

Is there a micro-macro paradox in international aid, or do the data deceive?

Stephen Howes, Sabit Otor and Cate Rogers¹

Abstract

In 1986, Mosely first drew attention to an apparent paradox in the performance of international aid. Microeconomic data from evaluations of aid financed projects showed a majority of projects were successful, whereas macroeconomic data from regressions of aid on growth were discouraging. The paradox, if real, implied that the aggregate impact of aid was less the sum of its parts. Mosely asked whether the paradox was real or whether the “data deceived.” This question, which has come to be equated with the issue of whether aid works, has been the subject of numerous cross-country regressions to test whether aid has an impact on growth (or related variables). But the regression results have been inconclusive, and the methodology at least in part discredited. Evidence from case studies offers an alternative test. The most prominent case study approach is by Picciotto (2009), which claimed to find strong evidence for the existence of the paradox, namely the fact that one third of World Bank country assistance program evaluations show success at the project (micro) level but not at the country (macro) level. This paper re-evaluates Picciotto’s claimed findings. Only about one-third of the disconnects survive critical scrutiny, and the source of these remaining disconnects has nothing to do with negative effects of aggregate aid. Although in the Picciotto case, the data do indeed deceive, we conclude that country-level aid studies are nevertheless a useful tool for donors to use to guard against possible, albeit uncertain, negative impacts of aid at the country level.

Introduction

In 1986 Mosley drew attention to an apparent paradox in the performance of international aid. Microeconomic data from evaluations of aid financed projects showed a majority of projects were successful, whereas macroeconomic data from regressions of aid on growth were discouraging (Mosely 1986, p.22). Mosley posed the following questions:

What is going on? Is it true as the data suggest that aid projects are succeeding while aid as a whole is failing, if so how? Or do the data in fact deceive? (1986, p.22)

¹ Centre for Development Policy, Crawford School, Australian National University.

The existence of a micro-macro paradox has come to be largely synonymous with the question of whether aid is effective. If there was strong evidence for it, it would markedly weaken the case for aid, and it would also force donors to invest much more in a search for ways to preserve micro-level success while weakening adverse macro-level consequences. There is in fact probably no more important question in the field of aid than whether the micro-macro paradox of aid actually exists.

Various reasons have been put forward for the existence of a macro-micro paradox. These include:

- The possibility of fungibility. If aid projects succeed, but would have occurred even without aid funding, then the aggregate impact of aid could be less than the aggregate impact of the projects, if the actual use to which the aid funding is put is of less value.
- The possibility of exchange rate appreciation. To the extent that aid is used to purchase non-traded goods, it will put upward pressure on the real exchange rate, with possibly negative growth impacts.
- The possibility of institutional deterioration. Aid may lead to an 'aid curse' similar to the 'resource curse'. Societies might consume resources trying to obtain aid which they would otherwise deploy for entrepreneurial purposes. In extreme cases, the prospect of aid can lead to corrupt and even violent behaviour. It can also reduce the expectation of citizens from their government, and then reduce the demand for good governance.
- The possibility of high transaction costs. The cumulative impact of a large number of aid projects might weaken government. It might distract civil servants from their own duties, it might make budgeting and planning difficult (especially if aid is volatile and/or unpredictable), and it might lead to a brain-drain from the civil service to the typically-better-paying donor community, all with negative consequences for economic management and performance.

All of these are possible and plausible pathways by which, regardless of success at the project level, aggregate aid would negatively impact on development. The essence of the micro-macro paradox is that the aggregate impact of aid is less than the sum of its parts. The aggregate impact of aid – taking into account both the impact of aggregate aid, and the combined impact of individual projects -- might still be positive, but might also be very small or even negative, depending on the relative contribution of each.

Given the importance of the paradox, and the *prima facie* plausibility of the causal mechanisms which would give rise to it, it is not surprising that the search for the micro-macro paradox has been the source of major research programs. Most of this research has been in the nature of cross-

country regressions. However, as summarized in the next section, this literature has disappointed, and left few if any firm results in its wake.

An alternative approach would be to use case-studies. This approach was used most famously in 2009 by Robert Picciotto, former head of the World Bank's Independent Evaluation Group. In his paper titled "Development Effectiveness: An Evaluation Perspective" Picciotto argues that the case-study evidence points to the existence of a significant micro-macro paradox in World Bank aid: in sixteen² of the 55 country-level evaluations, aid is rated a success at the project (micro) level, but not at the country (macro) level and in three it is considered a success at the country level despite being unsatisfactory at the project level. The Picciotto study is an important one, and is the central focus of this paper. First, however, we consider the cross-country literature, as it is disappointment with this literature which has given rise to the need to pursue a case-study approach of the type Picciotto deploys.

Cross-Country Regressions

Mosely (1986) used data from 1961-1981 to perform a series of regressions of aid on growth. He found that when other determinants of growth such as savings rates and export growth were held constant he could not confirm for any continent the significant and positive relationship of aid on improved standards of living (taken as being growth) that micro level data from World Bank project performance data might suggest.

Post-Mosely, the idea that projects on average work has been little contested. Cassen (1994, cited in Doucouliagos & Paldam 2009, p. 438) finds that about 50 per cent of all development projects work and very few of the remaining projects cause harm even if they fail. Aggregating these results leads to a modest case for aid working. Riddell in his 2007 book on aid concludes that 'The available evidence suggests, quite strongly, that the clear majority of official aid projects achieve their immediate objectives.' (pp. 192-3)

But whether aid works at the macro level has been much contested. A massive literature built on cross-country growth-aid regressions has been developed. In their survey, Doucouliagos and Paldam count some 97 papers to the end of 2004 (2009, p.435). Unfortunately, this extensive effort has not led to a consensus. There are two main attempts to survey the literature: McGillivray et al (2005), and Doucouliagos and Paldam (2008, 2009). McGillivray et al conclude that post 1998 there seems to

² At one point, Piciotto says there are fifteen such cases (p. 198) but his count of sixteen in the previous paragraph on the same page is consistent with this table reproduced as Table 1 in this paper.

be agreement that aid works, to the extent that in its absence growth would be lower (McGillivray et al, 2005).

Doucouliagos and Paldam, however, reach the opposite conclusion. These two authors have conducted a number of surveys into aspects of aid effectiveness literature. Doucouliagos and Paldam (2009) conclude that "After 40 years of development aid, the preponderance of the evidence indicates that aid has not been effective."

The limitations of the cross-country literature extend beyond its inconclusive nature. There are at least three other problems with the literature

First, the cross-country growth regression literature, within which the aid-growth literature is situated, is itself indecisive, and lacks legitimacy. It is not only aid where the profession has struggled to find a robust linkage with growth. Easterly (2009) concludes that 'we have failed to identify' (p. 122) *any* variables which provide a robust explanation of growth across countries. According to him, 'We have learnt something from the failure of growth regressions: that there is no universal factor X that works everywhere to reliably raise growth' (p. 129). Growth regressions are no longer widely trusted. Angus Deaton writes that 'the econometric studies that use international evidence to examine aid effectiveness currently have low professional status' (Deaton 2009, p.2).

Second, while it is unclear whether more aid leads to more growth, it is clear that more growth leads to less aid (Roodman 2008, p.7), as donors come to perceive the successfully-growing recipient as being less and less in need of aid. To overcome this reverse causality, the relationship between aid and growth has to be studied using 'instruments' for aid, that is, variables which are thought to be correlated with aid, but not with growth. But it is unclear which variables should be used as instruments and different choices will lead to different results (Roodman, 2008, p.10). This leads Roodman to conclude that 'there appear to be almost no findings in the contemporary literature that a) find a significant effect of aid on growth, b) are robust and c) are free of methodological problems [described here]'. (Roodman, 2008. p.17) Third, Arndt, Jones and Tarp (2009) restate a concern that the timeframe over which growth effects may be expected to accrue from aid is greater than many aid growth models allow. For example, the impact of improvements in education could be expected to influence growth only after a very substantial lag, perhaps ten years or more (Arndt, Jones & Tarp 2009, p.5). Despite these difficulties, the aid-growth literature continues to thrive. One growing strand of this literature is to test directly for a micro-macro paradox by examining directly whether aid generates paradoxical outcomes, such as higher corruption, or higher exchange rates. However, these sub-strands suffer from the same deficiencies as the broader aid-growth literature.

They deliver conflicting results, they lack credibility in the broader profession, and they require the use of instrumental variables for aid. To illustrate, Knack (1999) and Djankov (2007) conclude on the basis of regression analysis that aid weakens institutions. But Ear (2007) critiques Knack (1999) and argues that 'the causal link between aid dependence and worsening quality of governance may be tenuous at best and sensitive to alternative specifications'.

Another approach has been to use cross-country regressions at the sectoral level, for example to look for a link between sectoral (or total) aid and sectoral outcomes. For example, Dreher, Nunnenkamp, and Theile (2006) estimate a series of equations in which public expenditure, education outcomes and institutional quality are jointly determined. Their results show aid significantly increases primary school enrolment (Dreher, Nunnenkamp & Theile 2006, p.20). Mishra and Newhouse (2007) find that health aid has a significant positive effect on reducing infant mortality (p.6). There are too few studies of this kind to draw overall conclusions at this stage. It may be that they also turn out to be fragile. Moreover, they do not rule out the macro-micro paradox. It could be that aid improves health and education indicators, but slows growth.

Case Studies and Picciotto's approach

Given the increasing recognition of the short-comings of the cross-country regression approach, increasing use is being made of case studies. This is not only true for aid, but for understanding growth and development more generally. Darlauf et al, in their survey of growth econometrics in the *Handbook of Economic Growth* (2005) write 'We would also argue that an important contribution of growth econometrics has been the clarification of the limits that exist in employing statistical methods to address growth questions. One implication of these limits is that narrative and historical approaches ... have a lasting role to play in empirical growth analysis.'

There are a growing though still small number of individual case-studies examining the country-level impact of aid. Arndt, Jones and Tarp's 2007 study of aid to Mozambique uses growth accounting to evaluate the impact of aid on different drivers of growth. They find that aid has played a 'determinant' role in growth and poverty reduction since 1992. In their view aid played a critical role in building infrastructure and expanding access to health and education. On the negative side, they found that aid had generated important governance and economic management challenges – raising questions about the sustainability of its impact (Arndt, Jones & Tarp 2007, p.79) Moss, Pettersson and van de Walle (2008) conclude from their qualitative review of Africa that 'a large and sustained volume of aid can have negative effects on the development of public good institutions in low-

income countries.’ (p. 274) though the authors themselves concede that this is only a ‘tentative claim.’

The distinguishing feature of Picciotto’s paper (2009) is the strength or firmness of its findings. He examines the extent of the disconnect between performance ratings provided in IEG Country Assistance Evaluations (CAEs) on Bank Country Assistance Strategies (CASs) and project portfolios. He examines 55 CASs subject to independent evaluation and finds that a ‘fully fledged micro-macro paradox was found to prevail in a third of the cases.’ (Picciotto 2009,p.198)

Table 1 below reproduces Picciotto’s Box 8.2. The year next to the country name refers to the year of the country assistance evaluation. Years in brackets refer to a specific sub-period of time within the overall timeframe of the Country Assistance Evaluation (CAE), for which separate program-level evaluations are provided by some CAEs. Hence Mexico 2001 appears three times in the satisfactory column, with different years rated separately. This is typically done when the period covers a crisis: Russia in 1998 or Mexico in 2001-2. Finally, the large number of asterisks alongside country assistance strategies indicate that either or both of the country and aggregate project ratings were considered to be either moderately (or marginally) satisfactory or unsatisfactory.

The IEG started producing CAEs in 1995. Just as IEG project level evaluations rate all Bank projects at their closure as highly or moderately satisfactory or unsatisfactory, CAEs apply the same rating scheme to country assistance strategies, which are the strategies produced by the Bank to guide its assistance to each recipient country, typically for a period of about four years (World Bank 2009, ES). Box 1 summarises aspects of the IEG approach to country assistance evaluations. Picciotto uses this CAS performance rating to judge the macro impact of aid.

The IEG also undertakes a number of other country-level ratings, including an aggregate project rating. This is used by Picciotto to judge the micro impact of aid, that is, whether, on average, Bank-funded projects work. It is not clear from IEG documentation how this aggregate project rating is arrived at. The clearest explanation is that it is an aggregation of project outcomes in the country during the CAE period. Presumably, what this means is that a satisfactory rating is given if and only if half or more of projects which close in the period being covered are rated satisfactory by the IEG in the course of their project-level evaluations. (The IEG rates each completed Bank project.)

Box 1 Country Assistance Evaluation (CAE): approach and ratings.

A formal CAE methodology was adopted in 2000. It draws on the Country Assistance Strategy (CAS) for its initial evaluative framework. The CAE determines the extent to which the CAS's major strategic objectives were relevant (judged against major development constraints) and achieved. According to the IEG, three dimensions of the country assistance program and its evaluation are examined when coming up with the rating of performance against CAS objectives:

Produces and services dimension, involving a 'bottom-up' analysis of major loans, advisory and analytical assistance, and partnership and coordination.

Development impact dimension, involving a 'top-down' analysis of the program objectives for relevance, efficacy, outcome, sustainability and institutional impact.

Attribution dimension, where the evaluator assigns responsibility for the program outcome to four categories of actors: (a) the client; (b) the Bank; (c) partners and other stakeholders; and (d) exogenous forces (e.g. international economic shocks, natural disasters etc).

Bank performance is measured predominately on the basis of actions that the Bank can directly control; however the development outcome of the CAS is an aggregation of the actions of the four sets of actors noted above. Hence it is possible that if the actions of one of the actors are negative, the positive contributions of others may be overwhelmed leading to an unsatisfactory CAS performance rating.

The rating scale for CAS performance is given below. Note that this is known as the rating for CAS 'outcomes.' Separate ratings are given at the country level for sustainability and institutional impact as well.

Ratings Scale:

Highly Satisfactory: the assistance program achieved at least acceptable progress towards all major relevant objectives, and had best practice development impact on one or more of them. No major shortcomings (such as safeguard violations) were identified.

Satisfactory: the assistance program achieved acceptable progress toward all major relevant objectives. No best practice achievements or major shortcomings were identified.

Moderately Satisfactory: The assistance program achieved acceptable progress toward most of its major relevant objectives. No major shortcomings were identified.

Moderately Unsatisfactory: The assistance program did not make acceptable progress toward most of its major relevant objectives, or made acceptable progress on all of them but (a) did not take into account a key development constraint or (b) produced a major shortcoming, such as a safeguard violation.

Unsatisfactory: The assistance program did not make acceptable progress toward most of its major relevant objectives, and either (a) did not take into adequate account a key development constraint and (b) produced a major shortcoming, such as a safeguard violation.

Highly unsatisfactory: The assistance program did not make acceptable progress toward any of its major relevant objectives and did not take adequate account a key development constraint while also producing at least one major shortcoming, such as a safeguard violation.

Source: World Bank (2005)

As is evident from Table 1, there is a positive association between the ratings ascribed to project results and country assistance outcomes in 36 instances, and a disconnect in one third, or 19 of the 55 country strategies. In sixteen of these, project portfolio performance is rated satisfactory, but country strategy performance unsatisfactory, and in the other three it is the other way round.

The Picciotto approach is certainly innovative, and important, but suffers from two shortcomings. First, he accepts the IEG findings at face value. He does not apply any independent scrutiny. Second, he at no stage asks whether any of the typical causal mechanisms behind a paradox – the ones we listed earlier, and which he also lists – explains the disconnects that he finds.

In what follows we address these failings by independently scrutinizing the IEG findings, and by examining the reasons for those for which, even after independent scrutiny, there is a disconnect. We do this for both groups of disconnect countries: both paradox and reverse-paradox countries.

Scrutiny of IEG ratings

A qualitative review was undertaken of the 19 cases identified by Picciotto as being instances where the micro-macro paradox was evident. The CAE reports were examined primarily to determine whether the ratings applied to portfolio performance were justified. CAS performance ratings require more judgement, and so are harder to scrutinize, but we checked that the rating was consistent with the IEG methodology (Box 1)

There are several limitations to the approach that we have undertaken. First, we were working off public documents. It is possible that there is more information available regarding project performance in particular that we were not able to access. Second, to some extent this approach is a *de facto* assessment of the quality of evaluation reports (particularly in relation to how well/badly they use evidence to support their conclusions) rather than the actual performance of CASs and portfolios. Nevertheless, our analysis does raise important questions about Picciotto's claims.

Table 2 contains summary findings of our review.

Table 2: Review of IEG project portfolio and country assistance ratings

Country	Time period for evaluation	Project performance Rating	Country Assistance Strategy performance rating	Our assessment
Morocco 1997	1997-2000	Satisfactory	Marginally Unsatisfactory	Agree
Bulgaria 2002	1991-1997	Satisfactory	Unsatisfactory	Agree
Costa Rica 2000	1990s	Satisfactory	Unsatisfactory	Disagree
Ecuador 1999	1990s	Satisfactory	Unsatisfactory	Disagree
Haiti 2002	1986-2000	Satisfactory	Unsatisfactory	Disagree
Jamaica 1999	1993 CAS	Satisfactory	Unsatisfactory	Disagree
Lesotho 2002	1994,1996 &1998 CASs	Satisfactory	Moderately Unsatisfactory	Disagree
Mexico (1992-94) 2001	1992-1994	Satisfactory	Partially Unsatisfactory	Agree
Nepal 1999	1990s	Satisfactory	Unsatisfactory	Disagree
Paraguay 2001	1993 & 1997 CASs	Satisfactory	Unsatisfactory	Disagree
Peru 2003	1997-2000	Satisfactory	Unsatisfactory	Agree
Russia 2002 (1992-98)	1992-1998	Satisfactory	Unsatisfactory	Disagree
Ukraine 1999	1992-1996	Satisfactory	Unsatisfactory	Disagree
Yemen 1999	1970-1996 CAS	Satisfactory	Marginally Unsatisfactory	Agree
Zambia 2003	1996-2001	Satisfactory	Unsatisfactory	Disagree
Zimbabwe 2003	1990-2000	Satisfactory	Unsatisfactory	Disagree
Ethiopia 1999	1990s	Unsatisfactory	Satisfactory	Disagree
Ghana 2000	1995 & 1997 CASs	Unsatisfactory	Satisfactory	Disagree
Russia 2002 (1999-2001)	1999-2001	Unsatisfactory	Satisfactory	Agree

Note: The disagreement is with the project rating in all cases except Costa Rica.

Source: CAEs and Authors' assessments

Table 2 shows ten cases where the ratings on project performance were rated as satisfactory but where CAEs themselves provided evidence that, in our view, contradicted this assessment. This evidence is summarized in Annex B. Reasons for disagreement with the CAE assessment include the following (with the countries for which the particular reasons apply given in brackets):

- *Failure to take into account the CAE's own findings.* The aggregate project ratings do not appear in the CAEs themselves, and do not appear to take into account analysis contained in the CAEs. In nearly all cases where we question the rating, the CAE itself makes damning statements about project performance or provides information which questions a rating of satisfactory (Paraguay, Haiti, Jamaica, Zambia, Lesotho, Nepal, Ukraine, Russia, Zimbabwe)
- *Excessive reliance on performance ratings of closed projects.* Completed projects are individually rated by the IEG, and so provide good evidence for project performance. As discussed earlier, these IEG ratings seem to be the basis of the aggregate project rating. However, the projects which close over a, say, five year period may not be representative of project performance over that five years, which will also be influenced, and probably more influenced, by new and ongoing projects. (Paraguay)
- *Low sustainability.* Projects might on average have satisfactory ratings on closing, but if they have low sustainability ratings, it is likely that, when assessing performance at a later date (i.e., at time time of the CAE), their performance should be more harshly judged. Very low sustainability ratings were taken by us as reasons for lowering aggregate project ratings (Ecuador, Haiti, Lesotho, Nepal, Zambia, Zimbabwe).
- *Lack of supporting evidence.* In some cases, no evidence is available in the CAE to support the project rating (Russia, Ukraine)
- *Use of unweighted rather than weighted average.* In some cases, more than 50% of closed projects had a satisfactory rating by number, but not by size (commitment). It is not clear which should be used. The CAE seems to typically only give weight to the former, but we took a weighted-average success rate of less than 50 per cent as a good reason for an unsatisfactory rating overall (Paraguay, Jamaica, Ecuador), especially when the unweighted average was around 50%, as it was for these three countries, on the grounds that it matters more if the larger projects fail.

The evidence summarized in Annex B appears to us to provide compelling grounds for disagreement. Some of the ratings we challenge seem to be simple errors. For example, in Russia the aggregate performance rating is satisfactory, though only 47 per cent of projects by number and 16 per cent by commitment obtained a satisfactory rating. No other evidence is available in the CAE to justify this positive rating.

Other cases are more complex, but no less compelling. Paraguay was given a satisfactory project rating presumably because 50 per cent of projects are rated satisfactory in their ICRs. However, this

is 50 per cent of only three projects,³ and weighting by project size even this indicator falls to 35 per cent. Moreover, apart from the three projects which closed (and therefore had ICRs) over the period of review (1992-1999), there were a larger number of other projects underway over the review period. The 2001 Paraguay CAE tells us that of the 9 operations approved between 1992 and 1997, at the time of writing only one had closed, and five of the remaining eight were problem projects. The CAE also tells us that 'Projects approved since 1992 have had serious difficulties' and that in August 1999 the Paraguay portfolio was rated among the Bank's ten worst. Surely, even if one wants to attach some weight to the 50 per cent rating for closed projects (rather than the 35 per cent for the weighted average for these same projects) the poor performance of the non-closed projects should tip the aggregate project rating to unsatisfactory.

Zimbabwe is another case where the headline number is misleading. 6 out of 8 of projects approved between 1990 and 2002 had satisfactory ratings. However, the CAE notes that the two structural adjustment credits 'did not achieve their major objectives.' (p.12) If so, then at most 55 per cent of projects by volume should be regarded as satisfactory. For most of these projects (by volume and size) the rating is marginally satisfactory, with unclear or unlikely sustainability, suggesting that overall project performance could not be regarded as satisfactory.

In some other cases, there is more room for different views, but we still think the CAE rating should be challenged. In the case of Nepal, 65% of completed projects were rated satisfactory, and 85 per cent by volume. But only 16 per cent of projects were rated sustainable, and 27 per cent of commitments were cancelled. According to the CAE itself, projects were designed without due consideration to the policy environment. With these major shortcomings, an aggregate project rating of satisfactory no longer seems warranted.

We did not re-assess country ratings, except in one case, where it seemed to us that IEG had not followed its own methodology. In the case of Costa Rica, the CAE notes significant progress towards the objectives of the Country Assistance Strategy, and that the primary driver of this was the government. The document also notes that the Bank's strategy was relevant. However it is harsh on the Bank's judgement in relation to the pace of reform and indicates that inappropriate instruments were used. This seems to drive the unsatisfactory CAS performance rating. However, as per the IEG methodology, an unsatisfactory CAS performance rating requires that that the assistance program **did not** make acceptable progress toward most of its major relevant objectives. By IEG's own analysis, this was not the case for Costa Rica. The unsatisfactory rating appears to us to be a rating of

³ It is unclear how the rating can be 50% if there are only 3 projects, but this is what is provided in the CAE (Table 6, p. 28).

Bank performance rather than of the achievement of country assistance objectives, which, according to the IEG itself, should be the basis for CAS performance ratings.

Turning to the three reverse paradox countries, for Ethiopia 1999 and Ghana 2000, project portfolio performance was rated as unsatisfactory, yet both these countries had high project ratings: unweighted average satisfactory ratings of 79 per cent and 78 per cent respectively, and significantly higher than average ratings. In the case of Russia 2002 (1999-2000), the unsatisfactory rating on project performance appears justified.

Thus from an initial list of 19 cases of disconnect, only six stand up to scrutiny. It is possible that some of the 12 which do not stand up to scrutiny could be cases of disconnect, but at a minimum one should have a high level of doubt as to whether these are disconnect countries, and, with a lower level of confidence, one can hold that they are not.

What is the reason for the paradox in the remaining six cases? There appear to be two main ones. First, the disconnect is often caused by over-ambitious CAS objectives (Morocco 1997, Bulgaria 2002), often related to a poor understanding of the political context and an overestimation of either ability or willingness to reform (Yemen 1999, Peru 2003). Second, in the case of both Mexico 2001 (1992-94) (the Tequila Crisis) and Russia 2002 (1992-98) (the collapse of the Soviet Union and Asian Financial Crisis), crises led to the disconnect. In the former case, the external shock was judged to have prevented project level achievements from translating into country objectives. In the latter case, the Bank's response to the crisis was through to have enabled country objectives to be achieved even in the face of poor project performance.⁴

Neither of these factors are among which could lead to a macro-micro paradox. By definition, an external shock is a non-aid factor. Any disconnect due to an external shock says nothing about whether aggregate aid has an impact which differs from the aggregated effect of individual aid projects. Rather, the role of the shock is to break the link, which would be expected under normal circumstances, between project and country level performance.

The other factor, over-ambitious objectives, is equally unconnected to the question of whether aid impacts the recipient economy or society other than through the impact of individual projects. It

⁴ The IEG has done its own retrospective on CAEs. Using a different sample, it also finds one-third of cases having a disconnect. It explains this by noting that "The CAE is a comprehensive evaluation of the Bank's program in a country that comprises both projects and analytical and advisory activities. Moreover, CAEs must make an assessment of overall Bank strategy, including size, sectoral composition, and type of lending. For example, the CAE outcome may be unsatisfactory if there are critical omissions in the Bank's overall assistance strategy, even if the outcomes of individual projects are rated satisfactory. (World Bank 2005, p3)

simply tells us that the Bank is sometimes too optimistic in what it expects at the country level from the achievement of its project objectives.

Conclusion

To return to Mosely's 1986 question, in the case of Picciotto (2009) unfortunately the data do deceive. Not a third, but only just over 10 per cent of the Bank country evaluations show a disconnect between aggregate project and strategy performance when subject to scrutiny (as against the one-third before scrutiny). And none of the remaining cases of disconnect provide any evidence at all of a macro-micro paradox.

Indeed it is clear, after closer examination, that use of the particular World Bank evidence base used by Picciotto, though innovative, is unlikely to pay dividends. The time period is too short (if aid has negative impacts, it is likely due to the cumulative impact of aid, not just due to aid given in that period), and all aid needs to be examined, not just aid from one donor. Indeed, discussion of the aggregate impact of aid, and possible negative effects, is markedly absent from most of the CAEs.

Does this mean that country assistance strategies and their evaluation are a waste of time? The appropriate level of evaluation is indeed a matter of debate. Easterly, for example, favours project level evaluations. What is needed, he argues, is 'not overall sweeping evaluations of a whole nationwide development program, but specific and continuous evaluation of particular interventions' (2006, p. 194). But others dispute this. Collier (2002, p.2) argues that: 'project-level performance is an inadequate instrument for attaining donor objectives,' that a project-level focus is both costly and ineffective, and that aid should be assessed by its ability to promote reform, which would presumably be at the country level.

We would argue that country assistance strategies and their evaluation are indeed important, even if they are unable to establish conclusively whether a micro-macro paradox exists. Individual aid interventions might succeed, but they may not be the most important interventions. Interventions might succeed but strategic gaps might lead to failure at the country level. And projects might succeed better if there were fewer of them, in fewer sectors. Projects might succeed, but governments might be nevertheless overwhelmed by them. These country-level health-checks are critical for effective aid delivery, but can be only carried out through some form of country-level evaluation. They will not emerge from intervention-level analysis.

Country-level assessments also serve as useful accountability tools for donors, in particular forcing them to wrestle with the unfortunate reality that in many recipient countries, projects succeed even

as nations fail. They are also useful tools for harmonization and alignment with aid recipient countries.

At the same time, a number of lessons arise from the analysis of this paper with regard to the design of country assistance strategies and their evaluation. The first is the old message that strategies should adopt more realistic objectives. These objectives should indeed be aligned with the country's own goals but should also be set at a level where it is possible for the Bank and possibly other partners (if a joint strategy) to show some direct contribution. For example, instead of an objective of universal access to maternal health facilities, it may be appropriate to aim for universal access in provinces where there are Bank-supported projects.

The second is that that country assistance evaluations should pay more attention to possible paradox transmission mechanisms. Nowhere in the CAEs that we examined did we see any attempt to examine whether aid had led to rent-seeking, or exchange rate appreciation, or a heavy administrative burden, for example.

The third is that a more careful and rigorous methodology is needed for defining aggregate project performance in the context of a country assistance evaluation. In our view, the IEG gets over half its project ratings wrong. For the reasons discussed above, it does not make sense to use a mechanical rating based on the average rating of closed projects, as the IEG seems to.

To close, our inability over the last quarter of a century to provide a definitive answer to the questions Mosely posed in 1986 raises the distinct possibility that we may never know whether aid does involve a macro-micro paradox. It might all be too complex for either cross-country regressions or a case-study approach to pick up. The systemic impacts of aid, to the extent they exist, are likely to accumulate slowly, over many years. There is no clear counter-factual. It is very hard to envisage what aid-dependent countries, where presumably paradox impacts are strongest, would look like in the absence of aid.

Given this irreducible uncertainty, perhaps the best advice to donors is to act as if the paradox is indeed real, and to constantly examine and guard against transmission mechanisms which might prevent the aggregate effect of aid for being less than the sum of its parts. To this end, both country assistance evaluations, and in-depth country case studies, which are more likely to have a longer time horizon and to be of all aid rather than just focused on aid from one donor, even if they cannot themselves shed light on whether the micro-macro paradox is real, can help reduce its possible effects.

Annex A: List of Country Assistance Evaluations (with dates, where possible, available at http://www.worldbank.org/ieg/countries/cae/completed_cae.html (viewed 22 January 2011))

Costa Rica November 16, 2000 (Report no. 21391)
Ecuador June 4, 1999 (Report no.)
Ethiopia November 30, 2000 (Report no.21450)
Ghana April 18, 2000 (Report no.20328)
Haiti February 12, 2002 (Report no.23637)
Jamaica December 21, 1998 (Report no. 19356)
Lesotho, 2002
Mexico June 28, 2001 (Report no. 22498)
Morocco, 2001
Nepal November 1, 1999 (Report no.19850)
Paraguay, 2001
Peru, September 25, 2002 (Report no. 24898)
Republic of Bulgaria March 7, 2002 (Report no. 23809)
Russian Federation September 23, 2002 (Report no.24875)
Ukraine November 8, 2000 (Report no. 21358)
Yemen, January 2001 (Report no. 21787)
Zambia November 7, 2002 (Report no. 25075)
Zimbabwe May 21, 2004 (Report no. 29058)

ANNEX B: Summary Table Evidence of Micro-Macro Paradox (cross referenced where possible with CAEs)

Micro-Macro Paradox Country	Country Assistance Strategy Performance Rating CAE	Aggregate Project Performance Rating	Agree/Disagree with categorisation	Project Performance (weighted by value where available)	Evidence to support/challenge a change in categorisation (Disagreements, where they exist, are with the project rating, except for Costa Rica.)
Morocco 1997 (1997-2000)	Marginally unsatisfactory	Satisfactory	Agree	Outcome rating of 57 per cent (Annex table 5). Likely sustainability 82 per cent.	.
Bulgaria 2002 (1991-2000)	Unsatisfactory	Satisfactory	Agree	Outcome rating of 95 per cent (para 2.16), Likely sustainability rating of 81 per cent.	
Costa Rica 2000 (1990s)	Unsatisfactory	Satisfactory	Disagree	Outcome rating of 100 per cent (based on 3 projects). Likely sustainability 100 per cent (1 project).	Costa Rica achieved significant progress towards the objectives of the Country Assistance Strategy. The primary driver of this was the government (para 4.8). Given this, the CAS should have been rated as moderately satisfactory or better. The CAE suggests that the Bank did not adequately consider timeframes for reform, but this is a criticism of the Bank, not an indication that CAS outcomes were not achieved..
Ecuador 1999 (1994-1998)	Unsatisfactory	Satisfactory	Disagree	Outcome rating from From 1991-99, 44 per cent by commitment (50 per cent by voume). (Annex A, Table A.1).	See performance ratings.

				Sustainability low at 42 per cent (para3.2)	
Haiti 2002 (1986 – 2000)	Unsatisfactory	Satisfactory	Disagree	Satisfactory ratings for 63 per cent Sustainability rating of 21 per cent.	The evaluation report notes that projects in Haiti have unusually low ratings for outcome, institutional development and sustainability (para 3.6). There were only two projects rated between 1994-97. There were suspensions in lending and projects (1991-94 and 1994-97) due to the coup and related political events. The DG's memo notes that "The efficacy of the Bank's program has been negligible, and its efficiency, low... The development impact of Bank assistance to Haiti since 1986 has been severely limited. Based on both its impact and the ratings of its individual components, the outcome of the assistance program is rated unsatisfactory (if not highly so)." This would seem to explicitly rule out a disconnect.
Jamaica 1999 (1993 CAS)	Unsatisfactory	Satisfactory	Disagree	Satisfactory outcomes for 49 per cent of projects if weighted, 51 per cent is unweighted. Likely sustainability of 44 per cent.	See ratings. CAE suggests that 'The poor performance of the project portfolio is attributed to overambitious designs that did not reflect implementation capacity '(para 2.18).
Lesotho 2002 (1994,1996 and 1998 CAS)	Moderately unsatisfactory	Satisfactory	Disagree	Satisfactory outcomes for 67 per cent. Likelihood of sustainability 22 per cent. (Table 3.4, p.22)	The evaluation report noted that World Bank programs were ineffective with the exception of education. (E.S, page xvi). Sustainability ratings extremely low
Mexico (1992-94)	Partially unsatisfactory	Satisfactory	Agree	Satisfactory outcomes in 87 per cent of value by commitment.	The evaluation report covers four CAS periods between 1989-2000. Of these, one period of two years was considered partially unsuccessful in relation to CAS performance.
Nepal 1999	Unsatisfactory	Satisfactory	Disagree.	Satisfactory	"As a consequence of a poor enabling environment, Bank

(1990s)				outcomes in 82 per cent of projects, but sustainability only 16 per cent.	projects had limited impact on their broader objectives; suffered from a range of implementation problems; and there are serious doubts about their sustainability. Frequent changes in key decisionmakers due to political instability, inadequate management, and lack of counterpart funds undermined project implementation and sustainability." (Memorandum from IEG DG in Nepal CAE). The report notes that 27 per cent of original commitments were cancelled, this was the highest in the region (para 2.4).
Paraguay 2001 (1993 and 1997 CAS)	Unsatisfactory	Satisfactory	Disagree	Satisfactory outcomes in 35 per cent of net commitments, however based on only three evaluations between 1991-1999.	There were only three project evaluations over the relevant period. This is too small a number to draw conclusions that project performance was satisfactory. In Of the 9 operations approved between 1992 and 1997, at the time of writing only one had closed, and five of the remaining eight were problem projects. The CAE also tells us that 'Projects approved since 1992 have had serious difficulties' and that in August 1999 the Paraguay portfolio was rated among the Bank's ten worst.
Peru 2003 (1997-2000)	Unsatisfactory	Satisfactory	Agree	Satisfactory outcomes in 98 per cent (projects approved between 1991-2002, weighted by commitment)	
Russia 2002 (1992-98)	Unsatisfactory	Satisfactory	Disagree	Satisfactory outcomes in 28 per cent weighted by commitment. (p.13) Likely sustainability of 78 per cent.	The good results in sustainability cannot counter the poor results in outcomes.
Ukraine 1999 (1992-1996)	Unsatisfactory	Satisfactory	No evidence	No project ratings were presented in	"For the 12 ongoing projects at the end of 1998, 3 were rated "at risk", 5 were rated as : "potentially at risk " and

				the Country Assistance Evaluation	only 4 were considered "non risky", (see Annex 5). The disbursement ratio is one-third of the Bank-wide average, and significantly below the ECA region average. (p.8).
Yemen 1999 (IDA from 1970 1996 CAS)	Marginally Unsatisfactory (IDA lending between 1990-95)	Satisfactory	Agree	Satisfactory outcomes in 81 per cent of projects (weighted by value). This is higher than the Bank wide average of 74 per cent.	
Zambia 2003 (1996-2001)	Unsatisfactory	Satisfactory	Disagree	Satisfactory outcomes for 61 per cent (weighted by commitment), and sustainability 29%	The CAE states in its conclusion that "Outcomes of many Bank operations, and of the overall program, were unsatisfactory.(p. 27) which appears to explicitly rule out a conclusion of a disconnect. The introduction gives the same message: "Outcomes of the Bank's program during this period are judged unsatisfactory, based on a "bottom-up" evaluation of the Bank's products and services (Chapter II) as well as a "top-down" assessment of aggregate economic and social indicators (Chapter III)."(p.1) Five structural adjustment credits made up three-quarters of closed commitments for the review period. Two of them were rated unsatisfactory. "In hindsight, even those adjustment operations initially rated marginally satisfactory have had less robust outcomes than expected." (p. 15)

Zimbabwe 2003 (1990-2000, 1992, 1994 and 1997 CASs)	Unsatisfactory	Satisfactory	Agree	Satisfactory or moderately satisfactory outcomes for 81 per cent (weighted by net commitments)	The two structural adjustment credits “did not achieve their major objectives”, leaving at most 55% of projects by commitment satisfactory. The great bulk of these (by number and volume) have marginally (or moderately) satisfactory ratings, and have unlikely or unclear sustainability.
Ghana 2000 (1995 and 1997 CASs)	Satisfactory	Unsatisfactory	Disagree	Satisfactory outcomes for 78 per cent (by number). Sustainability 45 per cent.	CAE also comments that “the portfolio of ongoing projects is rated as very satisfactory” (Memo from IEG DG)
Ethiopia 1999 (1990s)	Satisfactory	Unsatisfactory	Disagree	Satisfactory outcomes for 79 per cent (by number), and sustainability 53 per cent	See ratings, which show strong performance compared to Bank wide averages . The CAE concludes that ‘the Bank’s lending and non-lending assistance have been well directed. Significant progress has been made in many areas...The Bank-supported projects have performed well by African and Bank standards.’ (p. 8) The very large (\$150m) Emergency Recovery and Reconstruction Credit is rated highly satisfactory, as was the one structural adjustment credit for \$250 m, so 86% of projects are satisfactory by volume (Table 5b).
Russia 2002* (1999-2001)	Satisfactory	Unsatisfactory	Agree	Satisfactory outcomes in 28 per cent weighted by commitment.	Project performance poor (see earlier discussion of Russia)

Annex C: References

- Arndt, C., Jones, S. and Tarp, F. (2009) 'Aid and Growth: Have We Come Full Circle?' *Discussion Paper No. 2009/05* United Nations University
- Boone, P. (1995) 'Politics and the Effectiveness of Foreign Aid', *Centre for Economic Performance Discussion Paper No 272*, London School of Economics and Political Science
- Collier, P. (2002) 'Making Aid Smart: institutional incentives facing donor organizations and their implications for aid effectiveness' http://pdf.dec.org/pdf_docs/Pnacp768.pdf
- Deaton, A. (2009) 'Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development', Research Program in Development Studies, Center for Health and Wellbeing, Princeton University
- Doucouliafos, H., Paldam, M. (2009) 'The Aid Effectiveness Literature: The Sad Results of 40 Years of Research', *Journal of Economic Surveys* (2009) Vol. 23 No. 3 pp- 433-461
- Doucouliafos, H., Paldam, M. (2008) 'Aid effectiveness on growth: a meta-study', *European Journal of Political Economy*, 24 (2008) 1-24
- Dreher, A., Nunnenkamp, P. and Theile, R. (2006) 'Does Aid for Education Educate? Evidence from Panel Data', No 146 ETH
- Durlauf, S., Johnson, P. and Temple, J. (2005) 'Growth Econometrics' Chapter 8 in Aghion, P and Durlauf, S.N. *Handbook of Economic Growth*, Elsevier B.V.
- Easterly, W. (2006) *The White Man's Burden: why the West's efforts to aid the rest have done so much ill and so little good*, The Penguin Press.
- Easterly, W. (2009) 'The Indomitable in Pursuit of the Inexplicable: The *World Development Reports*' Failure to Comprehend Economic Growth Despite Determined Attempts, 1978-2008' Commentary in Yusuf, S. at al *Development Economics through the Decades: a critical look at 30 years of the World Development Report*.
- Lawson, A., Booth, D., Msuya, M., Wangwe, S. and Williamson, T. (2005) *Evidence from Tanzania*, Overseas Development Institute and Daima Associates, Dar es Salaam
- McGillivray, M., Feeny, S., Hermes, N., Lensink, R. (2006) 'Controversies over the impact of development aid: it works; it doesn't; it can, but that depends', *Journal of International Development* 18: 1031-1050
- Michaelowa, K., Weber, A. (2007) *Aid Effectiveness in Primary, Secondary and Tertiary Education*, University of Zurich, Background paper for Education for All Global Monitoring Report 2008
- Mishra, M. and Newhouse, D. (2007) 'Health Aid and Infant Mortality', IMF Working Paper wp.07/100
- Mosely, P. (1986) 'Aid Effectiveness : the micro-macro paradox' *Institute of Development Studies Bulletin* 17: 214-225
- Mosely, P., Hudson, J. and Horrell, S. (1987) 'Aid, the Public Sector and the Market in Less Developed Countries', *The Economic Journal* 97 pp 616-641

Moss, T., Pettersson, G. and van de Walle, N. (2008) 'An Aid-Institutions Paradox? A Review Essay on Aid Dependency and State Building in sub-Saharan Africa' Chapter 8 in Easterly, W. (ed.) *Reinventing Foreign Aid*, MIT Press.

Picciotto, R. (2009) 'Development Effectiveness: An Evaluation Perspective', Chapter 8 in Mavrotas, G. and McGillivray, M. (eds) *Development Aid a Fresh Look*, United Nations University

Riddell, R. (2007) *Does Foreign Aid Really Work*, OUP.

Roodman, D. (2007) 'Macro Aid Effectiveness Research: A Guide for the Perplexed', *Centre for Global Development Working Paper 134*

Roodman, D. (2008) 'Through the Looking Glass, and What OLS Found There: On Growth, Foreign Aid and Reverse Causality', *Centre for Global Development Working Paper 137*

Stern, E., Altinger L., Feinstein, O., Marañón, M., Ruegenberg, D., Schulz, N., Neilsen, N., (2008) 'Thematic Study on the Paris Declaration, Aid Effectiveness and Development Effectiveness', Development Assistance Research Associates, November 2008

Quibria, M.G. (2004) 'Development Effectiveness: What Does Recent Research Tell Us', ADB OED Working Paper No.1 October 2004

World Bank (2005) *Country Assistance Evaluation Retrospective: An OED Self-Evaluation*, World Bank, Washington

World Bank (2009) *Country Assistance Strategies: Retrospective and Future Directions*, World Bank, Washington