

La refonte de l'enquête Emploi (2017-2021)

Documents de travail

N° 2023-22 - Octobre 2023



La refonte de l'enquête Emploi (2017-2021)

Département de l'emploi et des revenus d'activité

Ce document regroupe la contribution de nombreux acteurs ayant participé au projet de refonte de l'enquête Emploi : Romain Cantignaux, Eliette Castelain, Claude Delevallez, Marianne Fabre, Marguerite Garnero, François Guillaumat-Tailliet, Yves Jauneau, Perrine Kauffmann, Sylvain Larrieu, Ingrid Lefebvre-Hoang, Alexis Montaut, Julien Moreau, Anne-Sophie Pichavant, Olivia Sautory, Chloé Tavan, Antoine Van Assche, Klara Vinceneux.

Sa rédaction a été coordonnée par Laura Castell.

Les auteurs remercient Christel Colin et Vladimir Passeron pour leur appui et leur relecture attentive.

Introduction

En 2021, l'enquête Emploi a fait l'objet d'une nouvelle refonte majeure, consistant en une rénovation de son questionnaire, de son protocole et de sa méthode de pondération.

Cette refonte a été rendue nécessaire par l'adoption d'un nouveau règlement cadre européen sur les enquêtes sociales (IESS – Integrated European Social Statistics) qui pousse les États membres à harmoniser davantage leurs questionnaires. Ce projet de refonte a été l'occasion d'embarquer d'autres évolutions, dont l'introduction du mode de collecte Internet dans le protocole de l'enquête. Cette modernisation du protocole est en phase avec le développement des enquêtes sur Internet dans les enquêtes auprès des ménages à l'Insee comme dans la plupart des instituts nationaux de statistique.

Pour mener à bien cette refonte d'ampleur, l'Insee a mis en place dès 2017 un projet de rénovation de l'enquête Emploi, le projet « Mélopée » (pour Mise en ligne et optimisation de l'enquête Emploi). Il a complété le projet expérimental « Muse » qui entre 2014 et 2018 a été mené à l'Insee dans le but de préparer l'introduction de la collecte par Internet dans l'enquête Emploi en continu.

Ce document de travail présente les grands changements liés à cette refonte et l'organisation mise en place pour les mettre en œuvre. Il est constitué de diverses contributions dont le niveau de détail peut différer selon les sujets et n'est pas proportionnel à l'ampleur du travail consacré. Par exemple, la refonte applicative et informatique nécessaire aux divers changements n'est pas présentée en détail alors même qu'elle a représenté une part conséquente des travaux menés dans le cadre de ce projet.

Sommaire

Introduction.....	2
I. Le projet Mélopée pour refondre l'enquête Emploi.....	4
1. Le contexte.....	4
2. Les objectifs du projet de refonte.....	6
3. L'organisation du projet.....	7
4. Un pilote d'envergure : pourquoi ?.....	9
II. Les principales évolutions de l'enquête Emploi.....	12
1. La rénovation du questionnaire.....	12
2. La modernisation du protocole.....	19
2.1. Un protocole multimode en réinterrogation.....	19
2.2. Les règles de non-remise en collecte.....	25
3. La rénovation de la méthode de pondération.....	30
3.1. L'apport de la nouvelle base de sondage.....	31
3.2. La nouvelle méthode de pondération depuis 2021.....	36
3.3. De nouvelles règles de rattachement.....	38
4. Une refonte jusqu'à l'aval et la diffusion des données.....	39
5. Des outils adaptés à ces évolutions.....	41
III. Mesure des ruptures de série et rétropolation.....	45
1. Un ajustement des séries longues sur les nouvelles estimations de population.....	45
2. Méthode générale de mesure des ruptures et de rétropolation des séries longues.....	48
2.1. Un Pilote pour mesurer les ruptures de séries.....	48
2.2. La mesure des ruptures de séries.....	49
2.3. La méthode générale de rétropolation des séries longues annuelles et trimestrielles.....	50
3. Les ruptures de séries et rétropolations sur le statut BIT.....	53
4. La révision des poids dans les micro-données.....	57
IV. Bilan du projet.....	60
1. Un périmètre revu partiellement.....	60
2. Bilan de la réalisation du projet.....	60
3. Enseignements pour les projets futurs.....	65
Bibliographie.....	67
Annexes.....	68
Annexe 1 : Ruptures de série et rétropolation des séries par thématique (hors statut BIT).....	68
Annexe 2 : Séries longues particulières.....	103

I. Le projet Mélopée pour refondre l'enquête Emploi

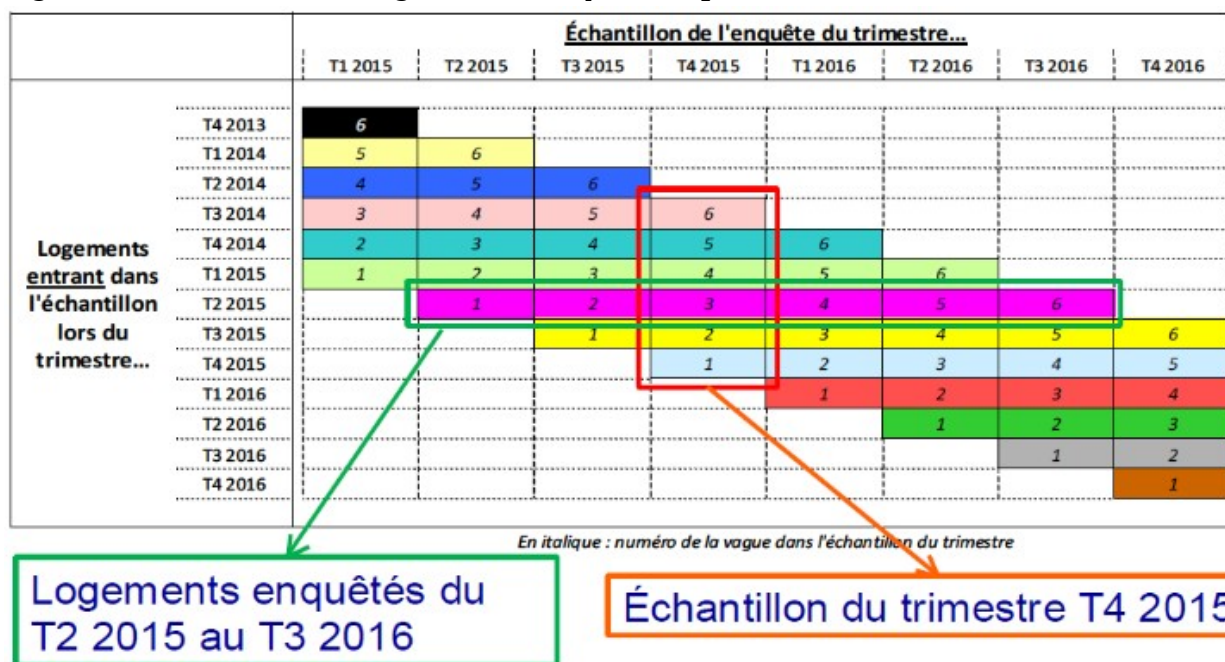
1. Le contexte

L'enquête Emploi est née en France en 1950 pour permettre une mesure régulière de l'emploi et du chômage. Depuis 1960, elle est la version française de l'enquête européenne sur la force de travail (Labour Force Survey – LFS). Elle évolue désormais en lien avec les préoccupations européennes d'harmonisation des enquêtes pour produire des résultats les plus comparables possibles sur le fonctionnement du marché du travail dans les différents États membres.

L'enquête Emploi a connu, au fil de son existence, de nombreuses refontes, de nature méthodologique ou conceptuelle. Devenue continue en France en 2003, conformément au règlement européen, la collecte de l'enquête Emploi est depuis lors réalisée sur toutes les semaines de l'année, permettant à l'Insee de produire des résultats trimestriels.

Le protocole consiste à interroger les habitants des logements échantillonnés durant 6 trimestres consécutifs (par la suite on parlera de rang d'interrogation ; *figure 1*). Jusqu'en 2021, la collecte était entièrement intermédiée par le réseau des enquêteurs de l'Insee, réparti sur l'ensemble du territoire : la première et la dernière interrogation étaient réalisées en face à face et les interrogations intermédiaires par téléphone.

Figure 1 : Protocole d'interrogation de l'enquête Emploi en continu



La précédente refonte de l'enquête Emploi (projet « Reflee », pour REFonte de l'Enquête Emploi) a eu lieu en 2013. Elle avait pour objectif d'améliorer le questionnaire en lien avec la réglementation européenne et de nouveaux besoins des utilisateurs, de refondre l'application de gestion propre à l'enquête Emploi et d'intégrer les Dom (hors Mayotte) dans le dispositif en continu.

Chaque modification de l'enquête peut affecter la mesure des indicateurs, nécessitant ainsi un travail important de réropolation, c'est-à-dire de construction de séries sur longue période qui

offrent une vision cohérente du marché du travail. Pour cette raison, l'enquête Emploi n'évolue pas chaque année, mais tous les dix ans environ, dans le cadre de lourds exercices de refonte.

Le projet de rénovation de l'enquête Emploi en 2021 est la dernière grande refonte mise en œuvre pour l'enquête Emploi. Elle s'inscrit dans le cadre de deux évolutions majeures. L'enjeu de faire coïncider la mise en œuvre de ces deux évolutions majeures, non liées l'une à l'autre, est de ne provoquer qu'une seule rupture de série sur les indicateurs diffusés, dont le taux de chômage trimestriel.

La première de ces évolutions est la mise en place d'un **nouveau règlement européen sur les enquêtes sociales (IESS)**¹. En discussion depuis 2013, ce règlement est mis en œuvre à partir de 2021. Il a pour principal objectif d'harmoniser davantage les questionnaires des différentes enquêtes sociales, entre pays mais également entre enquêtes (Cases, 2019).

Ce règlement impose les informations que doit collecter l'enquête, contraignant ainsi le questionnaire, y compris dans son déroulé. Il impose également certains principes méthodologiques, en matière d'échantillonnage ou de redressement, et de protocole. Si les contraintes méthodologiques et de protocole ne sont pas fondamentalement modifiées par le nouveau règlement, il vise une harmonisation plus poussée des questionnaires et une modularité accrue de l'enquête. Jusque-là, Eurostat imposait une harmonisation par les outputs, laissant libres les pays de la manière de les calculer. Le nouveau règlement européen va plus loin en matière d'harmonisation entre pays pour améliorer la comparabilité des résultats. Ainsi pour les variables servant à définir le statut d'activité au sens du BIT, l'ordre des questions et le déroulé du questionnaire sont désormais imposés.

La seconde évolution majeure concerne le **développement du recours à Internet pour les enquêtes auprès des ménages**. De fait, l'Insee a lancé un programme d'expérimentation au début des années 2010 qui visait en priorité l'évaluation de l'introduction d'Internet dans l'enquête Emploi en continu. Les objectifs de cette stratégie sont multiples : réduire les coûts de certaines opérations – l'enquête Emploi étant la plus coûteuse des enquêtes ménages – pour réorienter le travail des enquêteurs vers les enquêtes et les enquêtés nécessitant l'intervention d'enquêteurs ; moderniser les protocoles de collecte pour répondre aux attentes d'une population utilisant de plus en plus Internet dans ses démarches quotidiennes ; augmenter la participation en proposant un mode de collecte complémentaire plus souple et moderne.

Ainsi, entre 2014 et 2018, le projet « Muse » (Multimode Sur l'Emploi), qui n'était pas de refonte mais expérimental, a permis d'expérimenter le recours à Internet dans l'enquête Emploi afin de proposer un protocole multimode intégrant Internet comme nouveau mode de collecte (Garnero, 2019). Dans ce cadre, plusieurs expérimentations ont été réalisées, à la fois qualitatives et quantitatives, portant spécifiquement sur le questionnement par Internet ou plus largement sur le protocole multimode. Ces expérimentations ont permis de conclure sur le protocole le plus adapté pour introduire Internet comme nouveau mode de collecte avec les contraintes de l'enquête Emploi.

1 Le règlement cadre IESS encadre l'ensemble des enquêtes sociales européennes <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2019:261I:TOC>. Les actes spécifiques aux LFS comprennent un règlement délégué fixant la liste des variables (<https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32020R0257>) et un règlement d'exécution précise les définitions de ces variables ainsi que les éléments de protocole et de méthodologie à respecter (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.336.01.0059.01.ENG).

Ainsi, ce projet Muse a confirmé que le questionnaire de réinterrogation de l'enquête Emploi pouvait être réalisé par Internet et qu'une solution multimode était la seule à pouvoir maintenir des taux de collecte et une représentativité suffisants. Le projet a permis d'avancer en matière d'ergonomie du questionnaire web, et d'affiner les protocoles de prise de contact et de relance afin de maximiser le taux de réponse.

Ce contexte a conduit à réfléchir à une nouvelle refonte de l'enquête Emploi pour profiter de la nécessaire refonte liée aux impératifs européens pour réaliser les modifications de questionnement, de protocole et d'ingénierie statistique souhaitées.

2. Les objectifs du projet de refonte

Les objectifs affichés dans l'expression des besoins du projet étaient les suivants :

- **Produire une enquête Emploi en continu conforme au règlement européen IESS**, tant du point de vue du questionnaire, du protocole, du champ géographique, que des exigences en matière de précision régionale.
 - Pour le questionnaire, il s'agissait d'en produire un nouveau, harmonisé sur les questions européennes et concerté avec les principaux utilisateurs de l'enquête sur les parties du questionnaire laissées à l'initiative nationale.
 - Pour le protocole, il s'agissait de respecter un certain nombre de contraintes, comme en matière de durée de collecte par exemple. Ces contraintes étaient déjà en grande partie respectées par la France.
 - Pour le champ, il s'agissait de couvrir l'ensemble du territoire national, ce qui n'était pas le cas de l'enquête Emploi en continu qui ne couvre pas le département de Mayotte (et n'était toujours pas le cas à l'issue du projet, cet axe ayant été dépriorisé pour des raisons indépendantes de l'équipe projet).
 - Pour la précision, ces nouvelles exigences ont été prises en compte antérieurement à la mise en œuvre du projet, dès la mise en place du nouvel échantillon en 2019 (Vincent et Paliot, 2022).
- **Produire une enquête plus moderne et moins onéreuse** en mettant en place un protocole multimode concurrentiel en réinterrogation, en introduisant le mode de collecte Internet en première intention, en plus du téléphone.
 - Cet objectif nécessitait de mobiliser des outils transversaux développés à l'Insee pour concevoir, programmer et gérer une enquête multimode.
 - Il impliquait également de modifier le rôle des différents acteurs de la collecte : enquêteurs, gestionnaires en région, pôle EEC (Enquête Emploi en Continu) qui gère la collecte au niveau national.
- **Produire une enquête garantissant un bon taux de réponse**, gage de précision de l'enquête, et assurant la qualité des indicateurs produits. Atteindre cet objectif suppose de limiter l'attrition, alors même que le lien enquêteur-enquêtés pourrait se distendre avec une

collecte pour partie auto-administrée. Cet objectif peut se détailler en plusieurs objectifs opérationnels :

- Assurer la bascule internet/téléphone dans les délais contraints de collecte et en restant aussi proche que possible de la période de référence pour limiter les effets de mémoire.
- Assurer la qualité en mode auto-administré en mettant en place une aide en ligne, en assurant un haut niveau d'assistance auprès des enquêtés, et en mettant en place des relances ciblées.
- Promouvoir et organiser la collecte multimode dans les établissements régionaux et avec le réseau des enquêteurs.
- Assurer les conditions permettant d'agréger les données collectées selon différents modes, corriger de la non-réponse et d'éventuels effets de mode.
- Fournir aux utilisateurs des données de l'enquête Emploi des séries rétropolées corrigées des ruptures de série sur les indicateurs principaux de l'enquête.
- **Optimiser la chaîne de production** en aval de la collecte et l'adapter au nouveau contexte (protocole multimode, nouveau questionnaire) :
 - Procéder à une rationalisation de la chaîne informatique de gestion en tenant compte des dernières évolutions et de l'écosystème informatique plus général à l'Insee.
 - Examiner le partage entre l'application de gestion et la chaîne aval.
 - Rationaliser et sécuriser la chaîne statistique aval en introduisant davantage de contrôles et en adaptant le langage aux directives générales en vigueur à l'Insee.

3. L'organisation du projet

La maîtrise d'ouvrage du projet a été assurée par le chef du département emploi et revenus d'activité à l'Insee. Dès 2016, avant le lancement officiel du projet, une équipe dédiée a été constituée pour assurer la maîtrise d'ouvrage déléguée et mener à bien le projet sous ses diverses dimensions. Le programme d'évolution de l'enquête Emploi comprenait 3 personnes dédiées à la fois au projet Muse et au projet Mélopée : un responsable de programme, un chef de projet statistique et un méthodologue. S'est ensuite adjoint un autre poste dédié à la refonte du questionnaire, au sein de la division Emploi.

Le programme a mené le projet d'évolution de l'enquête Emploi en étroite collaboration avec la division Emploi, responsable de l'enquête Emploi, et la division Synthèse conjoncturelle du marché du travail, responsable de la publication des chiffres conjoncturels sur l'emploi et le chômage, ainsi que des séries longues sur le marché du travail. L'équipe statistique a également été accompagnée par le pôle enquête Emploi en continu, en charge de la mise en œuvre opérationnelle de l'enquête Emploi, et les pôles d'expertise et de reprise de la nomenclature des professions et catégories socioprofessionnelles (PCS) et des diplômes et spécialités de formation. Le département des méthodes statistiques (division Sondages et division Recueil et traitement de l'information) a fourni également son appui méthodologique lors des différentes phases du projet.

L'équipe informatique du projet a été constituée d'un chef de projet informatique. Celui-ci s'est appuyé sur plusieurs équipes informatiques, celles en charge des outils Blaise (questionnaire), Capi (collecte assistée par informatique) et de la maintenance de l'application Emploi.

L'équipe projet, dans sa composante métier et informatique, a fonctionné en mode agile.

Par ailleurs, un comité directeur du programme a été mis en place pour valider les orientations du programme, suivre les développements sur les différents chantiers et prendre les décisions nécessaires sur les principaux enjeux. Il a été présidé par la directrice des statistiques démographiques et sociales, comprenait différents départements de l'Insee et associait la Dares, le service statistique du ministère en charge de l'emploi. Il a établi un niveau de coordination du programme avec Eurostat et avec les acteurs en interne responsables du développement des outils et procédures d'enquêtes multimode à l'Insee.

Le comité s'est appuyé sur plusieurs groupes de travail (GT) :

- Un **GT protocole**, qui visait à discuter des différents aspects de protocoles, principalement l'introduction d'Internet comme nouveau mode de collecte. Son mandat consistait à définir le mode de collecte aux différents rangs, encadrer la réponse par proxy, la réinterrogation des logements non résidences principales, la gestion de questionnaires incomplets ; préciser la gestion des fiches adresses, des contacts, des relances ; préciser l'organisation du travail des enquêteurs, des établissements régionaux, en lien avec le pôle EEC. Il a aussi été en charge d'assurer le lien avec les travaux de groupes plus généraux relatifs au multimode et à la maîtrise d'œuvre organisationnelle.
- Un **GT questionnaire**, qui visait à discuter du contenu du questionnaire. Son mandat consistait à produire entre 2017 et 2018 un questionnaire harmonisé sur les questions européennes et concerté avec les principaux utilisateurs sur les parties laissées à l'initiative nationale. Il y a eu 9 sous-GT thématiques : questionnaire logement et questions géographiques ; statut d'activité ; inscription à Pôle Emploi ; pluriactivité ; profession et secteur d'activité ; contrat et ancienneté professionnelle ; temps de travail et horaires ; formation ; revenus et allocations. Les participants étaient membres du service statistique public, experts sur les différents sujets traités.
- Un **GT rétopolation**, qui visait à préparer la stratégie d'estimation et de correction des ruptures de série. Son mandat consistait à prévoir les modalités des tests d'impact nécessaires, les calibrer et les inscrire dans le cadre de la programmation triennale. Il a ainsi proposé plusieurs scénarios d'enquête pilote à mener l'année précédant l'atterrissage de l'enquête.
- Un **GT EEC Mayotte**, qui visait à instruire les questions méthodologiques et organisationnelles liées à l'intégration du département de Mayotte dans l'enquête Emploi en continu. Son mandat était de construire et préparer la mise en place de l'enquête Emploi en continu à Mayotte après 2021, en remplacement de l'enquête annuelle existante (échantillonnage, organisation de la collecte en continu). Si ce volet a été reporté après l'atterrissage du projet Mélopée, les travaux ont été très utiles pour une mise en œuvre ultérieure, à compter de 2023.

- Un **GT pondération**, qui visait à s'assurer de respecter les contraintes européennes et à proposer des pistes de rénovation de la méthode de correction des poids de sondages. Son mandat consistait à définir une nouvelle stratégie de pondération utilisant toutes les informations disponibles dans la nouvelle base de sondage, et gérer les différents niveaux de non-réponse.

Le projet a été initié en 2016, de façon distincte du projet d'expérimentations Muse, pour atterrir en 2021, avec la mise en production en janvier de l'enquête rénovée. Les grandes lignes du calendrier ont été les suivantes :

- 2016 : mise en place des équipes
- 2017 : préparation du projet (expression des besoins, séminaire de lancement).
- 2018 : test en face-à-face du questionnaire de première interrogation auprès d'un échantillon de 1 000 personnes (juin) ; test par Internet du questionnaire de réinterrogation auprès de personnes volontaires (décembre).
- 2019 : test de l'enquête rénovée en deux vagues avec une première interrogation en face-à-face puis une réinterrogation en multimode auprès d'un échantillon de 1 400 ménages ; passage du pilote au comité du Label de la statistique publique.
- 2020 : pilote de grande envergure, en situation de pré-production ; passage de la nouvelle enquête Emploi au comité du Label.
- 2021 : entrée en production de la nouvelle enquête Emploi en continu au 1^{er} trimestre (T1) et arrêt de l'ancienne enquête au T2.

4. Un pilote d'envergure : pourquoi ?

L'expérience de la dernière refonte de l'enquête Emploi, qui a atterri en 2013, incite à ne pas négliger le risque induit par des modifications même jugées a priori mineures du questionnaire de l'enquête et à se donner les moyens de se préparer aux ruptures de série. Lors de la précédente refonte, le choix d'une bascule « one shot », c'est-à-dire d'une bascule de l'ancienne vers la nouvelle enquête sans pilote préalable avait été fait. L'exploitation des données du T1 2013 a cependant rapidement mis en évidence des ruptures de série. Les instructions nécessaires pour en déterminer l'origine (changement du questionnaire, problème informatique, conséquence de la baisse du taux de collecte consécutive à la mise en place concomitante des nouvelles conditions d'emploi des enquêteurs (NCEE)) ont conduit à un report de la publication du T1 2013. En effet, les délais de production de l'enquête Emploi sont particulièrement courts. Les instructions avaient conclu à un effet du changement de questionnaire. L'impact de la rupture de série a pu être quantifié *ex post*, et les séries ont été rétropolées. Il s'est avéré par ailleurs que cet impact avait un caractère saisonnier.

Lors de la refonte Mélopée, pour se préparer au mieux au passage à la nouvelle enquête et être en mesure de respecter le calendrier de publication, il a été rapidement acté de réaliser un Pilote,

consistant à passer un sous-échantillon en production en nouvelle enquête sur l'ensemble de l'année précédant l'atterrissage.

Un Pilote a plusieurs intérêts au premier rang desquels le fait de disposer des données nécessaires pour évaluer et corriger des ruptures de série liées à la refonte du questionnaire et de son protocole. Un Pilote a également pour intérêt de sécuriser le processus en amont de sa mise en production effective. De fait, il permet de s'assurer du bon fonctionnement des outils et de préparer les traitements aval pour limiter les délais de diffusion des résultats de la nouvelle enquête.

Ce Pilote a été financé en partie par Eurostat, ce dernier ayant incité les pays à mettre en œuvre des protocoles ambitieux pour être en mesure d'estimer des ruptures de série.

Le périmètre et le protocole du pilote ont été décidés de manière à assurer la mesure des ruptures de série tout en limitant les coûts de l'opération. Tout d'abord, le choix d'un pilote sur l'ensemble des trimestres de l'année 2020 a été fait pour être en mesure d'évaluer des impacts dont les indicateurs sont en partie soumis à des évolutions saisonnières. Une solution idéale consiste à réaliser deux enquêtes en parallèle, une en ancienne enquête et une en nouvelle enquête, avec des tailles identiques. Cependant, le coût d'une telle opération aurait été prohibitif. Pour limiter les coûts du pilote tout en assurant un niveau de précision suffisant pour évaluer des ruptures de série et pour assurer la diffusion des indicateurs en ancienne enquête en 2020, une partie de l'échantillon de l'ancienne enquête a été ponctionné pour passer en nouvelle enquête. Il a ainsi permis de comparer, pendant cinq trimestres consécutifs, l'ancienne et la nouvelle version de l'enquête.

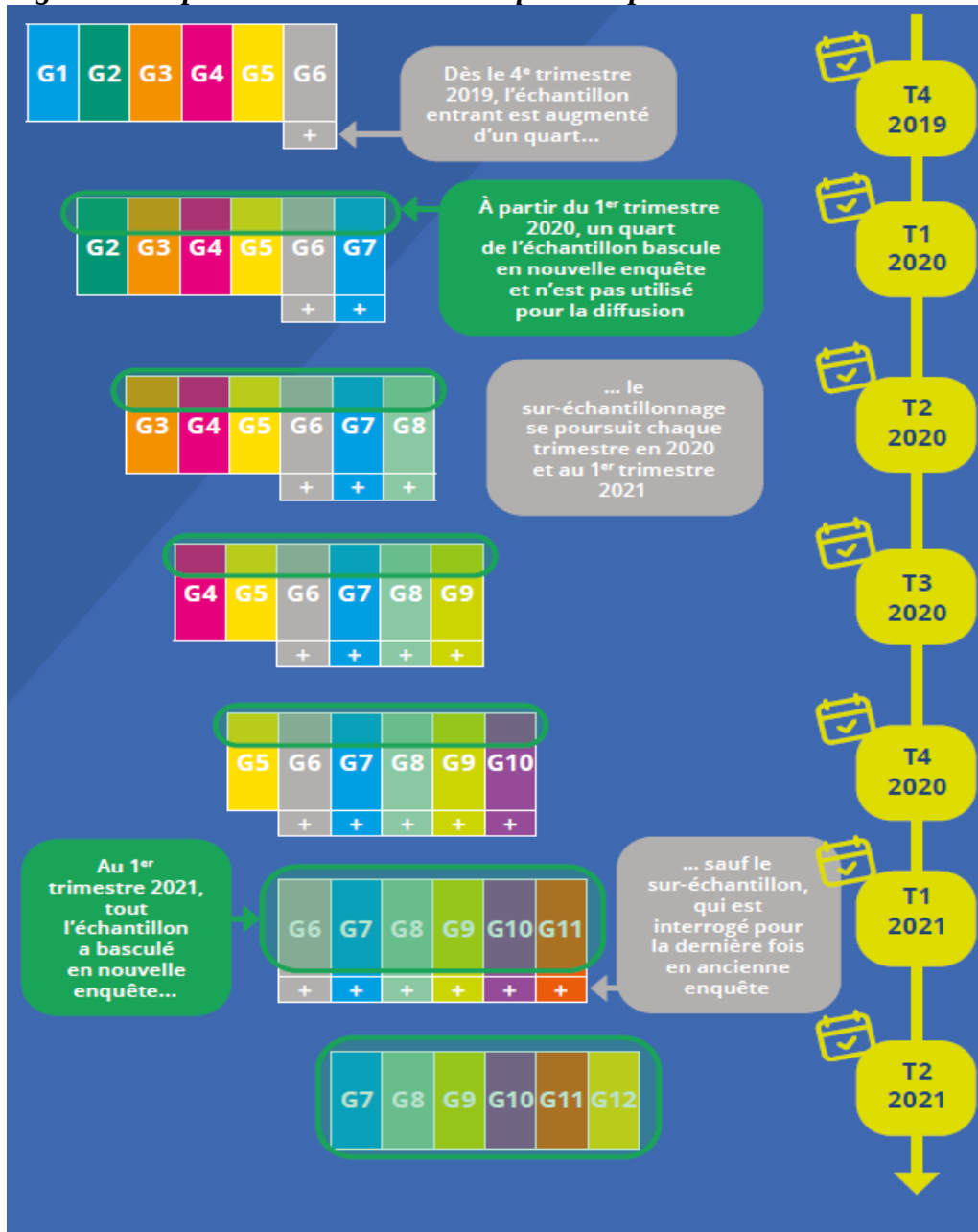
Concrètement, un quart de l'échantillon de chaque rang d'interrogation a basculé en nouvelle enquête au T1 2020 puis pour les échantillons entrants tout au long de l'année (*figure 2*). L'échantillon de l'enquête Emploi encore en cours a alors été amputé d'un quart par cette bascule en nouvelle enquête. Cependant, les échantillons entrants à partir du T4 2019 et jusqu'au T1 2021 ont été sur-échantillonnés d'un quart. Ce sur-échantillonnage a permis de compenser pour partie l'amputation de l'enquête en cours par le Pilote et de disposer d'un sur-échantillon complet réalisé en ancienne enquête au T1 2021.

Au premier trimestre 2021, la bascule en nouvelle enquête a été effectuée pour le reste de l'échantillon ne faisant pas partie du pilote en 2020. Les sur-échantillons tirés depuis le T4 2019 sont cependant restés en ancienne enquête pour disposer d'un trimestre supplémentaire de comparaison des deux enquêtes. Ces sur-échantillons n'ont plus été collectés à partir du T2 2021.

Ce type de Pilote comporte quelques risques par rapport aux objectifs : l'amputation d'un quart de l'échantillon en ancienne enquête conduit à une incertitude plus grande sur les résultats de l'enquête en 2020, résultats faisant référence sur cette période ; deux versions de questionnaires et de protocole sont sur le terrain en même temps, ce qui peut créer une éventuelle confusion et des difficultés dans la gestion de la collecte pour les enquêteurs et les gestionnaires en région ; on fait l'hypothèse que le fait d'avoir déjà répondu à l'enquête dans son ancienne version n'affecte pas la façon de répondre ultérieurement dans sa nouvelle version.

En revanche, il comporte de nombreux avantages : limiter le coût de l'opération ; disposer d'estimateurs en ancienne et en nouvelle enquête sur des échantillons et à une période comparable ; disposer d'estimateurs des probabilités de transition entre ancienne et nouvelle enquête pour les mêmes individus lorsqu'ils basculent de l'ancienne à la nouvelle enquête.

Figure 2 : Le pilote et la bascule de l'enquête Emploi en 2021



Source : « Une nouvelle enquête Emploi en 2021 », Courrier des Statistiques n°6, 2021.

Au total, le Pilote de l'enquête Emploi dans le projet Mélopée a donc été constitué :

- des 4 trimestres de 2020 en nouvelle enquête, sur l'équivalent d'1/4 de l'échantillon ;
- du 1^{er} trimestre de l'année 2021 en ancienne enquête sur 1/4 de l'échantillon également.

II. Les principales évolutions de l'enquête Emploi

1. La rénovation du questionnaire

Concernant le questionnaire de l'enquête Emploi, la refonte visait trois objectifs principaux.

Le premier était de **se conformer au nouveau règlement européen IESS** qui visait lui-même :

- une standardisation accrue entre les enquêtes ménages sous règlement européen. Le règlement définit ainsi un jeu de variables standardisées communes à différentes enquêtes et des concepts communs de logements ordinaires, de résidence habituelle et de ménages.
- une harmonisation accrue entre les LFS des différents pays, qui prend une forme plus contraignante : non seulement une liste de variables qui doivent être transmises, mais aussi des prescriptions sur la façon dont les informations doivent être obtenues (respect de l'enchaînement des questions).
- une intégration plus poussée des modules ad hoc qui deviennent des variables à longue périodicité : IESS a retenu 6 thèmes qui seront abordés tous les 8 ans ; deux créneaux sont réservés à des thèmes ponctuels. Les thèmes récurrents retenus sont : organisation du travail et aménagement du temps de travail (2019), accidents au travail et autres problèmes de santé liés au travail (2020), situation sur le marché du travail des immigrés et de leurs descendants (2021), retraite et participation au marché du travail (2023), les jeunes sur le marché du travail (2024), conciliation entre vie professionnelle et vie familiale (2025).

Le second objectif était de **prendre en considération les attentes françaises et internationales**, dès lors qu'elles sont compatibles avec les exigences européennes. De fait, au-delà des questions nécessaires pour les variables européennes, le questionnaire comporte un certain nombre de questions spécifiquement françaises. Pour garantir une continuité des séries, les évolutions du questionnaire sont en général limitées et ont lieu à l'occasion de refontes plus larges de l'enquête. Les principales attentes nationales portaient sur les conditions de travail, les nouvelles formes de l'emploi en lien avec les recommandations du Cnis (Gazier *et alii*, 2016), la formation non formelle à but professionnel et la mise en œuvre de la nouvelle nomenclature des professions et catégories socio-professionnelles (PCS) à la suite de sa rénovation (Amossé, 2020). Par ailleurs, au niveau international, l'objectif était d'anticiper autant que possible la mise en œuvre de la nouvelle classification internationale de la situation dans la profession (CISP, ou *International Classification of Status in Employment* (ICSE)) qui vise à mieux appréhender les nouvelles formes d'emploi, à la frontière entre salariat et non-salariat (en définissant notamment les *dependent contractors* – indépendants économiquement dépendants).

Enfin, le troisième objectif était de **développer un questionnaire adapté aux différents modes de collecte**, notamment à l'auto-administré sur Internet. Cela invitait à concevoir un questionnaire plus court et plus simple. Les principaux changements en ce sens ont porté sur le recueil de la profession et du diplôme, basé jusque-là sur une codification embarquée nécessitant une formation des enquêteurs, et donc peu compatible avec une interrogation en autoadministré. Des choix ont également été faits pour alléger le questionnaire de ré-interrogation pour faciliter et fluidifier la passation du questionnaire par Internet. Le principe a été posé de disposer d'un seul questionnaire quel que soit le mode de collecte, pour limiter d'éventuels effets de mode.

Pour élaborer ce nouveau questionnaire, un groupe de travail dédié a été mis en place, décliné en plusieurs sous-groupes thématiques, regroupant les experts du service statistique public (SSP) en la matière. Au préalable, des réunions de concertation ont eu lieu pour recueillir les besoins et les attentes sur la nouvelle enquête Emploi auprès d'un public plus large que celui du SSP. Ces réunions ont fait appel d'une part aux partenaires sociaux et d'autre part aux chercheurs utilisateurs de l'enquête Emploi.

Par ailleurs, pour influencer au maximum sur les contraintes européennes, la France a participé à l'ensemble des task force sur le sujet. Cette implication est importante pour faire converger les attentes nationales avec les impératifs européens.

Globalement, les concepts d'emploi, de chômage et d'inactivité demeurent ceux définis par la résolution de la 13^e Conférence internationale des statisticiens du travail (CIST) de 1982. Cependant, avec IESS, Eurostat en propose une définition opérationnelle renouvelée plus précise et impose un schéma de questionnaire encadrant la collecte des informations nécessaires pour reconstruire le statut d'activité, imposant désormais une harmonisation par les *inputs*. La population de référence, par le passé les personnes âgées de 15 ans ou plus, a été limitée aux personnes âgées de 15 à 89 ans. Ainsi, il n'y a plus de questionnaire individuel pour les personnes âgées de 90 ans ou plus. Pour ces personnes, seules les informations recueillies dans le questionnaire logement (sexe, âge, situation de logement, situation familiale...) sont disponibles ; leur statut d'activité est imputé à inactif.

En matière d'architecture, le questionnaire reste structuré autour d'un questionnaire logement, qui recueille des informations relatives au repérage, vérifie que le logement est bien dans le champ et recense les individus vivant dans le logement, et des questionnaires individuels, décrivant la situation de tous les individus âgés de 15 à 89 ans (*figure 3*). Ce questionnaire individuel reste organisé autour de différents modules dont l'architecture reste similaire à la précédente, hormis pour le module E (sur la situation un an auparavant) qui est supprimé. En effet, les variables relatives à la situation un an auparavant, qui peuvent être reconstruites par une exploitation longitudinale de l'enquête, ne sont plus requises par Eurostat.

Figure 3 : Structure du questionnaire de l'enquête Emploi avant et après refonte



Source : « Une nouvelle enquête Emploi en 2021 », Courrier des Statistiques n°6, 2021.

Les principaux changements sur le questionnaire sont présentés dans la suite par module

- Questionnaire logement : les principales modifications ont consisté à harmoniser le questionnement avec d'une part celui du tronc commun des ménages utilisé par les autres enquêtes de l'Insee et d'autre part les règles du recensement (questionnement de repérage, règles de rattachement, définition de la personne de référence, gestion de l'homoparentalité, relation de couple hors logement, etc.). Il ajoute également la notion de budget séparé pour définir les ménages, pour se conformer aux concepts standardisés européens.
- Module A sur la situation sur le marché du travail : Ce module est le plus central car il vise à déterminer le statut d'activité au sens du BIT : emploi, chômage, inactivité. Il s'agit ainsi du module le plus soumis à l'harmonisation européenne. Eurostat impose désormais les

questions et l'ordonnement de ces questions et plus seulement une harmonisation par les outputs (*figure 4*). Cependant, les changements de questions et de déroulé sont relativement limités par rapport à l'ancienne version de l'enquête Emploi en France. Les nouvelles définitions d'activité, d'emploi et de chômage sont désormais précisées dans une note méthodologique².

- Le changement principal est lié à une clarification par Eurostat de l'interprétation des critères du BIT qui conduit à modifier le périmètre de l'emploi. En effet, pour améliorer la comparabilité entre les pays, Eurostat propose des définitions opérationnelles des concepts établis par le BIT avec des critères plus précis encore pour classer la population entre emploi, chômage ou inactivité. Ces critères concernent, par exemple, les raisons d'absence pour les personnes ayant un emploi mais n'ayant pas travaillé pendant une semaine donnée, dite de référence, ou encore une liste précise de démarches de recherche d'emploi pour les personnes sans emploi. Avec l'entrée en vigueur du nouveau règlement européen, sont désormais classées en emploi les personnes qui déclarent avoir un emploi mais en être absentes en raison d'un congé maladie, et ce quelle que soit la durée de l'absence. Précédemment, elles n'étaient considérées en emploi qu'en cas d'absence prévisionnelle d'un an ou moins. De même, les personnes se déclarant absentes de leur emploi en raison d'un congé parental sont désormais classées en emploi si leur absence prévisionnelle est inférieure ou égale à 3 mois ou si elles perçoivent un revenu compensatoire, comme la prestation partagée d'éducation de l'enfant (Prepa) et dans ce cas, sans condition de durée. Précédemment, elles n'étaient considérées en emploi qu'en cas d'absence de 3 mois ou moins. Dans les deux cas, les évolutions conduisent à classer davantage de personnes en emploi qu'avant.
- Concernant les personnes non classées en emploi, les questions sur la recherche d'emploi et le souhait d'avoir un emploi sont inversées, revenant à la situation qui prévalait avant 2013. Les démarches de recherche d'emploi sont également moins nombreuses avec un questionnement plus restreint qui a été précisé par le règlement. Par ailleurs, les questions de rattrapage sont reformulées, notamment pour donner davantage d'exemples sur les petits boulots.
- Concernant les personnes en emploi, le module A se conclut désormais par une question visant à déterminer le nombre d'emplois, un emploi étant défini par une profession sous un statut (salarié / non-salarié) donné et pour un employeur donné (pour les salariés). La pluriactivité est donc désormais identifiée par une question unique et directe, conformément aux nouvelles recommandations européennes.
- Module B sur la description des emplois actuels : Les informations requises par Eurostat portent sur l'emploi principal et l'emploi secondaire. L'architecture du module B est simplifiée en décrivant d'abord l'emploi principal, puis l'emploi secondaire, avec des contraintes moins fortes en matière de traitement de la multiactivité.
 - De manière générale, la façon de repérer les personnes en emploi ou le fait qu'elles considèrent leur emploi comme un « petit boulot » ne filtre plus le questionnaire : l'ensemble des personnes en emploi sont invitées à décrire l'emploi principal exercé pendant la semaine de référence. La question subjective visant à qualifier l'emploi

2 <https://www.insee.fr/fr/metadonnees/source/operation/s2022/documentation-methodologique>

comme une « activité régulière ou plutôt une activité temporaire ou d'appoint » est supprimée, de nombreuses questions plus factuelles permettant de décrire l'emploi.

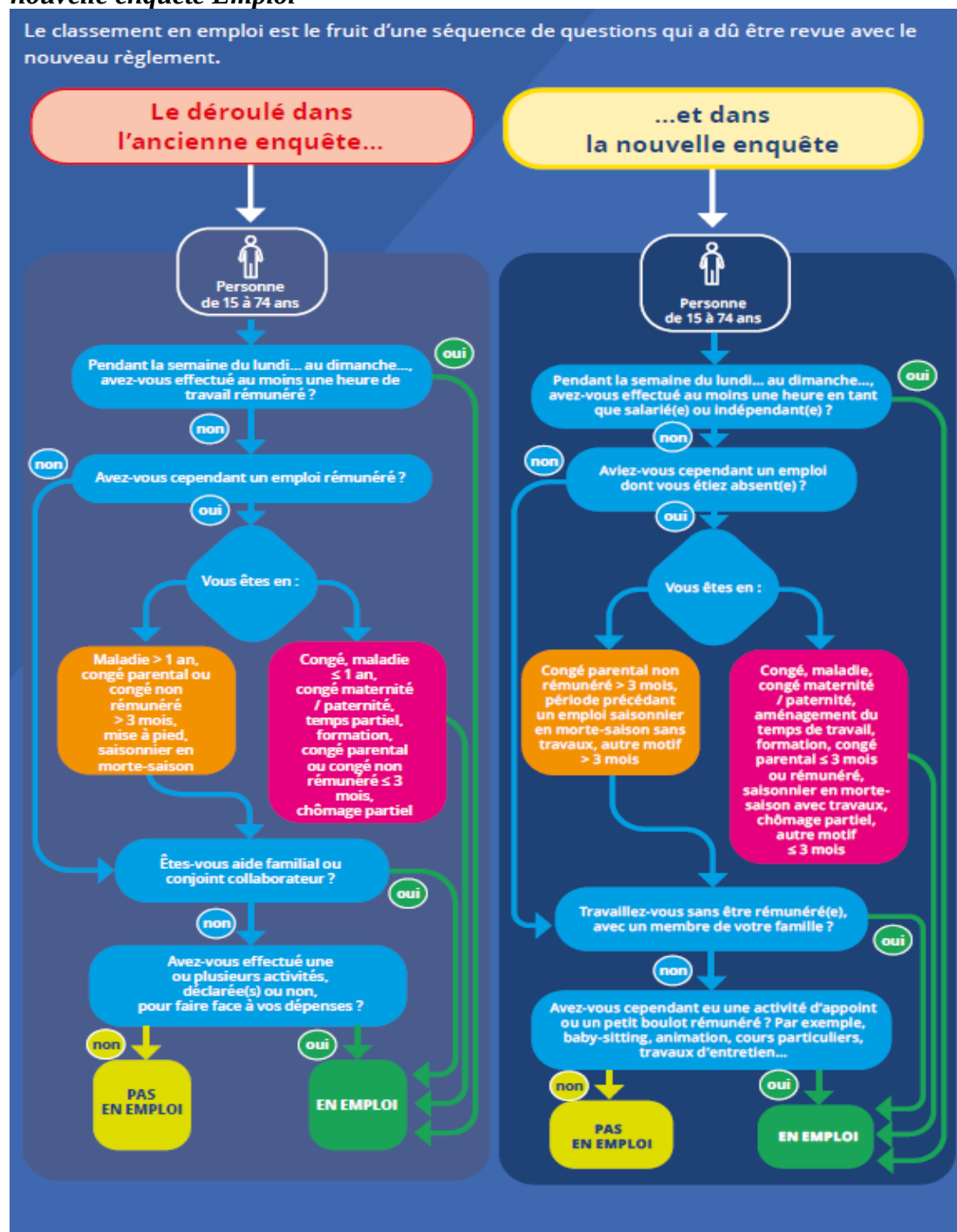
- Concernant la profession, le statut et l'employeur, la principale évolution concerne le mode de recueil du libellé de profession. D'une saisie en clair, peu normalisée et qui nécessite une formation des enquêteurs, on passe à une saisie sur liste, adaptée à l'auto-administré et permettant de recueillir des libellés plus riches et normalisés. Les évolutions de cette partie du questionnaire se sont appuyées sur les réflexions menées dans le cadre de la refonte de la PCS 2020³.
- Concernant le contrat de travail pour les salariés, le repérage des différentes formes de contrat est amélioré suivant les recommandations du rapport du Cnis sur la diversité des formes d'emploi. La date de début dans l'entreprise est demandée en référence à la date de début du contrat, permettant de mieux identifier les cas de récurrence (personnes en contrat en durée limitée déjà en contrat dans l'entreprise le mois précédant le début du contrat en cours). Les circonstances de fin de l'emploi précédent sont demandées pour l'ensemble des personnes ayant connu une fin d'emploi, y compris celles qui ont pu connaître une période de chômage ou d'inactivité entre-temps.
- Concernant le temps de travail et l'organisation du travail, l'ordre des questions est revu pour limiter les changements de période de référence et des informations nouvelles sont collectées, pour répondre à des demandes européennes (heures contractuelles, qui décide des heures de travail) ou nationales (télétravail).
- La notion de dépendance économique des indépendants est mieux appréhendée pour anticiper les réflexions sur les nouvelles formes de l'emploi et la future mise en œuvre de la refonte de la classification internationale de la situation dans la profession.
- Module C sur l'emploi antérieur : Le champ est désormais limité aux personnes qui ne sont pas classées en emploi pendant la semaine de référence, en complément du module B. Un changement conceptuel important a été apporté par le règlement puisque les expériences professionnelles qui se limitent à des petits boulots sont désormais incluses comme expérience professionnelle passée. Des questions sur le statut de micro-entrepreneur et sur le contrat de travail (type et durée) sont ajoutées.
- Module D et E sur la formation : Les contraintes européennes relatives à cette partie ont évolué en dédoublant les périodes de référence (4 dernières semaines / 12 derniers mois) et en réduisant les informations requises sur les formations non formelles. Le module est ainsi simplifié, tout en améliorant le repérage des formations formelles pour les plus jeunes. Ainsi, le diplôme est désormais recueilli sur liste contre une saisie en clair auparavant. La spécialité de la formation formelle en cours n'est plus demandée. Par ailleurs, une séquence de questions sur les expériences professionnelles pendant le cursus menant au plus haut diplôme a été ajoutée à la demande d'Eurostat.
- Module F sur les allocations : Le questionnement est nettement allégé, notamment en matière de recueil des montants reçus.
- Module G sur la santé : Ce module, composé de quelques questions, est très contraint par le nouveau règlement européen puisqu'il est constitué de variables standardisées à l'ensemble des enquêtes auprès des ménages.

3 Pour plus d'informations sur cette refonte, voir le site dédié à la nomenclature des PCS : <https://www.nomenclature-pcs.fr/>

- Module H sur l'origine géographique et sociale : Des questions sur l'histoire migratoire (raison de la migration, pays de résidence antérieure), désormais requises par Eurostat, ont été ajoutées. Le codage de la profession des parents est amélioré en s'inspirant du questionnement de l'enquête Formation et Qualification Professionnelle. Le questionnement permet également de gérer les situations d'homoparentalité.
- Module I sur la situation principale : Ce module a été simplifié pour se concentrer sur la situation déclarée de l'enquêté et des questions générales sur l'évolution de la situation sur le marché du travail dans les 12 derniers mois. Des questions sur le statut de micro-entrepreneur et sur le contrat ont été ajoutées pour l'emploi occupé un an auparavant.

Un autre changement dans le questionnaire concerne le sous-échantillonnage puisque toutes les variables requises au niveau européen ne sont pas demandées à la même fréquence. Ainsi les modules complémentaires sont désormais des variables *8-yearly* (définies comme des variables annuelles, et non trimestrielles, une année sur 8). En France, le choix a été fait de collecter les modules complémentaires en grappes entrantes uniquement.

Figure 4 : Déroulé du questionnement sur le statut d'activité dans l'ancienne et la nouvelle enquête Emploi



Source : « Une nouvelle enquête Emploi en 2021 », Courrier des Statistiques n°6, 2021.

2. La modernisation du protocole

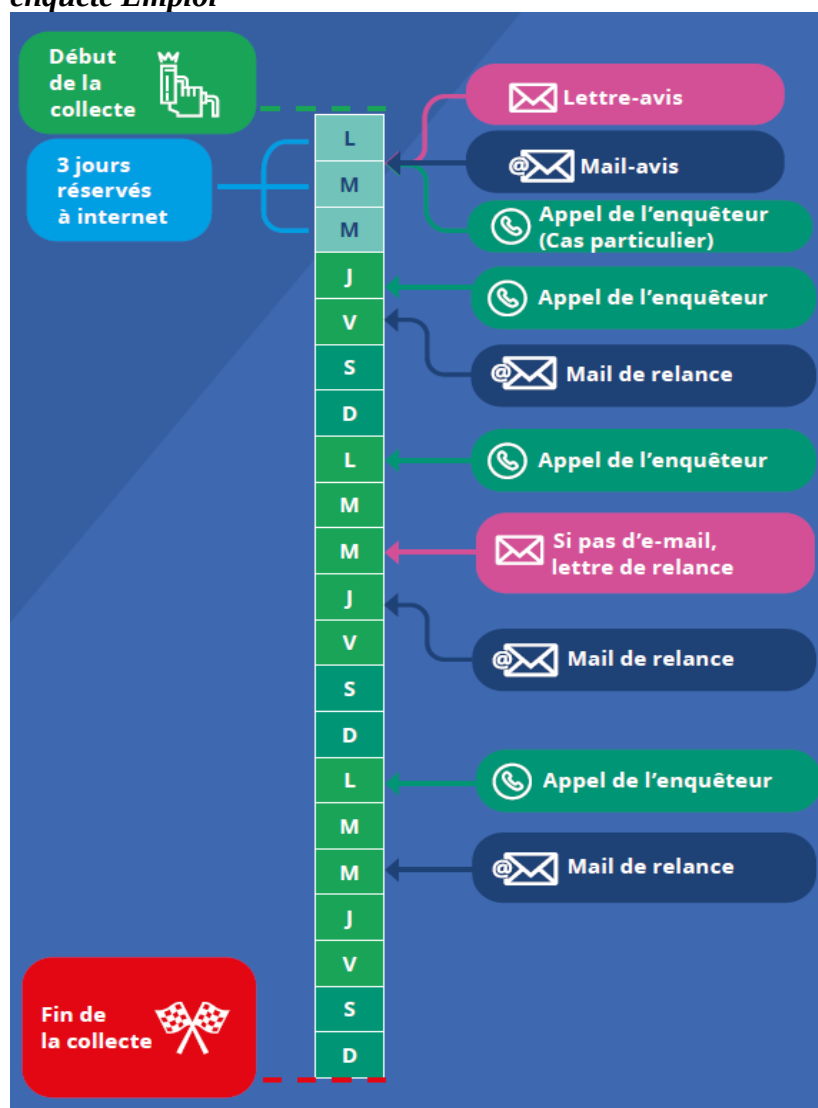
2.1. Un protocole multimode en réinterrogation

La refonte majeure du protocole consiste en l'introduction du mode de collecte Internet pour répondre lors des réinterrogations nécessitant la mise en œuvre d'un protocole multimode en réinterrogation.

Le protocole en première interrogation n'est pas modifié et repose sur un entretien en face à face avec une phase de repérage des logements à enquêter. Ce choix a été fait à la suite des expérimentations Muse, pour permettre une première interrogation plus longue, améliorer la compréhension des concepts pour les interrogations suivantes et fidéliser les enquêtés pour les réinterrogations.

Le protocole de collecte en réinterrogation est quant à lui profondément modifié. Tout d'abord, le dernier rang d'interrogation suit désormais le même protocole que les vagues intermédiaires alors que jusqu'à présent, en ancienne enquête il suivait le protocole de la première interrogation (face à face). Pour l'ensemble des réinterrogations, un protocole multimode Internet-téléphone est proposé (*figure 5*). Dans la plupart des cas, les 3 premiers jours de collecte sont réservés à la collecte par Internet, de façon à décharger les enquêteurs des réponses que certains enquêtés apportent sans difficulté et sans délai par eux-mêmes. À partir du jeudi de la première semaine de collecte, les enquêteurs interviennent pour la collecte par téléphone. Commence alors un protocole multimode concurrentiel, puisque la collecte par Internet reste possible jusqu'au dernier jour de collecte.

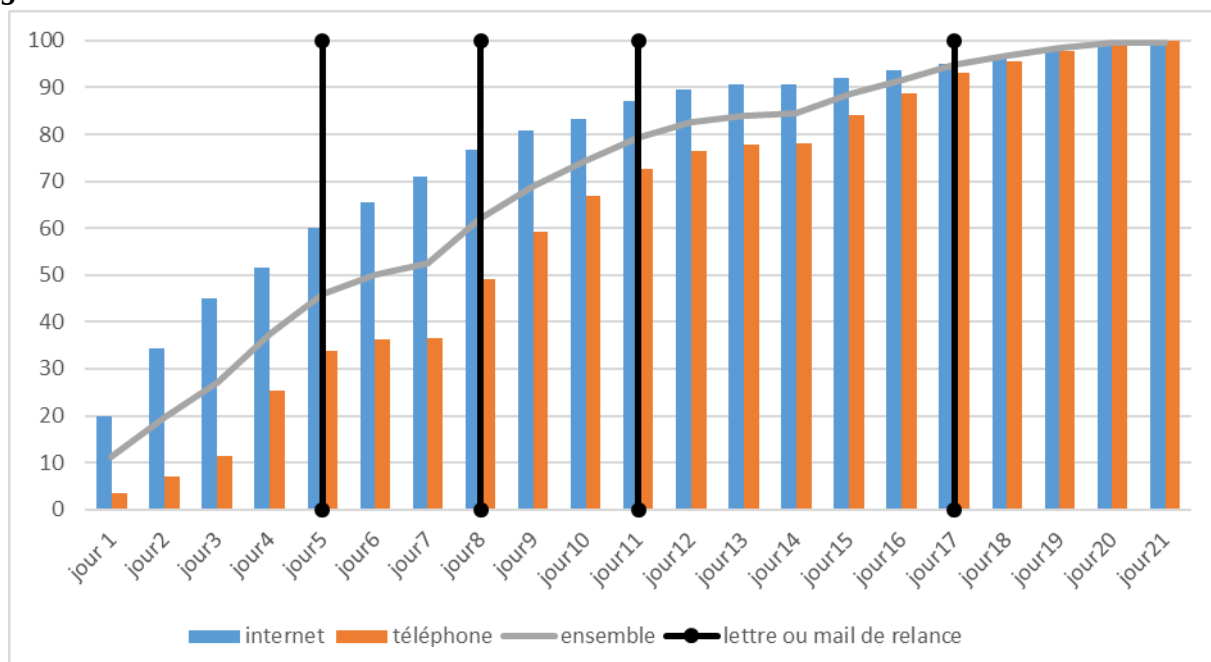
Figure 5 : Le protocole de réinterrogation dans la nouvelle enquête Emploi



Source : « Une nouvelle enquête Emploi en 2021 », Courrier des Statistiques n°6, 2021.

Pour laisser un minimum de temps à la réponse Internet sans pour autant mettre en péril la relance téléphonique et le taux de réponse, la durée de collecte est allongée et passe de 2 semaines et 3 jours à 3 semaines complètes. Cette durée reste inférieure à la limite fixée par Eurostat de 5 semaines. Le choix des 3 premiers jours réservés à la collecte par Internet provient des enseignements du projet Muse. De fait, les enquêtés par Internet sont très réactifs lorsqu'ils reçoivent une invitation à répondre ou une relance, notamment lorsqu'elle est faite par mail avec un lien direct vers le questionnaire à renseigner. Mais le taux de réponse diminue très rapidement après ces relances (les relances ont lieu aux jours 5, 11 et 17 de la collecte). Ce choix s'est avéré concluant puisque près de la moitié des questionnaires renseignés par Internet le sont dans les 3 premiers jours, 70 % le sont dans la première semaine (*figures 6a et 6b*).

Figure 6a : Jour de collecte selon le mode utilisé en 2021

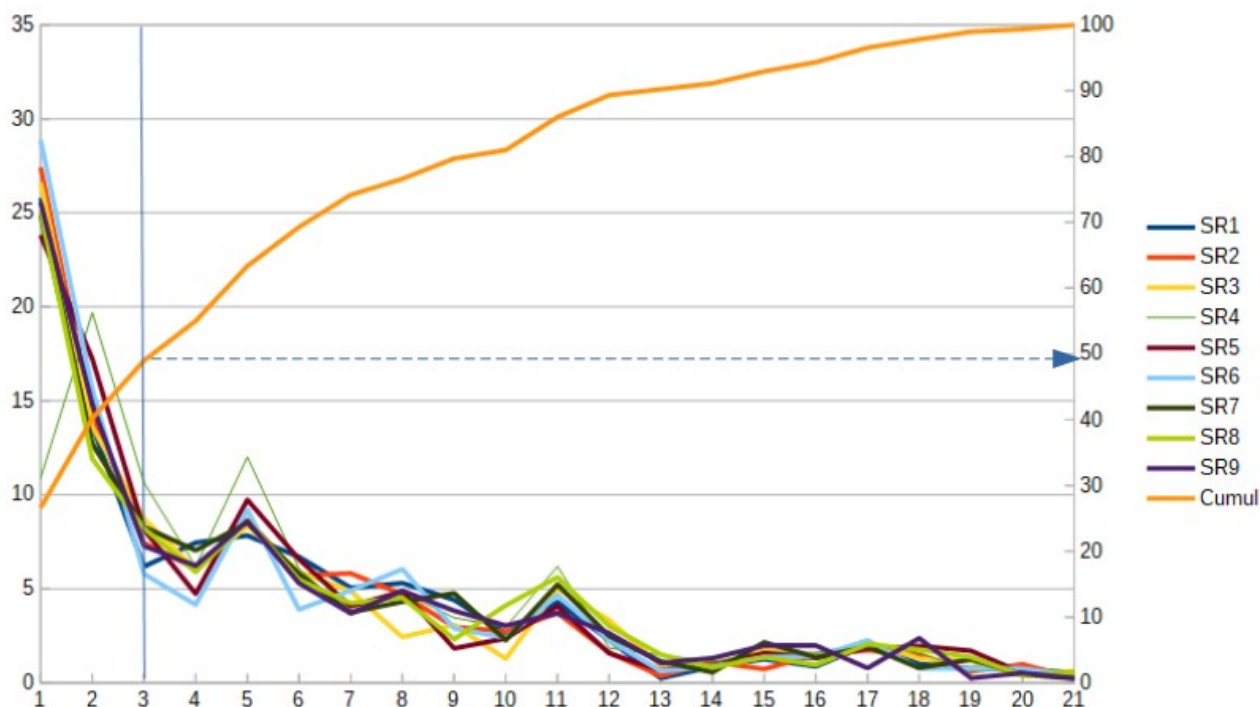


Lecture : 45 % des logements enquêtés par internet l'ont été après 3 jours de collecte.

Source : Insee, enquête Emploi en continu 2021.

Champ : ensemble des logements enquêtés en réinterrogation.

Figure 6b : Jour de collecte par Internet au premier trimestre 2021



Lecture : Au bout de 3 jours de collecte, 50 % des logements enquêtés par Internet ont répondu. Pour la 1^{re} semaine de référence (SR1), 25 % des logements enquêtés par Internet répondent le 1^{er} jour de la collecte (échelle de gauche). Le cumul de la collecte réalisée par Internet est représenté sur l'échelle de droite.

Source : Insee, enquête Emploi en continu T1 2021.

Champ : ensemble des logements enquêtés par Internet en réinterrogation.

Dans certains cas, il n'est pas utile de laisser ces 3 premiers jours exclusivement à la réponse par Internet. Trois raisons sont identifiées et autorisent les enquêteurs à intervenir dès le premier jour de collecte pour ne pas perdre de temps :

- le ménage a indiqué à l'interrogation précédente qu'il ne souhaitait/pouvait pas répondre par Internet (cas 1). Cette situation concerne 6,1 % des fiches-adresses (FA) en collecte en réinterrogation en 2021 (figure 7). De fait, seulement 0,1 % de ces ménages répondent effectivement par Internet lors de l'interrogation suivante. Il ne sert donc à rien de laisser 3 jours de collecte au mode de collecte Internet pour ces ménages.
- un questionnaire réalisé par Internet est détecté en apurement comme de mauvaise qualité (cas 2). Cette qualification est de la responsabilité des gestionnaires en région. Les questionnaires internet réalisés partiellement sont automatiquement affectés à cette situation. Dans ce cas, on souhaite éviter que le ménage réponde de nouveau par Internet. Cette situation concerne 1,2 % des FA en réinterrogation en 2021. Cependant, dans les faits, près de la moitié de ces ménages continuent à répondre par Internet le trimestre suivant. Des modifications, par exemple en rendant impossible la réponse par Internet – éventuellement sur une partie de la durée de la collecte –, sont à envisager pour améliorer ce protocole.
- en cas d'absence anticipée de l'enquêteur sur une partie de la période de collecte, le gestionnaire en région peut demander à ce que l'enquêteur puisse intervenir dès le premier jour de collecte sur la grappe concernée pour qu'il dispose de suffisamment de temps pour réaliser un maximum d'entretiens (cas 3). Cette situation concerne 6,9 % des FA en réinterrogation en 2021. On peut voir que cette gestion conduit à des résultats similaires à ceux pour lesquels s'applique la règle des 3 jours réservés à internet : la part d'enquêtes non réalisées est relativement similaire, comme la répartition des enquêtes réalisées par internet ou par un enquêteur.

Figure 7 : Devenir des FA de réinterrogation en fonction des cas identifiés le trimestre précédent

	Résultat le trimestre suivant				Ensemble
	Part	Enquête réalisée par Internet	Enquête réalisée par enquêteur	Enquête non réalisée	
Cas général	85,8 %	40,7 %	47,4 %	11,9 %	100 %
Cas 1 (demande de l'enquêté)	6,1 %	0,1 %	91,6 %	8,3 %	100 %
Cas 2 (mauvaise qualité)	1,2 %	45,8 %	32,8 %	21,4 %	100 %
Cas 3 (raison d'organisation)	6,9 %	37,4 %	50,0 %	12,6 %	100 %
Ensemble des FA de réinterrogation	100 %	38,0 %	50,1 %	11,9 %	100 %

Lecture : 6,9 % des fiche-adresses (FA) en 2021 sont dans le protocole correspondant au cas 3 (intervention de l'enquêteur dès le premier jour pour raison d'organisation). Parmi elles, 37,4 % donnent lieu à une enquête réalisée par internet le trimestre suivant.

Source : Insee, enquête Emploi en continu 2021.

Champ : ensemble des fiches-adresse (FA) remises en collecte en réinterrogation.

D'un point de vue technique et organisationnel, proposer un multimode concurrentiel avec un mode auto-administré et un mode intermédié est relativement complexe et nécessite une très bonne communication et diffusion de l'information. D'une part pour que les enquêteurs disposent de l'information nécessaire sur l'état d'avancement du questionnaire. D'autre part pour éviter des relances à tort auprès d'enquêtés qui auraient déjà répondu. Cet objectif de mettre en place un protocole multimode concurrentiel était particulièrement ambitieux et constituait un des enjeux majeurs du projet d'un point de vue informatique. Ce point est détaillé dans la partie 5.

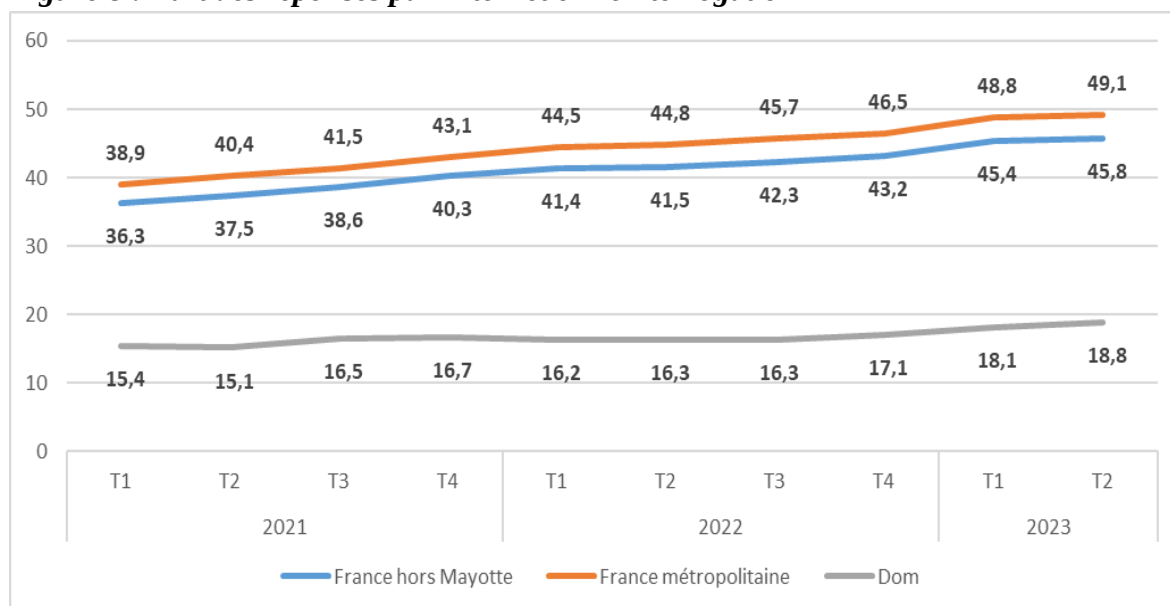
Au-delà de l'aspect informatique, ce nouveau protocole modifie le travail des différents acteurs de la collecte, au premier rang desquels les enquêteurs. De fait, l'enquêteur est désormais responsable de convaincre les enquêtés de répondre par Internet et les relancer dans ce but. Même si l'assistance aux enquêtés sur Internet est confiée à Insee contact, l'enquêteur, ayant tissé une relation de confiance en première interrogation, est souvent sollicité dans les faits. La gestion de sa collecte dépend de davantage de facteurs qu'auparavant. Pour cela il dispose d'informations dans le poste de collecte et alimentées par des flux réguliers : pour savoir s'il peut intervenir ou non dès le 1^{er} jour de collecte selon les FA ; pour connaître l'état d'avancement des questionnaires sur Internet ; pour connaître les dates des courriers et mails d'annonce et de relance. L'enquête Emploi est la première enquête à l'Insee à proposer un protocole multimode concurrentiel.

À noter aussi qu'au moment de la refonte, les enquêtes commencées par Internet devaient être recommencées dans leur intégralité au début lorsque l'enquêteur reprenait la main, or les questionnaires étaient parfois partiellement remplis. À partir de janvier 2023, une nouvelle fonctionnalité est disponible et les enquêteurs peuvent récupérer les informations saisies par les enquêtés et finaliser l'enquête sans devoir recommencer au début, ce qui diminue nettement la charge des enquêtés dans cette situation et facilite le travail des enquêteurs.

Le travail des gestionnaires dans les établissements régionaux est également modifié : l'envoi des courriers et mails d'annonce et de relances est centralisé et automatisé ; le cadencement des relances par mail impose un apurement au fil de l'eau ; le poste de gestion est enrichi de toutes les informations de collecte (quel que soit le mode) avec de nouvelles fonctionnalités pour un suivi regroupé dans l'application dédiée à l'enquête Emploi. Par ailleurs, le pôle EEC joue encore plus qu'auparavant un rôle de chef d'orchestre de la collecte et constitue le point de contact privilégié des gestionnaires en région.

L'impact de l'introduction d'Internet est encore en cours d'étude mais de premiers résultats peuvent être fournis. Tout d'abord, le taux de collecte par Internet parmi les FA en réinterrogation est nettement plus élevé qu'attendu, et a augmenté de façon régulière : en moyenne au premier trimestre 2023, 45,4 % des réinterrogations sont réalisées par Internet soit 9 points de plus qu'au premier trimestre 2021 (*figure 8*). Ce taux est nettement moindre dans les DOM (18,1 %), mais pas moins qu'attendu.

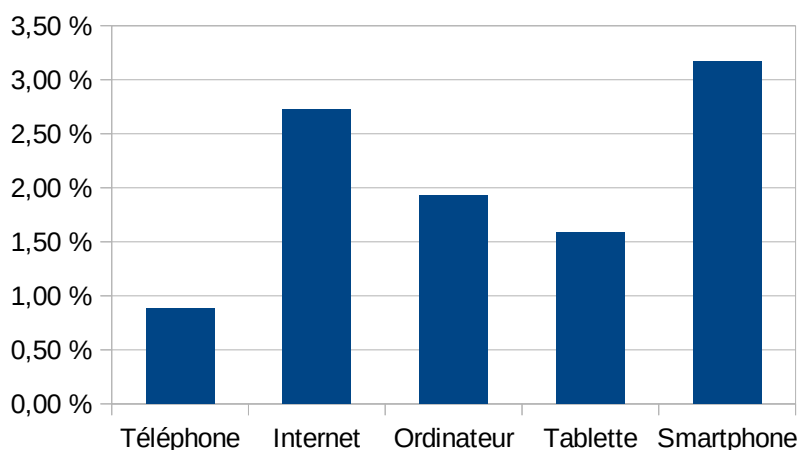
Figure 8 : Part des réponses par Internet en réinterrogation



Lecture : au T1 2021, 36,3 % des FA de réinterrogation sont réalisées par internet en France hors Mayotte.
 Source : Insee, enquête Emploi en continu.
 Champ : ensemble des logements enquêtés en réinterrogation.

La réponse par Internet conduit à davantage de non-réponse partielle (*figure 9*). Alors qu'elle est très rare au téléphone (moins de 1 % des enquêtes), elle est trois fois plus importante par Internet. C'est notamment le cas pour les réponses par smartphone, ce qui peut signaler des problèmes techniques spécifiques.

Figure 9 : Part de la non-réponse partielle selon le mode de collecte et l'outil utilisé



Lecture : en 2021, 2,75 % des logements enquêtés par internet font l'objet d'une non-réponse partielle.
 Source : Insee, enquête Emploi en continu 2021.
 Champ : ensemble des logements enquêtés en réinterrogation.

La réponse par smartphone est un enjeu important et constitue un défi technique à part entière. De fait, la réponse sur smartphone représente un quart des réponses par Internet fin 2021. Or la collecte sur cet outil nécessite que le questionnaire soit *responsive design*, c'est-à-dire qu'il s'adapte à la taille de l'écran. Bien que le questionnaire ait été revu pour s'adapter à la réponse par Internet, des questions longues ou avec beaucoup de modalités persistent. Surtout, certaines questions nécessitent

de naviguer dans une liste ; or la collecte actuelle de l'enquête pose encore des difficultés sur certains navigateurs relatives à cette collecte sur liste, ce qui peut induire des abandons plus fréquents sur smartphone.

Un autre effet de la collecte par Internet est une baisse de la réponse par proxy (un autre individu du ménage répond à la place de l'individu à enquêter). En 2021, elle représente 22 % des questionnaires individuels contre 25 % en 2019. Cette différence est essentiellement liée aux réinterrogations et au fait que le recours au proxy est nettement moins fréquent sur Internet (12 % en réinterrogation en 2021) que par téléphone (33 %). De fait, le mode de collecte est plus souple que le téléphone et permet plus facilement de répondre en plusieurs fois par différents individus du ménage. Cet effet de la collecte par Internet est bénéfique puisqu'on peut supposer des données collectées de meilleure qualité lorsque l'individu répond pour sa situation. La baisse de la réponse par proxy est particulièrement notable pour les jeunes : la part des questionnaires réalisés par un proxy pour les 15-24 ans passe de 65 à 53 % en nouvelle enquête (Dera, 2021). Cette baisse peut être attribuée à la réponse par Internet mais également aux changements de règles de rattachement pour les étudiants (voir partie 3). Or les jeunes ont plus souvent des petits boulots dont les parents (le plus souvent le proxy est un parent du jeune) n'ont pas forcément connaissance. Ainsi, parmi les jeunes qui se déclarent spontanément étudiants, le taux d'emploi est relevé de 5 points du fait de cette baisse de proxy. En outre, les jeunes sont les mieux placés pour exprimer leurs souhaits ou préciser leur recherche d'emploi, variables déterminantes pour les classer au regard du chômage ou du halo pour ceux qui n'ont pas d'emploi.

2.2. Les règles de non-remise en collecte

Dans l'enquête Emploi, certaines fiches-adresses (FA) ne sont pas remises en collecte aux vagues intermédiaires et sortantes, soit qu'on considère que le logement est durablement hors champ, soit que la situation du ménage est supposée ne plus changer. Les règles de remise en collecte ont été revues lors de la refonte de l'enquête Emploi en 2021, en lien avec de nouvelles règles autorisées par Eurostat, qui ne prescrivait aucune règle auparavant.

Eurostat autorise désormais de ne plus réinterroger deux catégories de population et d'imputer leur réponse à l'identique de la première interrogation : les personnes inactives âgées de 70 ans ou plus et les personnes inactives en situation de handicap.

Au-delà de ces règles, le changement de protocole des vagues sortantes, désormais traitées comme des vagues intermédiaires avec un protocole multimode Internet-Téléphone, conduit à modifier les règles de non remises en collecte. De fait, avec une interrogation en face à face pour les vagues sortantes auparavant, la remise en collecte de logements antérieurement hors champ était peu coûteuse du fait de la structure en grappe de l'échantillon qui limite les coûts de repérage. Ce n'est plus le cas avec un protocole multimode qui ne nécessite plus de déplacement sur le terrain en vague sortante.

L'objectif des nouvelles règles de remise en collecte consiste donc à remettre en collecte les FA dont le statut est susceptible de changer au cours des trimestres, tout en limitant la charge d'enquête. Pour évaluer cela, une analyse a été réalisée au préalable sur les FA entrantes en 2016. Ces règles sont résumées dans le tableau suivant :

Tableau : Logements remis en collecte selon la situation au trimestre précédent

Statut du logement le trimestre précédent	Protocole en ancienne enquête		Protocole en nouvelle enquête
	Grappes intermédiaires (RGA=2-5)	Grappes sortantes (RGA=6)	Grappes intermédiaires ou sortantes (RGA=2-6)
Logements vacants	X	X	X
Logements occasionnels	X	X	X
Résidences secondaires		X	
Logements non ordinaires		X	
FA de seniors inactifs*		X	
FA d'inactifs handicapés	X	X	X (uniquement si moins de 50 ans et a travaillé il y a moins de 5 ans)
Logements détruits			
FA ayant exprimé un refus	X	X	X (uniquement si 1 seul refus ou 1 réponse)

Lecture : les croix (X) signifient que la FA est remise en collecte.

* le ciblage des seniors n'est plus le même : 65 ans ou plus en ancienne enquête ; 70 ans ou plus en nouvelle enquête.

- **Les logements vacants**

Parmi les logements vacants ou occasionnels⁴ en rang 1 en 2016, 8,5 % sont devenus résidences principales en rang 2 (8,6 % pour les vacants et 7,8 % pour les occasionnels) et 24 % sont devenus résidences principales au cours des cinq vagues suivantes. Par ailleurs, ces FA présentent un profil particulier en termes d'âge ou de statut d'activité au sens du BIT : les individus qui les occupent sont plus jeunes et plus souvent au chômage. La non remise en collecte de ces logements peut ainsi conduire à un biais sur la mesure des indicateurs principaux de l'enquête. Le choix a donc été fait de remettre en collecte chaque trimestre les logements vacants.

Tableau : Poids des questionnaires individuels de logements vacants ou occasionnels selon l'âge et le statut d'activité

	15-24 ans	24-49 ans	50 ans ou plus	Ensemble
Emploi	3,6 %	1,6 %	0,5 %	1,4 %
Chômage	3,4 %	2,7 %	1,2 %	2,5 %
Inactivité	1,7 %	2,2 %	0,4 %	0,9 %
Ensemble	2,4 %	1,8 %	0,5 %	1,3 %

Champ : personnes répondantes appartenant à une FA de rang intermédiaire ou sortante en 2016.

Lecture : 3,6 % des personnes en emploi âgées de 15-24 ans appartiennent à un logement "vacant" ou "occasionnel".

- **Les résidences secondaires**

Parmi les résidences secondaires en rang 1 en 2016, 2,6 % sont devenues résidences principales en rang 6. Par ailleurs, leur profil en termes d'âge ou de statut d'activité au sens du BIT présente peu

4 Les logements occasionnels sont traités dans la nouvelle enquête comme les résidences principales du fait de la difficulté à définir ces logements. Ainsi, le champ du logement est défini plus précisément lors du questionnement sur les lieux de vie, ce qui permet de réduire les risques de placer hors champ ces logements à tort.

de particularités. Il aurait été très coûteux de revenir vérifier que le statut de ces résidences n'a pas changé : le choix a donc été fait de ne pas remettre en collecte les résidences secondaires.

Tableau : Poids des questionnaires individuels de logements secondaires selon l'âge et le statut d'activité

	15-24 ans	24-49 ans	50 ans ou plus	Ensemble
Emploi	0,3 %	0,2 %	0,2 %	0,2 %
Chômage	0,5 %	0,0 %	0,2 %	0,2 %
Inactivité	0,1 %	0,2 %	0,3 %	0,3 %
Ensemble	0,2 %	0,2 %	0,3 %	0,2 %

Champ : personnes répondantes appartenant à une FA de grappe sortante (RGA=6) en 2016.

Lecture : 0,3 % des personnes en emploi âgées de 15-24 ans appartiennent à une FA "résidence secondaire".

- **Les logements non ordinaires**

Parmi les logements non ordinaires (foyer de travailleur, centre d'hébergement, maison de retraite, local professionnel, etc.) en rang 1 en 2016, 0 % sont devenus résidence principale en rang 6. Le choix a donc été fait de ne pas remettre en collecte les logements non ordinaires.

- **Les logements ordinaires donnant lieu à un refus**

L'enquête comportant 6 interrogations espacées chacune d'un trimestre, il peut être délicat de remettre en collecte des FA de ménages très réfractaires. Parmi les FA « déjà enquêtées » en première interrogation (rang 1), 93 % ont été classées en « refus » ou en « déjà enquêté » l'interrogation suivante. Parmi les FA classées en refus en rang 1, 43 % ont à nouveau été classées en refus le rang suivant. La confirmation d'un refus réduit les chances que le ménage réponde : 67 % des FA classées en refus en rangs 1 et 2 l'ont encore été en rang 3, contre 12 % qui l'étaient en rang 1 et pas en rang 2. Le choix a donc été fait de ne pas remettre en collecte les logements ordinaires dont le ménage dit avoir répondu à une autre enquête et ceux dont le ménage a exprimé au moins deux fois le refus de répondre à l'enquête sans jamais avoir répondu.

Tableau : Nombre de FA et part répondant au moins une fois au cours des 6 trimestres selon leur nombre de refus

Refus au cours des 6 trimestres	1	2	3	4	5	6
Nombre de FA	2655	767	436	350	343	405
Part répondant au moins une fois	75,2 %	54,4 %	49,8 %	35,4 %	22,2 %	0,0 %
Part répondant au moins une fois après le dernier refus	74,7 %	14,8 %	5,7 %	2,6 %	2,2 %	0,0 %

Champ : FA refusant de répondre en 2016

Lecture : 2655 FA ont refusé de répondre une fois sur leurs six interrogations, 75,2% d'entre elles ont répondu au moins une fois et 74,7 % d'entre elles ont répondu au moins une fois après leur dernier refus.

- **Les logements ordinaires avec seniors inactifs**

Les nouvelles règles d'Eurostat permettent de ne pas réinterroger les inactifs de 70 ans ou plus. Parmi les logements concernés en 2016, 1,7 % sont occupés par un nouveau ménage en rang 6. La part de logements occupés par un nouveau ménage en rang 6 augmente avec l'âge de la personne de référence : elle est de 0,6 % pour les 70-79 ans, 1 % pour les 80-84 ans, 1,7 % pour les 85-89 ans et de 2,4 % pour les 90 ans ou plus. Il aurait été possible de ne remettre en collecte que les ménages de

seniors inactifs de 80 ans ou plus, un an après leur première interrogation mais avec un coût d'enquête relativement élevé.

Par ailleurs, les FA d'inactifs de 70 ans ou plus pèsent peu dans la mesure de l'emploi et du chômage. Le choix a donc été fait de ne pas réinterroger les logements composés exclusivement d'inactifs de 70 ans ou plus. Ces individus sont par la suite considérés comme répondants et imputés comme inactifs. Les autres informations du questionnaire ne sont pas renseignées.

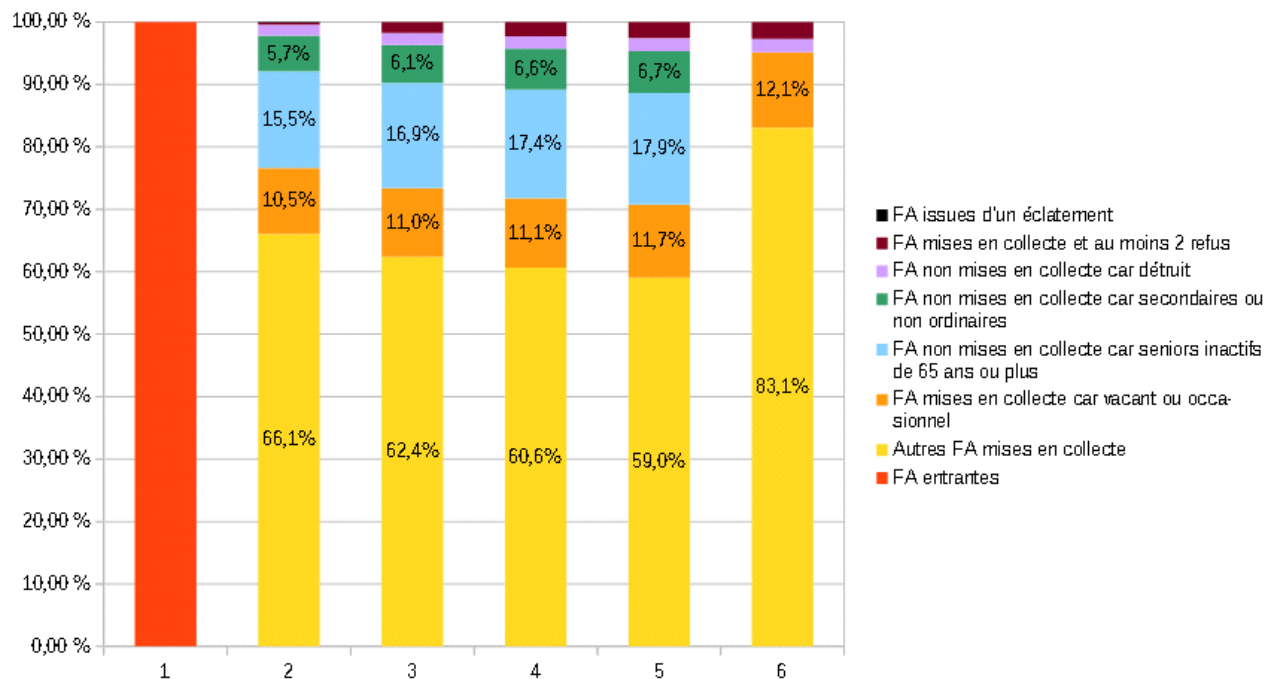
- **Les logements ordinaires avec inactifs en situation de handicap**

Les nouvelles règles d'Eurostat permettent de ne pas réinterroger les personnes inactives, déclarées comme en incapacité de travailler en raison d'un handicap ou d'un problème de santé durable. En 2016, parmi les personnes inactives handicapées âgées de 15-69 ans en première interrogation, 7,5 % sont au moins une fois actives sur les cinq vagues restantes. Des analyses ont été menées pour voir si cette part variait pour des sous-groupes, afin de voir s'il était pertinent d'écarter toute cette population ou s'il était préférable de cibler une population plus restreinte. Les chances de devenir actif dépendent peu de l'état de santé ressenti, mais varient selon l'âge et le fait d'avoir déjà travaillé. Ainsi, les chances de devenir actif sont plus réduites pour les 50-69 ans (4,4 %), qui représentent 62 % des inactifs handicapés de 15-69 ans. Avant 50 ans, les personnes handicapées inactives n'ayant jamais travaillé ou ne travaillant plus depuis au moins cinq ans, qui représentent 30 % des personnes handicapées inactives de 15-69 ans, sont 8 % à devenir actives au cours des cinq vagues restantes. Les jeunes peuvent être plus souvent dans cette situation tout en ayant une probabilité de devenir actif plus importante, car ils peuvent ne jamais avoir travaillé pour une autre raison que leur handicap. 16 % des personnes handicapées inactives de 15-29 ans deviennent actives au cours des cinq vagues restantes. Mais ce n'est le cas que de 9 % de ceux qui n'ont jamais travaillé ou n'ayant pas travaillé depuis au moins cinq ans.

Le choix a donc été fait de ne pas réinterroger les inactifs en situation de handicap âgés de 50-69 ans et ceux de moins de 50 ans n'ayant jamais travaillé ou n'ayant pas travaillé depuis au moins cinq ans. Les logements composés exclusivement d'individus suivant ces critères ne sont pas remis en collecte. Comme pour les seniors inactifs, ils sont par la suite considérés comme répondants et imputés comme inactifs. Au final, 92 % des personnes handicapées inactives de 15-69 ans ne sont plus réinterrogées. Parmi elles, 5,6 % sont devenus actives au cours de l'une des réinterrogations suivantes, contre 30 % des 8 % restants.

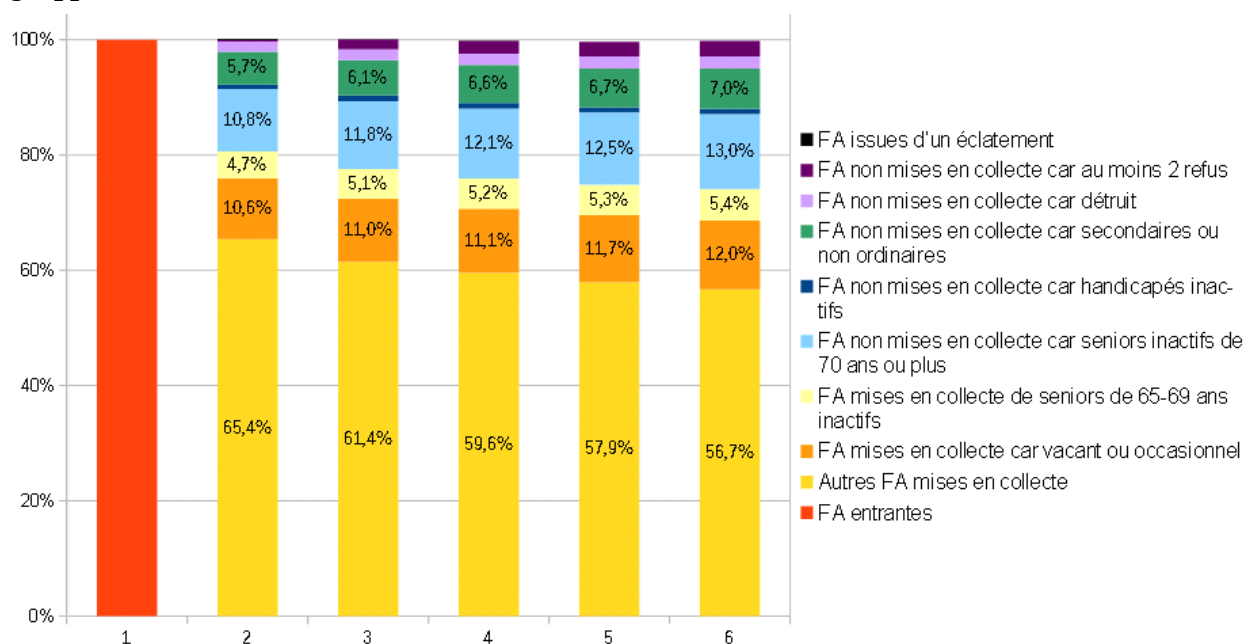
Au final, par rapport au protocole avant refonte, le nombre de FA remises en collecte pour les rangs 2 à 5 augmente légèrement, notamment du fait d'une moindre restriction du champ des seniors non remis en collecte (le seuil passant de 65 à 70 ans) ; cela est compensé dans la nouvelle enquête par la non-remise en collecte des logements composés d'inactifs en situation de handicap et de ceux avec au moins 2 refus exprimés (*figures 10 et 11*). Cependant, du fait du changement de protocole pour les sortantes, le nombre de logements non remis en collecte pour ce dernier rang d'interrogation est nettement plus important. En 2016, le nouveau protocole aurait conduit à remettre en collecte en dernière interrogation 74 % des FA, contre 95 % dans l'ancien protocole. Au total, les nouvelles règles se traduisent par une charge d'enquête relativement équivalente.

Figure 10 : FA remises en collecte en 2016 en ancienne enquête selon le rang de la grappe



Lecture : au rang 6, 12,1 % des FA sont mises en collecte car vacant ou occasionnel avec les anciennes règles.
 Source : Insee, enquête Emploi 2016.

Figure 11 : FA remises en collecte en 2016 avec les règles de la nouvelle enquête, selon le rang de la grappe



Lecture : au rang 6, 12,0 % des FA sont mises en collecte car vacant ou occasionnel avec les nouvelles règles.
 Source : Insee, enquête Emploi 2016.

3. La rénovation de la méthode de pondération

En parallèle de la refonte liée au nouveau règlement IESS, un nouvel échantillon a été tiré dans une nouvelle base de sondage issue de données fiscales ([la base des fichiers démographiques sur les logements et les individus - Fidéli](#)), et a été progressivement mis en production, à compter du 3^e trimestre 2019, l'échantillon étant totalement renouvelé depuis le 4^e trimestre 2020 (Vincent et Paliod, 2022).

Dans ce contexte, il a été décidé de faire évoluer la méthode de pondération de l'enquête Emploi en continu à compter de 2021, au moment de la bascule de la refonte. Ce chantier a été l'occasion de mettre en conformité la méthode de pondération avec le nouveau cadre européen, mais également de réinterroger certains aspects de la méthode utilisée précédemment : stratégie globale de pondération, variables à utiliser pour corriger du biais éventuel d'échantillonnage et de la non-réponse. Il a permis de mettre en évidence l'apport de la nouvelle base de sondage Fidéli, *via* notamment la présence de données sur les revenus d'activité, sur la pondération de l'EEC.

Cette refonte de la pondération de l'EEC ne concerne que la France métropolitaine. En effet, la base de sondage de l'échantillon dans les Dom est le Recensement de la population, et la méthode de pondération y est différente (méthode en deux étapes : correction de la non-réponse par un modèle de régression, puis calage sur marges)⁵.

Les travaux sur la refonte de la pondération ont démarré dans le cadre d'un groupe de travail, puis la nouvelle méthode de pondération a été finalisée en 2020 au moment de la collecte du Pilote de l'enquête Emploi, pour être mise en production début 2021.

La méthode de pondération avant 2021 consiste en un calage sur marges, en une seule étape, permettant à la fois de corriger de la non-réponse, et de caler l'enquête sur des marges externes (population des ménages ordinaires, parc de logements).

Le calage prend en compte les logements hors champ et les logements répondants, car c'est sur l'ensemble des logements qu'on dispose de marges correspondantes dans la base de sondage. Le calage se fait rang par rang. Les marges de calage sont de deux types :

- des marges issues de la base de sondage :
 - la localisation géographique : Rural / Semi-urbain / Urbain ;
 - des caractéristiques du logement : Résidence principale / Résidence secondaire / Logement vacant (selon la base de sondage) ; Maison / Appartement ; Propriétaire / Autre (Locataire, logé gratuitement, etc.) ; HLM / Autre logement ; Nombre de pièces ;
 - revenus perçus dans le logement : décile de revenu total perçu dans le ménage.
- des marges externes sur la population et les logements :
 - population des ménages ordinaires, par sexe et âge quinquennal (avec regroupement pour les 75 ans ou plus), issues des estimations de population ;

5 Dans les Dom, de très légères modifications ont été effectuées : calage rang par rang (au lieu de calage rangs 1, 2-5, 6) et modification du détail de tranches d'âge dans les marges de population. Ces modifications n'ont quasiment aucun impact sur les principaux indicateurs.

- population des ménages ordinaires par ancienne région (NUTS2) et par sexe et âge. Le niveau de détail sexe x âge dépend de la taille de la région ;
- nombre de logements et nombre de résidences principales ;
- nombre de logements construits après 2000 (marge « logements neufs »).

Le calage se fait au niveau logement, les données relatives à la population (calage sur marges externes) étant sommées au niveau logement. Tous les individus d'un logement ont le même poids. On passe de la pondération logement (EXTRILOG) à la pondération individuelle (EXTRI) par une translation des poids : $EXTRI = EXTRILOG \times (\text{nombre d'individus du logement dans le champ}) / (\text{nombre d'individus du logement dans le champ et répondant}^6)$.

Les pistes d'amélioration identifiées

La méthode globale en une seule étape rend cette dernière sensible aux révisions sur les marges extérieures. En effet, un logement répondant contribue à la fois au calage pour ses caractéristiques dans la base de sondage et pour celles à l'enquête (population, nombre de logements). Dans les strates avec peu de logements (par exemple, pour les jeunes), toute variation, même infime d'une marge externe (population), peut modifier de façon importante les poids lors des révisions. En pratique, depuis 2014, la révision annuelle des marges de population entraîne des révisions non négligeables sur les indicateurs relatifs aux jeunes, sans qu'il n'y ait une justification autre que l'aléa statistique.

Par ailleurs la méthode de calage en une seule étape, avec de très fortes contraintes en termes de marges du fait d'un nombre de marges important, augmente le risque de poids élevés.

Enfin, la diminution de la taille de l'échantillon (- 15 %) permise par la nouvelle méthode d'échantillonnage, et le changement de structure de celui-ci par région afin de respecter des seuils minima de précision exigés par Eurostat (par exemple, doublement de la taille de l'échantillon en Corse), a nécessité de réexaminer le niveau de détail des marges externes par région.

3.1. L'apport de la nouvelle base de sondage

Par rapport à l'ancienne base de sondage (taxe d'habitation - TH), le fichier Fidéli comporte des informations plus détaillées sur les occupants des logements, issues d'un appariement avec les fichiers d'impôt sur le revenu. On dispose en particulier d'informations sur les différents revenus perçus par chacun des individus (salaires, revenus non-salariés, allocations chômage, pensions, etc.), qui peuvent être utilisées de différentes façons au niveau logement : indicatrices de perception, montants totaux, etc. Auparavant, dans la TH, on ne disposait que de l'information sur le décile de revenu total perçu dans le ménage. Il est à noter que pour près de 20 % de l'échantillon de l'EEC, on n'a, par construction, aucune information sur les habitants (cas des résidences non principales : résidences secondaires, logements vacants).

Après différents essais, il a été choisi de travailler sur une **variable agrégée de revenus d'activité** (salaires + revenus non-salariés) perçus dans le ménage, cette variable étant ensuite regroupée en « décile détaillé ». Il s'agit d'un regroupement en décile de revenu avec un détail plus important sur le bas et le haut de la distribution. Le détail dans le bas de la distribution a été effectué car on

6 La non-réponse individuelle est très rare dans l'EEC (1%).

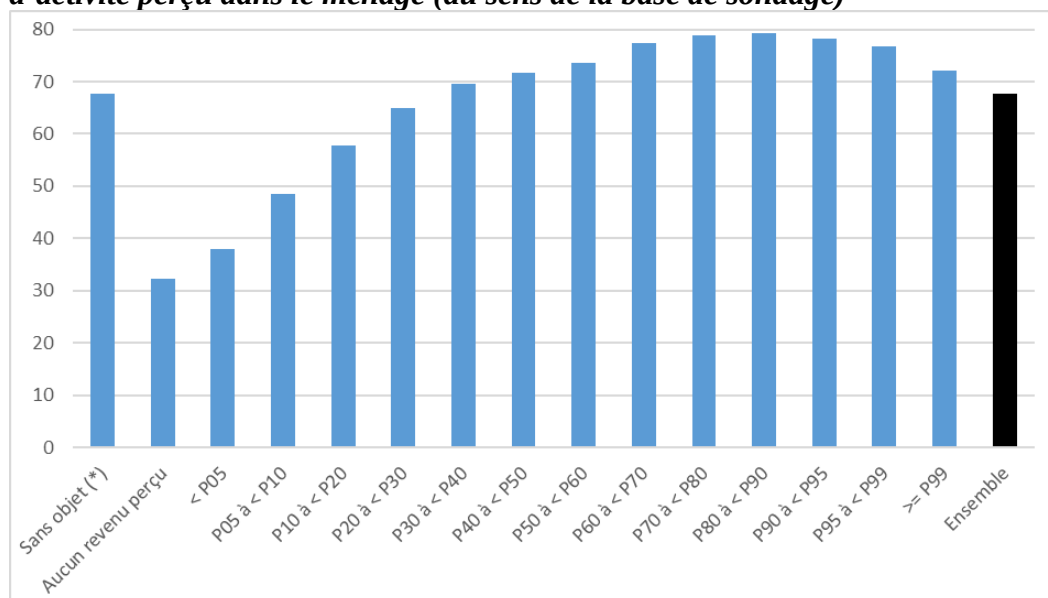
observe des disparités de statut BIT. Pour le haut de la distribution, c'est plutôt en termes de correction de la non-réponse que l'intérêt réside (voir plus loin).

Sans surprise, la variable de revenu d'activité est très fortement corrélée avec le statut BIT mesuré dans l'enquête. En effet, le taux d'emploi des 15-64 ans augmente fortement avec le niveau de revenu d'activité mesuré dans la base de sondage (figure 12), alors que le taux de chômage diminue (figure 13).

La variable sur les allocations chômage n'a pas été utilisée car son pouvoir explicatif sur la probabilité d'être en emploi est faible une fois la variable de revenus d'activité utilisée. En effet, une partie importante des chômeurs au sens du BIT ne sont pas indemnisés par l'Assurance chômage ; une partie est bénéficiaire de minima sociaux, variable non disponible directement dans Fidéli et qui pourrait être appariée mais avec trop de décalage temporel (deux à trois ans). Des travaux ont montré que l'ajout de cette variable à celle sur les revenus d'activité n'avait pas d'impact important sur les résultats.

Par ailleurs, le revenu d'activité est lié à la probabilité de répondre à l'enquête. Outre les résidences non principales au sens de la base de sondage, dont le taux de réponse avoisine 56 % en 2021 (contre 77 % sur l'ensemble des logements dans le champ), le taux de réponse augmente avec le niveau de revenus d'activité, sans toutefois présenter d'énormes disparités. Il se stabilise pour les très hauts revenus d'activité, voire diminue pour le dernier centile (figure 14).

Figure 12 : taux d'emploi des 15-64 ans dans l'EEC en 2021, selon le décile détaillé de revenu d'activité perçu dans le ménage (au sens de la base de sondage)



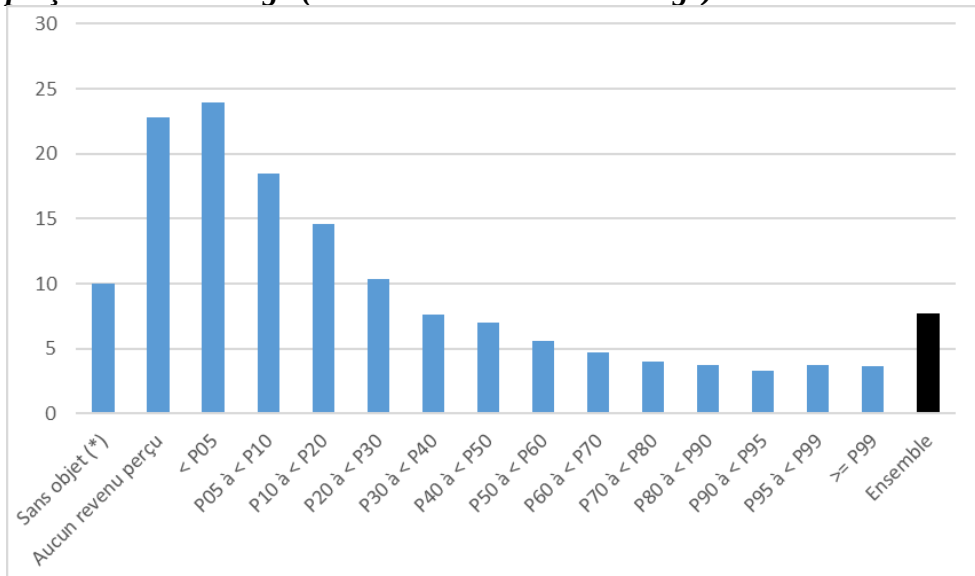
Lecture : en moyenne en 2021, le taux d'emploi des 15-64 ans résidant dans un logement dont le revenu d'activité dans le fichier Fidéli 2020 est supérieur au 99^e percentile est de 72,1 %.

(*) Résidence non principale (secondaire, vacant) au sens de la base de sondage.

Champ : France métropolitaine, logement ordinaire, personnes de 15 à 64 ans.

Source : Insee, enquête Emploi 2021, appariée avec Fidéli 2020

Figure 13 : taux de chômage dans l'EEC en 2021, selon le décile détaillé de revenu d'activité perçu dans le ménage (au sens de la base de sondage)



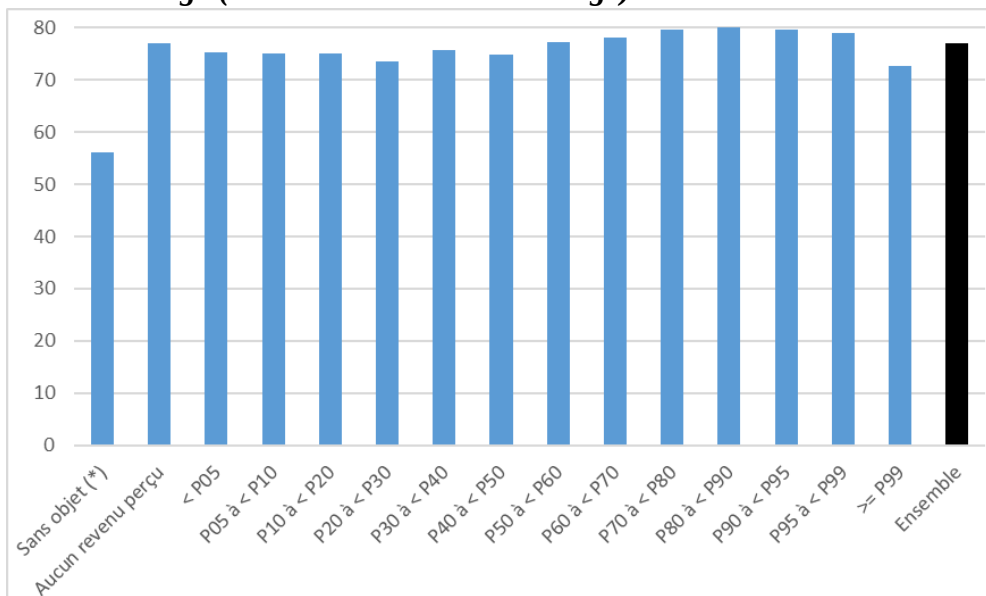
Lecture : en moyenne en 2021, le taux de chômage dans un logement dont le revenu d'activité dans le fichier Fidéli 2020 est supérieur au 99^e percentile est de 3,8 %.

(*) Résidence non principale (secondaire, vacant) au sens de la base de sondage

Champ : France métropolitaine, logement ordinaire, personnes actives.

Source : Insee, enquête Emploi 2021, appariée avec Insee, Fidéli 2020.

Figure 14 : Taux de réponse à l'EEC en 2021, selon le décile détaillé de revenu d'activité perçu dans le ménage (au sens de la base de sondage)



Note : le taux de réponse rapporte le nombre de logements répondants au nombre de logements dans le champ de l'enquête (logements ordinaires).

Lecture : en moyenne en 2021, le taux de réponse d'un logement dont le revenu d'activité dans le fichier Fidéli 2020 est supérieur au 99^e percentile est de 72,6 %.

(*) Résidence non principale (secondaire, vacant) au sens de la base de sondage

Champ : France métropolitaine, logement ordinaire (dans le champ de l'enquête).

Source : Insee, enquête Emploi 2021, appariée avec Insee, Fidéli 2020.

Les autres déterminants de la non-réponse

D'autres facteurs, déjà identifiés avec l'ancienne base de sondage, influent sur la probabilité de répondre à l'EEC. Ainsi, le taux de réponse est plus faible dans les résidences non principales au sens de la base de sondage, à Paris, chez les locataires ou en quartiers prioritaires de la ville (QPV). Il est en revanche plus élevé hors QPV, en maison, en Alsace ou en Bretagne (*figure 16*). Toutes choses égales par ailleurs, trois variables semblent très explicatives de la probabilité de répondre à l'enquête : région, type de logement (principal, secondaire, vacant) et nombre de pièces du logement. Le décile détaillé de revenu d'activité a un impact plus mesuré, comme le suggère la *figure 13*.

Toutefois, la variable de revenu d'activité de la base de sondage a une contribution écrasante (2^e et 3^e colonnes de la *figure 15*) lorsque l'on étudie la probabilité d'être en emploi ou au chômage dans l'enquête, selon les caractéristiques de la base de sondage. Il semble donc très intéressant de garder cette variable dans le modèle de correction de la non-réponse.

À l'inverse, d'autres variables, auparavant utilisées, ont été abandonnées, comme le statut d'occupation du logement. Ces variables avaient une contribution relativement importante au fait d'être répondant, mais étaient très faiblement corrélées au statut BIT.

Figure 15 : Contribution de différentes variables dans un modèle qualitatif mesurant la probabilité de trois événements (être répondant, en emploi ou au chômage) selon les caractéristiques de la base de sondage

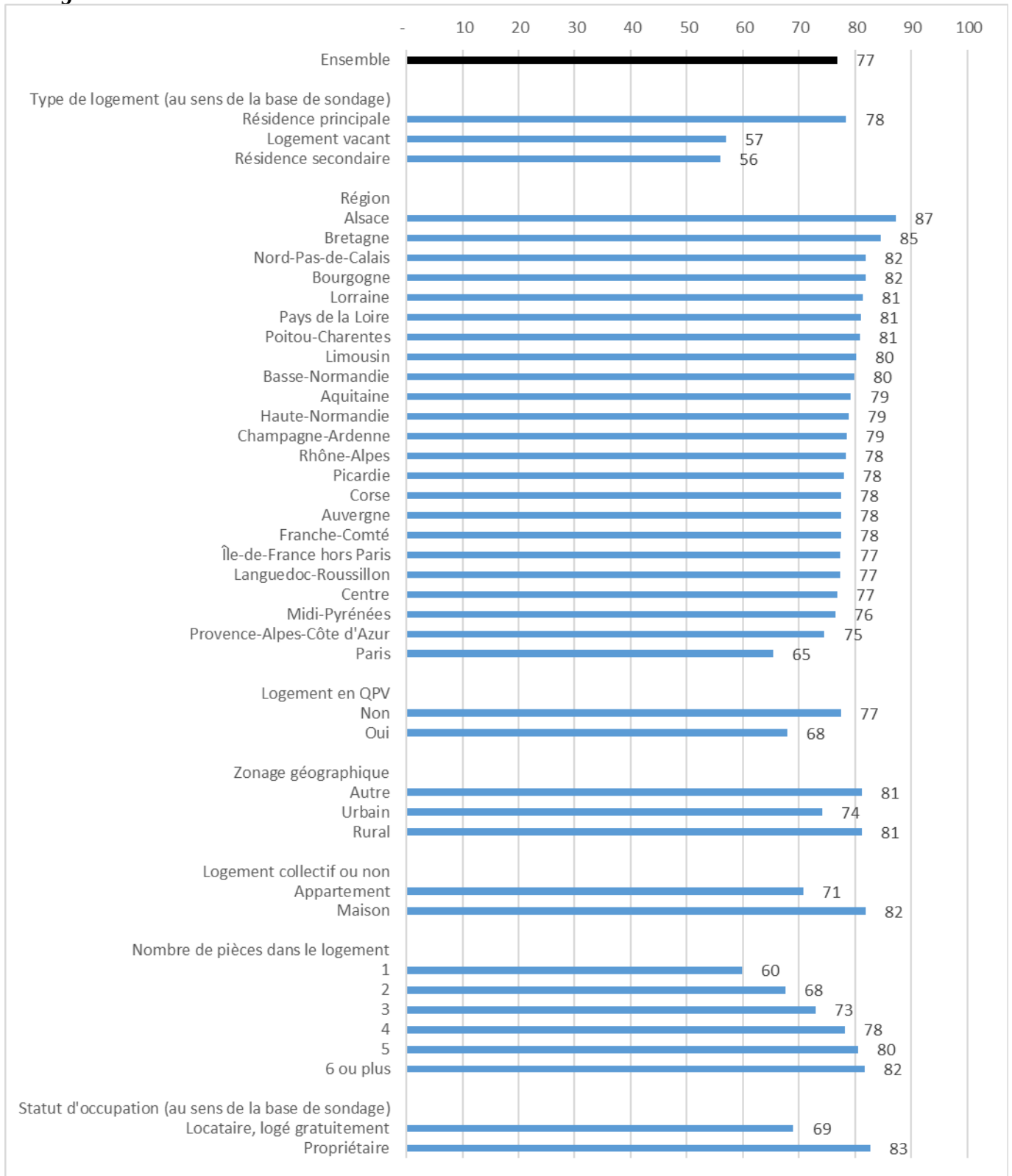
	Être répondant	Emploi	Chômage
Semaine de référence	12,0	0,5	1,9
Rang d'interrogation	4,5	0,0	0,2
Région	27,9	0,9	3,2
Zone urbaine	0,2	0,6	0,9
QPV	3,1	2,2	3,0
Résidence principale / secondaire / vacant	22,3	0,0	1,3
Nombre de pièces du logement	16,9	5,3	0,7
Maison / appartement	4,8	0,4	2,3
Logement neuf	0,1	1,5	0,1
Décile détaillé de revenu d'activité	8,2	88,6	86,5
Total	100,0	100,0	100,0

Lecture : trois modèles qualitatifs ont été estimés mesurant la probabilité d'être répondant à l'enquête, d'être en emploi (parmi les 15-64 ans) ou d'être au chômage (parmi les 15-64 ans) selon les variables de la base de sondage présentées dans le tableau. Pour chacun ces trois modèles, la « contribution » est calculée comme le poids relatif de la variable dans le Khi-2 de Wald.

Champ : France métropolitaine, logement ordinaire (dans le champ de l'enquête).

Source : Insee, enquête Emploi 2021, appariée avec Insee, Fidéli 2020.

Figure 16 : Taux de réponse à l'EEC en 2021, selon diverses caractéristiques de la base de sondage



Note : le taux de réponse rapporte le nombre de logements répondants au nombre de logements dans le champ de l'enquête (logements ordinaires).

(*) Résidence non principale (secondaire, vacant) au sens de la base de sondage.

Lecture : en moyenne en 2021, le taux de chômage dans un logement situé en QPV est de 68 %.

Champ : France métropolitaine, logement ordinaire (dans le champ de l'enquête).

Source : Insee, enquête Emploi 2021, appariée avec Insee, Fidéli 2020.

3.2. La nouvelle méthode de pondération depuis 2021

À la suite des différents travaux menés, il a été décidé de diviser la nouvelle méthode de pondération en deux étapes. La méthode en deux étapes permet de bien distinguer ce qui relève de la correction de la non-réponse, du biais d'échantillonnage et du calage sur marges externes, notamment en cas de révision des marges externes. La première étape est elle-même divisée en deux sous-étapes :

- **étape 1a** - calage de l'échantillon trimestriel sur des marges issues de la base de sondage de l'année N-1 :
 - par rang de la grappe (soit 6 strates) : nombre total de logements ; répartition par type de logement (résidence principale, résidence secondaire, logement vacant) ; zonage (urbain/rural/autre) ; zone de résidence (Paris, Île-de-France hors Paris, ZEAT sinon) ; implantation en QPV (oui/non) ; statut d'occupation (propriétaire/autre) ; type d'habitat (maison/appartement) ; année de construction (plus de 10 ans, entre 5 et 10 ans, moins de 5 ans) ; nombre de personnes dans le ménage, nombre de personnes de nationalité étrangère ; somme des revenus perçus dans l'année (salaires, revenus agricoles, revenus non salariés (hors revenus agricoles), pensions de retraite, allocations chômage).
 - pour l'ensemble des rangs : décile de revenu d'activité. Pour cette variable, afin de garder une information intéressante en termes de profil sur les modalités extrêmes (très faibles ou très forts revenus), il n'a pas été possible de caler rang par rang du fait d'effectifs trop faibles.

Par ailleurs, dans la même étape, l'échantillon est calé de façon uniforme sur chacune des semaines de références, afin que chaque semaine ait le même poids, en termes de nombre de logements.

- **étape 1b** - modèle économétrique mesurant la probabilité de réponse selon les caractéristiques suivantes : semaine de référence, rang de la grappe, région, zonage, résidence en QPV, type d'habitat, nombre de pièces, année de construction, décile de revenu d'activité. Les probabilités estimées sont ensuite regroupées par groupe de réponse homogène (on attribue la moyenne des probabilités à l'ensemble des observations appartenant à un même groupe), dans le but de limiter la dispersion des poids et le risque de poids extrêmes.

L'étape 1a permet d'avoir pour chaque trimestre et chaque rang un échantillon « parfaitement » représentatif de la base de sondage, et donc de corriger d'un éventuel biais d'échantillonnage. Le fait d'utiliser un modèle économétrique dans l'étape 1b permet de corriger d'un biais de non-réponse lié à ces caractéristiques.

L'**étape 2** consiste en un calage sur données externes :

- par rang d'interrogation : population vivant en logement ordinaire par âge quinquennal et sexe ; population vivant en logement ordinaire par région (au niveau NUTS2), âge (de deux à quatre tranches d'âge selon les régions) et sexe ; nombre de résidences principales
- par semaine de référence : population en quatre tranches d'âge.

L'étape 2 permet de corriger les biais résiduels liés à la non-réponse en augmentant la précision des variables d'intérêt liées aux variables de calage mais également d'assurer une cohérence avec des sources externes de référence.

Comme auparavant, le calage se fait au niveau logement, les données relatives à la population (calage sur marges externes) étant sommées au niveau logement, et seuls les individus répondants dans le champ ont une pondération. On passe de la pondération logement à la pondération individuelle par la même translation de poids qu'auparavant.

Le critère de complétude d'un questionnaire pour considérer un individu comme répondant a été en revanche modifié : tous les questionnaires avec un statut BIT rempli sont pris en compte, alors que dans l'ancienne méthode de pondération, on ajoutait un critère supplémentaire (le fait d'avoir rempli l'une des dernières questions : statut d'activité déclaré, pour garantir au mieux la complétude du questionnaire individuel). En effet, du fait d'une non-réponse plus élevée avec la réponse par Internet, le maintien de ce critère supplémentaire excluait trop de questionnaires du champ. L'inconvénient de ce choix, et de la collecte par internet en général, est qu'il y a davantage de non-réponse partielle dans l'EEC (sur des individus avec une pondération positive).

Un impact à la hausse sur l'emploi, neutre sur le chômage

Pour mesurer l'impact de la nouvelle méthode de pondération, ont été rejouées sur les données de l'ancienne enquête, l'ancienne et la nouvelle méthode de pondération⁷. L'exercice est réalisé sur les cinq trimestres du T1 2020 au T1 2021, et l'impact est mesuré de façon globale sur le cumul de cette période.

Le principal impact de la nouvelle méthode de pondération est une hausse du taux d'emploi, alors que l'effet est quasi nul sur le taux de chômage (*figure 17*). Le taux d'emploi des 15-64 ans augmente ainsi de 0,3 point du seul effet de changement de pondération, la hausse étant plus marquée chez les 25-49 ans (+0,4 point). L'impact sur les caractéristiques des personnes et des emplois est très limité. On observe une légère hausse de la part des CDD du fait de la nouvelle méthode de pondération (+0,3 point).

Figure 17 : impact de la nouvelle méthode de pondération de l'EEC (en points de %)

Taux d'emploi des 15-64 ans	0.3
15-24 ans	0.0
25-49 ans	0.4
50-64 ans	0.2
Taux de chômage	-0.1
15-24 ans	0.0
25-49 ans	-0.1
50 ans ou plus	0.0
Indicateurs sur l'emploi	
Part du temps partiel	0.0
Part des indépendants	-0.1
Part des CDD	0.3

Champ : France hors Mayotte, personnes vivant en logement ordinaire.

Source : Insee, enquête Emploi T1 2020-T1 2021.

⁷ On aurait pu tout aussi bien reproduire l'ancienne méthode de pondération sur les données de la nouvelle EEC, mais cela était techniquement plus délicat notamment du fait de l'appariement avec d'anciens fichiers de la TH.

L'ajout de la variable de revenus d'activité dans le redressement de la non-réponse semble expliquer la majeure partie de l'effet à la hausse sur l'emploi. L'ajout de cette variable – à autres paramètres inchangés – contribue pour +0,4 point sur le taux d'emploi des 15-64 ans (*figure 18*), alors que les effets restants ont, dans leur globalité, un impact faible (-0,1 point) sur ce même indicateur. Ces effets restants incluent notamment le changement de méthode de pondération (passage en plusieurs étapes) ou encore les légères modifications faites dans le choix des variables dans le redressement de la non-réponse ou le calage sur marges. Sur les 50-64 ans, ces effets « restants » semblent avoir un impact plus important sur le taux d'emploi (-0,4 point), alors que l'ajout de la variable de revenus d'activité a un impact de signe inversé, encore plus important que sur l'ensemble (+0,6 point).

Figure 18 : décomposition de l'impact de la nouvelle méthode de pondération de l'EEC (en points)

	Total	dont : ajout de la variable revenus d'activité	dont : effet restant
Taux d'emploi des 15-64 ans	0,3	0,4	-0,1
15-24 ans	0,0	0,1	-0,1
25-49 ans	0,4	0,3	0,1
50-64 ans	0,2	0,6	-0,4
Taux de chômage	-0,1	-0,1	0,0
15-24 ans	0,0	-0,1	0,1
25-49 ans	-0,1	-0,1	0,0
50 ans ou plus	0,0	-0,1	0,1

Champ : France hors Mayotte, personnes vivant en logement ordinaire.

Source : Insee, enquête Emploi T1 2020-T1 2021.

3.3. De nouvelles règles de rattachement

Les règles de rattachement des individus visent à gérer les cas de multi-résidence pour respecter le champ des enquêtes ménages et traiter les doubles comptes. Ces règles définissent la population d'intérêt, dans le champ de l'enquête. Au début du questionnaire, l'ensemble des personnes vivant, même occasionnellement, dans le logement sont identifiées, mais toutes ne remplissent pas un questionnaire individuel et toutes ne participent pas au calcul des indicateurs : soit parce qu'elles ne relèvent pas du champ de l'enquête (par exemple si elles résident dans une collectivité), soit pour éviter les doubles comptes (par exemple un enfant vivant en garde alternée chez ses parents séparés).

La plupart des règles de rattachement sont définies par le règlement européen. Ainsi, lorsqu'une personne vit régulièrement dans plusieurs résidences au cours de l'année, elle est rattachée à celle où elle vit le plus de temps. Cette règle générale n'est pas valable pour les cas suivants :

- Les personnes en double résidence logement ordinaire / communauté sont exclues du champ, à l'exception des mineurs en internat ou cité U.
- Les personnes qui travaillent hors du domicile familial pendant la semaine et qui reviennent habituellement au domicile familial le week-end doivent considérer leur domicile familial comme leur lieu de résidence habituelle.

- Les mineurs qui vivent chez un de leurs parents et dans un autre logement qui n'est pas celui d'un de leurs parents (logement pour les études par exemple) sont rattachés au logement où ils vivent avec au moins un de leurs parents.
- Pour les enfants en garde alternée, la règle retenue est celle du logement où l'enfant a passé une nuit donnée. Le logement ainsi déterminé est conservé pour toutes les vagues du panel.

Un autre cas particulier est celui des étudiants majeurs vivant à la fois chez leurs parents et dans un autre logement ordinaire. Le règlement européen ne statue pas précisément sur ce cas pour les enquêtes sur l'emploi. À l'occasion de la refonte de l'enquête, le questionnaire a évolué pour mieux repérer des situations de double résidence, et certaines règles de rattachement ont été revues, pour l'essentiel celle concernant les étudiants majeurs vivant à la fois dans un logement pour leurs études et chez leurs parents. Avant 2021, ces étudiants étaient essentiellement rattachés au logement de leurs parents ; ceci permettait d'améliorer leur taux de réponse (les parents étant plus facilement joignables et prompts à répondre à l'enquête), mais générait une part élevée de réponse par proxy (l'étudiant étant fréquemment absent lors du passage de l'enquêteur chez ses parents, ses parents répondaient à sa place). Or, d'une part, la réponse par un tiers est de moindre qualité, notamment pour les questions servant à déterminer le statut d'activité, si la personne réalise des petits boulots par exemple. Ainsi, à âge et niveau de formation donnés, le taux d'emploi des étudiants en double résidence est plus élevé lorsqu'ils sont enquêtés dans le logement de leurs études que dans celui de leurs parents. D'autre part, ce traitement n'était pas cohérent avec celui retenu dans les enquêtes annuelles de recensement ; toutefois l'application d'une telle règle dans le recensement, qui a lieu chaque année en janvier, hors vacances scolaires, conduit à une forte volatilité des taux de réponse pour une enquête en continu comme l'EEC, avec de faibles taux de réponses pendant les vacances scolaires. À partir de 2021, pour combiner les avantages de ces deux règles, une solution mixte est mise en œuvre : les étudiants en double résidence sont interrogés au titre de leurs deux logements, ils sont donc rattachés à leurs deux logements. Ce double-compte est ensuite traité lors de la repondération, en divisant leur poids par deux.

4. Une refonte jusqu'à l'aval et la diffusion des données

Au-delà des modifications de la chaîne aval rendues nécessaires par la refonte du questionnaire, la nouvelle liste de variables européennes et les évolutions du protocole (notamment concernant les règles de réinterrogation), il a été décidé d'en profiter pour mener une rénovation plus large de l'aval de la collecte et de la diffusion des données. L'objectif de cette rénovation est d'une part d'avoir un aval harmonisé avec l'aval européen et davantage sécurisé, et d'autre part de mettre à disposition des utilisateurs des produits répondant davantage à leurs besoins. La rénovation de la chaîne aval va de pair avec une évolution de la documentation associée, notamment du dictionnaire des codes.

Ainsi, la refonte de la chaîne aval a comporté plusieurs axes d'amélioration :

- Le **langage de programmation**. L'objectif était de profiter de cette refonte pour s'inscrire dans les orientations de moyen terme de l'Insee avec la sortie de SAS, permettant ainsi de sécuriser la pérennité de la chaîne aval en l'écrivant en R. Concrètement le changement de langage de programmation a démarré après l'atterrissage proprement dit du projet Mélopée, la refonte de la chaîne s'est donc faite en deux temps, avec une première écriture en SAS,

puis une traduction en R. Cette stratégie est possible car il n'y a pas d'adhérence entre le langage et le résultat de la chaîne, qu'il s'agisse de la construction des données ou des programmes de redressement.

- **L'architecture de la chaîne aval.** Celle-ci est peu modifiée et se décompose en plusieurs modules. Le module spécifique aux variables de formation est supprimé, et ces variables sont directement intégrées au module de la chaîne aval qui recalcule les variables individuelles.
- **Le format des tables.** Celui-ci est imposé pour les tables fournies à Eurostat, et contrôlé par un processus de validation automatisé. Concernant les tables de diffusion françaises, plusieurs types de tables sont construites : des tables brutes non diffusées ; des tables « Z » avec des informations indirectement identifiantes diffusées au CASD avec un accès restreint et sécurisé ; des tables de production et de recherche (FPR) non identifiantes diffusées via le réseau Quetelet-Progedo notamment ; un fichier détail beaucoup plus sommaire diffusé en accès libre sur Insee.fr. L'architecture des FPR a été revue pour simplifier son utilisation, en passant de 3 tables à 2 tables.
- **Le contenu des tables.** Le contenu a été revu pour faciliter l'utilisation des tables de diffusion : intégration des variables européennes dans les tables de diffusion, plutôt que dans un fichier distinct ; ajout de variables construites permettant de calculer directement les principaux indicateurs à la fois français et européens ; suppression des variables purement intermédiaires ; harmonisation des variables françaises et européennes dans leur construction ; augmentation du nombre de variables redressées en lien avec une réponse partielle plus importante du fait du mode Internet. Par ailleurs, la frontière des variables disponibles au CASD et pas dans les FPR a été revue pour correspondre davantage aux préconisations et aux choix faits dans d'autres enquêtes. Le format des variables a également été revu pour distinguer la non-réponse du hors champ, pour harmoniser le codage de variables avec les mêmes modalités de réponse (oui/non par exemple), harmoniser les noms de variables, etc. Enfin, les variables des modules complémentaires sont désormais pleinement intégrées aux tables de l'enquête alors qu'une diffusion distincte était faite auparavant. Cette intégration a l'avantage de limiter les procédures d'archivage et de diffusion et de faciliter l'accès à ces données pour les utilisateurs.
- **La sécurisation de la chaîne aval.** L'objectif était de mettre en place davantage de contrôles. De son côté, Eurostat met désormais en œuvre des contrôles de cohérence plus poussés.
- **La documentation.** La documentation associée aux tables de diffusion a été rénovée pour améliorer la transparence et l'utilisation des données. Ainsi, des tableaux de codage pour les principaux indicateurs sont diffusés et l'information fournie est harmonisée dans l'ensemble du dictionnaire des codes. Les noms de variables sont revus pour être davantage signifiant : par exemple les variables préfixées par LFS_ correspondent aux variables européennes.

En termes de calendrier, les programmes de la chaîne aval ont été prêts pour le Pilote en 2020. L'objectif était d'anticiper la production du T1 2021 pour laquelle la chaîne aval devait tourner rapidement et permettre de calculer les indicateurs principaux de l'enquête. Les fichiers pour Eurostat devaient également être prêts sous le format requis pour être livrés dans les temps impartis.

Pour diffuser les données de l'enquête Emploi 2021 dans des délais similaires aux années précédentes, il a été décidé de réaliser exceptionnellement deux diffusions. Une première diffusion a eu lieu en juin 2022, une seconde début 2023. Cette double diffusion a permis de ne pas rallonger le calendrier de mise à disposition des utilisateurs. Elle avait également pour objectif de profiter de premiers retours d'utilisateurs pour améliorer le contenu des fichiers et de la documentation associée lors de la deuxième diffusion.

5. Des outils adaptés à ces évolutions

Pour répondre à ces principaux changements, notamment à l'implémentation d'un protocole multimode concurrentiel, il a été nécessaire d'adapter les outils. Le processus métier cible conçu dans le cadre du projet Mélopée est resté proche du processus de collecte de l'ancienne enquête Emploi. Pour autant, les besoins du projet ont un impact sur les différentes phases du processus métier comme le montre la répartition au regard de la grille du modèle générique GSBPM (Generic Statistical Business Process Model) :

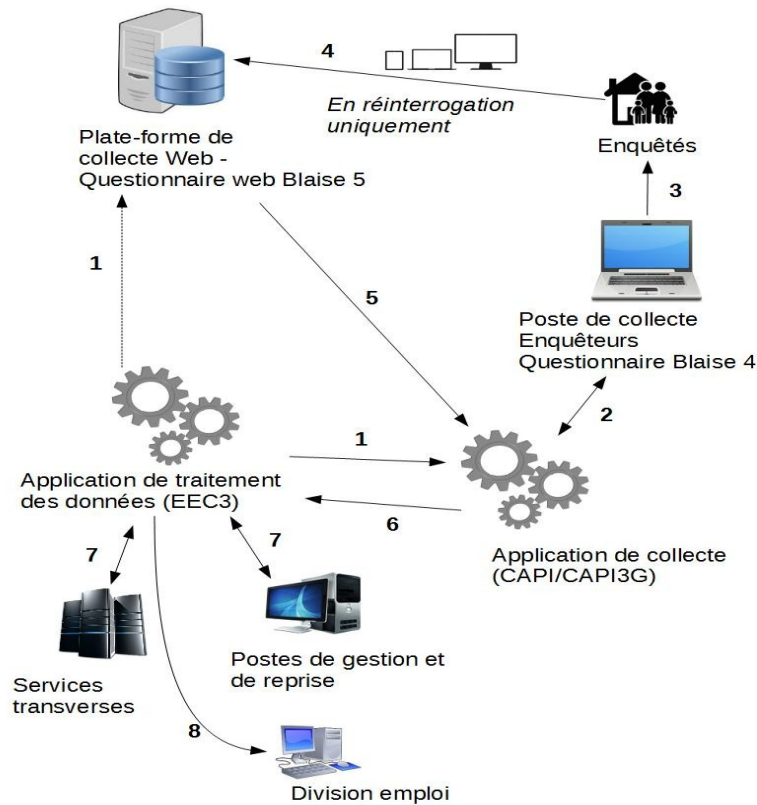
Grille GSBPM	Besoins
Concevoir	→ Spécifications d'un nouveau questionnaire en cohérence avec le règlement européen
Collecter	→ Développement d'une plateforme de collecte Web pour répondre au besoin de collecte par internet → Adaptation aux besoins multimode de l'application actuelle de collecte et de gestion (CAPI) → Adaptation de la chaîne de traitement des données de l'enquête Emploi actuelle (de l'application EEC2 à EEC3)
Traiter les données	→ Développement d'une nouvelle chaîne de traitements des données statistiques (nécessaire à la suite de la refonte du questionnaire et l'introduction du multimode) en self, gérée par la division Emploi
Analyser	→ Analyse de la rupture de série
Diffuser	→ Rétropolation des données

En termes d'outils et de développement informatique, la phase de collecte est plus particulièrement affectée. La nouvelle enquête Emploi repose sur trois briques applicatives majeures pour mener à bien la phase de collecte des données (schéma ci-dessous) :

- une application de traitement des données de l'enquête Emploi (dénommée par la suite « EEC3 », en remplacement de l'application « EEC2 » utilisée jusqu'alors) ;
- une application de collecte (Capi) avec un datamodel enquêteurs en Blaise4, Blaise étant le langage de programmation des questionnaires dans la filière d'enquête en vigueur au moment de la refonte ;
- une plateforme de collecte Web (portail d'authentification, transverse à plusieurs enquêtes par internet, et questionnaire internet en Blaise5).

Ces briques applicatives ont été conçues pour intégrer un système d'information pré-existant aux processus d'enquête ménages : outil de planification des activités liées aux enquêtes, utilisation de services transverses pour l'envoi de courrier, la codification automatique, la reprise manuelle, etc.

Schéma du processus de collecte multimode de la nouvelle enquête Emploi



- **La plateforme de collecte web**

Cette plate-forme prend la forme d'un site web dédié, se composant d'un portail d'authentification et du questionnaire de l'enquête Internet.

Le portail d'authentification assure deux groupes de fonctions : un portail d'accueil et une partie assistance.

Contrairement au questionnaire pour la collecte par enquêteur développé en Blaise 4, le questionnaire Internet a été développé en langage Blaise 5, plus à même de répondre aux besoins spécifiques de l'enquête en ligne que ne le pouvait Blaise IS (pendant Internet de Blaise 4). Les besoins en termes d'ergonomie et d'aide à la passation ne sont en effet pas les mêmes selon que la collecte se déroule par l'intermédiaire d'un enquêteur ou en auto-administré sur Internet. Les consignes aux enquêteurs doivent ainsi se transformer en rubriques d'aide à la passation en auto-administré sur Internet, et l'ergonomie change complètement. Le besoin fort de maintenir un seul code source, le souhait de conserver une structure de datamodel identique pour les deux modes, et les outils mis à disposition dans la suite logicielle Blaise ont conduit au choix de développement suivant : l'intégralité des développements sont réalisés dans le code source Blaise 4, y compris ceux spécifiques au questionnaire Internet. Ce code source intègre donc également des instructions en

langage Blaise 5, sous forme de commentaires et donc "non exprimées" dans le datamodel Blaise 4. Un applicatif de conversion de Blaise 4 en Blaise 5 est ensuite utilisé pour obtenir directement un datamodel Blaise 5. En outre, le langage Blaise 5 permet également le travail en *responsive design*, et donc d'adapter l'ergonomie du questionnaire selon le type de matériel utilisé lors de la connexion (PC portable, tablette, smartphone).

- **L'application de collecte (CAPI)**

L'application de collecte, utilisée par les acteurs locaux de la collecte, comprend un poste de collecte et un poste de gestion.

Le poste de collecte utilisé par l'enquêteur comprend l'échantillon des logements à enquêter, le questionnaire et des fonctionnalités de gestion et de suivi de la collecte. Le développement du multimode nécessite des fonctionnalités spécifiques : informer l'enquêteur que l'enquêté a commencé à remplir le questionnaire sur Internet ; désaffecter du poste de l'enquêteur les questionnaires validés sur Internet ; désaffecter du poste de l'enquêteur les FA web faisant l'objet d'un déménagement. Une mise à jour du casier enquêteur a lieu toutes les heures, et le carnet de tournée est mis à jour dès lors que le poste de l'enquêteur est connecté et que celui-ci en fait la demande.

Le gestionnaire de l'enquête Emploi en région dispose d'un poste de gestion contenant l'échantillon régional, le questionnaire et des fonctionnalités de gestion et de suivi de la collecte. Le développement du multimode nécessite des fonctionnalités spécifiques : récupérer des informations sur les questionnaires entamés ou validés sur Internet et les transmettre quotidiennement à l'application EEC3 ; concaténer dans une même base les FA issues de la collecte des enquêteurs et celles issues de la collecte web ; récupérer les informations de tentatives de contact entrées par les enquêteurs dans le datamodel ; permettre la réaffectation des FA web faisant l'objet d'un déménagement soit au même enquêteur (si le logement figure dans sa zone d'action) soit à un autre enquêteur.

- **L'application de traitement des données EEC3**

L'application de traitement des données permet de gérer les échantillons et de recevoir les données issues de la collecte. L'application précédente EEC2, entièrement réécrite lors de la précédente refonte de l'EEC en 2013, était peu sensible aux évolutions de questionnaire et répondait déjà à l'essentiel des besoins de traitement de la nouvelle enquête Emploi. Une refonte d'envergure de l'application n'a donc pas été nécessaire, mais une maintenance évolutive a été réalisée pour en moderniser l'architecture et l'adapter à une collecte multimode. Ainsi des fonctionnalités complémentaires ont été développées du fait du multimode en réinterrogation et pour centraliser le pilotage de gestion éditique et de mailing : appel aux web-services de génération identifiant/mot de passe et intégration à l'annuaire général ; envoi centralisé des courriers ; envoi centralisé des mails.

Un web-service est appelé par l'application EEC3, au moment de la constitution de l'échantillon Internet mensuel. Il permet de générer pour chaque unité à enquêter un couple d'identifiants (identifiant, mot de passe) pour permettre la connexion au site de collecte, ainsi que l'intégration de ces individus à l'annuaire général.

La centralisation des envois de courriers répond à deux objectifs : d'une part cela représente un gain de temps pour les enquêteurs et d'autre part cela garantit que le courrier est expédié à la bonne date,

ce qui est important du fait de la brièveté de la période de collecte. Les envois de courriers comme de mails sont pilotés de façon centralisée par l'application. Il a été fait le choix d'envois automatiques avec un calendrier défini au préalable.

Le positionnement de la frontière entre la partie de la chaîne gérée par l'application et celle réalisée en self au sein de la division Emploi a été réexaminé. Elle est finalement peu modifiée par rapport à la situation précédente, et ce pour plusieurs raisons. Principalement, les statisticiens de la division Emploi ont besoin de garder une certaine réactivité sur la chaîne, en particulier le temps de stabiliser les changements dus à l'évolution du questionnaire, qui ne se fera que lors des premières collectes. Par ailleurs, l'application Emploi est conçue pour stocker les données des différentes interrogations et s'adapter facilement aux évolutions du questionnaire d'une année sur l'autre car, en dehors des variables de gestion, le nom des variables n'apparaît pas dans le code de l'application. Pour conserver cet avantage, il est apparu plus opportun de maintenir les traitements statistiques en dehors de l'application.

III. Mesure des ruptures de série et rétopolation

La mise en œuvre d'un Pilote du T1 2020 au T1 2021 permet de mesurer les ruptures de série liées aux différents changements entre l'ancienne et la nouvelle enquête Emploi en comparant les principaux indicateurs du marché du travail dans les deux enquêtes. Cette comparaison est opérée sur cinq trimestres consécutifs, afin d'accroître la précision de la mesure de la rupture, ramenée en moyenne annuelle, et de détecter une éventuelle saisonnalité de ces ruptures. Si l'objectif premier était de mesurer les ruptures globales, un travail visant à quantifier les contributions des différentes évolutions de l'enquête à cette rupture a pu toutefois être mené, notamment en ce qui concerne les changements conceptuels ou les méthodes de pondération, la distinction entre les effets de mode et ceux du changement de questionnaire étant plus délicate.

Cette partie s'attache à décrire en détail les méthodologies de mesure des ruptures et de rétopolation utilisées. Les principaux indicateurs du marché de travail sont examinés un à un afin d'exposer leurs éventuelles spécificités quant à la méthodologie de rétopolation adoptée, ainsi que de tenter d'identifier et d'expliquer les différentes sources de rupture. Une première analyse des principaux résultats des indicateurs conjoncturels du marché du travail a été publiée dès juin 2021 (Dera, 2021), à l'occasion de l'estimation des résultats sur les indicateurs conjoncturels du premier trimestre 2021. Les ruptures les plus fortes (et donc les séries longues les plus fortement révisées) sont celles relatives à l'emploi (avec un taux d'emploi globalement révisé à la hausse), alors que les révisions sur les séries du chômage sont très faibles. Puis, en 2022 les travaux ont été poursuivis avec l'estimation et la diffusion des séries « structurelles » (celles qui ne sont pas diffusées à fréquence trimestrielle mais sur un pas annuel, dans les Insee Résultats dédiés). Il faut aussi noter que les travaux menés l'ont été sous contrainte de temps, afin de ne pas retarder les délais habituels de diffusion ; ils ne prétendent donc pas à l'exhaustivité. Notamment, d'autres analyses pourront être menées dans les années à venir, sur d'autres indicateurs ou sur l'impact spécifique de l'introduction d'Internet comme mode de réponse possible à l'enquête Emploi en réinterrogation.

1. Un ajustement des séries longues sur les nouvelles estimations de population

Même si cela ne faisait pas partie en tant que tel du projet de rénovation de l'enquête Emploi et des séries longues associées, la révision des séries longues du marché du travail en 2021 a été l'occasion de prendre en compte la nette **révision des estimations démographiques intervenue en 2018**, à la suite d'une modification dans le questionnaire du recensement de la population visant à améliorer la connaissance des situations de multi-résidence. L'objectif était effectivement de ne pas introduire de rupture de série en dehors de la période de refonte de l'enquête et d'évaluer et corriger l'ensemble des sources de ruptures au même moment.

En 2018, l'Insee a rénové une partie du questionnaire du recensement de la population, afin de mieux prendre en compte les situations de multi-résidence grâce à une meilleure connaissance des liens familiaux qui unissent les personnes habitant un même logement, ainsi que des lieux d'habitation des enfants dont les parents sont séparés⁸. Alors que certaines personnes en multi-résidence pouvaient auparavant être comptées à tort deux fois dans le recensement, ce nouveau questionnaire affectant la feuille de logement améliore la mesure de la population, avec un impact à

8 Voir https://www.insee.fr/fr/statistiques/fichier/2383177/fiche_complete.pdf

la baisse sur l'estimation du nombre total d'habitants en France, l'impact étant très concentré sur les moins de 20 ans, et en particulier chez les 10-19 ans, qui résident plus fréquemment que les autres dans deux résidences alternées du fait de la séparation de leurs parents.

La méthode du recensement annuel conduit à ce que l'effet total du changement de questionnaire, désigné par le terme « ajustement », monte en charge sur plusieurs années, affectant progressivement l'estimation des populations entre 2016 et 2022.

Les démographes n'ont pas réalisé de rétopolation « statistique » des séries de population totale, qui aurait neutralisé cet effet (mais ont fourni des ajustements statistiques année après année permettant de le faire). Or chaque trimestre, l'enquête Emploi est calée sur les estimations démographiques de population vivant en logement ordinaire. Pour éviter l'apparition de rupture sur les indicateurs économiques, notamment ceux mesurés en niveau, l'enquête Emploi a dans un premier temps continué d'être calée sur des marges *ad hoc* de population en « ancienne mesure » jusqu'à 2020, sur la base d'informations précises sur l'ajustement statistique, année par année et ventilées par sexe et classe d'âge. Puis dans un second temps une série de population « statistique » a été élaborée pour les besoins de l'enquête Emploi renouvelée.

La production d'une série longue « statistique » de population, à concepts inchangés dans le temps et cohérente avec les nouvelles estimations de population « à terme »

À l'occasion de la refonte de l'EEC de 2021, une rétopolation des populations⁹ a été réalisée afin que l'enquête Emploi soit calée sur des marges de population en « nouvelle mesure », « à terme » et à « concepts et méthodes inchangés » : il s'agissait de simuler les niveaux de population chaque année qui auraient été obtenus si la mesure avait été identique à celle adoptée depuis la rénovation de la feuille de logement ; mais comme la montée en charge des effets de l'ajustement s'étale de 2016 à 2022, les populations de l'enquête Emploi et celles des estimations de population ne coïncident à nouveau qu'à compter de 2023.

Ces nouvelles séries de population en « nouvelle mesure à terme » sont trimestrielles, débutent en 1975, et couvrent le champ des personnes de 15 ans ou plus, vivant en logement ordinaire en France hors Mayotte. Elles sont ventilées sur tous les croisements de sexe et d'âge quinquennal (jusqu'à 89 ans, puis avec la classe des 90 ans ou plus) et, à partir de 2014, sont disponibles par région¹⁰. Pour estimer ces nouvelles populations, des facteurs correctifs pour supprimer les doubles comptes sont calculés par sexe et tranche d'âge quinquennale et appliqués aux séries de populations en « ancienne mesure » allant de 1975 à 2019.

Les facteurs correctifs sont calculés à partir de l'ajustement annuel de 2016 à 2022, par âges fins et champs géographiques, estimés à partir des enquêtes annuelles du recensement par les démographes. Sur le champ France hors Mayotte, les ajustements agrégés par sexe et tranche d'âge quinquennale sont estimés sur toutes les années pour connaître l'ajustement total permettant de passer de l'ancienne mesure à la nouvelle mesure « à terme » (2023), résumé dans la *figure 19*.

9 Le champ de l'EEC est relatif à la population des logements ordinaires

10 Ce calage supplémentaire par région permet de recalibrer les micro-données de l'EEC sur ces nouvelles marges de population, dans le cadre de l'exercice de rétopolation des poids individuels de l'EEC pour 2014-2020.

Figure 19 : Ajustement estimé en nombre et en part de la population

	Hommes		Femmes	
	Ajustement		Ajustement	
âge quinq	En nb	En % pop	En nb	En % pop
00-04 ans	-18085		-17883	
05-09 ans	-52638		-51879	
10-14 ans	-66035		-65150	
15-19 ans	-53095	-2,65%	-50860	-2,64%
20-24 ans	-22196	-1,23%	-20788	-1,15%
25-29 ans	-10139	-0,54%	-9912	-0,50%
30-34 ans	-6108	-0,31%	-6033	-0,29%
35-39 ans	-4540	-0,23%	-4388	-0,21%
40-44 ans	-4466	-0,21%	-4267	-0,19%
45-49 ans	-5126	-0,23%	-4938	-0,22%
50-54 ans	-4486	-0,21%	-4399	-0,19%
55-59 ans	-3011	-0,15%	-2998	-0,14%
60-64 ans	-1564	-0,08%	-1610	-0,08%
65-69 ans	-969	-0,05%	-1005	-0,05%
70-74 ans	-428	-0,04%	-498	-0,04%
75 ans et+	-700	-0,03%	-991	-0,03%

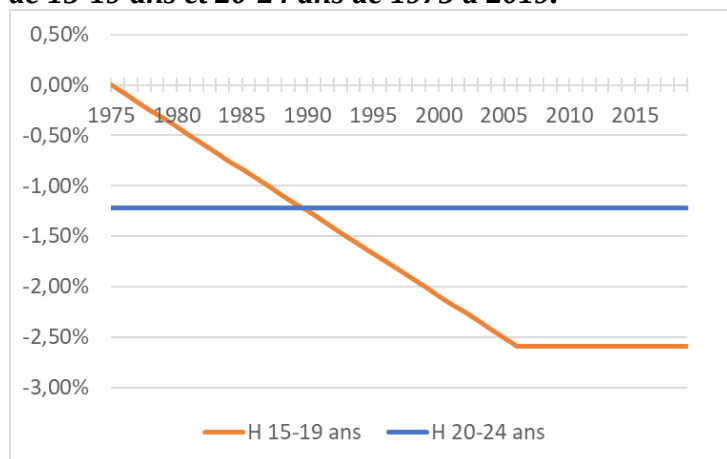
Lecture : à terme (2023), les révisions de population démographiques consécutives à la modification de la feuille logement du RP conduisent à une baisse du nombre d'hommes âgés de 15 à 19 ans de 53 095, soit une baisse de 2,65 % par rapport au niveau de 2016.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

Pour appliquer ces facteurs correctifs « plein régime » sur la période 1975-2019, des hypothèses supplémentaires ont été faites pour certaines tranches d'âge : en effet, les jeunes de moins de 20 ans sont les plus concernés par l'ajustement en raison des doubles comptes résultant des résidences alternées pour ceux dont les parents sont divorcés. Le nombre de doubles comptes dans cette tranche d'âge devrait donc être positivement corrélé au nombre de divorces. Ce dernier a plus que doublé entre 1975 et 2006 (Bellamy, 2016), passant de 56 000 en 1975 à 139 000 en 2006, et s'est stabilisé depuis. L'hypothèse d'une évolution linéaire de la part de l'ajustement dans la population des 15-19 ans entre 1975 (où l'ajustement serait nul) et 2006 (où l'ajustement atteint sa valeur de plein régime) a donc été faite (*figure 20*). Depuis 2006, l'ajustement est stable et fixé à sa valeur de plein régime.

Pour les autres tranches d'âge, l'ajustement est moins important, et ses déterminants beaucoup plus incertains. En l'absence d'informations plus précises sur son origine, le facteur correctif de plein régime est appliqué de manière constante sur toute la période 1975-2019.

Figure 20 : Facteurs correctifs de l'ajustement appliqués aux séries de population des hommes de 15-19 ans et 20-24 ans de 1975 à 2019.



Lecture : Pour les hommes de 15-19 ans, le facteur correctif appliqué aux populations passe de 0 à -2,65% entre 1975 et 2005, puis reste stable à cette dernière valeur à partir de 2006.

Champ : France hors Mayotte, personnes de 15-24 ans, vivant en logement ordinaire

Au total, les révisions démographiques, qui concernent surtout les enfants, ont un impact parfois important sur les indicateurs en niveaux mais négligeable sur les indicateurs du marché du travail calculés en taux, comme le taux de chômage ou le taux d'emploi.

2. Méthode générale de mesure des ruptures et de rétopolation des séries longues

2.1. Un Pilote pour mesurer les ruptures de séries

La mise en œuvre du Pilote permet d'avoir cinq trimestres (du 1^{er} trimestre 2020 au 1^{er} trimestre 2021) de « double » mesure des indicateurs, en ancienne et nouvelle enquête. Ceci permet de **mesurer l'effet global** des ruptures de séries, qui portent sur plusieurs dimensions : changements de concepts et de questionnaire ; introduction d'Internet en ré-interrogation qui produit des effets de mesure (les enquêtés répondent différemment sur Internet), de proxy (réponse des enquêtés directement plutôt que par proxy) et de sélection (nouveaux profils captés) ; nouvelle méthode de pondération (nouvelle base de sondage et amélioration du calage), nouvelles règles de rattachement et nouvelles populations (à la suite à la refonte de la feuille de logement du Recensement). Une analyse précise de **certains de ces effets** a été en outre réalisée pour identifier les différents mécanismes à l'œuvre (Dera, 2021).

Si ce Pilote constitue un outil de qualité inestimable pour l'étude des ruptures, il présente néanmoins l'inconvénient d'être de taille réduite : l'échantillon quatre fois plus petit en 2020 implique une précision deux fois moindre par rapport à l'enquête Emploi, ce qui pose problème pour certaines variables présentant des modalités rares dans la population. En outre, le hasard de calendrier a fait coïncider en partie la période de recouvrement du Pilote et la crise sanitaire, cette dernière ayant entraîné d'une part une baisse des taux de collecte, et, d'autre part, un bouleversement de certains indicateurs du marché du travail, notamment au deuxième trimestre 2020 (baisse en trompe-l'œil du chômage, forte hausse du chômage partiel, etc.). Après analyse spécifique, les ruptures n'apparaissant pas plus marquées au deuxième trimestre 2020 (qui a été

particulièrement marqué par le premier confinement de la population) qu’au premier trimestre, aucun traitement spécifique n’a été introduit pour prendre en compte ce paramètre. De même, les ruptures ne sont pas ressorties spécifiques au troisième trimestre 2020, qui est souvent le plus affecté par les variations saisonnières liées aux congés d’été. Au final, les ruptures de série ont été mesurées sur l’ensemble du champ de double couverture (cinq trimestres), sans estimer spécifiquement de paramètre saisonnier dans les ruptures ni de paramètre lié au premier confinement.

Ainsi, la comparaison directe entre le Pilote et les enquêtes en production sur la période de recouvrement a permis de quantifier précisément les ruptures de mesure introduite par le passage à la nouvelle enquête sur les principaux indicateurs macro-économiques du marché du travail, et de rétropoler en cette nouvelle norme les séries longues sur le marché du travail.

2.2. La mesure des ruptures de séries

La rupture de séries globale est mesurée, comme indiqué précédemment, en comparant les données en ancienne et nouvelle enquêtes en cumul du T1 2020 au T1 2021.

Les ruptures de mesure constatées peuvent être de différentes natures. Aussi, certaines estimations complémentaires ont été réalisées, afin de regarder l’impact de certains des changements.

Type de rupture	Méthode d’estimation
Changement de concept dans la mesure du statut BIT	Estimation via les micro-données en simulant les changements de questionnaire (par exemple : intégration ou non d’une absence d’emploi pour longue maladie dans l’emploi).
Questionnaire	Comparaison de l’ancienne / nouvelle enquête en rang 1 (première interrogation en face-à-face), en gardant la même méthode de rattachement/pondération pour les deux enquêtes. En effet, le rang 1 n’est pas affecté par les changements de protocole (la collecte demeure en face à face).
Règles de rattachement	Simulation, dans la nouvelle enquête, des nouvelles règles de rattachement, et comparaison avec les anciennes règles de rattachement.
Pondération	Application de la nouvelle méthode de pondération à l’ancienne enquête (à règles de rattachement inchangées, pour estimer le seul effet pondération).
Nouvelles marges démographiques	Calage des données de la nouvelle enquête selon les deux concepts de marges démographiques (voir 1.).
Protocole	Effet résiduel une fois les autres effets contrôlés

L’estimation du nouveau protocole et en particulier l’impact de l’introduction d’Internet en réinterrogation est plus complexe. On peut identifier en théorie trois composantes, qui sont difficiles à distinguer en pratique :

- un effet de mesure : variations de réponse pour un même enquêté entre les modes de collecte,
- un effet proxy : écart de réponse lorsque l'enquêté ciblé répond directement avec le cas où un autre membre du logement répond pour lui car il n'est pas disponible,
- un effet de sélection : nouvelles personnes répondantes.

Par ailleurs, les effets liés à l'introduction d'Internet peuvent être difficilement dissociables de ceux du changement de questionnaire, dans la mesure où les internautes peuvent par exemple lire les consignes inhérentes à cette question (un changement de consigne peut être soit un effet Internet soit un effet de questionnaire) : à l'inverse de l'effet pur du changement de questionnaire qui peut être estimé grâce à l'enquête Pilote, celui d'Internet est donc plus difficile à isoler.

Toutes les analyses sur l'impact de la contribution d'Internet à la rupture de série globale n'ont pu être menées à ce jour et vont se poursuivre dans les années à venir. Néanmoins, les principaux écarts constatés sur les indicateurs conjoncturels de chômage, emploi, halo et inactivité ont été analysés et présentés dans Dera, 2021.

2.3. La méthode générale de rétopolation des séries longues annuelles et trimestrielles

Les séries longues annuelles d'indicateurs sur le marché du travail sont publiées chaque trimestre dans une Informations Rapides et, sur un spectre souvent plus large d'indicateurs, chaque année dans un Insee Résultats, et visent à quantifier l'évolution de différents phénomènes économiques sur longue période. Afin de s'assurer que ces dernières ne sont pas affectées par des changements de mesure, il est nécessaire de rétopoler ces séries longues à chaque refonte de l'enquête Emploi. La rétopolation consiste en effet à ajuster les observations passées afin qu'elles reflètent les valeurs qui auraient été mesurées **si l'enquête Emploi avait toujours été telle que dans sa version la plus récente** : les séries rétopolées sont ainsi à mesure inchangée, cohérentes avec les concepts, questionnaires, protocoles et méthodes de pondération actuels y compris dans le passé. Il s'agit donc d'un exercice théorique : Internet par exemple n'existait pas en 1975. La suite de cette partie expose les éléments de méthode retenus.

- **La dilatation**

La méthode générale de rétopolation des séries longues en 2021 repose sur l'hypothèse d'un ratio constant dans le temps entre la mesure en nouvelle norme (dite « EEC3 ») et celle en ancienne norme (« EEC2 »). Il suffit alors de « dilater » les anciennes séries en les multipliant par ce ratio, nommé coefficient de dilatation, pour obtenir leur valeur en nouvelle norme. Le calcul de ce coefficient de dilatation est réalisé sur les cinq trimestres avec double mesure. Ainsi, pour chaque sous-population étudiée (souvent un croisement entre le sexe et un regroupement d'âges), le coefficient de dilatation de l'indicateur mesuré est donné par la formule suivante :

$$\beta = \frac{indic_{EEC3}}{indic_{EEC2}}$$

Et pour la série historique (en norme EEC2) de cet indicateur, à chaque date t on a :

$$indic_t^{dilaté} = \beta \times indic_t^{EEC2}$$

Dans de très rares cas où la valeur de l'indicateur mesuré pour un croisement de sexe et d'âge serait nulle en ancienne enquête (par exemple pour les hommes de 15-24 ans agriculteurs au chômage), le coefficient de dilatation est arbitrairement fixé à 1.

Il est important de différencier la rupture dans la mesure de l'indicateur selon ses variables explicatives, sexe et âge par exemple : la hausse de l'emploi liée à la requalification de certains congés parentaux en emploi ne touche pas uniformément les deux sexes et toutes les classes d'âge. Désagréger le calcul de la rupture comporte toutefois certaines limites, d'autant plus dans le contexte d'un Pilote à l'effectif réduit : la volatilité peut être importante pour certaines classes aux effectifs faibles, et l'écart entre ancienne et nouvelle enquêtes important sans que cela ne traduise une réalité économique ou statistique.

Ces faibles effectifs du Pilote incitent, pour plus de robustesse, à calculer le coefficient de dilatation en moyenne sur les 5 trimestres consécutifs (T1 2020 à T1 2021 inclus) où les deux mesures en ancienne et nouvelle enquêtes sont possibles, puis à l'appliquer aux indicateurs en moyenne annuelle. Certains indicateurs étant également diffusés au niveau trimestriel, le calcul de coefficients de dilatation trimestriels peut en théorie permettre de détecter une rupture différenciée selon les trimestres, consécutive à la saisonnalité de certains phénomènes économiques. Néanmoins, l'effectif du Pilote utilisé pour le calcul de chaque rupture étant divisé par 4 par rapport à une enquête en production habituelle, les écarts constatés entre les coefficients de dilatation trimestriels sont relativement volatiles. En tenant compte de cette imprécision, ils ont toutefois été analysés pour déterminer si les coefficients du troisième trimestre (le plus marqué habituellement par les variations saisonnières, liés aux congés estivaux) apparaissaient nettement différents des autres, ce qui n'est ressorti pour aucun indicateur. Pour ces raisons, les ruptures annuelles sont appliquées uniformément à tous les trimestres des séries trimestrielles.

- **Le lissage**

Afin d'accroître encore la robustesse des coefficients de dilatation calculés, les valeurs de ces derniers sont lissées entre les classes d'âge quinquennal consécutives sans signe distinctif particulier : en effet, pour tous les indicateurs déclinés par âge quinquennal, si les plus jeunes et les plus âgés peuvent présenter des comportements particuliers qui justifient des coefficients de dilatation plus dispersés, ce n'est pas le cas pour les âges intermédiaires consécutifs, pour lesquels d'éventuels sauts entre les valeurs de β s'expliquent plus probablement par un bruit statistique que par des comportements atypiques. Ainsi, les coefficients de dilatation par âge quinquennal sont lissés par sexe à l'aide de moyennes mobiles d'ordre 3 pour les 25-64 ans (à l'exception des 25-29 ans pour lesquels la moyenne n'est faite qu'avec le coefficient des 30-34 ans, et des 60-64 ans avec celui des 55-59 ans). Lorsque ce lissage ne suffit pas, notamment dans de rares cas de valeurs aberrantes (reposant sur de très petits échantillons), ou que les classes d'âge plus agrégées ne le justifient pas, les coefficients de dilatation ont pu être arbitrairement modifiés.

- **Méthode spécifique**

Il arrive néanmoins que la méthode de rétopolation générale ne puisse être appliquée telle quelle sur l'ensemble des années depuis 1975 : c'est le cas par exemple de la mesure de l'emploi, qui augmente en nouvelle enquête en 2020 en raison notamment de la requalification en emploi des congés parentaux supérieurs à 3 mois mais avec perception d'un revenu compensatoire. Le coefficient de dilatation entre l'emploi 2020 en ancienne norme et celui en nouvelle norme inclut

donc ce changement de concept. Le congé parental rémunéré n'ayant été introduit en France qu'en 1984 et ayant subi de nombreuses réformes depuis, l'hypothèse d'un ratio fixe dans le temps entre ancienne et nouvelle norme depuis 1975 n'est pas crédible, et des ajustements doivent être faits pour retracer plus fidèlement l'évolution du congé parental dans les séries historiques d'emploi (cf. *infra*).

- **Le calage**

Les séries d'emploi, chômage et inactivité par sexe et âge quinquennal sont rétropolées en premier et recalées sur la population totale, tandis que les séries longues des autres indicateurs (par exemple : emploi par catégories socioprofessionnelles) ne sont rétropolées que dans un deuxième temps afin de pouvoir elles-mêmes être calées sur l'emploi, le chômage et l'inactivité.

Sauf rares exceptions, ce calage a été fait en ajustant toutes les sous-séries concernées proportionnellement à l'effectif de chaque croisement : par exemple, pour le statut d'activité $a \in \{empl, chom, inact\}$, chaque année t et pour chaque sexe s et âge quinquennal q , l'effectif E calé est obtenu de la manière suivante :

$$E_{a_t^{s \times q}}^{calé} = E_{a_t^{s \times q}}^{dilaté} \times \frac{PopulationTotale_t^{s \times q}}{\sum_a E_{a_t^{s \times q}}^{dilaté}}$$

L'alternative qui aurait consisté à faire porter tous les écarts d'ajustement sur une seule sous-série (par exemple ici en ajoutant à l'inactivité le solde entre la population totale et la somme des effectifs dilatés) n'a pas été retenue, car elle n'assurait pas la fiabilité de chacune des sous-séries : les écarts entre les valeurs dilatées et recalées de la série choisie pour supporter l'ajustement pourraient en effet être importants notamment pour les croisements avec de petits effectifs.

Lorsque la série qui doit être rétropolée est à un pas trimestriel, la rétropolation de la série correspondante annuelle est d'abord effectuée, afin que le recalage de la série trimestrielle dilatée puisse être fait non seulement sur les séries mères trimestrielles, mais également sur la nouvelle série annuelle en nouvelle norme, en faisant coïncider la moyenne annuelle de la série trimestrielle avec la série annuelle rétropolée. Pour ce faire, les écarts entre la série annuelle rétropolée et la moyenne annuelle de la série trimestrielle dilatée sont trimestrialisés en utilisant la méthode de Denton-Cholette (Denton, 1971, corrigée par Cholette, 1984). Le facteur correctif qui en résulte est alors additif, ce qui assure une cohérence parfaite entre la série annuelle rétropolée et la série trimestrialisée.

Notons que dans certains rares cas, les classes d'âge des séries trimestrielles historiques sont plus fines que celles des séries annuelles : dans ce cas, la série trimestrielle rétropolée telle que décrit ci-avant est celle avec les mêmes agrégats d'âge que la série annuelle, puis une étape supplémentaire est nécessaire pour estimer la répartition entre les sous-classes d'âge, à l'aide de la méthode dite des « parts logistiques ». Il faut d'abord convertir la série trimestrielle en ancienne norme (EEC2) en part p de chaque sous-classe d'âge a dans l'agrégat supérieur à chaque trimestre. Cette même part p_a est calculée au niveau annuel (pour plus de robustesse) en 2020 en ancienne (EEC2) et nouvelle norme (EEC3). Pour chaque sous-classe d'âge, une « marche » est calculée en logarithme pour s'assurer de rester sur des valeurs comprises entre 0 et 1 :

$$\text{marche}_a = \log\left(\frac{p_a^{EEC3}}{1-p_a^{EEC3}}\right) - \log\left(\frac{p_a^{EEC22020}}{1-p_a^{EEC22020}}\right)$$

C'est cette marche qui permet de rétopoler la série des parts trimestrielles de l'ancienne à la nouvelle norme, chaque trimestre de chaque année, de telle façon que

$$\log\left(\frac{p_a^{EEC2t}}{1-p_a^{EEC2t}}\right) + \text{marche}_a = \log\left(\frac{p_a^{EEC3t}}{1-p_a^{EEC3t}}\right)$$

C'est-à-dire :

$$p_a^{EEC3t} = 1 - \frac{1}{1 + \exp\left(\log\left(\frac{p_a^{EEC2t}}{1-p_a^{EEC2t}}\right) + \text{marche}_a\right)}$$

Ces nouvelles parts sont ensuite recalées pour s'assurer que dans chaque agrégat d'âge les parts somment bien à 1. Une fois cette série de part trimestrielle de chaque sous-classe d'âge dans l'agrégat supérieur convertie en norme EEC3, elle peut être appliquée à la série trimestrielle rétopolée par âge agrégé en niveaux pour aboutir à la série trimestrielle rétopolée par âge fin.

Dans le cas général, les étapes de la rétopolation peuvent donc être résumées ainsi :

- Calcul du coefficient de dilatation à partir du Pilote permettant de comparer sur 5 trimestres ancienne et nouvelle normes
- Lissage des coefficients de dilatation pour les âges intermédiaires si la série est ventilée par âges quinquennaux
- Dilatation de la série annuelle en ancienne norme
- Recalage de la série annuelle dilatée sur la série annuelle « mère » (marges démographiques ou séries de statut BIT)
- Si la série historique trimestrielle existe, elle est dilatée en utilisant les coefficients de dilatation annuels ; ses écarts en moyenne annuelle avec la série rétopolée annuelle sont trimestrialisés pour que la série trimestrielle soit bien calée sur la série annuelle
- Recalage de la série trimestrielle dilatée et corrigée sur la série trimestrielle « mère » (marges démographiques)

Toute la rétopolation concerne des séries en données brutes (non corrigées des variations saisonnières, CVS) ; puis, les estimations des séries CVS ont été recalculées dans un second temps.

3. Les ruptures de séries et rétopolations sur le statut BIT

Dans cette partie, les ruptures de séries et la rétopolation associée sont présentées pour le statut BIT (emploi, chômage ou inactivité). Les autres séries rétopolées sont présentées en annexe (Annexe 1).

Les séries relatives au statut BIT en niveau sont rétopolées par sexe et âge quinquennal, d'abord à un pas annuel : les coefficients de dilatation calculés sont lissés avant d'être appliquées aux trois

séries historiques, puis les séries dilatées qui en résultent sont recalées pour que, dans chaque classe de sexe et âge quinquennal, la somme des personnes en emploi, au chômage et inactives corresponde bien à la population totale.

Figure 21 : Coefficients de dilatation entre ancienne et nouvelle enquêtes sur l'emploi, le chômage et l'inactivité par sexe et âge quinquennal



Lecture : Les hommes de 15-19 ans au chômage sont 1,5 fois plus nombreux en nouvelle enquête que dans l'ancienne enquête sur la période de recouvrement.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

Source : Insee, enquêtes Emploi

Pour ces différents statuts d'activité, les coefficients de dilatation apparaissent proches de 1,0 pour les classes d'âge intermédiaire, bien plus éloignés pour les classes d'âges « extrêmes » (figure 21) ; s'il y a quelques raisons liées au changement de questionnaire, il est nécessaire de rappeler : d'une part que le statut d'emploi ou de chômage concerne très peu de personnes parmi les plus jeunes (15-19 ans) ou parmi les seniors (au-delà de 65 ans), ce qui rend particulièrement peu précis les calculs de coefficients de dilatation (sans conséquence majeure, trois fois rien reste peu) ; d'autre part que, contrairement aux autres classes d'âges, les coefficients de dilatation n'ont pu être lissés avec les classes d'âge les plus proches.

Pour les classes d'âges davantage concernées par l'activité, l'emploi augmente (d'une norme à l'autre) légèrement pour les femmes d'âge intermédiaire ainsi que pour les hommes et les femmes de 55 à 64 ans (jusqu'à 69 ans chez les femmes), en raison notamment de la requalification en emploi d'une partie des congés parentaux et des congés maladie. Pour les jeunes de 20 ans ou plus, la hausse de l'emploi peut s'expliquer à la fois par une modification des règles de rattachement des étudiants en double-résidence, par la réduction des recours au proxy grâce à Internet¹¹, ainsi que par un effet questionnaire qui semble mieux recueillir les petits boulots ou stages rémunérés.

Enfin, l'emploi des 70 ans et plus baisse nettement (coefficient très nettement inférieur à 1). Outre les éléments de méthodes mentionnés, ceci peut s'expliquer par la non-remise en collecte des logements habités par des inactifs de 70 ans et plus : ils sont désormais considérés comme

11 Ainsi, le taux global de réponse par proxy dans l'EEC est passé de 28.5% à 24.1% entre 2020 et 2021. Le taux de recours par proxy est de 32.6% en réinterrogation hors Internet, contre 11.3% en réinterrogation sur Internet.

répondants lors des rangs suivants et leurs réponses sont imputées comme identiques à l'interrogation précédente. Les éventuels seniors qui rentreraient à nouveau dans l'emploi ne sont donc plus comptabilisés. Le taux d'emploi des 70 ans ou plus étant très faible (2,0 % en 2021, en première interrogation), ce changement de protocole n'a pas d'effet sur le taux d'emploi des 15 ans ou plus.

Cette modification du protocole de collecte explique sans doute également la baisse du chômage parmi les 70 ans et plus. Pour les jeunes, la hausse du chômage peut s'expliquer à la fois par la baisse du recours au proxy (connaissance plus détaillée de ses propres démarches de recherche d'emploi) et par les modifications du questionnaire (consignes avec des exemples de démarches de recherche).

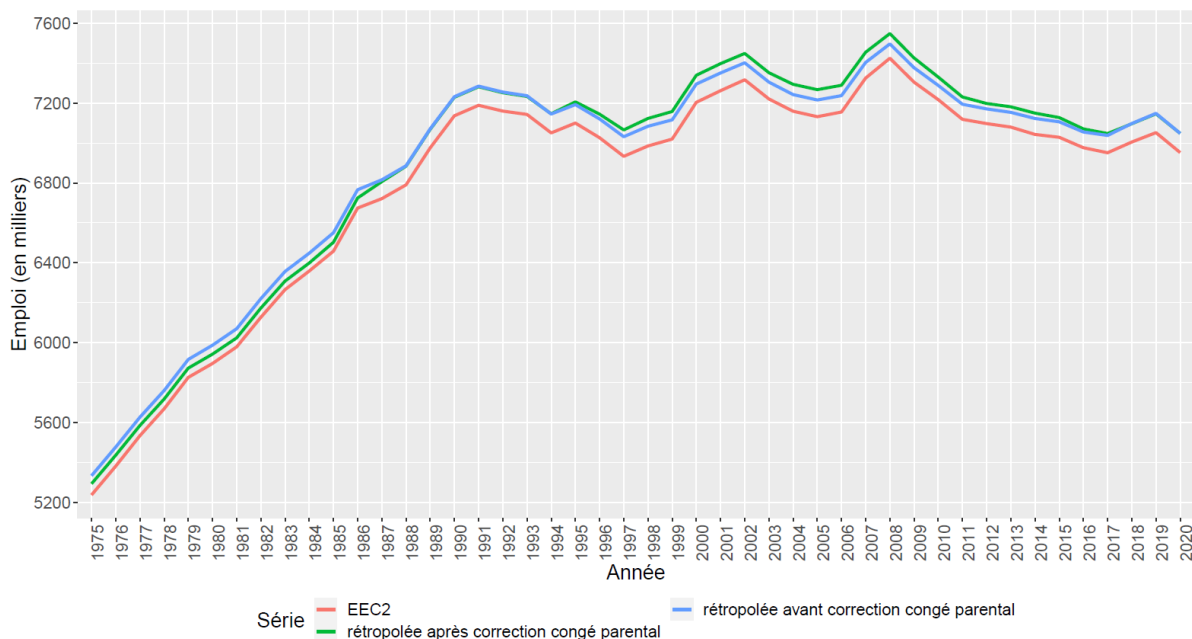
Enfin, pour les femmes d'âge intermédiaire notamment, la baisse de l'inactivité est la contrepartie directe de la hausse de l'emploi et du chômage.

Chez les femmes d'âge intermédiaire, la rupture sur l'emploi s'explique en partie par un changement conceptuel relatif à la considération du congé parental comme emploi : c'est désormais le cas pour tous les congés parentaux rémunérés quelle qu'en soit la durée, alors qu'il fallait auparavant qu'ils soient inférieurs à 3 mois. Or de 1975 à 1984, le congé parental rémunéré n'existait pas, les femmes qui voulaient s'arrêter pour élever leur enfant devaient quitter leur emploi, ce changement de concept n'aurait donc pas eu d'effet sur l'emploi s'il avait été mis en œuvre sur cette période. Depuis 1984, le périmètre, les conditions financières, de reprise d'emploi et la durée du congé parental ont évolué, modifiant les incitations à le prendre et donc le nombre de personnes concernées. Les effets sur l'emploi du changement de concept doivent donc être ajustés après la dilatation, ce qui nécessite une estimation du nombre de congés parentaux sur toute la période.

Depuis 2006, il existe dans l'enquête Emploi une modalité qui distingue le congé parental des congés non rémunérés parmi les raisons d'absence la semaine de référence pour les personnes ayant un emploi : le nombre de personnes basculant en emploi du fait du changement de concept peut donc être déterminé de façon exacte sur 2006-2021. Sur la période 1984-2005, seul le point 1992 est connu avec certitude, une enquête de la Dares évaluant à 100 000 le nombre de congés parentaux cette année-là (Renaudat, 1993). L'interpolation des points manquants a été faite à partir de la série historique du nombre de bénéficiaires de l'allocation parentale d'éducation (APE, qui deviendra ensuite la CLCA puis la PREPARE), destinée à un parent de 3 enfants ou plus qu'il soit en emploi ou non avant la naissance : l'hypothèse est celle d'une corrélation entre le nombre de bénéficiaire de l'APE et celui du congé parental.

La série ainsi créée doit ensuite être désagrégée par sexe et âge quinquennal : dans la mesure où en 2020, quasiment tous les congés parentaux sont pris par des femmes, on suppose que c'était le cas dans le passé ; en revanche les femmes bénéficiant du congé parental en 2020 sont majoritairement âgées de 30 à 39 ans, mais l'âge moyen du premier enfant s'est décalé de 5 ans entre 1975 (24 ans) et 2020 (29 ans). Il est donc probable qu'il en soit de même pour l'âge des femmes bénéficiant du congé parental. La distribution par âge des congés parentaux est donc créée en décalant progressivement les parts de chaque âge quinquennal dans le total des femmes en congé parental sur 1975-2006. Les effectifs ainsi obtenus sont retirés de l'inactivité et ajoutés à l'emploi, ce qui permet de corriger la méthode de rétroproportionnement générale de l'emploi des femmes de 20 à 44 ans (*figure 22*).

Figure 22 : Prise en compte du congé parental dans la rétropolation de l'emploi des femmes de 20-44 ans



Lecture : En 2008, en ancienne enquête, 7 425 000 femmes entre 20 et 44 ans sont en emploi. Après dilatation et calage sur la population totale, ce nombre s'élève à 7 496 000. En nouvelle norme de l'enquête, c'est-à-dire après correction des congés parentaux, il atteint 7 548 000.

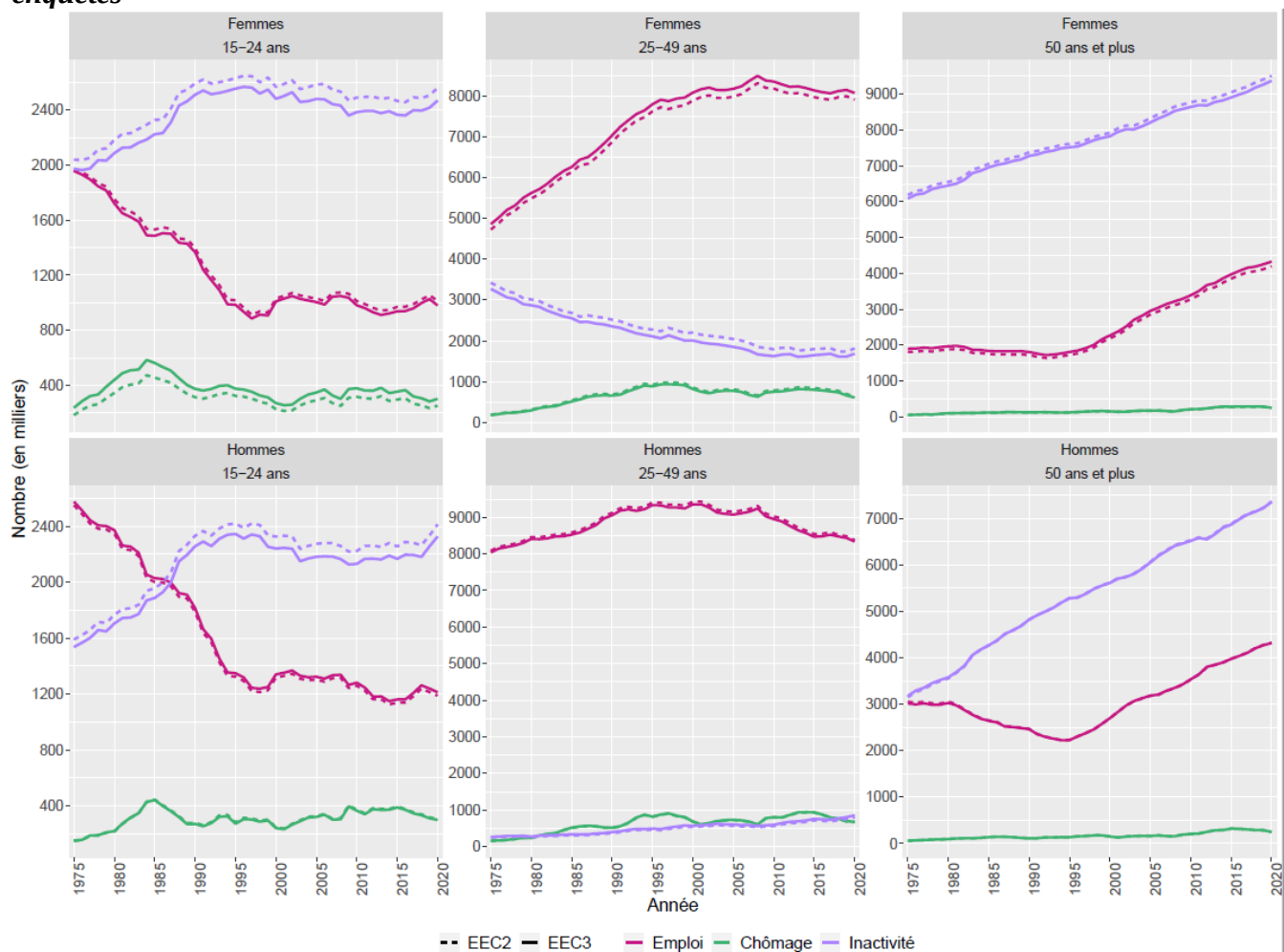
Champ : France hors Mayotte, femmes de 20 à 44 ans, vivant en logement ordinaire.

Source : Insee, enquêtes Emploi.

Les séries annuelles de statut d'activité par sexe et âge agrégé avant et après la rétropolation sont présentées *figure 23*.

Les séries trimestrielles d'emploi, chômage et inactivité par sexe et âge quinquennal peuvent ensuite être rétropolées à leur tour en suivant la méthode générale.

Figure 23 : Séries longues des statuts d'activité par sexe et âge regroupé en ancienne et nouvelle enquêtes



Lecture : En 2020, en ancienne enquête (EEC2), 2 562 000 femmes de 15-24 ans sont inactives, contre 2 470 000 en nouvelle enquête (EEC3).

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

4. La révision des poids dans les micro-données

En plus de la rétopolation des séries agrégées sur les principaux indicateurs du marché du travail de l'enquête Emploi, une pondération individuelle rétopolée a été calculée. Le principal objectif de cette rétopolation des pondérations individuelles est de permettre à l'utilisateur des micro-données de retrouver les séries agrégées préalablement rétopolées, en agrégeant les données individuelles. Il est également utile de pouvoir calculer des évolutions sur d'autres indicateurs de l'enquête ou sur d'autres sous-populations, qui n'ont pas fait l'objet de rétopolation agrégées. De fait, la mesure des ruptures de série et la rétopolation n'a pu se faire que sur un nombre restreint d'indicateurs, les principaux indicateurs du marché du travail, mais pas sur l'ensemble des variables de l'enquête. L'application d'un jeu de pondérations individuelles permet de conserver la cohérence entre ces calculs sur les autres indicateurs de l'enquête et les indicateurs qui ont fait l'objet d'une rétopolation.

La méthode mise en place consiste à modifier les poids des années précédant la refonte de façon à ce qu'ils satisfassent des marges sur les niveaux rétopolés des principaux indicateurs. Le principe de rétopolation est le même que celui opéré lors de la dernière refonte de l'enquête Emploi en 2013, lorsque les pondérations individuelles pour 2003-2012 avaient été rendues cohérentes avec les nouvelles séries longues démarrant en 2013. Pour le nouvel exercice, seuls les millésimes de 2014 à 2020 ont été rétopolés sur les nouvelles séries longues démarrant en 2021, du fait de la restriction de l'EEC à la France métropolitaine jusque 2013.

D'un point de vue pratique, un recalage a été effectué pour chaque trimestre entre le 1^{er} trimestre 2014 et le 4^e trimestre 2020, à partir des variables initiales de pondération. Ces variables ont été recalées sur un ensemble d'indicateurs de façon à retrouver les mêmes résultats que pour les séries longues rétopolées sur cette période. La liste de ces indicateurs et le niveau de détail pris en compte pour les marges sont présentés dans la *figure 24*. Par ailleurs, comme indiqué précédemment, en même temps que la révision des indicateurs liés au marché du travail, une forte révision des marges démographiques est opérée, ce qui implique un calage concomitant avec de nouvelles marges démographiques.

Figure 24 : Les marges prises en compte dans la pondération rétopolée

Indicateur	Source	Calage par rang	Niveau de détail
Statut BIT x âge x sexe	Nouvelles séries longues du marché du travail	Non	Âge quinquennal Emploi : Regroupé au-dessus de 75 et plus Chômage : Regroupé au-dessus de 65 et plus
Statut BIT par Dom x âge x sexe	Nouvelles séries longues du marché du travail	Non	Âge en 4 tranches (15-24, 25-49, 50-64, 65+)
Emploi par statut (<i>nouvelle variable STATUDET</i>)	Nouvelles séries longues du marché du travail	Non	Aucun
Emploi par groupe socioprofessionnel x sexe	Nouvelles séries longues du marché du travail	Non	PCS à 1 chiffre avec détail qualifié / peu qualifié pour employés et ouvriers
Temps partiel par âge x sexe	Nouvelles séries longues du marché du travail	Non	Âge en 3 tranches (15-24, 25-49, 50+)
Sous-emploi par âge x sexe x composante	Nouvelles séries longues du marché du travail	Non	Âge en 3 tranches (15-24, 25-49, 50+) Composante en 3 modalités
Halo par âge x sexe	Nouvelles séries longues du marché du travail	Non	Âge en 4 tranches (15-24, 25-49, 50-64, 65+)
Halo par composante	Nouvelles séries longues du marché du travail	Non	Composante en 3 modalités
Halo par Dom x âge x sexe	Nouvelles séries longues du marché du travail	Non	Âge en 4 tranches (15-24, 25-49, 50-64, 65+)
Chômage de longue durée par âge x sexe	Nouvelles séries longues du marché du travail	Non	Âge en 3 tranches (15-24, 25-49, 50+)
NEET par âge x sexe	Nouvelles séries longues du marché du travail	Non	Âge en 3 tranches (15-19, 20-24, 25-29)
Population totale par âge x sexe	Démographie	Oui	Âge quinquennal Regroupé au-dessus de 75 ans et plus
Population totale par région	Démographie	Oui	Anciennes régions (26)

Même si le nombre d'indicateurs pris en compte dans la rétopolation de 2021 est important, il n'est pas possible de traiter l'ensemble des ruptures de séries induites par la refonte. D'une part, il n'est pas possible de distinguer chacune de ces ruptures, dont les causes peuvent être très variées (changement de questionnaire, impact du protocole ou du nouveau règlement, etc.). D'autre part, la prise en compte d'un nombre trop important d'indicateurs dans le calage entraînerait une trop forte proportion de poids élevés, ce qui nuirait à la fiabilité des résultats issus de la pondération rétopolée. Dans le cas présent, malgré les nombreuses contraintes de calage supplémentaires imposées, la rétopolation des poids ne conduit pas à une dispersion des poids trop importante : 99 % des poids rétopolés sont inférieurs à 1 883, contre 1 801 pour les poids avant rétopolation ; aucun poids rétopolé n'est supérieur à 10 000 ; dans 80 % des cas, le ratio entre poids après/avant rétopolation est compris entre 0,8 et 1,2.

L'utilisation de la pondération rétopolée peut modifier les résultats pour des indicateurs et pour des sous-populations autres que ceux pris en compte explicitement comme marges. Par exemple, avec la refonte, on observe une hausse de la proportion des personnes se déclarant dans l'EEC inscrites à Pôle emploi. Cet indicateur n'est pas rétopolé en tant que tel, mais son niveau augmente avec l'utilisation de la pondération rétopolée, car il est très lié au halo autour du chômage (qui est lui rétopolé en tant que tel).

Cependant, l'utilisation des pondérations rétopolées ne permet pas de corriger toutes les ruptures potentielles sur les différentes variables de l'enquête. Il reste donc important pour les utilisateurs des données individuelles d'analyser précisément l'existence d'une rupture de série éventuelle entre 2020 et 2021 sur les indicateurs non rétopolés.

Par ailleurs, la rétopolation n'a pas conduit à modifier les caractéristiques au niveau individuel, mais seulement les pondérations : les calculs effectués par exemple sur les transitions individuelles entre statuts d'activité, dont les concepts ont été redéfinis avec la refonte, ne sont plus pertinentes si l'on compare des périodes avant et après 2021 (et ce quelle que soit la pondération).

IV. Bilan du projet

1. Un périmètre revu partiellement

Le périmètre de la refonte a été réévalué au cours du projet, prenant en compte des contraintes externes et une priorisation des objectifs au vu de la charge et du calendrier du projet. Parmi ces principaux changements, plusieurs volets ont été remis à plus tard : l'intégration de Mayotte dans l'enquête Emploi en continu ; la refonte de la chaîne aval en R. Par ailleurs l'ambition d'intégrer les outils de la nouvelle filière des enquêtes ménages de l'Insee (programme « Métallica ») n'a dès le départ pas été retenue dans le projet Mélopée mais fait l'objet d'un nouveau projet, commencé juste après.

Concernant Mayotte, le règlement IESS impose de réaliser l'enquête sur tout le territoire national selon les règles européennes. Ce n'est pas le cas de la France puisque le département de Mayotte n'a pas intégré l'enquête Emploi en continu. De fait, une enquête annuelle, comme c'était le cas en métropole jusqu'en 2002 et dans les autres Dom jusqu'en 2013, a lieu à Mayotte au printemps de chaque année, avec un questionnaire différent. L'intégration de Mayotte dans l'enquête Emploi en continu dans le cadre du projet Mélopée a été abandonnée du fait d'une charge d'enquête trop importante à Mayotte avant 2021, d'autres opérations étant également mises en place pour la première fois sur ce territoire. Une dérogation au règlement a été obtenue par la France jusqu'en 2023. Un nouveau projet (« Mayottica ») a pris la suite du projet Mélopée pour mettre en œuvre l'intégration de Mayotte dans l'enquête Emploi en continu à partir de janvier 2024 ; le même protocole que pour l'intégration des 4 DOM historiques en 2014 est mis en œuvre, avec une montée en charge dès janvier 2023.

Concernant les outils de la nouvelle filière des enquêtes ménages de l'Insee, l'inscription de l'enquête Emploi dans ces nouveaux outils dans le cadre du projet Mélopée n'a pas été retenue car s'est vite avérée trop risquée au vu du calendrier imposé par la mise en place du nouveau règlement européen et du périmètre des fonctionnalités opérationnelles dans la nouvelle filière à cet horizon. Mélopée garde toutefois une adhérence avec le programme Métallica, puisque l'enquête Emploi utilise le portail d'enquêtes ménages de l'Insee, Coleman, pour la collecte sur Internet. Le nouveau projet (Mayottica) qui a pris la suite du projet Mélopée comprend, outre l'intégration de Mayotte à l'enquête Emploi en continu, l'intégration aux outils de la nouvelle filière.

Concernant la refonte en R de la chaîne aval, il s'est avéré plus sûr de refondre la chaîne dans le langage pré-existant (SAS) dans un premier temps. De fait, une refonte du langage de programmation concomitant avec la refonte du contenu de la chaîne constitue un coût important puisqu'il est également nécessaire de tester l'absence d'impact de ce changement de langage sur les résultats : modifier le langage de programmation aurait constitué un risque sur la capacité à produire les résultats trimestriels dans le calendrier requis. Ce projet de refonte en R de la chaîne aval constitue un chantier à part après la stabilisation de la chaîne et de sa documentation ; il a été entamé fin 2022.

2. Bilan de la réalisation du projet

Les enjeux du projet initial pour assurer sa réussite étaient de plusieurs ordres :

- un enjeu de continuité de la production de l'enquête, l'enquête étant réalisée en continu sur toute l'année, ce qui suppose :
 - l'adaptation de la chaîne informatique de l'enquête (introduction du multimode, recours aux postes génériques de reprise de la codification) ;
 - la gestion de la transition entre l'ancienne enquête et la nouvelle, et la migration des données ;
- un enjeu de gestion de la rupture de série, qui passe par le calibrage, la préparation et la réalisation d'un Pilote d'envergure en situation de pré-production l'année précédant l'atterrissage afin d'anticiper la mesure des ruptures de série et les rétopolations.
- un enjeu de qualité de l'enquête, qui se traduit notamment par le maintien d'un taux de réponse élevé, et qui soulève des enjeux organisationnels (sur le travail des enquêteurs notamment).

En ce qui concerne la **continuité de la production de l'enquête**, le bilan est très positif puisqu'aucune interruption n'a eu lieu et les données de l'ancienne enquête ont bien été récupérées pour les interrogations en nouvelle enquête. Par ailleurs, la première publication trimestrielle sur le chômage au sens du BIT et les indicateurs sur le marché du travail, comprenant les principales séries rétopolées, a été diffusée fin juin 2021, soit un mois et demi plus tard qu'habituellement seulement, ce délai, resté compatible avec les impératifs européens, ayant été pris par précaution pour se donner le temps d'achever l'ensemble des travaux d'atterrissage de la refonte de l'enquête. Les autres publications ont ensuite suivi le rythme habituel. Au niveau européen, la France a été le premier pays à fournir des données en nouvelle enquête avec des séries rétopolées cohérentes dès juillet 2021.

Malgré un contexte d'incertitude forte sur les outils au moment de la bascule (le principal risque au moment du séminaire de lancement pointant l'incertitude autour de la mise en place de Capi3G, qui devait remplacer Capi, bascule présentée initialement comme nécessaire pour effectuer du multimode en concurrentiel) tout au long du projet, les **développements informatiques** ont permis d'être prêt pour réaliser une enquête en multimode concurrentiel dès janvier 2020. De fait, les solutions techniques théoriquement idoines pour répondre aux besoins d'une collecte multimode avec un protocole concurrentiel ont été écartées du fait de difficultés de mise en œuvre et de calendrier de déploiement incompatibles avec le calendrier imposé par le règlement européen pour la date d'atterrissage du projet. D'une part, les outils de la nouvelle filière, basés sur le langage XML, étaient à leur début au moment du lancement du projet Mélopée. Les fonctionnalités nécessaires à l'enquête Emploi étant relativement complexes, l'implémentation de la nouvelle enquête dans ces outils s'est vite avérée incompatible avec le calendrier de la refonte. D'autre part, s'il était initialement envisagé de s'appuyer sur l'application de collecte Capi3G, les difficultés lors des premières mises en production d'enquêtes au premier semestre 2018 sous cette application ont conduit à suspendre toute nouvelle intégration avec cet outil, le temps d'une expertise approfondie de son fonctionnement. Un plan d'action a par ailleurs été mis en place, dont l'un des axes portait sur l'analyse de risque à conserver les enquêtes sous Capi au-delà de 2020 et la possibilité de répondre aux besoins multimode de Mélopée à partir de Capi. Ces analyses menées au cours de l'année 2018 ont permis de s'assurer de la possibilité d'utiliser l'application Capi pour Mélopée, moyennant un certain nombre de développements ad hoc. Cette solution, moins risquée à court terme et seule à permettre de tenir le calendrier, est cependant une solution de court terme puisque la solution Capi ne sera pas pérennisée. Elle fait donc peser un enjeu fort sur le calendrier du projet

Mayottica, dont un des objectifs est de sortir des outils Capi et Blaise pour intégrer les outils de la nouvelle filière qui les remplacent.

L'adaptation à la collecte par smartphone a constitué également un enjeu de la refonte pour les développements informatiques. Une question qui se pose quand on propose un mode d'interrogation par internet est celle d'adapter la réponse à tous les supports, y compris le smartphone, ou bien de ne pas l'adapter pour favoriser une collecte sur ordinateur, considérée comme moins sujette à un biais de *satisficing*. Cependant, le smartphone est un support largement utilisé, qui serait utilisé même si le questionnaire n'était pas adapté. Pour ne pas risquer d'abaisser le taux de collecte, il a donc été rapidement décidé d'adapter le questionnaire à la collecte par smartphone. Cependant, des difficultés ont subsisté sur smartphone, comme la collecte sur liste qui, dans certains cas, ne fonctionne pas ou rend la navigation dans le questionnaire très difficile. L'usage de ce support comme mode de collecte a sans doute été initialement sous-estimé ; sa prise en compte dans les développements informatiques est un sujet à part entière dont il est aujourd'hui difficile de se passer.

Concernant la **transition entre l'ancienne et la nouvelle enquête**, le sujet de la stratégie à adopter et de la migration des données s'est avéré complexe à gérer. De fait, au T1 2020 puis au T1 2021, des ménages interrogés en ancienne enquête le trimestre précédent ont dû être interrogés en nouvelle enquête. Le questionnaire étant fortement modifié, plusieurs possibilités étaient envisageables : poser un questionnaire de première interrogation en nouvelle enquête à tous les ménages, y compris ceux en réinterrogation ; basculer les informations de l'ancienne enquête et ne reposer que les questions absolument nécessaires en plus du questionnaire de réinterrogation classique. La première méthode est très coûteuse puisqu'elle nécessite une interrogation en face à face et le questionnaire est plus long. Par ailleurs, elle serait difficilement acceptée par les enquêtés, puisqu'elle consiste à faire comme si les interrogations précédentes n'avaient pas eu lieu. Ainsi, le choix a été fait de basculer en nouvelle enquête directement avec le questionnaire de réinterrogation pour les ménages concernés. Cette méthode pose alors la question de la récupération des variables antérieures, dont certaines peuvent être identiques en ancienne et nouvelle enquête, certaines peuvent être reconstruites et d'autres ne sont plus du tout comparables. Dans ce dernier cas, seules les questions absolument nécessaires sont repostées ; ainsi, petit à petit un questionnaire de bascule spécifique a été nécessaire en nouvelle enquête, intermédiaire entre questionnaire de première interrogation et questionnaire de réinterrogation.

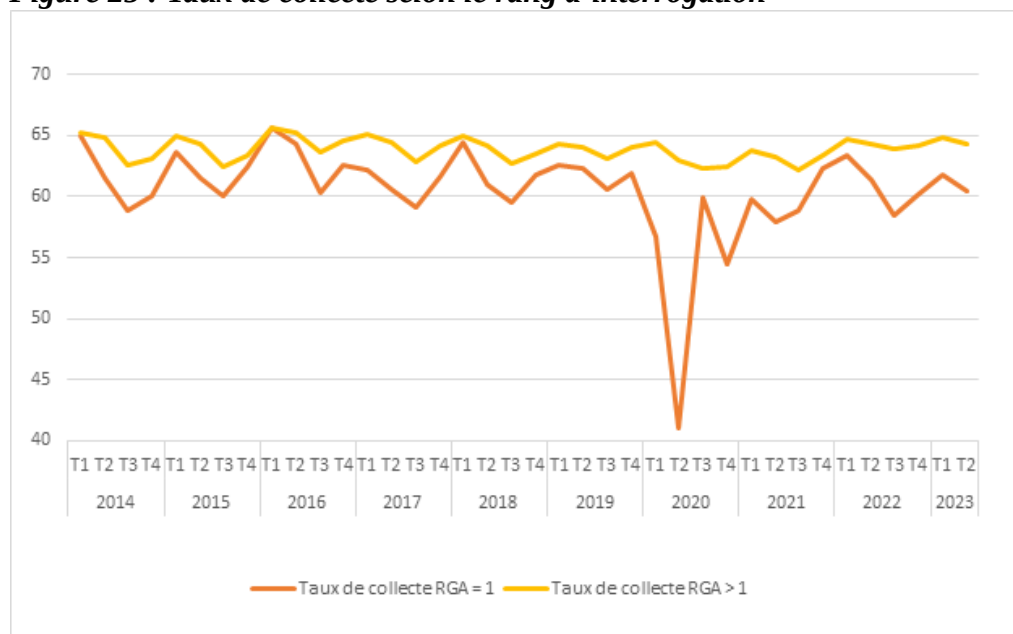
Cette bascule est apparue comme la seule solution pertinente opérationnellement, sous contraintes de coûts. Cependant, elle pose des difficultés techniques jusqu'à l'aval du processus. En amont, il est nécessaire de récupérer les données antérieures, de les reconstruire éventuellement et de les transférer dans un autre système d'information. À l'aval, certaines variables peuvent être manquantes pour les logements non remis en collecte lorsque la dernière interrogation a eu lieu en ancienne enquête. Ce phénomène peut concerner des ménages jusqu'au T2 2022, pour les ménages interrogés pour la première et dernière fois au T4 2020. Dans le cas où les informations collectées en ancienne enquête ne permettent pas de reconstruire les variables de diffusion en nouvelle enquête, alors l'information est manquante. Ces variables sont dites temporairement « yearly », c'est-à-dire qu'elles ne peuvent être exploitées dans la nouvelle enquête que pour les ménages en première interrogation. À partir du T3 2022, toutes les premières interrogations ont eu lieu en nouvelle enquête ; toutes les informations sont donc disponibles. Les fichiers de diffusion étant diffusés annuellement, les variables concernées seront exploitables de façon trimestrielle à partir de l'enquête Emploi 2023 seulement. Cette gestion de la bascule complexifie considérablement les

programmes aval, qui ont pourtant été faits pour être perennes et donc fonctionner bien après la fin des effets de cette bascule. Ces programmes seront toilettés de cet effet bascule à l'occasion de leur réécriture en R.

Concernant la **gestion de la rupture de série**, comme évoqué précédemment, le Pilote et l'anticipation de la mesure des ruptures de série ont permis de produire des séries rétopolées pour les principaux indicateurs de l'enquête Emploi diffusés chaque trimestre dans un temps très court. De fait, la première publication n'a été décalée que d'un mois et demi. La comparaison entre ancienne et nouvelle enquête grâce au pilote de 2020 a été cependant affectée par la concomitance avec la crise sanitaire. Cette crise a affecté à la fois le protocole de collecte puisque les déplacements n'étaient plus possibles, en période de confinement notamment, mais également la situation réelle sur le marché du travail. Enfin, la taille du Pilote a pu poser des difficultés pour mesurer de façon suffisamment robuste certaines ruptures de série liées à la refonte. Par exemple cette taille n'a ainsi pas été suffisante pour évaluer de manière robuste des ruptures de série propres aux DOM, malgré leur suréchantillon. Pour les mêmes raisons, la stratégie d'estimation de la rupture a aussi été modifiée s'agissant des estimations de taux d'emploi et de chômage par catégories socioprofessionnelles (cf. Annexe 1).

En ce qui concerne la **qualité de l'enquête** et le maintien des taux de réponse, il est encore difficile de juger de l'efficacité de la refonte compte tenu du contexte particulier des dernières années. De fait, ces taux ont été fortement affectés par la crise sanitaire malgré les protocoles ad hoc mis en place pour y répondre. Cependant, malgré la crise, les taux de collecte et de réponse de l'enquête Emploi sont très satisfaisants et sont revenus dès 2022 à des niveaux d'avant crise (*figure 25*). L'évolution forte du taux de réponse par Internet en réinterrogation (celui-ci a gagné presque 10 points en 2 ans et demi) montre que ce nouveau mode de collecte répond à un besoin des enquêtés. Les changements de protocole, et notamment l'introduction d'Internet comme nouveau mode de collecte, ont conduit à des changements importants en termes d'organisation de la collecte et du travail des enquêteurs, des gestionnaires et d'autres acteurs de l'Insee ayant une fonction d'assistance aux enquêtés (Insee contact, pôle EEC). Pour autant, tous les acteurs ont su trouver leur place et modifier leurs pratiques en s'adaptant au nouveau protocole de l'enquête. Un bilan a été réalisé auprès des DEM et d'un échantillon d'enquêteurs au T1 2021. Ce bilan met en avant une complexité à gérer plus grande par les différents acteurs du fait du protocole multimode, mais une complexité qui est globalement maîtrisée. Le protocole de l'enquête est ainsi bien plus exigeant, tous les outils techniques liés à la plateforme de collecte par Internet devant être opérationnels en continu, comme l'enquête ; depuis sa mise en place début 2020, de rares incidents techniques ont conduit à des interruptions de services, incidents qui ont nécessité une grande réactivité.

Figure 25 : Taux de collecte selon le rang d'interrogation



Lecture : Au T2 2023, le taux de collecte (part des fiches-adresses réussies sur l'ensemble de l'échantillon) est de 60,5 % en première interrogation et de 64,3 % en réinterrogation.

Champ : France hors Mayotte.

Source : Insee, enquêtes Emploi.

Enfin, un enjeu du projet consistait également à bien **communiquer auprès des utilisateurs** des données de l'enquête Emploi. Ces derniers sont nombreux tant au niveau national qu'international, tant parmi les chargés d'études statistiques que dans la recherche. Les refontes de l'enquête Emploi ne sont pas courantes et il est particulièrement important de bien communiquer sur les changements pour assurer une utilisation adéquate des données. Ainsi, un comité utilisateurs a été relancé début 2022 avec une présentation de la refonte à tous ses niveaux et de son impact sur les données et leur usage. Ce comité est à destination de tous les types d'utilisateurs. Plusieurs présentations de la refonte et de la méthode de réropolation ont été faites en interne comme au Cnis et à Eurostat. Par ailleurs, deux publications de référence sur la refonte ont été publiées très rapidement : un Insee Analyses fin juin 2021 sur l'impact de la refonte sur les principaux indicateurs de l'enquête ; un article dans la revue Courrier des Statistiques en juillet 2021 sur les objectifs et le contenu de cette nouvelle refonte de l'enquête Emploi.

La refonte de l'enquête Emploi en 2021 est également source d'enseignements pour d'autres projets en dehors de l'enquête en elle-même. De fait, l'enquête Emploi est la première enquête auprès des ménages à mettre en œuvre un protocole multimode concurrentiel et sur un temps de collecte très court. Dans un contexte de passage à des protocoles multimodes pour la plupart des enquêtes de l'Insee auprès des ménages, les résultats de l'enquête Emploi en termes de taux de collecte, d'organisation de la collecte (rythme et type d'annonce et relances ; reprise de questionnaire internet partiel par les enquêteurs ; organisation des enquêteurs et des gestionnaires, etc...), de qualité de l'information recueillie pourront servir pour d'autres enquêtes multimodes à venir.

Par ailleurs, l'enquête Emploi 2021 est la première enquête à mettre en œuvre la nouvelle nomenclature des professions et catégories socioprofessionnelles (PCS 2020). Cette nouvelle PCS a vocation à être implémentée dans toutes les sources utilisant jusque-là la PCS 2003. Son

implémentation dans l'enquête Emploi permet de disposer d'outils fonctionnels permettant de recueillir la PCS 2020 et d'une connaissance acquise sur ces premières données.

3. Enseignements pour les projets futurs

L'un des principaux enseignements de la dernière refonte d'ampleur de l'enquête Emploi en 2013 concerne l'importance de la mise en place d'un **Pilote** pour mesurer les éventuelles ruptures de série. Le pilote mis en œuvre pour la refonte de 2021 s'est avéré très utile et efficace pour rétroscander les séries longues sur le marché du travail, et ce dans un calendrier très serré permettant de publier très rapidement de nouvelles séries sur les principaux indicateurs du marché du travail. L'ampleur du Pilote n'a en revanche pas suffi pour mesurer avec assurance des ruptures de séries pour des agrégats à des niveaux trop fins, par exemple les indicateurs du marché du travail dans les DOM.

La gestion de la **bascule** des données de l'ancienne enquête pour leur utilisation en nouvelle enquête a été complexe et a eu des impacts multiples à tous les niveaux de la chaîne jusqu'à la mise à disposition des données. Cette gestion difficile est principalement liée au changement dans le questionnaire. Il est donc nécessaire de récupérer le maximum d'informations de l'ancienne enquête pour pouvoir collecter le nouveau questionnaire. Ce choix implique un travail important de recodification des variables antérieures. Par ailleurs, dans certains cas, une recodification n'était pas possible et il faut alors choisir entre poser la nouvelle question ou ne pas avoir l'information. Ainsi, pour certaines variables, elles ne sont pas disponibles pour l'ensemble de l'échantillon de l'enquête Emploi et ne le sont qu'à partir du moment où tous les enquêtés sont interrogés uniquement en nouvelle enquête, soit à partir du T3 2022.

Concernant les **outils**, il y a un arbitrage à avoir entre modifier le strict nécessaire en repartant de l'existant et repartir de zéro. Par exemple, concernant l'application EEC3, le choix a été fait de repartir de l'existant et de l'adapter, essentiellement à la gestion du protocole multimode. Ce choix, moins coûteux, a des désavantages dans la mesure où des développements obsolètes peuvent perdurer et des contradictions poindre. Il est donc important, sans forcément repartir de zéro, de remettre à plat les besoins et la façon d'y répondre avec les outils existants et à construire, pour les rationaliser au mieux. Par ailleurs, il peut parfois être difficile de bien spécifier les besoins en amont, avant des changements de protocole importants qui peuvent conduire à des changements dans les pratiques et donc dans les besoins. Un bilan post-refonte peut alors s'avérer utile pour identifier d'éventuels nouveaux besoins et adapter les nouveaux outils.

Enfin, les **contraintes européennes** en termes de calendrier constituent une difficulté importante en matière d'organisation d'un projet de refonte. De fait, au moment du lancement du projet, il était prévu que le règlement européen soit mis en place dès 2019. Ce calendrier n'étant pas tenable pour un certain nombre de pays, il a été finalement repoussé en 2021, ce qui avait été assez tôt anticipé comme la date la plus probable, avec toutefois une incertitude durant toute la première période du projet (avec même la possibilité que la date soit ultérieure). Cette contrainte imposant une date d'atterrissage a conduit à arbitrer par les délais. Il a été nécessaire de prévoir dans le projet un MVP (Minimal Viable Product) pour être en mesure de s'assurer de pouvoir répondre au règlement européen à date dite. Dans le cadre de la première partie du projet Mélopée, un scénario de secours a été étudié pour modifier uniquement le questionnaire dans le cas où le projet ne serait pas suffisamment avancé pour le passage au nouveau protocole multimode. Par ailleurs, malgré l'incertitude pesant sur la date d'implémentation et le contenu précis du règlement, il est

indispensable de lancer le projet et de mettre en œuvre les tests nécessaires très en amont. Ainsi, sans les expérimentations Muse, on aurait manqué d'éléments pour construire la refonte. Enfin, il apparaît indispensable de bien s'intégrer aux discussions européennes pour tenter de faire converger au maximum les attentes nationales et les contraintes européennes ainsi que pour bien suivre l'avancement du règlement et de faire valoir les contraintes de calendrier nationales.

Bibliographie

Dera, 2021. [L'enquête Emploi se rénove en 2021 : des raisons de sa refonte aux impacts sur la mesure de l'emploi et du chômage](#). Insee Analyses, n° 65.

Insee, 2021. [Au premier trimestre 2021, le taux de chômage est quasi stable à 8,1 %](#). Informations Rapides, n° 167.

AMOSSÉ T., 2020. [La nomenclature socioprofessionnelle 2020 : Continuité et innovation, pour des usages renforcés](#). In : Courrier des statistiques. Insee. N° N4, pp. 62-80.

BELLAMY. V., 2016. [123 500 divorces en 2014 : Des divorces en légère baisse depuis 2010](#). Insee Première, n°1599.

CASES, C., 2019. [IESS : l'Europe harmonise ses statistiques sociales pour mieux éclairer les politiques](#). In : Courrier des statistiques. Insee. N° N3, pp. 125-139.

CHEVALIER, M., COSTA, L., DELTA, L., DEROYON, T., FAVRE-MARTINOZ, C., GIVOIS, S., GUILLO, C., MERLY-ALPA, T., PALIOD, N., PENDOLI, P-A., SAUVAGET, T. et VINCENT, L., 2022. [Le renouvellement de l'échantillon-maître des enquêtes auprès des ménages et de l'échantillon de l'enquête Emploi de l'Insee](#). Insee. Insee Méthodes, n° 141.

CHOLETTE, P. A., 1984. Adjusting Sub-annual Series to Yearly Benchmarks. *Survey Methodology*, vol. 10, p. 35–49.

DENTON, F. T., 1971. Adjustment of monthly or Quarterly Series to Annual Totals : An Approach based on Quadratic Minimization. *Journal of the American Statistical Association*, vol. 66, No. 333, p. 99-102.

DESTANDAU, S., 2014. Bilan du projet REFLEE. Insee. Documents de travail, n° C2014-02.

GARNERO, M., 2019. [Le projet Muse : 5 ans d'expérimentations pour préparer l'introduction d'Internet dans l'enquête Emploi](#). Insee. Documents de travail, n° F1907.

GAZIER, B., MINNI, C. et PICART, C., 2016. [La diversité des formes d'emploi](#). Cnis, rapport de groupe de travail, n° 142.

GUILLAUMAT-TAILLIET, F. et TAVAN, C., 2021. [Une nouvelle enquête Emploi en 2021](#). In : Courrier des statistiques. Insee. N° N6, pp. 7-27.

RENAUDAT, E., 1993. Les salariés en congé parental. *Recherches et Prévisions*, n°32, pp. 9-17.

VINCENEUX, K., 2018. [Mode de collecte et questionnaire, quels impacts sur les indicateurs européens de l'enquête Emploi ?](#) Insee. Documents de travail, n° F1804.

Annexes

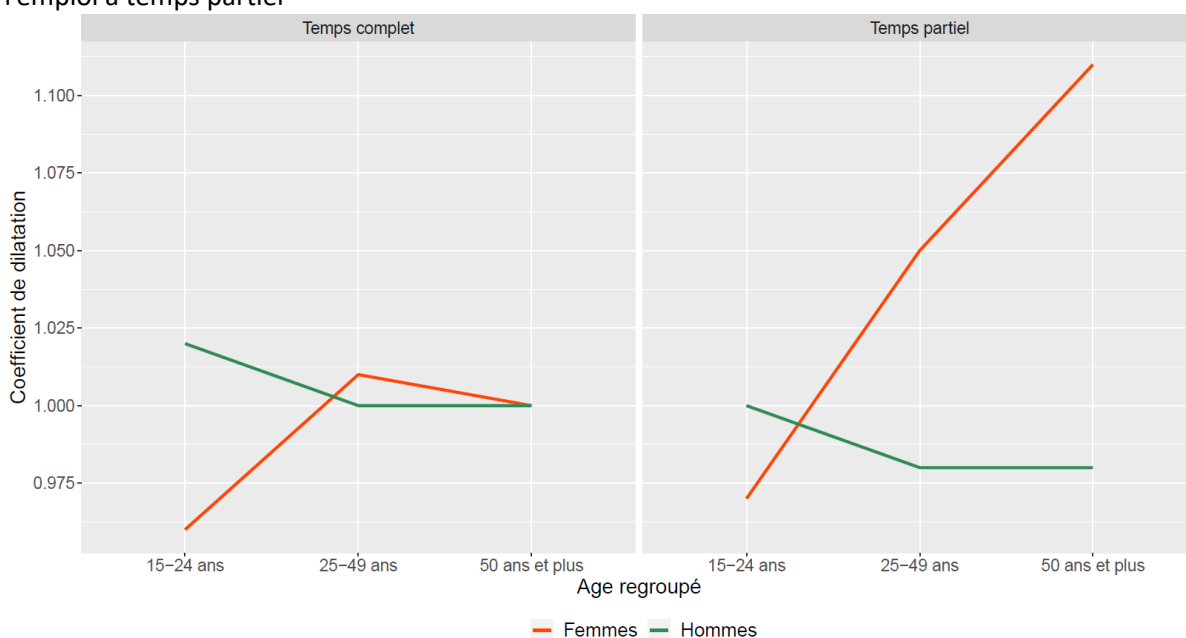
Annexe 1 : Ruptures de série et rétopolation des séries par thématique (hors statut BIT)

a) Temps de travail : temps complet et temps partiel

Les séries longues d'emploi à temps complet et à temps partiel sont publiées depuis 1975 pour les séries annuelles et 2003 pour les séries trimestrielles par sexe et âge regroupé en 3 classes agrégées (15-24 ans, 25-49 ans et 50 ans et plus).

La rétopolation des séries annuelles est faite selon la méthode générale, sans lissage des coefficients de dilatation puisque les âges sont regroupés, et le calage est fait sur l'emploi total rétopolé préalablement.

Figure 1 : Coefficients de dilatation entre ancienne et nouvelle enquête de l'emploi à temps complet et de l'emploi à temps partiel



Lecture : Les femmes de 25-49 ans à temps partiel sont 1,05 fois plus nombreuses en nouvelle enquête que dans l'ancienne enquête sur la période de recouvrement.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

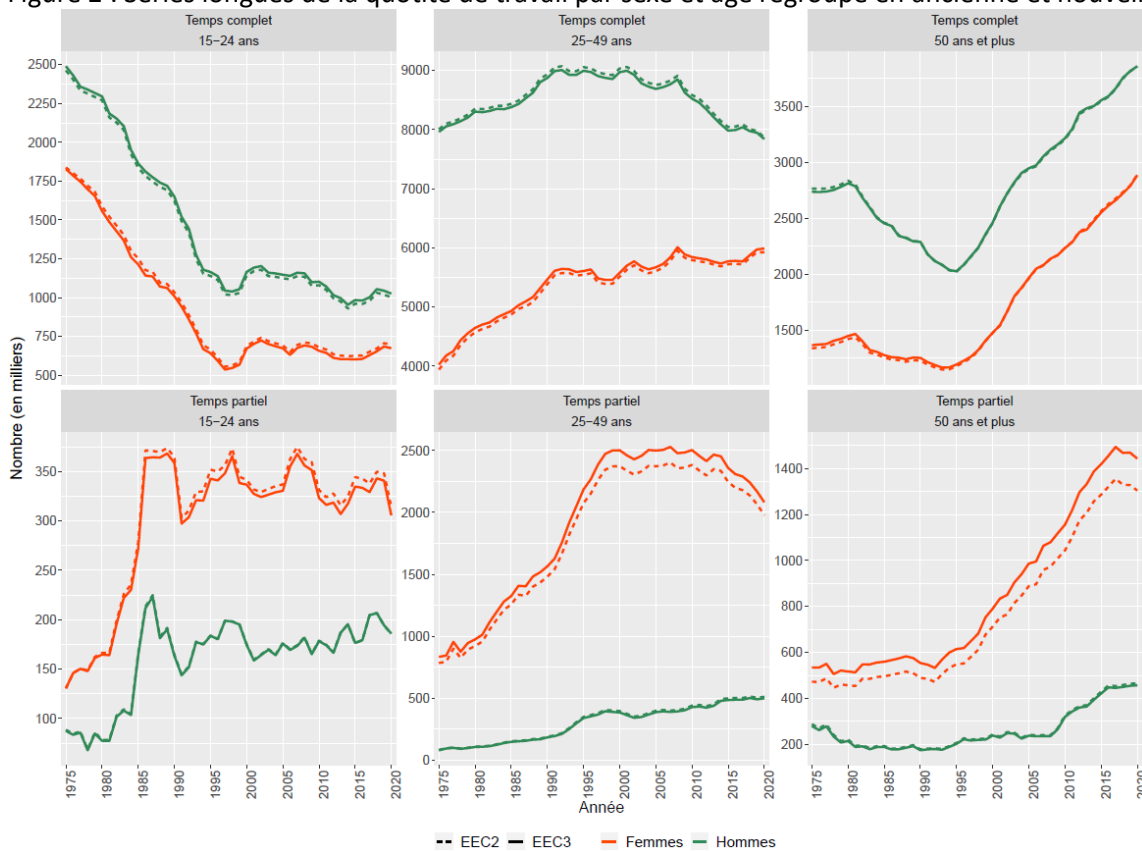
Source : Insee, enquêtes Emploi

La Figure 1 montre, pour les jeunes femmes, une baisse de l'emploi à temps complet comme à temps partiel, conséquence de la baisse de l'emploi total dans cette classe d'âge, tandis que chez les jeunes hommes toute la hausse de l'emploi concerne du temps complet.

Parmi les plus de 25 ans, le temps partiel augmente fortement pour les femmes, alors qu'il diminue légèrement pour les hommes. Le changement de concept dans la mesure du statut BIT et notamment la plus forte intégration des congés parentaux dans l'emploi est un premier élément explicatif (ce type d'emploi, plus fréquemment occupé par des femmes et des employés, est davantage à temps partiel, la personne décrivant dans ce cas son contrat de travail même si elle ne travaille pas). De façon générale sur l'ensemble des classes d'âge, l'autre facteur possible d'explication est que le nouveau questionnaire permet

de mieux capter les petits boulots (qui sont souvent à temps partiel) notamment via la baisse du recours au proxy.

Figure 2 : Séries longues de la quotité de travail par sexe et âge regroupé en ancienne et nouvelle enquêtes



Lecture : En 2020, en ancienne enquête (EEC2), 1 304 000 femmes de 50 ans et plus sont à temps partiel, contre 1 443 000 en nouvelle enquête (EEC3).

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

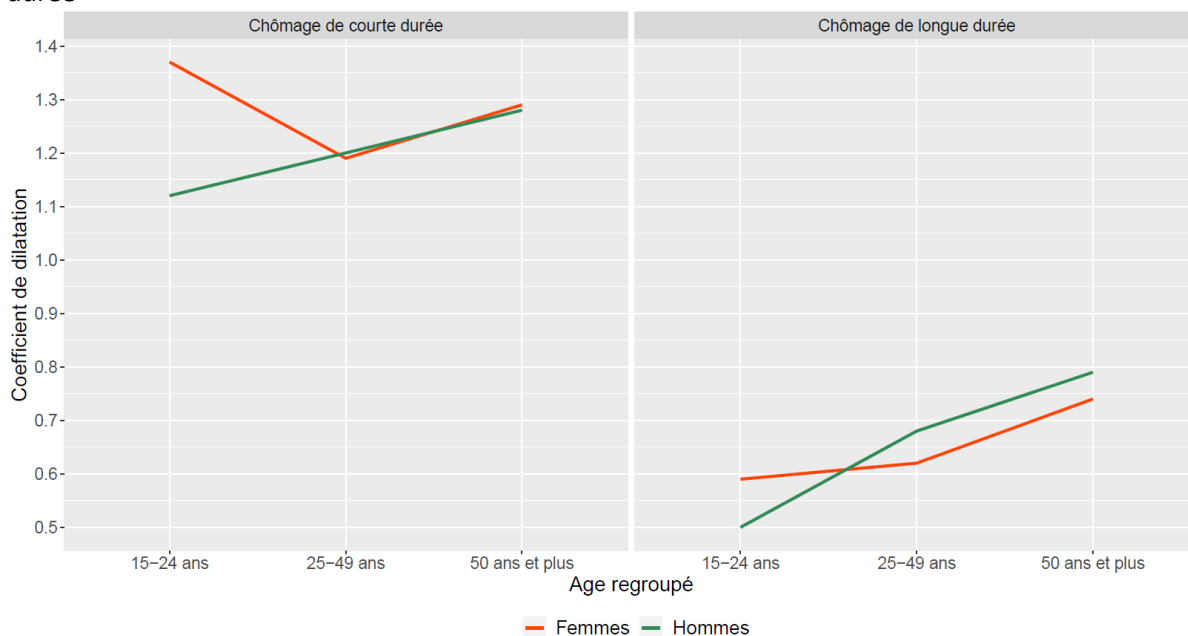
Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

Les séries annuelles rétrolées de temps complet et temps partiel par sexe et âge agrégé, représentées Figure 2, sont ensuite utilisées pour rétroler les séries trimestrielles par âge regroupé.

b) Chômage de longue durée

Un chômeur est considéré « de longue durée » lorsque qu'il est au chômage depuis 1 an ou plus. L'Insee publie des séries annuelles et trimestrielles du nombre de chômeurs de courte et longue durée, depuis 1975 pour les séries annuelles et 2003 pour les séries trimestrielles, par sexe et âge regroupé en 3 classes (15-24 ans, 25-49 ans et 50 ans et plus). La rétrolation a été faite selon la méthode générale, et les coefficients de dilution obtenus sont présentés Figure 3.

Figure 3 : Coefficients de dilatation entre ancienne et nouvelle enquêtes sur le chômage de courte et longue durée



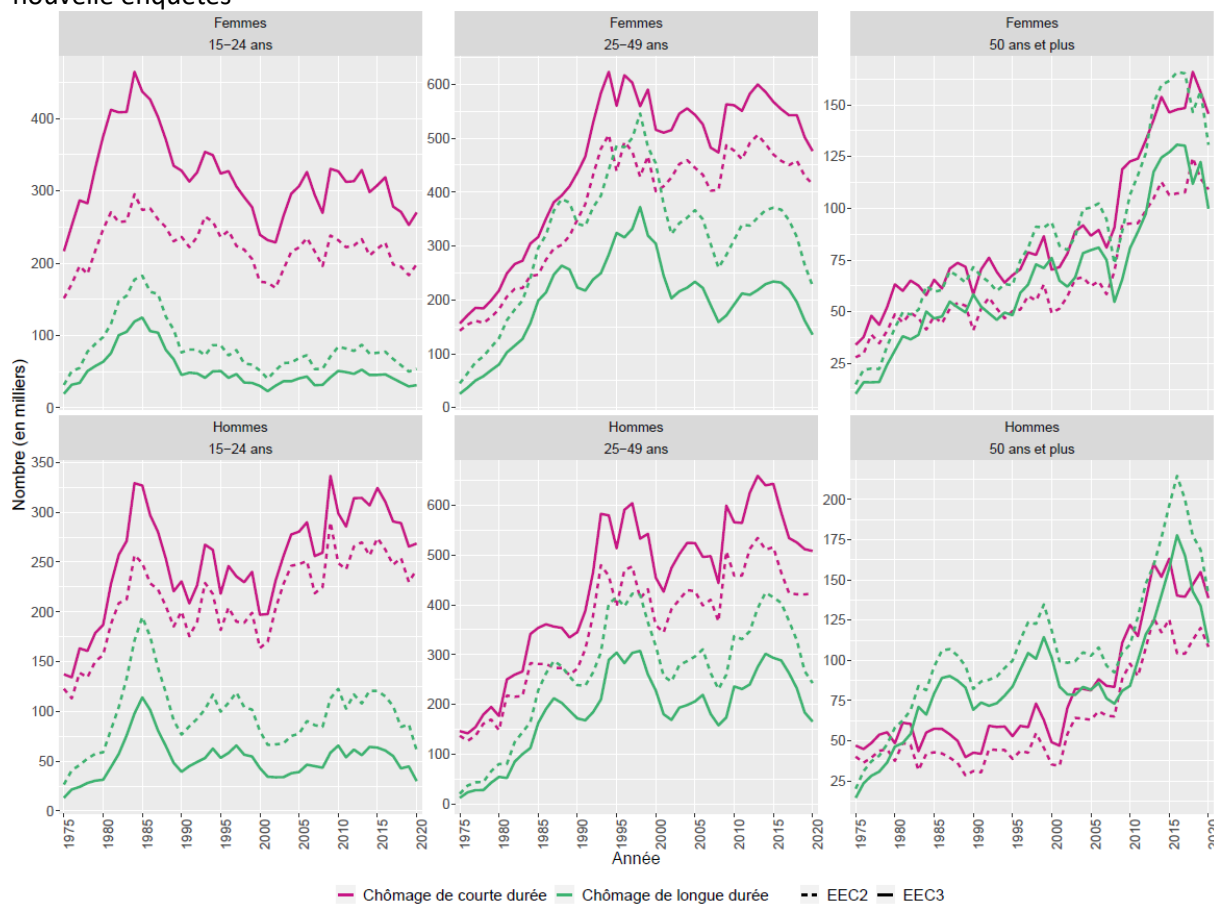
Lecture : Le nombre d'hommes de 15-24 ans au chômage de longue durée est divisé par 2 dans le Pilote par rapport à l'ancienne sur la période de recouvrement.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

Source : Insee, enquêtes Emploi

Quels que soient le sexe et l'âge, le nombre de chômeurs de longue durée baisse fortement lors du passage en nouvelle enquête. Pour l'expliquer, il est nécessaire de revenir aux durées de recherche d'emploi et écoulée depuis le dernier emploi, puisque l'ancienneté au chômage est définie comme le minimum des deux. Dans les deux cas, ces durées ont été revues à la baisse en nouvelle enquête. Le raccourcissement de la durée écoulée depuis le dernier emploi est dû au fait que les petits boulots sont désormais pris en compte dans le calcul, alors qu'auparavant uniquement le dernier emploi « régulier » servait de référence pour calculer cette durée. La baisse de la durée de recherche d'emploi est plus difficile à expliquer, cela pourrait éventuellement venir d'un effet du mode de réponse par Internet.

Figure 4 : Séries longues des chômeurs de courte et longue durée par sexe et âge regroupé en ancienne et nouvelle enquêtes



Lecture : En 2020, en ancienne enquête (EEC2), 226 000 femmes de 25-49 ans sont chômeuses de longue durée, contre 135 000 en nouvelle enquête (EEC3).

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

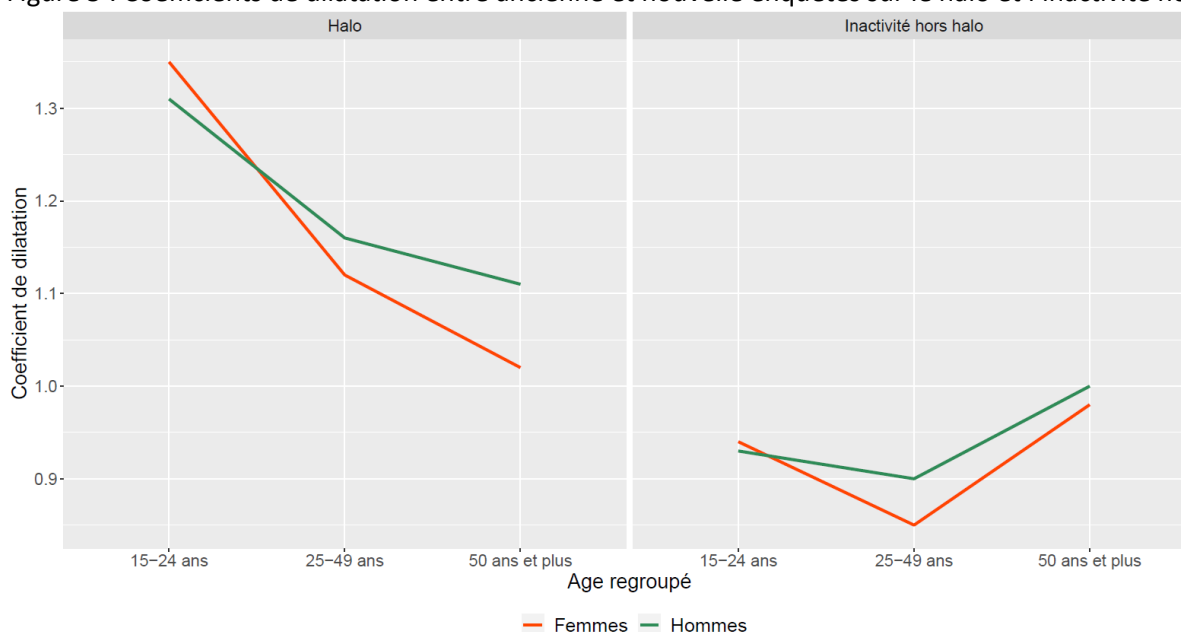
Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

Comme pour la quotité de travail, les séries annuelles rétopolées du chômage de courte et longue durée (Figure 4) sont utilisées pour rétopoler les séries trimestrielles par âge regroupé.

c) Halo et composantes du halo

Le premier type de séries diffusées concernant le halo sont celles donnant le nombre de personnes dans le halo et le nombre de celles inactives hors halo, par sexe et âge regroupé. La méthode générale de rétopolation a été adoptée pour ces séries.

Figure 5 : Coefficients de dilatation entre ancienne et nouvelle enquêtes sur le halo et l'inactivité hors halo.



Lecture : Le nombre d'hommes de 25-49 ans inactifs hors halo est 0,9 fois plus faible en nouvelle enquête qu'en ancienne enquête sur la période de recouvrement.

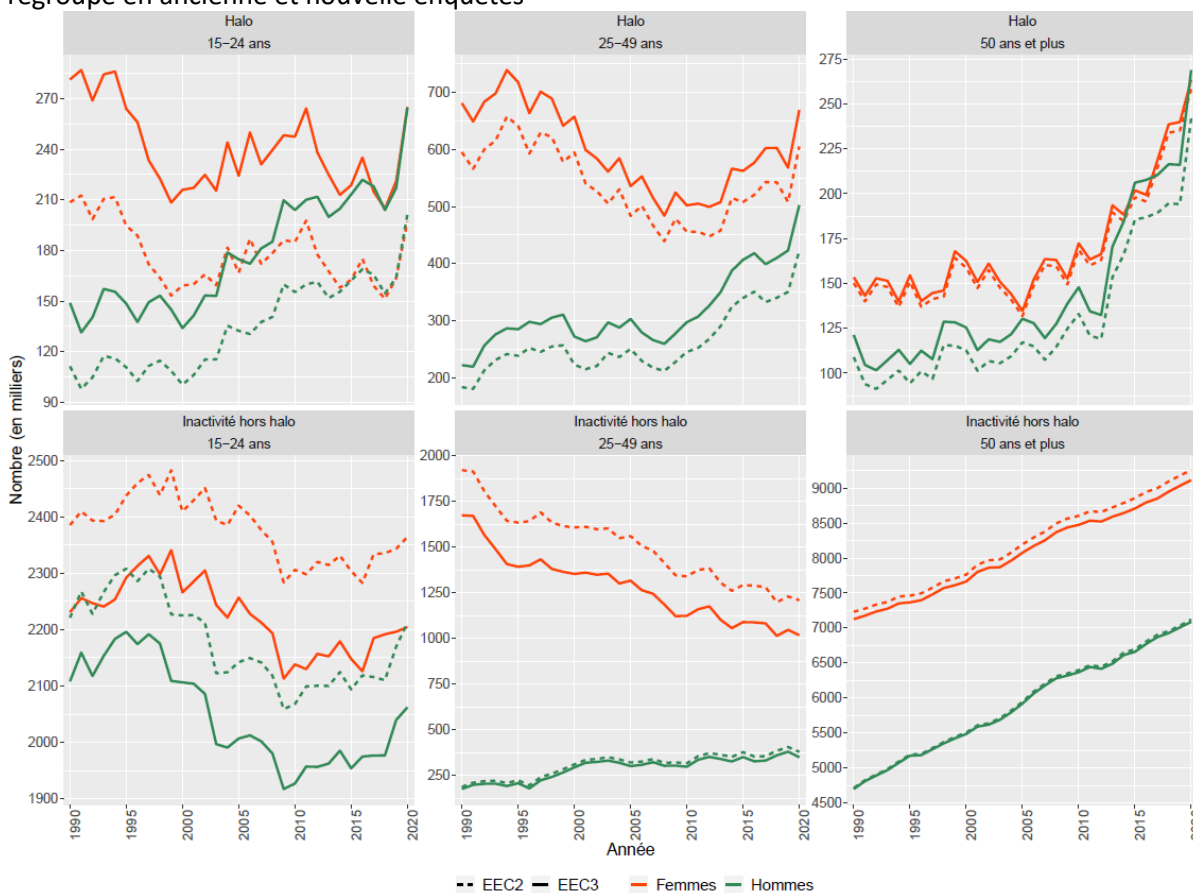
Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

Source : Insee, enquête Emploi

La Figure 5 montre une très forte hausse du halo autour du chômage lors du passage à la nouvelle enquête. Cette hausse s'explique surtout par des effets questionnaire et protocole : davantage d'enquêtés déclarent avoir effectué des recherches d'emploi dans la nouvelle enquête (sans nécessairement être disponibles). Les exemples de démarche de recherche ajoutés dans les consignes du questionnaire Internet peuvent avoir influencé leur réponse. Ils sont aussi un peu plus nombreux, lorsqu'ils n'ont pas fait de recherches d'emploi, à déclarer souhaiter néanmoins travailler, la réponse par Internet pouvant diminuer les éventuelles réticences à se décrire dans cette situation (biais de désirabilité).

La rupture de série à la hausse sur le halo est principalement marquée pour les jeunes. Pour ces derniers, la diminution du recours au proxy explique sans doute cet effet marqué : on peut supposer qu'un enquêté est plus à même de déclarer souhaiter travailler, que lorsqu'un proxy répond pour lui. Par ailleurs, plus de jeunes sont désormais en situation d'activité, d'emploi notamment (cf. supra), la baisse de l'inactivité totale (halo et hors halo) est plus faible pour les jeunes.

Figure 6 : Séries longues du nombre de personnes dans le halo et inactives hors halo par sexe et âge regroupé en ancienne et nouvelle enquêtes



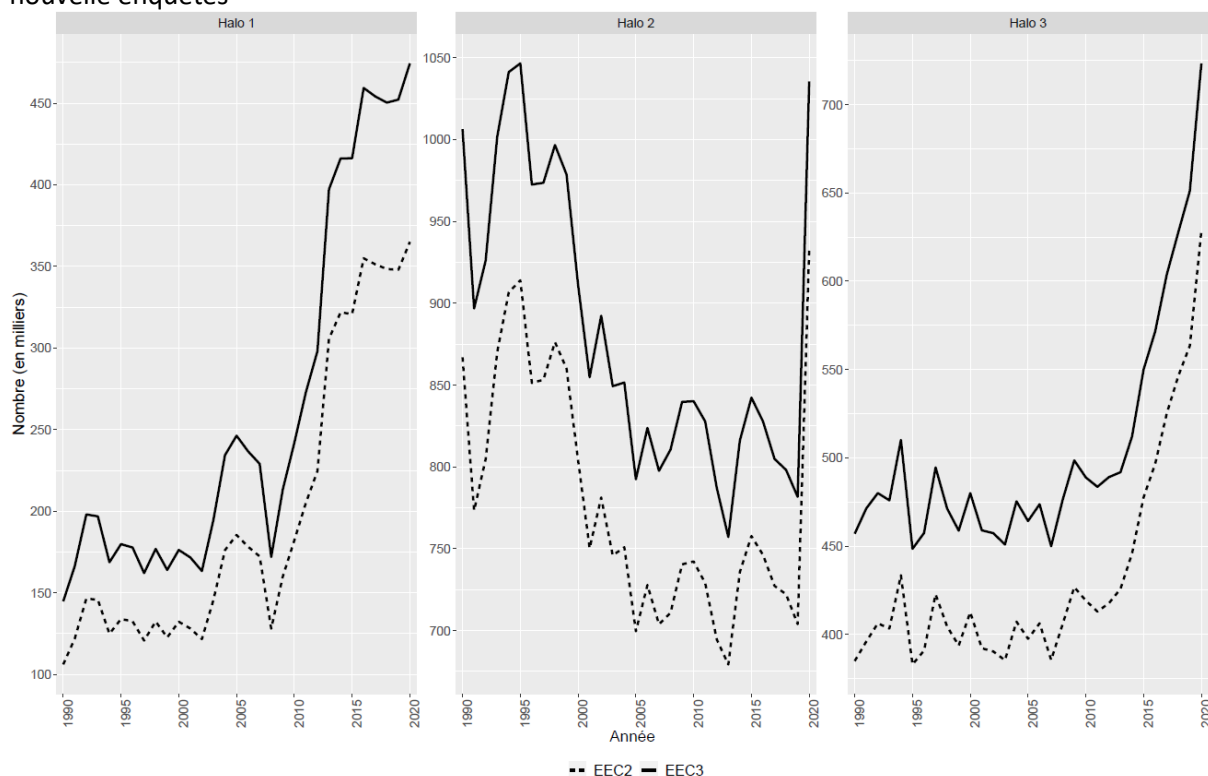
Lecture : En 2020, en ancienne enquête (EEC2), 423 000 hommes de 25-49 ans sont dans le halo, contre 502 000 en nouvelle enquête (EEC3).

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

Les séries de halo et d'inactivité hors halo sont ensuite recalées afin que la somme des deux corresponde bien à l'inactivité totale, ce qui permet d'obtenir les séries annuelles présentées en Figure 6. Elles permettent ensuite de construire le deuxième type de séries concernant le halo : le nombre total de personnes dans chacune des trois composantes du halo, tous sexes et âges confondus. Pour cela, la méthode des parts logistiques est utilisée pour désagréger non pas les âges, mais le halo total, en s'appliquant aux parts de chaque composante dans le halo total depuis 1990.

Figure 7 : Séries longues du nombre de personnes dans chaque composante du halo en ancienne et nouvelle enquêtes



Lecture : En 2020, en ancienne enquête (EEC2), 365 000 personnes sont dans la première composante du halo, contre 474 000 en nouvelle enquête (EEC3).

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

Sur la Figure 7, la rupture de série à la hausse concerne les trois composantes du halo (+100 000 personnes pour chacune). La hausse du halo 1 (personnes sans emploi, souhaitant travailler, ayant effectué des démarches de recherche d'emploi mais pas disponible pour travailler) peut s'expliquer par la modification du questionnaire qui donne des exemples de recherche d'emploi, et par la baisse du recours au proxy. Le halo 2 (personnes sans emploi, souhaitant travailler, disponibles pour cela mais n'ayant pas effectué de démarches de recherche d'emploi) n'a pas pu être affecté par ce paramètre, et résulte plutôt de la plus forte disponibilité annoncée en nouvelle enquête, ainsi que de la facilité accrue à avouer ne pas avoir fait de recherche d'emploi en répondant par Internet. Quant à la hausse du halo 3 (personne sans emploi, souhaitant travailler mais pas disponible pour cela et n'ayant pas non plus effectué de démarche active de recherche), elle est sans doute largement la conséquence de la hausse du halo total et du choix méthodologique d'utiliser la méthode des parts logistiques pour rétopoler les séries de composantes du halo ; néanmoins, un moindre recours au proxy peut aussi expliquer un souhait affiché de travailler plus fréquent.

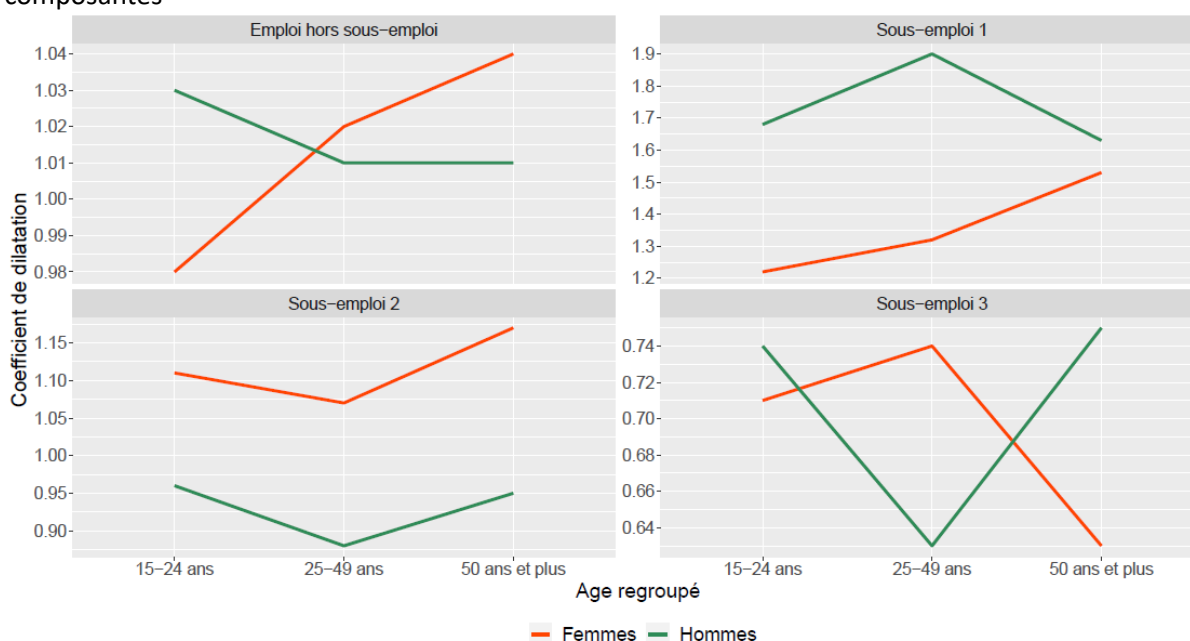
Après avoir rétopolé les deux types de séries annuelles de halo, les séries trimestrielles peuvent à leur tour être rétopolées. Le halo et l'inactivité hors halo sont d'abord rétopolés par sexe et âge en 3 classes selon la méthode générale, puis la dernière classe d'âge (50 ans et plus) est éclatée en deux (50-64 ans et 65 ans et plus) en utilisant la méthode des parts logistiques.

La série trimestrielle du nombre de personnes dans chacune des trois composantes du halo est, comme pour la série annuelle, créée en utilisant la méthode des parts logistiques appliquée aux parts trimestrielles de chacune des trois composantes du halo dans le halo total.

d) Sous-emploi et composantes du sous-emploi

Une personne est considérée comme étant en sous-emploi si elle occupe un emploi à temps partiel, souhaite travailler davantage et est disponible pour cela, qu'elle ait effectué des démarches actives de recherche d'emploi (1^{re} composante du sous-emploi) ou non (2^e composante du sous-emploi). Les personnes en emploi à temps complet mais ayant involontairement travaillé moins que d'habitude (chômage partiel ou technique) sont également considérées en sous-emploi et constituent la 3^e composante de cette catégorie. Quatre séries historiques concernant le sous-emploi ont donc été rétropolées au niveau annuel en nouvelle norme depuis 1990 : le nombre de personnes dans chacune de ses trois composantes, ainsi que le nombre de personnes en emploi hors sous-emploi. Les coefficients de dilatation obtenus sont présentés Figure 8.

Figure 8 : Coefficients de dilatation entre ancienne et nouvelle enquête sur le sous-emploi et ses composantes



Lecture : Le nombre d'hommes de 25-49 ans en sous-emploi de type 1 a été multiplié par 1,9 dans la nouvelle enquête par rapport à l'ancienne enquête sur la période de recouvrement.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

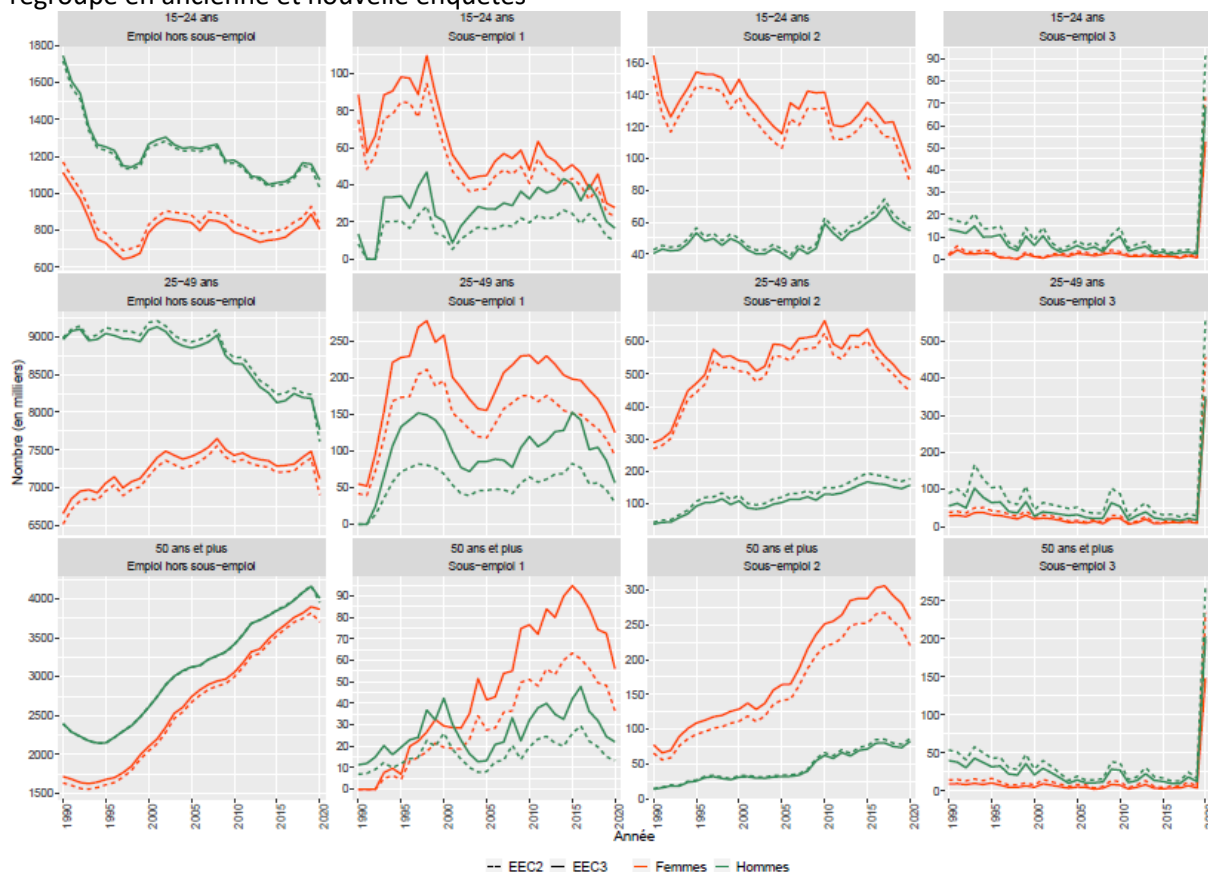
Source : Insee, enquête Emploi

La baisse du sous-emploi de type 3 semble principalement résulter de la modification du questionnaire, puisque le chômage partiel est désormais repéré au moment des questions sur les raisons d'absence (dont la liste a été modifiée et développée), alors qu'il l'était dans l'ancienne enquête au moment du questionnement sur les causes de modification des horaires la semaine de référence.

Le nombre de personnes en sous-emploi de type 1 est fortement revu à la hausse dans la nouvelle enquête, en particulier pour les hommes. Le sous-emploi de type 2 a également augmenté pour les femmes mais dans des proportions moindres par rapport au type 1, tandis que pour les hommes il a au contraire diminué. Cela suggère que la hausse du sous-emploi de type 1 pour les hommes s'explique surtout par des déclarations plus fréquentes de recherche d'emploi (avec en contrepartie une baisse de la 2^e composante), tandis que pour les femmes, la hausse de l'emploi à temps partiel étant plus importante, malgré l'effet de bascule, la deuxième composante du sous-emploi augmente aussi.

Enfin, l'emploi hors sous-emploi augmente pour toutes les personnes en âge de travailler à l'exception des femmes de 15 à 24 ans, pour lesquelles il est revu à la baisse en nouvelle enquête, en raison de la baisse de l'emploi total dans cette classe d'âge.

Figure 9 : Séries longues des composantes du sous-emploi et de l'emploi hors sous-emploi par sexe et âge regroupé en ancienne et nouvelle enquêtes



Lecture : En 2020, en ancienne enquête (EEC2), 36 500 femmes de 50 ans et plus sont en sous-emploi de type 1, contre 56 100 en nouvelle enquête (EEC3).

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

Les 4 séries sont ensuite recalées afin que leur somme corresponde bien à l'emploi total rétrolé, et les séries qui en résultent sont présentées Figure 9. La très forte hausse du sous-emploi de type 3 en 2020 s'explique par les conditions exceptionnelles relatives à la crise sanitaire et aux confinements cette année-là, mais la révision du questionnaire en minore l'ampleur (cf. supra).

Les séries trimestrielles sont rétrolées par sexe et âge en 3 classes à partir des séries annuelles rétrolées selon la méthode générale, puis la méthode des parts logistiques est appliquée à la dernière classe afin de distinguer, parmi les 50 ans et plus, les 50-64 ans, 65-74 ans et 75 ans et plus, et parmi les 15-24ans, les 15-19 ans des 20-24 ans. Les séries trimestrielles à ce niveau de détail d'âge en ancienne norme n'existaient pas auparavant, il a donc fallu créer des séries de part de chacun des détails d'âge dans l'agrégat, à partir des données brutes depuis 2003, qui ne prennent donc pas en compte la rupture de 2013. Le champ utilisé pour calculer ces parts a été restreint à la France métropolitaine, afin d'inclure dans le calcul de la marche de la méthode des parts logistique l'intégration des DOM (qui a en réalité été fait en 2014).

e) Taux d'emploi sous-jacent et espérance d'activité

Le taux d'emploi sous-jacent (moyenne arithmétique des taux d'emploi par âge détaillé) est utilisé pour annuler les éventuels poids démographiques inégaux des âges détaillés dans le calcul du taux d'emploi des 15-24 ans, 50-64 ans et 55-64 ans. L'espérance d'activité traduit, elle, le nombre d'année moyen d'activité restant aux 50-64 ans et 55-64 ans, si ces générations restaient soumises aux conditions d'activité observées

sur la période courante. Le calcul de cette grandeur nécessite également de connaître l'emploi par âge détaillé (par millésime), puisqu'il correspond à la somme des taux d'emploi par âge détaillé à partir de l'âge d'intérêt.

Ces deux indicateurs sont diffusés par sexe et uniquement dans des séries trimestrielles. Il faut donc rétropolier l'emploi trimestriel par sexe et âge détaillé. Calculer une rupture à partir du Pilote à ce niveau de détail semble imprudent au regard de la petite taille de cet échantillon : les déformations par âges détaillés risqueraient d'incorporer trop de bruit pour refléter un véritable phénomène économique. De ce fait, les parts annuelles de chaque âge détaillé dans l'emploi par âge quinquennal en ancienne enquête sont appliquées telles quelles à la série d'emploi annuelle rétropolée par âge quinquennal en nouvelle enquête, ce qui permet d'obtenir des séries annuelles de taux d'emploi et taux d'activité rétropolées en nouvelle norme. La même méthode est utilisée pour les séries trimestrielles, ce qui les rend ainsi cohérentes avec les séries annuelles.

f) Emploi en équivalent temps-plein

Le taux d'emploi en équivalent temps plein (ETP) correspond aux effectifs en emploi convertis en équivalent temps plein (calculé ici avec la quotité éventuelle de temps partiel¹²) dans l'emploi principal rapporté à la population totale. Il est diffusé uniquement à un niveau trimestriel et depuis 2003. Cet indicateur ne dépend pas du nombre d'heures effectivement travaillées : ainsi, une personne en emploi à temps plein mais n'ayant pas travaillé une semaine (congés, arrêts maladie, chômage partiel) compte malgré tout pour « 1 » dans le calcul d'ETP.

L'emploi en ETP est affecté par deux paramètres, aux effets inverses : le volume total de l'emploi, et la part de temps partiel dans l'emploi. Dans la mesure où ces deux indicateurs ont augmenté avec la nouvelle enquête, l'effet final sur l'emploi ETP n'est pas immédiat.

L'emploi en ETP a légèrement augmenté pour les hommes de moins de 25 ans, il est à peu près stable pour les hommes d'âge intermédiaire, et a diminué pour les plus de 50 ans (Figure 15). Parmi les femmes, il a diminué pour les moins de 25 ans et augmenté pour les femmes de 25 ans ou plus. La comparaison du rapport entre l'emploi ETP et l'effectif d'emploi total permet de montrer la déformation de la structure des emplois entre les deux enquêtes (Figure 10).

Ainsi, la baisse de l'emploi chez les jeunes femmes concerne surtout les emplois les plus partiels (ce qui fait que le ratio remonte pour elles), tandis que chez les femmes de plus de 25 ans, la hausse de l'emploi semble davantage concerner des emplois à faible quotité de travail (ce qui fait que ce ratio baisse).

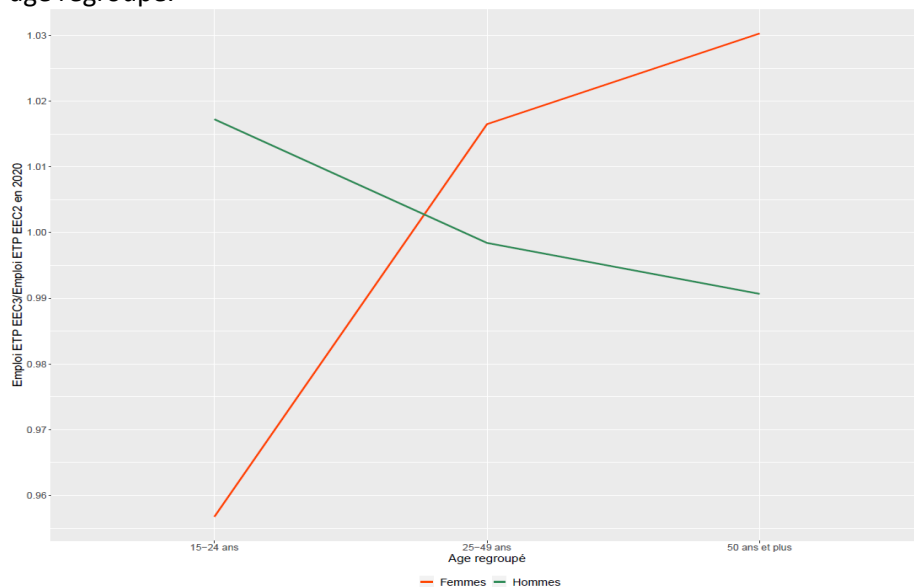
Pour les hommes, il ne semble pas y avoir une telle déformation de la structure des emplois, si ce n'est très marginalement chez les hommes d'âge intermédiaire pour lesquels le passage en nouvelle enquête semble avoir légèrement rehaussé la quotité de travail des personnes en emploi.

12 Pour chaque enquêté, son ETP est calculé à partir de la tranche de quotité de temps partiel déclarée et en fonction de son statut (salarié ou non) :

- les temps partiels de moins de 50 % ont un ETP de 0,25 pour les salariés et 0,30 pour les non-salariés ;
- les temps partiels à 50 % ont un ETP de 0,5 quel que soit le statut ;
- les temps partiels entre 50 et 80 % ont un ETP de 0,65 pour les salariés et 0,63 pour les non-salariés ;
- Les personnes à 80 % ont un ETP de 0,8 quel que soit leur statut ;
- Les temps partiels à plus de 80 % se voient affectés un ETP de 0,81 lorsqu'il s'agit de non-salariés, et 0,90 sinon ;

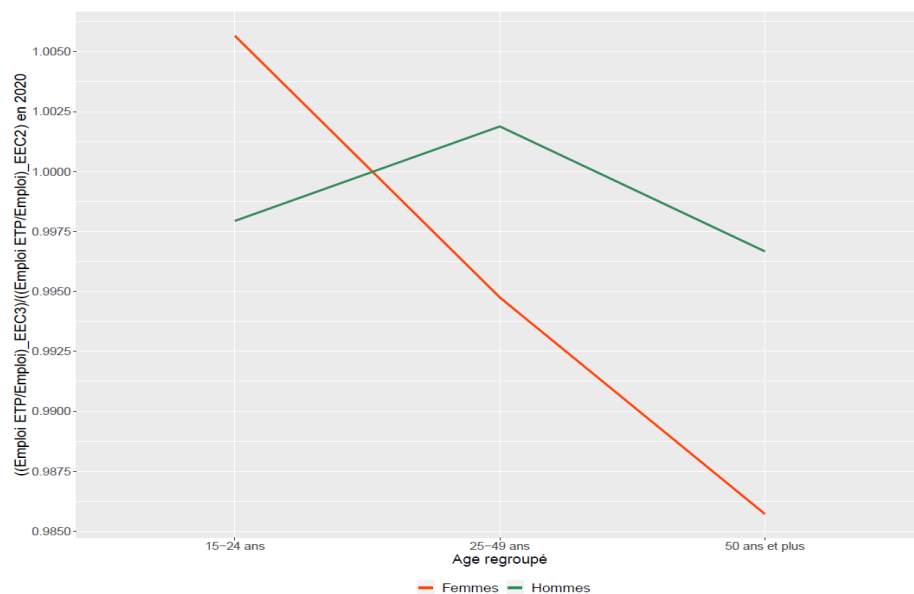
Enfin, les temps pleins ont un ETP de 1.

Figure 10 : Rapport entre l'emploi en ETP dans le Pilote et l'emploi en ETP dans l'EEC2 en 2020 par sexe et âge regroupé.



Lecture : L'emploi en ETP des femmes de 50 ans est plus a été multiplié par 1,03 dans la nouvelle enquête en 2020.
 Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire
 Source : Insee, enquête Emploi

Figure 11 : rapport entre l'emploi en ETP divisé par le niveau de l'emploi dans le Pilote et dans l'EEC2 en 2020 par sexe et âge regroupé.



Lecture : Le rapport entre l'emploi en ETP et l'emploi des femmes de 50 ans est plus a été multiplié par 0,986 dans la nouvelle enquête en 2020.
 Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire
 Source : Insee, enquête Emploi

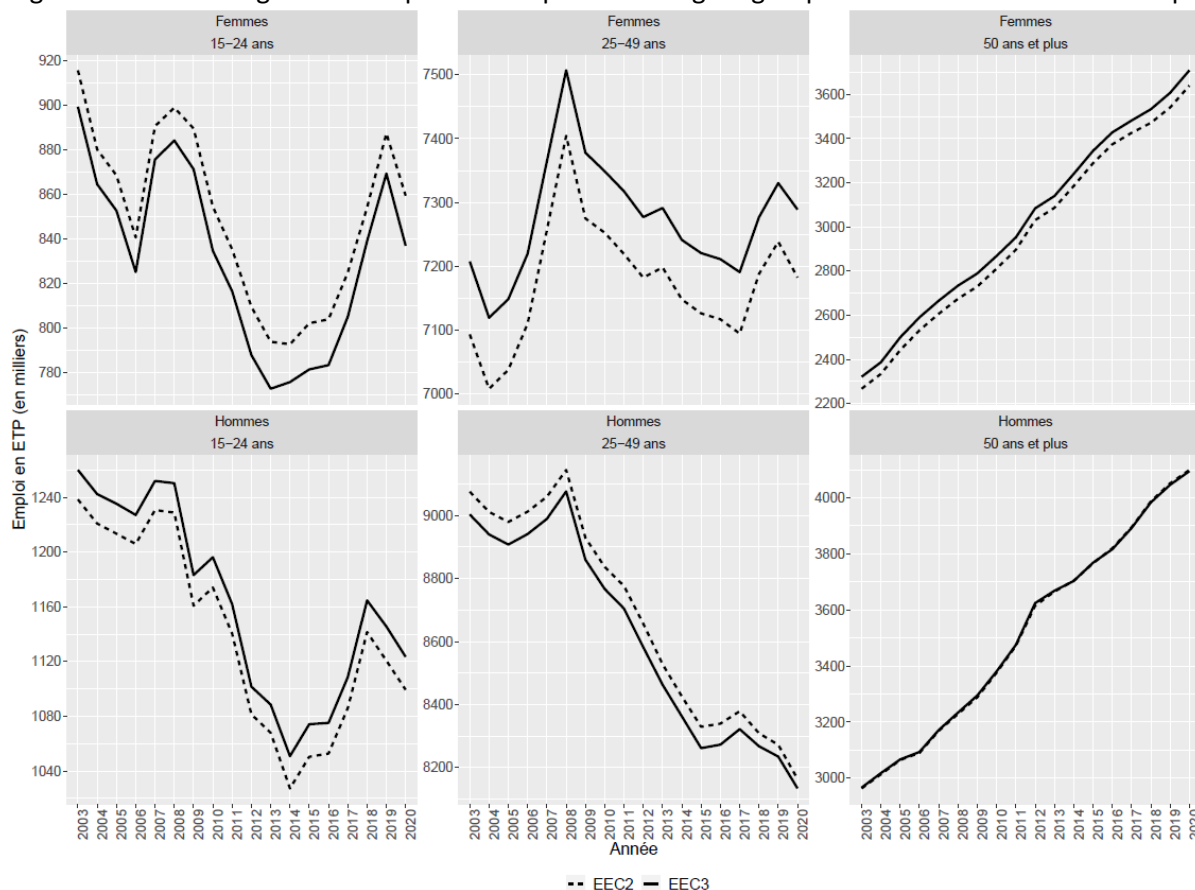
La rétopolation de la série trimestrielle d'emploi en ETP par sexe et âge quinquennal s'appuie sur la série trimestrielle d'emploi préalablement rétopolée en nouvelle norme : il s'agit d'associer à chaque emploi un taux d'ETP, afin de justement tenir compte de cette déformation de la structure des emplois.

La méthode des parts logistiques a pour cela été appliquée au rapport emploi ETP/emploi trimestriel dans les séries anciennes, à laquelle est appliquée la marche logistique calculée pour plus de robustesse. Une fois

ce rapport rétropolé en ancienne norme, il est appliqué à la nouvelle série trimestrielle d'emploi par sexe et âge quinquennal pour en déduire l'emploi ETP.

Des séries annuelles d'emploi en ETP en nouvelle norme ont ensuite pu être construites en faisant une simple moyenne annuelle de la série trimestrielle ainsi rétropolée (Figure 12).

Figure 12 : Séries longues de l'emploi en ETP par sexe et âge regroupé en ancienne et nouvelle enquête.



Lecture : En 2020, en ancienne enquête (EEC2), l'emploi en ETP des femmes de 25-49 ans est de 7 182 000, contre 7 289 000 en nouvelle enquête (EEC3).

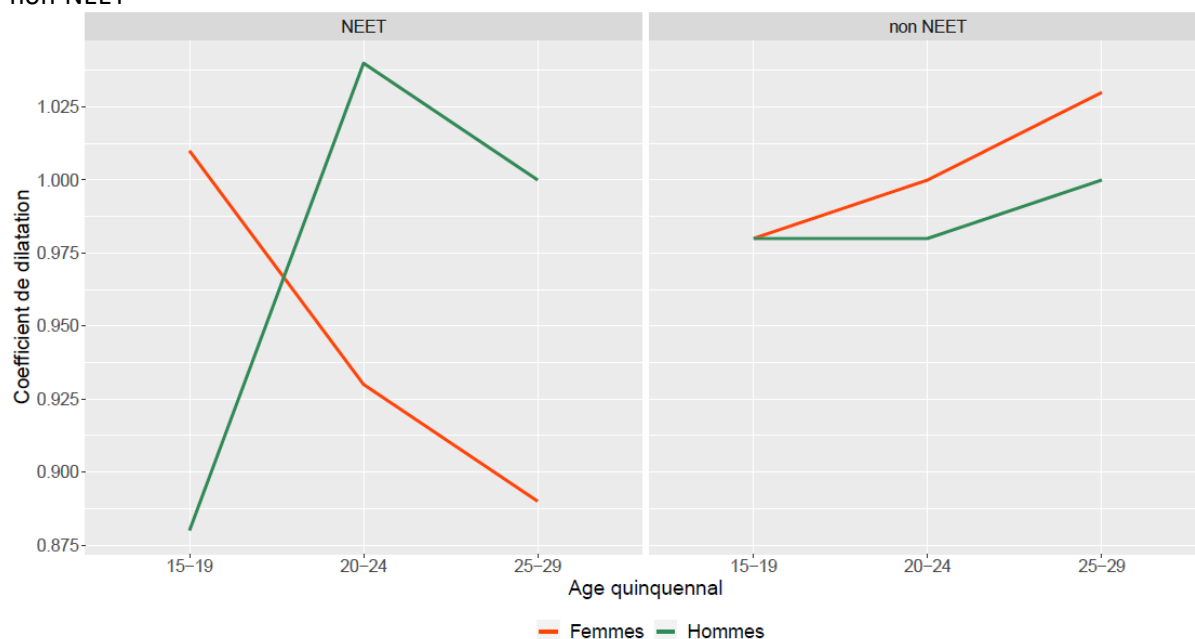
Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

g) NEET

Le nombre de jeunes de moins de 30 ans ni en emploi ni en formation (formelle ou non) (Neet, Not in Education, Employment or Training) est diffusé à fréquences annuelles et trimestrielles depuis 2003. La série annuelle en ancienne norme a été rétropolée via la méthode générale utilisant un coefficient de dilatation, ce dernier étant calculé pour 2 séries en parallèle : le nombre de Neet et le nombre de non-Neet.

Figure 13 : Coefficients de dilatation entre ancienne et nouvelle enquête sur le nombre de jeunes NEET et non-NEET



Lecture : Le nombre d'hommes de 20-24 ans ni en emploi, ni en études, ni en formation a été multiplié par 1,04 dans la nouvelle enquête par rapport à l'ancienne sur la période de recouvrement.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

Source : Insee, enquête Emploi

Comme le montre la Figure 13, le nombre de jeunes femmes ni en emploi ni en formation est très faiblement revu à la hausse chez les 15-19 ans, mais il diminue plus fortement chez les 20-24 ans et encore plus chez les 25-29 ans. C'est l'inverse qui se produit chez les hommes, puisque le nombre de jeunes hommes de 15-19 ans est très fortement revu à la baisse, tandis qu'il augmente chez les 20-24 ans et est inchangé chez les 25-29 ans.

Le nombre de jeunes non NEET (soit en emploi, soit en formation) varie moins : il est stable pour les hommes de 20-24 ans et les femmes de 25-29 ans, diminue légèrement pour les femmes de moins de 25 ans et pour les hommes de 15-19 ans, et augmente faiblement pour les hommes de 25-29 ans.

La hausse du nombre de NEET, qui concerne donc essentiellement les hommes de 20-24 ans, est surtout portée par la baisse du nombre de jeunes en formation dans cette tranche d'âge en nouvelle enquête. La forte baisse des jeunes femmes ni en emploi, ni en formation, dans la même classe d'âge a la même origine (ie la hausse du nombre de celles en formation). Il est difficile de trouver une explication à la révision opposée de la formation entre les hommes et les femmes de 20-24 ans, cette dernière concernant en particulier les formations non formelles tout en étant en partie contrebalancée par la hausse de l'emploi. Pour les femmes de 25-29 ans, une hausse de l'emploi intervient également, expliquée pour partie par la requalification des congés parentaux en emploi. La hausse de l'emploi chez les jeunes hommes de 15-19 ans à l'origine de la forte baisse du nombre de NEET est, elle, sans doute plus la conséquence du moindre recours au proxy ainsi que de la modification de la pondération via les règles de rattachement.

Les coefficients de dilatation obtenus ne sont pas lissés en raison du faible nombre de classes d'âge ainsi que des fortes spécificités de chacune. Ils sont appliqués à la série annuelle en ancienne enquête, puis les séries de NEET et non-NEET obtenues sont recalées afin de correspondre avec la population totale par sexe et âge quinquennal.

Les séries annuelles en nouvelle enquête ainsi obtenues (Figure 14) sont utilisées pour rétro-poler les séries trimestrielles selon la méthode générale décrite préalablement.

Figure 14 : Séries longues du nombre de jeunes ni en emploi, ni en études ni en formation par sexe et âge quinquennal en ancienne et nouvelle enquêtes



Lecture : En 2020, en ancienne enquête, le nombre de femmes de 25-29 ans ni en emploi, ni en études, ni en formation est de 415 000, contre 369 000 en nouvelle enquête.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

h) Formes d'emploi (CDI, CDD, etc.)

Les séries annuelle et trimestrielles du nombre de personnes dans chaque forme d'emploi sont produites par sexe et âge en 6 classes (15-19 ans, 20-24 ans, 25-49 ans, 50-64 ans, 65-74 ans et 75 ans et plus) depuis 1982. Jusqu'en 2020, les formes d'emploi publiées dans les séries longues de l'Insee regroupaient 5 catégories : les CDD et contrats saisonniers, les intérimaires, les apprentis, les CDI et les indépendants. Dans les séries longues jusque 2021, les personnes sans contrat ni stage étaient regroupées avec les CDI.

Lors de la refonte, il a été décidé d'isoler spécifiquement les « sans contrat » comme forme d'emploi. Il a donc fallu créer une série longue comptabilisant le nombre de personnes en emploi salarié mais sans contrat depuis 1982. Par ailleurs, pour les séries longues, il a été décidé de rétopoler une variable relativement agrégée, répartissant l'emploi salarié en 5 formes d'emploi : CDI¹³ ou fonctionnaire, CDD, intérim, alternance ou stage, sans contrat, indéterminé (variable STATUDET dans les fichiers de l'EEC à partir de 2021). En pratique, dans les séries longues publiées en 2021, les modalités « CDI ou fonctionnaire » et « indéterminé » sont agrégées, mais l'information détaillée reste disponible dans les micro-données depuis 2021.

Création d'une série longue sur les salariés sans contrat

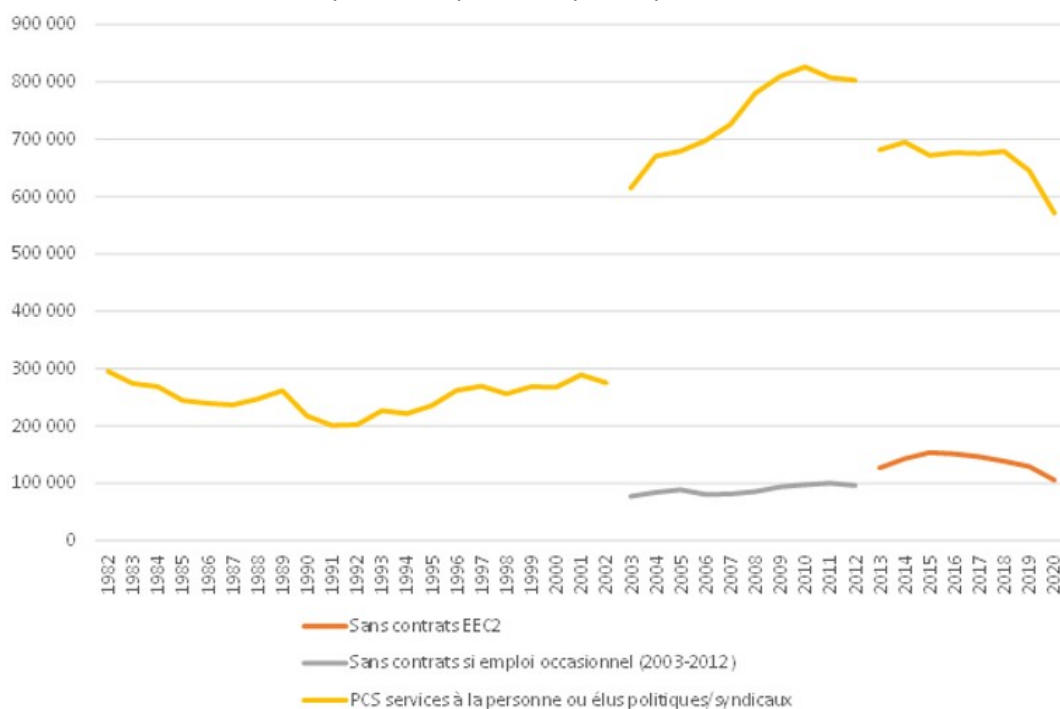
13 La variable CDI inclut ici les CDI intérimaires, cette information détaillée étant disponible dans l'EEC à partir de 2021 (variable SALDET)

Sur la période 2013-2020, il existait dans l'enquête Emploi une variable permettant d'identifier ces personnes¹⁴. Entre 2003 et 2012, la modalité « Pas de contrat de travail » (« 0 » dans la variable CONTRA) n'existait pas, et l'existence d'un contrat de travail (variable ACONTR) n'était posée qu'aux personnes occupant un emploi occasionnel, le périmètre était donc beaucoup plus restreint.

Une deuxième approche consiste à étudier les PCS des personnes dans la catégorie « Sans contrat » les plus fréquentes dans le Pilote. Ainsi, trois professions regroupent 37 % des salariés sans contrat dans le Pilote : les services d'aide à la personne (56E1 pour les employés de maison et personnels de ménage chez des particuliers ; 56D2 pour les autres employés d'aide à domicile et accueillants familiaux) et les responsables politiques et syndicaux (33E1). Même si ces professions ne sont que minoritairement occupées par des salariés sans contrat (12 %), leur évolution dans le temps peut constituer un proxy, en l'absence de toute information sur les salariés sans contrat avant 2002. En effet, comme le montre la Figure 20, sur les périodes où les informations sur les sans contrat existent dans l'EEC, l'évolution du nombre de sans contrat qu'elles reflètent va dans le même sens que celle du nombre de personnes en emploi dans ces PCS, même si les variations sont de plus faible ampleur.

L'effectif historique de ces 3 PCS (qui correspondaient alors aux professions 335A, 563B, 563C dans la nomenclature de 2003, et 3318 et 5632 dans celle de 1982) semble valider l'hypothèse d'une stabilité au cours du temps.

Figure 15 : Nombre non rétropolé de personnes sans contrat et dans les PCS de services à la personne à domicile et d'élus recensés par les enquêtes Emploi depuis 1982.



Lecture : En 1982, l'enquête emploi comptait près de 300 000 personnes en emploi d'élus politiques ou syndicaux, ou travaillant dans les services à la personne à domicile.

Champ : France métropolitaine jusqu'en 2013, France hors Mayotte à partir de 2014, personnes de 15 ans ou plus, vivant en logement ordinaire.

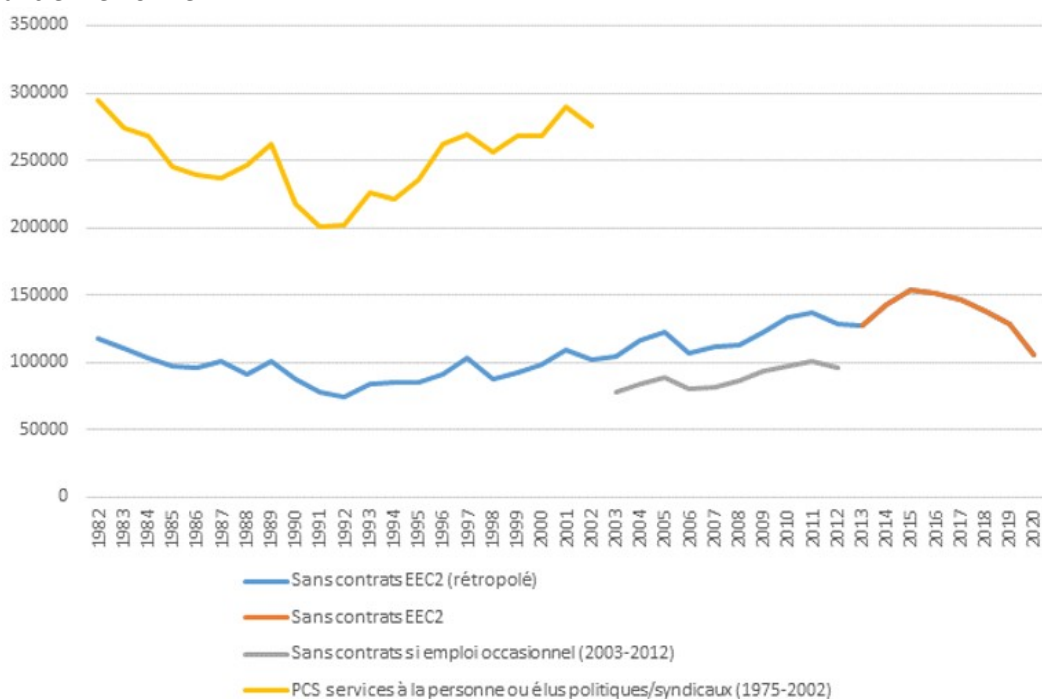
Sources : Insee, enquêtes Emploi.

14 En pratique il s'agit du filtre suivant (ACTEU == "1" & CONTRA == "0" & STATUTR == "5" & STAGE == "2" & TITC != "4" & TYPC != "2")

Pour constituer la série longue sur les salariés sans contrat, la première étape consiste à raccrocher la courbe des personnes sans contrat en emploi occasionnel de 2003 à 2013 (courbe grise Figure 15) avec celle des 3 PCS sélectionnées de 1982 à 2002 (courbe jaune). Pour cela, un coefficient multiplicatif a été déterminé pour chaque croisement de sexe et d'âge regroupé et appliqué à la série de PCS de 1982 à 2002, de manière à obtenir une variation entre 2002 et 2003 cohérente avec celle observée entre 2001 et 2002 et celle observée entre 2003 et 2004.

La même méthode est appliquée pour raccrocher la nouvelle série 1982-2013 ainsi créée avec celle de 2014-2020 (recalculée sur la France métropolitaine uniquement, pour ne prendre en compte que la rupture de mesure des sans contrats sans mélanger ses effets avec l'intégration des DOM). Ce n'est que dans un deuxième temps qu'un coefficient multiplicatif correctif est appliqué à la série 1982-2020 alors obtenue afin de corriger l'élargissement aux DOM du champ de l'enquête. Ce coefficient de passage, calculé par sexe et classe d'âge, correspond au rapport moyen entre les effectifs des sans contrat avec et sans les DOM sur la période 2014-2020. Par la suite, la période 2014-2020 est mise à jour avec la vraie série des sans contrats en France entière depuis 2014, ce qui conduit à une série en ancienne norme du nombre de personnes en emploi sans contrat présentée en Figure 16.

Figure 16 : Construction d'une réropolante annuelle du nombre de personnes en emploi sans contrat en ancienne norme



Lecture : En 1982, 117 000 personnes sont en emploi sans contrat au sens de l'ancienne enquête (EEC2).

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Source : Insee, enquêtes Emploi.

Une réropolante trimestrielle du nombre de sans contrat est construite en reproduisant exactement la méthode adoptée pour construire la série annuelle. Ces deux réropolantes en effectifs de sans contrats sont ensuite converties en part des sans contrats dans l'emploi, pour que les ruptures de niveau de l'emploi qui ont été réropolées dans les séries longues puissent être prises en compte.

Création d'une série longue sur les salariés en alternance ou en stage

Par rapport aux séries longues existant jusqu'en 2020, la modification majeure consiste en un élargissement de la modalité « Apprentis » afin d'y inclure toutes les formes d'alternance (y compris contrats professionnels) et stages, auparavant principalement comptabilisées parmi les CDD. La difficulté réside dans l'évolution au cours du temps de ces concepts et de leur dénomination. Si les apprentis ont toujours été

définis par le contrat d'apprentissage (créé en 1927)¹⁵ ce n'est pas le cas du contrat de professionnalisation, créé en 2004, qui synthétise et remplace les 3 contrats en alternance créés en 1993 : contrat d'adaptation, contrat de qualification et contrat d'orientation et (Dassault, 2007), ce dernier succédant au stage d'insertion dans la vie professionnelle (SIVP) apparu en 1983 tandis que le contrat de qualification succédait au contrat emploi-formation (CEF) institué en 1975 (Gouip et Trimouille, 1982). Les contrats de professionnalisation et ses ancêtres sont repérés dans l'enquête Emploi en continu¹⁶ et auparavant dans l'enquête Emploi annuelle¹⁷. Néanmoins, Minni et Picart (2016) soulignent la forte sous-estimation du nombre de contrats de professionnalisation par l'enquête Emploi de 2003 à 2014.

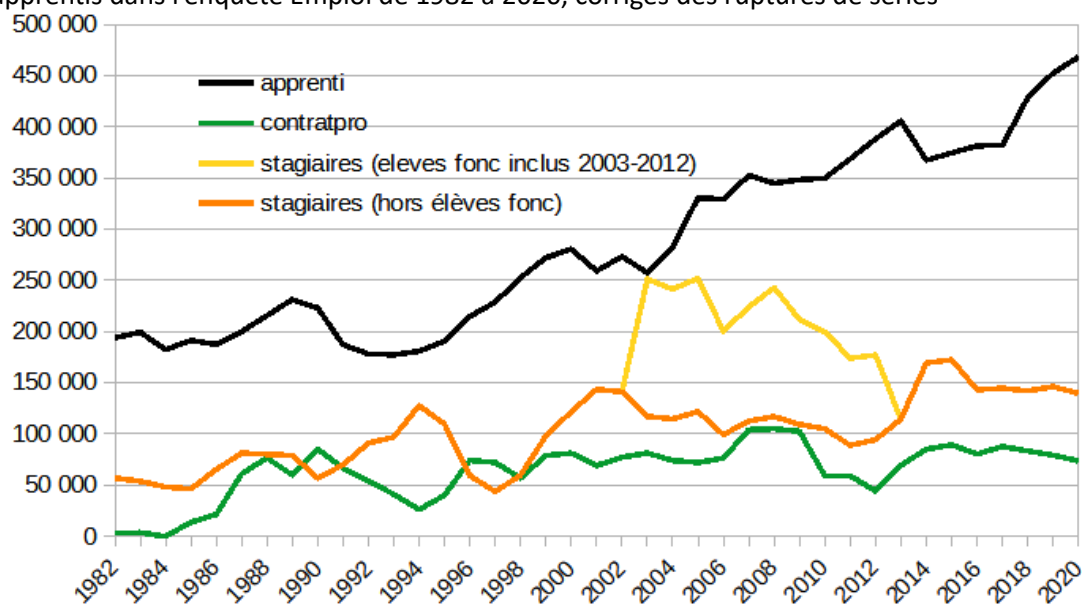
Enfin, les stages, désormais inclus dans la nouvelle catégorie de forme d'emploi, peuvent être identifiés dans l'enquête Emploi annuelle¹⁸ et dans l'enquête en continu¹⁹.

Plusieurs corrections sont apportées lorsque des anomalies ou ruptures semblent être présentes :

- en 1984, aucun « contrat de professionnalisation » n'est trouvé alors même qu'il y en a en 1983 et 1985 ; il est donc imputé comme la moyenne entre celui de ces deux années ;
- entre 1989 et 1990 ainsi qu'entre 2002 et 2003, il semble y avoir des ruptures inexplicables dans le nombre de stagiaires ; les séries avant 1989, puis avant 2002, sont translatées pour que l'évolution ces années-là soit du même ordre que celle du nombre d'apprentis dans chaque croisement de sexe et âge regroupé.

Par ailleurs sur la période 2003-2013, il n'est pas possible de distinguer les élèves-fonctionnaires des stagiaires dans le public, ce qui provoque une rupture à la hausse dans la reconstitution des courbes en niveau sur cette période (Figure 17, courbe jaune).

Figure 17 : Reconstitution du nombre de stagiaires, contrats de professionnalisation (et ancêtres) et apprentis dans l'enquête Emploi de 1982 à 2020, corrigés des ruptures de séries



Lecture : En 1982, près de 200 000 personnes sont en emploi en apprentissage sans contrat au sens de l'ancienne enquête.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Source : Insee, enquêtes Emploi.

15 Identifiés dans l'enquête emploi par la modalité 22 de la variable STATUT.

16 Par la variable TYPC de 2003 à 2020 (modalité « 2 » à partir de 2013, modalité « 1 » de 2010 à 2012, modalités « 1 », « 2 » et « 3 » et « 12 » de 2003 à 2009)

17 Par la variable STAGE : modalités 23 à 25 de 1994 à 2002, 11 et 23 à 25 de 1990 à 1993, et 17, 33 et 14 de 1982 à 1989)

18 Comme les modalités restantes de la variable STAGE pour les personnes identifiées en emploi et en stage par à la fois la variable STAGE et la variable STATUT, auxquelles sont préalablement retirées tous les contrats aidés hors alternance (ces derniers correspondent aux modalités 15, 26 et 27 de 1982 à 1989, 21 et 22 de 1993 à 1993, et 21, 22 et 26 de 1994 à 2002.

19 Par les variables STAGE (dans le privé) et TITC (dans le public, hors élèves fonctionnaires) à partir de 2014, et par les variables NATSTG (non vide) et TITC avant.

Pour corriger cela, la part moyenne sur 2014-2020 des fonctionnaires-stagiaires dans l'ensemble des salariés du public est calculée pour chaque croisement de sexe et d'âge regroupé, puis est appliquée sur 2003-2013 au nombre de personnes travaillant dans le public par sexe et âge pour estimer, chaque année, le nombre de fonctionnaires-stagiaires à retirer du total des stagiaires. Le nombre de stagiaires alors obtenu est représenté par la courbe orange de la Figure 17.

Enfin, comme pour les séries de personnes en emploi sans contrat, le nombre de personnes en stage et en contrat de professionnalisation avant 2014 est converti de France métropolitaine à France entière à partir de coefficients de passage calculés par sexe et âge regroupé en moyenne sur la période 2014-2020.

Ainsi, comme pour les sans-contrats, une série annuelle en ancienne norme du nombre d'alternants/stagiaires/contrats de professionnalisation a pu être construite de 1982 à 2020. En adoptant des étapes identiques à celles décrites précédemment, une série trimestrielle en norme EEC2 est également construite, puis les quatre séries en niveau sont converties en part de l'emploi afin de s'affranchir des ruptures de séries dans le niveau d'emploi.

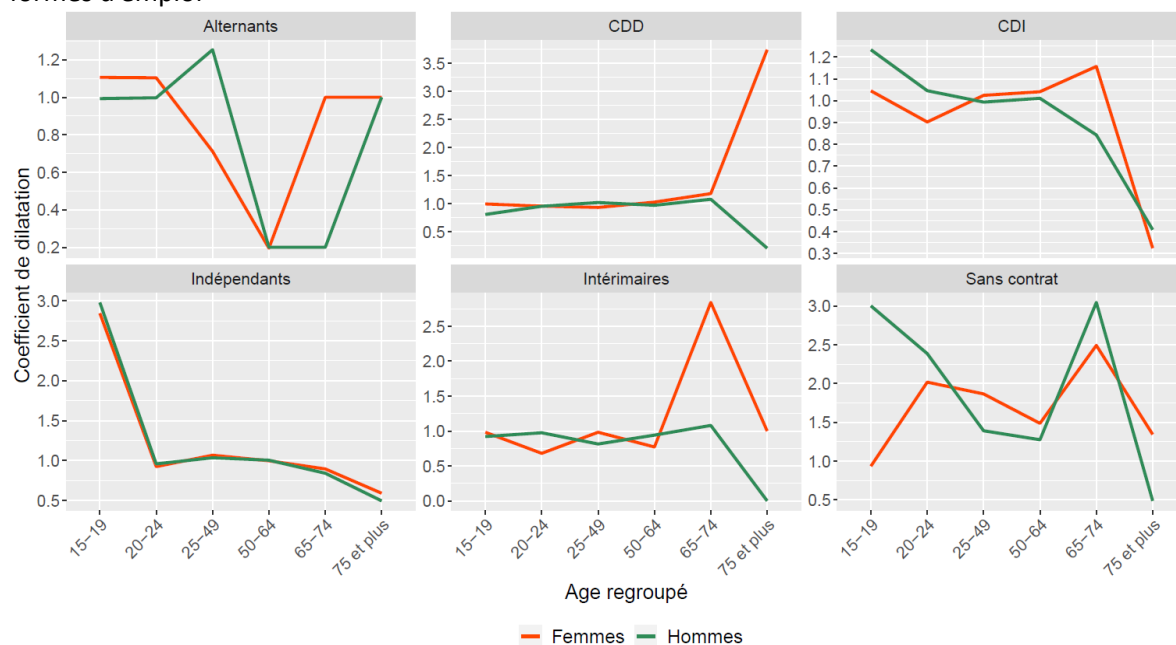
Une fois ces rétopolantes construites, les effectifs en EEC2 ont été calculés selon la nouvelle nomenclature des formes d'emplois, c'est-à-dire après transferts des contrats de professionnalisation et stages des CDD vers les alternants et création des sans contrats retirés aux CDI. Puis, les séries historiques de forme d'emploi ont été rétopolées en adoptant la méthode générale.

Les personnes pour lesquelles le statut n'est pas renseigné sont imputées en CDI, cette dernière forme d'emploi étant la plus fréquente. Les coefficients de dilatation obtenus ont été légèrement modifiés pour certaines catégories en raison de valeurs aberrantes obtenues :

- Pour les hommes de 50 à 64 ans et 65 à 74 ans, il n'y a pas d'alternants/stages/contrats de professionnalisation dans le Pilote, ce qui conduit à un coefficient de dilatation nul. Pour éviter de passer à 0 tout l'historique des séries, le coefficient de dilatation pour ces classes d'âge est conventionnellement fixé au même niveau que celui obtenu pour les femmes de 50-64 ans (0,2).
- Le coefficient de dilatation des hommes sans contrats de 15 à 19 ans est supérieur à 5, ce qui est peu crédible : dans la mesure où c'est la prise en compte du cinquième trimestre de recouvrement (T1 2021) qui fait augmenter ce coefficient, il est fixé à sa valeur calculée uniquement sur les 4 trimestres de 2020, qui correspond à une multiplication par 3.

Les coefficients de dilatation ainsi obtenus sont présentés Figure 18.

Figure 18 : Coefficients de dilatation entre ancienne et nouvelle enquête sur le nombre des différentes formes d'emploi



Lecture : Il y a 1,1 fois plus de femmes de 15-19 ans en alternance dans la nouvelle enquête par rapport à l'ancienne sur la période de recouvrement.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Source : Insee, enquête Emploi.

Les valeurs très élevées ou très faibles des coefficients de dilatation obtenus pour les 65-74 ans et les 75 ans et plus s'expliquent par le faible effectif de personnes en emploi dans ces catégories d'âge, qui conduit à une forte volatilité de chaque forme d'emploi dans le Pilote comme dans l'enquête Emploi. S'ils ne sont pas toujours crédibles, une valeur un peu biaisée n'a pas d'impact très significatif puisqu'il s'applique à de très faibles effectifs. C'est également le cas pour des formes d'emploi à certains âges, comme les indépendants chez les 15-19 ans.

La hausse du nombre de CDI pour les femmes d'âge intermédiaire peut s'expliquer en partie par la requalification des congés parentaux et maladie en emploi, ces derniers étant souvent occupés par des personnes en contrat à durée indéterminée ou des fonctionnaires. Cette hausse est néanmoins faible et nulle chez les hommes, sans doute car compensée en partie par le retrait des personnes sans contrat de la catégorie CDI. De la même manière, la hausse de l'emploi total ne paraît pas dans le nombre de CDD qui diminue même très légèrement, notamment en raison du retrait des stages et contrats de professionnalisation, désormais comptabilisés dans la catégorie Alternants. Ces derniers sont revus légèrement à la hausse chez les femmes de 15-24 ans, alors qu'ils n'évoluent pas chez les hommes dans cette classe d'âge, la hausse se situant pour eux chez les 15-49 ans.

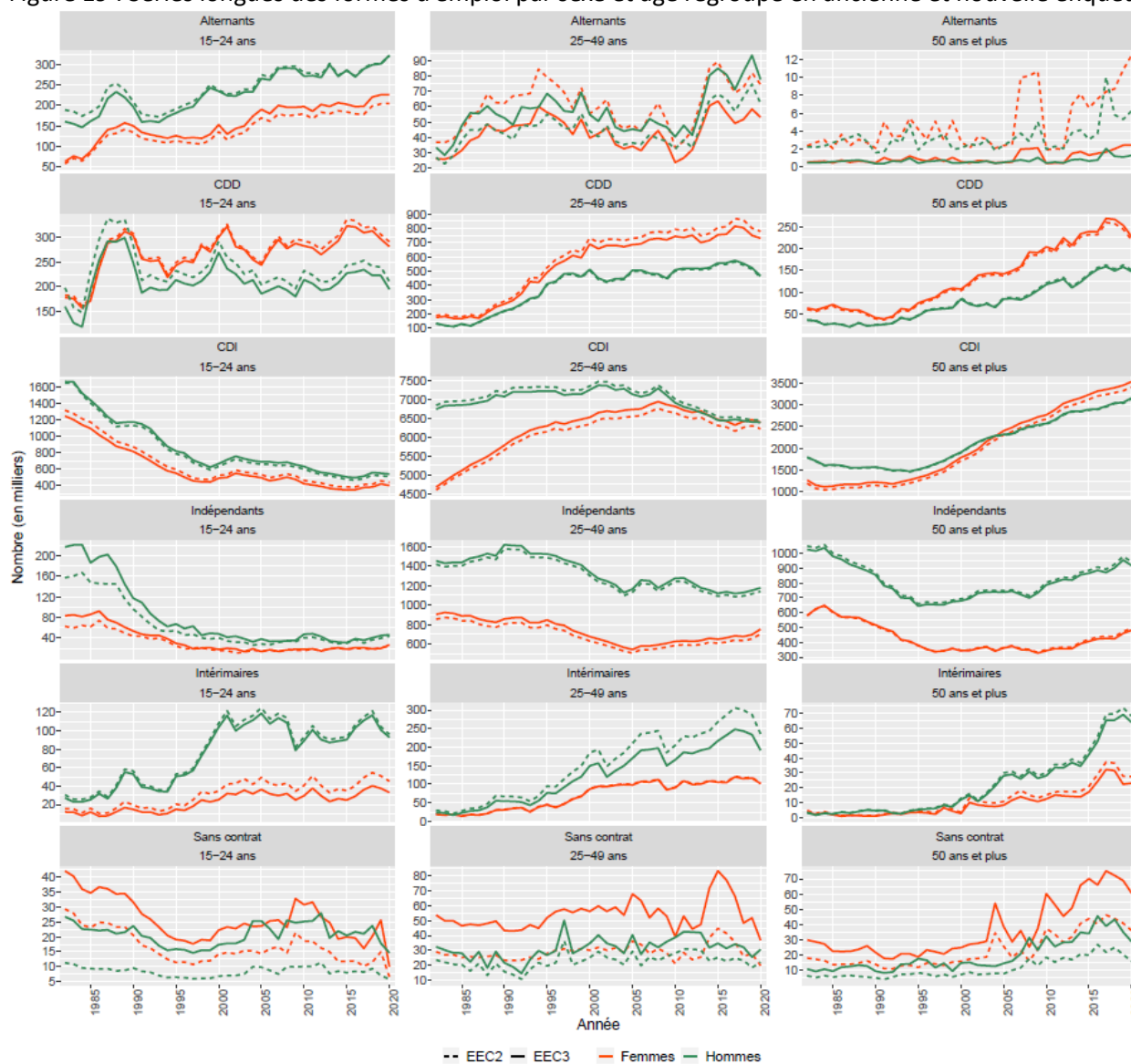
Le nombre de personnes sans contrat a beaucoup augmenté pour quasiment tous les croisements de sexe et âge, ce qui peut venir du fait que cette modalité est désormais incluse dans les variables de statut calculées et redressées à l'aide de nombreuses informations, alors que dans l'ancienne enquête elle ne pouvait être identifiée aussi précisément.

Ces coefficients de dilatation sont appliqués aux séries historiques des formes particulières d'emploi, la modalité des sans contrat construite précédemment (cf. supra) est ajoutée (et retirée de celle sur les CDI), les stages et contrats de professionnalisation sont ôtés des CDD pour être ajoutés aux alternants. Chaque année, les effectifs de sans contrat, contrats professionnels et stagiaires à ajouter et retrancher sont calculés

en appliquant les parts représentées dans l'emploi calculées précédemment aux dernières séries d'emploi en ancienne norme, pour lesquelles toutes les ruptures antérieures à 2020 ont été rétropolées.

Les séries annuelles ainsi obtenues sont présentés en Figure 19, dans laquelle les tranches d'âges ont été agrégées de 6 à 3 classes. Elles sont utilisées pour rétropoler les séries trimestrielles, à l'aide de la méthode générale de trimestrialisation après avoir corrigé les séries trimestrielles en ancienne norme de sans contrats, stagiaires et contrats de professionnalisation (cf. supra).

Figure 19 : Séries longues des formes d'emploi par sexe et âge regroupé en ancienne et nouvelle enquêtes



Lecture : En 2020, en ancienne enquête, 205 000 femmes de 15-24 ans sont en emploi en alternance, contre 226 000 en nouvelle enquête.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

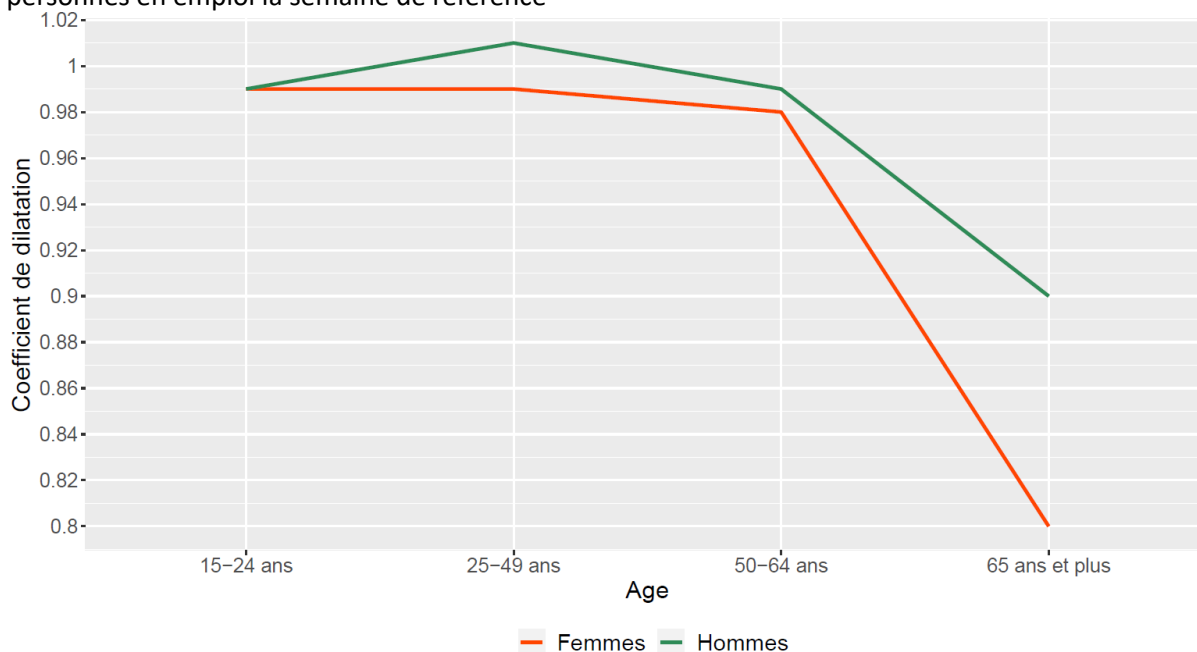
Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

i) Nombre moyen d'heures effectivement travaillées par semaine sur l'ensemble des emplois

Lors du passage en nouvelle enquête, l'imputation de la variable renseignant le nombre moyen d'heures travaillées la semaine de référence²⁰ a été intégrée à la chaîne aval ; la série historique est calculée par sexe et âge en 4 classes (15-24 ans, 25-49 ans, 50-64 ans, 65 ans et plus) depuis 2014, et publiée uniquement au niveau trimestriel pour l'ensemble des 15-64 ans.

Les coefficients de dilatation obtenus par le rapport des heures hebdomadaires travaillées sont présentés Figure 20.

Figure 20 : Coefficients de dilatation du nombre moyen d'heures effectivement travaillées pour les personnes en emploi la semaine de référence



Lecture : Entre l'ancienne et la nouvelle enquête, sur la période de recouvrement, le nombre moyen d'heures effectivement travaillées par les femmes de 50-64 ans a diminué avec un facteur multiplicatif de 0,98.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Source : Insee, enquête Emploi.

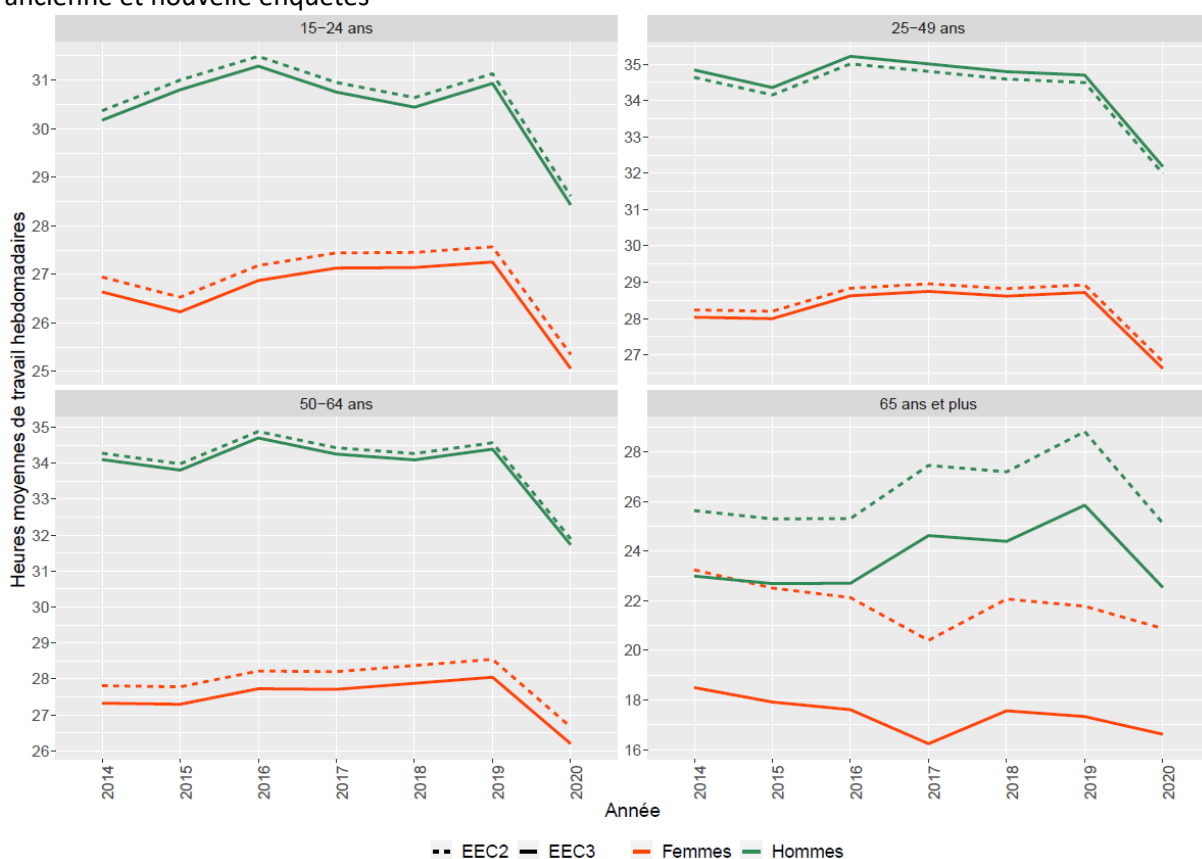
La légère baisse du nombre d'heures travaillées pour les femmes de 25-49 ans peut s'expliquer par la requalification en emploi des congés parentaux, les personnes concernées ayant alors 0 heure travaillée dans la semaine. De la même manière, les arrêts maladie de longue durée désormais considérés comme de l'emploi contribuent sans doute à la baisse des heures effectivement travaillées parmi les hommes et femmes de 50 ans et plus. En revanche, pour les hommes de 25-49 ans, le nombre d'heures travaillées est légèrement rehaussé en nouvelle enquête, ce qui peut résulter à la fois d'un effet proxy (surtout pour les personnes ayant travaillé un très faible nombre d'heures dans la semaine, que le proxy peut résumer à 0) mais également d'un effet de mode, les personnes travaillant un nombre élevé d'heures hebdomadaires pouvant être plus enclines à répondre par Internet, ayant davantage de temps pour les estimer que dans une interrogation avec enquêteur.

20 En ancienne enquête, c'est la variable TOTNBH qui donne le nombre moyen d'heures travaillées la semaine de référence. Lorsqu'elle était non renseignée, c'est le nombre d'heures travaillées dans l'emploi principal (EMPNBH) qui était utilisée. En nouvelle enquête, elle est remplacée par la variable HEFFTOT, qui est redressée de la non-réponse.

Enfin, la baisse du nombre d'heures travaillées chez les 15-24 ans résulte en partie d'un effet de composition, puisque l'emploi des 15-19 ans a augmenté alors que celui des 20-24 ans a diminué pour les femmes et a augmenté mais dans une moindre mesure pour les hommes ; or les 15-19 ans sont plus concernés par les petits boulots ou temps partiels que les 20-24 ans, ce qui contribue à faire chuter la moyenne du nombre d'heures travaillées.

Les séries en ancienne mesure sont ensuite dilatées à l'aide de ces coefficients de dilatation, sans qu'il ne soit nécessaire d'effectuer ensuite aucun recalage puisqu'il n'y a pas de séries de référence concernant le nombre moyen d'heures de travail hebdomadaire. Les séries annuelles obtenues sont présentées en Figure 21.

Figure 21 : Séries longues du nombre moyen d'heures travaillées par semaine, par sexe et âge regroupé en ancienne et nouvelle enquêtes



Lecture : En 2020, en ancienne enquête, le nombre moyen d'heures hebdomadaires effectivement travaillées par les femmes de 50-64 ans était de 26,7, contre 26,2 en nouvelle enquête.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

Ces séries ont ensuite permis de recalibrer les séries trimestrielles rétro-polées selon la méthode générale.

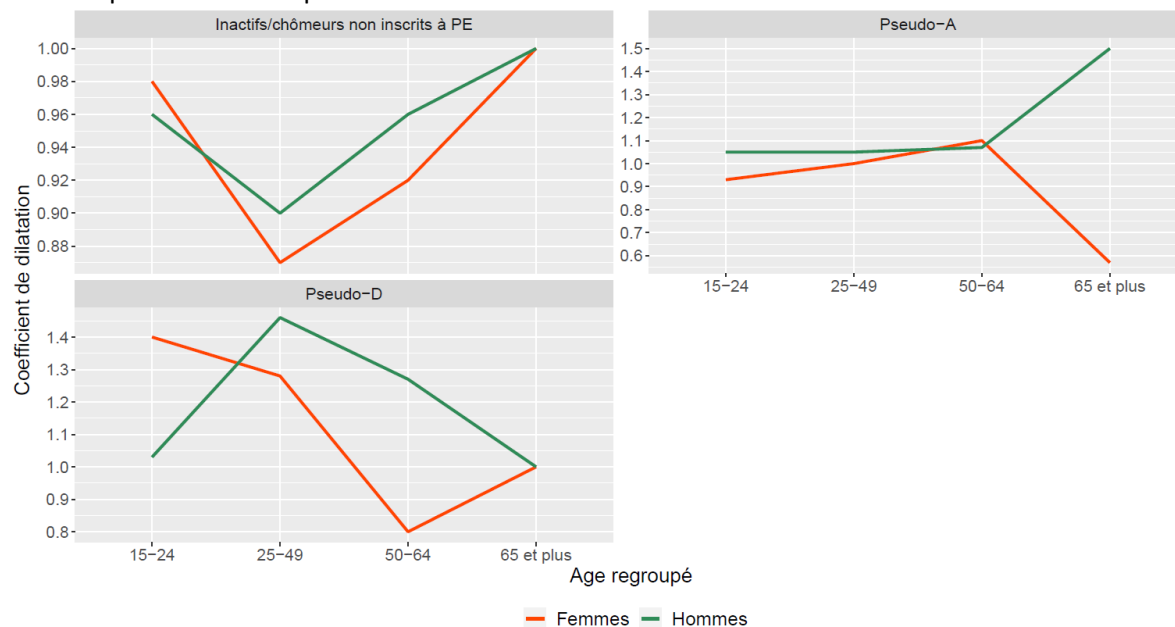
j) Inscription à Pôle Emploi

Depuis 2003 l'enquête Emploi pose la question de l'inscription à Pôle Emploi (anciennement ANPE)²¹, il est possible de construire des séries reconstituant les différents types de demandeurs d'emploi inscrits en fin de mois à Pôle Emploi (DEFM). Sont ainsi construites des séries d'effectifs dans les catégories « Pseudo-A » (qui approche la catégorie des DEFM A : personnes inscrites à Pôle Emploi mais ni en emploi ni en formation formelle ou informelle au cours des 4 dernières semaines) et « Pseudo-D » (qui approche la catégorie des

21 Avec la variable OFFICC.

DEFM D : personnes inscrites à Pôle Emploi, sans emploi mais ayant suivi une formation formelle ou informelle au cours des 4 dernières semaines), ainsi que la série des complémentaires des pseudo A et D dans l'ensemble chômage et inactivité, c'est-à-dire les chômeurs ou inactifs qui ne sont pas inscrits à Pôle Emploi. Ces séries sont ventilées par sexe et âge regroupé en 4 classes.

Figure 22 : Coefficients de dilatation du nombre de chômeurs et inactifs non-inscrits à Pôle Emploi, et des DEFM « pseudo-A » et « pseudo-D »



Lecture : En nouvelle enquête, sur la période de recouvrement, le nombre d'hommes de 15-24 ans demandeurs d'emploi en fin de mois « Pseudo-A » a été multiplié par 1,05 par rapport à l'ancienne enquête.

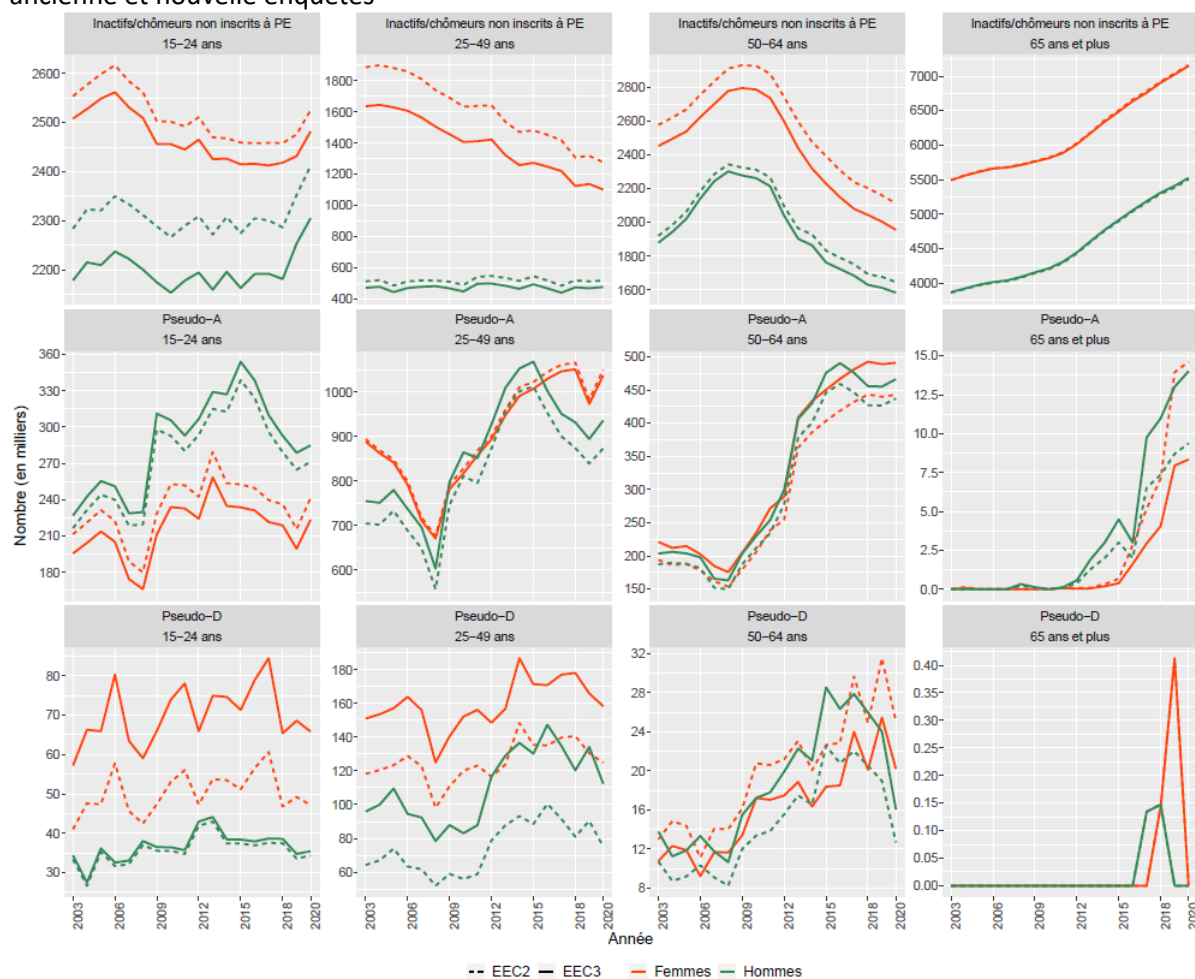
Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Source : Insee, enquête Emploi.

Les coefficients de dilatation présentés Figure 22 montrent que le nombre de chômeurs et d'inactifs non-inscrits à Pôle Emploi a diminué, en particulier pour les 25-49 ans : cela peut découler d'un effet conjugué d'une baisse du recours au proxy, mais surtout d'un effet Internet qui réduit le biais de désirabilité sociale et la réticence que peuvent avoir certains enquêtés à se présenter comme inscrits à Pôle Emploi face à un enquêteur. Ils se sont principalement reportés sur la catégorie des pseudo-D, notamment en raison d'une modification du questionnaire (reformulation des questions qui pourrait avoir affecté les réponses concernant les formations non formelles). Le nombre de pseudo-A a lui été relativement moins affecté par le changement d'enquête, il diminue pour les jeunes femmes et augmente pour les hommes.

La rétopolation des séries suit la méthode générale, en calant les séries dilatées pour retrouver les effectifs rétopolés des deux statuts chômeurs et inactifs regroupés. Elle conduit aux séries en nouvelle norme présentées en Figure 23. Elles sont ensuite utilisées pour la rétopolation des séries trimestrielles.

Figure 23 : Séries longues des demandeurs d'emploi en fin de mois (DEFM), par sexe et âge regroupé, en ancienne et nouvelle enquêtes



Lecture : En 2020, en ancienne enquête, le nombre d'hommes de 25-49 ans DEFM Pseudo-A est de 873 000, contre 936 000 en nouvelle enquête.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

k) Statut d'activité au sens du BIT selon le plus haut niveau de diplôme obtenu (1^{ère} version : série 2022)

L'objectif est de pouvoir décliner par diplôme les séries du nombre de chômeurs, de personnes inactives et en emploi, la population totale, le taux d'emploi, le taux de chômage, et la part du diplôme dans l'emploi, par sexe et âge regroupé en 3 classes, qui sont publiées chaque année dans des séries débutant en 1982. Dans la mesure où des indicateurs supplémentaires à fournir à Eurostat nécessitent des croisements d'âge plus fins, les effectifs de personnes en emploi, au chômage ou inactives par diplôme sont rétro-polées depuis 1982 par sexe et âge quinquennal.

Le diplôme était auparavant décliné en 6 modalités²² (supérieur à Bac+2 ; équivalent Bac+2 ; Bac ou équivalent ; CAP/BEP ou équivalent ; Brevet ; Aucun diplôme ou certificat d'études primaires). Pour les séries longues en nouvelle norme, il a été décidé de fusionner les deux dernières modalités, la distinction étant peu pertinente et les effectifs devenus très faibles conduisant à une forte volatilité.

22 Variable DDIPL.

Pour de rares individus, le niveau de diplôme n'est pas renseigné : 0,3% des 15-89 ans en 2021. Dans l'exercice de réropolation, le recalage sur l'emploi, le chômage et l'inactivité au sens du BIT permet de répartir ces non-répondants dans les niveaux de diplômes renseignés, ce qui conduit à de légères différences de niveau en 2021 entre séries réropolées et données agrégées à partir des données individuelles.

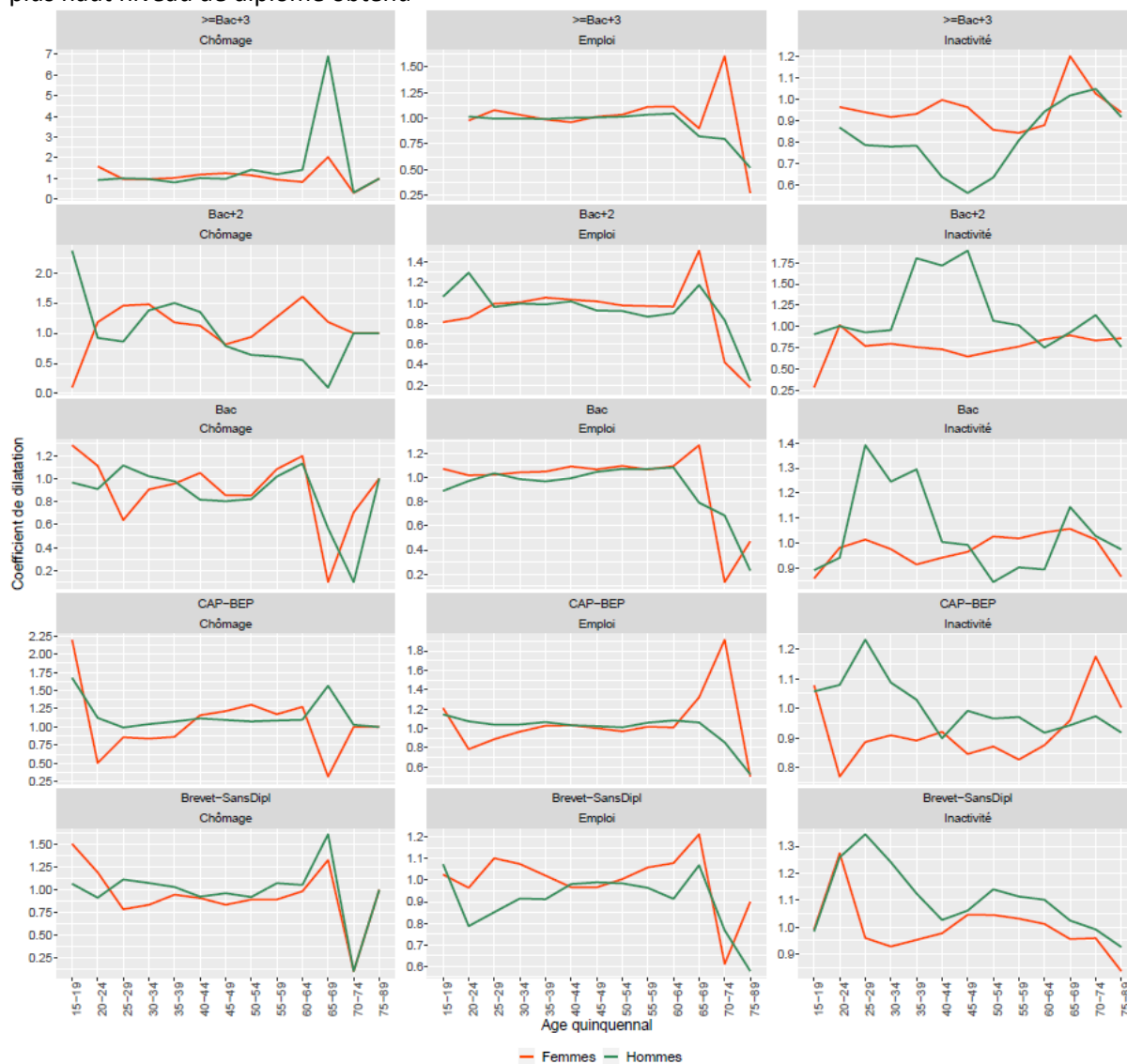
Du fait de la restriction du champ de passation du questionnaire individuel de l'EEC en 2021 (15-89 ans au lieu de 15 ans ou plus avant), il n'est plus possible de produire des séries sur les 15 ans ou plus concernant les indicateurs liés au diplôme. Le recalage sur la population totale ou l'inactivité totale des séries réropolées ne peut donc pas être fait pour les 75 ans et plus. Il faut préalablement retirer aux inactifs de 75 ans et plus la population totale de 90 ans et plus, effectif qui peut être retrouvé dans les marges de population de l'enquête Emploi.

Le calcul des coefficients de dilatation a nécessité plusieurs ajustements afin de prendre en compte les petits effectifs de certains croisements :

- Pour les 15-19 ans ayant un diplôme de niveau Bac+3 ou plus, les coefficients de dilatation multiplicatifs pour l'emploi, le chômage et l'inactivité créent une forte volatilité sur ces séries aux effectifs faibles ; ils ont donc été remplacés par des coefficients additifs ;
- De même, en raison de la petite taille d'échantillons dans le Pilote, plusieurs croisements de sexe, âge et niveau de diplôme n'y sont pas représentés alors qu'ils sont non vides en ancienne enquête sur la même période, en particulier pour le chômage. Plutôt que de fixer les coefficients de dilatation à 0, la valeur du plus petit coefficient de dilatation calculé sur les autres croisements a été retenue.

Les coefficients de dilatation obtenus, lissés par moyenne mobile pour les 25-64 ans, sont présentés Figure 24.

Figure 24 : Coefficients de dilatation du nombre de personnes en emploi, au chômage ou inactives selon le plus haut niveau de diplôme obtenu



Lecture : En nouvelle enquête, sur la période de recouvrement, le nombre de femmes de 40-44 ans, ayant un diplôme de niveau Bac, en emploi, a été multiplié par 1,09 par rapport à l'ancienne enquête.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Source : Insee, enquête Emploi.

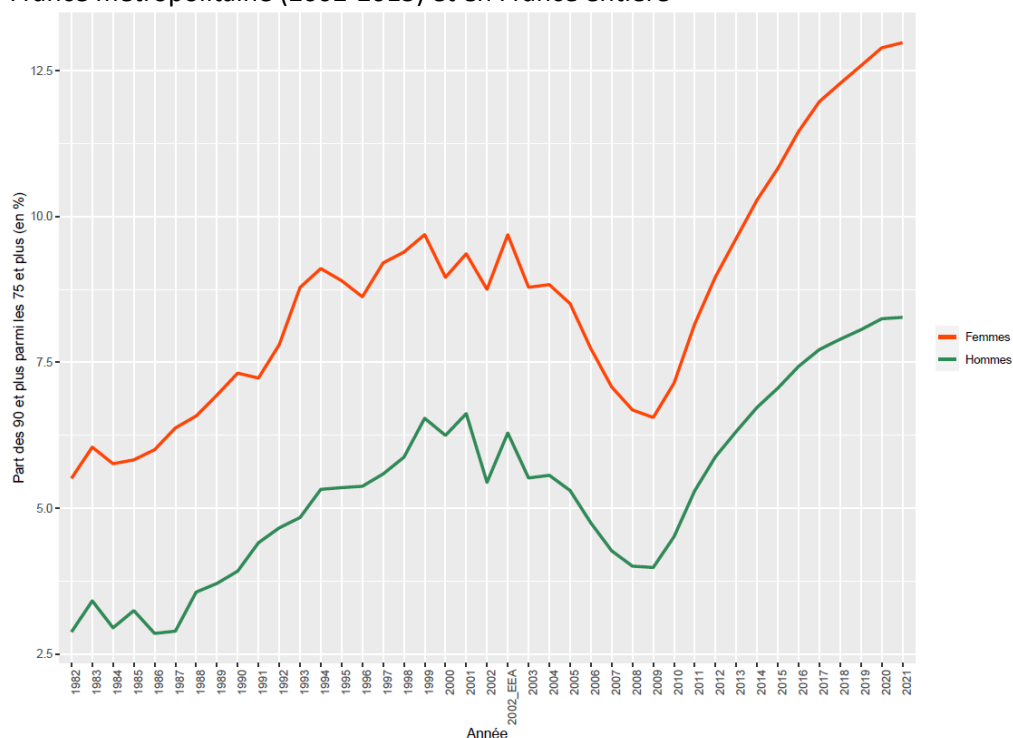
Les plus grosses ruptures concernent l'activité des 15-24 ans de niveau bac+2, difficiles à expliquer d'autant plus qu'elles sont de sens contraire entre les femmes et les hommes. Une grande partie résulte sans doute des petits effectifs dans le Pilote, d'où la décision de remplacer le coefficient multiplicatif par un coefficient additif dans certains cas : cela permet d'éviter d'introduire artificiellement de la volatilité sur les séries passées. Une forte révision à la baisse de l'emploi des hommes de 60 ans ou plus, à l'exception de ceux de niveau Bac+2, est également notable, alors qu'il est revu à la hausse pour les femmes de 60-69 ans.

Toujours pour les personnes de niveau Bac+2, le chômage des femmes de 50-69 ans est plus élevé en nouvelle enquête, alors que c'est l'inverse pour les hommes, et que femmes et hommes d'âge intermédiaire ont vu leur chômage rehaussé. Là encore, ces ruptures sont difficiles à expliquer, puisqu'au niveau du diplôme la seule modification consiste en l'intégration d'une liste déroulante de diplômes.

Les séries longues en ancienne norme par diplôme depuis 1982 sont dilatées grâce aux coefficients présentés ci-dessus (à l'exception de celles des 15-19 ans de diplôme Bac+3 ou plus pour ce sont des

coefficients additifs qui sont utilisés). Avant d'être recalée sur l'emploi, le chômage et l'inactivité par sexe et âge quinquennal rétopolés en nouvelle norme EEC3, l'inactivité des 75 ans ou plus doit être transformée en inactivité des 75-89 ans en retirant les personnes de 90 ans ou plus renseignées dans les marges de population. Ces dernières existent à partir de 2002 pour la France métropolitaine, et de 2014 pour les DOM. La Figure 25 montre la part des 90 ans et plus dans la classe d'âge des 75 ans et plus, calculée par sexe (à partir des marges France entière entre 2014 et 2020, des marges France métró entre 2002 et 2013, et des microdonnées micro de l'enquête Emploi annuelle (EEA) entre 1982 et 2002 (il y a donc deux sources pour l'année 2002 : la part calculée à partir de l'EEA et celle calculée à partir des marges de population en France métropolitaine)).

Figure 25 : Part des personnes de 90 ans et plus parmi celles de 75 ans et plus dans l'EEA (avant 2002), en France métropolitaine (2002-2013) et en France entière



Lecture : En 2016, 7,5% des hommes de 75 ans et plus sont âgés de 90 ans et plus.

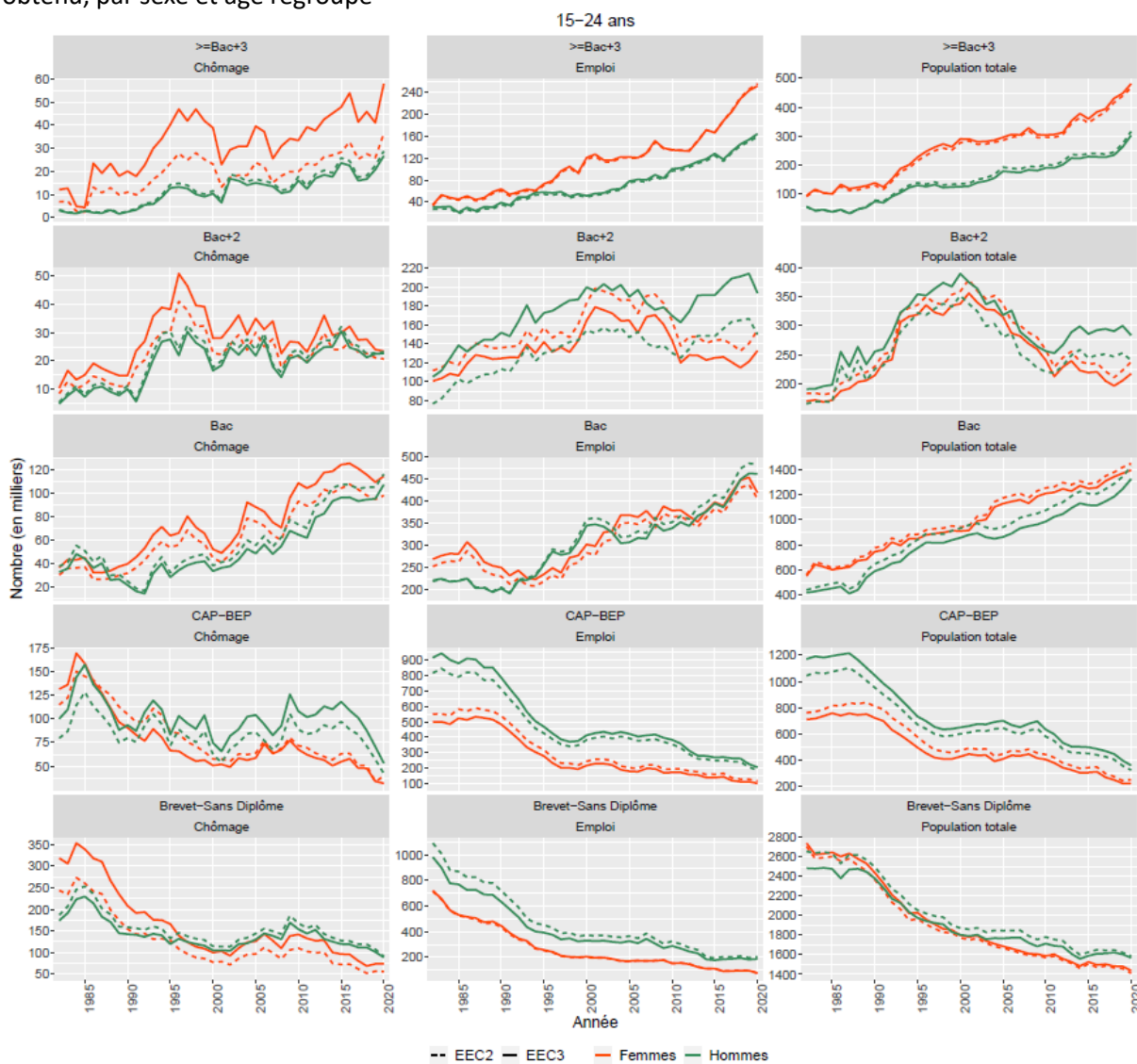
Champ : France métropolitaine jusqu'en 2013, France hors Mayotte à partir de 2014, personnes de 15 ans ou plus, vivant en logement ordinaire.

Source : Insee, enquêtes Emploi.

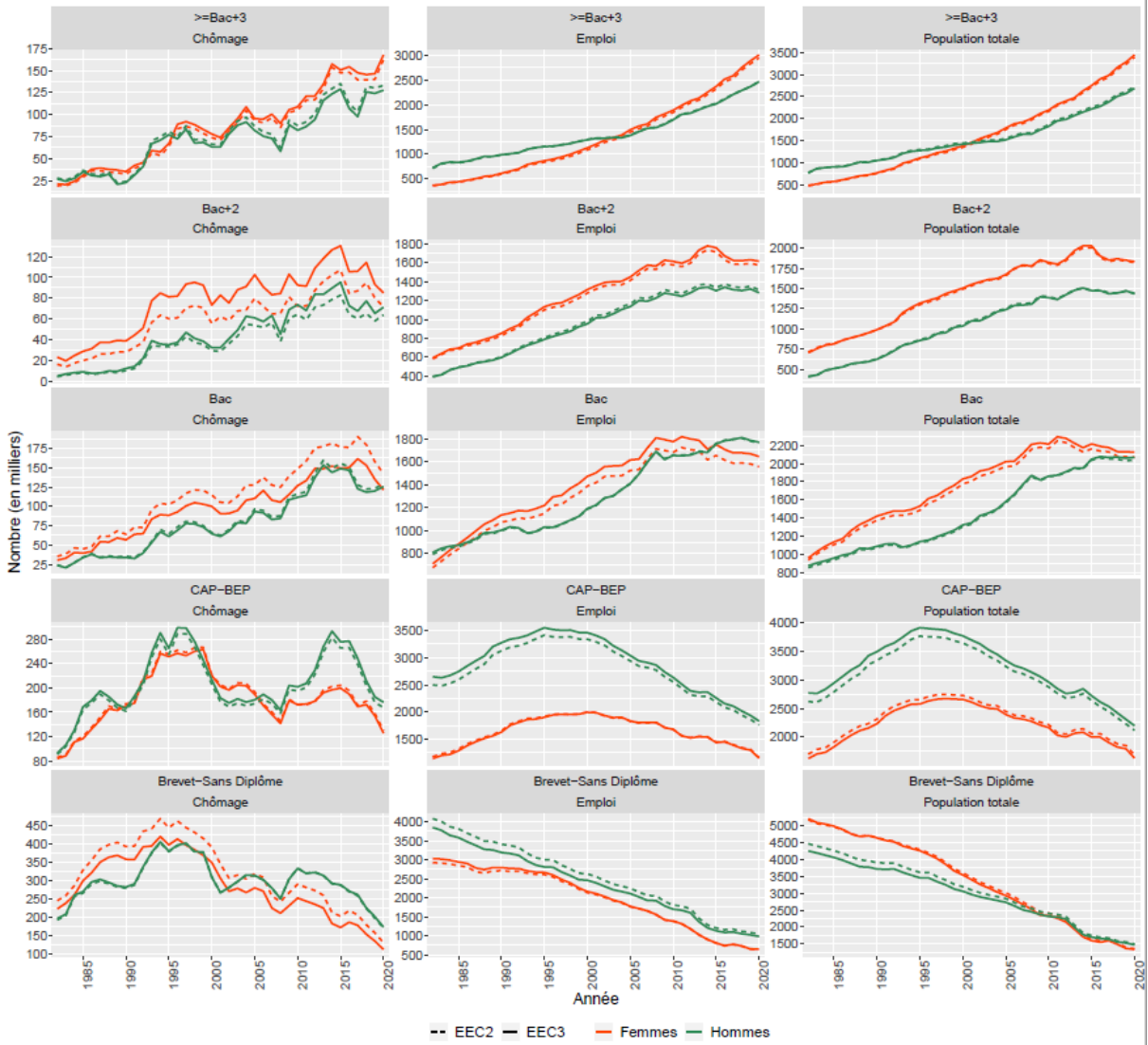
Si le passage de France métropolitaine à France entière pour le calcul de cette part en 2014 ne semble pas introduire de rupture de série, ce n'est pas le cas en 2002, lors du passage de l'EEA aux marges de population comme base de calcul : en choisissant l'une ou l'autre des deux mesures, une rupture se crée soit à droite, soit au gauche. C'est la solution la plus simple, bien qu'approximative, qui a été adoptée, à savoir lisser simplement la transition en choisissant pour 2002 la moyenne entre les deux types de mesures disponibles.

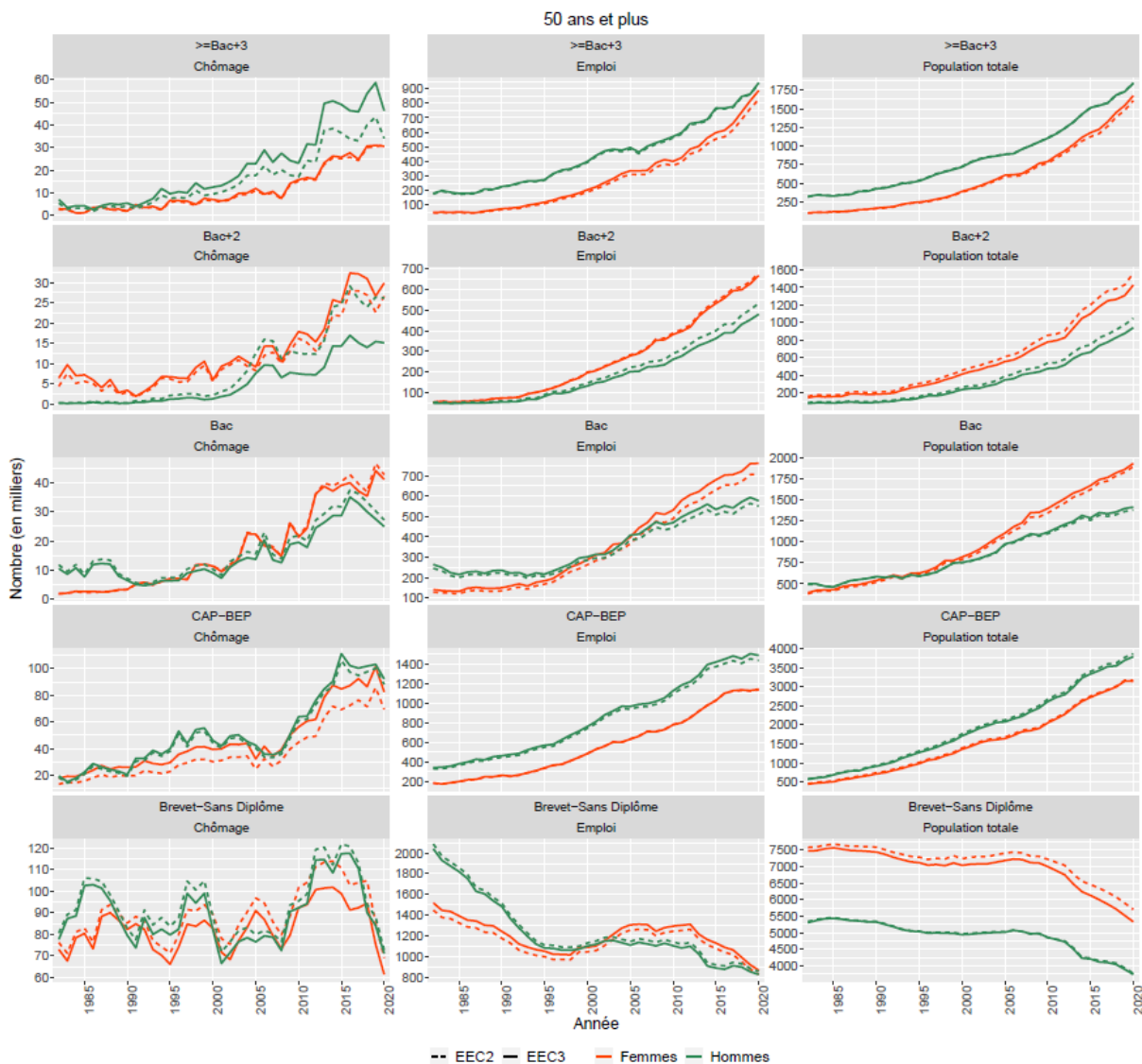
Une fois la part des 90 ans et plus dans les 75 ans et plus depuis 1982 reconstituée, le nombre de 90 ans et plus est retranché du nombre d'inactifs de 75 ans et plus rétopolé en nouvelle norme, et les séries dilatées annuelles par diplôme peuvent être recalées sur celles d'emploi, chômage et inactivité. Elles sont présentées en âge agrégé en trois classes dans la Figure 26.

Figure 26 : Séries longues de l'emploi, du chômage et de la population selon le plus haut niveau de diplôme obtenu, par sexe et âge regroupé



25-49 ans





Lecture : En 2020, en ancienne enquête, 142 000 femmes de 25-49 ans, ayant un diplôme de niveau Bac, sont au chômage, contre 121 000 en nouvelle enquête.

Champ : France hors Mayotte, personnes de 15 89 ans, vivant en logement ordinaire.

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

Ces séries annuelles sont ensuite utilisées pour le recalage lors de la rétopolation des séries trimestrielles. Ces dernières ne sont pas publiées par l'Insee, mais correspondent au pas demandé pour les données envoyées à Eurostat. La méthode générale de rétopolation est utilisée, avec simplement l'ajout de l'étape de suppression des 90 ans ou plus parmi les inactifs BIT chaque trimestre depuis 2003, et l'utilisation du coefficient additif et non multiplicatif pour les jeunes de 15-19 ans de niveau Bac+3 (autrement dit c'est la différence annuelle qui est utilisée).

I) Emploi et chômage selon la catégorie socioprofessionnelle (1^{ère} version : série 2022)

Le passage à la nouvelle enquête Emploi a été l'occasion d'introduire la nouvelle nomenclature des professions et catégories socioprofessionnelles (PCS), rénovée pour déboucher sur une nouvelle version, la « PCS 2020 ». Si les niveaux les plus détaillés de la PCS ont été modifiés pour mieux rendre compte des transformations des métiers, les niveaux plus agrégés, correspondant aux catégories et groupes socioprofessionnels, sont inchangés. Dans les séries longues du marché du travail jusque 2020, l'emploi et le

chômage sont déclinés selon la catégorie socioprofessionnelle (CS) par sexe et âge en 3 classes dans des séries longues débutant en 1982. La CS utilisée pour l'emploi est en 9 classes, correspondant au 1^{er} niveau de la nomenclature PCS, c'est-à-dire au 1^{er} chiffre du code PCS (cadres, ouvriers, etc.) avec toutefois une distinction qualifiés/peu qualifiés pour les employés et ouvriers, ainsi qu'une catégorie « autres » regroupant les professions non renseignées et les militaires du contingent), et en 8 classes pour le chômage (où la distinction qualifiés/peu qualifiés n'est faite que pour les ouvriers). L'actualisation de la nomenclature des PCS devrait donc *a priori* être neutre sur les séries longues d'emploi et de chômage.

En ancienne enquête le premier chiffre de la CS ainsi que le partage entre ouvriers qualifiés et peu qualifiés sont donnés directement (par la variable CSTOT), tandis que le partage entre employés qualifiés et peu qualifiés nécessite de connaître par ailleurs la profession (variable P). Pour les chômeurs, la variable CSR suffit au codage de la CS en 8 classes. En nouvelle enquête, une variable construite fournit toutes les informations directement, y compris le partage qualifiés/peu qualifiés pour les ouvriers comme pour les employés (variable PCS1Q).

Les coefficients de dilatation sont calculés avec la méthode générale sur la période de recouvrement entre l'ancienne et la nouvelle enquête, ils sont remplacés par des coefficients additifs pour les croisements suivants :

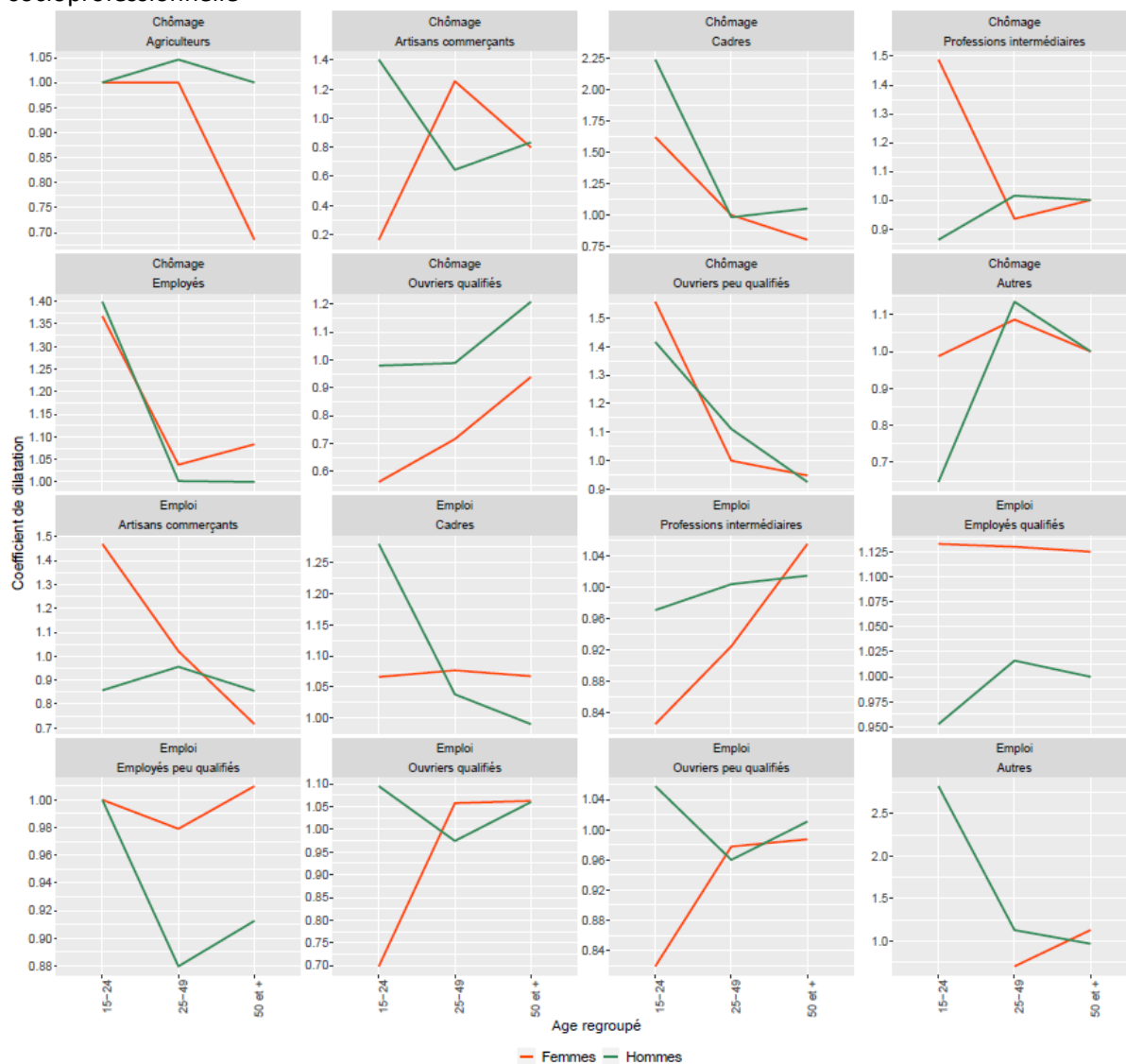
- L'emploi des agriculteurs quel que soit l'âge ou le sexe : les effectifs par sexe et tranche d'âge des agriculteurs sont faibles et le coefficient de dilatation ressort élevé, ce qui traduirait un aléa d'échantillonnage ; or les effectifs d'agriculteurs ont fortement baissé sur toute la période. L'utilisation du coefficient multiplicatif démultiplie alors l'aléa de 2020 dans le passé et aurait mécaniquement conduit à une part très élevée d'agriculteur en 1982, que rien ne justifie ;
- L'emploi des femmes de 15-24 ans dont la CS est inconnue : la CS des jeunes femmes est de manière inexplicable bien plus souvent manquante en nouvelle enquête qu'en ancienne, ce qui fait exploser leur nombre sur toute la période passée ; le scénario additif apparaît plus prudent.

D'autres coefficients de dilatation ont été fixés à 1, car ils conduisaient à une forte réévaluation de l'effectif remise en question par les dernières données datant de 2021, plus proches du niveau de l'EEC2 que du pilote ; c'est le cas de :

- l'emploi des employés peu qualifiés de 15-24 ans ;
- l'emploi des hommes de 50 ans et plus employés qualifiés ;
- le chômage des professions intermédiaires et des catégories « autres » de 50 ans et plus ;
- le chômage des hommes de 50 ans et plus employés et agriculteurs ;
- le chômage des femmes de 25-49 ans ouvrières peu qualifiées et agricultrices.

Les coefficients de dilatation obtenus (dans les cas multiplicatifs) sont présentés Figure 27.

Figure 27 : Coefficients de dilatation du nombre de personnes en emploi ou au chômage selon la catégorie socioprofessionnelle



Lecture : En nouvelle enquête, sur la période de recouvrement, le nombre d'hommes de 25-49 ans, employés peu qualifiés, en emploi, a diminué par rapport à l'ancienne enquête, d'un facteur multiplicatif de 0,88.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Source : Insee, enquête Emploi.

Les coefficients de dilatation très élevés ou très faibles sur le chômage des agriculteurs, des jeunes et des femmes artisans-commerçants, des jeunes cadres, des jeunes femmes ouvrières qualifiées ainsi que de la catégorie « autres » sont difficilement interprétables, dans la mesure où ils portent sur de très petits effectifs et relèvent en grande partie de l'aléa statistique. C'est aussi le cas de l'emploi des jeunes artisans-commerçants, ainsi que des jeunes femmes ouvrières qualifiées.

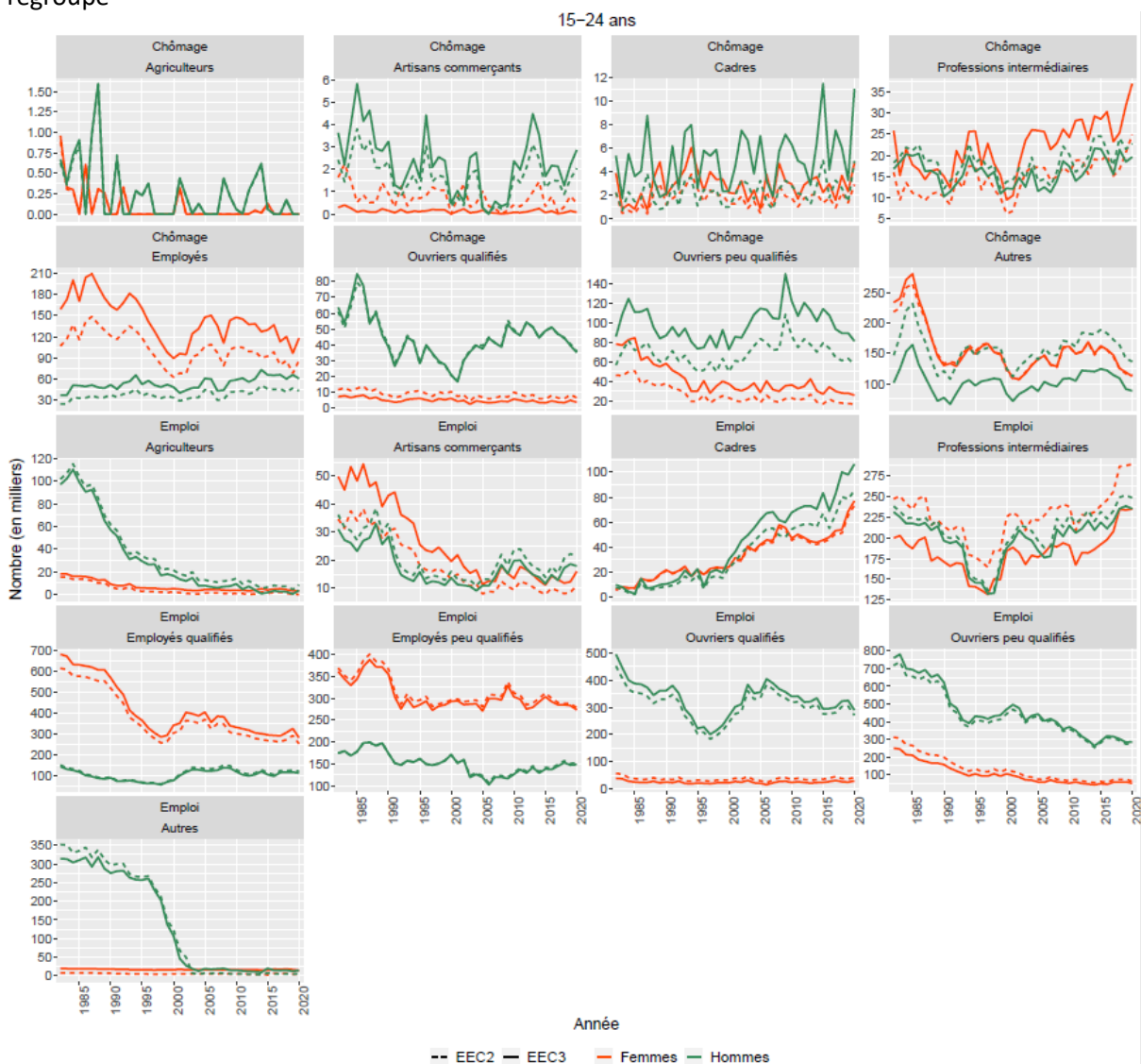
Le chômage des hommes de 25 ans et plus artisans commerçants est revu à la baisse, alors qu'il est quasiment stable chez les cadres, professions intermédiaires et employés. Pour les ouvriers qualifiés, il augmente chez les plus âgés, alors que pour les ouvriers peu qualifiés c'est le cas pour les plus jeunes. Chez les femmes, le chômage des cadres est plus faible pour les 50 ans et plus, il diminue également pour les ouvrières qualifiées, alors qu'il augmente beaucoup pour les jeunes femmes de profession intermédiaire, employées et ouvrières peu qualifiées. Ces modifications reflètent notamment la forte rupture à la hausse sur le chômage des jeunes femmes peu diplômées.

Sur l'emploi, il diminue par symétrie avec le chômage chez les jeunes femmes de profession intermédiaire et ouvrières peu qualifiées, en revanche l'emploi des femmes cadres et employées qualifiées augmente. Chez les hommes, il est revu à la baisse pour les artisans-commerçants et employés peu qualifiés, il varie très peu dans les autres professions.

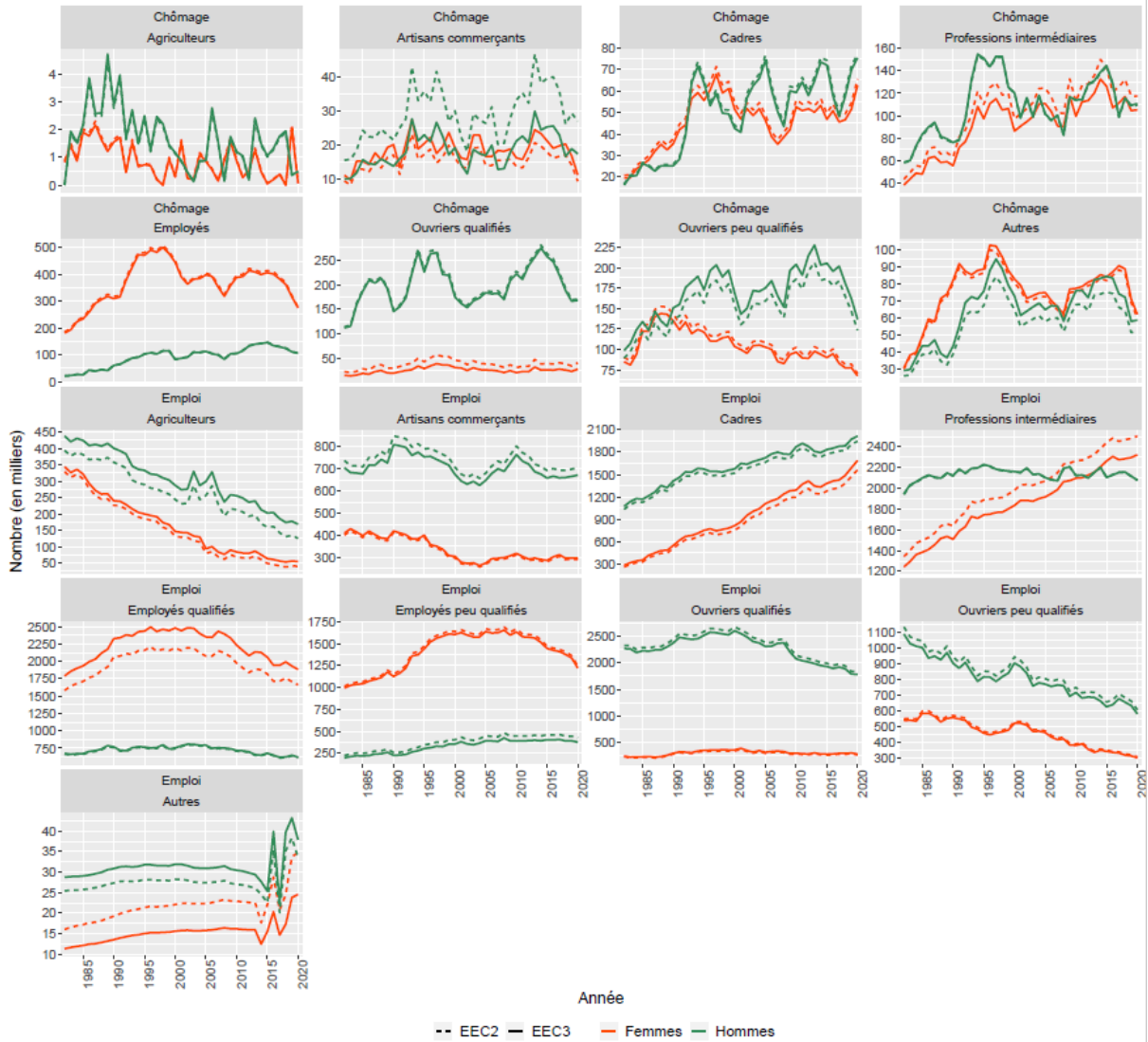
Après dilatation (ou translation dans les cas additifs), les séries longues sont recalées sur l'emploi et le chômage EEC3, mais la méthode de calage est légèrement modifiée pour l'emploi des hommes de 15-24 ans. En effet, leur coefficient de dilatation dans la catégorie « autres » est très élevé, depuis 2003 il s'applique à de petits effectifs donc les impacts sont mineurs, mais avant 2002 l'effectif « autres » incluant les militaires du contingent, il était beaucoup plus élevé et explose donc après dilatation. De ce fait, sur 1982-2002, le solde entre l'emploi EEC3 des hommes de 15-24 ans et la somme de leur emploi par CS est reporté entièrement sur la série des « autres », au lieu d'être réparti sur toutes les CS comme ce qui est fait d'ordinaire. A partir de 2003, la méthode générale de calage est également appliquée aux hommes de 15-24 ans.

Les séries annuelles rétrolées qui en résultent sont présentées Figure 28.

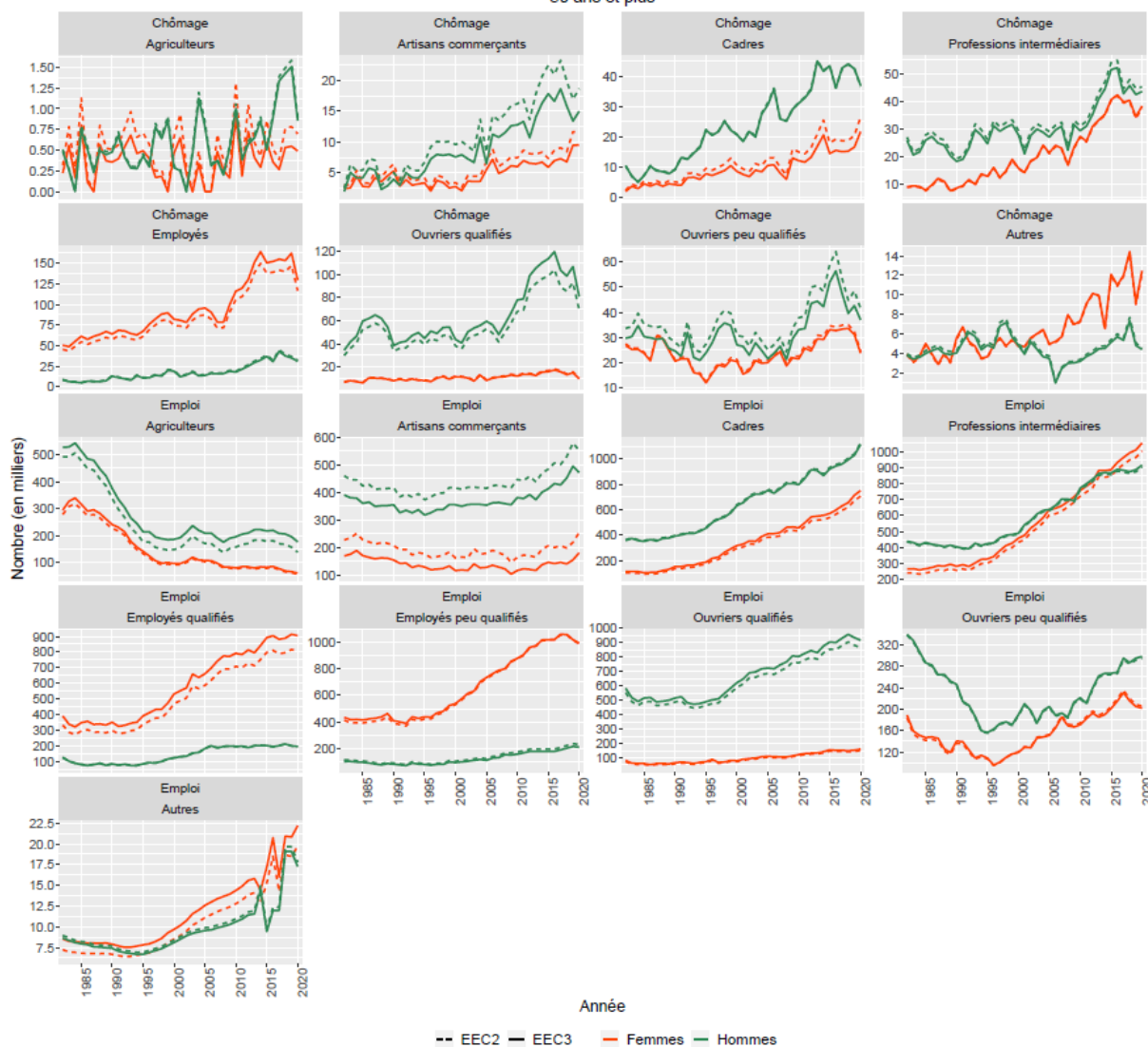
Figure 28 : Séries longues de l'emploi et du chômage selon la catégorie socioprofessionnelle, par sexe et âge regroupé



25-49 ans



50 ans et plus



Lecture : En 2020, en ancienne enquête, 1 661 000 femmes de 25-49 ans, employées qualifiées, sont en emploi, contre 1 884 000 en nouvelle enquête.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

Annexe 2 : Séries longues particulières

a) Ruptures et rétopolation de l'activité dans les DOM

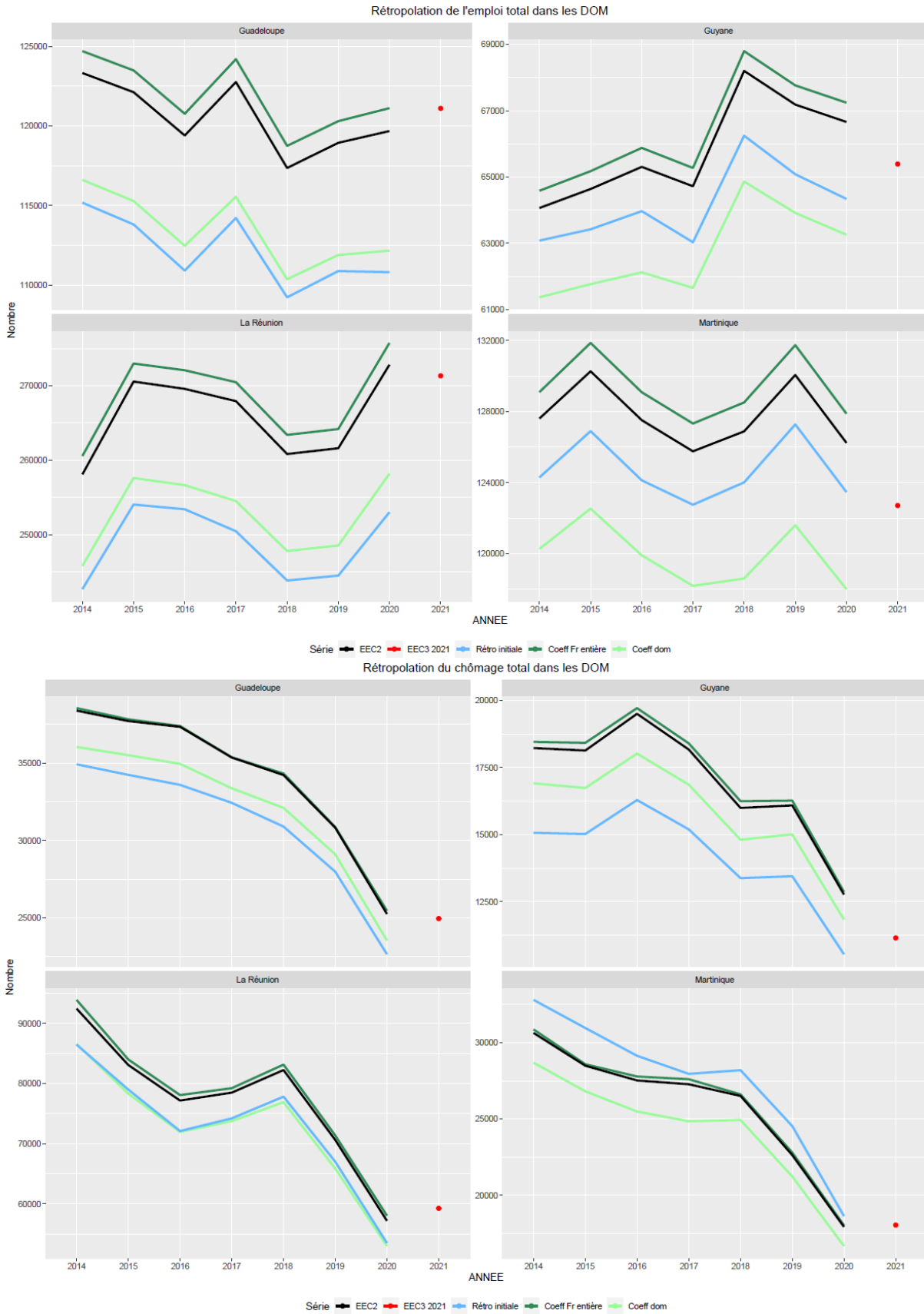
Le taux de sondage dans les Dom est beaucoup plus élevé qu'en France métropolitaine, ce qui permet une représentativité de l'EEC pour chacun des 4 DOM (hors Mayotte) et donc le calcul d'indicateurs à leur échelle, ce qui n'est pas possible pour chaque département métropolitain. C'est le cas des séries d'emploi, de halo et de chômage, qui sont calculées par sexe et âge regroupé en 3 classes.

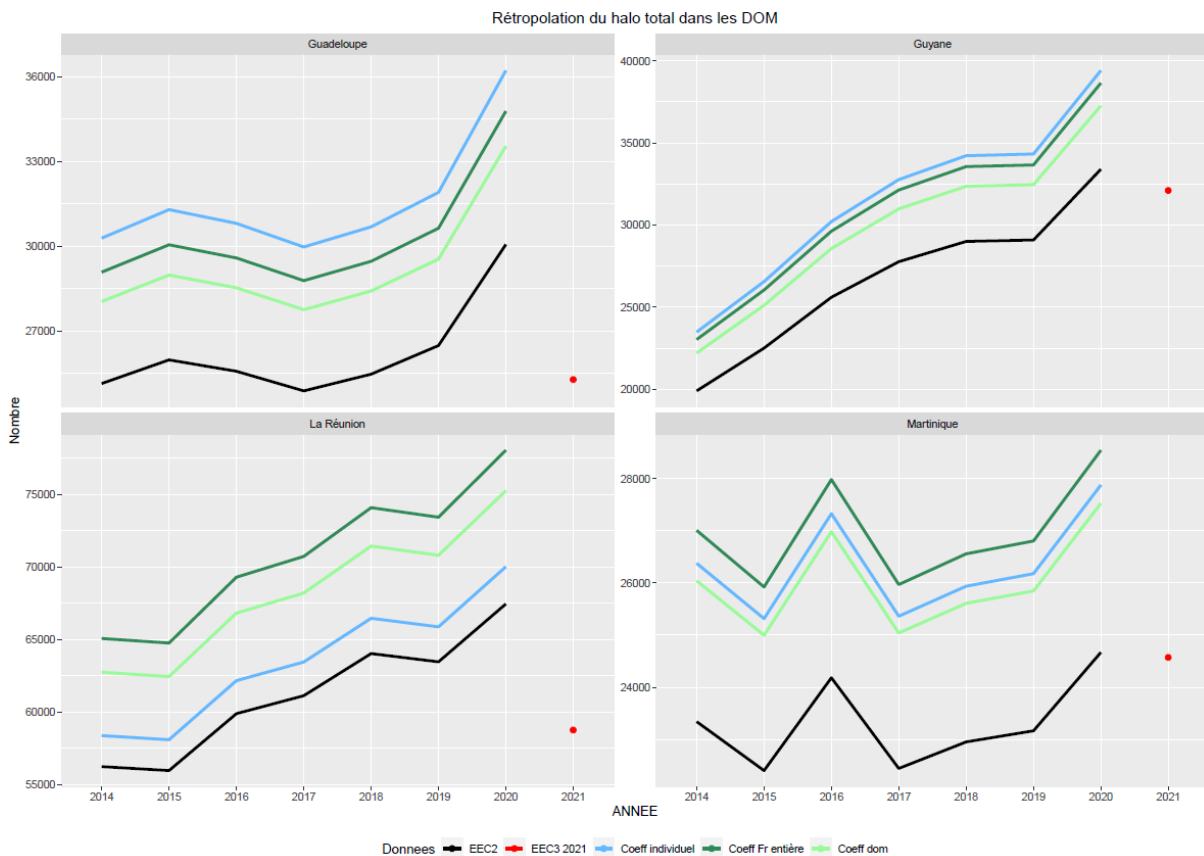
Ces séries démarrent en 2014 ; elles doivent donc être rétopolées à part, d'autant plus que les ruptures sont susceptibles d'y être différentes de celles constatées au niveau France entière : en effet, contrairement à la métropole, il n'y a pas eu de refonte de la pondération dans les DOM, et l'effet de mode consécutif à l'introduction d'Internet comme mode de réponse supplémentaire en ré-interrogation est sans doute beaucoup plus faible puisque le taux de réponse par Internet y est bien plus faible.

Une première rétopolation suivant la méthode générale a pu être réalisée en calculant un coefficient de dilatation pour chacun des 4 DOM et pour la France métropolitaine à partir de la comparaison entre ancienne et nouvelle enquête sur les 5 trimestres de recouvrement. Pour chaque indicateur, les 5 séries étaient ensuite recalées sur celle de la France entière. Néanmoins, les ruptures mesurées étaient très variables entre les DOM et selon les croisements de sexe et âge, sans qu'elle ne puissent être expliquées par les motifs de ruptures connus avancés et mesurés au niveau national : changements de concepts, effet questionnaire, effet Internet, etc. Cette fragilité était sans doute liée au fait que l'exploitation des données du Pilote Dom par Dom (et même pour l'ensemble des Dom) pouvait s'avérer très fragile, du fait de la taille des échantillons pris en compte.

Au printemps 2022, les données collectées en nouvelle enquête en 2021 ont pu être confrontées aux séries par DOM en ancienne enquête, à celles rétopolées selon la méthode générale, ainsi qu'à deux méthodes alternatives de rétopolation, l'une consistant à calculer un coefficient de dilatation commun aux quatre DOM (qui ne permet pas de prendre en compte les spécificités de chaque DOM mais les distingue de la France entière tout en augmentant la taille des échantillons du Pilote pour calculer chaque coefficient), et l'autre appliquant des coefficients de dilatation calculés par sexe et âge regroupé en 3 classes sur la France entière. La Figure 29 résume les résultats obtenus après réagrégation de chaque série sur l'ensemble des sexes et âges.

Figure 29 : Comparaison des différentes méthodes de rétropolation de l'emploi, du chômage et du halo dans les DOM avec les séries en ancienne enquête et le niveau 2021 en nouvelle enquête.





Lecture : En 2021, à la Réunion, le niveau de l'emploi s'élève à 271 300. En 2020, c'est le niveau en ancienne enquête qui s'en approche le plus (272 800), il est surévalué par la méthode de réropolation consistant à appliquer le coefficient de dilatation calculé sur la France entière (275 800), et sous-évalué par la méthode générale de réropolation (253 000) et celle consistant à appliquer un coefficient de dilatation commun à tous les DOM (258 200).

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Sources : Insee, enquête Emploi.

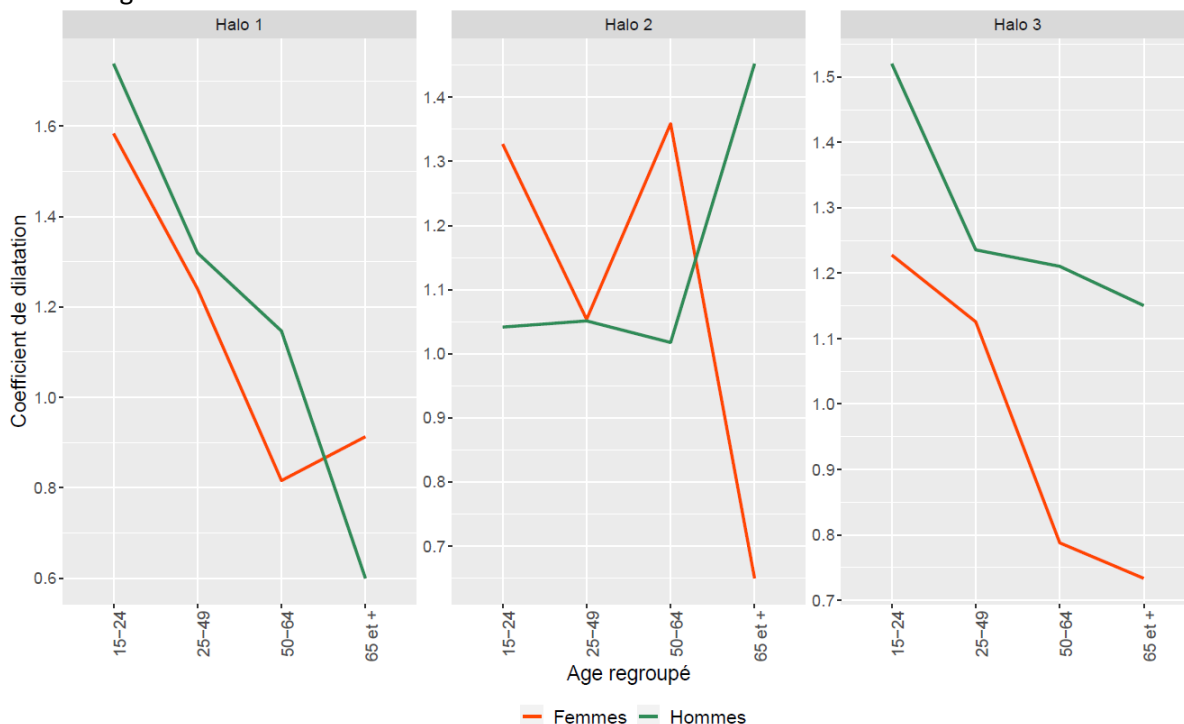
Il ressort de cette comparaison que, quelle que soit la méthode de réropolation retenue, les indicateurs non réropolés sont souvent plus cohérents avec les observations 2021, en particulier sur l'emploi et le halo. La mobilisation de données administratives sur l'emploi et les demandeurs d'emploi n'a pas non plus permis de mieux quantifier les ruptures dans chaque DOM.

Il a donc été décidé de ne pas réropoler l'emploi, le chômage et le halo dans les DOM, et d'estimer les nouvelles séries longues en se basant sur les indicateurs en ancienne mesure, recalés par tranche d'âge et sexe sur les nouvelles marges démographiques. Autrement dit, même en considérant les DOM comme un tout, le dimensionnement du Pilote s'est avéré trop faible pour apporter une information utile pour estimer une potentielle rupture de série spécifique aux DOM.

b) Création des séries de composantes du halo par sexe et âge regroupé

Pour compléter les séries de halo par composantes (tous âges et sexes confondus) et de halo total par sexe et âge en 3 classes (4 classes pour les séries trimestrielles), des séries croisant les composantes du halo par sexe et l'âge en 4 classes agrégées, débutant en 2014, ont été construites, en repartant des données individuelles et des informations dans l'ancienne enquête, puis en utilisant le Pilote pour calculer les coefficients de dilatation présentés en Figure 30.

Figure 30 : Coefficients de dilatation du nombre de personnes dans chaque composante du halo selon le sexe et l'âge



Lecture : En nouvelle enquête, sur la période de recouvrement, le nombre d'hommes de 15-24 ans, dans la première composante du halo a été multiplié par 1,74 par rapport à l'ancienne enquête.

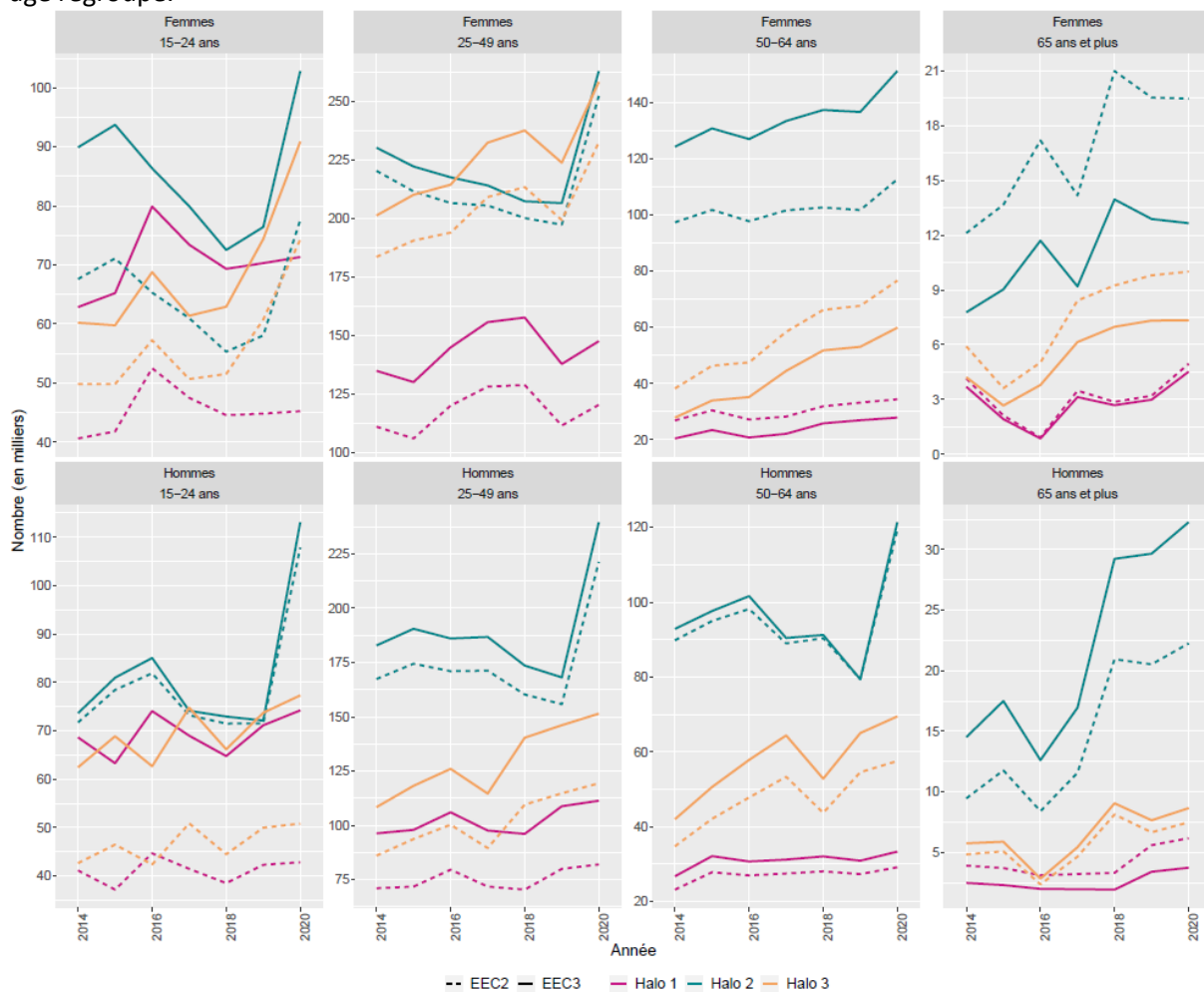
Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Source : Insee, enquête Emploi.

Les coefficients de dilatation élevés pour les 3 composantes chez les 15-49 ans découlent d'abord de la rupture à la hausse du halo total dans ces classes d'âge. Elle concerne surtout la première composante du halo, conséquence de la baisse du recours au proxy et du changement de questionnaire. Chez les 50-64 ans, seule la disponibilité pour travailler semble augmenter (halo 2) tandis que chez les plus de 65 ans, les halos 2 et 3 augmentent chez les femmes uniquement, mais les effectifs concernés sont très faibles et volatils.

Ces coefficients de dilatation sont appliqués aux séries trimestrielles en ancienne norme. Un calage sur marges est ensuite réalisé pour la rendre cohérente avec les séries de halo par composante et de halo par sexe et âge préalablement rétopolées en nouvelle norme : chaque trimestre depuis 2014 est recalé sur les marges des composantes du halo et du halo par sexe et âge en 4 classes. La série trimestrielle ainsi obtenue, en nouvelle norme, est transformée en série annuelle via une moyenne annuelle, représentée Figure 31.

Figure 31 : Séries longues du nombre de personnes inactives dans chaque composante du halo, par sexe et âge regroupé.



Lecture : En 2020, en ancienne enquête, 82 000 hommes de 25-49 ans sont dans la première composante du halo, contre 111 000 en nouvelle enquête.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

c) Chômage selon le plus haut niveau de diplôme obtenu (2^e version : série 2023)

En avril 2023, à l'occasion de l'actualisation des poids individuels sur 2018-2022 suite à la mise à jour des marges démographiques sur lesquelles l'enquête Emploi est calée, il a été constaté que les séries de chômage et taux de chômage par plus haut niveau de diplôme en découlant, construites à partir de la série rétrolée en 2022, présentaient alors une rupture de série entre 2017 et 2018. En effet, le nombre de chômeurs par diplôme n'étant pas inclus dans les marges de calage des poids individuels rétrolés, série longue rétrolée et série longue construite à partir des poids individuels rétrolés ne coïncident pas nécessairement.

Il y avait donc un arbitrage pour ces séries entre :

- d'une part la rétrolation précédente, basée sur une exploitation agrégée du Pilote de l'enquête Emploi (en 2020 et au premier trimestre 2021)

- d'autre part une rétopolation basée sur les pondérations individuelles redressées (pour tenir compte des ruptures de série sur les variables plus agrégées) et recalées sur les nouvelles marges de population sur les 4 derniers millésimes (comme habituellement)

Le premier jeu de pondération a l'avantage d'une méthode standardisée, en donnant le primat de la rétopolation au Pilote comme pour les autres variables; il a toutefois l'inconvénient de ne pas tenir compte des nouvelles marges de population d'une part ; de mobiliser le Pilote a des niveaux parfois très fins (diplôme*sexe*tranche quinquennale d'âge), potentiellement trop fins au regard de la taille de ce Pilote (cf exemple de la rétopolation des DOM pour lesquels le Pilote s'est avéré peu utilisable).

A noter enfin que cet arbitrage aurait pu se poser l'année précédente, même sans nouveau jeu de marges de population. Par ailleurs cet arbitrage ne se pose pas pour les variables utilisées dans les marges de calage des poids individuels rétopolés, les deux rétopolations étant par construction convergentes.

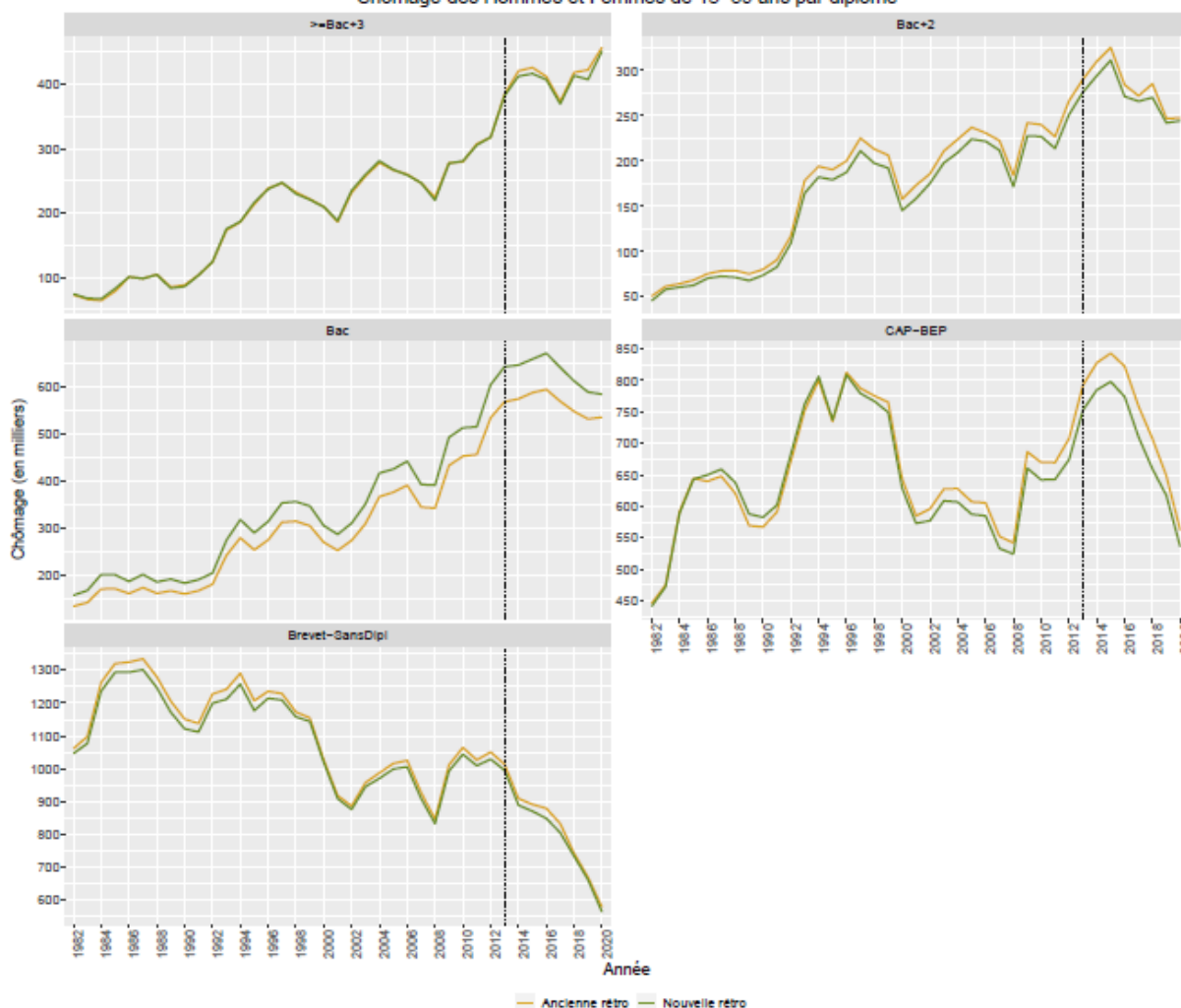
En outre, en revenant aux données du Pilote EEC, il a été constaté que la rupture de série entre ancienne et nouvelle enquête en 2021 relativement à ces indicateurs était relativement fragile et pas homogène selon les cinq trimestres de recouvrements des deux enquêtes (ancienne et nouvelle EEC).

Aussi, il a été décidé de revoir la méthode de rétopolation de la série longue du nombre de chômeurs par plus haut niveau de diplôme :

- Sur 2014-2020, la nouvelle série longue du nombre de chômeurs par diplôme, sexe, et âge quinquennal est construite en tabulant dans les données individuelles de l'enquête Emploi en utilisant les poids individuels rétopolés. Cela permet de corriger les principales ruptures de série (population, statut BIT) sans faire d'hypothèse *a priori* sur la rupture de série du taux de chômage par diplôme, compte-tenu de la difficulté à mesurer de façon nette la rupture de série (cf infra) ;
- En 2014, un coefficient multiplicatif permettant de passer du nombre de chômeurs par diplôme de la série longue rétopolée en 2022 à celui obtenu en tabulant les poids individuels rétopolés est calculé pour chaque croisement de sexe, âge quinquennal et plus haut niveau de diplôme. Avant d'être appliqués aux séries rétopolées construites en 2022, ces coefficients multiplicatifs sont préalablement lissés par moyennes mobiles pour les 20-64 ans, comme cela été fait dans la méthode générale de rétopolation. Le nombre de chômeurs qui en découle est recalé pour coïncider avec le nombre total de chômeurs par sexe et âge quinquennal de la série longue BIT, en faisant porter l'ajustement sur toutes les sous-séries par diplôme proportionnellement à l'effectif de chaque croisement, comme cela avait été fait dans la méthode générale de rétopolation des séries longues.

La Figure 32 résume les écarts du nombre total de chômeurs par plus haut niveau de diplôme entre les deux méthodes de rétopolation.

Figure 32 : Comparaison des séries longues du nombre de chômeurs selon le plus haut niveau de diplôme, obtenues lors de la première méthode de rétropolation en 2022 et suite à sa correction en 2023.
Chômage des Hommes et Femmes de 15-89 ans par diplôme



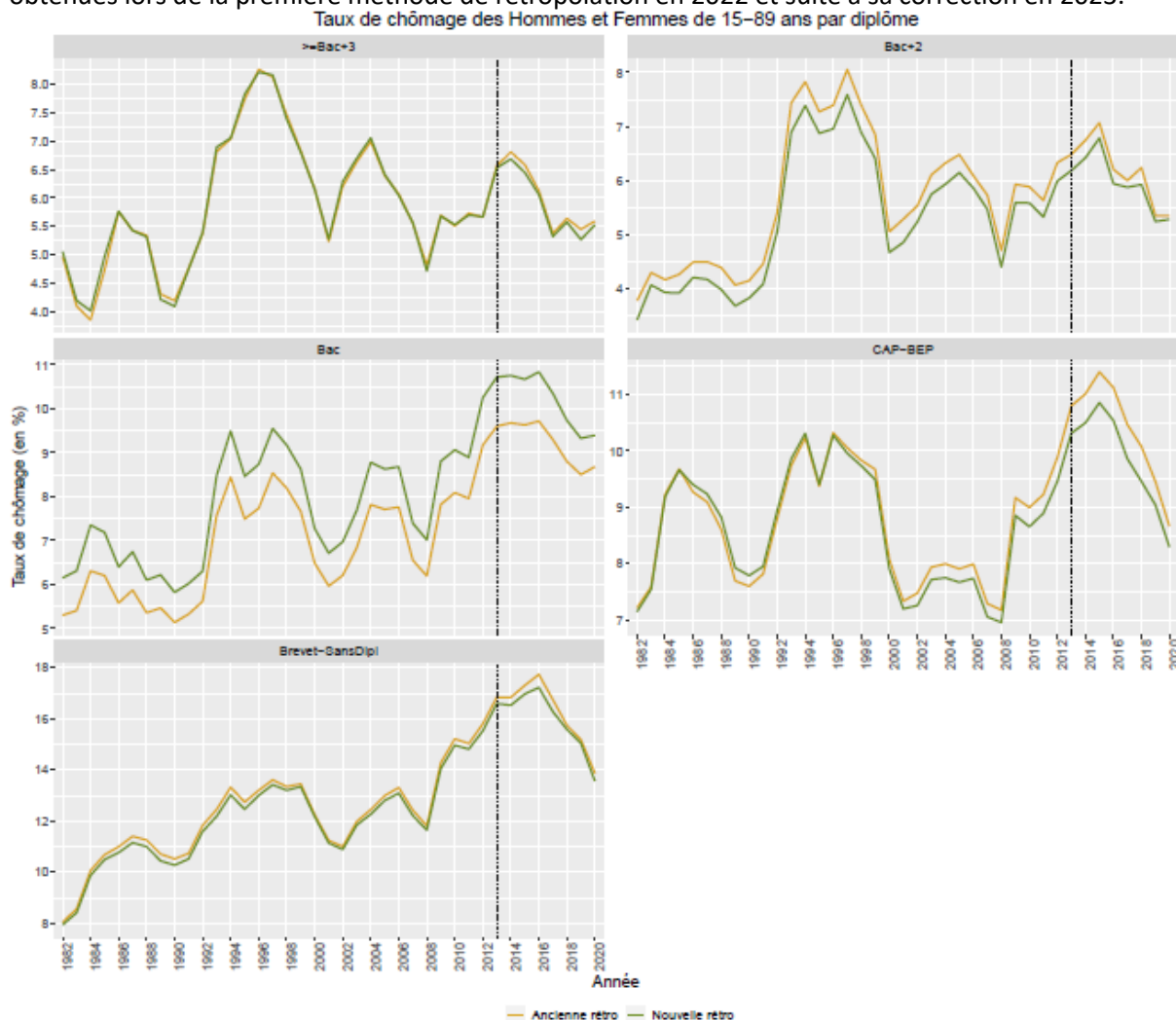
Lecture : En 2020, dans la nouvelle version de la série rétroplée, le nombre de chômeurs de 15-89 ans dont le plus haut diplôme est de niveau Bac est de 584 000, alors qu'il était de 535 000 dans la première version de la série rétroplée.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

La nouvelle série ainsi rétroplée du nombre de chômeurs par plus haut niveau de diplôme est utilisée pour calculer celle de taux de chômage par diplôme (Figure 33), à l'aide également de la série d'emploi par plus haut niveau de diplôme rétroplée en 2022. Dans la mesure où cette dernière ne fait pas non plus partie des marges de calage utilisées pour rétropléer les poids individuels, cette dernière série de taux de chômage continue à différer légèrement des taux de chômage par diplôme calculés en tabulant les données individuelles sur 2014-2020. L'écart est néanmoins très faible.

Figure 33 : Comparaison des séries longues du taux de chômage selon le plus haut niveau de diplôme, obtenues lors de la première méthode de réropolation en 2022 et suite à sa correction en 2023.



Lecture : En 2020, dans la nouvelle version de la série réropolée, le taux de chômage des 15-89 ans dont le plus haut diplôme est de niveau Bac est de 9,4%, alors qu'il était de 8,7% dans la première version de la série réropolée.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

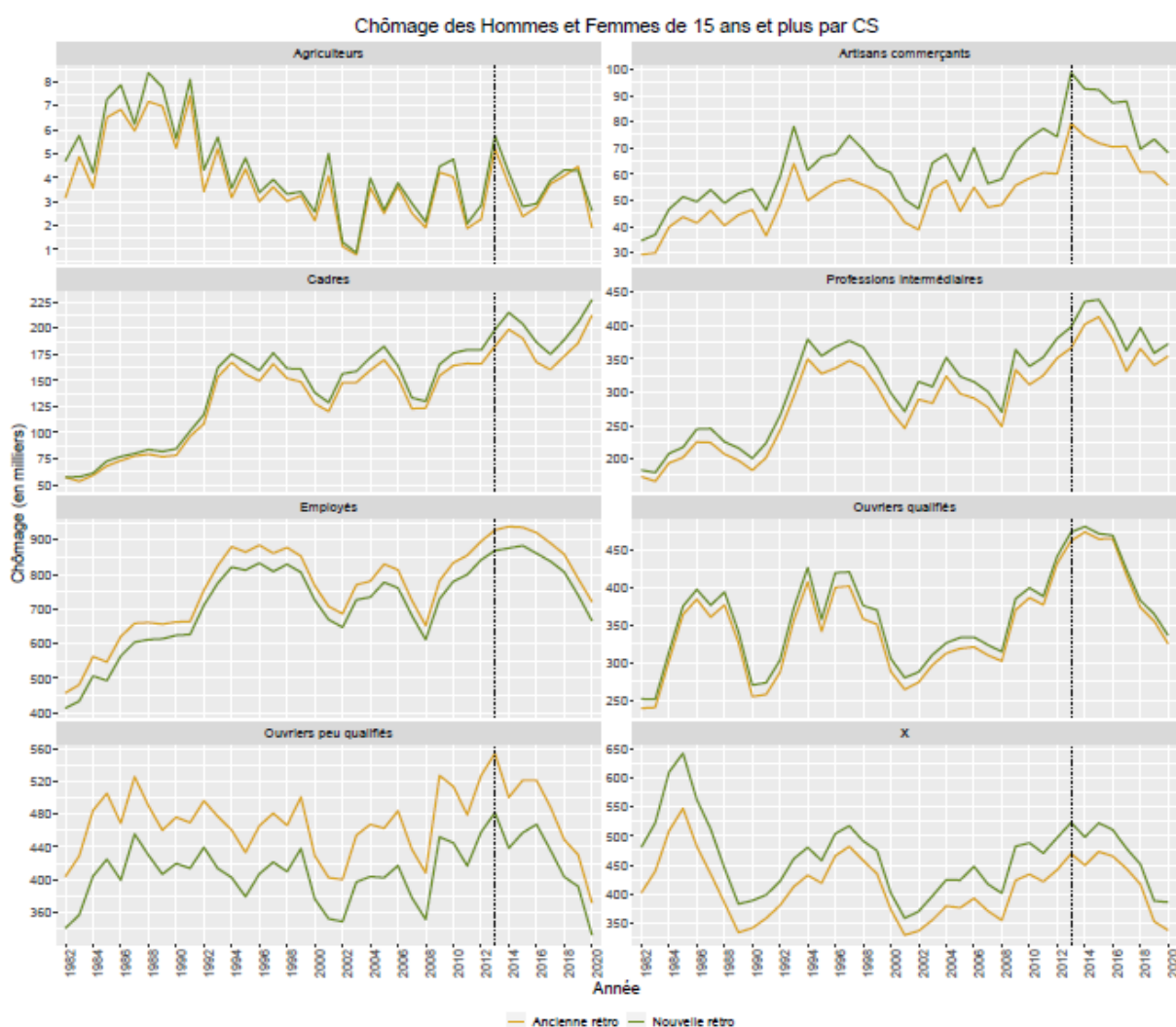
Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

d) Chômage selon la catégorie socioprofessionnelle (2^e version : série 2023)

Comme pour la série de chômage par plus haut niveau de diplôme, le chômage par CS n'est pas inclus dans les marges de calage des poids individuels réropolés, et une rupture de série est apparue en 2023 lors de leur actualisation. La même méthode que celle décrite ci-avant a donc été adoptée pour revoir cette série longue,

La Figure 34 résume les écarts du nombre total de chômeurs par CS entre les deux méthodes de réropolation.

Figure 34 : Comparaison des séries longues du nombre de chômeurs selon la catégorie socioprofessionnelle obtenues lors de la première méthode de rétropolation en 2022 et suite à sa correction en 2023.



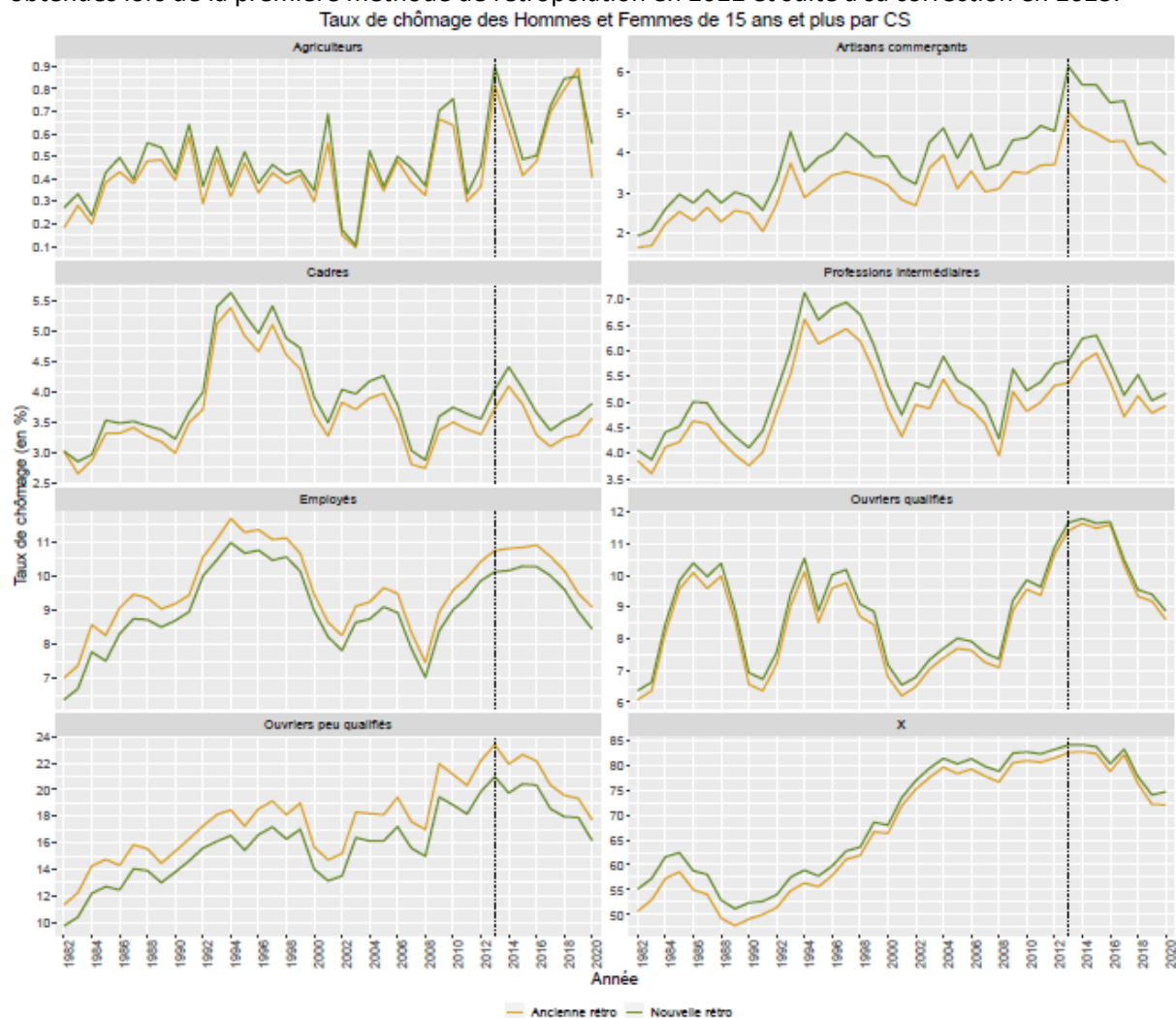
Lecture : En 2020, dans la nouvelle version de la série rétropolée, le nombre de cadres de 15 ans ou plus au chômage est de 227 000, alors qu'il était de 212 000 dans la première version de la série rétropolée.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.

La nouvelle série rétropolée du taux de chômage par CS découle alors directement de cette nouvelle version ainsi que de la série longue rétropolée de l'emploi par CS (Figure 35).

Figure 35 : Comparaison des séries longues de taux de chômage selon la catégorie socioprofessionnelle obtenues lors de la première méthode de rétopolation en 2022 et suite à sa correction en 2023.



Lecture : En 2020, dans la nouvelle version de la série rétopolée, le taux de chômage des cadres de 15 ans ou plus est de 3,8%, alors qu'il était de 3,6% dans la première version de la série rétopolée.

Champ : France hors Mayotte, personnes de 15 ans ou plus, vivant en logement ordinaire.

Sources : Insee, enquête Emploi, séries longues sur le marché du travail.