

Correction des ruptures de série suite au passage en 2003 de l'enquête Emploi annuelle à l'enquête Emploi en continu

- janvier 2008 -

Le suivi de séries longues à partir de l'enquête Emploi nécessite quelques précautions. Les évolutions de l'enquête au cours du temps peuvent avoir un impact sur les grandeurs mesurées, qu'il importe de corriger pour pouvoir interpréter des séries sur longues périodes.

Ces évolutions peuvent être de deux natures :

- Il peut s'agir de modifications du questionnaire d'enquête et/ou dans les modalités d'interrogation et de collecte (comme par exemple, le passage à un questionnaire informatisé ou encore l'introduction d'outils de codage automatique). Le principal changement a eu lieu lors du passage de l'enquête Emploi annuelle (EEA) à l'enquête Emploi en continu (EEC) en 2003¹. Auparavant, ces modifications étaient regroupées sur les années de recensement. Ainsi, les sous-périodes entre deux recensements (1968-1974 ; 1975-1981 ; 1982-1989 ; 1990-2002 ; 2003-aujourd'hui) donnent des estimations homogènes.
- Le passage à l'enquête en continu a aussi conduit à changer la période de référence de l'enquête : les données de l'enquête Emploi en continu conduite depuis 2003 portent sur la totalité de l'année, alors que l'enquête Emploi annuelle portait sur la situation en mars de l'année (et en janvier les années de recensement). Cette différence est importante, les variations saisonnières de l'emploi, du chômage et de l'activité étant de l'ordre de plusieurs centaines de milliers de personnes².

Cette note méthodologique détaille les corrections effectuées sur les données des enquêtes Emploi, suite au passage en 2003 de l'enquête Emploi annuelle à l'enquête Emploi en continu, pour construire les séries de l'*Insee Résultats* « Marché du travail - séries longues ». Ces corrections sont effectuées de manière à rendre les résultats des enquêtes Emploi annuelles homogènes à ceux de l'enquête Emploi en continu : **les séries longues qui s'en déduisent sont construites pour être représentatives de moyennes annuelles portant sur la population des personnes de 15 ans et plus en France métropolitaine résidant en ménages ordinaires** (population des ménages), c'est-à-dire *hors collectivités* (foyers, internats, hôpitaux, etc.). L'enquête Emploi n'est pas menée dans les collectivités.

Rupture sur la définition des variables pertinentes

En 2003, la principale rupture de définition concerne le statut d'activité au sens du BIT. A partir des données de la validité 2003, l'Insee utilise en effet la même interprétation qu'Eurostat des critères du BIT définissant le chômage, alors que l'interprétation française était légèrement différente auparavant (interprétation retenue dans les données de l'enquête

¹ Pour une documentation complète sur le passage à l'enquête Emploi en continu : « la nouvelle enquête Emploi, l'activité et le chômage », dossier de la revue *Economie et statistique*, n°362, 2003

² cf. « Les quatre saisons de l'emploi, une partition pour étudiants », Elise Coudin et Magali Befly, *Insee Première*, n° 1119, janvier 2007

annuelle)³. Ce changement est non négligeable. Il induit une rupture de niveau de l'ordre de 0,7 point du taux de chômage.

La rupture de définition du statut BIT est traitée de la même manière que les autres ruptures de niveau (cf. *infra*). Pour les autres variables, on a essayé, autant que possible (c'est-à-dire en utilisant toutes les variables disponibles), de reconstruire des définitions homogènes sur l'ensemble des périodes, en recalculant, éclatant ou regroupant certaines modalités.

Les choix effectués sont décrits en annexe 1.

Détermination de la population de référence

La population de référence de l'EEA comme de l'EEC est celle des ménages ordinaires de France métropolitaine (individus âgés de 15 ans ou plus en fin d'année). L'enquête annuelle s'appuie sur une population en mars de l'année, alors que l'enquête en continu permet de calculer des niveaux moyens dans l'année, c'est-à-dire des niveaux calés sur une population moyenne dans l'année ou plus exactement, des niveaux moyens sur les quatre trimestres, chaque trimestre étant calé sur une population moyenne sur le trimestre.

Pour pouvoir comparer les niveaux issus de l'enquête emploi annuelle et ceux issus de l'enquête en continu (moyenne des quatre enquêtes trimestrielles), il est donc nécessaire d'appuyer les résultats de l'enquête emploi annuelle sur une population des ménages de France métropolitaine **en moyenne annuelle**.

Aussi, **pour le suivi des séries longues, on utilise les données de l'enquête Emploi annuelle uniquement en structure (parts dans la population, dans la population active ou dans l'emploi). Ces données sont ensuite appliquées à des niveaux de population représentatifs de la population en moyenne annuelle des personnes de 15 ans et plus, vivant en ménages ordinaires en France métropolitaine.**

Ces populations en moyenne annuelle ne sont pas disponibles pour les années 1975-2002. On doit donc les reconstruire. Pour chaque catégorie de sexe et d'âge quinquennal :

- 1) on calcule le rapport moyen entre la population en moyenne annuelle et la population moyenne sur le premier semestre à partir des données de l'EEC de 2003 à 2006.
- 2) on applique ce rapport aux populations en mars pour les années antérieures à 2002, ces populations en mars étant directement prises dans l'EEA.

On obtient donc des populations des ménages ordinaires *moyennes dans l'année* pour les années 1975-2002, et ce, sous l'hypothèse que le rapport pour une catégorie entre la population moyenne dans l'année et population en mars est constant sur 1975-2002. Les populations moyennes de 2003 à 2006 viennent, quant à elles, directement de l'EEC.

Rupture de niveau due au passage à l'enquête Emploi en continu

La saisonnalité de l'activité et de ses diverses composantes fait que les niveaux calculés en mars (cas de l'enquête Emploi annuelle) ne sont pas comparables à des niveaux calculés comme moyenne sur l'année. A l'écart dû à la saisonnalité s'ajoutent des écarts dus aux modifications de questionnaire.

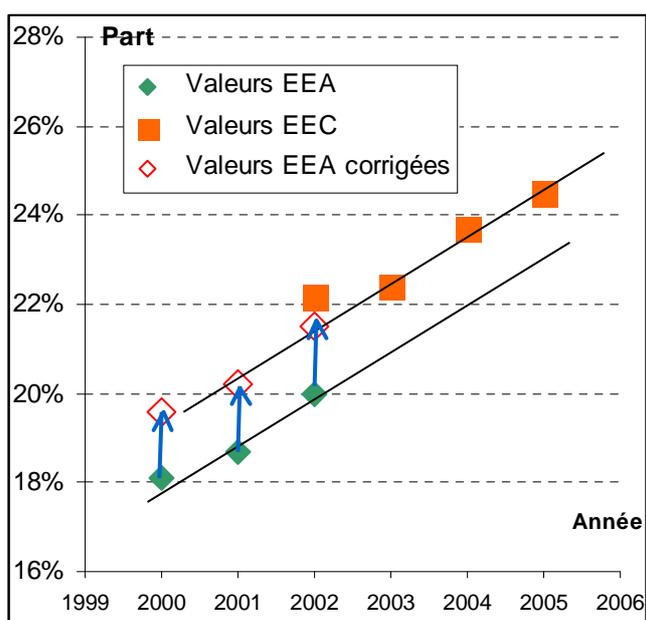
On corrige ces deux types de rupture en appliquant, sur toutes les valeurs issues de l'enquête annuelle, un coefficient additif censé représenter l'écart moyen entre les valeurs dans l'enquête en continu et dans l'enquête annuelle.

³ cf. *Insee Première* n°1164, « Résultats de l'enquête Emploi : le chômage baisse depuis début 2006 », novembre 2007



Par exemple, pour le taux d'activité, si on estime que le taux d'activité en moyenne annuelle est, en moyenne, de 1 point plus élevé qu'en mars, on ajoute un point de taux d'activité à toutes les valeurs observées dans les enquêtes Emploi annuelles, afin d'obtenir un taux d'activité qui puisse être considéré comme une moyenne annuelle.

Le graphique suivant illustre la méthode sur un exemple fictif (concernant, par exemple, la part d'une catégorie donnée de personnes en emploi). Les losanges représentent les valeurs tirées de l'enquête Emploi annuelle pour les années 2000 à 2002. Les carrés représentent les valeurs tirées de l'enquête Emploi en continu de 2002⁴ à 2005. Nous considérons que la tendance de la grandeur mesurée est localement affine (c'est-à-dire que sur une courte période, entre 2000 et 2005, les points sont *grosso modo* alignés sur le graphique), avec juste un écart de niveau entre les observations de l'enquête Emploi annuelle et de l'enquête en continu (cf. les deux droites parallèles sur le graphique). Le coefficient additif de correction est calculé comme cet écart moyen : il correspond à la flèche bleue sur le graphique.



Cette méthode a déjà été appliquée par l'Insee en 2006, dans le cadre des projections de population active⁵. Des populations actives « en moyenne annuelle » ont ainsi été calculées pour toute la période 1975-2002, pour chaque sexe et catégorie d'âge quinquennal.

Dans certains cas où les valeurs que l'on cherche à corriger sont proches de zéro, l'hypothèse de correction « additive », présentée ci-dessus, peut conduire à des valeurs corrigées inférieures à zéro, donc aberrantes. Dans ces cas, on effectue au préalable une transformation logistique des données. Cela revient *grosso modo* à effectuer une hypothèse de correction multiplicative plutôt qu'additive. Les cas où cette transformation a été appliquée sont signalés dans l'annexe 2.

Par ailleurs, une contrainte technique s'ajoute : lorsqu'il existe plusieurs niveaux d'agrégation pour une variable (par classe d'âge, par secteur, etc.), les résultats ne sont pas identiques selon que l'on corrige séparément chacune des parties ou que l'on corrige directement l'agrégat.

Pour assurer la cohérence entre les séries désagrégées et les agrégats correspondants dans les séries longues sur le marché du travail, on corrige les parts selon une procédure qui respecte les phases suivantes :

⁴ L'EEC a été mise en place en 2002 mais celle de 2002 n'a pas été utilisée dans les publications.

⁵ « Projections 2005-2050 - Des actifs en nombre stable pour une population âgée toujours plus nombreuse 1092 », Elise Coudin, *Insee Première* n°1092, juillet 2006



- les séries de population active corrigées par sexe et âge quinquennal retenues dans les séries longues sont d'abord calculées selon la méthodologie utilisée dans le cadre des projections de population active (Insee, 2006). Les valeurs sont néanmoins légèrement différentes de celles de cette publication, du fait de la modification à l'automne 2007 de la définition du statut BIT.
- on corrige ensuite de la rupture de série les taux de chômage par sexe et âge quinquennal. En appliquant ces taux aux populations actives calculées à l'étape précédente, on obtient des séries d'emploi et de chômage par sexe et âge quinquennal.
- on corrige enfin la répartition selon chacune des variables d'intérêt (répartition par diplôme, par CS, par type de contrat de travail, etc.) au sein de chaque catégorie de sexe, d'âge regroupé (c'est-à-dire : 15 à 24 ans, 25 à 49 ans, 50 ans et plus) et de statut d'activité BIT. Les niveaux d'emploi ou de chômage pour chacune de ces caractéristiques sont ensuite calculés en appliquant la répartition « corrigée » au niveau d'emploi par sexe et âge regroupé (lui-même également corrigé).
- Les taux de chômage, taux d'emploi, etc. par catégories (CS, diplôme, etc.) sont calculés à partir des niveaux d'emploi et de chômage estimés à l'étape précédente.

La correction, illustrée par le schéma précédent, est fondée sur l'hypothèse que chaque variable d'intérêt (par exemple, la part d'une catégorie de diplôme dans l'emploi) est localement affine. La méthode consiste à régresser les valeurs observées de part et d'autre de la date de passage à l'enquête Emploi en continu (par exemple de 2001 à 2005, avec 2 observations annuelles 2001 et 2002 et 3 observations en moyenne annuelle 2003 à 2005) sur une tendance affine et une indicatrice « enquête annuelle ».

$$VAR(t) = a + b * t + marche * (t < 2003) + résidu$$

Les valeurs « corrigées » avant 2002 (inclus) correspondent aux valeurs des enquêtes annuelles auxquelles on soustrait la marche estimée par la méthode précédente.

$$VAR_corrigée(t) = VAR(t) - marche \quad \text{pour } t < 2003$$

Les valeurs de l'enquête Emploi en continu sont ensuite prises telles quelles à partir de 2003.

$$VAR_corrigée(t) = VAR(t) \quad \text{pour } t \geq 2003$$

Cette méthodologie générale est adaptée selon les cas. Les modèles utilisés pour chaque variable sont présentés dans l'annexe 2.

Un calage est enfin effectué pour que la somme des parts de chacune des modalités d'une variable soit bien égale à un.



Annexe 1 : Reconstruction et choix effectués sur certaines variables

Les noms des variables font référence aux noms des variables dans les fichiers détaillés de l'enquête Emploi.

Activité au sens du BIT (ACT)

On retient les militaires du contingent parmi les actifs occupés.

La mise en œuvre de l'interprétation européenne des critères du BIT (appliquée par l'Insee pour les données de l'enquête Emploi en continu depuis la validité 2003) pose problème pour les années les plus anciennes de l'enquête Emploi annuelle⁶. Il a donc été décidé de mesurer le chômage avec l'interprétation européenne uniquement dans les enquêtes en continu à partir de 2003, et avec l'ancienne interprétation de l'Insee dans les enquêtes annuelles jusqu'en 2002. L'écart entre les deux interprétations est « capté » par la correction de rupture de niveau, en même temps que l'écart dû à la saisonnalité (valeur en mars vs. moyenne annuelle).

Âge

On retient l'âge atteint au 31/12 comme concept d'âge pour toutes les ventilations par âge (à l'exception des comparaisons internationales). Ce choix est homogène à toutes les publications de l'Insee (série mensuelle de chômage, *Insee Première* sur l'enquête Emploi, projections de population active, ...), mais peut différer de celui fait par divers organismes internationaux (OCDE, Eurostat).

On regroupe dans la même catégorie les 70 ans et plus de même sexe, ce qui induit une petite différence avec les fichiers détaillés de l'enquête Emploi, qui distinguent les 70-74 ans des 75 ans et plus.

Catégorie socioprofessionnelle (CS)

On ne retient la ventilation que selon le premier chiffre de la CS. On utilise la variable CSTOT, qui classe tous les individus, quel que soit leur statut d'activité. Pour les chômeurs, la CS est celle du dernier emploi occupé, une catégorie spéciale étant réservée aux chômeurs n'ayant jamais travaillé.

CSTOT est identique aux deux premiers caractères de la profession P sur le champ des actifs occupés. Elle est également identique à la variable CSE (CS pour les actifs) sur le champ des actifs.

On mélange les deux catégories 7 (inactifs ayant déjà travaillé) et 8 (chômeurs, militaires du contingent ou inactifs n'ayant jamais travaillé) en une catégorie X (chômeurs n'ayant jamais travaillé, militaires du contingent et inactifs). Les CS inconnues sont également comptabilisées dans cette catégorie.

Diplôme

On retient les modalités de la variable DDIPL. La série commence donc en 1982 (les modalités retenues étaient très différentes sur la période 1975-1981 ; les deux versions avant et après 1982 sont maintenues dans les enquêtes de 1983 à 1989, mais pas après).

⁶ Cf. note n°2727/DG75-F204 du 14 novembre 2007, « Rétropolation avant 2002 de la série trimestrielle de chômage au sens du BIT »



Formes particulières d'emploi

La variable STATUT a beaucoup évolué au cours du temps, notamment du fait d'une prise en compte différente du secteur public, et du fait de l'évolution des dispositifs légaux de stage ou de contrats aidés. Une tentative de définition « stable » dans le temps a été proposée.

Ses caractéristiques principales sont :

- pour l'EEC : on utilise STATUT plutôt que STATOEP (cette dernière variable a des évolutions difficilement explicables en 2005). On recode les modalités afin de rassembler public et privé en ce qui concerne les CDD, intérimaires, stagiaires et contrats aidés (par homogénéité avec l'enquête annuelle). On recode aussi les modalités pour utiliser les mêmes codes que l'enquête emploi annuelle.
- Pour 1993-2002 : on rebascule dans la catégorie idoine les CDD, intérimaires, stagiaires et contrats aidés du public (État et collectivités territoriales). De même les contractuels à durée limitée, les vacataires, auxiliaires et pigistes sont rebasculés parmi les CDD.
- Pour 1982-1989 : les stagiaires (hors fonctionnaires-stagiaires) de la fonction publique sont recodés en « stages et contrats aidés », plutôt qu'en « contractuels / vacataires / etc. ».

Pour le calcul de la part dans l'emploi dans l'enquête annuelle, on réintègre tous les stagiaires (STJ='1') parmi les actifs occupés (le stock de stagiaires est en effet similaire entre l'EEA et l'EEC en 2002, soit environ 500 000 personnes, mais quasiment tous les stagiaires sont des actifs occupés dans l'EEC alors que 100 000 ne le sont pas dans l'EEA).

Certaines ruptures de séries ne peuvent pas être corrigées par des filtres : ainsi, les salariés en période d'essai ou en stage d'entrée dans la profession n'étaient pas interrogés sur la nature de leur contrat de travail dans l'enquête annuelle, et donc jamais classés en CDD. Dans l'enquête en continu, les salariés en CDD et en période d'essai (ou en stage d'entrée dans la profession) sont environ 135 000 en 2002T1. Cela pourrait expliquer l'écart observé de 110 000 sur le nombre de CDD (hors contrats aidés) entre l'EEA et l'EEC.

Une autre rupture concerne l'intérim : les intérimaires sont environ 90 000 de plus dans l'EEA que dans l'EEC au premier trimestre 2002. Ceci pourrait être due aux différences de formulation entre les questions (« M. est-il intérimaire ? » dans l'EEA, « Quel est le type de votre contrat de travail ? » dans l'EEC).

Sous-emploi

Cette variable apparaît dans l'enquête annuelle avec la validité 1998. On recrée une variable sous-emploi depuis 1990 en appliquant le filtre servant après 1998. Avant 1990, les variables servant dans ce filtre n'existent pas et la série ne peut donc pas être reconstruite.

Comme pour le statut d'activité au sens du BIT, l'Insee a modifié son interprétation des critères permettant de définir le sous-emploi pour les données de l'enquête Emploi en continu depuis la validité 2003. Cette modification ne change pas le total des personnes en sous-emploi, mais la répartition entre les catégories 1 et 2. Cette cause de rupture de niveau n'a pas été corrigée dans la définition des variables utilisées, elle est donc traitée au moment de la correction de rupture de niveau, en même temps que l'écart dû à la saisonnalité (cf. annexe 2).



Annexe 2 : Modèles et hypothèses utilisés pour la correction de la rupture de niveau en 2003

Étape 1 : Population active au sens du BIT

Les séries de population active retenues dans les séries longues sont calculées selon la même méthodologie que celle de l'exercice de projections de population active (Insee, 2006). La seule différence avec cette publication provient de la modification par l'Insee, à l'automne 2007, de son interprétation des critères servant à déterminer le statut au sens du BIT.

La méthodologie de correction est détaillée dans le document de travail « [Projections de population active 2006-2050 : des actifs en nombre stable, une population âgée toujours plus nombreuse](#) »⁷ d'Élise Coudin (n°F0704, mars 2007). Cependant, le nouveau jeu de données a conduit à des spécifications légèrement différentes que celles utilisées dans le document de travail « projections » (présence ou non d'autocorrélation selon le résultat des tests, inclusion ou non d'une rupture de série selon la significativité du coefficient, et les critères de spécification du modèle de régression).

Étape 2 : Statut des actifs au sens du BIT

Afin d'obtenir des séries du nombre d'actifs occupés et de chômeurs au sens du BIT, on corrige la rupture de série sur les taux de chômage par sexe et âge quinquennal.

Le modèle estimé pour la rupture de série dans chacune de ces catégories est le suivant :

$$\text{Log} (TxChom / (1 - TxChom)) = a + b * t + rupture_niveau * EEA + \text{résidu}$$

TxChom désigne le taux de chômage pour la catégorie de sexe et d'âge quinquennal considérée, prise en mars pour l'enquête Emploi annuelle, et en moyenne sur l'année pour l'enquête Emploi en continu. *EEA* est une indicatrice qui vaut 1 si la mesure de *TxChom* provient d'une enquête Emploi annuelle, et 0 sinon. *t* désigne l'année.

Le modèle est estimé autour de l'année de rupture de série, c'est-à-dire en pratique sur la période 2001-2004⁸ (d'où deux points de l'enquête Emploi annuelle : 2001 et 2002, et trois points de l'enquête Emploi en continu : 2002, 2003 et 2004).

De 1975 à 2002, le taux de chômage corrigé est ensuite calculé de manière à ce que :

$$\text{Log} (TxChomCorrigé / (1-TxChomCorrigé)) = \text{Log} (TxChom / (1-TxChom)) - rupture_niveau$$

Quelques traitements supplémentaires ont été effectués :

Comme il a déjà été indiqué précédemment, la mise en œuvre de l'interprétation européenne des critères du BIT pour les années les plus anciennes de l'enquête Emploi pose problème. Il a été décidé de mesurer le taux de chômage *TxChom* avec l'interprétation européenne uniquement dans les enquêtes en continu à partir de 2003, et avec l'ancienne interprétation française dans les enquêtes annuelles jusqu'en 2002. L'écart entre les deux interprétations est donc « capté » par la correction de rupture de niveau, c'est-à-dire en pratique par le coefficient *rupture_niveau*.

En ce qui concerne le chômage, on dispose d'une information externe fournie par l'ancienne série mensuelle de chômage au sens du BIT (elle-même calée sur les valeurs annuelles, en mars ou en moyenne annuelle, de l'enquête Emploi). L'information sur le profil infra-annuel

⁷ http://www.insee.fr/fr/publications-et-services/docs_doc_travail/f0704.pdf

⁸ Le choix de cette période a pour but de satisfaire l'hypothèse que le taux de chômage est localement affine sur la période d'observation considérée. L'année 2001 est en effet une année de retournement conjoncturel : le taux de chômage diminue jusqu'à cette année, et réaugmente à partir de 2002 (il est ensuite stable en 2004 et 2005). Inclure dans la régression à la fois des observations avant et après 2001 aurait fragilisé l'hypothèse de taux de chômage localement affine, puisque la tendance n'est pas la même sur ces deux sous-périodes.



du chômage contenue dans cette ancienne série a été utilisée de la manière suivante : chaque taux de chômage a été multiplié au préalable par le ratio entre la valeur en moyenne annuelle et la valeur en mars du taux de chômage. Ce ratio est calculé pour chaque année et dans la catégorie de sexe et d'âge regroupé correspondante.

Au moment où la rétopolation des séries de taux de chômage a été effectuée, les nouveaux fichiers de l'enquête Emploi en continu de 2002 n'étaient pas encore disponibles. Disposer de l'information pour 2002 est cependant important pour l'estimation, dans la mesure où c'est entre 2002 et 2003 que l'augmentation du chômage a été la plus forte. Une estimation du taux de chômage en 2002 a donc été reconstruite, à partir de l'estimation de 2003 (tirée de l'enquête Emploi) et une estimation de l'évolution entre 2002 et 2003, calculée sur le champ des chômeurs PSERE (sans emploi à la recherche d'un emploi) et avec les données non pondérées de l'enquête Emploi. Ce calcul est effectué uniquement pour le taux de chômage. Pour toutes les autres variables, et donc pour toutes les régressions présentées ci-dessous, les données de l'enquête Emploi en continu ne sont utilisées qu'à partir de 2003 (inclus).

Enfin, du fait de valeurs aberrantes, ce point de 2002 n'a pas été utilisé pour les estimations relatives aux seniors (toutes les classes d'âge de plus de 50 ans). Par ailleurs, la correction a été contrainte à 0 pour les classes d'âge de plus de 65 ans, pour lesquelles le taux de chômage est de toute façon peu significatif.

Étape 3 : Part des catégories (CS, diplôme, sous-emploi, formes particulières d'emploi) au sein des classes par sexe, âge regroupé et statut BIT

Diplôme

Dans le calcul des parts de chaque diplôme pour chaque catégorie, on ne prend pas en compte les diplômes inconnus. Ces parts sont donc calculées uniquement pour les répondants. Elles sont extrapolées ensuite à l'ensemble.

Il existe une forte rupture de série entre l'EEA et l'EEC, indépendamment de la saisonnalité (rupture observable entre l'EEA et les premiers trimestres de chaque année dans l'enquête en continu). Cette rupture est due au fait que le diplôme était indiqué librement par les répondants dans l'EEA, alors qu'il est maintenant choisi dans une liste dans l'EEC.

Par ailleurs, les parts dans l'emploi de certaines catégories de diplôme sont très faibles. Une correction additive pourrait aboutir à des parts nulles (négatives) avant 2003. On effectue donc une transformation logistique de la part avant de procéder à la correction, afin d'obtenir des parts corrigées strictement positives.

On utilise à la fois les observations de l'enquête Emploi en continu en moyenne annuelle et au cours du premier trimestre de chaque année. Le but est d'estimer séparément la rupture due au changement de questionnaire entre l'EEA et l'EEC et la rupture liée à la saisonnalité. On augmente donc le modèle pour prendre en compte : 1) le fait d'être EEC ou EEA ; 2) le fait d'être en moyenne annuelle plutôt qu'en mars. On fait l'hypothèse que la saisonnalité de mars (EEA) est la même que celle du premier trimestre (dans l'EEC). Le modèle de correction de rupture de série sur la part s'écrit finalement :

$$\text{Log} (P / (1-P)) = a + b*t + \text{rupture_niveau} * \text{EEA} + \text{saison} * \text{MA} + \text{résidu}$$

Où P = part du diplôme dans l'emploi par exemple, MA est une indicatrice valant 1 si l'observation est en moyenne annuelle et 0 sinon (EEA ou premier trimestre de l'EEC), et EEA est une indicatrice valant 1 pour une observation tirée de l'EEA et 0 pour une observation de l'EEC. L'équation est estimée sur la période allant de 2000 à 2005 (inclus) pour chacune des classes de sexe et d'âge regroupé (15-24 ans ; 25-49 ans ; 50 ans et plus).

Il existe par ailleurs une rupture de série importante en 1990, due au fait qu'on avait beaucoup de diplômes non renseignés avant, notamment parmi les jeunes et les inactifs. Cette rupture existe toujours lorsqu'on ne calcule les parts dans chaque catégorie que parmi



les seuls répondants. On corrige donc d'une seconde rupture de série en 1990, en appliquant la même méthode que pour la rupture de 2003 (avec juste une indicatrice avant/après).

$$\text{Log} (P / (1-P)) = c + d*t + \text{rupt90} * \text{APRES1990} + \text{résidu}$$

Où P = part du diplôme dans l'emploi, APRES1990 est une indicatrice qui vaut 1 après 1990. L'équation est estimée sur la période allant de 1987 à 1992 (inclus) pour chacune des classes de sexe et d'âge regroupé (15-24 ans ; 25-49 ans ; 50 ans et plus).

Pour les deux corrections de rupture de niveau, une contrainte de calage est ajoutée pour que la somme des parts corrigées des diplômes dans l'emploi et dans le chômage soit bien égale à 1. Ce sera également le cas pour toutes les autres variables ci-dessous.

Catégorie socioprofessionnelle (CS)

Les valeurs observées en 2003 semblent instables pour certaines catégories, peut-être du fait du changement de nomenclature au cours de cette année. On exclut donc les valeurs 2003 pour le calcul du coefficient de correction de la rupture de série (on utilise uniquement 2001-2002 et 2004-2006).

Il semble que le changement d'enquête ou de nomenclature ait un fort effet sur le niveau de certaines catégories (notamment les 4, 5 et 6). Utiliser les premiers trimestres de l'EEC comme des « équivalents EEA » (comme si l'essentiel de la rupture de série provenait de la seule saisonnalité) aboutirait donc à des résultats biaisés. Dans le cas des CS, on n'utilise donc pas les premiers trimestres de l'EEC. Autrement dit, on ne cherche pas à identifier séparément l'effet de changement d'enquête et celui de la différence de saisonnalité entre les deux enquêtes.

Pour résumer, le coefficient de correction EEA/EEC est estimé à partir du modèle suivant :

$$\text{Part de la CS dans l'emploi} (t) = a + b * t + c * \text{indicatrice2003} + \text{rupture_niveau} * \text{EEA} + \text{résidu}$$

Sur la période allant de 2000 à 2006 (inclus), pour chacune des classes de sexe et d'âge regroupé (15-24 ans ; 25-49 ans ; 50 ans et plus).

On n'a pas retenu de rupture de niveau (*rupture_niveau*) dans l'équation pour les 15-24 ans qui sont cadres (cs=3) et chômeurs (act=2) (hommes et femmes), ainsi que chez les femmes (sexe=2), artisans (cs=2), chômeuses (act=2). Son inclusion conduisait à des valeurs aberrantes (parts dans le chômage nulles ou négatives par le passé).

Une contrainte de calage est ajoutée pour que la somme des parts corrigées des CS dans l'emploi et dans le chômage soit bien égale à 1.

Sous-emploi

La rupture de série sur le sous-emploi est estimée en utilisant les observations de l'enquête Emploi en continu en moyenne annuelle et au cours du premier trimestre selon le modèle suivant :

$$\text{Part des actifs occupés en sous emploi} (t) = a + b*t + \text{rupture_niveau} * \text{EEA} + \text{saison} * \text{MA} + \text{résidu}$$

Où MA est une indicatrice valant 1 si l'observation est en moyenne annuelle et 0 sinon (EEA ou premier trimestre de l'EEC), et EEA est une indicatrice valant 1 pour une observation tirée de l'EEA et 0 pour une observation de l'EEC. L'équation est estimée sur la période allant de 2001 à 2004 (inclus) pour chacune des classes de sexe et d'âge regroupé (15-24 ans ; 25-49 ans ; 50 ans et plus).



Formes particulières d'emploi

La rupture de série sur les formes particulières d'emploi est estimée en utilisant les observations de l'enquête Emploi en continu en moyenne annuelle et au cours du premier trimestre de chaque année.

Les parts dans l'emploi des formes particulières d'emploi (CDD, intérim, contrats aidés, ...) étant généralement faibles, notamment chez les plus de 25 ans, une correction additive pourrait aboutir à des parts négatives avant 2003. On corrige donc après une transformation logistique de la part, pour conserver des parts positives.

Le modèle suivant est donc estimé :

$$\text{Log} (P / (1-P)) = a + b*t + \text{rupture_niveau} * \text{EEA} + \text{saison} * \text{MA} + \text{résidu}$$

Où P désigne la part de la forme particulière considérée dans l'emploi à la date t , MA est une indicatrice valant 1 si l'observation est en moyenne annuelle et 0 sinon (EEA ou premier trimestre de l'EEC), et EEA est une indicatrice valant 1 pour une observation tirée de l'EEA et 0 pour une observation de l'EEC. L'équation est estimée sur la période allant de 2001 à 2005 (inclus) pour chacune des classes de sexe et d'âge regroupé (15-24 ans ; 25-49 ans ; 50 ans et plus).

