

# UN NOUVEL INDICATEUR SYNTHÉTIQUE POUR LES ENQUÊTES D'OPINION

Olga A. Vasechko <sup>1</sup>, Michel Grun-Réhomme <sup>2</sup>

## RÉSUMÉ

Les enquêtes d'opinion, répétitives dans le temps ont pour but d'estimer l'évolution de l'opinion sur une problématique donnée. Plus précisément, on espère obtenir un savoir sur l'évolution d'un facteur subjectif, inobservable quantitativement, à partir de différentes questions qualitatives ordonnées dont on suppose qu'elles sont corrélées avec ce facteur et entre elles. Ces différentes questions (ou séries temporelles) sont alors résumées par un indicateur qui décrit en une variable unique la composante commune à ces séries et qui estime ce facteur inobservable.

Dans cet article, on propose un nouveau mode de calcul d'un indice synthétique robuste, qui utilise une statistique de rang basée sur la médiane et qui est fondée sur une relation d'ordre lexicographique à partir des données individuelles. Un intervalle de confiance est construit autour de cette statistique. Cet indicateur est mis en œuvre sur des données conjoncturelles ukrainiennes dans l'industrie et comparé à l'indicateur synthétique classique obtenu à l'aide d'une analyse factorielle dynamique. Ce procédé est généralisable à des enquêtes dont les questions sont qualitatives ordonnées et corrélées entre elles.

MOTS CLES: Enquêtes conjoncturelles; indicateur synthétique; séries temporelles ; statistique d'ordre.

## ABSTRACT

Temporal repetitive surveys of opinion prove useful at estimating trends for opinions on given problems. More precisely, ordered qualitative variables provide insight about unobservable factor. Previous research shows that the ordered variables are often likely correlated. The presence of correlation between the variables enables the identification of an aggregate index.

In this paper, we propose a new method of calculating robust aggregative indexes that uses a rank statistics, specifically medians, based on a lexicographic order relation of the individual data. A confidence interval is constructed around this statistics. We implement this method using Ukrainian short-term time series in manufacturing and compared the findings with the classical synthetic indicator calculated with a dynamic factorial analysis. This procedure can be applied to surveys that contain correlated ordered qualitative responses.

KEY WORDS: Indicators, Robust Statistics, Short-term Survey, Time Series,

## 1. INTRODUCTION

De nombreuses enquêtes d'opinion sont réalisées dans différents domaines. Ces enquêtes, répétitives dans le temps, souvent à intervalles réguliers, ont pour but d'estimer l'évolution de l'opinion sur une problématique donnée. Il s'agit de mesurer l'évolution d'un sentiment, d'une impression, d'un facteur subjectif par une approche qualitative. Plus précisément, on espère obtenir un savoir sur l'évolution de ce facteur, inobservable quantitativement, à partir de différentes questions qualitatives ordonnées dont on suppose qu'elles sont corrélées avec ce facteur et entre elles. Ces différentes questions (ou séries temporelles, d'un point de vue économétrique) sont alors résumées par un indicateur qui décrit en une variable unique la composante commune à ces séries et qui estime ce facteur inobservable.

<sup>1</sup> Vasechko Olga, Research Institute of Statistics, 3 Shota Rustaveli str., 01023 Kyiv, Ukraine, E-Mail: [O.Vasechko@ukrstat.gov.ua](mailto:O.Vasechko@ukrstat.gov.ua)

<sup>2</sup> Grun-Réhomme, Michel, Université Paris 2, ERMES-UMR7181-CNRS, 12 place du Panthéon, 75005 Paris, France,  
E-Mail [grun@u-paris2.fr](mailto:grun@u-paris2.fr)

### 1.1 Les enquêtes conjoncturelles

Ces enquêtes constituent une source d'information précieuse pour l'analyse des fluctuations économiques à court terme. Elles permettent d'obtenir des informations économiques sur le passé récent, la situation présente et les perspectives à court terme. Elles sont en général réalisées par sondage stratifié à probabilités inégales auprès des chefs d'entreprise de différents secteurs d'activité (industrie, construction, commerce, transport, services). Ces enquêtes mensuelles ou trimestrielles sont harmonisées au niveau européen. Les enquêtes trimestrielles complètent les enquêtes mensuelles par des questions supplémentaires.

Les enquêtes de conjoncture présentent trois avantages :

l'information est recueillie directement auprès des entreprises, elles sont très peu révisées et la diffusion de l'information est rapide. Les opinions des chefs d'entreprise sur leur activité donnent une bonne estimation des grands mouvements de la conjoncture.

Dans ces enquêtes d'opinion, la plupart des questions sont qualitatives et concernent les évolutions de différents facteurs économiques de l'entreprise. Pour un même secteur d'activité (par exemple l'industrie), plusieurs questions sont posées mensuellement. Elles portent sur l'évolution de la production de l'entreprise, le carnet de commandes personnel et globalement pour le secteur, le niveau des stocks de produits finis, l'évolution probable des prix de vente, les perspectives de l'activité,... Dans les enquêtes trimestrielles, on ajoute des questions relatives à l'évolution des commandes (sur les trois derniers mois), l'évolution des effectifs totaux dans l'entreprise, les perspectives globales des salaires,...

En général, le répondant a le choix entre trois modalités : évolution à la hausse (amélioration, favorable, niveau supérieur à la normale), stabilité (normale) ou évolution à la baisse (diminution, défavorable, niveau inférieur à la normale).

Pour analyser ces enquêtes, les méthodes sont bien banalisées et utilisent à la fois l'approche multidimensionnelle et les séries temporelles. Les analyses sont macro économiques, que ce soit à travers les séries des soldes d'opinion ou des comparaisons avec des séries d'indicateurs économiques (par régression linéaire). Les résultats sont agrégés selon les strates ou plus globalement. La confidentialité étant par ailleurs toujours de rigueur.

## **1.2 Objectif et Plan**

Dans cet article, on propose un nouveau mode de calcul d'un indice synthétique robuste, qui utilise une statistique de rang avec la médiane et qui est fondé sur une relation d'ordre lexicographique à partir des données individuelles. Un classement des observations est ainsi possible. Un intervalle de confiance est construit autour de cette statistique. Cet indicateur a l'avantage d'être simple à calculer. Reste à prouver que son efficacité est au moins du même ordre que les autres méthodes. Il est mis en œuvre sur des données conjoncturelles ukrainiennes. Ce procédé est généralisable à des enquêtes dont les réponses sont qualitatives ordonnées.

Ce travail est organisé comme suit. La section 2 rappelle les différents principes de construction des indicateurs synthétiques pour les enquêtes conjoncturelles. La section 3 présente la construction de ce nouvel indicateur synthétique et d'un intervalle de confiance associé. La section 4 concerne la comparaison de ces méthodes sur les soldes d'opinion à partir de données des enquêtes conjoncturelles dans l'industrie en Ukraine. Une conclusion et une courte bibliographie terminent cette présentation.

## **2. LES MÉTHODES CLASSIQUES**

Pour les enquêtes conjoncturelles, la méthode classique, qui repose sur ces données agrégées, consiste à calculer pour chaque question un solde d'opinion (différence entre le pourcentage de réponses « en hausse » et le pourcentage de réponses « en baisse »). Cette démarche revient à calculer une simple moyenne arithmétique, où une réponse à la hausse correspond à la valeur +1, une réponse de stabilité à 0 et une réponse en baisse à -1. Ces moyennes peuvent être pondérées par l'effectif salarié. Les séries sont ensuite corrigées des variations saisonnières. Puis cette information est alors résumée par un seul indicateur synthétique à l'aide d'une analyse factorielle dynamique (Gouriéroux C., Monfort A., 1997, Doz C., Lenglart F., 1995, 1999), avec une estimation par le filtre de Kalman. Cet indicateur résume les corrélations entre les variables et permet d'effectuer des prévisions sur la tendance à court terme.

Tout récemment, d'autres auteurs (Mitchell J., Smith R.J., Weale M.R., 2004, 2005) ont proposé une alternative basée sur des données individuelles. Ces nouveaux indices MSW (Mitchell, Smith, Weale) peuvent remplacer les soldes d'opinion. Ils

ont comme objectif, outre celui de synthétiser l'évolution conjoncturelle, d'effectuer une prévision sur l'indice de production (par exemple l'IPI dans l'industrie), qui ne sera connu que plus tard à travers les comptes trimestriels. Pour construire l'indice MSW à une date  $t$ , les valeurs passées de cet indice de production ainsi que le nombre de fois où l'entreprise a donné la même réponse qu'à cette date  $t$  sont prises en compte. La suite de la démarche est la même que celle utilisée avec les soldes d'opinion. Cette démarche est encouragée par Eurostat (European Commission, 2004). Les travaux actuels montrent que cet indice donne de meilleurs résultats que les soldes d'opinion sur les données allemandes et anglaises, mais de moins bons résultats sur les données suédoises et portugaises. Des études sont encore en cours.

### 3. CONSTRUCTION DU NOUVEL INDICATEUR

Chaque mois (ou trimestre)  $t$ , l'enquête d'opinion comporte  $p$  questions qualitatives ayant chacune  $m$  modalités ordonnées. On suppose que les questions retenues ont la même importance dans une perspective d'analyse et de prévision à court terme, qu'elles sont fortement corrélées et que les répondants fournissent une réponse à toutes les questions (dans le cas contraire une imputation peut être utilisée pour les non réponses partielles).

#### 3.1 Définitions et notations

On note  $x_{i,t}^j$  la réponse de l'individu  $i$  le mois (ou le trimestre)  $t$  à la question  $j$ ,

Pour  $i = 1, \dots, n_t$ , où  $n_t$  correspond au nombre d'unités répondantes à la date  $t$ . D'une date à l'autre, la valeur de  $n_t$  est sensiblement la même.

$j = 1, \dots, p$  correspond au numéro de la question posée.

$t = 1, \dots, T$ , où  $T$  correspond à la dernière période d'observation.

$x_{i,t}^j \in \{1, 2, \dots, m\}$ , où 1 correspond à la première modalité, la plus défavorable. Chaque variable est codée de la même façon.

Dans un premier temps, on fixe  $t$  (et on n'indique plus cette référence dans les notations pour simplifier les expressions).

On note encore  $x_i = (x_i^{(1)}, \dots, x_i^{(p)})$  le vecteur des réponses « ordonnées » de l'individu  $i$ , à savoir  $x_i^{(j)} \leq x_i^{(j+1)}$  (et donc  $x_i^{(j)}$  ne correspond pas nécessairement à la réponse de la question  $j$ , mais à la  $j^{\text{ème}}$  réponse triée dans l'ordre croissant).

Cette approche a un sens avec l'hypothèse de corrélations entre les questions.

Puis, on définit sur l'ensemble des vecteurs de  $p$  réponses ordonnées, une relation d'ordre qui correspond à l'ordre lexicographique.

#### Définition

Soient deux  $p$ -uplets,  $x = (x_1, \dots, x_p)$  et  $y = (y_1, \dots, y_p)$ , où  $x_i \leq x_{i+1}$  et  $y_i \leq y_{i+1}$ .

On dit que  $x < y$  si et seulement si :  $\exists j \in \{1, \dots, p\}$  tel que  $x_j < y_j$  et  $\forall i \in \{1, \dots, j-1\}$   $x_i = y_i$ .

On a  $x = y$  si et seulement si  $\forall i \in \{1, \dots, p\}$   $x_i = y_i$ . On obtient ainsi une relation d'ordre totale.

Dans ce cas, le nombre de  $p$ -uplets ordonnés (« mots ») différents que l'on peut obtenir correspond au nombre de combinaisons avec répétition, soit  $C_{m+p-1}^p$ .

#### 3.2 Intervalle de confiance pour un rang

Soit  $X$  la variable réponse pour un échantillon de taille  $n$  et  $x_i$  la réponse ordonnée de l'unité  $i$ .

$X$  est une variable polytomique ordonnée, que l'on peut assimiler à une variable numérique discrète. On note  $x_{(i)}$  la  $i^{\text{ème}}$  observation de  $X$  rangée dans l'ordre croissant, pour tout  $i = 1, \dots, n$ .

On cherche un intervalle de confiance au niveau  $1 - \varepsilon$  ( $0 < \varepsilon$  petit) pour un rang  $r$  donné de la distribution  $X$ . Aucune hypothèse n'est posée sur la loi de probabilité de  $X$ .

Dans cette démarche, on cherche un intervalle de confiance autour d'un rang (par exemple pour l'observation de rang  $n/2$  si  $n$  pair) et non directement un intervalle de confiance pour la médiane de la distribution. La technique présentée ici n'est pas nouvelle, mais elle ne figure pas sous cette forme de statistique de rang dans les manuels et dans cette présentation l'intervalle de confiance est construit à partir du rang considéré et non des valeurs de la distribution.

D'un point de vue général, pour construire un intervalle de confiance autour de l'observation de rang  $r$ , on cherche deux rangs  $l$  et  $u$  tels que :

$$P(x_{(l)} \leq x_{(r)} \leq x_{(u)}) \geq 1 - \varepsilon$$

ou encore

$$P(x_{(l)} \leq x_{(r)}) - P(x_{(u)} \leq x_{(r)}) \geq 1 - \varepsilon$$

On introduit des variables de Bernoulli indépendantes de paramètre  $q$  définies par :

$$Y(i) = \begin{cases} 1 & \text{si } x_{(i)} \leq x_{(r)} \\ 0 & \text{si } x_{(i)} > x_{(r)} \end{cases}$$

Soit  $Z = \sum_{i=1}^n Y(i)$ ,  $Z$  suit une loi binomiale  $B(n, q)$ , où  $q = \frac{r}{n}$ . On obtient :

$$P(x_{(l)} \leq x_{(r)}) = P(Z \geq l) = \sum_{i=l}^n P(Z = i)$$

D'où

$$P(x_{(l)} \leq x_{(r)} \leq x_{(u)}) = \sum_{i=l}^{i=u-1} P(Z = i)$$

Le problème revient donc à chercher les rangs  $l$  et  $u$  tels que

$$\sum_{i=l}^{i=u-1} C_n^i q^i (1-q)^{n-i} \geq 1 - \varepsilon$$

On peut remarquer que le terme  $C_n^i (q)^i (1-q)^{n-i}$  est maximal pour  $i = qn = r$ .

On cherche alors un intervalle de confiance symétrique pour l'estimation du quantile  $q$ , il s'agit de trouver numériquement le plus petit entier naturel  $x$  tel que :

$$\sum_{i=r-x}^{r+x-1} C_n^i q^i (1-q)^{n-i} \geq 1 - \varepsilon$$

#### 4. LES SÉRIES DE SOLDES D'OPINION

Les données utilisées concernent 5 séries trimestrielles ( $X1-X5$ ) de l'industrie en Ukraine, obtenues à partir des enquêtes conjoncturelles. Elles correspondent au carnet de commandes, au niveau des stocks, à la solvabilité des clients, au personnel salarié et à l'opinion général sur la conjoncture à court terme. Chaque question a bien sûr trois niveaux de réponse. La période d'observation s'étend de quatrième trimestre 2001 au quatrième trimestre 2005. Avant cette date, les données ne sont pas de bonne qualité.

## 4.1 Comparaison

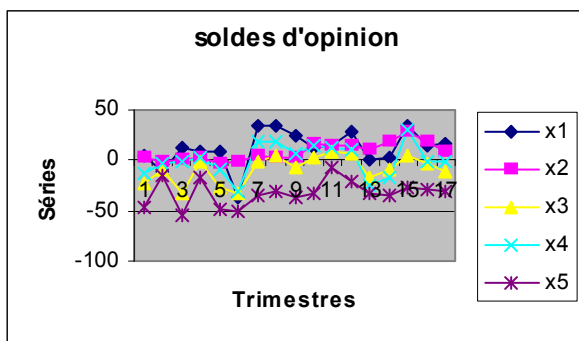
Une analyse factorielle est réalisée sur ces cinq séries (cf. graphique 1) qui sont corrélées positivement. Un seul facteur commun ( $F_c$ ) suffit à rendre compte de ces séries, dont l'expression est donnée par:

$$F_c = (0,631 \times x_1 + 0,443 \times x_2 + 0,913 \times x_3 + 0,773 \times x_4 + 0,520 \times x_5) / 3,28.$$

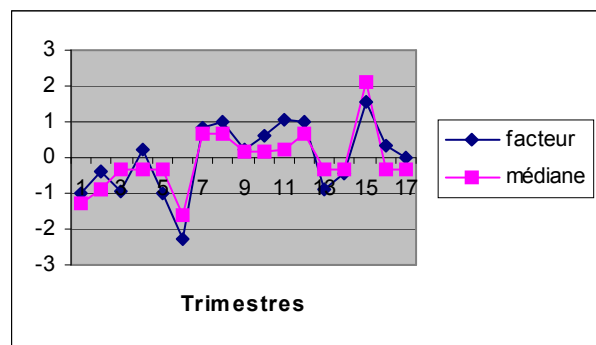
Ce facteur commun peut être modélisé par un processus ARMA(1,1). Il apparaît comme un indicateur du climat général dans l'industrie.

Pour construire ce nouvel indicateur, on effectue un classement des entreprises pour chaque date  $t$ . On peut obtenir ainsi deux séries qui correspondent à la classe médiane ou à la médiane obtenue par interpolation. Bien évidemment les deux séries sont très proches, avec davantage de paliers pour la première série.

Pour comparer la série du facteur commun des soldes d'opinion avec celle de la médiane des rangs, les deux séries sont au préalable centrées et réduites (graphique 2).



Graphique 1 : Séries des soldes d'opinion

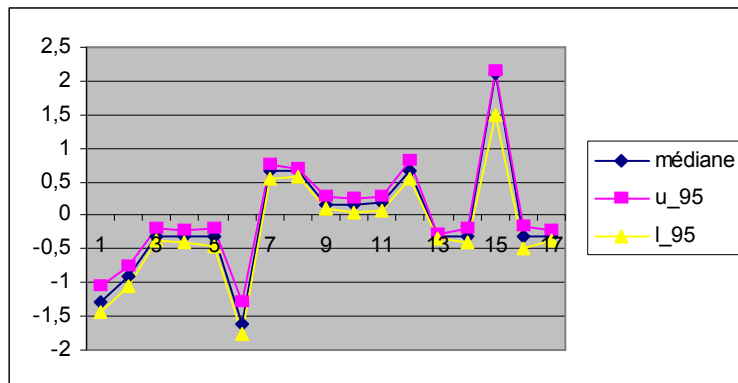


Graphique 2 : Facteur commun et médiane

Le coefficient de corrélation temporelle, calculé entre les différences premières de ces séries, est égal à 0,84, ce qui laisse à penser que les tendances locales de ces deux séries sont similaires en direction et en taux de variation.

## 4.2 Intervalle de confiance

Le graphique suivant représente la série de la médiane avec un intervalle de confiance au seuil de 95%. Si l'intervalle est symétrique par rapport au rang des observations, il n'est pas nécessairement symétrique par rapport à la classe médiane (aux valeurs de la série).



### 4.3 Application numérique

Le tableau suivant donne un intervalle de confiance presque symétrique pour un niveau de confiance de 95% ( $\varepsilon = 0.05$ ) selon la taille  $n$  de l'échantillon et le quantile  $q$  (les cases du tableau correspondent aux valeurs de  $x$ ).

<b>n \ q</b>	<b>0,05</b>	<b>0,25</b>	<b>0,5</b>	<b>0,75</b>	<b>0,95</b>
<b>50</b>	1	3	6	3	1
<b>100</b>	4	8	9	8	4
<b>500</b>	9	18	21	18	9
<b>1000</b>	13	26	30	26	13

Par exemple, pour un échantillon de taille 1000, on a 95% de chance que la médiane quantile se trouve entre l'observation de rang 479 et celle de rang 521.

## 5. CONCLUSIONS

L'analyse conjoncturelle est indispensable à la conduite de la politique économique. Les enquêtes de conjoncture fournissent des informations essentielles sur les évolutions récentes de l'économie. Elles ont un excellent rapport qualité/prix, elles sont simples, peu coûteuses, rapides et fiables.

Notre indicateur synthétique présente des qualités analogues à l'indicateur basé sur le facteur commun, avec toutefois une stabilité plus importante. Il est plus robuste aux variations instantanées et davantage corrélée à la tendance. Au niveau des prévisions, la difficulté réside dans la prévision des points de retournement. Cet indicateur basé sur la médiane ne semble pas être adéquat pour cette problématique. On peut suggérer d'utiliser d'autres quantiles comme indicateurs de retournement pour justement capter les points de retournement des cycles. L'historique de nos données est trop restreint pour faire cette étude.

## RÉFÉRENCES

- Doz, C., Lenglart, F. (1995). « Une grille de lecture pour l'enquête mensuelle de l'industrie ». *Note de conjoncture de décembre 1995, Insee*.
- Doz, C., Lenglart, F. (1999). « Analyse factorielle dynamique : test du nombre de facteurs, estimation et application à l'enquête de conjoncture dans l'industrie ». *Annales d'économie et de statistique*, **54**, 91-126
- Gouriéroux, C., Monfort, A. (1997). *Time Series and Dynamic Models*. Cambridge University Press, Cambridge, New York.
- Mitchell, J., Smith, R.J., Weale, M.R. (2004-1). « Aggregate versus Disaggregate Survey-Based Indicators of Economic Activity ». *27th CIRET conference, Warsaw, September*.
- Mitchell, J., Smith, R.J., Weale, M.R. (2004-2). « The impact of survey aggregation methods on the quality of business survey indicators ». *Report for the European Commission, ECFIN/2003/A3-04, November*.
- Mitchell, J., Smith, R.J., Weale, M.R. (2005). « Forecasting manufacturing output growth using firm-level survey data ». *Journal of business and Economic Statistics*, **20**, 147-162.