

# ***ANALYSE FACTORIELLE ET MODÈLES À COMPOSANTES INOBSERVABLES : APPLICATION À L'ÉTUDE DE L'ENQUÊTE DE CONJONCTURE DANS L'INDUSTRIE***

*Catherine Doz et Fabrice Lengart*

## **1 Introduction**

Les enquêtes de conjoncture offrent un type d'information particulièrement utile pour l'analyse conjoncturelle. En effet, parmi l'ensemble des indicateurs infra-annuels, elles cumulent nombre d'avantages : elles fournissent un message recueilli directement auprès des acteurs économiques ; elles sont disponibles relativement rapidement (environ un mois après l'envoi des questionnaires) ; elles sont publiées à intervalles rapprochés et réguliers ; enfin, elles ne sont pas révisées.

Cependant, le nombre et la diversité des questions posées rendent souvent délicate l'interprétation des résultats obtenus. C'est pourquoi il semble utile de chercher à construire un indicateur unique à partir de l'ensemble des réponses fournies par une enquête, qui pourrait constituer une sorte de résumé de l'information qu'elle contient. Une façon simple de le faire consiste à calculer la moyenne simple de plusieurs soldes d'opinion relatifs à l'état de l'activité économique. C'est ainsi que l'institut allemand IFO publie un « indicateur des affaires », qui est la moyenne des soldes d'opinion concernant la production passée et futures des entreprises interrogées. Un autre exemple est celui de l'indice NAPM américain, résultat de la moyenne de cinq soldes concernant les entrées de commandes, la production, les stocks, l'emploi et les délais de livraisons. De telles méthodes présentent l'avantage de la simplicité, mais le choix des questions retenues comme celui des pondérations attribuées à chaque question demeurent *ad hoc*.

Une approche plus satisfaisante consiste à supposer que chaque variable peut être décomposée en deux composantes orthogonales entre elles : l'une commune à l'ensemble des séries, et l'autre spécifique à la variable considérée. La composante commune est alors assimilée à l'indice composite recherché. Les modèles à facteurs dynamiques (*dynamic factor models*) constituent le cadre d'analyse approprié pour formaliser de telles hypothèses. Ce type de modèles a été utilisé dans des contextes variés par Geweke (1977), Sargent et Sims (1977), Geweke et Singleton (1981),

Engle et Watson (1981, 1983), Watson et Kraft (1984) ou, plus récemment, par Stock et Watson (1989, 1991, 1993), Quah et Sargent (1993), Forni et Lippi (1995).

Deux méthodes peuvent être utilisées pour estimer ce genre de modèles. La première se place dans le domaine des fréquences. Dans ce cas, la dynamique du modèle n'a pas besoin d'être spécifiée : les méthodes standard d'analyse factorielle (encore appelés analyse en composantes communes et spécifiques (AFCS)) peuvent être utilisées, de façon à décomposer la matrice de densité spectrale. La deuxième méthode relève plus directement du domaine temporel : dès lors que la dynamique des différentes composantes a été spécifiée, le modèle peut être mis sous une forme espace-état et estimé par le filtre de Kalman.

Dans cette étude, nous utilisons des représentations ARMA pour modéliser les différentes composantes et nous mettons en oeuvre la deuxième méthode d'estimation. La représentation espace-état présente l'avantage d'être très « souple » : en particulier, elle permet d'utiliser simultanément des données à périodicités différentes, mensuelle et trimestrielle.

Néanmoins, nous appliquons également aux données la technique standard de l'analyse factorielle. Certes, cette méthode n'est pas *a priori* appropriée dans un cadre d'analyse dynamique (elle a été créée au départ pour étudier des données individuelles). Mais nous montrons qu'elle fournit des estimateurs convergents des paramètres du modèle, même dans le cas où il y a présence d'autocorrélation temporelle des variables et où cette autocorrélation n'est pas prise en compte. En définitive, les programmes d'analyse factorielle standard peuvent donc être utilisés. De surcroît, ils offrent des éléments statistiques qui aident à choisir le nombre de facteurs communs à retenir. Ils offrent également des procédures de rotations d'axes qui facilitent l'interprétation des résultats obtenus lorsque plusieurs facteurs communs sont nécessaires pour décrire les données.

Les résultats obtenus par l'une ou l'autre méthode sont toujours très proches, ce qui renforce leur crédibilité. Nous commençons par estimer le modèle sur l'enquête mensuelle de conjoncture de l'INSEE dans l'industrie. Il apparaît qu'un seul facteur commun suffit à rendre compte de l'évolution commune des données, de sorte qu'il constitue un indicateur composite de l'enquête. Cet indicateur est estimé par les deux méthodes, et nous utilisons les deux estimations indifféremment pour commenter les fluctuations économiques dont elles rendent compte. Un tel indicateur peut être vu comme une sorte d'indicateur de « climat des affaires ». Dans une deuxième étape, nous tentons d'interpréter les mouvements spécifiques à chaque question de l'enquête, ce qui permet d'affiner le diagnostic conjoncturel. Enfin, nous commentons les résultats obtenus lorsque l'on mélange les enquêtes mensuelle et trimestrielle dans l'industrie.

La dernière partie du papier analyse plus à fond la structure d'information contenue dans l'indicateur synthétique, en étudiant les soldes d'opinion à un niveau plus désagrégé, celui des grandes branches industrielles (biens intermédiaires, biens d'équipement, biens de consommation). Des techniques de rotation axiale permettent de montrer que cet indicateur de climat n'est pas la résultante fortuite de conjonctures sectorielles indépendantes les unes des autres, mais qu'il est en fait présent dans chaque branche, même s'il s'y reflète suivant des modalités différentes.

## 2 Le cadre d'analyse théorique

### 2.1 Présentation de l'analyse factorielle standard

L'analyse factorielle a pour but d'offrir une description parcimonieuse d'un ensemble d'observations. Elle tente de représenter les variables étudiées dans un cadre linéaire, en fonction d'un ensemble de variables latentes, appelées *facteurs*, ou *facteurs communs*. C'est une technique qui est appropriée lorsqu'un petit nombre de facteurs peut rendre compte d'un part importante de l'information contenue dans l'ensemble des variables initiales.

Plus précisément, soit  $I$  le nombre de variables étudiées, soit  $T$  le nombre d'observations dont on dispose pour chaque variable, et soit  $y_{it}$  la valeur prise par la variable  $y_i$  à la date  $t$ <sup>1</sup>. Un modèle décrivant les variables  $y_1, \dots, y_I$  en fonction de  $J$  facteurs communs  $F_1, \dots, F_J$ ,  $J < I$ , s'écrit :

$$y_{it} = \lambda_{i1} F_{1t} + \dots + \lambda_{iJ} F_{Jt} + u_{it}, \text{ for } i = 1 \text{ to } I,$$

où les processus  $(u_{it})_{t \in Z}$  sont supposés indépendants entre eux et indépendant des facteurs, et où toutes les variables sont supposées centrées. Dans le cadre de l'analyse factorielle classique, *chacun des processus  $(F_{jt})$  et  $(u_{it})$  est supposé sans autocorrélation* (nous abandonnerons cette restriction dans la section 2.2).

Ce modèle peut s'écrire de façon plus concise, en utilisant des notations matricielles.

---

<sup>1</sup> Comme nous l'avons précisé en introduction, ce type de modèle est généralement utilisé pour étudier  $I$  variables observées sur un échantillon de  $N$  individus, mais nous nous situons ici d'emblée dans un contexte temporel.

Soient :  $y_t = (y_{1t}, \dots, y_{It})'$ ,  $F_t = (F_{1t}, \dots, F_{Jt})'$ ,  $u_t = (u_{1t}, \dots, u_{It})'$ , et

$$\Lambda = (\lambda_{ij})_{\substack{1 \leq i \leq I \\ 1 \leq j \leq J}}$$

On en déduit :

$$y_t = \Lambda F_t + u_t,$$

avec :  $EF_t = 0$ ,  $Eu_t = 0$ ,  $E(u_t u_t') = D = \text{diag}(\sigma_1^2, \dots, \sigma_I^2)$ ,

$$\forall (t, \tau) \quad E(F_t u_\tau') = 0,$$

$$\forall (t, \tau), t \neq \tau, \quad E(F_t F_\tau') = 0$$

$$\forall (t, \tau), t \neq \tau, \quad E(u_t u_\tau') = 0$$

Un tel modèle n'est pertinent que lorsque les variables sont fortement corrélées entre elles ; dans ce cas, il permet d'analyser cette structure de corrélation. Plus précisément, les facteurs rendent compte des corrélations entre variables, alors que chaque  $u_{it}$  représente une source de variation affectant la seule variable  $y_{it}$  : chaque  $u_{it}$  est appelé *facteur* ou *composante spécifique* de la variable  $y_i$ . Les  $\lambda_{ij}$  sont appelés les *pondérations (loadings)* des facteurs communs : chaque  $\lambda_{ij}$  représente la contribution du facteur  $F_j$  à l'évolution de la variable  $y_i$ .

Il apparaît clairement que, dans une telle formulation, les facteurs ne sont définis qu'à une transformation linéaire près, sous réserve de modifier les pondérations. Dans le modèle le plus classique, les facteurs sont supposés non corrélés entre eux (cette hypothèse peut être ensuite levée, voir plus loin section 4.1) et avoir des variances unitaires (ce qui ne restreint pas la généralité du modèle puisqu'ils n'interviennent qu'à un facteur multiplicatif près). Cependant, même dans ce cas de figure, s'il y a plus d'un facteur à estimer, les facteurs ne sont pas définis de façon univoque : ils peuvent être modifiés via n'importe quelle rotation. Dans la pratique, lors de l'estimation, une solution unique est obtenue en ajoutant des contraintes identifiantes sur les facteurs. Cette solution peut ensuite subir une rotation (ou une simple transformation linéaire) si une telle opération facilite l'interprétation des résultats obtenus.

Concentrons nous maintenant sur le modèle de base, c'est-à-dire sur le cas où les facteurs sont non corrélés entre eux et sont de variance unitaire. Ce modèle conduit

à une interprétation simple en terme de variance et covariances des variables. De fait, on a dans ce cas :

$$y_t = \Lambda F_t + u_t,$$

$$\text{avec } EF_t = 0, Eu_t = 0, E(F_T F_T') = Id, E(u_t u_t') = D,$$

$$\forall (t, \tau) E(F_t u_\tau') = 0,$$

$$\forall (t, \tau), t \neq \tau, E(F_t F_\tau') = 0$$

$$\forall (t, \tau), t \neq \tau, E(u_t u_\tau') = 0$$

On obtient alors les relations suivantes :

$$Vy_t = \Lambda' \Lambda + D,$$

$$\text{et : } Vy_{it} = \sum_{j=1}^J \lambda_{ij}^2 + \sigma_i^2 \text{ pour } i = 1 \text{ à } I, \text{ et pour tout } t.$$

Chaque  $\lambda_{ij}^2$  représente la part de la variance de  $y_{it}$  qui est expliquée par le facteur  $F_j$ , et  $h_i^2 = \sum_{j=1}^J \lambda_{ij}^2$  représente la contribution totale des facteurs à la variance de  $y_{it}$  ( $h_i^2$  est quelquefois appelée *communauté* (*communality*) de la variable  $y_i$ ). La variance de  $u_{it}$ ,  $\sigma_i^2$ , apparaît clairement comme la part de la variance de  $y_{it}$  qui ne peut pas être expliquée par les facteurs, puisque :  $Vy_{it} = h_i^2 + \sigma_i^2$ . Enfin,  $V_j = \sum_{i=1}^I \lambda_{ij}^2$  mesure la contribution totale du facteur  $F_j$  aux variances de l'ensemble des variables.

Il existe deux méthodes principales pour estimer le modèle : l'analyse factorielle principale (*principal factor analysis*) et la méthode du maximum de vraisemblance sous hypothèse de normalité (voir Lawley et Maxwell (1971) pour une présentation complète). La première ne nécessite aucune connaissance sur le nombre de facteurs à retenir, alors que ce nombre doit être donné *a priori* pour mettre en oeuvre la seconde. En revanche, la méthode de maximum de vraisemblance fournit des estimateurs convergents des paramètres, ce qui n'est pas le cas de l'analyse factorielle principale.

En pratique, les deux méthodes sont utilisées. Lors d'une première étape, l'analyse factorielle principale fournit un critère permettant de choisir le nombre de facteurs à retenir. En fait, sa mise en oeuvre revient à effectuer une analyse en composantes principales (ACP) sur une matrice particulière : les termes hors diagonale de cette matrice sont égaux aux corrélations entre variables, mais les éléments diagonaux sont fixés, en première approximation, aux valeurs des corrélations canoniques de chaque variable avec l'ensemble des autres. On est alors conduit à calculer les valeurs propres de cette matrice, et le nombre de facteurs à retenir est choisi en fonction de la taille de ces valeurs propres. Dans une deuxième étape, il est alors possible de mettre en oeuvre la méthode de maximum de vraisemblance. En outre, un test du rapport de vraisemblance permet de contrôler que le nombre de facteur retenu est correct.<sup>2</sup>

Signalons enfin que, pour ce qui est des deux méthodes, les paramètres  $\lambda_{ij}$  sont estimés en premier, et que les valeurs prises par les facteurs communs  $F_{jt}$  (les *scores*) sont, dans un deuxième temps, approximés comme des combinaisons linéaires des variables initiales - ceci est fait par le biais de techniques de régression visant à minimiser la variance de l'écart entre chaque facteur et son approximation linéaire.

Dans ce papier, nous nous intéresserons particulièrement au cas où un facteur suffit pour avoir une bonne approximation de l'ensemble des variables observées. En effet, dans ce cas, cela a un sens de considérer le facteur  $F_t$  comme une sorte d'indicateur coïncident, puisqu'il rend compte d'une part importante de la variance de chaque variable. Le modèle prend alors la forme plus simple suivante :

$$y_{it} = \lambda_i F_t + u_{it}, \text{ pour tout } i \text{ et tout } t,$$

c'est-à-dire :

$$y_t = \lambda F_t + u_t,$$

$$\text{où : } y_t = (y_{1t}, \dots, y_{nt})', \lambda = (\lambda_1, \dots, \lambda_n)', u_t = (u_{1t}, \dots, u_{nt})'$$

$$EF_t = 0, Eu_t = 0, VF_t = 1, E(u_t u_t') = D,$$

$$\forall (t, \tau) E(F_t u_\tau) = 0,$$

---

2 Ce test n'est cependant plus valide dans le cas où les processus sont temporellement autocorrélés.

$$\forall (t, \tau), t \neq \tau, E(F_t F_\tau) = 0$$

$$\forall (t, \tau), t \neq \tau, E(u_t u_\tau) = 0$$

## 2.2 Convergence, dans un cadre dynamique, des estimateurs obtenus dans l'analyse factorielle classique (statique)

On suppose que chacun des processus réels  $(F_{it})$  et  $(u_{it})$  peut présenter de l'autocorrélation, mais que le modèle est estimé par une procédure standard de maximum de vraisemblance, comme s'il n'y avait pas d'autocorrélation.

Soit  $\theta$  le vecteur des paramètres :  $\theta = (\lambda_{ij}, i = 1 \dots I, j = 1, \dots, J, \sigma_i^2, i = 1 \dots I)$ .

L'estimateur  $\hat{\theta}_T$  ainsi obtenu est alors un  $M$ -estimateur de  $\theta$ . Nous montrons que cet estimateur est convergent (les démonstrations des différents lemmes sont renvoyés en annexe). Il en résulte que les scores calculés à l'aide de cet estimateur ont les mêmes propriétés asymptotiques d'optimalité que dans le cas classique sans autocorrélation. En revanche, dans un tel contexte, on ne peut choisir le nombre de facteurs à retenir à l'aide de la procédure de test standard - le calcul de la matrice de variance-covariance asymptotique de  $\hat{\theta}_T$ , peut permettre d'élaborer un test adapté dans ce contexte, mais nous laissons ce point pour une recherche ultérieure.

Soit  $A$  la matrice de variance-covariance empirique des observations, et soit  $C = \Lambda\Lambda' + D$  la matrice de variance-covariance théorique. La vraisemblance du modèle, calculée sous hypothèse de normalité, lorsque les facteurs et les composantes spécifiques ne présentent pas d'autocorrélation, s'écrit (à un terme constant près) : 
$$\sum_i \ln l_i = -\frac{T-1}{2} \ln \det C - \frac{T-1}{2} \text{tr } C^{-1} A$$

(voir par exemple Lawley et Maxwell (1971)).

Soit  $\theta_0$  la vraie valeur du paramètre  $\theta$ . On suppose que  $\theta_0$  vérifie :

$$\theta_0 = (\lambda_{ij0}, i = 1 \dots I, j = 1, \dots, J, \sigma_{i0}^2, i = 1 \dots I), \text{ avec } \sigma_{i0}^2 \neq 0, \text{ pour } i = 1 \dots I.$$

Nous montrons ci-dessous que, si  $\theta_0$  vérifie cette hypothèse, alors le  $M$ -estimateur  $\hat{\theta}_T$ , obtenu par maximisation de la pseudo-vraisemblance précédente, vérifie un ensemble de conditions qui suffit à assurer sa convergence.

Dans toute la suite, on suppose que  $\theta$  varie dans une région de la forme  $R^I \times [\alpha, +\infty[^I$  contenant  $\theta_0$ . Ceci entraîne notamment que  $\theta$  vérifie lui-même :  $\sigma_i^2 \neq 0$ , pour  $i = 1 \dots I$ . Cette dernière condition est en particulier une condition suffisante pour que  $C$  soit inversible, donc pour que la pseudo-vraisemblance soit définie.

Pour tout  $t$  fixé, on note  $x_{it} = y_{it} - \bar{y}_i$  et  $x_t = (x_{1t}, \dots, x_{It})'$ . La fonction à maximiser, que nous noterons par la suite  $Q_T(y, \theta)$ , s'écrit alors :

$$Q_T(y, \theta) = -\frac{1}{2} \text{Ln det } C - \frac{1}{2} \text{tr} \left( C^{-1} \left( \frac{1}{T} \sum_t x_t x_t' \right) \right)$$

Enfin, pour toute matrice  $M$  on note :  $\|M\| = \max_{i,j} |m_{ij}|$ .

*Lemme 1 :*

i) Si  $\|C\| \rightarrow +\infty$ , alors  $Q_T(y, \theta) \rightarrow -\infty$ .

Il résulte de ce lemme que, lorsqu'on cherche à maximiser  $Q_T(y, \theta)$  sur une région de la forme  $R^I \times [\alpha, +\infty[^I$ , on peut se restreindre au cas où  $\theta$  varie dans un compact inclus dans cette région.

Dans toute la suite, on suppose donc que  $\theta$  appartient à un compact  $\Theta$  contenant  $\theta_0$  et inclus dans une région de la forme  $R^I \times [\alpha, +\infty[^I$ , où  $\alpha$  est un réel fixé.

*Lemme 2 :*  $Q_T(y, \theta)$  converge en probabilité, uniformément pour  $\theta \in \Theta$ , vers :

$$Q_0(\theta) = -\frac{1}{2} \text{Ln det } C - \frac{1}{2} \text{tr } C^{-1} C_0,$$

où  $C_0 = \Lambda_0 \Lambda_0' + D_0$  est la matrice de variance covariance associée à la vraie valeur  $\theta_0$  du paramètre.



*Lemme 3* : La fonction  $Q_0$  admet un maximum unique au point  $\theta_0$ .

*Lemme 4* : La fonction  $Q_0$  est continue sur  $\Theta$ .

*Proposition* : Le  $M$ -estimateur  $\hat{\theta}_T$  est convergent.

*Preuve* : Par application du lemme 1, on a vu qu'on pouvait supposer que  $\theta \in \Theta$ , où  $\Theta$  est un compact inclus dans  $R^I \times (R^{**})^I$ . Le résultat découle alors des lemmes 2 à 4 et des théorèmes généraux sur la convergence des  $M$ -estimateurs (voir par exemple Newey et Mac Fadden (1994), théorème 2.1).

### 2.3 Les modèles à composantes inobservables

L'analyse factorielle classique est une méthode rapide et très simple d'emploi. Cependant, elle présente l'inconvénient de ne pas préciser la dynamique des variables étudiées - même si, comme nous l'avons vu, cette méthode peut être utilisée dans un cadre dynamique - de sorte que, par exemple, elle ne peut pas être utilisée dans un but de prévision. De plus, les procédures standard d'analyse factorielle ne peuvent pas traiter le cas de variables dont les périodicités sont différentes.

Comme nous l'avons dit en introduction, dès lors que la dynamique des différentes composantes est spécifiée, un modèle dynamique à facteurs peut être écrit à l'aide d'une représentation espace-état. Comme les facteurs sont des variables inobservables, un tel modèle appartient à la classe des modèles à composantes inobservables (M.C.I.), et peut être estimé en utilisant la technique du filtre de Kalman.

Plaçons nous dans la situation où un seul facteur commun suffit à expliquer une grande part de la variance des variables - ceci doit être le cas lorsqu'on veut construire un indicateur coïncident. Des travaux antérieurs nous ayant conduit à spécifier un modèle ARMA(2,1) pour le facteur commun, et des modèles AR(1) pour les composantes spécifiques, on obtient alors le modèle suivant :

$$y_{it} = \lambda_i F_t + u_{it}, \text{ pour } i = 1 \text{ à } I, \text{ pour tout } t,$$

$$F_t = \phi_1 F_{t-1} + \phi_2 F_{t-2} + \varepsilon_t - \theta \varepsilon_{t-1}, \text{ pour tout } t,$$

$$u_{it} = \rho_i u_{i,t-1} + \varepsilon_{it}, \text{ pour } i = 1 \text{ à } I, \text{ pour tout } t,$$

où  $\varepsilon_t$  et  $\varepsilon_{it}$  sont les innovations de  $F_t$  et de  $u_{it}$  à chaque date  $t$ , et où les processus  $(\varepsilon_t)$  et  $(\varepsilon_{it})$  sont supposés indépendants.

Ce type de modèle admet une représentation espace-état. Rappelons que la forme générale des modèles espace-état est la suivante :

$$\begin{aligned} y_t &= Z_t \alpha_t + d_t + e_t & (M) \\ \alpha_t &= A_t \alpha_{t-1} + c_t + R_t \eta_t & (T) \end{aligned}$$

dans laquelle les processus  $(e_t)$  et  $(\eta_t)$  sont non autocorrélés, non corrélés entre eux, et vérifient :  $Ee_t = 0 \quad Ve_t = H_t \quad E\eta_t = 0 \quad V\eta_t = Q_t$ .

Dans un tel cadre, le vecteur  $\alpha_t$ , qui est appelé le *vecteur d'état*, est inobservable. L'équation (M) est appelée *équation de mesure* : elle donne la relation, à chaque date  $t$ , entre le vecteur  $y_t$  des variables observées et le vecteur d'état. L'équation (T), ou *équation de transition*, décrit la dynamique du vecteur d'état.

Les matrices  $Z_t, A_t, d_t, c_t, R_t, H_t$  et  $Q_t$  sont généralement non stochastiques, mais elles peuvent dépendre du temps de façon déterministe. Elles peuvent aussi dépendre des paramètres du modèle.

Si l'on définit le vecteur d'état par :  $\alpha_t = (F_t, F_{t-1}, \varepsilon_t, u_{1t}, \dots, u_{It})'$ , on peut écrire le modèle sous forme espace-état en utilisant les relations suivantes :

$$y_t = \begin{pmatrix} \lambda_1 & 0 & 0 & 1 & & \\ \vdots & \vdots & \vdots & & \ddots & \\ \lambda_I & 0 & 0 & & & I \end{pmatrix} \begin{pmatrix} F_t \\ F_{t-1} \\ \varepsilon_t \\ u_{1t} \\ \vdots \\ u_{It} \end{pmatrix},$$

$$\text{et : } \begin{pmatrix} F_t \\ F_{t-1} \\ \varepsilon_t \\ u_{1t} \\ \vdots \\ u_{It} \end{pmatrix} = \left( \begin{array}{ccc|ccc} \varphi_1 & \varphi_2 & -\theta & & & \\ 1 & 0 & 0 & & & 0 \\ 0 & 0 & 0 & & & \\ & & & \rho_1 & & \\ & & & & \ddots & \\ & & & & & \rho_I \end{array} \right) \begin{pmatrix} F_{t-1} \\ F_{t-2} \\ \varepsilon_{t-1} \\ u_{1,t-1} \\ \vdots \\ u_{I,t-1} \end{pmatrix} + \left( \begin{array}{c|ccc} 1 & & & \\ 0 & & 0 & \\ \hline 1 & & & \\ 0 & & & 1 \end{array} \right) \begin{pmatrix} \varepsilon_t \\ \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{It} \end{pmatrix}$$

Ceci peut s'écrire de façon plus concise, en utilisant les notations introduites précédemment, et en notant de plus :  $\tilde{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{It})'$ ,  $R = \text{diag}(\rho_1, \dots, \rho_I)$ , et

$\alpha_t = (F_t, F_{t-1}, \varepsilon_t, u_t)'$ . On obtient :

$$y_t = (\lambda \ 0 \ 0 \ Id) \begin{pmatrix} F_t \\ F_{t-1} \\ \varepsilon_t \\ u_t \end{pmatrix} \quad (M)$$

$$\begin{pmatrix} F_t \\ F_{t-1} \\ \varepsilon_t \\ u_t \end{pmatrix} = \left( \begin{array}{ccc|c} \Phi_1 & \Phi_2 & -\theta & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & R \end{array} \right) \begin{pmatrix} F_{t-1} \\ F_{t-2} \\ \varepsilon_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & Id \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \tilde{\varepsilon}_t \end{pmatrix} \quad (T),$$

où  $(\varepsilon_t, \tilde{\varepsilon}_t)'$  a une matrice de variance-covariance diagonale.

Un tel modèle peut être estimé de diverses façons. Dans ce papier, nous utilisons une procédure du maximum de vraisemblance à l'aide de la technique du filtre de Kalman. Rappelons brièvement les aspects principaux de cette méthode - une présentation complète figure, par exemple, dans Harvey (1989) ou Gouriéroux & Monfort (1990). Pour tout ensemble fixé de valeurs des paramètres, le filtre de Kalman en tant que tel consiste à déterminer la meilleure prévision du vecteur d'état à chaque date, compte tenu de l'information disponible à cette date. Pour tout  $t$ , les prévisions optimales :  $a_{t|t-1} = E(\alpha_t | I_{t-1})$  et  $a_t = E(\alpha_t | I_t)$ , sont ainsi calculées par une procédure itérative.

Sous l'hypothèse de normalité des perturbations, il est alors possible de calculer la vraisemblance du modèle :  $\ell(y_1, \dots, y_T, \Psi) = \prod_{t=1}^T f(y_t / I_{t-1})$ , où  $f(y_t / I_{t-1})$  désigne la densité conditionnelle de  $y_t$  sachant  $(y_1, \dots, y_{t-1})$ , et où  $\Psi$  désigne le vecteur des paramètres. En effet,  $f(y_t / I_{t-1})$  est alors une loi  $N(Z_t a_{t|t-1}, W_t)$  et la matrice de variance-covariance  $W_t$  est une matrice qui est calculée au cours de la procédure du filtre de Kalman. La vraisemblance peut ainsi être obtenue en tout point de l'espace des paramètres, de sorte qu'on peut la maximiser à l'aide d'une procédure quelconque d'optimisation, de façon à obtenir les estimateurs du maximum de vraisemblance des paramètres. Sous des hypothèses standard, ces estimateurs sont asymptotiquement normaux.

Enfin, on peut réappliquer le filtre de Kalman en utilisant les valeurs estimées des paramètres, et on obtient ainsi l'approximation correspondante du vecteur d'état à chaque date  $t$  ; mais on peut aussi améliorer cette approximation et déterminer la prévision optimale du vecteur d'état à chaque date  $t$ , sachant l'ensemble de l'information disponible. Ceci consiste à calculer  $E(\alpha_t | I_T)$ , et ce calcul peut être mené à l'aide d'une nouvelle procédure itérative, appelée procédure de lissage de Kalman.

## 2.4 Application à des données de périodicités différentes

Les modèles à composantes inobservables offrent aussi un cadre bien adapté pour estimer un modèle à facteurs lorsque les données n'ont pas toutes la même périodicité. Considérons le cas où l'on s'intéresse à la fois à un ensemble de données mensuelles :  $y_t = (y_{1t}, \dots, y_{I_1t})'$ , et à un ensemble de données trimestrielles :

$z_t = (z_{1t}, \dots, z_{I_2t})'$ . On suppose que  $t$  désigne l'indice du mois et que  $z_t$  n'est observé que lorsque  $t \equiv 0 [3]$ .

Nous faisons ici l'hypothèse que le même facteur commun peut rendre compte des deux ensembles de données. Avec les notations précédentes, le modèle à estimer est donc le suivant :

$$\underset{(I_1, t)}{y_t} = \lambda F_t + u_t$$

$$\underset{(I_2, t)}{z_t} = \mu F_t + v_t$$

où  $\mu = (\mu_1, \dots, \mu_{I_2})'$ , et  $v_t = (v_{1t}, \dots, v_{I_2t})'$ .

Les procédures standard d'analyse factorielle ne permettent pas d'estimer un tel modèle. Cependant, si l'on spécifie la dynamique des variables, il est possible d'écrire un modèle à composantes inobservables, comme nous le montrons ci-dessous.

On spécifie ici la dynamique des variables sous la même forme que précédemment :  $F_t$  est un processus ARMA(2,1), les  $u_{it}$  et les  $v_{it}$  sont des processus AR(1). On conserve les mêmes notations que précédemment pour les dynamiques de  $F_t$  et des  $u_{it}$ , et on suppose que :

$$v_{it} = \tau_i v_{i,t-1} + \zeta_{it}, \text{ pour tout } i, \text{ pour tout } t,$$

où  $\zeta_{it}$  est l'innovation de  $v_{it}$  et est non corrélé avec les autres perturbations.

On pose de plus :  $\zeta_t = (\zeta_{1t}, \dots, \zeta_{I_2t})'$  et  $T = \text{diag}(\tau_1, \dots, \tau_{I_2})$

Définissons maintenant une variable  $Y_t$  dont la taille varie selon la date :

$$Y_t = \begin{pmatrix} y_t \\ z_t \end{pmatrix} \text{ si } t \equiv 0 [3],$$

$$Y_t = y_t \text{ sinon.}$$

On peut alors obtenir une représentation espace-état du modèle sous la forme suivante :

$$- \alpha_t = (F_t, F_{t-1}, \varepsilon_t, u_t', v_t')',$$

$$- Y_t = Z_t \alpha_t \text{ où } : Z_t = \begin{pmatrix} \lambda & 0 & 0 & \text{Id} & 0 \\ \mu & 0 & 0 & 0 & \text{Id} \end{pmatrix} \text{ si } t \equiv 0 [3]$$

$$Z_t = (\lambda \ 0 \ 0 \ \text{Id} \ 0) \text{ sinon,}$$

$$- \alpha_t = A_t \alpha_{t-1} + R_t \eta_t, \text{ où } :$$

$$\eta_t = \begin{pmatrix} \varepsilon_t \\ \tilde{\varepsilon}_t \\ \zeta_t \end{pmatrix}, A_t = \left( \begin{array}{ccc|cc} \varphi_1 & \varphi_2 & -\theta & & \\ 1 & 0 & 0 & 0 & \\ 0 & 0 & 0 & & \\ \hline & & & R & 0 \\ & & & 0 & T \end{array} \right), R_t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \text{Id} & 0 \\ 0 & 0 & \text{Id} \end{pmatrix},$$

et où  $\eta_t$  a une matrice de variance-covariance diagonale et est non-autocorrélé.

## **3 Application aux données de l'enquête de conjoncture dans l'industrie**

### ***3.1 Les données***

On considère ici les réponses à six questions de l'enquête de conjoncture de l'INSEE dans l'industrie, sur la période janvier 1963-janvier 1996. Ces questions sont : opinion sur la tendance passée de la production personnelle (TPPA), opinion sur la tendance future de la production personnelle (TPPRE), opinion sur la demande et les carnets de commande globaux (OSCD), opinion sur la demande et les carnets de commande en provenance de l'étranger (OSCDE), opinion sur le niveau des stocks (OSSK) et perspectives générales d'activité (PGP).

Dans cette enquête, pour chaque question, les industriels peuvent répondre suivant trois modalités : le niveau est plus élevé, moins élevé, ou comparable à celui de la période précédente. Les réponses sont agrégées en pondérant par la taille de l'entreprise répondante. Le résultat publié est le solde d'opinion, c'est-à-dire la différence entre le pourcentage de réponses « plus élevé » et celui de réponses « moins élevé ».

Un simple coup d'oeil au *graphique 1* des six soldes (en annexe) suffit pour se convaincre que les profils de ces séries sont extrêmement similaires. D'où l'idée de chercher à résumer en un seul indice synthétique l'information commune qu'elles contiennent.

Chaque solde d'opinion présente une dynamique « cyclique » très prononcée, en ce sens qu'il peut être considéré comme stationnaire, mais qu'il présente une structure d'autocorrélation assurant une persistance à ses variations. En effet, les autocorrélogrammes empiriques montrent que les autocorrélations sont significativement positives jusqu'à l'ordre 8. Néanmoins, ces séries n'en sont pas pour autant intégrées : le test Dickey-Fuller augmenté refuse l'hypothèse nulle de la présence d'une racine unité dans la modélisation ARMA. Tout ceci justifie l'utilisation du modèle à composantes inobservables présenté plus haut, dans lequel toutes les variables sont stationnaires.

### ***3.2 - Résultats généraux obtenus par les deux méthodes***

L'analyse factorielle appliquée aux six soldes d'opinion conduit à retenir un seul facteur commun. En effet, ce facteur suffit à expliquer 82% de la variance totale des séries. Plus précisément, les pondérations des facteurs et les communautés sont :

Question	Pondérations	Communautés
TPPA	0.96	91.7 %
TPPRE	0.89	78.5 %
OSCD	0.97	93.2 %
OSCDE	0.89	79.2 %
OSSK	-0.88	77.1 %
PGP	0.86	73.6 %

Lorsqu'on estime le modèle dynamique à facteur présenté précédemment, en utilisant la technique du filtre de Kalman, on obtient les résultats suivants :

- les estimations obtenues pour les paramètres du facteur commun conduisent à l'équation suivante :

$$F_t = \underset{(0.04)}{1.88} F_{t-1} - \underset{(0.04)}{0.89} F_{t-2} + u_t - \underset{(0.07)}{0.60} u_{t-1}$$

(il faut ici noter que la valeur de  $\sigma_u$  a été fixée :  $\sigma_u = 0.5$  ; en effet le modèle est défini à un paramètre d'échelle près, et il faut donc fixer la valeur de l'un des paramètres pour le rendre identifiable).

- les estimations des autres paramètres figurent dans le tableau suivant :

Question	$\lambda_i$		$\rho_i$		$\sigma_i$	
	estimation	écart-type	estimation	écart-type	estimation	écart-type
TPPA	3.10	0.18	0.69	0.03	3.35	0.09
TPPRE	2.34	0.21	0.86	0.03	3.07	0.10
OSCD	4.00	0.20	0.88	0.04	2.57	0.14
OSCDE	3.50	0.26	0.92	0.02	4.00	0.16
OSSK	-2.37	0.16	0.82	0.04	3.18	0.11
PGP	5.08	0.38	0.89	0.03	6.72	0.19

On peut remarquer que l'estimation obtenue pour le polynôme autorégressif du facteur commun a des racines proches de l'unité. Cependant, il semble difficile de

tester directement la présence d'une racine unité dans un tel contexte. En effet, sous l'hypothèse nulle, la loi est non standard et les tests de Dickey-Fuller ne peuvent être directement appliqués dans le cadre non-linéaire des modèles à composantes inobservables. Cependant, les tests de racines unité menés sur les séries initiales ont conduit à rejeter l'hypothèse de non-stationnarité, et le facteur commun est ici nécessairement de même nature. Ce type de résultat n'est d'ailleurs pas spécifique à notre étude ; au contraire, il semble que l'on obtienne souvent des résultats analogues lorsqu'on étudie des séries présentant un fort caractère cyclique, ceci étant dû au caractère fortement persistant de telles séries<sup>3</sup>.

Comme nous travaillons dans un cadre stationnaire, les statistiques usuelles peuvent être utilisées. Cependant, le problème de la significativité des paramètres d'écart-type est un peu plus complexe puisque la valeur zéro est sur la frontière de l'espace admissible des paramètres. Dans ce cas, le carré de la statistique de Student suit un mélange de lois :  $1/2\chi_0^2 + 1/2\chi_1^2$ <sup>4</sup>. L'étude des résultats conduit, ici encore, à conclure que ces paramètres sont tous très significatifs.

Notons enfin que les facteurs communs obtenus avec les deux techniques sont extrêmement proches (cf. *graphique 2* en annexe). Ceci nous conduit à les utiliser indifféremment pour commenter les fluctuations économiques qu'ils reflètent.

### 3.3 Interprétation du facteur commun

Lorsque l'on rapproche l'évolution de l'indicateur obtenu du glissement annuel de la production manufacturière (tirée des comptes trimestriels), on constate que les retournements des deux séries se produisent le plus souvent à des dates très voisines<sup>5</sup> (cf. *graphique 3* en annexe). Il ne faudrait pas, cependant, se méprendre sur le sens de cette comparaison : on se borne à constater que le regard porté par les industriels sur l'activité de leur secteur (au travers de leurs réponses à l'enquête mensuelle) offre un reflet fidèle des grands mouvements et des inflexions de la

---

3 Par exemple, Watson (1986) estime la composante cyclique du PIB américain sous une forme AR(2) dans le cadre d'un M.C.I. et il obtient :  $C_t = 1.501 C_{t-1} - 0.577 C_{t-2} + u_t$ . Ici aussi,  $\phi_1 + \phi_2 = 1$ , ce qui signifie que les racines du polynôme retard sont en module proches de l'unité. Watson relève ce fait, mais ne teste pas l'hypothèse de non stationnarité.

4 Voir par exemple Harvey (1989) : un test de niveau 5% est obtenu en comparant le Student à 1.6.

5 La comparaison, a priori plus naturelle, entre la variable synthétique et le glissement annuel de la production de l'ensemble de l'industrie fournit le même type d'enseignement, mais l'aspect heurté de l'activité de la branche énergie rend le rapprochement un peu moins lisible. Ceci explique qu'on ait choisi comme illustration le glissement annuel de la seule production manufacturière.



conjoncture industrielle. En outre, les variations au mois le mois de cette variable apparaissent moins heurtées que celles des soldes d'opinion relatifs aux différentes questions de l'enquête, ce qui permet en principe de détecter plus facilement et plus rapidement les inflexions qu'elle peut connaître.

A l'observer sur une période relativement longue, cet indicateur synthétique permet ainsi de suivre les principales fluctuations qu'a connues le secteur industriel français. On y lit, dans les années 60, 70 et au début des années 80, des variations souvent en partie dues aux alternances des politiques gouvernementales de relance et de freinage de l'activité, ainsi que l'impact des deux chocs pétroliers (en 1974-75 puis 1979-80). L'expansion de la fin des années 80, initiée par le contre-choc pétrolier, apparaît très forte, puis le climat général rend compte du ralentissement entamé dans le courant de 1990, marqué par le faux redémarrage du printemps 1992 et par la récession profonde de 1993. Enfin, la remontée en flèche de 1994 traduit la vigueur et la brièveté de la dernière reprise industrielle.

### ***3.4 Interprétation des composantes spécifiques***

L'indice composite ici calculé rend compte de la plus grande part de l'information contenue dans les différents soldes. Cependant, si l'on veut analyser toute l'information fournie par l'enquête, il faut également interpréter les composantes spécifiques de chaque question. En fait, l'intérêt des enquêtes de conjoncture est double : d'une part, les résultats fournis peuvent être utilisés pour tenter de construire des prévisions quantitatives concernant l'évolution de certains agrégats économiques ; d'autre part, ils contribuent à affiner le diagnostic conjoncturel. L'utilisation de nos travaux à des fins d'estimation quantitatives fera l'objet de travaux ultérieurs. Nous nous contentons ici de montrer que, durant les périodes où la composante spécifique d'un solde d'opinion présente un mouvement prononcé et durable, il est souvent possible d'interpréter cette information d'un point de vue qualitatif.

Il est très important ici de ne pas se laisser abuser par les expressions employées : le terme d'information spécifique désigne l'information complémentaire apportée par une question donnée ; en particulier, il ne signifie pas qu'il faille se fonder uniquement sur elle pour analyser les résultats fournis par cette question. Bien au contraire, on a vu que l'indicateur de climat, cette information commune à toutes les questions, est responsable d'une part très importante de l'évolution du profil de chacune d'entre elles. Si bien que, pour prendre un exemple, en période de climat favorable, si une composante spécifique apparaît négative de façon durable, on jugera, non pas que l'opinion des industriels sur le sujet abordé est défavorable, mais qu'elle est plutôt moins favorable qu'elle ne pourrait -ou ne devrait- l'être dans le contexte conjoncturel étudié. Cette précaution prise, on peut maintenant tenter, à la

lumière des réponses aux questions de l'enquête mensuelle, de préciser la description des fluctuations conjoncturelles que l'industrie française a traversées.

La réponse relative au **niveau de la production passée** (cf *graphique 4.1* en annexe) est très proche du climat général, si bien qu'elle constitue à elle seule une bonne approximation de l'information commune sous-jacente aux six questions de l'enquête, à condition de la lisser quelque peu. Cet avantage a pour contrepartie évidente le fait que l'information spécifique apportée par l'opinion sur l'activité passée, animée de mouvements de faible amplitude, généralement irréguliers et contradictoires, ne semble pas interprétable.

En ce qui concerne les **perspectives personnelles de production** (cf *graphique 4.2* en annexe), si l'on veut voir dans l'information spécifique fournie par cette question une variable anticipée de la conjoncture, le bilan que l'on peut dresser est pour le moins contrasté. En fait, il semble, en première approximation, que, jusqu'à la fin des années 70, cette information spécifique tendait à positionner le solde d'opinion au dessus du climat général en période de conjoncture ascendante et en dessous dans le cas contraire, si bien que le croisement des deux courbes signalait qu'un véritable retournement était en cours ; en ce sens, on pouvait donc estimer que les grandes inflexions de l'activité pouvaient être décelées au travers des anticipations des entrepreneurs. Ceci peut signifier que la moindre internationalisation d'alors de l'économie française et la plus grande efficacité des politiques gouvernementales de régulation rendaient plus aisée la compréhension de la conjoncture et de ses mouvements à venir. La situation semble s'être obscurcie depuis le début des années 80. Néanmoins, on peut noter que la composante spécifique semble avoir retrouvé un certain pouvoir prédicteur lors du dernier cycle (93-95).

C'est généralement en période de haut ou de bas de cycle que l'information spécifique apportée par la réponse sur le **niveau de la demande et des carnets globaux** (cf *graphique 4.3* en annexe) se distingue du climat général des affaires. Une interprétation possible est la suivante : lorsque l'information spécifique est positive, cela signifie que l'état de la demande adressée aux industriels interrogés est jugé meilleur que ce que laisserait prévoir le contexte conjoncturel qu'ils décrivent : ils font donc sans doute preuve d'un pessimisme personnel relatif lorsqu'ils rendent compte du climat général. Ce cas de figure semble avoir prévalu lors des années de forte expansion de 1989 et 1990. Si ce point de vue est juste, il faut mettre en regard cette période et l'année 1994, au cours de laquelle la composante est, au contraire, faiblement négative ; ceci signifierait que les industriels ont perçu, lors de la dernière reprise, une réalité plutôt plus favorable que ce qu'indiquaient les carnets, se montrant cette fois plutôt optimistes.

L'information spécifique liée à la question sur **la demande et les carnets étrangers** (cf *graphique 4.4* en annexe) offre un éclairage extrêmement intéressant sur les décalages conjoncturels éventuels entre la France et l'étranger. On y voit par

exemple que la France n'a pas bénéficié pleinement de la reprise mondiale de 1984. On peut néanmoins noter que cette composante a tendance sur l'espace de trente ans à se réduire, de sorte que les évolutions du solde d'opinion semblent se rapprocher globalement de celles suivies par le climat général. Ceci peut sans doute s'interpréter comme un signe tangible de l'ouverture progressive de l'économie, et plus spécifiquement du secteur industriel, vers l'extérieur, si bien que les fluctuations conjoncturelles majeures suivies par le pays apparaissent de plus en plus reliées à celles que traversent ses principaux partenaires. Depuis 1991, l'activité industrielle de la France apparaît, à cet égard, plutôt en phase avec celle de l'Europe occidentale considérée dans son ensemble.

Un intérêt de la décomposition en informations commune et spécifique du solde d'opinion relatif au **niveau des stocks** (cf. *graphique 4.5* en annexe) pourrait être de proposer une lecture plus précise de la réponse à cette question qui est, en effet, toujours difficile à analyser. On demande aux industriels s'ils jugent le niveau de leurs stocks supérieur, inférieur ou conforme à la normale. La difficulté porte donc sur l'interprétation à donner à ce terme : la "normale" désigne-t-elle un niveau moyen de stocks, constant au cours du temps, ou bien un niveau de stocks variable, dépendant de la conjoncture du moment ? Dans le schéma présenté ici, les mouvements dus aux fluctuations du climat général traduiraient le comportement de stockage lié aux aléas conjoncturels proprement dits : en phase ascendante, pour des raisons tenant à la fois à une activité passée et à des anticipations de demande favorables, le niveau des stocks a tendance à être jugé de plus en plus léger et les industriels sont donc plutôt enclins à reconstituer leurs stocks, alors que les variations inverses se produisent en phase descendante. Dans ces conditions, l'information spécifiquement apportée par le solde d'opinion relatif à la question sur les stocks fournirait un renseignement sur l'écart de comportement par rapport à ce scénario de référence.

Les fluctuations de l'information spécifiquement apportée par la réponse sur les **perspectives générales de production** (cf. *graphique 6* en annexe) sont amples et persistantes. La remarque faite à propos de la question sur les perspectives personnelles de production demeure valable : il semble que le pouvoir prédictif de cette information se soit amoindri depuis le début des années 80, de sorte qu'en faisant passer le solde d'opinion au dessus ou en dessous du climat général qui se dégage de l'enquête, elle constitue sans doute plus, à présent, une sorte de miroir du discours ambiant sur l'état de la conjoncture, et de ses excès éventuels.

### ***3.5 Le mélange des enquêtes mensuelles et trimestrielles***

Comme on l'a vu dans la partie 2.3, la représentation espace-état permet facilement de mélanger les soldes d'opinion relatifs au même secteur (l'industrie), mais ayant des périodicités différentes. L'enquête trimestrielle fournit (entre autres) les

réponses à neuf questions : les opinions relatives à la demande globale ou étrangère passée ou future, celles concernant les effectifs passés et futurs, le taux d'utilisation des capacités de production, le jugement sur ces capacités de production, enfin l'opinion sur les goulots de production. Les données sont disponibles depuis le premier trimestre de 1976. Nous estimons un modèle à composantes inobservables par filtre de Kalman, en utilisant ces neuf soldes trimestriels plus les six soldes mensuels précédents, sur la période 1976:T1-1996:T1.

Les résultats obtenus confirment le lien entre les deux enquêtes, car le facteur commun est extrêmement proche de l'indicateur précédent (calculé à partir de la seule enquête mensuelle) ; il explique au moins 50% de la variance de chacun des soldes d'opinion. Les paramètres gouvernant la dynamique de l'indicateur sont quasiment identiques à ceux obtenus lors de la précédente estimation (et ce, bien que la période d'estimation ne soit pas la même) :

$$F_t = \underset{(0,05)}{1,88} F_{t-1} - \underset{(0,05)}{0,90} F_{t-2} + u_t - \underset{(0,11)}{0,61} u_{t-1}$$

De plus, ce « nouvel » indicateur de climat est un peu plus lisse que le précédent, puisque le ratio de la variance de l'innovation sur la variance de l'indicateur est inférieur de 14% ; ceci n'est d'ailleurs pas réellement surprenant, l'estimation tenant compte d'une quantité plus importante d'information. Enfin, les composantes spécifiques obtenues peuvent là encore recevoir une interprétation, exercice auquel nous ne nous livrons pas ici car il serait quelque peu fastidieux. Contentons-nous de remarquer que les questions relatives aux effectifs ou aux capacités de production fournissent des informations spécifiques particulièrement utiles s'agissant des évolutions possible de l'emploi ou de l'investissement.

## 4 Étude à un niveau plus désagrégé

### 4.1 Analyse factorielle : cas de plusieurs facteurs communs

Comme nous l'avons mentionné précédemment, lorsque plusieurs facteurs sont nécessaires pour rendre compte des données étudiées, les facteurs initialement calculés par la procédure ne sont pas toujours aisément interprétables. Il est souvent nécessaire de les soumettre à une rotation, afin d'obtenir de nouveaux facteurs ayant davantage de signification économique. La rotation en question peut même être une simple transformation linéaire (rotation oblique), c'est-à-dire que l'hypothèse de non-corrélation des facteurs peut être abandonnée.

Bien sûr, le mot "interprétable" contient une part de jugement subjectif. Mais il est cependant possible de donner un contenu statistique à ce mot. En effet, plus les facteurs représentent un groupe donné de variables, plus ils sont faciles à interpréter. Il est clair, par ailleurs, qu'un facteur  $F_j$  représente un groupe de variables  $y_{i_1}, \dots, y_{i_p}$ , si deux conditions sont vérifiées :

- pour  $r = 1, \dots, p$ , les pondérations  $\lambda_{i_r, j}$  sont nettement plus importants que les pondérations  $\lambda_{i_r, k}$  correspondant à l'un quelconque des autres facteurs  $F_k$  ;
- ils sont aussi beaucoup plus importants que les pondérations  $\lambda_{ij}$  associés à l'une quelconque des autres variables  $y_i$ , pour  $i \notin \{i_1, \dots, i_p\}$ .

En pratique, toutes les procédures existantes consistent à effectuer une rotation des facteurs initiaux de façon à ce que les facteurs obtenus à l'issue de cette rotation soient associés au plus grand nombre possible de pondérations nulles (ou, au moins négligeables). Si l'on note  $F_j^*$  les facteurs obtenus après rotation, le modèle devient :

$$y_{it} = \lambda_{i1}^* F_{1t}^* + \dots + \lambda_{ij}^* F_{jt}^* + u_{it},$$

dans lequel  $\Lambda^* = \left( \lambda_{ij}^* \right)_{\substack{1 \leq i \leq I \\ 1 \leq j \leq J}}$  est une matrice contenant de nombreux termes nuls ou voisins de zéro, et les facteurs  $F_{jt}^*$  sont généralement corrélés (mais restent, bien sûr, non corrélés avec les  $u_{it}$ ).

Dans les logiciels d'analyse factorielle, plusieurs procédures sont généralement disponibles pour effectuer de telles rotations - sans rentrer dans le détail de ces procédures, précisons simplement qu'elles reposent le plus souvent sur la maximisation d'un critère quadratique. Nous avons, pour notre part, utilisé la méthode PROMAX (voir Cureton et Mulaik, 1975), disponible dans le logiciel SAS.

## ***4.2 Application à l'enquête mensuelle au niveau des grandes branches industrielles***

L'analyse factorielle a montré qu'une part importante de l'information contenue par les réponses de l'enquête mensuelle dans l'industrie pouvait être résumée par le comportement d'une seule variable cachée, que nous avons appelé « climat général des affaires ». Néanmoins, une telle dénomination n'est réellement justifiée que si ce

climat se trouve présent à un niveau plus désagrégé, parmi les branches du secteur industriel : si cet indicateur ne constituait, en définitive, que la résultante de conjonctures sectorielles très indépendantes les unes des autres, une analyse menée au niveau le plus global se révélerait plus formelle que véritablement pertinente : parler d'un climat général des affaires aurait, dans ces conditions, moins d'intérêt. Mieux vaudrait se concentrer directement sur les variations observées au niveau des branches.

C'est pourquoi il semble nécessaire d'approfondir quelque peu la nature de l'indicateur de climat en modélisant ensemble les variations de cinq soldes d'opinion <sup>6</sup> pris au niveau de trois grands secteurs, à savoir celui des biens intermédiaires, celui des biens d'équipement et celui des biens de consommation. La période d'estimation est mars 1976-janvier 1996.

L'analyse factorielle offre un critère pour déterminer le nombre de facteurs communs, en calculant les valeurs propres successives de la matrice  $AA'$ . Les résultats obtenus ainsi que la proportion des communautés totales expliquée par chaque valeur propre sont :

	valeur propre	proportion
1	91.81	79.8%
2	12.02	10.6%
3	7.27	6.4%
4	2.15	1.9%
5	1.34	0.67%
...	...	...

Deux enseignements peuvent être tirés de ces résultats. D'une part, on est amené à retenir trois facteurs communs. D'autre part, le premier de ces trois facteurs est largement prépondérant.

Lorsque l'on compare les variations de ce premier facteur avec l'indicateur de climat des affaires précédemment obtenu (cf *graphique 5* en annexe), les deux variables se révèlent extrêmement proches l'une de l'autre. Ceci signifie précisément que l'indicateur de climat, qui a été estimé à l'aide des soldes d'opinion

---

<sup>6</sup> Ces soldes d'opinion concernent les niveaux de production passé et futur, les niveaux des carnets globaux et des carnets étrangers, enfin le niveau des stocks. Les réponses relatives aux perspectives générales d'activité ne sont pas disponibles par branche.

au niveau le plus agrégé, est « réellement » et fortement présent au sein des trois branches étudiées ; ce qui justifie *a posteriori* la dénomination que nous lui avons donnée.

Néanmoins, cet indicateur de climat ne suffit pas pour rendre compte de l'information commune contenue dans les quinze soldes d'opinion : d'autres liens existent lorsque l'analyse est menée à un niveau plus fin. D'où le problème d'interpréter les deux autres facteurs. Pour ce faire, on est amené à appliquer une transformation linéaire aux facteurs, de façon à ce que les nouveaux facteurs ainsi obtenus puissent faire l'objet d'une interprétation. Dans le cas présent, la recombinaison des trois facteurs communs initiaux fournit des résultats facilement explicables : une fois appliquée la procédure PROMAX, il apparaît clairement que les nouveaux facteurs obtenus sont relatifs à chacune des branches (cf *graphique 7* en annexe). En effet, lorsque l'on mène séparément sur chaque secteur une analyse factorielle, le facteur commun obtenu se trouve très proche de l'un des trois facteurs fournis par le modèle d'ensemble.

Pour résumer ce qui vient d'être dit, il apparaît en définitive que la construction d'un indicateur de climat pour l'industrie prise dans son ensemble comme pour chaque grande branche a du sens : le climat des affaires de l'industrie est présent à un niveau plus désagrégé, mais il se décline suivant des modalités différentes. En d'autres termes, on est fondé à parler d'un climat pour chaque grande branche, dont les variations diffèrent certes d'une branche à l'autre, mais présentent tout de même une grande cohérence : c'est cette cohérence dont rend compte le climat d'ensemble.

### ***4.3 Interprétation des facteurs communs et spécifiques au niveau des branches***

Nous présentons maintenant les résultats obtenus dans les différentes branches, en commentant tout d'abord les climats de chacune d'entre elles, avant d'en venir à la composante spécifique à chaque question.

Si l'on en croit les évolutions comparées des climats branche par branche (cf *graphique 6* en annexe), les effets du second choc pétrolier ont été moins prononcés dans la branche des biens d'équipement que dans les deux autres branches. Les industries de biens de consommation ont, très logiquement, bénéficié d'un rebond plus important que celles des biens intermédiaires lors du plan de relance gouvernemental de 1981. Ces dernières, sans doute plus sensibles à la conjoncture internationale, ont en revanche été tirées plus tôt par la reprise mondiale du milieu des années 80 (dès 1983 aux Etats-Unis, en 1984 en RFA), et semblent en avance sur les autres branches lors du cycle d'expansion initié par le contre-choc pétrolier, mais aussi au moment du ralentissement qui a suivi.

La décélération en deux temps (ralentissement en 90-91, fausse reprise en 92 et récession profonde en 93) est visible dans les trois secteurs, mais on peut noter que les biens d'équipement avaient touché un point bas dès 1991 (le niveau de l'indicateur correspondant est inférieur dès cette date à celui du début des années 80, ce qui n'est pas le cas pour les autres branches), ce qui constitue une illustration du recul de l'investissement trois années durant. La reprise de 1994 semble simultanée dans les trois branches, mais son ampleur apparaît plus marquée dans le secteur des biens intermédiaires : le retournement provenait pour une grande part d'un mouvement sur les stocks, qui stimula le commerce interindustriel.

L'analyse de la composante spécifique de chaque réponse recueillie au niveau des branches est quelquefois rendue difficile par le profil instable que certaines d'entre elles peuvent avoir (instabilité en général plus heurtée qu'au niveau de l'industrie dans son ensemble). Néanmoins, l'étude systématique des questions pour lesquelles l'information spécifique est importante peut aider le conjoncturiste dans la formulation de son diagnostic. Plutôt que de se livrer à un commentaire conjoncturel des quinze soldes d'opinion (cinq questions, trois branches), on se contentera ici de dégager quelques faits saillants.

Au niveau de l'industrie prise dans son ensemble, la composante commune semblait très proche des de la réponse portant sur la **tendance passée de la production** (cf *graphique 8.1* en annexe) et, dans une moindre mesure, de celle portant sur le **niveau des carnets de commande** (cf *graphique 8.3* en annexe). La branche des biens d'équipement semble toutefois sortir nettement de ce schéma : les carnets de commande y jouent un rôle beaucoup plus important que la tendance passée de la production. La longueur et le coût des processus de production, sans doute plus importants pour ce qui concerne les biens d'équipement, peuvent expliquer que la réponse concernant la tendance passée de la production y soit nettement moins proche de l'indicateur de climat que dans les autres secteurs : la conjoncture y apparaît bien plus dépendante du nombre de commandes d'ores et déjà enregistrées, et donc de l'état des carnets, que de la production effectuée dans les mois écoulés.

Cette particularité explique peut-être aussi la spécificité de la question relative aux **stocks dans les industries de biens d'équipement** (cf *graphique 8.5* en annexe). C'est en effet la seule question qui soit "mal expliquée" (en terme de décomposition de variance) par la composante commune : le niveau des stocks est sans doute bien moins rapidement adaptable aux aléas de la conjoncture qu'il ne l'est dans les deux autres secteurs.

Au-delà d'un intérêt d'ordre conjoncturel, l'amplitude moyenne de la composante spécifique relative à la question sur **la demande et les carnets étrangers** (cf *graphique 8.4* en annexe) constitue une mesure indirecte du poids des marchés étrangers dans la détermination de l'activité du secteur. On constate ainsi sur longue période que la branche pour laquelle ce solde d'opinion est le plus proche du climat



sectoriel correspondant est celle des biens d'équipement, suivie par celles des biens intermédiaires. Cette plus grande influence de l'étranger sur l'activité est conforme avec les données de la comptabilité nationale, puisque la part de la production exportée se montait en 1994 à environ 45% dans la branche des biens d'équipement professionnels, à 37% dans celle des biens intermédiaires et à seulement 29% pour ce qui concerne les biens de consommation.

## **Conclusion**

Dans cette étude, nous estimons des modèles factoriels dynamiques en utilisant aussi bien des techniques standard d'analyse factorielle que le filtre de Kalman. Nous montrons que ce type de modèle peut être utilisé pour aider à analyser l'information contenue dans les enquêtes de conjoncture.

Une extension naturelle de ce travail consisterait à mélanger des données provenant de différents secteurs de façon à construire un indicateur de climat plus général. La modélisation espace-état permet aisément de résoudre la difficulté technique relative au mélange de données de périodicités différentes.

Une autre extension possible serait d'utiliser le même genre de techniques à des fins prévisionnelles. En fait, l'indicateur construit ici semble plutôt coïncider. Une étape supplémentaire consisterait à chercher à construire un indicateur avancé d'activité.

---

## BIBLIOGRAPHIE

---

- Cureton E. E., Mulaik S. A. (1975). « The weighted Varimax rotation and the Promax rotation », *Psychometrika*, 40, 183-195.
- Engle R.F., Watson M.W., (1981). « A one factor multivariate time series model of metropolitan wage rates », *Journal of American Statistical Association*, 76, 774-781.
- Engle R.F., Watson M.W., (1983). « Alternative algorithms for the estimation of dynamic factor, MIMIC, and varying coefficient regression models », *Journal of Econometrics*, 23, 385-400.
- Forni, M., Reichlin L. (1995). « Let's get real : a dynamic factor analytical approach to disaggregated business cycle », *Center for Economic Policy Research Discussion Paper n° 1244*.
- Geweke J. (1977). « Labor turnover and employment dynamics in US manufacturing », in *New methods in business cycles research*, Sims Ed., Minneapolis : Federal Reserve Bank of Minneapolis.
- Geweke J., Singleton K.J. (1981). « Maximum likelihood "confirmatory" factor analysis of economic time series », *International Economic Review*, 22 (1), 37-54.
- Gourieroux C., Monfort A. (1990) *Séries temporelles et modèles dynamiques*, Economica
- Harvey, A.C. (1989). *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press.
- Lawley, D.N., Maxwell A.E. (1971). *Factor Analysis as a Statistical Method*, New York Macmillan Publishing Co., Inc.
- Newey W.K., Mac Fadden D. (1994). « Large Sample Estimation and Hypothesis testing », in *Handbook of Econometrics*, vol. 4, 2113-2245.
- Quah D., Sargent T.J. (1993). « A dynamic index model for large cross-sections », in *Business cycles, indicators and forecasting*, J.H. Stock and M.W. Watson Ed, University of Chicago Press.
- Sargent, T.J., Sims S.A. (1977). « Business cycle modelling without pretending to have too much *a priori* economic theory », in *New methods in business cycles research*, Sims Ed., Minneapolis : Federal Reserve Bank of Minneapolis.

Stock, J.H., Watson M.W. (1989). « New indexes of coincident and leading indicators », in *NBER Macroeconomics Annual*, Blanchard & Fisher Ed., MIT Press, Cambridge.

Stock J.H., Watson M.W. (1991). « A probability model of the coincident economic indicators », in *Leading economic indicators : new approaches and forecasting records*, K.Lahiri and G.H.Moore Ed, Cambridge University Press.

Stock J.H., Watson M.W. (1993). « A procedure for predicting recessions with leading indicators : econometric issues and recent experience », in *Business cycles, indicators and forecasting*, J.H.Stock and M.W.Watson Ed, University of Chicago Press.

Watson M.W., Kraft D.F. (1984). « Testing the interpretation of indices in a macroeconomic index model », *Journal of Monetary Economics*, 13, p 165-181

Watson M. W. (1986). « Univariate detrending methods with stochastic trends », *Journal of Monetary Economics*, 18, 49-75

---

## ANNEXES

---

*Preuve du lemme 1:*

i) Il s'agit de vérifier que si l'un des paramètres  $\lambda_{ij}, i = 1 \dots I, j = 1, \dots, J$  ou  $\sigma_i^2, i = 1 \dots I$  tend vers l'infini alors  $Q_T(y, \theta) \rightarrow -\infty$ . Nous allons en effet montrer que, dans ce cas,  $\det C \rightarrow +\infty$ , alors que  $\text{tr } C^{-1}A$  reste borné, ce qui entraîne le résultat annoncé.

- Montrons d'abord que : si  $\|C\| \rightarrow +\infty$ , alors  $\det C \rightarrow +\infty$ .

Comme on a supposé que  $\sigma_i^2 \neq 0$ , pour  $i = 1 \dots I$ , la matrice  $D$  est définie positive. Comme, en outre,  $C \geq D$ ,  $C$  est aussi définie positive. Pour montrer que  $\det C \rightarrow +\infty$ , il suffit donc de montrer que l'une au-moins des valeurs propres de  $C$  tend vers  $+\infty$ .

Or on sait que :  $\|M\|_2^2 = \text{Max}_{y \neq 0} \frac{y' M' M y}{y' y}$  définit une autre norme sur l'espace des matrices, et que, si l'on désigne par  $\alpha_i, i = 1 \dots I$  les valeurs propres de  $C$  matrice définie positive, cette norme vérifie :  $\|C\|_2 = \text{Max}_i \alpha_i$ .

Par équivalence des normes sur l'espace des matrices, on sait que  $\|C\|_2 \rightarrow +\infty$ , lorsque  $\|C\| \rightarrow +\infty$ . On obtient ainsi le résultat.

- Montrons maintenant que  $\text{tr } C^{-1}A$  reste borné. Tout d'abord, il est clair que :

$$|\text{tr } C^{-1}A| \leq I^2 \|C^{-1}\| \|A\|.$$

Si l'on note :  $\sigma^2 = \text{Min}_i \sigma_i^2$ , comme  $C \geq D > 0$ , on sait que les valeurs propres de  $C$  sont toutes supérieures à  $\sigma^2$ . D'après le résultat précédemment cité sur l'équivalence des normes de matrices, il existe donc un réel  $k$  tel que :  $\|C^{-1}\| \leq \frac{k}{\sigma^2}$ . Comme on a, de plus, supposé qu'il existe un réel  $\alpha > 0$ , tel que :  $\forall \theta \in \Theta, \forall i = 1, \dots, I, \sigma_i^2 \geq \alpha$ , on obtient :  $|\text{tr } C^{-1}A| \leq \frac{kI^2}{\alpha} \|A\|$ , d'où le résultat.

*Preuve du lemme 2 :* Il faut montrer que :  $\text{Max}_{\theta \in \Theta} |Q_T(y, \theta) - Q_0(\theta)| \xrightarrow{P} 0$ .

Or cette expression est égale à :  $\text{Max}_{\theta \in \Theta} \left| \text{tr} C^{-1} \left( \frac{1}{T} \sum_t x_t x_t' \right) - \text{tr} C^{-1} C_0 \right|$ , et on peut écrire :

$$\left| \text{tr} C^{-1} \left( \frac{1}{T} \sum_t x_t x_t' \right) - \text{tr} C^{-1} C_0 \right| \leq \left| \text{tr} C^{-1} \left( \frac{1}{T} \sum_t x_t x_t' - C_0 \right) \right| \leq I^2 \|C^{-1}\| \left\| \frac{1}{T} \sum_t x_t x_t' - C_0 \right\|$$

En raisonnant comme précédemment, on obtient :

$$|Q_T(y, \theta) - Q_0(\theta)| \leq \frac{kI^2}{\alpha} \left\| \frac{1}{T} \sum_t x_t x_t' - C_0 \right\|,$$

et cette inégalité est valable quel que soit  $\theta \in \Theta$ .

Comme par hypothèse  $(y_t)$  est stationnaire, et comme  $\theta_0$  est la vraie valeur du paramètre, on sait que :  $\left( \frac{1}{T} \sum_t x_t x_t' \right) \xrightarrow{P} C_0$ .

On déduit alors de ce qui précède que :  $|Q_T(y_t, \theta) - Q_0(\theta)| \xrightarrow{P} 0$ , uniformément sur  $\Theta$ .

*Preuve du lemme 3 :* On considère un processus  $(z_t)$  non autocorrélé, tel que pour tout  $t$ , la loi de  $z_t$  est la loi  $N(0, C)$ , où  $C = \Lambda \Lambda' + D$  est définie comme précédemment, et où  $C_0 = \Lambda_0 \Lambda_0' + D_0$  est la matrice de variance-covariance associée à la vraie valeur des paramètres. On note  $f(z_t, \theta)$  et  $f(z_t, \theta_0)$  les densités associées.

Les propriétés de l'information de Kullback conduisent à l'inégalité suivante :

$$E_{\theta_0} \left( \text{Ln} \left( \frac{f(z_t, \theta)}{f(z_t, \theta_0)} \right) \right) < 0 \text{ si } \theta \neq \theta_0.$$

Pour  $\theta \neq \theta_0$ , cette inégalité entraîne que :

$$E_{\theta_0} \left( \frac{1}{T-1} \sum_i \text{Ln}(f(z_i, \theta)) \right) < E_{\theta_0} \left( \frac{1}{T-1} \sum_i \text{Ln}(f(z_i, \theta_0)) \right),$$

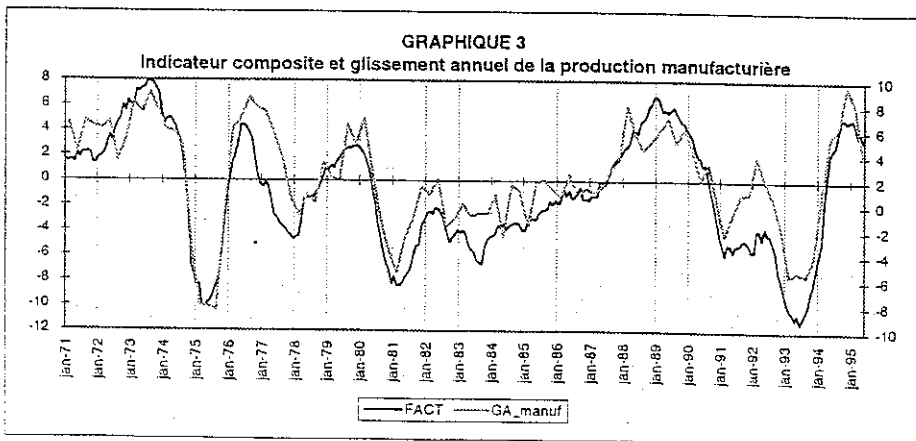
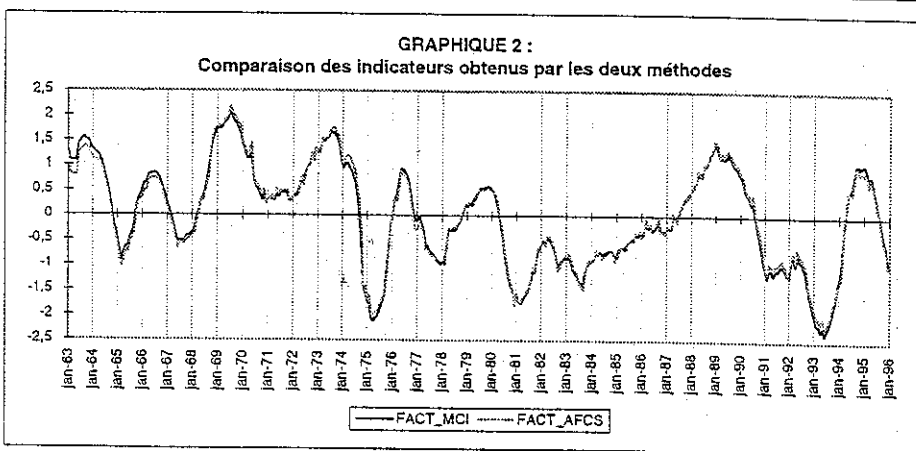
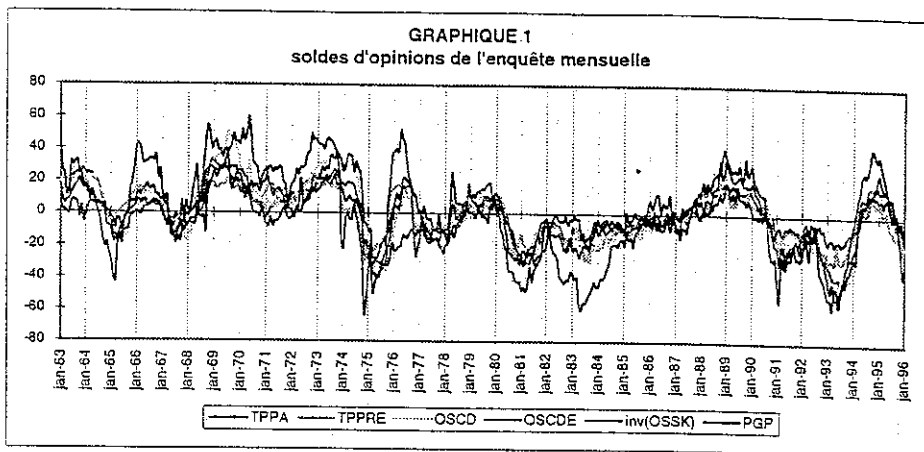
c'est-à-dire que :  $-\frac{1}{2} \text{Ln det } C - \frac{1}{2} \text{tr } C^{-1} C_0 < -\frac{1}{2} \text{Ln det } C_0 - \frac{1}{2} \text{tr } C_0^{-1} C_0,$

ou encore que :  $Q_0(\theta) < Q_0(\theta_0)$ <sup>7</sup>.

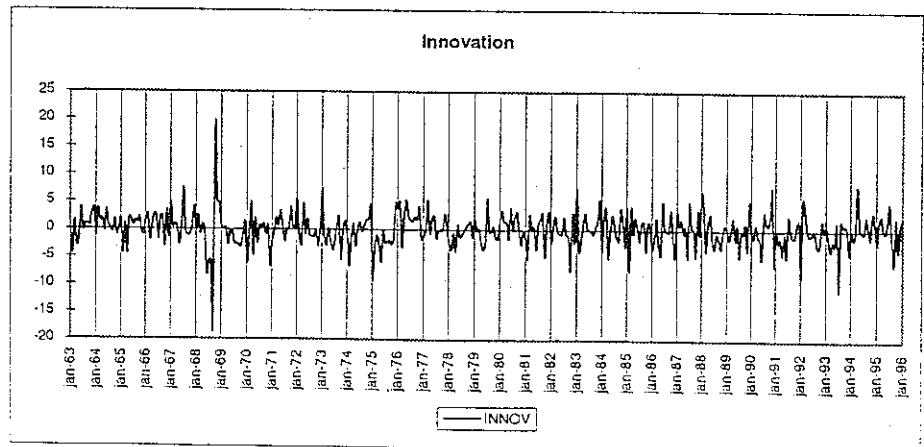
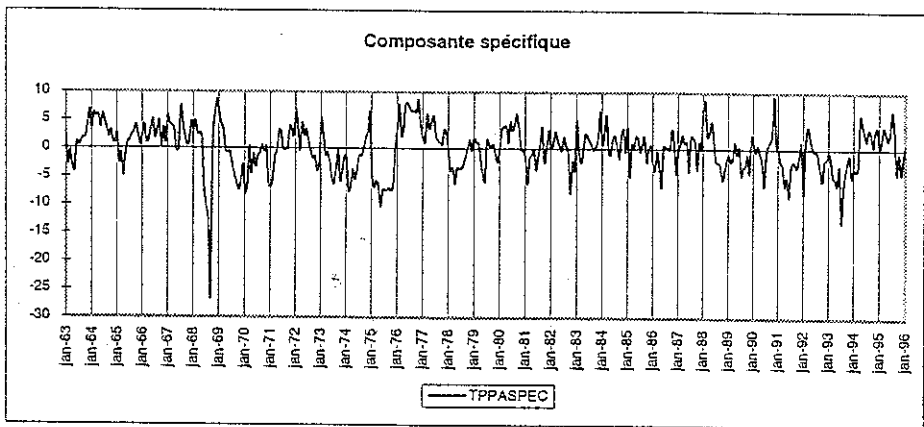
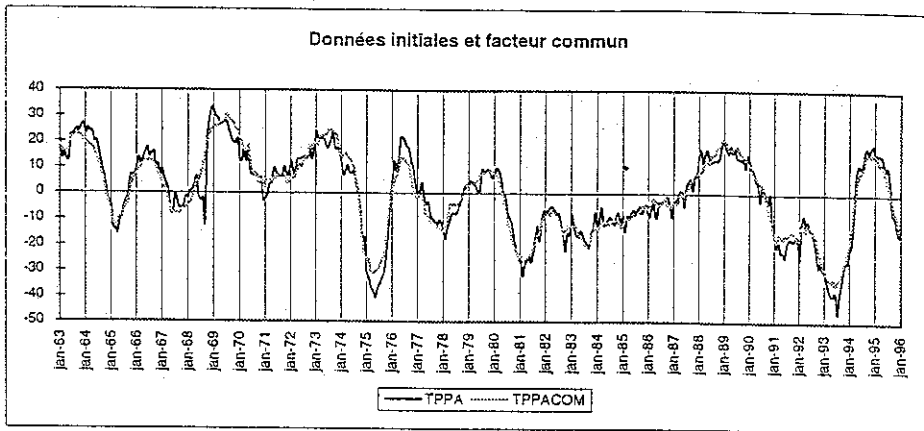
*Preuve du lemme 4 :* Ceci résulte de façon immédiate de la continuité de l'application qui à  $\theta$  associe  $C$ , des applications *trace* et *déterminant*, ainsi que de l'application qui à  $C$  associe  $C^{-1}$  (cette dernière application étant bien définie sur le compact  $\Theta$ ).

---

<sup>7</sup> On notera que le raisonnement qui est fait ici consiste à appliquer l'inégalité de l'information de Kullback à la pseudo-loi, en faisant momentanément comme si c'était la vraie loi. Cette inégalité est en effet une propriété de la loi étudiée, qu'elle soit ou non la vraie loi des observations.

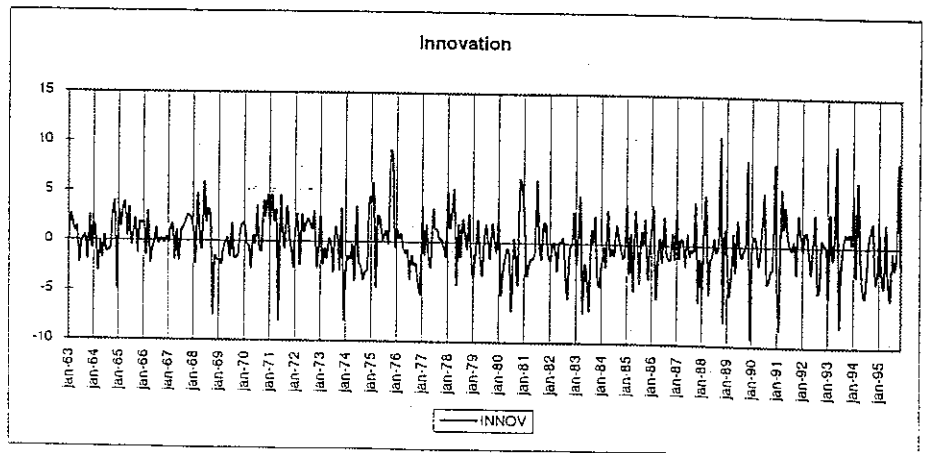
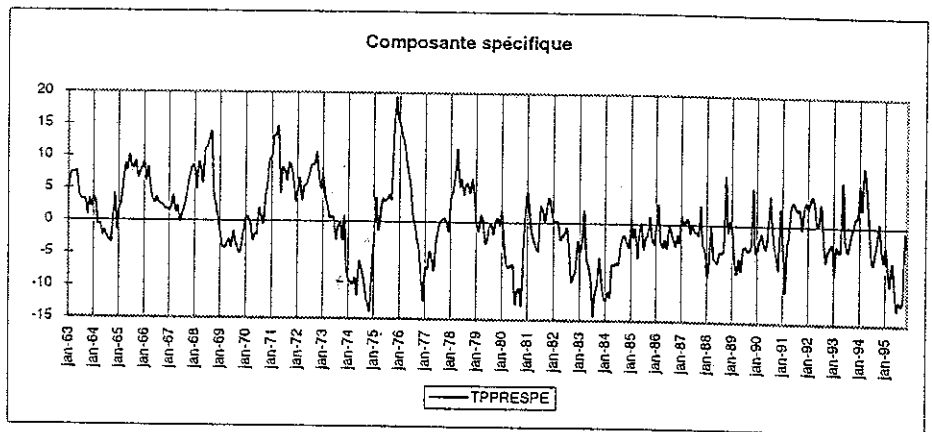
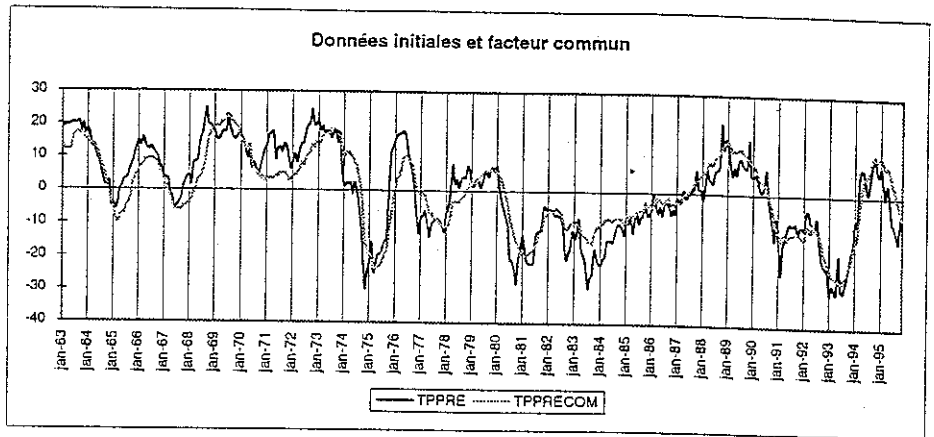


GRAPHIQUE 4.1  
TENDANCE DE LA PRODUCTION PASSEE

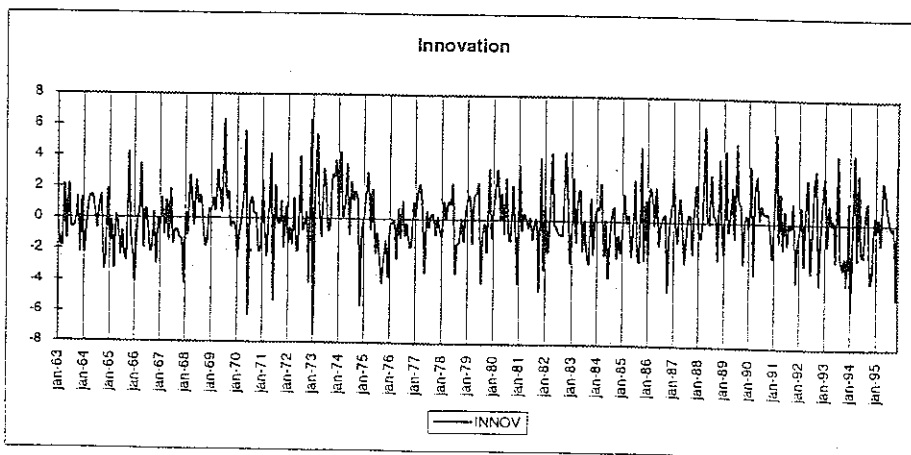
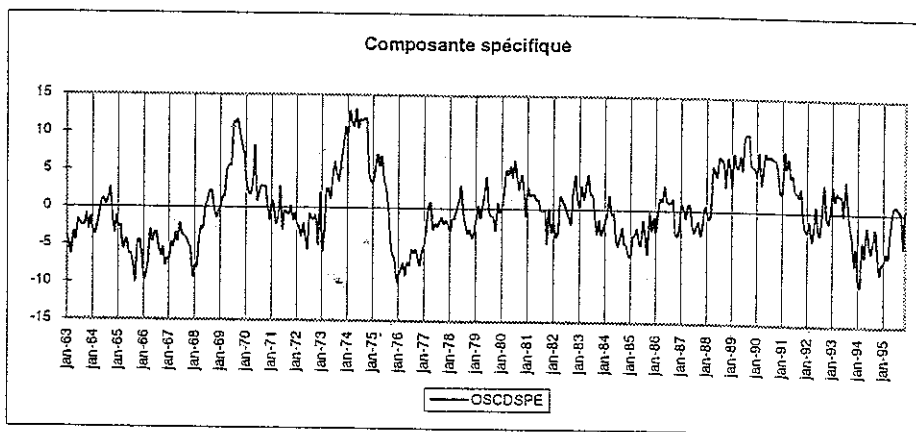
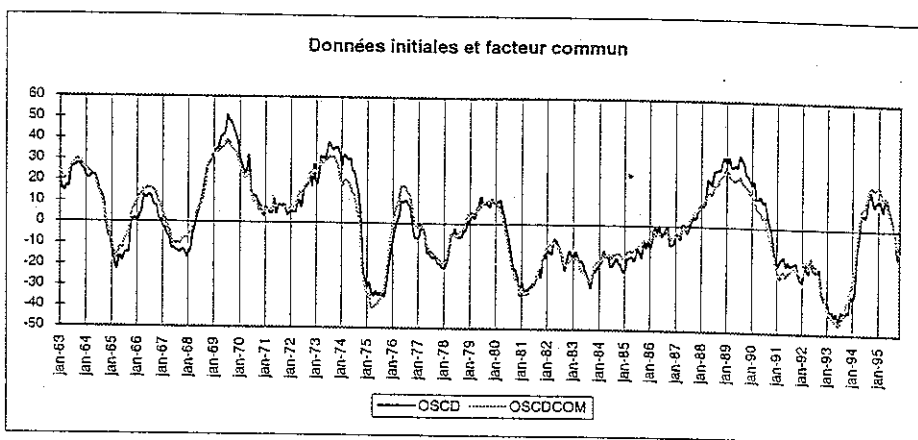




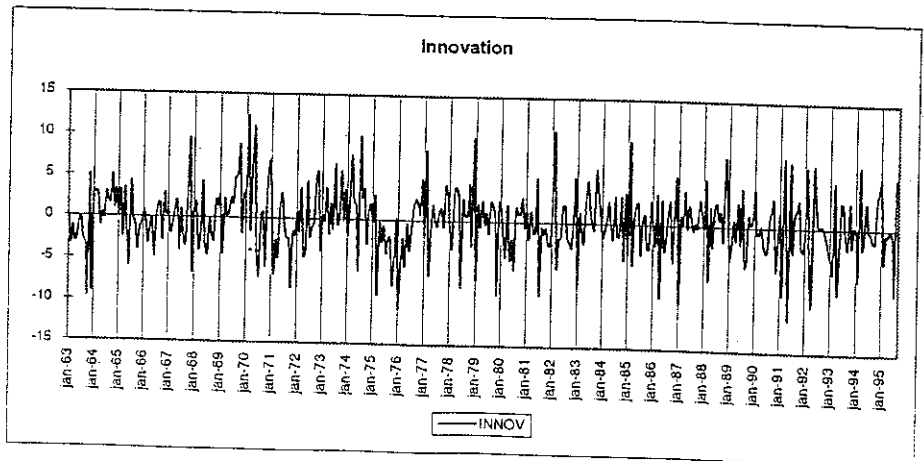
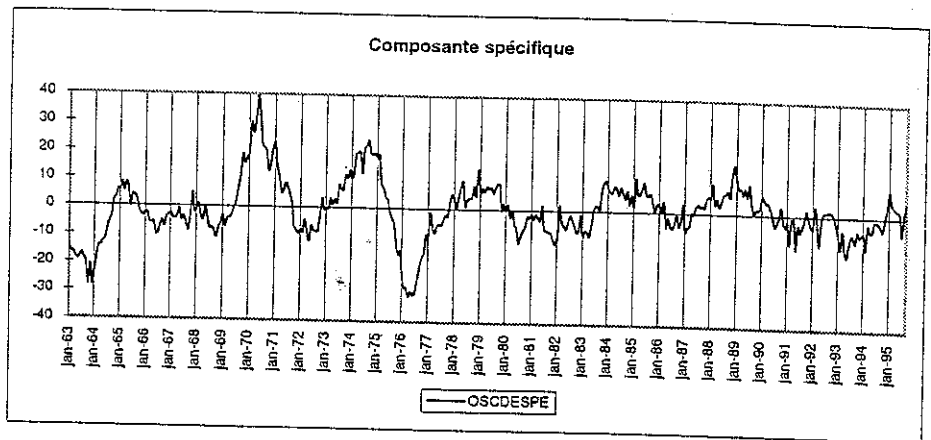
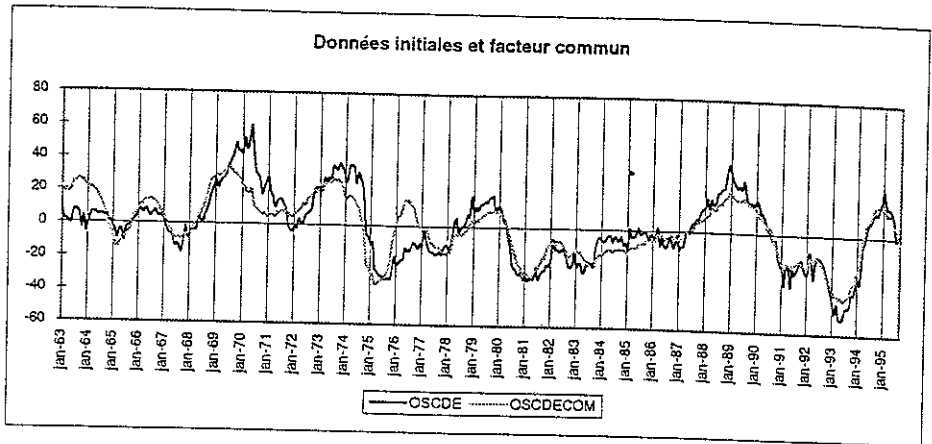
GRAPHIQUE 4.2  
TENDANCE DE LA PRODUCTION PREVUE



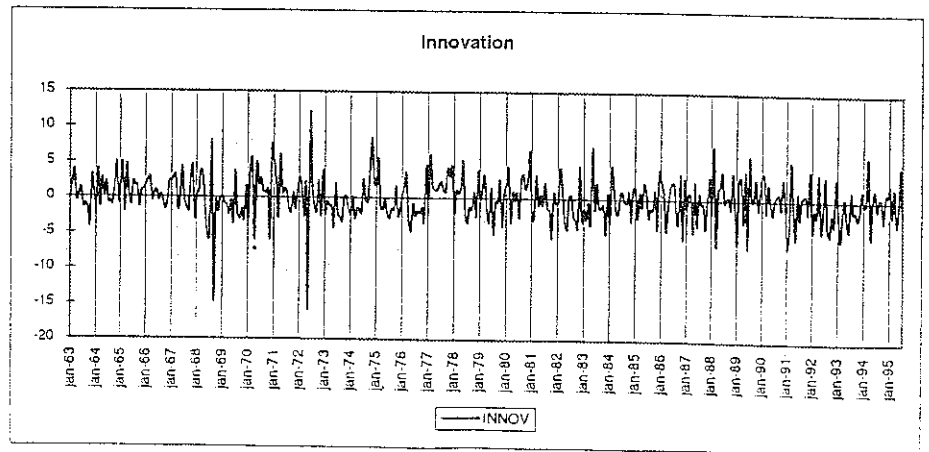
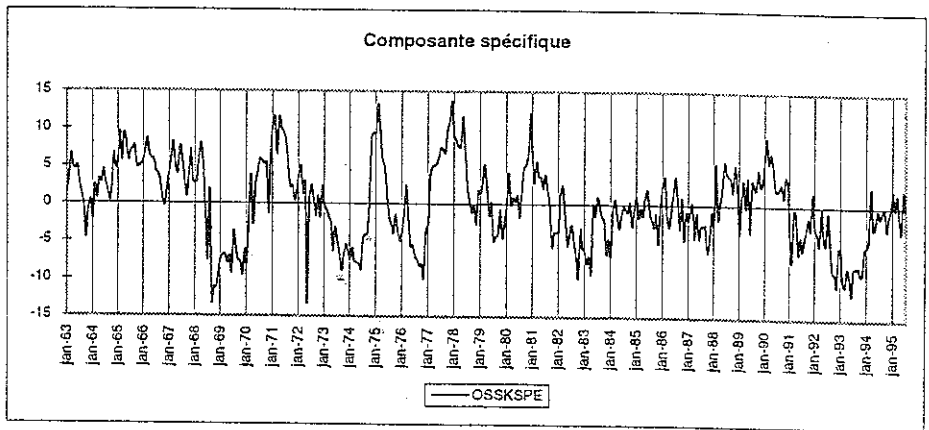
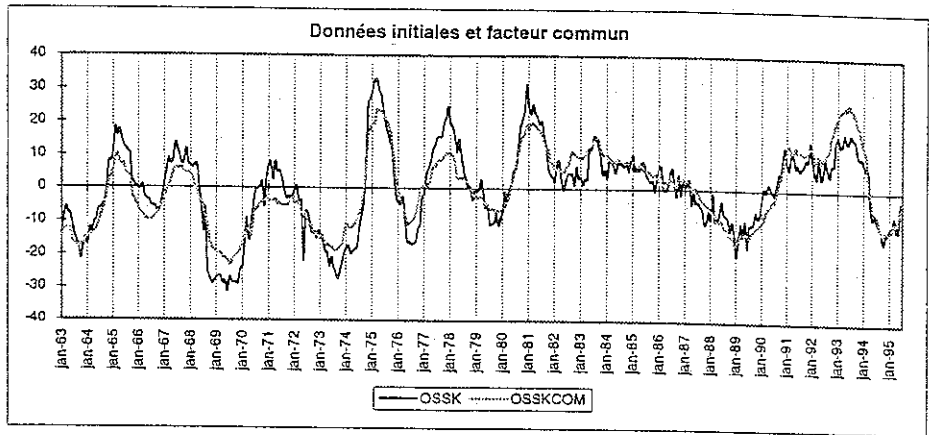
GRAPHIQUE 4.3  
CARNETS DE COMMANDE



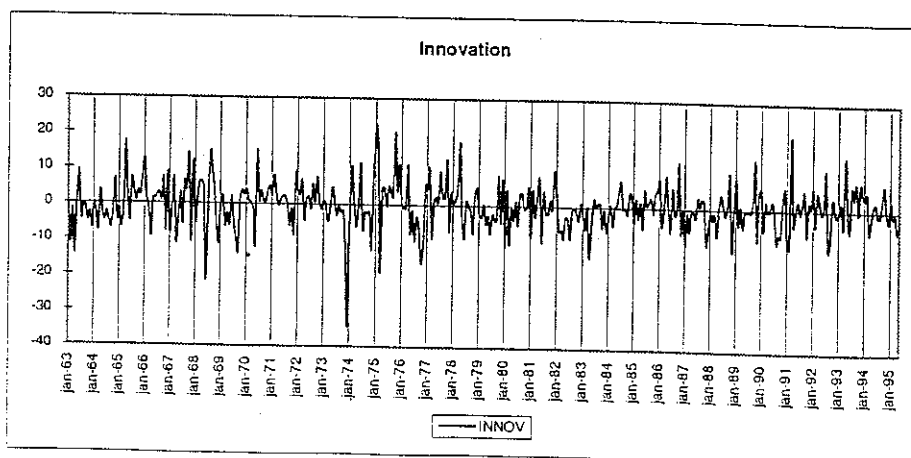
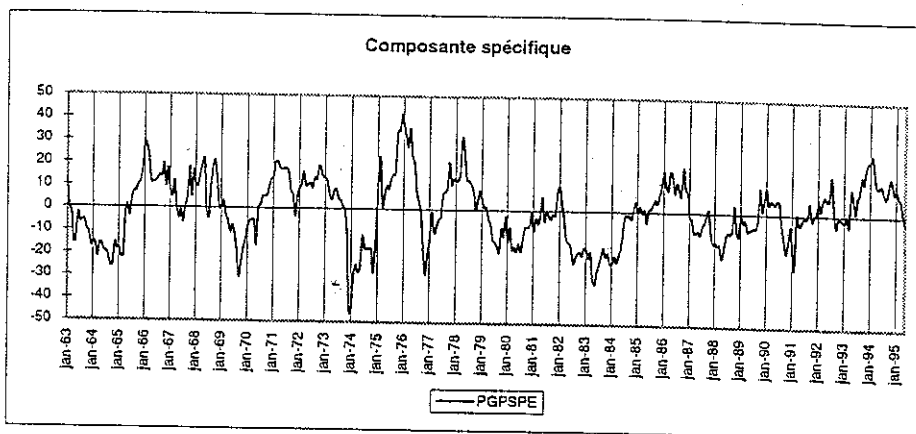
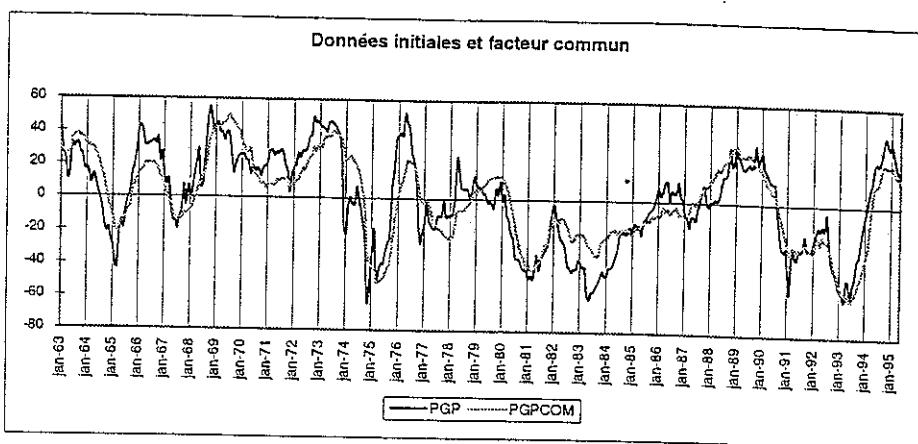
GRAPHIQUE 4.4  
CARNETS DE COMMANDE ETRANGERS

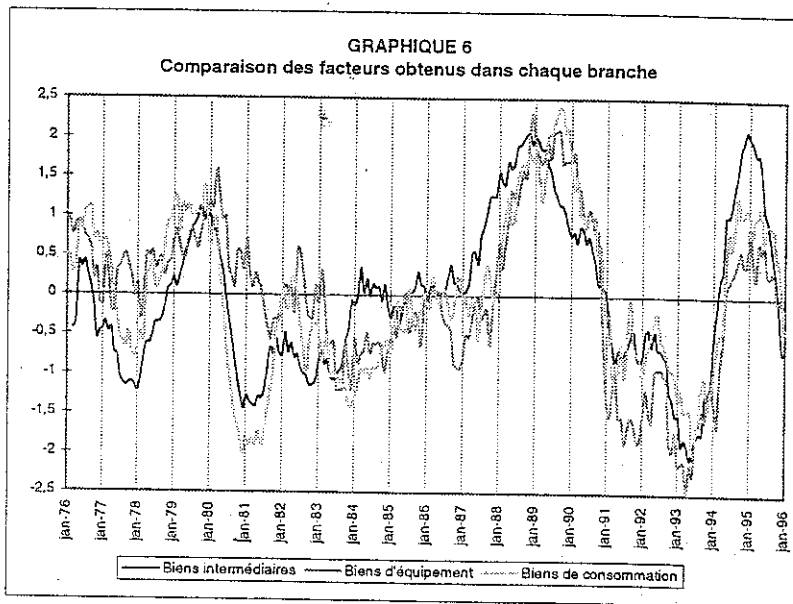
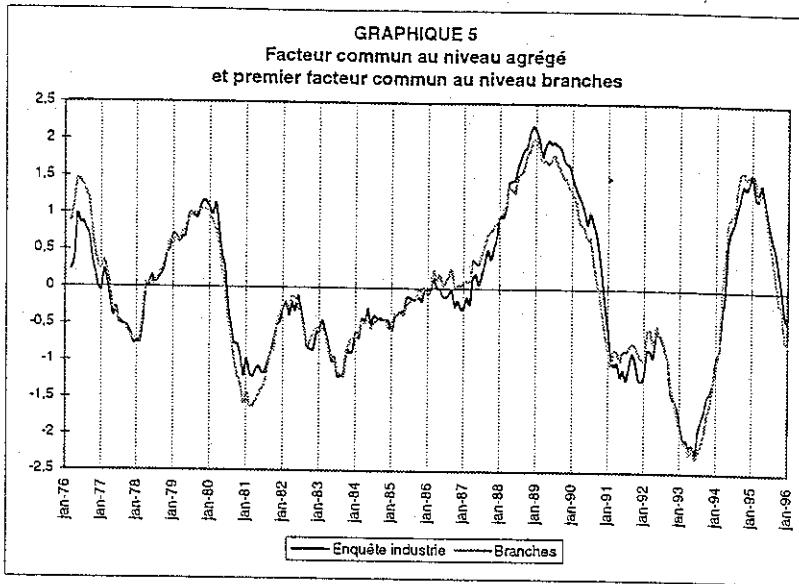


GRAPHIQUE 4.5  
OPINION SUR LES STOCKS

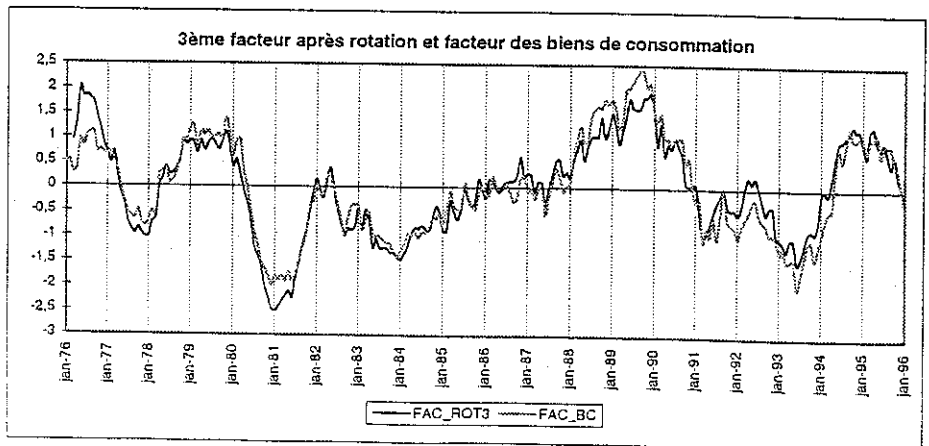
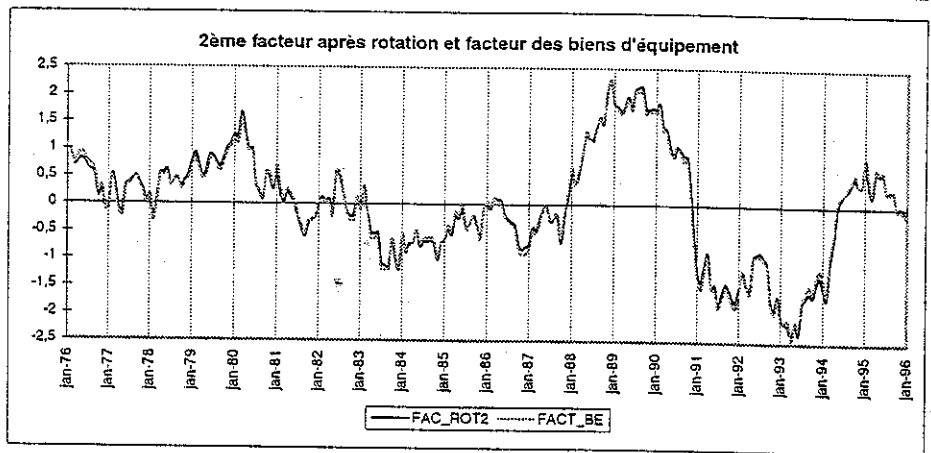
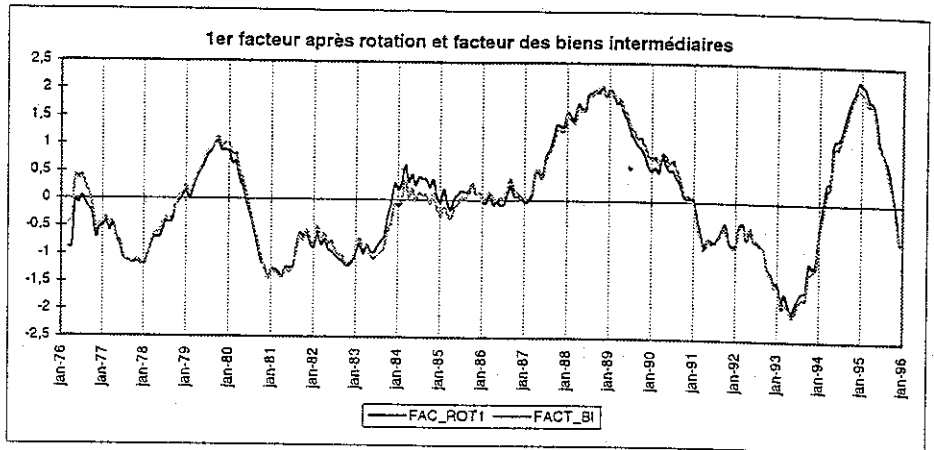


GRAPHIQUE 4,6  
PERSPECTIVES GÉNÉRALES DE PRODUCTION

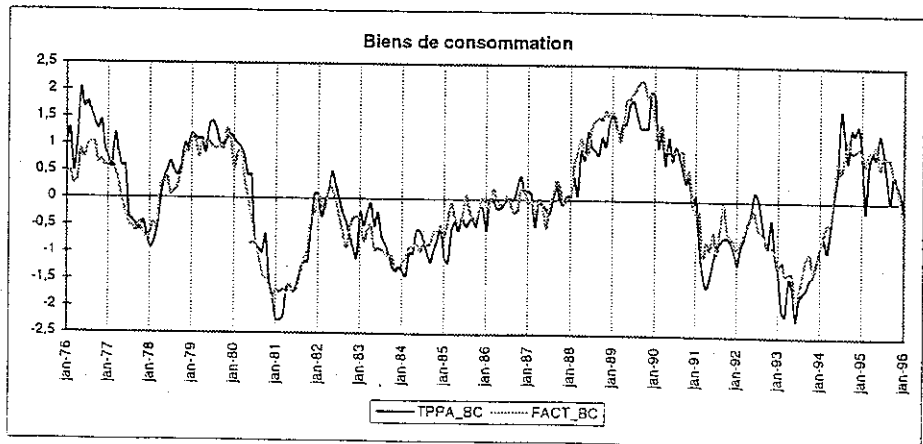
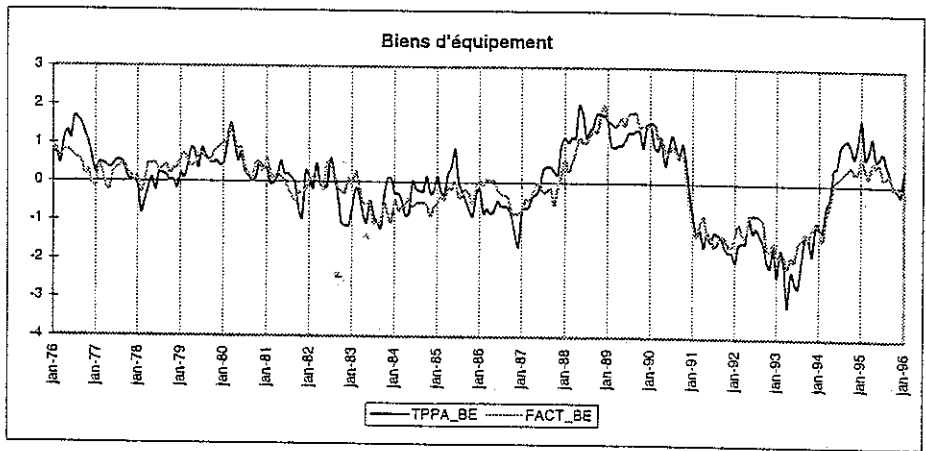
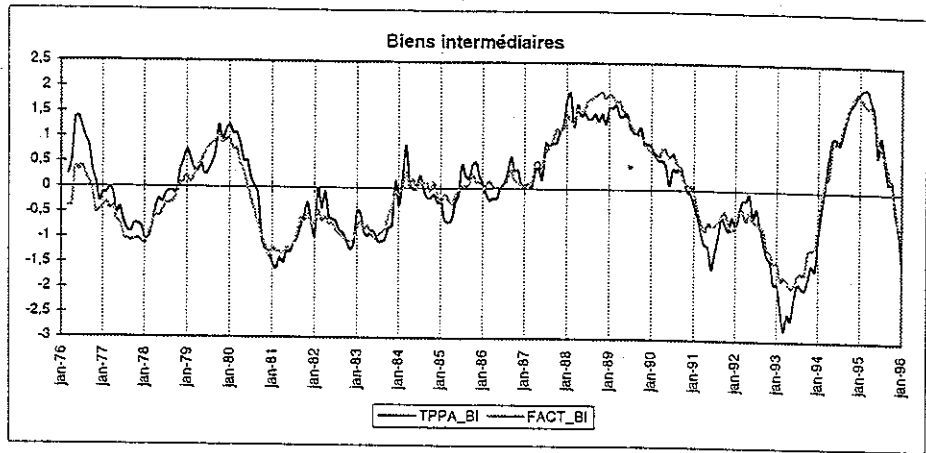




**GRAPHIQUE 7**  
**FACTEURS APRES ROTATION**  
**ET FACTEURS OBTENUS DANS CHAQUE BRANCHE**

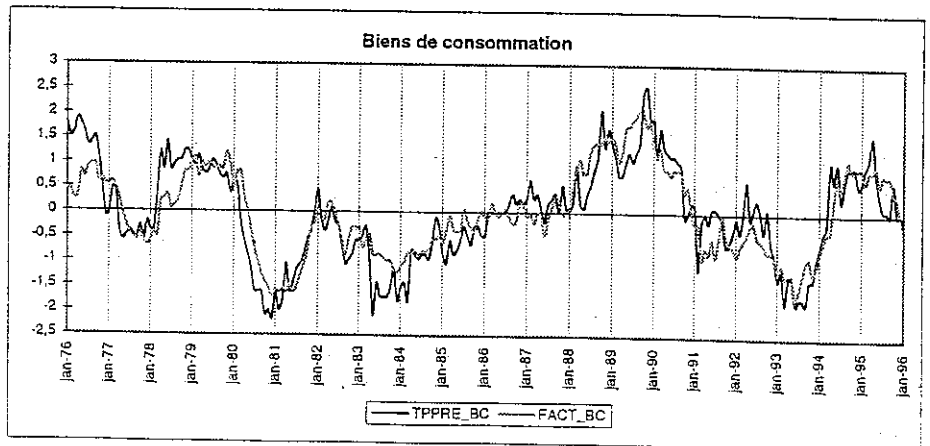
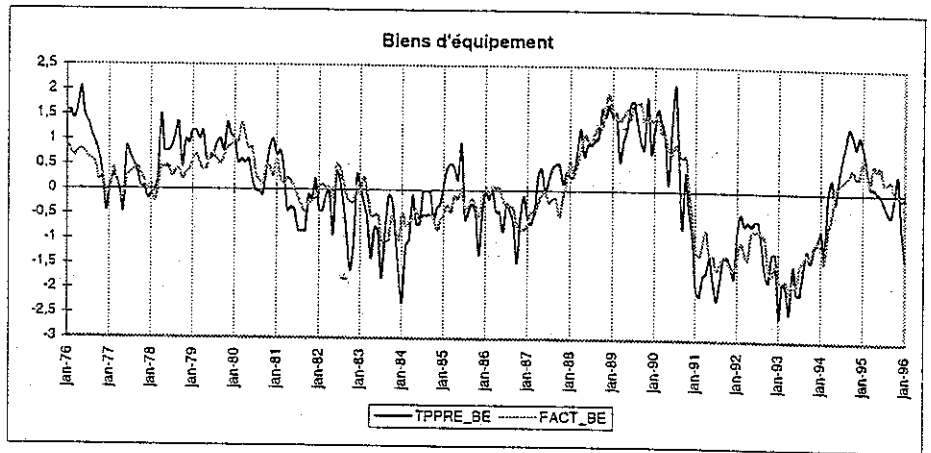
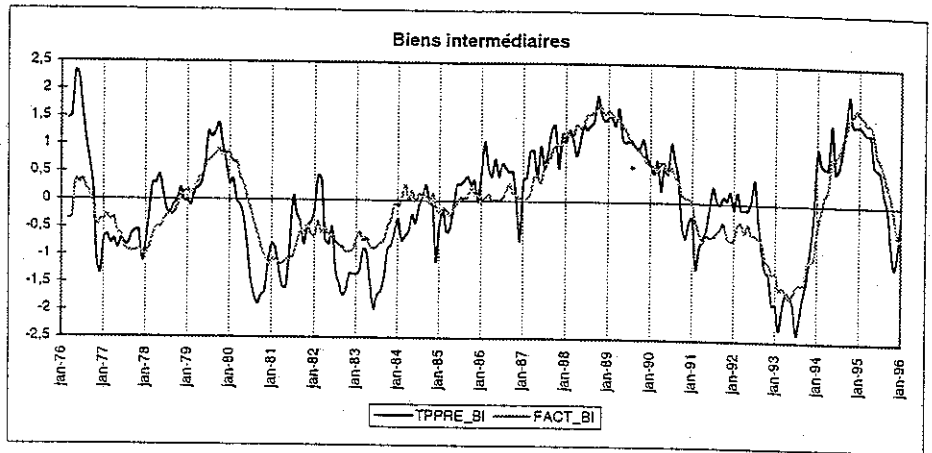


**GRAPHIQUE 8.1 : TENDANCE DE LA PRODUCTION PASSEE**  
Données initiales et facteurs par branches

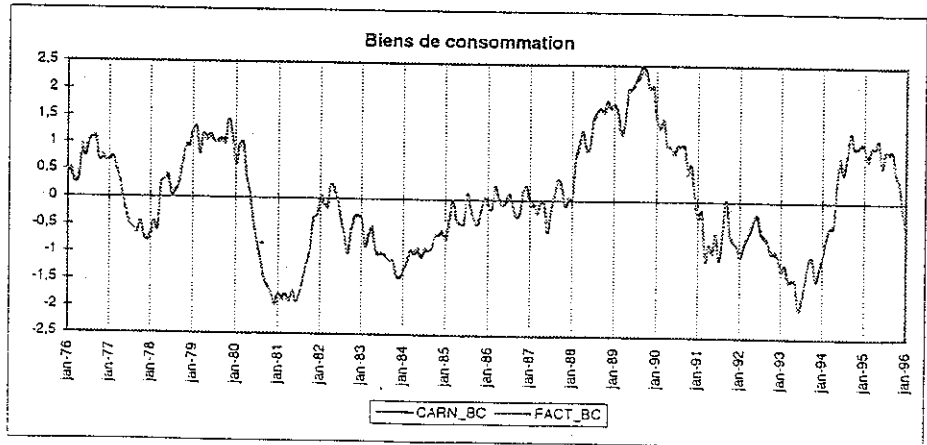
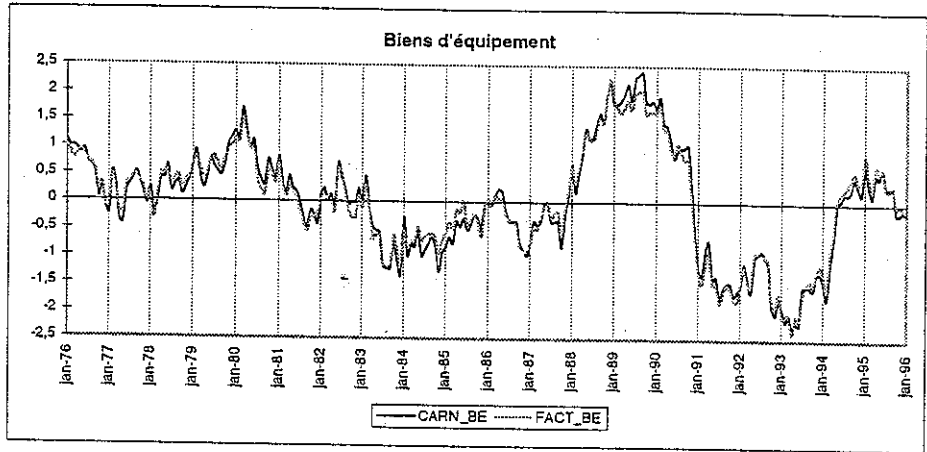
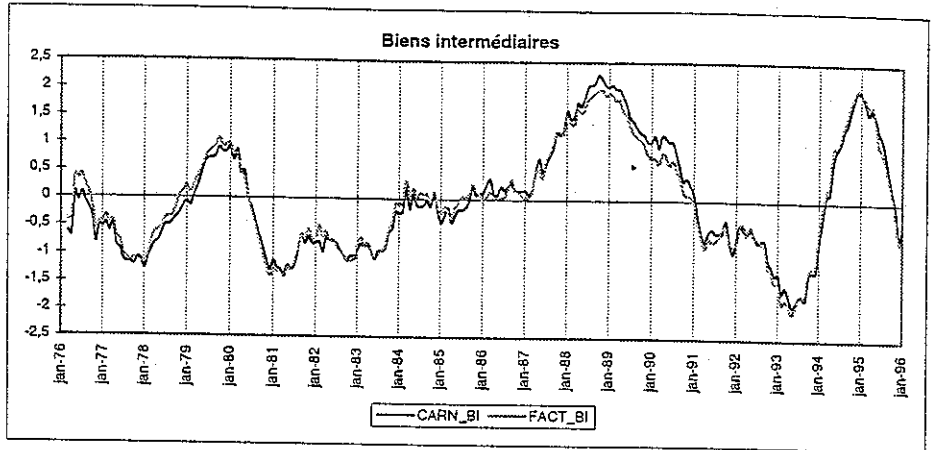




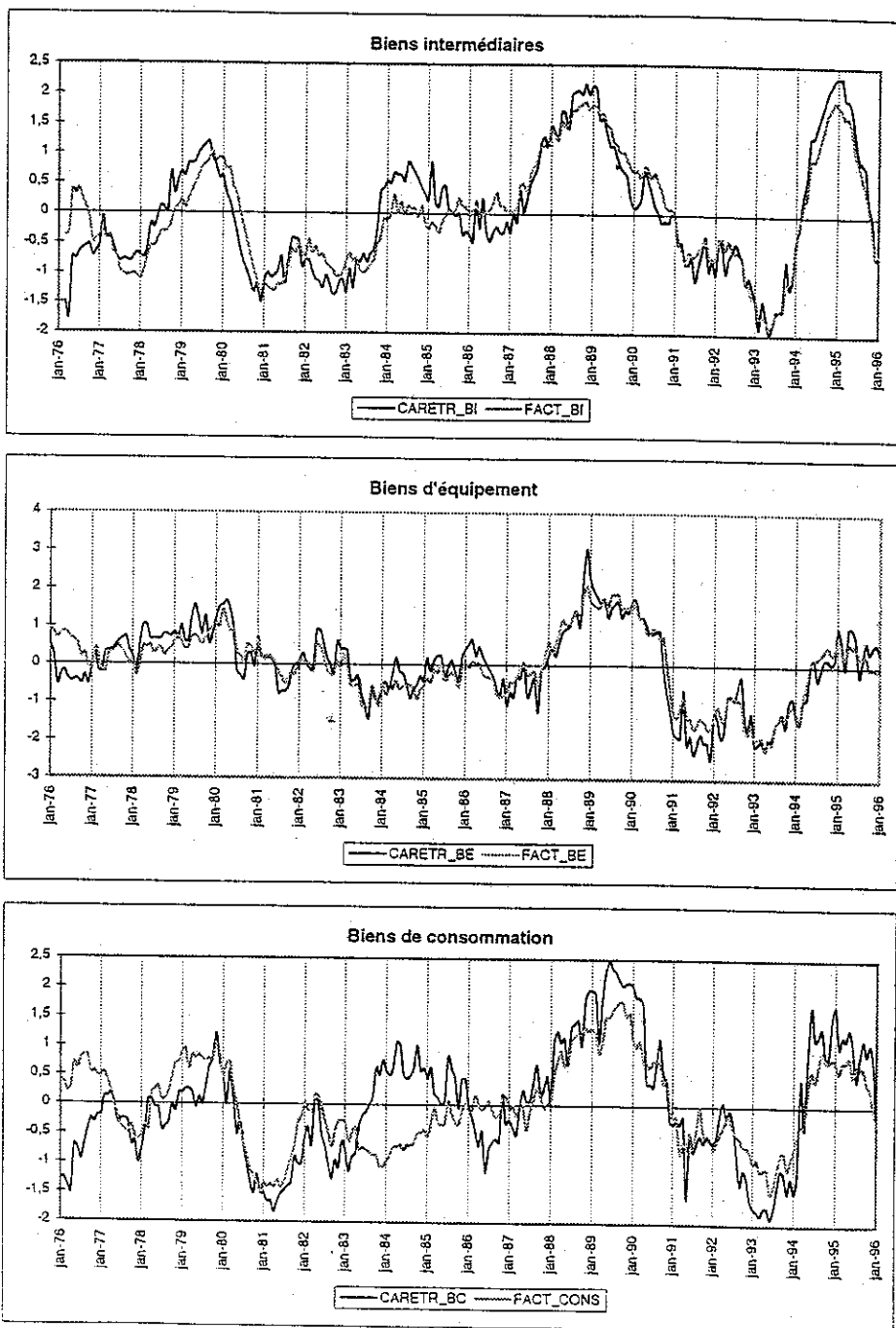
**GRAPHIQUE 8.2 : TENDANCE FUTURE DE LA PRODUCTION**  
Données initiales et facteurs communs par branches



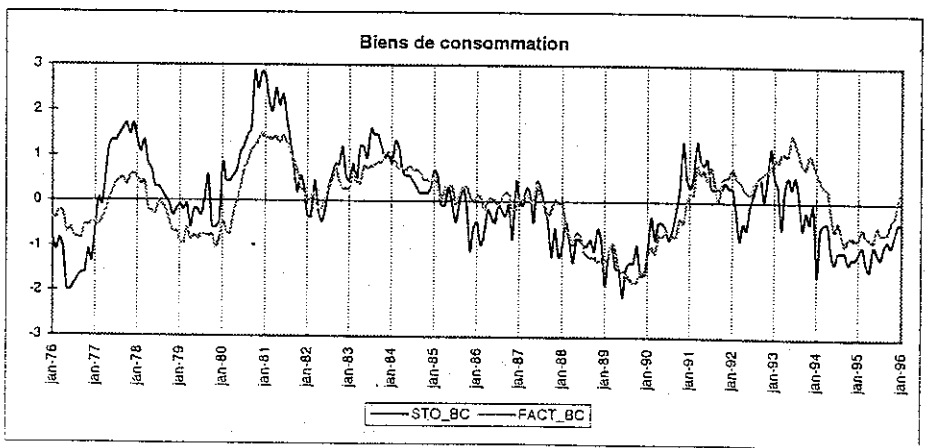
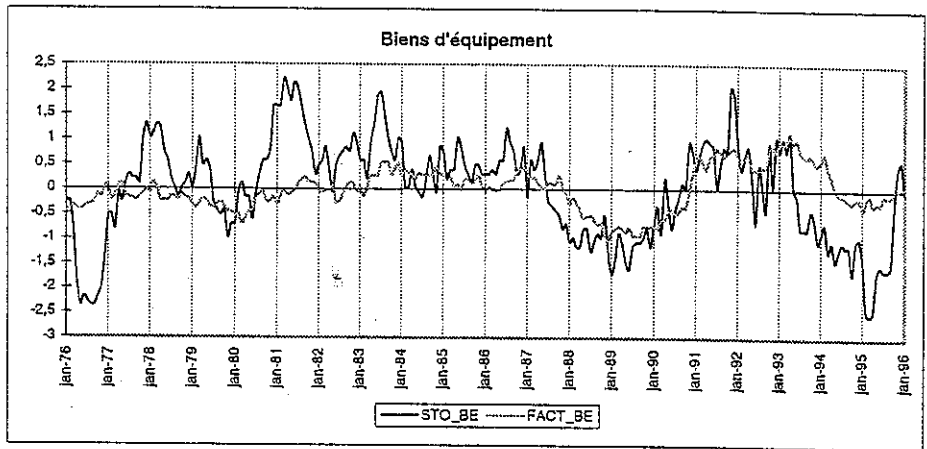
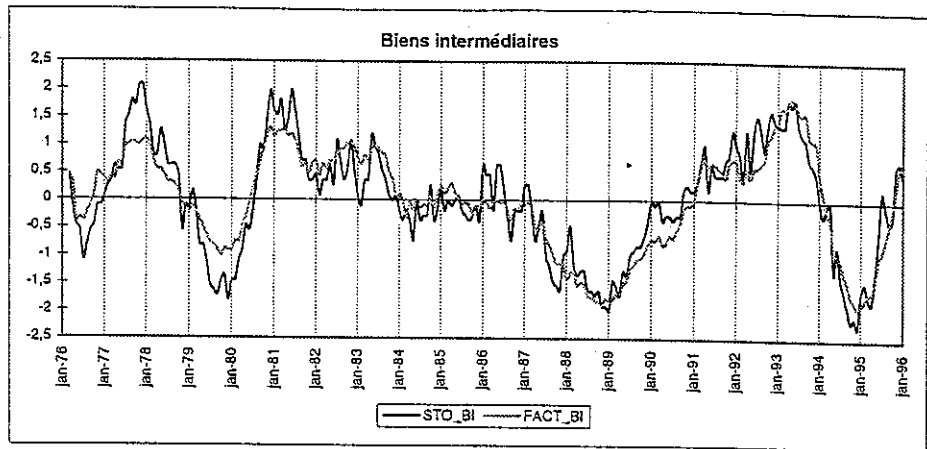
GRAPHIQUE 8.3 : CARNETS DE COMMANDE  
Données initiales et facteurs communs par branche



GRAPHIQUE 8.4 : CARNETS DE COMMANDE ETRANGERS  
Données initiales et facteurs communs par branche



GRAPHIQUE 8.5 : OPINION SUR LES STOCKS  
Données initiales et facteurs communs par branches



INSEE AI DG 12 1996