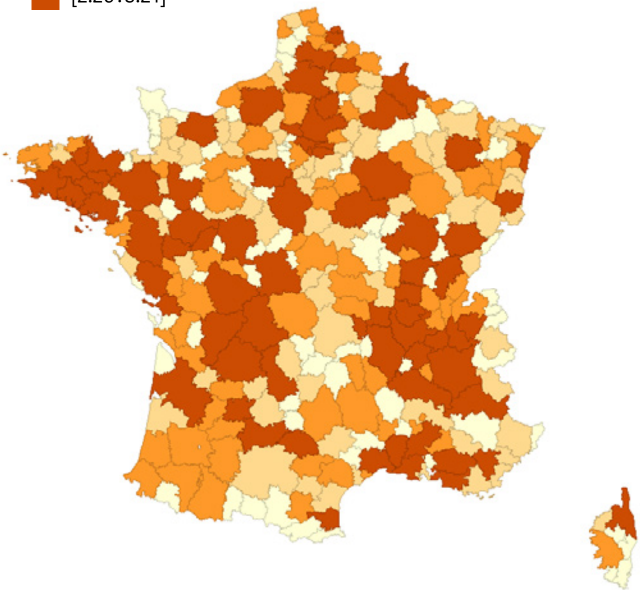
Une analyse critique d'indicateurs de spécialisation et de diversification sectorielle

Le quotient de localisation et les variétés non reliée et reliée

Distribution de la **variété non reliée** dans les zones d'emploi

Quantiles (4 classes de 76 zones d'emploi)

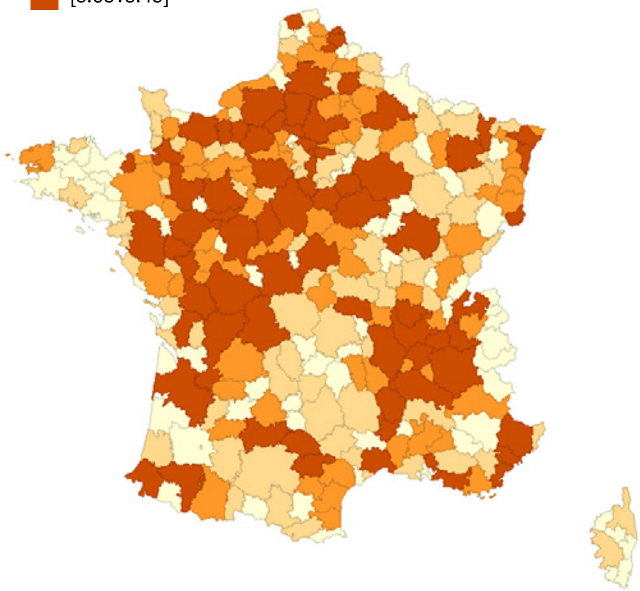
- [0.5 : 1.55]
- [1.55 : 1.9]
- [1.9 : 2.26]
- [2.26 : 3.21]



Distribution de la **variété reliée** dans les zones d'emploi

Quantiles (4 classes de 76 zones d'emploi)

- [1.27 : 2.49]
- [2.49 : 2.79]
- [2.79 : 3.03]
- [3.03 : 3.45]



Indicateurs calculés à partir des emplois salariés dans
les établissements relevant des 12 secteurs manufacturiers
Source : INSEE, DADS 2014

Auteurs : Olivier Pauly et Pierre Triboulet, UMR AGIR,
Université de Toulouse, INRAE, Castanet-Tolosan, France

INTRODUCTION

Pour mettre en évidence l'influence des externalités spatiales sur la performance à l'innovation des firmes (Galliano et al, 2015), il faut se poser les questions du choix des indicateurs pour mesurer ces externalités mais aussi de l'interprétation des valeurs fournies par ces indicateurs. Ceci nécessite de bien comprendre comment ils fonctionnent sur le plan mathématique et de caractériser leurs domaines de valeurs. Il faut également prendre en compte les interactions qui existent entre ces indicateurs. Ce sont ces questions auxquelles nous proposons de répondre dans ce document, en nous appuyant sur la nomenclature des activités françaises (NAF) qui permet de définir dans sa déclinaison la plus fine l'activité principale d'une entreprise (APE) ou d'un établissement (APET).

La littérature sur l'innovation a depuis longtemps montré l'importance de la dimension géographique dans les échanges de connaissance entre entreprises. Ainsi des regroupements spécialisés ou diversifiés d'entreprises peuvent constituer des environnements favorables au développement des innovations du fait que les connaissances circulent entre les entreprises, renvoyant aux travaux séminaux de Marshall (1890) et de Jacobs (1969). Pour Marshall, ce sont les regroupements d'entreprises d'un même secteur d'activités qui sont favorables aux échanges de connaissance et donc à l'innovation. Pour Jacobs, c'est plutôt la diversité des activités présentes sur un même territoire qui est source de nouvelles idées pour les entreprises. Le débat pour savoir si ce sont les externalités marshaliennes, dites de spécialisation, ou les externalités jacobiennes, dites de diversification, qui sont le plus propice à l'innovation reste largement ouvert (Beaudry and Schiffauerova, 2009[1] ; Galliano et al., 2015[2]), Plus récemment, le concept de « variété reliée » a été proposé pour mettre l'accent sur les activités présentant des proximités en termes de compétences et de savoir-faire (Frenken et al., 2007[3]). Cette proximité entre des activités diverses favoriserait le renouvellement des connaissances tout en limitant les risques d'une diversification vers des secteurs complétement nouveaux pour l'entreprise. Cette variété dite « reliée » vient donc compléter la vision « jacobienne » de la diversité qui elle, ne prend pas en compte la plus ou moins grande proximité existant entre les secteurs d'activités. Cette variété peut être qualifiée de « variété non reliée » (Frenken et al., 2007).

Pour mettre en évidence ces externalités spatiales, il faut disposer d'indicateurs permettant de décrire la spécialisation et la diversification sectorielle des activités au sein de zones géographiques. Pour la spécialisation, le quotient de localisation est l'indicateur le plus souvent retenu dans la littérature ; il permet de mesurer les externalités marshalliennes (Beaudry and Schiffauerova, 2009). Pour mesurer la diversification absolue versus relative des activités, les indicateurs de variété reliée et non reliée sont largement utilisés (Frenken et al., 2007).

1. Beaudry C. and Schiffauerova A., 2009. Who's right, Marshall or Jacobs? The localization versus urbanization debate. Research Policy, 38, 318-337.
2. Galliano D., Magrini M.-B., Triboulet P., 2015, Marshall's versus Jacobs' externalities in firm innovation performance: the case of French industry. Regional Studies, 49, 1840–1858.
3. Frenken K., Van Oort F., Verburg T., 2007. Related Variety, Unrelated Variety and Regional Economic Growth. Regional Studies, 41:5, 685-697.

Pour notre étude, nous avons donc choisi de mettre en lumière le fonctionnement de ces trois indicateurs :

QL : Quotient de Localisation
UV : Unrelated Variety (variété non reliée)
RV : Related Variety (variété reliée)

Le calcul des indicateurs est sensible à la classification sectorielle et au découpage géographique retenu (Beaudry and Schiffauerova, 2009). Nous n'abordons pas dans ce focus la question du découpage géographique. Pour la classification sectorielle, elle est mobilisée à deux niveaux emboités, le niveau des activités élémentaires et le niveau des secteurs d'activités. En première approche, il faut tenir compte que le quotient de localisation est calculé pour chaque secteur d'activité et que les variétés non reliée et reliée sont calculées de manière globale respectivement au niveau des secteurs d'activités et des activités élémentaires. De plus, les variétés reliée et non reliée peuvent être abordées comme contribuant à une mesure d'entropie globale des activités d'une zone géographique (Franken et al., 2007).

L'objectif de ce document est de préciser les modalités d'utilisation de ces indicateurs. Pour chacun d'eux, nous présentons tout d'abord la formule permettant de le calculer, puis nous décrivons les valeurs qu'il peut prendre. Enfin, nous présentons et discutons les interprétations que l'on peut en faire. Nous nous appuyons sur l'exemple de l'analyse de la spécialisation / diversification sectorielle de la main d'oeuvre dans l'industrie manufacturière, en considérant comme indiqué précédemment deux niveaux d'activités emboités : un niveau détaillé (activité élémentaire) et un niveau agrégé (secteur d'activités). Il s'agit pour le niveau détaillé de la Nomenclature d'Activités Française (NAF rév. 2) en 732 sous-classes et pour le niveau agrégé de la Nomenclature Agrégée de 2008 (NA – 2008) en 38 postes[4].

Ces indicateurs et cette méthodologie peuvent être utilisés dans d'autres domaines avec des données de nature différente.

Liste des abréviations / sigles utilisés
- **A38 : Nomenclature Agrégée en 38 postes**
- **A_j : Activité élémentaire indicée «j»**
- **APE : Activité Principale Exercée**
- **QL : Quotient de Localisation**
- **RV : Related Variety (variété reliée)**
- **S_k : Secteur (niveau d'activité agrégé) indicé «k»**
- **UV : Unrelated Variety (variété non reliée)**
- **Z_i : Zone / Unité spatiale indicée «i»**

Remarque : Toute activité appartient à un seul secteur, et tout secteur contient au moins une activité.

4. Voir sur le site INSEE : https://www.insee.fr/fr/information/2016811

Notations mathématiques

- = : Egal à / A peu près égal à (pas de distinction entre égalité stricte et approchée)
- $\forall$: Quel que soit ...
- $\in$: Appartient à ...
- $\sum$: Somme mathématique
- $\log_2$: Fonction logarithme en base 2
- $\boldsymbol{eff_i}$: Total des effectifs salariés de la zone Z_i, tous secteurs confondus
- $\boldsymbol{eff_k}$: Total des effectifs salariés du secteur S_k, toutes zones confondues
- $\boldsymbol{eff_{k,i}}$: Total des effectifs salariés du secteur S_k dans la zone Z_i
- $\boldsymbol{eff_{j,i}}$: Total des effectifs salariés de l'activité A_j dans la zone Z_i
- $\boldsymbol{eff_{TOT}}$: Total des effectifs salariés, toutes activités et zones confondues
- $\boldsymbol{p_k = eff_k \,/\, eff_{TOT}}$: Part des salariés du secteur S_k, toutes zones confondues
- $\boldsymbol{p_{k,i} = eff_{k,i} \,/\, eff_i}$: Part des salariés du secteur S_k dans la zone Z_i
- $\boldsymbol{p_{j,i} = eff_{j,i} \,/\, eff_i}$: Part des salariés de l'activité A_j dans la zone Z_i
- $\boldsymbol{p_{j,k,i} = eff_{j,i} \,/\, eff_{k,i}}$: Part des salariés de l'activité A_j, au sein des salariés du secteur S_k, dans la zone Z_i

VALEURS ET INTERPRÉTATIONS DU QUOTIENT DE LOCALISATION (QL)

Formule du quotient de localisation

Le **QL**, calculé pour un secteur dans une zone, permet d'apprécier la spécialisation de la zone dans ce secteur. Pour le secteur $\mathbf{S_k}$ dans la zone $\mathbf{Z_i}$, il se calcule de la manière suivante :

$$QL_{Z_i}^{S_k} = \frac{eff_{k,i} \,/\, eff_i}{eff_k \,/\, eff_{TOT}} \quad \rightarrow \quad \boldsymbol{QL_{k,i} = \frac{p_{k,i}}{p_k}}$$

Valeurs minimale et maximale pour le QL

Pour un secteur $\mathbf{S_k}$, seul $\mathbf{p_{k,i}}$ varie d'une zone à l'autre ($\mathbf{p_k}$ ne dépend pas de la zone). On observe trois valeurs particulières pour $\mathbf{QL_{k,i}}$ en fonction de celle de $\mathbf{p_{k,i}}$:

$p_{k,i} = 0 \quad \leftrightarrow \quad QL_{k,i} = 0$ (valeur minimale)

$p_{k,i} = p_k \quad \leftrightarrow \quad QL_{k,i} = 1$ (valeur «neutre»)

$p_{k,i} = 1 \quad \leftrightarrow \quad QL_{k,i} = \frac{1}{p_k} = p_k^{-1}$ (valeur maximale)

On note que la valeur maximale pour $\mathbf{QL_{k,i}}$ dépend de la valeur de $\mathbf{p_k}$. Par exemple :

➜ $\boldsymbol{p_l = 0.5}$ ➜ Quelle que soit la zone -Z_i, $\boldsymbol{QL_{l,i}} \in [\mathbf{0, 2}]$ (car $0.5^{-1} = 2$)

➜ $\boldsymbol{p_m = 0.1}$ ➜ Quelle que soit la zone -Z_i, $\boldsymbol{QL_{m,i}} \in [\mathbf{0, 10}]$ (car $0.1^{-1} = 10$)

$\mathbf{QL_{k,i}}$ prend donc ses valeurs dans l'intervalle $[\mathbf{0}\,, \boldsymbol{p_k^{-1}}]$

Interprétation des valeurs du QL

Les valeurs du $\mathbf{QL_{k,i}}$ s'interprètent de manière assez intuitive, relativement à $\mathbf{p_k}$:

$QL_{k,i} < 1 \quad \leftrightarrow \quad p_{k,i} < p_k \quad \leftrightarrow$ Secteur S_k sous-représenté dans la zone Z_i.

$QL_{k,i} = 1 \quad \leftrightarrow \quad p_{k,i} = p_k \quad \leftrightarrow$ Secteur S_k «à l'identique» dans la zone Z_i.

$QL_{k,i} > 1 \quad \leftrightarrow \quad p_{k,i} > p_k \quad \leftrightarrow$ Secteur S_k sur-représenté dans la zone Z_i.

BILAN SUR LE QL

La valeur du $\mathbf{QL_{k,i}}$, relativement à celle de $\mathbf{p_k}$, s'interprète de la manière suivante :

- Plus $\mathbf{QL_{k,i}}$ est proche de $\mathbf{p_k^{-1}}$, plus le secteur $\mathbf{S_k}$ est sur-représenté dans la zone $\mathbf{Z_i}$.
- Plus $\mathbf{QL_{k,i}}$ est proche de **1**, plus la part du secteur $\mathbf{S_k}$ dans la zone $\mathbf{Z_i}$ est proche de sa part totale (calculée dans l'ensemble des zones).
- Plus $\mathbf{QL_{k,i}}$ est proche de **0**, plus le secteur $\mathbf{S_k}$ est sous-représenté dans la $\mathbf{Z_i}$.

Par contre, la valeur maximale pour $\mathbf{QL_{k,i}}$ dépendant de la valeur de $\mathbf{p_k}$, il n'est pas prudent de comparer les valeurs de deux **QL** associés à des secteurs différents, sans prendre en compte les parts totales respectives de chacun de ces secteurs dans l'ensemble de la population étudiée (i.e. toutes zones confondues).

VALEURS ET INTERPRÉTATIONS DE LA VARIÉTÉ NON RELIÉE (UV)

Formule de la variété non reliée

L'**UV**, calculé pour une zone, permet d'apprécier la répartition de la main d'œuvre dans les différents secteurs qui la composent. On parle également d'entropie au niveau des secteurs dans la zone. Pour la zone $\mathbf{Z_i}$, il se calcule de la manière suivante :

$$UV_{Z_i} = \sum_{S_k \in Z_i} \left[\frac{eff_{k,i}}{eff_i} \times log_2\left(\frac{eff_i}{eff_{k,i}}\right) \right] \rightarrow \boldsymbol{UV_i} = \sum_{\boldsymbol{S_k \in Z_i}} [\, \boldsymbol{p_{k,i} \times log_2(p_{k,i}^{-1})}]$$

Remarque : La mention « $S_k \in Z_i$ » sous la somme indique que seuls les secteurs représentés dans la zone $\mathbf{Z_i}$ sont considérés.

Valeurs minimale et maximale pour l'UV

Cas d'un seul secteur représenté

Si un seul secteur $\mathbf{S_1}$ est présent dans la zone $\mathbf{Z_i}$, alors il n'y a qu'un élément dans la somme :

$$p_{1,i} = 1 \Rightarrow log_2(p_{1,i}^{-1}) = log_2(1) = 0 \Rightarrow UV_i = p_{1,i} * log_2(p_{1,i}^{-1}) = 0$$

Cas de K secteurs représentés de manière uniforme

Si **K** secteurs sont présents de manière uniforme dans la zone $\mathbf{Z_i}$ (i.e. $p_{k,i} = \frac{1}{K}, \forall k$), alors on a **K** éléments dans la somme, ce qui donne :

- $\Rightarrow p_{k,i} = \frac{1}{K} \quad \forall k \in \{1,2,\dots,K\}$
- $\Rightarrow log_2(p_{k,i}^{-1}) = log_2(K) \quad \forall k \in \{1,2,\dots,K\}$
- $\Rightarrow p_{k,i} \times log_2(p_{k,i}^{-1}) = \frac{1}{K} * log_2(K) \quad \forall k \in \{1,2,\dots,K\}$
- $\Rightarrow UV_i = \sum_k [\, p_{k,i} \times log_2(p_{k,i}^{-1})] = \sum_k \left[\, \frac{1}{K} * log_2(K)\right] = K * \left[\frac{1}{K} * log_2(K)\right] = log_2(K)$

Valeurs possibles

La valeur minimale pour l'UV est 0. Pour le démontrer, il suffit de montrer que **0** est un minorant ET une valeur possible. « C'est une valeur possible » a déjà été démontrée (lorsqu'un seul secteur est présent dans la zone). Pour montrer que c'est un minorant, on montre que chaque élément présent dans la somme est positif ou nul :

$$p_{k,i} \in]0,1] \Rightarrow p_{k,i}^{-1} \geq 1 \Rightarrow log_2(p_{k,i}^{-1}) \geq 0 \Rightarrow p_{k,i} * log_2(p_{k,i}^{-1}) \geq 0 \;\; \forall k \;\in \{1,2,\dots,K\}$$

La valeur maximale pour l'UV est $\boldsymbol{log_2(K)}$ (**K** étant le nombre de secteurs présents dans la zone). La démonstration ne sera pas faite ici car elle nécessite des outils mathématiques assez lourds. Pour le prouver, il faudrait résoudre le problème d'optimisation suivant :

Maximiser la somme $\sum_k [\, p_{k,i} \times log_2(p_{k,i}^{-1})]$ sous la contrainte $\begin{cases} p_{k,i} \in]0,1] \; \forall k \in \{1,2,\dots,K\} \\ \sum_k p_{k,i} = 1 \end{cases}$

$\boldsymbol{UV_i}$ **prend donc ses valeurs dans l'intervalle** $[0\,,\boldsymbol{log_2}(K)]$**.**

Exemples de valeurs pour l'UV et interprétations

Le tableau suivant permet d'apprécier la valeur de l'**UV** en fonction de la répartition de la main d'œuvre dans 12 secteurs désignés par **A**, **B**, …, **K** et **L**. Il s'agit de cas fictifs.

	A	B	C	D	E	F	G	H	I	J	K	L	UV
Cas 1	100,00%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	**0,0000**
Cas 2	80,00%	13,33%	6,67%	0%	0%	0%	0%	0%	0%	0%	0%	0%	**0,9056**
Cas 3	62,50%	37,50%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	**0,9544**
Cas 4	50,00%	50,00%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	**1,0000**
Cas 5	72,00%	16,00%	12,00%	0%	0%	0%	0%	0%	0%	0%	0%	0%	**1,1313**
Cas 6	41,82%	27,27%	18,18%	12,73%	0%	0%	0%	0%	0%	0%	0%	0%	**1,8629**
Cas 7	25,00%	25,00%	25,00%	25,00%	0%	0%	0%	0%	0%	0%	0%	0%	**2,0000**
Cas 8	35,71%	30,61%	15,31%	10,20%	5,10%	3,06%	0%	0%	0%	0%	0%	0%	**2,1768**
Cas 9	21,02%	19,89%	17,05%	11,36%	8,52%	7,39%	6,25%	5,68%	2,84%	0%	0%	0%	**2,9395**
Cas 10	12,50%	12,50%	12,50%	12,50%	12,50%	12,50%	12,50%	12,50%	0%	0%	0%	0%	**3,0000**
Cas 11	10,00%	10,00%	10,00%	10,00%	10,00%	10,00%	10,00%	10,00%	10,00%	10,00%	0%	0%	**3,3219**
Cas 12	8,33%	8,33%	8,33%	8,33%	8,33%	8,33%	8,33%	8,33%	8,33%	8,33%	8,33%	8,33%	**3,5850**

Tableau 1 - Exemples de valeurs d'UV selon la répartition de la main d'œuvre dans 12 secteurs

Lecture : le cas 2 montre une répartition de la main d'œuvre dans 3 secteurs et une valeur de l'**UV** de 0,9056. Le cas 3 présente un **UV** de 0,9544 avec une répartition de la main d'œuvre dans 2 secteurs.

On voit dans ce tableau un lien positif entre la valeur de l'**UV** et le nombre de secteurs représentés. De plus, à nombre de secteurs égal, plus la main d'œuvre se répartit de manière uniforme, plus l'**UV** est fort (comparaison des cas 3-4, 6-7 ou encore, 2-5).

Concernant les valeurs particulières, on observe ce qui a déjà été dit :

- Cas 1 : un seul secteur représenté ➜ **UV = 0**
- Cas 4, 7, 10, 11 et 12 : **K** secteurs représentés de manière uniforme ➜ **UV = $\log_2(K)$**

Concernant la valeur 1, elle ne présente pas de sens particulier pour l'**UV** (pour le **QL**, elle traduisait une part dans la zone équivalente à la part globale). Cette valeur est atteinte lorsque les effectifs salariés se répartissent de manière uniforme entre 2 secteurs (cas 4 dans le tableau). Ceci est dû à l'utilisation du logarithme en base 2 dans la formule (l'utilisation du logarithme en base 10 entrainerait une valeur de 1 lorsque les effectifs salariés se répartissent de manière uniforme entre 10 secteurs). Précisons qu'un changement de base pour le logarithme employé dans la formule de l'**UV** ne fait que modifier l'échelle de l'indicateur de manière proportionnelle (ce résultat est également valable pour le **RV**).

BILAN SUR L'UV

La valeur de l'UV_i dans la zone Z_i comportant K secteurs s'interprète de la manière suivante :

- **Plus l'UV_i est proche de $\log_2(K)$, plus la zone Z_i est diversifiée dans les K secteurs.**
- **Plus l'UV_i se rapproche de 0, moins la zone Z_i est diversifiée.**
- **Plus le nombre de secteurs est important, plus la valeur maximale théorique de l'UV_i est élevée.**

VALEURS ET INTERPRÉTATION DE LA VARIÉTÉ RELIÉE (RV)

Formule de la variété reliée

Le **RV**, calculé pour une zone, permet de tenir compte de la diversification des activités au sein de chaque secteur. Deux paramètres vont jouer sur la valeur du **RV**, tout d'abord le nombre d'activités pour chaque secteur et ensuite la manière dont les effectifs salariés s'y répartissent. Pour la zone $\mathbf{Z_i}$, il se calcule de la manière suivante :

$$RV_{Z_i} = \sum_{S_k \subset Z_i} \left[\frac{eff_{k,i}}{eff_i} \times \sum_{A_j \subset S_k * Z_i} \left[\frac{eff_{j,i}}{eff_{k,i}} \times log_2\left(\frac{eff_{k,i}}{eff_{j,i}}\right)\right]\right]$$

$$\rightarrow \boldsymbol{RV_i} = \sum_{S_k \in Z_i}\left[\boldsymbol{p_{k,i}} \times \sum_{A_j \in S_k * Z_i}\left[\boldsymbol{p_{j,k,i}} \times \boldsymbol{log_2}\left(\boldsymbol{p_{j,k,i}^{-1}}\right)\right]\right] = \sum_{S_k \in Z_i}\left[\boldsymbol{p_{k,i}} \times \boldsymbol{H_{k,i}}\right]$$

$$\text{avec } \; H_{k,i} = \sum_{A_j \subset S_k * Z_i}\left[p_{j,k,i} \times log_2(p_{j,k,i}^{-1})\right]$$

Remarque 1 : La mention « $S_k \in Z_i$ » sous la première somme indique que seuls les secteurs représentés dans la zone $\mathbf{Z_i}$ y sont considérés. La mention « $A_j \in S_k * Z_i$ » sous la seconde somme indique que seules les activités appartenant au secteur $\mathbf{S_k}$ et représentées dans la zone $\mathbf{Z_i}$ sont considérées.

Remarque 2 : $\mathbf{H_{k,i}}$ est une mesure d'entropie de la zone **i** similaire à $\mathbf{UV_i}$. Le premier mesure l'entropie au niveau des activités au sein d'un secteur. Le second mesure l'entropie au niveau des secteurs. La littérature a montré que l'entropie pouvait être décomposée (Jacquemin and Berry, 1979[5]). Ainsi, l'entropie $\mathbf{E_i}$ mesurée au niveau élémentaire des activités, peut être décomposée de la manière suivante : $\mathbf{E_i = UV_i + RV_i = UV_i + somme\,(H_{k,i})}$. Voir aussi encadré « Approfondissement de la notion d'entropie ».

$\mathbf{H_{k,i}}$ sera désigné dans la suite du document d'**entropie du secteur $\mathbf{S_k}$ dans la zone $\mathbf{Z_i}$.**

SOMME VERSUS MOYENNE

La formule ci-dessus décrit le $\mathbf{RV_i}$ comme la <u>somme</u> des $\mathbf{H_{k,i}}$ pondérés par les $\mathbf{p_{k,i}}$. Cette somme correspond également à la <u>moyenne</u> des $\mathbf{H_{k,i}}$ pondérés par les $\mathbf{p_{k,i}}$. En effet, la moyenne pondérée vaut la somme pondérée divisée par la somme des poids. En développant la formule de la moyenne des $\mathbf{H_{k,i}}$ pondérée par les $\mathbf{p_{k,i}}$, on observe qu'on retombe bien sur la formule du $\mathbf{RV_i}$ (puisque la somme des poids $\mathbf{p_{k,i}}$ vaut 1) :

$$\frac{\textit{Somme pondérée}}{\textit{Somme des poids}} = \frac{\sum_{S_k \in Z_i}\left[p_{k,i} \times H_{k,i}\right]}{\sum_{S_k \in Z_i}\left[p_{k,i}\right]} = \frac{\sum_{S_k \in Z_i}\left[p_{k,i} \times H_{k,i}\right]}{1} = \sum_{S_k \subset Z_i}\left[p_{k,i} \times H_{k,i}\right] = RV_i$$

[5] Jacquemin A.P. and Berry C.H., 1979. Entropy measure of diversification and corporate growth. Journal of Industrial Economics, 27, 359-369.

Il y a deux intérêts à qualifier le RV$_i$ de moyenne pondérée des H$_{k,i}$:

1. **Une moyenne est du même ordre de grandeur que les éléments utilisés pour la calculer, quels que soient le nombre d'éléments et l'ordre de grandeur des poids utilisés. De ce fait, qualifier le RV$_i$ de moyenne pondérée des H$_{k,i}$ permet d'apprécier d'ores et déjà que son ordre de grandeur ne dépend pas du nombre de secteurs présents dans la zone Z$_i$ (i.e. du nombre de H$_{k,i}$ présents dans la somme), mais de l'entropie au sein de chaque secteur présent dans la zone (i.e. de la valeur des H$_{k,i}$ présents dans la somme).**
2. **Une moyenne est bornée par le minimum et le maximum des valeurs utilisées pour la calculer. Cette propriété facilite l'analyse numérique du RV.**

Ainsi, considérer le RV$_i$ comme la moyenne pondérée des H$_{k,i}$ permet de mieux apprécier son ordre de grandeur et de faciliter son analyse numérique.

Valeurs minimale et maximale pour le RV

a. Valeurs pour les H$_{k,i}$

Pour chaque **H$_{k,i}$**, on peut affirmer les mêmes propriétés que celles établies pour l'**UV$_i$** (de par leurs structures mathématiques similaires) :

- **La valeur minimum du H$_{k,i}$ est 0.** Ceci arrive lorsqu'une seule activité du secteur **S$_k$** est représentée dans la zone **Z$_i$**.
- **La valeur maximum du H$_{k,i}$ est log$_2$(n$_k$)**, où **n$_k$** est le nombre d'activités qui composent le secteur **S$_k$**. Ceci arrive lorsque les **n$_k$** activités du secteur **S$_k$** sont représentées de manière uniforme dans la zone **Z$_i$**.

b. Valeurs pour le RV

Pour déterminer les valeurs minimum et maximum du **RV**, le raisonnement sera le même que pour l'**UV** : si une valeur est à la fois un minorant (respectivement un majorant) et une valeur possible, alors, c'est la valeur minimum (respectivement maximum).

Valeur minimum du RV$_i$

- **Minorant** : $H_{k,i} \geq 0,\ \forall\, k \quad \rightarrow \quad RV_i = moyenne_k(H_{k,i}) \geq 0$
- **Valeur possible** : Si dans la zone **Z$_i$**, un seul secteur est représenté (➡ un seul **H$_{k,i}$** dans la somme) et au sein de ce secteur, une seule activité est représentée (➡ **H$_{k,i}$ = 0**) alors, le **RV$_i$** vaut **0** (car il est la moyenne d'une seule valeur valant 0).
- **BILAN** : **0** est un minorant et une valeur possible, c'est donc la valeur minimum.

Valeur maximum du RV$_i$

Cette partie nécessite l'introduction de trois notations :

- **K** : nombre de secteurs représentés dans la zone **Z$_i$**
- **n$_k$** : nombre d'activités pour le secteur **S$_k$** $(k \in \{1,2,\ldots,K\})$
- **n$_{max}$** : nombre d'activités du secteur ayant le plus grand nombre d'activités. En d'autres termes, **n$_{max}$ = max (n$_1$, n$_2$, …, n$_k$)**

- **Majorant** :
$H_{k,i} \leq log(n_k) \leq log(n_{max}), \forall k \;\rightarrow\; RV_i = moyenne_k(H_{k,i}) \leq log(n_{max})$
- **Valeur possible** : Si dans la zone $\mathbf{Z_i}$, un seul secteur est représenté (➡ un seul $\mathbf{H_{k,i}}$ dans la somme) et ce secteur est composé de $\mathbf{n_{max}}$ activités qui se partagent les effectifs salariés de manière uniforme (➡ $\mathbf{H_{k,i} = log(n_{max})}$) alors, le $\mathbf{RV_i}$ vaut $\mathbf{log(n_{max})}$ (car il est la moyenne d'une seule valeur valant $\mathbf{log(n_{max})}$).
- **BILAN** : $\mathbf{log(n_{max})}$ est un majorant et une valeur possible, c'est donc la valeur maximum.

RV_i prend donc ses valeurs dans l'intervalle $[0, log_2(n_{max})]$.

Exemples de valeurs pour le RV et interprétations

On rappelle que dans ce document, au sein d'une zone $\mathbf{Z_i}$, on définit **l'entropie du secteur S_k** comme la manière dont se répartit la main d'œuvre entre ses différentes activités. L'entropie du secteur $\mathbf{S_k}$ dans la zone $\mathbf{Z_i}$ est mesurée par $\mathbf{H_{k,i}}$.

Nous montrons et discutons des valeurs possibles du **RV** en utilisant pour les secteurs, les 12 postes relevant de l'industrie manufacturière dans la nomenclature agrégée d'activités en 38 postes (A38), et pour les activités, l'APE (Activité Principale Exercée) relevant du niveau 5 de la nomenclature d'activité française NAF rèv. 2. Le tableau suivant liste ces 12 secteurs, le nombre d'APE qui les composent et les valeurs maximales que peuvent prendre leurs $\mathbf{H_{k,i}}$ respectifs :

Secteur	CA	CB	CC	CE	CF	CG	CH	CI	CJ	CK	CL	CM
Nombre d'APE	45	21	21	17	2	31	37	11	10	23	12	27
$H_{k,i}$ max	5,49	4,39	4,39	4,09	1	4,95	5,21	3,46	3,32	4,52	3,58	4,75

Tableau 2 - Postes A38 de l'industrie manufacturière, nombre d'APE et $H_{k,i}$ maximum

Ce tableau permet de constater l'influence du nombre d'APE sur la valeur maximale théorique de la mesure d'entropie du fait de la forte hétérogénéité existant dans la classification élémentaire des activités (de 2 à 45 APE). Si l'on exclue le secteur très particulier CF (composé de seulement 2 APE), on voit que le $\mathbf{H_{k,i}}$ maximum par secteur varie entre 3.32 (secteur CJ, 10 APE) et 5.49 (secteur CA, 45 APE). Un secteur atteint son entropie maximale quand les effectifs se répartissent de manière uniforme entre toutes les APE qui le composent. Et le **RV** maximum serait atteint si toute la main d'oeuvre était concentrée dans le secteur CA (et répartie de manière uniforme dans les 45 APE qui le composent).

De fait, les valeurs prises par le RV vont dépendre d'un ensemble de paramètres (valeur maximale possible pour les $\mathbf{H_{k,i}}$, répartition des activités au sein de chaque secteur, poids relatif des secteurs, ...). Nous pouvons donner quelques exemples permettant de mieux comprendre l'influence de ces paramètres sur les valeurs prises par le **RV**.

Le tableau suivant propose différentes configurations de zones, en indiquant dans chaque cas, le poids et le $\mathbf{H_{k,i}}$ associés à chaque secteur A38, et le RV qui en découle. En tête de tableau, le $\mathbf{H_{k,i}}$ maximum de chaque secteur est rappelé (colonne «$\mathbf{H_{k,i}}$ max») afin d'apprécier chaque mesure d'entropie par rapport à sa valeur maximale théorique.

	Secteur	CA	CB	CC	CE	CF	CG	CH	CI	CJ	CK	CL	CM	RV
	$H_{k,i}$ max	5,49	4,39	4,39	4,09	1	4,95	5,21	3,46	3,32	4,52	3,58	4,75	
Config. 1	Poids	0%	0%	0%	0%	0%	0%	0%	0%	80%	20%	0%	0%	1,22
	$H_{k,i}$	-	-	-	-	-	-	-	-	1,1	1,7	-	-	
Config. 2	Poids	20%	5%	10%	10%	20%	0%	0%	5%	10%	0%	0%	20%	1,22
	$H_{k,i}$	1,2	2	1,3	2	0,7	-	-	1,8	1,8	-	-	0,7	
Config. 3	Poids	15%	10%	20%	5%	10%	0%	10%	0%	10%	10%	0%	10%	1,24
	$H_{k,i}$	4	1,1	0,9	1,2	0	-	1	-	0	0,9	-	1	
Config. 4	Poids	10%	10%	10%	10%	0%	10%	10%	10%	10%	10%	5%	5%	2,65
	$H_{k,i}$	2,8	3	1,9	2,7	-	3	2,1	2,6	2,8	3	3	2,2	
Config. 5	Poids	25%	25%	0%	0%	25%	0%	0%	0%	25%	0%	0%	0%	2,3
	$H_{k,i}$	3,7	3,4	-	-	0,9	-	-	-	1,2	-	-	-	
Config. 6	Poids	50%	0%	0%	0%	0%	50%	0%	0%	0%	0%	0%	0%	3,3
	$H_{k,i}$	3,4	-	-	-	-	3,2	-	-	-	-	-	-	
Config. 7	Poids	10%	10%	10%	10%	0%	10%	10%	10%	0%	10%	10%	10%	3,61
	$H_{k,i}$	4	3,6	4,1	3,9	-	3,5	4	3,2	-	3,6	3,2	3	
Config. 8	Poids	25%	20%	20%	0%	10%	10%	5%	10%	0%	0%	0%	0%	3,16
	$H_{k,i}$	4,5	3,5	3,6	-	0	3,7	3,3	0,8	-	-	-	-	

Tableau 3 - Exemples de valeurs de RV, calculés sur la base des 12 postes A38 du secteur manufacturier, à partir de leurs poids et de leurs mesures d'entropie

Ce tableau permet d'illustrer les propriétés du **RV**.

Tout d'abord, on note que la valeur du **RV** ne dépend pas du nombre de secteurs représentés dans une zone mais bien de la valeur des $\mathbf{H_{k,i}}$, c'est-à-dire de l'entropie mesurée au sein de chacun des secteurs représentés dans une zone.

RV faibles (config 1 à 3)

3 configurations sont présentées : les deux premières présentent des entropies faibles ($\mathbf{H_{k,i}}$) avec deux secteurs représentés (config 1) et 8 secteurs représentés (config 2). Le **RV** qui est la moyenne des entropies par secteur est donc faible. Dans la config 3, un secteur présente une entropie élevée, le secteur CA avec un poids de 15%. Les autres entropies étant faibles, le **RV** calculé présente une valeur faible.

RV intermédiaires (config 4 et 5)

La configuration 4 présente des entropies intermédiaires pour 11 secteurs. Il en résulte un **RV** du même ordre de grandeur. La configuration 5 permet de montrer qu'un **RV** intermédiaire peut résulter d'une configuration avec des entropies fortes pour certains secteurs et faibles pour d'autres. De ce fait, le **RV** associé présente une valeur intermédiaire, représentative d'aucun des quatre secteurs (ceci est bien entendu un cas extrême, mais intéressant à relever).

RV élevés (config 6 à 8)

3 configurations sont présentées : les deux premières présentent des entropies élevées avec deux secteurs représentés (config 6) et 10 secteurs représentés (config 7). Le **RV** résultant est donc un **RV** fort. La configuration 8 met en avant le fait qu'une zone présentant un **RV** élevé peut voir un ou plusieurs de ses secteurs présenter une faible entropie (secteurs CF et CI avec un poids cumulé de 10 %). En effet, les autres secteurs, disposant d'une forte entropie, représentent un poids cumulé de 90 %. Le **RV** calculé est donc forcément plus proche des valeurs hautes que des valeurs faibles.

Ainsi, contrairement aux **RV** faibles qui peuvent être atteints quels que soient les secteurs représentés, les **RV** élevés ne seront atteints que si les secteurs fortement représentés dans la zone sont ceux regroupant un grand nombre d'activités et avec une répartition homogène des emplois dans les différentes activités.

En conclusion, la valeur du **RV** reste complexe à appréhender car :

- La valeur maximale du **RV** traduit un cas très spécifique, dans lequel seul le secteur comprenant le plus d'activités est présent (et au sein duquel l'entropie est maximale).
- Le nombre de secteurs présents dans une zone n'a pas d'impact sur la valeur du **RV**. Ce sont les mesures d'entropie au sein des secteurs représentés qui sont importantes.
- Le **RV** est une moyenne, de ce fait, sa valeur peut être proche ou éloignée des $\mathbf{H}_{k,i}$ utilisés pour le calculer.

BILAN SUR LE RV

Le $\mathbf{RV}_i$ mesure l'entropie moyenne des secteurs représentés dans la zone $\mathbf{Z}_i$. Ses valeurs s'interprètent de la manière suivante :

- **Plus la valeur du $\mathbf{RV}_i$ est élevée, plus les secteurs représentés dans la zone $\mathbf{Z}_i$ présentent (en moyenne) une entropie élevée.**
- **Plus la valeur du $\mathbf{RV}_i$ est faible, plus les secteurs représentés dans la zone $\mathbf{Z}_i$ présentent (en moyenne) une entropie faible.**

CONCLUSION

Ce document permet de formaliser l'interprétation que l'on peut faire des valeurs numériques proposées par les trois indicateurs **QL**, **UV** et **RV**, et des points de vigilance à prendre en compte.

Concernant le quotient de localisation (**QL**) associé à un secteur k dans une zone, nous avons montré qu'il est borné par $\mathbf{p_k^{-1}}$ (l'inverse du pourcentage de main d'œuvre global dans le secteur **k**) et que cette valeur varie d'un secteur à l'autre. Il permet de classer les zones géographiques selon leur degré de spécialisation dans ce secteur et les valeurs les plus élevées du **QL** correspondent aux zones géographiques présentant les plus fort potentiels d'externalités spatiales. Cependant, comme les valeurs les plus élevées du **QL** dépendent des caractéristiques du secteur, il faut être prudent dans l'interprétation des valeurs de **QL** quand on compare les **QL** entre différents secteurs. En effet, les secteurs ayant un faible poids au niveau global présentent des bornes supérieures plus faibles que celles associées aux secteurs ayant un poids plus important au niveau global

La variété non reliée (**UV**) visant à mesurer l'entropie des secteurs d'activité dans une zone, ne présente quant à elle pas de difficulté majeure d'interprétation. Dans chaque zone, elle est bornée par $\mathbf{\log_2(K)}$ (**K** étant le nombre de secteurs d'activité représentés dans la zone). Plus ces secteurs d'activité se répartissent de manière uniforme, plus l'**UV** associé se rapprochera de sa valeur maximale théorique. Des valeurs fortes suggèrent donc une répartition équilibrée entre un nombre important de secteurs, tandis que des valeurs faibles mettront en évidence des zones dans lesquelles la main d'œuvre se concentre dans un nombre restreint de secteurs.

La variété reliée (**RV**) est probablement l'indicateur le plus complexe à apprécier. Il permet de mesurer l'entropie moyenne au sein de chaque secteur d'activité (les $\mathbf{H_{k,i}}$ qui mesurent la répartition de la main d'œuvre dans les activités du secteur **k** dans la zone **i**). La difficulté d'interprétation majeure dans notre cadre d'analyse est lié au nombre variable d'activités en lequel se décompose chaque secteur. En effet, la valeur maximale théorique d'un $\mathbf{H_{k,i}}$ augmente avec le nombre d'activités qui composent le secteur **k** (et vaut $\mathbf{\log_2(n_k)}$, $\mathbf{n_k}$ étant le nombre d'activité qui composent le secteur **k**). De ce fait, une zone avec des secteurs comportant un faible nombre d'activités présentera un **RV** probablement plus faible que celui associé à une zone avec des secteurs regroupant un grand nombre d'activités. Cette difficulté pourrait cependant être contournée en regroupant les secteurs d'activités disposant de peu d'activités avec d'autres secteurs (en particulier le secteur CF dans notre cas qui ne regroupe que deux activités). Cela permettrait de disposer de secteurs composés d'un nombre d'activités comparable, et donc de $\mathbf{H_{k,i}}$ variant selon des échelles similaires quel que soit le secteur.

Approfondissement de la notion d'entropie

Dans la littérature, la notion d'entropie renvoie à la mesure de la répartition de la main d'œuvre dans les différentes activités d'une zone. Une mesure d'entropie est établie sur une population, en considérant un niveau de nomenclature d'activités contenant **N** catégories d'activité.

Une fois la population et le niveau de nomenclature d'activité choisis, on définit $\mathbf{x_n}$ comme la part de main d'œuvre travaillant dans la catégorie «**n**» de la nomenclature choisie (en se restreignant aux catégories d'activité représentée dans la population étudiée), ces $\mathbf{x_n}$ vérifiant les deux propriétés suivantes :

- $x_n \in]0,1] \ \forall \ n \in \{1,2,\dots,N\}$ ➡ Chaque catégorie d'activité présente dans la population représente une part de la main d'œuvre de cette population, comprise entre 0 et 1, 0 exclu.
- $\sum_n x_n = 1$ ➡ La somme des parts des catégories d'activité vaut 1.

Ainsi, on peut définir la mesure d'entropie sur la base de cette nomenclature dans la population étudiée, de la manière suivante :

$$\text{ENTROPIE} = \sum_{\mathrm{n}} [\, x_n \times log_2(x_n^{-1})]$$

Cette formule mathématique se retrouve dans deux mesures étudiées précédemment :

- $\mathbf{UV_i}$ qui mesure l'entropie au niveau des secteurs dans la zone $\mathbf{Z_i}$
- $\mathbf{H_{k,i}}$ qui mesure l'entropie au niveau des activités dans la zone $\mathbf{Z_i}$ restreinte au secteur $\mathbf{S_k}$

Si on définit $\mathbf{E_i}$ comme l'entropie au niveau des activités dans la zone $\mathbf{Z_i}$:

$$E_i = \sum_{\mathrm{k}} [\, p_{j,i} \times log_2(p_{j,i}^{-1})]$$

alors, on observe la propriété suivante : $\mathbf{E_i = UV_i + RV_i}$

C'est pour cette raison que l'on parle de décomposition de l'entropie (tout comme la variance inter-groupe et la variance intra-groupe représentent la décomposition de la variance totale).

> Remarque : Attention, dans la formule de $\mathbf{E_i}$, ce sont les parts des activités au sein de la zone $\mathbf{Z_i}$ qui sont considérées (les $\mathbf{p_{j,i}}$) alors que dans la formule de $\mathbf{H_{k,i}}$, ce sont les parts des activités au sein de la zone $\mathbf{Z_i}$ restreinte au secteur $\mathbf{S_k}$ qui le sont (**les** $\mathbf{p_{j,k,i}}$).

$$\sum_k \left[p_{k,i} \times log_2\left(p_{k,i}^{-1}\right)\right]$$

Pour aller plus loin

Ce travail d'analyse fait partie du projet de recherche REPRO-INNOV, qui étudie les réorganisations productives et les innovations dans les filières agroalimentaires. Un projet co-financé par INRAE et la Région Occitanie dans le cadre du programme PSDR4 Occitanie (2016-2021).

Plus d'informations sur le programme PSDR4 Occitanie et le projet REPRO-INNOV : www.psdr-occitanie.fr

$$\left[\frac{eff_{k,i}}{eff_i} \times log_2\left(\frac{eff_i}{eff_{k,i}}\right)\right] \rightarrow UV_i = \sum_{S_k \in Z_i} \left[p_{k,i} \times log_2\left(p_{k,i}^{-1}\right)\right]$$

$$QL_{Z_i}^{S_k} = \frac{eff_{k,i} \,/\, eff_i}{eff_k \,/\, eff_{TOT}} \rightarrow QL_{k,i} = \frac{p_{k,i}}{p_k}$$

Auteurs

Olivier Pauly et **Pierre Triboulet**,
UMR AGIR, Université de Toulouse, INRAE, Castanet-Tolosan, France.

Contacts : olivier.pauly@inrae.fr
et pierre.triboulet@inrae.fr

Coordination éditoriale : Lucie Viou, PSDR4 Occitanie
Conception graphique : Clara Luneau