

# II- L'analyse de données spatiales

Analyse de variance et analyse structurelle-résiduelle

# Introduction

- Enjeu de ces méthodes : savoir interpréter les comparaisons globales entre espaces ou groupes d'espaces
  - › Disparités entre espaces : les effets de structure
    - Disparités de taux de création d'entreprises entre régions
    - Disparités de taux de croissance de l'emploi entre zones d'emploi
    - Disparités de taux de chômage entre quartiers
  - › Différences de moyennes entre groupes d'espaces
    - Problème de l'agrégation de données hétérogènes = les différences de moyenne entre catégories ont-elles un sens ?
    - Tester la pertinence d'une typologie d'espaces : met-elle à jour des trajectoires/comportements spécifiques ?

# A) L'analyse structurelle-résiduelle

## A-1) Approche intuitive

→ Deux régions A et B, composées chacune de deux secteurs, 1 et 2

<b>t=0</b>	<b>S<sub>1</sub></b>	<b>S<sub>2</sub></b>	<b>Total</b>
<b>R<sub>A</sub></b>	200	100	300
<b>R<sub>B</sub></b>	100	200	300
<b>Total</b>	300	300	600

→ Calcul de la disparité moyenne de chaque région :

Si les deux secteurs avaient connu le taux de croissance global moyen, chaque région aurait connu un accroissement d'emploi de +60 :

$$D_A = 70 - 60 = +10$$

$$D_B = 50 - 60 = -10$$

→ Calcul de l'effet structurel :

Croissance théorique : Si les deux secteurs avaient connu le taux de croissance sectoriel moyen

$$ES_A = 90 - 60 = +30$$

$$ES_B = 30 - 60 = -30$$

→ Calcul de l'effet résiduel :

Différence entre croissance réelle et croissance théorique

$$ER_A = 70 - 90 = -20$$

$$ER_B = 50 - 30 = +20$$

On vérifie que :

$$D_A = ER_A + ES_A \quad D_B = ER_B + ES_B$$

t=0	S <sub>1</sub>	S <sub>2</sub>	Total
R <sub>A</sub>	200	100	300
R <sub>B</sub>	100	200	300
<b>Total</b>	300	300	600

t=1	S <sub>1</sub>	S <sub>2</sub>	Total
R <sub>A</sub>	280	90	370
R <sub>B</sub>	170	180	350
<b>Total</b>	450	270	720
Taux de croissance	50%	-10%	20%

Emploi théorique	S <sub>1</sub>	S <sub>2</sub>	Total
R <sub>A</sub>	300	90	390
R <sub>B</sub>	150	180	330
<b>Total</b>	450	270	720

# B-2 l'égalité structurelle-résiduelle

→ La décomposition structurelle-résiduelle s'applique à des « **taux** »: taux de croissance de l'emploi, taux de chômage, taux de création d'entreprises...

→ Notations :

› Les régions sont notées en *minuscule* :

$e_i$  : emploi initial régional dans le secteur  $i$

$\Delta e_i$  : variation de l'emploi régional dans le secteur  $i$

$g_i = \Delta e_i / e_i$  : taux de croissance régional du secteur  $i$

$g = \Delta e / e$  : taux de croissance régional

› Le niveau agrégé (national) est noté en *majuscule* :

$E_i$  : emploi initial total dans le secteur  $i$

$\Delta E_i$  : variation de l'emploi total dans le secteur  $i$

$G_i = \Delta E_i / E_i$  : taux de croissance national du secteur  $i$

$G = \Delta E / E$  : taux de croissance national

# A-2 l'égalité structurelle-résiduelle

→ Pour chaque secteur, on peut écrire l'égalité suivante :  $g_i - G = (G_i - G) + (g_i - G_i)$

→ En effectuant la moyenne régionale des trois termes précédents :

$$\sum_{i=1}^K \frac{e_i}{e} (g_i - G) = \sum_{i=1}^K \frac{e_i}{e} (G_i - G) + \sum_{i=1}^K \frac{e_i}{e} (g_i - G_i)$$

*or*

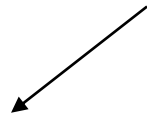
$$\sum_{i=1}^K \frac{e_i}{e} (g_i - G) = \sum_{i=1}^K \frac{e_i}{e} g_i - \sum_{i=1}^K \frac{e_i}{e} G = g - G$$

# A-2 l'égalité structurelle-résiduelle

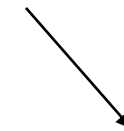
→ On obtient l'égalité structurelle-résiduelle :

$$g - G = \underbrace{\sum_{i=1}^K \frac{e_i}{e} (G_i - G)}_{\text{Effet structurel ou de composition}} + \underbrace{\sum_{i=1}^K \frac{e_i}{e} (g_i - G_i)}_{\text{Effet résiduel ou géographique}}$$

Effet structurel ou  
de composition



Effet résiduel  
ou  
géographique



Il n'intègre que des taux nationaux : les différences entre régions ne sont imputables qu'aux disparités de composition sectorielle

Moyenne régionale des écarts de croissance spécifiques à la région

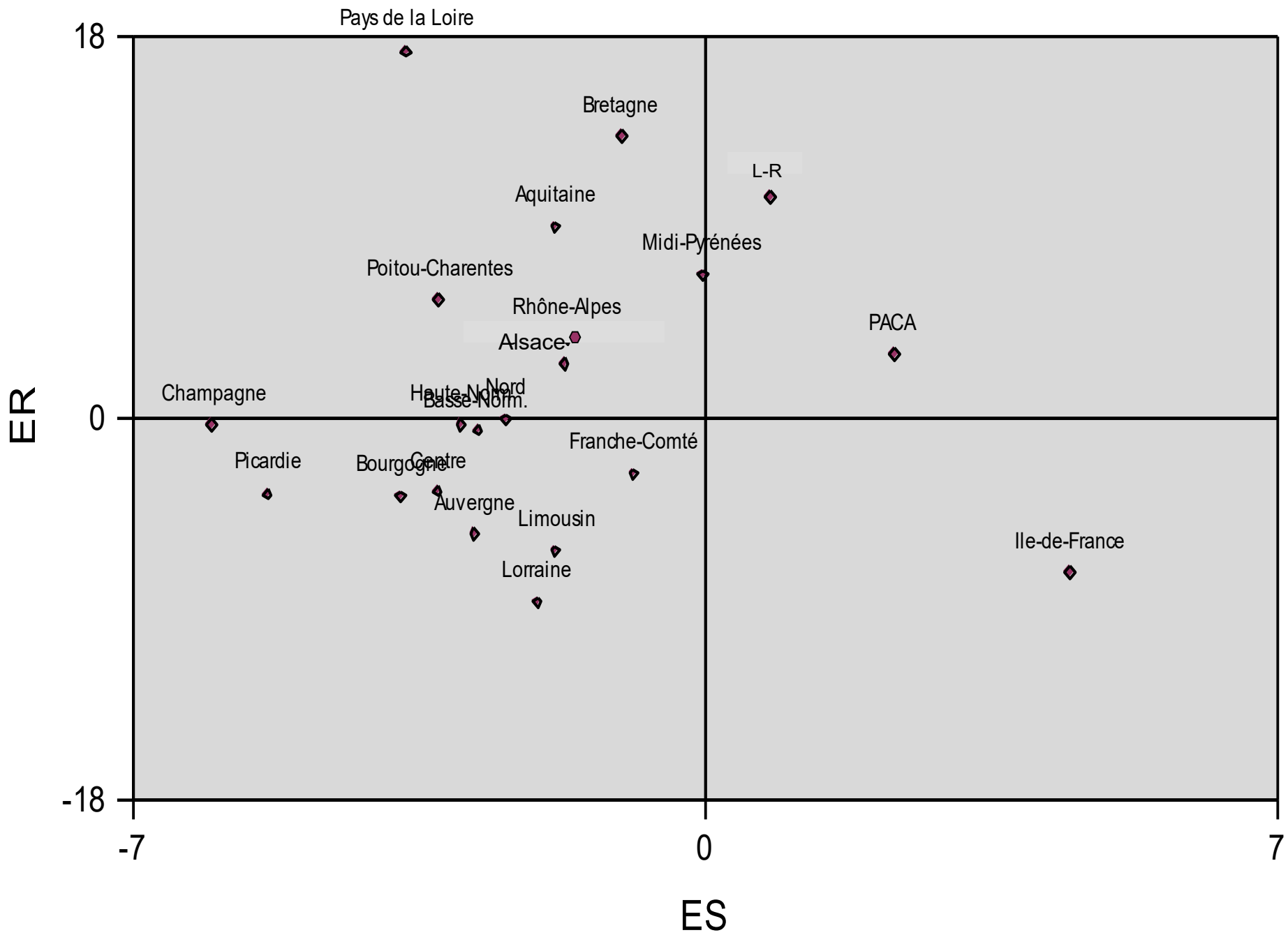
# A-2 l'égalité structurelle-résiduelle

- Application : Evolutions régionales de la valeur ajoutée entre 1990 et 2002
- Effet structurel : secteurs NES 14
- Effet géographique : appartenance régionale



Régions	croissance VA 1990-2002	Disparités régionales
Ile-de-France	45,6%	-2,83%
Rhône-Alpes	50,5%	2,05%
Provence-Côtes-d'Azur	53,8%	5,33%
Nord-Pas-de-Calais	45,9%	-2,51%
Pays de la Loire	62,1%	13,66%
Aquitaine	55,6%	7,19%
Bretagne	60,7%	12,31%
Midi-Pyrénées	55,2%	6,72%
Centre	41,7%	-6,72%
Lorraine	37,7%	-10,75%
Languedoc-Roussillon	59,7%	11,21%
Alsace	49,3%	0,82%
Haute-Normandie	45,2%	-3,26%
Picardie	39,5%	-8,94%
Bourgogne	41,0%	-7,45%
Poitou-Charentes	50,8%	2,45%
Champagne-Ardennes	42,1%	-6,37%
Basse-Normandie	45,1%	-3,35%
Auvergne	40,2%	-8,26%
Franche-Comté	44,9%	-3,51%
Limousin	40,3%	-8,11%
France	48,4%	

Régions	D	ES	ER
Ile-de-France	-2,83	4,46	-7,29
Rhône-Alpes	2,05	-1,7	3,76
Provence-Côtes-d'Azur	5,33	2,31	3,02
Nord-Pas-de-Calais	-2,51	-2,44	-0,07
Pays de la Loire	13,66	-3,67	17,33
Aquitaine	7,19	-1,84	9,02
Bretagne	12,31	-1,03	13,34
Midi-Pyrénées	6,72	-0,04	6,76
Centre	-6,72	-3,28	-3,44
Lorraine	-10,75	-2,06	-8,69
Languedoc-Roussillon	11,21	0,8	10,42
Alsace	0,82	-1,73	2,56
Haute-Normandie	-3,26	-2,99	-0,28
Picardie	-8,94	-5,36	-3,58
Bourgogne	-7,45	-3,73	-3,71
Poitou-Charentes	2,45	-3,27	5,61
Champagne-Ardennes	-6,37	-6,04	-0,33
Basse-Normandie	-3,35	-2,79	-0,56
Auvergne	-8,26	-2,83	-5,43
Franche-Comté	-3,51	-0,88	-2,63
Limousin	-8,11	-1,84	-6,26

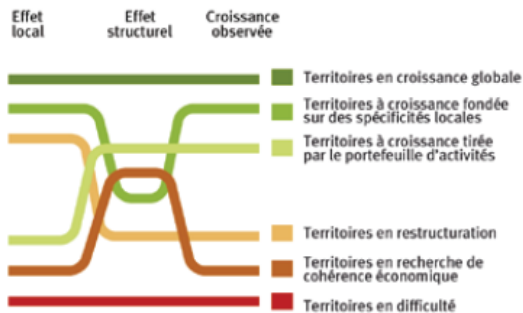


<b>Aquitaine</b>	<b>Quotient de localisation</b>	<b>ES</b>	<b>ER</b>
<b>Agriculture</b>	2,2	-3,41	-0,2
<b>Industries agricoles et alimentaires</b>	1,07	-0,26	0,27
<b>Industries de biens de consomm.</b>	0,64	-0,63	0,47
<b>Industrie automobile</b>	0,41	0,14	0,12
<b>Industries de biens d'équipement</b>	0,86	-0,95	1,62
<b>Industries de biens intermédiaires</b>	0,74	-1,8	1,15
<b>Energie</b>	1,21	-0,73	-1,24
<b>Construction</b>	1,03	-1,27	1,47
<b>Commerce</b>	1,13	-1,79	0,38
<b>Transport</b>	0,79	0,17	-0,02
<b>Activités financières et immobilières</b>	0,93	0,57	2,33
<b>Services aux entreprises</b>	0,8	3,47	2,67
<b>Services aux particuliers</b>	0,87	0,85	0,63
<b>Education, santé, administration</b>	1,17	3,82	-0,67
<b>Total Aquitaine</b>		-1,84	9,02

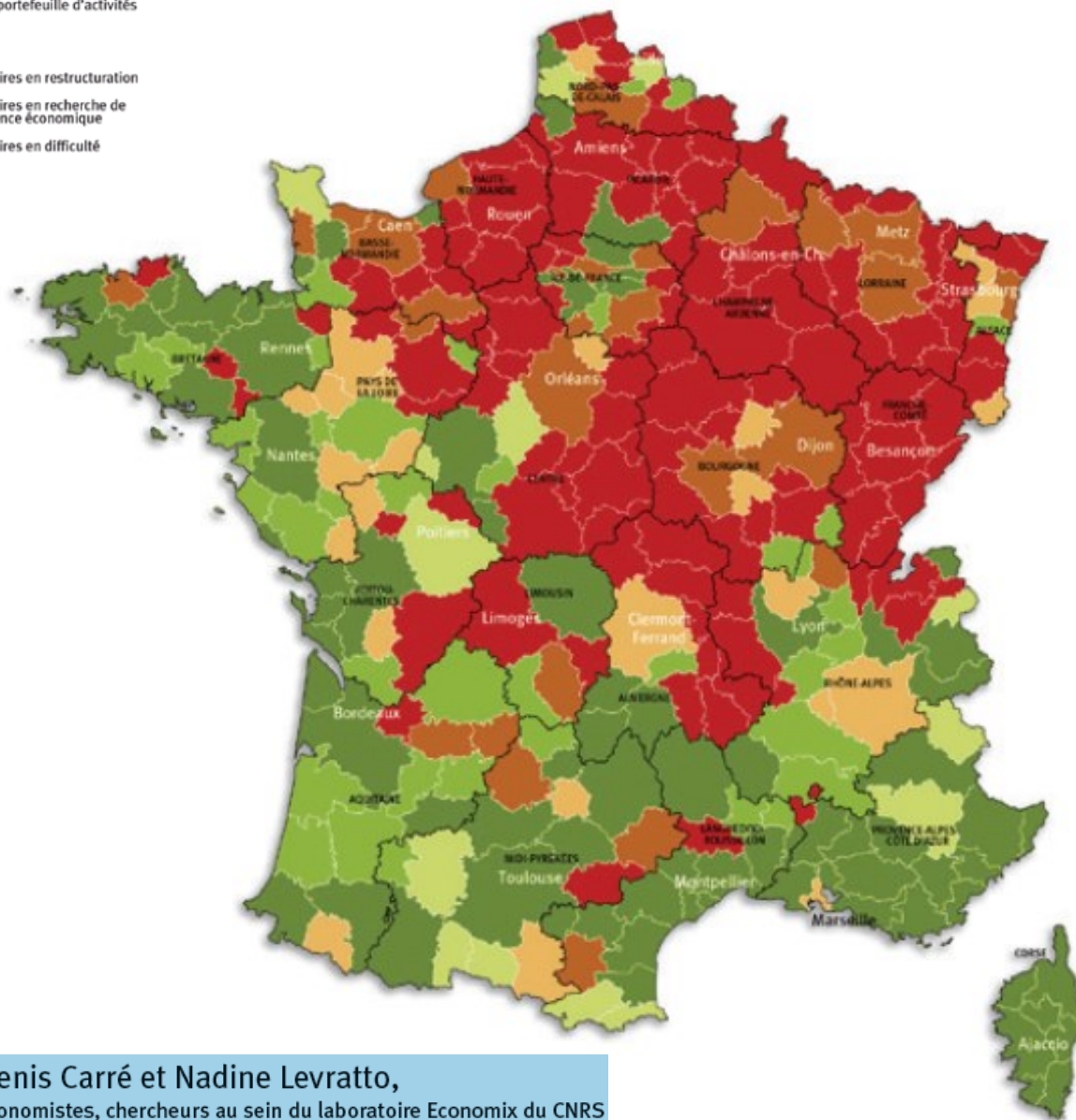
<b>Ile-de-France</b>	<b>Quotient de localisation</b>	<b>ES</b>	<b>ER</b>
<b>Agriculture</b>	0,07	-0,11	-0,06
<b>Industries agricoles et alimentaires</b>	0,45	-0,11	-0,17
<b>Industries de biens de consomm.</b>	1,31	-1,3	-0,51
<b>Industrie automobile</b>	0,95	0,33	-0,01
<b>Industries de biens d'équipement</b>	0,95	-1,05	-0,77
<b>Industries de biens intermédiaires</b>	0,54	-1,32	-1,51
<b>Energie</b>	1,04	-0,63	-0,81
<b>Construction</b>	0,95	-1,17	-2,22
<b>Commerce</b>	0,95	-1,5	-0,94
<b>Transport</b>	1,28	0,28	-0,16
<b>Activités financières et immobilières</b>	1,13	0,69	-0,22
<b>Services aux entreprises</b>	1,47	6,38	1,44
<b>Services aux particuliers</b>	1,43	1,4	-1
<b>Education, santé, administration</b>	0,79	2,58	-0,35
<b>Total Ile-de-France</b>		4,46	-7,29

## 6 profils de territoire : lecture de la typologie

Typologie des zones d'emploi



## Typologie des zones d'emploi selon les combinaisons d'effets structurels et locaux (2000-2009)\*



**1 Les territoires en croissance globale**  
Ces zones d'emploi se distinguent par une croissance relative qui reflète à la fois un portefeuille d'activité porteur et un effet local positif. Au total, 82 zones d'emploi appartiennent à cette catégorie (soit 27 % de l'ensemble des zones d'emploi de France métropolitaine).

**2 Les territoires à croissance fondée sur des spécificités locales**  
Alors qu'elles sont pénalisées par un portefeuille d'activités a priori peu porteur, les 35 zones d'emploi appartenant à cette catégorie réalisent une « sur-performance » en termes d'emplois créés par rapport aux résultats attendus. L'exploitation d'atouts locaux, et des modes d'organisation pertinents expliquent en général ce bon résultat.

**3 Les territoires à croissance tirée par le portefeuille d'activités**  
Cette classe de territoires regroupe 14 zones d'emploi caractérisées par des valeurs de croissance extrêmement proches des valeurs moyennes. Si la croissance est positivement tirée par le profil du portefeuille d'activités, les spécificités locales sont soit insuffisantes pour différencier ce territoire de la moyenne nationale, soit insuffisamment valorisées.

**4 Les territoires en restructuration**  
Ce profil de territoires regroupe 24 zones d'emploi dont l'effet local est positif mais trop faible pour contrebalancer le déclin ou la fragilité des secteurs composant l'appareil productif. Dans ces bassins d'emploi, le repli de l'emploi industriel pèse sur la croissance et la création d'emplois induits par la base compétitive. L'exploitation des capacités d'organisation et de coopération autour de spécialisations porteuses d'économies d'agglomération n'a pas encore suffi à sortir ces zones d'emploi de la crise qu'elles traversent.

**5 Les territoires en recherche de cohérence économique**  
Cette classe de territoires est constituée de 27 zones d'emploi qui, tout en bénéficiant d'un appareil productif composé de secteurs d'activité porteurs, rencontrent des difficultés de croissance et de développement : la croissance de l'emploi observée y est inférieure à la moyenne nationale. L'« effet local » apparaît ici négatif.

**6 Les territoires en difficulté**  
Très dominante numériquement, la catégorie des territoires en difficulté regroupe 122 zones d'emploi dans lesquelles la croissance des effectifs est nettement inférieure à la moyenne nationale. Concernés par une double trajectoire (structurelle et géographique) défavorable, la plupart de ces zones d'emploi se situent au nord de l'axe Avranches – St-Claude. Dans ces régions du Nord-Est, la place considérable de l'industrie explique pour l'essentiel la faible dynamique structurelle de ces zones.



Denis Carré et Nadine Levratto, économistes, chercheurs au sein du laboratoire Economix du CNRS

\* 2000-2009

# 4/ ANALYSES STRUCTURELLE-RÉSIDUELLE : LES EFFETS LOCAUX PRÉDOMINENT TOUJOURS

**TABLEAU 3** RÉSULTATS DE L'ANALYSE STRUCTURELLE-RÉSIDUELLE 2019T4-2020T3 POUR LES RÉGIONS DE FRANCE MÉTROPOLITAINE

RÉGION	EFFET NATIONAL	EFFET STRUCTUREL	EFFET LOCAL	TAUX DE CROISSANCE
Auvergne-Rhône-Alpes	-1,50%	-0,07%	0,24%	-1,41%
Bourgogne-Franche-Comté	-1,50%	-0,16%	-0,35%	-2,09%
Bretagne	-1,50%	0,01%	0,55%	-1,02%
Centre-Val de Loire	-1,50%	-0,12%	0,13%	-1,57%
Corse	-1,50%	-0,01%	-2,36%	-3,95%
Grand Est	-1,50%	-0,05%	-0,12%	-1,75%
Hauts-de-France	-1,50%	0,00%	0,57%	-1,01%
Île-de-France	-1,50%	0,10%	-0,43%	-1,01%
Normandie	-1,50%	-0,04%	-0,25%	-1,80%
<b>Nouvelle-Aquitaine</b>	<b>-1,50%</b>	<b>-0,01%</b>	<b>0,38%</b>	<b>-1,21%</b>
Occitanie	-1,50%	0,00%	0,00%	-1,50%
Pays de la Loire	-1,50%	-0,09%	0,27%	-1,40%
Provence-Alpes-Côte d'Azur	-1,50%	-0,06%	-0,09%	-1,73%

**TABLEAU 4** RÉSULTATS DE L'ANALYSE STRUCTURELLE-RÉSIDUELLE 2019T4-2020T3 POUR LES ZONES D'EMPLOI DE NOUVELLE-AQUITAINE

ZONE D'EMPLOI	EFFET NATIONAL	EFFET STRUCTUREL	EFFET LOCAL	TAUX DE CROISSANCE
Marmande	-1,50%	0,12%	1,31%	-0,14%
Bayonne	-1,50%	-0,13%	1,53%	-0,10%
Libourne	-1,50%	0,07%	1,31%	-0,20%
Bressuire	-1,50%	-0,33%	1,69%	-0,22%
Périgueux	-1,50%	0,20%	1,09%	-0,29%
Dax	-1,50%	-0,06%	1,29%	-0,35%
Mont-de-Marsan	-1,50%	0,01%	1,16%	-0,41%
La Rochelle	-1,50%	-0,05%	1,13%	-0,50%
Guéret	-1,50%	0,24%	0,50%	-0,76%
Tulle	-1,50%	0,21%	0,52%	-0,85%
Saintes - Saint-Jean-d'Angely	-1,50%	0,22%	0,48%	-0,80%
Bordeaux	-1,50%	0,14%	0,46%	-0,90%
Bergerac	-1,50%	0,22%	0,25%	-1,11%
Jonzac - Barbezieux-Saint-Hilaire	-1,50%	0,25%	0,21%	-1,12%
Villeneuve-sur-Lot	-1,50%	0,00%	0,20%	-1,22%
Niort	-1,50%	0,02%	0,29%	-1,27%
Limoges	-1,50%	-0,01%	0,20%	-1,30%
Brive-la-Gaillarde	-1,50%	0,05%	0,21%	-1,32%
Agen	-1,50%	0,22%	-0,09%	-1,45%
Angoulême	-1,50%	0,17%	-0,06%	-1,47%
Cognac	-1,50%	-0,03%	-0,08%	-1,68%
La Teste-de-Buch	-1,50%	-0,35%	0,13%	-1,80%
Ussel	-1,50%	0,37%	-0,63%	-1,84%
Châtelleraut	-1,50%	-0,54%	0,23%	-1,89%
Oloron-Sainte-Marie	-1,50%	0,24%	-0,72%	-2,06%
Paulliac	-1,50%	0,09%	-0,57%	-2,07%
Thouars - Loudun	-1,50%	-0,25%	-0,30%	-2,20%
Pau	-1,50%	0,29%	-0,97%	-2,26%
Poitiers	-1,50%	0,06%	-1,06%	-2,57%
Royan	-1,50%	0,07%	-1,26%	-2,76%
Rochefort	-1,50%	-0,13%	-2,25%	-3,96%
Sarlat-la-Canéda	-1,50%	-0,08%	-2,56%	-4,22%
Parthenay	-1,50%	-0,36%	-3,72%	-5,65%

**DATAR**

Délégation à l'Aménagement du Territoire et à l'Attractivité Régionale

Direction de l'Intelligence Territoriale et de la Prospective



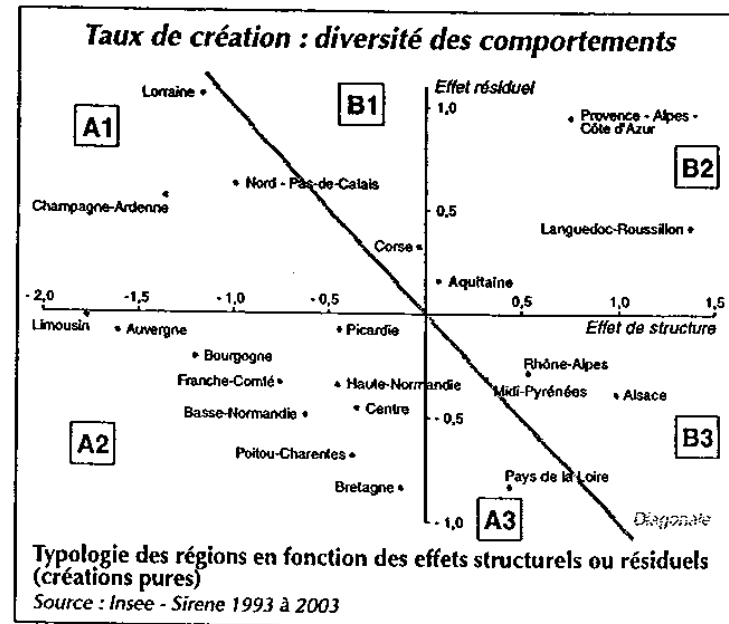
[nouvelle-aquitaine.fr](http://nouvelle-aquitaine.fr)

## Le dynamisme de l'Aquitaine stimule son taux de création

Classement des régions selon l'effet résiduel	Taux de création	Écart avec taux national	Taux attendu	Effet structurel	Effet résiduel
France (hors Île-de-France)	6,7	///	///	///	///
Lorraine	6,6	-0,1	5,6	-1,2	1,1
Provence - Alpes - Côte-d'Azur	8,4	1,7	7,5	0,7	1,0
Champagne-Ardenne	5,9	-0,8	5,4	-1,4	0,6
Nord - Pas-de-Calais	6,4	-0,4	5,7	-1,0	0,6
Languedoc-Roussillon	8,5	1,8	8,1	1,4	0,4
Corse	7,0	0,3	6,7	0,0	0,3
<b>Aquitaine</b>	<b>7,0</b>	<b>0,2</b>	<b>6,8</b>	<b>0,0</b>	<b>0,2</b>
Limousin	4,9	-1,8	5,0	-1,8	0,0
Picardie	6,2	-0,5	6,3	-0,4	-0,1
Auvergne	5,0	-1,7	5,1	-1,6	-0,1
Bourgogne	5,3	-1,4	5,5	-1,2	-0,2
Haute-Normandie	5,9	-0,8	6,3	-0,5	-0,3
Franche-Comté	5,6	-1,1	6,0	-0,8	-0,3
Midi-Pyrénées	6,7	0,0	7,0	0,3	-0,3
Rhône-Alpes	7,0	0,3	7,3	0,6	-0,3
Centre	5,9	-0,8	6,4	-0,4	-0,4
Alsace	7,3	0,6	7,7	1,0	-0,4
Basse-Normandie	5,6	-1,1	6,1	-0,6	-0,5
Poitou-Charentes	5,7	-1,1	6,4	-0,4	-0,7
Pays de la Loire	6,3	-0,4	7,2	0,4	-0,8
Bretagne	5,8	-1,0	6,6	-0,1	-0,9

## Taux de création (créations pures) : Analyse structurelle-résiduelle - régions hors Île-de-France

Source : Insee - Sirene 1993-2003





# Exemple d'application: la création d'entreprises en Aquitaine

## *Le dynamisme local favorable à la création pour six zones*

Classement des zones d'emploi selon l'effet résiduel	Taux de création	Ecart avec taux régional	Taux attendu	Effet structurel	Effet résiduel
Aquitaine .....	7,0	///	///	///	///
Bordeaux-Médoc .....	7,7	0,8	6,6	-0,3	1,1
Bordeaux-Zone centrale .....	8,4	1,5	7,7	0,7	0,8
Villeneuve-sur-Lot-Fumel .....	5,9	-1,1	5,5	-1,5	0,4
Libourne-Montpon-Ste-Foy-la-Grande .....	6,4	-0,5	6,2	-0,7	0,2
Bordeaux-entre-Deux-Mers .....	6,8	-0,1	6,6	-0,3	0,2
Bergerac .....	6,0	-1,0	5,9	-1,1	0,1
Sarlat-la-Canéda .....	6,2	-0,8	6,2	-0,8	0,0
Marmande-Casteljaloux .....	5,3	-1,7	5,4	-1,6	-0,1
Terrasson .....	5,5	-1,5	5,6	-1,3	-0,2
Agen .....	6,2	-0,8	6,3	-0,6	-0,2
Langon-Bazas-La Réole .....	6,2	-0,8	6,4	-0,6	-0,2
Bayonne-Pyrénées .....	7,1	0,1	7,3	0,3	-0,2
Nord-Est de la Dordogne .....	4,3	-2,7	4,6	-2,4	-0,3
Bordeaux-Cubzacais .....	6,5	-0,4	6,8	-0,1	-0,3
Mont-de-Marsan-Haute-Lande .....	5,7	-1,3	6,0	-1,0	-0,3
Lacq-Orthez .....	5,2	-1,8	5,7	-1,3	-0,5
Périgueux .....	5,7	-1,3	6,3	-0,7	-0,6
Pau .....	6,2	-0,7	6,9	-0,1	-0,6
Bordeaux-Arcachonnais .....	8,7	1,8	9,3	2,4	-0,6
Dax .....	5,9	-1,0	6,6	-0,3	-0,7
Sud des Landes .....	8,7	1,7	9,4	2,4	-0,7
Oloron-Mauléon .....	4,4	-2,5	5,2	-1,7	-0,8
Mont-de-Marsan-Est des Landes .....	5,1	-1,8	6,2	-0,7	-1,1

**Taux de création (créations pures) : Analyse structurelle-résiduelle des zones d'emploi d'Aquitaine**

Source : Insee - Sirene 1993-2003

# A-2 l'égalité structurelle-résiduelle

- L'égalité structurelle peut être étendue à la prise en compte simultanée de plusieurs variables de structure
- Exemple : deux variables de structure, indicées  $i$  ( $K$  catégories) et  $j$  ( $J$  catégories)
  - › Ex : secteur d'activité et tranche de taille des établissements

Effets structurels des deux facteurs

$$g - G = \underbrace{\sum_{i=1}^K \frac{e_{i.}}{e} (G_{i.} - G) + \sum_{j=1}^J \frac{e_{.j}}{e} (G_{.j} - G)}_{\text{Effets structurels des deux facteurs}} + \underbrace{\sum_{j=1}^J \sum_{i=1}^K \frac{e_{ij}}{e} (g_{ij} - G_{i.} - G_{.j} - G)}_{\text{Effet résiduel}}$$

Effet résiduel

# Exemple de décomposition structurelle-résiduelle multifactorielle: Analyse des disparités territoriales de salaire

Niveau d'urbanisation	Effet structurel					Effet géographique	Disparité totale
	Global (1) =	CSP (2)	Secteur + (3)	Age (4) +	Sexe (5)	(6)	(7) = (1) + (6)
Zones rurales	- 10.3	- 10.0	- 0.2	- 1.2	1.1	- 7.4	- 17.7
2 000 à 5 000 habitants	- 8.4	- 9.2	4.0	- 0.4	0.3	- 6.0	- 14.4
5 000 à 10 000 habitants	- 7.3	- 7.5	0.7	- 0.5	0.1	- 5.3	- 12.6
10 000 à 20 000 habitants	- 5.9	- 6.2	0.7	- 0.3	- 0.1	- 3.4	- 9.3
20 000 à 50 000 habitants	- 5.6	- 4.9	0	- 0.4	- 0.3	- 4.1	- 9.7
10 000 à 100 000 habitants	- 4.3	- 4.2	0.3	- 0.3	- 0.2	- 3.1	- 7.5
100 000 à 200 000 habitants	- 2.1	- 2.4	0.2	- 0.1	0.1	- 1.7	- 3.8
200 000 à 2 000 000 habitants	- 0.3	0	- 0.4	0.1	0.1	- 0.8	- 1.2
Agglomération de Paris	13.8	13.5	- 0.2	1.1	- 0.5	10.4	24.5

Source: JAYETH., 1993, Analyse spatiale quantitative, Economica.

# B) L'analyse de variance

## B-1 Le théorème de Huygens : la décomposition de la variance

→ La variance d'une variable  $x$  dans une population comptant  $N$  individus est égale à la moyenne des carrés des écarts à la moyenne:

$$V(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

→ Si la population peut être scindée en deux sous-groupes  $G_1$  et  $G_2$  d'effectifs  $n_1$  et  $n_2$ :

*Variance intrapopulation*  
= *moyenne des variances*

*Variance interpopulation*  
= *variance des moyennes*

$$V(x) = \frac{1}{n} [n_1 \cdot V(x_1) + n_2 \cdot V(x_2)] + \frac{1}{n} [n_1 (\bar{x}_1 - \bar{x})^2 + n_2 (\bar{x}_2 - \bar{x})^2]$$

avec

$$V(x_i) = \frac{1}{n_i} \sum_{j=1}^k (x_j - \bar{x}_i)^2$$

# A-1 Le théorème de Huygens : la décomposition de la variance

→ Résultat généralisable quel que soit le nombre de sous-groupes :

$$\text{Variance totale} = \text{variance intergroupes} + \text{variance intragroupes}$$

→ Interprétation :

- › La variance intergroupes = la partie de la dispersion des valeurs du phénomène imputable aux différences entre groupes
  - › La variance intragroupes = variance résiduelle
- Intuitivement, la capacité explicative de la partition en groupes va dépendre du rapport entre ces deux variables

# B-1 Le théorème de Huygens : la décomposition de la variance

- Exemple : calcul d'un indice de ségrégation sur les 137 aires urbaines de plus de 50 000 habitants en 1999
  - › Répartition entre 7 groupes de taille des aires urbaines
- Décomposition de la variance :
  - › Variance intragroupes= moyenne des variances de chaque groupe  
 $= (4*0,00025+...+29*0,00052)/137=0,00052$
  - › Variance intergroupes= variance des moyennes  
 $= [4*(0,199-0,142)^2+...+29(0,121)^2]/137=0,0003$
  - › Environ 37% de la variance totale est « expliquée » par la taille des aires urbaines

<i>Groupes</i>	<i>Nombre d'échantillons</i>	<i>Moyenne</i>	<i>Variance</i>
+ de 1000	4	0,199	0,00025
500 -1000	10	0,169	0,00049
300 -500	12	0,161	0,00024
200 -300	16	0,151	0,00066
100 -200	36	0,143	0,00038
70 -100	30	0,132	0,00079
50 -70	29	0,121	0,00052
ensemble	137	0,142	0,00082

# B-2 l'analyse de variance à un facteur

- Comment comparer les variances intragroupes et intergroupes ?
- On forme un ratio  $F$ , égal au rapport entre variance intergroupes et variances intragroupes, ajusté par les « degrés de liberté » :

$$F = \frac{\frac{\text{Variance intergroupes}}{p-1}}{\frac{\text{Variance intragroupes}}{n-p}}$$

- Ce ratio est distribué selon la loi dite de Fisher-Snédecor
- La décision consiste à tester (au niveau de risque  $\alpha=5\%$ ) l'hypothèse nulle (dite  $H_0$ ) d'absence de différences significatives entre groupes :
  - › Si  $F > F^{\text{théorique}}$  : rejet de l'hypothèse nulle
  - › Sinon, on accepte l'hypothèse nulle

# B-2 l'analyse de variance à un facteur

→ Application :

$$F = \frac{\frac{\text{Variance intergroupes}}{p-1}}{\frac{\text{Variance intragroupes}}{n-p}} = \frac{\frac{0,0003}{7-1}}{\frac{0,00052}{137-7}} \approx 12,56$$

Valeur critique du  $F(6, 130)=2,17$

→ Conclusion

- › Rejet de l'hypothèse nulle
- › Les différences entre groupes sont statistiquement significatives= effet significatif de la taille urbaine sur le niveau de ségrégation



# B-3 l'analyse de variance à n facteurs

→ Le théorème de Huygens se généralise au cas de k partitions croisées:

$$V = V_1^{\text{inter}} + V_2^{\text{inter}} + \dots + V_k^{\text{inter}} + \text{résidu}$$

→ Il suffit de calculer :

La variance globale  $V_k^{\text{inter}}$

Les k variances intergroupes

Les k ratio  $F_k$  :

$$F_k = \frac{\frac{V_k^{\text{inter}}}{p_k - 1}}{\frac{\text{résidu}}{(n-1) - (P_1 - 1) - \dots - (P_k - 1)}}$$

# B-3 l'analyse de variance à n facteurs

## → Le piège des variables non additives

- › La décomposition précédente n'est pertinente que si les variables de partition sont indépendantes les unes des autres : les effets sont alors dits additifs
- › En cas de corrélation, il est absolument nécessaire de prendre en compte les effets *d'interaction* entre les  $k$  facteurs
- › Les termes d'interaction se testent de manière analogue, en construisant un ratio  $F$

# B-4 Analyse structurelle-résiduelle et tests statistiques

- Dans quelle mesure les résultats d'une analyse structurelle-résiduelle sont-ils significatifs ?
  - › Pertinence des variables structurelles choisies (y compris le degré de décomposition)
  - › Les effets géographiques sont-ils significativement différenciés ?

# B-4 Analyse structurelle-résiduelle et tests statistiques

- L'égalité structurelle-résiduelle est strictement équivalente à une analyse de variance « pondérée » à un facteur = le facteur structurel
  - › Pondération des taux = les effectifs de base
  - › L'effet résiduel est alors estimé directement par le résidu de l'analyse de variance
- Intérêt
  - › tester la pertinence de la variable de structure (test du F)
  - › On peut recalculer les effets structurels et résiduels à partir du modèle de variance
- Limites
  - › On ne peut pas tester la pertinence du facteur géographique
  - › Le facteur géographique est considéré comme un résidu
    - Statistiquement, les effets géographiques sont supposés, par hypothèse, nuls en moyenne
    - On commente la non nullité de facteurs considérés statistiquement comme nuls...

# B-4 Analyse structurelle-résiduelle et tests statistiques

- Formulation alternative : modèle d'analyse de variance pondéré à deux facteurs
  - › Le facteur structurel
  - › Le facteur géographique
- Les résultats obtenus sont différents de la décomposition structurelle-résiduelle descriptive
- Intérêt
  - › Test global de la pertinence de chacun des facteurs
  - › Possibilité de tests individuels des différentes catégories composant chacun des facteurs
    - Quels sont les secteurs dont la dynamique est fortement différenciée ?
    - Quelles sont les régions présentant un dynamisme ou une répulsion marqués ?
- Possibilité d'introduire plusieurs variables structurelles

# B-4 Analyse structurelle-résiduelle et tests statistiques

- Exemple : analyse des disparités de taux de chômage en 1999 sur l'agglomération bordelaise à l'échelle des quartiers (IRIS)
- Population = chômeurs ayant déjà travaillé
- Variables structurelles :
  - › CSP des chômeurs (correspondant à leur dernier emploi)
  - › Sexe
- Effet résiduel = effet de quartier

# Pouvoir explicatif des différents facteurs

Facteurs	F de Fisher
Quartier	5,88863**
CSP	245,705**
SEXE	95,2709**

## Test des différentes modalités des deux facteurs de structure

<b>Facteur</b>	<b>Modalités</b>	<b>coefficients estimés</b>	<b>t de Student</b>
<b>Catégorie socio-professionnelle</b>			
	<b>artisans, commerçants, chefs d'entreprise</b>	<b>-0,0057</b>	<b>1,305</b>
	<b>cadres et professions intellectuelles supérieures</b>	<b>-0,0648</b>	<b>21,0**</b>
	<b>professions intermédiaires</b>	<b>-0,0223</b>	<b>8,84**</b>
	<b>employés</b>	<b>0,0272</b>	<b>10,51**</b>
	<b>ouvriers</b>	<b>0,0370</b>	<b>22,59**</b>
<b>Sexe</b>			
	<b>Hommes</b>	<b>-0,0143</b>	<b>9,76**</b>
	<b>femmes</b>	<b>0,01521</b>	<b>9,76**</b>



# Effets géographiques significatifs des quartiers de l'agglomération de Bordeaux

