

JOURNAL OF INTERDISCIPLINARY HISTORY OF IDEAS



2021

Volume 10 Issue 19

Item 3

– Section 2: Articles –

Comment évaluer la discrimination raciale et ethnique sur le marché du travail

L'usage de la régression multiple aux Etats-Unis depuis les
années 1960

par

Christophe A. Brochier



JIHI 2021

Volume 10 Issue 19

Section 1: Editorials

1. *Next Call for Papers: Hybridization in the History of Ideas* (JIHI)

Section 2: Articles

2. *A Complete Atheist: Jean Meslier's Political Philosophy* (C. Develennes)
3. *Comment évaluer la discrimination raciale et ethnique sur le marché du travail. L'usage de la régression multiple aux Etats-Unis depuis les années 1960* (C. A. Brochier)
4. *De l'intérêt égoïste à l'empathie : vers de nouveaux fondements pour l'économie politique?* (E. Blanc)

Section 3: Notes

5. *Research Report | The Radical Translations Project: Some Challenges in Using Translation as an Approach to Revolutionary History* (S. Perovic)
6. *Founding Stone: A discussion of Charles T. Wolfe's Lire le matérialisme* (P. D. Omodeo, C. Wolfe)

Section 4: Reviews

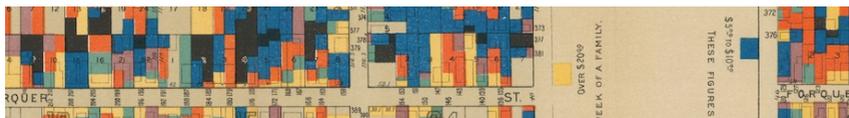
7. *Quelle est la nature du tournant économique? Essay/Review* (A. Tiran)
 8. *Book Reviews* (A. Mattana, M. Morabito, A. Skornicki, R. Tortajada)
-

Comment évaluer la discrimination raciale et ethnique sur le marché du travail

L'usage de la régression multiple aux Etats-Unis depuis les années 1960

Christophe A. Brochier *

This text aims at looking at some advances and above all some problems coming from the study of 'racial' discrimination ('whites' and 'non whites') in the labour market (hiring, promotion, wages) in the United States since the 1960s by statistical models of multiple regression. In a first part we describe the progressive generalization of this method in the 1960s and 1970s for this sujet, showing an evident lack of precision in the beginnings on methodological issues. In a second part we show that many important problems remain in spite of a tendency of methodological improvements. We suggest in conclusion a discussion about the usefulness, after half a century of practices, of trying to 'explain' discrimination with this kind of tools.



L'étude des discriminations « raciales » sur le marché du travail présente deux difficultés relevées par Herbert Blalock en 1961. Le chercheur doit montrer que la discrimination est un phénomène social assez général pour être signalé, mais aussi établir le lien qu'elle entretient avec les écarts (de revenus, de position, d'emploi) constatés entre les groupes qu'il veut comparer. Il est en effet aisé de prouver l'existence de comportements discriminatoires par des entretiens auprès d'employeurs (Bastide et Fernandes 1955, Dewey 1952), ou par des histoires

* Laboratoire CREDA (IHEAL-PARIS III) / Université Paris 8 (chrisbrochier@yahoo.com).

de vie (Terkel 2010). À défaut, l'utilisation aujourd'hui bien établie, de tests expérimentaux (faux CV envoyés à des recruteurs ou faux candidats auditionnés) permet de révéler des pratiques que les entreprises préfèrent cacher (Carlsson et Rooth 2006). Plus simplement, la mise en évidence d'attitudes racistes assez généralisées permet légitimement de supposer des effets discriminatoires sur le marché du travail. En revanche, montrer que des écarts de salaires ou de niveaux d'emplois sont dus à des traitements discriminatoires est plus délicat. Pour ce faire, il devient nécessaire de faire la preuve que d'autres facteurs n'expliquent pas mieux les faits étudiés. Dans les cas où l'existence de la discrimination au travail est sujette à débat, comme au Brésil, c'est d'ailleurs en proposant une évaluation chiffrée d'écarts « non explicables » que l'on cherche souvent à prouver l'existence du racisme dans le monde du travail (Silva, 1980). La question des méthodes est donc centrale dans l'étude de la discrimination. L'administration de la preuve doit pouvoir établir de façon convaincante, notamment en évaluant des effets causaux, le lien entre une caractéristique « raciale » et certaines situations d'emploi. Les enquêtes de type ethnographique ou psychosociologique mobilisées, par exemple par Myrdal dans sa synthèse de 1944, permettaient d'illustrer les différentes modalités d'existence du racisme aux Etats-Unis, mais elles ne pouvaient établir de façon concrète et précise l'effet du racisme sur les revenus et l'emploi. Ce sont les travaux statistiques qui se développent aux Etats-Unis à partir des années 1950, sur la base de données de recensement, qui vont donc compléter les connaissances. Le défi principal à relever était de faire la part entre les effets des caractéristiques sociales des membres du groupe défavorisé et les effets du racisme. Si les membres des minorités sont en moyenne moins payés, ce peut-être en raison de qualifications insuffisantes ou de comportements inadaptés aux attentes des employeurs. Il s'agit donc toujours de comparer le groupe favorisé et le groupe défavorisé en étudiant plusieurs variables qui décrivent les caractéristiques potentiellement valorisées par les employeurs. La méthode la plus utilisée par les sociologues (National Council 2004, Pager et Shepherd 2008) et une partie des économistes du travail¹

¹ Les économistes utilisent également d'autres approches (que je n'étudierai pas ici) fondées sur la mise en équation du fonctionnement du marché du travail à partir de d'hypothèses concernant les effets de la discrimination. Les prévisions obtenues sont alors comparés aux données : des écarts faibles servent d'indice de confirmation aux hypothèses (Kofi et Guryan 2011).

à partir des années 1960 a consisté à recourir à la régression multiple sur des échantillons tirés d'enquêtes démographiques. C'est à l'étude de l'application de cette technique que cet article est consacré. Mon approche consistera à combiner les perspectives de l'histoire des pratiques de recherche, déjà bien établie, et celle de la critique méthodologique. Cette dernière manque en général de recul historique car elle privilégie les recommandations de bon usage (par exemple : Becker 1958). L'histoire des méthodes telle qu'elle est habituellement conduite (par exemple : Goodson 2001, Platt 2002, Demetriou et Rodometof 2014) se limite essentiellement à décrire l'apparition d'une méthode et les formes ou les conditions de son extension éventuelle, voire de son déclin. La mise en lumière des limites ou des insuffisances des pratiques étudiées n'intervient généralement que par la prise en compte des débats qu'elles ont pu susciter. De cette façon, d'importants angles morts persistent : on s'interdit en effet de questionner les directions qu'ont prises les évolutions méthodologiques au regard des limites et des possibilités des pratiques. Certaines faiblesses des méthodes sont pourtant corrigées au fil des ans par les chercheurs alors que d'autres échappent à la critique. L'histoire des progrès méthodologique en sciences sociales (voir par exemple : Cook, 2008) tient compte de la correction des insuffisances par les praticiens mais ne saisit généralement les méthodes que comme des raisonnements en voie de perfectionnement. Elle ignore ainsi les résultats parfois fort variables et souvent critiquables, des divers utilisateurs.

A partir de ces constats, le but de cet article sera donc de montrer comment une méthode quantitative a été appliquée à l'étude d'une question sociologique (l'étude de la discrimination « raciale » sur le marché du travail) des origines à l'époque actuelle. Il s'agira de comprendre pour quels usages a débuté son application, avec quels présupposés, sur la base de quelles discussions techniques, puis comment se sont faites les évolutions par rapport aux résultats. Je regarderai comment certaines imperfections de la méthode ont été corrigées, ou bien, au contraire, se sont maintenues. Dans le cas qui nous intéresse, il est ainsi bien connu que la régression multiple présente de nombreuses difficultés (que j'examinerai) dès lors que l'on se propose de lui faire jouer un rôle explicatif (et pas seulement prédictif). Certaines de ces faiblesses, dont le problème de l'évaluation des effets causaux, étaient perçues dès le début du XX^e siècle, et ont été régulièrement rappelées par des statisticiens. On peut dès lors se demander comment les praticiens ont pris en compte ces problèmes et comment

ils ont adapté les possibilités de la méthode à l'étude de la discrimination. En passant en revue un échantillon de travaux importants depuis les années 1950¹, je me propose donc d'examiner quels problèmes les utilisateurs de la régression ont cherché à dépasser par l'usage de cette méthode, quels sont ceux qu'ils ont laissés en suspens et de quelles façons se sont concrétisées les évolutions techniques. Le but de l'article n'est donc pas de proposer une critique générale de l'application de la régression linéaire aux sciences sociales, ni de pointer les négligences d'un chercheur ou d'un courant en particulier, mais de montrer comment une méthode de traitement du sujet s'est développée dans le temps². Suivre de décennie en décennie les problèmes de fond qui ne sont que partiellement affrontés est un moyen de comprendre certains mécanismes à l'œuvre dans l'évolution des sciences sociales, notamment la place du ritualisme méthodologique, de l'inertie et des arguments d'autorité. Ces questionnements me paraissent essentiels dans le cas de la sociologie quantitative dans la mesure où demeure encore vive l'idée que les enquêtes produisant des chiffres pourraient se prévaloir d'une sorte de supériorité scientifique. En France, une partie importante des sociologues professionnels connaissent peu l'évolution des méthodes quantitatives ; les débats sur la validité des résultats de travaux statistiques sur le racisme sont donc nimbés d'une sorte de flou autour de la solidité des acquis par les techniques statistiques. Ce texte permettra donc en conclusion de reformuler en partie les débats portant sur les apports et les possibilités de l'étude de la discrimination par la régression multiple.

¹ Pour des raisons d'homogénéité et de pertinence, je restreindrai mon analyse aux travaux américains publiés dans des revues visibles et qui furent régulièrement cités. Pour des raisons de faisabilité, je regarderai essentiellement les recherches portant sur les écarts de salaires (« pay gaps ») et de taux d'emploi entre blancs et « non blancs », sans ignorer que de nombreuses autres situations de discrimination ont également été étudiées. Je tiens également à signaler que ce texte a bénéficié des remarques décisives de deux lecteurs anonymes du JIHL.

² Afin d'éviter les confusions je précise donc que mon article ne cherche pas à faire le bilan de la littérature sur la discrimination raciale aux Etats-Unis (tâche considérable). Je ne discuterai donc pas l'influence des théories d'auteurs comme Gary Becker ou Edmund Phelps. Il s'agit d'une limitation assumée de cet article dont les buts sont bien plus modestes.



1. Avantages et limites de la régression linéaire multiple

Le principal avantage de la régression multiple est qu'elle permet d'étudier des corrélations entre des variables d'un modèle en tenant les autres constantes. Ainsi, je peux considérer que le salaire (pour chaque individu i de mon échantillon) est décrit par l'équation suivante : $Y_i = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + u_i$. Les variables (X_1 , X_2 , etc.) sont celles que je pense corrélées au salaire (l'âge, le sexe, le secteur, etc.), u est une variable aléatoire qui résume tous les autres facteurs non connus, les coefficients (b_1 , b_2 , etc.) me disent de combien varie le salaire pour l'augmentation d'une unité de la variable, les autres tenues constantes.

Sans équation de régression, permettant de contrôler les variables, il est très difficile de comprendre ce qui fait varier le salaire : si les ruraux gagnent moins que les urbains c'est sans doute qu'ils sont moins scolarisés. Il faut donc tenir constant le niveau de scolarité pour évaluer l'effet de la ruralité. Mais il faudrait également tenir compte de beaucoup d'autres variables liées au fait d'habiter en ville (par exemple le marché de l'emploi ou la taille de la famille), ce qui pose des problèmes de calcul ou de présentation des résultats. Avant le développement de cette méthode, les sociologues et les économistes ne disposaient guère que des procédures de standardisation. Ainsi Palmer (1947) applique à la discrimination raciale à l'embauche un schéma de distribution conditionnelle. Il construit un tableau, croisant pour les noirs et les blancs, 8 catégories occupationnelles avec les variables maladie, sexe et âge. Il standardise ensuite les données pour l'un des groupes : il s'agit d'appliquer les taux de chômage des noirs à une population qui aurait les mêmes caractéristiques que celles des blancs. Le taux obtenu est de 12 points supérieur à celui réel des blancs : d'après l'auteur, si les noirs avaient les caractéristiques sociales des blancs, leur taux de chômage diminuerait à peine et resterait très supérieur à celui des blancs. Les 12% sont

pour lui une mesure de la discrimination bien que, dit-il, probablement grossie d'autres facteurs inconnus.

Palmer avait signalé un problème important que ne résoudra pas la régression multiple : d'autres facteurs non pris en compte agissent toujours sur la variable d'intérêt et les résultats sont donc toujours contestables dans une certaine mesure. Dans le cas de la régression multiple, les coefficients changent quand on ajoute des variables liées aux autres variables déjà incluses (et dans le cas des modèles logit, même si les variables ajoutées ne sont pas liées aux autres). Ainsi le coefficient lié à la durée de scolarisation est modifié par l'ajout d'une variable mesurant les acquis scolaires. Une conséquence importante est que les coefficients ne peuvent mesurer des effets causaux à moins que l'on apporte la preuve d'avoir affaire à des variables exogènes (c'est-à-dire non liées à des variables non prises en compte). L'application de la régression à l'étude de la société présente donc des différences notables avec son usage dans les sciences naturelles, qui est facilité par la mise en évidence plus fréquente de variables exogènes et donc d'effets causaux. Dans les enquêtes sur la discrimination, cela est bien évidemment impossible car on ignore la liste complète des facteurs à l'œuvre et les coefficients évaluent donc toujours des corrélations, dont on évalue mal la précision. Une deuxième conséquence de l'ignorance des mécanismes de génération des données est que l'on ne sait généralement pas dans quelle mesure le modèle « résumé » correctement la réalité, c'est-à-dire décrit correctement les phénomènes réels à l'œuvre. Ce n'est pas gênant dans le cas d'un modèle prédictif s'il est efficace (par exemple s'il donne sans trop d'erreur l'évolution probable d'un type de cancer chez un groupe de patients). Un modèle « explicatif », en revanche peut difficilement se justifier s'il ne mobilise que des corrélations entre variables choisies parce que des données sont disponibles et que les ordinateurs sont capables de les traiter rapidement. Or, et malgré les plaintes récurrentes des méthodologues et de certains directeurs de revues (Thompson 2002), beaucoup de praticiens se limitent à justifier leur modèle par les habitudes de leurs collègues. La régression permet donc (sous conditions) de voir si certaines variables sont plus corrélées (positivement ou négativement) que d'autres avec une variable d'intérêt, mais n'« explique » rien à proprement parler, en tout cas en l'absence d'un dispositif théorique complet permettant de relier mesures de façon convaincante mesures des corrélations et propositions explicatives. Ces limitations ont dès le départ été minimisées

par les pionniers de la pratique aux Etats-Unis. Introduite dès les années 1930 par William Ogburn à Chicago en sociologie électorale, la régression linéaire se diffuse en sociologie essentiellement dans le sillage des travaux de Herbert Blalock et de Otis Duncan au cours des années 1960. Les auteurs s'en tiennent alors en général au vocabulaire assez flou mais encourageant des manuels : les modèles « expliquent » les phénomènes en montrant l'« effet » des variables choisies sur la variable d'intérêt. Ils ne justifient d'ordinaire pas le choix des variables par d'autres raisons que la disponibilité ou l'évidence logique, ne tiennent pas compte des variables omises, et ne discutent pas les limites des modèles, en particulier dans l'étude des discriminations au travail.

2. L'« explication » de la discrimination raciale par la régression multiple avant 1975

L'étude de la discrimination salariale par régression s'est développée en économie et en sociologie dans les années 1960, d'abord dans le prolongement des travaux de Gary Becker, lui-même influencé par la thèse de Morton Zeman en 1955 (*A Quantitative Analysis of White-Non White Income Differentials in The United States*)¹. Ainsi Gilman en 1965 veut étudier l'effet de la couleur de peau sur les taux de chômage et conclut qu'il faut tenir constants les facteurs qui provoquent le chômage des noirs comme des blancs. Son texte montre la difficulté à sélectionner et mesurer les variables nécessaires. De la même manière, Siegel (1965) ne fournit pas de réflexion sur les variables quand il régresse la différence moyenne des revenus entre noirs et blancs pour plusieurs tranches d'âge en fonction de la catégorie d'emploi et de la région. Si l'usage de la méthode n'est pas encore hégémonique au début de la décennie (en 1962, Blalock étudie le cas du baseball professionnel sans y recourir), l'ouvrage de Blau et Duncan (1967) sur la structure sociale et l'article « Discrimination Against Negroes » (Duncan 1967) vont contribuer à le populariser. Duncan élabore une équation de régression pour étudier la durée de la scolarité à partir de données démographiques dans trois groupes d'âge, chez les blancs et les noirs, en considérant

¹ Les études sans régression deviennent rares en économie dans les années 1960. Celle de Rayack (1961), qui pointe certaines faiblesses des analyses de Becker en utilisant des comparaisons de ratios de revenus médians, et celle de Rasmussen (1971) sont des exceptions.

l'occupation et la scolarité du chef de famille, le nombre d'enfants et la stabilité de la famille (couple séparé ou non). Il est remarquable qu'il ne justifie pas le choix des variables (et donc pas leur exhaustivité pour un modèle « explicatif »), ne discute pas les problèmes de mesure et les conditions statistiques d'application, pas plus que la comparabilité des groupes, ni les avantages ou limites de la présentation des résultats (coefficients bruts ou ajustés). Le modèle est appelé « causal » sans justification théorique, sans doute car il permet de produire des conclusions chiffrées.

Sur cette base épistémologique fragile, les années 1970 verront la multiplication des études socio-économiques par régression sur la discrimination au travail, souvent dans le but de déterminer la liste des variables « expliquant » les écarts de revenus entre blancs et non blancs. Ainsi Weiss (1970) ajoute la condition maritale et le statut de vétéran. Blinder (1973) régresse le logarithme du salaire sur une douzaine de variables. Rexroat (1978) s'intéresse aux revenus des femmes noires. Dans d'autres cas, l'analyse est rendue plus complexe par le souci de vérification d'hypothèses. Ainsi, Shepherd et Levin (1973) examinent les « effets » de variables concernant la position de marchés des grandes firmes industrielles (et constatent des coefficients très faibles). Saunders et Coccarri (1973) étudient l'« effet » de l'étroitesse du marché du travail et le degré de diversification industrielle. Hill (1974) regarde l'« effet » de la syndicalisation (et constate une relation positive sans s'intéresser à la valeur des coefficients). D'autres études considèrent la réticence à l'embauche des hommes noirs : Kain (1968) en fonction du lieu d'habitation, Leigh et Rawlins (1974) en fonction de différentes caractéristiques personnelles plus ou moins protectrices contre le chômage. Il semble que le mouvement de recherche est donc alors porté par la volonté d'examiner un grand nombre de situations diverses sans s'assurer de la solidité ou même de l'utilité des chiffres produits. On constate ainsi en lisant les textes cités qu'il n'y a pas de règles précises gouvernant la quantité de variables retenues ni leur rapport à l'« explication » des différences à analyser. Certaines études ne mobilisent que deux ou trois variables comme le niveau d'éducation, l'expérience et l'ancienneté, d'autres plusieurs dizaines. Ainsi Flanagan (1974) ajoute au schéma de Blinder le statut matrimonial, l'emploi public et le nombre d'employeurs. Oaxaca (1973), apparemment dans un souci d'exhaustivité, propose un tableau de 54 lignes prétendant présenter les principales modalités des facteurs essentiels, selon lui, de variations des salaires, sans qu'il puisse vé-

ritablement au rôle déterminant de certaines. La plupart des auteurs écrivent alors explicitement qu'ils se sont conformés aux choix de leurs prédécesseurs ou qu'ils ont sélectionné les variables parce que les chiffres étaient disponibles (en général dans les *Current Population Surveys*). Il n'y a pas en général de discussion crédible sur l'effet de l'omission de variables importantes. Weiss (1970) fait remarquer qu'omettre tous les contrôles conditionnant l'effet de l'éducation (types d'écoles, diplômes, motivation, savoirs, incitations parentales, etc.) a pour effet de surestimer l'impact d'une année de scolarisation supplémentaire, mais il n'en tire pas conséquence pour sa propre étude.

On voit donc que l'absence d'un débat méthodologique de fond chez les précurseurs a été source d'un certain nombre de négligences méthodologiques. Ainsi, la plupart des auteurs cités évaluent la scolarité en nombre d'années, sans tenir compte des redoublements ni des véritables niveaux atteints. Wright (1978) est l'un des rares à découper la variable en niveaux sociologiquement significatifs (par exemple : « un peu de lycée », « lycée complet », etc.). La qualité des écoles et des diplômes ou leur spécialité n'est prise en compte dans aucune étude. Des problèmes du même type se rencontrent pour les catégories occupationnelles qui sont parfois réduites à 6, avec les risques d'hétérogénéité intra catégorielle qu'une telle réduction impose. L'expérience, de même, est évaluée en soustrayant à l'âge le nombre d'années d'école, sans enregistrer les années d'inactivité. La question de savoir s'il faut utiliser le salaire ou les revenus totaux divise également les auteurs, ou plutôt sépare ceux qui se soucient de cette question et les autres. On peut repérer des flottements comparables pour la catégorisation des secteurs économiques, pour le classement d'une carrière dans un secteur industriel ou pour le découpage en régions (qui se limite parfois à distinguer le Nord du Sud des Etats-Unis). Concernant la mesure des variables, aucun des auteurs cités n'a pris en compte les biais dans les déclarations de revenus. De même, les données concernant les caractéristiques individuelles sont en général tirées de recensements et dépendent de la qualité différentielle de l'enquête suivant les années et les quartiers (plus faible dans les quartiers pauvres et immigrés). Le flou des déclarations sur la profession ou le niveau d'éducation des parents est pourtant connu et ses effets potentiels ont été signalés par Bielby (et al. 1977). Pour chaque variable mobilisée, dans chaque étude, un problème évident de mesure apparaît : la mesure de l'état de santé, de la durée de la syndicalisation, de la mobilité, etc. La source de ces négligences vient ap-

paremment de l'habitude de travailler sur les données les plus communément accessibles. D'où des échantillons parfois incomplets pour les buts visés (ainsi le recensement américain de 1960 est très peu détaillé pour les femmes) et parfois trop réduits pour que la parcellisation puisse s'exercer de façon convaincante (quelques dizaines de personnes forment parfois certains sous échantillons produit par le contrôle des variables, notamment chez les populations noires). Ces inconvénients sont parfois augmentés par l'élimination d'une partie de la population, par exemple la population rurale ou les moins de 25 ans.

L'analyse de cette littérature montre que les auteurs étaient alors plus préoccupés de pouvoir construire des modèles à partir des données existantes que de prouver leur validité ou leur capacité explicative. Un indice en est que la valeur du coefficient de détermination est rarement discutée. Par exemple, Smith (et al. 1978) ne commentent pas le R^2 de 0,25 pour leur équation sur les noirs, Rexroat (1978) se tait sur son R^2 de 0,165 pour les femmes blanches; même attitude chez Hodge and Hodge (1965), Wright, Weiss, Leigh and Rawlins, Oaxaca, Flanagan. Une remarque similaire s'applique aux niveaux de significativité (à 5%) des variables étudiées qui sont souvent très faibles pour les noirs, avec des conséquences non discutées. Une même négligence se retrouve concernant la linéarité des variables. Ainsi la relation entre le nombre d'enfants dans la famille et le nombre d'années d'études complétées par le sujet ne peut être linéaire, mais Duncan (1967) n'en dit rien. Il en va de même pour la relation entre la scolarité et le taux de chômage (Gilman 1965, Leigh et Rawlins 1974), pour celle entre le taux de syndicalisation et revenu familial (Hill 1974), ou celle entre les résultats aux tests scolaires et le revenu annuel. Oaxaca (1973) est l'un des premiers à utiliser le logarithme du revenu comme variable d'intérêt et à inclure dans l'équation la valeur quadratique de l'expérience et de l'éducation pour compenser les effets de la non linéarité évidente. En 1980, Beck (et al.) recourent encore à une variable « expérience » qui est visiblement non linéaire (et qui est associée à un coefficient négatif). Si la linéarité n'est pas complètement examinée, il n'est pas surprenant que la question de la colinéarité forte entre les variables explicatives ne reçoive pas non plus d'attention (ici Weiss fait exception). Peu d'auteurs donnent d'ailleurs le tableau des corrélations entre variables. Enfin, l'imprécision des modèles s'étend au choix des coefficients. Sans discussion des vertus ou des limites des coefficients ajustés (« bêtas ») ou bruts, Duncan choisit d'utiliser les bêtas pour comparer l'intensité de l'« effet » de chacune des va-

riables sur le nombre d'années d'écoles, de même que Weiss qui compare le retour sur éducation des noirs et des blancs (alors même que les *t-ratios* sont insuffisants pour 3 classes d'âges sur 4 chez les noirs¹), et Rexroat qui isole grâce au bêta la variable la plus « déterminante ».

Une autre remarque est digne d'intérêt : en plus de l'utilisation de la régression pour les noirs et les blancs séparément, une pratique courante (par exemple, Smith et al. 1978) consiste à mesurer la discrimination en utilisant l'équation obtenue pour un groupe racial avec les données de l'autre groupe. On suppose qu'en l'absence de discrimination, les moyennes des variables obtenues pour les noirs, si elles sont incorporées à l'équation des blancs devraient donner la même moyenne de revenu. En principe ce n'est pas le cas, et la différence peut être conçue comme une évaluation des effets de la discrimination. Cette méthode repose sur le même principe que celle de Palmer et présente les mêmes défauts. Le groupe des blancs est pris comme un substitut de groupe de contrôle, alors même que les caractéristiques des deux groupes sont différentes en de nombreux aspects et sont par ailleurs liées, d'où le fait qu'il est impossible de dire ce qu'il se passerait en l'absence de discrimination. Par ailleurs, la méthode suppose que les différences entre blancs et noirs se limitent aux écarts des moyennes pour les variables choisies, plus la « race ». Cette idée est aussi peu crédible que le principe qui voudrait que si les noirs et les blancs avaient les mêmes caractéristiques sociales ou situationnelles moins la « race », le reste du fonctionnement de la société demeurerait inchangé (Brochier 2020).

3. Le « tournant descriptif » microsociologique des années 1975-80

L'année 1975 constitue un tournant avec la prise en compte croissante par les tribunaux américains des possibilités des analyses statistiques dans les cas d'accusation de discrimination au sein d'entreprises ou d'administrations². La

¹ Les *t-ratios* sont produits par les tests de significativité (pour une significativité à 5%, ils doivent être proches de 2).

² La loi sur les droits civique de 1964 impose la fin de la discrimination pour l'embauche, la rémunération et la promotion. Les procès ne deviennent nombreux qu'à partir des années 1970. L'usage des statistiques est visible dès 1971 (Meier 1986). Les avocats n'ont pas à démontrer la volonté de discriminer, le constat d'un traitement disproportionné suffit.

Harvard Law Review publie en effet une note technique prescrivant l'usage de la régression multiple qui, selon l'auteur, « détermine si un facteur particulier influence le caractère observé (...) et assigne un nombre qui dit approximativement de combien la valeur d'un facteur donné est pondérée » (Beyond the Prima Facie 1975, 397-98). En 1977, la Cour suprême officialise l'usage de techniques statistiques, ce qui ouvre la voie à une mathématisation croissante des plaidoiries, et notamment à la validation de l'emploi de la régression multiple pour les litiges du travail. Dans les années 1980, dans et hors les tribunaux, la régression est largement utilisée et ne peut plus être alors regardée comme une méthode parmi d'autres (Connolly 1991). Il n'en a pas pour autant découlé une amélioration de la rigueur d'application dans la mesure où les juges ont eu à arbitrer entre des arguments avancés par les statisticiens de parties opposées pliant les méthodes au gré de leurs intérêts¹. Ainsi les plaignants ont généralement tendance à mélanger tous les emplois au sein d'une entreprise, et la défense à les séparer (Barrett et Sansonetti 1988). Les défenseurs cherchent généralement à invalider le raisonnement du plaignant en arguant du poids des variables omises et en mettant en avant l'effet des catégories hiérarchiques. Les juges ont eu à prendre parti sur des problèmes aussi délicats que la validité des modèles et l'effet de démonstration², parfois contre l'avis d'une partie des statisticiens de métier (Meier et al. 1984). Dès 1977, le rôle des tests statistiques (et le niveau de leur significativité statistique) est affirmé par une décision de la cour suprême (Crews 2019). L'un des problèmes évident de l'usage judiciaire de la régression a été d'obscurcir le débat concernant la capacité explicative des modèles. Bien que la note de 1975 indiquait que l'analyse ne portait que sur des corrélations (et pas sur des effets causaux), les démonstrations d'avocats sont en général bien libellées en termes d'« effets » des caractéristiques des plaignants.

Le deuxième évènement notable de la période est l'invention, séparément par A. Birnbaum et H. Roberts en 1979 de la « régression inversée ». Roberts qui en fut le principal diffuseur, voulait résoudre un problème de « shunting »

¹ Aux Etats-Unis l'expert n'est pas un évaluateur neutre ; le plaignant et le défendant mobilisent chacun une expertise différente.

² Typiquement, le plaignant et le défendant recourent chacun à une équation différente sur des cohortes différentes. Les équations incorporent en général la variable race. Si le coefficient est non nul et statistiquement significatif au seuil de 5%, les juges concluent souvent à l'existence d'une discrimination (Finkelstein 1980).

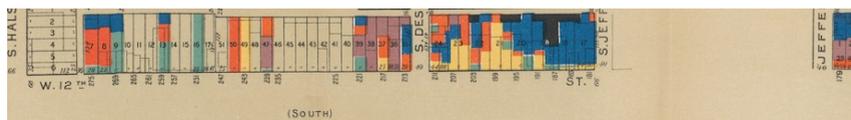
(l'embauche de minorités ou de femmes à des postes où ils/elles seraient surqualifiés) : alors que la discrimination envers les femmes semblait évidente, les hommes et les femmes de ses échantillons présentaient les mêmes moyennes pour les variables mesurant l'éducation (Roberts 1988). La régression inversée consiste donc non plus à examiner les différences de revenus à niveaux d'éducation comparables, mais l'inverse : les caractéristiques personnelles à salaires égaux. Le problème, mais aussi l'intérêt de la perspective, est que les conclusions des deux régressions divergent souvent. Ainsi Roberts a montré que si l'on mélange tous les salariés d'une entreprise et que (c'est souvent le cas) la moyenne des qualifications des femmes est inférieure à celle des hommes, une régression classique montrera un phénomène de discrimination. Ce constat disparaît lorsque l'on raisonne sur des postes de même niveau impliquant les mêmes tâches. L'avantage de la régression inversée est que le salaire est souvent mesuré avec plus de précision que les autres caractéristiques (ce qui évite l'exagération artificielle de la discrimination mesurée). Mais la discrimination agissant sur l'accès aux différents niveaux de salaire devient alors difficilement discernable. Les idées de Roberts se sont donc heurtées notamment à l'opposition des chercheurs s'inquiétant de voir affaiblis les soupçons de discrimination (par exemple : Ash 1986), ce qui nous permet de mieux discerner l'opposition entre une logique probabiliste (on ne peut pas modéliser de façon réaliste les salaires, on peut juste constater qu'ils lèsent les femmes ou les minorités) et une orientation « déterministe » (plus en on sait sur les mécanismes réels de formation des salaires, plus on peut étudier le rôle réel de la discrimination).

Les deux innovations des années 1975/1980 présentent le point commun de chercher à mieux décrire les facteurs réels qui déterminent les salaires et notamment la « productivité » dans les organisations. Cette orientation est également visible dans les articles publiés par les revues de gestion dans les années 1980, marquant une sorte de « tournant descriptif ». A la pure logique corrélationnelle/démographique dominante dans les années 1970, on voit s'ajouter, une recherche des « mécanismes réels » (voir par exemple Corcoran et Duncan 1979) et une critique des approches structurales (Bielby and Baron, 1980). L'une des raisons est sans doute le faible pouvoir démonstratif des études économiques et sociologiques précédentes. Non seulement la démarche consistant à faire la liste des variables disponibles potentiellement corrélées à la détermination des salaires est mise en question (en particulier si elle est associée à un

dédain pour l'évaluation du pouvoir descriptif du modèle), mais encore la capacité des recherches à imposer des « faits » (par exemple une mesure) sur le sujet s'est montrée très limitée. Un examen même rapide des bibliographies dans les articles des années 1980 montre que les études datées de plus de dix ans sont oubliées et que les auteurs cherchent en général à tester des hypothèses nouvelles. L'observateur est également étonné de retrouver année après année souvent les mêmes questions posées par des auteurs différents avec des échantillons et des modèles changeants, comme si la cumulativité des acquis restait limitée. Ainsi par exemple, Kaufman en 1983 s'interroge encore dans l'*American Journal of Sociology*, sur la nécessité de prendre en compte la position du salarié dans un secteur économique et dans un type de poste.

Ce que j'ai appelé le « tournant descriptif » trouve une matérialisation dans l'ouvrage de référence écrit pour les juristes en 1986 par Kay et Aickin. Le chapitre que McCabe (1986) y consacre à la régression insiste sur la nécessité de saisir l'« essence des procédures de détermination des salaires ». Pour cela, les variables mobilisées doivent, selon lui, permettre de tenir compte des caractéristiques du salarié, de celles du poste et de l'intersection entre les deux (durée dans l'entreprise et le poste, absences, performances, etc.). Il souligne donc la difficulté à décrire les mécanismes réels en absence de données concernant la façon dont les salariés se comportent au travail. Cette perspective tranche avec la position précédente (comme le montre la revue de Cain en 1976) selon laquelle les variables démographiques habituelles suffisent à décrire l'essentiel de la discrimination. L'article technique de Dempster en 1988 donne, de même, une traduction en termes statistiques de cet intérêt pour les mécanismes internes aux organisations (l'analyse statistique porte selon lui sur la décision par l'employeur de fournir une récompense à un salarié). On trouve d'ailleurs un écho de ces exigences y compris dans les textes qui mobilisent les variables démographiques habituelles : ainsi Ruhm (1989) remarque justement que la comparaison des revenus de salariés en poste ignore l'effet des comportements dans le temps et qu'une mesure plus réaliste de la discrimination est produite par la comparaison des conditions de la reprise d'emploi après un licenciement. Il ne faudrait pas cependant conclure que les années 1980 consacrent une rupture avec les flottements techniques des années 1970. De nouvelles questions sont posées (par exemple Mare and Winship [1984] regardent la probabilité d'embauche en fonction du service militaire), le niveau de précision augmente (par

exemple, Kaufman en 1983 croise 12 secteurs d'entreprise et 8 catégories occupationnelles), mais la réflexion sur les limites de la régression multiple reste rare (voir par exemple : Cotton 1988). Ainsi, la prise en compte des effets de l'éducation reste peu convaincante, les échantillons sont peu représentatifs ou trop réduits par la parcellisation, la spécification des modèles n'est pas ou peu justifiée (Heckman 1998), la significativité des variables n'est pas discutée, et le respect des règles statistiques reste souvent, en général, une question secondaire.



4. Des raffinements méthodologiques limités dans les années 2000

Les améliorations amorcées dans les années 1980 se sont prolongées dans les années 1990-2000. Il s'est d'abord agi d'intégrer plus régulièrement des variables (« proxys ») cherchant à décrire l'action des entreprises en matière de sélection des candidats ou de valorisation du travail fourni (Pager et Shepherd 2008). Sont concernés des facteurs concrets comme le paiement à la pièce (Heywood and O'Halloran 2005) ou plus vagues comme le « degré de contact avec les clients » (Goldsmith et al., 2006). Du côté des salariés, on a mesuré d'autres types de facteurs « pré marché », comme le ton de couleur de peau, l'« estime de soi » (Goldsmith et al., 2006), ou le lieu exact d'habitation (Black et al. 2013). Des facteurs plus diffus comme la « motivation » ou l'« intégration sociale » ont également été pris en compte par des variables comme le nombre de jours d'exclusion à l'école, le « sentiment de contrôle sur sa propre vie », les problèmes avec la justice (Ritter et Taylor 2011). Un effort a également été fait pour évaluer de façon plus réaliste l'effet probable de la scolarisation. Neal et Johnson (1996) dans un article célèbre ont proposé de remplacer les années de scolarisation (qui surestiment les capacités réelles des non blancs) par les résultats d'un test passé par

les hommes entre 17 et 19 ans pour l'armée (le *Armed Force Qualification Test*). La scolarisation correspond en effet à un *input* et pas à un *output* (on ignore les capacités réelles acquises) et son déroulement est lui-même affecté potentiellement par la discrimination. L'analyse montre alors qu'à niveaux de capacités égaux, les écarts de salaires postérieurs sont très réduits. L'utilisation de l'AFQT s'est diffusée dans les années suivantes, malgré de nombreuses critiques et des ajouts (Lang et Lehman 2012). Une alternative est l'addition d'une variable du type « perception de la qualité de l'école » par les sujets eux-mêmes (Goldsmith et al., 2006) malgré son évidente subjectivité et les problèmes de mesure. Au sujet des études universitaires, Black (et al. 2006) ont également tenté de corriger l'imprécision des niveaux de diplômes en regardant les choix d'options des étudiants en premier cycle universitaire (et ont montré que les noirs choisissaient plus souvent les filières les moins valorisées par le marché du travail). À l'intersection entre les caractéristiques des salariés et les critères des employeurs, certains auteurs ont regardé les préférences des chercheurs d'emploi. Maume jr (2004) a par exemple ajouté dans son modèle le statut social des emplois mesuré par la proportion d'homme blancs dans le secteur, les exigences cognitives et la dangerosité des tâches. Pager et Pedulla (2015) ont examiné l'éventail des emplois auxquels postulaient les blancs et les noirs.

En matière de données, la question de la valeur de l'échantillon a également reçu plus d'attention. Il est en effet évident que les échantillons sont souvent produits à partir de mécanismes de sélection non aléatoires qui nuisent à la qualité des résultats. Ainsi, les chômeurs, les auto entrepreneurs, les individus hors emploi, les plus de 50 ans sont souvent écartés. Pour la comparaison entre groupes « raciaux », les phénomènes affectant les mécanismes de sélection des individus étudiés sont souvent différents d'un groupe à l'autre. Par exemple les hommes blancs arrivent plus facilement que les noirs à accumuler un capital leur permettant de partir à la retraite précocement ou de se mettre à leur compte (Maume jr 2004). On a donc assisté à une diffusion de techniques de correction des biais d'échantillonnage comme la procédure à deux étapes de Heckman (1979). Les différences de dispersion des revenus au sein de l'échantillon ont également été mieux prises en compte, notamment par un usage plus fréquent de la régression par quantiles ou par des régressions répétées sur des fractions différentes de l'échantillon (Naito et Takagi 2017). De même, la comparaison d'un groupe blanc et d'un groupe non blanc, qui pose toujours des problèmes

en raison des différences structurelles de distributions des caractéristiques a pu être facilité par des procédures de « matching », en dépit de la réduction de la taille des échantillons (voir : Pager et Pedulla 2015, pour le chômage).

Si l'on compare les publications des années 2000-2010 avec celles des années 1990 (voir par exemple : Stratton 1993), la spécification des modèles se fait aujourd'hui selon des critères moins ritualistes : l'utilisation sans véritable questionnement de variables démographiques est de moins en moins acceptée. Les auteurs ont à la fois multiplié les variables et ont cherché à les rapprocher de la réalité. De même, ils se sont plus sérieusement souciés de la qualité des données. Mais toutes les précautions sont rarement prises en même temps. La relégation au second plan des questions techniques s'appuie généralement sur l'idée implicite que l'« important est ailleurs » dans la mesure où les auteurs essaient généralement de décrire des aspects peu étudiés comme la comparaison dans le temps (Mandel and Semyonov, 2016) ou l'effet d'un changement de politique (Wilson et al. 2015). En dépit des critiques de nombreux spécialistes, au premier rang desquels David Freedman (1995, 1997, 2006) des problèmes de fond se maintiennent.



5. Le maintien de problèmes de fond

1 Le plus essentiel de ces problèmes est celui du statut de l'« explication » que propose l'application de la régression à notre sujet. Dans la plupart des recherches cités dans ce texte, le lecteur ignore généralement ce que l'auteur entend par explication. Il est en fait sous-entendu que des propositions théoriques seront faites à partir de certaines corrélations entre variables et que l'explication consiste en cela (Abbott 1998). Le fondement épistémologique de la démarche n'est en général pas abordé, car les auteurs se réfèrent aux « usages » en vigueur. Ainsi la différence entre modèles prédictifs et modèles « explicatifs » n'est pas

prise en compte. Les auteurs se livrent donc à une sorte de jeu rhétorique avec la notion de cause. Parfois ils parlent de « conséquence », parfois « d'impact », ou d'« effets », c'est-à-dire qu'ils évitent d'utiliser le mot cause. En même temps, toute l'argumentation est structurée par le présupposé que l'« explication » est de nature causale, puisqu'il s'agit en général de dire ce qui provoque des écarts entre groupes. Pour donner un aperçu récent de ces hésitations sémantiques : Anderson et Shapiro (1996), estiment avoir analysé les « déterminants » du salaire des femmes de l'échantillon ; Restifo (et al. 2013, 918) disent que les variables choisies dans leur modèle « *aident à expliquer* les variations sur le marché du travail » ; Mac Call (2001, 535) après avoir parlé d'« association » dans le corps de l'article conclut que l'immigration est « la *source* principale » de différences salariales avec les blancs pour « les asiatiques et les latinos » (d'autres variables étant vues comme « expliquant » ou montrant un « impact ») ; Mouw (2000), après avoir respecté le vocabulaire de la corrélation, parle en conclusion d'« impact » ; Pager et Pedulla (2015, 1033) après avoir rappelé qu'il « est difficile de faire des affirmations appuyées de processus de médiation causale » concluent que leurs résultats sont « hautement suggestifs » dans le cadre d'une théorie des facteurs ; Mandel et Semyonov (2016) alternent les propos prudents (« nos résultats sont en accord avec la théorie selon laquelle... », 1058) et les affirmations risquées (« la figure 1 révèle la nature genrée de l'inégalité raciale », 1048). Comme l'a noté Holland (2003), le terme d'« effet » sert généralement de « pente glissante » vers l'idée de causalité. L'une des conséquences en est que le rapport qu'entretiennent les théories à valider et les modèles statistiques est flou. Ainsi, la mise en évidence du statut causal d'une variable ne pouvant venir du modèle lui-même, elle ne peut venir que des théories, mais on se demande alors comment les résultats pourraient valider la théorie.

2 Le deuxième problème le plus courant concerne l'usage des coefficients. L'étude de la discrimination n'échappe pas à l'habitude courante en sociologie et en économie de n'étudier les coefficients qu'en fonction des intentions théoriques des recherches. Il en résulte en général une appréciation très incomplète des tailles d'effets comme l'ont remarqué de nombreux critiques (Ziliak and McCloskey 2008, Cohen et Cohen 2009, Ellis 2010, Bernardi et al., 2016). Cela signifie en pratique que les coefficients sont étudiés pour leur signe, ou comparés entre eux sans que l'auteur ne donne explicitement la taille de l'effet constaté ni n'exprime de façon intelligible par le public la réalité de cet effet

dans le contexte social étudié. Ce problème est particulièrement saillant dans les études sur la discrimination car elles utilisent généralement des données démographiques et donc des échantillons de très grande taille, qui permettent de repérer des corrélations parfois très ténues entre les variables (Allison 1999a). Les auteurs se reportent généralement sur les tests de significativité pour démontrer « l'importance des liens » et donc la valeur de ce qu'ils ont mis à jour. Cet usage indu des tests a pourtant été signalé tout au long de la période étudiée par des statisticiens professionnels¹. Rappelons que le test de student (généralement utilisé) donne la probabilité d'obtenir les résultats auxquels on aboutit (ici la mesure d'une corrélation) dans l'hypothèse où ce seraient t l'échantillonnage ou des erreurs de mesure qui produiraient l'apparence d'une véritable relation. Les tests ne donnent donc pas le degré de probabilité de l'existence d'un lien et encore moins une mesure de sa force, ou une confirmation de son importance sociologique. La confusion est pourtant courante. Des phrases ambiguës du type « notre modèle est hautement significatif », ou « nous avons montré des relations hautement significatives » ne font que cacher l'absence d'analyse des tailles d'effets et de leur signification réelle. Beaucoup d'auteurs se contentent de ne discuter que les variables qui sont significatives statistiquement, comme si elles étaient sociologiquement décisives et sans réfléchir aux raisons de la non significativité des autres. Dans le domaine de la sociologie des relations raciales, les groupes de « non blancs » présentent ainsi souvent des variables non significatives, ce qui indique que leur hétérogénéité de est sans doute forte et que la connaissance que l'on en a est insuffisante.

3 Un autre problème méthodologique concerne la sélection des variables dans le modèle. Même si l'on laisse de côté les auteurs qui ne retiennent que les variables qui sont significatives au seuil de 0,05, la sélection donne fréquemment au lecteur une sensation d'arbitraire. Duncan, d'ailleurs, à la fin de sa carrière avait moqué l'« illusion qu'en décomposant les covariations d'un ensemble de variables choisies presque au hasard, on peut pratiquement justifier non seulement un modèle 'causal', mais aussi, admirez l'adjectif, un modèle de mesures »

¹ Dès les années 1940 par des mathématiciens et les psychologues, et pour la sociologie dès 1957 par Hanan Selvin. L'article de Selvin a été suivi par une série de débats tout au long des années 1960 qui sont résumés par Henkel et Morrison (1970). Cette discussion a été continuée notamment par David Freedman dans les années 1980/90 et par Ziliak et Mc Closkey dans les années 1990/2000.

(Duncan 1984 : 226). Il est en effet évident que les modèles d'analyse de la discrimination sont encore largement « tirés par les données » : les variables sont en général utilisées car elles sont disponibles ou usuelles, sans véritable connaissance des mécanismes causaux (Abbott 1998, Bernardi et al. 2017). La synthèse du *National Council* (2004) fournit un bon exemple de cette situation. L'auteur y affirme que les modèles causaux sont indispensables mais qu'ils sont difficiles à construire. Il reconnaît pourtant que les variables n'étant pas manipulables (comme en laboratoire), on ne peut parler que de corrélation et pas de cause. Il montre une double affiliation contradictoire à l'activisme social (les plaignants et les militants doivent pouvoir continuer à se servir de modèles pour dénoncer la discrimination) et à la rigueur statisticienne. Il en résulte un moyen terme qui correspond à l'usage habituel depuis les années 1960 : les modèles doivent selon lui faire intervenir les « variables essentielles », ce qui montre l'adhésion au paradigme descriptif. Pourtant il parle tantôt de « toutes les variables qui agissent sur la variable étudiée et sont liées à la race », tantôt d'éviter d'« omettre des contrôles importants », acceptant ainsi les limites évidentes de la démarche exhaustive. Finalement (p.137) il faut que « la structure du modèle régressif (les variables et leur forme fonctionnelle) saisisse les mécanismes causaux avec une précision suffisante ». Cela revient à dire finalement que pour expliquer la discrimination il faut déjà posséder l'explication (les « variables essentielles »). Les autres sont souvent saisies par des « *proxys* » plus ou moins convaincants censés mieux décrire les situations réelles (« degré de motivation », « estime de soi », etc.), voire des indices composites (par exemple Pager et Grodsky 2001).

4 La spécification des modèles se heurte aussi au problème de travailler avec une ou plusieurs équations. Si l'on utilise une équation pour chaque « groupe racial », il faut pouvoir les comparer (donc la composition des échantillons doit être similaire, et la confiance accordée aux coefficients et aux variables doit être de même niveau). Or, on l'a vu, les groupes blancs et non blancs n'ont pas les mêmes caractéristiques sociales. Certaines distributions des variables ont des effets différents (quand la relation avec Y n'est pas véritablement linéaire). Si par exemple, les noirs sont plus jeunes, l'effet de l'expérience sera plus grand (Gray 1993), et il en va de même pour l'effet de la scolarisation pour un groupe où les diplômés sont plus nombreux. La solution limitée du *matching* réduit la taille des échantillons. Si l'on inclut, à l'inverse, dans l'équation la variable race (par exemple pour « mesurer » la discrimination), il faut aussi en bonne

logique inclure toutes les variables qui peuvent agir sur la variable dépendante (par exemple le salaire) à travers la « race » (Lieberson 1985). Cette question renvoie à celle des interactions entre variables puisqu'il est évident que, pour des raisons de lisibilité, la plupart des modèles limitent leur prise en compte (Kaufman 1983). Ainsi le sexe, l'éducation ou la « race » par exemple peuvent avoir des effets différents selon les niveaux des autres variables. Ragin (1984) a ainsi montré que la description des situations d'embauche était très appauvrie sans la prise en compte des interactions, mais que le calcul de toutes les interactions était pratiquement impossible même avec un échantillon considérable. Dans les modèles logit et probit qui regardent l'effet de variables sur la situation d'une personne type, le problème est accentué par le fait que les coefficients sont affectés par les variables omises même quand elles ne sont pas corrélées aux autres variables du modèle (Mood 2010). Les comparaisons entre groupes ou entre modèles pour les ne peuvent se faire que si l'on suppose que l'hétérogénéité non observée reste constante (Allison 1999b, Karlson et al. 2012), ce qui représente une hypothèse très forte en général, et en sociologie des relations raciales, très peu probable.

Les déficiences de la spécification posent donc la question de la capacité des modèles à décrire correctement les variations du phénomène étudié puisque dans la plupart des études il n'y a pas d'analyse convaincante (voire pas d'analyse du tout) du R^2 , ni des erreurs-types et le caractère structurel de l'équation n'est jamais prouvé. Rappelons que pour cela, il faut faire la preuve d'une faible variation des coefficients d'un échantillon à l'autre sur une certaine période de temps, tout en montrant que les coefficients restent stables après intervention. Ces conditions imposeraient un dispositif proche de la situation expérimentale. Freedman (2006 : 698) a ainsi résumé la situation : « Les modèles supposent, sans garantie, que les effets sont linéaires et additifs sur l'échelle choisie avec des paramètres qui sont constants d'un sujet à l'autre et invariants à l'intervention. De façon à justifier les formules usuelles sur les erreurs types et les niveaux de significativité – et l'allégation habituelle que les estimations de la régression sont non biaisées – des hypothèses supplémentaires sont indispensables, par exemple, des termes d'erreurs indépendants d'un sujet à l'autre et indépendants des variables explicatives dans l'équation ».

5 Un autre problème important et généralement occulté est celui de la mesure. Ainsi, la mesure des revenus par déclaration dans des « surveys » n'a aucune

chance d'être également précise et juste d'une catégorie à l'autre, la mesure de l'éducation est faussée par les affirmations des répondants voulant cacher leur ignorance, il en va de même pour l'expérience et les antécédents criminels. La profession des parents est toujours très mal mesurée par les enquêtes démographiques (Peneff 1984) et l'on n'a aucune garantie sur la façon dont sont remplis les formulaires d'enquête sur le terrain (Roth 1966, Becker 2017). Les variables plus complexes mesurant des données qualitatives n'ont quant à elles que la précision que l'on veut bien leur accorder (« tons de peau », « degré d'exigence cognitive de l'emploi », etc.) et de nombreuses variables clefs des promotions et des déroulements de carrière sont à peine mesurables (Altonji et Pierret 2001). Le codage de la couleur ou de l'ethnicité des répondants est quant à lui une telle source de complications potentielles que le problème n'est pratiquement jamais évoqué (Kofi et Guryan 2011). Or, en l'absence d'une mesure exacte des variables, les évaluations de la discrimination sont difficiles, voire biaisées. Toute variable indépendante dont la mesure est exagérée par rapport à la réalité grossit l'importance de la discrimination supposée (par exemple les années d'éducation pour les noirs, ou l'expérience pour les femmes mesurée à partir de leur âge).

6 L'assimilation de la discrimination à un résidu « non expliqué », problème ancien, toujours évoqué mais rarement affronté, se maintient. Tant que l'on ne connaît pas le mécanisme de génération des données, ni les situations réelles de formation des revenus, d'accès à l'emploi ou de comportement au travail, la définition de la discrimination comme « ce qui n'est pas expliqué » n'est pas satisfaisante. Elle correspond clairement à une dépendance de l'analyse statistiques vis-à-vis de positions politiques : prouver la discrimination raciale devient équivalent produire un modèle qui laisse une grande part de variance non contrôlée. Le recours à des méthodes de décomposition renforce ensuite artificiellement la conviction que la discrimination peut être précisément mesurée malgré les problèmes de spécification et de logique de décomposition. Comme celle consistant à calculer le salaire que les noirs auraient s'ils avaient les caractéristiques des blancs (Palmer, 1947, Anderson and Shapiro, 1996) la décomposition de Blinder (1973) ou celle d'Oaxaca (1973) reposent sur des situations contrefactuelles irréalistes et incalculables (Brochier 2020).

Les problèmes que j'ai indiqués ne se retrouvent bien évidemment pas de manière similaire dans toutes les recherches, mais on constate facilement leur

maintien dans le temps, comme l'illustrera un exemple récent. Dans un article à paraître, Budig et al. (2021) cherchent à étudier l'effet des diplômes sur la réduction des disparités salariales selon le genre et la « race ». Le sujet est classique, mais son intérêt est rehaussé par l'étude des choix de spécialités à l'université. Pour produire les données, les auteurs utilisent le NLSY79, un échantillon probabiliste qui suit jusqu'à aujourd'hui une cohorte de 12686 collégiens de 1979. L'échantillon s'est considérablement réduit depuis l'origine et les auteurs le réduisent encore en retirant les hispaniques, les non employés, les non déclarants de salaires et ceux qui n'ont pas achevé le lycée (au total $N=2828$). Une partie des régressions sont faites sur l'année 2012, d'autres sur toutes les années. Les auteurs ne discutent pas des effets possibles de cet échantillonnage sur les résultats et se fondent apparemment sur l'indication des niveaux de significativité des variables, qui ne sont d'ailleurs pas commentés. Le logarithme du salaire horaire est pris comme variable d'intérêt. Les principales variables indépendantes étudiées concernent la scolarité. Le niveau d'études est donné par la « plus haute classe achevée », et pas le « plus haut diplôme atteint » car cette dernière donnée ne distingue pas les années qui n'ont pas abouti à un diplôme, des autres. Le problème est cependant que les auteurs n'ont pas les diplômes obtenus. Les spécialités sont codées très sommairement de la façon suivante : sciences et ingénierie, sciences sociales, management, médecine et droit, humanités. Les auteurs ne commentent pas le manque de signification réelle de ces rapprochements (droit associé à médecine, architecture associée à sciences sociales, études militaires associées à humanités, etc.), ni le fait que ces codages ne correspondent pas à des diplômes identifiables dans des établissements hiérarchisés. Les variables de contrôle sont 5 variables de capital humain, 5 variables des caractéristiques de l'emploi, 11 secteurs économiques, et 7 variables démographiques (dont le test AFQT). L'intérêt de ces variables (qui sont habituelles dans le domaine) est seulement justifié par la littérature. Leur capacité à contrôler correctement les effets de la scolarité n'est pas discutée. Par exemple, la scolarité des parents est donnée par le « plus haut diplôme atteint » (contrairement à celles des enfants) et se trouve mesurée par une moyenne du niveau de prestige éducationnel et du diplôme des deux parents. La question de la mesure correcte des variables n'est pas discutée non plus, pas plus que le R^2 , et la valeur du modèle semble se fonder sur son caractère « habituel » au regard de la littérature sur le sujet. Une analyse descriptive des variables est ensuite proposée mais la signification

sociologique des écarts importants entre groupes n'est pas discutée. Les conséquences à attendre de l'application du même modèle à des groupes différents ne sont pas discutées non plus. Un tableau donnant les résultats de la régression pour les variables d'éducation (mais pas les variables de contrôle) est ensuite donné. Dans la discussion, les coefficients sont traduits en % d'effets (en passant par l'exponentielle) ce qui facilite l'interprétation des résultats en termes de taille d'effets. L'analyse des résultats consiste cependant essentiellement à comparer pour deux modèles (avec plus ou moins de contrôles) les coefficients des quatre groupes de genre/race. Les coefficients traduits en termes d'effets sur le salaire ne sont pas comparés à ceux d'autres études (les études citées de toute manière ne sont pas examinées en fonction de la qualité de leur modèle). Les résultats que les auteurs retiennent sont ceux qui correspondent à leurs attentes : les femmes ont un plus faible retour sur étude que les hommes (en tout cas avant les études doctorales), les hommes noirs ont à peu près les mêmes retours que les hommes blancs, les femmes noires sont nettement désavantagées. Les cas qui ne correspondent pas aux attentes ne sont pas vraiment discutés, et le problème de la non significativité de certains coefficients n'est pas regardé non plus. L'apport des spécialités est peu exploité, sans doute en raison de l'abondance des chiffres. Pourtant il apparaît que certaines spécialités dont les sciences sociales et le droit/santé favorisent les minorités. Ces résultats ne sont pas analysés en fonction des phénomènes qu'ils pourraient impliquer dans la réalité vécue par les acteurs. Les auteurs parviennent toutefois à la conclusion qu'ils avaient prévue : « la réussite éducative n'aide pas à annuler l'écart de salaire initial entre les genres et les races » (p.14). Ils n'indiquent à pas de quelle manière leurs calculs sur leur échantillon peut les conduire à de telles conclusions sur la population totale, comme si la construction d'un modèle selon les normes habituelles suffisait. A partir de là, les auteurs signalent les théories qui sont en accord avec leur conclusion et qui reposent toutes sur l'idée d'une discrimination volontaire par les employeurs. Ils ne discutent d'aucune théorie rivale qui serait en accord avec les données, bien qu'ils reconnaissent être dépourvus de toute information sur le fonctionnement concret des carrières dans les lieux de travail. L'article se conclut par des propositions de politiques de discrimination positive.



6. Synthèse et débats

L'examen de la littérature sur 60 ans nous montre comment a évolué l'usage de la régression multiple pour l'étude de la discrimination « raciale ». Son développement initial s'est fait dans un contexte intellectuel incitant à mettre en évidence les différents facteurs expliquant la position des Afro-américains dans le monde du travail, notamment dans des buts militants et d'intervention. Mais il n'y a pas eu de véritable discussion méthodologique à l'origine. Le mouvement de recherche initial s'est traduit par l'effort pour identifier le rôle d'un nombre croissant de variables, à mesure que des données plus complètes devenaient disponibles. En ce sens, l'usage n'a cessé d'être en partie au moins « guidé par les données » : ce sont les enquêtes démographiques disponibles qui ont fourni les variables et leur mesure. La plupart des chercheurs posaient des questions limitées, comme la comparaison entre elles de deux ou trois variables clefs pour deux groupes. A partir du milieu des années 1970, l'extension de la méthode aux litiges juridiques a contribué à renforcer le statut implicitement causal des coefficients ainsi que certains usages ritualisés comme la valorisation du R^2 ou des tests de significativité. Les plaignants se servant des modèles pour prouver que leur « race » ou leur sexe sont la cause d'une forme de discrimination, les débats ont porté sur le choix des variables ou des échantillons plus que sur les réelles possibilités de la méthode. L'obligation faite employeurs de se justifier a ensuite été concomitante d'une tendance à mieux prendre en compte dans les années 1980 les pratiques réelles des entreprises. Dans les décennies qui ont suivi, l'utilisation de la régression est devenue lentement plus rigoureuses à mesure que certains auteurs ont pris l'initiative de chercher à combler certaines lacunes (la nécessité du logarithme du salaire, puis le problème de l'omission de variables clefs, puis la prise en compte de mécanismes réels, ou au contraire l'inadéquation de certaines variables comme les années d'école), sans pour autant résoudre les problèmes de fond dont j'ai donné une liste non exhaustive.

Surtout, le flou entourant les limites de la méthode a permis de l'appliquer à des questions de plus en plus variées, en particulier grâce à la disponibilité de programmes informatiques qui guident souvent plus qu'ils ne servent les opérations de recherche. Mais les problèmes de méthode et d'échantillon laissant toujours planer un doute sur la validité, l'interprétation et la portée des résultats, la cumulativité des acquis est restée faible. Les chiffres les plus souvent cités sont ceux qui donnent l'évolution des écarts globaux entre groupes et qui ne proviennent donc pas de modèles par régression. Les autres sont rarement considérés comme assez probants pour être cités d'un article à l'autre. Il n'y a ainsi pas de raison d'adopter une perspective finaliste postulant que désormais « les pratiques sont au point » car moins improvisées que dans les années 1970. Les cas récents que nous avons examinés montrent que les conclusions sociologiques proposées entretiennent des rapports discutables avec les dispositifs méthodologiques, dans un climat intellectuel fortement marqué par ce que l'on a pu appeler le « statactivisme » (Bruno et Didier 2021). Beaucoup de recherches prétendent valider des options théoriques pour lesquelles ils ont des préférences politiques sans que la capacité des modèles à décrire correctement une réalité complexe et mal connue ne soit véritablement établie.

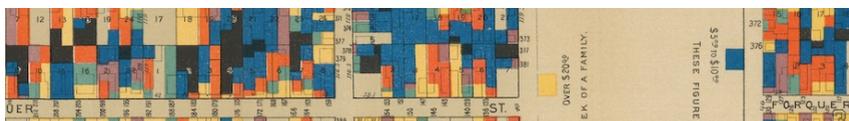
A partir de ces constats, ce texte peut mener à trois débats que je me contenterai d'esquisser ici. Le premier, qui intéressera les historiens des méthodes, porte sur les raisons expliquant le maintien de problèmes de fond et l'adoption de stratégies d'évitement, notamment par des procédures rhétoriques. Cette question rejoint celle posée par Howard Becker (2017) au sujet du maintien de procédures que l'on sait insatisfaisantes en sciences sociales. L'une des réponses les plus évidentes peut se formuler en termes d'inertie. Les chercheurs ont intérêt à adhérer à l'idée d'une « validité apparente » des méthodes en vigueur, même s'ils connaissent les problèmes. Un autre aspect tient aux arguments d'autorité et aux statuts. Si des sociologues et des économistes de renom fondent leurs conclusions sur la discrimination raciale sur des coefficients de régression depuis les années 1960, sur la base de quelle autorité peut-on contester la pratique ? Les critiques sont d'autant plus inaudibles que la régression multiple s'est étendue à pratiquement tous les thèmes de recherche en sciences sociales. Enfin, les limitations de la méthode constituent l'un des moteurs de la démultiplication de son usage : un auteur peut toujours prétendre apporter une correction partielle ou une amélioration aux travaux de ses prédécesseurs.

Un deuxième débat possible concerne les spécialistes des méthodes quantitatives. Si l'application de la régression multiple est délicate et que ses capacités d'explication sont plus incertaines que ses possibilités de prévision en sciences sociales (Lieberson 1985), quelle place peut-on lui donner dans l'étude de la discrimination ? Beaucoup de praticiens agissent comme si leurs modèles pouvaient leur donner les mêmes certitudes que ceux utilisés dans les sciences de la nature, alors que ce n'est évidemment pas le cas¹. Une suggestion serait de continuer à utiliser les modèles par régression multiple pour étudier les données et non pour expliquer les phénomènes sous-jacents (pour un exemple : Mayer et al. 2020). Il s'agirait alors seulement de générer des questions ou de faire émerger des aspects peu ou mal connus. Ainsi Lang et Lehman (2012) ont montré que les écarts de salaires entre blancs et noirs diminuent en haut de la pyramide des diplômés et de « l'intelligence » mais restent forts chez les travailleurs manuels. Cette question ouvre des pistes de recherche intéressantes à condition que l'on ne fonde pas de trop grands espoirs sur l'addition de contrôles plus ou moins bien définis à une équation mobilisant les variables habituelles de capital humain. Les modèles sont également particulièrement utiles quand ils permettent de comparer les résultats obtenus pour des échantillons différents pour lesquels on possède assez de renseignements sur les modes de génération des données. Un autre usage pourrait être celui d'indicateur. Il s'agirait alors de fournir un repère comparable dans le temps de la corrélation entre plusieurs variables démographiques importantes ; par exemple les résultats scolaires, l'âge, la profession des parents et les revenus ou le chômage pour les mêmes localités tous les 5 ou 10 ans, à partir d'un échantillon standardisé. Les résultats n'auraient pas prétention à expliquer mais à repérer des évolutions dans le temps, rendues impossibles par les changements permanents d'échantillons et de variables d'une enquête à l'autre.

Le troisième débat possible concerne plus généralement les spécialistes de l'étude de la discrimination. Si les modèles multivariés fournissent à chaque fois des chiffres différents, est-il utile d'en produire toujours davantage dans le but de « comprendre » les phénomènes ? Une remise en cause de la démarche ne serait-elle pas une occasion de déterminer ce que le futur de la recherche en

¹ Cette situation dépasse d'ailleurs le sujet qui nous intéresse et concerne par exemple aussi la méthode expérimentale popularisée par Esther Duflo (Jatteau & Labrousse, 2020).

ce domaine doit chercher à obtenir ? On peut ainsi se demander, à partir des idées de Blalock, s'il est important de produire des chiffres pour expliquer la discrimination. Dans les sociétés où le racisme est une réalité tangible et permanente, les évaluations chiffrées de l'« effet » de la « race » sur les salaires, la promotion ou le chômage sont-elles utiles si elles sont méthodologiquement contestables ? De même, le calcul d'effets de discrimination en tant que résidus inexpliqués permet-il de mieux comprendre la réalité ? L'étude que j'ai proposée montre certains angles morts de la recherche sur la discrimination. Ainsi, les coefficients calculés sur des grands échantillons obscurcissent la façon dont différentes combinaisons de situations, de caractéristiques et d'intentions produisent des résultats au niveau vécu des individus. De là, on peut regretter que la recherche quantitative se suffise parfois à elle-même en fixant des objectifs fondés sur la disponibilité de techniques économétriques et de données démographiques. L'étude des discriminations gagnerait très certainement à être guidée par des analyses qualitatives de cas qui montreraient dans quelles situations, des gens précis, dotés de caractéristiques saisies selon les logiques des acteurs, agissent en fonction de considérations raciales. Ainsi, alors que Restifo (et al.) utilisent la recherche qualitative comme un complément de la recherche quantitative, on peut adopter la démarche inverse. Les études de cas dégageraient des mécanismes qui seraient ensuite intégrés à des modèles permettant de les questionner davantage. Cela signifie par exemple qu'au lieu de décider grâce à un coefficient dans quelle mesure le revenu des « non-blancs » est corrélé à leur niveau d'éducation mesuré en années de scolarité, on peut partir de situations d'éducation vécues, examiner des groupes réels pour constater les parcours, se servir d'entretiens ou d'observations en situation pour dégager les logiques à l'œuvre (celles des acteurs ou les effets de contexte), construire une théorie de portée limitée et tester par des moyens statistiques son champ d'application. L'intégration des méthodes quantitatives (qui n'ont pas de raison de n'être appliquées qu'à des grands échantillons mal connus) et des méthodes qualitatives (qui ne sont pas obligées d'étudier des cas uniques) est si peu développée que des champs importants restent à défricher. Le préalable, et c'est là un objectif de ce texte, est d'accepter l'idée que l'histoire critique des méthodes peut inciter à des remises en cause du ritualisme des pratiques.



Références

- Allison, Paul. 1999a. *Multiple Regression : A Primer*. Thousands Oaks : Pine Forge Press.
- . 1999b. « Comparing Logit and Probit Coefficients Across Groups ». *Sociological Methods and Research*, 28 : 186-208.
- Abbott, Andrew. 1998. « The Causal Devolution ». *Sociological Method et Research* 27 : 148-181.
- Altonji, Joseph et Charles Pierret. 2001. « Employeur Learning and Statistical Discrimination ». *The Quarterly Journal of Economics* 116(1) : 313-350.
- Ash, Arlene. « The Perverse Logic of Reverse Regression ». In *Statistics Methods in Discrimination Litigation*, édité par David Kaye et Mikel Aickin, 85-106. New York : Dekker, 1986.
- Barrett, Gerald et Donna Sansonetti. 1988. « Issues Concerning the Use of Regression Analysis in Salary Discrimination Cases ». *Personal Psychology* 41 : 503-16.
- Bastide, Roger et Florestan Fernandes. 1955. *Relações raciais entre brancos e negros em São Paulo*. São Paulo : Anhembi.
- Beck, Elwood, Patrick Horan et Charles Tolbert. 1980. « Industrial segmentation and labor market discrimination ». *Social Problems* 28 (2) : 113-130.
- Becker, Gary. 1957. *The Economics of Discrimination*. Chicago : Chicago UP.
- Becker, Howard S. 2017. *Evidence*. Chicago : Chicago UP.
- Bernardi, Fabrizio, Lela Chakhaia et Liliya Leopold. 2016. « ‘Sing Me a Song with Social Significance’ : The (Mis)Use of Statistical Significance Testing in European Sociological Research ». *European Sociological Review* 33(1) : 1-15.
- « Beyond the Prima Facie Case in Employment Discrimination Law : Statistical Proof and Rebuttal ». 1975. *Harvard Law Review* 89 (2) : 387-422.
- Bielby, William et James Baron. 1980. « Men and Women at Work : Sex segregation and Statistics Discrimination ». *American Journal of Sociology* 91(4) : 759-99.
- Bielby, William, Robert Hauser et David Featherman. 1977. « Response Errors of Black and Non Black Males in Models of the Intergenerational Transmission of Socioeconomic Status ». *American Journal of Sociology* 82(6) : 1242-88

- Black, Dan, Amelia Haviland, Seth Sanders et Lowell Taylor. 2006. « Why Do Minorities Earn Less? A Study of Wage Difference Among the Highly Educated ». *The Review of Economics and Statistics* 88(2) : 300-13.
- Black, Dan, Natalia Kolesnikova, Seth Sanders et Lowell Taylor. 2013. « The Role of Location in Evaluating Racial Wage Disparity ». *IZA Journal of Labor Economics* 2. <https://doi.org/10.1186/2193-8997-2-2>.
- Blalock, Hubert. 1961. « Theory, Measurement and Replication ». *American Journal of Sociology* 66(4) : 342-347.
- . 1962. « Occupational Discrimination : Some Theoretical Propositions ». *Social Problems* 9(3) : 240-247.
- Blinder, Alan. 1973. « Wage Discrimination : Reduced Form and Structural Estimates ». *Journal of Human Resources* 8(4) : 436-55.
- Brochier, Christophe. 2020. « Mesurer la discrimination ? Une critique radicale de la décomposition de Oaxaca-Blinder ». *Bulletin de méthodologie sociologique* 145(1) : 61-79.
- Didier, Emmanuel et Isabelle Bruno. 2021. « Le “statactivisme” comme usage militant de la quantification ». *Sociologias* 23(56) : 82-109.
- Budig, Michelle, Misun Lim et Melissa Hodges. 2021. « Racial and Gender Pay Disparities : The Role of Education ». *Social Science Research*, sous presse.
- Cain, Glen. 1976. « The Challenge of Segmented Labor Market Theories to Orthodox Theories : a Survey ». *Journal of Economic Literature* 14(4) : 1215-57.
- Carlsson, Magnus and Dan-Olof Rooth. 2006. « Evidence of Ethnic Discrimination in the Swedish Labor Market Using Experimental Data ». *IZA Discussion Paper* 2281 : 1-26.
- Cohen Jacob et Patricia Cohen. 2009. *Applied Multiple Regression Correlation Analysis for the Behavioural Sciences*. New York : Psychology Press.
- Connolly, Catherine. 1991. « The Use of Multiple Regression in Employment Discrimination Cases ». *Population Research and Policy Review* 10 : 117-35.
- Cook, Thomas. 2008. « Waiting for Life to Arrive : a History of the Regression Discontinuity Design in Psychology, Statistics and Economic ». *Journal of Econometrics* 142 : 636-54.
- Corcoran, Mary et Greg Duncan. 1979. « Work History, Labor Force Attachment, and Earnings Differentials Between the Races and Sexes ». *Journal of Human Resources* 14(1) : 3-20.
- Cotton, Jeremiah. 1988. « On the Decomposition of Wage Differentials ». *The Review of Economics and Statistics* 70(2) : 236-43.
- Crews, Aaron. « The Big Move Toward Big Data in Employment ». In *Data-Driven Law : Data Analytics and the New Legal Services*, édité par Edward Walters, 59-102. New York : Taylor and Francis, 2019.

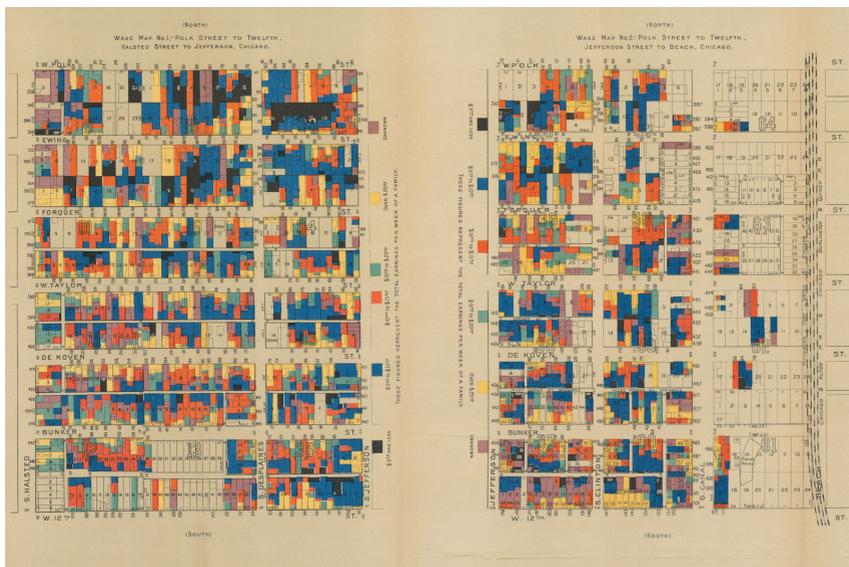
- Demetriou, Charles et Victor Roodometof. 2014. « The History of Historical-Comparative Methods in Sociology ». In *The Oxford Handbook of Qualitative Research*, edited by Patricia Leavy, 43-61. Oxford : Oxford UP.
- Dempster, Arthur. 1988. « Employment Discrimination and Statistical Science ». *Statistical Science* 3(2) : 149-61.
- Dewey, Donald 1952. « Negro Employment in Southern Industry ». *The Journal of Political Economy* 60(4) : 279-93.
- Duncan, Otis. 1984. *Notes on Social Measurement Historical and Critical*. Beverly Hills : Sage.
- . 1967. « Discrimination Against Negroes ». *The Annals of the American Academy of Political and Social Science*, 371 : 85-103.
- Duncan, Otis et Peter Blau. 1967. *The American Occupational Structure*. New York : Wiley.
- Ellis, Paul, 2010. *The Essential Guide to Effect Sizes*. Cambridge : Cambridge UP.
- Finkelstein, Michael. 1980. « The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Case ». *Columbia Law Review*, 80(4) : 737-54.
- Finkelstein, Michael et Bruce Levin. 2015. *Statistics for Lawyers*. New York : Springer.
- Flanagan, Robert. 1974. « Labor Force Experience, Job Turnover and Racial Wage Differentials ». *The Review of Economics and Statistics* 56 (4) : 521-529.
- Freedman, David. 2006. « Statistical Models for Causation : What Inferential Leverage do they Provide ». *Evaluation Review* 30 : 694-715.
- . 1997. « From Causation to Regression ». *Advances in Applied Mathematics* 18 : 59-110.
- . 1995. « Some Issues in the Foundation of Statistics ». *Foundations of Science* 1 : 19-39.
- Gilman, Harry. 1965. « Economic Discrimination and Unemployment ». *The American Economic Review* 55 : 1077-96.
- Goldsmith, Arthur, Darrick Hamilton et William Darity Jr. 2006. « Shades of Discrimination : Skin Tones and Wages ». *The American Economic Review* 96(2) : 242-45.
- Goodson, Ivor. 2001. « Story of Life History. Origins of the Life History Method in Sociology ». *Identity* 1 : 129-42.
- Gray, Mary. 1993. « Can Statistics Tell us What We Do not Want to Hear? The Case of Complex Salary Structures ». *Statistical Science* 8(2) : 144-58.
- Heckman, James. 1998. « Detecting Discrimination ». *Journal of Economic Perspectives* 12(2) : 101-16.
- . 1979. « Sample Selection Bias as a Specification Error ». *Econometrica* 47(1) : 153-61.
- Henkel, Ramon et Denton Morrison. 1970. *The Significance Test Controversy*. Chicago : Aldine.
- Heywood, John et Patrick O'Halloran. 2005. « Earning Differentials and Performance Pay ». *The Journal of Human Resources* 40(2) : 435-52.

- Hill, Richard. 1974. « Unionization and Racial Income Inequality in the Metropolis ». *American Sociological Review* 39(4) : 507-22.
- Hisahiro Naito et Yu Takagi. 2017. « Is Racial Salary Discrimination Disappearing in the NBA? Evidence From Data During 1985-2015 ». *International Review of Applied Economics* 3(5) : 651-69.
- Hodge, Robert et Patricia Hodge. 1965. « Occupational Assimilation as a Competitive Process ». *American Journal of Sociology*, 71(3) : 249-64.
- Holland, Paul. 2003. « Causation and Race ». Research report. Princeton Educational Testing Service.
- Jatteau, Arthur et Agnès Labrousse. 2020. « Les dix limites de la méthode Duflo ». *Alternatives Économiques* 88 : 93-106.
- Kain, John. 1968. « Housing Segregation, Negro Employment and Metropolitan Decentralization ». *The Quarterly Journal of Education* 82(2) : 175-97.
- Karlson, Kristian, Anders Holm et Richard Breen. 2012. « Comparing Regression Coefficients Between Same-sample Nested Models Using Logit and Probit : A New Method ». *Sociological Methodology* 42 : 286-313.
- Kaufman, Robert. 1983. « A Structural Decomposition of Black-White Earnings Differentials ». *American Journal of Sociology* 89(3) : 585-611.
- Klein, David. 1987. « Causation in Sociology Today : a Revised View ». *Sociological Theory* 5(1) : 19-26.
- Kofi, Charles K. et Jonathan Guryan. 2011. « Studying Discrimination : Fundamental Challenges and Recent Progress ». *Annual Review of Economics* 3(1). <http://doi.org/10.1146/annurev.economics.102308.124448>
- Lang, Kevin et Jee-Yeon Lehman. 2012. « Racial Discrimination in The Labor Markets, Theory and Empirics ». *Journal of Economic Literature* 50(54) : 1-48.
- Leigh, Duane et Lane Rawlins. 1974. « Racial Differentials in Male Unemployment Rates : Evidence From Low Income Urban Areas ». *The Review of Economics and Statistics* 56(2) : 150-57.
- McCabe, George. 1986. « Regression Analysis in Discrimination cases ». In *Statistics Methods in Discrimination Litigation*, édité par David Kaye et Mikel Aickin, 69-84. New York : Dekker.
- Lieberman, Stanley. 1985. *Making it Count*. Los Angeles : University of California Press.
- Mandel, Hadas et Moshe Semyonov. 2016. « Going Back in Time ? Gender Differences in Trends And Sources of Racial Pay Gap ». *American Sociological Review* 81(5) : 1039-68.
- Mare, Robert et Christopher Winship. 1984. « The Paradox of Lessening Racial Inequality and Joblessness among Black Youth : Enrollment, Enlistment and Employment, 1964-1981 ». *American Sociological Review* 49(1) : 39-55.

- Maume, David Jr. 2004. « Wage Discrimination Over the Life Course : a Comparison of Explanations ». *Social Problems* 51(4) : 505-27.
- Mayer, Nonna, Guy Michelat, Vincent Tiberj et Tommaso Vitale. « Evolution et structure des préjugés : Le regard des chercheurs ». In *La lutte contre le racisme, l'antisémitisme et la xénophobie. Année 2019*, 55-75. Paris : La Documentation française.
- Meier, Paul. 1986. « Damned Liars and Expert Witnesses ». *Journal of the American Statistician* 81(394) : 269-76.
- Meier, Paul, Jerome Sacks et Sindy Zabell. 1984. « What Happened in Hazeland : Statistics, Employment Discrimination and the 80% Rule ». *The American Bar Foundation Research Journal* 9(1) : 139-86.
- Mc Call, Leslie. 2001. « Sources of Racial Wage Inequality in Metropolitan Labor Markets : Racial, Ethnic and Gender Differences ». *American Sociological Review* 66(4) : 520-41.
- Mood, Carina. 2010. « Logistic Regression : Why We Cannot Do What We Think We Can Do, and What We Can Do About It. ». *European Sociological Review* 26(1) : 67-82.
- Mouw, Ted. 2000. « Job Relocation and the Racial Gap in Unemployment in Detroit and Chicago, 1980 to 1990. *American Sociological Review* 65(5) : 730-53.
- National Council. 2004. *Mesuring Racial Discrimination*. Washington : National Academy Press.
- Neal, Derek et William Johnson. 1996. « The Role of Market Factors in Black-White Wage Differentials ». *Journal of Political Economy* 104(5) : 869-95.
- Oaxaca, Ronald. 1973. « Male-Female Wage Differences in Urban Labor Markets ». *International Economics Review* 14(3) : 693-701.
- Pager, Devah et Eric Grodsky. 2001. The Structure of Disadvantage : Individual and Occupational Determinants of The Black/White Wage Gap ». *American Sociological Review* 66(4) : 542-67.
- Pager, Devah et David Pedulla. « Race, Self Selection and The Job Search Process ». *American Journal of Sociology*, 120, 4 (2015) : 1005-1054.
- Pager Devah et Hana Shepherd. 2008. « The Sociology of Discrimination : Racial Discrimination in Employment, Housing, Credit and Consumer Markets ». *Annual Review of Sociology*, 34 : 181-209.
- Palmer, Edward. 1947. « Discrimination in Urban Employment ». *American Journal of Sociology* 52(4) : 357-61.
- Peneff, Jean. 1984. « La Fabrication Statistique ou le Métier du Père. » *Sociologie du travail* 26(2) : 195-211.
- Platt, Jennifer. 2002. « The History of the Interview ». In *Handbook of Interview Research*, édité par Jaber Gubrium et James Holstein, 9-26. London : Sage.
- Ragin, Charles, Susan Mayer et Kriss Drass. 1984. « Assessing Discrimination : A Boolean Approach ». *American Sociological Review*, 49(2) : 221-34.

- Rasmussen, David. 1971. « Discrimination and the Income of Non White Males ». *The American Journal of Economics and Sociology* 30(4) : 377-382.
- Rayack, Elton. 1961. « Discrimination and the Occupational Progress of Negroes ». *Review of Economics and Statistics* 43) : 209-14.
- Restifo, Salvatore, Vincent Roscigno et Zenchao Quian. 2013. « Segmented Assimilation, Split Labor Markets and Racial/Ethnic Inequality : The Case of Early 20th-Century New York ». *American Sociological Review* 78(5) : 897-924.
- Rexroat, Cynthia. 1978. « The Changing cost of Being a Black Woman ». *Sociology of Work and Occupation* 5(3) : 341-59.
- Ritter, Joseph and Taylor Lowell. 2011. « Racial Disparity in Employment ». *The Review of Economics and Statistics* 93(1) : 30-42.
- Roberts, Harry. 1988. « Employment Discrimination and Statistical Science : A Comment ». *Statistical Science* 3(2) : 167-70.
- Roth, Julius. 1966. « Hired Hand Research ». *American Sociologist* 1(4) : 190-96.
- Rhum, Christopher. 1989. « Labor Market Discrimination in the US ». In *Affirmative Action in Perspective*, édité par Fletcher Blanchard et Faye Crosby, 149-58. New York : Springer.
- Saunders, Robert et Ronald Coccari. 1973. « Racial Earnings Differentials : Some Economic Factors ». *The American Journal of Economics and Sociology* 52(2) : 150-59.
- Selvin, Hanan. 1957. « Critique of Tests of Significance in Survey Research ». *American Sociological Review* 22(5) : 519-527.
- Shepherd William et Sharon Levin. 1973. « Discrimination in Large Firms ». *The Review of Economics and Statistics* 52(2) : 150-159
- Siegel, Paul. 1965. On the Cost of Being a Negro. *Social Inquiry* (35) : 41-57.
- Silva Do Valle, Nelson. 1980. « O preço da cor : Diferenciais raciais na distribuição da renda no Brasil ». *Pesquisa e planejamento econômico* 10(1) : 21-44.
- Smith, Lewis, Vernon Briggs, Brian Rungeling et James Smith Jr. 1978. « Wage Occupational Differences Between Blacks and Whites ». *Southern Economic Journal* 45(1) : 250-57.
- Sterkel, Studs. 2010.. Race. Histoires orales d'une obsession américaine. Paris : Ed. Amsterdam
- Stratton, Leslie. 1993. « Racial Differences in Men's Unemployment ». *Industrial and Labor Relation Review* 46(3) : 451-63.
- Thompson, Bruce. 2002. « What Future Quantitative Social Science Research Could Look Like : Confidence Intervals for Effect Sizes ». *Educational Researcher* 31(3) : 25-32.
- Turner, Ralph. 1952. « Foci of Discrimination in the Employment of Non White ». *American Journal of Sociology* 58(3) : 247-56.

- Weiss, Randall. 1970. «The Effect of Education on the Earnings of Blacks and Whites». *The Review of Economics and Statistics* 52(2) : 150-59.
- Wright, Erik. 1978. «Race, Class and Income Inequality». *American Journal of Sociology* 83(6) : 1368-97.
- Ziliak Stephen et Deirdre McCloskey. 2008. *The Cult of Statistical Significance : How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor : University of Michigan Press.



Hull House Wage Maps (*part.*), 1895
(Persuasive Cartography. The P.J. Mode Collection, Cornell University Library,
<https://digital.library.cornell.edu/catalog/ss:3293790>).