CHAPITRE 1-2

L'INTERPRÉTATION DES GRANDEURS

Plan

CHAPITRE 1-2
L'INTERPRÉTATION DES GRANDEURS

Au-delà du problème de la mesure surgit celui de l'interprétation des grandeurs, c'est-à-dire du sens à donner aux nombres. Nous aborderons ici trois sujets. Nous examinerons d'abord deux techniques numériques fréquemment utilisées pour faciliter l'interprétation des grandeurs : la construction d'une mesure relative et l'analyse de décomposition. Dans chaque cas, la méthode sera illustrée par une technique largement répandue en sciences régionales et en études urbaines. L'accent sera mis sur les limites de ces outils. Ensuite, nous traiterons de la mesure de la croissance ou, plus généralement, de la manière de résumer l'évolution d'une grandeur dans le temps.

1-2.1 Mesures relatives : l'exemple du quotient de localisation

- En 1600, l'Angleterre comptait une population de l'ordre de cinq millions d'habitants [1]. Peut-on dire que l'Angleterre était alors densément peuplée ?
- À Berlin, vers 1800, une famille de maçon de cinq personnes consacrait 44,2 % de son budget à l'achat de pain [2]. Pour l'époque, était-ce normal ?
- Le prix réel du blé en France, traduit en heures de travail de manoeuvre, a été de moins de 100 heures le quintal au 15^e siècle et au 16^e jusqu'en 1543, puis au-dessus de 100 jusqu'en 1883, environ [3]. Qu'est-ce que cela signifie pour le niveau de vie ?

En somme, « Est-ce beaucoup ? » : c'est la réaction de la plupart d'entre nous quand on nous cite un chiffre dans un domaine qui ne nous est pas familier. Dans les exemples qui précèdent, ce ne sont pas les unités de mesure qui font problème. C'est le manque de point de référence.

---

1 Braudel (1979, p. 49). Braudel compare l'Angleterre à la France (20 millions à la même époque) et conclut que la France était surpeuplée, puisque « Si l'un et l'autre pays s'étaient agrandis au rythme moyen du monde, l'Angleterre devrait compter aujourd'hui 40 millions d'habitants, la France 160 », ce qui est loin des chiffres actuels : en 2001, la France comptait 59,6 millions d'habitants et le Royaume-Uni, 58,9 (PNUD, 2003)

2 Braudel (1979, p. 142). L'ensemble de la nourriture représente 72,7 % du budget. Le pain compte donc pour 60,8 % de la dépense alimentaire de la famille, « proportion énorme étant donné le prix relatif des céréales ». L'auteur fait la comparaison avec la dépense alimentaire du Parisien en 1788 et 1854 : « Le blé, premier fournisseur d'énergie, n'arrive qu'au troisième poste des dépenses, après la viande et le vin (17 % seulement, dans les deux cas, de la dépense totale) » (p. 143-144).

3 Braudel (1979, p. 145) explique : « Un travailleur accomplit *approximativement* 3 000 heures de travail, chaque année; sa famille (4 personnes) consomme *approximativement* 12 quintaux par an... Franchir la ligne des 100 heures pour un quintal [1 quintal = 100 kg] est toujours grave; celle des 200 signale une cote d'alerte; à 300, c'est la famine ».

En somme, si mesurer, c'est comparer, l'interprétation des grandeurs requiert une « méta-comparaison », une comparaison avec une grandeur qui a un sens pour l'observateur, afin de mettre les données en perspective et de saisir l'ordre de grandeur des chiffres.

La méthode qui est peut-être la plus courante pour donner à l'observateur un point de repère pour l'interprétation des grandeurs est la représentation graphique avec comparaison. Le lecteur est invité à consulter à cet égard le passage « Du bon et du mauvais usage des graphes » dans Wonnacott et Wonnacott (1992, p. 61-69).

Il est également important, pour interpréter correctement une grandeur, de connaître son domaine de variation. Nous reviendrons sur ce thème lorsque nous examinerons les mesures d'inégalité et les mesures de similarité/dissimilarité.

Cela dit, il est parfois utile d'aller plus loin et de formaliser davantage cette méta-comparaison en construisant une mesure relative, qui est le rapport de deux valeurs. C'est ce que fait le quotient de localisation.

LE QUOTIENT DE LOCALISATION

Réf. : Page-Patton, 1991, ch. 14 ; Polèse, 1994, p. 128-129

Les quotients de localisation, aussi appelés *indices de concentration relative*, sont des mesures de l'importance relative de l'emploi d'une branche d'activité dans une ville ou une région [4]. Ils s'appliquent donc aux données d'un tableau de l'emploi par branche et par ville ou région. Voici un exemple numérique fictif :

Emploi par zone et par branche

BRANCHE	*B1*	*B2*	*B3*	Total
ZONE				
Z1	48	325	287	660
Z2	27	185	148	360
Z3	45	90	45	180
Total	120	600	480	1200

Un tableau ce ce type s'appelle un tableau de contingence (voir Gilles, 1994, section 6.3). Nous voulons pouvoir répondre à des questions comme : « 48 emplois de la branche *B1* dans la zone *Z1*, est-ce peu ? » ou « 325 emplois de la branche *B2* dans la zone *Z1*, est-ce beaucoup ? ».

4 Ils appartiennent à la famille de ce que Jayet (1993, p. 18) appelle les « indicateurs de spécificité ».

Comme première étape, nous pouvons examiner les distributions.

Distribution de l'emploi des branches entre zones

BRANCHE / ZONE	*B1*	*B2*	*B3*	Total
Z1	0,400	0,542	0,598	0,550
Z2	0,225	0,308	0,308	0,300
Z3	0,375	0,150	0,094	0,150
Total	**1,000**	**1,000**	**1,000**	**1,000**

48 emplois de la branche *B1* dans la zone *Z1*, est-ce peu ? L'examen de la distribution de l'emploi entre les zones montre que ces 48 emplois constituent 40 % de l'emploi total de la branche *B1* : la zone *Z1* est celle où l'on trouve le plus grand nombre d'emplois ce cette branche. Par contre, la zone *Z1* contient 55 % de l'emploi, toutes branches confondues : 48 emplois, ce n'est donc pas beaucoup, compte tenu de la taille de la zone *Z1*.

Distribution de l'emploi des zones entre branches

BRANCHE / ZONE	*B1*	*B2*	*B3*	Total
Z1	0,073	0,492	0,435	**1,000**
Z2	0,075	0,514	0,411	**1,000**
Z3	0,250	0,500	0,250	**1,000**
Total	0,100	0,500	0,400	**1,000**

325 emplois de la branche *B2* dans la zone *Z1*, est-ce beaucoup ? L'examen de la distribution de l'emploi entre les branches montre que ces 325 emplois représentent près de la moitié (49 %) de l'emploi de la zone *Z1*. Mais dans l'ensemble de l'économie, la branche *B2* compte pour la moitié : 325 emplois, c'est donc « normal ».

Le calcul des quotients de localisation est une manière de formaliser ce genre de raisonnement. Dans le premier cas (48 emplois de la branche *B1* dans la zone *Z1*), l'importance de la zone est mesurée par la part de cette zone dans l'emploi total de la branche (48 / 120 = 0,4 ou 40 %) [5]. Mais pour interpréter ce 40 %, nous nous sommes référés au pourcentage correspondant de l'ensemble des activités (660 / 1200 = 0,55 ou 55 %). La mesure relative que nous avons utilisée implicitement pour apprécier l'importance de *Z1* pour *B1* est le rapport 0,40 / 0,55 : un quotient de localisation, c'est ça ! Dans le second cas (325 emplois de la branche *B2* dans la zone *Z1*), nous avons procédé de manière analogue. L'importance de la branche est mesurée par la part de cette branche dans l'emploi total de la zone (325 / 660 = 0,492 ou 49 %). Mais

[5] En un sens, c'est déjà là une mesure relative, puisque cette part est donnée par le rapport de deux nombres comparables.

pour interpréter de 49 %, nous nous sommes référés au pourcentage correspondant de l'ensemble des zones (600 / 1200 = 0,5 ou 50 %). La mesure relative que nous avons utilitée implicitement pour apprécier l'importance de *B2* pour *Z1* est le rapport 0,49 / 0,5 : ça aussi, c'est un quotient de localisation. Ainsi, le quotient de localisation compare deux points correspondants sur deux distributions (deux points correspondants, et non pas deux distributions : les distributions sont des objets multidimensionnels ; nous verrons au chapitre 1-4 comment on peut les comparer). Voyons cela plus formellement.

Tableaux de contingence : notation et identités fondamentales

Avant de faire une présentation plus formelle des quotients de localisation, établissons un système de notation approprié et rappelons les identités fondamentales qui se vérifient dans un tableau de contingence comme celle de l'emploi par zone et par branche.

Notation

x_{ij}	nombre d'emplois de la branche *j* dans la zone *i*
$x_{\bullet j} = \sum_i x_{ij}$	nombre total d'emplois de la branche *j*
$x_{i\bullet} = \sum_j x_{ij}$	nombre total d'emplois dans la zone *i*
$x_{\bullet\bullet} = \sum_i \sum_j x_{ij}$	nombre total d'emplois de toutes branches dans toutes zones
$p_{ij} = x_{ij} / x_{\bullet\bullet}$	fraction de l'emploi total global qui appartient à la branche *j* **et** est situé dans la zone *i*
$p_{\bullet j} = \sum_i p_{ij}$	fraction de l'emploi total global qui appartient à la branche *j*
$p_{i\bullet} = \sum_j p_{ij}$	fraction de l'emploi total global qui est situé dans la zone *i*
$p_{j/i\bullet} = p_{ij} / p_{i\bullet}$	fraction de l'emploi total de la zone *i* qui appartient à la branche *j*
$p_{i/\bullet j} = p_{ij} / p_{\bullet j}$	fraction de l'emploi total de la branche *j* qui est situé dans la zone *i*

On aura reconnu dans cette notation que $p_{\bullet j}$ et $p_{i\bullet}$ sont des *probabilités marginales* : $p_{\bullet j}$ est la probabilité qu'un emploi tiré au hasard parmi les $x_{\bullet\bullet}$ emplois dénombrés appartienne à la branche *j* ; $p_{i\bullet}$ est la probabilité qu'un emploi tiré au hasard soit situé dans la zone *i*. On aura aussi reconnu que $p_{j/i\bullet}$ et $p_{i/\bullet j}$ sont des *probabilités conditionnelles* : $p_{j/i\bullet}$ est la probabilité qu'un emploi tiré au hasard appartienne à la branche *j*, *étant donné* qu'il est situé dans la zone *i* ; $p_{i/\bullet j}$ est la probabilité qu'un emploi tiré au hasard soit situé dans la zone *i*, *étant donné* qu'il appartient à la branche *j*.

On a naturellement les identités suivantes :

$$p_{\bullet j} = \sum_i p_{ij} = \sum_i x_{ij}/x_{\bullet\bullet} = x_{\bullet j}/x_{\bullet\bullet} \text{ et } p_{i\bullet} = \sum_j p_{ij} = \sum_j x_{ij}/x_{\bullet\bullet} = x_{i\bullet}/x_{\bullet\bullet}$$

$$p_{j/i\bullet} = p_{ij}/p_{i\bullet} = x_{ij}/x_{i\bullet} \text{ et } p_{i/\bullet j} = p_{ij}/p_{\bullet j} = x_{ij}/x_{\bullet j}$$

$$\sum_i \sum_j p_{ij} = \sum_i p_{i\bullet} = \sum_j p_{\bullet j} = 1$$

$$\sum_j p_{j/i\bullet} = \sum_j p_{ij} \Big/ p_{i\bullet} = 1 \text{ et } \sum_i p_{i/\bullet j} = \sum_i p_{ij} \Big/ p_{\bullet j} = 1$$

Le quotient de localisation : formalisation

Le quotient de localisation peut être défini aussi bien à partir de la distribution de l'emploi entre branches qu'à partir de la distribution entre zones. À partir de la distribution entre zones, le quotient de localisation de l'activité *j* dans la zone *i* est défini comme

$$QL_{ij} = \frac{\text{Fraction de l'emploi total de la branche } j \text{ situé dans la zone } i}{\text{Fraction de l'emploi total global situé dans la zone } i}$$

$$QL_{ij} = \frac{p_{i/\bullet j}}{p_{i\bullet}} = \frac{x_{ij}/x_{\bullet j}}{x_{i\bullet}/x_{\bullet\bullet}}$$

Par exemple,

$$QL_{21} = 0{,}225 \,/\, 0{,}300 = 0{,}750$$

De façon équivalente, à partir de la distribution entre branches, le quotient de localisation de l'activité *j* dans la zone *i* est défini comme

$$QL_{ij} = \frac{\text{Fraction de l'emploi total de la zone } i \text{ appartenant à la branche } j}{\text{Fraction de l'emploi total global appartenant à la branche } j}$$

$$QL_{ij} = \frac{p_{j/i\bullet}}{p_{\bullet j}} = \frac{x_{ij}/x_{i\bullet}}{x_{\bullet j}/x_{\bullet\bullet}}$$

Par exemple,

$$QL_{21} = 0{,}075 \ / \ 0{,}100 = 0{,}750$$

Ce n'est pas par accident que les deux calculs donnent le même résultat, puisque

$$\frac{x_{ij}/x_{\bullet j}}{x_{i\bullet}/x_{\bullet\bullet}} = \frac{x_{ij}/x_{i\bullet}}{x_{\bullet j}/x_{\bullet\bullet}} = \frac{x_{ij}\ x_{\bullet\bullet}}{x_{i\bullet}\ x_{\bullet j}}$$

Dans notre exemple,

$$QL_{21} = \frac{x_{21}/x_{\bullet 1}}{x_{2\bullet}/x_{\bullet\bullet}} = \frac{x_{21}/x_{2\bullet}}{x_{\bullet 1}/x_{\bullet\bullet}} = \frac{x_{21}\ x_{\bullet\bullet}}{x_{2\bullet}\ x_{\bullet 1}}$$

$$QL_{21} = \frac{27/120}{360/1200} = \frac{27/360}{120/1200} = \frac{27 \times 1200}{360 \times 120} = 0{,}75$$

Il y a bel et bien équivalence.

Quotients de localisation

BRANCHE / ZONE	*B1*	*B2*	*B3*
Z1	0,727	0,985	1,087
Z2	0,750	1,028	1,028
Z3	2,500	1,000	0,625

Les quotients de localisation peuvent prendre des valeurs entre zéro et l'infini [6]. Lorsque $x_{ij} = 0$, le quotient de localisation atteint sa valeur minimum : $QL_{ij} = 0$. Par ailleurs, il atteint la valeur la plus élevée possible lorsque $x_{ij} = x_{\bullet j} = x_{i\bullet}$, c'est-à-dire lorsque la totalité des emplois de l'activité *j* sont situés dans la zone *i* et qu'il ne se trouve aucune autre activité dans cette zone ; dans ces conditions, $QL_{ij} = {x_{\bullet\bullet}}/{x_{ij}}$

Dans l'expression qui précède, $x_{ij} = x_{\bullet j} = x_{i\bullet} \geq 1$, sans quoi la branche *j* n'existerait pas. Il s'ensuit que la valeur maximale de QL_{ij} est $x_{\bullet\bullet}$: cette valeur n'a pas de limite théorique et c'est pourquoi on dit que le quotient de localisation peut prendre des valeurs jusqu'à l'infini ; en pratique, le maximum est néanmoins limité par les valeurs observées.

Le point de repère naturel pour interpréter le quotient de localisation, est 1,0. Et comme le montrent les formules qui précèdent, on peut faire deux lectures du quotient de localisation :

- selon la première lecture, si $QL_{ij} > 1$, on dit que l'activité *j* est *relativement* concentrée [7] dans la zone *i* ; « relativement », c'est-à-dire en comparaison des autres activités, parce que la fraction de l'emploi qui est situé dans la zone *i* est *plus* importante pour l'activité *j que pour les autres activités* ; plus exactement, on dirait que la zone *i* est une zone de concentration relative pour cette activité, parce qu'il peut y avoir d'autres zones de concentration relative de cette même activité ;

Par exemple, $QL_{23} = 1{,}028$: l'activité *B3* est *relativement* concentrée dans la zone *Z2* ; pour autant, ce n'est pas dans la zone *Z2* qu'il y a le plus d'emploi de cette branche : c'est dans la zone *Z1*.

---

6 Certains auteurs normalisent le quotient de localisation à l'aide de la transformation $\frac{QL_{ij} - 1}{QL_{ij} + 1}$. Ce rapport varie de -1 à $+1$.

7 D'où, la justesse de l'expression « indice de concentration relative » pour désigner le quotient de localisation.

- selon la seconde lecture, si $QL_{ij} > 1$, on dira aussi que la zone i est *relativement* spécialisée dans l'activité *j* ; « relativement », c'est-à-dire en comparaison des autres zones, parce que l'activité *j* occupe dans la zone *i* une place *plus* importante *qu'ailleurs* ;

Par exemple, $QL_{31} = 2{,}500$: la zone *Z3* est *relativement* spécialisée dans l'activité *B1* ; pour autant, ce n'est pas la branche *B1* qui compte le plus grand nombre d'emplois de la zone *Z3* : c'est la branche *B2*.

- si au contraire $QL_{ij} < 1$, on dit que l'activité *j* est *relativement* moins présente dans la zone *i* qu'ailleurs : l'activité *j* n'est pas *relativement* concentrée dans la zone *i*, et la zone *i* n'est pas *relativement* spécialisée dans l'activité *j*.

Par exemple, $QL_{12} = 0{,}985$: la branche *B2* est *relativement* moins présente dans la zone *Z1*, bien qu'elle y soit la branche avec le plus grand nombre d'emplois et bien que ce soit en *Z1* que cette branche ait le plus grand nombre d'emplois (325 est le nombre le plus grand de sa ligne et de sa colonne).

Les exemples donnés montrent que l'adverbe « relativement » est important dans les énoncés d'interprétation qui précèdent. Voyons cela d'un point de vue plus général. Si la zone *i* est petite par rapport aux autres zones ($p_{i\bullet}$ petit), il se peut, même quand $QL_{ij} > 1$, que la fraction de l'emploi de l'activité *j* qui se trouve dans la zone *i* ($p_{i/\bullet j}$) ne soit pas importante. En effet,

$$QL_{ij} = \frac{\text{Fraction de l'emploi total de la branche } j \text{ situé dans la zone } i}{\text{Fraction de l'emploi total global situé dans la zone } i}$$

$$QL_{ij} = \frac{p_{i/\bullet j}}{p_{i\bullet}} = \frac{x_{ij}/x_{\bullet j}}{x_{i\bullet}/x_{\bullet\bullet}}$$

de sorte que si $p_{i\bullet}$ est petit, il est possible que $QL_{ij} > 1$ même si $p_{i/\bullet j}$ est petit, pourvu que $p_{i\bullet}$ soit encore plus petit. Dans une telle situation, il serait évidemment inexact de prétendre que l'activité *j* est concentrée (en termes *absolus*) dans la zone *i*.

De même, si l'activité *j* est d'importance mineure dans l'économie ($p_{\bullet j}$ petit), il se peut, même quand $QL_{ij} > 1$, que la part de l'emploi de l'activité *j* dans l'emploi total de la zone *i* ($p_{j/i\bullet}$) ne soit pas importante. En effet,

$$QL_{ij} = \frac{\text{Fraction de l' emploi total de la zone } i \text{ appartenant à la branche } j}{\text{Fraction de l' emploi total global appartenant à la branche } j}$$

$$QL_{ij} = \frac{p_{j/i\bullet}}{p_{\bullet j}} = \frac{x_{ij}/x_{i\bullet}}{x_{\bullet j}/x_{\bullet\bullet}}$$

de sorte que si $p_{\bullet j}$ est petit, il est possible que $QL_{ij} > 1$ même si $p_{j/i\bullet}$ est petit, pourvu que $p_{\bullet j}$ soit encore plus petit. Dans une telle situation, il serait évidemment inexact de prétendre que la zone *i* est spécialisée (en termes *absolus*) dans l'activité *j*.

Correctement interprétés, les quotients de localisation peuvent servir notamment à l'analyse descriptive de données d'emploi (voir Lemelin et Polèse, 1993).

NOTE : Il est mathématiquement impossible que $QL_{ik} > 1$ pour toutes les zones *i* en même temps (ou, symétriquement que $QL_{ik} < 1$ pour toutes les zones *i* en même temps). En effet, puisque $QL_{ik} = \frac{p_{i/\bullet k}}{p_{i\bullet}}$, cela impliquerait que $p_{i/\bullet k} > p_{i\bullet}$ pour chaque *i*, de sorte que l'on aurait $\sum_i p_{i/\bullet k} > \sum_i p_{i\bullet}$, ce qui est manifestement impossible, étant donné que les deux sommations doivent être égales à 1.

De même, il est mathématiquement impossible que $QL_{kj} > 1$ pour toutes les activités *j* en même temps (ou, symétriquement, que $QL_{kj} < 1$ pour toutes les activités *j* en même temps). En effet, puisque $QL_{kj} = \frac{p_{j/k\bullet}}{p_{\bullet j}}$, cela impliquerait que $p_{j/k\bullet} > p_{\bullet j}$ pour chaque *j*, de sorte que l'on aurait $\sum_j p_{j/k\bullet} > \sum_j p_{\bullet j}$, ce qui est manifestement impossible, étant donné que les deux termes de la comparaison doivent être égaux à 1.

Il est utile de se rappeler ces règles : si l'on obtient un tel résultat, c'est qu'il y a une erreur dans les calculs...

ESTIMATION DE L'EMPLOI EXPORTATEUR AU MOYEN DU QUOTIENT DE LOCALISATION

On utilise aussi les quotients de localisation dans le cadre de la théorie de la base économique (Polèse, 1994, p. 125-138), pour estimer l'emploi « exportateur » (pour un exemple, voir Polèse et Stafford, 1982). Vu la rareté de données sur les échanges interrégionaux, cette possibilité est attrayante. Mais l'estimation de l'emploi exportateur au moyen des quotients de localisation repose sur des hypothèses plutôt restrictives (Isserman, 1980, p. 157) :

1. La productivité du travail est égale entre villes ou régions.
2. L'absorption (utilisation locale) du produit par emploi dans l'économie locale est égale entre villes ou régions [8].
3. Il n'y a pas d'importations ou d'exportations nettes de l'ensemble du pays.
4. La demande locale s'approvisionne en priorité auprès des producteurs locaux ; cela implique qu'il n'y a pas de flux croisés entre villes ou régions (« cross-hauling »).

Sous ces conditions, on peut interpréter l'excédent du quotient de localisation par rapport à 1,0 comme une mesure de l'emploi exportateur. Plus exactement, EXP_{ij}, l'emploi « exportateur » de la branche j, qui appartient à la *base économique* de la région i, peut alors s'estimer au moyen de la formule

$$EXP_{ij} = \begin{cases} x_{ij} \dfrac{(QL_{ij} - 1)}{QL_{ij}}, \text{si } QL_{ij} > 1 \\ 0 \text{ autrement} \end{cases}$$

Par exemple, $EXP_{31} = 45 \times \frac{2{,}5 - 1}{2{,}5} = 27$ des 45 emplois de la branche $B1$ dans la zone $Z3$.

La fraction de x_{ij} qui appartient à l'emploi exportateur est la fraction de QL_{ij} qui excède 1. Quand $QL_{ij} < 1$, il n'y a pas d'exportations de l'activité j à partir de la région i et, en conséquence, l'emploi exportateur est nul.

Pour comprendre plus facilement la signification de ce calcul, on substitue QL_{ij} et on simplifie, de façon à obtenir

$$EXP_{ij} = x_{ij} - \left(x_{i\bullet} \big/ x_{\bullet\bullet} \right) x_{\bullet j} = x_{ij} - p_{i\bullet} x_{\bullet j} \text{ , si } x_{ij} > p_{i\bullet} x_{\bullet j}$$

---

8 Isserman (1980) et Norcliffe (1983) utilisent le terme « consommation » pour désigner l'utilisation à la fois par la demande finale et par la demande intermédiaire. Le terme « absorption » semble plus exact.

Par exemple, $EXP_{31} = 45 - \left(180 / 1200\right) 120 = 27$

On voit alors que l'emploi exportateur est la différence entre la valeur observée x_{ij} et la valeur hypothétique que prendrait le chiffre de l'emploi si la région *i* produisait seulement « sa part » de *j* (c'est-à-dire $p_{i\bullet}$, auquel cas le quotient de localisation QL_{ij} serait égal à 1).

Pour voir comment interviennent les hypothèses énoncées précédemment, récrivons la formule sous la forme suivante :

$$EXP_{ij} = \left[\left(x_{ij} / x_{\bullet j}\right) - \left(x_{i\bullet} / x_{\bullet\bullet}\right)\right] x_{\bullet j} = \left(p_{i/\bullet j} - p_{i\bullet}\right) x_{\bullet j} \text{, si } p_{i/\bullet j} > p_{i\bullet}$$

Par exemple, $EXP_{31} = \left[\left(45 / 120\right) - \left(180 / 1200\right)\right] 120 = 27$

La **première hypothèse** concerne le rapport $p_{i/\bullet j}$ ou $x_{ij} / x_{\bullet j}$; ce rapport est la part de la région *i* dans l'*emploi* de l'activité *j* ; la première hypothèse permet de considérer ce rapport comme une approximation de la part de la région dans la *production* du bien *j*.

La **seconde hypothèse** concerne le rapport $p_{i\bullet}$ ou $x_{i\bullet} / x_{\bullet\bullet}$; ce rapport est la part de la région *i* dans l'*emploi total ;* la seconde hypothèse permet de considérer ce rapport comme une approximation de la part de la région dans l'*utilisation totale* (absorption) du bien *i*.

Les deux autres hypothèses permettent d'interpréter la différence comme la part de l'emploi national de la branche *j* qui appartient à la base économique de la région *i*. Ainsi, la **troisième hypothèse** dit que les importations et exportations internationales sont nulles : il s'ensuit que l'identité

$$\boxed{\begin{array}{c} \textit{Production} \\ + \\ \textit{Importations des autres régions} \\ + \\ \textit{Importations internationales} \end{array}} = \boxed{\begin{array}{c} \textit{Absorption} \\ + \\ \textit{Exportations aux autres régions} \\ + \\ \textit{Exportations internationales} \end{array}}$$

devient

$$\boxed{\begin{array}{c}\textit{Production}\\ +\\ \textit{Importations des autres régions}\end{array}} = \boxed{\begin{array}{c}\textit{Absorption}\\ +\\ \textit{Exportations aux autres régions}\end{array}}$$

c'est-à-dire

$$\boxed{\textit{Production} - \textit{Absorption}} = \boxed{\textit{Exportations nettes aux autres régions}}$$

lorsque l'excédent de la production sur l'absorption est positif.

La **quatrième hypothèse** enfin dit que, s'il y a des exportations vers les autres régions, il n'y a pas d'importations en provenance d'autres régions, et vice-versa. Par conséquent, quand les exportations nettes sont positives, elles sont égales aux exportations brutes.

Après avoir estimé l'emploi exportateur de chaque branche, il suffit de faire la somme pour obtenir un estimé de la « base » économique de la région *i* (on dit aussi emploi *basique*) :

$$\text{Base exportatrice} = \sum_{j \text{ lorsque } QL_{ij} > 1} EXP_{ij}$$

Dans le modèle de la base économique, on fait l'hypothèse que le rapport

$$\theta = \frac{\sum_j x_{ij}}{\sum_j EXP_{ij}}$$

est constant. Le modèle prédit que pour chaque emploi exportateur qui se crée (ou qui disparaît), l'emploi total augmente (ou diminue) de θ emplois. Le facteur θ s'appelle le *multiplicateur de la base économique* [9].

1-2.2 L'analyse de décomposition additive et multiplicative des variations

PRINCIPE

L'analyse « shift-share » est un cas particulier d'une technique plus générale, l'analyse de décomposition des variations [10]. L'analyse de décomposition des variations peut s'appliquer en

---

9 On a proposé de multiples variantes du quotient de localisation pour relâcher les hypothèses très contraignantes sur lesquelles repose la méthode.

10 L'article, bien connu en sciences régionales, de Williamson (1965) en donne un autre exemple : il propose une décomposition de l'évolution dans le temps d'une mesure d'inégalité interrégionale.

principe à tout écart entre deux valeurs observées d'une même variable. Il peut s'agir de deux observations d'un même objet à des moments différents dans le temps, ou d'observations effectuées sur deux objets distincts.

L'analyse de décomposition des variations consiste à décomposer l'écart entre deux valeurs d'une mesure en une somme de termes (décomposition additive) ou en un produit de facteurs (décomposition multiplicative). Une telle décomposition est toujours une tautologie du genre

$$x - y = (x - a) + (a - b) + (b - c) + (c - y)$$

ou

$$x/y = (x/a)\ (a/b)\ (b/c)\ (c/y)$$

c'est-à-dire

$$\log x - \log y = (\log x - \log a) + (\log a - \log b) + (\log b - \log c) + (\log c - \log y)$$

L'utilité de la décomposition vient donc de l'interprétation que l'on peut donner aux termes d'une décomposition additive ou aux facteurs d'une décomposition multiplicative. Cette interprétation repose sur un modèle, qui demeure souvent implicite. Le langage utilisé (« effet ceci », « facteur cela ») laisse parfois percer des connotations de causalité qui ne sont pas toujours justifiées.

APPLICATION À L'ANALYSE « SHIFT-SHARE »

Réf. : Page-Patton, ch.9 ; Coffey et Polèse (1988) ; Polèse, 1994, p. 349-357

L'analyse « shift-share » [11] est une méthode d'analyse de décomposition bien connue des praticiens des sciences régionales. Elle consiste à décomposer la variation de l'emploi d'une ville ou d'une région. Nous allons examiner successivement la méthode de décomposition de la variation de l'emploi d'une activité, puis celle de la variation de l'emploi d'un ensemble d'activités.

Décomposition de la variation de l'emploi d'une activité

Pour illustrer l'analyse shift-share, nous allons nous servir de l'exemple numérique fictif suivant.

---

[11] Jayet (1993, p. 29-34) emploie l'expression « analyse structurelle-géographique ». Je préfère encore l'expression de Bonnet (1995) : « analyse structurelle-résiduelle ».

Emploi par zone et par branche

	An 1				An 2			
BRANCHE	*B1*	*B2*	*B3*	Total	*B1*	*B2*	*B3*	Total
ZONE								
Z1	48	325	287	660	24	388	300	712
Z2	27	185	148	360	11	173	200	384
Z3	45	90	45	180	25	99	52	176
Total	120	600	480	1200	60	660	552	1272

Variation de l'emploi par zone et par branche entre l'An 1 et l'An 2

	Différences				Taux de variation			
BRANCHE	*B1*	*B2*	*B3*	Total	*B1*	*B2*	*B3*	Total
ZONE								
*Z*1	-24	63	13	52	-50,00%	19,38%	4,53%	7,88%
*Z*2	-16	-12	52	24	-59,26%	-6,49%	35,14%	6,67%
*Z*3	-20	9	7	-4	-44,44%	10,00%	15,56%	-2,22%
Total	-60	60	72	72	-50,00%	10,00%	15,00%	6,00%

Penchons-nous sur la variation de l'emploi de la branche *B1* dans la zone *Z2*.

L'analyse shift-share est basée sur la comparaison de trois scénarios :

1. Quelle aurait été la variation si l'emploi de *B1* en *Z2* avait évolué au même taux que l'emploi total (toutes branches et toutes zones) ?
   – Taux = 6 %
   – Nombre = 6 % de 27 = 1,62
2. Quelle aurait été la variation si l'emploi de *B1* en *Z2* avait évolué au même taux que l'emploi de l'ensemble de la branche *B1* ?
   – Taux = –50 %
   – Nombre = –50 % de 27 = –13,50
3. Quelle a été la variation observée de l'emploi de *B1* en *Z2* ?
   – Taux = –59,26 %
   – Nombre = –59,26 % de 27 = –16

La comparaison de ces trois scénarios conduit à la décomposition additive suivante :

1. Effet national = scénario 1 :
   – Taux = 6 %
   – Nombre = 6 % de 27 = 1,62
2. Effet proportionnel (ou sectoriel) = écart entre scénario 2 et scénario 1 :
   – Taux = –50 % – 6 % = –56 %

- Nombre = –56 % de 27 = –15,12 = –13,5 – 1,62

3. Effet résiduel (ou régional) = écart entre scénario 3 et scénario 2 :
   - Taux = –59,26 % – (–50 %) = –9,26 %
   - Nombre = –9,26 % de 27 = –2,5 = –16 – (–13,5)

On peut vérifier que la somme des trois « effets » est bien égale à la variation observée :

- Taux = 6 % + (–56 %) + (–9,26 %) = –59,26 %
- Nombre = 1,62 + (–15,12) + (–2,5) = –16

Cette méthode de décomposition peut se formaliser à l'aide de la notation suivante :

x_{ij}^t	l'emploi de la branche *j* dans la région *i* au temps *t*
$x_{\bullet j}^t = \sum_i x_{ij}^t$	l'emploi de la branche *j* dans l'ensemble des régions au temps *t*
$x_{\bullet\bullet}^t = \sum_i \sum_j x_{ij}^t$	l'emploi de toutes les branches dans l'ensemble des régions au temps *t*

On a l'identité suivante :

$$\frac{x_{ij}^t}{x_{ij}^0} = \left(\frac{x_{ij}^t}{x_{ij}^0} - \frac{x_{\bullet j}^t}{x_{\bullet j}^0}\right) + \left(\frac{x_{\bullet j}^t}{x_{\bullet j}^0} - \frac{x_{\bullet\bullet}^t}{x_{\bullet\bullet}^0}\right) + \frac{x_{\bullet\bullet}^t}{x_{\bullet\bullet}^0}$$

Dénotons les taux d'accroissement par

$$r_{ij} = \frac{x_{ij}^t}{x_{ij}^0} - 1$$

L'identité précédente peut alors se récrire en termes des taux d'accroissement

$$\left(\frac{x_{ij}^t}{x_{ij}^0} - 1\right) = \left[\left(\frac{x_{ij}^t}{x_{ij}^0} - 1\right) - \left(\frac{x_{\bullet j}^t}{x_{\bullet j}^0} - 1\right)\right] - \left[\left(\frac{x_{\bullet j}^t}{x_{\bullet j}^0} - 1\right) - \left(\frac{x_{\bullet\bullet}^t}{x_{\bullet\bullet}^0} - 1\right)\right] - \left(\frac{x_{\bullet\bullet}^t}{x_{\bullet\bullet}^0} - 1\right)$$

c'est-à-dire

$$r_{ij} = \left(r_{ij} - r_{\bullet j}\right) + \left(r_{\bullet j} - r_{\bullet\bullet}\right) + r_{\bullet\bullet}$$

ou alors en nombres d'emplois

$$r_{ij}\, x_{ij}^0 = \left(r_{ij} - r_{\bullet j}\right) x_{ij}^0 + \left(r_{\bullet j} - r_{\bullet\bullet}\right) x_{ij}^0 + r_{\bullet\bullet}\, x_{ij}^0$$

Dans cette décomposition,

- $r_{\bullet\bullet} x_{ij}^{0}$ est l'effet national (« national share effect ») : c'est l'accroissement qui se serait réalisé si l'emploi de la branche *j* dans la région *i* avait augmenté au même taux que l'emploi total au pays (scénario 1) ;
- $\left(r_{\bullet j} - r_{\bullet\bullet}\right)x_{ij}^{0}$ est l'effet sectoriel ou effet de déplacement proportionnel (« proportional shift effect ») : c'est l'accroissement *supplémentaire* (positif ou négatif) de l'emploi qui se serait réalisé si l'emploi de la branche *j* dans la région *i* avait augmenté au même taux que l'emploi de la branche *j* dans l'ensemble du pays (c'est donc la différence entre le scénario 2 et le scénario 1) ; l'effet sectoriel se rattache à la question de savoir si, comparée au reste de l'économie, la branche *j* est dynamique, si elle jouit d'une croissance accélérée ;
- $\left(r_{ij} - r_{\bullet j}\right)x_{ij}^{0}$ est l'effet régional ou effet de déplacement différentiel (« differential shift effect ») : c'est l'écart résiduel entre l'accroissement observé et l'accroissement résultant de l'application de l'effet de part et l'effet de déplacement proportionnel.

La somme de l'effet de déplacement proportionnel et de l'effet de déplacement différentiel est l'effet de déplacement total ou net (« total shift » ou « net shift ») pour une activité *j* dans une région *i*.

Le déplacement différentiel est souvent interprété comme une mesure de la *compétitivité* de la branche *j* de la région *i* (« competitive effect »). Cette utilisation est fort contestable. En effet, supposons que l'emploi de la branche *j* croît plus rapidement dans la région *i* que dans la région *k* ; cela veut-il dire que la *production* de cette branche croît plus rapidement dans la région *i* que dans la région *k* ? Pas nécessairement, si la proportion entre la main-d'oeuvre et les autres facteurs de production varie d'une région à l'autre et dans le temps (en réponse aux changements des prix relatifs) [12]. Nous verrons dans un moment que ce n'est pas la seule raison de douter de la validité de l'interprétation de l'effet régional comme mesure de la compétitivité.

Décomposition de la variation de l'emploi total d'une région

En faisant la sommation sur l'ensemble des branches de chacun des trois termes de la décomposition, on obtient

---

[12] Une démonstration formelle de cette proposition exigerait que l'on développe un modèle d'équilibre général de deux économies.

$$\sum_j r_{ij}\, x_{ij}^0 = \sum_j \left(r_{ij} - r_{\bullet j}\right) x_{ij}^0 + \sum_j \left(r_{\bullet j} - r_{\bullet\bullet}\right) x_{ij}^0 + \sum_j r_{\bullet\bullet}\, x_{ij}^0$$

où

$\sum_j r_{\bullet\bullet}\, x_{ij}^0 = r_{\bullet\bullet} \sum_j x_{ij}^0$ est l'effet national

$\sum_j \left(r_{\bullet j} - r_{\bullet\bullet}\right) x_{ij}^0$ est l'effet de structure

$\sum_j \left(r_{ij} - r_{\bullet j}\right) x_{ij}^0$ est l'effet régional

Les quatre tableaux qui suivent complètent les calculs pour notre exemple.

Analyse shift-share par branche

Branche *B1*

	Nombre d'emplois				Taux de variation			
	Effet national	Effet sectoriel	Effet résiduel	Effet total	Effet national	Effet sectoriel	Effet résiduel	Effet total
ZONE								
Z1	2,88	-26,88	0	-24	6,00%	-56,00%	0,00%	-50,00%
Z2	1,62	-15,12	-2,5	-16	6,00%	-56,00%	-9,26%	-59,26%
Z3	2,7	-25,2	2,5	-20	6,00%	-56,00%	5,56%	-44,44%

Branche *B2*

	Nombre d'emplois				Taux de variation			
	Effet national	Effet sectoriel	Effet résiduel	Effet total	Effet national	Effet sectoriel	Effet résiduel	Effet total
ZONE								
Z1	19,5	13	30,5	63	6,00%	4,00%	9,38%	19,38%
Z2	11,1	7,4	-30,5	-12	6,00%	4,00%	-16,49%	-6,49%
Z3	5,4	3,6	0,0	9	6,00%	4,00%	0,00%	10,00%

Branche *B3*

p	Nombre d'emplois				Taux de variation			
	Effet national	Effet sectoriel	Effet résiduel	Effet total	Effet national	Effet sectoriel	Effet résiduel	Effet total
ZONE								
Z1	17,22	25,83	-30,05	13	6,00%	9,00%	-10,47%	4,53%
Z2	8,88	13,32	29,8	52	6,00%	9,00%	20,14%	35,14%
Z3	2,7	4,05	0,25	7	6,00%	9,00%	0,56%	15,56%

Ensemble des branches (classification à trois branches)

	Nombre d'emplois				Taux de variation			
	Effet national	Effet sectoriel	Effet résiduel	Effet total	Effet national	Effet sectoriel	Effet résiduel	Effet total
ZONE								
Z1	39,6	11,95	0,45	52	6,00%	1,81%	0,07%	7,88%
Z2	21,6	5,6	-3,2	24	6,00%	1,56%	-0,89%	6,67%
Z3	10,8	-17,55	2,75	-4	6,00%	-9,75%	1,53%	-2,22%

L'effet de structure est interprété comme l'effet de la structure économique ou industrielle de la région, c'est-à-dire de la composition de sa production industrielle (« industry mix effect »). L'effet régional est souvent interprété comme une mesure de la compétitivité de la région *i* : cette interprétation appelle les mêmes réserves que précédemment. Mais il y a plus grave.

En effet, lorsqu'on décompose la variation du niveau *global* de l'emploi d'une région, l'importance de l'effet de structure dépend du niveau d'agrégation de la classification des activités. Et comme l'effet régional est calculé de façon résiduelle, ce dernier dépend aussi du niveau d'agrégation. Pour une région particulière, le passage d'une classification à une autre peut augmenter ou diminuer l'effet régional, de sorte qu'il peut en résulter des interversions de rang entre les régions : une région qui paraissait plus compétitive qu'une autre dans une classification donnée peut paraître moins compétitive dans une autre classification ! Cela restreint encore la validité de l'interprétation de l'effet régional comme mesure de la compétitivité d'une région.

Ce phénomène est illustré au moyen de l'aggrégation des branches *B1* et *B2*. On peut constater que les résultats de la décomposition pour l'ensemble des branches sont différents de ceux qui ont été obtenus précédemment avec une classification à trois branches.

Branches *B1* et *B2* agrégées

	Nombre d'emplois				Taux de variation			
	Effet national	Effet sectoriel	Effet résiduel	Effet total	Effet national	Effet sectoriel	Effet résiduel	Effet total
ZONE								
Z1	22,38	-22,38	39	39	6,00%	-6,00%	10,46%	10,46%
Z2	12,72	-12,72	-28	-28	6,00%	-6,00%	-13,21%	-13,21%
Z3	8,1	-8,1	-11	-11	6,00%	-6,00%	-8,15%	-8,15%

Ensemble des branches (classification à deux branches)

	Nombre d'emplois				Taux de variation			
	Effet national	Effet sectoriel	Effet résiduel	Effet total	Effet national	Effet sectoriel	Effet résiduel	Effet total
ZONE								
Z1	39,6	**3,45**	**8,95**	52	6,00%	**0,52%**	**1,36%**	7,88%
Z2	21,6	**0,6**	**1,8**	24	6,00%	**0,17%**	**0,50%**	6,67%
Z3	10,8	**-4,05**	**-10,75**	-4	6,00%	**-2,25%**	**-5,97%**	-2,22%

Cette faiblesse touche également l'analyse de décomposition de la variation du niveau d'une seule activité, puisque l'activité en question est définie selon une classification qui comporte inévitablement un certain degré d'agrégation. En d'autres mots, s'agissant d'une seule activité, l'effet régional calculé contient l'effet structurel associé aux déplacements entre les sous-branches qui composent l'activité considérée. Il est donc abusif, même lorsqu'on ne considère qu'une seule branche, d'interpréter l'effet résiduel comme une mesure de la compétitivité.

1-2.3 La mesure de la croissance (le calcul du taux de variation dans le temps)

D'une certaine manière, l'analyse d'une série chronologique pose le problème de la multidimensionnalité, dont il sera question plus loin. En effet, le concept de « croissance » ou de « variation dans le temps » comporte de multiples dimensions. Car sauf dans le cas où le taux de variation est constant (la croissance est uniforme), le concept a autant de dimensions que l'on a d'observations sur la croissance.

TAUX DE CROISSANCE PAR PÉRIODE

Considérons par exemple l'évolution de l'indice des prix à la consommation (IPC) au Canada de 1984 à 1992 [13] :

1984	92,4
1985	96,0
1986	100,0
1987	104,4
1988	108,6
1989	114,0
1990	119,5
1991	126,2
1992	128,1

Entre chaque moment et le suivant, on peut calculer un *taux de croissance pour cet intervalle*. Ainsi, entre 1984 et 1985, le taux de croissance pour la période a été de

$$\frac{96{,}0 - 92{,}4}{92{,}4} = 0{,}039$$

soit 3,9 %. Avec neuf observations consécutives (de 1984 à 1992), on peut calculer huit taux de croissance par période :

de...	à...	taux
1984	1985	0,039
1985	1986	0,042
1986	1987	0,044
1987	1988	0,040
1988	1989	0,050
1989	1990	0,048
1990	1991	0,056
1991	1992	0,015

13 Moyenne annuelle non désaisonnalisée (Statistique Canada 62-210).

En général, avec une série de T+1 observations, de la période 0 à la période T,

$x_0, x_1, x_2, ..., x_t, ..., x_T$

on peut calculer T valeurs du taux de croissance r_t d'une période par rapport à la précédente :

$$r_t = \frac{x_t - x_{t-1}}{x_{t-1}} = \frac{x_t}{x_{t-1}} - \frac{x_{t-1}}{x_{t-1}} = \frac{x_t}{x_{t-1}} - 1$$

Par exemple, pour $t = 1985$,

$$r_{1985} = \frac{96{,}0}{92{,}4} - 1 = 0{,}039 = 3{,}9\ \%$$

où t varie de 1 à T. On dit alors qu'entre t–1 et t, x s'est accru de R_t pourcent, où

$R_t = 100 \times r_t$

> Note : si x_t est inférieur à x_{t-1}, il y a eu *décroissance* : r_t est *négatif*. On parle alors de *croissance négative*.

La série

$r_1, r_2, ..., r_t, ..., r_T$

décrit l'évolution dans le temps de la variable x. En effet, si on inverse la formule de calcul du taux de croissance par période, on obtient

$x_t = (1+r_t)\ x_{t-1}$

Mais on a également

$x_{t-1} = (1+r_{t-1})\ x_{t-2}$

de sorte que

$x_t = (1+r_t)\ (1+r_{t-1})\ x_{t-2}$

et, en procédant à des substitutions successives, on a

$x_t = (1+r_t)\ (1+r_{t-1})\ ...\ (1+r_2)\ (1+r_1)\ x_0$

En somme, si l'on connaît les r_t et x_0, on peut reconstituer la série des x_t. Il est donc bien vrai que la série des taux de croissance par période

$r_1, r_2, ..., r_t, ..., r_T$

décrit l'évolution dans le temps de la variable x [14]. Mais comment peut-on résumer cette évolution en un seul chiffre ?

D'une certaine manière, les T valeurs de r_t constituent autant de *dimensions* de l'évolution dans le temps de la variable x. C'est en ce sens que la mesure de la croissance s'apparente au problème de la multidimensionnalité et de la construction des nombres indices.

MOYENNE DES TAUX DE CROISSANCE PAR PÉRIODE

Pour résumer l'évolution de la variable x dans le temps, on peut prendre la moyenne arithmétique des taux de croissance par période :

$$\frac{1}{T}\sum_{t=1}^{T} r_t = \frac{r_1 + r_2 + \cdots + r_t + \cdots + r_T}{T}$$

Dans le cas de l'IPC entre 1984 et 1992, la moyenne des taux de croissance par période est de 0,042 (c'est-à-dire 4,2 %).

Cette façon de résumer l'évolution de la variable x dans le temps a cependant l'inconvénient de ne tenir aucun compte de la variabilité des taux de croissance. Or, pour un même taux moyen, plus les taux sont uniformes, plus la croissance cumulée est forte [15]. Nous n'allons pas démontrer cette propriété : un exemple suffira à l'illustrer. Comparons les deux séries suivantes :

100, 110, 121

et

100, 100, 120

Dans les deux cas, la moyenne des taux de croissance par période est égale à 0,1 (c'est-à-dire 10 %). Pourtant, la croissance cumulée sur les deux périodes est de 21 % dans le premier cas, mais de seulement 20 % dans le second. La question plus générale que cela soulève est donc la suivante : dans quelle mesure la moyenne des taux par période est-elle représentative de l'évolution d'une série lorsque les taux par période sont variables ?

---

[14] En pratique cependant, on utilise pour les taux de croissance des valeurs arrondies, de sorte que l'on ne pourrait pas reconstituer avec exactitude la série originale.

[15] Si la croissance cumulée dépend de la variabilité des taux, elle ne dépend pas en revanche de l'ordre chronologique entre les différents taux de croissance. On peut le voir en constatant dans la formule suivante que la valeur de x_t demeure inchangée si l'on change l'ordre des facteurs du membre de droite :

$$x_t = (1+r_t)\,(1+r_{t-1})\ \ldots\ (1+r_2)\,(1+r_1)\,x_0.$$

CALCUL D'UN TAUX DE CROISSANCE EXPONENTIEL

Le taux de croissance exponentiel est une autre façon de résumer l'évolution de la variable x dans le temps. On peut le définir de deux façons équivalentes.

La première définition du taux de croissance exponentiel fait appel à la moyenne géométrique [16]. C'est le taux de croissance r obtenu à partir de la moyenne géométrique des facteurs [17] de croissance par période :

$$1+r = \left[(1+r_T)(1+r_{T-1})\cdots(1+r_2)(1+r_1)\right]^{1/T} = \sqrt[T]{(1+r_T)(1+r_{T-1})\cdots(1+r_2)(1+r_1)}$$

C'est-à-dire, sous forme logarithmique :

$$\log(1+r) = \frac{1}{T}\sum_{t=1}^{T}\log(1+r_t)$$

La mesure précédente, au contraire, était donnée par la moyenne *arithmétique* des taux de croissance par période. On peut simplifier le calcul du taux de croissance exponentiel en exploitant la relation

$$x_T = (1+r_T)\ (1+r_{T-1})\ \ldots\ (1+r_2)\ (1+r_1)\ x_0$$

Or selon la définition de la moyenne géométrique

$$(1+r)^T = (1+r_T)(1+r_{T-1})\cdots(1+r_2)(1+r_1)$$

de sorte que

$$x_T = (1+r)^T\, x_0$$

où x_T et x_0 sont des valeurs connues et où r est l'inconnue. En développant cette formule, on arrive à une méthode de calcul du taux de croissance exponentiel.

$$(1+r)^T = \frac{x_T}{x_0}$$

$$\log(1+r)^T = \log\left(\frac{x_T}{x_0}\right)$$

$$T\log(1+r) = \log(x_T) - \log(x_0)$$

---

16 À propos de la moyenne géométrique et de ses applications, voir Wonnacott et Wonnacott (1991, p. 755).

17 Noter la distinction entre le *taux* de croissance r et le *facteur* de croissance $(1+r)$.

$$\log(1+r) = \frac{\log(x_T) - \log(x_0)}{T}$$

$$1 + r = \text{antilog}\left(\frac{\log(x_T) - \log(x_0)}{T}\right)$$

où antilog $z = e^z$ ou 10^z, selon qu'on a pris le logarithme népérien de base e (= 2,71828…) ou le logarithme commun de base 10.

$$r = \text{antilog}\left(\frac{\log(x_T) - \log(x_0)}{T}\right) - 1$$

C'est-à-dire, avec les logarithmes communs,

$$r = 10^{\frac{\log x_T - \log x_0}{T}} - 1$$

et avec les logarithmes népériens.

$$r = e^{\frac{\ln x_T - \ln x_0}{T}} - 1 = \exp\left(\frac{\ln x_T - \ln x_0}{T}\right) - 1$$

Par exemple, le taux de croissance exponentiel de l'IPC de 1984 à 1992 se calcule de la façon suivante :

$x_0 = 92{,}4$ et $\log_e x_0 = 4{,}526126979$

$x_T = 128{,}1$ et $\log_e x_T = 4{,}852811209$

$T = 8$

$$r = \exp\left(\frac{4{,}852811209 - 4{,}526126979}{8}\right) - 1 = 0{,}042 \text{, c'est-à-dire } 4{,}2\ \%$$

Il y a une autre définition du taux de croissance exponentiel, qui contient sa propre interprétation. En effet, nous venons de voir que le taux de croissance exponentiel r est la solution de l'équation

$$x_T = (1+r)^T x_0$$

Le taux de croissance exponentiel peut donc être vu comme un taux hypothétique : c'est la réponse à la question « si la variable x avait évolué à un taux par période constant, à quel taux eût-il fallu qu'elle crût pour que sa valeur terminale soit égale à la valeur terminale observée ? ». Par définition, le taux de croissance exponentiel est donc le taux de croissance par période uniforme qui donne la même croissance cumulée que la série

$r_1, r_2, ..., r_t, ..., r_T$

En ce sens, le taux de croissance exponentiel résume, comme la moyenne arithmétique des taux par période, l'évolution de la variable x dans le temps. Mais le taux de croissance exponentiel ne tient compte que de la première et de la dernière valeur de la série. Cela constitue un inconvénient si la première valeur de la série, x_0, ou la dernière, x_T, est exceptionnelle (hors tendance) : dans ce cas, le taux de croissance exponentiel pourra être trompeur.

En tant qu'indice, c'est en quelque sorte un indice tronqué, puisqu'il n'utilise pas toute l'information disponible. Les indices tronqués peuvent parfois avoir leur utilité et leurs faibles exigences en information leur font certainement des adeptes : qu'on pense au « Big Mac cost-of-living index » du périodique *The Economist*.

ENTRE DEUX MAUX...

En tant qu'indices de l'évolution chronologique d'une variable, la moyenne des taux par période et le taux de croissance exponentiel présentent tous deux des inconvénients. Il s'agit donc de choisir le moindre mal. Lequel ? Cela dépend évidemment de l'utilisation que l'on veut en faire. Par exemple, est-il important que le taux de croissance retenu permette de « prédire » avec exactitude (c'est-à-dire de reproduire) la valeur terminale à partir de la valeur initiale ? À cet égard, le taux de croissance exponentiel semble préférable. Ou est-il plus important au contraire que le taux de croissance retenu soit représentatif de la tendance ? Dans ce cas, rien ne garantit *a priori* que les valeurs initiale et terminale ne soient pas hors tendance et que, par conséquent, le taux exponentiel ne soit pas trompeur.

Mais nous avons vu que, de son côté, la moyenne des taux par période a le défaut de ne pas tenir compte de la variabilité des taux par période. Ce défaut est-il important ? Ça dépend. Ainsi, dans le cas de l'IPC 1984-1992 au Canada, si le taux de croissance avait été constant, égal à la moyenne des taux de croissance par période, la valeur de l'IPC en 1992 aurait été de 128,16, au lieu de 128,1. L'écart est minime (six centièmes d'un pourcent !).

L'écart serait-il plus considérable sur un grand nombre de périodes, avec un taux de croissance plus volatil ? Par exemple, la valeur de fermeture de l'indice boursier Standard & Poor's 500 était de 470,34 le 17 février 1994, et de 656,37 le 9 février 1996, 499 séances de marché plus tard [18]. La moyenne des taux de croissance par période (d'une séance de marché sur la précédente) a été de 0,0007 (0,07 %), avec un écart type de 0,0057 (0,57 %), ce qui représente une volatilité importante (le coefficient de variation est de 8,37). Or quelle aurait été la valeur de clôture le 9 février 1996 si l'indice avait crû à un taux constant égal à la moyenne des taux par période ? Elle aurait été de 661,76... Encore une fois, l'écart est mince (0,82 %) : dans ce cas encore, la moyenne des taux de croissance par période est assez représentative de la tendance.

AJUSTEMENT D'UNE COURBE DE TENDANCE

Quoi qu'il en soit, il y a une manière plus exacte de résumer l'évolution d'une série : c'est de lui ajuster une courbe de tendance. On choisit pour cela un *modèle* de l'évolution de la série. Par exemple, on peut prendre le modèle linéraire simple

$$x_t = a + bt$$

ou le modèle exponentiel simple

$$\mathrm{x}_t = \mathrm{a}\ \mathrm{b}^t$$

qui devient un modèle linéaire simple lorsqu'on prend les logarithmes :

$$\log x_t = \log a + (\log b)\ t$$

Le modèle exponentiel simple peut être considéré comme une version améliorée du taux de croissance exponentiel. En effet, on peut établir un parallèle entre le paramètre a et la valeur initiale x_0 et entre b et le facteur de croissance exponentielle $(1+r)$. L'estimation du modèle exponentiel simple consiste à rechercher les valeurs de x_0^* $(= a^*)$ et de $(1+r^*)$ $(= b^*)$ qui feront que les valeurs x_t^* « prédites » par la relation

$$x_t^* = x_0^*\ (1+r^*)^t = a^*\ (b^*)^t$$

« colleront » le mieux possible aux valeurs observées. De cette manière, on aura calculé un taux de croissance exponentiel qui risque moins d'être influencé par des valeurs hors tendance [19].

---

[18] Source : www.fortitude.com/data.htm.
[19] Voir Wonnacott et Wonnacott (1992, p. 513-523).

Plus généralement, après avoir choisi un modèle, on choisit les valeurs des paramètres qui rapprochent le modèle le plus possible de la réalité observée, en appliquant les méthodes statistiques appropriées. La régression linéaire est une technique qui permet de trouver les « meilleures » valeurs pour a^* et b^*. Nous y reviendrons dans la troisième partie de cet ouvrage.

QUE RETENIR ?

Nous examinerons bientôt la question de la multidimensionnalité dans la mesure. La mesure de la croissance, apparemment simple, illustre déjà quelques-unes des difficultés soulevées par la multidimensionnalité. Nous avons examiné deux façons simples de résumer en un taux unique l'évolution d'une variable dans le temps (la moyenne des taux de croissance par période et le taux de croissance exponentiel) : chacune des deux mesures présente des inconvénients. Nous avons évoqué une autre technique, l'ajustement d'une courbe de tendance, qui semble exempte des défauts des deux autres. Mais l'utilisation de cette technique a son prix : une plus grande complexité, une moindre transparence et des calculs plus lourds. Voilà bien une illustration des difficultés qui font que la mesure n'est pas qu'une science, elle est aussi un art :

- Il n'y a guère de mesure parfaite.
- Les mesures moins imparfaites sont généralement plus complexes et plus lourdes à utiliser.