

ESTIMATIONS LOCALES DU NOMBRE DE DESCENDANTS D'IMMIGRES

Yves Breem¹ & Yannick Croguennec²

¹ *Département des Statistiques, des Etudes et de la Documentation, Secrétariat Général à l'Immigration et à l'Intégration, Ministère de l'Intérieur, Place Beauveau 75800 Paris cedex 08. yves.breem@immigration-integration.gouv.fr*

² *Direction de l'Animation, de la Recherche, des Etudes et des Statistiques, Ministère du Travail, de l'Emploi et de la Santé, 39-43 quai André Citroën 75902 Paris cedex 15. yannick.croguennec@travail.gouv.fr*

Selon les dernières estimations réalisées à l'aide de l'Enquête Emploi (EEC) et de l'enquête "Trajectoires et Origines", la part des descendants d'immigrés dans la population est de l'ordre de 10 % (Breem¹, 2010 ; Borrel, 2010). Il existe d'autres sources importantes d'information sur la population ayant un lien à la migration parmi lesquelles figure le Recensement de la Population (RP). Mais ce-dernier, a contrario des enquêtes ménages, ne permet pas de comptabiliser les descendants d'immigrés. Or, il est la seule source dont la taille de l'échantillon est adaptée à une estimation locale de population. L'objectif « avoué » de notre étude sera de déterminer à l'aide des sources disponibles, une solution méthodologique permettant d'estimer la population des descendants d'immigrés par zone infra nationale.

I - Définitions

Pour comprendre le concept de descendants d'immigrés, il est nécessaire de rappeler la définition française de l'immigré. D'après la définition officialisée par le Haut Conseil à l'Intégration en 1991, un immigré est une personne née étrangère à l'étranger. Cette définition est propre à la France : la plupart des autres pays, ainsi que des organisations internationales (OCDE, Commission européenne) lui préfère la notion de personne née à l'étranger. La notion de descendant d'immigré est une définition plus officieuse, mais elle a fait l'objet d'un quasi-consensus au sein de la statistique publique, qui cherchait à mieux connaître la situation des enfants d'immigrés à la suite des nombreuses demandes des pouvoirs publics. La définition couramment utilisé est donc la suivante : est descendant d'immigré toute personne née en France d'au moins un parent immigré.

Les enquêtes ménages contiennent depuis peu les questions nécessaires à l'identification des descendants d'immigrés : le pays de naissance et la nationalité à la naissance des deux parents. Dans l'Enquête Emploi, ces questions ont été introduites en 2005.

Rares sont les études qui citent le nombre de descendants d'immigrés, et les comparaisons entre sources sont particulièrement délicates. Ceci pour plusieurs raisons :

- tout d'abord, les quelques estimations déjà réalisées sont souvent assez anciennes et les décalages de date rendent complexe toute comparaison, le nombre de descendants étant croissant dans le temps (si l'on suit les tendances récentes d'évolution des flux migratoires et de la fécondité) ;
- en second lieu, les tailles des échantillons sont également très différentes d'une source à l'autre, sachant qu'un échantillon de trop petite taille ne pourra être exploité ;

¹ « Les descendants d'immigrés », *Infos Migrations n°15*, juillet 2010

- enfin, les enquêtes portent rarement sur des champs semblables ou comparables. Ainsi, l'âge de la population interrogée diffère d'une enquête à l'autre.

Le tableau suivant présente un récapitulatif des estimations « récentes » des descendants d'immigrés à travers les sources.

<i>Nom</i>	<i>Auteur</i>	<i>Date</i>	<i>Age</i>	<i>Esti- mation</i>	<i>Particularité</i>
EHF: Histoire de Famille	Borrel, Simon	1999	Tous	4.5 M	Estimation ancienne
FQP: Formation et Qualification Professionnelle		2003	17 – 65	2.6 M	Ancien et petit échantillon
TeO: Trajectoires et Origines	Beauchemin, Hamel, Simon et al.	2008	18 – 50	3.1 M	Enquête spécifique sur les immigrés et les descendants
EEC: Enquête Emploi en Continu	Breem	2010	Tous	6.4 M	

Le nombre de descendants d'immigrés d'après l'EEC 2010 est bien supérieur à celui de toutes les enquêtes ménages faites avant 2005. Cette différence s'explique surtout par les facteurs de date et de taille d'échantillon. Etant donné les 10 ans de différence avec EHF par exemple, il est probable que le chiffre de 4,5 millions en 1999 soit cohérent avec celui de 2010. Autre exemple, l'enquête FQP ayant un échantillon près de 9 fois plus faible que l'EEC, le chiffre de l'EEC est plus vraisemblable.

Toutes les estimations précédemment citées sont basées sur des enquêtes de population générale. L'enquête TeO (Trajectoires et Origines, INED-INSEE, 2008), est la seule enquête récente à ce jour permettant entre autres d'interroger spécifiquement la population des descendants. A champ constant (18 – 50 ans), l'estimation EEC est cette fois plus faible que l'estimation TeO (2,7 millions contre 3,1). Le fait que l'enquête TeO contienne des questions supplémentaires qui permettent de capter plus précisément l'origine des parents (meilleure distinction entre immigrés et expatriés) explique en partie cette différence. Le risque de sous-estimation des descendants d'immigrés Algériens, relativement fort dans n'importe quelle source, est donc plus limité dans TeO (58 % de descendants Maghrébins en moins dans l'EEC par rapport à TeO). Toutefois, cette explication est insuffisante. Il serait opportun de se demander, par exemple, quelle part de cet écart est la conséquence directe de différences de méthode de calage entre ces deux enquêtes.

Toutes ces estimations sont des estimations nationales. La seule estimation d'un nombre de descendants d'immigrés au niveau local a été réalisée récemment pour la fiche 4.5 de l'INSEE Référence « Immigrés et descendants d'immigrés en France »². A partir de l'empilement des vingt dernières Enquêtes Emploi trimestrielles (de 2006 à 2010), la fiche 4.5 présente entre autre la part des descendants d'immigrés dans la population active par zone d'emploi³. Par l'empilement de données d'enquête sur cinq ans, cette estimation évite l'écueil de la non-significativité des effectifs et permet de mieux appréhender la répartition géographique des descendants à un niveau fin. Toutefois, cette méthode n'est pas exempte de défauts. Tout d'abord, à cause de ces problèmes d'effectifs, elle ne permet pas des estimations annuelles, qui sont d'un grand intérêt. Mais surtout, on juge d'une manière générale que l'EEC n'est

² INSEE, « Insee Références Immigrés et descendants d'immigrés en France », édition 2012

³ Une zone d'emploi est un espace géographique à l'intérieur duquel la plupart des actifs résident et travaillent, et dans lequel les établissements peuvent trouver l'essentiel de la main-d'œuvre nécessaire pour occuper les emplois offerts. Il y a 321 zones d'emploi en France métropolitaine.

significative qu'au niveau Ile de France / hors Ile de France. Une répartition géographique aussi fine à partir de l'EEC peut donc être soumise à caution étant donné que le lieu de résidence n'est pas intégré dans le calcul de la pondération. La seule répartition géographique fine jugée satisfaisante est celle estimée par le Recensement de la Population. Mais le Recensement ne pose pas de questions sur l'origine des parents, ce qui fait qu'on ne peut estimer les descendants d'immigrés par ce biais.

II - Ecart observés entre Recensement de la Population et Enquête Emploi

Le recensement (RP) est la seule source fiable pour faire des estimations locales. L'Enquête Emploi (EEC) est la source la plus fiable pour estimer les descendants d'immigrés à une fréquence régulière. Si l'on souhaite donc estimer à un niveau local le nombre de descendants d'immigrés, il faut modéliser le nombre de descendants via le recensement à partir des données existantes dans l'Enquête Emploi. Pour ce faire, il faut en premier lieu observer la cohérence des sources en comparant les variables explicatives.

Le champ de notre étude doit être comparable entre les deux sources. Pour ce faire, il porte uniquement sur:

- la période 2006-2010, ce qui correspond au Recensement 2008 et à l'empilement des EEC 2006-2010.
- la France métropolitaine (les DOM-COM ne sont pas dans l'EEC).
- les personnes ayant entre 18 et 64 ans inclus (âge « actif »).
- les personnes vivant en ménage ordinaire (les communautés ne sont pas enquêtées dans l'EEC)

Les variables d'intérêt pour notre étude présentes à la fois dans le recensement et l'enquête emploi sont les suivantes :

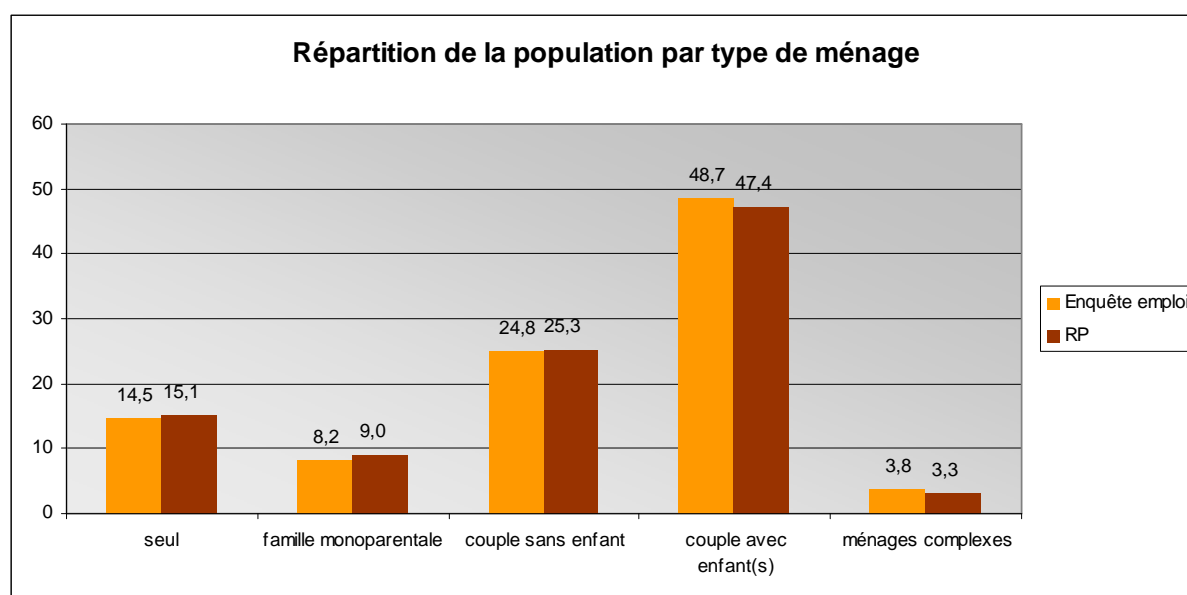
- o Sexe
- o Âge révolu (pas d'âge atteint dans l'année dans le RP)
- o Type de ménage :
 - > Personne seule
 - > Famille monoparentale
 - > Couple sans enfant
 - > Couple avec enfant(s)
 - > Ménage complexe
- o Niveau de diplôme :
 - > Licence et +
 - > BAC +2
 - > BAC
 - > Inférieur au BAC (CAP, BEP, brevet)
 - > Aucun diplôme ou CEP
- o Catégories socioprofessionnelles :
 - > Agriculteurs
 - > Artisans, Commerçants et Chefs d'entreprise (ACC)
 - > Cadres et Professions Intellectuelles Supérieures (CPIS)
 - > Professions intermédiaires
 - > Employés
 - > Ouvriers
- o Statut d'occupation des ménages :
 - > Propriétaire

- > Locataire HLM
- > Locataire privé
- > Logé gratuitement

Pour des raisons de non-comparabilité des concepts, nous n'avons pas intégré certaines variables d'intérêt dans notre étude. C'est le cas du chômage et de l'inactivité, qui sont calculés par croisement de questions précises dans l'EEC (d'après la définition du BIT⁴ (Bureau International du Travail)), alors qu'ils sont estimés de manière strictement déclarative dans le recensement, sans vérification du respect des critères du BIT. C'est également le cas du lieu de résidence en ZUS, donnée inutilisable dans le RP pour des raisons techniques.

Outre ces variables d'intérêt, notre fichier inclut les variables correspondant à différents niveaux géographiques, à savoir la Zone d'Etude et d'Aménagement du Territoire (variable ZEAT), la région (variable REG), et le département (variable DEP).

Nous avons systématiquement comparé les distributions des différentes variables dans chaque source utilisée pour chaque niveau géographique disponible (y compris France entière), exception faite de la variable sexe. Celle-ci servant de variable de calage entre les deux sources, les répartitions de la population par sexe dans chacune des sources sont logiquement quasi-identiques. Seuls les exemples les plus prégnants d'écarts entre recensement et Enquête Emploi sont présentés ici.

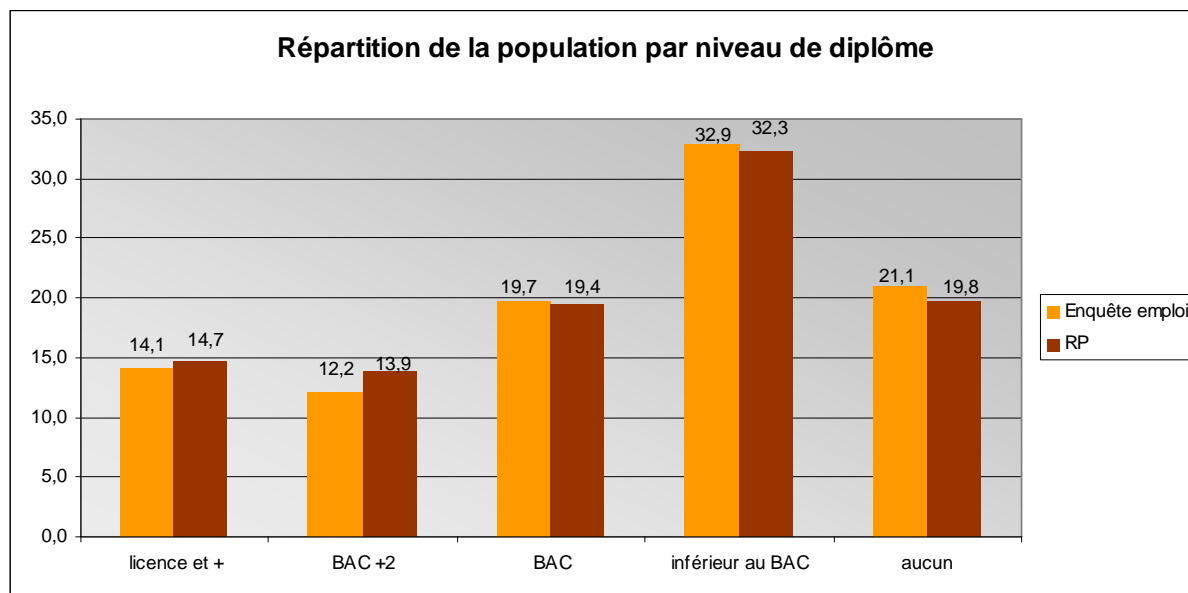


L'EEC surreprésente les couples, et particulièrement les couples avec enfant (+1,3 point) au détriment des personnes seules (avec ou sans enfant). En terme d'effectifs bruts, on constate que les ménages complexes sont les types de ménage les plus surestimés dans l'EEC (+ 17 %).

⁴ Au sens du Bureau International du Travail (BIT) :

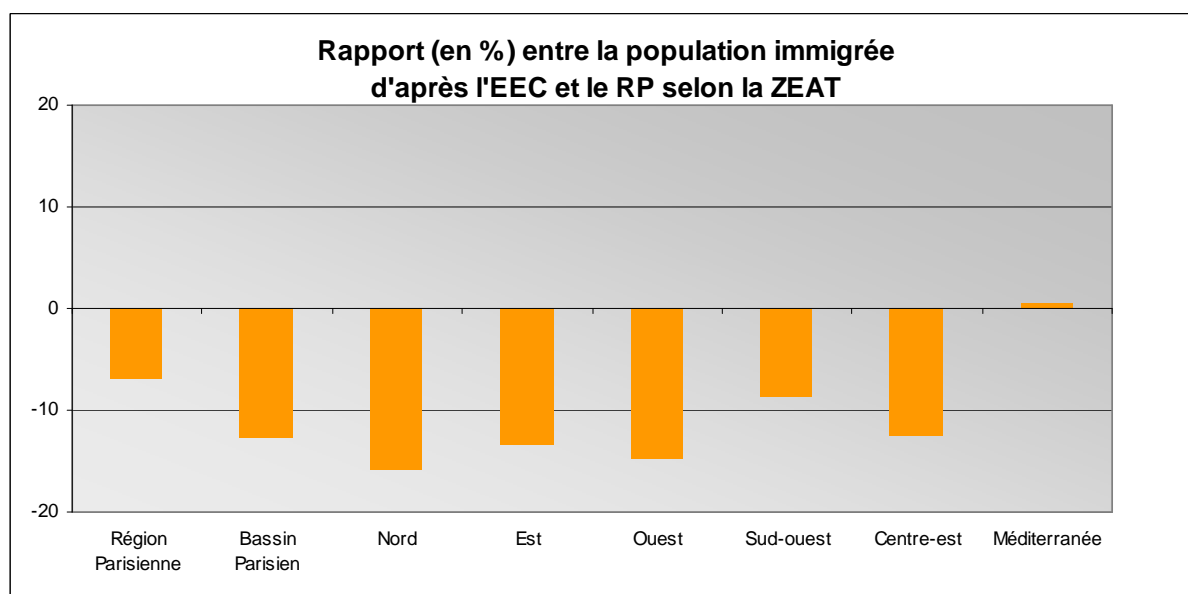
- un actif est un individu de 15 ans ou plus ayant travaillé (ne serait-ce qu'une heure) au cours d'une semaine de référence. Les personnes ayant un emploi mais qui sont temporairement absentes pour un motif tel que les congés payés ou une formation (entre autres) sont également actives occupées.
- Un chômeur est un individu de 15 ans ou plus qui (1) n'est pas actif occupé et (2) est disponible pour travailler dans les deux semaines et (3) a entrepris des démarches actives de recherche d'emploi dans le mois précédent ou (3') a trouvé un emploi qui commence dans les trois mois.
- un inactif est un individu qui n'a pas d'emploi et n'est pas chômeur.

Ils restent toutefois un type minoritaire, ce qui n'est pas le cas des familles monoparentales, sous-estimées de 9 % dans l'EEC.

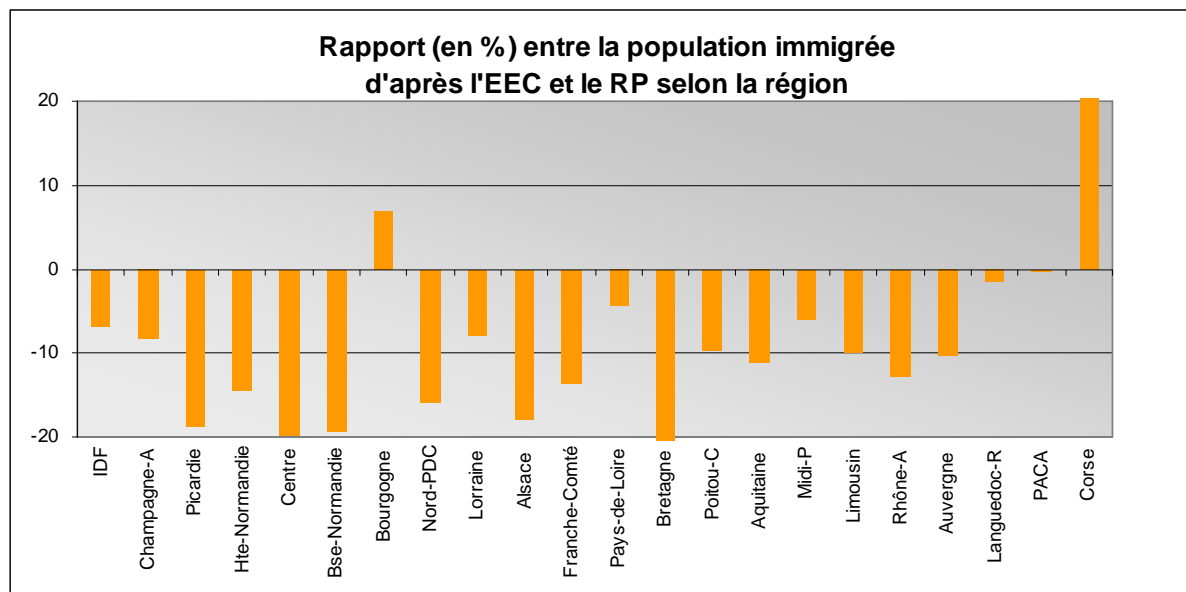


La répartition par diplôme est très différente selon la source. Les plus diplômés sont sous-représentés dans l'EEC (-1,7 point chez les Bac + 2) par rapport aux moins diplômés. Les sans-diplômes sont particulièrement nombreux dans l'EEC (+1,2 point).

La part d'immigrés est plus faible de 0,9 point dans l'EEC par rapport au RP (-9 %). Cette sous-estimation est plus flagrante lorsque l'on descend à un niveau géographique plus fin.

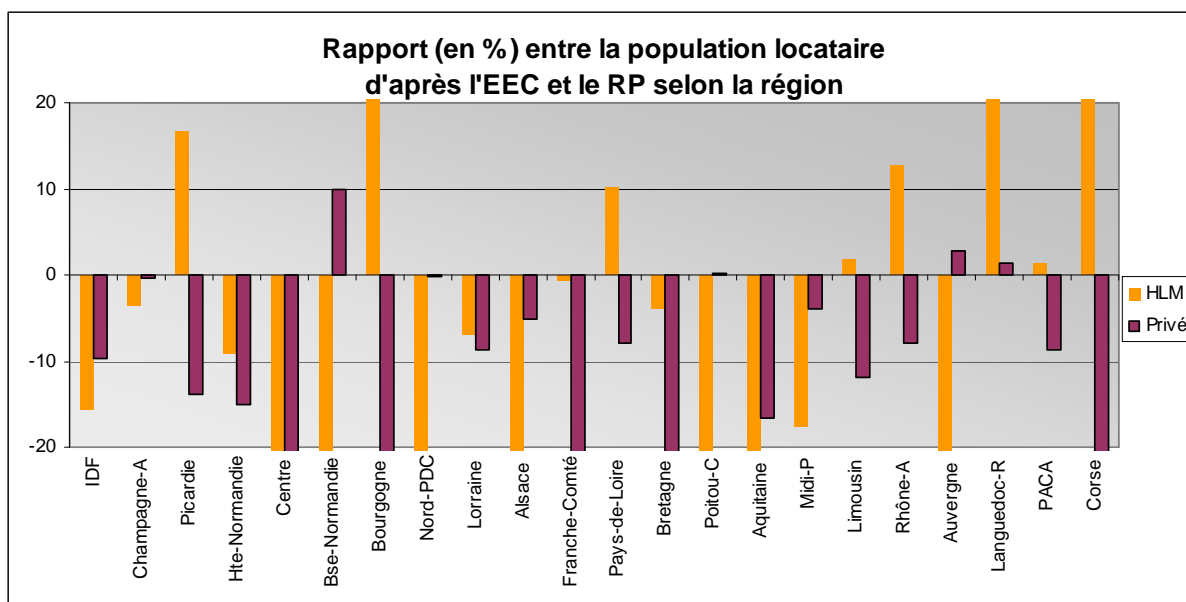


A part en Méditerranée, où les chiffres sont similaires, le nombre d'immigrés est en effet largement sous-estimé d'après l'EEC dans les autres ZEAT, l'écart atteignant 16 % de moins dans le Nord et 15 % dans l'Ouest. Cette sous-estimation n'est que de 7 % en région parisienne.



Les écarts observés entre l'EEC et le RP par région correspondent souvent aux écarts observés au sein de la ZEAT, mais certaines régions au sein de la ZEAT peuvent avoir des écarts très différents, ce qui montre qu'il n'y a pas foncièrement de sous représentation chronique dans toutes les régions au sein d'une ZEAT. A titre d'exemple, les immigrés sont sous-représentés d'environ 20 % dans toutes les régions du bassin parisien, à l'exception de la Bourgogne où les immigrés sont plus nombreux dans l'EEC. Hormis ces régions, les immigrés sont les plus sous-représentés en Bretagne (-25 %), en Alsace (-18 %) et dans le Nord Pas de Calais (-16 %). A l'exception de cette dernière région, les immigrés semblent moins sous-représentés dans les régions fortement peuplées (PACA, Rhône-Alpes, Ile de France (IDF)). A noter enfin qu'ils sont 26 % plus nombreux en Corse dans l'EEC, ce qui semble confirmer que les écarts sont plus forts dans les régions où la population est plus faible, mais pas foncièrement dans les régions où la population immigrée est moins présente.

De grandes différences de répartition ont lieu entre les deux sources en ce qui concerne le statut d'occupation au niveau national. L'écart est particulièrement important pour les propriétaires, qui sont largement surestimés dans l'EEC (+3,7 points) au détriment des locataires, notamment ceux du secteur privé (-2,2). En termes d'effectifs bruts, les propriétaires sont la seule catégorie de statut d'occupation surreprésentée dans l'EEC (+6 %). Les locataires sont sous-représentés de 8 à 9 %, les logés gratuitement (catégorie numériquement faible) de plus de 15 %. Ici, descendre à un niveau géographique plus fin exacerbe les écarts entre les deux sources.



La sous-estimation du nombre de locataires à la faveur du nombre de propriétaires cache parfois des tendances différentes selon le secteur de location. Dans la majorité des régions, cette sous-estimation se fait au détriment des locataires du secteur privé et des locataires du secteur public, mais de manière inégale selon le lieu. Dans la plupart des régions, la sous-estimation concerne surtout les locataires du secteur public : -32 % en Alsace, -38 % en Aquitaine, et autour de -20 % pour le Poitou-Charentes, l'Auvergne, le Nord-Pas de Calais, le Midi-Pyrénées, etc. Mais elle se fait parfois au détriment du secteur privé (Bretagne) ou bien concerne les deux secteurs équitablement (-20 % minimum dans la région Centre, entre -10 et -15 % en Ile de France). Cette sous-estimation globale des locataires ne n'observe toutefois pas dans toute la France : en Languedoc-Roussillon, les locataires HLM sont même surestimés (+25 %). Dans le Rhône-Alpes, en PACA, en Picardie, en Bourgogne et en Corse, on constate de même une forte surestimation des locataires HLM compensée par une sous-estimation tout aussi forte des locataires du privé. Le processus inverse (sous-estimation du secteur social compensé par une surestimation du secteur privé) est moins courant. Devant tant de variété de situation, aucune tendance régionale expliquant les écarts ne semble se dégager, ces-derniers étant d'ailleurs plus faibles dans les régions peuplées (même si on observe 13 % de locataires en moins en Ile de France d'après l'EEC).

Les variables pour lesquelles la distribution des effectifs pondérés varie le plus entre RP et EEC au niveau national sont donc le statut d'occupation, suivi de la présence immigrée, du niveau de diplôme puis du type de ménage. A un niveau géographique plus fin, certaines sous-populations de tailles significatives sont particulièrement concernées par ces écarts. Les locataires du secteur social sont de loin les plus touchés par ces différences RP-EEC. Ensuite, les plus gros écarts concernent les diplômés d'un Bac + 2, les immigrés et les familles monoparentales, ces dernières étant au même niveau de différence que les locataires du secteur privé.

III - Modélisations au niveau « régions »

Ce travail préliminaire a permis d'analyser la cohérence à un niveau local des deux sources pour chaque variable explicative. Nous nous proposons désormais de réaliser, à partir de l'Enquête Emploi en Continu (EEC), une modélisation "à la Fay et Herriot"⁵ de la population d'intérêt (les descendants d'immigrés) au niveau région en utilisant les régresseurs communs aux deux enquêtes décrits au cours de la partie précédente.

Le nombre de descendants d'immigrés (Y) est modélisé via une régression linéaire $Y = AX + \varepsilon$ où les coefficients A sont estimés à partir de l'EEC $Y_{eec} = \hat{A}X_{eec}$. Après le test des différents modèles possibles, le nombre de descendants d'immigrés par région sera estimé en appliquant le meilleur modèle sur les données du Recensement $\hat{A}X_{rp}$.

Dans le but d'estimer au mieux les descendants d'immigrés de 18-64 ans, notre premier travail consiste à créer une fonction du nombre de descendants d'immigrés en utilisant les régresseurs un par un.

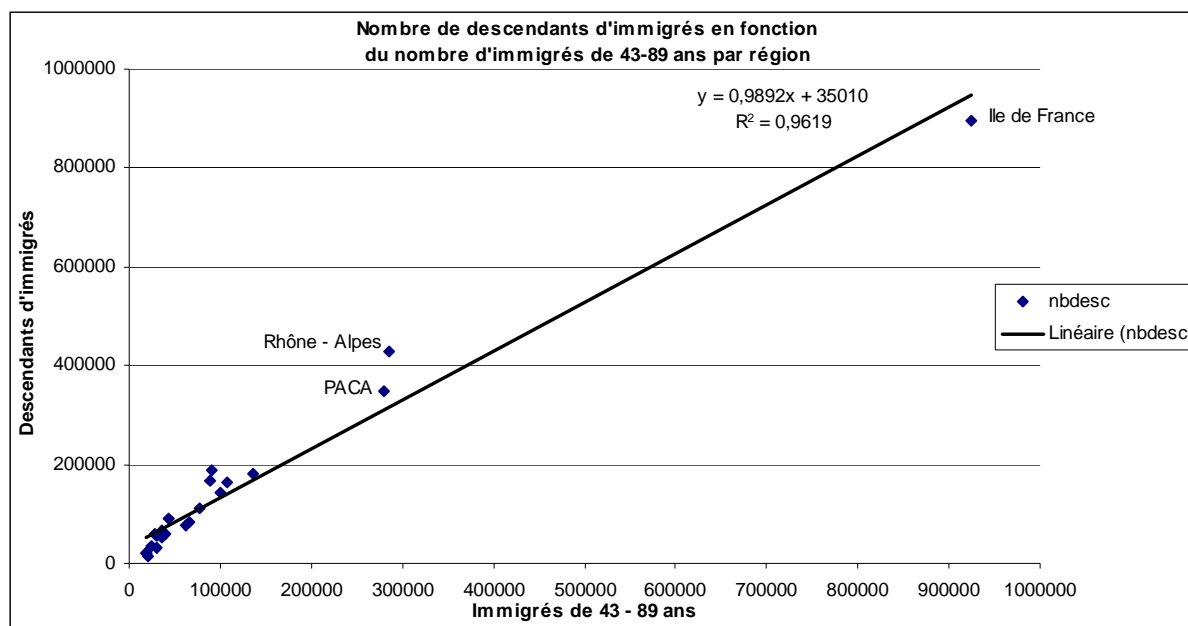
Modèle (DI = a+bx+e)	a	b	R ² (ajusté)	Signif var
x				
Immigrés	51707	0,6	0,915	***
Immigrés de 43 - 75 ans	40566	1,057	0,948	***
Immigrés de 43 - 89 ans	35010	0,989	0,96	***
Personnes seules	-51782	0,816	0,922	***
Couples sans enfant	-100478	0,591	0,848	***
Familles monoparentales	-41285	1,373	0,961	***
Couples avec enfant	-56303	0,248	0,909	***
Sans diplôme	-82885	0,647	0,909	***
Diplômes inférieurs au Bac	-104918	0,277	0,864	***
Licences ou plus	23977	0,521	0,902	***
Ouvriers et employés	-94994	0,371	0,867	***
Ouvriers	-115926	0,916	0,703	***
Employés	-67630	0,589	0,920	***
CPIS et Prof. intermédiaires	-5893	0,321	0,917	***
Locataires	-30683	0,277	0,957	***
Locataires HLM	9508	0,595	0,902	***
Propriétaires	-87798	0,235	0,845	***

Seuil de significativité du chi² : *** =< 0,001 ; ** =< 0,01 ; * =< 0,05 ; ns (non significatif) > 0,05

A l'exception du nombre de propriétaires, de couples avec enfants, d'ouvriers et de peu diplômés, le coefficient de corrélation (R²) dépasse 0,9 pour tous les modèles, tous les paramètres (variables incluses) étant significatifs au niveau régional. Avec un R² ajusté proche de 0,960, les modèles expliquant le nombre de descendants d'immigrés par le nombre de personnes en familles monoparentales ou par le nombre de locataires semblent les plus robustes.

⁵ http://www.insee.fr/fr/publications-et-services/docs_doc_travail/m0602.pdf : panorama des principales méthodes d'estimation sur les petits domaines, P. Ardilly

Par rapport à ces deux variables, le nombre d'immigrés semble être un régresseur moins significatif. ($R^2=0,915$) Ce phénomène s'explique logiquement car, si tous les descendants d'immigrés sont des enfants d'immigrés, ils ne sont pas les enfants de tous les immigrés, mais juste de ceux venus en France au cours de la génération précédente. Ainsi, le nombre de descendants d'immigrés d'un âge x devrait être en théorie fonction du nombre d'immigrés de l'âge " $x + \text{âge moyen à la naissance}$ ". Cet âge est complexe à établir car il n'existe pas de données d'âge moyen à la naissance selon le statut d'immigré. Les différentes enquêtes disponibles (EHF, RP) montrent cependant que la fécondité des immigrées est maximale entre 24 et 30 ans. L'âge des immigrés à la naissance de leur enfant n'étant pas le sujet central de notre étude, nous n'irons pas plus loin dans son estimation. Par facilité, nous considérerons donc qu'ils ont en moyenne leur enfant à 25 ans. Les descendants d'immigrés de 18-64 ans (âge atteint dans l'année) sont (en prenant l'année de référence 2008) nés entre 1944 et 1990. En théorie, leurs parents immigrés ont donc actuellement entre 43 et 89 ans. Bien entendu, cette fonction est purement théorique dans la mesure où nombre de parents de descendants d'immigrés ont disparu (décès, retour dans le pays d'origine), ou bien sont non-immigrés (la moitié des descendants d'immigrés sont nés de couple "mixte" natif-immigré). Toutefois, cette fonction théorique semble tout à fait significative. En effet, en ne considérant que les immigrés de 43-89 ans (plutôt que l'ensemble), le R^2 passe de 0,915 à 0,960, soit un niveau très élevé. Devant cette conclusion, il apparaît judicieux de prendre la variable « immigré de 43 – 89 ans » comme référence, plutôt que la variable « immigrés de 18 – 64 ans ».



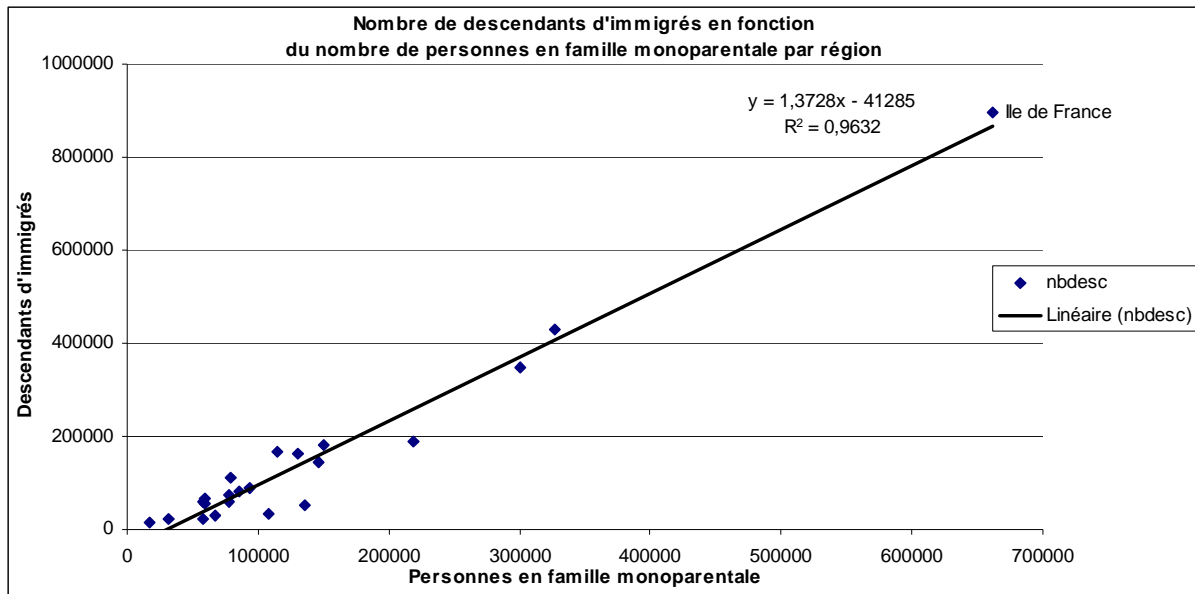
La courbe de tendance nous montre que l'estimation des descendants en fonction du nombre d'immigrés de 43 – 89 ans est assez éloignée de celle observée dans l'Enquête Emploi pour certaines régions. Ceci est particulièrement vrai dans les régions les plus peuplées (Rhône-Alpes, -113 000 par rapport à l'EEC, le Nord-Pas de Calais -63 000, l'Île de France -53 000, la Lorraine -43 000). Mais c'est également le cas dans des régions peu peuplées, où l'écart en différence absolue est plus faible mais celui en différence relative particulièrement prégnant. Ainsi, en Corse, l'estimation des descendants est 4 fois plus forte que la valeur observée dans l'EEC (56 000 au lieu de 13 000). En Basse Normandie, dans le Poitou-Charentes ou dans le Limousin, les descendants estimés sont également deux fois plus nombreux que dans l'EEC (+ 30 000). Estimer les descendants d'immigrés à l'aide du seul nombre d'immigrés de 43-89 ans

semble donc insuffisant : le modèle doit être complété par d'autres variables, présentées dans le tableau suivant.

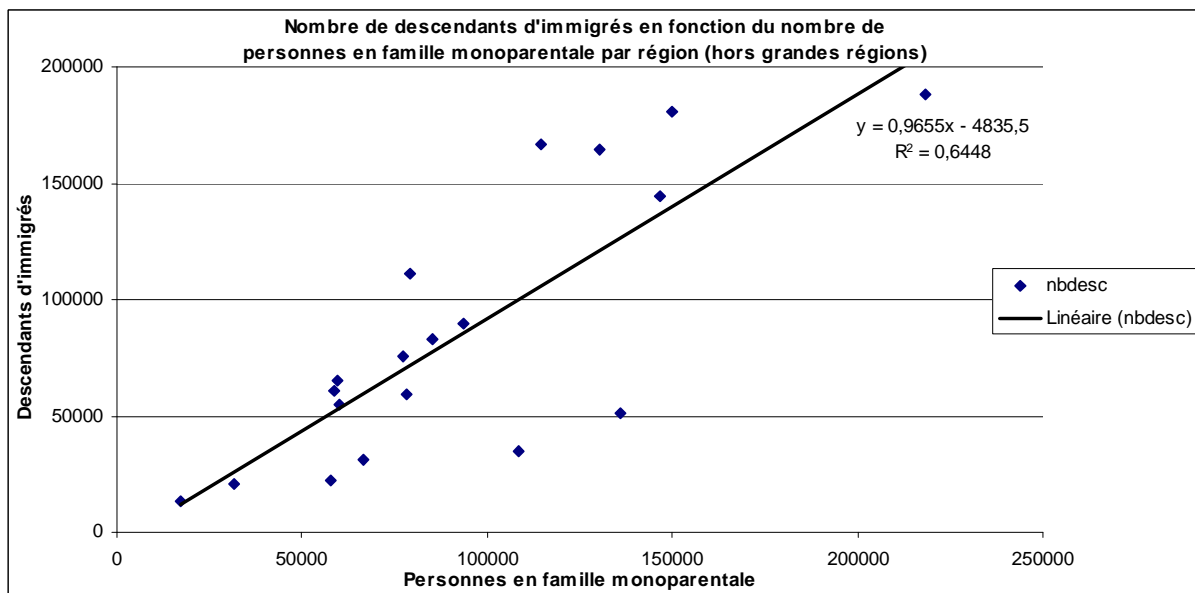
Modèle (DI = a+b1Immi4389+b2x2+e) x2	a	b1	b2	R ²	Signif var
personnes seules	11329	0,751	0,208	0,963	ns
Couples sans enfant	-2176	0,785	0,144	0,969	*
Familles monoparentales	-6759	0,496	0,711	0,979	***
Couples avec enfant	5740	0,725	0,072	0,967	*
Sans diplômes	-8846	0,684	0,22	0,974	**
Diplômes inférieurs au Bac	-10922	0,742	0,081	0,974	**
Licences ou plus	39676	1,256	-0,148	0,961	ns
Ouvriers et employés	-6800	0,746	0,106	0,973	**
Ouvriers	-9343	0,831	0,216	0,975	**
Employés	-1770	0,685	0,195	0,970	**
CPIS et Prof. intermédiaires	31673	0,921	0,023	0,958	ns
Locataires	2220	0,531	0,132	0,97	*
Locataires HLM	31232	0,883	0,068	0,959	ns
Propriétaires	2947	0,798	0,054	0,967	*

Pour toutes les variables d'intérêt, le R² reste fort mais certains paramètres des modèles deviennent non significatifs. Comme sur les modèles avec une seule variable, le meilleur modèle est celui incluant immigrés de 43-89 ans et personnes en familles monoparentales (R²= 0,979, toutes variables étant significatives au seuil de 1%). La majorité des autres modèles ont un R² proche de 0,97, avec de meilleurs résultats en incluant les peu diplômés et les ouvriers. Les variables ne sont significatives qu'à 95 % quand on inclut les locataires.

Ne pas estimer les descendants d'immigrés en fonction de la présence immigrée à minima peut surprendre. Cependant, les résultats des modèles en fonction exclusivement des personnes en famille monoparentale et en fonction du nombre de locataires sont si élevés (parfois plus que pour les immigrés de 43 – 89 ans) qu'il nous a paru indispensable d'inclure ces variables. Ceci est corroboré par la surreprésentation observée des descendants d'immigrés dans ces catégories de population. Ainsi, sur la période 2008-2010, 47 % des descendants d'immigrés de 15 – 64 ans sont locataires contre 35 % des Français de parents nés français. De même, 12 % des descendants d'immigrés de ces âges vivent en famille monoparentale contre 9 % des personnes sans lien à la migration.



Le nombre de personnes en famille monoparentale explique assez bien le nombre de descendants d'immigrés, qui vivent un peu plus souvent dans ce type de ménage. Cependant, on observe toujours de forts écarts selon les régions. Dans ce modèle, les plus forts écarts ne sont pas pour les régions peuplées, mais plutôt pour les régions où la présence immigrée est faible, particulièrement les régions de l'ouest, où l'estimation est trois fois plus importante que l'observation dans l'EEC (Pays de la Loire, + 94 000, Bretagne, + 72 000). Les écarts sont également élevés dans certaines régions comme le Nord-Pas de Calais (+ 70 000) ou la Lorraine (-51 000). Bien entendu, les forts écarts observés sont influencés par l'importance des grandes régions dans la régression linéaire. Mais en retirant les 3 régions les plus peuplées (Ile de France, Rhône-Alpes, PACA), le modèle estimé en fonction du nombre de personnes vivant en famille monoparentale est encore moins bon.



Avec un R^2 de 0,64, le modèle n'est plus assez robuste pour estimer les descendants d'immigrés. Ce constat montre qu'il serait peut être plus efficace d'appliquer des modèles différents selon la taille de la région.

Modèle (DI = a+b1monop+b2x2+e) x2	a	b1	b2	R ²	Signif var
Sans diplômes	-16581	2,038	-0,326	0,966	ns
Diplômes inférieurs au Bac	4655	2,13	-0,164	0,975	**
Licenciés ou plus	-32422	1,153	0,09	0,962	ns
Ouvriers et employés	-8783	1,987	-0,179	0,971	*
Ouvriers	-9693	1,664	-0,248	0,97	*
Employés	-21834	2,182	-0,358	0,966	ns
CPIS et Prof. intermédiaires	-37910	1,2	0,043	0,96	ns
Locataires	-38361	0,929	0,09	0,961	ns
Locataires HLM	-36161	1,199	0,081	0,961	ns
Propriétaires	-13178	1,923	-0,104	0,971	*

Les modèles incluant obligatoirement le nombre de personnes en famille monoparentale ne donnent pas d'aussi bons résultats quand on ajoute un autre paramètre. Tous les modèles ont des problèmes de significativité des variables, sauf le modèle incluant les peu diplômés (R²=0,97). Le nombre de personnes en famille monoparentale est donc une bonne fonction du nombre de descendants d'immigrés, mais ce constat n'est plus vrai dès qu'on rajoute d'autres types de variables pour améliorer le modèle.

On tire la même conclusion lorsqu'on estime le nombre de descendants d'immigrés en fonction exclusivement du nombre de locataires. La conclusion étant la même, nous ne développerons que très brièvement ce modèle. Le nombre de descendants d'immigrés s'avère globalement une bonne fonction croissante du nombre de locataires. Cependant, certaines estimations régionales sont très éloignées de celle de l'EEC. C'est le cas notamment dans l'ouest de la France où la part de locataires est particulièrement faible. Toutefois, cet argument ne semble pas suffisant pour expliquer le manque de robustesse du modèle pour ces régions, car le nombre de descendants est également surestimé de 25 % dans le Nord-Pas de Calais où le nombre de locataires est fort (idem dans les régions du sud). Le nombre élevé de régions pour lesquelles le modèle ne convient pas confirme que le modèle ne peut se baser en premier lieu sur le nombre de locataires et qu'il faudrait utiliser plusieurs modèles selon la région. Ces deux exemples montrent l'intérêt de privilégier le nombre d'immigrés de 43 – 89 ans dans les modèles.

Après avoir testé 134 modèles, les meilleures estimations se limitent à trois variables d'intérêt, comprenant nécessairement le nombre d'immigrés de 43 – 89 ans. Voici un récapitulatif des modèles les plus significatifs, les meilleurs modèles étant surlignés en orange :

Modèle (DI = a+b1immi43-89+ b2+b3x3+e)	a	b1	b2	b3	R ²	Signif var
b2 : Famille monop. b3 : Licences ou plus	-4926	0,879	0,816	-0,252	0,987	**
b2 : Famille monop. b3 : CPIS et Professions intermédiaires	-2168	0,7	0,931	-0,122	0,984	*
b2 : Ouvriers b3 : Couples avec enfant	-16996	1,312	-0,229	0,707	0,982	**
b2 : Ouvriers b3 : Licences ou plus	-15050	1,374	0,295	-0,332	0,987	***
b2 : Ouvriers b3 : Propriétaires	-6853	1,047	0,624	-0,144	0,981	*
b2 : Locataires b3 : Licences ou plus	-11044	0,964	0,243	-0,452	0,989	***
b2 : Locataires b3 : CPIS et Professions intermédiaires	-2902	0,703	0,305	-0,264	0,982	**

Trois modèles ont des R² se rapprochant de 0,99. Tous ces modèles incluent comme variable d'intérêt le nombre d'immigrés de 43 – 89 ans et le nombre de diplômés du supérieur (licence ou plus). La troisième variable d'intérêt séparant ces modèles est soit le nombre de personnes en famille monoparentale (modèle 1, n'est significatif qu'au seuil de 1 %), soit le nombre d'ouvriers (modèle 2), soit le nombre de locataires (modèle 3, ayant le meilleur R²).

Le modèle 1 donne des résultats par région assez proche de ce qui est observé dans l'EEC. Toutefois, on retrouve des écarts numériques plus importants sur les régions peuplées du sud de la France (-40 000 en Rhône-Alpes et Midi-Pyrénées, + 40 000 en PACA). Comme pour tous les modèles, on obtient une surestimation relative dans les régions de l'ouest mais sans que l'écart ne soit trop important (sauf en Pays de la Loire + 31 000). Enfin, on constate également une sous-estimation de 35 000 en Lorraine. On retrouve donc avec ce modèle les écarts par région observés dans tous les modèles, en moins prégnant, mais aussi un écart plus marqué dans les régions du sud, alors que l'estimation de l'Ile de France est similaire à celle de l'EEC.

Le modèle 2 donne des résultats globalement similaires au modèle 1 avec des écarts plus diffus et seulement importants dans certaines régions. Le nombre de descendants d'immigrés est ainsi deux fois supérieur par rapport à celui de l'EEC dans les Pays de la Loire (+ 45 000). Il est par contre fortement sous-estimé dans le Midi-Pyrénées (-52 000) et dans le Nord-Pas de Calais (-32 000). Aucune tendance géographique ne se distingue donc vraiment.

Comme le laissait supposer son R², le modèle 3 (immigrés de 43 – 89 ans, diplômés d'une licence ou plus, locataires) nous permet d'obtenir la meilleure prédiction. Une zone géographique reste néanmoins problématique : les régions peuplées du sud, avec une forte sous-estimation dans le Midi-Pyrénées (-42 000) et dans le Rhône-Alpes (-34 000), couplé à une surestimation en PACA (+ 33 000). En écart relatif, certaines régions de l'ouest (dont les Pays de la Loire mais surtout la Bretagne) sont surestimées de plus d'un tiers, mais les chiffres concernés sont plus faibles. A noter que l'estimation pour l'Ile de France est très proche de l'EEC, quel que soit le modèle.

Les écarts régionaux selon les modèles interrogent. Faudrait-il différents modèles selon les régions pour arriver à une estimation proche des résultats de l'EEC ? La question se pose dans la mesure où les caractéristiques sociodémographiques sont propres à chaque région, ce qui

induit que la présence d'origine immigrée a une identité unique dans chaque région. Le tableau ci-dessous résume les modèles donnant les estimations les plus fidèles pour chacune de ces régions. On constate que le modèle 3, le plus fiable au niveau national, est le meilleur pour 6 régions, le modèle 1 pour 5 et le modèle 2 pour 4. La Champagne-Ardenne et le Midi-Pyrénées ont les meilleures estimations avec des modèles n'incluant pas les immigrés de 43 – 89 ans. Aucune tendance géographique ou démographique n'explique qu'un modèle marche mieux dans tel région plutôt que dans une autre. A noter que quel que soit le modèle national, aucune estimation proche de l'EEC n'est possible pour la Basse Normandie et les Pays de la Loire, régions pour lesquelles on observe au mieux une surestimation de 39 %.

Meilleur modèle d'estimation	
Immigrés de 43-89 ans, licence et +, locataires	Picardie, Nord-Pas de Calais, Franche Comté, Bretagne, Poitou-Charentes, Languedoc-Roussillon
Immigrés de 43-89 ans, licence et +, famille monoparentale	Ile de France, Haute Normandie, Centre, Aquitaine
Immigrés de 43-89 ans, licence et +, ouvriers	Basse Normandie, Lorraine, Limousin, Rhône-Alpes
Immigrés de 43-89 ans, locataires	Bourgogne, Auvergne
Immigrés de 43-89 ans	Alsace, Pays de la Loire
Locataires	Champagne-Ardenne, Midi-Pyrénées
Autres modèles	PACA, Corse

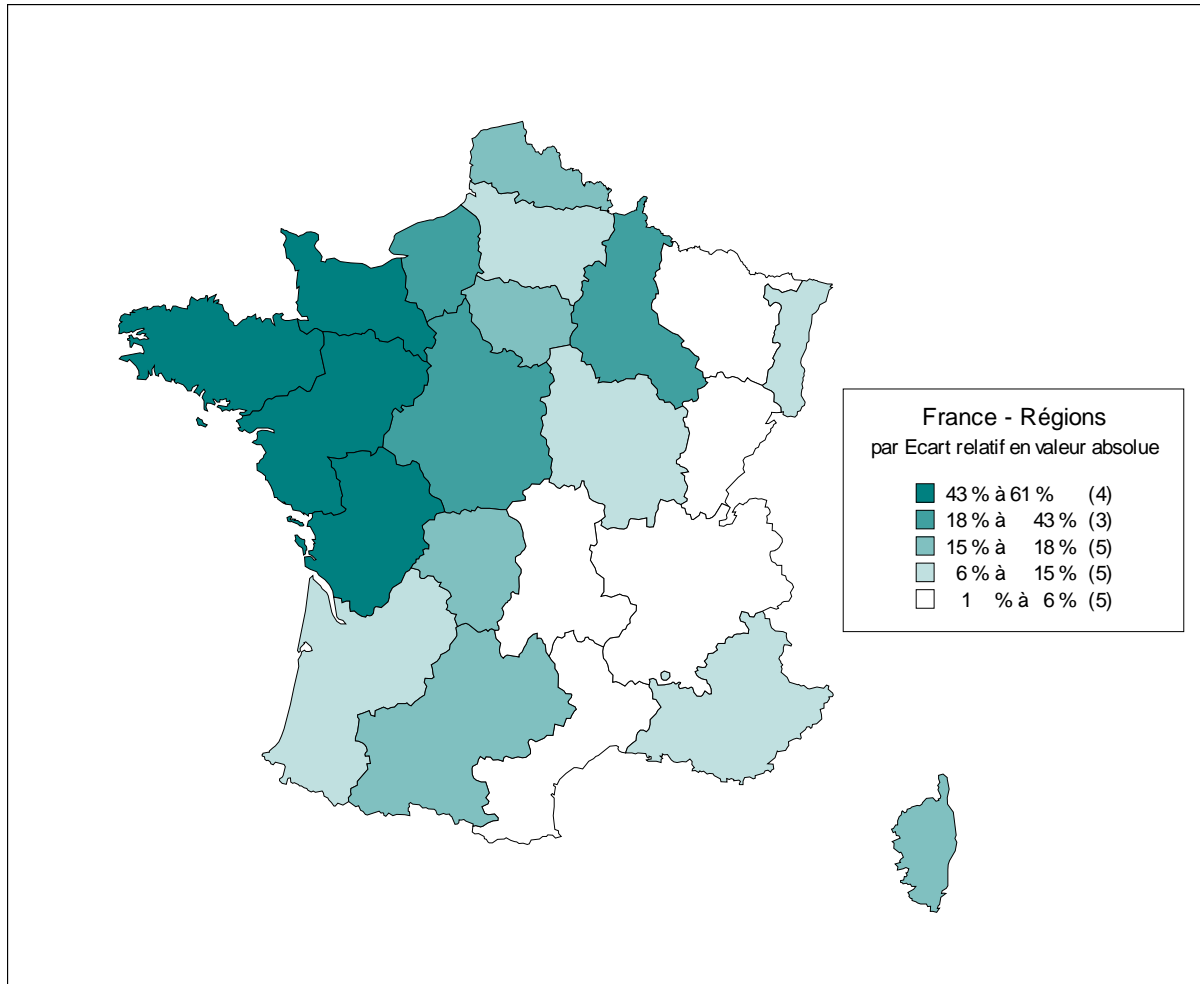
Des travaux futurs sur le sujet devront prendre en compte des modèles différents selon le niveau géographique. En attendant ces nouvelles avancées, le modèle 3 (immigrés de 43 – 89 ans, diplômés d'une licence ou plus, locataires) reste le meilleur modèle pour estimer le nombre de descendants d'immigrés au niveau local sur l'ensemble du territoire métropolitain.

IV – Application des modèles au Recensement de la Population

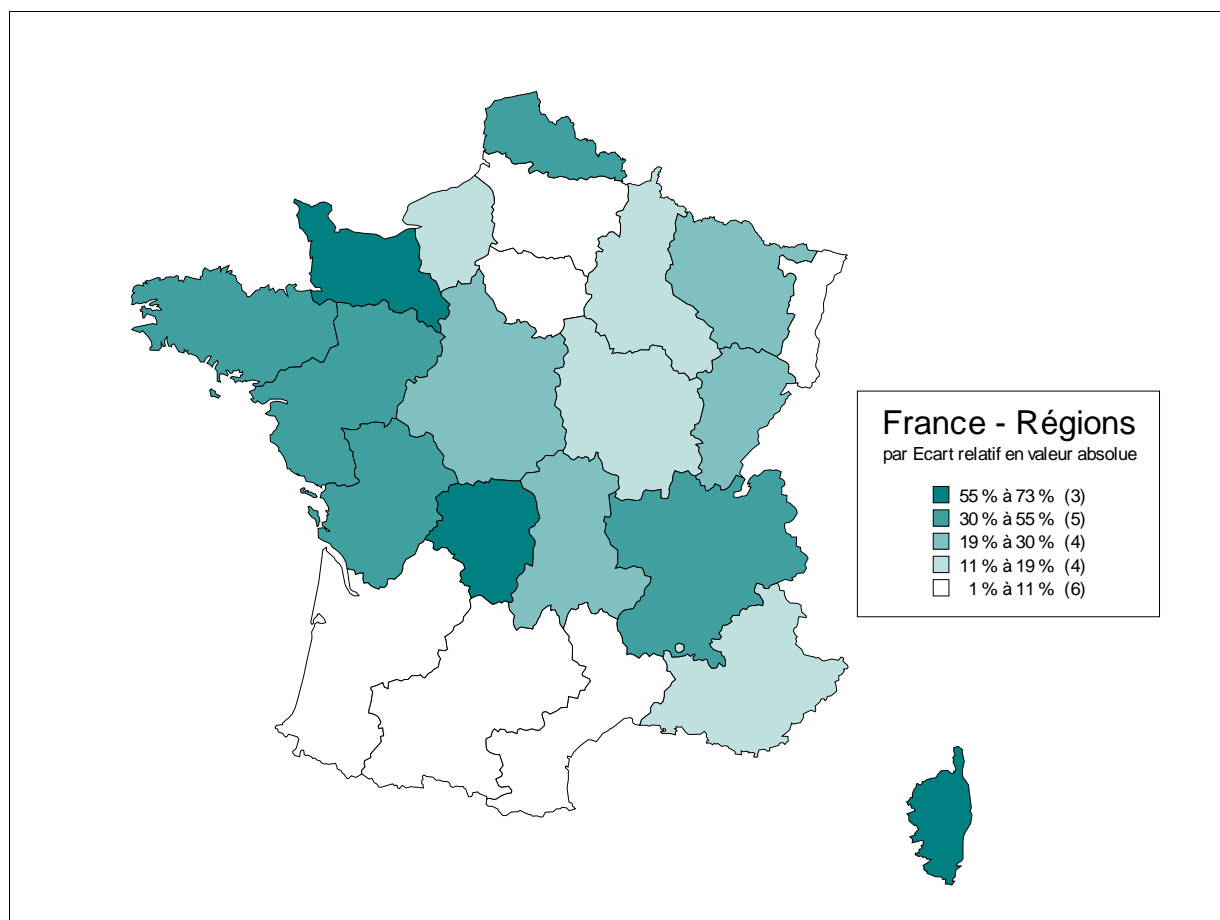
Les résultats obtenus par région sont les suivants :

Région	Nombre de diplômés de licence et +	Nombre de locataire	Nombre d'immigrés âgés	Estimation du nombre de descendants d'immigrés	Nombre de descendants d'immigrés dans l'EEC	Ecart relatif
CORSE	21 514	74 538	13 741	11 262	13 127	-17%
LIMOUSIN	45 908	141 307	20 835	24 349	20 457	16%
BASSE-NORMANDIE	82 872	327 105	20 018	55 566	21 904	61%
POITOU-CHARENTES	106 522	334 353	34 669	60 433	31 122	49%
BRETAGNE	228 212	577 394	34 033	61 786	35 014	43%
PAYS DE LA LOIRE	230 494	689 375	40 894	99 376	51 074	49%
AUVERGNE	84 927	263 179	37 312	55 022	54 574	1%
HAUTE-NORMANDIE	111 820	441 006	36 243	89 214	59 385	33%
FRANCHE-COMTE	73 998	248 734	40 371	60 195	60 631	-1%
CHAMPAGNE-ARDENNE	73 981	331 176	40 293	83 294	65 092	22%
BOURGOGNE	94 351	325 800	56 327	87 880	75 282	14%
CENTRE	157 796	517 801	75 614	127 973	83 186	35%
PICARDIE	107 511	405 067	49 362	95 525	89 596	6%
ALSACE	143 992	426 980	90 823	126 271	110 809	12%
AQUITAINE	242 346	679 057	108 741	162 769	144 349	11%
MIDI-PYRENEES	264 538	609 883	115 928	138 944	164 491	-18%
LORRAINE	149 780	511 497	105 269	162 365	167 006	-3%
LANGUEDOC-ROUSSILLON	202 727	588 570	133 632	185 604	180 607	3%
NORD-PAS-DE-CALAIS	263 355	965 801	99 248	221 748	188 431	15%
PROVENCE-ALPES-COTE D'AZUR	417 575	1 222 699	271 981	395 174	348 819	12%
RHÔNES-ALPES	554 936	1 442 682	297 881	409 965	429 592	-5%
ILE-DE-FRANCE	1 807 216	3 601 810	977 793	1 063 835	895 422	16%

D'après ce modèle, le nombre de descendants d'immigrés est estimé à 3,7 millions environ, une estimation plus proche de celle de TeO (ce qui s'explique amplement par le fait que TeO soit calée sur le recensement). Les quatre régions (voir carte ci-dessous) pour lesquelles les écarts relatifs sont les plus importants ($\geq 43\%$) sont les régions concentrées dans l'ouest de la France et dont la population de descendants d'immigrés reste faible d'après l'EEC. Il s'agit de la Bretagne, de la Basse-Normandie, des Pays de la Loire et de la région Poitou-Charentes. En contrepartie, les régions de l'est, qu'elles soient du nord ou du sud, ont très peu d'écarts avec les estimations de l'EEC.



Pour donner un autre exemple de répartition géographique à partir d'un autre modèle, nous avons appliqué celui où le nombre de descendants d'immigrés n'est fonction que du nombre d'immigrés âgés de 43 à 89 ans (voir carte ci-dessous). Les écarts régionaux sont toujours présents, mais dans des régions plus diverses. Les cinq régions pour lesquelles les écarts relatifs sont les plus importants ($\geq 55\%$) sont les régions où la population de descendants d'immigrés est la plus faible d'après l'Enquête Emploi. Ajoutons que pour le Nord-Pas de Calais, l'écart est également élevé puisqu'il atteint à 41 %. A contrario, pour les régions du sud-ouest (Aquitaine, Midi-Pyrénées et Languedoc-Roussillon) ainsi qu'en Ile de France, cet écart est inférieur ou égal à 10 %.



Cette étude a clairement mis en exergue les difficultés à faire un modèle unique applicable à toutes les régions. Pour confirmer nos interrogations à ce sujet, on construit cette-fois-ci plusieurs modèles à partir de l'EEC sur les cinq régions Basse Normandie, Pays de la Loire, Bretagne, Poitou-Charentes et Limousin, du fait de la faible qualité de prédiction des modèles précédents concernant ces régions. Le modèle le plus significatif pour ces régions est celui où le nombre de descendants d'immigrés est fonction uniquement du nombre de personnes vivant en famille monoparentale. Après avoir appliqué ce modèle aux cinq mêmes régions dans le Recensement de la Population, les résultats obtenus sont les suivants :

Région	Nombre d'immigrés vivant dans des familles monoparentales	Estimation du nombre de descendants d'immigrés	Nombre de descendants d'immigrés dans l'EEC	Ecart relatif
BASSE-NORMANDIE	65 399	27 804	21 904	21%
PAYS DE LA LOIRE	134 856	47 267	51 074	-8%
BRETAGNE	133 767	46 962	35 014	25%
POITOU-CHARENTES	76 772	30 991	31 122	0%
LIMOUSIN	33 008	18 728	20 457	-9%

Par rapport au modèle « France entière », les écarts entre l'estimation et le nombre de descendants ont sensiblement diminué dans toutes les régions, et ont même quasiment disparu dans les Pays de la Loire et en Poitou-Charentes. La qualité du modèle employé (R^2 ajusté = 0,79) est plutôt satisfaisante.

Réaliser une estimation « petits domaines » d'une population sur à partir d'une source à l'aide de modèles construits via une autre source s'avère un exercice complexe qui se trouve vite limité par les problèmes de comparabilité des dites-sources. Comme nous avons pu le démontrer, aucun modèle n'est parfait pour exprimer la « réalité locale » de chaque région. Cette difficulté oblige à s'interroger sur les travaux à venir pour améliorer l'estimation.

Une des pistes consiste à appliquer plusieurs modèles différents selon la région de manière à se rapprocher de l'estimation « France entière » de l'EEC, qui reste la source la plus fiable pour quantifier au niveau national le nombre de descendants d'immigrés. Il serait aussi très intéressant de prolonger le travail précédent en essayant de descendre à un niveau local plus fin (départements, voire Zone d'Emploi). Cela permettrait de consolider les résultats précédents, si les problèmes rencontrés s'avèrent être les mêmes ; ou justement de pousser la recherche dans une autre direction s'ils s'avèrent différents. Il serait également particulièrement intéressant de relier les modèles régionaux à une ambition d'estimation des descendants selon l'origine des parents, en prenant de grands groupes nationaux d'origine des descendants (Europe du Sud et Maghreb en particulier).

Toutefois, au vu des résultats de cette étude, une autre direction de recherche peut être envisagée. En effet, les méthodes appliquées ici se basent sur l'hypothèse forte que l'on peut appliquer des modèles estimés via l'EEC sur les données du recensement. Si cette hypothèse est théoriquement pertinente (l'EEC est « calée » sur le recensement), nos travaux préliminaires présentés en deuxième partie ont montré que ce calage ne donne pas les mêmes estimations de population immigrée selon les deux sources, ce qui a nécessairement une incidence sur l'estimation de leurs descendants. La pondération de l'EEC est calée sur le Recensement pour être représentative de la répartition de la population nationale par grandes caractéristiques démographiques (sexe, âge,...). Cependant, elle n'est ni représentative au niveau local, ni représentative d'autres sous-population. A partir de ce constat, il faudrait réfléchir à calculer pour l'EEC un jeu de pondération spécifique pour mieux correspondre aux résultats du recensement concernant nos sous-populations d'intérêt (immigrés, locataires, etc.). Au vu des résultats présentés ici, faire ces pondérations spécifiques semble l'axe de recherche prioritaire pour des travaux futurs.

Résumé. Selon les dernières estimations réalisées à l'aide de l'Enquête Emploi (EEC) et de l'enquête "Trajectoires et Origines", la part des descendants d'immigrés dans la population est de l'ordre de 10 % (Breem, 2010 ; Borrel, 2010). Il existe d'autres sources importantes d'information sur la population ayant un lien à la migration parmi lesquelles figure le Recensement de la Population (RP). Mais ce-dernier, a contrario des enquêtes ménages, ne permet pas de comptabiliser les descendants d'immigrés. Or, il est la seule source dont la taille de l'échantillon est adaptée à une estimation locale de population. L'objectif de l'étude est de déterminer, à l'aide de ces deux sources disponibles, une solution méthodologique permettant d'estimer la population des descendants d'immigrés par zone infra nationale.

Pour s'assurer de la cohérence entre RP et EEC, il s'agit de comparer les "explicatives". A un niveau géographique fin, la distribution des effectifs pondérés varie le plus entre les deux sources pour certaines sous-populations, particulièrement les locataires du secteur social, les diplômés d'un Bac + 2, les immigrés et les familles monoparentales.

Ceci réalisé, les auteurs se proposent de réaliser, à partir de l'EEC, une modélisation "à la Fay et Herriot" de la population d'intérêt (les descendants d'immigrés) au niveau région en utilisant les régresseurs communs aux deux sources. La meilleure modélisation est celle où le nombre de descendants d'immigrés est une fonction du nombre d'immigrés de 43 – 89 ans, de diplômés d'une licence ou plus et de locataires. Appliqué au RP, ce modèle permet d'estimer

le nombre de descendants d'immigrés à 3,7 millions environ. Cependant, les écarts observés selon les régions entre EEC et RP restent importants avec ce modèle.

Une des pistes de réflexion future consiste donc à appliquer plusieurs modèles différents selon la région de manière à se rapprocher de l'estimation « France entière » de l'EEC. Il serait aussi très intéressant de relier les modèles régionaux à une ambition d'estimation à un niveau local plus fin (départements, voire Zone d'Emploi) ou pour estimer des groupes de descendants selon l'origine des parents (Europe du Sud et Maghreb en particulier). Toutefois, au vu des résultats de cette étude, une autre direction de recherche peut être envisagée. En effet, les méthodes appliquées ici se basent sur l'hypothèse forte que l'on peut appliquer des modèles estimés via l'EEC sur les données du recensement. Si cette hypothèse est théoriquement pertinente (l'EEC est « calée » sur le recensement), cette étude montre que ce calage ne donne pas les mêmes estimations de population immigrée selon les deux sources, ce qui a nécessairement une incidence sur l'estimation de leurs descendants. La pondération de l'EEC n'est en effet ni représentative au niveau local, ni représentative d'autres sous-populations. A partir de ce constat, il faudrait réfléchir à calculer pour l'EEC un jeu de pondération spécifique pour mieux correspondre aux résultats du recensement concernant nos sous-populations d'intérêt (immigrés, locataires, etc.).

Bibliographie

(2012). Fiche 4.5 : Activité par zone d'emploi, Immigrés et descendants d'immigrés en France, INSEE Références, 188-189.

Ardilly, P. (2006). Panorama des principales méthodes d'estimation sur les petits domaines, Documents de Travail INSEE, M0602.

Borrel, C., (2012). Synthèse sur le thème des Migrations internationales, Note de l'INSEE, 773/DG75-F110.

Borrel, C. et Simon, P. (2005). Les origines des Français, Histoire de familles, histoires familiales, les résultats de l'enquête, Cahiers de l'INED, 425-441.

Breem, Y. (2010). Les descendants d'immigrés, Infos Migration, 15.

Equipe TeO (2010). Trajectoires et Origines, Enquête sur la diversité des populations en France, Documents de Travail INED, 168.