

Credible Methods for Evaluating the Impacts of Amil Programs

Muhamad Said Fathurrohman

Universitas Airlangga (Unair)

Paper to be presented at International Conference of Zakat 2018
15-16 November, MM UGM Hotel, Yogyakarta, Indonesia

ABSTRACT

Existing studies on the impacts of amil programs are still rare and potentially contain bias in their impact estimates. This paper explains the selection problem common to simple difference method employed by these studies. Then this paper introduces two evaluation designs, difference-in-differences and randomized experiment, that can more credibly estimate program impacts, and still simple enough to be independently conducted by amil institutions.

Keywords: impact evaluation, zakat, amil

INTRODUCTION

The amount of zakat and infaq collected nationwide in Indonesia has grown rapidly at an average rate of 24 percent per year for the last decade. The growth might show the increasing awareness of people of the duty of zakat payment, their willingness to share their wealth, and their trust to zakat amil institutions (*Lembaga Amil Zakat - LAZ*).

Many zakat amil institutions published their financial reports so public can monitor how their donation money spent. However, none have conducted a rigorous study on the impacts of those programs on the recipients.

Such studies will benefit the amil institutions themselves, donors, and the benefit recipients of amil programs. Having measures on the size of the impacts of every program allow amil institutions to better allocate the collected fund. It also help donors to decide where to donate their money based on the outcomes of amil's programs. The current and future program recipients will have a chance to participate into programs that has been proven to increase

their welfare.

There are studies on the impacts of LAZ programs, but they are mainly focused on productive zakat programs, only represent population from limited areas in some part of Java, and are likely to contain some bias since they are not designed to avoid it. This paper argues that we need to improve the accuracy of the impact estimates by designing the evaluation studies that eliminates the potential confounding effects.

This paper introduces two well-known impact evaluation designs that can deliver credible estimates without involving sophisticated calculation. Hence, amil institutions would be able to independently conduct the impact evaluation of their programs, although a supervision from and/or collaboration with researchers will be preferred.

The next section will briefly review previous literature on the impact of zakat on macro and micro level. The third section will review the common methodological weakness of these studies. Two simple but credible methods for estimating program's impact will be introduced in the following two sections. Last section will conclude and

suggest the way forward for conducting impact evaluation on amil programs.

LITERATURE REVIEW

The literature on the impact of zakat can be grouped into two general categories: studies that examine the macro effects of zakat on overall area level variables, and those of micro level that measures the impacts on individual recipients' well-being. Using ARDL-ECM model, Suprayitno et al. (2017) found that the amount of zakat in Malaysian states has positive Granger causality on school enrollment in those states. Applying OLS regression on Sudan data during period from 1990 to 2009, Abdelmawla (2014) found that an increase in zakat by one percent can be associated with reduction in poverty by 0.26 to 0.28 percent, depending on the definition of controlled education variable.

The types of outcomes that get examined by micro level studies includes conventional measure of welfare, such as HDI (Nurzaman, 2016), as well as an Islamic measure such as CIBEST model (Beik and Pratama, 2017; Beik and Arsyianti, 2016). While there are two models of zakat disbursement, including consumptive and productive models, most micro level empirical studies examined zakat programs of productive type.

Beik and Arsyianti (2016) took a survey from February to May 2015 to the recipients of BAZIS DKI Jakarta and Dompét Dhuafa's productive zakat programs in Jakarta and Bogor. Although they measured outcomes using CIBEST model, the underlying calculation shows that the number recipient households who are below the official poverty line, has reduced from 159 to 62 households since they participated in the programs.

Beik and Pratama (2017) collected similar survey to the participants of Dompét

Dhuafa's Masyarakat Mandiri program in Bogor regency during February-March 2015. They found that the recipient households' average monthly income has increased 47 percent from previously Rp 874 thousands before participating in the programs.

Pratiwi (2016) also conducted similar study in Serang city during January-March 2016 and found that the average monthly income of BAZNAS program participant increased by 62 percent, while that of Dompét Dhuafa program increased by almost three times of the pre program income.

Nurzaman (2016) looked at the impacts of productive zakat programs in Jakarta on human development index (HDI) measures. He found that while at the baseline survey zakat recipients have lower HDI than the control group of commercial funding recipient, one year later the zakat group's HDI is then better than that of control group. Sohag et al. (2015) examined the impacts of Hasana projects, a productive zakat model, organized by Masjid Council for Community Advancement (MACCA) in Manikganj district, Bangladesh. Using propensity score matching, they estimated an impact on the participants' income ranges from BDT 501 to BDT 1065 depending on the chosen matching method.

We can see above that most empirical studies on zakat impact employed a simple comparison of the beneficiaries' welfare before and after participating in zakat program. Nurzaman improved the analysis by having a control group from non beneficiaries to compare the change in the beneficiary group. The latter design resemble difference in differences (DiD) design, only that Nurzaman did not explicitly take difference between both groups' change of outcomes to get the estimate of program's impact.

Sohag et al.'s matching method has the advantage of controlling the observable differences between program participants

and non participants. However since he compared the level of income, rather than the change, part of the difference in the income might be explained by the initial difference before the program introduced.

Most of mentioned before-after studies did not measure the effects of size and duration of the programs. Regression is needed to obtain elasticity estimates. Overall, there are still rooms for potential improvements toward the existing studies on the impacts of LAZ programs.

SELECTION PROBLEMS

Getting accurate estimates of program impacts takes more than just looking at the difference in outcome before-after the program or between recipient and non recipient of the program. Employing multiple regression with a bunch of control variables is often still not enough to prevent bias.

The primary cause of the bias is that the selection of program recipient involves factors that often unobserved by researchers, and those factors also determine outcomes. When these factors are not controlled, their effects on the outcomes will get captured by the estimates of the program's impacts, i.e. making the estimates biased.

Take an example of a program that is common among zakat amil institutions, such as financial aid for poor students. The students selected into the program are those who are considered the poorest families observed by amil in their operating area. Comparing those students with other students who are expectedly less poor will result in a biased estimates. Since less poor students are expected to perform better had the poorer students not receive aids, the estimates coming from simple comparison of performances of both groups is likely to be smaller than it actually is.

Another common approach to

measure program outcomes is the comparison of the recipient situation before and after they receive program benefit. If something else changed within the comparison period, and that thing also has impacts on outcomes, then those effects will also be captured by the simple before-after difference.

If it is believed that there is no big change to the recipient students other than getting program benefit, and there is no significant change either to their peers at schools, then a change in student performance might be safely attributed to the program if the change is significant. A small change would be harder to be confidently attributed to the program since it might as well come from a normal performance fluctuation that would happen anyway without the program.

DIFFERENCE IN DIFFERENCES

While simple comparison of before-after or recipient-non recipient outcomes has a lot of issues, the combination of the two could greatly improve the reliability of the estimates. All that is needed is to have data of outcomes for recipient and non recipient from both before and after the program is executed, and then subtracting pre program differences between recipients and non recipient from the post program difference. This method is known as difference in differences (DID) or double differences (DD).

This method have the advantages of each of the two simple difference method, while removing each's drawbacks by combining the two. Differencing across time has the advantage of removing the effects of time invariant factors, but it captures the effects of any changes that happen within the same period. While differencing across group could remove the effects of changes across

time that common to both groups, however it will capture the effects of differences between groups that does not come from the program. Combining both differences would remove the effects of group differences and of common changes across time

Difference in differences is among the most popular method in program evaluation due to its simplicity while still delivering reliable result as long as its assumption is assured. DD requires that the size of difference across recipient and non recipient is relatively stable over time. In order to convince audience that the assumption is held, researchers need more pre treatment data to be able to show that the difference across time is indeed stable. If it turns out that the differences are not stable over time, then DD estimate would lose its reliability. The researcher would need to sort to other method of program evaluation.

RANDOMIZATION

A better way to make sure that selection problem does not contaminate the impact estimate is by randomizing the selection itself. This randomization based method is known as field experiment or randomized control trial (RCT). Since treatment is randomly allocated in an experiment, the distribution of observed and unobserved characteristics are expected to be similar between treatment and control group. Hence the differences in outcomes between treatment and control can be surely attributed to the differential treatment they received.

RCT studies are strong in internal validity, but they lack of external validity. The results of field experiment under one setting and at one place might not be replicated elsewhere and under another setting.

However, the external validity of field experiments can be improved by replicating

the experiments in different places and with another settings. With enough replication, pattern can be drawn and conclusion can be made about what make program works in one setting but not in another. This will help to design program that can work in most settings and how to adjust it in different settings.

Another problem of RCT studies is on the ethical issue. Field experiments require the researchers to keep subjects in the control group from getting program benefits. This might be a serious issue when there is actually enough resources to provide all subjects with program benefits.

However, zakat and infaq funds are scarce. Even without randomization, there are still many program target population who still do not get the benefit due to the resource limitation. What researcher does is only replacing the allocation mechanism that is currently exercised to a randomization based mechanism.

RCT study design is increasingly used by government around the world to evaluate programs and policies. Governments have a lot more resources than non governmental organizations such as zakat amil institutions. Even governments do not have enough resources to provide all program benefit to all citizens at the same time. Both the lack of resources and the importance of spending resources wisely only to programs that are scientifically proven to be effective has motivated governments to design program pilots as a randomized control trial. Those reasons are even stronger for zakat amil institutions.

Although most existing programs were not designed as experiment, their expansion can still be designed as so. There are still regions and areas that has not been covered by amil institutions or their programs. Under resource limitation, the selection of new recipients in the existing area or the selection of the new area itself can

be randomized.

Some programs are not fit for experiment. Quick response programs to natural disaster cannot be limited only for a group of the victims while keeping it from the other. However, longer term programs for recovery and rebuilding the impacted communities are not as urgent, require larger budgets, and are not clear if they are better than their alternatives.

CONCLUSIONS

This paper reviews the drawbacks of the common methodology used by existing studies on evaluating amil programs. This paper then introduces two simple yet credible methods that has been standard in the literature of program evaluation. If the readers are interested to learn more techniques for program evaluation, they might resort to the canonical books in program evaluation, such as Morgan and Winship (2015), Angrist and Pischke (2008), as well as Imbens and Rubin (2015).

The two evaluation designs introduced in this paper are simple enough to be exercised independently by amil institutions to evaluate their own programs. However, amil institutions have nothing to lose and will mutually benefit from a collaboration with external researchers.

Researchers need to collaborate with amil institutions to integrate the planning of studies and the zakat programs. Experimental studies absolutely requires the amil to agree to allocate its programs randomly over the chosen population. Baseline survey must be taken before a program is actually rolled out in the targeted area, so researchers need to know amils' program planning in advance. Asking retrospective question about pre program situation after the program started would provide less accurate information.

With a carefully designed studies, the

impacts of LAZ programs can be accurately estimated. This will allow a true comparison of the cost effectiveness of each program that will help to prioritize the use of zakat and infaq funds. Besides improving the quality of the evaluation, the number and coverage of evaluation also needs to be expanded to enable the comparison.

The author believe that both the quantity and the quality of these studies are still less than socially optimal, hence more resources need to be allocated for them. Since there is huge potential resource savings from having such studies, the resources spent on them are well justified.

REFERENCES

- Abdelmawla, M. A. (2014). The Impacts of Zakat and Knowledge on Poverty Alleviation in Sudan: An Empirical Investigation (1990-2009). *Journal of Economic Cooperation and Development*, 35(4), 61–84.
- Anis, F. M., & Kassim, S. H. (2016). Effectiveness of Zakat-based Programs on Poverty Alleviation and Economic Empowerment of Poor Women: A Case Study of Bangladesh. *Journal of Islamic Monetary Economics and Finance*, 1(2), 229–258.
- Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Beik, I. S., & Pratama, C. (2017). Zakat Impact on Poverty and Welfare of Mustahik: A CIBEST Model Approach. *AFEBI Islamic Finance and Economic Review*, 1(01).
- Beik, Irfan Syauqi, & Arsyianti, Laily Dwi. (2016). Measuring Zakat Impact on Poverty and Welfare Using CIBEST Model. *Journal of Islamic Monetary Economics and Finance*, 1(2), 141–160.

- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference*. Cambridge University Press.
- Nurzaman, M. S. (2016). Evaluating the Impact of Productive Based Zakat in The Perspective of Human Development Index: A Comparative Analysis. *Kyoto Bulletin of Islamic Area Studies*, 9(29), 42–62.
- Pratiwi, E. (2016). *Analisis Pendayagunaan Zakat Produktif Sebagai Pengurang Kemiskinan Berdasarkan Model CIBEST*. Institut Pertanian Bogor.
- Saad, N., & Abdullah, N. (2014). Is Zakat Capable of Alleviating Poverty?: An Analysis on the Distribution of Zakat Fund in Malaysia. *Journal of Islamic Economics, Banking and Finance*, 113(3250), 1–27.
- Sohag, K., Mahmud, K. T., Alam, M. F., & Samargandi, N. (2015). Can Zakat System Alleviate Rural Poverty in Bangladesh? A Propensity Score Matching Approach. *Journal of Poverty*, 19(3), 261–277.
- Suprayitno, E., Aslam, M., & Harun, A. (2017). Zakat and SDGs: Impact Zakat on Human Development in the Five States of Malaysia. *International Journal of Zakat*, 2(1), 61–69.
- Muhamad Said Fathurrohman
Universitas Airlangga (Unair)
Indonesia
said@feb.unair.ac.id