

LA RÉTROPOLATION ÉCONOMÉTRIQUE DE L'INDICE DE COMMANDES FRANÇAIS

Laïla BENTOUDJA (), Cynthia GABORIEAU-FAIVRE (**)¹*

() SSP, Bureau des statistiques sur les productions et les comptabilités agricoles*

*(**) INSEE, Unité Méthodologie Statistique entreprises*

Introduction

La nécessité de fournir à la Banque Centrale Européenne des indicateurs conjoncturels relatifs à l'industrie et à la construction pour élaborer un diagnostic conjoncturel a conduit à la mise en place du règlement européen de mars 2001 sur les statistiques de court terme.

Suite à ce règlement, l'indice de commandes français a été bâti afin de permettre d'appréhender les évolutions des entrées de commandes dans l'industrie et la construction. L'un des buts poursuivis était de créer un indicateur avancé de la production industrielle de chacun des états membres.

L'indice couvre (ainsi que le stipule le règlement) cinq zones géographiques: les commandes totales par branches, les commandes à l'exportation, les commandes d'origine France, et les commandes à destination respectivement de la zone euro et hors zone euro.

Par ailleurs selon le pan de l'industrie couvert, les données par branches débutent à des années distinctes, comprises la plupart du temps entre 2001 et 2003.

En particulier l'année de base: 2000, qui devait servir de référence à l'indice, n'avait pas donné lieu à un recueil suffisant des commandes. Par conséquent son estimation s'était avérée incontournable pour de nombreuses branches.

En 2007, le recul dont on dispose - supérieur à trois ans - permet désormais d'asseoir des modèles économétriques sur les données recueillies depuis 2001, puis de les mobiliser à des fins de rétropolations de la série (année 1998 comprise afin de se conformer à la lettre du règlement).

L'amélioration qualitative de l'indice de commandes a posé de nombreux problèmes méthodologiques: choix des pondérations des séries élémentaires en vue d'une agrégation, amélioration de la lisibilité du message transmis par la prise en compte des grosses commandes passées. Le travail qui fait l'objet de cet article ne constitue donc qu'un aspect d'un effort méthodologique plus large pour former un indice de qualité à partir d'un nombre réduit d'observations.

1. Nature du travail effectué

On s'intéresse ici à la rétropolation des séries élémentaires. On a exclu d'emblée du champ de la rétropolation les gros biens d'équipement, pour lesquelles les fluctuations d'activité sont trop importantes pour utiliser un modèle économétrique (grosses commandes, ...) ainsi que quelques cas particuliers (cas où aucun rétropoleur pertinent n'est disponible, etc.).

Parmi les branches restantes, celles pour lesquelles le recueil des données débutait plus tardivement qu'en 2000, soit 220 séries, ont **donné lieu à une rétropolation de type économétrique**, le choix des rétropoleurs étant dès le départ axé sur une mobilisation d'information aussi pertinente que possible. La nomenclature « native » dans laquelle étaient présents les différents candidats rétropoleurs a joué un rôle important, puisqu'elle a conduit, pour les commandes totales, à favoriser le choix des facturations au détriment des chiffres d'affaires pourtant disponibles sans intermédiaires, car les facturations étaient déjà exprimées en Naf Rev1 et par branches. La ventilation géographique des facturations n'existant pas sur le passé, on a utilisé les exportations des douanes. Les données mensuelles sur la période 1998-2006 ont été agrégées selon la NAF par les Douanes à la demande de la division ICA.

¹ Les auteurs remercient Philippe Brion, Emmanuel Gros, Johara Khélif et Jean-François Loué pour l'aide précieuse qu'ils leur ont apportée. À l'époque où le travail a été réalisé, les deux auteurs se trouvaient en poste à la division ICA (« Indicateurs conjoncturels d'activité ») de la DSE.

Commandes totales Nouvelle rétropolation	Commandes France		
	Dédites par différence		
	Commandes export Nouvelle rétropolation	Commandes « zone euro »	Inférées par le partage « zone euro » / « hors zone euro » de la Comptabilité Nationale
		Commandes « hors zone euro »	Inférées par le partage « zone euro » / « hors zone euro » de la Comptabilité Nationale

Tableau 1 montrant l'articulation des différentes opérations pour les cinq « zones géographiques »

Les contraintes de calage: respectivement des commandes France et export sur les commandes totales et des commandes « zone euro » et « hors zone euro » sur les commandes export, conjuguées à l'absence d'un rétopoleur pertinent pour les commandes françaises, ont conduit naturellement à rétopoler d'une part les commandes à l'export (à l'aide des exportations des Douanes agrégées selon la naf Rev1) et d'autre part les commandes totales (à l'aide des facturations), puis seulement à en déduire par différence les commandes françaises.

Si par malchance les commandes françaises ainsi produites se trouvaient être négatives, alors les commandes totales et à l'export étaient égalées, de sorte que les commandes françaises soient nulles au final pour la date concernée. Rétrospectivement ce cas a été extrêmement marginal (occurrence pour deux dates seulement).

Idéalement il aurait fallu rétopoler économétriquement trois « zones géographiques » sur cinq (par exemple total, export et « zone euro ») pour en déduire les deux restantes par différence.

Or aucune information n'avait été jugée pertinente pour procéder à une rétropolation économétrique pour les séries « zone euro » ou « hors zone euro ». Ce sont donc simplement deux « zones géographiques » qui ont été concernées par les rétopolations économétriques.

Pour ce qui est des commandes « zone euro » et « hors zone euro », qui devaient une fois agrégées correspondre aux commandes à l'export, la ventilation a été effectuée de sorte à respecter les glissements annuels des parts des exportations « zone euro » et « hors zone euro » observées par la Comptabilité Nationale.

2. Mode de rétopolation des séries total et export

Dans tous les cas, le premier modèle testé a été un modèle de cointégration reliant variable rétopolée (les commandes pour une zone géographique spécifique: l'export ou le champ total) et variable rétopolante. En effet il avait été constaté que les séries de commandes, tout comme de nombreuses séries économiques, n'étaient pas stationnaires.

Or de telles séries présentent alors des tendances stochastiques ou déterministes pouvant causer des « régressions fallacieuses », pour lesquelles les statistiques usuelles ne sont pas interprétables (cas du T de Student) et peuvent éventuellement donner une appréciation trop positive, ce qui est le cas des statistiques de qualité globale de la modélisation.

Une première possibilité consiste alors à considérer l'ensemble des séries en jeu en différences premières, mais à l'usage il s'avère que cette méthode induit une perte non négligeable d'information sur le long-terme.

Le traitement classique pour de telles séries est alors d'utiliser la théorie de la cointégration (dont les fondations ont été posées par l'article d'Engle et Granger, 1987²) qui stipule qu'il existe une combinaison linéaire de ces deux séries qui est stationnaire. De tels modèles décrivent en fait l'ajustement à long terme vers cette relation. À court terme, les « erreurs » entre la réalité et cette relation de long terme sont « corrigées » partiellement par la force de rappel du modèle, de façon à ne pas rompre sa convergence à long terme vers cette relation.

Ici le modèle est intuitivement bien adapté, car les facturations doivent correspondre grosso modo aux commandes, du moins à long terme. Dans ce cas, l'élasticité à long terme des commandes aux

² Cf. la référence [3] dans la bibliographie.

facturations devrait être unitaire. Au demeurant, la différence entre commandes et facturations ne devrait pas présenter de structure particulière si ce n'est une suite de chocs « d'amplitude contrôlée » correspondant aux annulations de commandes non-observées ou à la variation des taxes. L'introduction d'une constante et d'une tendance peut se justifier dans le cas où une partie de l'activité ne se ferait pas à la commande, ou inversement, dans le cas où une partie des commandes donne lieu ultérieurement à des annulations (qui ne sont pas observées).

Après examen graphique superposant séries rétopolées et séries rétopolantes, la mise en évidence de « tendances communes » dans de nombreux cas a permis d'avaliser le choix de la méthode mettant en œuvre la théorie de la cointégration, qui conduisait à des modèles économétriques interprétables.

C'est la théorie de la cointégration telle qu'elle a évolué avec l'apport de Johansen et Juselius (cf. [5] par exemple) et non celle développée initialement par Engle et Granger (cf. [3]) qui fonde tous les développements qui suivent, car elle permet en plus de tenir compte de la présence probable d'autocorrélation dans les résidus de la régression de long-terme (si elle n'est pas prise en compte, cela peut induire un biais sensible dans les estimations).

2.1. Modèles utilisés

On forme l'hypothèse que Y est intégré d'ordre 1, c'est-à-dire qu'il est non stationnaire mais que sa différence première l'est.

Le modèle le plus général synthétisant la dynamique de court terme est donné ci-dessous, Y désignant le vecteur des logarithmes des séries:

$$\Delta Y_t = \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \mu + \gamma^* t - \Pi \begin{bmatrix} Y_{t-1} \\ 1 \\ t \end{bmatrix} + \varepsilon_t$$

où le vecteur Y regroupe commandes et facturations. Autrement dit, le différentiel du logarithme de Y dépend des variations passées de ce différentiel et de la valeur précédente du logarithme de Y considéré en niveau.

On s'est limité ici à un retard car l'inflation du nombre de paramètres avec le nombre de retards n'aurait pas permis d'estimer de façon assez robuste les paramètres au vu du nombre de dates connues.

Le modèle se réécrit donc:

$$\Delta Y_t = \Gamma_1 \Delta Y_{t-1} + \mu + \gamma^* t - \Pi \begin{bmatrix} Y_{t-1} \\ 1 \\ t \end{bmatrix} + \varepsilon_t$$

C'est ce premier modèle qui est estimé en première étape dans tout ce qui suit. Précisons son origine:

Soit ΔY_t stationnaire, suffisamment régulière pour admettre la représentation de Wold infinie suivante:

$$\Delta Y_t = \partial + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \partial + \psi(L) \varepsilon_t \quad (*)$$

$$\text{Soit } \psi^* \text{ défini par: } \psi(L) = \psi(1) + (1-L)\psi^*(L) \quad (**)$$

En itérant (*) il vient $Y_t = Y_0 + \partial t + \psi(1)(\varepsilon_1 + \dots + \varepsilon_t) + \psi^*(L)\varepsilon_t$. (* bis) , en supposant $\varepsilon_0 = 0$

Alors $\beta' Y_t$ est stationnaire si
$$\begin{cases} \beta' \psi(1) = 0 \\ \beta' \partial = 0 \end{cases} \quad (***)$$

La condition $\beta' \psi(1) = 0$ correspond au fait que $\psi(1)(\beta' \varepsilon_1 + \dots + \beta' \varepsilon_t)$ soit stationnaire³, car en multipliant par β' chaque membre de (* bis) il vient:

$$\beta' Y_t = \beta' Y_0 + \beta' \partial t + \psi(1)(\beta' \varepsilon_1 + \dots + \beta' \varepsilon_t) + \beta' \psi^*(L)\varepsilon_t,$$

Or sa variance est proportionnelle au nombre de termes t et au carré de la norme de $\psi(1)\beta'$, donc nécessairement $\psi(1)\beta'$ est nul, d'où il s'ensuit en prenant sa norme, celle de β étant non nulle, que:

$|\psi(1)| = 0$ ce qui signifie que ψ a une racine unitaire et donc que $\psi(L)^{-1}$ n'existe pas.

☞ **Conclusion:** un système cointégré ne peut être représenté par un VAR fini en différence, car $\psi(L)^{-1}$ n'existe pas, donc (*) ne peut avoir une écriture AR infinie qui pourrait s'approximer par une écriture AR finie.

Supposons à présent que le niveau de Y_t puisse être représenté par le VAR non stationnaire suivant:

$\Theta(L)Y_t = \mu + \varepsilon_t$ et admette toujours la même représentation de Wold, il vient:

$$(1-L)\Theta(L)Y_t = \Theta(1)\partial + \Theta(L)\psi(L)\varepsilon_t$$

$$(1-L)(\mu + \varepsilon_t) = \Theta(1)\partial + \Theta(L)\psi(L)\varepsilon_t$$

ce qui implique, en identifiant respectivement coefficients unitaires et de ε_t :

$$\Theta(1)\partial = 0 \text{ et } (1-L)I_N = \Theta(L)\psi(L).$$

En particulier en estimant le second terme en 1 on a: $\Theta(1)\psi(1) = 0$.

Soit π' une ligne de $\Theta(1)$, alors

$$\pi' \psi(1) = 0 \text{ et } \pi' \partial = 0.$$

D'après (***) , π' est donc un vecteur de cointégration.

Notant $(\beta_1 \dots \beta_r)$ une base linéaire de l'espace vectoriel des vecteurs de cointégration, il existe donc a de dimension $(r,1)$ tel que $\pi = [\beta_1, \dots, \beta_r]a$ soit $\pi' = a' \beta'$.

Si on applique le résultat à toutes les lignes de $\Theta(1)$ il vient: $\Theta(1) = \alpha \beta'$.

D'après (**) on a: $\Theta(1)Y_t + (I-L)\Theta^*(L)Y_t = \mu + \varepsilon_t$ soit:

$$\Theta(1)Y_t + \Theta^*(L)\Delta Y_t = \mu + \varepsilon_t \text{ en permutant les polynômes en } L.$$

On a finalement l'écriture:

$$\Delta Y_t = u_1 \Delta Y_{t-1} + \dots + u_{p-1} \Delta Y_{t-p+1} + \mu - \alpha \beta' Y_{t-1} + \varepsilon_t$$

d'où le **Théorème de représentation de Granger**.

Encadré 1 donnant l'origine du modèle général utilisé dans tout l'article

³ et en particulier de variance constante

Johansen montre que l'hypothèse de cointégration porte sur le rang de la matrice Π . Le nombre de vecteurs de cointégration d'un système est le nombre de colonnes linéairement indépendantes de Π (rang de Π) ou de manière équivalente, le nombre de valeurs propres de Π différentes de 0. La méthode de Johansen consiste donc à tester le rang de Π dans le modèle ci-dessus. Le modèle ensuite estimé est adapté au rang retenu:

2.1.1. rang(Π)=0

Il n'existe aucun vecteur de cointégration, l'analyse appropriée est celle d'un VAR en différences premières. En effet Π est alors la matrice nulle.

$$\begin{pmatrix} \Delta ICom_t \\ \Delta Factu_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} * t + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} * \begin{pmatrix} \Delta ICom_{t-1} \\ \Delta Factu_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

2.1.2. rang(Π)=2

Au contraire si Π est de plein rang alors Y est nécessairement stationnaire, si bien qu'un modèle Var (Vectoriel autorégressif) est tout à fait satisfaisant:

$$\begin{pmatrix} ICom_t \\ Factu_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} * t + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} * \begin{pmatrix} ICom_{t-1} \\ Factu_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

2.1.3. $0 \leq r \leq N$ soit ici rang(Π)=1

Nous allons nous attarder sur le cas où Π n'est pas de plein rang, tout en n'étant pas nulle pour autant. Elle peut alors s'écrire:

$$\Pi = \alpha\beta' \text{ où } \alpha \text{ et } \beta \text{ sont deux vecteurs colonne de dimension } (2,1).$$

Y n'est pas stationnaire, de sorte qu'il peut incorporer une tendance stochastique en sus d'une tendance déterministe. Or ΠY_{t-1} est stationnaire (car Y est intégrée d'ordre 1 donc sa différence première est stationnaire) de sorte que toute tendance stochastique présente dans Y doit nécessairement être contenue dans l'orthogonal de l'espace de cointégration⁴.

Quant à la constante et à la tendance qui apparaissent dans l'équation, on peut également introduire leur décomposition sur l'espace de cointégration et son orthogonal. De la sorte l'équation générale devient:

$$\begin{bmatrix} \Delta ICom_t \\ \Delta Factu_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \gamma^1_1 \\ \gamma^1_2 \end{bmatrix} * t + \Gamma * \begin{bmatrix} \Delta ICom_{t-1} \\ \Delta Factu_{t-1} \end{bmatrix} + \alpha\beta' \begin{bmatrix} ICom_{t-1} \\ Factu_{t-1} \\ 1 \\ t-1 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

(Equation A)

⁴ En effet, on a: $\Pi T_t = 0$ (où T_t désigne la tendance stochastique) d'où il s'ensuit en prenant les transposées que: $T_t' \Pi' = 0$, donc T' est orthogonal à l'image de Π' , donc à $\text{Vect}(\beta\alpha')$, d'où le résultat.

On appelle alors relation de cointégration ou relation d'équilibre de long terme le vecteur:

$$\beta' \begin{bmatrix} ICom_{t-1} \\ Factu_{t-1} \\ 1 \\ t-1 \end{bmatrix}. \alpha \text{ regroupe alors les « paramètres d'ajustement à la relation de long terme ».}$$

Les relations de cointégration peuvent inclure des tendances déterministes (sauf dans des cas très particuliers où les tendances déterministes présentes dans les données s'annulent naturellement dans la combinaison qui est formée), et également une constante.

En pratique, le vecteur de cointégration β est normalisé (c'est-à-dire que son premier coefficient β_1 vaut 1. Dans ce cas, α prend des valeurs négatives si bien que le mécanisme de correction d'erreur agit de façon intuitive à la manière d'un ressort: lorsque l'indice de commandes est

légèrement trop bas pour satisfaire la relation de long terme, un terme positif ($\alpha\beta' \begin{bmatrix} ICom_{t-1} \\ Factu_{t-1} \\ 1 \\ t-1 \end{bmatrix}$) est

entre autres ajouté à l'accroissement des commandes à la date t , favorisant ainsi un « rattrapage » dynamique vers la relation de long terme. Formulé autrement, les relations de cointégration agissent à la manière d'attracteurs si bien que tout écart à la situation d'équilibre sera rapidement résorbé.

Le modèle est estimé par une méthode de maximisation de la vraisemblance. Douze paramètres doivent alors être estimés (hors paramètres de nuisance⁵).

Cinq cas peuvent découler de la dernière formulation (A) selon la structure déterministe (tendances et constantes) qui y est effectivement incorporée⁶:

- Cas 1: le modèle interdit toute présence de termes déterministes. Cela peut difficilement permettre la modélisation de données réelles.
- Cas 2: le modèle interdit toute présence de tendance dans les données mais rend possible la présence d'une constante à condition qu'elle appartienne à la relation de cointégration.
- Cas 3: il n'y a pas de tendance présente dans le modèle. Le modèle est compatible avec une série présentant des accroissements moyens non nuls.
- Cas 4: $\gamma^1 = 0$, si bien que la tendance se trouve uniquement dans la relation de cointégration. Un tel modèle peut convenir pour des séries présentant une tendance, mais en aucun cas pour des séries quadratiques.
- Cas 5: aucune restriction n'est apportée à la présence de termes déterministes. Le modèle est alors compatible à la présence d'une tendance linéaire dans les accroissements et peut donc être ajusté sur des séries quadratiques. De tels modèles sont d'un usage délicat, car bien qu'ils puissent améliorer l'ajustement ils donnent souvent lieu à des résultats surprenants lorsqu'ils sont utilisés à des fins de prévisions.

Nous allons ici modéliser tout d'abord la série selon l'équation (A). La théorie de la cointégration développée par Johansen et Juselius nous permettra d'en déduire le modèle ad hoc.

⁵ Paramètres estimés lors de la maximisation de la log-vraisemblance sans être pour autant des paramètres d'intérêt, en l'occurrence les paramètres de second ordre du résidu.

⁶ Cette présentation est tirée de Henry D. F., Juselius K. (2000).

2.2. Tests réalisés

Une fois la structure déterministe posée, on peut définir une partition des différents cas possibles fondée sur le rang de la matrice Π ainsi que l'a montré Johansen. **Le nombre de vecteurs de cointégration d'un système est le nombre de colonnes linéairement indépendantes de Π (rang de Π) ou de manière équivalente, le nombre de valeurs propres de Π différentes de 0.** La méthode de Johansen consiste donc à tester le rang de Π dans le modèle ci-dessus. La valeur du rang détermine le type de modèle à utiliser.

Il existe deux tests de cointégration de Johansen permettant de déterminer le rang de la matrice Π , à savoir:

Le test de la λ -trace et le test du λ -max.

Pour le test de la λ -trace, l'hypothèse nulle est la suivante:

(H0): $r_0 \leq r$ (où r_0 désigne le rang de la matrice Π) tandis que l'alternative est:

vs (H1): $r_0 > r$

Quant au test du λ -max, l'hypothèse nulle est la suivante ⁷:

(H0): $r_0 = r$ tandis que l'alternative est:

vs (H1): $r_0 = r + 1$

Les valeurs critiques pour les tests dépendent du nombre de variables (dans notre cas, 2 variables: commandes et facturations ou commandes et exportations en valeur) et de la structure déterministe ajoutée au modèle à correction d'erreur (régresseur, tendance...) à la fois dans le VAR et dans les équations de cointégration. Il y a cinq tables de valeurs critiques. Elles correspondent aux cinq cas de figure présentés supra.

Cependant, les conclusions de ces tests peuvent diverger. On a préféré une approche plus générale utilisant les critères d'information AIC et BIC que l'on a cherché à minimiser à option donnée, tout en respectant le principe de parcimonie.

On a affaire lorsque l'on emploie le logiciel Eviews à un tableau comme ci-dessous:

	Option 1	Option 2	Option 3	Option 4	Option 5
Rang = 0					
Rang = 1					
Rang = 2					

avec pour chacun des croisements la valeur de la log-vraisemblance ainsi que deux critères d'information.

On opère une présélection des cas à étudier dans ce tableau en fonction du critère de parcimonie (options 1 à 3 retenues en priorité) et des résultats des tests du rang ou de la trace, ou encore en fonction des critères d'informations, critères que l'on cherche à minimiser.

2.3. Choix du modèle ad hoc en vertu du test et selon la structure déterministe

Selon le rang de la matrice Π nous avons:

⁷ Les conclusions de ces tests peuvent ne pas coïncider, ce qui nous a conduit à une présélection assez large de modèles en nous laissant guider par les critères d'information AIC et BIC.

A - Lorsque le rang de la matrice Π vaut 0, il n'existe aucun vecteur de cointégration, l'analyse appropriée est celle d'un VAR en différences premières. L'équation générale devient:

$$\begin{bmatrix} DLCom(k) \\ DLFactu(k) \end{bmatrix} = \begin{bmatrix} c(1,1) & c(1,2) & c(1,3) & c(1,4) \\ c(2,1) & c(2,2) & c(2,3) & c(2,4) \end{bmatrix} \begin{bmatrix} DLCom(k-1) \\ DLFactu(k-1) \\ temps(k-1) \\ 1 \end{bmatrix}$$

Certains coefficients peuvent être nuls selon les options choisies. Ici nous avons en fonction de la structure déterministe choisie (cas 1, ...)

	C(1,1) et C(2,1)	C(1,2) et C(2,2)	C(1,3) et C(2,3)	C(1,4) et C(2,4)
Cas 1			=0	=0
Cas 2			=0	
Cas 3			=0	
Cas 4				
Cas 5				

B - Lorsque le rang de Π vaut 1, il existe un vecteur de cointégration β et on peut factoriser Π de la façon suivante: $\Pi = \alpha * \beta'$. L'équation générale devient alors:

$$\begin{bmatrix} DLCom(k) \\ DLFactu(k) \end{bmatrix} = \begin{bmatrix} a(1,1) \\ a(1,2) \end{bmatrix} \begin{bmatrix} b(1,1) & b(1,2) & b(1,3) & b(1,4) \end{bmatrix} \begin{bmatrix} LCom(k-1) \\ LFactu(k-1) \\ temps(k-1) \\ 1 \end{bmatrix} + \begin{bmatrix} C(1,1) & C(1,2) & C(1,3) & C(1,4) \\ C(2,1) & C(2,2) & C(2,3) & C(2,4) \end{bmatrix} \begin{bmatrix} DLCom(k-1) \\ DLFactu(k-1) \\ 1 \\ temps(k-1) \end{bmatrix}$$

Dans ce cas, il existe une relation de cointégration entre lcom et F: on estime un modèle à correction d'erreur. Certains coefficients peuvent être nuls selon les options choisies. Plus précisément:

	A(1,1)	B(1,1)	B(1,2)	B(1,3)	B(1,4)	C(1,1)	C(1,2)	C(1,3)	C(1,4)
Cas 1				=0	=0			=0	=0
Cas 2				=0				=0	=0
Cas 3				=0					=0
Cas 4									=0
Cas 5									

(On n'a présenté ici que les restrictions portant sur les coefficients de la première ligne. Les autres s'en déduisent par symétrie.)

C'est uniquement dans ce cas que les cinq options d'intégration de termes déterministes se distinguent bien entre elles.

C - Lorsque le rang de la matrice Π vaut 2, la matrice est de plein rang, ce qui implique que toutes les variables sont stationnaires et que l'on peut se contenter d'estimer un VAR en niveau. En effet il existe alors deux combinaisons linéaires indépendantes des facturations et des commandes qui sont stationnaires.

L'équation générale devient:

$$\begin{bmatrix} LCom(k) \\ LFactu(k) \end{bmatrix} = \begin{bmatrix} c(1,1) & c(1,2) & c(1,3) & c(1,4) \\ c(2,1) & c(2,2) & c(2,3) & c(2,4) \end{bmatrix} \begin{bmatrix} LCom(k-1) \\ LFactu(k-1) \\ temps(k-1) \\ 1 \end{bmatrix}$$

Nous avons alors un tableau analogue:

	C(1,1)	C(1,2)	C(1,3)	C(1,4)
Cas 1			=0	
Cas 2			=0	
Cas 3			=0	
Cas 4				
Cas 5				

(On n'a présenté ici que les restrictions portant sur les coefficients de la première ligne. Les autres s'en déduisent par symétrie.)

On effectue alors une comparaison interne à cette présélection au vu de critères de pertinence statistique et économique des modèles.

2.4. Critères économiques et statistiques d'acceptabilité

Les équations obtenues spontanément pour décrire les commandes n'étant pas toujours acceptables, on explicite ci-dessous les conditions supplémentaires qui ont été posées pour accepter une équation.

2.4.1. L'élasticité des commandes aux facturations doit avoir une valeur acceptable

L'équation sélectionnée avec la méthode de Johansen décrit les commandes (passées) en fonction des facturations (futures). Chaque variable étant modélisée en logarithme, on s'attend à ce que cette équation conduise à long terme à une élasticité unitaire des commandes (passées) par rapport aux facturations (futures) (ce qui correspond à un écart entre commandes et facturations qui procède essentiellement d'un changement d'unité). En pratique, nous avons admis des modèles dont l'élasticité était comprise entre 0 et 2. Retenir un modèle avec une élasticité nulle revient dans les faits à prolonger les séries existantes par une tendance.

Notons:

- F : le logarithme des facturations,
- $Icom$: le logarithme de l'indice des commandes.

Le temps ayant été inversé, le modèle estimé s'écrit: $P(L)(F) = Q(L)(Icom)$, où P et Q désignent des polynômes et L l'opérateur retard: $L(Y_t) = Y_{t-1}$.

Lorsque le modèle est en taux de croissance on peut écrire: $P(L) = (1-L)P1(L)$ et $Q(L) = (1-L)Q1(L)$ où P1 et Q1 sont des polynômes.

L'élasticité de long terme des commandes aux facturations vaut $Q(1)/(P(1))$ si le modèle est écrit en niveau et $Q1(1)/P1(1)$ si le modèle est écrit en taux de croissance

2.4.2. Lorsque l'on retient un modèle en niveau, sa dynamique doit être convergente

Lorsque le modèle est en niveau, il peut être écrit sous forme d'un modèle à correction d'erreur, c'est à dire:

$$(1-L)(Icom) = \alpha L [Icom - P(1)/Q(1).(F) + b + c * TEMPS] + Q2(L)(1-L)(F) + P2(L)(1-L)(Icom)$$

La dynamique de ce modèle est convergente vers la cible de long terme si et seulement si $-2 < \alpha < 0$. La convergence est monotone si $-1 < \alpha < 0$, elle est oscillatoire si $-2 < \alpha < -1$.

- Pour $\alpha > 0$, le modèle diverge: tout écart à la cible de long terme entraîne un écart de même signe plus grand les trimestres suivants.
- Pour $\alpha = -2$, le modèle est oscillatoire et ne converge pas.
- Pour $\alpha < -2$, le modèle diverge de façon oscillatoire: tout écart à la cible induit un écart plus grand de signe opposé le trimestre suivant.

2.4.3. Les délais d'ajustement doivent être interprétables

Le délai moyen d'ajustement des commandes aux facturations (DMA) s'écrit:

- $DMA = P'(1)/P(1) - Q'(1)/Q(1)$, lorsque le modèle est en niveau
- $DMA = P1'(1)/P1(1) - Q1'(1)/Q1(1)$, lorsque le modèle est en taux.

En effet si on a:

$$Y = P(L)X$$

Alors $DMA = P'(1)/P(1)$, car par définition:

$$DMA = \frac{\text{somme _ pondérée _ des _ délais _ par _ les _ rattrapages}}{\text{rattrapage _ cumulé}}$$

Or

$$Y_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p}$$

donc on a:

$$DMA = \frac{0 * a_0 + 1 * a_1 + 2 * a_2 + 3 * a_3 + \dots + p * a_p}{a_0 + a_1 + \dots + a_p}$$

En effet au vu de l'expression de Y en fonction du polynôme autorégressif de X, il vient que si X subit un choc de 1 à la date 0, alors un premier rattrapage a lieu immédiatement, d'importance a_0 . Puis à la date $t=1$ un second rattrapage a lieu, valant $(a_0 + a_1) - a_0 = a_1$, etc.

En identifiant formellement les dérivées de la fraction rationnelle P/Q et de son expression polynômiale « infinie », on trouve bien le résultat souhaité.

L'unité du DMA est l'unité de temps qui rythme le recueil des données.

Le temps ayant été inversé, les délais estimés représentent les délais d'ajustement des facturations aux commandes ; ils représentent peu ou prou la longueur des cycles de production et les délais moyens de paiement. Un délai d'ajustement anormalement élevé peut être lié à une mauvaise spécification. Est considéré anormalement élevé ici un délai qui serait supérieur au cycle de production, soit 6 mois en général.

2.4.4. Les transformations opérées pour obtenir une équation acceptable

Lorsque la démarche précédente ne permet pas d'obtenir une équation acceptable, les opérations suivantes sont tentées successivement jusqu'à obtenir une élasticité acceptable des commandes aux facturations et écarter les modèles divergents en niveaux:

- si l'élasticité est négative ou supérieure à 2, la contraindre à 0, supprimer les variables non significatives: on revient alors à un modèle autorégressif sur les commandes ;
- traiter les points aberrants et au besoin, raccourcir la période d'estimation,
- pour des modèles en niveau à dynamique explosive, traiter les points aberrants, contraindre l'élasticité à long terme, au besoin raccourcir la période d'estimation, voire revenir à un modèle en taux de croissance.

2.4.5. Récapitulatif de la démarche employée:

La démarche se décompose en huit étapes:

1. On désaisonnalise la série à rétopoler et la série rétopolante ;
2. On inverse le temps ;
3. On teste la cointégration des séries, ce qui permet de déterminer le rang de la matrice de cointégration et le système d'équations à estimer. A cette étape, les critères BIC et AIC sont également pris en compte ; parmi les quinze cas possibles un maximum de quatre cas sélectionnés grâce à ces critères sera étudié ensuite.
4. On estime le modèle retenu par la méthode du maximum de vraisemblance, conditionnellement au nombre de relations de cointégration indépendantes⁸ ;
5. On vérifie l'interprétabilité du modèle et on contrôle l'allure graphique de la série ;
6. On utilise le modèle estimé en prévision pour calculer la série rétopolée sur le passé ;
7. On rétablit l'ordre chronologique ;
8. On réestime la série brute.

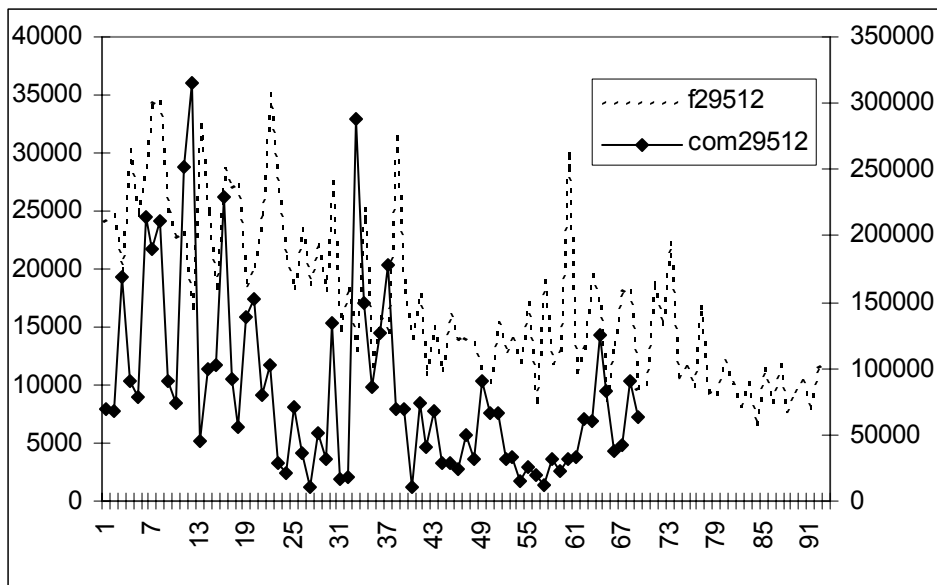
3. Exemple pratique de rétopolation d'une série: cas des commandes de machines pour la sidérurgie

La série des commandes à l'exportation de la branche « machines pour la sidérurgie C295A01 » (série codée 29512 pour le logiciel Eviews) est l'une des séries élémentaires de l'ICOM qui débute en 2001⁹. La série d'exportations de machines pour la sidérurgie issue des douanes était candidate pour la rétopoler.

Avant toute tentative de modélisation, il a été procédé à un examen graphique des séries afin de juger de la pertinence du candidat au rôle de série rétopolante. Un certain parallélisme des deux courbes laisse augurer une bonne qualité de la rétopolation (voir graphique 1).

⁸ la méthode du MV pour ce type de modèles a été développée Johansen et Juselius.

⁹ Certaines séries rétopolées ne commençaient qu'en 2002 voire 2003. Les séries connues seulement à partir de 2004 ont été exclues d'emblée du champ de la rétopolation faute de disposer de d'une période d'observation suffisante (3 ans).



Graphique 1: exportations et commandes de « machines pour la sidérurgie »¹⁰

Le logiciel Eviews fournit un récapitulatif des log vraisemblances des modèles à équations multiples qui peuvent être envisagés selon le rang de la matrice et l'option retenue, ainsi que les critères BIC et AIC.

¹⁰ Note d'explication: F29512 désigne les exportations et com29512 désigne les commandes.

Sample: 1990:01 1999:01					
Included observations: 67					
Series: LCOM29512 LFACTU29512					
Lags interval: 1 to 1					
Option	1	2	3	4	5
Data Trend:	None	None	Linear	Linear	Quadratic
Rank or	No Intercept	Intercept	Intercept	Intercept	Intercept
CE Models	No Trend	No Trend	No Trend	Trend	Trend
		Log Likelihood by Model and Rank			
0	-99.70418	-99.70418	-99.61553	-99.61553	-99.60306
1	-88.76307	-87.38680	-87.34707	-84.31890	-84.31459
2	-88.67442	-84.07492	-84.07492	-75.03645	-75.03645
	Akaike Information Criteria by Model and Rank				
0	3.095647	3.095647	3.152702	3.152702	3.212032
1	2.888450	2.877218	2.905883	2.845340	2.875063
2	3.005207	2.927609	2.927609	2.717506	2.717506
		Schwarz Criteria by Model and Rank			
0	3.227271	3.227271	3.350137	3.350137	3.475278
1	3.151697	3.173371	3.234941	3.207305	3.269933
2	3.400077	3.388292	3.388292	3.244000	3.244000
L.R. Test:	Rank = 1	Rank = 1	Rank = 2	Rank = 2	Rank = 2

Tableau 2: comment procéder à la présélection des modèles

Chacun de ces deux critères a été examiné pour chaque série et une présélection de modèles a ainsi pu être opérée en retenant ici deux modèles parmi ceux obtenant les meilleurs scores (i.e. les plus faibles valeurs) pour les critères BIC et AIC et en respectant le principe de parcimonie.

Parmi les modèles sans tendance (options 1, 2 et 3), les tests amènent à retenir un rang de cointégration égal à 1 ou 2. Lorsque l'on teste des modèles avec constante et tendance (option 4) ou avec constante et tendance quadratique (option 5), on retient un rang de cointégration égal à 2. Comme il est difficile de justifier l'intégration d'une tendance quadratique dans les données (option 5) on écarte ce modèle au profit du modèle à tendance linéaire dont les résultats sont aussi bons. Le critère d'information AIC est minimum pour le rang 2 combiné à l'option 4. Le critère d'information BIC

est minimum pour le rang 1 combiné à l'option 1. Ce sont donc finalement les modèles avec rang 2 et option 4 d'une part, et rang 1 et option 1 d'autre part, qui ont été présélectionnés. L'estimation statistique de ces deux modèles conduit aux résultats ci-dessous:

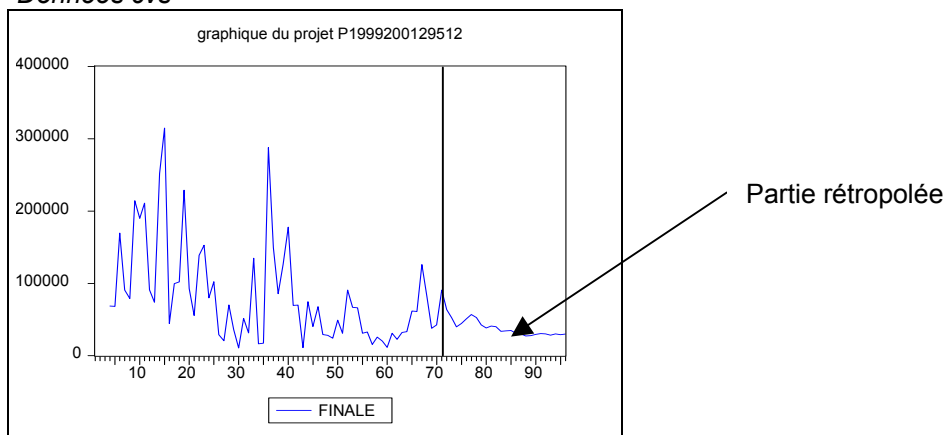
R101			R204			
Cointegrating Eq:	CointEq1			LCOM29512	LFACTU29512	@TREND
LCOM29512(-1)	1.000000		LCOM29512(-1)	0.294451	0.026931	8.06E-14
				(0.12022)	(0.04445)	(2.4E-14)
LFACTU29512(-1)	-1.124816			(2.44930)	(0.60592)	(3.32113)
	(0.01236)					
	(-90.9921)		LFACTU29512(-1)	-0.095252	-0.076214	-5.12E-13
Error Correction:	D(LCOM29512)	D(LFACTU29512)		(0.33849)	(0.12515)	(6.8E-14)
CointEq1	-0.566799	0.158413		(-0.28140)	(-0.60900)	(-7.48857)
	(0.14837)	(0.06179)				
	(-3.82028)	(2.56358)	@TREND	-0.014383	-0.012788	1.000000
				(0.00635)	(0.00235)	(1.3E-15)
D(LCOM29512(-1))	-0.098571	-0.109992		(-2.26588)	(-5.44916)	(7.8E+14)
	(0.12928)	(0.05384)				
	(-0.76247)	(-2.04280)	C	9.180836	10.68517	1.000000
				(3.63901)	(1.34542)	(7.3E-13)
D(LFACTU29512(-1))	-0.307728	-0.487479				
	(0.25243)	(0.10513)		(2.52290)	(7.94191)	(1.4E+12)
	(-1.21908)	(-4.63672)				
R-squared	0.311353	0.373431	R-squared	0.270747	0.469589	1.000000
Adj. R-squared	0.289833	0.353850	Adj. R-squared	0.236564	0.444726	1.000000
Sum sq. resids	35.84624	6.218125	Sum sq. resids	33.66534	4.601828	1.37E-24
S.E. equation	0.748397	0.311702	S.E. equation	0.725273	0.268148	1.46E-13
F-statistic	14.46791	19.07177	F-statistic	7.920365	18.88705	4.08E+29
Log likelihood	-74.11617	-15.43187	Log likelihood	-72.58449	-4.923982	
Akaike AIC	2.301975	0.550205	Akaike AIC	2.252485	0.262470	
Schwarz SC	2.400693	0.648923	Schwarz SC	2.383044	0.393029	
Mean dependent	-0.001029	-0.014107	Mean dependent	11.00435	9.804850	34.50000
S.D. dependent	0.888079	0.387769	S.D. dependent	0.830071	0.359850	19.77372
Determinant Residual Covariance		0.048504	Determinant Residual Covariance		6.75E-28	
Log Likelihood		-88.76307	Log Likelihood		1837.686	
Akaike Information Criteria		2.888450	Akaike Information Criteria		-53.69665	
Schwarz Criteria		3.151697	Schwarz Criteria		-53.30497	
D(LCOM29512) = - 0.5667989397*(LCOM29512(-1) - 1.12481595*LFACTU29512(-1)) + 0.09857092551*D(LCOM29512(-1)) + 0.307728373*D(LFACTU29512(-1))						
D(LFACTU29512) = 0.1584125468*(LCOM29512(-1) - 1.12481595*LFACTU29512(-1)) + 0.1099920993*D(LCOM29512(-1)) + 0.4874789213*D(LFACTU29512(-1))						

Tableau 3: les résultats de l'estimation par maximum de vraisemblance des deux modèles

On retient le modèle de rang 1 option 1 qui est surligné ci-dessus. En effet, l'élasticité des commandes aux facturations à long terme (1,13) est proche de 1 dans ce modèle, alors qu'elle est très faible et de mauvais signe $(-0,12 = -0.0952/(1-0.2944))$ dans l'autre modèle. Le modèle retenu étant un modèle à correction d'erreur, l'élasticité de long terme apparaissant dans le terme de rappel. Le coefficient du terme de rappel assure la convergence dynamique du modèle. Celui-ci s'écrit:

$$\begin{pmatrix} \Delta LICom_t \\ \Delta LF_t \end{pmatrix} = \begin{bmatrix} -0,1 & -0,1 \\ -0,3 & -0,5 \end{bmatrix} \cdot \begin{pmatrix} \Delta LICom_{t-1} \\ \Delta LF_{t-1} \end{pmatrix} + \begin{bmatrix} -0,6 \\ 0,2 \end{bmatrix} \cdot [1,0 \quad -1,13 \quad 0 \quad 0] \cdot \begin{bmatrix} LICom_{t-1} \\ LF_{t-1} \\ 1 \\ t-1 \end{bmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Données cvs

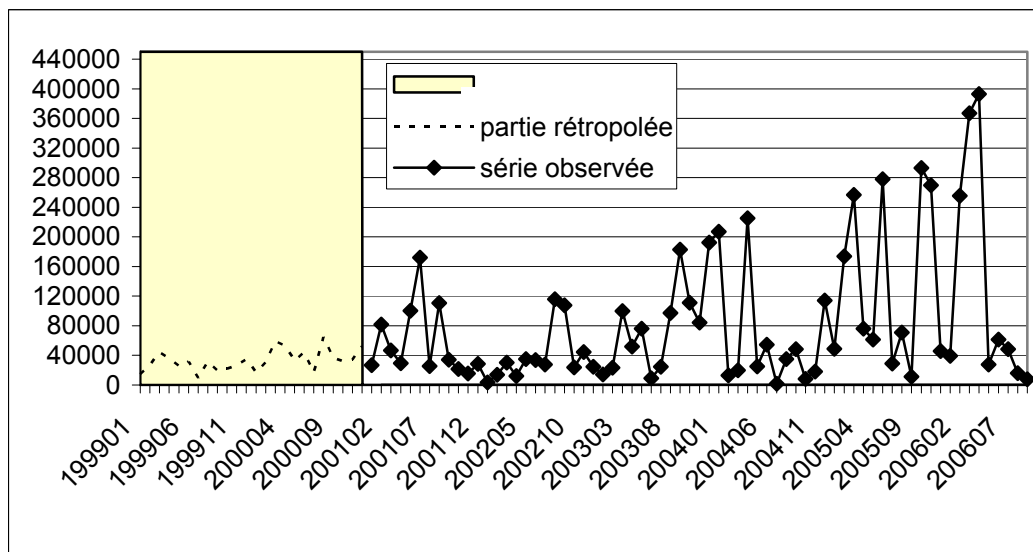


Graphique 2: Les commandes de machines sidérurgiques à l'exportation observées et rétropolées

On vérifie (Graphique 2) qu'il n'y a pas rupture de la tendance entre les données rétropolées et les données observées. Toutefois, la partie rétropolée de la série est plus lisse que la partie observée. Ce défaut apparaît assez souvent dans les rétropolations économétriques qui ont été menées.

Enfin on rétablit l'ordre chronologique, et on applique à la partie rétropolée les coefficients saisonniers moyens de la série brute sur la période d'observation (Graphique 3).

Données brutes



Graphique 3: les commandes à l'exportation de "machines pour la sidérurgie" C295A01

Bilan de la rétropolation

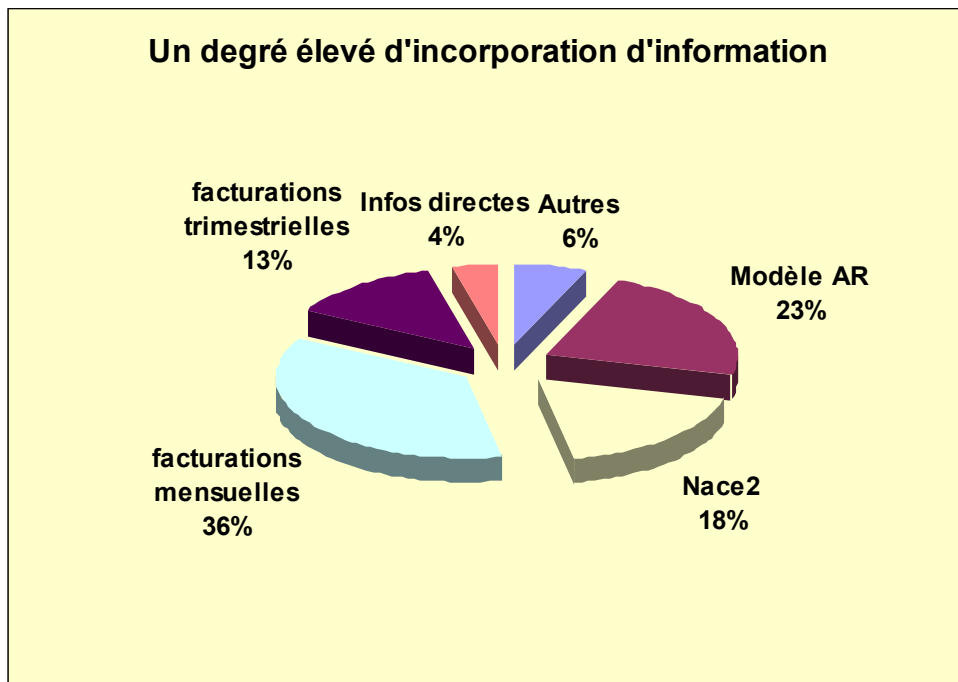
Au final on a abouti au décompte suivant des types de traitements utilisés pour les différentes séries:

Nombre de séries rétropolées

		Année de début de collecte					Total	
		≤2000	2001	2002	2003	≥2004		
Série rétropolante	Régression économétrique	Facturations mensuelles		32	15	3		50
		Facturations trimestrielles		6	11	-		17
		Douanes mensuelles		52	35	3		90
		Modèle auto-régressif		36	17	10		63
		IPI_IPVI			2			2
		Grosses commandes (presse)				10		10
		Nace2 ¹¹		16	10		10	36
		Sans objet	192		2	4		198
		Total	192	142	92	30	10	466

Tableau 4 de synthèse des traitements réalisés sur les séries de l'ICom

Concernant les commandes totales, le graphique suivant est éloquent: plus des trois quarts des rétropolations des séries ont pris en compte une information externe.



Graphique 4: une information externe très largement mobilisée

Par ailleurs, la volatilité (et sans doute bientôt son corollaire à savoir les révisions de l'indice) s'est trouvée sensiblement atténuée grâce notamment à l'abandon des pondérations de la Naf35, de sorte que l'indice se trouve désormais bien plus cohérent avec les évolutions combinées de l'indice de production industrielle et l'indice de prix de vente industriels.

¹¹ Cas où l'évolution de l'agrégat des séries déjà rétropolées a été reprise pour mener à bien la rétropolation (pas de série cointégrante satisfaisante)

Conclusion

Le travail présenté dans cet article n'a constitué qu'une partie d'un effort méthodologique plus large visant à améliorer la qualité de l'indice. La réflexion conduite a également abordé la question de l'exhaustivité des sources et du choix des pondérations élémentaires (pour plus de détails sur ses derniers, se référer à la référence [1] donnée en bibliographie).

L'amélioration sera poursuivie à l'avenir à mesure que de nouvelles données seront disponibles.

Bibliographie

[1] Bentoudja L., Gaborieau-Faivre C., « The french order index backcasting through the use of cointegration theory », *NTTS conference, February 2009*

[2] Bernard A., Roy G., « Etude des distorsions de niveau des tests de Johansen pour la cointégration », *Cahier de recherche 03-08 Département d'économie, Université de Sherbrooke, 2003*

[3] Engle R.F., Granger C.W.J. "Co-Integration and Error Correction: Representation, Estimation, and Testing", *Econometrica*, 1987

[4] Henry D. F., Juselius K. « Explaining cointegration analysis: part II », *Discussion papers, Department of Economics, University of Copenhagen, 2000*

[5] Johansen S. Juselius K. "Maximum likelihood estimation and inference on cointegration - with applications to the demand for money", *Oxford bulletin of Economics and Statistics*, 1990