

University of Groningen

Impact evaluations, bias, and bias reduction

Eriksen, Steffen

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Eriksen, S. (2018). *Impact evaluations, bias, and bias reduction: Non-experimental methods, and their identification strategies*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen, SOM research school.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 6

Conclusion

6.1 General discussion

‘We do not want to restrict our learning about what works to interventions in which an RCT proves possible.’ This statement, written by Howard White (2014), stands as a reminder that just because an RCT is not feasible in a given setting, it does not follow that a proper impact evaluation cannot be done. In an era where RCTs hold the status as the gold standard for drawing causal inference, questions still persist about whether non-experimental designs can and should be used to answer attribution questions. The central question being if non-experimental designs can adequately address selection bias. The findings of this thesis addresses the need for continued interest and improvement of non-experimental designs, arguing that, if designed appropriately, non-experimental designs can control for selection bias and thus provide proper impact estimates.

In 2016, at the “State of Economics, State of the World” conference, Esther Duflo gave a talk, where she discussed the influence of RCTs as a tool for policy evaluation, pointing out ‘that there has been a very rapid increase in the number of RCTs in development since mid-1990s’ and that ‘RCTs have increased the standards of non-experimental work’ (Duflo et al., 2016). That is to say, the rapid increase in the number of RCTs, as documented by Cameron et al. (2016), has pushed researchers to develop better and more reliable non-experimental designs, as more attention has been paid to the potential biases that arises when the chosen methodology is non-experimental. Thus, although RCTs hold a clear identification advantage, it is fine to apply alternative designs, as long as one understands their shortcomings.

Not only has there been a rapid increase in RCTs in development, but there has been a general growth in the number of impact evaluations. RCTs only account for a fraction of the total number of impact studies, reflecting the continued importance of research into non-experimental designs. Despite this considerable growth, the challenge is still to produce more studies. ‘Global policy should not be based on a single study from a single country, but on a large number of studies confirming whether an intervention works or not, and how that impact varies according to context’ (White, 2014). A 1000 impact evaluations might sound like a lot, however, when you look at the bigger picture, and realize that there are more than 100 developing countries, each with 20 sectors or more, then a 1000 impact

CONCLUSION

evaluations are suddenly not a lot (Savedoff, 2013). It is therefore important to continue to produce impact evaluations. Not only applying RCTs, but also study the application of non-experimental techniques to further understand the impact of different interventions.

The non-experimental techniques discussed in this thesis far from represents the complete set of available non-experimental techniques. Rather, they represent a selected subset of techniques applicable for the different evaluation scenarios encountered throughout the chapters of this thesis. They add to the growing amount of impact evaluations, showing that a (convincing) causal relationship can be established in the absence of a randomized design. Furthermore, they emphasize the need for the continued development of non-experimental designs as an alternative to randomized designs.

The sequence of the chapters in this thesis shows, in a step by step fashion, the different challenges associated with the introduction of bias, and the possible remedies. Starting at the macro level, each step gradually zooms in to the individual level. The first step considers the challenge of evaluating the impact of macroeconomic policies (Chapter 2), finding that healthcare financing reforms curb total healthcare expenditures. The scope of the following two chapters is then set for data at the household level, considering the impact of different micro credit programs, with a common factor, that both credit markets are not absent ex-ante, and randomization is therefore not possible. The first of the studies (chapter 3), applies the expansion plans of the MFI to develop an identification strategy to assess the impact, finding an overall limited impact for the program. However, finding substantial differences between regions. The second study (chapter 4), identifies the causal impact of a microfinance project using a retrospective non-experimental evaluation design, finding small impacts of the program. The chapter then investigates how the loan proceedings of the households are spend in an attempt to explain the limited impacts, revealing that loan proceedings were not spent productively, as otherwise indicated by the beneficiaries. The last chapter zooms in to the individual level, finding that social desirable and opportunistic behaviour affect the revealed support for Farmers' Market Organizations. No matter the setting, and possible cause of bias, these findings emphasize the need for a continued interest in non-experimental techniques, and thus not restrict our learning to randomized settings. The next sections review the main lessons we learned from each of the chapters, discussing the resulting policy implications, and their implications for future research.

6.2 Evaluating macroeconomic interventions

In chapter 2, we analyse whether healthcare financing privatisations curb total healthcare expenditures. Applying a propensity score matching methodology, we find that healthcare financing privatisations lead to cost savings in total healthcare expenditures in our sample of OECD countries. The results suggest an annual cost saving of 0.09 percentage points of GDP per year. Accumulated this means that about 0.45 percentage points of GDP are saved over 5 years following a privatisation. The results also show that savings in total HCE are large in the beginning of the post reform period, but decrease continually approaching a zero effect after five years. That is, the yearly effect decreases as a function of time after the reforms. The results presented in chapter 2 are robust to various sensitivity tests, which leads to the conclusion that healthcare financing privatisations seems like a valid approach to reduce government expenditures. However, the results presented here must be taken with caution as outlined below.

The main policy implication from chapter 2 is that healthcare financing privatisations curb aggregate healthcare expenditures in advanced economies. Such gradual shifts from public to private financing can lead to lower aggregate healthcare consumption. However, we are not able to draw any conclusions regarding efficiency. The cost savings that we observe might come from negative effects of the privatisations. Low-income groups are more likely to consume less than the optimal level of healthcare due to budget constraints. Due to data restrictions, we cannot at this point assess whether the privatisations in question cause health equality and/or quality to decrease. An increase in private payments may make the consumers choose lower quality of healthcare. We are unable to assess whether the reform we analyse have an impact on the overall quality of healthcare.

The results given in chapter 2 should not be interpreted as if healthcare financing privatisations will deliver a positive outcome in general equilibrium, but only that these privatisations leads to cost savings in total healthcare expenditures. Part of the population might be excluded from healthcare, while another group will choose for a lower quality of healthcare, thus decreasing the overall health quality of the population. It is therefore not settled if healthcare privatisations are optimal from a welfare perspective. This is clearly a topic for future research.

6.3 Expansion plans and impact

In chapter 3, we zoom into the household level, where we evaluate the impact of microfinance loans from a Bolivian microfinance institute in two different regions within Bolivia. To isolate the causal

CONCLUSION

effect of the microcredit loans, our identification strategy apply the potential expansion plans of the MFI. The expansion plans enable us to follow and extend the method of Coleman (1999), by applying a difference-in-difference in space, where we also forecast the composition of clients and non-clients in the planned expansion area as well as the area where the MFI is currently active. The findings in chapter 3 give us a lesson in how much results can differ depending on the region under consideration. For the Yungas region, little to no impact is found, whereas the story is much different for the Chuquisaca region, where significant impacts are found on multiple outcome indicators. The impacts in the Chuquisaca region are both negative as well as positive. The negative impacts are for agricultural outcomes, and the positive impacts are observed for the business outcomes. This combined with an observed shift in the income distribution could signal that the loans provided by the MFI under study, to some extent, finance a shift in the income generating activities of households from agriculture to business.

The methodology in chapter 3 provides the policymaker with a tool to evaluate the impact of a microcredit program relatively cheaply and fast, relying only on one round of data collection. However, it relies on that the MFI under study has plans to undergo an expansion in the near future, and have such expansion plans available.

Future research would benefit from additional analyse on the substitution effects and complementary effects of a household's loans, as well as the intra household substitution that might take place when a member of the household takes on a new loan. The microcredit market in Bolivia is one of the densest markets in the world, and therefore the likelihood that households already have one or more loans from different institutes is very high. If on one hand, a relatively large substitution effect is present between a new microfinance loan and the households' pre-existing loans, little to no significant impacts can be expected of a subsequent impact evaluation. On the other hand, if the additional microfinance loan act as a complement to the pre-existing loans, it gives reason to expect significant impacts of a subsequent impact evaluation. The findings in chapter 3 differ significantly between the two regions, and the potential difference in substitution and complementary effects of the additional loan, could explain this difference, as the households in the Yungas regions have significant larger amount of pre-existing loans compared to the households in the Chuquisaca region. Additionally, future research applying this methodology would benefit from a second survey round, to be able to study how the composition of future borrowers and non-borrowers turned out in the expansion area.

6.4 Impact of ongoing projects

In Chapter 4, we consider a retrospective, non-experimental, impact evaluation of a microcredit program implemented by Ghanaian microfinance institute (MFI). In addition to providing new results about the impact of microcredit, the study in chapter 4 suggests a non-experimental evaluation strategy when an RCT is not possible. The methodology consists of a propensity score matching approach combined with a double difference methodology, where pre-intervention data is constructed using recall. By calculating the propensity scores and constructing a common space, microfinance clients and non-clients that exhibit substantial differences in their observable characteristics are excluded from the analysis. Even after ignoring households outside this common space, the balancing tests show persistent differences between the control and treatment households. These observed differences signals the need for further controlling for selection bias, and hence the importance of applying a double difference methodology. The findings are in line with recent experimental research investigating the impact of microcredit. That is, we observe minor impacts of the provision of microcredit, at least in the short run. In an attempt to explain the results of the impact evaluation, a list experiment is implemented to investigate the effect of social desirable behaviour on the reported loan use of the household's microfinance loan. An important aspect of the underlying theory of change is that loan proceedings are mainly spent productively. Similar to Karlan and Zinman (2012), we find that almost 50 per cent of the microfinance members spent their loan proceedings on consumer items rather than being spent productively. The results here clearly reveal a discrepancy between the respondent's actions and their survey response, and thus provide a possible explanation for the results of the impact evaluation.

From the point of view of policymakers, the findings in chapter 4 regarding the impact of microcredit suggest a less costly and faster approach, when it comes to evaluating the impact of their programs. The outcome of the list experiment clearly indicates the challenge associated with obtaining information about sensitive questions through direct surveys. This suggests that future surveys should be equipped to deal with questions regarding sensitive issues. The main strength but also the limitation of chapter 4 is with the use of recall. The crucial importance of baseline data, which is needed to employ a double difference methodology, indicates a clear benefit of using recall. Yet, if the treatment group is more precise in recalling than the control group, the recall method may lead to bias of the results. Additional research is needed to determine whether the potential bias due to recall outweigh the clear benefits. The extent to which the outcome of the list experiment might also invalid the answers to other questions is unclear. However, we expect little to no bias, because most of the

CONCLUSION

other questions cannot be regarded as sensitive. Yet, more research about potential bias related to direct questioning appears important.

6.5 Unrevealing true preferences

In chapter 5, we investigate the farmers' perceptions in rural Ethiopia regarding the Farmers' Market Organizations. Reliable information about farmers' perceptions is important as it may help donors and governments to focus their activities towards vital initiatives. However, if target groups are simply asked whether they 'like it or not', it is appealing for many farmers to simply confirm, leading to the false impression that they will actually use these services. This chapter provides two main contributions. First, we test the influence of social desirability and opportunism on the answers by the farmers to direct questions regarding the importance of the Farmers' Market Organizations operating in the area. Using a list experiment, we support our argument that farmers may feel social and/or opportunistic pressure to express positive opinions concerning the Farmers' Market Organization. The revealed actions are thus not in line with the opinion they express when asked directly. The opinions derived from a direct question therefore give us a biased upper bound of support for the activities of the Farmers' Market Organization, while the revealed actions of the farmers provide a bias lower bound. Second, the list experiment method is extended, enabling us to distinguish various characteristics. Specifically, we investigate differences in the response bias between members versus non-members, and supporters (believers) versus non-believers.

We learned that social desirability and opportunistic behaviour, bias the responses given to questions regarding the strength of the Farmers' Market Organization in the area. In particular, those respondents who express support for the organizations when asked directly tend to lie more often. The results in chapter 5 help us to understand why further growth of Farmers' Market Organizations in Africa is lagging behind expectations. The importance of this topic lies in the fact that Farmers' Market Organizations and their donors are misled if support for the Farmers' Market Organization is based on direct questions. Advocates of Farmers' Market Organizations interpret the favourable opinions as support and defend their initiatives to establish Farmers' Market Organizations, while opponents refer to the limited use of the services. This ambiguity hampers managers of the organizations, policy-makers, and donors who try to base their actions on realistic prognoses of existing support by targeted members. If donor/government investments are not matched by local fees or commitments, it is relatively easy for interviewees to provide an opportunistic answer or to give in to social pressure and provide the desired answer. Thus, the results presented in chapter 5

imply that managers of the Farmers' Market Organizations, policy makers and donors have to take into account the fact that perceptions regarding local support for their initiatives are not easily measured.

Future research into this area would benefit from taking the next step and introduce a significant fee and/or commitments to be able to disentangle the true perceptions of the farmers, and thus be able to close the gap that famers' market organizations have to similar organizations other places around the world.

6.6 Final remarks

This thesis set out to investigate bias and bias reduction, when applying non-experimental designs to study the impact of a certain policy or intervention. Starting at the macro level, and step by step zooming in to the individual level, different scenarios were studied, highlighting the need for a whole toolkit of different identification strategies. Despite the rapid growth in the number of impact evaluations, there is still a lot to be learned about what works and what does not. This thesis has helped answering these questions, by adding to the number of impact evaluations, exploring interesting scenarios. One question that jumps to mind after reading this thesis is, "if randomization is not possible. What is the second best alternative?" While some researchers might agree, I still believe that there is no clear answer, as it depends on the given scenario. I would instead argue that each intervention needs an assessment of what techniques are available, and then select the most appropriate one. I hope to see researchers continue to study non-experimental designs as a fundamental alternative to randomized designs. Future research might then bring us to a point where there is a clear answer to what the second best alternative is.

'People say nothing is impossible, but I do nothing every day.' (Milne, 1926).

