La régression sur discontinuité

Denis Fougère, Nicolas Jacquemet

▶ **To cite this version:**

Denis Fougère, Nicolas Jacquemet. La régression sur discontinuité. LIEPP Fiche méthodologique n°13, 2023. hal-04103335

HAL Id: hal-04103335

https://sciencespo.hal.science/hal-04103335

Submitted on 23 May 2023

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

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

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

SciencesPo
LABORATORY FOR INTERDISCIPLINARY EVALUATION OF PUBLIC POLICIES

LIEPP FICHE MÉTHODOLOGIQUE n°13

2023

Cette fiche fait partie d'un ensemble de fiches méthodologiques publiées par le LIEPP. A ce titre, elle bénéficie du soutien apporté par l'ANR et l'État au titre du programme d'Investissements d'avenir dans le cadre de l'IdEx Université Paris Cité (ANR-18-IDEX-0001).

La régression sur discontinuité

Denis FOUGÈRE (Sciences Po, CRIS, CNRS, LIEPP, CEPR et IZA)
denis.fougere@sciencespo.fr

Nicolas JACQUEMET (Université Paris 1, CES, PSE)
nicolas.jacquemet@univ-paris1.fr

Partage selon les Conditions Initiales 4.0 International License

www.sciencespo.fr/liepp

Comment citer cette publication :
FOUGERE, Denis, et JACQUEMET Nicolas, **La régression sur discontinuité**, *LIEPP Fiche méthodologique* n°13, 2023-05

This publication is also available in English:
FOUGERE, Denis, and JACQUEMET Nicolas, **The regression discontinuity design**, *LIEPP Methods Brief* n°14, 2023-05

EN BREF :

La régression sur discontinuité est une méthode quantitative quasi-expérimentale qui évalue l'impact d'une intervention en comparant les observations qui se situent à proximité d'un seuil d'éligibilité fixé par les autorités en charge de la politique étudiée. L'existence d'un tel seuil (par exemple, l'éligibilité à la politique à partir de tel âge, ou en dessous de tel niveau de revenu) permet en effet de reconstituer un groupe cible et un groupe de contrôle, de façon analogue à une démarche expérimentale.

Mots-clés : Méthodes quantitatives, méthodes quasi-expérimentales, seuil d'éligibilité, variable de forçage, régression sur discontinuité stricte/floue, fenêtre d'observation, amplitude d'intervalle, monotonicité, personnes observantes

I. En quoi cette méthode est-elle utile pour l'évaluation des politiques publiques ?

Lorsque l'on souhaite réaliser l'évaluation quantitative des effets d'une politique publique, la difficulté principale consiste à trouver un groupe de comparaison (appelé groupe de contrôle) dont la situation puisse servir de point de référence (i.e., « de contrefactuel » : voir la fiche 'Méthode des doubles différences') à celle que connaissent les bénéficiaires de l'intervention (groupe cible ou groupe de traitement). L'expérimentation randomisée, dans laquelle les bénéficiaires et les non-bénéficiaires sont choisis aléatoirement dans une population éligible donnée, constitue le cadre de référence idéal pour définir un groupe de contrôle valide : par construction, si l'on dispose d'un échantillon assez grand, le groupe de contrôle aura les mêmes caractéristiques (de genre, d'âge, de niveau d'éducation, etc.) que le groupe de traitement.

Les méthodes quasi-expérimentales visent à pallier l'absence d'expériences contrôlées en s'appuyant sur des variations qui surviennent spontanément (en général, par décision des pouvoirs publics) et produisent des observations qui se rapprochent d'une situation expérimentale. Les méthodes d'appariement ou d'estimation par doubles différences exploitent les cas dans lesquels la mise en place d'une politique publique produit naturellement deux groupes dont la comparaison permet, sous certaines conditions, d'en mesurer l'effet. La méthode de régression sur discontinuité, quant à elle, exploite l'application d'un seuil d'éligibilité pour décider du déclenchement de l'intervention, qui produit l'équivalent d'une expérience randomisée locale au voisinage du seuil.

II. En quoi consiste cette méthode ?

Lorsque l'accès à une intervention ou à une politique publique est conditionné par un seuil fixé par les autorités en charge de cette politique, l'intervention produit deux groupes dont l'un seulement bénéficie de l'intervention. Mais ces groupes ne sont pas directement comparables puisqu'ils diffèrent par construction en raison de la valeur de la caractéristique définissant le seuil. Ce seuil peut être une condition d'âge (pour un départ en retraite par exemple), de taille pour les entreprises (par exemple, une politique de réduction des charges destinée aux entreprises de moins de 20 salariés) ou encore un niveau de ressources donnant accès à une bourse d'études ou à un crédit d'impôt. Comme le montrent ces exemples, l'hypothèse selon laquelle la variable à laquelle le seuil s'applique (par exemple l'âge, ou la taille de l'entreprise), communément appelée « variable de forçage » (ou *forcing variable*) n'aurait pas d'influence sur la variable de résultat de l'intervention, est en général peu crédible. Le départ en retraite va de pair avec une augmentation de l'âge qui a par lui-même de nombreuses conséquences sur l'état de santé, les habitudes de consommation, la vie sociale, etc. Les entreprises de grande taille interviennent dans des secteurs généralement distincts de ceux dans lesquels opèrent les PME, leur structure et leur activité sont souvent très différentes. Le niveau de revenu a évidemment un impact

majeur sur de nombreuses décisions des ménages. Dans ces conditions, les deux groupes ainsi formés ne permettent pas de réaliser une évaluation de l'effet de l'intervention en comparant directement la valeur de la variable de résultat entre les bénéficiaires et les non-bénéficiaires.

L'application d'un seuil d'éligibilité produit en revanche une discontinuité soudaine dans la répartition des observations qui se situent à **proximité du seuil** : les observations dont la variable de forçage se trouve juste au-dessous du seuil bénéficient de l'intervention tandis que leurs voisins, dont la variable de forçage est située juste en-dessus, en sont exclus. La régression sur discontinuité exploite cette propriété en faisant l'hypothèse que les variations de faible ampleur de la variable de forçage autour du seuil résultent d'un pur aléa, similaire à un tirage au sort, qui détermine l'accès à l'intervention d'observations par ailleurs identiques. A proximité du seuil, l'affectation d'une observation au groupe de traitement s'apparente ainsi à une expérimentation randomisée. Lorsque les observations sont ordonnées en fonction de la valeur croissante de la variable de forçage, toute rupture dans la valeur de la variable de résultat une fois le seuil franchi peut, sous cette hypothèse, être interprétée comme une mesure de l'effet de l'intervention.

Dans sa forme la plus simple, la méthode de régression sur discontinuité mesure donc l'effet d'une politique en comparant la valeur moyenne de la variable de résultat dans le groupe des personnes bénéficiaires, par exemple celles dont le revenu ou l'âge est juste inférieur au seuil d'éligibilité, et la valeur moyenne de cette variable dans le groupe de contrôle comparable, formé des personnes dont le revenu ou l'âge est juste supérieur à ce seuil. L'hypothèse sous-jacente est que, pour des personnes ayant par ailleurs les mêmes caractéristiques du point de vue de la qualification, du niveau d'éducation ou du genre, celles situées juste en-dessous et au-dessus du seuil sont identiques. Cette mise en œuvre de la méthode nécessite donc de définir l'intervalle au sein duquel les observations sont conservées pour l'analyse. Cette « fenêtre d'observation » est choisie en réalisant un arbitrage entre la qualité de l'analyse statistique permise par un échantillon de plus grande taille et la fragilisation de cette hypothèse qui résulte d'un élargissement de l'intervalle. Imbens et Kalyanaraman (2012) proposent une méthode pour choisir de manière optimale l'amplitude de cet intervalle (bandwidth).

Cette régression sur discontinuité est dite stricte (« *sharp regression discontinuity design* ») lorsque l'affectation au groupe pouvant bénéficier de l'intervention ou du dispositif public est obligatoire et strictement déclenchée par la valeur de la variable de forçage. Si l'éligibilité est fondée, par exemple, sur un critère d'âge, et appliquée par une autorité qui dispose d'un recensement exhaustif de la population, alors la probabilité de bénéficier de l'intervention est égale à 1 dès lors que la condition d'âge est remplie ; et cette probabilité est égale à 0 sinon, de sorte que l'affectation en fonction du seuil est un événement certain. Prenons l'exemple d'un programme de formation pour les personnes en recherche d'emploi ayant 25 ans révolus. Le principe est alors de comparer la moyenne de la variable de résultat (la variable dépendante, par exemple, le salaire d'embauche au moment du retour à l'emploi) pour les personnes en recherche d'emploi qui sont juste au-dessus du seuil d'âge, par exemple âgées de 25 ou 26 ans, et le salaire moyen d'embauche pour celles qui sont âgées de 23 ou 24 ans, qui n'ont pu bénéficier du programme de formation.

La régression sur discontinuité floue (*fuzzy regression discontinuity design*) correspond à l'inverse aux situations dans lesquelles ce seuil est moins contraignant, si bien qu'il existe de part et d'autre du seuil des observations qui sont ou non bénéficiaires de l'intervention. Dans l'exemple du programme de formation destiné aux personnes en recherche d'emploi de 25 ans et plus développé ci-dessus, supposons que, dans une localité donnée, cette formation ne puisse être dispensée qu'à 100 personnes âgées de 25 ou 26 ans en raison de contraintes budgétaires, et que cette formation ne soit pas obligatoire, de sorte que seuls 80 de ces 100 personnes éligibles (soit 80% d'entre elles) acceptent effectivement de participer à la formation. L'agence locale pour l'emploi propose alors les 20 places restantes à des personnes de 23 ou 24 ans qui sont également au nombre de 100. Seules 10 d'entre elles (soit 10%) acceptent de participer à la formation. Plutôt qu'à un changement soudain du statut vis-à-vis de

l'intervention, la notion de discontinuité fait donc référence ici au « saut » que subit la probabilité de bénéficier de l'intervention lorsque le seuil d'éligibilité (25 ans) est franchi. L'objectif de l'approche est donc de mesurer l'effet de l'intervention en se restreignant à la variation de la variable de résultat qui résulte de ce « saut » dans la probabilité de bénéficier de l'intervention évaluée. Cette approche repose sur une hypothèse forte, appelée hypothèse de monotonicité (« *monotonicity* ») : cette hypothèse implique que parmi les chômeuses et chômeurs qui ne participent pas au programme de formation parce que leur âge est inférieur à 25 ans, il existe un sous-groupe d'individus qui accepteraient d'y participer si leur âge était de 25 ans révolus. En termes techniques, ces observations sont appelées « *compliers* », terme que l'on peut traduire par « personnes observantes », « qui obtempèrent » ou « qui se conforment à la prescription ». Par construction, la régression sur discontinuité floue mesure l'effet de l'intervention uniquement pour ce sous-groupe. Outre que ce sous-groupe peut être parfois de taille très restreinte, il exclut deux groupes importants : les individus qui sont prêts à participer à l'intervention, quelle que soit la valeur de la variable de forçage (les « *always takers* »), ainsi que ceux qui ne souhaitent pas y participer en toutes circonstances (les « *never takers* »).

III. Deux exemples d'utilisation de cette méthode dans le secteur éducatif

Les variations de prix de l'immobilier entre quartiers reflètent la disposition à payer des contribuables pour l'ensemble des services et aménités (caractéristiques de l'environnement) auquel un logement donne accès. L'une de ces aménités est bien sûr la qualité de l'école de secteur à laquelle les enfants des résidents pourront accéder. Les tentatives d'estimation des effets de la qualité des écoles sur le prix des logements sont souvent peu convaincantes car les meilleures écoles ont tendance à se trouver dans les meilleurs quartiers. Ces deux aspects contribuent conjointement à une élévation du prix de l'immobilier et les évaluations qui ne tiennent pas suffisamment compte des caractéristiques du quartier tendent donc à surestimer la valeur des écoles qui y sont situées. Pour contourner cette difficulté, Black (1999) recourt à une application particulièrement originale de la méthode de la régression sur discontinuité stricte, en s'appuyant sur un seuil correspondant aux contours de la carte scolaire à Boston. L'étude consiste à estimer la valeur que les parents accordent à la qualité de l'école publique de secteur, en comparant les logements qui sont situés de part et d'autre des limites géographiques d'un secteur scolaire incluant les écoles publiques auxquelles les enfants sont affectés. Le fait que les résultats moyens obtenus par les élèves d'écoles de secteurs différents mais voisins varient parfois fortement, alors que les caractéristiques des logements situés des deux côtés des divisions scolaires changent par définition assez peu, permet d'isoler par discontinuité la relation entre les résultats scolaires (interprétés comme la qualité des écoles) et les prix des logements. Les estimations suggèrent qu'une augmentation d'un point de la moyenne obtenue aux tests scolaires entraîne une augmentation de 1,3% à 1,6% du niveau des prix des logements situés à la limite du secteur scolaire.

L'étude de Matsudaira (2008) constitue un exemple de mise en œuvre de la méthode de la régression sur discontinuité floue, également appliquée à la réussite scolaire. L'étude exploite un ensemble de données administratives provenant d'un grand district scolaire des États-Unis. Dans ce district, les élèves accèdent à la classe suivante si leurs notes sont supérieures à des seuils prédéfinis. Les élèves obtenant des notes inférieures à ces seuils doivent participer à une école d'été de quatre à six semaines afin d'éviter le redoublement. Étant donné que les caractéristiques observées des élèves au voisinage des seuils sont presque identiques, les différences de résultats scolaires entre les élèves qui se situent juste en dessous et juste au-dessus des seuils peuvent être attribuées à l'impact causal de l'école d'été. L'échantillon est restreint aux élèves scolarisés entre le cours élémentaire de deuxième année (CE2, à l'âge de huit ans environ) et la classe de cinquième (à l'âge de 12 ans environ). Les résultats des élèves ont été enregistrés aux examens de mathématiques et de lecture aux printemps 2001 et 2002, ce qui donne un échantillon d'analyse de 338 608 élèves. Cependant, la régression sur discontinuité est ici floue : la relation entre les résultats aux tests de fin d'année et la fréquentation de l'école d'été n'est pas

absolument obligatoire. Certains élèves dont les notes étaient inférieures aux seuils n'ont pas suivi cette école d'été, alors que d'autres, dont les notes étaient supérieures aux seuils, s'y sont inscrits. Seuls 38% des élèves de CE2 et de cours moyen de deuxième année (CM2) dont les notes en mathématiques étaient inférieures aux prérequis à la fin de l'année scolaire 2000-2001 se sont inscrits à l'école d'été de 2001. Les estimations provenant de la mise en œuvre de la régression sur discontinuité floue suggèrent que les notes des élèves « observants » de CE2 ont augmenté de 12,8% l'année suivante, alors que celles des élèves « observants » de CM2 participant à cette école d'été ont augmenté de 24,1%. Les effets se sont avérés faibles pour les élèves « observants » de sixième, voire inexistants pour les élèves « observants » de cours moyen de première année (CM1) et de cinquième.

IV. Quels sont les critères permettant de juger de la qualité de la mobilisation de cette méthode ?

Pour que la discontinuité s'apparente à une expérimentation locale, il est important que la variable de forçage soit une donnée objective qui échappe au contrôle des populations concernées par l'intervention. S'il est possible de « manipuler » le positionnement de cette variable par rapport au seuil, l'affectation au groupe de traitement devient alors une variable de choix. L'exemple classique est celui d'une politique publique qui accorde des aides à l'emploi aux entreprises de moins de 20 salariés. La réaction naturelle de certaines entreprises dont l'effectif s'approche du seuil est d'intensifier le recrutement d'intérimaires, afin d'augmenter leur effectif sans que cette augmentation n'apparaisse dans les déclarations fiscales auxquelles elles sont soumises, de manière à continuer à bénéficier des aides à l'emploi. Pour détecter de telles manipulations du seuil, McCrary (2008) propose un test statistique simple, qui repose sur un raisonnement agrégé. Les entreprises qui emploient en réalité plus de 20 salariés (21 ou 22 salariés par exemple), mais dont la taille déclarée est inférieure à 20 salariés (soit 19 ou 20), vont faire croître artificiellement la proportion d'entreprises de moins de 20 salariés dans le secteur et simultanément faire diminuer la proportion d'entreprises employant 21 ou 22 salariés. L'existence de manipulations en réaction au seuil d'éligibilité a donc des conséquences sur la distribution de la taille des entreprises, qui peut être inspectée à l'aide d'un histogramme. En théorie, cet histogramme ne devrait pas faire apparaître de discontinuité juste avant et juste après le seuil de 20 salariés. Si tel était toutefois le cas, et cela peut être testé statistiquement, on pourrait alors suspecter un comportement de « manipulation » de la part des entreprises.

Pour éviter de réduire trop fortement la fenêtre d'observation utilisée, il est fréquent d'ajouter des variables explicatives autres que la variable de forçage : cela permet de tenir compte des variations de la variable de résultat qui sont dues à certaines caractéristiques observables. Le revenu, par exemple, a tendance à croître avec l'âge, de sorte qu'un élargissement de la fenêtre autour du seuil d'âge conduit à inclure des observations dont la variable de résultat change en raison des variations de revenu. La prise en compte de cet effet du revenu permet d'éliminer cette différence entre les groupes. Il est important de vérifier que la distribution de ces variables ne présente pas de discontinuité au voisinage du seuil considéré. Dans le cas contraire, cela signifie que l'intervention que l'on souhaite évaluer a des effets non seulement sur la variable de résultat mais aussi sur ces variables de contrôle. Prendre en compte ces variables dans l'analyse statistique conduit à des estimations biaisées de l'effet de l'intervention sur la variable de résultat, car ces variables sont elles-mêmes expliquées par la mise en œuvre de l'intervention.

V. Quels sont les atouts et les limites de cette méthode par rapport à d'autres ?

La principale difficulté à laquelle se heurtent la plupart des méthodes quasi-expérimentales est qu'elles reposent sur des hypothèses fortes, souvent remises en cause, quant à la comparabilité du groupe contrôle et du groupe cible après intervention. C'est ce qui conduit, par exemple, lorsque l'on souhaite mettre en œuvre la méthode des doubles différences, à s'assurer à la fois que la variable de résultat a

connu par le passé la même évolution dans les deux groupes et que leurs caractéristiques observables sont semblables. La difficulté est la même lorsque l'on souhaite recourir à une méthode d'appariement d'échantillon (*matching*) : il convient pour ce faire de trouver des observations servant de groupe de contrôle qui présentent certes des caractéristiques observables similaires à celles des observations du groupe cible, mais qui ont surtout une probabilité non nulle d'être éligibles à la politique évaluée. La méthode de régression sur discontinuité échappe à cette difficulté car elle repose sur un principe d'affectation aléatoire au sein de la sous-population qui se situe au voisinage du seuil. A l'instar d'une expérimentation contrôlée, la comparabilité entre les observations appartenant à chacun des deux groupes repose sur un argument statistique : si la taille de l'échantillon est suffisante, la distribution de toutes les caractéristiques qui sont pertinentes pour expliquer les variations de la variable de résultat est similaire entre les deux groupes.

Cette assimilation de la discontinuité à une expérimentation aléatoire est d'autant plus convaincante que l'intervalle à l'intérieur duquel elle est supposée se dérouler est étroit, ce qui conduit à restreindre l'effet mesuré à une sous-population très particulière, qui se caractérise par la proximité de la variable de forçage par rapport au seuil. La mesure fournie par cette expérimentation aléatoire locale est donc spécifique à cette sous-population. Dès lors que l'effet de l'intervention varie fortement entre différents sous-groupes, la mesure qui en découle est donc elle-même locale et seulement valable au voisinage du seuil retenu (ce qui correspond à un effet local moyen, ou LATE « *local average treatment effect* »). L'extrapolation des résultats obtenus à des sous-populations éloignées du seuil (qui définirait la « validité externe » de la mesure obtenue) n'a que très peu de pertinence. Cette limite de la méthode est encore amplifiée dans le cas d'une régression sur discontinuité floue, dont l'effet local est spécifique aux seules entités « observantes ». Ce manque de validité externe est d'autant plus problématique que les seuils sont souvent fixés en fonction du bénéfice attendu de l'intervention dans les populations ciblées. Un programme de formation à destination des personnes qui connaissent une situation de chômage de longue durée vise par exemple à contrecarrer les effets des pertes en capital humain dues à des épisodes de chômage prolongés. La fixation d'un seuil permettant de distinguer les épisodes de chômage de longue durée est en partie fondée sur le fait que cette perte de capital humain reste minime lorsque les épisodes sont suffisamment courts : l'estimation de l'effet d'un tel programme par la méthode de la régression sur discontinuité revient donc à centrer l'analyse sur la partie de la population pour laquelle l'effet du programme est très probablement le moins fort, à savoir les personnes en recherche d'emploi dont les épisodes de chômage sont relativement plus courts.

Le lecteur intéressé trouvera d'excellentes synthèses sur la méthode de la régression sur discontinuité, par exemple, dans l'article de Lee et Lemieux (2010), ou bien dans l'ouvrage de Cattaneo, Idrobo et Titiunik (2019)

Quelques références bibliographiques pour aller plus loin

Black, Sandra E. 1999. « Do Better Schools Matter? Parental Valuation of Elementary Education ». *Quarterly Journal of Economics* 114 (2): 577-99. https://doi.org/10.1162/003355399556070

Cattaneo, Matias D., et Nicolás Idrobo, et Rocío Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press. https://doi.org/10.1017/9781108684606

Imbens, Guido, et Karthik Kalyanaraman. 2012. « Optimal Bandwidth Choice for the Regression Discontinuity Estimator », *Review of Economic Studies,* 79 (3), 933-59. https://doi.org/10.1093/restud/rdr043

Lee, David S., et Thomas Lemieux. 2010. « Regression Discontinuity Designs in Economics ». *Journal of Economic Literature* 48 (2): 281-355. https://doi.org/10.1257/jel.48.2.281

Matsudaira, Jordan D. 2008. « Mandatory Summer School and Student Achievement ». *Journal of Econometrics*, The régression discontinuity design: Theory and applications, 142 (2): 829-50. https://doi.org/10.1016/j.jeconom.2007.05.015

McCrary, Justin. 2008. « Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test ». *Journal of Econometrics*, 142 (2): 698-714. https://doi.org/10.1016/j.jeconom.2007.05.005

Des ressources pour mettre en œuvre cette méthode avec les logiciels Stata et R

Cunningham, Scott. 2021, *Causal Inference: The Mixtape*. Yale University Press: New Haven and London. Disponible en libre accès sur le site https://mixtape.scunning.com/index.html

Huntington-Klein, Nick. 2022, *The Effect: An Introduction to Research Design and Causality*, Chapitre 20. Chapman and Hall/CRC Press: Boca Raton, Florida. Disponible en libre accès sur le site https://theeffectbook.net/ch-RegressionDiscontinuity.html