

# L'intégration des nouveaux migrants : impact de la naturalisation et des formations linguistiques

Anaëlle Solnon

► **To cite this version:**

Anaëlle Solnon. L'intégration des nouveaux migrants : impact de la naturalisation et des formations linguistiques. Economies et finances. 2014. dumas-01109910

**HAL Id: dumas-01109910**

**<https://dumas.ccsd.cnrs.fr/dumas-01109910>**

Submitted on 27 Jan 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

UNIVERSITÉ PARIS 1 - UFR02 SCIENCES ÉCONOMIQUES

---

ECONOMIE QUANTITATIVE  
MASTER 2 RECHERCHE  
ECONOMIE THÉORIQUE & EMPIRIQUE

---

L'INTÉGRATION DES NOUVEAUX MIGRANTS :  
IMPACT DE LA NATURALISATION ET DES  
FORMATIONS LINGUISTIQUES

---

*Présenté et soutenu par :*  
Anaëlle Solnon

*Directeur de soutenance :*  
M. Hippolyte d'Albis

6 juin 2014

L'université de Paris 1 Panthéon-Sorbonne n'entend donner aucune approbation ni désapprobation aux opinions émises dans ce mémoire ; elles doivent être considérées comme propres à leur auteur.

## **Abstract**

En dépit de la masse grandissante de données publiques disponibles sur les populations immigrées et issues de l'immigration, les études de type quantitatif sur ces dernières sont très rares en France. En particulier et bien que le déclassement et la pauvreté des nouveaux immigrés soit un sujet de recherche désormais classique, peu de travaux s'intéressent aux moyens pour ces nouveaux migrants de s'extraire de leur désavantage économique. Nous nous proposons dans ce travail d'évaluer l'impact de deux instruments de politique migratoire susceptibles de faciliter l'intégration des immigrés : les formations linguistiques récemment mises en place dans le cadre du Contrat d'Accueil et d'Intégration et la naturalisation en tant qu'étape du processus d'intégration. L'enjeu méthodologique est de taille car la naturalisation comme la formation linguistique ne sont pas des traitements aléatoires : les migrants naturalisés sont positivement sélectionnés sur des caractéristiques observables mais également inobservables à l'économétricien. Nous utilisons successivement une régression de discontinuité, une estimation de double-différence et une estimation de panel pour neutraliser ces effets de sélection. Nous trouvons un léger et contestable impact de la formation linguistique sur le niveau en langue française des migrants mais nous n'observons pas d'effets sur des résultats plus larges tels que la probabilité de sortie du chômage, le revenu, le nombre de mois travaillés ou encore la volonté de s'installer définitivement. Avec un modèle à effets fixes nous ne trouvons pas d'impact significatif de la naturalisation.

# 1 Introduction

En dépit de la masse grandissante de données publiques disponibles sur les populations immigrées et issues de l’immigration, les études de type quantitatif sur ces dernières sont très rares en France. En particulier, et alors que l’intégration des nouveaux migrants est un sujet brûlant en France<sup>1</sup>, peu de travaux s’attachent à évaluer l’efficacité des outils de politiques migratoires visant à faciliter l’intégration des nouveaux arrivants. Dans ce travail, nous nous proposons d’analyser l’effet de deux de ces outils : d’une part, la naturalisation en tant qu’instrument d’intégration des migrants sur le marché du travail et d’autre part, les formations en français proposées aux nouveaux arrivants. Nous utilisons pour cela les données de l’enquête ELIPA (Enquête Longitudinale sur l’Intégration des Primo-Arrivants). Il s’agit d’une enquête longitudinale menée par le département « statistiques » du Ministère de l’intérieur français, qui vise à observer le parcours des nouveaux arrivants signataires d’un Contrat d’Accueil et d’Intégration<sup>2</sup> en 2009 : 6107 « nouveaux migrants » ont été interrogés en 2010, juste après l’obtention de leur premier titre de séjour, 4756 ont ensuite été réinterrogés en 2011 et 3573 ont finalement été réinterrogés en 2013.

Globalement, la part de la population immigrée en France est relativement stable depuis le début des années 60 : elle fluctue de 6% à 8% et était en légère hausse ces dernières années (+2% par an en valeur absolue de 1999 à 2008 (fiches thématiques INSEE, 2012 [14])). En 2010, 8,5% de la population française (5,5 millions d’individus en valeur absolue) est immigrée au sens de l’INSEE<sup>3</sup>, un chiffre légèrement en deçà des 11,1% calculés par les Nations-Unies qui possèdent une définition moins restrictive<sup>4</sup> du statut de migrant. En 2010, Gilles Pison trouve à partir de ces chiffres que la France se situe au 54e rang (sur 197) des pays accueillant le plus d’immigrés (Gilles Pison, 2010 [22]). Toutefois, avec environ 40% de sa population immigrée qui a acquis la nationalité française, la France se situe en dessous de la moyenne des pays de l’OCDE ([21]). De plus, alors que la tendance générale en Europe est à la hausse du nombre de naturalisations, en France, le nombre d’immigrés accédant à la naturalisation tend à décroître légèrement depuis 2004 (passant d’environ 168 000 acquisitions en 2004 à 137 000 en 2008). Ceci s’explique par un durcissement des conditions de

---

1. En témoigne la récente proposition de réforme du « Contrat d’Accueil et d’Intégration » ou la plus ancienne création de l’éphémère « Ministère de l’Immigration, de l’Intégration, de l’Identité nationale et du Codéveloppement ».

2. Le Contrat d’Accueil et d’Intégration (CAI) est géré par l’Office Français de l’Immigration et de l’Intégration (OFII). Expérimenté dans certains départements à partir de 2003 puis rendu obligatoire en 2007 (loi du 24 juillet 2006 relative à l’immigration et à l’intégration), il vise les engagements réciproques d’un étranger nouvellement arrivé légalement en France et les engagements des autorités françaises dans un contrat individuel. Le non-respect de ce contrat par l’étranger peut entraîner le non renouvellement de sa carte de séjour. Ce contrat comporte deux volets : d’une part, un contrat type commun à tous les publics comportant les engagements réciproques : respecter les lois et les valeurs de la République et suivre la formation civique pour le nouvel arrivant et organiser l’accès aux droits individuels et à l’apprentissage de la langue pour l’État français; d’autre part, une annexe personnalisée faisant état de l’engagement à suivre, si nécessaire, une formation linguistique et/ou une formation supplémentaire à la connaissance de la vie en France et proposant, si nécessaire, un référent social.

3. Au sens de l’INSEE, un immigré est un étranger né à l’étranger et résidant en France.

4. Au sens de l’ONU, un immigré est une personne née dans un autre pays que celui où elle réside.

l'accès à la nationalité française.

Nous nous proposons dans ce travail d'évaluer l'impact de deux instruments de politique migratoire susceptibles de faciliter l'intégration des immigrés : les formations linguistiques récemment mises en place dans le cadre du Contrat d'Accueil et d'Intégration et la naturalisation en tant qu'étape dans le processus d'intégration. L'enjeu méthodologique est de taille car la naturalisation comme la formation linguistique ne sont pas des traitements aléatoires : les migrants naturalisés sont positivement sélectionnés sur des caractéristiques inobservables (au moins en partie) à l'économétricien alors que les migrants qui bénéficient de la formation linguistique sont négativement sélectionnés. Nous utilisons successivement une régression de discontinuité, une estimation de double différence et une estimation de panel pour neutraliser ces effets de sélection. Nous trouvons un léger impact de la formation linguistique sur le niveau en Français des migrants mais nous n'observons pas d'effets sur des résultats plus larges tels que la probabilité de sortie du chômage, le nombre de mois travaillés ou encore l'accès au logement. En estimant une équation de salaire à l'aide d'un modèle de panel à effets individuels fixes, nous ne trouvons pas non plus d'impact significatif de la naturalisation sur notre échantillon. La suite de l'étude est organisée de la façon suivante : la section 2 est une revue de la littérature. Les données sont présentées dans la section 3. La section 4 commence par évaluer l'impact des formations en français d'abord sur le niveau de français des migrants puis plus largement sur l'emploi et l'intégration en général. La section 5 s'intéresse à l'impact de la naturalisation. La section 6 conclut.

## 2 Revue de la littérature

En 2012, le taux de chômage de la population immigrée était de près de 7 points supérieur à celui de la population native (Insee, Enquête Emploi 2012). En 2007, ils étaient 36% à vivre sous le seuil de pauvreté contre 11% de la population non immigrée (Lombardo et Pujol, 2010 [18]). Ce désavantage économique prégnant ne résulte pas seulement d'un effet de sélection : Lombardo et Pujol montrent qu'à caractéristiques sociodémographiques identiques, les écarts de niveau de vie entre ménages immigrés et ménages non immigrés restent marqués. Face à ce constat, tout un pan de la littérature récente s'est attaché à comprendre et identifier les facteurs de ce déclassement des immigrés (Kazemipur et Halli, 2001 [16]; de Bustillo et Antón, 2011 [6]; Sullivan et Ziegert, 2008 [24]). Ont été passés en revue l'origine ethnique, la région de provenance, l'éducation antérieure, la durée de présence dans le pays d'accueil, la maîtrise de la langue, ainsi que des facteurs plus classiques et dont l'impact sur la pauvreté est reconnu : le niveau d'éducation ou encore la structure familiale.

Si le déclassement des populations immigrées est à présent communément admis et l'objet d'une vaste littérature, il n'en reste pas moins que peu d'études s'intéressent aux moyens qui peuvent aider les migrants à sortir de la pauvreté. Nous nous intéressons dans ce travail à l'effet de deux

outils : les formations linguistiques proposées à l'arrivée et la naturalisation en tant qu'instrument d'intégration. Nous commençons cette revue de la littérature par les travaux s'intéressant aux effets des formations linguistiques puis nous passerons aux travaux analysant l'impact de la naturalisation.

La maîtrise de la langue officielle joue un rôle majeur dans l'intégration des nouveaux migrants sur le marché du travail (Dustmann, 1994 [12]; Delander, Hammarstedt, Mansson et Nyberg, 2004 [7]; Zhen, 2013 [26]). On peut imaginer au moins deux canaux par lesquels le niveau linguistique améliore l'intégration économique des immigrants : premièrement, la maîtrise de la langue affecte directement la probabilité d'emploi en aidant les immigrants à trouver un travail en adéquation avec leurs qualifications antérieures et en leur permettant une communication plus aisée dans le monde du travail (Chiswick et Miller, 2003 [5]). Elle améliore ensuite indirectement la productivité des immigrants en accroissant le rendement économique du capital humain acquis à l'étranger (e.g. niveau d'éducation et expérience). Les pays d'accueil, eux, améliorent directement leur performance économique puisque les immigrants sont alors plus aptes à y travailler et plus adaptés à leur marché du travail.

Dans ce cadre, il est naturel de se demander quel serait l'effet d'une formation linguistique proposée aux migrants les plus fragiles à leur arrivée. De nombreux pays de l'Union Européenne, ainsi que le Canada ont mis en place de telles formations à l'intention des nouveaux migrants. En France, les signataires du CAI passent un test de langue et, suivant le niveau estimé, se voient proposer jusqu'à 400 heures de cours de français. La formation est prescrite à 20% des signataires du CAI. Pour autant, les formations linguistiques ne sont évidemment pas les uniques moyens de progresser dans la langue du pays d'accueil : les nouveaux arrivants progressent essentiellement en dialoguant avec leurs amis, voisins ou collègues de travail ou en suivant les médias (van Tubergen et Kalmijn, 2009 [25]). Tout l'enjeu est donc de savoir si la formation linguistique permet une progression plus rapide que ces moyens plus classiques et informels. A priori, en les exposant davantage à la langue de leur pays d'accueil, en les insérant dans une structure administrative et en leur permettant d'acquérir non seulement, des compétences indispensables au quotidien (savoir lire, écrire, accueillir des gens, faire ses courses...) mais aussi des compétences spécifiques au marché du travail (comment chercher un emploi, passer un entretien d'embauche dans le nouveau pays. . .), la formation doit donner les progrès les plus rapides. On peut aussi envisager ces cours comme des lieux de rencontre importants où les immigrants forment leurs premiers liens forts et durables avec leur nouvelle société. Enfin, on peut faire l'hypothèse que ces cours peuvent aussi permettre aux immigrants d'avoir une meilleure estime d'eux-mêmes. Les études sur les bénéfices de telles formations sont limitées et contradictoires. Une étude qualitative menée par Duff et al (2002 [11]) examine l'efficacité du Resident Care Home Support Attendant program en Colombie britannique au Canada, qui offre des cours d'anglais aux immigrants non anglophones. Les chercheurs montrent que deux ans après la fin de la formation, de nombreux participants peinent encore à trouver un emploi à temps complet. Ces résultats amènent à s'interroger sur l'efficacité de telles formations. Plus récemment,

Akresh (2007 [1]) montre que les cours d'anglais n'ont pas eu d'effets significatifs sur l'augmentation des revenus des nouveaux migrants aux Etats-Unis. D'autres études quantitatives amènent à ces mêmes conclusions. A contrario, dans leur évaluation d'un programme pilote de formation linguistique destinée aux immigrants sans emplois en Suède, Delander et al. (2005 [7]) concluent que les participants aux programmes trouvent un emploi plus rapidement que les non participants. Renaud et Cayn (2007 [23]), trouvent également que participer à une formation de Français a un impact positif sur l'accès au marché du travail des migrants dans les 18 mois suivant leur arrivée au Québec. Enfin, tout récemment, Lisa Kaida (2013 [15]) trouve, après soigneuse prise en compte de l'hétérogénéité inobservée et de la sélection des migrants dans ces formations, que la probabilité de sortie de la pauvreté croît significativement pour les migrants bénéficiaires. Tout l'enjeu de ces travaux tient donc en la méthodologie utilisée pour composer avec l'hétérogénéité inobservée et les biais de sélection qui résulte du fait que les migrants bénéficiant de formations sont non seulement les plus faibles dans la langue du pays d'accueil mais aussi et souvent, les plus faibles à tous points de vue (éducation, revenus, expérience...).

Les chercheurs en sciences sociales ont fait depuis longtemps de la naturalisation un objet d'analyse sociologique et politique alors que les économistes n'ont commencé à s'intéresser au sujet que tout récemment. Dans une des toutes premières contributions, Chiswick (1978 [4]) analyse l'intégration économique des immigrants à l'aune de la naturalisation. A partir de données en coupe issue du recensement américain de l'année 1970, il examine le processus d'intégration économique des migrants en comparant leurs revenus à ceux des natifs et montre qu'à caractéristiques socioéconomiques égales, les immigrés naturalisés touchent aux Etats-Unis des salaires en moyenne 15% plus élevés que les étrangers non naturalisés. Toutefois, il trouve que cet effet tombe à 7% et n'est plus significatif si l'on tient compte de la durée de séjour. Ces dernières années, de nombreux travaux se sont attelés à la question de la naturalisation. Cependant, la plupart d'entre eux analysent la situation des Etats-Unis ou du Canada (Bratsberg et al. (2002 [3]), DeVoretz et Pivnenko (2005 [9]), DeVoretz et Pivnenko (2005 [10]), DeVoretz (2008 [8]), Mazzolari (2007 [19])). Pour les pays européens, on ne trouve que quelques papiers empiriques sur l'impact économique des naturalisations à l'image du travail de Kogan (2003 [17]) pour l'Autriche et la Suède ou de Bevelander (2006 [2]) pour les Pays-Bas. La majeure partie de ces études se base sur des données et ne tiennent qu'improprement en compte des biais de sélection existant dans le processus de naturalisation. Bratsberg, Ragan et Nasir (2002 [3]) sont les premiers à utiliser des données de panel pour estimer l'impact de la naturalisation sur l'emploi et le salaire. Parce qu'elles comprennent plusieurs observations à divers moments, les données de panel permettent de contrôler l'effet des caractéristiques inobservables. Ainsi, Bratsberg et Ragan montrent que la naturalisation entraîne une hausse accrue du salaire et que cet effet est encore plus fort pour les immigrés originaires d'un pays en voie de développement. Enfin, plus récemment et sur des données françaises, Mirna Safi et Denis Fougère (2005 [13]) utilisent un probit bivarié pour neutraliser l'effet de sélection et montrent qu'après la



naturalisation, la probabilité que les immigrants trouvent un emploi augmente en moyenne de plus de 20 points de pourcentage. Ces résultats suggèrent que la naturalisation compense significativement l'ampleur de certaines pratiques discriminatoires sur le marché du travail en France. Les auteurs voient leur résultat comme la première preuve empirique, pour la France, de l'efficacité de la naturalisation comme instrument politique d'intégration des immigrants sur le marché du travail.

On le voit, l'étude de l'effet des formations linguistiques comme celui de l'impact de la naturalisation revêt un enjeu méthodologique majeur : la prise en compte de la sélection des migrants au sein de ces traitements. En effet, les migrants naturalisés comme ceux bénéficiaires d'une formation linguistique ne sont pas un échantillon aléatoire de la population immigrée dans sa totalité. A priori, les migrants bénéficiaires d'une formation sont négativement sélectionnés parmi la population de migrants et ceux naturalisés sont positivement sélectionnés. La majeure partie des travaux cités précédemment tente de contrôler ces biais. Notre travail se distingue de deux manières :

- Premièrement, nous tentons d'évaluer l'impact de la formation linguistique de manière tout à fait originale : nous nous servons d'un indicateur construit du niveau de français afin de retrouver le seuil en deçà duquel les migrants se sont vus prescrire la formation linguistique. Dès lors, il devient possible de mener une régression de discontinuité afin d'estimer l'effet de la formations sur les « formés ».

- Deuxièmement, la qualité des données nous permet d'estimer pour la première fois en France l'effet de la naturalisation avec interprétation causale. Soulignons d'ailleurs qu'une des seules estimations de l'effet de la naturalisation sur données de panel (Bratsberg et al. 2002 [3]) repose sur un panel de 304 immigrants dont 34 naturalisés. Nous disposons avec ELIPA de plus de 100 naturalisations (sur un total de plus de 3500 migrants) survenues depuis la première interrogation.

### 3 Présentation des données

Les données utilisées sont issues de l'enquête ELIPA. Cette dernière, menée auprès de 6107 nouveaux migrants en première interrogation, possède deux avantages :

- Premièrement, il s'agit d'une enquête longitudinale qui va donc nous permettre de créer un panel nous permettant d'estimer proprement l'effet de la naturalisation, en neutralisant l'effet des caractéristiques inobservables. Nous revenons en détails sur ces problèmes dans la partie 5.

- Deuxièmement, il s'agit d'une enquête extrêmement récente, les données utilisées pour la dernière vague ont été récoltées en 2013. En ce sens, les résultats obtenus sont des résultats originaux.

TABLE 1 – Statistiques descriptives

	Vague 1				Vague 2		Vague 3		Formés	
	N	%	dt formés (%)	Indic	N	%	N	%	N	%
<b>Sexe</b>										
Homme	3045	49,9	10,15	0,03	2352	49,5	1754	49,1	309	42,5
Femme	3062	50,1	13,65	-0,03	2404	50,5	1819	50,9	418	57,5
Total	6107	100	11,9	0	4756	100	3573	100	727	100
	Mean	-	-		Mean		Mean		Mean	
<b>Niveau de vie par uc</b>	973,74 (613,19)	-	-		1092,11 (634,61)		1182,11 (670,78)		856,69 (518,17)	
<b>Age</b>	33,28 (8,97)	-	-		34,55 (8,99)		36,86 (8,83)		34,05 (10)	
<b>Durée de présence en France</b>	48,83 (54,44)	-	-		60,21 (54,9)		73,03 (55,67)		30,97 (39,96)	
<b>Age à l'arrivée</b>	29,01 (8,98)	-	-		29,29 (9,03)		29,53 (8,74)		31,23 (9,4)	
<b>Nombre de mois travaillés</b>	-	-	-		6,42 (5,41)		7,47 (5,07)		-	
	N	%	dt formés (%)	Indic	N	%	N	%	N	%
<b>Nationalité</b>										
Europe (hors France)	71	1,2	15,49	-0,17	62	1,3	48	1,3	11	1,5
Russie et CEI	327	5,4	29,05	-0,37	275	5,8	214	6	95	13,1
Turquie	365	6	52,05	-1,18	253	5,3	163	4,6	190	26,1
Chine	269	4,4	26,39	-0,63	196	4,1	155	4,3	71	9,8
Autre Asie	516	8,4	39,73	-0,93	395	8,3	273	7,6	205	28,2
Algérie	1167	19,1	4,63	0,38	913	19,2	685	19,2	54	7,4
Maroc et Tunisie	835	13,7	3,71	-0,03	641	13,5	461	12,9	31	4,3
Afrique subsaharienne	1700	27,8	1	0,29	1354	28,5	1069	29,9	17	2,3
Autre Afrique	462	7,6	7,36	0,24	364	7,7	286	8	34	4,7
Amérique et Océanie	395	6,5	4,81	0,34	303	6,4	219	6,1	19	2,6
Total	6107	100	11,9	0	4756	100	3573	100	727	100
<b>Diplôme le plus élevé obtenu</b>										
Jamais scolarisé	321	5,3	12,46	-0,7	241	5,1	179	5	40	5,5
Aucun diplôme	1451	23,9	12,13	-0,15	1124	23,8	845	23,8	176	24,3
CEP	330	5,4	16,97	-0,16	247	5,2	176	5	56	7,7
Brevet des collèges	709	11,7	14,81	-0,02	553	11,7	405	11,4	105	14,5
CAP, BEP	666	11	6,76	0,22	530	11,2	409	11,5	45	6,2
Bac pro/techno	387	6,4	17,57	-0,02	308	6,5	245	6,9	68	9,4
Bac général	682	11,3	13,64	0,08	537	11,4	392	11	93	12,9
1er cycle universitaire	538	8,9	9,85	0,2	426	9	338	9,5	53	7,3
2e ou 3e cycle universitaire	976	16,1	8,91	0,21	758	16	562	15,8	87	12
Total	6060	100	11,9	0	4724	100	3551	100	723	100
<b>Motif d'admission au séjour</b>										
Migration de travail	729	11,9	4,12	0,08	549	11,5	448	12,5	30	4,1
Famille	4228	69,2	11	0,06	3315	69,7	2495	69,8	465	64
Réfugiés et apatrides	850	13,9	26,59	-0,45	667	14	472	13,2	226	31,1
Autres	300	4,9	2	0,23	225	4,7	158	4,4	6	0,8
Total	6107	100	11,9	0	4756	100	3573	100	727	100
<b>Activité de l'enquêté</b>										
En emploi	2869	47	8,71	0,06	2663	56,1	2270	63,6	250	34,4
Etudiant	511	8,4	9,39	0,24	338	7,1	136	3,8	48	6,6
Au chômage	1620	26,5	9,81	0,09	914	19,2	616	17,3	159	21,9
Au foyer	883	14,5	25,14	-0,42	678	14,3	448	12,5	222	30,5
Autres inactifs	220	3,6	21,82	-0,36	156	3,3	100	2,8	48	6,6
Total	6103	100	11,9	0	4749	100	3570	100	727	100
<b>Sous seuil de pauvreté</b>										
Non	2722	53,3	10,36	0,04	2704	63,1	2274	68,6	282	45,4
Oui	2388	46,7	14,2	-0,05	1579	36,9	1041	31,4	339	54,6
Total	5110	100	11,9	0	4283	100	3315	100	621	100
<b>Désir de s'installer en France</b>										
Non	1724	28,8	10,44	0,07	1055	22,6	699	19,9	180	25,4
Oui	4268	71,2	12,42	-0,03	3616	77,4	2819	80,1	530	74,6
Total	5992	100	11,85	0	4671	100	3518	100	710	100
<b>Niveau de français autodéclaré</b>										
Très mauvais	356	5,8	63,2	-1,67	159	3,3	137	3,8	225	30,9
Plutôt mauvais	1364	22,3	28,01	-0,69	834	17,5	562	15,7	382	52,5
Plutôt bon	2600	42,6	4,38	0,21	2072	43,6	1733	48,5	114	15,7
Très bon	1787	29,3	0,34	0,51	1691	35,6	1141	31,9	6	0,8

Total	6107	100	11,9	0	4756	100	3573	100	727	100
<b>Type de logement</b>										
Logement d'appoint	1818	29,8	11,83	-0,07	1066	22,5	589	16,5	215	29,7
Propriétaire ou bailleur	4277	70,2	11,92	0,03	3673	77,5	2970	83,5	510	70,3
Total	6095	100	11,9	0	4739	100	3559	100	725	100

Nous commençons par une analyse descriptive de la base de donnée en insistant sur les caractéristiques spécifiques des migrants qui se sont vus prescrire la formation linguistique (nous les appellerons par la suite « les formés »). La table 1 donne la composition de l'échantillon à chaque vague d'interrogation ainsi que la composition des formés. Les colonnes 3 et 4 donnent, pour chaque sous-groupes de population le pourcentage de formés ainsi que leur niveau en français résumé par un indicateur centré réduit. Nous reviendrons sur la construction de cet indicateur dans la suite du travail (voir en particulier l'annexe A) mais nous l'introduisons d'ores et déjà. Il va nous permettre de répondre à la question suivante : existe-il un effet de sélection allant au-delà de la simple prescription de la formation aux migrants les plus faibles en français ? Globalement, les migrants obtenant leur 1er titre de séjour en 2009 viennent majoritairement du continent africain (environ 70%), sont près de 45% à vivre sous le seuil de pauvreté (contre moins de 13% de la population native), sont relativement jeunes (33 ans de moyenne d'âge), venus à près de 70% pour motifs familiaux et résident en France depuis à peu près 4 ans en moyenne. Les bénéficiaires de la formation linguistique sont une population plus fragile encore à tous points de vue : ils sont plus pauvres (plus de 100 euros de différence de niveau de vie par unité de consommation entre les formés et l'ensemble de la population), moins diplômés en moyenne (ils sont environ 32% à posséder le baccalauréat au minimum contre plus de 36% de la population totale). Les réfugiés et apatrides sont surreprésentés parmi les formés (ils représentent 14% de la population immigrée totale mais plus de 31% des formés). De la même manière les individus « au foyer » constitue 30% des formés contre seulement 15% de la population de départ. Comme on pouvait s'y attendre les nationalités majoritairement francophones sont sous-représentées parmi les formés (les algériens constituent 20% de la population récemment immigrée mais seulement 7% des formés). Il est intéressant de noter que le taux de prescription de la formation n'est pas parfaitement indexé sur notre indicateur du niveau de français : ainsi les migrants originaires d'Europe de l'Est bien qu'ayant un niveau de français en moyenne légèrement supérieur à celui des migrants chinois (indicateur centré réduit à -0,37 vs -0,63) se sont davantage vus prescrire la formation. Ceci est un premier indice de la forte hétérogénéité inobservée qui caractérise la population des formés : il ne s'agit pas seulement de la population au niveau de français le plus faible mais également de la population la plus fragile à tous points de vue.

## 4 Impact de la formation en français

La formation linguistique a été prescrite sur la base d'un test que tous les migrants ont été dans l'obligation de passer le jour de leur signature du CAI. S'ils obtenaient un score supérieur à un certain seuil prédéterminé par arrêté ministériel, ils se voyaient prescrire une formation allant de 100h à 400h. Si nous disposions de la variable (dite variable d'assignation) donnant le score à ce test obtenu par chacun des migrants, nous aurions pu mener une régression de discontinuité afin d'identifier proprement l'effet de la formation sur divers résultats avec, en premier lieu le niveau de français à l'issue de la formation<sup>5</sup>.

Or, nous n'avons malheureusement pas réussi à obtenir le résultat de ce test sous forme d'une variable continue. Notre objectif premier et original était donc de mener une régression de discontinuité sur la base de cette variable. L'hypothèse sur laquelle se base la régression de discontinuité est que les individus obtenant un score proche du seuil possèdent des caractéristiques inobservables similaires. Comparer les résultats obtenus par les individus se situant juste au-dessus ou juste en dessous de ce seuil permet donc de neutraliser, au moins en partie, l'effet de sélection : les individus bénéficiaires de la formation (donc dont le score obtenu se situe juste en dessous du seuil) sont considérés comme un échantillon aléatoire tiré sur la population de migrants obtenant un score autour du seuil. Si la formation a effectivement un impact, nous espérons observer un saut, au niveau du seuil, dans l'outcome étudié (niveau en français après 1 an par exemple).

En l'absence de cette variable, nous avons fait le choix de construire un indicateur continu du niveau de français et de mener une régression de discontinuité floue. Le mode de construction de cet indicateur est présenté en annexe A. Il se base sur 3 manières différentes d'observer le niveau de français des migrants à leur arrivée :

- Le score obtenu à une partie des tests de l'OFII en trois modalités
- L'auto-évaluation du niveau de français par le migrant
- Le résultat au test de français proposé dans le questionnaire ELIPA

La régression de discontinuité floue se rapproche très fortement de la régression de discontinuité à ceci près que le seuil dans la variable d'assignation (notre indicateur du niveau de français) ne détermine plus précisément le traitement ( $T = 1$  si indicateur de niveau en français  $>$  seuil  $s$ ) mais est plus modestement fortement corrélé au traitement. On utilise ensuite cette variable d'assignation en tant que variable instrumentale de la participation à la formation.

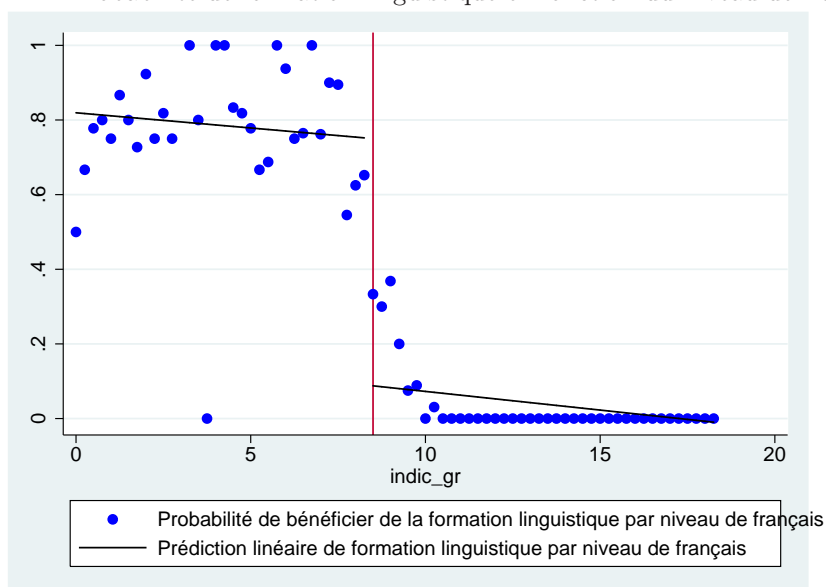
La figure 1 présente la probabilité de traitement (probabilité de bénéficier ou non de la formation) distribuée en fonction de notre indicateur de français. Pour que l'indicateur soit exploitable, il faut que nous observions une discontinuité nette dans la probabilité de traitement en fonction de

---

5. L'enquête ELIPA comprend un volet « Compréhension de la langue française » qui consiste en un test écrit et oral, identique au moins pour les vague 1 et 2. En attribuant 1 point par bonne réponse, nous avons obtenu un score rapporté sur 20 pour chacune des 2 vagues.

la variable d'assignation.

FIGURE 1 – Probabilité de formation linguistique en fonction du niveau de français



Nous observons effectivement une discontinuité : la probabilité de bénéficier d'une formation linguistique passe d'environ 0.8 à presque 0 lorsque l'individu possède un niveau de français supérieur à 8,5 selon notre indicateur. Afin de tester la significativité de ce seuil nous menons la régression suivante :

$$formation_i = \alpha + \beta_1 \cdot above8 + \beta_2 gapabove + \beta_3 gapbelow$$

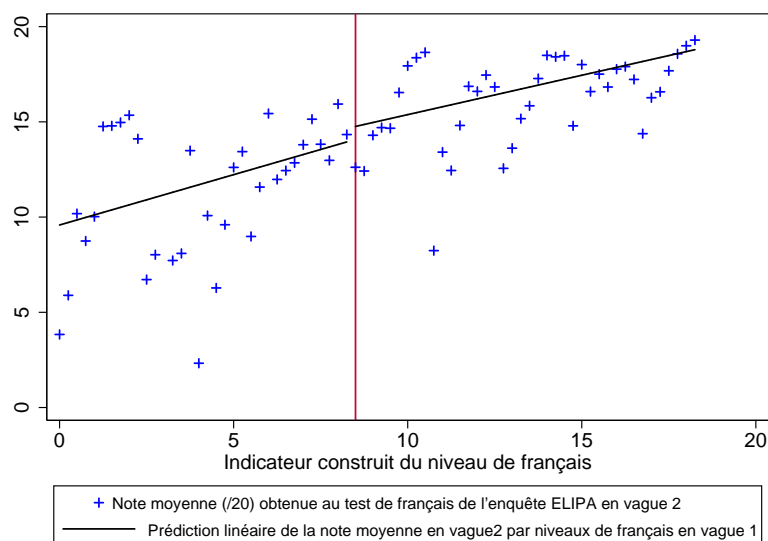
Où *formation* est une dummy égale à 1 si l'individu suit la formation linguistique et à 0 sinon, *above8* est une dummy égale à 1 si l'individu possède un indicateur du niveau de français supérieur à 8,5, *gapabove* est égal, pour chaque individu, à l'éloignement du seuil (indicateur en français de l'individu moins 8,5) si l'individu se situe au-dessus du seuil et *gapbelow* est égal à l'éloignement du seuil si l'individu se situe en dessous ce seuil. Le coefficient de *above8* est significatif à 0,1% et indique un saut d'environ -0,66 dans la probabilité de bénéficier de la formation au passage du seuil. Nous passons donc à la régression de discontinuité proprement dite. Nous cherchons l'impact de la formation linguistique sur la note obtenue au test linguistique ELIPA en vague 2. Nous procéderons de deux manières : d'abord par la stratégie classique qui repose sur l'instrumentalisation de *formation* par *gapabove*, *gapbelow* et *above8* et ensuite en utilisant la commande *rd* sous Stata créée par Austin Nichols (2012 [20]).

Les résultats obtenus sont présentés dans la table ci-après.

	Rd command	IV strategy
Impact de la formation	-1.02	-1.75
$p$ - value	0.351	0.149

Nous trouvons un coefficient négatif : les formés obtiennent, à l'issue de la formation un score de 1 à 2 points inférieur au score des non formés. L'instabilité des résultats ainsi que leur peu de significativité tend à nous convaincre de la faiblesse de l'impact de la formation sur le niveau de français tel que mesuré par l'enquête ELIPA en vague 2. Par ailleurs, les coefficients sont de signes opposés à celui attendu. Nous traçons le graphique de la régression de discontinuité (figure 2) afin d'avoir une meilleure vision de la situation.

FIGURE 2 – Régression de discontinuité



On voit qu'effectivement nous n'observons pas de rupture de trend au niveau du seuil autour du lequel la probabilité de formation chute.

Pour compléter l'analyse, nous décidons de mener des régressions différence de différence. Là encore, la double différence doit permettre de neutraliser l'effet de sélection à l'intérieur de la formation. L'idée est la suivante : soit  $T$  la variable de traitement. On a  $T = 1$  lorsque l'individu se voit prescrire la formation. Notons  $Y$  le niveau en français mesuré par le test proposé par ELIPA. Indiquons par 1 le niveau en français  $Y$  lorsque celui-ci est observé après la formation et par 0 lorsqu'il est observé avant (ou sans formation). Ainsi  $Y_1|T_1 = 1$  fait référence au niveau de français

du groupe sélectionné pour la formation ( $T_1 = 1$ ) après qu'ils aient reçu la formation ( $Y_1$ ). L'effet du traitement sur les traités correspond à la différence entre le niveau de français moyen avec traitement  $E(Y_1|T = 1)$ , qu'il faut comparer avec le niveau de français qu'ils auraient eu s'ils n'avaient pas été traités  $E(Y_0|T = 1)$ . Si le premier est connu, le deuxième ne l'est évidemment pas, et il nous faut trouver un estimateur crédible de ce niveau de français potentiel contrefactuel.

L'estimateur le plus simple, qu'on peut qualifier de "naïf", de l'effet moyen du traitement consiste à comparer le niveau de français des bénéficiaires  $E(Y_1|T = 1)$  avec celui des non-bénéficiaires  $E(Y_0|T = 0)$ . Or dans notre cas, les bénéficiaires ne le sont pas par hasard, ils ont été affectés à la formation précisément à cause de leur faible niveau en français. Les deux groupes ne sont donc pas comparables, l'estimateur naïf risque de sous-estimer grandement l'impact réel de la formation. L'estimateur naïf s'écrit :

$$\begin{aligned} E(Y|T = 1) - E(Y|T = 0) &= E(Y_1|T = 1) - E(Y_0|T = 0) \\ &= \underbrace{E(Y_1|T = 1) - E(Y_0|T = 1)}_{\text{effet que l'on cherche à estimer}} + \underbrace{E(Y_0|T = 1) - E(Y_0|T = 0)}_{\text{biais}} \end{aligned}$$

Entre l'estimateur naïf et ce que l'on cherche réellement à estimer (i.e. la différence qu'on aurait observée si les traités n'avaient pas bénéficié de la formation), il y a un biais égal à  $E(Y_0|T = 1) - E(Y_0|T = 0)$ . Ce biais correspond donc à la différence de capacité linguistique moyenne qui serait observée entre le groupe des traités ( $T = 1$ ) et le groupe des non-traités ( $T = 0$ ), si aucun des deux n'avait bénéficié de la formation ( $Y_0$ ). Ce biais n'est pas nul dès lors que la sélection au sein du traitement n'est pas aléatoire, ce qui est notre cas.

Une seconde approche pourrait consister à retrancher le niveau en français des formés avant la formation de leur niveau de formation après la formation cette méthode est également biaisée car elle ignore les effets de la dynamique de l'apprentissage de la langue.

L'idée de la différence de différence est de combiner ces deux approches. On exploite les données sur le groupe traité et sur le groupe de contrôle avant (date  $t_0$ ) et après (date  $t_1$ ) le traitement. On commence par retrancher le niveau de français avant la formation du niveau après pour chacun des groupes, ce qui permet d'éliminer les effets individuels. Puis on fait la différence des deux résultats obtenus, ce qui permet d'éliminer les effets temporels. Les tables 2 et 3 présentent les résultats pour le niveau de français, mesuré de deux manières différentes.

TABLE 2 – Difference in differences - Niveau de français vague 1 à 2

Outcome Variable	Vague 1			Vague 2			DIFF-IN-DIFF
	Control	Treated	Diff(BL)	Control	Treated	Diff(FU)	
Résultats au test de français ELIPA	17,556	10,166	-7,389	17,571	12,298	-5,273	2,116
Part d'individus déclarant parler bien ou très bien français	0,000	0,000	0,000***	0,000	0,000	0,000***	0,000***
	0,797	0,153	-0,644	0,854	0,25	-0,604	0,039
	0,000	0,000	0,000***	0,000	0,000	0,000***	0,195

\* Means and Standard Errors are estimated by linear regression  
 \*\*Inference : \*\*\* p<0,01 ; \*\* p<0,05 ; \* p<0,1

TABLE 3 – Difference in differences - Niveau de français vague 1 à 3

Outcome Variable	Vague 1			Vague 3			DIFF-IN-DIFF
	Control	Treated	Diff(BL)	Control	Treated	Diff(FU)	
Résultats au test de français ELIPA	17,556	10,166	-7,389	12,847	5,877	-6,97	0,419
Part d'individus déclarant parler bien ou très bien français	0,000	0,000	0,000***	0,000	0,000	0,000***	0,248
	0,797	0,153	-0,644	0,857	0,324	-0,533	0,111
	0,000	0,000	0,000***	0,000	0,000	0,000***	0,000***

\* Means and Standard Errors are estimated by linear regression  
 \*\*Inference : \*\*\* p<0,01 ; \*\* p<0,05 ; \* p<0,1

L'estimateur de double différence entre les résultats en vague 1 et en vague 2 est positif et significatif pour le test de français ELIPA. Ce test de français a été proposé de manière identique en vague 1 puis 1 an après en vague 2. Il est donc particulièrement utile pour mesurer les progrès réalisés en français. Nous trouvons un effet de la formation de plus 2,1 points en un an sur ce test. La proportion d'individus, se déclarant bons ou très bon en français, elle n'est pas suffisamment impactée par la formation et l'estimateur de double différence n'est pas significatif. Par contre il



l'est pour les résultats entre la vague 1 à 3 : la différence de proportion entre les formés et les non-formés était de 64 points de pourcentage en vague 1 (80% des non formés déclaraient posséder un bon ou très bon niveau de français contre seulement 15% des formés) ; elle tombe à 53 points en vague 3, soit un estimateur diff-in-diff de 0,11 significatif à 0,01%. Toutefois ces résultats sont à nuancer : en effet l'estimation de double-différence repose sur l'hypothèse que le trend temporel entre les 2 estimations est commun. Or il fort probable que ce ne soit pas le cas. Les non-formés disposant déjà d'un très bon niveau en français tendent à ne plus progresser beaucoup, quand la marge de progression des formés est naturellement beaucoup plus importante. L'interprétation en termes de causalité est donc à prendre avec précautions.

Explorons à présent plus avant l'effet de la formation sur des résultats plus larges que le niveau de français : la table 4 donne l'estimateur de double différence pour le nombre de mois travaillés, le niveau de vie par unité de consommation, la proportion de migrants désirant s'installer en France, la probabilité de se faire de nouveaux amis entre les deux vagues, la probabilité de sortie du chômage et un indicateur d'intégration dont le calcul est détaillé en annexes (voir l'annexe B).

TABLE 4 – Difference in differences - Autres outcomes

Outcome Variable	Vague 1			Vague 3			DIFF-IN-DIFF
	Control	Treated	Diff(BL)	Control	Treated	Diff(FU)	
Nombre de mois travaillés	6,79 0,000	3,989 0,000	-2,801 0,000***	7,71 0,000	5,257 0,000	-2,453 0,000***	0,348 0,475
Niveau de vie par uc	987,024 0,000	824,546 0,000	-162,478 0,000***	1196,509 0,000	1065,335 0,000	-131,174 0,003***	31,303 0,62
Désir d'installation en France	0,714 0,000	0,78 0,000	0,066 0,020**	0,798 0,000	0,828 0,000	0,03 0,284	-0,035 0,377
Probalité de nouveaux amis d'origines diverses	0,761 0,000	0,477 0,000	-0,283 0,000***	0,723 0,000	0,457 0,000	-0,266 0,000***	0,018 0,616
Probabilité de sortie du chômage	0,000 1	0,000 1	0,000 1	0,465 0,000	0,343 0,000	-0,122 0,000***	-0,122 0,002***
Indicateur d'intégration	7,482 0,000	5,67 0,000	-1,812 0,000***	7,482 0,000	5,71 0,000	-1,772 0,000***	0,04 0,785

\* Means and Standard Errors are estimated by linear regression  
 \*\*Inference : \*\*\* p<0,01 ; \*\* p<0,05 ; \* p<0,1

Aucun des coefficients obtenus n'est significatif. Nous ne trouvons donc pas d'impact de la formation linguistique au-delà d'un léger (voire contestable) effet sur le niveau de français 1 ans après la formation.

Nous pouvons avancer deux raisons pour expliquer la faiblesse des résultats obtenus sur l'effet de la formation linguistique :

- Premièrement, il se peut que l'hétérogénéité inobservée soit si forte qu'il est quasiment impossible de l'éliminer par des méthodes économétriques. En effet, il semble que la formation ait été prescrite non seulement en fonction du niveau en français à l'arrivée mais également en fonction de caractéristiques plus abstraites : en particulier, à niveaux de français égal, la formation a été significativement plus prescrite aux inactifs et aux migrants originaires d'Asie (65% des formés sont asiatiques contre moins de 20% des migrants dans leur ensemble), deux populations qui diffèrent de l'ensemble des migrants par des caractéristiques qui vont largement au-delà du simple niveau en français.

- Deuxièmement, il est également probable que la formation en elle-même manque d'ambition : que ce soit pour apprendre à parler ou à écrire le français, les nouveaux migrants jugent sévèrement la formation linguistique : ils sont 70% à penser qu'elle était insuffisante et autant à souhaiter un prolongement de celle-ci.

## 5 Impact de la naturalisation

Dans cette partie, nous tentons de répondre à la question suivante : la naturalisation est-elle un instrument au service de l'intégration ou plutôt le point final d'un parcours d'intégration réussi ?

Nous pouvons imaginer deux types d'effets de la naturalisation sur l'intégration des immigrés sur le marché du travail. D'une part, la naturalisation permet aux immigrés d'accéder à des professions qui leur étaient auparavant fermées (notamment les emplois de la Fonction publique, mais aussi de nombreuses professions libérales ou indépendantes). D'autre part, elle lève certains obstacles discriminatoires au cours de la procédure de recrutement. Comme nous l'avons expliqué en introduction, la mesure de l'effet propre de la naturalisation est cependant difficile à mettre en place : les immigrés qui acquièrent la nationalité française ne sont pas un échantillon aléatoire, pris au hasard dans la population des immigrés vivant sur le sol français. La naturalisation est un choix, le résultat d'un processus d'autosélection parmi les immigrés et pour cette raison les naturalisés diffèrent des autres par des caractéristiques observables (le niveau d'éducation par exemple), mais certainement aussi par d'autres caractéristiques, celles-ci inobservables. Or ces caractéristiques affectent également leurs possibilités d'accéder à l'emploi, ce dont il faudra tenir compte. Enfin, il ne faut pas oublier que les naturalisés diffèrent également du reste de la population, ne serait-ce que par la batterie de conditions à remplir pour pouvoir faire une demande de naturalisation (durée de résidence dans le pays, niveau de langue, test de citoyenneté...).

Avant de commencer l'analyse, nous examinons les caractéristiques spécifiques des migrants naturalisés en vague d'interrogation 3.

TABLE 5 – Naturalisation : statistiques descriptives

	N	%	Niveau de vie par uc	Durée de présence en France (mois)	Indicateur niveau de français	Indicateur intégration	% de diplômés bac ou +	Taux d'activité
Naturalisés en vague 3	3485	97,26	1439 (857)	82 (53)	0,45 (0,69)	0,30 (0,83)	46%	64%
Non naturalisés	98	2,74	1176 (664)	73 (56)	0,03 (0,96)	0 (1,0)	36%	64%

La table 5 montre bien que les migrants naturalisés sont positivement sélectionnés à tous points de vue : ils ont un niveau de vie par unité de consommation supérieur de près de 300 euros au niveau de vie des non-naturalisés, sont meilleurs en français, mieux « intégrés », plus diplômés en moyenne (46% sont titulaires du bac général ou technologique ou d'un diplôme du supérieur contre 36% des non-naturalisés) et résident en France depuis en moyenne plus longtemps (6 ans et 10 mois contre 6 ans pour les non-naturalisés). Les naturalisés sont donc une population bien spécifique susceptible de diverger des non-naturalisés non seulement par ces caractéristiques observables mais également par un grand nombre de « caractéristiques inobservables ».

Une manière de neutraliser l'impact de ces caractéristiques inobservables consiste à estimer un modèle de panel. L'observation répétée dans le temps d'un même individu permet précisément de capter ces caractéristiques inobservables ou, au moins, celles qui ne varient pas dans le temps. Nous commençons l'analyse par une simple estimation MCO sur données empilées. Ici nous ne contrôlons donc pas l'hétérogénéité inobservable car les observations sont « empilées » sans que l'on ne tienne compte du fait que plusieurs observations concernent un même individu. L'équation de base du modèle est la suivante :

$$\ln w_{it} = \alpha_0 + \alpha_1 N_{it} + \alpha_2 H_{it} + \alpha_3 M_{it} + \alpha_4 Z_{it} + \varepsilon_{it}$$

Où  $\ln w_{it}$  donne le log des ressources mensuelles de l'individu  $i$  au moment de la vague d'interrogation  $t$  ( $t = 1, 2, 3$ ). La naturalisation est captée par le terme  $N_{it}$  qui est une dummy égale à 1 si l'individu s'est fait naturalisé durant l'année précédant la vague d'interrogation et qui reste égale à 1 ensuite. Cette variable est notre variable d'intérêt, c'est elle qui capture la « prime de naturalisation ».  $H_{it}$  comprend deux variables décrivant le capital humain de l'individu (une dummy égale à 1 s'il parle couramment ou presque couramment l'anglais et une dummy égale à 1 s'il est diplômé de l'enseignement supérieur).  $M_{it}$  est un ensemble de variables résumant son parcours migratoire (âge d'arrivée, des indicatrices pour le motif de migration, pour la nationalité (en 10 postes), une variable continue donnant la durée de présence en France en mois et enfin une dummy égale à 1 si l'individu se déclare à l'aise à très à l'aise en langue française). Enfin,  $Z_{it}$  est un ensemble de

variables socio-démographiques (sexe, âge, âge au carré, statut matrimonial, nombre d'enfants, lieu de vie durant l'enfance, activité). Les résultats sont présentés dans la table 6.

TABLE 6 – Effet de la naturalisation sur les salaires : estimation d'un modèle empilé

	Model A	Model B	Model C	Model D	Model E
Naturalisation	0,259*** (4,63)	0,248*** (4,40)	0,238*** (4,25)	0,190*** (3,36)	0,148 (1,11)
Sexe	0,0723*** (4,73)	0,0766*** (4,98)	0,0784*** (5,11)	0,0707*** (4,59)	0,100*** (5,65)
Age (différence de millésime)	-0,0192*** (-3,59)	-0,0140* (-2,51)	-0,0140* (-2,51)	0,0923*** (8,50)	0,0217 (0,79)
Age	0,000184** (2,76)	0,000181** (2,72)	0,000178** (2,66)	0,000189** (2,83)	0,000182* (2,33)
Situation matrimoniale	0,411*** (20,71)	0,423*** (20,87)	0,425*** (21,05)	0,423*** (21,01)	0,429*** (18,13)
Nombre d'enfants	0,0155* (2,44)	0,0116 (1,76)	0,0114 (1,73)	0,00793 (1,20)	0,000106 (0,01)
Diplômé du supérieur	0,0205*** (6,33)	0,0218*** (6,71)	0,0203*** (6,28)	0,0188*** (5,81)	0,0177*** (4,97)
Actif	0,355*** (25,04)	0,351*** (24,15)	0,347*** (23,96)	0,341*** (23,81)	0,323*** (18,58)
Anglais parlé	0,0628*** (6,58)	0,0630*** (6,59)	0,0598*** (6,23)	0,0577*** (6,00)	0,0541*** (5,01)
Migration de travail (ref.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
Famille	-0,0386 (-1,71)	-0,0220 (-0,92)	-0,0278 (-1,17)	-0,0460 (-1,92)	-0,0724* (-2,57)
Réfugiés et apatrides	-0,241*** (-8,58)	-0,218*** (-7,32)	-0,218*** (-7,35)	-0,238*** (-8,00)	-0,272*** (-8,31)
Autres	-0,0853* (-2,27)	-0,102** (-2,71)	-0,107** (-2,84)	-0,102** (-2,68)	-0,0745 (-1,67)
Europe	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
Russie et CEI	-0,128 (-1,79)	-0,128 (-1,77)	-0,119 (-1,65)	-0,110 (-1,53)	-0,119 (-1,70)
Turquie	-0,0649 (-0,92)	-0,0675 (-0,95)	-0,0483 (-0,68)	-0,0492 (-0,69)	-0,0211 (-0,30)
Chine	0,0567 (0,83)	0,0521 (0,75)	0,0751 (1,09)	0,0870 (1,25)	0,106 (1,53)
Autre Asie	-0,129 (-1,90)	-0,127 (-1,86)	-0,109 (-1,59)	-0,102 (-1,48)	-0,0723 (-1,08)
Algérie	-0,231*** (-3,59)	-0,228*** (-3,50)	-0,242*** (-3,71)	-0,236*** (-3,62)	-0,272*** (-4,32)
Maroc et Tunisie	-0,167* (-2,53)	-0,166* (-2,48)	-0,175** (-2,62)	-0,170* (-2,54)	-0,211** (-3,20)
Afrique subsaharienne	-0,141* (-2,20)	-0,139* (-2,15)	-0,152* (-2,34)	-0,147* (-2,26)	-0,192** (-3,08)
Autre Afrique	-0,171* (-2,53)	-0,172* (-2,52)	-0,176** (-2,59)	-0,168* (-2,46)	-0,197** (-2,94)
Amérique et Océanie	0,0249 (0,36)	0,0265 (0,38)	0,0209 (0,30)	0,0265 (0,38)	-0,0282 (-0,41)
Lieu de vie durant l'enfance	0,00321 (0,37)	0,00159 (0,18)	0,00390 (0,45)	0,00395 (0,46)	0,0130 (1,35)

Age d'arrivée		-0,00556** (-2,82)	-0,00482* (-2,46)	-0,112*** (-11,65)	-0,0401 (-1,49)
Niveau de français			0,0728*** (4,36)	0,0712*** (4,29)	0,0381* (2,05)
Durée de présence en France				-0,00920*** (-10,79)	-0,00319 (-1,42)
Intégration					0,0385*** (8,73)
Constant	7,245*** (57,39)	7,208*** (56,93)	7,152*** (56,55)	7,165*** (56,53)	6,910*** (49,99)
Observations	9464	9437	9437	9437	6177

*t* statistics in parentheses

\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

Le coefficient de la variable d'intérêt  $N_{it}$  est significatif et possède le signe attendu. Les résultats indiquent que la naturalisation est corrélée avec une hausse des revenus d'environ 20% (puisque nous sommes dans un modèle log-niveau). Toutefois, et comme nous l'avons déjà dit, ce simple modèle MCO ne tient pas compte des biais de sélection. Il se peut que les naturalisés de notre échantillon, (qui touchent donc un salaire près de 20% supérieur à celui des non-naturalisés), diffèrent également de ces derniers par des caractéristiques qui ne sont pas prises en compte par notre modèle : la motivation, l'habileté, la vivacité, la confiance en soi, la témérité, la détermination voire même encore l'apparence physique. Ces caractéristiques inobservables par l'économétricien, sont susceptibles de jouer à la fois sur le salaire et sur l'accès à la naturalisation. La « prime de naturalisation » pourrait donc en fait être la conséquence de ces caractéristiques inobservables, sans que la naturalisation à proprement parler n'y soit pour quelque chose. L'addition progressive de contrôles confirme cette hypothèse : la taille du coefficient diminue à chaque nouveau contrôle. Surtout, l'addition de notre dernier régresseur, notre indicateur du niveau d'intégration construit à partir d'un ensemble de questions de l'enquête ELIPA (voir l'annexe B) réduit le coefficient de naturalisation de près de 25% et celui-ci devient non-significatif. Cette variable « intégration », bien que maladroitement construite à partir des seules questions posées à la fois en vague 1 et vague 2, semble la plus à même de précisément capter une partie au moins de l'hétérogénéité inobservable. Le fait qu'elle réduise autant le coefficient de la variable « naturalisation » nous informe bien sur le caractère biaisé du coefficient : il devient alors évident que le large effet de la naturalisation sur les niveaux de salaire est essentiellement une conséquence de l'effet de caractéristiques inobservables.

Pour confirmer cette hypothèse, il nous faut maintenant estimer un modèle tenant compte de ces caractéristiques inobservables propres à chaque individu. C'est ce que va faire un panel à effets fixes. Afin de confirmer le choix d'un modèle à effets fixes et de le départager d'avec un modèle à effets aléatoires, nous menons un test de Hausman (voir l'annexe D). Le test rejette fortement l'hypothèse nulle de non-corrélation entre les effets individuels et les régresseurs avec un  $\chi_2(10)$  à 188,30 ; nous en concluons qu'il existe une très forte hétérogénéité individuelle inobservée dans

notre échantillon et faisons le choix de contrôler cette hétérogénéité par des effets fixes individuels. Dans ce type de modèle, on ajoute une indicatrice par individu : on suppose donc que les caractéristiques inobservables (motivation, ambition..) restent exactement les mêmes au cours des 3 périodes. L'estimation est présentée dans la table 7.

TABLE 7 – Effet de la naturalisation sur les salaires : panel à effets fixes

	(1) lrev	(2) lrev	(3) lrev
Naturalisation	0,150*** (0,0373)	0,0843 (0,0570)	-0,0237 (0,0581)
Actif	0,315*** (0,0149)	0,344*** (0,0157)	0,298*** (0,0155)
Conjoint		0,219*** (0,0278)	0,257*** (0,0288)
Nombre d'enfants		0,131*** (0,00930)	0,0596*** (0,0107)
Durée de présence en France en mois			0,00755*** (0,000553)
Constante	7,222*** (0,00920)	6,892*** (0,0225)	6,494*** (0,0375)
<i>N</i>	13629	9544	9517
adj. <i>R</i> <sup>2</sup> within	0,069	0,135	0,168

Standard errors in parentheses  
 \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

Avec pour seuls contrôles les effets fixes individuels et le statut (actif=1 si l'individu possède un emploi au moment de l'interrogation), nous trouvons que la naturalisation s'accompagne d'une hausse des salaires de 15%. L'ajout de contrôles sociodémographique diminue le coefficient et lui fait perdre sa significativité. Enfin, si l'on contrôle par la durée de présence en France, le coefficient devient même négatif (et toujours largement non significatif). Ceci confirme bien, que la « prime de naturalisation » était simplement liée à un effet de sélection. Dès lors que l'on intègre des effets individuels et que l'on introduit un nombre suffisamment conséquent de caractéristiques observables dans le modèle, l'impact devient nul et non significatif. La naturalisation en tant que telle ne semble pas apporter une croissance des salaires significative sur notre échantillon.

## 6 Conclusion

Dans cette étude nous nous sommes attachés à évaluer l'impact de la formation linguistique proposée aux nouveaux migrants depuis 2009 ainsi que l'effet de la naturalisation. Nous avons pris soin de ne pas négliger les biais de sélection qui surviennent lorsque des individus sont traités de manière non-aléatoire. En contrôlant pour ces biais, nous trouvons que ni la formation linguistique, ni la naturalisation n'ont d'impact significatif sur le devenir des migrants. En fait, il semble que

la très forte hétérogénéité inobservée caractérisant les bénéficiaires de la formation linguistique ne nous ait pas permis d'estimer proprement son impact. Toutefois, nous pensons malgré tout que l'efficacité et l'ambition de cette formation sont à questionner : pour preuve, à l'issue de la formation, plus de 70% des migrants interrogés estiment que la formation était insuffisante pour apprendre à parler et à écrire français. Ils sont également plus de 70% à souhaiter bénéficier d'heures de cours supplémentaires.

Les résultats que nous obtenons sur l'impact de la naturalisation sont en contradiction avec la littérature récente. Nous ne trouvons pas de « prime à la naturalisation » et celle-ci serait donc davantage l'aboutissement d'un parcours d'intégration réussie qu'un moyen d'intégration per se. Nous pouvons avancer deux raisons pour expliquer ces résultats : premièrement une grande partie des travaux mettant en évidence une hausse des salaires consécutive à la naturalisation ne tenaient qu'improprement compte de l'existence d'effets de sélection : seul le travail de Bratsberg et al. (2002 [3]) utilise un modèle de panel à effets fixes. Deuxièmement, les conditions d'accès à la naturalisation se sont considérablement endurcies ces dernières années<sup>6</sup> en France. Aussi les migrants qui ont accédé à la nationalité française en 2010, 2011 ou 2012 étaient davantage « sélectionnés » : ils parlaient parfaitement français, disposaient de revenus largement supérieurs à ceux du reste des immigrés, étaient plus diplômés et vivaient en France depuis en moyenne plus longtemps. Cette sélection ne nous a pas permis de mettre en évidence un effet propre de la naturalisation. Toutefois, nous ne pouvons écarter l'existence d'un tel effet : il se peut, et il est même très probable, que sur des populations plus fragiles la naturalisation eût eu un impact sur le niveau de salaire et l'« intégration » en général. L'ensemble de ces résultats suggère la nécessité de politiques plus courageuses si l'on veut réellement faciliter l'intégration des nouveaux migrants et réduire le large différentiel qu'il existe entre le niveau de vie des immigrés et celui des natifs.

---

6. Ainsi à partir du 1er janvier 2012, les étrangers désirant l'acquérir devaient prouver qu'ils maîtrisaient le français au niveau « B1 oral ».

## A Construction d'un indicateur du niveau de français

Afin de pouvoir mener une régression de discontinuité floue, nous avons du construire un indicateur du niveau de français présentant un seuil assez net au-delà duquel la probabilité de formation se chute brutalement. Notre indicateur est la simple moyenne pondérée des « notes » obtenues à chacun de trois tests différents :

- Le **test Elipa de vague 1** : 16 questions rapportant chacune 0, 1, 2 ou 3 points selon la réponse données (note dépendant du degré de compréhension de la question et de la langue de réponse). Nous obtenons un score sur 43 que nous rapportons sur 20.

- Le **test de l'OFII**. Nous disposons des résultats partiels du test mené par l'OFFI qui a véritablement présidé à la prescription de la formation. Nous avons une variable en 3 modalités donnant le résultat à un test oral et un variable en 5 modalités donnant le résultat à un test écrit. Nous les sommons et les rapportons sur 20.

- Enfin nous construisons **un indicateur du niveau de français autodéclaré** à partir de 4 questions de l'enquête ELIPA :

- **Comment parlez-vous français ?** Très bien/Plutôt bien/Pas très bien/Peu ou pas du tout

- **Difficultés pour téléphoner pour prendre rendez-vous avec un médecin, obtenir un renseignement ?** Toujours/Parfois/Ça ne vous pose pas de difficultés/Vous ne le faites jamais

- **Difficultés pour demander son chemin dans la rue ?** Toujours/Parfois/Ça ne vous pose pas de difficultés/Vous ne le faites jamais

- **Difficultés pour écrire seul une lettre en français, remplir un document administratif ?** Toujours/Parfois/Ça ne vous pose pas de difficultés/Vous ne le faites jamais

En attribuant 0, 1, 2 ou 3 à chacune des réponses, puis en sommant, nous obtenons une « note » sur 9 que nous rapportons sur 20.

Enfin nous faisons la moyenne de ces trois notes en mettant un poids important sur le résultat au test de l'OFII, de manière à observer une discontinuité dans la probabilité de traitement. Nous obtenons les résultats suivants :



FIGURE 3 – Niveau de français des migrants

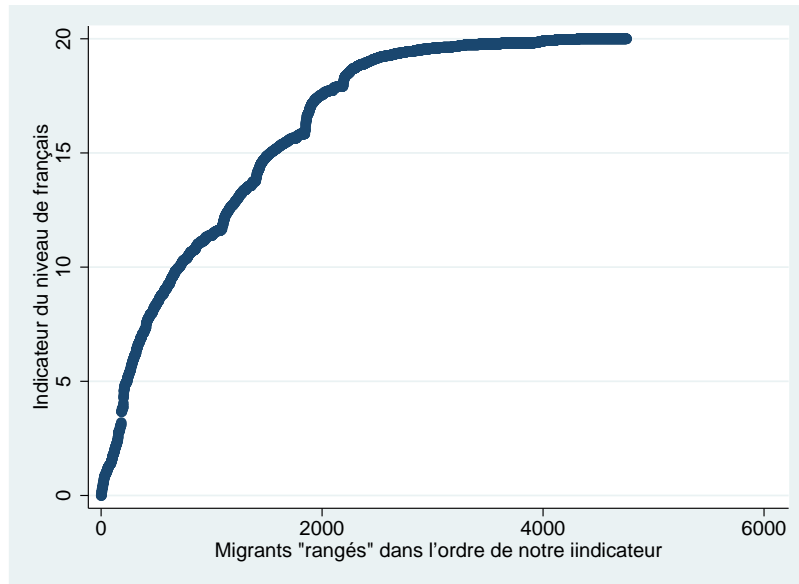


FIGURE 4 – Répartition de l'indicateur construit du niveau de français

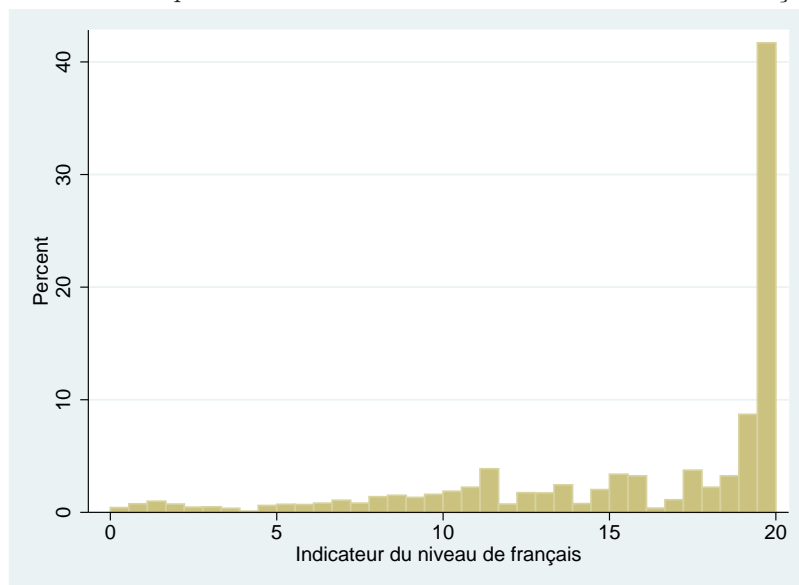


FIGURE 5 – Probabilité de formation linguistique en fonction du niveau de français

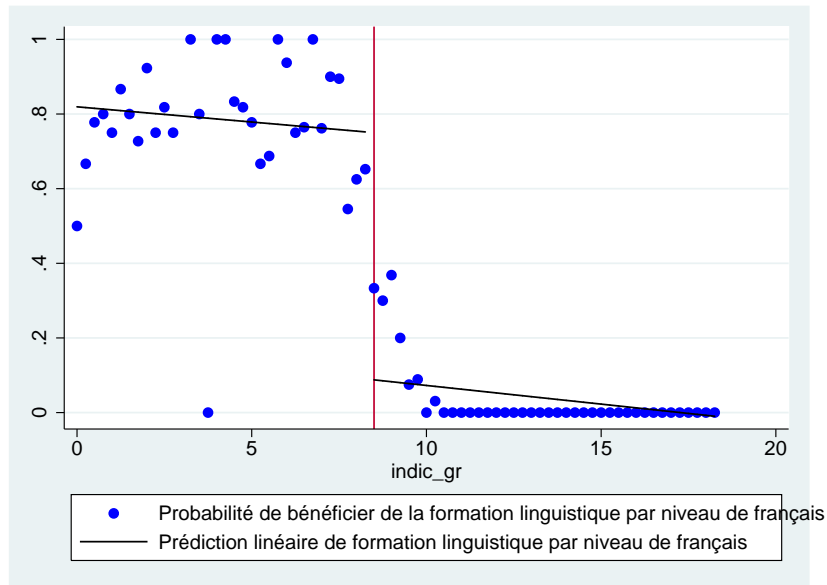


FIGURE 6 – Matrice de corrélation des sous-indicateurs

```
. pwcorr indic_autodeclar indic_OFII indic_elipa
```

	indic~ar	indic~I	indic~a
indic_auto~r	1.0000		
indic_OFII	0.5739	1.0000	
indic_elipa	0.6038	0.5798	1.0000

La variable « indic » présente dans la table descriptive 1 correspond à la version centrée réduite de cet indicateur.

## B Construction d'un indicateur du niveau d'intégration

Afin d'intégrer dans nos analyses une variable décrivant de manière plus subjective le parcours des migrants et tenter de capter, une partie au moins de l'hétérogénéité inobservée nous avons construit une variable « intégration ». Il s'agit d'un indicateur construit à partir de deux sous-indicateurs représentant chacun une dimension de l'"intégration". Nous tenons à insister sur le caractère extrêmement subjectif et en aucun normatif de ces indicateurs. Nous avons juste tenté de construire à partir des seules questions posées de la même manière en vague 1 et en vague 2 de construire un indicateur qui capterait une dimension non encore prise en compte par notre modèle. Nous la nommons "intégration" pour plus de commodité. L'indicateur d'"intégration" opère donc

des classements des individus selon leur comportement ou leur opinion et leur attribuent une « note » plus ou moins élevée selon leur position dans ce classement. Or, nous tenons à souligner qu'il n'y a pas d'individu « moins intégré » ou « plus intégré » : l'indicateur ne présente pas de connotation normative ; il s'agit pour nous d'utiliser une convention opérationnelle qui nous permettra de mener une étude quantitative.

Notre indicateur possède donc deux dimensions :

### **Mixité et relations sociales**

Cet indicateur correspond à la somme des 5 variables binaires suivantes (obtenues à partir des mêmes questions disponibles en vague 1 et en vague 2) :

- $y_1 = 1$  si le conjoint est d'une nationalité différente
- $y_2 = 1$  si les collègues de travail sont de nationalités diverses
- $y_3 = 1$  si les voisins sont de nationalités diverses
- $y_4 = 1$  si le migrant mange régulièrement chez des amis d'une origine différente de la sienne
- $y_5 = 1$  si le migrant s'est fait de nouveaux amis d'une origine différente que la sienne au cours de la période précédant l'interrogation

### **Modernité et vie en société**

Cet indicateur correspond à la somme des 5 variables binaires suivantes (obtenues à partir des mêmes questions disponibles en vague 1 et en vague 2) :

- $y_1 = 0$  si le migrant déclare « ne pas s'adapter/ne pas apprécier les modes de vie français » , 1 sinon
- $y_2 = 1$  si le migrant dispose d'une connexion internet
- $y_3 = 1$  si le migrant ne cuisine pas uniquement de la cuisine traditionnelle
- $y_4 = 0$  si le migrant déclare respecter la totalité des interdits de sa religion, 1 sinon
- $y_5 = 1$  si le migrant ne parle pas uniquement sa langue maternelle avec son conjoint

Encore une fois, nous insistons sur la connotation non normative de cet indicateur. Il s'agissait seulement pour nous de synthétiser un ensemble de questions présentes dans les 2 vagues du questionnaire et présentant un certain degré de corrélation.

## C Régression de discontinuité : résultats bruts

FIGURE 7 – Régression de discontinuité floue par variables instrumentales

### First-stage regressions

```

Number of obs =      2536
N. of clusters =      623
F(   3,   2532) =    383.98
Prob > F      =      0.0000
R-squared     =      0.6329
Adj R-squared =      0.6325
Root MSE     =      0.2066
  
```

fling	Robust		t	P> t	[95% Conf. Interval]	
	Coef.	Std. Err.				
gapabove	-.0168486	.0033981	-4.96	0.000	-.0235118	-.0101853
gapbelow	.0000246	.0085851	0.00	0.998	-.0168099	.0168591
above	-.6284665	.053794	-11.68	0.000	-.7339512	-.5229817
_cons	.7955444	.0432469	18.40	0.000	.7107416	.8803473

### Instrumental variables (2SLS) regression

```

Number of obs =      2536
Wald chi2(3)  =    159.42
Prob > chi2   =      0.0000
R-squared     =      0.2906
Root MSE     =      3.7153
  
```

(Std. Err. adjusted for 58 clusters in indic\_gr)

testelipa2	Robust		z	P> z	[95% Conf. Interval]	
	Coef.	Std. Err.				
fling	-1.747729	1.209752	-1.44	0.149	-4.1188	.6233419
gapabove	.3509754	.0691472	5.08	0.000	.2154493	.4865014
gapbelow	.4285529	.1882954	2.28	0.023	.0595008	.7976051
_cons	14.81131	.6395358	23.16	0.000	13.55784	16.06477

Instrumented: fling

Instruments: gapabove gapbelow above

FIGURE 8 – Régression de discontinuité par commande Austin Nichols

```
. rd testelipa2 indic_gr, z0(`s') graph noscatter
Two variables specified; treatment is
assumed to jump from zero to one at Z=8.5.

Assignment variable Z is indic_gr
Treatment variable X_T unspecified
Outcome variable y is testelipa2

Command used for graph: lpoly; Kernel used: triangle (default)
Bandwidth: 2.7220818; loc Wald Estimate: -1.0197188
Bandwidth: 1.3610409; loc Wald Estimate: .93813777
Bandwidth: 5.4441635; loc Wald Estimate: -.77113519
(1037 missing values generated)
(1037 missing values generated)
(1037 missing values generated)
Estimating for bandwidth 2.72208175203811
Estimating for bandwidth 1.361040876019055
Estimating for bandwidth 5.44416350407622
```

testelipa2	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lwald	-1.019719	1.094278	-0.93	0.351	-3.164464	1.125027
lwald50	.9381378	1.508517	0.62	0.534	-2.018502	3.894778
lwald200	-.7711352	.7682042	-1.00	0.315	-2.276788	.7345173

## D Test de Hausman

FIGURE 9 – Test d’Hausman

. hausman FE RE

	Coefficients		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) FE	(B) RE		
naturalisa~n	.0497559	.1366312	-.0868753	.0815581
age	-.0701455	.0328678	-.1030134	.1004801
age2	-.0004537	.0001657	-.0006194	.0005623
conjoint	.2215359	.3965818	-.175046	.0423822
nb_enf	.0117132	.0070782	.004635	.0165657
dip	-.0195947	.0204603	-.040055	.0216023
actif	.3038116	.3301687	-.0263571	.0133561
bon	-.0181733	.0337845	-.0519578	.0161721
duree_mois	.0179579	-.0038679	.0218258	.007949
integ	.0026212	.0262693	-.0236481	.0037072

b = consistent under Ho and Ha; obtained from xtreg  
 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(10) = (b-B)'[(V\_b-V\_B)^(-1)](b-B)  
 = 188.30  
 Prob>chi2 = 0.0000

## Table des figures

1	Probabilité de formation linguistique en fonction du niveau de français . . . . .	11
2	Régression de discontinuité . . . . .	12
3	Niveau de français des migrants . . . . .	23
4	Répartition de l'indicateur construit du niveau de français . . . . .	23
5	Probabilité de formation linguistique en fonction du niveau de français . . . . .	23
6	Matrice de corrélation des sous-indicateurs . . . . .	24
7	Régression de discontinuité floue par variables instrumentales . . . . .	26
8	Régression de discontinuité par commande Austin Nichols . . . . .	26
9	Test d'Hausman . . . . .	27

## Liste des tableaux

1	Statistiques descriptives . . . . .	8
2	Difference in differences - Niveau de français vague 1 à 2 . . . . .	14
3	Difference in differences - Niveau de français vague 1 à 3 . . . . .	14
4	Difference in differences - Autres outcomes . . . . .	15
5	Naturalisation : statistiques descriptives . . . . .	17
6	Effet de la naturalisation sur les salaires : estimation d'un modèle empilé . . . . .	18
7	Effet de la naturalisation sur les salaires : panel à effets fixes . . . . .	20

## Références

- [1] Ilana Redstone AKRESH : US immigrants' labor market adjustment : Additional human capital investment and earnings growth. *Demography*, 44(4):865–881, 2007.
- [2] Pieter BEVELANDER et Justus VEENMAN : Naturalisation and socioeconomic integration : The case of the Netherlands. Rapport technique, IZA Discussion Papers, 2006.
- [3] Bernt BRATSBERG, James F RAGAN, Jr et Zafar M NASIR : The effect of naturalization on wage growth : A panel study of young male immigrants. *Journal of Labor Economics*, 20(3):568–597, 2002.
- [4] Barry R CHISWICK : The effect of americanization on the earnings of foreign-born men. *The Journal of Political Economy*, 86(5):897, 1978.
- [5] Barry R CHISWICK et Paul W MILLER : The complementarity of language and other human capital : Immigrant earnings in Canada. *Economics of Education review*, 22(5):469–480, 2003.

- [6] Rafael Munoz de BUSTILLO et José-Ignacio ANTÓN : From rags to riches ? Immigration and poverty in Spain. *Population Research and Policy Review*, 30(5):661–676, 2011.
- [7] Lennart DELANDER, Mats HAMMARSTEDT, Jonas MÅNSSON et Erik NYBERG : Integration of immigrants the role of language proficiency and experience. *Evaluation Review*, 29(1):24–41, 2005.
- [8] Don J DEVORETZ : The economics of citizenship : a common intellectual ground for social scientists ? *Journal of Ethnic and Migration Studies*, 34(4):679–693, 2008.
- [9] Don J DEVORETZ et Sergiy PIVNENKO : The economic causes and consequences of canadian citizenship. *Journal of International Migration and Integration/Revue de l'integration et de la migration internationale*, 6(3-4):435–468, 2005.
- [10] Don J DEVORETZ et Sergiy PIVNENKO : Self-selection, immigrant public finance performance and canadian citizenship. Rapport technique, IZA Discussion paper series, 2005.
- [11] Patricia A DUFF, Ping WONG et Margaret EARLY : Learning language for work and life : The linguistic socialization of immigrant canadians seeking careers in healthcare. *The Modern Language Journal*, 86(3):397–422, 2002.
- [12] Christian DUSTMANN : Speaking fluency, writing fluency and earnings of migrants. *Journal of Population economics*, 7(2):133–156, 1994.
- [13] Denis FOUGÈRE et Mirna SAFI : L'acquisition de la nationalité française : quels effets sur l'accès à l'emploi des immigrés ? *France, portrait social, 2005-2006*, 2005.
- [14] INSEE : Immigrés et descendants d'immigrés en France, 2012.
- [15] Lisa KAIDA : Do host country education and language training help recent immigrants exit poverty ? *Social science research*, 42(3):726–741, 2013.
- [16] Abdolmohammed KAZEMIPUR et Shiva S HALLI : Immigrants and 'new poverty' : The case of Canada. *International Migration Review*, 35(4):1129–1156, 2001.
- [17] Irena KOGAN : Ex-yugoslavs in the austrian and swedish labour markets : the significance of the period of migration and the effect of citizenship acquisition. *Journal of Ethnic and Migration Studies*, 29(4):595–622, 2003.
- [18] Philippe LOMBARDO et Jérôme PUJOL : Niveau de vie et pauvreté des immigrés en 2007. *INSEE Références*, 2010.
- [19] Francesca MAZZOLARI : Dual citizenship rights : do they make more and better citizens ? 2007.



- [20] Austin NICHOLS : rd : Stata module for regression discontinuity estimation. *Statistical Software Components*, 2012.
- [21] OCDE : *Naturalisation : A Passport for the Better Integration of Immigrants ?* OECD Publishing, 2011.
- [22] Gilles PISON : Le nombre et la part des immigrés dans la population : comparaisons internationales. Rapport technique, Institut National d'Études Démographiques (INED), 2010.
- [23] Jean RENAUD et Tristan CAYN : Jobs commensurate with their skills? Selected workers and skilled job access in Québec. *Journal of International Migration and Integration/Revue de l'intégration et de la migration internationale*, 8(4):375–389, 2007.
- [24] Dennis H SULLIVAN et Andrea L ZIEGERT : Hispanic immigrant poverty : does ethnic origin matter? *Population Research and Policy Review*, 27(6):667–687, 2008.
- [25] Frank VAN TUBERGEN et Matthijs KALMIJN : A dynamic approach to the determinants of immigrants' language proficiency : The United States, 1980–2001. *International Migration Review*, 43(3):519–543, 2009.
- [26] Ying ZHEN : The effects of english proficiency on earnings of US foreign-born immigrants : Does gender matter? *Journal of Economics & Finance*, 1(1):27–41, 2013.