UNITED NATIONS UNIVERSITY

UNU-WIDER

World Institute for Development Economics Research

Working Paper No. 2010/96

Aid, Growth, and Development

Have We Come Full Circle?

Channing Arndt[1], Sam Jones[1], and Finn Tarp[1,2]

September 2010

Abstract

The micro-macro paradox has been revived. Despite broadly positive evaluations at the micro and meso-levels, recent literature doubts the ability of foreign aid to foster economic growth and development. This paper assesses the aid-growth literature and, taking inspiration from the program evaluation literature, we re-examine key hypotheses. In our findings, aid has a positive and statistically significant causal effect on growth over the long run, with confidence intervals conforming to levels suggested by growth theory. Aid remains a key tool for enhancing the development prospects of poor countries.

Keywords: foreign aid, growth, aid effectiveness, causal effects.
JEL classification: O1, O4, F35, C21

[1]University of Copenhagen (channingarndt@gmail.com) (sam.jones@econ.ku.dk); [2]UNU-WIDER Helsinki (tarp@wider.unu.edu).

This study has been prepared within the UNU-WIDER project on Foreign Aid: Research and Communication.

UNU-WIDER gratefully acknowledges the financial contributions to the research programme by the governments of Denmark (Royal Ministry of Foreign Affairs), Finland (Ministry for Foreign Affairs), Sweden (Swedish International Development Cooperation Agency—Sida) and the United Kingdom (Department for International Development—DFID).

ISSN 1798-7237 ISBN 978-92-9230-334-1

Acknowledgements

We thank Tony Addison, Ernest Aryeetey, Pranab Bardhan, Bruce Bolnick, Imed Drine, Jan Willem Gunning, Gerry Helleiner, Paul Isenman, Homi Kharas, Dirk Krueger, David Roodman, Erik Thorbecke, Alan Winters, and Adrian Wood for encouragement and most valuable comments. The same goes for the participants at conferences and seminars held by the African Economic Research Consortium (AERC), the Bergen Resource Center for International Development at CMI Norway, the Brookings Institution, central ministries in Mozambique, Tanzania and Vietnam, Cornell University, the European Union Development Network (EUDN) (organized by CERDI, Clermont-Ferrand, France), the Helsinki Center for Economic Research (HECER), NORAD Norway, OECD Paris, the United Nations HQ (organized by UNU-ONY), the University of Copenhagen, the University of Ghana, the UNU Conference of Directors (CONDIR), and UNU-WIDER. 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.

UNU World Institute for Development Economics Research (UNU-WIDER)
Katajanokanlaituri 6 B, 00160 Helsinki, Finland

Typescript prepared by Lorraine Telfer-Taivainen at UNU-WIDER

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 labeled this the 'micro–macro paradox'. Now, after two decades of intense analytical work using new theory, new data and new empirical methodologies, it would appear that the paradox has been revived. At the macro-level, Rajan and Subramanian (2008) conclude 'it is difficult to discern any systematic effect of aid on growth'. Nevertheless, evaluations of aid effectiveness at the microeconomic level continue to indicate positive rates of return (World Bank 2008). Also, an increasing number of rigorous microeconomic impact evaluations have demonstrated the potential for well-designed project interventions to generate positive results (Banerjee and Duflo 2009).

This paper has two objectives. First, we attempt to provide a balanced and up-to-date assessment of the aid–growth literature. We observe that, while aid may have very high returns in specific circumstances, expectations surrounding the average potency of aid have been excessive.[1] 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. In the spirit of Angrist and Pischke (2010), we add new insights by applying recent methodological advances in the programme evaluation literature to a macroeconomic question—the causal effect of aid on growth. In particular, we provide an extension of the doubly robust methodology of Robins and Rotznitzky (1995). We find that, on average, foreign aid exerts a long-run positive effect on growth in developing countries.

The remainder of this paper is structured as follows. Section 2 provides a literature review. Section 3 outlines our counterfactual model for evaluating aid's impact on growth, while Section 4 details the estimation approach and presents principal empirical findings. A final section summarizes and concludes. Further information on countries and data sources are found in the Appendices.

2 Literature review

Scholarship on the relationship between aid and growth is voluminous and convoluted. Given the existence of numerous detailed literature reviews (e.g., Tsikata 1998; Hansen and Tarp 2000; Easterly 2003; Kanbur 2006; Roodman 2007; Thorbecke 2007), this section provides a relatively brief survey of how the aid–growth debate has evolved.

2.1 Earlier generations

Studies of the aid–growth relationship from the 1970s until recently can be classified into three generations, each influenced by dominant theoretical paradigms as well as available

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

empirical tools. The first two generations were inspired by simple models of the growth process, i.e. 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 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 increased savings and investment in recipient countries (see Papanek 1972, 1973). As the detailed survey in Hansen and Tarp (2000) testifies, first generation studies show that aid tends 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, a second generation of literature explored 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 these studies consistently indicate a positive link between aid and investment. While a majority of the aid–growth studies of this generation also suggested a positive impact, the result that captured attention was Paul Mosley's micro–macro paradox. This puzzle raised doubts concerning the appropriateness of the underlying growth model and the empirical techniques used. 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 is the argument that growth is less related to physical capital investment 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 broader 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. 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.

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 panel data, allowing analysts to look at changes both across and within countries over time. 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, the contribution by Burnside and Dollar (2000) came to exert a significant influence on policy. These authors made an argument for conditional aid effectiveness, specifically: '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). However, these results were subject to substantial criticism. For example, Hansen and Tarp (2001) found 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 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. 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 useable information regarding the causal impact of aid.

2.2 Recent studies

More recently, a fourth generation of literature has emerged. A distinctive aspect of this generation is the view that aid's aggregate impact on economic growth is non-existent. A leading paper that appears to establish this result is Rajan and Subramanian (2008). These authors find no systematic effect of aid on growth regardless of the estimation approach, the time period and the type of aid analyzed. Explanations for non-positive aggregate effects of aid often refer to political economy dynamics. Aid inflows can 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.

Fourth generation scholars have also become increasingly sceptical about our ability to make valid causal inferences with respect to complex aggregate phenomena, such as the determinants of economic growth. In particular, previous methods used to deal with endogeneity have been subject to criticism. For example, there is increasing awareness that dynamic panel (system) GMM methods—frequently employed in the third generation of the aid literature—are not a panacea. The concern that weak instruments typically bias coefficient estimates towards their unadjusted counterparts (e.g., OLS or panel fixed effects estimates) applies as much to panel GMM as to cross-section estimators. Bun and Windmeijer (2010) show that the weak instrument problem (previously attributed mainly to the Arellano-Bond estimator) may be equally problematic in the system approach. Also, for the Blundell-Bond (system GMM) 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, Hauk and Wacziarg (2009) 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) alerts that the Blundell-Bond estimator also 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. Finally, internal instruments do not prevent bias arising from systematic measurement error in the endogenous regressors. This is an important limitation in the context of aid–growth regressions.

In response to these concerns, alternative methods to assess aid effectiveness have come to the fore. These often eschew cross-country macroeconomic analysis in favor of specific micro- and meso-outcomes (Temple 2010; Riddell 2007). 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. Alongside cautious optimism surrounding the potential efficacy of microeconomic policy interventions, financed either directly by donors through projects or indirectly via budget or sector-wide support, these findings give new sustenance to the micro–macro paradox. Indeed, with few exceptions (e.g., Sachs 2005, 2006), findings at the micro- and meso-levels have not been deployed to argue that aid is effective on aggregate. This is despite increasing evidence that meso-level outcomes can add up to substantial macroeconomic effects (Cohen and Soto 2007).

2.3 Rationale for continued macro-analysis

In light of the above methodological concerns, it is helpful to reflect on the value of assessing the macroeconomic impact of foreign aid. In the first place, serious empirical challenges have not dissuaded economists from investigating other complex questions (Angrist and Pischke 2010). For example, there are numerous parallels between the problem of estimating the causal impact of aid on growth and the causal impact of schooling on earnings, which has also generated a controversial and voluminous literature. Both problems are likely to be characterized by endogenous selection, heterogeneous treatment responses, and mis-measurement of treatment input (both in terms of quality and quantity). Considerable effort has been expended in the analysis of large, high-quality schooling datasets by some of the most skilled econometricians in the profession. Even so, 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).

If the profession has experienced serious difficulties estimating the causal effect of schooling on earnings in developed countries, then it should not be surprising that estimating the impact of aid on growth in developing countries is contentious. However, it is difficult to deny that the aid–growth issue is both compelling and relevant. In developed countries, policy-makers and the wider public continue to ask whether aid is a cost effective use of taxpayer money on aggregate. Today, 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. The financial crisis of 2008-09 has also highlighted the importance of public spending to stabilize and stimulate economic activity. While foreign aid has multiple objectives, economic growth is central among them. If the economics profession as a whole were to abandon the question of aid's impact on growth, it would leave the issue open to speculative and potentially unhelpful contributions.

2.4 Formulating an appropriate prior

A key aim of empirical analysis is to falsify or discriminate between competing hypotheses. Consequently, it is necessary to make explicit the prior upon which empirical testing is focused. With respect to the effect of foreign aid on economic growth, relatively few studies address the issue of an appropriate prior. A recent exception is Rajan and Subramanian (2008) who consider aid in a standard neoclassical growth model. Assuming that aid only augments physical capital investment and has no effect on productivity; they derive that a one percentage point increase in the ratio of aid to GDP should be expected to raise the growth rate of per capita GDP 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 point 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. This is considerably less than the predictions based on Harrod-Domar models. In sum, growth theory points towards more modest expectations with respect to the potency of aid (see also Dalgaard and Erickson 2009).

A related issue is the appropriate time-frame over which any growth effects accruing from aid can be expected to materialize. Various factors may exert a cumulative but not immediate impact on the rate of income growth. For example, changes in education move only slowly at the aggregate level and have 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 can take 30 years or more for per capita incomes to return to pre-eradication levels. They also find that significant increases in life expectancy at birth only begin to have a modest positive effect on incomes after about a 35 year lag.

Ashraf et al. (2008) focus on demographic trends as a result of disease eradication. Productivity effects, demand effects, and complementary policies may speed the realization of growth benefits from health gains. Nevertheless, a series of considerations indicate that the aid–growth relationship is only likely to emerge over a long time horizon. As noted, many aid investments, such as in education, health, and institution building are long term in nature; and growth theory indicates that the contribution of these investments to growth is likely to be relatively modest. When these observations are combined with the volatility of growth in most developing countries and the high degree of measurement error inherent in nearly all the variables of interest, relatively long time frames would appear to be necessary to reliably detect the aid–growth relationship.

3 A counterfactual model for evaluating aid

3.1 Basic framework

In this section and in Section 4, 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 Costas Dias (2009), 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). 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 $i = 1, \ldots, N$ 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 is either treated or is not, only one of these outcomes is 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 above requirements and motivates instrumental variable (IV) methods. Where IVs are used, analogous versions of these requirements can be set out. 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 potential outcomes. Using the standard econometric vocabulary, these requirements imply that the instruments should be relevant and valid (i.e., exclusion restrictions hold).

3.2 Extended doubly robust estimator

In the treatment and control framework, established techniques exist to deal with relevant causal challenges. In particular, doubly robust estimators (Robins and Rotznitzky 1995) correspond to what Imbens and Wooldridge describe as 'best practice' (2009: 25). A number of such 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. As long as *either* the propensity score *or* the linear regression is correctly specified, consistent estimates are generated.

More formally, propensity scores can be derived from estimates on a logistic form:

(1) $$\pi(x_i) = \Pr(W_i = 1 | X_i = x_i) = \exp\left(X_i^{'}\lambda\right) / \left[1 + \exp\left(X_i^{'}\lambda\right)\right]$$

where W represents a binary treatment variable and X a vector of controls. The predicted probabilities ($\hat{\pi}_i$) are then used as inverse probability weights in the least squares problem:

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

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

Doubly robust estimators do not automatically lend themselves to investigation of the aid–growth issue. First, in the evaluation literature, microeconomic units, such as individuals exposed to labor market programmes, are often in 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. Second, the predominant focus of the literature is on binary as opposed to continuous treatment variables. 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). Third, existing doubly robust estimators are only suitable for situations in which there is random selection into treatment, or at least 'selection on observables'. They cannot be applied to cases where there is endogenous selection into treatment, such as in the aid–growth case.

In the remainder of this section we extend the IPWLS estimator to an instrumental variables setting, thereby allowing doubly robust techniques to be applied to the aid–growth issue. To do so, we begin with a standard 2SLS set-up containing an endogenous treatment variable (W), a (presumed valid) single continuous instrumental variable (Z^*), an outcome of interest (Y), and a vector of additional controls (X):

(3a) $$W_i = \rho Z_i^* + X_i^{'}\theta + \upsilon_i$$

(3b) $$Y_i = \alpha \widehat{W}_i + X_i^{'}\beta + \varepsilon_i$$

where υ and ε are independent, mean zero error terms. The critical step required to extend the IPWLS estimator requires we dichotomize the instrument, creating a binary 'assignment-to-treatment' variable. In the case of aid, the resulting variable (Z) can be thought of as dividing the sample into small and large aid recipients, according to the chosen instrument for aid.

More formally, dichotomization is a simple rule:

$$Z_i = g(Z_i^*) = \begin{cases} 1 \text{ if } & Z_i^* > c \\ 0 \text{ otherwise} \end{cases}$$

where c is a threshold indicating 'high' aid recipients. Assuming the parameter for α in equation (3b) is consistently identified, dichotomization of the instrument is not problematic. By the properties of expectations, we assume any function of the instrument is orthogonal to the errors (ε) in the outcome equation of interest. Thus, it follows that:

$$\mathrm{E}[Z_i^* \varepsilon_i \,|X_i] = \mathrm{E}[g(Z_i^*)\varepsilon_i \,|X_i] = \mathrm{E}[Z_i \varepsilon_i \,|X_i] = 0.$$

Following this logic, equations (1) and (2) can be modified by replacing the treatment variable W with the binary instrument Z. This is simply the reduced form of a (weighted) two stage least squares problem where Z instruments for the endogenous aid variable. To see this, note that the modified version of equation (3a), using the dichotomous instrument, looks like:

(4) $$W_i = \rho Z_i + X_i^{'}\theta + \nu_i$$

which yields the reduced form of the problem:

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

Ignoring weights, these coefficients are directly comparable to those in equation (2), where W_i has been replaced by Z_i. Specifically we now have: $\delta = \alpha\rho$; 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. Nevertheless, to compare against (unweighted) 2SLS results, we need to extract α from δ. This is calculated by dividing through by ρ, estimated from the first stage (with the same weights). It thus follows that the IV counterpart of the IPWLS estimator is weighted two-stage least squares, employing a dichotomous instrument and weights estimated from a propensity score procedure applied to the binary instrument. This new estimator is denoted IV-IPWLS, and is a useful basis from which to explore the relationship running from aid to growth.

4 Empirical analysis

We now turn to the empirical analysis. As indicated in Section 2, Rajan and Subramanian (2008) (henceforth RS08) present a thoughtful and highly influential contribution. It also achieves a number of methodological advances on previous studies. RS08 is widely understood as having re-established that aid has no impact on growth thus reviving the micro–macro paradox. However, a number of concerns question this fundamental conclusion. Once these concerns are addressed, the aid–growth relationship is shown to conform to appropriate priors (Section 2.4); and the micro–macro paradox disappears once again.

Following the recommendation of Temple (2010) to build explicitly on existing empirical work, our starting point is RS08. In this vein, a brief summary of their approach and core results are presented in Section 4.1. We then proceed to construct our estimator on the basis of progressive changes to the principal estimator and aid instrument employed by RS08. These changes are detailed in Sections 4.2 to 4.4. Section 4.4 contains our preferred approach and estimates. In Section 4.5, we subject these preferred results to a range of sensitivity tests,[2] and we finish in Section 4.6 by subjecting the aid instrument to validity tests.

4.1 RS08

In their principal approach, RS08 consider four periods: 1960-2000; 1970-2000; 1980-2000; and 1990-2000. In each period, their instrument for aid is generated from a preliminary stage regression estimated at the bilateral donor-recipient level using supply-side factors. These include past colonial relations, relative population sizes and interaction terms. The predicted Aid/GDP ratio estimated from this regression is aggregated across donors to give a fitted average ratio for each recipient. This then enters an instrumental variables estimation equivalent to the set-up in equations (3a) and (3b). Using the same notation, the fitted instrument corresponds to Z^*; the outcome variable (Y) is average real (PPP-adjusted) per capita GDP growth (over the entire period); and the treatment variable (W) is the continuous Aid/GDP measure.

Results from RS08 for the 1970-2000 period are replicated in column I of Table 1. This is a core result of their study. The generated instrument appears reasonably strong according to conventional measures, such as the first stage partial F-statistic; also, the coefficient on Aid/GDP is exactly in line with the prediction from their growth model but is not significant. RS08 conclude that there is no systematic (causal) effect of aid on growth. This is shown to hold when alternative 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 our view, RS08 advance the aid–growth literature in two ways. First, their principal approach investigates the aid–growth relationship using long-run averages, thus turning focus away from dynamic panel methods. We concur. Dynamic panel methods are subject to doubt

---

[2] Throughout, our focus is on the average relationship between aid and growth. We do not provide a detailed discussion of other growth determinants; however, our results are broadly in line with the extant literature.

given the expected cumulative effect of aid and corresponding concerns regarding the validity of internal instruments in GMM estimators (see Section 2.2). Their approach echoes the

Table 1: Alternative IV estimators and specifications, 1970-2000

	(I) 2SLS	(II) IV-IPWLS	(III) IV-LIML	(IV) IV-LIML	(V) IV-IPWLS
Aid / GDP	0.10	0.15*	0.11*	0.10	0.10**
	(0.07)	(0.08)	(0.07)	(0.06)	(0.04)
Initial per capita GDP	-1.41***	-1.67***	-0.84*	-1.24***	-1.38***
	(0.43)	(0.34)	(0.45)	(0.34)	(0.28)
Initial level of policy	2.14***	2.32***	2.65***	2.17***	2.47***
	(0.62)	(0.58)	(0.58)	(0.47)	(0.49)
Initial life expectancy	0.08*	0.05	0.11***	0.03	0.04
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)
Geography	0.61**	0.67***	0.59***	0.21	0.25
	(0.26)	(0.23)	(0.20)	(0.22)	(0.22)
Institutional quality	4.08*	5.11**	-	-	-
	(2.33)	(2.24)			
Revolutions	-1.41**	-1.10*	-	-	-
	(0.66)	(0.66)			
Coastal pop. density	-	-	-	0.00**	0.00***
				(0.00)	(0.00)
Primary schooling	-	-	-	2.50**	2.09**
				(1.16)	(1.01)
Malaria risk	-	-	-	-1.62**	-1.11
				(0.79)	(0.68)
Price of invest. goods	-	-	-	-0.01	-0.01**
				(0.00)	(0.00)
Civil liberties in 1972	-	-	-	-0.98	-0.66
				(0.60)	(0.50)
Air distance (log.)	-	-	-	0.12	0.18
				(0.38)	(0.33)
Scale of excluded instrument	Continuous	Binary	Continuous	Continuous	Binary
Regional dummies	SSA, EA	SSA, EA	SSA, A, LA	SSA, A, LA	SSA, A, LA
N	78	78	78	78	78
R-squared	0.59	0.66	0.50	0.65	0.73
Weak identification stat.	31.60	18.28	31.80	22.71	20.47
Stock-Wright LM S stat.	-	4.74	3.23	3.89	5.35
(probability)	-	0.030	0.072	0.048	0.021

Notes: Significance levels: *10%; **5%; ***1%. Endogenous variable is Aid/GDP, instrumented as per Rajan and Subramanian (2008; RS08); column (I) replicates RS08; column (II) implements an IV version of the inverse probability weighted least squares estimator (IV-IPWLS); column (III) drops 'bad controls' and redundant variables; column (IV) adds additional covariates from Sala-i-Martin et al. (2004) (a small number of observations are imputed to maintain sample size); column (V) applies the IV-IPWLS estimator to the specification in column (IV); cols (I) and (II) include regional dummies for sub-Saharan Africa (SSA) & East Asia (EA); cols (III) to (V) include regional dummies for SSA, Asia (A) and Latin America & the Caribbean (LA); intercept not shown; weak identification statistic is the first stage partial-F statistic in column (I), and the Kleibergen-Paap Wald F statistic in all others; standard errors, given in parentheses, are robust to arbitrary heteroskedasticity; dependent variable is mean real growth rate.

Source: Authors' estimates; Appendix C.

average OLS estimator proposed by Mankiw et al. (1992) (see also 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. Second, their supply-side approach to instrumentation represents the state of the art in the aid–growth literature and provides a potentially fruitful basis for exploring the long-run effects of aid.[3] Consequently, their contribution is an appropriate starting point for further empirical analysis.

4.2 Application of the doubly robust estimator

The extended doubly robust estimator assigns greater weight to countries which have the characteristics of recipients of substantial (negligible) aid inflows, but do not (do) receive large amounts of aid.[4] These countries are regarded as being more informative. As detailed in Section 3.2, application of the IV-IPWLS estimator requires a binary instrument. To derive this we take the fitted instrument from RS08's preliminary stage regression, sort countries in ascending order (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' ($Z_i = 1$).[5] 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.

Besides permitting application of doubly robust techniques, 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. Consequently, possible non-linear effects due to diminishing returns to aid are addressed by this dichotomization. Finally, as the instrument is derived from a preliminary stage regression (as per RS08), dichotomization provides a check against measurement error or misspecification in the preliminary stage.

In the development of our estimates, we focus on the 1970-2000 period. This allows the effects of aid to be considered over a generation of elapsed time. We believe that the 1970-2000 is the most relevant for consideration of the empirical question at hand. The shorter periods (1980-2000 and 1990-2000) may not allow sufficient time for the aid growth relationship to emerge. With respect to 1960-2000, many countries had not attained independence by 1960, particularly those in Africa. Further, even though the majority of French colonies achieved independence in 1960, the shift to independent administration was very gradual in most cases (Berg 1993). In contrast, by 1970 the large majority of developing countries had achieved independence and had operated independently for at least a few years, with Portuguese colonies being the prominent exception. While the 1970-2000 period is the

[3] Clemens and Bazzi (2009) provide a critique of the RS08 instrumentation strategy. This issue is taken up in Section 4.6.

[4] We employ a flexible doubly robust estimator as a sensitivity analysis in Section 4.5.

[5] We recognise up-front 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 the weighting scheme are presented in Section 4.4.

predominant focus of our analysis, results for our preferred approaches are presented for all time periods considered by RS08.

Using the same data, specification, and instrument as RS08, column II of Table 1 presents results from the (instrumental variables) IV-IPWLS estimator for the 1970-2000 period.[6] While greater weight is placed on more informative observations, one notes only small changes to the estimated coefficients compared to the (unweighted) 2SLS results given in column I. Dichotomization of the instrument leads to some loss of power, as shown by the Kleibergen-Paap Wald F weak identification statistic. Nevertheless, the overall explanatory power of the model, as measured by the R-squared, is improved and the standard error on the treatment effect is nearly the same. The end result is a moderate upward shift in the Aid/GDP point estimate to 0.15 (column II), which is significantly different from zero at the 10 per cent level. Thus, based on the same data, specification and instruments as RS08, application of a robust counterfactual method yields results that are both comparable to the point estimates obtained by RS08 and statistically significant.

4.3 Specification issues

Changes to the specification used by RS08 are justified for four reasons. First, given the relatively small sample size, inclusion of redundant variables may lead to a loss of efficiency and/or contribute to undesirable multicollinearity. In the present case, we note that the three macroeconomic initial conditions (inflation, money supply, budget balance) as well as ethnic fractionalization are insignificant in RS08's cross-section outcome regressions for all periods.

Second, 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 also 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.[7] 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. Such variables also may introduce unwanted reverse causality.

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 an ex post 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 relationships. A priori, a more plausible approach is to include a fuller set of regional dummies. Fourth, it is relevant to include additional variables that reflect initial socioeconomic conditions such as

[6] Note that in the implementation of the IV-IPWLS estimator it is the aid instrument that is dichotomised. The endogenous variable (Aid/GDP) is held continuous, thus one can directly compare the IV-LIML and IV-IPWLS coefficient estimates.

[7] Initial institutional conditions, on the other hand, are important and are captured (at least partially) by the Sachs and Warner trade policy index for the beginning of each period analysed as well as other proxies (see Appendix C).

education and health indicators, as well as additional geographic characteristics such trading distances. These variables are frequently seen as important determinants of growth and may also proxy for initial conditions; as such, they may explain some of the variation in the expected growth returns to foreign aid.

Columns III to V of Table 1 present results for these alternative specifications, retaining the same estimation strategy and sample as RS08.[8] Columns III and IV employ the limited information maximum likelihood instrumental variables (IV-LIML) estimator, with robust standard errors. This is chosen as it is generally taken to be more robust to weak instruments than 2SLS or cross-section GMM approaches (see Baum et al. 2007).[9] We also report additional test statistics that support inference in the presence of (weak) instruments. Column III revisits RS08's original specification dropping the contemporaneous outcome covariates and redundant variables, and adding an alternative set of regional dummies.[10] 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. We include variables identified by the authors which are among those with the highest posterior probability of inclusion and refer to initial conditions (see Appendix C for further details). To this we add civil liberties in 1972 and air distance to major ports. The first of these captures some dimensions of initial institutional quality, as well as the ability of citizens to bring the government to account (also often deemed relevant for aid effectiveness). Air distance is associated with export transaction costs and ease of access to developed markets and has recently been identified by Moral-Benito (2009) as a robust correlate of growth. Finally, in column V we apply the IV-IPWLS estimator to this modified specification.

The point estimates for Aid/GDP remain consistently in the domain suggested by theory and in all but one case are significantly different from zero. Although removal of the endogenous covariates (column III) leads to some loss of overall explanatory power, inclusion of additional controls (column IV) substantially boosts the power of the specification returning it to a level similar to the original specification (column I). The additional test statistics confirm the fitted instrument for aid remains strong. For example, the Stock-Wright S statistic, which is based on the reduced form regression and is robust to the presence of weak instruments (see Baum et al. 2007), shows a significant non-zero (positive) correlation between the endogenous regressor and the outcome variable.

4.4 Instrumentation strategy

As a third empirical step, we review the instrumentation strategy underlying the foregoing results. Our concerns with the RS08 instrumentation strategy lead to five modifications. First, there are potential errors in the calculation of average Aid/GDP flows used in the preliminary

---

8 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.

9 In the case of a single instrument and endogenous regressor, 2SLS, GMM and IV-LIML converge to the same point estimates. While CUE-GMM may be preferable (for efficiency reasons) in the presence of heteroskedasticity, IV-LIML is a computationally simpler estimator in the context of multiple instruments. We employ this IV estimator throughout alongside robust standard errors (using STATA's ivreg2 -liml robust- options). Nevertheless, we find immaterial differences in our main results if a CUE-GMM estimator is used in place of IV-LIML.

10 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.

stage regression. 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.[11] RS08 incorrectly treat these as missing. This is material because it distorts estimates for average bilateral aid flows over time. Consequently, it is necessary to re-estimate the bilateral aid dataset and calculate period averages for Aid/GDP and aid per capita setting missing entries to zero. The effect of this modification on the preliminary stage regression is shown in Table 2. Column I replicates the RS08 specification (only selected coefficients shown); column II employs the revised dependent variable in which missing Aid/GDP values are set to zero. The impact of this change appears moderate; and the pair wise correlation between the fitted values from these two models is 0.83.

Table 2: Alternative preliminary stage regressions, 1970-2000

	(I) OLS	(II) OLS	(III) OLS	(IV) OLS	(V) Heckman
Colonial relationship (dummy)	1.65***	2.09***	11.95***	-0.55	-0.88
	(0.24)	(0.19)	(1.62)	(2.08)	(2.20)
Currently a colony (dummy)	-0.97*	0.63	9.88***	14.14	24.48
	(0.56)	(0.45)	(3.81)	(21.15)	(36.71)
Common language (dummy)	0.07*	0.09***	1.36***	1.30**	1.30*
	(0.04)	(0.03)	(0.27)	(0.60)	(0.67)
Ratio of (initial) log. population	0.09***	0.05***	0.40***	0.32***	0.45***
	(0.01)	(0.00)	(0.04)	(0.06)	(0.08)
Ratio of log. population x colony	0.62***	0.77***	7.16***	3.32***	3.36***
	(0.11)	(0.08)	(0.69)	(0.72)	(0.77)
Dependent variable	Aid/GDP	Aid/GDP	Aid p.c.	Aid p.c.	Aid p.c.
Treatment of 'missing' aid values	Unknown	Zero	Zero	Zero	Zero
Metropole fixed effects & interactions	Yes	Yes	Yes	No	No
Donor fixed effects	No	No	No	Yes	Yes
Outcome and selection independence	-	-	-	-	9.56***
Number of obs.	3288	3286	3328	3328	3328
R-squared	0.42	0.31	0.26	0.21	-
F statistic	185.93	113.55	90.49	10.65	-

Notes: Significance levels: *10%; **5%; ***1%. Column (I) replicates Rajan and Subramanian's preliminary stage regression (2008, Table 4); columns (II) and (III) retain the same RHS specification, but alter the dependent variable (denoted in the table); column (IV) revises the specification, dropping metro pole (colony-specific) fixed effects and interactions (coefficients not shown); column (V) implements a Heckman correction, based on the specification in column (IV); 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; intercept not shown; standard errors are robust to arbitrary heteroskedasticity and intra-group correlation between aid recipients (except for columns I to III where standard errors assume homoskedasticity in order to replicate Rajan and Subramanian, 2008).

Source: Authors' estimates; Appendix C.

Second, in the RS08 strategy, recipient GDP occurs in the denominator of the dependent variable in the preliminary stage regressions. Following Kronmal (1993), inappropriate use of ratio variables may lead to substantial misinterpretation (or bias) in least squares regressions. This may arise if the denominator of the dependent variable is correlated with the RHS

[11] Confirmed in correspondence with the OECD-DAC Secretariat.

variables independently of the numerator of the dependent variable. 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.

Third, 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 include specific colonial relations. These are the colony dummies and population interactions for the major imperial powers of the eighteenth and nineteenth centuries. The institutional transplants and broader colonizing strategies pursued by these imperial powers were not alike, and they may have a persistent effect on income levels to the present day. This notion 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. The colonial relations variables are not orthogonal to growth and therefore should not be included in the preliminary stage regression explaining aid.

Fourth, it is apparent that individual donor countries exhibit distinct attitudes to giving foreign aid (Alesina and Dollar 2000), which reflect 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.

The last four concerns inform further modifications to the preliminary stage regression. Specifically, in place of Aid/GDP we use aid per capita (Aid/POP) 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 can also drop the colony-specific variables (and interactions) and employ only the aggregate RHS variables in RS08's specification. We also add donor-specific fixed effects. Thus, our preliminary stage regression emerges as follows:

$$
\begin{aligned}
\frac{Aid_{dr}}{POP_r} = \beta_0 &+ \beta_1 CURCOL_{dr} + \beta_2 COLONY_{dr} + \beta_3 COMLANG_{dr} \\
&+ \beta_4 \log(POP_d/POP_r) + \beta_5 COLONY_{dr} \times \log(POP_d/POP_r) \\
&+ \vartheta_d DONOR_d + \varepsilon_{dr}
\end{aligned} \tag{6}
$$

where the subscripts *d* and *r* represent donors and recipients respectively, CURCOL indicates whether the recipient is a current colony of the donor, COLONY indicates whether the recipient was a former colony of the donor, POP represents population, and DONOR is a donor specific dummy variable.

Results from this specification are given in columns III and IV of Table 2. Column III retains the original RHS specification but introduces aid per capita as the dependent variable (with missing aid values set to zero). All core coefficients retain the same sign and significance, and there is only a minor fall in explanatory power, indicating there may have been some unwanted independent correlation between GDP in the dependent variable and the RHS variables. Column IV employs the new RHS specification, as per equation (6). Again, there is a small loss of explanatory power, but the population ratio and its interaction with the colony dummy remain highly significant. Also, the donor fixed effects (coefficients not shown in the table) vary in sign and many are significant. Overall, the RHS variables continue to explain a reasonable share of observed aid allocations.

The existence of zero-value aid inflows points to a final possible weakness. 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 supply.[12] The distribution of bilateral aid flows reflects an unobserved selection process. In the absence of an explicit model, one way to address potential bias from unobserved selection effects is to use Heckman's correction (Heckman 1979). Column V of Table 2 employs a Heckman selection model (estimated by full information maximum likelihood) to the specification in column IV, where the existence of zero or non-zero aid flows is used as the binary selection variable.[13] Despite these changes, the direction of the results and their interpretation are largely unchanged. However, we reject the hypothesis that there is no selection bias. We therefore retain the Heckman estimator employed in Column V as our preferred preliminary stage regression.

We now verify the impact of these alternative preliminary stage regressions on estimates of the aid–growth relationship. As per RS08, fitted values from the preliminary stage regressions are aggregated over donors to give total predicted aid for each recipient in the period of interest (1970-2000); this is used as the aggregate aid instrument. Also, in place of the original Aid/GDP variable in the aggregate regression, we employ Aid/GDP estimates based on setting null values to zero. Table 3 reports results using different combinations of instruments, specifications and estimators. Column I employs the instrument generated from column II of Table 2 using the RS08 specification and IV-LIML estimator. Column II is similar, but now uses the instrument based on aid per capita from column III of Table 2. Column III of Table 3 switches to the preferred instrumentation strategy, retaining the RS08 specification; column IV employs the IV-IPWLS estimator. Finally, columns V and VI respectively apply the IV-LIML and IV-IPWLS estimators alongside the new specification and the preferred instrument. Column VI presents our preferred estimator, specification, and instrumentation strategy.

Instrument test statistics indicate these alternative instruments continue to perform strongly.[14] Under-identification tests, which can be interpreted as testing the null hypothesis of a zero correlation between the instruments and the endogenous regressors, are rejected. The weak identification test (reported in Table 3), 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 (see Table 1, column I). Perhaps more importantly, the Stock-Wright S statistic finds a significant (partial) correlation between the instrument and the outcome of interest. Excluding column I,

---

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

[13] 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.

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

which is the least coherent of the models, all other specifications reject the null hypothesis that the effect of aid on growth is zero. Rather, the point estimates consistently suggest a positive effect.

Table 3: Alternative instrumental and endogenous variables, 1970-2000

	(I)	(II)	(III)	(IV)	(V)	(VI)
	IV-LIML	IV-LIML	IV-LIML	IV-IPWLS	IV-LIML	IV-IPWLS
Aid/GDP	0.17	0.18*	0.22*	0.21*	0.25**	0.13***
	(0.11)	(0.11)	(0.12)	(0.13)	(0.12)	(0.05)
Initial per capita GDP	-1.38***	-1.38***	-1.34***	-1.92***	-1.03***	-1.33***
	(0.40)	(0.40)	(0.40)	(0.39)	(0.38)	(0.27)
Initial level of policy	2.16***	2.16***	2.14***	2.58***	2.12***	2.44***
	(0.58)	(0.58)	(0.60)	(0.62)	(0.54)	(0.46)
Initial life expectancy	0.08**	0.08**	0.09**	0.05	0.04	0.03
	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
Geography	0.60**	0.60***	0.63**	0.48**	0.29	0.25
	(0.23)	(0.23)	(0.25)	(0.24)	(0.26)	(0.21)
Coastal pop. density					0.00**	0.00***
					(0.00)	(0.00)
Primary schooling					2.58**	2.26**
					(1.15)	(0.88)
Malaria risk					-1.50*	-1.06*
					(0.85)	(0.58)
Price of invest. goods					-0.01	-0.01
					(0.00)	(0.00)
Civil liberties in 1972					-1.28*	-0.98*
					(0.70)	(0.50)
Air distance (log.)					0.09	-0.03
					(0.38)	(0.33)
Source of aid instrument	Tab.2 col II	Tab.2 col III	Tab.2 col V	Tab.2 col V	Tab.2 col V	Tab.2 col V
Scale of excluded instrument	Continuous	Continuous	Continuous	Binary	Continuous	Binary
Regional dummies	SSA, EA	SSA, EA	SSA, EA	SSA, EA	SSA, A, LA	SSA, A, LA
N	78	78	78	78	78	78
R-squared	0.60	0.60	0.57	0.70	0.59	0.77
Kleibergen-Paap Wald F stat.	21.87	25.97	29.48	24.42	17.28	39.78
Stock-Wright LM S stat.	3.22	3.49	4.33	3.53	5.77	6.49
(probability)	0.073	0.062	0.037	0.060	0.016	0.011

Notes: significance levels: *10%; **5%; ***1%. The endogenous variable is Aid/GDP, re-estimated from OECD-DAC (2008) data treating possible missing values as zeroes; excluded instrument is denoted in the row 'Source of aid instrument'; in columns (I) to (IV) the specification replicates Rajan and Subramanian (2008) (only selected covariates shown); columns (V) and (VI) use the modified specification as per Table 1, column (IV); intercept not shown; standard errors, given in parentheses, are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Source: Authors' estimates; Appendix C.

4.5 Sensitivity tests

The preferred estimates presented in Column VI of Table 3 represent a new estimator, a new specification, and a new instrumentation strategy. The previous sub-sections quantified the individual and combined impact of these new approaches as compared to the RS08 cross-section results. This process reveals robust empirical support for a positive aid–growth relationship for the 1970-2000 period. Further robustness and sensitivity tests are now considered.

First, as a robustness check on the IV-IPWLS results we employ a more flexible doubly robust estimator. This relaxes the assumption in equation (2) 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 (2) 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.[15] Table 4 summarizes the results from this estimator (denoted IV-FDR) for different combinations of instruments and specifications. These are compared to the estimates from standard (unweighted) IV-LIML estimators extracted from previous tables. Each cell of the table shows the lower and upper bounds of the 90 per cent confidence interval for the Aid/GDP coefficient (in parentheses), as well as the point estimate (in bold). The result that emerges from this approach is unchanged. For both estimators, the vast mass of this interval lies in the positive domain. In this context, the preferred estimator is the IV-FDR estimator with our specification and instrument (both denoted as AJT). For this case, the point estimates are highly consistent with the preferred results in Column 6 and Table 3. The 90 per cent confidence interval lies strictly in the positive domain, and the same goes for the IV-LIML estimator.

Table 4: Summary of Flexible Doubly Robust (FDR) results, 1970-2000

Instrument	Specification	Estimator	
		2SLS / IV-LIML	IV-FDR
RS08	RS08	(-0.02) **0.10** (0.21)	(0.02) **0.16** (0.31)
	AJT	(0.00) **0.10** (0.20)	(-0.01) **0.12** (0.25)
AJT	RS08	(0.02) **0.22** (0.41)	(0.03) **0.23** (0.43)
	AJT	(0.05) **0.25** (0.46)	(0.03) **0.17** (0.31)

Notes: AJT refers to our preferred instrument (Table 2 Column V) and specification (Table 1 Column V). Cells show 95 per cent confidence intervals for the coefficient on Aid/GDP from IV regressions involving different combinations of specifications, instruments and estimators; in each cell the lower and upper bounds of the interval are given in parentheses and the point estimate is in bold; column (I) provides estimates from standard IV estimators (2SLS or IV-LIML); column (II) employs the flexible doubly robust estimator as described in the text; estimates in column (I) are taken directly from results in Tables 1 and 3; standard errors used to calculate confidence intervals are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Source: Authors' estimates; Appendix C.

[15] Evidently, in estimating equation (2) across groups defined by W_i or Z_i these terms do not enter the RHS. In deriving standard errors for this estimator, we note that the estimated 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. 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.

Both the IV-IPWLS and the IV-FDR estimators rely on an estimated vector of weights (applied at the country level). It is therefore helpful to consider the extent to which our results are driven by specific countries or weights, and whether the results are robust to the exclusion of influential observations. To get a sense of the distribution of the weights, Figure A1 (Appendix A) gives a scatter plot of the estimated weights plotted against the residual from an OLS regression of the growth rate against core control variables (excluding Aid/GDP).[16] Panel (a) refers to the specification and instrument (from which the weights are derived) from column II of Table 1; panel (b) uses the specification and instrument from column VI of Table 3, which combines our preferred instrument, specification and estimator. Three points can be noted from the figure. First, in both panels, countries with higher weights typically lie towards the middle of the range of the *x*-axis, and thus do not refer to extreme (unexplained) growth rates. Second, there is a distinct shift in weights between the two models owing to the different sets of covariates used. Third, in panel (b) there are slightly fewer countries with very high weights—thus, the median of the weights declines from 1.25 in model (a) to 1.20 in model (b) while the interquartile range is stable. This gives some support to our use of the new specification.

The figure does not give a sense of the effect of these weights on the regression results. In order to identify the extent of dependence on individual observations, we re-estimate 12 different models (i.e., combinations of specifications, instruments and estimators) excluding one country (observation) at a time out of the total of 78 observations. Table 5 presents a summary. Cells of the table correspond to a single model and show (i) in the 'beta' rows, the minimum, mean and maximum values of the point estimates on the (endogenous) aid variable, and (ii) in the 'prob.' rows, the minimum, mean and maximum probabilities that the estimate is not different from zero (p-value). Four main results merit mention. First, the point estimate of the effect of aid on growth is positive in all of the 12×78 = 936 regressions encompassed by the table. Second, in none of the regressions does the same point estimate fall outside the 90 per cent confidence interval established in the corresponding full sample estimate (not shown). Third, there *are* important country observations. In only one of the 12 models considered does the impact of aid remain significantly different from zero at the 10 per cent threshold level when observations are dropped sequentially. Thus, in the remaining 11 models, dropping important observations can lead to a p-value greater than the 10 per cent level. Nevertheless, the existence of important observations cuts both ways. Dropping an important observation can also lead to greater significance (i.e., lower probability that the parameter is not different from zero). After dropping the observation that contributes most importantly to a lack of significance, the aid variable is at least significant at the 10 per cent level in all models and is significant at the 5 per cent level in 10 of the 12 (both of which use the original RS08 instrument). Note also that the average probabilities are also lower than 10 per cent in all models employing the modified instrument (denoted AJT).

Three further checks on influential observations are investigated. First, we re-estimate our preferred model with the IV-IPWLS estimator, but exclude observations falling in the top 10 per cent of the estimated weight distribution. Results from this specification are given in column I of Table 6. This confirms the preceding exercise—the point estimate of interest is broadly unchanged; however, it is no longer significant due to a much larger standard error, which reflects a weakening of the (binary) instrument. Second, we re-estimate the same model but now only include the bottom and top 30 countries in the fitted aid per capita distribution. This provides a sensitivity check to the cut-point used to derive the binary

[16] Appendix B provides a detailed list of in-sample countries, variables and estimated weights.

instrument. Third, we exclude some of the largest and most dynamic economies in the sample (India, Brazil and China), as well as Israel and Egypt, which are often taken to be special foreign aid cases due to their links with the USA. Results from these restricted samples are shown in columns II and III of Table 6. They reinforce the conclusion that although the IV-IPWLS results are broadly robust, there is some dependence on inclusion of the most informative observations. Nevertheless, the point estimates appear to be stable, suggesting that it is the confidence intervals that are most sensitive to the specific sample chosen.

Table 5: Summary of sequential exclusion procedure, 1970-2000

	Specification ->	RS08		AJT	
	Instrument ->	RS08	AJT	RS08	AJT
IV-LIML	beta	.06 [.10] .13	.15 [.22] .27	.06 [.10] .14	.17 [.25] .32
	prob.	.07 [.14] .25	.05 [.07] .13	.08 [.11] .18	.04 [.05] .08
IV-IPWLS	beta	.07 [.15] .20	.11 [.23] .32	.08 [.10] .12	.09 [.12] .15
	prob.	.01 [.05] .11	.04 [.09] .25	.01 [.02] .16	.00 [.02] .22
IV-FDR	beta	.11 [.16] .20	.12 [.23] .31	.08 [.13] .21	.13 [.17] .24
	prob.	.01 [.05] .10	.04 [.06] .23	.03 [.11] .18	.01 [.07] .21

Notes: AJT refers to our preferred instrument (Table 2 Column V) and specification (Table 1 Column V); 'beta' rows show the minimum [mean] maximum of the point estimates on the endogenous variable of interest (aid) from a vector of estimates based on a single model in which one observation (country) is excluded and the estimation repeated; 'prob.' rows show the minimum [mean] maximum of the vector of probabilities that the beta point estimates are equal to zero; standard errors used to calculate probabilities are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Source: Authors' estimates; Appendix C.

We also explore the sensitivity of the unweighted regression. Specifically, it is useful to consider further alterations to the underlying data and/or specification, where the choice of modifications reflects potentially important sources of fragility. First, we note that in the preliminary stage regression used to generate the instrument, a number of very small states are included which do not appear in the aggregate aid–growth sample. These are potentially influential observations with respect to donor-beneficiary population differences. Thus, in Column IV of Table 6 we exclude all countries with populations under 500,000 persons from the preliminary stage. Second, we note from the previous analysis of the IV weights that under both the RS08 and AJT specifications, some of the largest weights are attributed to large natural resource exporters (e.g., Venezuela, Nigeria). Thus, in column V we add to the specification a dummy variable taking the value of one if the country was an oil exporter in 1960.[17] Lastly, in column VI we replace the endogenous Aid/GDP variable with the same aid per capita variable used in the (modified) preliminary stage regressions (column V, Table 3),

[17] We have also run the full set of different regressions including hydrocarbon deposits in 1993 as an additional covariate. All our main results hold if this is the case. However, due to concerns regarding endogeneity, this variable was not included in previous sections.

Table 6: Robustness and sensitivity tests, 1970-2000

	(I) IV-IPWLS	(II) IV-IPWLS	(III) IV-IPWLS	(IV) IV-LIML	(V) IV-LIML	(VI) IV-LIML
Aid measure	0.15	0.15***	0.12**	0.24**	0.42**	13.93***
	(0.24)	(0.04)	(0.05)	(0.12)	(0.19)	(4.26)
Initial per capita GDP	-1.14	-0.83***	-1.26***	-1.05***	-0.84**	-1.74***
	(0.74)	(0.29)	(0.29)	(0.37)	(0.43)	(0.30)
Initial level of policy	2.00***	2.49***	2.71***	2.12***	1.94***	2.28***
	(0.49)	(0.48)	(0.50)	(0.53)	(0.64)	(0.45)
Initial life expectancy	0.03	0.01	0.03	0.03	0.08	0.00
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
Geography	0.17	0.12	0.15	0.29	0.24	0.46*
	(0.24)	(0.22)	(0.25)	(0.25)	(0.29)	(0.25)
Coastal pop. density	0.00**	0.00***	0.00***	0.00**	0.00**	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Primary schooling	2.14*	1.66	2.22**	2.56**	2.35**	2.28***
	(1.22)	(1.09)	(0.94)	(1.12)	(1.20)	(0.84)
Malaria risk	-1.28	-1.27**	-1.14*	-1.49*	-1.64*	-1.66**
	(0.79)	(0.59)	(0.65)	(0.84)	(0.97)	(0.78)
Price of invest. goods	-0.01*	-0.01**	-0.01**	-0.01	-0.00	-0.01*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Civil liberties in 1972	-1.06	-0.71	-0.75	-1.24*	-1.30*	-1.43**
	(0.95)	(0.49)	(0.48)	(0.68)	(0.75)	(0.64)
Air distance (log.)	-0.02	0.10	-0.16	0.09	-0.21	-0.01
	(0.41)	(0.36)	(0.38)	(0.38)	(0.42)	(0.34)
Oil producer in 1960 (1=yes)	-	-	-	-	1.41**	-
					(0.59)	
Scale of excluded instrument	Binary	Binary	Binary	Continuous	Continuous	Continuous
N	71	60	73	78	78	78
R-squared	0.64	0.84	0.75	0.60	0.45	0.64
Kleibergen-Paap Wald F stat.	5.01	63.69	31.64	14.53	11.72	21.65
Stock-Wright LM S stat.	0.56	10.18	4.13	5.64	7.92	5.77
(probability)	0.453	0.001	0.042	0.018	0.005	0.016

Notes: Significance levels: *10%; **5%; ***1%. Unless otherwise indicated, the endogenous variable is Aid/GDP, measured as per the models in Table 3; column (I) excludes the top 10 per cent of observations according to their estimated weight in the regression; column (II) excludes the middle 18 countries in the fitted Aid/GDP distribution; column (III) excludes 5 possible ‘special cases’; in column (IV) the excluded instrument is modified by dropping small states from the preliminary stage regression (not shown); in column (V) a 1960 oil production dummy is added as an additional included instrument; in column (VI) the endogenous variable is measured as aid per capita; intercept and regional dummies not shown; standard errors, given in parentheses, are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Source: Authors’ estimates; Appendix C.

thereby excluding any possible influence arising from changes in the denominator. As can be seen from the table, none of these modifications changes our core results. In all models the treatment effect is positive, is in the expected range, and remains statistically significant.

Finally, we return to the other periods analyzed in RS08 (1960-2000, 1980-2000 and 1990-2000). Using our preferred specification and instrumentation strategy as well as both the IV-LIML and IV-IPWLS estimators, results for each alternative period are presented in Appendix Table A1.[18] 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 3 for the 1970-2000 period. 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.

4.6 Instrument validity

Before concluding, it is pertinent to address a critique of the instrumentation strategy we pursued in this paper, which is grounded in the approach of RS08.[19] In a recent working paper Clemens and Bazzi (2009) question a large number of the instrumental variables that have been employed in the applied growth regression literature. They note that different sets of authors have used the same variables as exogenous instruments for numerous different endogenous variables. This raises the possibility that these exogenous instruments are correlated with other omitted variables, thereby invalidating the exclusion restriction necessary for valid causal inference. Among other papers they direct attention to the reliance of the RS08 (fitted) instrument on the natural logarithm of aid recipient population size. They find that log population has a 'statistically significant partial relationship with several variables that [both] are plausible growth determinants' (2009: 11) and are omitted from RS08's specification.

These concerns merit attention. We also note that while the existence of a partial correlation between omitted explanatory variables and the chosen instrument indicates that the coefficients in a regression specification *may be* biased, the extent to which this is material ultimately is an empirical matter. This is recognized by Clemens and Bazzi (2009), leading them to advocate application of a range of empirical tests for instrument validity. On the face of it, however, straightforward validity checks based on Sargan or Hansen tests are not possible in the RS08 case (or our modification) as the regression model is 'just identified'—i.e., the number of excluded instruments equals the number of endogenous variables. Nonetheless, recalling that the preliminary stage of the RS08 approach generates a single instrument as a linear combination of 'exogenous' variables, it is possible to use modified versions of these same RHS variables as excluded instruments directly in the aggregate aid–growth regressions. This provides for a large number of potential instruments, and therefore opens the way to undertake Hansen tests either on the full set of instruments, or specific

---

[18] 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.

[19] We are grateful to an anonymous referee for raising this issue and making constructive empirical suggestions.

subsets of them. Intuitively, these tests essentially amount to regressing the residuals from the aid–growth regression on the excluded instrument set. If a significant relationship is encountered, then this is taken as evidence against instrument validity as identification turns on the assumption of a mean zero error term, conditional on the (excluded and included) instrument set. Additionally, difference-in-Hansen tests can be calculated, which can be understood as comparing the aggregate Hansen statistics for a full versus restricted set of the excluded instruments. If specific sets of variables are found to contribute substantially to the difference statistic, then they may be particularly suspect—for further details see Baum et al. (2007); also Roodman and Morduch (2009) for an application.[20] Following this logic, the first step is to transform the explanatory variables used in the bilateral preliminary stage regressions such that they can be employed at a more aggregate level. As the unit of observation in the aggregate regressions is aid recipients, the simplest approach is to collapse the bilateral dataset along the donor dimension (only). Thus, for continuous preliminary stage regressors such as the donor-recipient population ratio, the corresponding 'aggregate' instrument is the mean of the population ratio for each recipient across all donors. For dummy regressors, such as specific colonial relationships, it is more appropriate to take the maximum value of the dummy for a given recipient (again, across all donors). Ignoring relatively unimportant variables such as currently being a colony and the population-colony interaction terms employed in RS08, this yields a set of eight possible instruments of which a more restricted subset were used in the modified specification presented in Section 4.4. These are presented in the rows of Table 7.

Table 7: Instrument validity checks, 1970-2000

	Fitted coefficients		Residual coefficients		RS08 model		AJT model	
	RS08	AJT	RS08	AJT	C stat.	Prob.	C stat.	Prob.
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
Population ratio	0.85***	0.59***	-0.18	-0.03	0.40	0.53	0.24	0.63
Colony (ever)	0.10**	0.01	-0.19	-0.25	1.11	0.29	1.96	0.16
Pop. ratio × colony	0.11	0.32***	0.10	0.09	0.48	0.49	0.06	0.81
Common language	0.04	0.21***	0.25*	0.26*	2.36	0.12	0.27	0.60
Spanish colony	-0.08	0.03	0.26	0.18	0.29	0.59	1.36	0.24
Portuguese colony	0.03	0.04*	0.27**	0.08	3.28	0.07	1.77	0.18
French colony	-0.01	0.11***	0.53**	0.71***	2.14	0.14	8.13	0.00
UK colony	-0.25***	0.11**	0.35	0.42**	0.02	0.89	1.02	0.31
R-squared	0.95	0.99	0.12	0.23	-	-	-	-

Notes: Significance levels: *10%; **5%; ***1%. Columns (I) and (II) report standardized OLS regression coefficients in which the dependent variable is the fitted aid instrument taken from Table 2 columns (I) and (V) respectively; columns (III) and (V) report standardized OLS regression coefficients from regressions of residuals saved from columns (III) and (IV) of Table 8 against the row variables; columns (V) to (VIII) report individual difference-in-Hansen C statistics and corresponding probabilities associated with each individual row instrument (relative to the full instrument set); all OLS regression specifications incorporate the relevant set of included instruments as additional controls (not reported); significance from OLS regressions are based on robust standard errors.

Source: Authors' estimates; Appendix C.

[20] These procedures are implemented using the ivreg2 command in STATA. As the LIML-IV estimator is used here, Hansen J statistics and associated tests are reported, corresponding to the overall validity of the full instrument set; C (difference-in-Hansen) statistics also are reported where appropriate, based on the orthog- option of the command.

To verify that these aggregate instruments are good proxies for the fitted aid variables generated in the preliminary stage regressions, the first two columns of Table 7 report standardized coefficients from simple OLS regressions of the fitted aid instruments against the full set of these aggregate instruments. Column I employs RS08's long-run instrument as the dependent variable (as per Table 2, column I); column II employs our preferred modification (as per Table 2, column V). As expected, the explanatory power of each regression is extremely high. Moreover, underlining the contention of Clemens and Bazzi (2009), a driving force being the fitted aid instruments appears to be the population ratio term. This is confirmed by re-running the aid–growth regression models, adding the aggregate population ratio instrument as an included instrument (a standard RHS covariate). These results are given in columns I and II of Table 8, again referring to RS08's and our own preferred models. As can be seen from the Kleibergen-Paap Wald F statistic, used to evaluate instrument relevance, the strength of the fitted aid instrument is now extremely low, meaning that the model may be effectively unidentified (reflected by the negative R-squared statistics). Thus, a fundamental issue for the RS08 instrumentation strategy is the validity of the exclusion restriction as it applies to the population-based instruments. Nevertheless, the results of Table 7 indicate that other variables make some (albeit smaller) contribution to the overall fitted instrument.

To test instrument validity, we commence by estimating both the RS08 and our own aid–growth models employing the full set of eight aggregate instruments. These results are reported in columns III and IV of Table 8. Whilst they broadly confirm previous findings based on the generated single instrument, we note that overall instrument strength is moderate at best. Next, and following the intuition of Sargan-type tests, we save the residuals and regress these against the same set of excluded instruments. Although inference on these latter OLS results is unlikely to be accurate, it provides initial insight as to which variables in the instrument set may be suspect. Standardized coefficients from these regressions are given in columns III and IV of Table 7. They show that neither the population ratio term nor its interaction with the (ever being a) colony dummy is significantly correlated with the unexplained components of the growth models. In contrast, excluding Spain, all of the colony-specific terms are significant in at least one specification. The common language variable is also significant at the 10 per cent level in both cases.

Following this simple but intuitive OLS approach, columns V to VIII of Table 7 report formal tests of the orthogonality of each of the individual aggregate instruments to the growth regression errors. Specifically, we report the difference-in-Hansen C statistic associated with excluding each row instrument (individually) from the full set. For example, in the first row the C statistic corresponds to the reduction in the overall Hansen J test statistic when the population ratio term is excluded from the instrument set; the corresponding probability is also shown. These findings corroborate the residual-based OLS results. We find that the French colony and Portuguese colony instruments fail difference-in-Hansen tests and therefore should not be used as exogenous regressors. This supports our decision, discussed in Section 4.4, to drop colony-specific variables from the preliminary stage regressions. In both the RS08 and our own specifications, however, the population ratio variables do not give cause for concern (yielding 0.53 and 0.63 test probability levels respectively). This provides comfort as to their suitability as exogenous instruments in these models.

Table 8: Aid–growth regressions using aggregate instruments, 1970-2000

	(I) RS08	(II) AJT	(III) RS08	(IV) AJT	(V) RS08	(VI) AJT
Aid/GDP	0.75	-1.43	0.10	0.12	0.07	0.21*
	(2.60)	(1.22)	(0.08)	(0.10)	(0.06)	(0.11)
Initial per capita GDP	0.57	-4.29**	-1.40***	-1.44***	-1.44***	-1.34***
	(8.24)	(2.08)	(0.39)	(0.30)	(0.37)	(0.33)
Initial level of policy	2.45*	2.38*	2.13***	2.28***	2.16***	2.29***
	(1.34)	(1.41)	(0.56)	(0.46)	(0.52)	(0.52)
Initial life expectancy	0.31	-0.17	0.08**	0.04	0.07**	0.05
	(0.93)	(0.17)	(0.04)	(0.04)	(0.03)	(0.04)
Geography	0.60	0.96	0.62**	0.25	0.58***	0.29
	(0.55)	(0.65)	(0.24)	(0.22)	(0.22)	(0.24)
Population ratio (mean)	-1.28	1.98	-	-	-	-
	(5.05)	(1.32)				
Regional dummies	SSA, EA	SSA, A, LA	SSA, EA	SSA, A, LA	SSA, EA	SSA, A, LA
N	78	78	78	78	78	78
R-sq.	-1.14	-1.12	0.58	0.69	0.62	0.65
No. of excluded instruments	1	1	8	8	3	3
Kleibergen-Paap Wald F stat.	0.11	0.96	5.50	4.32	10.40	5.06
Stock-Wright LM S stat.	0.89	7.32	11.85	18.17	3.05	5.40
(probability)	0.347	0.007	0.158	0.020	0.384	0.145
Hansen J stat.	-	-	7.72	11.88	0.29	0.31
(probability)	-	-	0.358	0.104	0.865	0.857
Difference-in-Hansen C stat.	-	-	0.40	0.23	0.01	0.00
(probability)	-	-	0.529	0.628	0.936	0.980

Notes: Significance level: *10%; **5%; ***1%. Columns (I) and (II) replicate the models reported in Table 1 column (I) and Table 3 column (V) respectively, but adding the aggregate population ratio variable used in the preliminary stage regressions; columns (III) and (IV) also replicate these models but employ the full set of aggregate instruments (Table 7) in place of the generated aid instrument; columns (V) and (VI) replicate columns (III) and (IV) using as excluded instruments only the mean population ratio, colony dummy and their interaction; difference-in Hansen tests refer to the mean population ratio only; all regressions use the (unweighted) IV-LIML estimator; intercept, regional dummies and additional covariates not shown; standard errors, given in parentheses, are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Source: Authors' estimates; Appendix C.

As a last exercise we rerun the aid–growth model using a smaller and 'less suspect' subset of the aggregate instruments, namely the population ratio, the colony dummy and their interaction. These results are given in columns V and VI of Table 8. The overall Hansen J test and the difference-in-Hansen test associated with the population ratio term (only) are also reported. The results are encouraging. Not only are these tests passed with a very high level of confidence, but the regression coefficient estimates also remain exactly in the domain of previous results. For example, the coefficient on Aid/GDP using our preferred specification and endogenous aid variable is 0.21, and is significant at the 10 per cent level. While none of this evidence can be deemed conclusive as regards the validity of the instrumentation strategy pursued herein, it provides considerable support to the overall robustness of our results.

Finally, it is worth keeping in mind that this exercise provides a clear indication of the trade-off between efficiency and transparency in instrument selection. Using a preliminary stage to generate a single instrument is likely to be more efficient, especially in small samples such as those faced by (static) cross-country regressions. Nevertheless, it limits the extent to which one can easily test whether some or all of the factors used to generate the instrument are valid. The approach developed in this subsection thus provides a very simple robustness check that can be applied in such cases.

5 Conclusion

To conclude, we return to the question posed in the title to this paper: has the aid and growth literature gone full circle? Our response is 'no'. While in the most recent literature the pendulum has swung to deep skepticism concerning the ability of aid to contribute to economic growth, a series of important points of consensus have emerged. First, methodological advances in the programme evaluation literature have improved the profession's capacity to identify causal effects in economic phenomena. These advances are beginning to be applied at the more aggregate level, as pursued here. Second, methodological advances also highlight the serious challenges that must be surmounted in order to derive robust causal conclusions from non-experimental 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 growth theory. Finally, there is increasing recognition that many of the key interventions pursued by foreign aid will only result in positive growth outcomes over long time horizons.

In line with Temple (2010), we started by replicating RS08. Subsequently, we developed a preferred specification, instrumentation strategy, and estimator. The preferred specification contains a fuller set of regional fixed effects and indicators of initial human capital and geographic conditions. These were drawn from theory and previous research. They 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. With respect to the preliminary stage instrumentation, 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; (iii) introduced donor-specific fixed effects; and (iv) accounted for selection bias through a Heckman correction. Finally, we deployed robust regression estimators which adjust for heterogeneity across countries. This involved introducing a new doubly robust estimator that can be used in instrumental variable contexts. A variety of robustness and validity checks, including of the underlying instrument, provide support to our approach.

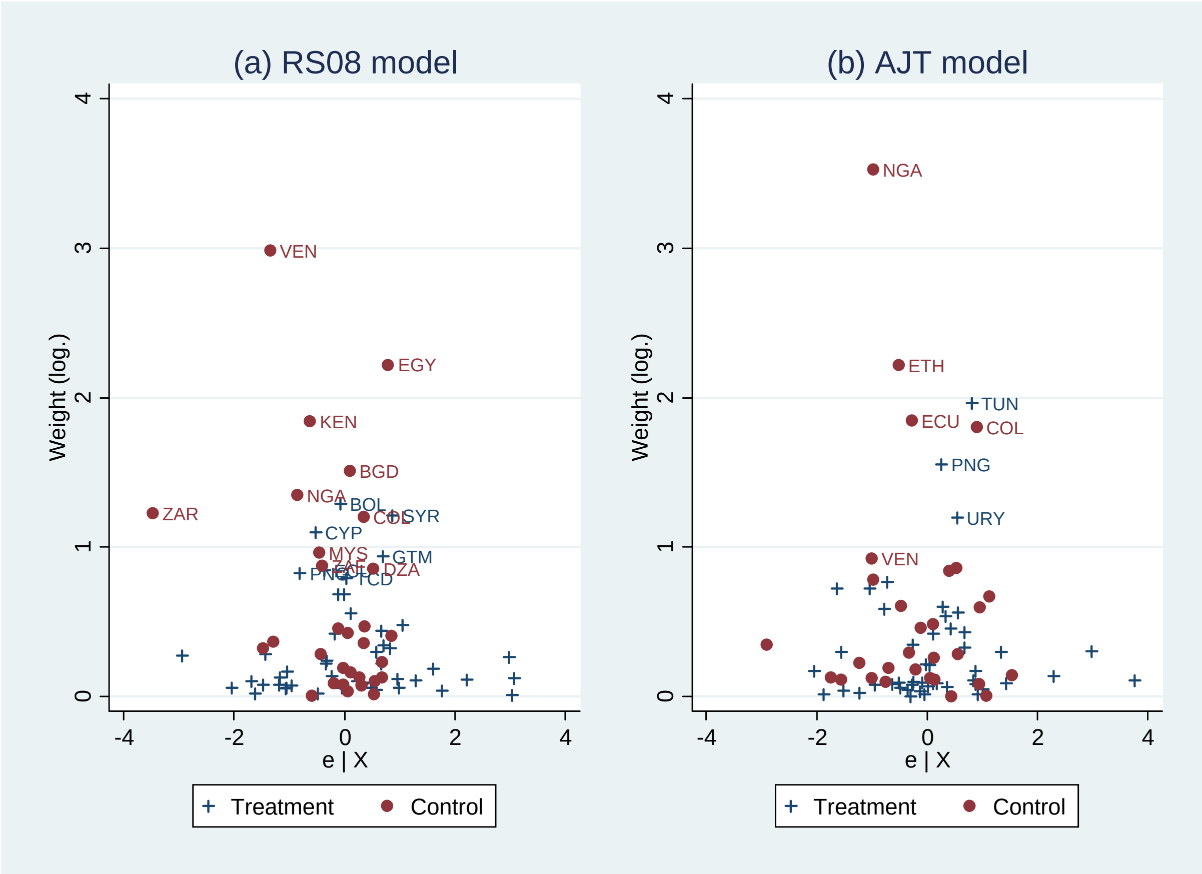
Overall, we believe our approach represents the most carefully developed empirical strategy employed in the aid–growth literature to date. Our results provide solid support for the view that the average treatment effect of aid on growth is positive in both the 1970-2000 and 1960-2000 periods. Our preferred doubly robust estimators place the point estimate of the long run elasticity of growth with respect to the share of aid in recipient GDP at 0.13 (IV-IPWLS) and 0.17 (IV-FDR), which is below an unweighted point estimate of 0.25 (IV-LIML). These suggest that an inflow on the order of ten percent of GDP spurs per capita growth rate by more than one percentage point per annum in the long run. These estimates are consistent with the view that foreign aid stimulates aggregate investment and may also contribute to productivity growth, despite some fraction of aid being allocated to consumption. The 95 per cent confidence interval around these estimates lies in the strictly positive domain and contains the prior, suggested in RS08 that the long-run elasticity of growth to foreign aid should be around 0.1. In the shorter term, our analysis indicates that the impact of aid is difficult to discern. Nevertheless, combining the longer run macro evidence with the evidence at the micro- and meso-levels, a consistent case for aid effectiveness emerges. There is no micro–macro paradox.

On the whole, 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 favors 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.

Finally, 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. 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.

Appendix A: Additional figures and tables

Figure A1: Scatter plot of IV-IPWLS regression weights versus growth residuals



Notes: panel (a) refers to the model in Table 1, column II; panel (b) refers to the model in Table 3, column VI; *y*-axis plots the log. of (inverse propensity score) estimated weights from these models (transformed by natural logarithms to aid clarity); *x*-axis plots the residual from an OLS regression of the growth rate against core control variables (excluding Aid/GDP).

Source: Authors' estimates; Appendix C.

Table A1: Modified regressions, alternative periods

	1960-2000		1980-2000		1990-2000	
	(I)	(II)	(III)	(IV)	(V)	(VI)
	IV-LIML	IPWLS	IV-LIML	IPWLS	IV-LIML	IPWLS
Aid / GDP	0.16*	0.09**	0.02	0.05	-0.11	0.11
	(0.08)	(0.04)	(0.14)	(0.10)	(0.19)	(0.14)
Initial per cap. GDP	-0.67**	-0.83***	-1.41***	-1.36***	-0.72	-0.12
	(0.31)	(0.24)	(0.42)	(0.35)	(0.74)	(0.57)
Initial level of policy	1.88***	2.33***	2.10***	1.51	0.65	0.99*
	(0.45)	(0.42)	(0.71)	(1.01)	(0.53)	(0.52)
Initial life expectancy	0.01	0.02	0.06	0.10**	0.12	0.13**
	(0.02)	(0.03)	(0.04)	(0.04)	(0.07)	(0.06)
Geography	0.26	0.21	0.48**	0.33	0.13	0.23
	(0.19)	(0.17)	(0.21)	(0.23)	(0.41)	(0.38)
Coastal pop. density	0.00***	0.00***	0.00*	0.00	0.00	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Primary schooling	2.76***	2.13***	1.45	1.15	-1.38	-2.08
	(0.91)	(0.74)	(1.00)	(1.04)	(1.93)	(1.68)
Malaria risk	-1.20**	-1.03**	-0.99	-1.09	-2.36**	-2.48**
	(0.57)	(0.40)	(0.79)	(0.77)	(0.96)	(1.02)
Price of invest. goods	-0.01**	-0.01**	0.00	-0.00	-0.01	-0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Civil liberties in 1972	-0.33	-0.40	-0.34	-0.19	0.67	0.74
	(0.26)	(0.27)	(0.31)	(0.35)	(0.64)	(0.62)
Air distance (log.)	0.31	0.28	1.19*	1.27	2.11**	1.78**
	(0.42)	(0.39)	(0.69)	(0.82)	(0.90)	(0.75)
Scale of excluded instrument	Binary	Continuous	Binary	Continuous	Binary	Continuous
N	74	74	75	75	70	69
R-squared	0.71	0.84	0.64	0.60	0.52	0.51
Kleibergen-Paap Wald F stat.	15.82	42.80	19.31	22.01	17.58	20.04
Stock-Wright LM S stat.	4.30	4.30	0.03	0.20	0.30	0.63
(probability)	0.038	0.038	0.869	0.654	0.584	0.428

Notes: Significance level: *10%; **5%; ***1%. The endogenous variable is Aid/GDP, measured as per the models in Table 3; intercept and regional dummies not shown; standard errors, given in parentheses, are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Source: Authors' estimates; Appendix C.

Appendix B: Summary of countries, variables, and estimated weights, 1970-2000

Country	Growth rate	Endogenous Aid/GDP		Fitted aid measure		Estimated weight	
		RS08	AJT	RS08	AJT	RS08	AJT
Algeria	1.18	0.54	0.53	2.24	59.11	2.35	2.36
Argentina	0.57	0.08	0.05	-0.02	54.53	1.17	1.30
Bangladesh	1.41	4.92	2.91	-1.17	48.92	4.52	1.12
Benin	0.35	9.76	5.46	8.47	76.39	1.06	1.07
Bolivia	0.29	7.17	4.08	4.91	67.92	3.62	2.06
Botswana	6.36	3.83	6.04	9.44	91.12	1.13	1.11
Brazil	2.29	0.09	0.06	-7.66	45.39	1.50	2.32
Burkina Faso	1.19	12.73	7.87	5.79	69.83	1.24	1.09
Burundi	-1.61	16.36	7.47	7.26	68.49	1.11	1.04
Cameroon	0.85	3.88	2.70	5.65	72.91	1.10	1.08
Chad	-0.87	14.29	6.33	7.37	73.68	2.20	1.39
Chile	2.43	0.29	0.18	2.60	61.26	1.04	1.20
China	5.09	0.41	0.18	-6.18	38.23	1.08	1.01
Colombia	1.78	0.31	0.34	0.15	54.93	3.32	6.06
Congo, Dem. Rep.	-4.90	3.73	2.33	2.68	55.97	3.41	1.42
Congo, Rep.	2.22	7.04	4.87	11.24	83.40	1.12	1.09
Costa Rica	1.13	1.92	1.55	7.43	75.73	1.08	1.41
Cote d'Ivoire	-0.82	4.34	2.81	5.87	70.02	1.08	1.06
Cyprus	4.22	0.76	0.74	9.01	84.99	2.99	1.23
Dominican Rep.	3.20	1.26	0.89	4.31	63.14	1.11	1.95
Ecuador	1.38	1.21	0.81	3.92	64.99	2.33	6.33
Egypt, Arab Rep.	2.51	3.25	3.68	0.79	59.48	9.21	1.16
El Salvador	0.23	4.26	3.50	5.36	69.27	1.35	1.79
Ethiopia	0.15	8.71	3.99	0.49	52.86	1.53	9.19
Fiji	1.59	3.04	2.63	9.90	93.10	1.98	1.82
Gabon	0.68	1.90	1.94	14.59	92.09	1.01	1.15
Gambia, The	0.30	22.66	10.15	10.16	94.22	1.06	1.11
Ghana	0.17	3.99	3.45	3.59	67.40	1.14	1.08
Guatemala	0.90	1.52	1.09	4.29	66.07	2.55	1.71
Guinea Bissau	2.43	27.63	21.83	20.38	84.51	1.41	1.07
Guyana	1.37	9.43	5.03	9.20	90.09	1.27	1.53
Haiti	3.31	9.48	5.26	6.60	71.78	1.31	1.35
Honduras	0.32	7.28	4.09	6.29	72.14	1.18	1.24
Hungary	2.21	0.53	0.00	2.54	58.00	1.38	1.25
India	2.79	0.76	0.46	-5.74	42.71	1.11	1.33
Indonesia	4.03	1.18	1.50	-2.01	44.21	1.57	1.10
Iran, Islamic Rep.	0.46	0.12	0.06	0.52	52.95	1.21	1.62
Israel	2.17	2.41	3.02	5.98	76.61	1.52	1.57
Jamaica	-0.15	2.77	2.55	7.03	80.87	1.06	1.35
Kenya	1.39	7.22	4.55	2.94	65.07	6.31	1.01
Korea, Rep.	5.89	0.09	0.50	0.81	57.19	1.13	1.00

Country	Growth rate	Endogenous Aid/GDP		Fitted aid measure		Estimated weight	
		RS08	AJT	RS08	AJT	RS08	AJT
Lesotho	1.96	15.67	10.95	8.23	82.96	1.99	2.06
Madagascar	-1.40	8.95	4.87	5.07	68.11	1.02	1.09
Malawi	1.81	19.54	9.08	5.04	72.88	1.08	1.03
Malaysia	4.12	0.42	0.44	2.55	60.06	2.62	1.12
Mali	0.71	15.35	9.62	5.99	70.31	1.11	1.09
Mauritania	-1.24	17.58	7.92	11.09	77.24	1.02	1.10
Mauritius	4.16	2.23	1.70	8.86	88.61	1.04	1.01
Mexico	1.54	0.08	0.05	-2.15	49.85	1.43	1.81
Morocco	1.66	1.85	1.53	2.15	61.35	1.44	1.21
Namibia	-0.23	5.41	2.01	8.42	80.41	1.20	1.19
Nicaragua	-2.71	16.79	8.83	5.77	67.20	1.31	1.19
Niger	-1.84	13.43	7.98	6.90	72.53	1.08	1.08
Nigeria	-1.51	0.42	0.29	-0.50	54.36	3.85	33.99
Pakistan	2.52	2.34	1.70	-0.79	53.59	1.09	1.09
Panama	1.54	1.05	0.85	7.83	77.03	1.12	1.54
Papua New Guinea	0.07	10.28	11.44	7.97	72.25	2.29	4.73
Paraguay	1.63	1.38	1.17	6.57	73.01	1.24	2.16
Peru	-0.07	1.00	0.90	1.67	58.75	1.60	2.19
Philippines	1.19	1.52	1.34	3.25	60.86	1.01	1.13
Romania	2.45	0.47	0.00	1.20	54.61	1.00	1.14
Rwanda	0.03	15.49	11.00	7.55	72.62	1.06	1.09
Senegal	-0.01	11.16	6.31	6.90	72.52	1.05	1.05
Sierra Leone	-1.87	9.75	5.35	6.24	77.63	1.15	1.10
Singapore	5.97	0.08	0.21	6.79	79.90	1.62	1.00
South Africa	0.31	0.43	0.06	1.48	60.16	2.40	1.83
Sri Lanka	2.50	5.29	3.94	2.27	59.13	1.26	1.59
Syrian Arab Rep.	3.04	0.95	0.66	5.11	65.58	3.36	1.76
Thailand	4.42	0.71	0.69	0.58	56.62	1.33	1.13
Togo	-1.58	10.56	5.96	9.54	79.08	1.06	1.03
Trinidad & Tobago	1.76	0.26	0.08	8.50	87.06	1.38	1.35
Tunisia	3.23	2.00	2.20	5.84	67.29	1.55	7.11
Turkey	2.12	0.34	0.31	1.02	45.36	1.07	1.34
Uganda	1.46	6.58	3.79	3.30	66.34	1.14	1.10
Uruguay	1.50	0.40	0.24	5.22	65.64	1.74	3.31
Venezuela, RB	-1.65	0.07	0.03	2.26	60.32	19.74	2.52
Zambia	-1.35	13.19	8.57	5.21	73.54	1.33	1.02
Zimbabwe	0.48	4.71	2.38	4.42	79.47	1.06	1.05

Notes: Countries listed in alphabetical order; all RS08 columns refer to variables and estimated weights used in Table 1 column (II); all AJT columns refer to variables and estimated weights from Table 3 column (VI) for AJT; note that for RS08 the fitted aid measure is (100×Aid)/GDP, but for AJT it is aid per capita; growth rate is the dependent variable used throughout.

Source: Authors' estimates; Appendix C.

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

Additional covariates taken from Sala-i-Martin et al. (2004) [21]	Description
Coastal population density	Coastal (within 100 km of coastline) in 1965
Primary schooling	Enrolment rate in primary education in 1960
Price of investment goods	Average investment price level between 1960 and 1964 on purchasing power parity basis
Malaria risk	Index of malaria prevalence in 1966
Civil liberties in 1972	Index of civil liberties in 1972
Air distance (log.)	Logarithm of minimal distance (in km) from New York, Rotterdam, or Tokyo

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 database online (2009) 22
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$	World Development Indicators (WDI) CD ROM; World Bank (2008)

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

22 Accessible at: http://www.oecd.org/dac/stats

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

Angrist, J.D., and J.-S. Pischke (2010). 'The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con Out of Econometrics', *NBER Working Papers,* No. 15794.

Ashraf, Q.H., A. Lester, and D.N. Weil (2008). 'When Does Improving Health Raise GDP?', *NBER Working Papers,* No. 14449.

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, UNDP: 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 (2009). 'Alternative Approaches to Evaluation in Empirical Microeconomics', *Journal of Human Resources,* 44(3*).*

Bun, M., and F. Windmeijer (2010). 'The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models', *Econometrics Journal*, 13(1): 95-126.

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

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

Clemens, M., and S. Bazzi (2009). 'Blunt Instruments: On Establishing the Causes of Economic Growth', *Center for Global Development Working Papers,* No.171.

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

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.

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

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. (2003). ‘Can Foreign Aid Buy Growth?’, *Journal of Economic Perspectives,* 17(3): 23-48.

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

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

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

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.

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

Imbens, G.W. (2004). ‘Non-parametric Estimation of Average Treatment Effects Under Exogeneity: A Review’, *The Review of Economics and Statistics,* 86(1): 4-29.

Imbens, G.W., and J.M. Wooldridge (2009). ‘Recent Developments in the Econometrics of Programme 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), North Holland Elsevier: Amsterdam.

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., 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 Papers,* No. 05/100.

Mishra, P., and D.L. Newhouse (2007). ‘Health Aid and Infant Mortality’, *International Monetary Fund Working Papers,* No. 07/100.

Moral-Benito, E. (2009). ‘Determinants of Economic Growth: A Bayesian Panel Data Approach’, *World Bank Policy Research Working Papers,* No. 4830

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

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

Nunn, N. (2008). ‘The Long-Term Effects of Africa’s 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 University Press: Oxford.

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

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

Roodman, D., and J. Morduch (2009). ‘The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence’, *Center for Global Development Working Papers,* No. 174

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

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

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

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

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

Temple, J.R.W. (2010). ‘Aid and Conditionality’, in D. Rodrik and M. Rosenzweig (eds) *Handbook of Development Economics*, 5: 4415-523.

Thorbecke, E. (2007). ‘The Evolution of the Development Doctrine, 1950-2005’, in G. Mavrotas and A. Shorrocks (eds) *Advancing Development: Core Themes in Global Economics*, Palgrave Macmillan: New York.

Tsikata, T. (1998). ‘Aid Effectiveness: A Survey of the Recent Empirical Literature’, *International Monetary Fund Working Papers,* No. 98/1.

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