

# L'estimation des taux de chômage au niveau infra-annuel national

Thierry LACROIX

Insee

L'utilisation d'un taux de chômage comme indicateur synthétique conjoncturel des déséquilibres affectant le marché du travail s'est imposée dans les principaux pays développés. Le calcul et la publication d'indicateurs "officiels" sont pourtant intervenus plus tardivement en France que dans les pays anglo-saxons. Il a en effet fallu attendre 1986 pour que soit diffusé chaque mois un taux de chômage national, même si des estimations trimestrielles infra-nationales étaient déjà établies depuis l'année précédente.

Outre une certaine sacralisation des observations directes portant sur des chiffres en niveau issus de sources exhaustives (le nombre des demandeurs d'emploi inscrits à l'ANPE), sans doute plus marquée dans les sociétés latines, la difficulté de construction et de gestion d'un indicateur de taux de chômage explique probablement la lenteur qui a présidé à son élaboration en France.

## I - CONTRAINTES ET OBJECTIFS DES TAUX DE CHOMAGE

Un taux de chômage est un indicateur synthétique du déséquilibre entre l'offre et la demande d'emploi pour un groupe ou un marché donné, disons pour une population  $P$  donnée ayant une activité professionnelle. Si on le note  $T$  et si  $E$  et  $C$  représentent respectivement les effectifs de  $P$  ayant un emploi ou se trouvant en chômage alors :

$$T = \frac{C}{P} = \frac{C}{E + C} \quad (1)$$

Pour que (1) soit valide il faut que les variables figurant au numérateur et au dénominateur ( $E$  et  $C$ ) soient homogènes et en particulier

évaluées à l'aide des mêmes concepts.  $P = E + C$  doit ainsi pouvoir être considérée comme une évaluation de population active.

L'objectif d'un taux de chômage est triple :

- pouvoir apprécier l'évolution du déséquilibre du marché du travail dans le temps pour une population donnée en s'affranchissant de l'effet taille ;
- pouvoir comparer entre elles des valeurs de chômage observées pour des populations de taille différente ;
- pouvoir comparer leurs évolutions.

Contrairement à tout indicateur de niveau, un taux de chômage constitue donc un indicateur opérationnel qui peut fonder (pas à lui seul bien sûr !) une politique de lutte contre les inégalités en matière d'accès à l'emploi. Un exemple typique de ce type d'utilisation est fourni par la politique sociale de la CEE qui fait dépendre explicitement l'éligibilité des régions communautaires aux programmes d'aide économique (Fonds Social Européen notamment) de leur taux de chômage harmonisé.

## II - LES TAUX DE CHOMAGE CONJONCTURELS CALCULES A L'INSEE

Pour des raisons de coût, de disponibilité des données et d'intérêt conjoncturel, l'INSEE calcule trois types de taux de chômage à rythme infra-annuel :

- des taux de chômage nationaux mensuels par sexe et âge (3 tranches) ;
- des taux de chômage régionaux et départementaux trimestriels soit par rapport à la population active totale, soit par rapport à la population active salariée. Cette dernière option, moins pure conceptuellement, peut en effet se révéler plus pertinente, en particulier pour les départements très agricoles. Le degré d'exposition au risque de chômage des agriculteurs est en effet faible : calculer le taux de chômage par rapport à la population active globale conduit alors à minimiser les tensions réelles du marché du travail puisque celui-ci est en réalité beaucoup plus étroit que ne le suggère la seule donnée représentée par la population active globale. Apparaissent ainsi les limites permanentes de tout indicateur synthétique, auxquelles n'échappent pas les taux de chômage, dès lors qu'on l'applique à des populations hétérogènes.

- des taux de chômage infra-régionaux ou infra-départementaux par bassin ou zone d'emploi. Contrairement aux deux types de taux précédents établis par la Division Emploi de la Direction générale de l'INSEE, l'initiative de ces calculs portant sur des aires géographiques plus limitées et ne coïncidant souvent pas avec les découpages administratifs est laissée aux Services d'Etudes des Directions Régionales de l'INSEE, moyennant une cohérence entre les calculs effectués aux différents niveaux géographiques. Effectuées en données brutes, les estimations des divers taux de chômage sont ensuite désaisonnalisées pour jouer pleinement leur rôle d'indicateur conjoncturel.

Dans la suite de cette note on s'intéressera principalement aux taux de chômage régionaux et départementaux calculés par rapport à la population active totale. Leur mode d'estimation conduira, comme on le verra, à étudier au préalable l'évaluation des taux de chômage nationaux. La plus grande simplicité des calculs nationaux et la meilleure maîtrise des sources nationales (au moins vu de Paris) renforceront d'ailleurs la place donnée à l'analyse des estimations effectuées au niveau France entière.

### III - LES CONCEPTS RETENUS ET LE PRINCIPE DE L'ESTIMATION

Conformément à la pratique, ou au moins à l'objectif, des organismes internationaux ou des principaux pays développés, le choix a été de se rapprocher le plus possible de la définition internationale du taux de chômage préconisée par le Bureau International du Travail (BIT). Or il n'existe pas de source statistique exhaustive et qui plus est conjoncturelle, qui permette d'observer sans biais les taux de chômage définis dans la partie précédente. Il faut alors s'appuyer sur les différentes sources qui fournissent une information utile pour l'estimation de l'emploi et du chômage intervenant dans (1). Les sources sont fondamentalement de deux types : des enquêtes par sondage (SE) et des sources d'origine administrative exhaustives (SA). Toutes sont sujettes à des erreurs de mesure, les premières étant de plus affectées par des erreurs aléatoires.

On fait l'hypothèse que l'emploi (E) et le chômage (C) sont fonction de paramètres observables dans les diverses sources, soit :

$$C = FC (C_1^e, C_2^e, \dots, C_{ne}^e, C_1^a, C_2^a, \dots, C_{na}^a) \quad (2)$$

où les  $C_i^e$  sont observés dans les  $ne$  enquêtes et les  $C_i^a$  dans les  $na$  sources administratives,  $FC$  étant une fonction régulière.

$$\text{On notera pour simplifier : } C = FC \quad (C_1^e, ne; C_1^a, na) \quad (3)$$

$$\text{De même : } E = FE \quad (E_1^e, me; E_1^a, ma) \quad (4)$$

Alors l'estimation  $\hat{T}$  du taux de chômage sera obtenue par :

$$\hat{T} = \frac{\hat{C}}{\hat{C} + \hat{E}} = \frac{FC (\hat{C}_1^e, ne; C_1^a, na)}{FC (\hat{C}_1^e, ne; C_1^a, na) + FE (\hat{E}_1^e, me; E_1^a, ma)} \quad (5)$$

Dans (5) les sources utilisées pour les estimations des paramètres de l'emploi et du chômage peuvent être identiques ou non.

#### IV - L'ESTIMATION DU CHOMAGE ( $\hat{C}$ )

##### 1 - Les sources d'information

\* L'enquête emploi réalisée chaque année par l'INSEE au mois de mars auprès de 65 000 ménages est la seule source qui permet une observation directe du chômage selon les normes du BIT. Effectuée selon un plan de sondage aréolaire régional au 1/300e avec un renouvellement par 1/3 de l'échantillon elle est sujette à des erreurs aléatoires qui s'amplifient lorsqu'on passe du niveau national au niveau local, puisque le taux de sondage est actuellement uniforme. L'intervalle de confiance à 95 % associé à l'estimation du chômage national est de  $\pm 76\ 000$ , soit une précision de 3,3 % compte tenu d'un niveau égal à 2 308 000 en mars 1989<sup>1</sup>. Pour une région d'environ 2 millions d'habitants la précision n'est par contre que d'environ 17 %. Si on s'intéresse aux variations d'une année à l'autre l'aléa est très voisin de celui qui affecte les niveaux.

Portant sur les ménages ordinaires ou les personnes qui ont pu être réintégrées dans ces ménages (étudiants et militaires du contingent par exemple), le champ de l'enquête n'est pas tout à fait exhaustif. Le recensement de la population de 1982 (RP 82) permet d'estimer à 33 000 personnes le nombre des chômeurs ainsi omis à cette date.

Enfin l'enquête emploi ne fournit qu'une observation annuelle, disponible de plus avec un délai relativement long (4 mois pour la série

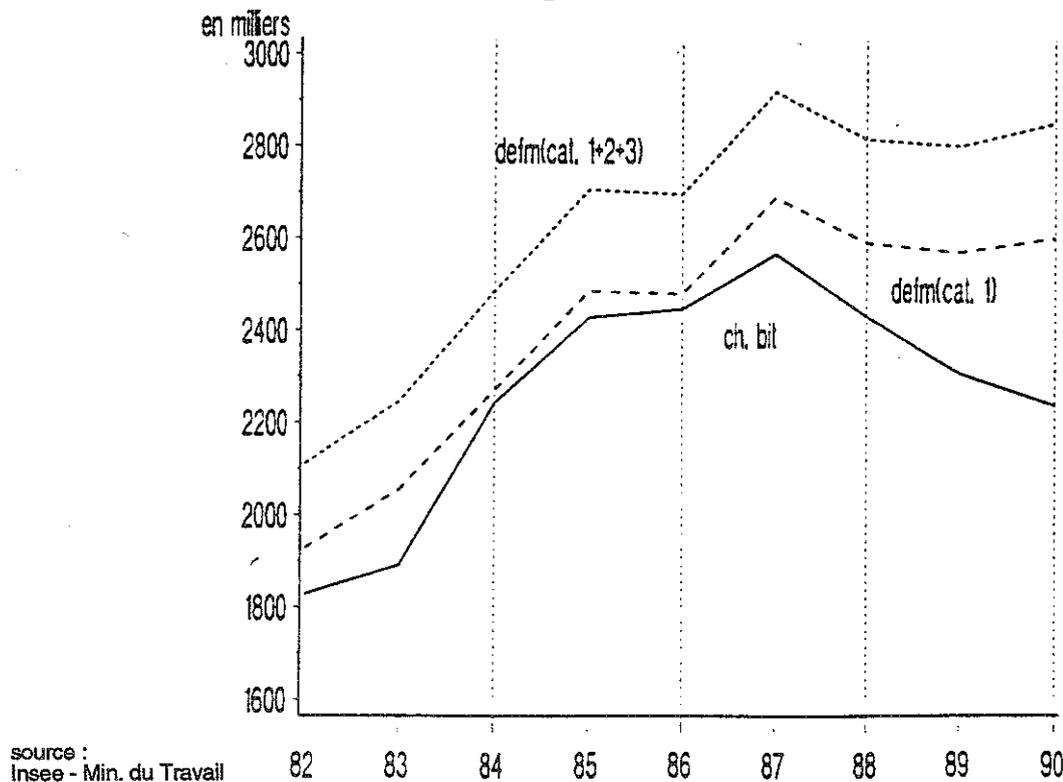
<sup>1</sup> On dira par la suite indifféremment que le niveau de chômage est de 2 308 000  $\pm 76\ 000$ , ou qu'il est connu à 3,3 % près, ou enfin que sa précision est de 76 000 ou de 3,3 %.

d'enquêtes 1982-1989) pour les objectifs conjoncturels, même s'ils sont courts pour une enquête de cette importance. Une source annexe doit donc être utilisée pour actualiser mensuellement les données tirées de l'enquête emploi. Cette source est la statistique administrative des demandeurs d'emploi en fin de mois (DEFM) inscrits à l'ANPE.

- \* La statistique des DEFM est exhaustive, mensuelle et exploitée à des niveaux géographiques très fins (département, zone d'emploi et même, à rythme trimestriel, commune de résidence). Les données sont traitées très rapidement (moins d'un mois). Même si elle est irremplaçable aujourd'hui en tant qu'indicateur conjoncturel, cette source présente des inconvénients sérieux.

Le premier tient à la différence de concept entre chômage au sens du BIT et DEFM. Celle-ci se traduit par des divergences en niveau et en évolution des deux grandeurs. Ainsi, en janvier 1990<sup>2</sup>, le nombre des chômeurs s'élevait à 2 237 000, celui des DEFM à 2 850 000. Au cours des quatre dernières années le nombre des chômeurs a évolué systématiquement plus favorablement que celui des DEFM, l'écart cumulé (sur les variations) atteignant 370 000 personnes (graphique 1). Le repérage, dans l'enquête emploi, des inscrits à l'ANPE (selon leur

**Graphique 1 : Chômage au sens du BIT et demandes d'emploi de 1982 à 1990 (à la date de l'enquête emploi)**



<sup>2</sup> La dernière enquête emploi a été réalisée en janvier en raison de la "concurrence" du recensement de la population effectué en mars.

propre déclaration) permet d'étudier le recouvrement entre chômage et demande d'emploi (tableau 1). D'un côté 15 % des chômeurs ne se disent pas inscrits à l'ANPE en janvier 1990, proportion qui varie peu depuis quelques années ; de l'autre 27 % des inscrits ne satisfont pas aux critères du BIT pour être chômeurs, soit qu'ils aient un travail (11 %), soit qu'ils ne soient pas immédiatement disponibles (5 %) soit enfin qu'ils ne recherchent pas ou plus d'emploi (11 %). La forte augmentation de cette dernière catégorie depuis 1986 explique comptablement les 2/3 de la divergence d'évolution entre chômage au sens du BIT et DEFM. Le 1/3 restant correspond à une différence d'appréciation sur l'évolution du chômage "enregistré", entre d'une part les inscrits à l'ANPE selon l'enquête et d'autre part les DEFM présents dans les fichiers de l'ANPE.

Les DEFM présentent également les inconvénients de toute statistique d'origine administrative, à savoir d'être tributaires de la réglementation en vigueur (conditions d'inscription à l'ANPE, de renouvellement de la demande d'emploi, d'indemnisation, avantages indirects liés à l'inscription) et de la qualité de gestion des fichiers dont elles sont un sous-produit. Le caractère politiquement très sensible de cet indicateur et la rapidité des traitements accentuent bien sûr ce handicap : l'évolution de la série est ainsi biaisée à certaines époques et elle est en permanence entâchée d'un bruit (erreur de mesure).

Une réforme du traitement des statistiques de DEFM intervenue en 1986 à la suite du rapport de M. Malinvaud a ainsi conduit à une révision à la hausse de la série sur la période 1982-1986, faible en début de période mais d'environ + 70 000 en fin de période. En retardant de quelques jours l'exploitation des données, elle a également permis de réduire significativement le bruit. La mise en place progressive d'un nouveau système de gestion informatique dans les unités de l'ANPE à partir de 1989 amène à nouveau des perturbations depuis quelques mois, dont l'importance est souvent difficile à chiffrer.

Enfin une particularité du mode d'élaboration des DEFM doit être signalée compte tenu des ruptures de série qu'elle entraîne aux niveaux infra-nationaux. Il s'agit de la procédure d'inventaire des fichiers qui consiste à confronter périodiquement (une fois par an en général) et par roulement le fichier local des unités de l'ANPE avec le fichier statistique central. Cette confrontation a pour but d'apurer le fichier central des anomalies liées à son mode de mise à jour, qui s'effectue sur la base de mouvements mensuels issus

Tableau 1 - Evolution du recouvrement entre chômage au sens du BIT  
et demandes d'emploi de 1982 à 1989

en milliers ou

	Avril -mai 1982	Mars 1984	Mars 1986	Mars 1987	Mars 1988	Mars 1989	Janvier 1990
Chômeurs au sens du BIT	1 827,5	2 244,5	2 447,5	2 567,3	2 431,3	2 308,3	2 236,8
dont :							
non inscrits à l'ANPE	320,7	356,9	352,7	344,5	320,5	313,3	327,7
inscrits à l'ANPE (1)	1 807,9	2 250,1	2 587,2	2 754,3	2 667,2	2 601,8	2 616,6
dont :							
non chômeurs au sens du BIT	302,5	362,6	492,4	531,5	556,4	606,8	707,4
- actifs occupés	150,7	181,8	290,8	281,2	290,0	309,6	288,6
- inactifs	151,8	180,9	201,6	250,3	266,4	297,2	418,8
* non disponibles	113,0	113,2	131,1	39,7	122,6	128,6	122,5
* absence de recherche d'emploi	38,8	67,7	70,6	110,6	143,9	168,6	296,3
DEFM (1+2+3) (2)	2 107,9 (e)	2 487,0	2 696,1	2 918,0	2 815,9	2 801,9	2 849,8
Z chômeurs BIT non inscrits	17,6	15,9	14,4	13,4	13,2	13,6	14,7
Z inscrits à l'ANPE non chômeurs	16,7	16,1	19,0	19,3	20,9	23,3	27,0
- actifs occupés	8,3	8,1	11,2	10,2	10,9	11,9	11,0
- inactifs	8,4	8,1	7,8	9,1	10,0	11,4	16,0
* non disponibles	6,3	5,0	5,1	5,1	4,6	4,9	4,7
* absence de recherche d'emploi	2,1	3,0	2,7	4,0	5,4	6,5	11,3
Chômeurs BIT/DEFM (Z)	86,6	90,2	90,8	88,0	86,3	82,4	76,5
Inscrits ANPE/DEFM (Z)	85,6	90,5	96,0	94,4	94,7	92,9	91,8

1 - Inscrits à l'ANPE : selon l'enquête emploi

2 - DEFM (1+2+3) : selon le Ministère du Travail

des fichiers locaux. L'étalement sur l'année des inventaires a peu d'effet, sauf problème spécifique, sur les statistiques nationales, mis à part un biais systématique de 20 000 à 30 000 personnes<sup>3</sup>. Il en va différemment en région et a fortiori en département où les inventaires conduisent à des "coups d'accordéon" d'autant plus violents que la qualité de gestion des fichiers locaux est médiocre.

## 2 - La méthode d'estimation

### a) Au niveau national

On utilisera les notations suivantes :

$\hat{C}_{a,m}^s$  = estimation du chômage pour la combinaison sexe-âge repérée par  $s$  pour l'année  $a$  et le mois  $m$ .

$\hat{CE}_{a,m}^s$  = même chose pour estimation fondée sur les observations de l'enquête emploi de l'année  $a$ , faite au mois  $m$ .  $m$  est la date moyenne de l'enquête, en général le 15 mars, soit  $m = 2,5$

$DEFM_{a,m}^s$  = idem pour les DEFM de catégorie 1

$CHC^s$  = estimation du nombre des chômeurs hors champ de l'enquête emploi effectuée en mars 1982 à partir du recensement de la population (RP82)

$CVS(X)$  = série  $X$  désaisonnalisée.

Toute la procédure d'estimation est menée indépendamment pour chacune des 6 valeurs de  $s$ , l'estimation globale étant ensuite obtenue par somme des 6 estimations élémentaires.

On commence par ajouter aux chômeurs estimés dans l'enquête emploi les chômeurs hors champ de l'enquête dont on suppose le nombre constant depuis le RP 82.

On passe ensuite de la date moyenne de l'enquête à la fin mars à l'aide de l'évolution des DEFM sur la période correspondante, elle même évaluée par interpolation linéaire sur données mensuelles.

On aboutit à une estimation  $\hat{C}_{a,3}^s$  fin mars :

$$\hat{C}_{a,3}^s = (\hat{CE}_{a,m}^s + CHC^s) \frac{DEFM_{a,3}^s}{DEFM_{a,m}^s} \quad (6)$$

<sup>3</sup> La procédure d'inventaire conduit généralement à radier plus de demandes du fichier statistique qu'on n'y en réintroduit à cette occasion.

Les niveaux mensuels de chômage sont ensuite évalués en appliquant à l'estimation issue de (6) le taux d'évolution des DEFM entre mars et le mois considéré.

$$\hat{C}_{a,m}^s = \hat{C}_{a,3}^s \frac{\text{DEFM}_{a,m}^s}{\text{DEFM}_{a,3}^s} \quad (7)$$

Lorsque sont disponibles les résultats de l'enquête emploi effectuée l'année suivante ( $a + 1$ , au mois  $m$ ) on procède à une révision de la série entre les dates mars  $a$  et mars  $a + 1$ . A cet effet on estime  $\hat{C}_{a+1,3}^s$  par (6) qui se trouve être différent de l'estimation antérieure déduite de (7). On procède alors à un recalage linéaire de la série, en supposant l'erreur d'estimation représentée par l'écart entre (6) et (7) en mars  $a + 1$  fonction linéaire du temps entre mars  $a$  et mars  $a + 1$  :

$$\hat{C}_{a,m}^s = \left[ \frac{\hat{C}_{a+1,3}^s}{\hat{C}_{a,3}^s} / \frac{\text{DEFM}_{a+1,3}^s}{\text{DEFM}_{a,3}^s} \right]^{n/12} \cdot \hat{C}_{a,3}^s \cdot \frac{\text{DEFM}_{a,m}^s}{\text{DEFM}_{a,3}^s} \quad (8)$$

où

$$\begin{aligned} n &= m - 3 \quad \text{si} \quad m \in a \\ &= m + 9 \quad \text{si} \quad m \in a + 1 \end{aligned}$$

Enfin, les estimations postérieures à mars  $a + 1$  sont révisées en appliquant (7) à  $\hat{C}_{a+1,3}^s$

Le passage à la série globale et aux séries CVS s'effectue de manière additive :

$$\hat{C}_{a,m} = \sum_s \hat{C}_{a,m}^s \quad (9)$$

$$\text{CVS} (\hat{C}_{a,m}^s) = x_{a,m}^s \hat{C}_{a,m}^s \quad (10.1)$$

$x_{a,m}^s$  est le coefficient saisonnier obtenu par une désaisonnalisation de type multiplicatif (variante du programme X11 du U.S. Bureau of the Census). La désaisonnalisation est en réalité effectuée sur les  $DEFM^s$  et non sur les  $\hat{C}^s$  mais ceci n'entraîne qu'une erreur très minime compte tenu de l'absence de saisonnalité résiduelle entre  $\hat{C}^s$  et  $DEFM^s$ , elle même liée à la méthode de recalage conduisant à (8).

$$\text{Enfin CVS}(\hat{C}_{a,m}) = \sum_s \text{CVS}(\hat{C}_{a,m}^s) \quad (10.2)$$

b) Aux niveaux infra-nationaux

Disposant d'une estimation mensuelle de chômage par sexe et âge on souhaite obtenir des estimations régionales et départementales<sup>4</sup> globales et trimestrielles. A cet effet on répartit les niveaux nationaux  $\hat{C}_{a,m}^s$  selon la structure géographique des DEFM par sexe et âge, puis on calcule l'estimation globale pour chaque région ou département par agrégation :

$$\hat{C}_{a,m}^D = \frac{\sum_s \hat{C}_{a,m}^s \text{DEFM}_{a,m}^{D,s}}{\text{DEFM}_{a,m}^s} \quad (11)$$

La stratification par sexe et âge a pour objectif de tenir compte de la différence de structure selon ces critères qui existe entre chômage BIT et DEFM : les chômeurs sont, en moyenne, un peu plus jeunes et plus souvent des femmes que les demandeurs d'emploi.

La combinaison de (7) et (9) conduit à exprimer simplement  $\hat{C}_{a,m}^D$  en fonction de sa valeur au mois  $m'$  antérieur à  $m$  :

$$\hat{C}_{a,m}^D = \frac{\sum_s \hat{C}_{a,m'}^{D,s} \text{DEFM}_{a,m}^{D,s}}{\text{DEFM}_{a,m'}^{D,s}} \quad (12)$$

<sup>4</sup> Le traitement des régions et des départements est ici tout à fait identique (sauf pour le calcul des CVS). On présente la méthode utilisée pour un département D sachant qu'elle est la même pour une région.

(12) correspond au mode de calcul opératoire du chômage au niveau infra-national de trimestre en trimestre, en année courante. L'avantage de (12) est de ne pas nécessiter pour l'estimation d'un niveau géographique donné des informations de niveau géographique supérieur.

La désaisonnalisation des séries s'effectue en 2 étapes, pour assurer la cohérence entre les corrections réalisées aux différents niveaux géographiques.

1<sup>ère</sup> étape : chaque série  $DEFM_{a,m}^D$  est désaisonnalisée (programme X11 multiplicatif) soit  $96 + 22 = 118$  désaisonnalisations.

2<sup>ème</sup> étape : les séries CVS ( $\hat{C}_{a,m}^D$ ) calculées en appliquant les coefficients saisonniers obtenus aux séries brutes  $\hat{C}_{a,m}^D$  sont recalées sur celles de niveau géographique immédiatement supérieur (régions sur France entière, départements sur région).

## V - L'ESTIMATION DE L'EMPLOI ( $\hat{E}$ )

Contrairement au chômage où le processus d'estimation va du national au départemental, l'estimation de l'emploi régional ou départemental obéit à un processus décentralisé qui privilégie le niveau régional.

- \* Les sources de base sur l'emploi sont les recensements de la population. Ceux-ci ont l'avantage d'être quasiment exhaustifs (les données sur l'emploi sont issues d'un sondage au 1/4) et de couvrir la totalité (à peu de chose près) de la population active.

L'inconvénient des recensements est bien sûr leur faible périodicité ainsi que la rusticité des concepts utilisés pour mesurer l'emploi, le chômage et l'inactivité. En fait le questionnement vis à vis de l'activité professionnelle suivi dans le recensement s'apparente aux déclarations spontanées effectuées par les personnes interrogées à l'enquête emploi, sans pouvoir être affiné ou corrigé par une interrogation plus fouillée comme c'est le cas dans l'enquête.

Attention : (12) n'est plus valable sur les séries révisées par (8). L'utilisation de (12) n'est en fait possible que lorsque les méthodes d'actualisation du chômage et de sa répartition spatiale reposent sur le même principe, ce qui est seulement le cas en année courante.

Compte tenu de l'objectif visé ici, l'estimation d'un taux de chômage, ces problèmes sont toutefois mineurs. Les différences entre estimations nationales de l'emploi selon le RP82 et l'enquête emploi de 1982 sont faibles (0,7 %) et leur impact sur le calcul du taux de chômage national est inférieur à 0,1 point.

- \* Les données sur l'emploi tirées du dernier recensement ont été rétopolées au 31 décembre de l'année précédente à l'aide de sources plus conjoncturelles. Cette base (au 31.12.81 pour la base actuelle) a été constituée par sexe, statut, secteur d'activité (NAP 40) et région. Proposée par la Division Emploi, elle a été négociée avec les Services d'Etudes Régionaux de l'INSEE qui pouvaient disposer d'informations propres à affiner les estimations (utilisation d'autres sources, connaissance de la qualité de traitement du recensement).

Une base d'emploi par département, cohérente avec la précédente, a également été constituée selon le même principe et sous la coordination de la Division Statistiques et Etudes Régionales.

- \* Les estimations annuelles d'emploi sont alors effectuées par application aux niveaux de la base d'indices chaînes annuels calculés dans le découpage précédent. Pour chaque strate du découpage ces indices sont établis sur la somme des effectifs estimés par les différentes sources qui couvrent la strate. Les sources d'actualisation annuelle sont pour la plupart d'origine administrative. Plus précisément il s'agit de sources émanant d'organismes de Sécurité Sociale (UNEDIC ou URSSAF selon les régions pour les salariés, ACOSS, CANAM, CANCAVA, MSA,... pour les non salariés). Pour les secteurs non ou mal couverts, elles sont complétées par des enquêtes statistiques (enquête sur le personnel de la Fonction Publique ou des hôpitaux publics, enquête emploi...) et des données obtenues auprès des grandes entreprises nationales (SNCF, Air France, EDF-GDF,...) ou d'établissements publics.

Les indices d'évolution annuelle de l'emploi font l'objet de négociation entre la Direction Générale de l'INSEE et les Directions Régionales, comme pour la constitution de la base initiale. L'emploi est pris au lieu de travail et non au lieu de résidence pour être cohérent avec la règle adoptée par les principales sources d'actualisation.

Remarque : La confrontation entre les estimations fondées sur la base antérieure au 31/12/74 et la nouvelle base au 31/12/81 a montré qu'il était préférable de calculer des indices à champ variable, et non à champ constant, pour limiter les risques de dérapage à moyen terme. Seule la dernière année est établie avec des indices à champ constant, de façon à faciliter les analyses conjoncturelles.

Avec les notations :

$E_{81,12}^{D.RP}$	=	Base départementale d'emploi au lieu de travail déduite du RP82
$\hat{E}_{a,12}^{DIt}$	=	Estimation départementale de l'emploi total au 31.12/a au lieu de travail (on notera FE pour le niveau France entière)
$\hat{E}_{a,12}^{Dir}$	=	Idem au lieu de résidence
$E_{adm.}^{a,12}$	=	Emploi dans les sources d'origine administrative
$\hat{E}_{enq}^{a,12}$	=	Estimation de l'emploi dans les sources de type enquête
w	=	Strate correspondant au découpage par sexe, statut et secteur d'activité
$\hat{I}_{a,12}^D(w)$	=	Estimation de l'indice d'évolution de l'emploi pour la strate w, le département D et l'année a.

Alors :

$$\hat{I}_{a,12}^D(w) = \frac{(E_{adm.}^{a,12}(w) + \hat{E}_{enq}^{a,12}(w))}{(E_{adm.}^{a-1,12}(w) + \hat{E}_{enq}^{a-1,12}(w))} \quad (13)$$

la sommation porte sur les sources utilisées pour le calcul de l'indice en ce qui concerne la strate w.

$$E_{a,12}^{DIt}(w) = E_{81,12}^{D.RP}(w) \prod_{x=82}^a \hat{I}_{x,12}^D(w) \quad (14)$$

$$\hat{E}_{a,12}^{DIt} = \sum_w \hat{E}_{a,12}^{DIt}(w) \quad (15)$$

Conformément aux sources utilisées pour estimer le chômage et parce qu'on s'intéresse à la population résidente, le repérage des actifs occupés doit se faire au lieu de résidence et non au lieu de travail. Il faut donc passer d'estimations de l'emploi au lieu de travail issues de (15) à des estimations de l'emploi au lieu de résidence.

L'emploi au lieu de résidence n'est observé que dans les recensements. On calcule donc des coefficients de passage entre emploi au lieu de travail et emploi au lieu de résidence par département  $N^D$  et on suppose ces coefficients constants depuis le dernier recensement :

$$N_{a,12}^D = N_{82,3}^D \quad (16)$$

L'hypothèse de constance temporelle peut être justifiée par la relative stabilité des coefficients observée entre les deux derniers recensements. Sauf pour une ou deux régions, la variation des coefficients entre le RP 75 et le RP 82 est en effet inférieure à 1 %.

On crée d'autre part une unité géographique fictive, la région "hors territoire", qui rassemble les résidents ayant une activité professionnelle à l'étranger : il s'agit essentiellement de travailleurs frontaliers en RFA et Suisse qui résident en Alsace et Rhône-Alpes. L'estimation des effectifs de cette unité  $E_{a,12}^{HT}$  est déduite de l'observation du RP 82 actualisée principalement par des données communiquées par les Directions Régionales de l'INSEE concernées par le travail frontalier :

$$E_{a,12}^{HT} = E_{81,12}^{HT} \prod_{x=82}^a I_{x,12}^{HT} \quad (17)$$

L'estimation de l'emploi départemental au lieu de résidence s'effectue par application des  $N^D$  aux  $\hat{E}^{Dit}$  puis recalage sur l'emploi résident France entière :

$$\hat{E}_{a,12}^{Dir} = \frac{\hat{E}_{a,12}^{Dit} N_{a,12}^D \left( \sum_D \hat{E}_{a,12}^{Dit} + E_{a,12}^{HT} \right)}{\sum_D \hat{E}_{a,12}^{Dit} N_{a,12}^D} \quad (18)$$

## Remarques :

1. L'emploi résident France entière est bien égal à la somme de l'emploi des résidents évaluée au lieu de travail et des effectifs de la région hors territoire (expression entre parenthèses dans (18)).
2. La procédure d'estimation de l'emploi est d'abord menée au niveau régional avant de l'être au niveau départemental. Aux différentes étapes les estimations départementales sont donc recalées sur les estimations régionales. De fait, les formules (13) à (18) décrivent le processus d'estimation régional. Le processus de décontraction départemental de l'emploi régional suit une même logique mais s'appuie sur des données moins riches qu'au niveau régional.
3. Les effectifs du contingent sont considérés ici comme faisant partie des actifs occupés. Leur méthode d'estimation est spécifique mais n'a que peu d'effet sur l'estimation du taux de chômage vu la faiblesse des effectifs concernés.

La procédure décrite par (13) à (18) est menée chaque automne  $n$ . Elle conduit à des estimations semi-définitives (SD) au 31/12/ $n-1$  qui seront révisées un an plus tard lorsque l'ensemble des sources seront disponibles (estimations définitives (D)).

Les estimations infra-annuelles  $\hat{E}_{a,m}$  sont calculées par interpolation linéaire entre  $\hat{E}_{a-1,12}$  et  $\hat{E}_{a,12}$ . Cette approximation a l'avantage de la simplicité, sachant que le choix du mode de trimestrialisation de l'emploi n'a que peu de conséquence sur l'estimation du taux de chômage, surtout en données CVS.

Plus importante est la question de l'estimation de l'emploi en année courante. Si  $a'$  est la dernière année pour laquelle on a pu utiliser la procédure (13) à (18), il s'agit d'estimer  $\hat{E}_{a,m}^{Dlt}$  pour  $(a,m) > (a',12)$ , c'est à dire pour une période qui s'étend sur 9 à 21 mois.

A cet effet on s'appuie sur des estimations nationales d'emploi fondées sur des sources nationales trimestrielles ou sur des prévisions d'emploi formulées dans les Notes de Conjoncture de l'INSEE. Seules les valeurs annuelles sont utilisées, soit  $\hat{E}_a^{FE} +1,12$  si on prend le cas  $a = a' + 1$ .

On actualise ensuite au 31/12/ $a$  les estimations régionales d'emploi issues de (18) à l'aide d'indices  $I_a^R$  "à dire d'expert" qui tiennent compte des estimations régionales trimestrielles, voire annuelles, d'emploi de l'UNEDIC (source qui couvre 13 000 000 de salariés sur un total d'environ 22 000 000 d'actifs occupés). Ces évaluations sont recalées sur l'estimation nationale.

Soit :

$$\hat{E}_{a,12}^{Rlr} = \frac{\hat{E}_{a,12}^{Rlr} I_a^R \hat{E}_{a,12}^{FE}}{\sum_R \hat{E}_{a,12}^{Rlr} I_a^R} \quad (19)$$

Les estimations régionales infra-annuelles sont calculées par interpolation linéaire<sup>5</sup>.

Les estimations départementales sont établies en conservant constante la structure départementale des effectifs régionaux au 31/12/a'.

$$\hat{E}_{a,m}^{Rlr} = \hat{E}_{a-1,12}^{Rlr} + \frac{m}{12} (\hat{E}_{a,12}^{Rlr} - \hat{E}_{a-1,12}^{Rlr}) \quad (20)$$

$$\hat{E}_{a,m}^{Dir} = \frac{\hat{E}_{a,12}^{Dir} \hat{E}_{a,m}^{Rlr}}{\hat{E}_{a,12}^{Rlr}} \quad (21)$$

## VI - REVISION DES SERIES

La fréquence de révision des séries obéit à un compromis entre qualité des estimations et coût des changements. Deux révisions annuelles sont ainsi effectuées, l'une en début d'année, l'autre à l'été.

En début d'année  $n$  on actualise la correction des variations saisonnières du chômage et on incorpore les estimations SD de l'année  $n - 2$  et D de l'année  $n - 3$  de l'emploi régional et départemental.

La révision de l'été tient compte des résultats de l'enquête emploi du début de l'année  $n$ . Des évaluations plus sûres des évolutions de l'emploi au cours des années  $n - 1$  et  $n$  sont également introduites.

<sup>5</sup> C'est pourquoi l'on doit disposer en année courante  $a$  d'une prévision d'emploi au 31/12/a.

## VII - QUALITE DES ESTIMATIONS

## 1 - Biais

## a) Biais sur l'estimation de l'emploi

La principale source antérieure de biais sur l'estimation  $\hat{E}$  a été corrigée. Elle venait de l'utilisation dans (14) d'indices chaînés  $\hat{I}$  calculés à champ constant d'une année à l'autre pour ce qui concerne la source d'actualisation principale (UNEDIC), méthode qui conduisait à une sous-estimation progressive des effectifs (120 000 sur 7 ans de 1974 à 1981 au niveau national).

Le concept utilisé pour la mesure de l'emploi (emploi au sens du recensement) est potentiellement source de biais. Mais, on l'a vu, la différence par rapport à la mesure fondée sur les critères du BIT est très mineure. Au total on peut donc considérer que l'estimation  $\hat{E}$  ne présente pas de biais particulier.

## b) Biais sur l'estimation du chômage

- au niveau national :

L'estimation  $\hat{C}$  est biaisée à court terme si l'évolution des DEFM n'est pas parallèle à celle du chômage au sens du BIT compte tenu de (7) et (8). L'évolution plus favorable du chômage au sens du BIT que des DEFM entre 1986 et 1990 a ainsi entraîné des révisions à la baisse de  $\hat{C}$  à l'été de chacune de ces années (90 000 en moyenne au niveau national), soit un biais de 0,33 point sur le taux de chômage.

Tenter de réduire le risque de dérive à court terme de l'estimation du chômage suppose de pouvoir expliquer la divergence d'évolution entre les deux indicateurs en fonction de paramètres connus. Il s'agit de substituer à la relation (7) la relation (22) :

$$\hat{C}_{a,m}^s = \hat{C}_{a,3}^s \cdot g(\text{DEFM}_{a,m}^s / \text{DEFM}_{a,3}^s, x_1, \dots, x_n) \quad (22)$$

où les paramètres  $x_1, \dots, x_n$  dépendent des dates (a,m) et (a,3).

Une voie naturelle pour le choix de  $g$  est d'estimer un modèle de régression linéaire sur données annuelles :

$$\frac{\hat{C}_{a+1,3}^s}{\hat{C}_{a,3}^s} = b^s \frac{DEFM_{a+1,3}^s}{DEFM_{a,3}} + \sum_{i=1}^{\infty} b_i x_{i,a}^s \quad (23)$$

Des tentatives d'estimation de modèles de ce type ont été réalisées sur la période 1975-1989 avec des variables explicatives  $x_i$  représentatives des tensions sur le marché du travail, de la flexibilité de l'emploi, de la politique de l'emploi ou enfin de certains facteurs institutionnels. Pour le moment les résultats obtenus ne sont pas très convaincants. La mise en évidence de facteurs économiques ou institutionnels qui expliqueraient les divergences d'évolution entre chômage BIT et DEFM est difficile. Les aléas de l'enquête emploi, les ruptures liées aux changements de série d'enquête d'une part, les perturbations de toute nature qui ont affecté la mesure des demandes d'emploi d'autre part obscurcissent sans doute les liaisons. Mais il faut reconnaître sur le fond que l'on a beaucoup de mal à donner une interprétation à la principale source de divergence identifiée pour les 4 dernières années, à savoir la montée du nombre des chômeurs "découragés" alors qu'on se situe dans une période de nette reprise des embauches.

Une autre voie pour estimer la fonction  $g$  de (22) pourrait être d'opérer une confrontation directe entre les deux mesures du chômage : étude du classement vis à vis des critères du BIT d'un échantillon de demandeurs d'emploi, repérage dans les fichiers de l'ANPE des personnes classées chômeurs ou s'étant déclarées inscrites à l'Agence dans l'enquête emploi. Une telle approche permettrait de s'affranchir des vicissitudes liées à chacune des deux sources. Réalisée ponctuellement, elle permettrait sans doute de mieux comprendre la liaison chômage BIT - DEFM mais peut-être moins son évolution temporelle. De ce fait cette 2ème voie pourrait être combinée avec la 1ère, en éclairant le choix des variables explicatives du modèle économétrique.

- aux niveaux infra-nationaux -

Il est difficile de se prononcer sur l'existence d'un biais "structurel"

lié à l'utilisation de (11) aux niveaux régionaux et départementaux. Ce biais existe si la répartition géographique du chômage diffère de celle des demandes d'emploi, test qui n'a jamais véritablement été effectué.

Il faudrait plus précisément tester l'égalité à 1 des ratios  $R_a^D$  :

$$R_a^D = \frac{\hat{C}_{a,3}^D}{\hat{C}_{a,3}} \cdot \frac{DEFM_{a,3}^D}{DEFM_{a,3}} \quad (24)$$

dans lesquels  $\hat{C}_{a,3}^D$  est déduit d'une observation directe de l'enquête emploi par une relation du type (6). Une confrontation opérée sur la seule année 1982 et au niveau régional entre les niveaux de chômage observés au RP, à l'enquête emploi et par les DEFM montre des divergences importantes. L'aléa sur les mesures tirées de l'enquête emploi est cependant tel lorsqu'on descend à des niveaux infra-nationaux (précision de 16 % pour une région moyenne, de 30 % pour une petite région ou un gros département) qu'il est difficile de conclure.

Une validation plus significative devrait en fait être menée sur la série homogène d'enquêtes emploi 1982-1989 (test  $R_a^D = 1$ ).

Au cas où, pour certaines unités géographiques, l'hypothèse  $R_a^D = 1$  devrait être rejetée, on pourrait alors chercher à formaliser la loi d'évolution temporelle de  $R_a^D$ .

$$\hat{R}_a^D = f^D(a) \quad (25)$$

l'estimation des niveaux de chômage  $\hat{C}_{a,3}^D$  pourrait alors être effectuée par :

$$\hat{C}_{a,3}^D = \hat{R}_a^D \cdot C_{a,3} \cdot \frac{DEFM_{a,3}^D}{DEFM_{a,3}} \quad (26)$$

Une telle estimation serait sans biais. Par contre elle risque d'être d'une précision médiocre si la liaison locale chômage BIT - DEFM est très instable, situation que l'on risque de rencontrer plus fréquemment pour des unités géographiques de taille réduite. On pourrait alors imposer la contrainte :

$$\hat{R}_a^D = R^D \quad (27)$$

Enfin l'estimation par la relation (26) doit être réservée au niveau régional : outre leur plus faible taille les départements présentent en effet le handicap de ne pas constituer un niveau de stratification de plan de sondage aréolaire de l'enquête emploi.

Une part de l'erreur d'estimation des niveaux de chômage infra-nationaux peut venir du type d'indicateur utilisé pour mesurer les demandes d'emploi. En effet les DEFM dont il a été fait état jusqu'à maintenant ne représentent que les demandes d'emploi à temps plein et à durée indéterminée, soit un peu moins de 90 % de l'ensemble des demandes d'emploi (DEFM'). Sachant que le concept DEFM' est formellement plus proche de celui du chômage au BIT que celui des DEFM, il serait légitime de l'utiliser dans les estimations infra-nationales de chômage. Pour que cette modification soit fondée il faudrait vérifier que la limitation des demandes d'emploi aux seuls DEFM entraîne effectivement une distorsion géographique de cet indicateur.

Etant donné le faible écart relatif entre DEFM et DEFM' et, au moins au niveau national et jusqu'en 1989 le parallélisme de leurs évolutions, il ne semble pas néanmoins que l'adoption de l'indicateur DEFM' soit de nature à améliorer sensiblement la qualité des estimations. Ceci mériterait toutefois d'être vérifié : on pourrait notamment étudier l'influence du choix de l'indicateur de demandes d'emploi sur les tests fondés sur la relation (22).

Enfin une méthode d'estimation sans biais plus directe du chômage infra-national consiste à reprendre directement l'évolution fournie par l'enquête emploi, moyennant quelques corrections mineures pour tenir compte des chômeurs hors champ de l'enquête et de sa date. Cette idée, qui est la plus naturelle, se heurte bien sûr à l'imprécision de telles estimations. Elle

sera toutefois étudiée dans le point suivant consacré justement à la précision des diverses estimations : une estimation directe du taux de chômage par évaluation simultanée du chômage et de l'emploi à partir du même échantillon peut en effet sous certaines conditions, se révéler performante.

## 2 - Précision

Dans le cas où  $\hat{C}$  et  $\hat{E}$  sont estimés à partir d'un même échantillon  $e$  d'enquête tiré selon une loi commune, la variance de l'estimateur  $\hat{T} = \hat{C}/(\hat{C} + \hat{E})$  peut être approximée en utilisant une procédure de linéarisation au voisinage des valeurs  $\hat{C}_e$  et  $\hat{E}_e$  estimées à partir de l'échantillon.

$$\text{Soit } Z = \begin{pmatrix} \frac{\partial T}{\partial C} \\ \frac{\partial T}{\partial E} \end{pmatrix}_e \hat{C} + \begin{pmatrix} \frac{\partial T}{\partial C} \\ \frac{\partial T}{\partial E} \end{pmatrix}_e \hat{E} \quad (28.1)$$

$$\text{alors} \quad \text{Var } \hat{T} \approx \text{Var } Z$$

Var  $\hat{T}$  se calcule alors à partir des valeurs  $Z_j$  prises sur l'échantillon  $e$  :

$$Z_j = \begin{pmatrix} \frac{\partial T}{\partial C} \\ \frac{\partial T}{\partial E} \end{pmatrix}_e C_j + \begin{pmatrix} \frac{\partial T}{\partial C} \\ \frac{\partial T}{\partial E} \end{pmatrix}_e E_j \quad (28.2)$$

où  $j = 1$  à  $n$ , taille de l'échantillon

$C_j = 1$  si l'individu est chômeur, 0 sinon

$E_j = 1$  si l'individu a un emploi, 0 sinon

$$\text{Soit : } Z_j = \frac{\hat{E}_e}{(\hat{C}_e + \hat{E}_e)^2} C_j - \frac{\hat{C}_e}{(\hat{C}_e + \hat{E}_e)^2} E_j$$

$$\text{ou : } Z_j = \hat{T}_e (1 - \hat{T}_e) \begin{pmatrix} C_j - E_j \\ \hat{C}_e \\ \hat{E}_e \end{pmatrix} \quad (29)$$

$$\text{Alors : } \text{Var} \begin{pmatrix} \hat{T} \\ \hat{T}_e \end{pmatrix} = (1 - \hat{T}_e)^2 \left[ \text{Var} \begin{pmatrix} \hat{C} \\ \hat{C}_e \end{pmatrix} + \text{Var} \begin{pmatrix} \hat{E} \\ \hat{E}_e \end{pmatrix} - 2 \text{Cov} \begin{pmatrix} \hat{C} & \hat{E} \\ \hat{C}_e & \hat{E}_e \end{pmatrix} \right] \quad (30)$$

Dans le cas où l'emploi  $\hat{E}$  n'est pas estimé à partir de l'échantillon e mais est obtenu à partir de sources administratives on peut appliquer (30) en considérant le terme  $\text{Var} \left( \frac{\sum_{i=1}^n \hat{e}_i}{n} \right)$  comme représentatif d'erreurs de mesure sur  $\hat{E}$  et en annulant le terme de covariance (indépendance des estimations de l'emploi et du chômage). De la même façon, en cas d'estimation "composite" de l'emploi ou du chômage on utilisera (30) en sommant les erreurs, supposées indépendantes, liées à chaque composante.

La relation (30) a donc une portée assez large. Selon cette spécification ce sont les précisions relatives (et non absolues) des estimations de l'emploi et du chômage qui conditionnent la précision de l'estimation du taux de chômage. Or, de manière générale, la précision relative des estimations de chômage est bien inférieure à celle des estimations d'emploi, que les erreurs soient de nature aléatoire ou liées à des problèmes de mesure. Dans le premier cas c'est la différence de taille des populations qui est en cause (rapport de 1 à 10 entre le chômage et l'emploi) ; dans le second c'est, par exemple, le flou sur la frontière séparant l'emploi du chômage qui génère des erreurs absolues identiques sur les deux composantes de la population active donc des erreurs relatives supérieures pour le chômage.

Ainsi c'est la qualité d'estimation du chômage qui est décisive pour la précision du taux de chômage, sauf dans un cas : celui où l'emploi est observé à partir du même échantillon que le chômage et où la corrélation entre les deux variables est forte (le terme en covariance de (30) est alors non négligeable).

a) Niveau national

- *Méthode actuelle*

Le tableau 2 fournit un ordre de grandeur approximatif des erreurs relatives sur  $\hat{C}$ ,  $\hat{E}$  et  $\hat{T}$  au niveau national, décomposées selon leur origine. Toutes les erreurs non aléatoires (erreurs de mesure) sont évaluées à "dire d'expert" : les chiffres indiqués sont donc nécessairement fragiles. Les erreurs sont évaluées en 1990 ce qui correspond à un cas de figure défavorable pour l'estimation  $\hat{E}$  (éloignement par rapport au dernier recensement). L'indépendance entre les diverses sources d'erreur est supposée. Cette commodité de calcul peut toutefois être discutée.

La précision du taux de chômage national est estimée à 3,6 % soit 0,33 point pour un taux égal à 9 %. Comme on s'y attendait, l'erreur vient essentiellement des étapes chômage (qui expliquent 85 % de la variance) et plus particulièrement de l'aléa lié à l'observation du chômage dans l'enquête emploi.

- Comparaison avec un estimateur où l'emploi est observé dans l'enquête-emploi

On peut comparer la précision de l'estimation "composite" actuelle à celle d'une estimation qui serait fondée sur une observation simultanée de l'emploi et du chômage dans l'enquête emploi, ce qui revient à modifier les étapes (13) et (14).

Tableau 2 - Précision de l'estimation nationale du taux de chômage - impact des différentes sources d'erreur

Indicateur	Etapes de l'estimation, relations ( )	Impact sur la précision relative de l'indicateur (en %)
$\hat{C}$	chômage observé à l'enquête emploi (EE) (6)	3,3
$\hat{C}$	addition des chômeurs hors champ de l'EE (6)	1
$\hat{C}$	actualisation par les DEFM (7)	1
$\hat{C}$	désaisonnalisation (10)	1
$\hat{E}$	constitution de la base d'emploi déduite du RP (14)	1
$\hat{E}$	actualisation annuelle de l'emploi par sources sectorielles et régionales (13)	1
$\hat{E}$	actualisation en année courante	0,5
$\hat{T}$	ensemble des étapes : - dont étapes chômage - dont étapes emploi	3,6 3,1 1,4

La relation (30) conduit alors à une précision de 3,8 %, qui n'est pas meilleure que celle de la méthode actuelle. Ce résultat paraît surprenant car on

aurait pu penser qu'une observation simultanée dans des concepts homogènes de l'emploi et du chômage améliorerait la précision de l'estimateur du taux de chômage.

La différence entre les 2 méthodes porte sur les termes d'erreurs de (30) liés aux estimations  $\hat{E}$  dans les étapes (13) et (14) (actualisation du RP d'un côté, observation par l'enquête emploi de l'autre).

Si l'on note respectivement  $\beta_E$ ,  $\beta_C$  les coefficients de variation des erreurs sur l'emploi (du type précédent), sur le chômage et  $\rho$  le coefficient de corrélation estimé de  $\hat{E}$  et  $\hat{C}$  dans l'enquête emploi on obtient les résultats suivants (tableau 3).

Tableau 3 : Impact de la méthode d'estimation de l'emploi sur la précision de l'estimation du taux de chômage national <sup>6</sup>

	$\rho$	$\beta_E^2$	$-2\rho\beta_E\beta_C$	$\beta_E^2 - 2\rho\beta_E\beta_C$
Méthode actuelle	0	5.10-5	0	5.10 <sup>-5</sup>
Méthode "enquête emploi"	- 0,51	1,1.10-5	5,7. 10-5	6,8.10-5

<sup>6</sup> : On se limite aux seules étapes distinctes des deux méthodes. La colonne de droite représente la partie correspondante de l'expression entre crochets dans la relation (30).

Les chiffres du tableau 3 montrent que, malgré sa précision supérieure de l'estimation de l'emploi, l'estimation du taux de chômage fondée sur une observation directe de l'emploi dans l'enquête emploi est moins bonne en raison de la corrélation négative des estimations  $\hat{E}$  et  $\hat{C}$  déduites de l'enquête. Celle-ci conduit en effet à un terme de covariance important qui contribue à augmenter nettement la variance totale.

L'origine de la covariance négative entre  $\hat{E}$  et  $\hat{C}$  vient surtout de ce que ces estimations sont établies après post-stratification sur des évaluations exogènes de population totale par sexe et âge quinquennal. Or la somme  $\hat{E} + \hat{C}$  qui représente la population active totale est fortement corrélée avec l'estimation de la population totale (avant post-stratification). Deux autres effets contradictoires jouent également. D'une part l'incompatibilité

d'appartenance, pour un individu donné, aux deux sous-populations des chômeurs et des actifs occupés qui conduit à une corrélation empirique négative (égale à - 0,21) dans le cas d'un tirage aléatoire simple de l'échantillon ; d'autre part, l'effet de grappe lié au sondage aréolaire pratiqué dans l'enquête emploi qui augmente au contraire fortement la corrélation de  $\hat{E}$  et  $\hat{C}$  (elle passe à + 0,31, avant la post-stratification).

Remarque : la corrélation positive entre  $\hat{E}$  et  $\hat{C}$  est, en soi, un facteur favorable pour la précision de l'estimation  $\hat{T}$  par (30). Mais ce gain est inférieur à la détérioration observée sur les estimations de  $\hat{C}$  et  $\hat{E}$ . On a donc bien intérêt à utiliser la post-stratification pour l'estimation de  $\hat{T}$  type "enquête emploi". Le gain total dû à la post-stratification est toutefois mineur (moins de 10 % sur la précision de  $\hat{T}$ ).

#### b) Niveaux infra-nationaux

Un calcul de même nature peut être effectué aux niveaux infra-nationaux, avec cependant beaucoup plus d'incertitude sur le chiffrage de chaque type d'erreur. Les résultats sont rassemblés dans le tableau 3.

Tableau 3 - Précision des estimations infra-nationales de taux de chômage - impact des différentes sources d'erreur

Indicateur	Etapes de l'estimation, relation ( )	Impact sur la précision relative de l'indicateur (en %)
$\hat{C}$	Chômage national observé à l'enquête emploi (EE) (0)	3,3
$\hat{C}$	Addition des chômeurs hors champ de l'EE (6)	1
$\hat{C}$	Actualisation nationale par les DEFM (7)	1
$\hat{C}$	Répartition spatiale par les DEFM (11)	2 à 5
$\hat{C}$	Désaisonnalisation infra-nationale	1 à 2
$\hat{C}$	Calage de la CVS sur le niveau national	1
$\hat{E}$	Constitution de la base d'emploi déduite du RP (14)	1 à 3
$\hat{E}$	Actualisation annuelle de l'emploi par sources sectorielles et régionales (13)	1 à 3
$\hat{E}$	Actualisation en année courante	1 à 2
$\hat{E}$	Passage emploi au lieu de travail/emploi au lieu de résidence (16)	1 à 3
$\hat{T}$	ensemble des étapes : - dont étapes chômage - dont étapes emploi	4,1 à 7,7 3,7 à 5,9 1,8 à 5,0

Comme prévu, la précision de l'estimation du taux de chômage diminue lorsqu'on descend au niveau de la région ou du département. La perte n'est cependant pas considérable sauf si les DEFM ou les dispositifs locaux d'estimation annuelle d'emploi sont très médiocres.

Remarque : Les évaluations de la précision des  $\hat{T}$  portent sur des niveaux. Les résultats seraient sans doute meilleurs sur les variations  $\Delta \hat{T}$  du taux de chômage d'une année à l'autre : les erreurs provenant de la répartition spatiale du chômage et l'ensemble des erreurs affectant l'emploi seraient nettement inférieures alors que les erreurs émanant d'autres étapes du calcul seraient du même ordre. La grande stabilité temporelle du classement des régions suivant leur taux de chômage confirme d'ailleurs cette bonne précision de  $\Delta \hat{T}$ . Les estimations seraient à l'inverse beaucoup moins précises si l'on s'appuyait sur des valeurs régionales  $\hat{C}$  ou  $\hat{E}$  tirées directement de l'enquête emploi (forte instabilité du classement selon le taux de chômage).