Aid and Growth

Have We Come Full Circle?

Arndt, Channing; Jones, Edward Samuel; Tarp, Finn

Discussion Papers
Department of Economics
University of Copenhagen

No. 09-22

Aid and Growth: Have We Come Full Circle?

Channing Arndt, Sam Jones,
and Finn Tarp

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (online)

Aid and Growth
Have We Come Full Circle?

Channing Arndt (Development Economics Research Group (DERG), Department of Economics, University of Copenhagen, e-mail: channing.arndt@econ.ku.dk)

Sam Jones (Development Economics Research Group (DERG), Department of Economics, University of Copenhagen, e-mail: sam.jones@econ.ku.dk)

Finn Tarp (Development Economics Research Group (DERG), Department of Economics, University of Copenhagen and UNU-WIDER, Helsinki, e-mail: tarp@wider.unu.edu)

October 2009

Abstract

The micro-macro paradox has been revived. Despite broadly positive evaluations at the micro and meso-levels, recent literature has turned decidedly pessimistic with respect to the ability of foreign aid to foster economic growth. Policy implications, such as the complete cessation of aid to Africa, are being drawn on the basis of fragile evidence. This paper first assesses the aid-growth literature with a focus on recent contributions. The aid-growth literature is then framed, for the first time, in terms of the Rubin Causal Model, applied at the macroeconomic level. Our results show that aid has a positive and statistically significant causal effect on growth over the long run with point estimates at levels suggested by growth theory. We conclude that aid remains an important tool for enhancing the development prospects of poor nations.

Keywords: foreign aid, growth, aid effectiveness, causal effects

JEL classification: O1, O4, F35, C21

Acknowledgements

We thank Tony Addison, Pranab Bardhan, Imed Drine, Paul Isenman, and Alan Winters for encouragement and valuable comments. Thanks are also due to Tseday Jemaneh Mekasha for excellent research assistance and to Raghuram G. Rajan and Arvind Subramanian for sharing their original data and STATA files. The usual caveats apply.

1 Introduction

The extent to which foreign aid can be a decisive factor in the economic development of low income countries remains controversial. In 1987 Paul Mosley suggested that while aid seems to be effective at the microeconomic level, any positive aggregate impact of aid is much harder to identify (Mosley 1987). He labelled this the micro-macro paradox, and it challenged the conclusions of the seminal works by Papanek (1972, 1973). Now, after more than twenty years, Rajan and Subramanian (2008) (hereafter RS08) conclude 'it is difficult to discern any systematic effect of aid on growth'. At the same time, microeconomic evaluations, including rigorous contributions to the programme evaluation literature by development economists, remain largely positive.[1] Thus, after two decades of intense analytical work using new theory, new data and new empirical methodologies, it would appear that the micro-macro paradox has been revived. Other similarities with the late 1980s and early 1990s exist, not least with respect to policy. In 1994, the Economist magazine concluded from the results of Boone (1994) that 'Aid [goes] Down the Rathole'. Today, in the midst of a serious global economic crisis, where aid is arguably more needed than ever, the attention of both the aid community and decision-makers is on 'Dead Aid' (Moyo 2009), which argues for a complete cessation of aid flows to Africa.

This paper has two main objectives. First, we attempt to provide a balanced assessment of the aid-growth literature, focusing on recent contributions. We observe that, while aid may have very high returns at times,[2] there is an emerging consensus that expectations surrounding the average potency of aid have been too high. Second, we turn our attention to the fundamental empirical evaluation challenge: identifying the counterfactual. Using observational data, there is no way of identifying a plausible counterfactual without making assumptions that are bound to be debatable, in theory and in practice. We add new insights by framing the aid-growth debate in terms of potential outcomes (counterfactuals) taking motivation from the programme evaluation literature. This helps clarify the conditions required for valid causal inference, and it motivates the application of robust empirical methods that have not been employed in the literature to date.

Our empirical analysis starts from the RS08 contribution. As noted, these authors do not find a significant long-run effect of aid on growth, leading to pessimism with respect to the ability of aid to contribute to economic growth in poor countries. Despite this conclusion, their long-run results point to a positive effect that is consistent with (and insignificantly different from) predictions of the causal effect of aid on growth that they derive from a simple neoclassical growth model. Following reproduction of the RS08 results, we progressively develop a counterfactual framework aimed at analyzing macroeconomic phenomena, including an extension of the double robust methodology of Robins and Rotnitzky (1995) along with other methodological modifications. We

---

1 The most rigorous evaluations in this area are undertaken by the World Bank, and reports from the Independent Evaluation Group (IEG) of the World Bank are encouraging (World Bank 2008); see also, Mosley (1987) and Cassen and Associates (1994). For discussion of the contribution of randomized field experiments, see Banerjee and Duflo (2009).

2 In fragile states and certain post-conflict situations foreign aid can play a decisive role in recovery and economic development (Collier and Hoeffler 2004).

find that the long-run positive effects implicit in the RS08 results are strongly reinforced and remain reasonably robust.

The remainder of this paper is structured as follows. Section 2 provides a literature review with emphasis on the recent literature. Section 3 presents the aid-growth issue through the lens of the counterfactual framework developed in the programme evaluation literature. Section 4 contains an application starting from the principal results obtained by RS08 and then casting the aid-growth problem in terms of the programme evaluation literature. The final section summarizes and concludes.

2 Literature review

This section provides a review of the voluminous literature on the relationship between aid and growth. Given the existence of various surveys (e.g., Tsikata 1998; Hansen and Tarp 2000; Easterly 2003; Kanbur 2006; Roodman 2007; Thorbecke 2007), primary attention is placed on the most recent studies not surveyed elsewhere. We also provide a summary of our view of the current state of the literature.

2.1 Earlier generations

Previous studies of the aid-growth relationship can be classified into three generations, each influenced by dominant theoretical paradigms as well as available empirical tools. The first two generations were inspired by relatively simple models of the growth process such as the Harrod-Domar model and the two-gap Chenery-Strout extension. The underlying idea behind the Harrod-Domar model is of a stable linear relationship between growth and investment in physical capital. Assuming all aid is invested, it is straightforward to calculate how much aid is required to achieve a target growth rate. The impact of aid is assumed to be positive and helps plug either a savings or a foreign exchange gap. Empirical studies in this tradition consequently focused on the extent to which aid increases savings and investment in recipient countries (Papanek 1972, 1973). As the detailed survey in Hansen and Tarp (2000) testifies, first generation studies generally concluded that aid does tend to increase total savings, but not by as much as the aid flow. Quite reasonably, this suggests a non-negligible proportion of aid is consumed rather than invested.

Retaining the focus on capital accumulation, the second generation of literature moved on to explore the impact of aid on growth via investment. Using data for a cross section of countries, a large number of studies of this kind were produced during the 1980s and early 1990s. Hansen and Tarp (2000) conclude that the findings from these studies consistently indicate a positive link between aid and investment. The more striking result, however, is that many studies do not establish a clear positive relationship between savings and growth across countries over time. This puzzle motivated doubts concerning the appropriateness of the underlying growth model and the techniques used for empirical analysis in these studies. Indeed, it is a tall order to expect both a constant output-capital relationship and that all aid is invested. A second line of critique of the Harrod-Domar and two-gap approach has been the argument that growth is less related to physical capital investment, including aid, than often assumed (Easterly 1999, 2003). If the productive impact of aid depends more on incentives and relative prices, as well as the policy environment more generally, then it becomes important to consider these potential wider effects. The second generation of studies also introduced the problem that poorly performing countries may receive more aid precisely *because* of their poor growth performance (e.g., Mosley et al. 1992). Empirical analyses that do not account for the endogeneity of aid will not reveal aid's causal impact. Many second generation studies, however, did not deal with this issue, which partly explains some of their puzzling results.

From the early 1990s a third generation of more sophisticated econometric studies came to dominate the academic and public discourse about aid. This was motivated by the availability of better data, allowing analysts to look at changes both across and within countries over time (i.e., panel data became available). Insights from new theories of economic growth, as well as a rapidly increasing number of general empirical growth studies, also influenced the research agenda. Mindful of the weaknesses of previous studies, the aid–growth relationship came to be perceived as (possibly) non-linear and the endogeneity of aid was taken more seriously. Among the numerous studies of this generation, a leading paper that came to exert a significant influence on aid policy is the study by Burnside and Dollar (1997, 2000). The authors argued that although aid has no impact on growth on average, it can work as long as recipients pursue 'good' policies. In their words: '… aid has a positive impact on growth in developing countries with good fiscal, monetary and trade policies ... [but] ... in the presence of poor policies, aid has no positive effect on growth' (2000: 847).

These results were subject to substantial criticism and were shown to be fragile. For example, Hansen and Tarp (2001) found early on that a story of diminishing returns to aid, captured by a squared aid term, best captured the non-linear relationship between aid and growth and is the empirical specification with most support in the data. In a later contribution Easterly et al. (2004) added that the Burnside-Dollar aid-policy result is fragile when the dataset is expanded (by years and countries). Dalgaard et al. (2004) found that aid has been far less effective in tropical areas over the last 30 years; but they also stressed that it is hard to believe that aid should be inherently less potent in the tropics. Thus, the real explanation for the aid-tropics link likely lays elsewhere and the authors called for further research to help disentangle the channels through which aid matters for productivity. In an empirical review of these contributions, Roodman (2007) argues that the results of this generation are extremely sensitive to methodological choices, concluding that while some aid is likely to increase investment and growth, aid 'is probably not a fundamentally decisive factor for development' (2007: 275). Moreover, due to multiple kinds of aid, and differences in the efficiency with which it may be put to use, the noise in the data may mask any valuable information regarding the causal impact of aid.

2.2 Recent studies

Recently, a fourth generation of literature has emerged. A distinctive aspect of this literature is that it starts from the proposition that foreign aid's aggregate impact on economic growth is either non-existent or negative. A leading paper, which is widely taken to establish this result, is RS08. Their approach and principal findings are discussed and reproduced in Section 4. In short, these authors do not find any systematic effect of aid on growth and this conclusion appears to hold for different estimation approaches, for different time periods and for different types of aid.

Following these findings, and as explicitly advocated in RS08 (2008: 660), a prevailing research theme is to explain why these negative or negligible results appear. The leading explanation for negative effects from aid refers to political economy dynamics, whereby aid inflows undermine or weaken governance, for example, by increasing the returns to corruption and/or increasing rent seeking activities. Djankov et al. (2008) argue that aid has analogous effects to a natural resource curse. Their core result is that foreign aid has a statistically significant negative effect on changes in political institutions (specifically democracy) and this effect is larger in magnitude than that caused by natural resource windfalls. Similarly, Rajan and Subramanian (2007) find that the rate of growth of value added by the manufacturing sector in developing countries has been undermined by a detrimental effect of aid inflows on governance quality.

Recent theoretical contributions have highlighted the need for modest expectations with respect to the magnitude of the impact of foreign aid on growth. Based on a standard neoclassical Cobb-Douglas production function framework, and assuming that aid only augments physical capital investment and has no effect on productivity, RS08 estimate that a one percentage point increase in the ratio of aid to GDP should be expected to raise the growth rate by around 0.16 percentage points on average. However, a sober assessment dictates that at least some aid is directed towards consumption or non-growth enhancing activities. As a result, Rajan and Subramanian place the expected growth return at around 0.1 percentage points for each percentage point of aid in GDP. Thus, the implied increase in the growth rate accruing from aid inflows at 10 per cent of GDP should be about 1 per cent. A higher impact is possible if aid also enhances productivity. A lower bound elasticity of 0.1, however, implies that increments to growth due to aid may be difficult to distinguish from business cycle fluctuations and internal/external shocks over short time frames. Indeed, as will be discussed in more detail below, factors such as data quality and the long time frames associated with many aid investments point to the need for three decades or more of elapsed time.

Using a more sophisticated theoretical approach, namely an augmented Solow-Swan growth model, Dalgaard and Erickson (2009) find that the predicted increase in the *level* of GDP per capita accruing from aid inflows to sub-Saharan Africa over the past 30 years is between 4 per cent and 7 per cent, depending on the assumed share of capital in value added. This converts to a considerably smaller elasticity of aid to growth than the value of 0.1 posited by RS08. Dalgaard and Erickson conclude that, viewed through the lens of a neoclassical growth model, the disappointing growth performance of Africa does not necessarily imply an aid effectiveness puzzle. Rather, expectations with respect to the potency of aid have been systematically too high.

Finally, a prominent feature of third generation studies was the use of (cross-country) panel data comprising time period averages of four or five years. Although either simple OLS or two stage least squares (2SLS) estimators were initially applied to these panels (e.g., Burnside and Dollar 2000), later contributions have employed dynamic panel GMM methods (exemplified by the Arellano-Bond and Blundell-Bond procedures). The advantages of GMM panel methods in this context are threefold. They (i) account for unit-level fixed effects, (ii) incorporate internal methods for dealing with endogenous regressors, and (iii) avoid the bias of standard panel estimators in dynamic settings. The last characteristic is important as any panel regression of growth on lagged income arithmetically derives from an autoregressive specification, which introduces substantial bias in the presence of unobserved unit-specific heterogeneity especially where the number of time periods is small. The use of transformations of lagged variables as internal instruments for the endogenous variables may also seem to be an attractive alternative to finding external instruments that remain valid and robust across all panels. Roodman (2009) notes that as a consequence of these properties, as well as the automation of panel GMM techniques in popular statistical packages, these methods have proliferated since the late 1990s. Among these, greater preference has been given to the Blundell-Bond 'system' GMM estimator as it is considered to be more efficient than the earlier Arellano-Bond 'first difference' estimator.

While dynamic panel GMM methods can be extremely useful, growing concerns over their application are particularly pertinent in the present case. First, the use of internal instruments is no guarantee of their strength or validity, particularly where there is substantial time series persistence. The concern that weak instruments typically bias estimates towards their unadjusted counterparts (e.g., OLS or panel fixed effects estimates) applies as much to panel GMM as to cross-section estimators. Indeed, Bun and Windmeijer (2007) have shown that the weak instrument problem, previously attributed mainly to the Arellano-Bond estimator, may be equally problematic in the system approach.

Perhaps more critically, for the Blundell-Bond estimator to be valid, both country fixed effects and omitted variables must be orthogonal to the lagged differences of the right-hand side (RHS) variables which are used as instruments for the level equation. This assumption cannot be tested and may be suspect given (i) the highly complex nature of the growth process, and (ii) that country fixed effects are expected to incorporate determinants of steady-state income levels that may correlate with growth along individual countries' steady-state transition paths. In a Monte Carlo investigation of the robustness of different panel estimators, this led Hauk and Wacziarg (2009) to conclude that the principle issue for system GMM is not one of strong or weak instruments but rather the validity of these moment conditions. Roodman (2009) also alerts that the Blundell-Bond estimator may give a false sense of certainty as a large number of internal instruments can over-fit the endogenous variables and weakens the Hansen/Sargan tests for instrument validity. Furthermore, internal instruments do not prevent bias arising from measurement error in the endogenous regressors. As indicated, measurement error is not immaterial in the context of aid-growth regressions.

A related issue, alluded to above, is the appropriate time-frame over which any growth effects accruing from aid (and other RHS regressors) are expected to materialize. Recently, attention has been given to long-run determinants of growth that have a cumulative but often not immediate impact on the rate of income growth. For example, changes in education move only slowly at the aggregate level and exert a positive influence on economic growth with a substantial lag. This derives from simple demographics whereby improvements in schooling indicators, for example at the primary level, can take many years to translate into noticeable increases in average education levels among working age adults. Changes in human capital due to improved health indicators may take even longer to translate into more rapid economic growth. Ashraf et al. (2008) and Acemoglu and Johnson (2007) find that the initial economic impact of gains in life expectancy from disease eradication may be a reduction in *per capita* incomes due to the increase in population and dependency ratios. The former authors find that it takes 30 to 40 years for per capita incomes to return to pre-eradication levels. They also find that significant increases in life expectancy at birth (from 40 to 60 years) only begin to have a modest positive effect on incomes after about a 35 year lag.

The simulation model employed by Ashraf et al. (2008) focuses on demographic impacts and resulting dependency ratios, capital/labour ratios and land/labour ratios. If, in addition to being a major cause of childhood death, diseases like malaria pose significant constraints to the growth of industries like tourism (by depressing demand) and food processing (by complicating recruitment of skilled labour to rural areas), the economic benefits of malaria eradication would be larger and would accrue more quickly. Furthermore, as Asharaf et al. point out, complementary policies such as population control and enhanced investment levels (potentially through foreign aid) could speed the realization of benefits from disease eradication and life expectancy improvement. Nevertheless, the point remains that for some important types of aid, the realization of growth benefits may require a full generation of elapsed time. From a causal evaluation perspective, the implication is that aid can have persistent effects, meaning that the exclusion restrictions assumed in GMM dynamic panels (that lagged levels of aid do not have an independent effect on outcomes) may be invalid. Also in contrast to GMM panel settings, it may be more meaningful to investigate the aid-growth relationship over much longer time horizons (e.g., 30-40 years).

In response to these concerns, alternative approaches are being explored. One strand of recent literature emphasizes the need to 'open the black box' (Bourguignon and Sundberg 2007) incorporating political economy aspects. Others have pursued non-growth (meso-level) aid outcomes, pointing to the multi-dimensional objectives of aid. Mishra and Newhouse (2007), for example, uncover a small but statistically significant effect of health aid on infant mortality. Masud and Yontcheva (2005) also find that aid helps reduce infant mortality, but this effect is only significant for aid provided by non-governmental organizations (NGOs) rather than bilateral aid. Easterly (2009) documents substantial improvements across a wide range of social indicators in sub-Saharan Africa since 1960, but he does not relate these improvements to aid. With few exceptions (e.g., Sachs 2005, 2006), findings at the meso-level have not been deployed to argue for aggregate aid effectiveness. This is despite increasing evidence that outcomes at this level do have substantial macroeconomic effects (Cohen and Soto 2007).

Another strand has adopted long horizon cross-section methods in place of (dynamic) panel approaches. This includes RS08 who, among other methods, look at the relationship between long-run averages of aid and growth, conditional on other covariates, using data spanning from 1960 to 2000 as well as shorter time frames. This approach echoes the average OLS estimator proposed by Mankiw et al. (1992) (Hauk and Wacziarg 2009) as well as the long difference approach used in the Acemoglu and Johnson (2007) analysis of the long-run effect of health on income in developing countries.

Although cross-country studies continue to dominate published research, there has been an increasing demand for case studies as a useful supplement (Collier and Gunning 1999). In this spirit, Arndt et al. (2007) provide a comprehensive case study of Mozambique in which they attempt to evaluate the effect of aid on different proximate drivers of growth. Starting with long-run growth accounting estimates, they find that aid has played a critical role in rebuilding infrastructure and expanding access to health and education. While aid has supported rapid reconstruction and has crowded-in private investment, it also has generated important governance and economic management challenges; thus there is no guarantee that higher growth associated with aid will be sustained over the long-term. This provides supporting evidence for the Collier and Hoeffler (2004) argument that aid can be particularly beneficial in post-conflict environments.

To summarize, a number of points of consensus can be highlighted from the recent literature. Motivated from neoclassical growth theory and ruling out negative effects by assumption, the expected impact of aid on growth is positive but small. It is certainly less than the values implied by constant incremental capital to output ratios. Also, to the extent that aid supports gains in health indicators that lead to a rise in population and the dependency ratio, the expected effect of this channel on per capita income may be negative in the 'short-term' and this may persist for three decades or more. Echoing themes from the third generation of the aid-growth literature, measurement issues and the endogeneity of aid complicate matters further. Low quality data and weak instruments bias the results towards zero (Tarp 2006); and the tendency to allocate aid towards poor growth performers leads to an unconditional negative association between aid and growth. Dynamic panel estimators, which rely on internal instruments, do not provide a definitive answer to these problems. As Deaton (2009) has argued, few studies have dealt with the endogeneity of aid in a convincing manner. The supply side approach of RS08 likely represents the state of the art, but even this approach has its weaknesses (discussed in Section 4).

Given the difficulties inherent in deriving robust causal conclusions from macro data, some scholars suggest we should only analyse micro- and meso-effects (e.g., Riddell 2007). While this view has foundation, it is apparent that policy-makers and the wider public continue to ask the broad question: does aid support economic growth? The present authors consider that this issue remains valid and cannot be disregarded. Without a substantial outward shift in the production possibilities frontier, few development objectives – including poverty reduction – are achievable. Thus, growth is a key objective of aid and must be evaluated as such. Research into the extent to which aid has and can achieve this high-level goal is critical to achieving better *overall* aid effectiveness and to validate the allocation of aid across countries and sectors within countries. Abandoning such questions only leaves them open to speculative and potentially unhelpful contributions.

3 A counterfactual model for evaluating aid

3.1 Basic framework

In this and the next section, we couch the aid-growth debate in an explicit causal framework. This forces the researcher to be rigorous regarding the counterfactual of interest as well as the conditions required to move from associational to causal inference. We take as a starting point the programme evaluation literature, carefully reviewed by Imbens and Wooldridge (2009), see also Blundell and Costa Dias (2008), and Angrist and Pischke (2008). This literature considers the effect of exposure of a unit to a treatment on an outcome variable, the objective being to isolate the causal effect of the treatment. In a series of seminal papers, Rubin (e.g., 1974, 1976, 1978) elaborated what has become known as the Rubin Causal Model (hereafter RCM; Holland 1986). In this model, causal effects are determined by comparing *potential outcomes* according to different levels of exposure to the treatment. The framework encompasses randomized experiments, where development economists have been particularly active in recent years. However, due principally to the nature of economic data and studies, which are often not amenable to randomization, economists normally apply the framework to observational data.

Important components and concepts of the RCM are sketched in the following paragraphs. To begin, we consider the RCM for the case of a binary treatment variable. This suggests two potential outcomes for unit *i*, where *i=1,…,N*, are considered. The first, denoted by $Y_i(0)$, corresponds to the potential level of the outcome variable in the absence of treatment. The second, denoted by $Y_i(1)$, corresponds to the potential outcome in the presence of treatment. Because the unit can either be treated or not, only one of these outcomes is actually observed. However, prior to the assignment of treatment, both potential outcomes are possible; thus, regardless of treatment assignment, the unrealized outcome is the counterfactual. An additional element of the RCM is the treatment assignment mechanism. This may be random, as in the random experiments literature; criteria based, as is frequently the case for job training programmes; or in large measure *ad hoc*, as in the allocation mechanisms for foreign aid viewed across providers.

Within the RCM, several conditions are necessary for meaningful causal inference. The essential idea is that where these hold, counterfactual outcomes can be validly inferred from the available sample, thus allowing the treatment effect to be estimated. These conditions are:

SUTVA (stable unit treatment value assumption). The value of outcome Y for unit *i* when exposed to treatment will be the same regardless of the mechanism used to assign the treatment to *i* and regardless of the treatments received by other units. This is analogous to the assumption of identically and independently distributed errors in a regression framework.

Unconfoundedness. The probability of treatment is conditionally independent of the potential outcomes given a vector of covariates X. This can be stated more formally as:

$$W_i \perp (Y_i(0), Y_i(1) | X)$$

where $W_i = 1$ if unit *i* is treated and zero otherwise.

Overlap in distributions. This relates to the covariates *X* and the specification of the functional relationship between the covariates and the outcome variable. While a linear relationship may be accurate locally near the average of the covariates, the linear approximation may not be accurate globally. If the means of the covariates between the treated and control groups differ substantially, misspecification of the functional relationship can lead to severe bias in the estimated treatment effect. In other words, the ideal is that the treatment and control groups be as similar as possible in all respects except the treatment.

Where exposure to the treatment is randomized or is (weakly) unconfounded conditional on observed covariates and there is acceptable overlap in distributions, the above framework applies straightforwardly. However in many cases, including that of aid and growth, the treatment is widely taken to depend upon potential outcomes, particularly over time. Unobserved or endogenous selection into the treatment group violates the requirements and motivates instrumental variable (IV) methods. Where IVs are used, analogous versions of the above requirements can be set out (Angrist et al. 1996). Critically, instruments should operate as random components in the treatment assignment rule with a non-zero causal effect on treatment status (*W*) but also with no independent effect on outcomes. Using the standard econometric vocabulary, these requirements imply that the instruments should be relevant and valid (i.e., exclusion restrictions hold). These are stringent criteria and demand that the instruments are uncorrelated with all omitted variables in the outcome equation of interest (Murray 2006). Furthermore, even if all these conditions are satisfied, the instrument may only effectively identify a subpopulation in the presence of heterogeneous treatment effects, suggesting the relevance of a local average treatment effects (LATE) interpretation (Imbens and Angrist 1994; Angrist 2004).

Careful examination of the extent to which chosen instruments meet the above requirements is critical in applied work. Weak instruments tend to be biased towards OLS estimates and may be inconsistent (Chao and Swanson 2005). Consistency proofs of IV estimators depend on asymptotic properties, but bias from using weak instruments may be acute in small samples (Choi and Phillips 1992). Also, the standard errors on IV estimates generally are larger than those in OLS and rapidly increase in size as the correlation between the instruments and the endogenous variables diminishes. Thus, in the presence of weak instruments, little may be gained from an IV approach and there is a risk of committing a Type II error arising from either the bias or large standard errors of the estimates.

3.2 The programme evaluation literature and the aid-growth debate

The previous sub-section outlined essential aspects of the potential outcomes approach. While there are innumerable parallels with standard econometric practices, including use of OLS and IV estimators, the framework retains distinctive practical and theoretical features. Specifically, interpretation of regression coefficients as causally meaningful parameters should be undertaken with extreme care, even in the presence of plausible instruments; and one must explicitly attend to the extent of overlap in the distributions between treatment and control groups. Turning to the aid-growth debate, it remains to be established that a counterfactual framework can be meaningfully applied. In the evaluation literature, microeconomic units such as individuals exposed to labour market programmes are often the focus. However, there is no inherent reason to restrict application to units at the microeconomic level. Indeed, Imbens and Wooldridge (2009) are explicit that units can be regions or countries. For the purposes of evaluating aid, however, a potentially more pertinent limitation is the predominant focus of the literature on binary as opposed to continuous treatment variables. This reflects primarily the state of the art within the evaluation literature as a whole. Imbens and Wooldridge (2009) note that, compared with the binary case, much less is known about settings with continuous treatment variables even though such settings are common in practice. Hence, the aid-growth debate is one application in ‘a [methodological] area with much ongoing work and considerable scope for further research’ (2009: 73).

Motivation can be taken from a well developed area of application of evaluation techniques. There are numerous parallels between the problem of estimating the causal impact of aid on growth and the causal impact of schooling on earnings. Looking generally, both problems are likely to be characterized by endogenous selection (selection bias), heterogeneous treatment responses (ability bias), and mis-measurement of treatment input (both in terms of quality and quantity). Furthermore, in both cases, the treatment is manifestly continuous; and the parallels continue. Leading applications in the returns to education literature rely on supply side factors to identify the underlying causal mechanisms (Card 2001). This is the main innovation introduced into the aid and growth literature by RS08. In addition, the units considered in the returns to education literature have all received some treatment. Hence, research focuses on the returns to incremental treatment. This is also true of the samples typically employed in the aid-growth literature. In most cases, all countries in the sample have received some foreign assistance.

Moreover, the returns to education literature points to the need for care in the interpretation of aid and growth regressions. While the aid and growth literature is voluminous, the returns to education literature is even more so. Considerable effort has been expended in the analysis of large high-quality datasets by some of the most skilled econometricians in the profession. Even so extensive debate persists with respect to the net bias of ordinary least squares (OLS) estimates of returns to education. As stated by Card, 'If one assumes on *a priori* grounds that OLS methods lead to upward biased estimates of the true causal effect of schooling, the even larger IV estimates obtained in many recent studies present something of a puzzle' (2001: 1155). It is, therefore, hardly surprising that estimates of the returns to aid remain controversial.

4 Application

In this section, we apply the counterfactual framework to the aid-growth debate. Our starting point is RS08, which is a recent, thoughtful and influential attempt to address the broad empirical question of interest – whether aid has a causal impact on economic growth over time. We begin with a brief summary of the RS08 empirical strategy and results before applying the insights from Section 3 to the same data. Our primary focus is on the average relationship between aid and growth. We do not provide a detailed discussion of other growth determinants, which are broadly in line with the extant literature.

4.1 Rajan and Subramanian

The stated objectives of RS08's study are to: (i) provide a comprehensive examination of the aid-growth relationship; and (ii) carefully deal with the endogeneity of aid. The latter leads the authors to deviate from a reliance on the dynamic panel GMM methods that dominated the third generation aid-growth literature. For many of the reasons discussed in Section 3, RS08 indicate their preference for external instruments that apply over long time horizons. Their instrumentation strategy, which echoes Frankel and Romer (1999) in the context of trade flows, is based on the supply-side characteristics of donors. Specifically, the authors suggest that bilateral aid from donor d to recipient r over time horizon h, where $h \in H$ = {1960–2000, 1970–2000, 1980–2000, 1990–2000}, is a function of colonial characteristics, relative population sizes and their interaction terms. Taking fitted values at the bilateral donor-recipient level from a preliminary stage regression (RS08 Table 3) for each recipient, they aggregate across donors to give a fitted Aid/GDP ratio in period h. Treating each period separately, standard 2SLS cross-section regressions are then undertaken where the outcome equation takes the form of:

(1) $$Y_i = \alpha W_i + X_i' \beta + \varepsilon_i$$

where Y_i is average real (PPP-adjusted) per capita GDP growth over the entire period; W_i is the Aid/GDP treatment variable, instrumented by the generated regressor from the preliminary stage regression; and X_i represents a vector of control variables reflecting previous contributions to the literature (e.g., Burnside and Dollar 2000).

Results from the two longest periods encompassed by these cross-section regressions are replicated in Table 1, where the estimate of α (the coefficient on the Aid/GDP ratio) is presented in the first row denoted 'treatment effect'. These are the core results of the RS08 study. In both cases, the generated instrument appears reasonably strong according to conventional measures such as the first stage partial F-statistic. The coefficients on Aid/GDP are small and positive, but also insignificant. This leads RS08 to conclude that there is no systematic (causal) effect of aid on growth; in turn, this is shown to hold when shorter sub-periods (RS08 Table 4), alternative growth horizons (RS08 Table 6), non-linear effects (RS08 Table 7) and different types of aid (RS08 Table 8) are considered. Similarly, the same basic results emerge when the question is considered in a dynamic panel setting (RS08 Table 10).

In light of the discussion in Section 3, we agree that long horizon cross-section estimates, using external instruments to address the endogeneity of aid amount to a potentially meaningful basis for exploring the long-run effects of aid. Even so, aside from the concern about aid's endogeneity, RS08 make no reference to a potential outcomes framework and the associated range of difficulties involved in moving from associational to causal inferences. Moreover, given how little is known about the growth process and the widely agreed fragility of past aid-growth results, the possibility of specification errors cannot be ruled out. In contrast, these concerns are central to the potential outcomes framework via its emphasis on distributional overlap and causally well-posed questions.

However, the aid-growth debate should not be reduced to a question of whether the point estimate for the coefficient on aid is statistically different from zero. Rather, it is more pertinent to explore whether the confidence interval for the effect of aid occupies an economically meaningful and plausible domain, as well as whether it is broadly consistent across different specifications and (robust) estimators.[3] In this sense, the results of RS08 are encouraging. When the endogeneity of aid is addressed, ruling out the OLS specifications, the coefficient on Aid/GDP is positive over all long horizon cross-section estimates and the vast mass of the 95 per cent confidence interval lies in the positive domain (e.g., see the columns pertaining to 1960-2000 and 1970-2000 in Tables 4 and 5).[4] Moreover, in all these cases the magnitude of aid's effect on growth is not statistically different from 0.1. Thus, based on RS08's reported results, we cannot reject the theoretical hypothesis that aid has a modest but positive effect on growth over the long run. This observation motivates further exploration of the data.

4.2 Robust estimators for causal effects

As discussed in Section 3, empirical applications of the potential outcomes framework are best developed for binary treatments. In such cases the counterfactual can be easily understood, the notion of a treatment effect can be interpreted in an intuitive and clear fashion, and established techniques exist to deal with relevant 'causal' challenges. For example, doubly robust estimators (Robins and Rotnitzky 1995) are robust to

3 This echoes Temple (2000), who places emphasis on finding plausible bounds for coefficients rather than point estimates.

4 The only exception to this rule is when Aid/GDP squared is added to the specification. However, calculated at the average level of Aid/GDP in their dataset, the combined effect of the two terms is 0.11 which is highly consistent with other specifications.

misspecification of either the propensity score or the outcome regression. They reflect what Imbens and Wooldridge describe as: 'best practice [that] is to combine linear regression with either propensity score or matching methods in ways that explicitly rely on local rather than global linear approximations to the regression functions.' (2009: 25). In principle, such techniques should be helpful for investigating the aid-growth problem. This is because the underlying causal model is not established and substantial differences exist between individual units (c.f., Bangladesh vs. Rwanda).

A number of doubly robust estimators have been proposed in the literature. One of the more straightforward of these, set out in Imbens (2004), can be described as inverse probability weighted least squares (IPWLS). This combines a standard inverse probability weighting estimator, which *only* uses propensity scores to estimate treatment effects, with a linear regression which controls for the observed covariates. Thus, as long as *either* the propensity score *or* the linear regression is correctly specified, consistent estimates are generated. More formally, and continuing the notation from equation (1) but now assuming the treatment variable is binary, propensity scores can be derived from estimates on a logistic form:

(2) $$\pi(x_i) = \Pr(W_i = 1 | X_i = x_i) = \exp(X_i'\lambda)/[1 + \exp(X_i'\lambda)]$$

The predicted probabilities ($\hat{\pi}_i$) are then used as inverse probability weights in the least squares problem:

(3) $$\min_{\delta,\zeta} \sum_{i=1}^{N} \frac{[Y_i - (\delta W_i + X_i'\zeta)]^2}{(W_i/\hat{\pi}_i) + (1-W_i)/(1-\hat{\pi}_i)}$$

where the estimate of δ represents the treatment effect of interest; and inference is made with robust standard errors.

Direct application of the IPWLS estimator to the aid-growth problem is complicated both by the continuous nature of the treatment effect of interest and its endogeneity. Nevertheless, we can proceed with some simple modifications. First, the generated instrument can be dichotomized to create a binary assignment-to-treatment variable. This step is not problematic as, by iterated expectations, we assume any function of the instrument is orthogonal to the errors in the outcome equation of interest. Dichotomization of the instrument also represents a useful robustness check – if results arising from the binary instrument were inconsistent with its continuous counterpart, this might indicate that the latter findings were driven by peculiarities in the distribution of the instrument. It also relaxes the assumption of a constant linear relationship between aid and growth, rather placing emphasis on the average difference between treatment and control groups regardless of the shape of growth's response to aid. Moreover, in the present case where the instrument is derived from a preliminary stage regression (as in RS08), dichotomization provides a check against measurement error or misspecification in the preliminary stage.

Given the above, equations (2) and (3) can be modified by substituting the binary treatment variable (W) with the dichotomous instrument (Z). From this we note that the subsequent results can be understood as the reduced form of a (weighted) two stage least squares problem where Z_i instruments for the endogenous aid variable. To see this, note that the first stage of the two stage least squares problem associated with equation (1) looks like:

(4) $$W_i = \gamma Z_i + X_i'\theta + \nu_i$$

which, substituting into equation (1), gives the reduced form:

(5) $$Y_i = \alpha\gamma Z_i + X_i'(\beta + \alpha\theta) + \{\varepsilon_i + \alpha\nu_i\}$$

Ignoring weights, these coefficients are directly comparable to those in the modified version of equation (3), where W_i has been replaced by Z_i. Specifically we now have: $\delta = \alpha\gamma$; and $\zeta = (\beta + \alpha\theta)$. The reduced form is of interest *per se* because failure to find a relationship here would be indicative of an absence of any treatment effect. Also, following from the preliminary stage, the dichotomized instrument represents groups of high- versus low-aid treatment and control groups. Thus, to some extent, the reduced form has a direct and natural interpretation.

While a weighted estimation of equation (5) provides substantial insight, it is not sufficient. For comparison against the (marginal) treatment effects estimated by RS08, we need to restate the reduced form results in units of the endogenous treatment variable; or, put differently, we need to extract α from δ. As can be seen from equation (5), the correct rescaling factor is the estimate for γ from the first stage, (estimated with the same weights).[5] It follows that the IV counterpart of the IPWLS estimator is weighted two-stage least squares, employing the dichotomous instrument and the propensity score weights. A particular advantage of seeing the problem in this way is that, compared to manual estimation of the reduced form and first stage, automated versions of 2SLS ensure correct calculation of standard errors.

Another doubly robust estimator is an extended version of IPWLS, which can be modified in similar fashion to allow for instrumental variables. This relaxes the assumption in equation (3) that the coefficients on the covariates are the same for treatment and control groups. Following Imbens and Wooldridge (2009), one can estimate versions of equation (3) separately for the treatment and control groups, this time with the covariates stated as deviations from overall sample means such that the treatment effect is given by the difference in the estimated intercept terms.[6] In an IV context one replaces the (binary) treatment variable with the dichotomous instrument, and the result can be interpreted as deriving from a reduced form. To extract the

---

5 A naïve approach would be to divide the reduced form estimate of δ by the difference in average Aid/GDP for each group of Z_i. Given we have a binary instrument; this difference is equivalent to the estimated coefficient on a linear regression of the endogenous treatment variable on the instrument which is a type of first stage regression excluding the set of covariates used in the reduced form.

6 Evidently, in estimating equation (3) across groups defined by W_i or Z_i these terms do not enter the RHS. Other versions of the doubly robust estimator, such as that given by Lunceford and Davidian (2004) yield basically equivalent results to those presented here; these are available on request from the authors.

treatment effect of interest, we divide the latter by the same first stage coefficient indicated in the IV modification of the IPWLS.[7]

Using RS08's data and specification, results for the 1960-2000 and 1970-2000 periods for these doubly robust estimators are given in columns III to VI of Table 1. To derive the binary instrument, we sort countries in ascending order of the instrument generated from the preliminary stage (from lowest to highest predicted aid shares) and then select the first 30 for the 'control' (Z_i = 0) and the rest for the 'treatment' (Z_i = 1). The motivation for this choice is to identify a subsample of countries with the smallest possible average value for predicted aid inflows while still maintaining statistical viability. In practice, the control group approximately corresponds to all countries falling below the 40th percentile.[8] Taking the results for the period beginning 1960, both the IV versions of the IPWLS and the flexible doubly robust (FDR) estimators give results that are almost equivalent to those in column I. Importantly, this shows that these new estimators provide only modest modifications to more standard 2SLS estimators. Even so, despite the fact that dichotomization of the instrument leads to some loss of power, as shown by the F-statistic on the excluded instruments (derived from the first stage), the overall explanatory power of the model is marginally improved and the standard error on the treatment effect is reduced. For the period beginning 1970, both the doubly robust estimators (columns IV and VI) produce a moderate upward shift in the treatment effect point estimate; although the corresponding standard errors do not shrink, the former shift is sufficient to push the estimate into a statistically significant domain.

Note that for columns V and VI, the row entries for the control covariates provide estimates of the extent of overlap in the distributions of the two groups of the dichotomized instrument. As suggested by Imbens and Wooldridge (2009), this is calculated as the standardized mean difference between the distributions for the two groups. As a rule of thumb, absolute values above 0.25 are taken to indicate considerable non-overlap. From these values one notes that the 'treatment' and 'control' (instrument) groups are distinctive, especially in terms of their geographic location, initial inflation and number of revolutions. Moreover, the extent of overlap appears to worsen over time, underlining the usefulness of using methods that attempt to balance the distribution of covariates across treatment and control groups. Thus, based on the same data, specification and instruments as RS08, robust counterfactual methods yield results that are comparable to the point estimates obtained by RS08, but also with statistical significance for the 1970-2000 period.

---

7 In deriving standard errors, we note that the resulting treatment effect is a ratio of estimates from two (independent) least squares procedures. The standard error of the treatment effect is generated numerically via a parametric bootstrap.

8 We recognize that the chosen cut point is somewhat arbitrary and potential inefficiencies are involved in moving from a continuous to a binary measurement scale. Sensitivity analysis on the cut off point and continuous approaches are presented in later sub-sections.

Table 1: Long horizon instrumental variable cross-section estimates

	RS Replications		IPWLS		FDR	
	(I)	(II)	(III)	(IV)	(V)	(VI)
	1960-2000	1970-2000	1960-2000	1970-2000	1960-2000	1970-2000
	b/se	b/se	b/se	b/se	*overlap*	*overlap*
Treatment effect	0.063	0.096	0.065	0.153	0.061	0.163
	(0.06)	(0.07)	(0.04)	(0.08)*	(0.04)	(0.08)**
Initial per capita GDP	-1.175	-1.409	-1.307	-1.672	*-0.006*	*0.118*
	(0.39)***	(0.43)***	(0.26)***	(0.34)***		
Initial level of policy	1.620	2.139	1.651	2.322	*0.181*	*0.186*
	(0.67)**	(0.62)***	(0.51)***	(0.58)***		
Initial life expectancy	0.059	0.076	0.047	0.047	*0.136*	*0.343*
	(0.03)**	(0.04)*	(0.02)**	(0.03)		
Geography	0.526	0.606	0.540	0.674	*0.507*	*0.556*
	(0.19)***	(0.26)**	(0.16)***	(0.23)***		
Institutional quality	4.558	4.077	4.203	5.114	*0.089*	*0.173*
	(1.70)***	(2.33)*	(1.43)***	(2.24)**		
Initial inflation	-0.003	-0.005	-0.004	-0.004	*0.327*	*0.317*
	(0.00)	(0.00)	(0.00)	(0.00)		
Initial M2/GDP	0.017	0.010	0.019	0.025	*0.184*	*0.262*
	(0.01)	(0.02)	(0.01)*	(0.02)		
Initial budget balance/GDP	0.016	0.016	0.005	-0.006	*0.134*	*0.171*
	(0.03)	(0.04)	(0.02)	(0.04)		
Revolutions	-1.144	-1.406	-1.195	-1.103	*0.401*	*0.371*
	(0.62)*	(0.66)**	(0.52)**	(0.66)*		
Ethnic fractionalization	0.712	0.788	0.482	0.538	*0.031*	*-0.006*
	(0.61)	(0.85)	(0.41)	(0.65)		
East Asia	0.552	0.577	0.342	0.237	*0.275*	*0.290*
	(0.45)	(0.51)	(0.41)	(0.54)		
Sub-Saharan Africa	-1.338	-1.672	-1.760	-2.605	*-0.471*	*-0.528*
	(0.48)***	(0.63)***	(0.41)***	(0.67)***		
Intercept	4.688	5.505	6.579	8.058		
	(2.98)	(3.53)	(1.74)***	(2.62)***		
Scale of excluded instrument	Continuous	Continuous	Binary	Binary	Binary	Binary
F-stat. on excluded instrument	25.2	31.6	19.7	18.3	-	-
Number of obs.	74	78	74	78	74	78
R-squared	0.66	0.59	0.72	0.66	-	-

significance level: * 10%; ** 5%; *** 1%

Notes: Columns (I) and (II) replicate Rajan and Subramanian's (2008) results; columns (III) and (IV) implement an instrumental variables version of the inverse probability weighted least squares estimator (IPWLS) discussed in the text; columns (V) and (VI) implement an instrumental variables version of the flexible doubly robust (FDR) estimator discussed in the text; for the FDR estimator, row entries for the treatment variable give the estimated treatment effect and corresponding standard error (calculated by parametric bootstrap), while for all control covariates rows estimate the extent of overlap in the distribution across groups of the (binary) instrument; variables and specification are as per Rajan and Subramanian (2008; Table 4); 'treatment effect' refers to endogenous Aid/GDP; standard errors, given in parentheses are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Sources: Authors' estimates, see Appendix A.

4.3 Specification issues

We now turn to a reconsideration of the specification used by RS08 and review results for alternative specifications using both continuous and binary instruments. This is justified for a number of reasons. First, given the relatively small sample size, inclusion of redundant variables may lead to a loss of efficiency and contribute to undesirable multicollinearity. In the present case, we note that the three macroeconomic initial conditions (inflation, money supply and budget balance) are insignificant in RS08's outcome regressions for all periods. As shown in Table 2, their removal makes no material changes to the remaining coefficients. Second, additional problems arise from including too many control variables in the treatment effects context. As Wooldridge (2005) clarifies, inclusion of contemporaneous outcome variables – i.e., variables which may also be affected by the level of treatment – can invalidate the unconfoundedness assumption required for valid causal inference (see the discussion about 'bad controls' in Angrist and Pischke 2008). This is pertinent as RS08's chosen specification includes two variables that capture average *outcomes* during the period of analysis – institutional quality and the number of revolutions.[9] Inclusion of these variables is puzzling in light of the literature which examines the effects of aid on growth through institutional performance. Controlling for such outcomes blocks potential channels through which aid may affect growth and thereby restricts the estimated coefficient on aid to a partial as opposed to a general effect.

Third, it is helpful to consider the appropriate role of regional fixed effects. In RS08's specification, only East Asia and sub-Saharan Africa are included as regional dummy variables. This appears to be a posterior choice in the sense that prior to the 1980s there was no particular reason to identify these as 'special' regions. Including regional dummy variables helps absorb intra-regional correlations and captures omitted spatial fixed effects such as those arising from geography, shared historical experiences and trade relations. *A priori*, a more plausible approach is to include a fuller set of regional dummies. Fourth, following the discussion in Section 3, it may be useful to include additional variables that reflect initial socio-economic conditions such as education and health indicators. In themselves these are frequently seen as important determinants of growth and may also proxy for initial institutional conditions; as such, they may explain some of the variation in the expected growth returns to foreign aid.

As a result of these considerations, Table 2 re-runs regressions for various alternative specifications. For the time being we focus the analysis on the 1970-2000 period, corresponding to our interest in the long run impact of aid. We place less emphasis on the 1960-2000 period. Many countries in the dataset, particularly those in Africa, had not attained independence by 1960. The majority of French colonies achieved independence in 1960; however, the shift to independent administration was in most cases very gradual (Berg 1993). In contrast, by 1970 the large majority of the countries in the sample had achieved independence and had operated independently for a period of at least a few years, with Portuguese colonies being the prominent exception. We return to results for the other periods considered by RS08 in sub-section 4.5.

---

9 Note that initial institutional conditions are captured (at least partially) by the Sachs and Warner trade policy index for the beginning of each period analysed (see Appendix A).

Table 2: Alternative specifications for 1970-2000 cross-section estimates

	(I) IV-LIML b/se	(II) IV-LIML b/se	(III) IV-LIML b/se	(IV) IV-LIML b/se	(V) IPWLS b/se	(VI) FDR *overlap*
Treatment effect	0.096	0.089	0.117	0.084	0.090	0.139
	(0.06)	(0.08)	(0.07)	(0.06)	(0.06)	(0.08)*
Per capita GDP	-1.409	-1.166	-0.858	-1.423	-1.515	*0.118*
	(0.39)***	(0.43)***	(0.45)*	(0.33)***	(0.29)***	
Level of policy	2.139	2.545	2.619	1.934	2.433	*0.186*
	(0.56)***	(0.66)***	(0.60)***	(0.48)***	(0.50)***	
Life expectancy	0.076	0.091	0.114	0.029	0.045	*0.343*
	(0.04)**	(0.04)**	(0.03)***	(0.04)	(0.04)	
Geography	0.606	0.735	0.617	0.322	0.342	*0.556*
	(0.23)***	(0.27)***	(0.22)***	(0.23)	(0.23)	
Ethnic fractionalization	0.788	0.241	0.329	0.285	0.019	*-0.006*
	(0.77)	(0.91)	(0.76)	(0.74)	(0.75)	
Institutional quality	4.077					
	(2.11)*					
Revolutions	-1.406					
	(0.59)**					
Coastal pop. density				0.001	0.001	*-0.093*
				(0.00)***	(0.00)***	
Primary schooling				2.572	2.147	*0.222*
				(1.14)**	(1.03)**	
Price of invest. goods				-0.007	-0.007	*-0.004*
				(0.00)*	(0.00)**	
Malaria risk				-1.486	-1.022	*-0.367*
				(0.84)*	(0.73)	
Scale of excl'd instrument	Continuous	Continuous	Continuous	Continuous	Binary	Binary
Regional dummies	SSA, EA	SSA, EA	SSA, A, LA	SSA, A, LA	SSA, A, LA	SSA, A, LA
Number of obs.	78	78	78	78	78	78
R-sq.	0.59	0.55	0.50	0.66	0.68	-
Weak identification stat.	36.12	31.56	29.06	22.12	17.65	-
Stock-Wright LM stat.	2.83*	1.71	3.33*	2.94*	2.64	-
Probability effect = 0.1	0.96	0.88	0.81	0.78	0.86	-

significance level: * 10%; ** 5%; *** 1%

Notes: Column (I) uses RS's chosen specification but is estimated using the IV-LIML estimator (calculated in STATA with *ivreg2*); column (II) drops 'bad controls' and redundant variables; column (III) alters the regional dummies; column (IV) adds additional covariates from Sala-i-Martin et al. (2004) (for which a small number of observations are imputed to retain the sample size); columns (V) and (VI) employ doubly robust estimators on the same specification in column (IV); columns (I) and (II) include regional dummies for sub-Saharan Africa (SSA) and East Asia (EA); columns (III) to (VI) include regional dummies for SSA, Asia (A) and Latin America and the Caribbean (LA); 'Probability effect = 0.1' refers to a Wald test of the null hypothesis that the treatment effect is equal to 0.1; intercept not shown; other reported statistics are as described in the text; for the FDR estimator, row entries for the control covariates indicate the extent of overlap in the distribution across groups of the (binary) instrument; 'treatment effect' refers to endogenous Aid/GDP; standard errors, given in parentheses, are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Sources: Authors' estimates, see Appendix A.

For the regressions in Table 2, the same instrumentation strategy, base variables and sample is retained as in Table 1.[10] Column I revisits RS08's original specification using the limited information maximum likelihood instrumental variables (IV-LIML) estimator with robust standard errors. This is chosen for analyzing the continuous treatment variable case as it is generally taken to be more robust to weak instruments than 2SLS or cross-section GMM approaches (Baum et al. 2007).[11] We also report additional test statistics that support inference in the presence of (weak) instruments. Column II of the table drops the outcome covariates and redundant variables from the RS08 specification; and column III adds an alternative set of regional dummies.[12] Column IV includes a set of additional controls based on the findings of Sala-i-Martin et al. (2004) who undertake comprehensive Bayesian averaging of long-run growth estimates. Excluding regional location and variables already found in column I, we include those variables identified by Sala-i-Martin et al. (2004) which refer to initial conditions and are among those with the highest posterior probability of inclusion (see Appendix A for further details). In columns V and VI, we use this new specification and redeploy the doubly robust estimators from the previous sub-section.

The results are informative and do nothing to challenge our prior that aid exerts a small positive effect on long-run growth. First, although removal of the endogenous covariates leads to some loss of overall explanatory power compared to the original specification (Table 2, column I), the inclusion of additional controls for initial conditions substantially boosts the power of the specification, returning it to similar levels as the original specification (Table 2, column IV). Second, although many of the point estimates are not significantly different from zero, in all continuous cases we cannot reject the hypothesis that the marginal effect of aid on growth is equal to 0.1. Third, the additional test statistics confirm that the instrument remains strong and indicate a non-zero (positive) correlation between the endogenous regressor and the outcome variable based on the reduced form regression (captured by the Stock-Wright S statistic, which is robust to the presence of weak instruments (Baum et al. 2007). Finally, the notion that the treatment effect occupies a positive domain is underlined by the flexible doubly robust estimator (FDR), which is significantly different from zero at the one per cent level.

4.4 Instrumentation strategy

As a third empirical step, we review the instrumentation strategy underlying the foregoing results. While there are weaknesses in the approach adopted by RS08, we consider their instrument does capture variation in the allocation of aid that is plausibly exogenous to economic performance.[13] At this stage, our objective is not to develop a radically new instrument. Rather, we strengthen the approach, and in so doing, verify

10 A wider range of results from alternative specifications are available from the authors. However, these add little in terms of substance and are omitted here for presentational clarity.

11 In the case of a single instrument and endogenous regressor, 2SLS, GMM and LIML converge to the same point estimates. However, LIML is preferred in the context of multiple instruments and we employ this IV estimator throughout.

12 Only a small number of regions are chosen due to the small sample size and to ensure, in the binary instrument case, that both treatment and control groups always contain countries from all regions.

13 Clemens and Bazzi (2009) develop a more strident critique of the RS08 instrument.

the robustness of the results to alterations in the preliminary stage regressions on which the instrument is ultimately based.

Five concerns with RS08's instrumentation strategy inform the modifications. First, in their preliminary stage regressions, recipient GDP occurs in the denominator of the dependent variable. Following Kronmal (1993), inappropriate use of ratio variables can lead to substantial misinterpretation (or bias) in least squares regressions. This may arise if the denominator of the dependent variable (recipient GDP) is correlated with the RHS variables independently of the numerator of the dependent variable (aid inflows). In the present case this could arise if donor decision rules do not target the Aid/GDP ratio, and/or if there is a direct association between recipient GDP levels and population size or past colonial experiences. Second, in light of the complexity of the growth process, it is always possible to find a reason why a given instrument may not be valid. Although this concern is acknowledged by RS08 (2008: 651), we believe it is most pertinent with respect to variables in the preliminary stage regression that refer to *specific* colonial relationships. These are the colony dummies and population interactions for the major imperial powers of the 18th and 19th centuries. The issue is that the institutional transplants and broader colonizing strategies pursued by these imperial powers were not alike, and it is plausible they have a persistent effect on income levels to the present day. This idea is at the heart of the debate concerning the effect of different legal origins (La Porta et al. 2008), historical events (e.g., Nunn 2008) and other institutional forms on contemporary economic outcomes. Third, it is apparent that individual donor countries exhibit distinct attitudes to giving foreign aid (Alesina and Dollar 2000), which can be traced to deep cultural and historical factors such as the social democratic heritage of the Nordic countries. These time-invariant influences can be understood as fixed effects and may be included as RHS variables in the preliminary stage regression. Notably, and unlike the RS08 explanatory variables, these fixed effects may explain a part of the variation in aid allocations that is unrelated to purely strategic or political motives. As such, their inclusion may strengthen the overall validity and interpretation of the generated instrument.

Taken together, the first three concerns amount to straightforward modifications to the preliminary stage regression. Specifically, in place of Aid/GDP we use average aid *per capita* as the dependent variable. This variable accords more closely with the explicit aid allocation rules used by donors, such as the World Bank (see for example Annex 1 of IDA15, 2008), and reduces the scope for misinterpretation from spurious correlation when GDP is in the denominator. We also drop the donor-specific interaction terms and employ only the aggregate RHS variables in RS08's original preliminary specification. In addition, we include donor-specific fixed effects. The modified preliminary stage regression emerges as follows:

$$\text{(6)} \quad \frac{Aid_{dr}}{POP_r} = \beta_0 + \beta_1 CURCOL_{dr} + \beta_2 COLONY_{dr} + \beta_3 COMLANG_{dr} + \beta_4 \log(POP_d/PO \\ + \beta_5 COLONY_{dr} \times \log(POP_d/POP_r) + \vartheta_d DONOR_d + \varepsilon_{dr}$$

Running these modified regressions does not lead to any major changes in the interpretation of the preliminary stage regressions. The estimates for all periods point to the same qualitative conclusions; therefore, in Table 3 we present the 1970-2000 findings only. For comparison, column I replicates RS08's original preliminary stage specification; column II uses aid per capita as the dependent variable with the RS08 specification; and column III regresses aid per capita on our new specification including donor fixed effects.[14] Of note is the drop in the R-squared statistic when we replace Aid/GDP (column I) with aid per capita (column II). This is expected, and indicates there may have been some unwanted independent correlation between GDP in the dependent variable and the RHS variables. Moreover, when we exclude the empire-specific colonizer variables, the colony dummy is no longer significant (column III). However, it remains the case that the explanatory variables explain a substantial share of observed aid allocations.

Two additional concerns motivate further modifications. The OECD-DAC aid dataset used for bilateral aid flows includes numerous missing values. While in some cases these genuinely refer to absent data, in most cases they represent unreported null values.[15] RS08 incorrectly treat these as missing values. This is a material omission, because it distorts estimates for average bilateral aid flows over time. Consequently, we re-estimate the required bilateral aid dataset and calculate period averages for Aid/GDP and aid per capita allowing for both possibilities – i.e., missing and zero entries.[16]

The existence of zero-value aid inflows points to a final possible concern – selection bias in observed aid flows. In principle, the decision by a donor to provide aid involves at least two distinct choices – (i) which recipients should receive aid; and (ii) how much to provide.[17] In the absence of a more detailed examination of this decision problem, one way to address the potential bias from unobserved selection effects is to use Heckman's correction (Heckman 1979). Thus, following the previous modifications, column IV of Table 3 presents the same specification as in Column III but with aid per capita now calculated to treat missing cell entries as zeroes. Column V gives the outcome equation from a Heckman selection model (estimated by full information maximum likelihood), where the existence of zero or non-zero aid flows is used as the binary selection variable.[18] Despite these changes, the direction of the results and their interpretation are largely unchanged. However, we do reject the hypothesis that there is no selection bias (i.e., the outcome and selection equations are independent). This supports our use of this estimator and is our preferred preliminary stage regression.

---

[14] See Appendix A for further details regarding data sources and variable construction. All these regressions are by OLS.

[15] We have confirmed this in correspondence with the OECD DAC Secretariat.

[16] Pairwise correlations show that the results from these two approaches are highly comparable both on aggregate and at the disaggregate levels.

[17] For discussion and elaboration of this two-stage decision rule see Tarp et al. (1999); also Berthélemy and Tichit (2004).

[18] To avoid a situation where identification relies on the assumption of multivariate normality alone, in the selection equation we include a set of dummy variables capturing the total number of colonial relationships experienced by each country (which ranges from zero to four). This is largely a technical issue and does not alter the nature or direction of results.

Table 3: Alternative preliminary stage regressions, 1970-2000

	(I) OLS	(II) OLS	(III) OLS	(IV) OLS	(V) Heckman
Colonial relationship (dummy)	1.653	8.652	-1.698	-0.551	-0.885
	(0.24)***	(5.32)	(2.39)	(2.08)	(2.20)
Currently in a colonial relationship (dummy)	-0.968	4.482	8.887	14.144	24.475
	(0.56)*	(15.45)	(21.31)	(21.15)	(36.71)
Common language (dummy)	0.074	1.942	1.971	1.299	1.302
	(0.04)*	(0.68)***	(0.78)**	(0.60)**	(0.67)*
Ratio of (initial) log. population	0.091	0.722	0.669	0.320	0.445
	(0.01)***	(0.09)***	(0.10)***	(0.06)***	(0.08)***
Ratio of log. population x colony	0.615	8.559	4.230	3.316	3.356
	(0.11)***	(5.32)	(0.94)***	(0.72)***	(0.77)***
Colonial relationship with UK	-1.954	-17.340			
	(0.32)***	(6.17)***			
Colonial relationship with France	-3.539	-31.703			
	(0.46)***	(9.89)***			
Colonial relationship with Spain	-2.906	-15.565			
	(0.43)***	(6.47)**			
Colonial relationship with Portugal	0.830	-1.758			
	(0.45)*	(10.74)			
Ratio of log. population x UK colony	-0.363	-4.798			
	(0.12)***	(5.49)			
Ratio of log. population x French colony	1.036	3.924			
	(0.17)***	(6.33)			
Ratio of log. population x Spanish colony	0.230	-4.638			
	(0.19)	(5.87)			
Ratio of log. population x Portuguese colony	2.902	4.147			
	(0.17)***	(7.15)			
Dependent variable	Aid/GDP	Aid p.c.	Aid p.c.	Aid p.c.	Aid p.c.
Treatment of ‘missing’ aid values	Per RS08	Missing	Missing	Zero	Zero
Donor fixed effects	No	No	Yes	Yes	Yes
Test for independence of outcome and select. eqs.	-	-	-	-	9.56***
Number of obs.	3288	2681	2681	3328	3328
R-squared	0.42	0.29	0.23	0.21	-
F statistic	185.93	12.25	13.52	11.59	-

significance level: * 10%; ** 5%; *** 1%

Notes: Column (I) replicates Rajan and Subramanian’s preliminary stage regression (2008, Table 4); differences in other column estimates depend on the specification, dependent variable and treatment of potentially missing aid values; intercept not shown; Heckman estimator uses full information maximum likelihood (FIML); the Heckman selection equation (not shown) includes all outcome covariates and a dummy for the number of colonial relationships experienced by the recipient; test for independence of outcome and selection equations refers to a Wald test that the correlation (*rho*) between the residuals in the two equations is equal to zero; variables are as per Rajan and Subramanian (2008; Table 3); standard errors are robust to arbitrary heteroskedasticity and intra-group correlation between aid recipients (except for column I where standard errors assume homoskedasticity in order to replicate Rajan and Subramanian 2008).

Sources: Authors’ estimates, see Appendix A.

In order to verify the respective strength of each of the four 'new' instruments (columns II to V of Table 3), we aggregate upwards from the bilateral data and run RS08's original specification. This is reported in Table 4, where columns I to IV employ the same endogenous aid variable as in previous sub-sections. In other words, and although the instrument is generated from aid per capita, the same Aid/GDP variable is taken as the endogenous treatment. In columns V and VI, the estimate for Aid/GDP is recalculated now also treating absent aid cells as zeros (as in the preliminary stage regression). As can be seen from the instrument test statistics reported in the table, all these new instruments perform strongly.[19] Under-identification tests, which can be interpreted as testing the null hypothesis of a zero correlation between the instruments and the endogenous regressors, are very clearly rejected. The weak identification test, which uses a finite-sample adjustment of the standard F-statistic to assess the strength of the partial correlation between the excluded instruments and the endogenous variables in first-stage regressions, not only exceeds critical values in all cases but is highly comparable to levels achieved using RS08's original approach (Table 2, column I). Perhaps more importantly, the Stock-Wright S statistic finds a significant (partial) correlation between the instrument and the outcome of interest; and in all specifications we cannot reject the hypothesis that the point estimate is equal to 0.1. Thus, the modified instrumentation strategy provides comparable results to those already presented, while resting on strengthened foundations.

Finally, Table 5 revisits the alternative specifications and estimators employed in Table 2. Here we rely only on the last of the new instruments (from column V, Table 3) as well as the preferred endogenous variable (from columns V and VI, Table 4). Again, the instruments remain strong and the estimated treatment effect is consistently positive but significant across all specifications and estimators, including the flexible doubly robust (FDR) estimator. Thus, the treatment effect remains squarely in the domain anticipated by theory with point estimates ranging from 0.12 to 0.30. With the recalculation of the treatment variable and the preferred instrumentation approach, the point estimates for the estimated treatment effect are somewhat larger than in Table 4. Possible explanations for this result are that: (i) we have reduced systematic, upwards measurement error in the endogenous variable by correctly dealing with zero-valued aid flows, (ii) there is a diminished influence of spurious correlation from endogenous demand-side factors; and (iii) we capture higher quality aid via the inclusions of donor fixed effects.

To give a visual impression of the main results, Figure 1 plots the distribution of the estimated treatment effects for our final result in column VI of Table 5 and the corresponding result in RS08 (assuming the effects are normally distributed).[20] This moves away from an exclusive focus on point estimates and emphasizes the richer information provided by confidence intervals. The figure underlines the point that the original results in RS08 are broadly consistent with a view that aid may have a small positive effect on growth over the long-run (sub-section 4.1). As the 'RS08 distribution' indicates, although the point estimate may be insignificantly different from

19 See Baum et al. (2007) for further discussion of these test statistics and their implementation in STATA.

20 Note these distributions are based on two cross-section regressions from similar datasets. They do not represent a comprehensive sample of specifications as in Sala-i-Martin et al. (2004).

Table 4: IV regressions with alternative endogenous and instrumental variables, 1970-2000

	(I) IV-LIML b/se	(II) IV-LIML b/se	(III) IV-LIML b/se	(IV) IV-LIML b/se	(V) IV-LIML b/se	(VI) IV-LIML b/se
Treatment effect	0.106	0.113	0.125	0.147	0.185	0.216
	(0.07)	(0.07)	(0.08)	(0.09)*	(0.11)*	(0.12)*
Initial per capita GDP	-1.395	-1.385	-1.366	-1.334	-1.372	-1.341
	(0.39)***	(0.38)***	(0.39)***	(0.41)***	(0.38)***	(0.40)***
Initial level of policy	0.079	0.081	0.086	0.093	0.079	0.085
	(0.04)**	(0.04)**	(0.04)**	(0.04)**	(0.04)**	(0.04)**
Initial life expectancy	2.131	2.126	2.115	2.098	2.154	2.143
	(0.57)***	(0.56)***	(0.58)***	(0.61)***	(0.58)***	(0.60)***
Geography	4.081	4.083	4.087	4.093	3.977	3.965
	(2.13)*	(2.14)*	(2.17)*	(2.21)*	(2.07)*	(2.10)*
Institutional quality	0.618	0.626	0.642	0.670	0.608	0.629
	(0.24)**	(0.25)**	(0.26)**	(0.27)**	(0.24)**	(0.25)**
Revolutions	-1.411	-1.415	-1.422	-1.434	-1.366	-1.368
	(0.61)**	(0.62)**	(0.64)**	(0.67)**	(0.57)**	(0.60)**
Ethnic fractionalization	0.855	0.899	0.984	1.132	0.849	0.969
	(0.79)	(0.81)	(0.85)	(0.92)	(0.73)	(0.76)
Instrument (col. of Table 3)	II	III	IV	V	IV	V
Scale of excld'd instrument	Contin.	Contin.	Contin.	Contin.	Contin.	Contin.
Missing Aid/GDP entries?	Missing	Missing	Missing	Missing	Zero	Zero
Number of obs.	78	78	78	78	78	78
R-square	0.58	0.57	0.55	0.52	0.60	0.57
Under-identification stat.	14.67***	14.07***	13.21***	12.46***	14.3***	14.31***
Weak identification stat.	29.64	28.12	27.1	22.69	34.68	29.48
Stock-Wright S stat.	3.15*	3.08*	3.48*	4.33**	3.48*	4.33**

significance level: * 10%; ** 5%; *** 1%

Notes: In columns (I) to (IV) the endogenous treatment variable is taken from Rajan and Subramanian (2008); in columns (V) to (VI) the endogenous treatment is re-estimated from OECD-DAC (2008) data treating possible missing values as zeroes (as indicated by 'Missing Aid/GDP entries?'); excluded instrument is denoted by reference to column numbers of Table 3, from which fitted values are estimated and aggregated across all donors for each recipient; in all columns Rajan and Subramanian's (2008) chosen specification is used (selected covariates shown); standard errors, given in parentheses, are robust to arbitrary heteroskedasticity; endogenous treatment variable is Aid/GDP; dependent variable is the average real growth rate.

Sources: Authors' estimates, see Appendix A.

Table 5: Combining alternative instruments and new specifications, 1970-2000

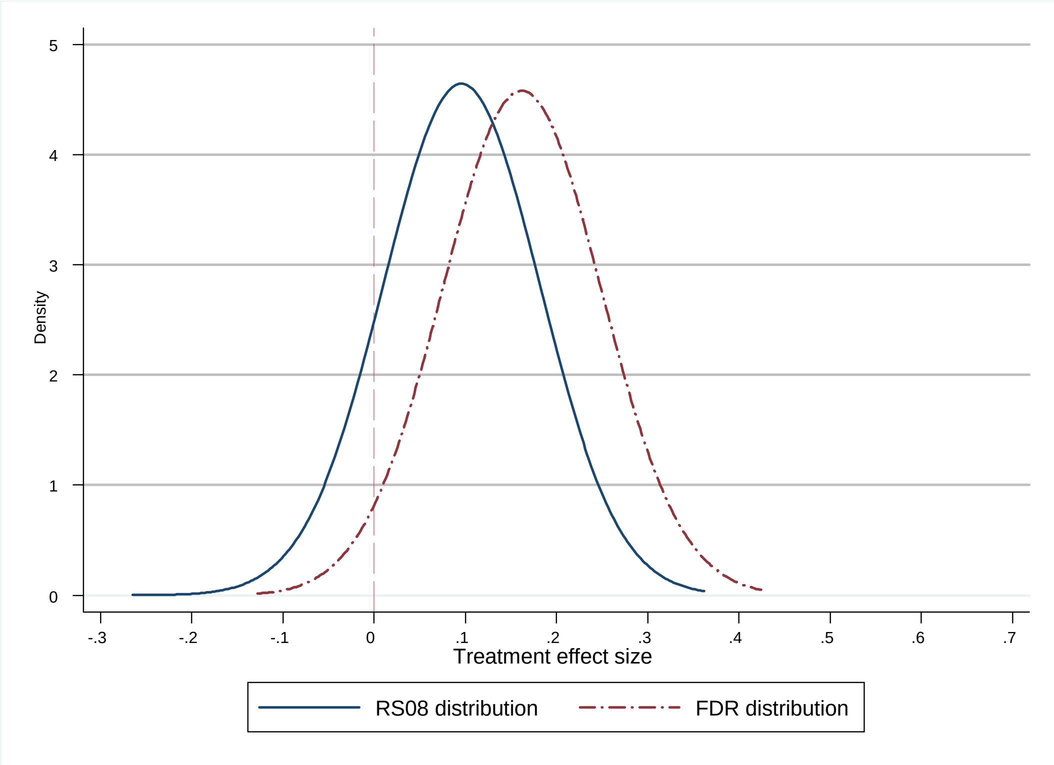
	(I)	(II)	(III)	(IV)	(V)	(VI)
	IV-LIML	IV-LIML	IV-LIML	IV-LIML	IPWLS	FDR
	b/se	b/se	b/se	b/se	b/se	*overlap*
Treatment effect	0.216	0.266	0.301	0.218	0.123	0.162
	(0.12)*	(0.14)*	(0.13)**	(0.11)*	(0.06)**	(0.07)**
Per capita GDP	-1.341	-1.044	-0.684	-1.265	-1.446	*0.233*
	(0.40)***	(0.47)**	(0.49)	(0.37)***	(0.29)***	
Level of policy	2.143	2.514	2.421	1.881	2.361	*0.177*
	(0.60)***	(0.70)***	(0.62)***	(0.53)***	(0.47)***	
Life expectancy	0.085	0.112	0.129	0.034	0.038	*0.419*
	(0.04)**	(0.04)**	(0.04)***	(0.04)	(0.04)	
Geography	0.629	0.798	0.712	0.446	0.305	*0.562*
	(0.25)**	(0.29)***	(0.26)***	(0.29)	(0.24)	
Ethnic fractionalization	0.969	0.685	0.677	0.609	-0.201	*-0.155*
	(0.76)	(0.93)	(0.77)	(0.74)	(0.73)	
Coastal pop. density				0.001	0.001	*-0.082*
				(0.00)***	(0.00)***	
Primary schooling				2.139	2.139	*0.343*
				(0.93)**	(0.93)**	
Price of invest. goods				-0.005	-0.005	*-0.132*
				(0.00)	(0.00)	
Scale of excl'd instrument	Continuous	Continuous	Continuous	Continuous	Binary	Binary
Number of obs.	78	78	78	78	78	78
R-square	0.57	0.48	0.42	0.61	0.72	-
Weak identification stat.	29.48	26.75	20.39	19.91	37.32	-
Stock-Wright S stat.	4.33**	4.34**	6.49**	5.14**	3.42*	-
Probability effect = 0.1	0.32	0.24	0.13	0.30	0.71	-

significance level: * 10%; ** 5%; *** 1%

Notes: Excluded instrument and endogenous treatment variables are based on those used in column VI of Table 4; only selected control covariates shown; column (I) repeats RS's chosen specification; column (II) drops 'bad controls' and redundant variables; column (III) alters the regional dummies; column (IV) adds additional covariates from Sala-i-Martin et al. (2004) (for which a small number of observations are imputed to retain the sample size); columns (V) and (VI) employ doubly robust estimators on the same specification in column (IV); columns (I) and (II) include regional dummies for sub-Saharan Africa (SSA) and East Asia (EA); columns (III) to (VI) include regional dummies for SSA, Asia (A) and Latin America and the Caribbean (LA); 'Probability effect = 0.1' refers to a Wald test of the null hypothesis that the treatment effect is equal to 0.1; other reported statistics are as described in the text; for the FDR estimator, row entries for the control covariates indicate the extent of overlap in the distribution across groups of the (binary) instrument; 'treatment effect' refers to endogenous Aid/GDP; standard errors, given in parentheses, are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Sources: Authors' estimates, see Appendix A.

Figure 1: Normal approximations of the distribution of treatment effects from alternative models, 1970-2000



Notes: Each plot gives an approximation to a normal distribution with mean at the point estimate and standard deviation equal to the standard error of estimated treatment effects. Plotted effects are from Table 1, column II ('RS08 distribution') and Table 5, column VI ('FDR distribution').

Source: Authors' estimates.

zero, the bulk of the estimated distribution of the treatment effect (>84 per cent) lies in the positive domain. The empirical modifications introduced in this paper have simply nudged this distribution to the right, diminishing substantially the estimated probability that the aggregate effect of aid on growth is less than or equal to zero.

4.5 Sensitivity tests

For the final empirical step we investigate the sensitivity of our results. First, we consider the impact of individual alterations to the underlying data and/or specification, where the choice of modifications reflects some of the main potential sources of fragility. Given the critical role of population factors in driving the generated instrument, we principally investigate whether different ways of handling population-related variables alter our results. Corresponding to the columns in Table 6, we proceed with alternative specifications as follows:[21] (I) in the preliminary stage regression we exclude very small states (with populations under 500,000) that do not appear in the final regression sample; (II) we exclude the largest and (recently) most dynamic

21 Unless otherwise stated, changes are made in the main IV regression only.

economies in the sample (India, Brazil, and China), as well as Israel and Egypt, which are often taken to be special cases of aid due to their links with the USA; (III) we include land area as an additional instrument and check the orthogonality of the fitted aid variable (see the Hansen J test); and (IV) we include land area as an additional covariate. Two more general checks are undertaken. We replace the treatment effect variable (Aid/GDP) with the aid per capita variable used in the (new) preliminary stage regressions (column V, Table 3). Finally, in column (VI) we repeat the flexible doubly robust procedure but redefine the control and treatment groups to include only the bottom and top 33 countries in the fitted aid per capita distribution. This ensures a substantial gap between the treatment and control groups. As can be seen in Table 6, none of these modifications changes our core results – in all models the treatment effect is positive and statistically significant.

Second, we return to the different periods analyzed in RS08 (namely: 1960-2000, 1980-2000 and 1990-2000). Using our preferred empirical approach and estimators, results for each alternative period are presented in Table 7.[22] For the 1960-2000 period (columns I and II) both the point estimate and variance of the estimated treatment effect are squarely in the domain found in Table 5 and depicted in Figure 1, which refers to 1970-2000. The long run impact of foreign aid comes across as well established. With respect to the shorter run effects of aid, given in columns (III) to (VI), we cannot reject the hypothesis that the treatment effect is zero. This is confirmed by the (very weak) relation in the reduced form given by the Stock-Wright S statistic. The plausible range for the treatment effect is much wider for these periods, reflected by larger standard errors on the treatment effect. This is most apparent for 1990-2000 where the standard error on the FDR treatment effect estimate is almost five times larger than that for the 1960-2000 period. As suggested from the discussion of Section 2, a meaningful and robust *average* short run effect of aid on growth may be very difficult to discern from the available empirical data.

5 Conclusion

To conclude, we return to the question posed in the title to this paper: has the aid and growth literature come full circle? Our response is ‘no’. While the pendulum has swung to deep scepticism concerning the ability of aid to contribute to economic growth in the most recent literature, a series of important points of agreement have emerged. First, methodological advances have improved the profession’s capacity to identify causal effects in economic phenomena. These advances in methods are beginning to be applied at the more aggregate level; in this regard, the supply side instrumentation approach of RS08 counts as a significant advance (with room for improvement as emphasized in Section 4). Second, these methodological advances highlight the serious challenges that must be surmounted in order to derive robust causal conclusions from observational

[22] Note that the endogenous treatment variable and generated instrument are re-estimated from the relevant bilateral data and time period in accordance with the procedure described in the previous sub-section. Where possible, covariate values also vary according to the period chosen; this is not the case with respect to the variables taken from Sala-i-Martin et al. (2004) as alternative initial values are not provided; however, this is not a major concern as the majority of these variables are highly persistent over time.

Table 6: Sensitivity tests, 1970-2000

	(I)	(II)	(III)	(IV)	(V)	(IV)
	IV-LIML	IV-LIML	IV-LIML	IV-LIML	IV-LIML	FDR
	b/se	b/se	b/se	b/se	b/se	*overlap*
Treatment effect	0.205	0.288	0.254	0.444	0.028	0.146
	(0.11)*	(0.13)**	(0.14)*	(0.22)**	(0.01)**	(0.09)*
Per capita GDP	-1.280	-0.968	-1.223	-1.115	-1.868	*0.219*
	(0.36)***	(0.44)**	(0.39)***	(0.52)**	(0.35)***	
Level of policy	1.896	2.202	1.840	1.792	2.029	*0.133*
	(0.51)***	(0.62)***	(0.57)***	(0.67)***	(0.45)***	
Life expectancy	0.032	0.044	0.038	0.080	0.002	*0.471*
	(0.04)	(0.05)	(0.04)	(0.06)	(0.04)	
Geography	0.434	0.432	0.479	0.340	0.580	*0.637*
	(0.28)	(0.32)	(0.31)	(0.36)	(0.29)**	
Ethnic fractionalization	0.560	0.875	0.745	0.838	0.394	*-0.170*
	(0.74)	(0.79)	(0.80)	(0.94)	(0.68)	
Coastal pop. density	0.001	0.001	0.001	0.002	0.001	*-0.098*
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	
Primary schooling	2.611	2.724	2.687	3.223	2.347	*0.356*
	(1.10)**	(1.20)**	(1.19)**	(1.35)**	(0.85)***	
Price of invest. goods	-0.005	-0.005	-0.004	-0.004	-0.006	*-0.164*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
Malaria risk	-1.407	-1.064	-1.437	-1.480	-1.492	*-0.532*
	(0.87)	(0.99)	(0.92)	(1.09)	(0.80)*	
Land area				0.328		
				(0.21)		
Scale of excl'd instrument	Continuous	Continuous	Continuous	Continuous	Continuous	Binary
Endogenous variable	Aid/GDP	Aid/GDP	Aid/GDP	Aid/GDP	Aid/pop.	Aid/GDP
Number of obs.	78	73	78	78	78	78
R-square	0.62	0.54	0.57	0.39	0.65	-
Weak identification stat.	16.29	17.54	9.86	13.00	22.09	-
Stock-Wright S stat.	4.98**	7.43**	5.37*	4.65**	5.14**	-
Probability effect = 0.1	0.39	0.17	0.30	0.13	-	-
Hansen J test (probability)	-	-	0.12	-	-	-

significance level: * 10%; ** 5%; *** 1%

Notes: In columns (II), (IV) and (VI) the excluded instrument and endogenous treatment variables are based on those used in column VI of Table 4; in column (I) the excluded instrument derives from a modification of the latter, dropping small states in the preliminary stage; in column (III) land area is added as an additional excluded instrument; in column (V) the excluded instrument and endogenous treatment variables are based on those used in column V of Table 3; only selected control covariates shown; ‘treatment effect’ refers to the chosen endogenous variable; for the FDR estimator, row entries for the control covariates indicate the extent of overlap in the distribution across groups of the (binary) instrument; ‘Probability effect = 0.1’ refers to a Wald test of the null hypothesis that the treatment effect is equal to 0.1; standard errors, given in parentheses, are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Sources: Authors’ estimates, see Appendix A.

Table 7: Preferred specification for alternative periods

	1960-2000		1980-2000		1990-2000	
	(I)	(II)	(III)	(IV)	(V)	(VI)
	IV-LIML	FDR	IV-LIML	FDR	IV-LIML	FDR
	b/se	*overlap*	b/se	*overlap*	b/se	*overlap*
Treatment effect	0.143	0.102	0.034	0.072	-0.182	0.341
	(0.08)*	(0.05)*	(0.12)	(0.11)	(0.19)	(0.23)
Per capita GDP	-0.825	*0.221*	-1.573	*0.181*	-0.938	*0.180*
	(0.30)***	.	(0.40)***	.	(0.71)	
Level of policy	1.811	*0.154*	2.060	*0.112*	0.601	*0.047*
	(0.45)***	.	(0.75)***	.	(0.55)	
Life expectancy	0.004	*0.301*	0.072	*0.427*	0.086	*0.392*
	(0.03)	.	(0.04)**	.	(0.07)	
Geography	0.337	*0.587*	0.671	*0.558*	0.178	*0.540*
	(0.21)	.	(0.28)**	.	(0.51)	
Ethnic fractionalization	0.398	*-0.196*	0.985	*-0.162*	-0.274	*-0.049*
	(0.56)	.	(0.72)	.	(1.52)	
Coastal pop. density	0.001	*-0.045*	0.001	*-0.086*	0.001	*-0.136*
	(0.00)***	.	(0.00)**	.	(0.00)**	
Primary schooling	2.731	*0.350*	1.580	*0.236*	-0.720	*0.082*
	(0.97)***	.	(1.03)	.	(1.75)	
Price of invest. goods	-0.005	*-0.066*	0.004	*-0.340*	-0.015	*-0.097*
	(0.00)	.	(0.00)	.	(0.01)	
Malaria risk	-1.180	*-0.544*	-0.962	*-0.451*	-2.093	*-0.300*
	(0.59)**	.	(0.86)	.	(0.88)**	
Scale of excl'd instrument	Continuous	Binary	Continuous	Binary	Continuous	Binary
Number of obs.	74	74	75	75	70	70
R-square	0.70	-	0.64	-	0.50	-
Weak identification stat.	18.24	-	27.96	-	19.44	-
Stock-Wright S stat.	4.58**	-	0.08	-	0.83	-
Probability effect = 0.1	0.58	-	0.59	-	0.14	-

significance level: * 10%; ** 5%; *** 1%

Notes: Excluded instrument and endogenous treatment variables follow those used in column VI of Table 4 but in each case they are based on the relevant sub-periods; only selected control covariates shown; all specifications include regional dummies for SSA, Asia and Latin America and the Caribbean; 'Probability effect = 0.1' refers to a Wald test of the null hypothesis that the treatment effect is equal to 0.1; other reported statistics are as described in the text; for the FDR estimator, row entries for the control covariates indicate the extent of overlap in the distribution across groups of the (binary) instrument; 'treatment effect' refers to endogenous Aid/GDP; standard errors, given in parentheses, are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Sources: Authors' estimates, see Appendix A.

data. In many important areas of inquiry, longstanding debates with respect to causal impacts persist despite improved methods and improved data availability. Third, the formation of reasonable expectations about the likely returns to foreign assistance has been greatly facilitated by the application of modern growth theory. Finally, there is an increasing recognition that many of the key interventions pursued by foreign aid will only result in positive growth outcomes over long time horizons.

We started by replicating RS08. Subsequently, we developed a preferred specification, with a fuller set of regional fixed effects and indicators of initial human capital and geographic conditions. These were drawn from modern growth theory and included primary schooling, coastal population density and malaria risks. Consistent with best practice in the programme evaluation literature we excluded covariates, such as revolutions and institutional performance, which represent potential channels through which aid affects growth. To account for endogeneity in aid allocations, we (i) corrected errors in the implementation of the RS08 instrumentation strategy; (ii) employed aid per capita in place of Aid/GDP to preclude spurious correlation with the chosen instruments (particularly colonial origin); (iii) introduced donor-specific fixed effects; and (iv) accounted for selection bias through a Heckman correction. In the process, we deployed robust regression estimators which adjust for heterogeneity across countries and introduced a new doubly robust estimator that can be used in instrumental variable contexts. Overall, we believe our approach represents the most carefully developed empirical strategy employed in the aid-growth literature to date.

Our results provide strong support for the view that the average treatment effect of aid on growth is positive in both the 1970-2000 and 1960-2000 periods. The flexible doubly robust estimator places the point estimate of the long run elasticity of growth with respect to the share of aid in recipient GDP in a range between 0.10 and 0.23. These point estimates are consistent with the view that foreign aid stimulates aggregate investment and also may contribute to productivity growth, despite some fraction of aid being allocated to consumption.

With respect to confidence intervals, there is no evidence that would lead us to reject the prior, suggested in RS08, that the elasticity of growth to foreign aid is around 0.1. While in some specifications the (95 per cent) confidence interval is not restricted to the positive domain, the preponderance of macro-evidence points to a positive causal impact of aid on growth over long time frames. When combined with the evidence at the micro- and meso-levels, a consistent case for aid effectiveness emerges. There is no paradox. We find ourselves in a similar position to Winters (2004) in his review of the implications of trade liberalization for growth. While he concludes that trade liberalization stimulates growth over the long-run and on average, he adds that: ‘For a variety of reasons, the level of proof remains a little less than one might wish but the preponderance of evidence certainly favours that conclusion.’ (2004: F18). Similarly, we conclude that the bleak pessimism of much of the recent aid-growth literature is unjustified and the associated policy implications drawn from this literature are often inappropriate and unhelpful. Aid has been and remains an important tool for enhancing the development prospects of poor nations.

Furthermore, unlike the relatively straightforward policy recommendation of maintaining low tariff barriers, the complex and idiosyncratic process of managing aid to spark and sustain growth is subject to considerable learning. Nearly all participants in the aid-growth debate, not least these authors, recognize the potential for aid to do better, particularly in fostering productivity growth. The evidence indicates that sustaining foreign assistance programmes at reasonable levels can be expected to enhance the living standards of more than a billion of the world's poorest people. Abolishing foreign aid, or drastically cutting it back, would be a mistake and is not warranted by any reasonable interpretation of the evidence. The challenge is to improve foreign assistance effectiveness so that living standards in poor countries are substantially advanced over the next three decades.

References

Acemoglu, D., and S. Johnson (2007). 'Disease and Development: The Effect of Life Expectancy on Economic Growth'. *Journal of Political Economy*, 115 (6): 925–85.

Alesina, A., and D. Dollar (2000). 'Who Gives Foreign Aid to Whom and Why?' *Journal of Economic Growth*, 5 (1): 33–63.

Angrist, J. D. (2004). 'Treatment Effect Heterogeneity in Theory and Practice'. *Economic Journal*, 114 (494): C52–C83.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). 'Identification of Causal Effects Using Instrumental Variables'. *Journal of the American Statistical Association*, 91 (434): 444–55.

Angrist, J. D., and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Arndt, C., S. Jones, and F. Tarp (2007). 'Aid and Development: The Mozambican Case'. In S. Lahiri (ed.), *Frontiers of Economics and Globalization: Theory and Practice of Foreign Aid*. Amsterdam: Elsevier.

Ashraf, Q. H., A. Lester, and D. N. Weil (2008). 'When Does Improving Health Raise GDP?'. *NBER Working Paper* 14449, National Bureau of Economic Research, Inc.

Banerjee, A. V., and E. Duflo (2009). 'The Experimental Approach to Development Economics'. *Annual Review of Economics*, 1: 151–78.

Baum, C. F., M. E. Schaffer, and S. Stillman (2007). 'Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing'. *Stata Journal*, 7 (4): 465–506.

Berg, E. (1993). 'Rethinking Technical Cooperation: Reforms for Capacity Building in Africa'. Technical Report, Regional Bureau for Africa, United Nations Development Programme, New York.

Berthélemy, J.-C., and A. Tichit (2004). 'Bilateral Donors' Aid Allocation Decisions – A Three-dimensional Panel Analysis'. *International Review of Economics and Finance*, 13 (3): 253–74.

Blundell, R., and M. Costa Dias (2008). 'Alternative Approaches to Evaluation in Empirical Microeconomics'. *IZA Discussion Paper* 3800.

Boone, P. (1994). 'The Impact of Foreign Aid on Savings and Growth'. *London School of Economics CEP Working Paper* (677).

Bourguignon, F., and M. Sundberg (2007). 'Aid Effectiveness: Opening the Black Box'. *American Economic Review*, 97 (2): 316–21.

Bun, M., and F. Windmeijer (2007). 'The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models'. *CeMMAP Working Paper* CWP08/07.

Burnside, C., and D. Dollar (2000). 'Aid, Policies, and Growth'. *American Economic Review*, 90 (4): 847–68.

Burnside, C., and D. Dollar (1997). 'Aid, Policies, and Growth'. Technical Report 1777, The World Bank.

Card, D. (2001). 'Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems'. *Econometrica*, 69 (5): 1127–60.

Cassen, R., and Associates (1994). *Does Aid Work?*. Oxford: Clarendon Press.

Chao, J. C., and N. R. Swanson (2005). 'Consistent Estimation with a Large Number of Weak Instruments'. *Econometrica*, 73 (5): 1673–92.

Choi, I., and P. C. B. Phillips (1992). 'Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations'. *Journal of Econometrics*, 51 (1-2): 113–50.

Clemens, M., and S. Bazzi (2009). 'Blunt Instruments: On Establishing the Causes of Economic Growth'. *Center for Global Development Working Paper*, 171.

Cohen, D., and M. Soto (2007). 'Growth and Human Capital: Good Data, Good Results'. *Journal of Economic Growth*, 12: 51–76.

Collier, P. (2007). *The Bottom Billion: Why the Poorest Countries are Failing and What Can be Done About It*. Oxford: Oxford University Press.

Collier, P., and J. W. Gunning (1999). 'Why Has Africa Grown Slowly?'. *The Journal of Economic Perspectives*, 13 (3): 3–22.

Collier, P., and A. Hoeffler (2004). 'Aid, Policy and Growth in Post-Conflict Societies'. *European Economic Review*, 48 (5): 1125–45.

Dalgaard, C.-J., and L. Erickson (2009). 'Reasonable Expectations and the First Millennium Development Goal: How Much Can Aid Achieve?'. *World Development*, 37 (7): 1170–81.

Dalgaard, C.-J., H. Hansen, and F. Tarp (2004). 'On The Empirics of Foreign Aid and Growth'. *Economic Journal*, 114 (496): F191–F216.

Deaton, A. (2009). 'Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development'. *NBER Working Paper* 14690, National Bureau of Economic Research, Inc.

Djankov, S., J. Montalvo, and M. Reynal-Querol (2008). 'The Curse of Aid'. *Journal of Economic Growth*, 13 (3): 169–94.

Easterly, W. (2009). 'How the Millennium Development Goals are Unfair to Africa'. *World Development*, 37 (1): 26–35.

Easterly, W. (2003). 'Can Foreign Aid Buy Growth?'. *Journal of Economic Perspectives*, 17 (3): 23–48.

Easterly, W. (1999). 'The Ghost of Financing Gap: Testing the Growth Model Used in the International Financial Institutions'. *Journal of Development Economics*, 60 (2): 423–38.

Easterly, W., R. Levine, and D. Roodman (2004). 'Aid, Policies, and Growth: Comment'. *American Economic Review*, 94 (3): 774–80.

Frankel, J. A., and D. Romer (1999). 'Does Trade Cause Growth?'. *American Economic Review*, 89 (3): 379–99.

Hansen, H., and F. Tarp (2001). 'Aid and Growth Regressions'. *Journal of Development Economics*, 64 (2): 547–70.

Hansen, H., and F. Tarp (2000). 'Aid Effectiveness Disputed'. *Journal of International Development*, 12 (3): 375–98.

Hauk, W., and R. Wacziarg (2009). 'A Monte Carlo Study of Growth Regressions'. *Journal of Economic Growth*, 14 (2): 103–47.

Heckman, J. J. (1979). 'Sample Selection Bias as a Specification Error'. *Econometrica*, 47 (1), 153–61.

Holland, P. W. (1986). 'Statistics and Causal Inference'. *Journal of the American Statistical Association*, 81 (396): 945–60.

IDA15 (2008). 'Additions to IDA Resources: Fifteenth Replenishment'. *Technical Report*, International Development Association.

Imbens, G. W. (2004). 'Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review'. *The Review of Economics and Statistics*, 86 (1): 4–29.

Imbens, G. W., and J. D. Angrist (1994). 'Identification and Estimation of Local Average Treatment Effects'. *Econometrica*, 62 (2): 467–75.

Imbens, G. W., and J. M. Wooldridge (2009). 'Recent Developments in the Econometrics of Program Evaluation'. *Journal of Economic Literature*, 47 (1): 5–86.

Kanbur, R. (2006). 'The Economics of International Aid'. In S. Kolm and J. Mercier Ythier (eds), *Handbook of the Economics of Giving, Reciprocity and Altruism* (Vol. 2). Amsterdam: North Holland, Elsevier.

Kronmal, R. A. (1993). 'Spurious Correlation and the Fallacy of the Ratio Standard Revisited'. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 156 (3): 379–92.

La Porta, R. L., F. L. de Silanes, and A. Shleifer (2008). 'The Economic Consequences of Legal Origins'. *Journal of Economic Literature*, 46 (2): 285–332.

Lunceford, J. K., and M. Davidian (2004). 'Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study'. *Statistics in Medicine*, 23: 2937–60.

Mankiw, N. G., D. Romer, and D. N. Weil (1992). 'A Contribution to the Empirics of Economic Growth'. *The Quarterly Journal of Economics*, 107 (2): 407–37.

Masud, N., and B. Yontcheva (2005). ‘Does Foreign Aid Reduce Poverty? Empirical Evidence from Nongovernmental and Bilateral Aid’. *International Monetary Fund Working Paper* 05/100, International Monetary Fund.

Mishra, P., and D. L. Newhouse (2007). ‘Health Aid and Infant Mortality’. *International Monetary Fund Working Paper* 07/100, International Monetary Fund.

Mosley, P. (1987). *Overseas Aid: Its Defence and Reform.* Brighton: Wheatshead Books.

Mosley, P., J. Hudson, and S. Horrell (1992). ‘Aid, the Public Sector and the Market in Less Developed Countries: A Return to the Scene of the Crime’. *Journal of International Development*, 4 (2): 139–50.

Moyo, D. (2009). *Dead Aid: Why Aid Is Not Working and How There Is a Better Way for Africa*. London: Allen Lane.

Murray, M. P. (2006). ‘Avoiding Invalid Instruments and Coping with Weak Instruments’. *Journal of Economic Perspectives*, 20 (4): 111–32.

Nunn, N. (2008). ‘The Long-Term Effects of Africas Slave Trades’. *The Quarterly Journal of Economics*, 123 (1): 139–76.

Papanek, G. F. (1972). ‘The Effect of Aid and Other Resource Transfers on Savings and Growth in Less Developed Countries’. *Economic Journal*, 82 (327): 935–50.

Papanek, G. F. (1973). ‘Aid, Foreign Private Investment, Savings, and Growth in Less Developed Countries’. *Journal of Political Economy*, 81 (1): 120–30.

Rajan, R., and A. Subramanian (2007). ‘Does Aid Affect Governance?’. *AEA Papers and Proceedings*, 97 (2): 322–27.

Rajan, R. G., and A. Subramanian (2008). ‘Aid and Growth: What Does the Cross-Country Evidence Really Show?’. *The Review of Economics and Statistics*, 90 (4): 643–65.

Riddell, R. C. (2007). *Does Foreign Aid Really Work?.* Oxford: Oxford University Press.

Robins, J. M., and A. Rotnitzky (1995). ‘Semiparametric Efficiency in Multivariate Regression Models with Missing Data’. *Journal of the American Statistical Association*, 90 (429): 122–29.

Roodman, D. (2009). ‘A Note on the Theme of Too Many Instruments’. *Oxford Bulletin of Economics and Statistics*, 71 (1): 135–58.

Roodman, D. (2007). ‘The Anarchy of Numbers: Aid, Development, and Cross-Country Empirics’. *World Bank Economic Review*, 21 (2): 255–77.

Rubin, D. B. (1978). ‘Bayesian Inference for Causal Effects: The Role of Randomization’. *The Annals of Statistics*, 6 (1): 34–58.

Rubin, D. B. (1976). ‘Inference and Missing Data’. *Biometrika*, 63 (3): 581–92.

Rubin, D. B. (1974). ‘Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies’. *Journal of Educational Psychology*, 66 (5): 688–701.

Sachs, J. (2005). *The End of Poverty: Economic Possibilities for Our Time*. New York: Penguin Press.

Sachs, J. (2006). 'Why Aid Does Work'. BBC News, available at http://news.bbc.co.uk/1/hi/sci/tech/4210122.stm.

Sala-i-Martin, X., G. Doppelhofer, and R. I. Miller (2004). 'Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach'. *American Economic Review*, 94 (4): 813–35.

Tarp, F. (2006). 'Aid and Development'. *Swedish Economic Policy Review*, 13 (2): 9–61.

Tarp, F., C. F. Bach, H. Hansen, and S. Baunsgaard (1999). 'Danish Aid Policy: Theory and Empirical Evidence'. In K. Gupta (ed.), *Foreign Aid: New Perspectives*. Boston: Kluwer Academic Publishers.

Temple, J. (2000). 'Growth Regressions and What the Textbooks Don't Tell You'. *Bulletin of Economic Research*, 52 (3): 181–205.

Thorbecke, E. (2007). 'The Evolution of the Development Doctrine, 1950-2005'. In G. Mavrotas and T. Shorrocks (eds), *Advancing Development: Core Themes in Global Economics*. New York: Palgrave Macmillan.

Tsikata, T. (1998). 'Aid Effectiveness – A Survey of the Recent Empirical Literature'. *International Monetary Fund Working Paper* 98/1, International Monetary Fund.

Winters, L. A. (2004). 'Trade Liberalisation and Economic Performance: An Overview'. *Economic Journal*, 114 (493): F4–F21.

Wooldridge, J. M. (2005). 'Violating Ignorability of Treatment by Controlling for Too Many Factors'. *Econometric Theory*, 21 (05): 1026–28.

World Bank (2008). *Annual Review of Development Effectiveness 2008: Shared Global Challenges*. Independent Evaluation Group, available at: http://go.worldbank.org/U2T30HQKG0

Appendix A: Data sources and variable description

The base data is from Rajan and Subramanian (2008), kindly supplied by the authors. Data from other sources are also used as deemed necessary. Explanatory variables from the preliminary stage regressions are all taken from RS (2008). Other variables and their respective sources are described as follows:

Variables from Rajan and Subramanian (2008)	Description	Source
Initial per capita GDP	Log of per capita (PPP) GDP at the beginning of the relevant time period.	Rajan and Subramanian (2008)
Initial level of policy	The Sachs-Warner trade policy index as updated by Wacziarg and Welch and prevailing at the beginning of the relevant time horizon or the year closest to it.	Rajan and Subramanian (2008)
Initial life expectancy	Life expectancy at birth in years at the beginning of the relevant time period.	Rajan and Subramanian (2008)
Geography	Average of number of frost days and tropical land area.	Rajan and Subramanian (2008)
Institutional quality	ICRGE index averaged over the period 1986–1995.	Rajan and Subramanian (2008)
Initial inflation	Average annual rate of growth of CPI-based inflation for the first five years of the relevant time horizon.	Rajan and Subramanian (2008)
Initial M2/GDP	The ratio of M2/GDP for the first five years of the relevant time horizon.	Rajan and Subramanian (2008)
Initial budget balance/GDP	The ratio of general government budget balance to GDP for the first five years of the relevant time horizon.	Rajan and Subramanian (2008)
Revolutions	Average number of revolutions per year in the relevant time horizon.	Rajan and Subramanian (2008)
Land area	Recipient land area.	Rajan and Subramanian (2008)
Ethnic fractionalization	Average of five different indices of ethno-linguistic fractionalization which is the probability of two random people in a country not speaking the same language.	Rajan and Subramanian (2008)

Additional covariates	Description	Source
Coastal population density	Coastal (within 100 km of coastline) in 1965.	Sala-i-Martin et al. (2004) [23]
Primary schooling	Enrolment rate in primary education in 1960.	Sala-i-Martin et al. (2004)
Price of investment goods	Average investment price level between 1960 and 1964 on purchasing power parity basis.	Sala-i-Martin et al. (2004)
Malaria risk	Index of malaria prevalence in 1966.	Sala-i-Martin et al. (2004)

Variables used to (re)calculate the treatment effect	Description	Source
Aid	Refers to Official Development Assistance (ODA), total net disbursements in current prices (USD millions). ODA is defined as flows to developing countries and multilateral institutions provided by official agencies, including state and local governments, or by their executive agencies.	OECD-DAC data base online (2009) [24]
Population	Total population, all residents except for refugees not permanently settled in the country of asylum, who are generally considered part of the population of their country of origin.	World Development Indicators (WDI) CD ROM; World Bank (2008)
Gross Domestic Product	GDP in current US dollars.	World Development Indicators (WDI) CD ROM; World Bank (2008)

[23] Accessible at: http://www.aeaweb.org/articles/issue_detail_datasets.php?journal=AER&volume=94&issue=4&issue_date=September%202004

[24] Accessible at: http://www.oecd.org/dac/stats