

M. VAN BELLINGHEN

**Estimation par la formule de reconstitution
des données de l'indice général de la
production industrielle**

Les cahiers de l'analyse des données, tome 3, n° 3 (1978),
p. 366-370

http://www.numdam.org/item?id=CAD_1978__3_3_366_0

© Les cahiers de l'analyse des données, Dunod, 1978, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

ESTIMATION PAR LA FORMULE DE RECONSTITUTION
DES DONNÉES DE L'INDICE GÉNÉRAL
DE LA PRODUCTION INDUSTRIELLE
[ESTIM. IND. PRODUCTION]

par M. Van Bellinghen ⁽¹⁾

Nous nous proposons de présenter deux méthodes d'estimation de l'indice général de la production industrielle de l'Institut National de la Statistique et des Etudes Economiques (I.N.S.E.E.) (*).

Lors de la publication de l'indice général de l'INSEE, les indices de production d'un certain nombre de séries élémentaires manquent. Ces séries industrielles manquantes correspondent à peu près à 20% de la valeur ajoutée de la production industrielle.

Le principe de l'estimation proposée dans cet article repose sur la formule de reconstitution des données en analyse des correspondances.

Les résultats obtenus ont été comparés à ceux de l'INSEE. Le procédé utilisé par l'INSEE ne tient pas compte de l'information que pourraient apporter des séries industrielles de profil voisin des séries à estimer, ou ayant des incidences sur celles-ci.

Par contre, les méthodes décrites ci-dessous et qui utilisent la formule de reconstitution des données tiennent compte de l'évolution de l'ensemble de l'industrie française.

1 Description des méthodes employées

Les tableaux de données étudiés peuvent être décrits de la façon suivante :

The diagram shows a rectangular table with dashed lines representing rows and columns. The vertical axis is labeled 'i' and the horizontal axis is labeled 'j'. A central cell is labeled $k(i,j)$.

A chaque ligne du tableau correspond une série industrielle.

A chaque colonne du tableau correspond un mois de la période étudiée.

Un élément du tableau $k(i,j)$ correspond à la valeur ajoutée de la $i^{\text{ème}}$ série durant le $j^{\text{ème}}$ mois.

(1) Assistante à l'université de Grenoble II.

(*) Michèle Van Bellinghen : 1978 "Etude statistique des indices de la production industrielle française de Janvier 1973 à Juin 1977". Thèse de 3^o cycle. Université Pierre et Marie Curie. Paris.

La formule de reconstitution des données s'énonce ainsi :

$$k(i,j) = \frac{k(i) k(j)}{k} [1 + \sum_{\alpha} \{\lambda_{\alpha}^{-12} F_{\alpha}(i) G_{\alpha}(j)\}].$$

$k(i) = \sum_j k(i,j)$ = valeur ajoutée correspondant à la production de la série industrielle i durant la période donnée.

$k(j) = \sum_i k(i,j)$ = valeur ajoutée correspondant à la production de l'industrie durant le mois j .

$k = \sum_{i,j} k(i,j)$

λ_{α} = valeur propre d'ordre α issue de l'analyse des correspondances du tableau des données.

$F_{\alpha}(i)$ = coordonnée sur l'axe factoriel α de la série industrielle i .

$G_{\alpha}(j)$ = coordonnée sur l'axe factoriel α du mois j .

La formule de reconstitution est exacte, si la somme \sum_{α} est étendue à l'ensemble des facteurs ; elle n'est qu'approchée si l'on se borne aux facteurs de rang supérieur. Dans les applications faites ici on a utilisé les 12 premiers facteurs ce qui correspond à plus de 80% de l'inertie totale des différents tableaux analysés. Les séries industrielles manquantes sont presque toujours les mêmes. Ces séries peuvent être départagées en deux groupes :

1 . Les séries pour lesquelles il ne manque que l'indice du dernier mois par rapport à la date de publication de l'indice.

2 . Les séries pour lesquelles il manque l'indice de deux mois ou plus par rapport à la date de publication de l'indice.

On supposera, pour simplifier, que le retard des séries du groupe 2 n'est que de deux mois.

On peut donc représenter le tableau de données de la façon suivante :

		1	2
α	}		
β			
γ			

A chaque ligne du tableau correspond une série industrielle.

A chaque colonne du tableau correspond un mois de la période Janvier 70-Décembre 75.

Les séries du groupe α sont les séries pour lesquelles il ne manque aucune donnée lors de la publication de l'indice.

Les séries du groupe β sont les séries pour lesquelles il ne manque que les indices du dernier mois de la période étudiée.

Les séries du groupe γ sont les séries pour lesquelles il manque les indices des deux derniers mois de la période étudiée.

Ce tableau peut être transformé de la façon suivante :

		1	2
α			
I			A
II		C	B

A la suite des séries du groupe α, on rajoute les lignes I et II . La ligne I est constituée en termes égaux, pour chaque mois, à la somme des valeurs ajoutées des séries industrielles du groupe β. La ligne II est constituée de termes égaux, pour chaque mois, à la somme des valeurs ajoutées des séries du groupe γ.

Les estimations de A, B et C permettent d'évaluer l'indice de la production industrielle des mois 1 et 2. Deux méthodes ont été testées.

1.1 Première méthode

Les "cases manquantes" du tableau A, B et C sont estimées à partir de l'analyse des correspondances du tableau X. La mise en éléments supplémentaires des lignes I et II et des colonnes 1 et 2 permet à l'aide de la formule de reconstitution des données de trouver une estimation de A, B et C.

		1	2
α			
	X		
I			A
II		C	B

1.2 Deuxième méthode

Elle est plus précise que la première et nécessite trois analyses de correspondances.

		1	2
α			
	γ		
I			A
II		C	B

		1	2
α			
	Z		
I			A
II		C	B

		1	2
α			
	T		
I			A
II		C	B

L'analyse des correspondances du tableau Y et la mise en éléments supplémentaires de la série II et du mois 1 donnent une estimation de C (grâce à la formule de reconstitution des données).

De même, l'analyse des correspondances du tableau Z et la mise en éléments supplémentaires du mois 2 et de la série I donnent une estimation de A.

Enfin, l'analyse des correspondances du tableau T et la mise en éléments supplémentaires du mois 2 et de la série industrielle II permettent d'estimer B.

2 Résultats

Ils sont résumés dans le tableau ci-joint. Les notations sont les suivantes.

Nous appelons indice estimé numéro un de l'INSEE, l'estimation de l'indice général au mois t , effectuée par l'INSEE le mois de la parution de l'indice. L'indice estimé numéro deux de l'INSEE est l'indice réestimé par l'INSEE le mois qui suit la publication de cet indice (si nous nous rapportons aux notations précédentes, nous aurons $t = 1$, et les séries du groupe γ sont seules inconnues).

L'estimation numéro un par la formule de reconstitution des données appliquée à cette étude est l'indice du mois t , les séries des groupes β et γ sont inconnues (le mois t sera le mois 2 du paragraphe précédent).

L'estimation numéro deux par la formule de reconstitution des données est l'estimation de l'indice du mois t , en supposant connus les indices des séries des groupes α et β (pour se référer aux notations du paragraphe précédent, le mois t joue le rôle du mois 1).

3 Conclusion

Les deux méthodes testées ont deux principaux avantages :

- leur rapidité
- selon le critère de l'erreur moyenne, les résultats sont meilleurs que ceux obtenus par l'INSEE.

Ces résultats pourraient encore être améliorés en utilisant ce principe d'estimation pour les séries élémentaires dont il manque l'indice (et non pour les séries agrégées constituées par la somme de ces séries manquantes).

Cet affinement cependant alourdirait la méthode.

Ajoutons que ces résultats sont d'autant plus intéressants qu'ils ont été obtenus pour l'année 75, année de grands mouvements économiques.

Indices	Mois														
	Décembre 1974	Janvier 1975	Février 1975	Mars 1975	Avril 1975	Mai 1975	Juin 1975	Juillet 1975	Août 1975	Septembre 1975	Octobre 1975	Novembre 1975	Décembre 1975	Erreur moyenne	
Indice définitif INSEE ; Juill.1977	121.7	122.0	115.0	117.6	119.1	108.2	115.9	101.0	68.0	114.2	125.5	116.1	127.9		
	Estimation n° 1 INSEE Erreur	123.0	123.0	116.5	118.4	119.2	109.5	115.6	99.8	67.7	111.7	122.0	115.1	126.5	1.24
		1.0	1.0	1.5	0.8	0.1	1.3	0.3	1.2	0.3	2.5	3.5	1.0	1.4	
Estimation n° 2 INSEE Erreur	122.9	123.7	116.1	117.6	119.0	108.5	115.5	100.4	67.5	113.3	123.8	115.4		0.77	
	1.2	1.7	1.1	0.0	0.1	0.3	0.4	0.6	0.5	0.9	1.7	0.7			
première méthode	Estim. n° 1 Reconstit. Erreur	122.4	122.4	115.7	118.4	120.1	109.1	117.7	102.1	69.4	118.9	125.9	116.9	128.7	1.17
		0.4	0.4	0.7	0.8	0.2	0.9	1.8	1.1	1.4	4.7	0.4	0.8	0.8	
	Estim. n° 2 Reconstit. Erreur	121.5	121.9	115.1	117.2	119.2	108.5	116.5	101.3	68.5	114.2	125.4	116.3		0.24
	0.2	0.1	0.1	0.4	0.1	0.3	0.6	0.3	0.5	0	0.1	0.2			
deuxième méthode	Estim. n° 1 Reconstit. Erreur	122.4	122.4	115.6	118.5	119.9	109.1	117.3	102.1	69.8	114.7	125.8	117.0	129.1	0.90
		0.4	0.4	0.6	0.9	0.8	0.9	1.4	1.1	1.8	0.5	0.3	0.9	1.2	
	Estim. n° 2 Reconstit. Erreur	121.5	122.0	115.1	117.8	119.2	109.4	116.6	101.4	68.5	114.2	125.4	116.3		0.31
	0.2	0.0	0.1	0.2	0.1	1.2	0.7	0.4	0.5	0.0	0.1	0.2			

L'indice général (B.T.P. exclus) en 1975 : tableau comparatif des estimations n°1 et n°2 suivant la méthode originelle de l'INSEE, et suivant les deux méthodes présentées dans l'article.