



**UNIVERSITÉ
DE GENÈVE**

**FACULTÉ DES SCIENCES
DE LA SOCIÉTÉ**

Faculté des Sciences de la société (SDS)
Master en Socioéconomie
Année académique 2021-2022
Cours: Analyse de données d'enquêtes
Professeur : Prof. Studer

Quelle est l'impact de la nationalité sur la probabilité de se retrouver en situation de qualification négative, surqualifié ou en inadéquation de qualification ?

Le rôle du réseau, des diplômes et de la connaissance du marché du travail Suisse.

Vestin Hategekimana & Julien Massard

13/12/2021

Table des matières

Introduction.....	2
Hypothèses.....	4
Données et opérationnalisation.....	6
Source des données.....	6
Variable indépendante principale: Nationalité.....	6
Variable dépendante: correspondance qualification-emploi.....	7
Variable médiatrice : âge de résidence.....	9
Variables de contrôle.....	10
Présentation du nombre de données manquantes par variable et globale.....	11
Analyse exploratoire.....	11
Khi deux et résidus standardisés.....	12
Analyse par arbre.....	13
Modélisation.....	16
Description des modèles.....	16
Modèles emboîtés.....	16
Choix du modèle.....	20
Discussion.....	21
Conclusion.....	23
Bibliographie.....	24
Annexe Étude des données manquantes.....	25

Introduction

La Suisse est un pays avec une très forte proportion d'immigrés dans sa population résidente permanente. Même si l'OCDE indique que la situation des travailleurs immigrés sur le marché du travail Suisse est "dans son ensemble largement favorable [...] Pour les hommes et pour les femmes, le taux d'emploi des immigrés est plus élevé que dans les autres pays" (Liebig et al., 2012) qui peut s'expliquer par l'introduction de la loi sur la libre circulation des personnes, il existe des disparités sur le marché du travail entre nationaux et immigrés ainsi qu'entre les immigrés. De nombreux obstacles se dressent sur le chemin des immigrés pour s'intégrer au marché du travail suisse. En effet, la sous-utilisation des qualifications de cette population est un problème connu en Suisse. Cette problématique entraîne d'une part un gaspillage de cerveaux pour les migrants comme pour la société suisse et d'autre part pose la question éthique d'égalité de traitement entre suisse et étrangers.

Un des enjeux d'équité dans l'accès au marché du travail, tout comme une des dimensions de la performance du marché du travail, est l'adéquation existant entre qualification des

travailleurs et l'emploi effectivement occupé par les individus. Ce lien peut être synthétisé par une typologie sommaire regroupée en deux pôles. D'un côté un lien que l'on peut qualifier de positif pour les personnes dont les qualifications correspondent à leurs emplois, ainsi que pour les travailleurs ayant des compétences inférieures à leurs postes. Cette dernière est considérée positive, car la personne en état de sous-qualifications n'a pas de besoin de fournir des compétences équivalentes normalement à un niveau d'emploi à salaire supérieur et, d'autre part, la société peut avoir un retour sur investissement éducatif avantageux, car l'état n'a pas besoin de fournir une formation permettant un salaire supérieur. De l'autre côté ce lien peut être considéré comme négatif pour les individus ayant des qualifications supérieures à leurs emplois ainsi qu'ayant des qualifications ne correspondant pas à leurs emplois. En effet, ces deux cas amènent une sous-utilisation des compétences tout comme ils peuvent provenir d'opportunités relativement faibles des individus. Dès lors nous pouvons nous demander comment ce lien positif ou négatif entre niveau de qualification et niveau d'emploi peut affecter de manière différencié les Suisses et les étrangers.

En effet, il est généralement admis dans la littérature que les étrangers ont tendance à être plus en situation de surqualification dans leur emploi que les Suisses (OFS, 2021). Nous pouvons lire par exemple dans le rapport de l'OCDE sur l'intégration des immigrants et de leurs enfants sur le marché du travail en Suisse ceci:

“Le taux d'emploi généralement élevé des immigrants en Suisse est également lié à une fréquente surqualification des migrants diplômés de pays non membres de l'OCDE. Les qualifications acquises à l'étranger sont insuffisamment valorisées sur le marché du travail suisse, et il n'y a que peu de cours passerelle [...] peu de programme de parrainage” (Liebig et al., 2012)

Une des causes de cette qualification négative peut être dû à la discrimination sur le marché de l'emploi. À ce sujet, une étude en Suisse par envoi de CV similaires pour postuler à un emploi a montré qu'il existait effectivement de la discrimination entre Suisses et Suisses avec une double nationalité (Zschirnt & Fibbi, 2019). Effectivement, ces derniers doivent envoyer en moyenne 30% de candidature en plus pour espérer obtenir leur premier entretien pour des apprentissages même si l'étude souligne que l'inégalité de traitement est moins prononcée dans des emplois qui nécessitent un niveau de formation tertiaire. Les auteurs parlent ici de discrimination par le goût, c'est-à-dire que l'éloignement culturel des étrangers serait à l'origine d'effet discriminatoire des suisses envers les étrangers, en particulier pour des formations supérieures. Dans cette perspective, les descendants de parents originaires d'un pays non-membre de l'UE sont principalement discriminés dans les emplois de qualification supérieure secondaire avec une hiérarchisation ethnique très marquée. Ainsi la population suisse issue de l'est de l'Europe est connue pour subir de la discrimination dans de nombreuses situations. On constate aussi, dans une moindre mesure, de la discrimination pour les descendants de parents originaires de l'UE, dans certains secteurs spécifiques. Cette étude confirme les résultats d'une précédente étude réalisée en 2002 (Fibbi et al., 2003) montrant que les candidats non suisses étaient souvent pris après le refus des Suisses, révélant en quelque sorte que la discrimination sur le marché du travail est courante en Suisse. Dans l'étude de 2002, ce sont les ressortissants de l'ex-Yougoslavie et les Turcs qui montraient des taux de discrimination plus forte.

Au-delà du facteur discriminatoire, d'autres effets peuvent expliquer le fait que les étrangers aient plus de chances de tomber dans des situation de surqualification ou dans des emplois ne correspondant pas à leurs formations: la difficulté de faire reconnaître un diplôme étranger, la connaissance du marché du travail Suisse et le réseau de connaissance (Liebig et al., 2012). De plus, le fait que les descendants d'immigrés et les personnes naturalisées s'en sorte mieux sur le marché du travail pourrait suggérer que le temps passé en Suisse joue un rôle tout comme l'âge d'arrivée en Suisse (Liebig et al., 2012).

Comme la littérature le suggère, nous formulons notre hypothèse principale de la manière suivante: Les travailleurs non-suisse ont bien plus de chance que leur homologues suisses de se retrouver dans une situation de qualification négative, c'est-à-dire, soit en surqualification, soit en emploi ne correspondant pas à leur qualification. À cela nous ajoutons, que selon l'origine, la probabilité sera plus ou moins forte, notamment que les personnes soient naturalisées, issues d'un pays frontalier, issues d'un pays de l'OCDE, issues d'un pays de l'Europe de l'est ou d'ailleurs. Toutefois, nous souhaitons contrôler l'âge d'arrivée en Suisse de notre population afin de prendre en compte l'effet du diplôme et de la scolarité suisse, ainsi que la connaissance du marché du travail et l'effet du réseau de la personne.

Afin de répondre à cette question, nous allons utiliser les données sur le Panel Suisse des Ménages, plus particulièrement la vague 22, soit l'année 2020. Notre population cible est constituée d'individus actifs (en emploi) suisses et étrangers. Les données étant encore récentes, nous ne disposons pas de variables de pondération et de stratification dans cette étude. Il est dès lors possible d'essayer certains biais quant à la distribution réelle des individus du panel dans la population suisse. Dans un premier temps, après une présentation des données, nous allons analyser la relation entre la qualification et la nationalité dans une analyse bivariée simple par la méthode du Khi Deux. Par la suite, nous allons complexifier le modèle en ajoutant une série de variables de contrôle usuelles avant d'ajouter une variable de médiation: l'âge de résidence. Finalement nous testerons des éventuels effets d'interaction avec notre variable de médiation. Ce travail permettra d'apporter des informations quant au lien entre l'origine et le degré d'adéquation de l'emploi à la qualification.

Hypothèses

Nous avons en tout 4 hypothèses qui peuvent être résumées dans le schéma suivant:

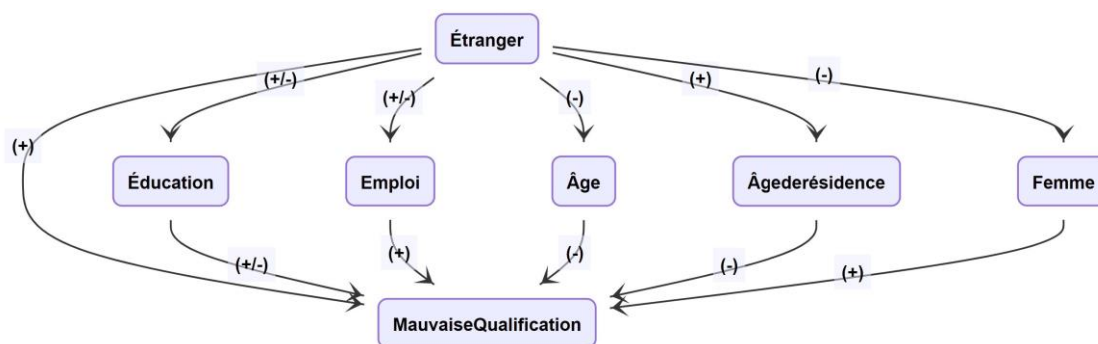


Schéma des variables (1)

Notre première hypothèse est qu'il existe un lien entre l'origine et la possibilité d'être en situation de mauvaise qualification. Effectivement, un premier article de l'OCDE (Liebig et al., 2012), indiquait déjà que la population étrangère avait tendance à être surqualifiée. Ces affirmations sont en accord avec les chiffres de l'Office fédérale de la statistique (OFS, 2021). Nous nous attendons donc, non seulement, à voir un lien entre l'origine et le fait de se retrouver en bonne/mauvaise qualification, notamment dans le test de Khi Deux, et nous nous attendons à voir le lien tenir après l'ajout de nos variables de contrôle et notre variable médiatrice dans nos modèles de régression logistique.

Notre seconde hypothèse est qu'il existe des hiérarchies entre les types de nationalité. De manière générale, plus un pays est éloigné culturellement, plus d'une part les individus peuvent être susceptibles d'être discriminés sur le marché du travail du fait des préjugés des natifs s'intensifiant par la distance d'origine. En particulier, les individus venant des pays de l'Europe de l'est pourraient subir plus de qualification négative si nous faisons le parallèle avec les travaux sur la discrimination en emploi (Fibbi et al., 2003; Zschirnt & Fibbi, 2019).

À côté de cela, trois autres variables peuvent affecter cet effet. Tout d'abord, une des ressources les plus importantes dans l'accès à l'emploi est le réseau que les individus sont capables de mobiliser. Ainsi les personnes étrangères peuvent avoir des difficultés à trouver un emploi correspondant à leurs qualifications du fait qu'elles n'ont pas eu autant de temps que les Suisses pour construire un réseau solide. Autrement dit, un déficit dans le capital social peut entraver l'accès à l'emploi et par extension un emploi équivalent aux qualifications. Ensuite, la non-reconnaissance des diplômes est également un effet qui peut expliquer des différences de qualification négative entre suisse et étranger. Par exemple, un grand nombre de pays ne font pas partie du programme de Bologne d'universalisation des études. Ainsi il est parfois difficile de faire valoir ses qualifications pour les étrangers (Liebig et al., 2012). Pour finir, l'insertion sur le marché du travail demande une connaissance certaine de son fonctionnement. De ce fait, les étrangers peuvent être défavorisés par rapport aux Suisse sans pour autant que cela soit de la discrimination à proprement parlée. Ce qui nous amène à notre troisième hypothèse: plus un étranger arrive tard dans le pays, plus il a de chances de subir une qualification négative selon les différents mécanismes susmentionnés.

Notre quatrième hypothèse s'intéresse aux liens entre nationalité, éducation et qualification. Tout d'abord, étant donné qu'un grand nombre d'étranger viennent en Suisse pour une place spécifique requérant un haut niveau d'étude, nous pouvons dès lors envisager un lien positif entre nationalité étrangère et niveau d'étude (OFS, 2021). Cependant, certains groupes d'étrangers ont tendance à être moins formés que les Suisses, notamment les réfugiés, donc le lien entre les deux peut varier selon le type d'étranger. Ensuite, l'éducation devrait affecter de manière ambivalente la qualification négative. En effet, les individus ayant un niveau de formation faible devraient moins être sujet à la surqualification étant donné leur niveau d'étude faible. Cela s'explique de manière mécanique: leurs niveau d'étude étant bas, il y a moins de possibilité que leur niveau de qualification soit au-dessus de la plupart des emplois disponibles sur le marché de l'emploi. Par contre, il y a plus de chances que ces personnes soient sujet à des emplois ne correspondant pas à leurs niveaux d'étude du fait de la plus grande interchangeabilité des formations ou encore des formes de précarisations obligeant l'acceptation d'emploi différent de leurs formations (niveau d'étude et précarisation étant

fortement documenté) (Liebig et al., 2012). Inversement les personnes ayant un niveau d'étude élevé devraient être plus affectées par la surqualification pour la même raison mécanique et devraient moins avoir des emplois qui ne concordent pas à leur qualification en raison d'un plus grand nombre d'opportunités étant donné le coût engendré par le temps de formation et les possibilités de mobilités plus grandes pour cette catégorie.

Données et opérationnalisation

Source des données

Lors de cette recherche, nous avons utilisé des données provenant du Panel Suisse des ménages de 2020. Cette enquête longitudinale a été créée en 1999 et est renouvelée chaque année. Elle cherche à évaluer l'évolution des conditions de vie en Suisse. Les thématiques que recouvrent ce panel sont très larges: marché du travail, éducation, origine, etc.

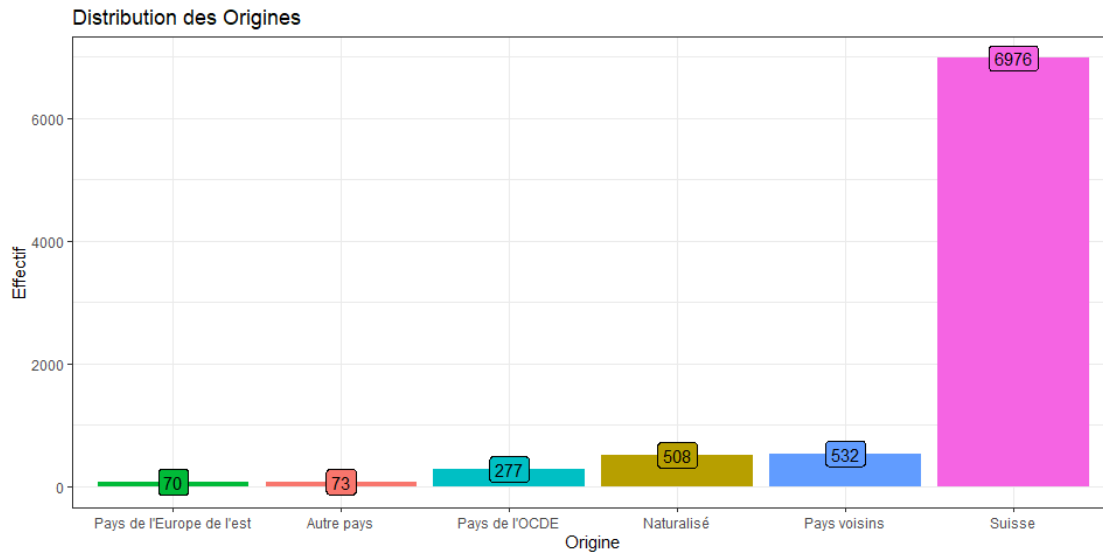
Pour notre recherche, nous avons utilisé uniquement les données personnelles, mettant ainsi de côté les données concernant le Covid, les données des ménages et les données sur l'origine sociale. Les données personnelles observées initialement dans l'enquête de 2020 sont de 24'727 observations. Nous avons sélectionné uniquement les individus étant en emploi du fait que notre recherche s'intéresse au marché du travail. En retirant les diverses valeurs manquantes présentes dans nos données, nous avons fini avec un jeu de données contenant 5216 observations pour 7 variables.

Variable indépendante principale: Nationalité

En ce qui concerne l'origine ou le fait d'être étranger, nous avons utilisé une variable s'intéressant aux nationalités des personnes. En effet, les possibles discriminations que peuvent subir les personnes sur le marché du travail sont susceptibles de provenir du pays d'origine des personnes. Ceci pouvant se comprendre comme un effet de l'éloignement culturel sur les préjugés des natifs et par extension sur les discriminations. Étant donnée que cette variable comprend l'ensemble des pays du monde (205), nous l'avons recodée en plusieurs catégories dans l'objectif de former des catégories de nationalités avec des chances et des niveaux de discriminations relativement homogènes. Ainsi nous avons créé quatre catégories dans lesquels on devrait observer une discrimination de plus en plus élevée: une catégorie "Suisse", une catégorie "Pays voisins" regroupant les pays ayant une frontière commune avec la Suisse, une catégorie "Pays de l'OCDE" composée des pays occidentaux, une catégorie "Pays d'Europe de l'est", du fait de la présence dans la littérature de discriminations envers ces individus (Fibbi et al., 2003; Zschirnt & Fibbi, 2019) et une catégorie "autre" qui regroupe les pays d'Afrique, d'Asie et d'Amérique du sud.

Nous nous sommes également intéressés à la naturalisation. Comme nous l'avons vu dans l'article de Zschirnt & Fibbi (2019), la discrimination se produit également pour les personnes de nationalité Suisse mais avec des origines étrangères. Ainsi les personnes avec une double nationalité la subissent également. Nous utilisons donc la variable sur l'année de réception de la naturalisation pour différencier deux catégories et mettre en évidence les étrangers. Puisque la question n'a pas été posée à tout le monde (par exemple aux étrangers), nous avons décidé de coder toute les personnes étant naturalisé comme "Naturalisé" et de laisser le reste des personnes conserver leur catégorie de la variable nationalité. Ainsi, toutes

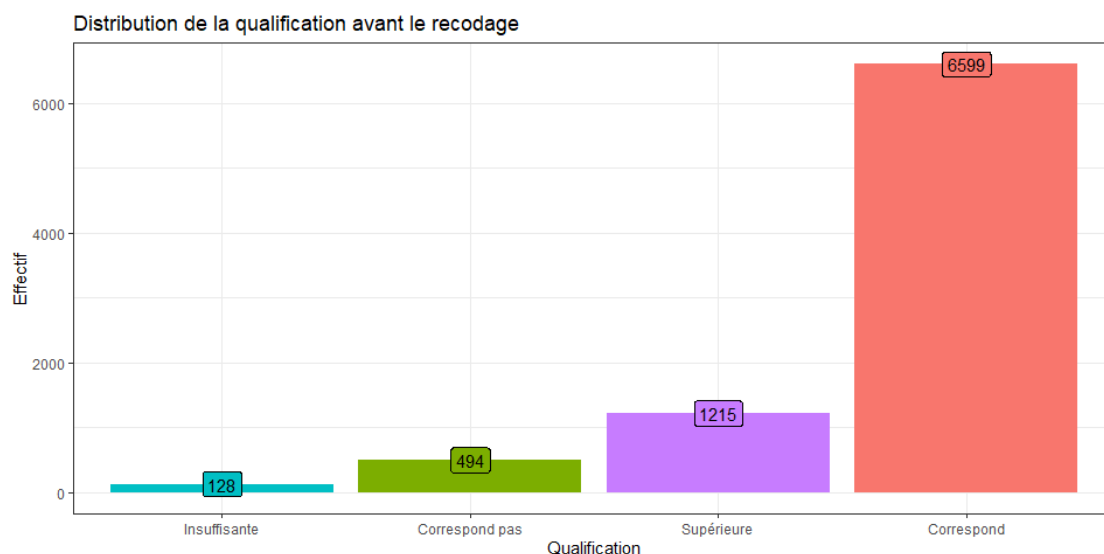
les personnes naturalisées sont passées de la catégorie “Suisse” à la catégorie “Naturalisé”. Le graphique ci-dessous montre la répartition des effectifs en fonction de l’origine. Ainsi sans surprise les nationalités les plus représentées sont dans l’ordre: les Suisses (largement), les pays voisins, les naturalisés, les ressortissants de l’OCDE, les autres pays et finalement les pays de l’Europe de l’est.



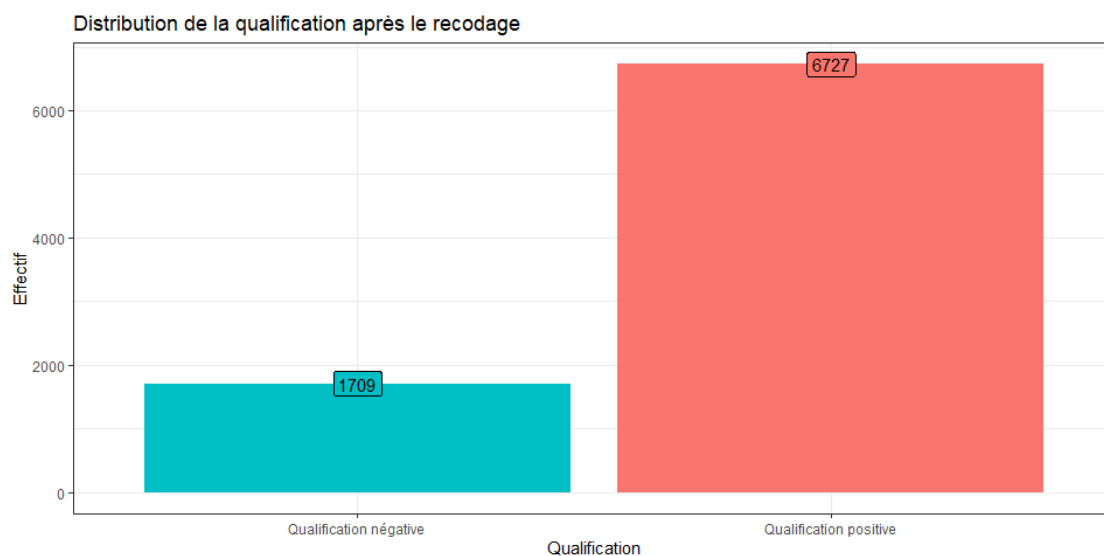
Nous pouvons observer que la catégorie suisse est largement surreprésentée dans nos données alors que les autres catégories représentent relativement peu d’observations. Dès lors, nous pouvons supposer que ce manque de données pour certaines catégories, en particulier les pays autres et les pays de l’Europe de l’est puissent poser des problèmes en termes de significativité.

Variable dépendante: correspondance qualification-emploi

Notre recherche s’intéressant aux liens entre nationalité et qualification négative, nous avons sélectionné une variable interrogeant la correspondance entre la qualification de l’individu et son emploi. Cette dernière nous permet d’opérationnaliser le concept de non-concordance de qualification, car elle possède quatre catégories de réponses: 1) la sous-qualification par rapport à l’emploi, 2) correspondances entre emploi et qualification, 3) surqualification et 4) emplois différents des qualifications. Nous avons donc recodé cette variable en deux catégories: la première est “la qualification positive” intégrant la sous-qualification et la correspondance qualification-emploi; la deuxième catégorie “qualification négative” est composée de la surqualification et d’un emploi différent des qualifications. En reprenant notre hypothèse principale, la qualification positive devrait être plus représentée par les Suisses. Par ailleurs, nous avons choisi de mettre la sous-qualification dans la catégorie qualification positive, car la sous qualification procède d’une forme de surestimation des qualifications et ne correspond pas à une perte utilitaire de prestations publique de formation. A l’inverse, la qualification négative devrait voir une surreprésentation d’individus étrangers. Nous pouvons voir dans l’illustration suivante l’effectif de chaque catégorie de notre variable avant le recodage, dans l’ordre de la plus peuplée à la moins peuplée: Correspond, Supérieur, Correspond pas, Insuffisante.



Le nombre très faible de données de sousqualification tout comme le nombre relativement peu élevé des catégories surqualification et non-correspondance nous enjoint de regrouper ces catégories ensemble, nous permettant par extension d'obtenir une variable binomiale et non plus catégorielle.



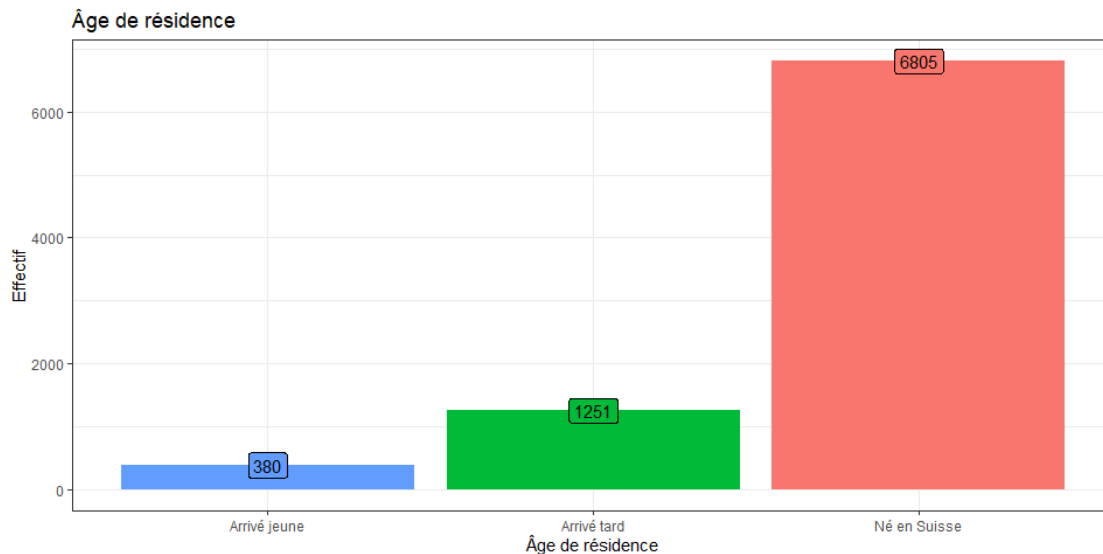
Malgré ce recodage, nous pouvons observer que le nombre qualification positive représente environ 80% des données totale, mais le nombre d'occurrence des qualifications négatives nous apparaît tout de même suffisante pour pouvoir avoir des résultats significatifs. Gardons en tête le fait que la plus grande part des qualifications négatives sont des surqualifications.

Même si cette variable peut nous permettre d'observer la qualification négative, c'est une variable subjective qui peut subir divers biais. Par exemple, la non-connaissance du marché du travail Suisse des étrangers peut amener à des réponses erronées ou encore la désirabilité sociale peut affecter les réponses.

Variable médiatrice : âge de résidence

Lorsque l'on cherche à mesurer la discrimination possible entre qualification négative et nationalité, trois effets concomittants peuvent rentrer en considération: le réseau mobilisable, la reconnaissance du diplôme et les connaissances du marché du travail suisse. Malheureusement, il nous manquait toutes ces variables dans nos données. Pour prendre en compte ces effets, nous avons construit une variable s'intéressant à l'âge d'arriver sur le sol Suisse des étrangers. En effet, si un étranger est arrivé très jeune, les différents effets susmentionnés seront moindres étant donné qu'il aura eu le temps de construire un réseau, d'avoir plus de chance d'obtenir un diplôme suisse et d'avoir des connaissances approfondies sur le marché du travail. Inversement, un individu arrivé plus tard dans son parcours de vie en Suisse n'aura pas la possibilité de faire ces choses et sera ainsi désavantagé par rapport aux Suisses dans l'accès à un emploi en concordance avec ses qualifications.

Par conséquent, nous avons construit une variable s'appelant âge de résidence grâce au variable "âge" et "année d'habitation". Ainsi l'âge de résidence correspond à l'âge de la personne soustrait par la différence entre l'année d'arrivée en Suisse et l'année d'enquête. Puisque la question n'a pas été posée à tout le monde (seulement aux personnes avec au moins une nationalité étrangère), nous sommes partie de l'hypothèse que les Suisses sont nés en Suisse. Nous avons alors construit trois catégories: la première étant "née en Suisse" pour les personnes nées en Suisse, la deuxième "arrivé jeune" correspondant aux individus s'étant installés en Suisse entre leur première et leur dix-huitième année et la troisième "arrivé tard" pour les individus étant arrivés en Suisse après leurs 18 ans. Cette variable nous permettra de répondre à notre troisième hypothèse. Dans cette optique, les personnes arrivées après 18 ans devraient avoir une proportion plus élevée d'observations de qualification négative que les individus étant arrivés avant leurs 18 ans et d'autant plus que les personnes nées en Suisse. Cette variable nous permet d'effectuer un proxy des effets concomittants susmentionnés, évidemment sujet à critique, afin d'isoler l'effet de discrimination.



Le nombre de personnes nées en suisse apparaît bien plus conséquent que les deux autres catégories. Ces dernières ont quant à elles un nombre de fréquences assez éloigné, mais le nombre relativement faible de la catégorie jeune pourrait poser problèmes en termes de significativité de nos résultats.

Variables de contrôle

Nous avons sélectionné des variables de contrôle usuelles en sciences sociales: l'âge, le sexe, l'éducation et l'emploi. Le but étant de vérifier nos diverses interprétations par ses variables qui peuvent avoir des incidences majeures sur les résultats. En effet, la discrimination genrée sur le marché du travail est largement documentée (Bonvin, 2009), d'autant plus dans le cas de l'immigration (Fibbi et al., 2003; Zschirnt & Fibbi, 2019). L'âge pourrait également avoir une incidence, surtout lorsqu'on pense à l'entrée sur le marché du travail des jeunes adultes et le risque plus élevé de ne pas avoir un emploi correspondant aux qualifications (stage etc...). D'autant plus que les chiffres de l'OFS sur la surqualification semble montrer que plus les personnes sont âgées, moins elles ont de chance d'être en situation de surqualification (OFS, 2021). L'éducation pourrait jouer un rôle du fait que d'une part les formations de faible degré (cycles) ont moins de chances de subir de la surqualification mais par contre elles ont plus de risques d'avoir des emplois qui ne correspondent pas. D'autre part, les formations de haut degré (bachelor, master, doctorat) ont des propensions plus élevées de subir de la surqualification alors que la non-correspondance est moins fréquente. Pour finir, le taux d'emploi peut également expliquer une part de la variance de nos données du fait que les emplois à temps partiels sont le plus souvent des emplois de basse qualité amenant des chances plus élevées de non-concordance de qualification avec l'emploi.

Le schéma suivant résume nos variables pour l'analyse:

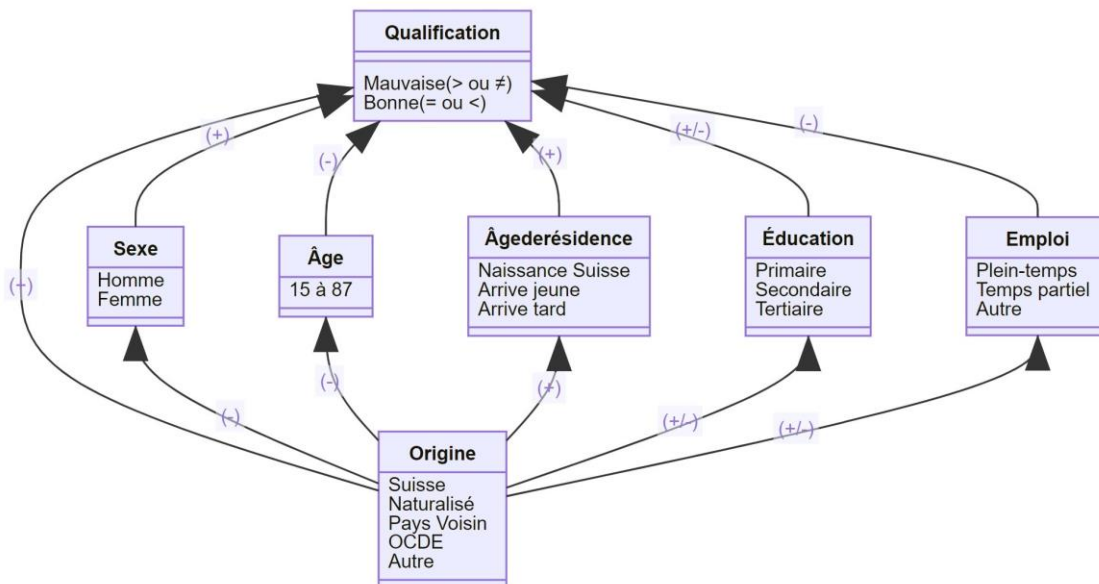
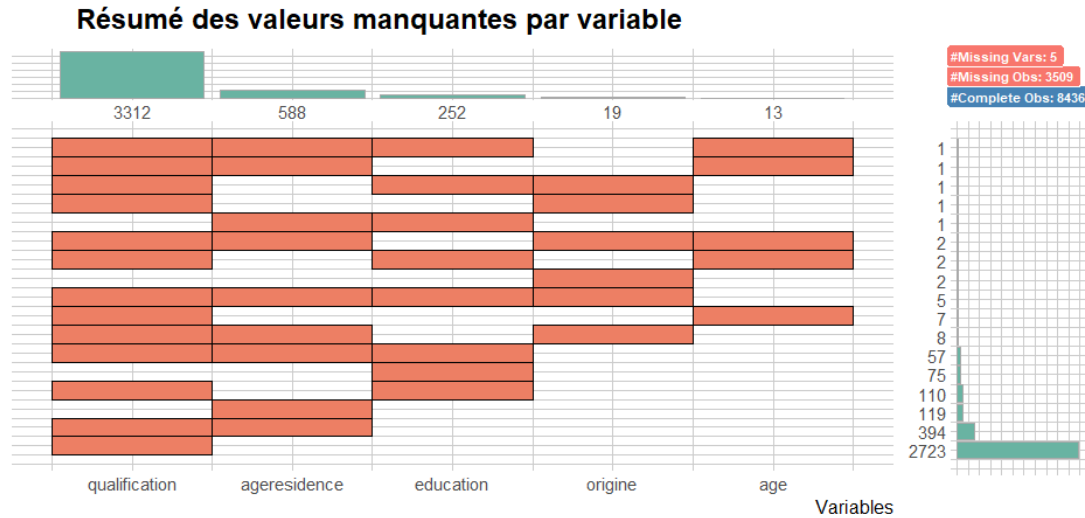


Schéma des variables (2)

Présentation du nombre de données manquantes par variable et globale

Nous observons ensuite le nombre de valeurs manquantes par variable, pour déterminer les éventuelles pertes. Avec la sélection de notre population, nous nous sommes retrouvés avec une population de 11'945 individus. Dans ces données nous avons les valeurs manquantes suivantes:

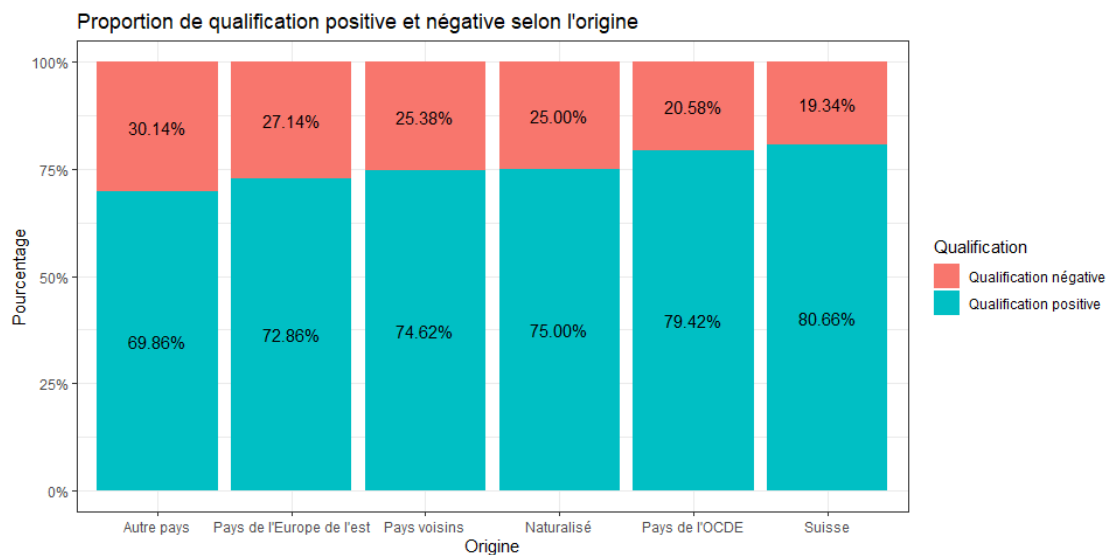


Nous remarquons que les variables sexe et emploi sont absente du graphique car elles n'ont pas de valeurs manquantes. Nous remarquons aussi que la variable avec le plus de valeurs manquantes est la qualification avec 3'312 valeurs manquantes dont 2'723 qu'elle a individuellement. Nous voyons que nous avons également 8'436 observations complètes (Compleat Obs). Pour notre analyse, nous allons exclure toutes les variables manquantes, ce qui nous laisse avec un nombre total d'observations pour 8'436 pour 7 variables.

Cela a une implication importante sur notre analyse, car cela veut dire que notre travail comporte potentiellement des biais. Une partie spécifique en annexe sera entièrement dédiée à l'analyse des valeurs manquantes et des profils types.

Analyse exploratoire

Afin de vérifier s'il existe une corrélation entre le pays d'origine et la qualification négative, nous avons effectué un tableau croisé en regardant les proportions d'observations entre qualification positive et négative selon le pays ou groupe de pays d'origine. Voici le graphique qui résume cette répartition:



L'association semble corroborer à première vue notre hypothèse principale à savoir l'augmentation de la qualification négative selon la nationalité. En effet, les Suisses et les pays de l'OCDE sont les plus sous-représentés dans la qualification négative (OCDE légèrement plus que la Suisse) et les pays de l'Est et autres sont les plus surreprésentés.

Ces résultats peuvent s'expliquer de plusieurs manières: tout d'abord, les personnes provenant des pays de l'Est et des autres pays du monde peuvent être surreprésentés dans la qualification négative du fait des effets de réseau, de reconnaissance du diplôme et de manque d'information sur le marché du travail suisse que nous avons évoqué précédemment. En sus, il pourrait s'agir des effets proprement de discrimination où l'éloignement culturel joue un rôle dans la manière de traiter, juger et évaluer les travailleurs provenant d'autres origines.

Par ailleurs, il est intéressant de noter que les catégories "pays voisins" et "naturalisés" ont des résultats légèrement différents qu'attendus, car ces catégories, donc proches culturellement, sont plus représentées dans qualification négative que les pays de l'OCDE.

De plus, la catégorie naturalisée est tout de même légèrement plus élevée en termes de qualification négative que les Suisse. Cela pourrait s'expliquer du fait que les 12 ans minimum passés en Suisse pour obtenir la nationalité Suisse peut compenser en partie les effets de réseau, diplôme et connaissances (Liebig, 2012). Par ailleurs, nous pouvons évidemment imaginer que des effets de discriminations peuvent expliquer ce taux dans l'autre sens.

Khi deux et résidus standardisés

Nous avons effectué un test du khi-deux et un V de Cramer pour observer s'il y avait un lien entre notre variable dépendante et notre variable explicative principale. Nous avons trouvé que le lien est effectivement significatif ($p=0.01634^*$) mais faible selon le V de Cramer. Avec principalement une sous représentation de Suisse dans les qualification négative, mais une surreprésentation de ressortissants de pays voisins, selon le tableau des résidus standardisés, nous voyons que seule une part de notre hypothèse semble validée. Le tableau suivant des résidus standardisés résume nos observations:

Tableau des résidus standardisé du Khi deux entre la nationalité et la qualification

Variables	Qualification positive	Qualification négative
suisse	4.60	-4.60
pays voisins	-3.03	3.03
Naturalisé	-2.74	2.74
pays de l'OCDE	-0.13	0.13
autre	-2.11	2.11
pays de l'est	-1.44	1.44

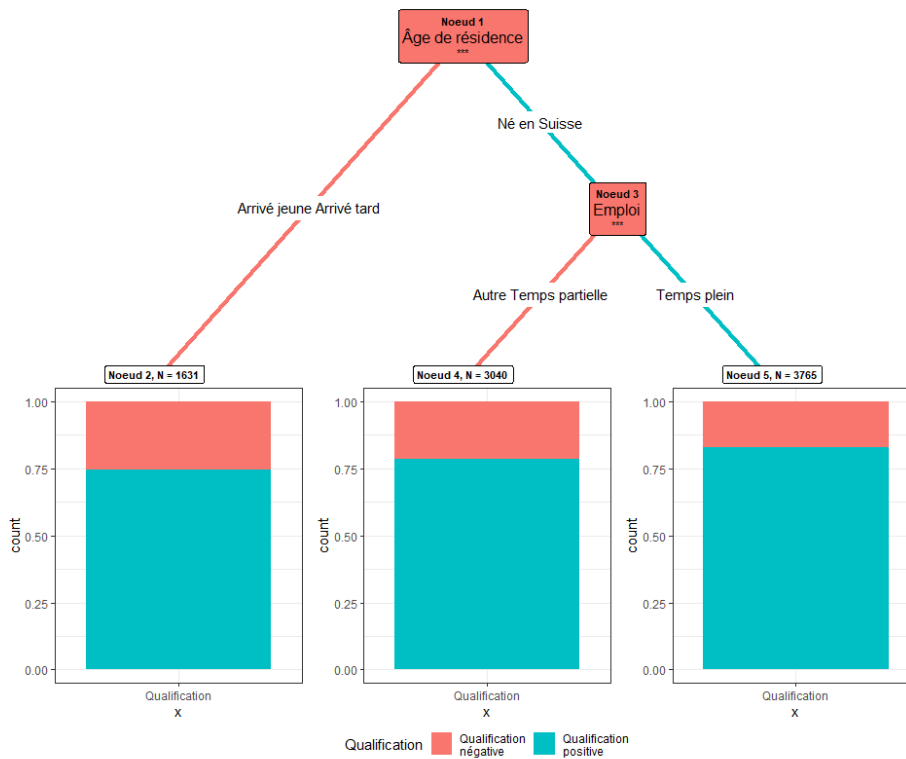
Khi2 p-value = 9.6006461441738e-05 ; V de Cramer = 0.06

Dans un tableau de résidus standardisé, les valeurs dépassant 1.96 indique une surreprésentation et les valeurs inférieure à -1.96 indique une sous représentation. Nous voyons donc que nous avons une sous-représentation des Suisses dans la qualification négative, ce qui corrobore nos hypothèses. Nous avons également une surreprésentation dans les catégories des pays voisins, des autres pays et des naturalisés. La surreprésentation semble plus forte du côté des pays voisins toutefois. Par contre, les autres groupes quant à eux ne sont pas significatifs étant donné que leurs résidus sont en valeurs absolue inférieure à 1.96. Aussi, la non-significativité pourrait s'expliquer dans le groupe "des "Pays de l'Europe de l'est" par un manque de données suffisantes. Effectivement, comme nous l'avons vu plus haut, la proportion de qualification négative est élevée dans ces groupes. Mais l'effet reste plus fort pour les pays voisins. Néanmoins, nous pouvons observer que notre hypothèse principale peut être généralisée à la population. La p-valeur étant inférieure à 0,05, nous pouvons dire que l'effet général de la nationalité sur la qualification négative est significatif. Nous devons vérifier si ce lien est toujours significatif en employant un modèle plus complexe prenant en compte plusieurs variables.

Analyse par arbre

Afin de détecter d'éventuels effets d'interaction, nous avons utilisé un arbre de classification. Dans le principe, l'arbre de classification a pour objectif de classer la population de la manière la plus distinctive possible en utilisant les variables dépendantes qu'on lui soumet pour regrouper nos observations en fonction du résultat sur la variable dépendante. Elle divise à chaque niveau la population en deux en utilisant une variable à un niveau distinctif (qu'elle soit catégorielle ou continue). Le tableau suivant illustre le résultat de cet arbre de classification:

Classification par arbre (1)

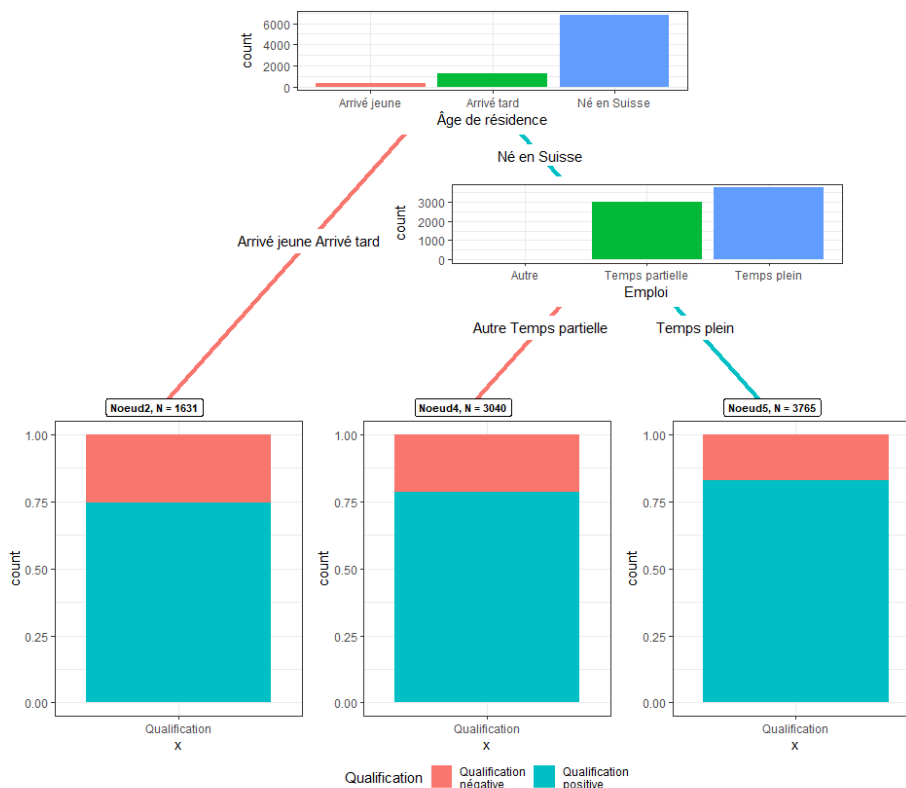


Lorsque nous observons ce découpage, nous pouvons voir que l'âge de résidence et l'emploi pourraient avoir un effet d'interaction. Nous devons tout de même relever que les différences ne sont pas très grandes d'une catégorie à une autre. Nous pouvons décrire les catégories de cette manière:

Les personnes n'étant pas nées en suisse, avec une plus grande proportion d'individus en situation de qualification négative (Noeud 2) Les personnes nées en Suisse ne travaillant pas à plein temps (Noeud 4) avec une proportion de personnes en situation de qualification négative légèrement plus basse Les personnes nées en Suisse travaillant à plein temps, avec la plus basse proportion de personnes en situation de qualification négative (Noeud 5)

Nous pouvons également constater que l'arbre de classification a fait un bon travail en ayant des catégories suffisamment grandes de taille relativement similaire. Cela est dû principalement par le regroupement de petites catégories.

Classification par arbre (2)



Il n'est pas étonnant de constater ces regroupements car la plus grande partie des travailleurs sur le marché de l'emploi sont suisses et travaillent à plein temps (Liebig et al., 2012). En nous basant sur les catégories produit par l'arbre de classification, nous obtenons le tableau suivant:

Distribution des groupes générés par l'arbre de classification par niveau de qualification

groupe	Qualification positive	Qualification négative
Né en Suisse à temps partiel	78.4%	21.6%
Pas né en Suisse	74.5%	25.5%
Suisse à plein temps	83.1%	16.9%

Nous voyons donc que la principale différence se trouve entre les personnes nées en Suisse à plein temps et les personnes n'étant pas nées, avec en proportion de personnes en situation de qualification négative respectivement 16.9 et 25.5%, soit 8.6 points de pourcentage de différence. Dans notre modélisation, nous allons donc créer un modèle prenant spécifiquement en compte l'interaction entre l'emploi et l'âge de résidence en Suisse.

Modélisation

Nous allons réaliser au total quatre modèles emboîtés, chacun avec leurs spécificités. Notre but est d'une part de vérifier nos hypothèses dont notre hypothèse principale sur le lien entre la nationalité et la probabilité de finir dans une situation de mauvaise qualification, d'autre part de tester l'effet d'interaction produit par l'arbre de classification au chapitre précédent. Etant donné que nos variables sont la plupart catégorielles nous allons utiliser une régression logistique. Notre variable est codée de sorte à estimer les coefficients pour les chances d'être dans une situation de qualification négative. Dans ce qui suit, nous allons décrire chaque modèle et leur spécificité.

Description des modèles

1. **Modèle simple bivarié:** Ce modèle a pour objectif d'estimer le lien simple avec notre variable dépendante, la qualification, et notre variable explicative principale, la nationalité.
2. **Modèle avec variables de contrôle:** Ce modèle se base sur le précédent, mais ajoute les variables de contrôle âge, sexe, niveau d'éducation et emploi. Les variables de l'âge et du sexe sont importantes car elles sont structurelles et influencent la trajectoire des individus de manière générale, surtout sur le marché de l'emploi (par exemple, les femmes ont plutôt tendance à travailler à temps partiel). Le niveau d'éducation et l'emploi sont importants à contrôler car ils ont un impact sur la possibilité d'être en situation de surqualification par exemple. Comme nous l'avons vu plus tôt, plus le niveau d'éducation est élevé, plus le nombre d'emplois où l'individu serait en situation de surqualification augmente. Concernant l'emploi, les métiers à temps partiel peuvent également représenter des emplois demandant un faible niveau de qualification.
3. **Modèle avec variable médiatrice:** Afin de ne pas surestimer l'effet de la nationalité sur la qualification négative, nous intégrons la variable de l'âge de résidence. En effet, comme nous l'avons évoqué précédemment les effets de réseau, de reconnaissance de diplôme et d'information concernant le marché du travail suisse peuvent être prépondérants pour expliquer des niveaux de qualifications négatives plus élevés dans la population étrangère. Ainsi l'âge d'arrivée en Suisse devrait influencer significativement la qualification négative: la catégorie "Arrivé après 18 ans" devrait avoir une proportion plus élevée de personnes en situation de qualification négative en comparaison des individus arrivés entre 18 ans et 1 an et encore plus que ceux étant nés en Suisse.
4. **Modèle avec un effet d'interaction:** dans ce dernier modèle nous ajoutons l'effet d'interaction mis en évidence par l'arbre des classifications, c'est-à-dire entre l'âge de résidence et l'emploi. Le but étant de voir si à période d'âge de résidence en Suisse différente, l'emploi influence différemment les chances de se retrouver en situation de qualification négative.

Modèles emboîtés

Le tableau suivant récapitule les résultats de nos quatre modèles:

	(1)	(2)	(3)	(4)
Pays voisin	0.350 *** (0.104)	0.372 *** (0.105)	0.015 (0.137)	0.005 (0.137)
Naturalisé	0.330 ** (0.107)	0.335 ** (0.108)	0.112 (0.122)	0.108 (0.122)
Pays de l'OCDE	0.078 (0.152)	0.096 (0.154)	-0.261 (0.177)	-0.265 (0.178)
Autres pays	0.587 * (0.257)	0.583 * (0.259)	0.191 (0.276)	0.169 (0.277)
Pays de l'est	0.441 (0.270)	0.453 (0.272)	0.140 (0.284)	0.134 (0.284)
Femme		0.021 (0.064)	0.010 (0.064)	0.011 (0.064)
Age		-0.004 (0.002)	-0.006 * (0.002)	-0.006 * (0.002)
Éducation Secondaire		-0.095 (0.125)	-0.076 (0.125)	-0.082 (0.125)
Éducation Tertiaire		0.001 (0.123)	-0.005 (0.124)	-0.010 (0.124)
Temps partiel		0.282 *** (0.064)	0.289 *** (0.064)	0.314 *** (0.071)
Autres emplois		1.257 ** (0.426)	1.291 ** (0.427)	1.220 * (0.488)
Arrivé après 18 ans			0.444 *** (0.109)	0.515 *** (0.128)

Arrivé avant 18 ans			0.248 (0.136)	0.229 (0.180)
Temps partiel x Arrivé après 18 ans				-0.170 (0.148)
Autres emplois x Arrivé après 18 ans				0.923 (1.258)
Temps partiel x Arrivé avant 18 ans				0.066 (0.254)
Autres emplois x Arrivé avant 18 ans				-11.236 (196.968)
<hr/>				
N	8436	8436	8436	8436
logLik	-4239.208	-4220.650	-4212.354	-4210.664
AIC	8490.415	8465.299	8452.709	8457.328

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

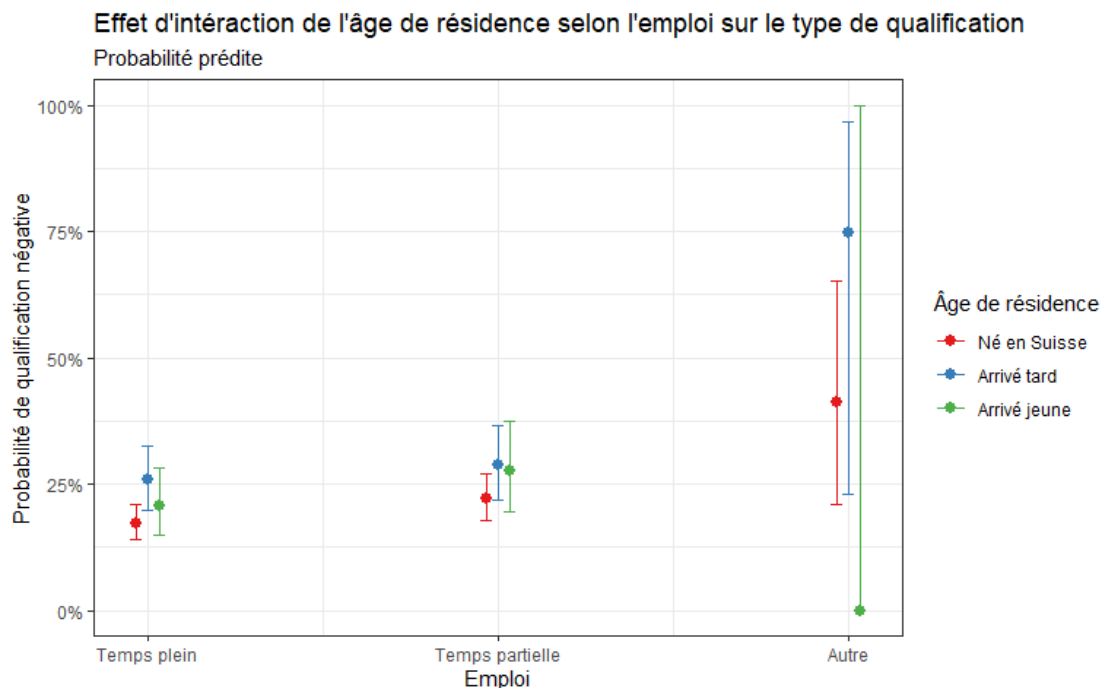
Premièrement, nous constatons que dans notre premier modèle, notre variable explicative principale, c'est à dire l'origine, est significative. Ainsi les travailleurs naturalisés, provenant de pays voisins ou d'autres pays ont plus de chance de se retrouver dans une situation de qualification négative avec des coefficient de, respectivement, 0.330, 0.350 et 0.587 log odds.

Nous soulignons toutefois que pour les travailleurs issus de l'Europe de l'est, le coefficient est élevé (0.441), bien plus que pour les ressortissants de pays voisins. Pourtant ce n'est pas significatif. Ce qui expliquerait cette situation est le petit effectif de cette population dans nos données. Effectivement, le coefficient "Pays de l'Europe de l'est" partage avec le coefficient "Autres pays", une erreur standard élevée, respectivement, de 0.270 et 0.257. Cela semble lié aux effectifs de chaque population, 70 pour le groupe provenant des pays de l'Europe de l'est et 73 pour le groupe provenant d'autres pays. Nous gardons cette remarque pour la partie de discussion de notre article.

Dans le second modèle, à l'ajout de nos variables de contrôle, nous voyons que l'effet reste significatif pour notre variable sur l'origine. Nous notons toutefois que l'effet de l'emploi est significatif. De plus, travailler à temps partiel ou travailler soit dans un atelier protégé pour personnes handicapées soit dans une entreprise familiale (Catégorie "Autre emploi"), augmente les chances de se retrouver dans une situation de mauvaise qualification par rapport à une personne travaillant à plein temps.

Dans le troisième modèle, lorsque nous ajoutons notre variable de médiation concernant l'âge de résidence, nous voyons que pas mal de choses changent. Premièrement, cette variable est significative. Ainsi plus la personne est présente en Suisse à un âge avancé, plus elle a de chance de se retrouver dans une situation de qualification négative. Les personnes étant arrivées après l'âge de 18 ans, ont plus de chance de se retrouver dans une position de mauvaise qualification que celles nées en Suisse, toutefois l'effet n'est pas significatif pour les personnes arrivées en Suisse entre 1 an et 18 ans. Nous remarquons également que l'ajout de cette variable a pour effet, à la fois de réduire les coefficients de la variable sur l'origine et de les rendre non significatifs. Cela correspond bien à un effet de médiation sur notre variable principale. Cela pourrait s'expliquer par le fait que la variable de médiation et notre variable explicative principale soient très liées, mais que ce soit notre variable médiatrice qui soit plus explicative. Ainsi, la majorité des Suisses ont tendance à naître en Suisse et les étrangers ont tendance à arriver après, mais ce serait le moment de la présence en Suisse qui détermine l'effet sur la qualification. Nous vérifierons cela plus tard. Notons toutefois que pour les ressortissants des pays de l'OCDE, le coefficient s'inverse lorsque nous prenons en considération l'âge de résidence en Suisse, faisant qu'ils auraient moins de chance que les Suisses d'être en situation de qualification négative. Puisque les coefficients ne sont pas significatifs, nous garderons cela pour la discussion. Finalement, la variable âge devient significative, avec un coefficient négatif. Cela signifie que plus une personne est âgée, moins elle a de chance de se retrouver dans une situation de qualification négative.

Finalement le quatrième modèle ajoutant l'effet d'interaction entre l'emploi et l'âge de résidence n'apporte rien d'important au modèle. Aucun coefficient de l'interaction n'est significatif. Nous pouvons voir dans notre graphique d'interaction que cela est dû aux petites variations entre les groupes et aux larges erreurs standards:



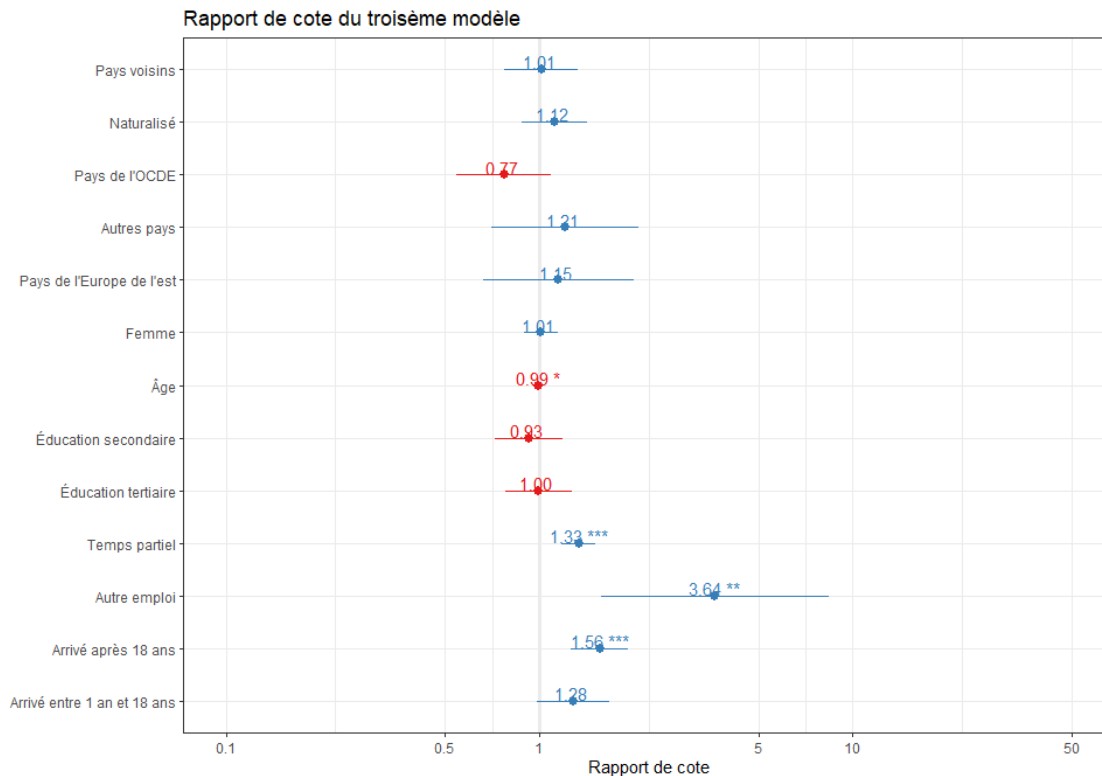
Choix du modèle

Nous avons réalisé deux tests pour choisir quel modèle est le plus adapté et de meilleure qualité. Nous avons premièrement réalisé un test de rapport de vraisemblance pour estimer la différence de déviance entre nos modèles. Voici un tableau synthétisant les résultats des deux tests:

Synthèse des tests sur les modèles

Modèles	BIC	p
Model 1	8,532.657	
Model 2	8,549.783	0.000001671567
Model 3	8,551.272	0.000249680167
Model 4	8,584.053	0.496296312165

Premièrement, nous pouvons afficher les rapports de cote pour nos coefficients pour avoir une première idée sur l'amplitude de nos coefficients. Nous illustrons cela avec le graphique suivant:



Les points en rouge indiquent un effet négatif et les points en bleu un effet positif. Dans ce graphique, il est intéressant de relever qu'une personne travaillant dans un autre type d'emploi (atelier protégé, compagnie familiale) ont 3.64 fois plus de chance de se retrouver

dans une situation de qualification négative que les individus travaillant à temps plein, ce qui est compréhensible au vu de la situation de ces personnes qui sont quelque part contraintes d'exercer ce type de profession. Pour les autres coefficients, il est difficile de dire quelque chose d'intéressant avec leur amplitude à l'aide des rapports de quote. Notons néanmoins que le coefficient pour l'âge est très petit. Nous allons donc utiliser les Average Mean Effect (AME) pour observer en moyenne la probabilité de chance de se retrouver en qualification négative qu'implique chaque coefficient significatif. Voici le tableau des AME en ne gardant que les coefficients significatifs:

Variables	AME	SE	z	p	lower	upper
Age	-0.001	0.000	-2.413	0.016	-0.002	0.000
Autre	0.263	0.104	2.521	0.012	0.058	0.467
Temps Partielle	0.047	0.011	4.442	0.000	0.026	0.067
Arrivé Tard	0.077	0.020	3.791	0.000	0.037	0.117

Ainsi, nous pouvons confirmer que l'effet de se retrouver dans un emploi de type autre augmente les chances de se retrouver en situation de qualification négative de 26.3%, ce qui est grand si on compare au temps partiel qui augmente les chances de 4.7%. Ensuite le fait d'arriver tard, soit après 18 ans en Suisse augmente les chances de se retrouver en qualification négative de 7.7%, ce qui est grand comme amplitude. Finalement, nous voyons que pour l'âge, une augmentation d'une année augmente les chances de 0.1% ce qui est très faible (1% pour 10 ans).

Discussion

L'impact de l'origine sur la surqualification et de la non-concordance entre qualification et emploi est un enjeu essentiel tant en termes d'égalité de traitement que d'efficacité d'allocation des compétences.

Comment expliquer ce lien, quelles en seraient les causes? La question de la discrimination par le goût, revenant à des formes de discrimination conscientes ou inconscientes des Suisses envers les étrangers du fait d'une distance culturelle, est un effet présent dans la littérature et nous avons essayé de mesurer si un écart existait. Ainsi en prenant le premier modèle de régression entre nationalité et qualification, nous pouvons observer que la qualification négative était significativement plus prononcée dans les catégories des pays voisins, autres pays que chez les Suisses corroborant notre hypothèse principale. Par contre, notre seconde hypothèse postulant une stratification de l'intensité de qualification négative selon l'éloignement culturel des pays semble ne pas tenir complètement. Notons tout de même le résultat inattendu des Naturalisés dans le premier modèle, qui se retrouvent avec plus de chance que les Suisses d'être en situation de qualification négative. Pour être naturalisé il faut au minimum 12 ans de résidence légale sur le territoire Suisse, ce qui montrerait une assez bonne connaissance du marché du travail suisse et potentiellement un bon réseau. Cependant, à l'ajout de la variable sur l'âge de résidence, il n'y a plus de significativité. Ce qui

peut suggérer que la plupart des naturalisés sont arrivés à un âge avancé et n'ont probablement pas eu de reconnaissance de diplômes.

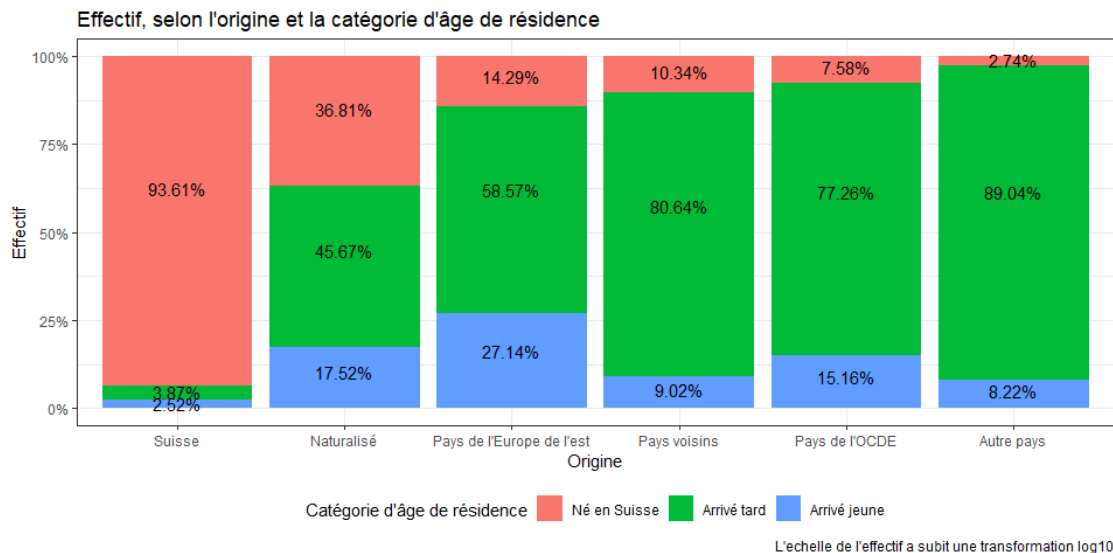
En effet, les pays d'Europe de l'Est et de l'OCDE ne montrent pas d'effet significatif. Même si pour les pays d'Europe de l'Est l'explication devrait probablement venir d'un jeu de données trop réduit, le cas de l'OCDE est assez surprenant, car cette catégorie présente un effet non significatif et quasiment nulle alors qu'on aurait du voir un effet plus prononcé dans les pays de l'OCDE que les pays voisins par exemple. Cependant, on peut tout de même voir un effet plus prononcé chez les pays "autre" donc plus éloignés culturellement (donc pouvant plus être sujet à la discrimination par le goût) que les pays voisins. Nous notons également le fait que pour les ressortissants des pays de l'OCDE, les proportions de qualification négative étaient très similaires à celles des Suisses et même si les coefficients n'étaient pas significatifs dans les modèles avancés, le coefficient devient négatif lorsque l'on prend en considération l'âge d'arrivée.

Evidemment un grand nombre de facteurs, autres que la discrimination peuvent expliquer le fait que les étrangers subissent plus de qualification négative. Ainsi dans notre deuxième modèle nous avons ajouté des variables de contrôles.

Notre quatrième hypothèse portant sur l'éducation n'est pas corroborée par nos données. En effet, les individus ayant une éducation tertiaire ou secondaire semblent moins affectés par la qualification négative que les personnes ayant une formation primaire, l'effet n'est pas significatif. Ce résultat pourrait s'expliquer par l'agglomération des deux catégories de qualification négative: non-concordance et surqualification qui répondent à des mécanismes différents et possiblement contradictoires. Effectivement, les individus d'éducation tertiaire seraient plus touchés par la surqualification alors que les personnes d'éducation primaire par la non-concordance. Mais les autres variables du modèle jouent aussi un rôle dans les valeurs de ces coefficients.

Au-delà de ces variables de contrôle, nous avons intégré une variable clé à notre troisième modèle de régression: l'âge de résidence. Pour rappel, elle est construite comme un proxy nous permettant de contrôler trois effets concomitants pouvant jouer un rôle clé dans le fait de subir une qualification négative: la possibilité de développer un réseau, la concordance des diplômes et l'acquisition de connaissances concernant le marché du travail Suisse. Nous avons postulé que plus un individu arrive tôt en Suisse, plus il aura de chances d'avoir la capacité de développer ces trois éléments. Même si l'ajout de cette variable ne change pas la significativité des variables de contrôles, elle enlève la significativité pour les groupes des pays voisins, des autres pays et des Naturalisés. Ce phénomène peut être interprété comme un effet de médiation entre de l'âge de résidence entre la nationalité et la qualification négative. En effet, cela signifie qu'une partie de la variance expliquée à la base par le lien entre nationalité et qualification négative provient au final de l'âge d'arrivée en Suisse. Ainsi nous pouvons affirmer que la troisième hypothèse est vérifiée: les individus arrivés après 18 ans ont significativement plus de chance de subir des qualifications négatives, tout comme ceux arrivés entre leurs 1ère et 18ème année que les individus étant nés en Suisse (même si ce n'est pas significatif pour le dernier cas). Par conséquent, nous pouvons dire que l'âge de résidence serait plus explicatif concernant la propension de subir de la qualification négative que la nationalité.

Toutefois, il faut mentionner que la variable nationalité et âge de résidence partagent une corrélation élevée. En effet, la catégorie née en Suisse est proportionnellement bien plus représentée par les suisses et inversement arrivée tôt et arrivée tard sont proportionnellement plus présentes dans les catégories étrangers. Ainsi il est difficile de pouvoir affirmer que les effets de réseau, de reconnaissance de diplôme et de connaissance du marché du travail sont réellement les facteurs qui expliquent plus que la discrimination par le goût la propension plus grande que des étrangers (ou naturalisé) à subir de la qualification négative.



Pour finir, il est complexe d'affirmer que nos résultats s'inscrivent dans la continuité des recherches sur la qualification négative. En effet, cette dernière a mis en évidence les effets de discrimination de goût alors que la notre tendrait avec notre troisième modèle à l'infirmer. Cependant en prenant en compte la colinéarité de l'âge de résidence avec la nationalité, cette dernière doit être interprétée avec précaution. Surtout lorsque l'on s'interroge de manière plus approfondie sur notre variable proxy au delà de la distribution proche de celle de nationalité: dans quelle mesure est-ce qu'elle interroge correctement les effets de réseau, reconnaissance de diplôme et connaissance du marché du travail et non une forme d'assimilation des étrangers dans la société suisse diminuant la distance culturelle avec les Suisses et, par extension, diminuant le risque d'être discriminé.

Conclusion

Après avoir effectué des analyses bi-variées à travers un tableau croisé, un khi-carré et un χ^2 de Fisher, l'effet de la nationalité sur la qualification négative était présent mais assez léger. Notre premier modèle de régression logistique nous a confirmé ces impressions: l'effet est significatif pour certaines catégories (pays voisins, naturalisé et autres) mais pas pour les pays de l'est et l'OCDE. L'ajout de notre variable d'âge de résidence permettant de vérifier les effets de réseau de réseau, de reconnaissance de diplôme et de connaissance du marché suisse, nous a montré qu'au final cette variable semble plus explicative que la nationalité, invalidant par là même notre première hypothèse au profit de la troisième.

L'interprétation de ces résultats est à prendre avec mesure, mais ils tendraient à ne pas corroborer forcément la littérature sur le sujet. En effet, alors que la discrimination par le goût semble être plus mis en avant comme facteur explicatif, nos résultats semblent plutôt indiquer que les effets de réseau de réseau, de reconnaissance de diplôme et de connaissance du marché suisse sont plus prépondérant.

Quatre grandes limites entravent notre recherche. Tout d'abord, le manque de données concernant un certain nombre de catégories d'étrangers. En effet, la catégorie pays de l'Europe de l'est en particulier possède peu d'observations rendant les effets possiblement perceptibles non significatifs. Pour cela une pondération aurait pu, selon nous, rendre ces effets significatifs. Ensuite, notre variable âge de résidence fonctionnant en proxy des effets de réseau, de reconnaissance de diplôme et de connaissance du marché de l'emploi n'est qu'une estimation de ces effets, surtout si l'on ajoute le fait que cette variable présente une forte corrélation avec la variable nationalité et peut également mesurer un effet d'assimilation des étrangers, réduisant la discrimination et non les effets voulu initialement. Pour mesurer ces effets correctement, il aurait été souhaitable de pouvoir les distinguer les uns des autres, par exemple en ayant le pays de provenance du diplôme, cependant ces données ne sont pas disponibles dans le Panel Ménage Suisse. Pour avoir une analyse plus fine, il aurait pu être intéressant de ne pas regrouper la non-concordance des qualifications et la surqualification. En effet, ces deux phénomènes sont impactés par des mécanismes différenciés en particulier en ce qui concerne le niveau d'éducation. Ainsi les populations qui subissent ces effets sont assez différentes. Pour cela, nous aurions pu ajouter une variable contrôlant le type d'emploi ou le secteur. Pour finir, nos données n'intègrent uniquement les étrangers travaillant en suisse et qui y résident, en sont donc exclus tous les frontaliers, alors qu'il aurait été intéressant d'observer leurs réponses à la qualification négative.

Malgré les nombreuses limites que comporte notre recherche, nous pensons qu'elle permet en partie de mettre la lumière sur un phénomène complexe qu'est la mauvaise qualification. Les causes peuvent être de différents ordres, notre travail a montré l'importance de l'âge de résidence et pointe vers des directions de recherche qui pourraient être intéressantes de creuser comme le réseau de la personne. Le sujet ayant une importance tant pour les travailleurs étrangers que pour la Suisse qui les accueille notamment en termes d'allocation adéquate des compétences, nous espérons y avoir suffisamment contribué à notre niveau.

Bibliographie

Bonvin, J.-M. (2009). La discrimination multiple sur le marché du travail suisse. *Tangram*, 23, 61-63.

Fibbi, R., Kaya, B., & Piguet, E. (2003). Le passeport ou le diplôme ? Étude des discriminations à l'embauche des jeunes issus de la migration. Neuchâtel: SFM, 1-64.

Liebig, T., Kohls, S., & Krause, K. (2012). L'intégration des immigrants et de leurs enfants sur le marché du travail en Suisse (OECD Social, Employment and Migration Working Papers No 128; OECD Social, Employment and Migration Working Papers, Vol. 128).

<https://doi.org/10.1787/5k9fjrtlwwl-fr>

OFS. (2021). Adéquation entre niveau de formation et activité exercée [Officiel]. admin.ch. <https://www.bfs.admin.ch/bfs/fr/home/statistiques/population/migration-integration/indicateurs-integration/indicateurs/surqualification.html>

Zschirnt, E., & Fibbi, R. (2019). Do Swiss Citizens of Immigrant Origin Face Hiring Discrimination in the Labour Market? National Center of Competence in Research – The Migration-Mobility Nexus, Working paper(20). https://nccr-onthefmove.ch/wp_live14/wp-content/uploads/2019/02/nccrotm-WP20-Zschirnt-Fibbi_Feb19.pdf

Annexe Étude des données manquantes

Afin de voir s'il y a un lien entre nos variables et les données manquantes, nous allons effectuer plusieurs tests statistiques d'association (principalement khi2 et corrélation de spearman). Si le lien est déterminé, c'est que nous avons une sous ou surreprésentation de certaines populations dans nos données manquantes. Cela a pour effet de générer du biais dans nos analyses. Voici le tableau résumant le résultat de ces multiples tests:

Variabes	p-value
Qualification	0.00215
Nationalité	0.00000
Âge de résidence	0.00000
Éducation	0.00000
Emploi	0.00000
Sexe	0.00000
Âge	0.00000

Comme nous pouvons le voir dans nos données, la relation entre les variables et les données manquantes est à chaque fois significative, ce qui veut dire que pour chaque variable, nous avons systématiquement une proportion importante de population manquante. Nous pouvons d'avance indiquer que nos résultats sont biaisés et difficilement généralisables.

Nous allons observer les profils manquants pour chaque variable afin de déterminer ce qui nous a manqué et les éventuels biais. Tous les tableaux qui vont suivre représentent les résidus standardisés. Pour rappel, les valeurs dépassant 1.96 indique une surreprésentation et les valeurs inférieure à -1.96 indique une sous représentation.

Premièrement, concernant la qualification, nous voyons que les personnes en situation de qualification négative ont tendance à être surreprésentées dans les variables manquantes. Cela veut dire que nous avons exclu proportionnellement plus de personnes en situation de qualification négative. Nous avons donc probablement un biais de représentativité.

Qualification	Complet	Manquant
Qualification positive	3.157959	-3.157959
Qualification négative	-3.157959	3.157959

Deuxièmement, concernant la nationalité, nous pouvons voir que toutes les nationalités étrangères sont surreprésentées dans les données manquantes.

Nationalité	Complet	Manquant
Autre	-11.722224	11.722224
Naturalisé	10.720410	-10.720410
Pays De L'est	-10.696115	10.696115
Pays De L'ocde	-8.890320	8.890320
Pays Voisins	-7.420983	7.420983
Suisse	11.018776	-11.018776

Troisièmement, concernant l'âge de résidence, les personnes nées en Suisse sont surreprésentées parmi les données manquantes.

Age de résidence	Complet	Manquant
Né en Suisse	-5.1092884	5.1092884
Arrivé tard	6.3284679	-6.3284679
Arrivé jeune	-0.9443557	0.9443557

Quatrièmement, concernant le niveau d'éducation, il y a une surreprésentation des personnes n'ayant pas fait d'étude de niveau tertiaire.

Éducation	Complet	Manquant
Éd. Primaire	-11.849444	11.849444
Éd. Secondaire	-6.660871	6.660871
Éd. Tertiaire	12.667711	-12.667711

Cinquièmement, concernant le taux d'emploi, les personnes n'étant pas en temps partiels sont surreprésentées parmi les données manquantes.

Emploi	Complet	Manquant
Temps plein	-7.667976	7.667976
Temps partiel	8.562349	-8.562349
Autre	-5.960734	5.960734

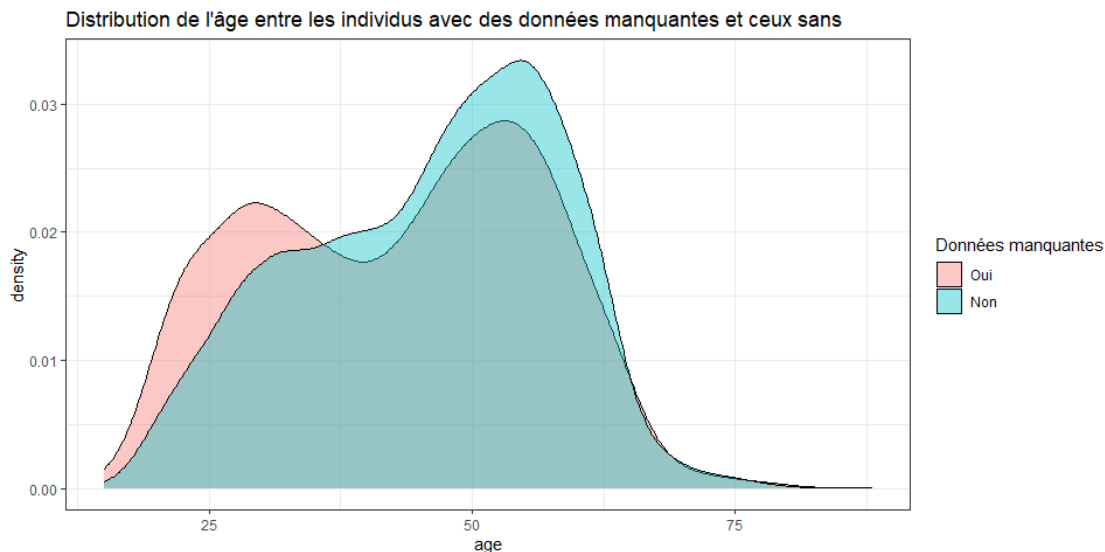
Ensuite, concernant le sexe, nous avons une surreprésentation d'hommes dans les données manquante.

Sexe	Complet	Manquant
Homme	-6.334454	6.334454
Femme	6.334454	-6.334454

Finalement, concernant l'âge nous avons un effet peu clair puisqu'elle est continue. Il semblerait que la population des personnes exclues soit plus jeune.

Manquant	Moyenne d'âge
Non	45.90031
Oui	43.10441

Nous pouvons vérifier cela en observant la distribution de l'âge parmi nos deux populations:



Ainsi, nous voyons effectivement que la population dans notre échantillon de données manquantes est plus jeune.

Nous pouvons donc affirmer que notre profil type de personne exclue de notre travail est une personne en situation de qualification négative, non-suisse, mais né en Suisse, n'ayant pas fait des études de niveau tertiaire, à temps pleins ou dans un autre type d'emploi, de sexe masculin et jeune. Sachant que cela représente une population de 3312 personnes soit environ 28% de notre population de référence, les biais sont importants.