

# Les indices de concentration géographique à l'épreuve de l'agrégation des données

**Emmanuel Auvray**<sup>1</sup>    **Salima Bouayad Agha**<sup>2</sup>

<sup>1</sup>*GAINS (TEPP)  
Le Mans Université*

<sup>2</sup>*GAINS (TEPP) et CREST  
Le Mans Université*

**Journées de Méthodologie Statistique de l'Insee**  
13 juin 2018



## Introduction (1/2)

La proximité géographique, étudiée depuis Marshall (1890), plus récemment par Porter (1990), permet :

- de réduire des coûts de transport,
- un meilleur échange technologique (potentiel effet *spillover*),
- une meilleure adaptation de la production,
- une adaptation de la main d'œuvre aux entreprises,
- une potentielle spécialisation des territoires.

Spécialisation potentielle du territoire en cas de forte concentration d'un secteur

▶ Terminologie

Ce phénomène de concentration pouvant aboutir à :

- Des économies d'échelle internes
- Des économies sectorielles
- Des économies d'urbanisation

Comment mesurer la proximité géographique des activités ?

# Introduction (1/2)

La proximité géographique, étudiée depuis Marshall (1890), plus récemment par Porter (1990), permet :

- de réduire des coûts de transport,
- un meilleur échange technologique (potentiel effet *spillover*),
- une meilleure adaptation de la production,
- une adaptation de la main d'œuvre aux entreprises,
- une potentielle spécialisation des territoires.

Spécialisation potentielle du territoire en cas de forte concentration d'un secteur

▶ Terminologie

Ce phénomène de concentration pouvant aboutir à :

- Des économies d'échelle internes
- Des économies sectorielles
- Des économies d'urbanisation

**Comment mesurer la proximité géographique des activités ?**

## Introduction (2/2)

Application d'indices spatiaux dont le but est d'étudier la localisation des activités sans modélisation explicite de l'espace :

- Sur données américaines : Ellison et Glaeser (1997)
- Sur données françaises : Maurel et Sédillot (1999), Barlet et al. (2008)
- Sur données britanniques : Devreux et al. (2004), Duranton et Overman (2005)

Des travaux axés sur le choix et la qualité des indices :

- Quelles propriétés doivent remplir ces indices ? (Combes et Overman (2004) et Fratesi (2008)) 
- Confrontation aux données (Fratesi (2008), Guimarães et al. (2011)).
- Comparaison d'indices (Billings et Johnson (2015wp), Palan (2010)).

## Introduction (2/2)

Application d'indices spatiaux dont le but est d'étudier la localisation des activités sans modélisation explicite de l'espace :

- Sur données américaines : Ellison et Glaeser (1997)
- Sur données françaises : Maurel et Sédillot (1999), Barlet et al. (2008)
- Sur données britanniques : Devreux et al. (2004), Duranton et Overman (2005)

Des travaux axés sur le choix et la qualité des indices :

- Quelles propriétés doivent remplir ces indices ? (Combes et Overman (2004) et Fratesi (2008)) 
- Confrontation aux données (Fratesi (2008), Guimarães et al. (2011)).
- Comparaison d'indices (Billings et Johnson (2015wp), Palan (2010)).

# Les indices de concentration

## Indices discrets

On résume l'espace à l'appartenance à une entité géographique

- 1 Indices de première génération, créés et utilisés dans d'autres domaines.

Indice de Gini (1912) ▶

Indice de Herfindahl (1950) ▶

- 2 Indices de seconde génération, spécifiquement créés pour étudier la concentration géographique des établissements

Ellison et Glaeser (1994) ▶

Maurel et Sédillot (1999) ▶

## Indices continus

Les frontières ne sont pas déterminantes car on regarde les distances. Nécessité d'avoir des données géolocalisées à l'adresse.

- Duranton et Overman (2002) ▶

- Marcon et Puech (2003) ▶

# Application sur données françaises

- Analyse de la répartition des activités en France métropolitaine à partir des données des stocks d'établissements à l'échelle de la commune (2014, INSEE).

Découpage	Fin				Agrégé
Géographique	Commune	Arrondissement	Département	Région	Région 2016
Sectoriel	NAF5	NAF4	NAF3	NAF2	NAF1

- Calcul des différents indices de concentration en considérant différentes échelles géographiques (communes, département, régions).
- Est-ce que les indices de seconde génération sont plus pertinents que les indices de première génération ?
- Existe-t-il une dépendance des mesures de concentration au choix géographique ?
- Pour un secteur, quelle est l'échelle géographique qui maximise la concentration des établissements ?

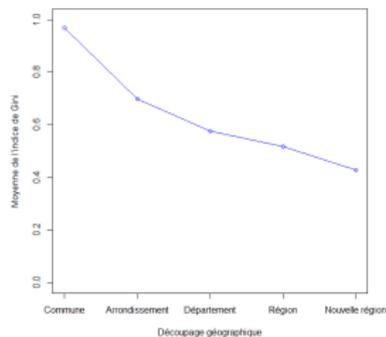
# Application sur données françaises

- Analyse de la répartition des activités en France métropolitaine à partir des données des stocks d'établissements à l'échelle de la commune (2014, INSEE).

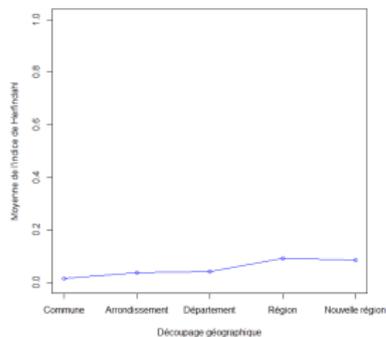
Découpage	Fin				Agrégé
Géographique	Commune	Arrondissement	Département	Région	Région 2016
Sectoriel	NAF5	NAF4	NAF3	NAF2	NAF1

- Calcul des différents indices de concentration en considérant différentes échelles géographiques (communes, département, régions).
- Est-ce que les indices de seconde génération sont plus pertinents que les indices de première génération ?
- Existe-t-il une dépendance des mesures de concentration au choix géographique ?
- Pour un secteur, quelle est l'échelle géographique qui maximise la concentration des établissements ?

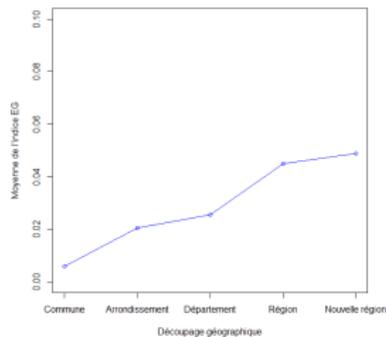
# Résultats sur données françaises



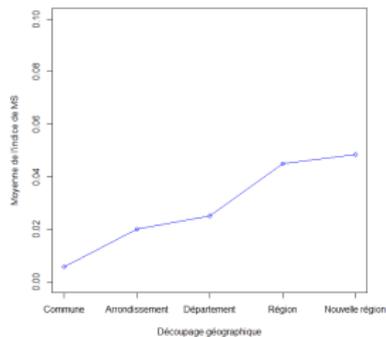
(a) Gini



(b) Herfindahl



(c) Ellison Glaeser



(d) Maurel Sédillot

## Explications possibles

- Nature des données
- Utilisation de la moyenne
- Mesures mises en œuvre

**Peut-on considérer que, quelles que soient les données utilisées et la répartition des activités prises en compte, les indices conduisent toujours à des conclusions identiques selon le niveau d'agrégation considéré ?**

# Effet théorique de l'agrégation géographique

TABLE – Effets de l'agrégation géographique sur les indices de concentration

	Effet empirique	Effet théorique
Gini	↘	↘ sous conditions
Herfindahl	↗	↗
MS	↗	↗ si la concentration du secteur augmente plus que la concentration de l'ensemble des activités
EG	↗	↗

Comment expliquer ces variations suite à l'agrégation géographique des données ?

# Effet théorique de l'agrégation géographique

TABLE – Effets de l'agrégation géographique sur les indices de concentration

	Effet empirique	Effet théorique
Gini	↘	↘ sous conditions
Herfindahl	↗	↗
MS	↗	↗ si la concentration du secteur augmente plus que la concentration de l'ensemble des activités
EG	↗	↗

Comment expliquer ces variations suite à l'agrégation géographique des données ?

# Le MAUP (*Modifiable Areal Unit Problem*)

Influence du découpage spatial sur les résultats de traitements statistiques ou de modélisation (Openshaw et Taylor (1979, 1981, 1984)).

## Le problème de l'échelle (*scale effect*)

Variation de l'information engendrée lorsqu'un jeu d'unités spatiales est agrégé afin de former **des unités moins nombreuses et plus grandes**. Si le nombre de zones est trop faible celles-ci seront trop homogènes.

## Le problème de zonage (*zone effect ou shape effect*)

Variation de l'information engendrée par **les différents schémas possibles d'agrégation à une même échelle**.

Si l'agrégation géographique a un impact systématique sur l'indice de concentration, cet impact est expliqué par l'effet d'échelle.

Comment identifier l'effet de zonage ?

# Le MAUP (*Modifiable Areal Unit Problem*)

Influence du découpage spatial sur les résultats de traitements statistiques ou de modélisation (Openshaw et Taylor (1979, 1981, 1984)).

## Le problème de l'échelle (*scale effect*)

Variation de l'information engendrée lorsqu'un jeu d'unités spatiales est agrégé afin de former **des unités moins nombreuses et plus grandes**. Si le nombre de zones est trop faible celles-ci seront trop homogènes.

## Le problème de zonage (*zone effect* ou *shape effect*)

Variation de l'information engendrée par **les différents schémas possibles d'agrégation à une même échelle**.

Si l'agrégation géographique a un impact systématique sur l'indice de concentration, cet impact est expliqué par l'effet d'échelle.

Comment identifier l'effet de zonage ?

# Le MAUP (*Modifiable Areal Unit Problem*)

Influence du découpage spatial sur les résultats de traitements statistiques ou de modélisation (Openshaw et Taylor (1979, 1981, 1984)).

## Le problème de l'échelle (*scale effect*)

Variation de l'information engendrée lorsqu'un jeu d'unités spatiales est agrégé afin de former **des unités moins nombreuses et plus grandes**. Si le nombre de zones est trop faible celles-ci seront trop homogènes.

## Le problème de zonage (*zone effect ou shape effect*)

Variation de l'information engendrée par **les différents schémas possibles d'agrégation à une même échelle**.

Si l'agrégation géographique a un impact systématique sur l'indice de concentration, cet impact est expliqué par l'effet d'échelle.

**Comment identifier l'effet de zonage ?**

# Identifier l'effet zonage

- 1 Répartir des établissements sur une territoire
- 2 Découper le territoire en deux régions
  - ▶ Une frontière linéaire
  - ▶ Une frontière dépend de son inclinaison et des aires des deux régions

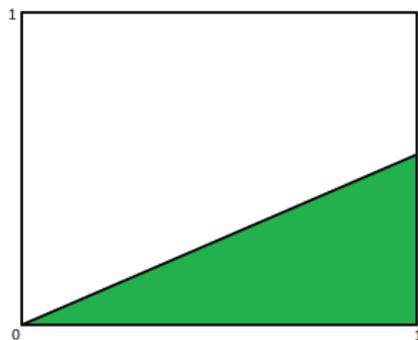
Cette méthode nous permet de vérifier l'effet zonage si, pour une même aire mais des inclinaisons différentes, l'indice de concentration varie.

# Identifier l'effet zonage

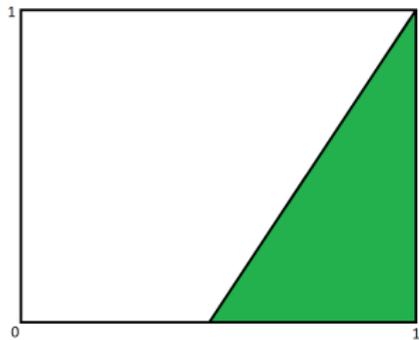
- 1 Répartir des établissements sur une territoire
- 2 Découper le territoire en deux régions
  - ▶ Une frontière linéaire
  - ▶ Une frontière dépend de son inclinaison et des aires des deux régions

Cette méthode nous permet de vérifier l'effet zonage si, pour une même aire mais des inclinaisons différentes, l'indice de concentration varie.

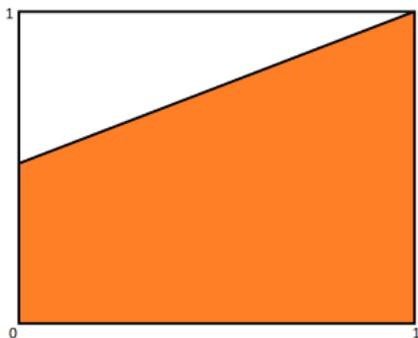
## Exemples de frontières



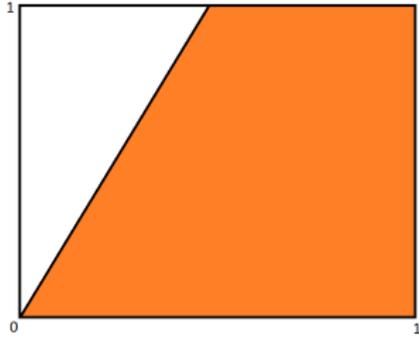
(a) Aire= 0,25 et Inclinaison de  $22,5^\circ$



(b) Aire= 0,25 et Inclinaison de  $67,5^\circ$

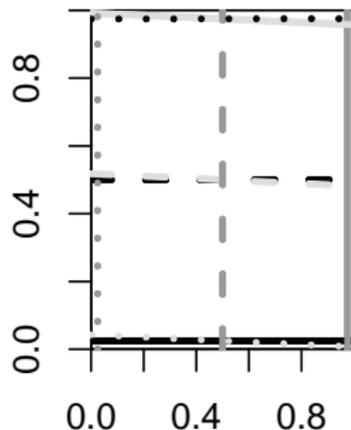
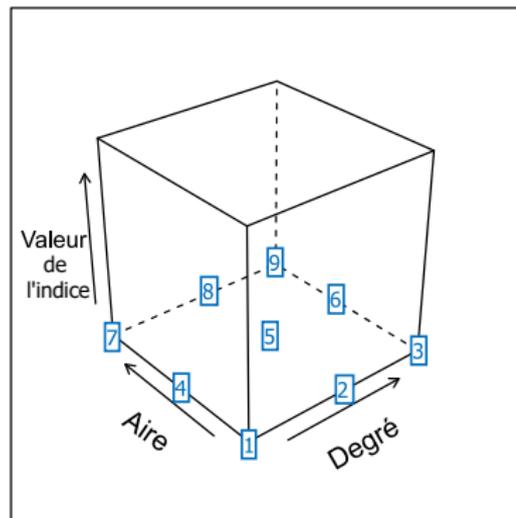


(c) Aire= 0,75 et Inclinaison de  $22,5^\circ$



(b) Aire= 0,75 et Inclinaison de  $67,5^\circ$

# Tracer des frontières automatiquement



	Degré = 0	Degré = 90	Degré = 179
Aire = 0,01	1 ———	2 ———	3 ———
Aire = 0,5	4 - - - -	5 - - - -	6 - - - -
Aire = 0,99	7 . . . .	8 . . . .	9 . . . .

# Illustration

Répartition homogène et affectation sectorielle arbitraire des établissements

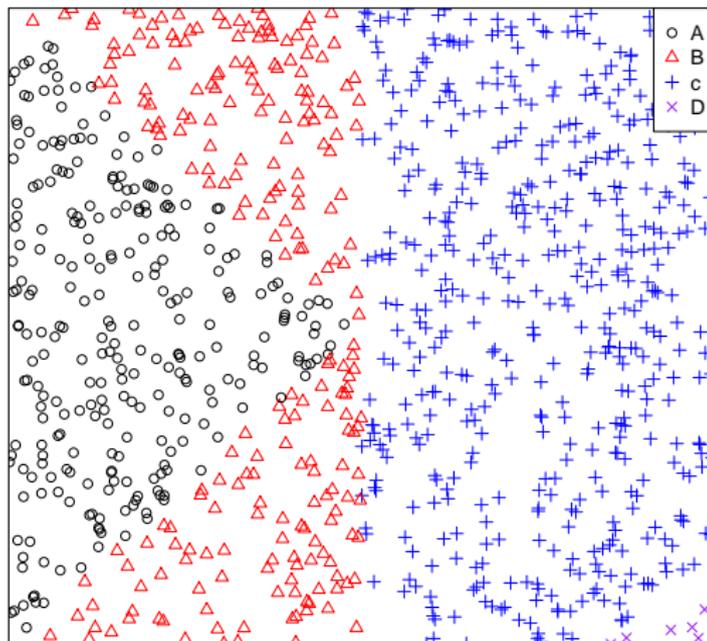
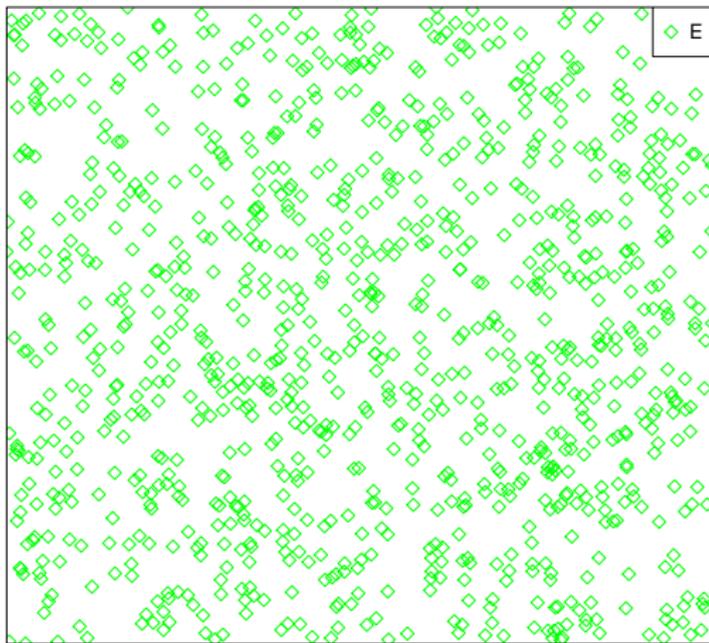


FIGURE – Répartition des 1000 établissements des secteurs A, B, C et D



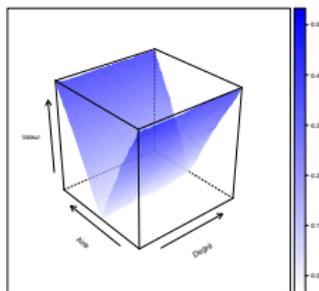
**FIGURE** – Répartition des 1000 établissements du secteur E.

# Statistiques descriptives

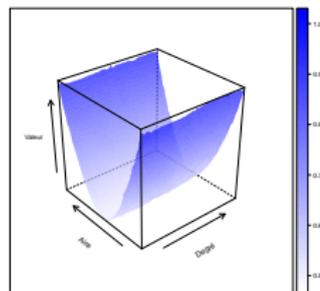
Indice	Secteur A	Secteur B	Secteur C	Secteur D	Secteur E
<b>Gini</b>	0.727 (0.320)	0.547 (0.361)	0.620 (0.326)	0.955 (0.169)	0.507 (0.288)
<b>WGini</b>	0.200 (0.156)	0.115 (0.010)	0.148 (0.113)	0.385 (0.270)	0.099 (0.047)
<b>Herfindahl</b>	0.631 (0.387)	0.430 (0.400)	0.491 (0.376)	0.940 (0.213)	0.340 (0.299)
<b>EG</b>	0.402 (0.422)	0.125 (0.167)	0.197 (0.204)	2.491 (8.000)	0.000 (0.001)
<b>MS</b>	0.345 (0.779)	0.138 (0.624)	0.190 (0.647)	0.746 (1.861)	0.025 (0.078)

Moyenne et écart-type par indice et par secteur

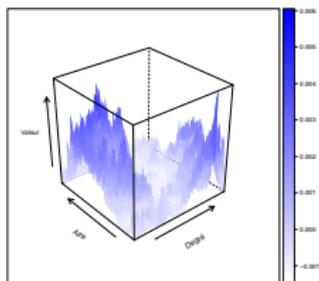
# Résultats pour le secteur E



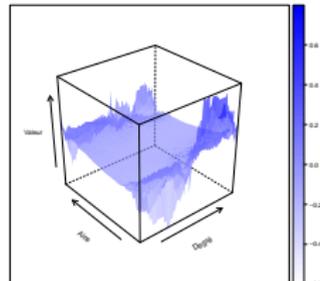
(a) Gini



(b) Herfindahl

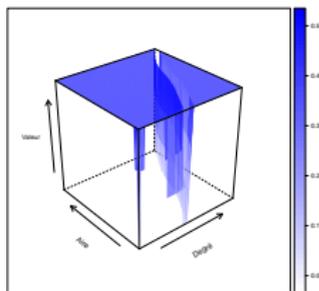


(c) Ellison Glaeser

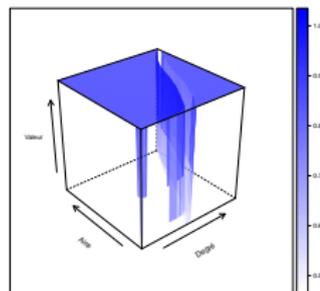


(d) Maurel Sédillot

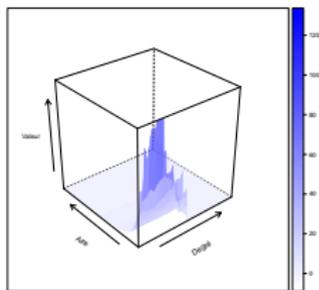
# Résultats pour le secteur D



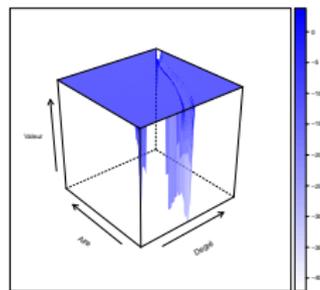
(a) Gini



(b) Herfindahl



(c) Ellison Glaeser



(d) Maurel Sédillot

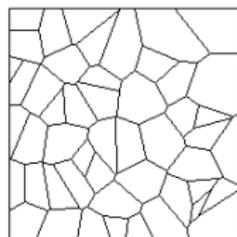
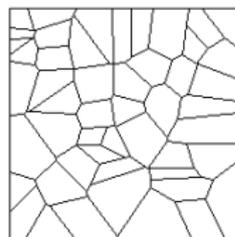
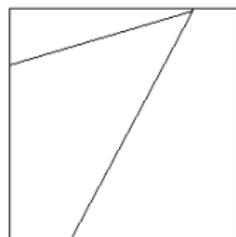
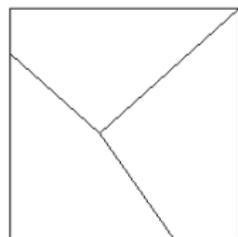
Il est possible que nos résultats soient sensibles :

- aux découpages en **deux** zones
- à la répartition contrôlée

# Variation du nombre de zones géographiques

Découpages géographiques à partir de diagrammes de Voronoï de deux à cinquante zones.

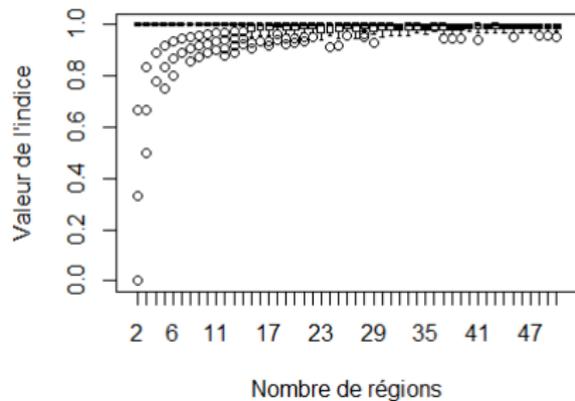
- On a 100 découpages géographiques pour un même nombre de zones.



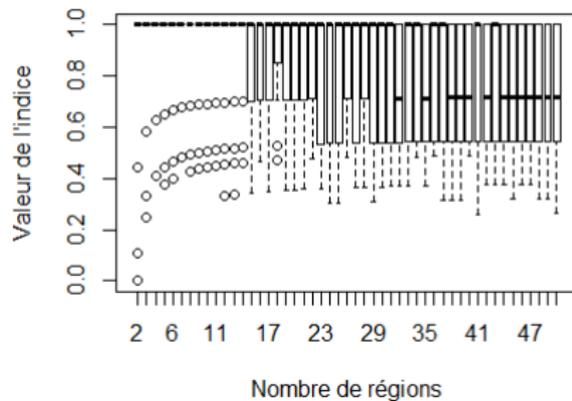
Exemples de diagrammes de Voronoï en trois et cinquante zones

Quelle est l'évolution de la moyenne des indices suite à l'augmentation du nombre de régions ?

# Comparaison des distributions de N à N+1 régions pour le secteur D

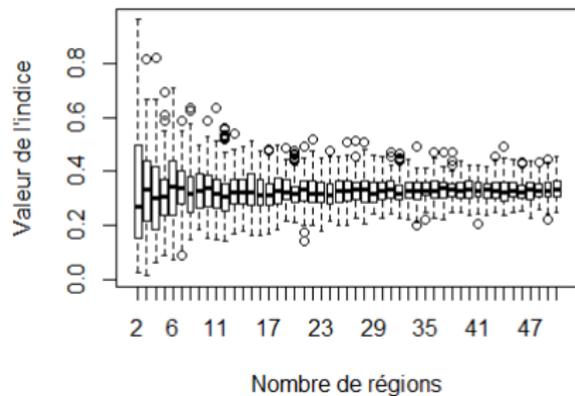


(a) Gini

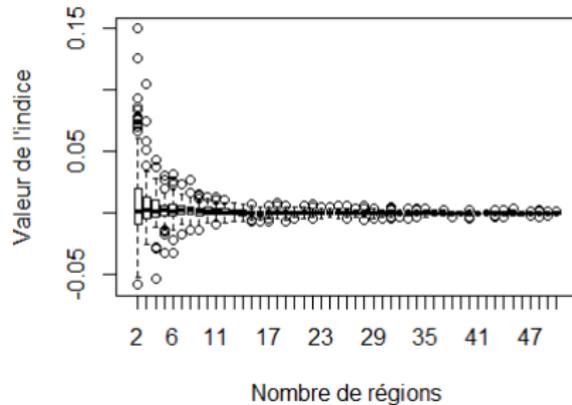


(b) Herfindahl

# Comparaison des distributions de $N$ à $N+1$ régions pour le secteur E



(a) Gini



(b) Maurel Sédillot

## Comparaison des distributions de N à N+1 régions

Indice	Secteur A	Secteur B	Secteur C	Secteur D	Secteur E	Moyenne
Gini	4	13	4	44	2	13
WGini	19	15	17	8	6	13
Herfindahl	13	13	8	58	10	20
EG	19	13	40	8	13	18
MS	10	6	8	4	4	7
Moyenne	13	12	15	25	7	14

Pourcentage de tests de Kolmogorov-Smirnov pour lesquels les distributions à N et N+1 régions sont significativement différentes

- L'indice de Herfindahl est le plus sensible et l'indice MS est le moins sensible à une augmentation du nombre de zones
- L'indice d'EG est presque aussi sensible que l'indice de Herfindahl à cause du secteur C
- L'utilisation de l'indice WGini n'améliore pas nécessairement la précision

**Certains indices sont plus sensibles à une augmentation du nombre de zones, cette sensibilité varie selon la logique de localisation du secteur**

## Variation des logiques de localisation

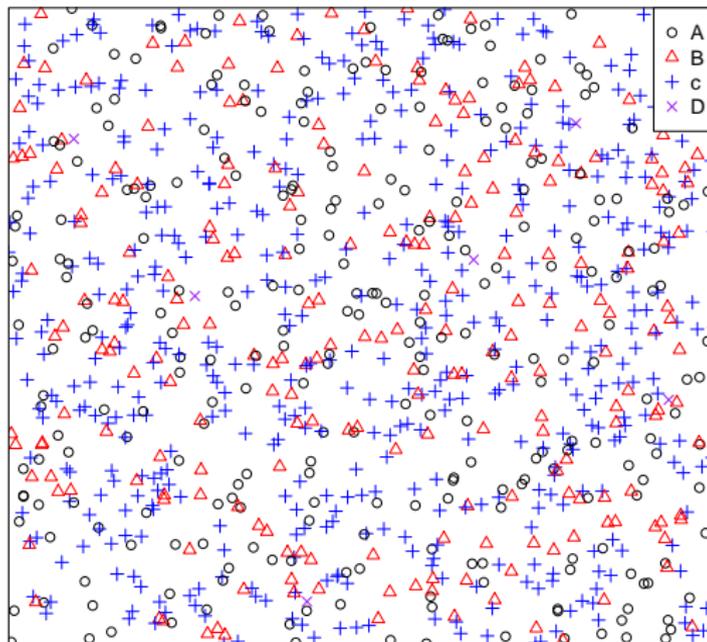
On procède à 16 logiques de localisations des secteurs A, B, C et D

- Répartition aléatoire
- Répartition concentrée en un ou plusieurs pôles

Le secteur E ne change pas de logique de localisation.

On procède uniquement à des découpages en deux zones.

# Illustration



Répartition aléatoire des 1000 établissements des secteurs A, B, C et D

## Statistiques descriptives pour des localisations aléatoires

	Secteur A	Secteur B	Secteur C	Secteur D	Secteur E
	Aléa.	Aléa.	Aléa.	Aléa.	Aléa.
<b>N</b>	271	231	492	6	1000
<b>Gini</b>	0.483 (0.287)	0.493 (0.293)	0.494 (0.286)	0.501 (0.317)	0.499 (0.289)
<b>WGini</b>	0.094 (0.045)	0.096 (0.046)	0.097 (0.046)	0.108 (0.077)	0.098 (0.046)
<b>Herfindahl</b>	0.316 (0.291)	0.329 (0.302)	0.326 (0.294)	0.351 (0.349)	0.333 (0.297)
<b>EG</b>	-0.002 (0.003)	-0.001 (0.003)	-0.001 (0.002)	-0.043 (0.237)	-0.001 ( $<0.001$ )
<b>MS</b>	-0.030 (0.132)	0.010 (0.143)	-0.008 (0.122)	-0.248 (1.260)	0.009 (0.045)

- À indice donné, les valeurs pour le secteur D sont les plus élevées en valeur absolue et avec les écarts-type les plus élevés.
- En revanche, l'augmentation du nombre d'établissements ne donne pas l'impression d'avoir un effet systématique.

## Impact de la concentration sur la valeur des indices

	Secteur A	Secteur B	Secteur C	Secteur D
<b>N</b>	Conc. 271	Conc. 231	Conc. 492	Conc. 6
<b>Gini</b>	0.719*** (0.298)	0.775*** (0.310)	0.513*** (0.405)	0.790*** (0.337)
<b>WGini</b>	0.191*** (0.137)	0.176*** (0.121)	0.110*** (0.113)	0.210*** (0.161)
<b>Herfindahl</b>	0.606*** (0.367)	0.697*** (0.380)	0.427*** (0.422)	0.737*** (0.389)
<b>EG</b>	0.317*** (0.325)	0.231*** (0.209)	0.184*** (0.174)	0.320*** (0.523)
<b>MS</b>	0.368*** (0.668)	0.608*** (0.494)	0.180*** (0.661)	0.555*** (0.722)

- La concentration augmente significativement la valeur des indices.
- Plus il y a d'établissements, moins la valeur est forte.

## Valeur de la concentration pour le secteur E suite à la concentration des autres secteurs

	A, B, C et D aléa.	Si A conc.	Si B conc.	Si C conc.	Si D conc.
N		271	231	492	6
Gini	0.499 (0.289)	0.499 (0.289)	0.499 (0.289)	0.499 (0.289)	0.499 (0.289)
WGini	0.098 (0.046)	0.100*** (0.050)	0.094*** (0.046)	0.105*** (0.055)	0.097*** (0.045)
Herfindahl	0.333 (0.297)	0.333 (0.297)	0.333 (0.297)	0.333 (0.297)	0.333 (0.297)
EG	-0.001 ( $<0.001$ )	0.008*** (0.010)	0.005*** (0.006)	0.018*** (0.018)	-0.001*** (0.001)
MS	0.009 (0.045)	0.008*** (0.138)	-0.042*** (0.085)	0.019*** (0.217)	0.015*** (0.060)

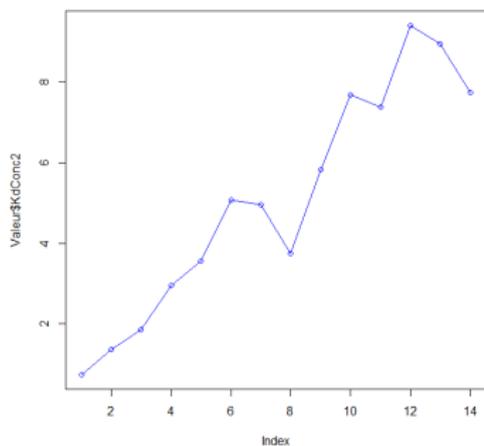
- Aucun effet pour les indices de Gini et de Herfindahl car il s'agit d'indice d'agglomération
- La pondération de l'indice de Gini en fait un indice de concentration
- Plus les établissements qui se concentrent sont nombreux, plus les valeurs sont significativement impactées.

# Conclusion

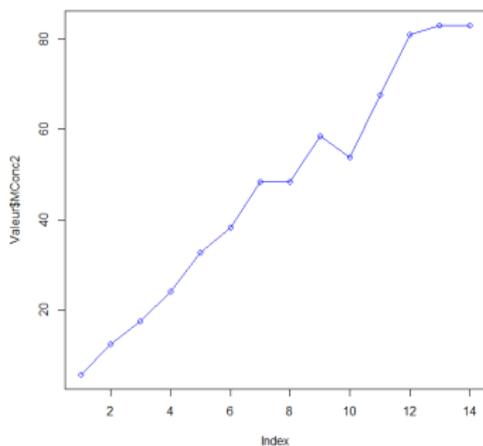
- Il existe un effet systématique de l'agrégation géographique des données sur les indices de concentration.
- En appliquant un découpage géographique de manière automatique, on constate que les indices discrets sont sensibles au MAUP.
- L'indice de Maurel et Sédillot n'est pas sensible à l'effet d'échelle mais est sensible à l'effet zonage. À l'inverse de l'indice d'Ellison et Glaeser.
- La logique de localisation d'un secteur ne peut être déduite des indices discrets.

**Merci de votre attention**

- Duranton et Overman (2002) et Marcon et Puech (2003)
- Ne dépendent pas du découpage géographique mais du choix de la distance pertinente

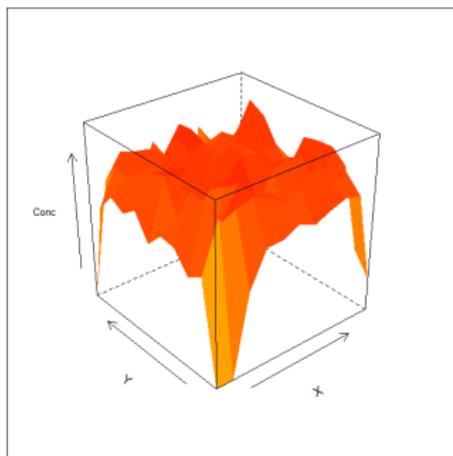


Duranton-Overman

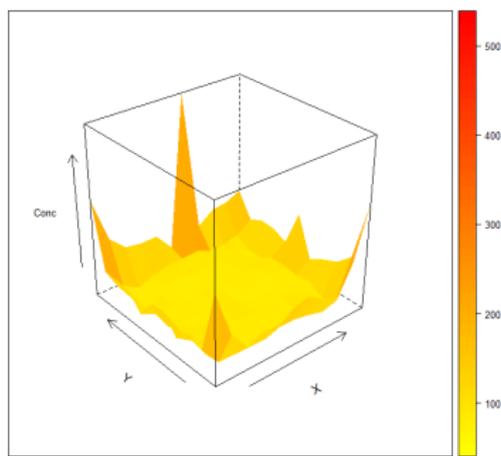


Marcon-Puech

- Ne dépendent pas du découpage géographique mais de la position relative d'un cluster



Duranton-Overman



Marcon-Puech

- Choix des localisations contrefactuelles
- Problème du seuil de distance pour présenter les résultats

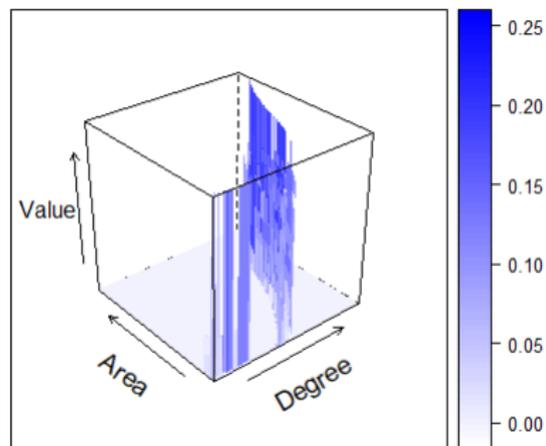
### Application par Barlet, Briant et Crusson (2008)

TAB. 1 – Part de secteurs manufacturiers concentrés aux distances de 5, 30 et 150 km

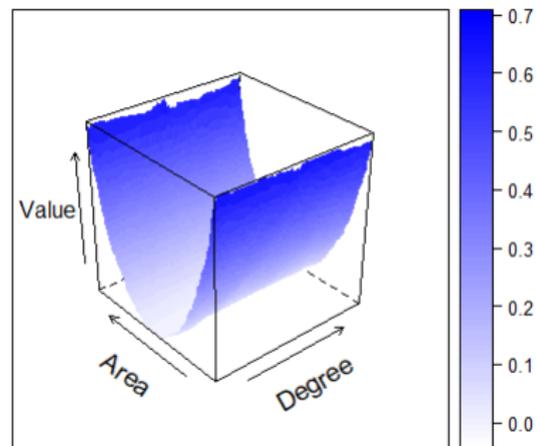
5km	5km seul	5 & 30 km seul	5 & 150 km seul	5 & 30 & 150 km
34 %	0 %	24 %	0 %	10 %
30km	30km seul	30 & 150 km seul		
38 %	2 %	2 %		
150km	150km seul			
22 %	10 %			

*Note de lecture : 34 % des secteurs manufacturiers sont significativement concentrés à 5 km. Parmi ces secteurs, aucun n'est concentré à 5 km uniquement. En revanche, 24 % des secteurs sont concentrés à la fois à 5 et 30 km. Source : Insee - Calculs des auteurs.*

# Indice de Theil



(a) Secteur D



(b) Secteur E

# Terminologie

## Localisation

La localisation désigne la présence d'une firme sur un territoire. La localisation ne désigne que l'implantation d'un établissement uniquement au sens physique du terme.

## Agglomération

L'agglomération désigne le regroupement d'activité (établissements et/ou secteurs) sur un territoire.

## Concentration

La concentration désigne une sur-représentation d'un secteur sur un territoire, cette sur-représentation étant trouvée à l'aide d'indices.

## Spécialisation

La spécialisation désigne le fait que l'activité d'un territoire est très marquée par un secteur.

# Gini

Pour cela, il faut ordonner de manière croissante les parts du secteur  $k$  dans chaque région.

Pour  $n$  régions, on a  $r_k^1 \leq \dots \leq r_k^{i-1} \leq r_k^i \leq r_k^{i+1} \leq \dots \leq r_k^n$ . La fonction cumulative de  $r_k^i$  est noté  $R_k^i$ , soit la part des  $i$ èmes régions ayant les plus faibles  $r_k^i$  parmi les  $n$  régions.

L'indice de Gini est la somme des écarts entre la répartition homogène et la courbe de Lorenz tel que :

$$G_k = 1 - 2 \times \sum_{i=1}^n (R_k^i) \quad (1)$$

Si la répartition empirique correspond à une répartition homogène,  $G_k = 0$ . Plus  $G_k$  augmente, moins la répartition est homogène (avec un maximum de  $G_k = 1$ <sup>1</sup>).

---

1. Une version alternative (et peut-être plus intuitive graphiquement) est  $G_k = 0,5 - \sum_{i=1}^n (R_k^i)$ ,  $G_k$  est alors compris entre 0 et 0,5.

La répartition est parfaitement homogène si l'aire  $A$  est nulle. La droite représente une répartition homogène des établissements sur le territoire, l'aire  $A$  est nulle si la courbe (répartition issue des données) se superpose à la droite et  $G_k = 2A$  si  $G_k$  est compris entre 0 et 1.

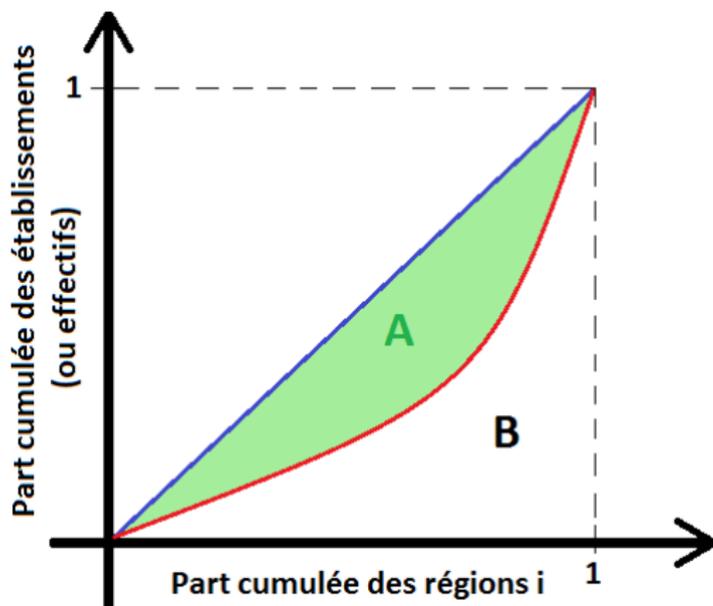


FIGURE – Courbe de Lorenz d'une répartition homogène

Cependant, des répartitions différentes peuvent avoir la même valeur.

La figure montre les courbes de Lorenz de deux répartitions différentes mais dont les indices de Gini sont égaux.

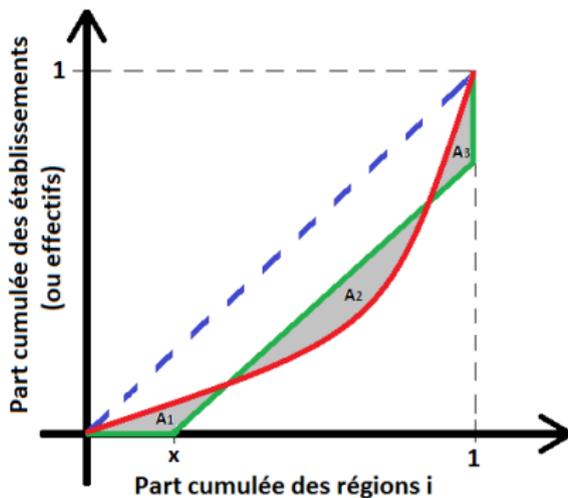


FIGURE – Courbes de Lorenz de deux répartitions



# Indice de Herfindahl

$$H_k = \sum_{i=1}^N s_k^{i2}$$

$s_k^i$  : la part des emplois du secteur  $k$  dans la zone  $i$   $N$  : le nombre de régions sur le territoire

Si  $H_k = 1$ , il y a une concentration totale du secteur.

Si  $H_k = 1/N$ , cela veut dire que les effectifs du secteur  $k$  sont répartis de manière homogène donc un phénomène de dispersion<sup>2</sup>.



---

2. On peut normaliser les valeurs entre 0 et 1.

## Indice d'Ellison et Glaeser

$$\gamma_{EG} = \frac{\frac{\sum_i (s_i - x_i)^2}{1 - \sum_i x_i^2} - H_{Etab_k}}{1 - H_{Etab_k}} \quad (2)$$

$s_i$  : la part de l'emploi du secteur dans une zone  $i$

$x_i$  : la part de l'unité géographique  $i$  dans l'emploi total

$H_{Etab_k}$  : concentration des emplois du secteur au sein des établissements du secteur

Si  $\gamma_{EG} > 0.05$ , il y a une concentration forte du secteur.

Si  $0.05 \geq \gamma_{EG} > 0.02$ , il y a une concentration modérée du secteur.

Si  $\gamma_{EG} = 0$ , il n'y a pas de concentration.

Si  $\gamma_{EG} < 0$ , il y a une dispersion des établissements.



## Indice de Maurel et Sédillot

$$\gamma_{MS} = \frac{\frac{\sum_i s_i^2 - \sum_i x_i^2}{1 - \sum_i x_i^2} - H_{Etab_k}}{1 - H_{Etab_k}} \quad (3)$$

$s_i$  : la part de l'emploi du secteur dans une zone  $i$

$x_i$  : la part de l'unité géographique  $i$  dans l'emploi total

$H_{Etab_k}$  : concentration des emplois du secteur au sein des établissements du secteur

Si  $\gamma_{MS} > 0.05$ , il y a une concentration forte du secteur.

Si  $0.05 \geq \gamma_{MS} > 0.02$ , il y a une concentration modérée du secteur.

Si  $\gamma_{MS} = 0$ , il n'y a pas de concentration.

Si  $\gamma_{MS} < 0$ , il y a une dispersion des établissements. 

## Indice de Duranton et Overman

Pour une industrie donnée avec  $n$  établissements, on calcule les  $\frac{n(n-1)}{2}$  distances **euclidiennes**. Soit  $d_{i,j}$  la distance euclidienne entre les établissements  $i$  et  $j$ . Pour estimer la densité ( $K$ -densité) à chaque point  $d$ , on utilise une estimation en utilisant le noyau de la densité de fenêtre  $h$ ,  $f$  est une fonction gaussienne.

$$\hat{K}(d) = \frac{1}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^n f\left(\frac{d - d_{i,j}}{h}\right)$$

Dans leur papier, DO se limitent volontairement à la médiane de toutes les distances obtenues. Si un établissement décide de se localiser en France, il ne tiendra compte que des établissements déjà existants dans un rayon limité de kilomètres. ▶

▶ Intervalle de confiance

## Indice de Marcon et Puech

La construction de cette mesure se base sur la question suivante : est-ce qu'à la distance maximum de  $r$ , il y a plus de chances de trouver des entreprises similaires par rapport à l'ensemble des autres entreprises ?

Il faut définir :

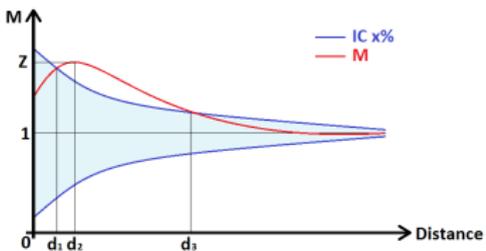
- Le type de point référence (entreprises du secteur automobile par exemple) ;
- Les voisins cibles ( $T$ ), du même secteur par exemple.
- La localisation de contrôle ( $N$ ), la localisation industrielle dans son ensemble.

À une distance inférieure à  $r$ , le nombre total de voisins cibles d'un point  $i$  est  $T_i$ , et le nombre total de voisins  $N_i$ .

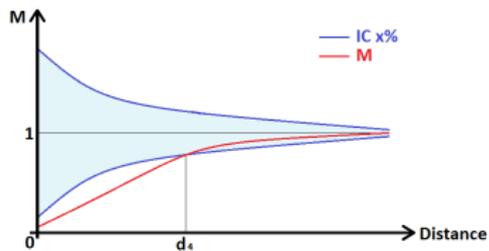
On obtient donc un ratio de voisins cibles  $T_i/N_i$  et le ratio moyen est noté  $\overline{T_i/N_i}$ . L'indice  $M$  est noté au final de la manière suivante :

$$M = \frac{\overline{T_i/N_i}}{T/N} \quad (4)$$

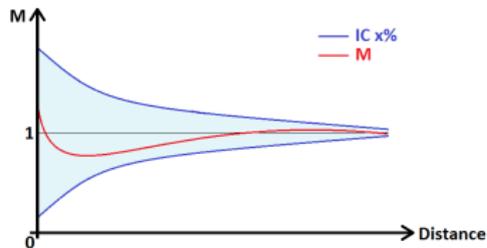
Si  $\overline{T_i/N_i} > T/N$ ,  $M > 1$  et cela veut dire qu'il y a proportionnellement plus d'établissements cibles à une distance  $r$  que sur l'ensemble du territoire. 



A



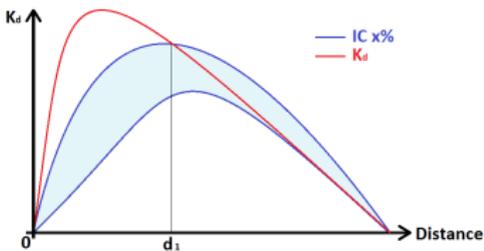
B



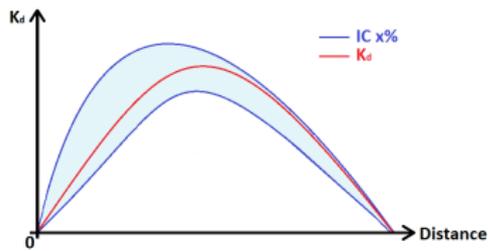
C

FIGURE – Exemples

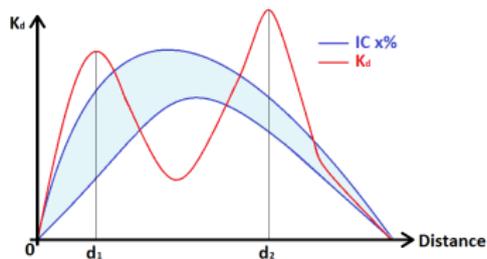




A



B

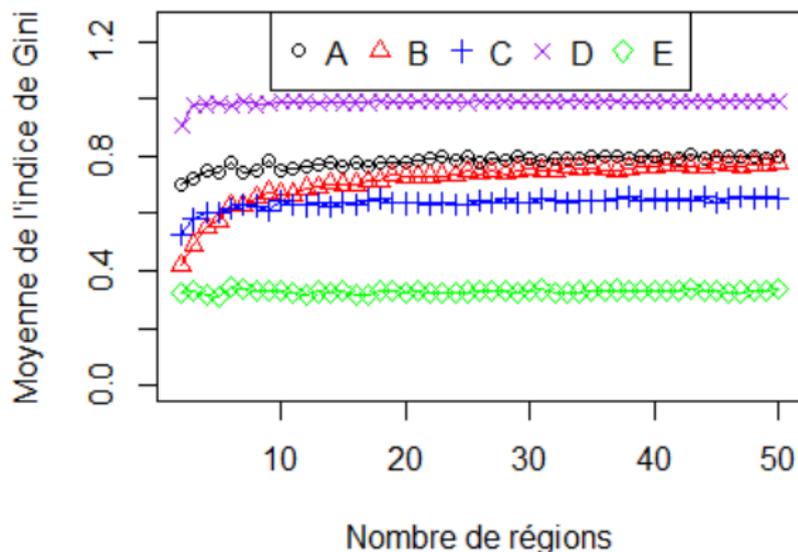


C

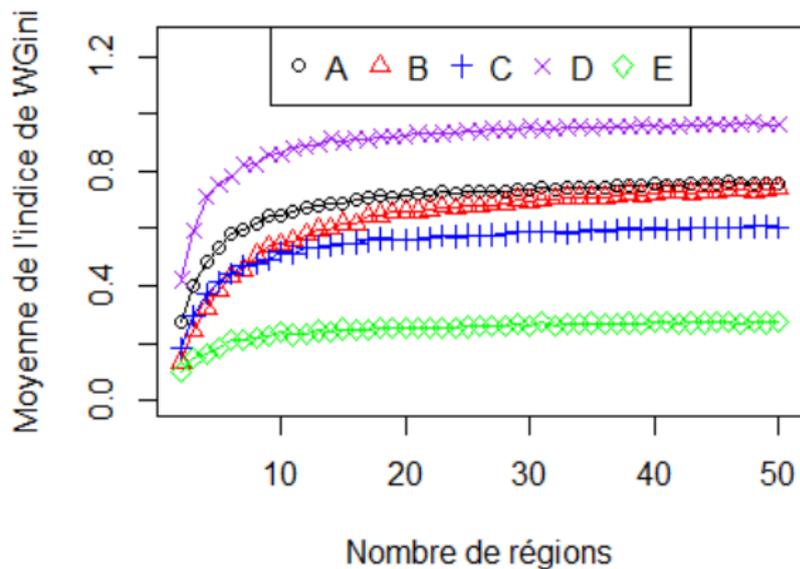
FIGURE – Exemples



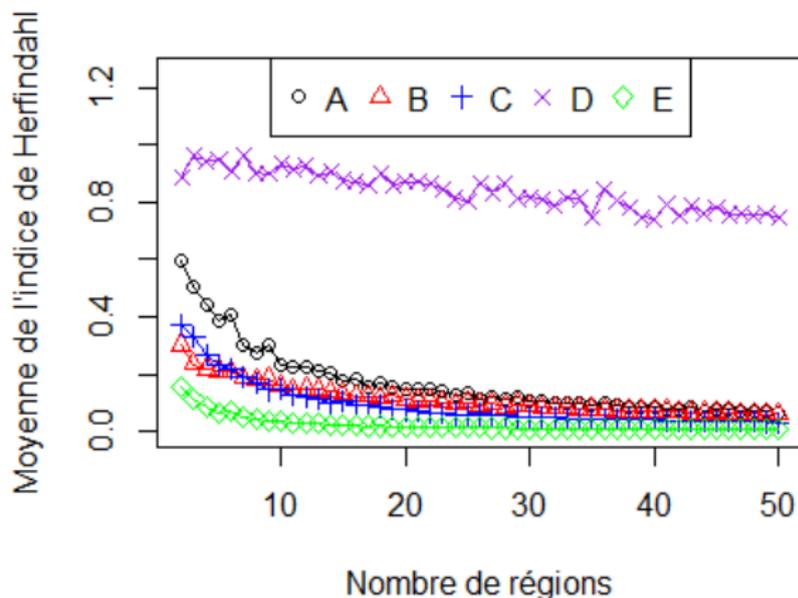
# Évolution de l'indice de Gini suite à l'augmentation du nombre de régions



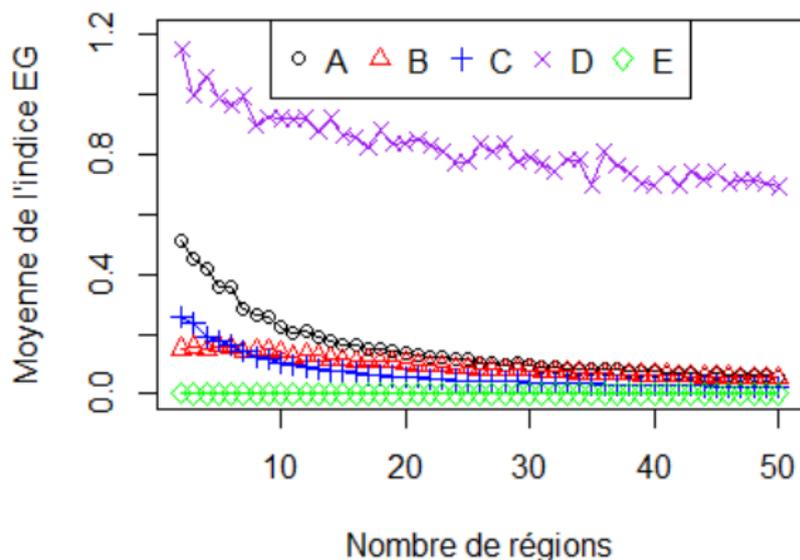
# Évolution de l'indice de WGini suite à l'augmentation du nombre de régions



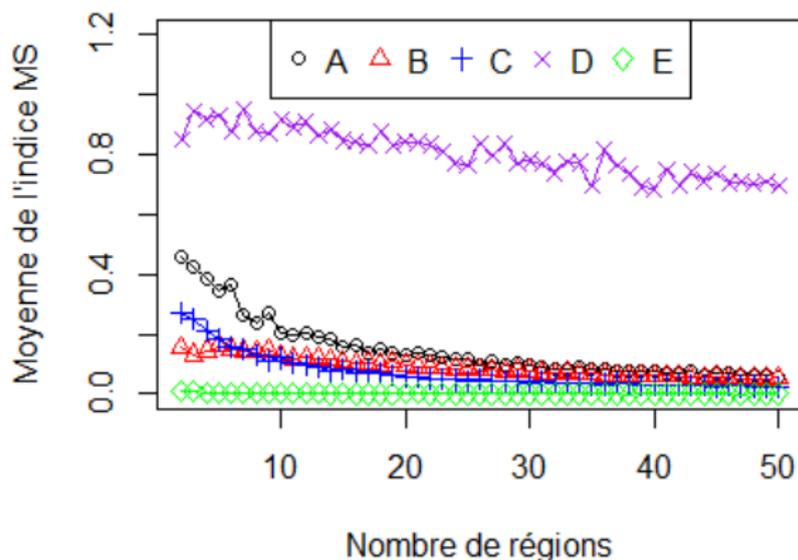
# Évolution de l'indice de Herfindahl suite à l'augmentation du nombre de régions



# Évolution de l'indice d'Ellison et Glaeser suite à l'augmentation du nombre de régions



# Évolution de l'indice de Maurel et Sédillot suite à l'augmentation du nombre de régions



# Propriétés d'un bon indice

- Doit corriger de l'agglomération de l'activité économique
- Doit corriger de la structure sectorielle
- Ne doit pas être sensible à une modification des frontières ou au choix de l'unité géographique
- Ne doit pas être sensible à une modification de la classification sectorielle
- L'hypothèse nulle de non concentration doit être testable significativement
- Doit permettre la comparaison entre les secteurs
- Doit permettre la comparaison entre des territoires différents
- Doit être facilement utilisable compte tenu des données disponibles

