



# How Impact Evaluation Is Shaping the Design of Labour Market Policies

*Verónica Escudero*

Support for evidence-based policy making and the microeconomic evaluation methods necessary to uncover causal effects has grown during the last two decades. Today, there is a growing appetite for credible and transparent evidence on whether a policy intervention achieves its expected outcomes. This concern is even more pressing as governments are increasingly held accountable for their decisions, and as the resources available for the implementation of policies are continuously scrutinized.

There has been an intense debate in Economics as to what methods are appropriate to ascertain causal effects and how much evidence is needed to assess credibly the effects of policies. The credibility crisis of the

---

I am grateful to Hilary Hoynes, Miguel Ángel Malo, Domenico Tabasso and the editors of this book for their comments.

---

V. Escudero (✉)

Research Department, International Labour Organization, Geneva, Switzerland

University of California Berkeley, CEGA, Berkeley, CA, USA

e-mail: [escudero@ilo.org](mailto:escudero@ilo.org)

© The Author(s), under exclusive license to Springer Nature  
Switzerland AG 2022

P. Goulart et al. (eds.), *Global Labour in Distress, Volume II*,  
Palgrave Readers in Economics,

[https://doi.org/10.1007/978-3-030-89265-4\\_26](https://doi.org/10.1007/978-3-030-89265-4_26)

1980s–1990s (Angrist & Pischke, 2010) pushed empirical economics in a spiral of ever-increasing rigour, settling the question of what econometric evaluation methods achieve better identification (Angrist & Pischke, 2014). This transformation also set the course for the upsurge in sophistication and diversity of the last decade (Abadie & Cattaneo, 2018).

However, no agreement has yet been reached (particularly among policy makers and practitioners), as to whether high standards of rigour in impact evaluation are needed, or appropriate, in all situations (Clemens & Demombynes, 2011). This is in part why the support for evidence-based policy making is not equally widespread in policy settings as in academic circles. Indeed, while the number of impact evaluation studies has rocketed during the last decade, a great number of labour market policies not based on evidence is still implemented today. How can impact evaluation be used further in the future to exploit its full potential?

This short article discusses this question. First, it reviews the progress achieved thus far. Then, it examines the obstacles that the impact evaluation profession needs to overcome to achieve an even wider use of evaluation techniques that continues shaping the design of labour market policies.

## 1 PROGRESS MADE IN LABOUR ECONOMICS TOWARDS THE USE OF CAUSAL INFERENCE FOR POLICY ANALYSIS

Labour economics has been traditionally a space for methodological innovation. The influential *Handbook of Labor Economics* (Ashenfelter et al., 1986, 1999, 2011) explains that since the 1970s, many of the innovations in econometric and statistical methods were developed with labour applications in mind (Angrist & Krueger, 1999; Moffitt, 1999; List & Rasul, 2011). Examples of this include sample selection models, non-parametric methods for censored data and survival analysis, quantile regression and panel data models, and so on.

Causal inference, the type of empirical research that “seeks to determine the effects of particular interventions or policies, or to estimate features of the behavioral relationships suggested by economic theory” (Angrist & Krueger, 1999, p. 1280) has also been at the heart of research in labour economics for various decades. In fact, the renewed interest in identification problems related to instrumental variables estimators and quasi-experimental methods of the 1970s took place in labour economics

(Angrist & Krueger, 1999). Already from the beginning of the 1990s, there was an explosion in the number of economics articles using quasi-experimental methods, including those based on fixed effects, matching methods, difference-in-differences, regression discontinuities, instrumental variables, and natural experiments—from 27 in the 1980s to close to 250 in the 1990s, peaking at over 660 in the 2000s and staying around that level in the following decade.<sup>1</sup> By 2000, these methods were considered part of the mainstream empirical research toolbox to measure causal effects.

Field experiments (i.e., which use randomization to define treatment and control groups) were added last to the labour economist toolkit. But some of the first analyses using experimental design in economics were also implemented to answer labour-related questions, already a century ago.<sup>2</sup> Today, although the use of experimental design is less common in economics than in other fields (such as medical research), and it might also be less common in labour economics than in other economics fields (such as development economics), its application has grown. Indeed, this method carries advantages, such as the possibility of using economic theory to craft the research hypotheses, engineering exogenous variations in local labour market settings, as well as using primary data which is key when no other data is available (Angrist & Krueger, 1999; List & Rasul, 2011). As these insights have gained support, field experiments have become more common in labour economics.

Reaching an understanding of the methods capable of telling apart true causal relationships from correlations has not been a smooth process. Empirical economics underwent a serious credibility crisis in the 1980s and 1990s, due to the lack of attention given in research design to identification of the causal effect of interest and robustness to changing assumptions (Angrist & Pischke, 2010; Stock, 2010). Edward Leamer observed in 1983, as he reflected on the state of the empirical economics

<sup>1</sup> Own calculations based on the data presented in Panhans and Singleton (2017), which was kindly provided by the authors. Data includes articles published in 11 all-field journals plus the top four economic journals.

<sup>2</sup> List and Rasul (2011) trace the first use of field experiments in labour economics to two historical examples: first, the Hawthorne plant experiment in the 1920s, that varied the amount of light for different groups in the workplace to assess the effect on productivity of female assemblers. Second, the large-scale social experiment carried out by the US government (and led by Heather Ross) starting in 1968, which explored the behavioural effects of negative income taxation.

profession, “Hardly anyone takes data analysis seriously. Or perhaps more accurately, hardly anyone takes anyone else’s data analysis seriously” (Leamer, 1983, p. 37). This prompted a push in the 1990s towards an increase in rigour, based on a greater emphasis on identification in econometric models (Abadie & Cattaneo, 2018). Improvements in empirical work were facilitated by more and better data and advances in computational power and estimation methods, but the driving force was an impetus for more robust research design.<sup>3</sup> The change in the nature of empirical economics was so profound that some scholars called this time the “credibility revolution” (Angrist & Pischke, 2010, p. 4). It increased not only the rigour of research but also its scientific impact and policy relevance.

As a result, the past two decades have seen an explosion in the number of impact evaluation studies, using experimental or quasi-experimental methods. Counting them all would be burdensome, but to give an idea of the progress achieved to date, we can look at the increase in impact evaluations of active labour market policies (ALMPs) included in the systematic reviews. In their seminal review two decades ago, Heckman et al. (1999) summarized approximately 75 microeconomic evaluation studies of ALMPs from advanced countries. In a more recent review, Kluge (2010) included nearly 100 separate studies from Europe alone, while Vooren et al. (2019) reviewed 57 experimental and quasi-experimental studies in only 12 advanced countries. In addition, Escudero et al. (2019) compiled and assessed 51 programme evaluations in Latin America and the Caribbean. On a geographically larger scale, the seminal review by Card et al. (2018) included 200 separate studies of ALMPs around the world. There are also a number of impact analyses of labour market programmes targeted to specific groups. Greenberg et al. (2003) surveyed 31 evaluations of government-funded programmes for the disadvantaged in the US. Meanwhile, Kluge et al. (2019) compiled 107 separate interventions that primarily targeted youth.

In addition to these studies that directly apply experimental and quasi-experimental methods, a great deal of research during the last decade was devoted to refining and expanding these methods, as well as developing solutions to address their constraints—for example, synthetic controls, variable selection methods such as machine learning methods and LASSO

<sup>3</sup>Landmark studies underlying this transformation include: Ashenfelter and Card (1985), Solon (1985), LaLonde (1986), Ashenfelter (1987), Angrist (1990), Angrist and Krueger (1991), Gruber (1994), Meyer (1995).

methods, design of high-dimensional experiments (Athey & Imbens, 2017; Cattaneo et al., 2018; Fougère & Jacquemet, 2020). It is safe to say that econometric evaluation methods have gotten more sophisticated and diverse with time.

## 2 THE FUTURE OF IMPACT EVALUATION FOR LABOUR MARKET POLICY ANALYSIS: EXPLOITING ITS FULL POTENTIAL

Data and methodological innovations have driven progress in the field of impact evaluation, but progress has also been facilitated by a mounting commitment to evaluation by governments and other institutions in many countries. Despite this progress, many labour market policies are implemented today, without regard to the available evidence on the effectiveness of these policies. The large support that exists today among academics for the use of impact evaluation methods has, therefore, not permeated equally into the policy making arena.

On reflection, this is not surprising. Impact evaluation is hard to implement—data and techniques are not accessible to everyone (often promoting scepticism about their validity or appropriateness), and their implementation is time-consuming and often costly. More importantly, it is not always clear to policy makers how to use the results of impact evaluations, making their benefits less evident.

With hindsight, we know the benefits of impact evaluation are extensive. First, impact evaluation increases the rigour of the findings generated—we can produce more credible causal evidence, but also understand better its implications. Second, impact evaluation has a disciplinary effect on policy makers, developing agencies and policy practitioners, as it increases transparency and ensures that scarce resources are not lost on ineffective programmes that look attractive on paper (Clemens & Demombynes, 2011). Third, impact evaluation offers a special opportunity to test innovations before adapting labour policy at a bigger scale.

However, there remain serious challenges that the social sciences professions need to overcome to leverage further the advantages of evidence-based policy. First, there are those who resist the implementation of impact evaluations on the basis of ethical or political concerns. Ethical concerns include questions such as: who benefits or not from an intervention, how to address potential negative unintended consequences or the methods

used to study subjects (Gertler et al., 2016). Taking into account ethical considerations in the implementation of an evaluation is indeed essential and should be an integral part of the evaluation plan. This is, however, different from questioning whether, in and of itself, impact evaluation is ethical, which is another point sometimes raised by detractors. This is also linked to the political concerns often raised, including the need to maintain positive narratives about programmes, because modifying or closing a popular policy or programme may cause social unrest or change the course of an election. I believe the useful starting point for this debate is to consider the ethics of implementing programmes (or continuing them), investing large amounts of public resources, without considering their effectiveness. It is the lack of evaluation that would be unethical in this context.

Second, despite the progress made in improving the rigour of impact evaluation methods and establishing standards for their appropriate use, there are many studies today that fail to abide by these norms. Investigator, publication, and political biases continue to taint the credibility of results today (Miguel, 2021). This is why a relatively new scholarly movement has emerged to advance the agenda of transparency and reproducibility of research findings (Christensen & Miguel, 2018; Christensen et al., 2019; Hoces de la Guardia et al., 2021; Miguel, 2021). This movement seeks to open the data and research practices to the wider community, so research objectives and strategies, as well as their findings, can be inspected, understood, and replicated. This would ensure the precision of estimated effects, contributing also to the credibility and applicability of impact evaluation (Clemens, 2017).

Third, with the increase in impact evaluations, policy makers around the world have a tremendous amount of evidence about “what works”. How to reconcile the various findings, especially since, as we know, this evidence is context-specific? The move towards openness can be a first key. It will make transparent how estimates are produced, how precise they are, and what are assumptions withholding the stability of results (Hoces de la Guardia et al., 2021). Moreover, improving the precision of estimates and making data and methodologies available to other researchers, would allow a broader production of research studies that aim to reconcile findings across individual studies. This is the case of meta-analyses and systematic reviews that use impact estimates from individual impact evaluations to determine overall trends and test the consistency of treatment effects across studies. Meanwhile, cost-effectiveness analyses are an important complementary tool to impact evaluations, in order to compare between programme alternatives (Gertler et al., 2016). Meta-analysis and

systematic reviews can improve the applicability of research findings; while cost-effectiveness analyses can help policy makers and practitioners differentiate among policies evaluated.

Finally, an effort to broaden the use of evidence-based policy will not be complete without a coordinated effort to foster closer collaborations between policy makers and policy practitioners, and researchers. Efforts should aim to agree on a research design that is robust but also applicable on the ground, collect the necessary data, and discuss the appropriate solutions to adjust the labour policies evaluated on the basis of the evaluation results. These closer collaborations would also ensure that the questions asked by impact evaluations are directly relevant to the issues that matter to policy makers and practitioners.

## REFERENCES

- Abadie, A., & Cattaneo, M. D. (2018). Econometric Methods for Program Evaluation. *Annual Review of Economics*, 10(1), 465–503. <https://doi.org/10.1146/annurev-economics-080217-053402>
- Angrist, J. D. (1990). Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. *The American Economic Review*, 80(3), 313–336.
- Angrist, J. D., & Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics*, 106(4), 979–1014. <https://doi.org/10.2307/2937954>
- Angrist, J. D., & Krueger, A. B. (1999). Chapter 23—Empirical Strategies in Labor Economics. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (pp. 1277–1366). Elsevier. [https://doi.org/10.1016/S1573-4463\(99\)03004-7](https://doi.org/10.1016/S1573-4463(99)03004-7)
- Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, 24(2), 3–30. <https://doi.org/10.1257/jep.24.2.3>
- Angrist, J. D., & Pischke, J.-S. (2014). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.
- Ashenfelter, O. (1987). The Case for Evaluating Training Programs with Randomized Trials. *Economics of Education Review*, 6(4), 333–338. [https://doi.org/10.1016/0272-7757\(87\)90016-1](https://doi.org/10.1016/0272-7757(87)90016-1)
- Ashenfelter, O., & Card, D. (1985). Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. *The Review of Economics and Statistics*, 67(4), 648–660. <https://doi.org/10.2307/1924810>
- Ashenfelter, O., Layard, R., & Card, D. E. (1986). *Handbook of Labor Economics* (Vols. 1–2). Elsevier Science Pub. Co.

- Ashenfelter, O., Layard, R., & Card, D. E. (1999). *Handbook of Labor Economics* (Vol. 3). Elsevier Science Pub. Co.
- Ashenfelter, O., Layard, R., & Card, D. E. (2011). *Handbook of Labor Economics* (Vol. 4). Elsevier Science Pub. Co.
- Athey, S., & Imbens, G. W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives*, 31(2), 3–32. <https://doi.org/10.1257/jep.31.2.3>
- Card, D., Kluge, J., & Weber, A. (2018). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association*, 16(3), 894–931.
- Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2018). *A Practical Introduction to Regression Discontinuity Designs II* (Vol. II, p. 106). Cambridge University Press.
- Christensen, G., Freese, J., & Miguel, E. (2019). *Transparent and Reproducible Social Science Research: How to Do Open Science*. University of California Press.
- Christensen, G., & Miguel, E. (2018). Transparency, Reproducibility, and the Credibility of Economics Research. *Journal of Economic Literature*, 56(3), 920–980. <https://doi.org/10.1257/jel.20171350>
- Clemens, M. A. (2017). The Meaning of Failed Replications: A Review and Proposal. *Journal of Economic Surveys*, 31(1), 326–342. <https://doi.org/10.1111/joes.12139>
- Clemens, M. A., & Demombynes, G. (2011). When does rigorous impact evaluation make a difference? The case of the Millennium Villages. *Journal of Development Effectiveness*, 3(3), 305–339. <https://doi.org/10.1080/19439342.2011.587017>
- Escudero, V., Kluge, J., Mourelloand, E. L., & Pignatti, C. (2019). Active Labour Market Programmes in Latin America and the Caribbean: Evidence from a Meta Analysis. *The Journal of Development Studies*, 55(12), 2644–2661.
- Fougère, D., & Jacquemet, N. (2020). *Policy Evaluation Using Causal Inference Methods*. 12922. IZA—Institute of Labor Economics. Retrieved July 21, 2021, from <https://www.iza.org/publications/dp/12922/policy-evaluation-using-causal-inference-methods>
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2016). *Impact Evaluation in Practice* (2nd ed.). International Bank for Reconstruction and Development/The World Bank. <https://openknowledge-worldbank-org.ilo.idm.oclc.org/handle/10986/25030>
- Greenberg, D. H., Michalopoulos, C., & Robins, P. K. (2003). A Meta-Analysis of Government-Sponsored Training Programs. *ILR Review*, 57(1), 31–53.
- Gruber, J. (1994). The Incidence of Mandated Maternity Benefits. *The American Economic Review*, 84(3), 622–641.
- Heckman, J. J., Lalonde, R. J., & Smith, J. A. (1999). The Economics and Econometrics of Active Labor Market Programs. In O. Ashenfelter, & D. Card (eds.), *Handbook of Labor Economics* (pp. 1865–2097). Retrieved February 18, 2018, from <https://experts.umich.edu/en/publications/chapter-31-the-economics-and-econometrics-of-active-labor-market>



- Hoces de la Guardia, F., Grant, S., & Miguel, E. (2021). A Framework for Open Policy Analysis. *Science and Public Policy*, 48(2), 154–163. <https://doi.org/10.1093/scipol/scaa067>
- Kluge, J. (2010). The Effectiveness of European Active Labor Market Programs. *Labour Economics*, 17(6), 904–918. <https://doi.org/10.1016/j.labeco.2010.02.004>
- Kluge, J., Puerto, S., Robalino, D., Romero, J. M., Rother, F., Stöterau, J., Weidenkaff, F., & Witte, M. (2019). Do Youth Employment Programs Improve Labor Market Outcomes? A Quantitative Review. *World Development*, 114, 237–253.
- LaLonde, R. J. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *The American Economic Review*, 76(4), 604–620.
- Leamer, E. E. (1983). Let's Take the Con Out of Econometrics. *The American Economic Review*, 73(1), 31–43.
- List, J. A., & Rasul, I. (2011). Chapter 2—Field Experiments in Labor Economics. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (pp. 103–228). Elsevier. [https://doi.org/10.1016/S0169-7218\(11\)00408-4](https://doi.org/10.1016/S0169-7218(11)00408-4)
- Meyer, B. D. (1995). Natural and Quasi-Experiments in Economics. *Journal of Business & Economic Statistics*, 13(2), 151–161. <https://doi.org/10.2307/1392369>
- Miguel, E. (2021). Evidence on Research Transparency in Economics. *Journal of Economic Perspectives*, 35(3), 193–214. <https://doi.org/10.1257/jep.35.3.193>
- Moffitt, R. A. (1999). Chapter 24—New Developments in Econometric Methods for Labor Market Analysis. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (pp. 1367–1397). Elsevier. [https://doi.org/10.1016/S1573-4463\(99\)03005-9](https://doi.org/10.1016/S1573-4463(99)03005-9)
- Panhans, M. T., & Singleton, J. D. (2017). The Empirical Economist's Toolkit: From Models to Methods. *History of Political Economy*, 49(Supplement), 127–157. <https://doi.org/10.1215/00182702-4166299>
- Solon, G. (1985). Work Incentive Effects of Taxing Unemployment Benefits. *Econometrica*, 53(2), 295–306. <https://doi.org/10.2307/1911237>
- Stock, J. H. (2010). The Other Transformation in Econometric Practice: Robust Tools for Inference. *Journal of Economic Perspectives*, 24(2), 83–94. <https://doi.org/10.1257/jep.24.2.83>
- Vooren, M., Haelermans, C., Groot, W., & van den Brink, H. M. (2019). The Effectiveness of Active Labor Market Policies: A Meta-Analysis. *Journal of Economic Surveys*, 33(1), 125–149. <https://doi.org/10.1111/joes.12269>