

L'abondance de données améliore l'évaluation des politiques publiques

À l'image de l'apprentissage automatique, la multiplication des données numériques disponibles ouvre de nouvelles voies méthodologiques pour jauger l'efficacité des mesures de politique économique. *Martin Huber*

Abrégé L'exploitation de données destinée à évaluer l'effet causal d'une mesure (comme un programme de qualification pour demandeurs d'emploi) sur le résultat visé (par exemple le retour à l'emploi) a connu d'importantes innovations ces dernières décennies. Il s'agit notamment de nouveaux procédés statistiques de mesure de la causalité et de l'utilisation de données plus informatives et étendues. Cet article donne un aperçu des techniques actuelles d'évaluation des politiques publiques, illustré notamment par des exemples pratiques en Suisse. Il évoque également de nouvelles méthodes liées à la progression exponentielle de la masse de données disponibles, méthodes qui permettent de combiner l'évaluation des politiques avec « l'apprentissage automatique », un sous-domaine de l'intelligence artificielle.

Quel est l'impact des cours de formation continue sur la réintégration des chômeurs, ou celui des services de garde d'enfants sur l'emploi des parents? Quelle influence peut avoir une réforme des rentes sur l'âge de la retraite? L'évaluation des politiques publiques permet de répondre à ce type de questions. À partir d'un ensemble de données, elle cherche à préciser l'effet d'une mesure déterminée (par exemple la formation continue) sur un résultat visé (comme le retour à l'emploi). Elle constitue souvent le fondement des analyses coûts-avantages des interventions des pouvoirs publics.

Pour définir avec précision l'effet d'une mesure, il faut que tous les paramètres pouvant influencer le résultat restent constants dans les groupes soumis à la mesure et dans ceux qui ne le sont pas. Par exemple, la comparaison des salaires moyens de deux groupes avec et sans formation continue ne livre d'information probante sur cet effet que si les caractéristiques pertinentes de ces groupes (âge, formation, expérience professionnelle, etc.), sont comparables. Cela reviendrait autrement à comparer des pommes et des poires en confondant l'effet des

mesures avec celui des caractéristiques. Dans les expériences, la comparabilité est assurée par le fait qu'une mesure s'applique de manière aléatoire, c'est-à-dire indépendamment des caractéristiques. Dans un contexte non expérimental, ce n'est généralement pas possible. Par exemple, les personnes qui suivent une formation continue diffèrent de celles qui ne la suivent pas en raison des particularités individuelles qui influencent leur choix d'y participer.

Diverses méthodes causales ont été développées pour l'évaluation des politiques, basées notamment sur le « modèle quasi expérimental », proche des conditions d'une expérience. En 2008, une étude lausannoise a par exemple ciblé l'influence du prolongement de l'allocation de chômage destinée aux plus de 50 ans en Autriche sur la durée du chômage¹. En comparant les résultats (la durée du chômage) des personnes de 50 ans et de 49 ans, situées de part et d'autre du seuil, il semble plausible, sous certaines hypothèses, que les deux groupes présentent des caractéristiques comparables compte tenu de la faible différence d'âge. L'effet d'une mesure sur les personnes proches d'un seuil peut ainsi être évalué de la même façon que dans une expérience. Cette approche est connue sous le nom de « régression par discontinuité ».

Programmes de formation tirés au sort

Dans une autre méthode, ce n'est pas la mesure elle-même, mais un instrument dit « quasi aléatoire » qui est appliqué et qui influence le résultat uniquement par son effet sur la mesure. Cette approche a été utilisée pour l'analyse du programme américain « Job Corps » en 2008,

1 Lalive (2008).

2 Schochet et al. (2008).

3 Berger et Lanz (2019).

lors de laquelle des jeunes défavorisés aux États-Unis ont été invités, sur la base d'une sélection aléatoire, à suivre une formation². L'invitation (instrument) a un effet sur la mesure (formation) parce qu'elle incite certaines personnes à participer. On suppose en outre qu'elle influence le résultat, par exemple le salaire, du seul fait de la décision de participer. Dans ce cas, l'effet d'une mesure sur le salaire peut être défini en divisant l'influence de l'invitation sur le salaire par l'influence de l'invitation sur la mesure.

Une autre approche est la méthode des « différences de différences ». Elle repose sur l'hypothèse que les résultats moyens de deux groupes (l'un avec la mesure, l'autre sans) auraient évolué de façon égale au fil du temps si aucun groupe n'avait bénéficié de la mesure. Une récente étude portant sur l'introduction d'un salaire

minimum dans certaines régions de Suisse et ses effets sur l'emploi en fournit une illustration³. Une « comparaison naïve » de l'emploi entre régions avec et sans salaire minimum ne renseigne pas sur l'effet de la mesure lorsque les régions diffèrent par des caractéristiques liées à l'emploi, comme la structure des branches. De même, la comparaison « avant-après » dans les régions ayant introduit un salaire minimum ne sera pas concluante si une fluctuation conjoncturelle est dans l'intervalle venue fausser l'effet de la mesure. Toutefois, si les régions avec et sans salaire minimum sont soumises en moyenne à la même évolution conjoncturelle, l'impact de celle-ci peut être précisé grâce à la différence « avant-après » dans les régions sans salaire minimum. L'effet de la mesure apparaît alors comme la différence entre la différence

Comment mesure-t-on l'efficacité d'un cours de langue pour les personnes en recherche d'emploi ?



KEystone

Les institutions et les entreprises embauchent de plus en plus d'économistes spécialisés dans l'approche quantitative pour optimiser leurs processus.

« avant-après » dans les régions avec un salaire minimum (effet de la mesure + conjoncture) et la différence « avant-après » dans les régions sans salaire minimum (conjoncture).

L'approche de la « sélection en fonction des caractéristiques observées » suppose également que toutes les caractéristiques qui influencent à la fois la mesure et le résultat sont prises en compte dans les données. La mesure s'applique dès lors de façon « quasi expérimentale » parmi les individus dont les traits observés sont comparables, et ne dépend pas de traits non observés qui influencent également le résultat. L'effet des mesures peut ainsi être identifié en formant des groupes avec et sans mesure dont les particularités observées sont comparables.

Davantage de souplesse grâce aux nouvelles méthodes

Ces dernières décennies, l'évaluation des politiques publiques a connu un développement dynamique au niveau de l'approche conceptuelle de la mesure de causalité, mais également des méthodes statistiques⁴. Les méthodes plus traditionnelles supposent que la relation entre le résultat et les caractéristiques observables et/ou la mesure corresponde à une certaine fonction, par exemple linéaire, où une année d'expérience supplémentaire sur le marché du travail aura toujours le même effet sur le salaire, indépendamment du niveau d'expérience acquise. Mais les méthodes plus récentes d'évaluation des politiques, comme celle de l'appariement (« matching »), rendent ces hypothèses superflues.

La méthode de l'appariement identifie dans les données disponibles des paires de personnes avec et sans mesure, dont les caractéristiques observables sont les plus semblables possibles. On obtient ainsi des groupes avec et sans mesure aux particularités comparables. L'appariement a par exemple été utilisé dès 2002 pour évaluer les programmes mis en place sur le marché du travail en Suisse⁵. Il est apparu que les subventions

salariales temporaires avaient des effets positifs sur l'emploi.

Une autre méthode consiste à pondérer les individus en fonction de leur probabilité d'être l'objet de la mesure, afin de rendre à nouveau comparables les groupes pondérés avec et sans mesure au regard des caractéristiques observées. La pondération peut également se combiner à une régression (apparentée à l'appariement). Puisqu'il suffit qu'au moins l'une des deux méthodes fonctionne correctement, cette approche est appelée « estimation doublement robuste ». Dans une étude menée en Suisse parue en 2016, cette méthode a par exemple permis de mettre en évidence une corrélation entre les ventes directes de médicaments par les médecins et l'augmentation des prix des médicaments⁶.

L'apprentissage automatique progresse

La masse croissante de données numériques disponibles a amené une autre innovation au service de l'évaluation des politiques : l'apprentissage automatique causal (AAC)⁷. Sur la base de ces données, les algorithmes d'apprentissage automatique apprennent à prédire le mieux possible une mesure ou un résultat en fonction des caractéristiques observées. Cette méthode est particulièrement utile lorsque les caractéristiques sont si nombreuses qu'elles échappent aux capacités d'analyse d'une personne. Lors d'une sélection basée sur les caractéristiques observées, on recourt donc à l'apprentissage automatique causal pour choisir les principales données pertinentes afin de rendre comparables les groupes avec et sans intervention. Cette méthode permet également de définir, à partir des données disponibles, des sous-groupes sur lesquels l'effet d'une mesure est particulièrement important ou faible par rapport aux autres sous-groupes, ce qui est précieux pour repérer des hétérogénéités dans les effets des politiques et adapter la mesure le cas échéant. Une analyse AAC des programmes de recherche d'emploi en Suisse a par exemple montré que l'effet sur l'emploi est plus important parmi les chômeurs sans grande expérience et peu formés que parmi les personnes ayant de meilleurs atouts sur le

4 Voir Imbens et Wooldridge (2009) pour un aperçu.

5 Gerfin et Lechner (2002).

6 Kaiser et Schmid (2016).

7 Chernozhukov et al. (2018).

marché du travail⁸, ces dernières se trouvant même défavorisées par la mesure.

En outre, cette méthode permet de définir l'assignation optimale d'une mesure disponible (limitée) afin de maximiser son impact⁹. Par exemple, un programme mis en place sur le marché du travail qui est organisé de manière optimale maximise le retour à l'emploi des chômeurs. Un apprentissage de la politique optimale peut également s'appliquer au secteur privé, notamment sous la forme de rabais ciblés offerts par les détaillants à certains consommateurs afin de maximiser leur chiffre d'affaires. Il n'est donc pas surprenant que les institutions publiques et internationales, tout comme les entreprises technologiques, embauchent de plus en plus d'économistes spécialisés dans l'approche quantitative pour optimiser leurs processus. Cette évolution va probablement gagner d'autres domaines de l'économie. Les

hautes écoles réagissent d'ailleurs également au nouveau profil d'exigences d'une formation économique associée aux méthodes quantitatives modernes: l'université de Saint-Gall a récemment restructuré son programme de master en économie quantitative et en finance, tandis que l'université de Fribourg proposera dès l'automne 2020 un master en économie et analyse des données.

8 Knaus et al. (2018).

9 Athey et Wager (2018).



Martin Huber

Professeur d'économétrie appliquée et d'évaluation des politiques publiques, Département d'économie politique, université de Fribourg

Bibliographie

- Athey S. et Wager S. (2018). *Efficient policy learning*, document de travail, université de Stanford.
- Berger M. et Lanz B. (2019). *Adjusting to minimum wage regulation: Evidence from a direct-democracy experiment in Switzerland*, document de travail, université de Neuchâtel.
- Chernozhukov V., Chetverikov D., Demirer M., Duflo E., Hansen C., Newey W. et Robins J. (2018). « Double/debiased machine learning for treatment and structural parameters ». *Econometrics Journal*, 21: C1–C68, février.
- Gerfin M. et Lechner M. (2002). « A microeconomic evaluation of the active labour market policy in Switzerland ». *The Economic Journal*, 112: 854–893.
- Imbens G. W. et Wooldridge J. M. (2009). « Recent developments in the econometrics of program evaluation ». *Journal of Economic Literature*, 47: 5–86.
- Kaiser B. et Schmid C. (2016). « Does physician dispensing increase drug expenditures? Empirical evidence from Switzerland ». *Health Economics*, 25: 71–90.
- Knaus M., Lechner M. et Strittmatter A. (2018). *Heterogeneous employment effects of job search programmes: A machine learning approach*, document de travail, université de Saint-Gall.
- Lalive R. (2008). « How do extended benefits affect unemployment duration? A regression discontinuity approach ». *Journal of Econometrics*, 142: 785–806.
- Schochet P. Z., Burghardt J. et McConnell S. (2008). « Does job corps work? Impact findings from the national Job Corps study ». *American Economic Review*, 98: 1864–1886.