SWEDISH ECONOMIC POLICY REVIEW 13 (2006) 165-167

Comment on Jan Willem Gunning: Aid evaluation: Pursuing development as if evidence matters

Per-Åke Andersson*

This very interesting paper proposes a new methodology for development evaluation. The methodology is based on statistical impact evaluation, but modifications are made to allow for heterogeneity, non-linearity and to some extent externalities. In addition, the method can be used at different levels of aggregation, at the project level, at the sector level or at the national level.

The general message of the paper is that development interventions should be tested much more thoroughly than what is done today. The author notes that multilateral and bilateral donor agencies have been reluctant to use available tools. There has also been a tendency towards evaluating processes instead of impacts. I am inclined to agree with the view that development interventions should be evaluated more comprehensively and as Savedoff et al (2006) and Easterly (2006) argue, there is a need for an independent evaluation of aid.

The first section of the paper discusses the impact evaluation tool. The importance of a counterfactual and reliable data is pinpointed. These impact evaluations are designed for binary solutions, assuming homogeneity of treatment as well as possible impact. They are used at the project level and there have been a number of applications in evaluations of social safety nets, schooling programs targeted at the poor, health interventions and even rural empowerment programs. However, they are not defined for the higher level of aggregation such as the sector or the national level. Thus, the current tendency for donors to move away from project financing towards sector budget support and budget support makes the impact evaluation tool less useful.

At the same time, evaluation tools for a higher level of aggregation have proven to have limitations. The cross-country regression approach, the log framework approach and the Computable General Equilibrium (CGE) models have problems such as econometric weaknesses, structural problems and black box problems.

* *Per-Åke Andersson is a Research Fellow at the Swedish Agency for Development Evaluation.*

The second section presents the innovative methodology. The author sees it as a bottom-up approach where impact evaluations at the project level are performed in a way allowing for conclusion to be drawn at a higher aggregated level. First, a representative sample of all activities by a Ministry or Government should be selected. Second, impact evaluations are performed for the selected activities at the project level. The results are aggregated and the end result of the evolution should be statements like "public spending in country X reduced poverty by Y per cent". The author notes that the end result is not aid specific, but I believe this would not be a problem. The attempts to implement the Paris declaration on improved donor coordination, harmonisation and alignment, indicate a new kind of seriousness from the donor community. This might also be the time to change expressions from aid evaluation to development evaluation.

Gunning cautions that the methodology has not yet been tried, but a number of evaluation studies are being performed by the evaluation agency of the Dutch Ministry of Foreign Affairs to test the feasibility of the approach. It will of course be very interesting to follow these attempts.

Before proceeding, I would like to make a comment on the selection of a representative sample. Different actors would surely have different views on which activities to include. Representatives of the Ministry as well as the donor community would naturally be interested in including more successful activities, especially, since the methodology will make it possible to compare effectiveness between different activities within a sector as well as between different sectors. Thus, the importance of independent evaluations is obvious. In addition, since we are no longer suggesting aid evaluations, but evaluations of total government expenditure, there is an apparent need to strengthen the evaluation capacity in partner countries to increase ownership.

The econometric fundaments of the methodology are panel data and fixed effect regression analysis. The most important feature is the acceptance of heterogeneity, both in interventions and impacts. Gunning presents an example from Tanzania, where there are three inputs (wells, sanitation and training) and three impact variables (cholera incidence, water fetching time and poverty). This design feels important since development and aid policy are characterised by goal congestion and the uneasy coexistence of explicit and implicit objectives. The heterogeneity feature implies that a control group is no longer needed. Instead, fixed-effect regression analysis based on differences in treatment history over time makes it possible to compare impacts of different interventions. The heterogeneity feature implies that interaction effects are captured and multiple impacts of the same intervention are accepted. The methodology also accepts non-linearity and externalities can be included in the design.

When assessing aid effectiveness, the ultimate impact should be considered. But we are also interested in the links between interventions and impacts. Unfortunately, the proposed methodology does not fully open this black box, since no explanations are offered on why interventions are effective. Gunning proposes that more informal and descriptive studies may be needed to capture this link.

The main problem of the methodology is data intensity. Time series data on interventions and impacts is needed. The example in the paper has data for 35 years, but this kind of data base is rare in developing countries. Future tests will indicate how sensitive the methodology is to data availability.

My final comment is connected to the delicacy of the evaluation design. First of all, we do not want to arrive at a situation where the evaluation design is affected by data availability, since we can expect selection bias in data access. Successful projects have longer histories and thus, more data. Second, we need to be aware of what the data is measuring. For example, a number of expenditure tracking exercises done lately have shown that dispersion of funds is common. It is therefore important to make sure that the intervention is measured correctly. If a specific school is evaluated, one of the interventions to be included has to be the amount that actually reached the school, not the budgeted amount at the provincial or ministerial level. This might imply that we revive the micro/macro paradox, which suggested that aid worked better at the micro than the macro level. Third, we need to examine when and where to include indicators of macro-economic performance such as inflation, since we need to avoid that the methodology only provides a more modern version of the micro-macro paradox.

References

Easterly, W. (2006), The White Man's Burden*,* the Penguin Press, New York.

Savedoff, W., Levine, R. and Birdsall, N. (2006), When will we ever learn: Improving lives through impact evaluation, Report of the Evaluation Gap Working group, Center for Global Development.