4, p. 667-683.
- BERZEG, Korhan. The empirical content of shift-share analysis. *Journal of regional science*, 1978, vol. 18, no 3, p. 463-468.
- CONFAIS, Josiane et LE GUEN, Monique. Premiers pas en régression linéaire avec SAS®. 2007.
- DAGNELIE, Pierre. *Statistique théorique et appliquée*. De Boeck Université, 1998.
- DUNN, Edgar S. A statistical and analytical technique for regional analysis. *Papers in Regional Science*, 1960, vol. 6, no 1, p. 97-112.
- ESTEBAN-MARQUILLAS, Joan Maria. I. A reinterpretation of shift-share analysis. *Regional and urban economics*, 1972, vol. 2, no 3, p. 249-255.
- GAIGNÉ, Carl, PIGUET, Virginie, et SCHMITT, Bertrand. Évolution récente de l'emploi industriel dans les territoires ruraux et urbains: une analyse structurelle-géographique sur données françaises. *Revue d'Économie Régionale & Urbaine*, 2005, no 1, p. 3-30.
- JAYET, Hubert. *Analyse spatiale quantitative: une introduction*. Economica, 1993.
- KNUDSEN, Daniel C. et BARFF, Richard. Shift-share analysis as a linear model. *Environment and Planning A*, 1991, vol. 23, no 3, p. 421-431.
- LE GALLO, Julie et KAMARIANAKIS, Yiannis. The evolution of regional productivity disparities in the European Union from 1975 to 2002: A combination of shift-share and spatial econometrics. *Regional Studies*, 2011, vol. 45, no 1, p. 123-139.
- LOVERIDGE, Scott et SELTING, Anne C. A review and comparison of shift-share identities. *International Regional Science Review*, 1998, vol. 21, no 1, p. 37-58.
- MAYOR, Matías et LÓPEZ-MENÉNDEZ, Ana. *The spatial shift-share analysis-new developments and some findings for the Spanish case*. European Regional Science Association, 2005.
- MAYOR, Matías et LÓPEZ, Ana Jesús. Spatial shift-share analysis versus spatial filtering: an application to Spanish employment data. In : *Spatial Econometrics*. Physica-Verlag HD, 2009. p. 123-142.
- NAZARA, Suahasil et HEWINGS, Geoffrey JD. Spatial Structure and Taxonomy of Decomposition in Shift-Share Analysis. *Growth and change*, 2004, vol. 35, no 4, p. 476-490.
- ROSENFELD, Félix. *Commentaire à l'exposé de MES Dunn sur une méthode statistique et analytique d'analyse régionale: présentation mathématique de la méthode*. 1959.
- ZACCOMER, Gian Pietro. Shift-share analysis with spatial structure: an application to Italian industrial districts. *Transition Studies Review*, 2006, vol. 13, no 1, p. 213-227.

