Essays on foreign aid and economic development: macro, meso and micro studies

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at The University of Copenhagen

Edward **Sam**uel **Jones**

June 2012

Department of Economics
Faculty of Social Sciences
University of Copenhagen

Acknowledgements

This thesis has been brewing for some time. Principal thanks, therefore, go to my wife – Flávia Celina Américo Macocoro Jones – for bearing the yoke of my studies for all these years. This thesis also would not be but for the joy of our son Matthew, as well as for the encouragement, support and patience of my parents. Thank you.

Very sincere gratitude goes to Finn Tarp, my principal advisor, for his guidance, friendship and steadfast support over many years, including the last few. Many thanks also to Channing Arndt and Peter Gibbon who, alongside Finn, have been my co-authors, mentors and good friends along this doctoral path. Additionally, I have gained a great deal from regular discussions with Thomas Barnebeck Anderson, Carl-Johan Dalgaard and others at the Department of Economics, University of Copenhagen, not to mention the regular Departmental PhD seminars.

The various chapters in this thesis have been very much improved due to numerous comments and probing questions. Particular mentions are due to Phil Abbott, Tony Addison, Ernest Aryeetey, Pranab Bardhan, Bruce Bolnick, Simon Bolwig, Imed Drine, Jan Willem Gunning, Henrik Hansen, Gerry Helleiner, Paul Isenman, Homi Kharas, Dirk Krueger, David Roodman, Erik Thorbecke, Alan Winters, and Adrian Wood. Thanks also to participants at various conferences and seminars organised by the Brookings Institution (Washington, D.C.), DIIS (Danish Institute for International Studies), NCDE (Nordic Conference on Development Economics), AERC (African Economics Research Consortium), DGPE (Danish Graduate Programme in Economics) and UNU-WIDER (UNU-World Institute for Development Economics Research).

Summary (English)

The matter of whether foreign aid is effective in enhancing social welfare in developing countries remains a quotidian controversy. The present thesis addresses this question empirically and from a variety of angles. Whilst our understanding of aid effectiveness is far less comprehensive and secure than we might wish – in part due to the technical challenges of attributing observed impacts uniquely to aid – the analysis assembled here points to consistent evidence of positive effects of aid at the macro-, meso- and micro-economic levels. Simply put, there are no grounds to support cynical opinions of foreign aid which call for it to be ceased. At the same time, much can be done to augment the effectiveness of foreign aid.

The constituent chapters of this thesis may be read independently. Nonetheless, a richer understanding of the complex and diverse topic of aid effectiveness is gained by taking them as a whole. Chapter 1 provides a more substantive introduction, including a brief overview of the nature of foreign aid, the particular impact evaluation challenges she presents, and a detailed discussion of the findings of the thesis. Chapter 2 provides a complementary literature review, which considers existing evidence for aid effectiveness at various levels of aggregation. In turn this helps pinpoint a number of open and relevant research questions which motivate the remaining chapters.

Chapters 3 and 4 address aid effectiveness from a comparatively aggregate standpoint. The former takes up the issue of whether aid supports economic growth over the long-run. The latter extends the analysis to a range of other outcomes, including social indicators, and provides a decomposition of the main transmission channels through which aid affects economic growth. Chapters 5 and 6 consider two microeconomic interventions, co-financed by bilateral aid, designed to stimulate organic agriculture in sub-Saharan Africa. Specifically, the chapters consider the household welfare effects associated with smallholder contract out-grower schemes for organic coffee and cocoa located in Uganda.

Chapter 7 considers aid from the supply-side. In part motivated by the recent financial crisis (2008/09), it seeks to identify the principal domestic determinants of how much aid is given by some of the major bilateral donors. Finally, Chapter 8 discusses the advantages and disadvantages of a number of innovative aid instruments. This reflects the fact that aid is not a single coherent ‘thing’, but rather has many facets and is evolving. For this reason, it will remain an important research topic.

Resumé (Dansk)

Spørgsmålet, om ulandsbistanden er effektiv i arbejdet med at forøge den sociale velfærd i udviklingslandene, forbliver kontroversielt. Denne afhandling behandler dette spørgsmål empirisk og fra forskellige vinkler. Mens vores forståelse af bistandens effektivitet er langt mindre omfattende og sikker end vi kunne ønske – delvis på grund af tekniske udfordringer i forbindelse med at tilskrive observerede virkninger entydigt til bistanden – peger den samlede analyse her på et konsistent billede af positive virkninger af bistanden på makro-, meso- og mikroøkonomisk niveau. Sagt mere direkte, der er ingen grund til at støtte kyniske udtalelser, om at bistanden bør indstilles. Samtidig kan meget gøres for at øge effektiviteten af den udenlandske bistand.

Kapitlerne i denne afhandling kan læses uafhængigt af hinanden. Imidlertid opnås en rigere forståelse af det komplekse og forskelligartede emne, som bistandens effektivitet er, ved at se på dem i deres helhed. Kapitel 1 giver en omfattende introduktion, herunder en kort beskrivelse af den udenlandske bistands natur, de særlige evaluerings udfordringer den stiller os overfor, og en detaljeret diskussion af resultaterne i afhandlingen. Kapitel 2 gennemgår supplerende litteratur, som omfatter dokumentation for bistandens effektivitet på forskellige aggregeringsniveauer. Dette hjælper med identifikationen af en række åbne og relevante forskningsspørgsmål, som inspirerer de resterende kapitler.

Kapitel 3 og 4 drejer sig om bistandseffektiviteten fra at aggregeret synspunkt. Kapitel 3 stiller spørgsmålet om bistanden øger den økonomiske vækst på langt sigt; mens Kapitel 4 udvider analysen til en række andre resultater, herunder virkningen på en række sociale indikatorer, og gennemgår derudover de vigtigste transmissionskanaler, hvorigennem bistanden påvirker den økonomiske vækst. Kapitel 5 og 6 drejer sig om to mikroøkonomiske interventioner, der medfinansieres af bilateral bistand, og som er designet til at fremme økologisk landbrug i Afrika syd for Sahara. De to kapitler undersøger, hvorledes husholdningernes velfærd påvirkes af kontrakter med småfarmerne ved dyrkning af henholdsvis økologisk kaffe og kakao i Uganda.

Kapitel 7 ser på bistanden fra udbudssiden. Delvist motiveret af den seneste finansielle krise (2008/09), søger dette kapitel at kortlægge de vigtigste indenlandske faktorer, der bestemmer omfanget af den bistand, der gives af nogle af de største bilaterale donorer. Endelig diskuterer kapitel 8 sandsynlige fordele og ulemper ved en række innovative bistandsinstrumenter. Dette afspejler det faktum, at udviklingsbistand ikke er en enkelt sammenhængende ‘størrelse’, men istedet har mange facetter og konstant forandrer sig. Af denne grund, vil bistanden forblive et vigtigt forskningsemne.

Chapters

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> *"This is the tragedy in which the the West spent $2.3 trillion on foreign aid over the last five decades and still had not managed to get twelve-cent medicines to children to prevent half of all malaria deaths. The West spent $2.3 trillion and still had not managed to get four-dollar bed nets to poor families ... It's a tragedy that so much well-meaning compassion did not bring these results for needy people."* (Easterly 2006: 4)

> *"Because it hasn't been tried, no one really knows whether poverty on a global scale can be overcome by a truly substantial amount of aid provided without political interference. ... the annual total amount of foreign aid for the world's approximately 3 billion poor people ... comes to only about $20 per person. Should we be surprised that this paltry sum hasn't ended poverty?"* (Singer 2009: 110-111)

The above quotations indicate something of the hope and frustration which surrounds foreign aid. Some see it as essentially damaging (e.g., Moyo, 2009), others see aid as a crucial means to realise development outcomes and support sustained growth in the poorest countries (e.g., Sachs, 2006), while others argue for a total reform of the aid system due its historical under-performance (e.g., Easterly, 2003, 2005, 2006). Today, foreign aid is under additional pressure as many developed country governments implement stringent austerity programmes in the aftermath of the global financial crisis of 2008/09.

The broad subject of this thesis is whether foreign aid is effective in promoting development. As even the most casual observer will appreciate, this covers an extremely broad area. Indeed, aside from ideologically charged positions, this subject does not admit a clear starting point for discussion. In part this is because foreign aid is a very complex and multi-faceted thing. Consequently, in order to contribute something to our understanding of aid effectiveness, a basic appreciation of the nature of foreign aid is indispensable. Section 1.1 of this opening chapter thus provides a brief primer on foreign aid, organised around some stylized facts. This prompts a statement of some of the main conceptual and empirical challenges that bedevil rigorous assessments of aid's effectiveness, which are set out in Section 1.2.

With these elements of context in place, and bolstered by a review of existing theoretical and empirical studies of foreign aid (found separately in Chapter 2), a number of specific, relevant and open questions about aid effectiveness can be identified. These provide the inspiration for each of the core chapters of this thesis, which are reviewed in Section 1.3. Finally, Section 1.4 draws this introduction to a close and reflects on the principal contributions of the thesis to the wider academic literature, as well as to our overall understanding of the effectiveness of foreign aid.

1.1 Stylized facts

A first stylized fact about foreign aid is that it is extremely heterogeneous. Generically, aid refers to Official Development Assistance (ODA), as well as financing raised by non-governmental organizations and philanthropic organizations channelled to activities in developing countries.[1] Among these sources, financing instruments and modalities vary. For example: aid encompasses grants and loans; aid may be tied to specific uses (e.g., imports from specific firms); aid may be given in kind (e.g., food aid); aid may fund individual projects; and aid can be provided as sector-wide or general budget support. Also, official aid is rarely unconditional. Donors often require recipient countries to pursue reforms or implement specific policies or programmes they believe will promote development. These conditions have been the subject of debate, with some scholars arguing they have been directly counter-productive (for discussion see Sahn et al., 1999).

The above suggests a second stylized fact – foreign aid is given for starkly different purposes. The aims of aid have ranged from building infrastructure, expanding access to education, responding to humanitarian emergencies and supporting democracy and governance (among others). It is also the case that aid may be given primarily with an eye to the interests of donors rather than recipients. This occurred, for example, during the Cold War, where aid from the Western and Eastern blocs helped support collaborative regimes. It is also prevalent in tied aid, which has been used to support firms with headquarters in donor countries. In a review of aid allocation patterns, McGillivray (2003) finds that past colonial links and political alliances are major determinants of foreign aid and that such strategic factors are at least as important as variables which reflect recipient needs (also Alesina and Dollar, 2000).

A third stylized fact is that foreign aid has tended to follow fashions regarding ‘what works’ in promoting growth and development. This would not be problematic *per se* if it were not for two characteristics of these fashions. First, they have changed substantively and frequently. As one experienced commentator puts it:

> “No area of economics has experienced as many abrupt changes in leading paradigm during the post Word War II era as has economic development. ... [The] specific form of argumentation has ... remained fundamentally the same: underdevelopment is due to constraint X; loosen X, and development will be the inevitable result ... The universal remedy for underdevelopment, X, thought to be both necessary and sufficient for inducing self-sustained economic development, has varied over time, and hence so have the recommendations for the optimal forms of state-market interactions and primary policy levers.” (Adelman, 2001: 103-104).

Second, these fashions have often lacked robust theoretical or empirical foundations. Indeed, this has

[1] According to the IMF (and OECD), ODA is defined as: “Flows of official financing administered with the promotion of the economic development and welfare of developing countries as the main objective, and which are concessional in character with a grant element of at least 25 percent (using a fixed 10 percent rate of discount). By convention, ODA flows comprise contributions of donor government agencies, at all levels, to developing countries (bilateral ODA) and to multilateral institutions.” (http://stats.oecd.org/glossary/detail.asp?ID=6043).

been a persistent theme of critiques directed at the IMF and World Bank (for discussion see Pritchett, 2002; Banerjee, 2007; Temple, 2010). For example, Joseph Stiglitz (2002; 2003) argues these institutions have based their policy advice on outdated and overly simplified economic theories and have continued to advocate polices, such as capital market liberalization, despite mounting evidence they are both ineffective and contribute to global financial instability.

A fourth and related stylized fact is that the (real) volume of aid flows has evolved over time according to shifting views regarding the role of foregn aid, as well as changing perceptions of need. This is evident from Figure 1.1, which shows the value of total aid flows to different developing countries, grouped by region, over the period 1980-2008. The measure of aid used here is the sum of the committed value in billions of constant 2000 US Dollars of all entries included in the AidData database for the given region/year combination, excluding aid given as debt forgiveness and projects without a specified recipient.[2] The figure indicates that the trend increase in aid flows observed during the 1980s (growing from around 60 to 80 billion USD) was replaced by a broadly flat trend following the end of the Cold War. However, over the period 2000-2008, total aid flows rose again reaching over USD 120 billion in 2008, coinciding with increased attention toward the Millennium Development Goals. The figure also indicates some broad shifts in the allocation of aid. Flows to East Asia and the Pacific (EAP) have fallen since the late 1990s, largely reflecting rising standards of living in the region. In contrast, when viewed over the entire period, Europe and Central Asia (ECA) and sub-Saharan Africa (SSA) have come to receive more aid in both absolute and relative terms. It is also notable that Latin America and the Carribean (LAC) have often been the largest overall recipients of aid. However, in many instances this can be traced to large short-term stabilization inflows (received from the IMF), triggered by domestic financial crises such as in Argentia (1999-2002).

An important story behind these broad patterns in aid flows is one of falling aid dependence. Glennie and Prizzon (2012) calculate that since 1990, the number of countries receiving aid (excluding short-term stabilization funding) in excess of 10% of GNI fell from 45 to just 27 in 2009. The fact that this number has continued to stay low in recent years, despite some reorientation of aid toward lower income countries, testifies to robust growth in many of these countries since 2000. Indeed, it is helpful to note that the volumes indicated in Figure 1.1, have been and continue to be relatively modest. The average total value of foreign aid has been substantially lower than that of inflows of foreign direct investment to developing countries since the mid-1990s.[3] The magnitude of aid is also modest when considered in annual per capita terms. At the upper end of the distribution of per capita aid flows are the low income aid dependent countries, typically defined as receiving aid in excess of 10% of GNI. The volume of aid received by these countries only amounts to around USD 30 per person per year (e.g., Ethiopia in recent years), an extreme case being Mozambique, which received approximately 100 USD per person in 2007.[4] The implication is that the majority of aid recipients receive much smaller amounts in per capita terms (for further discussion see Tarp, 2006).

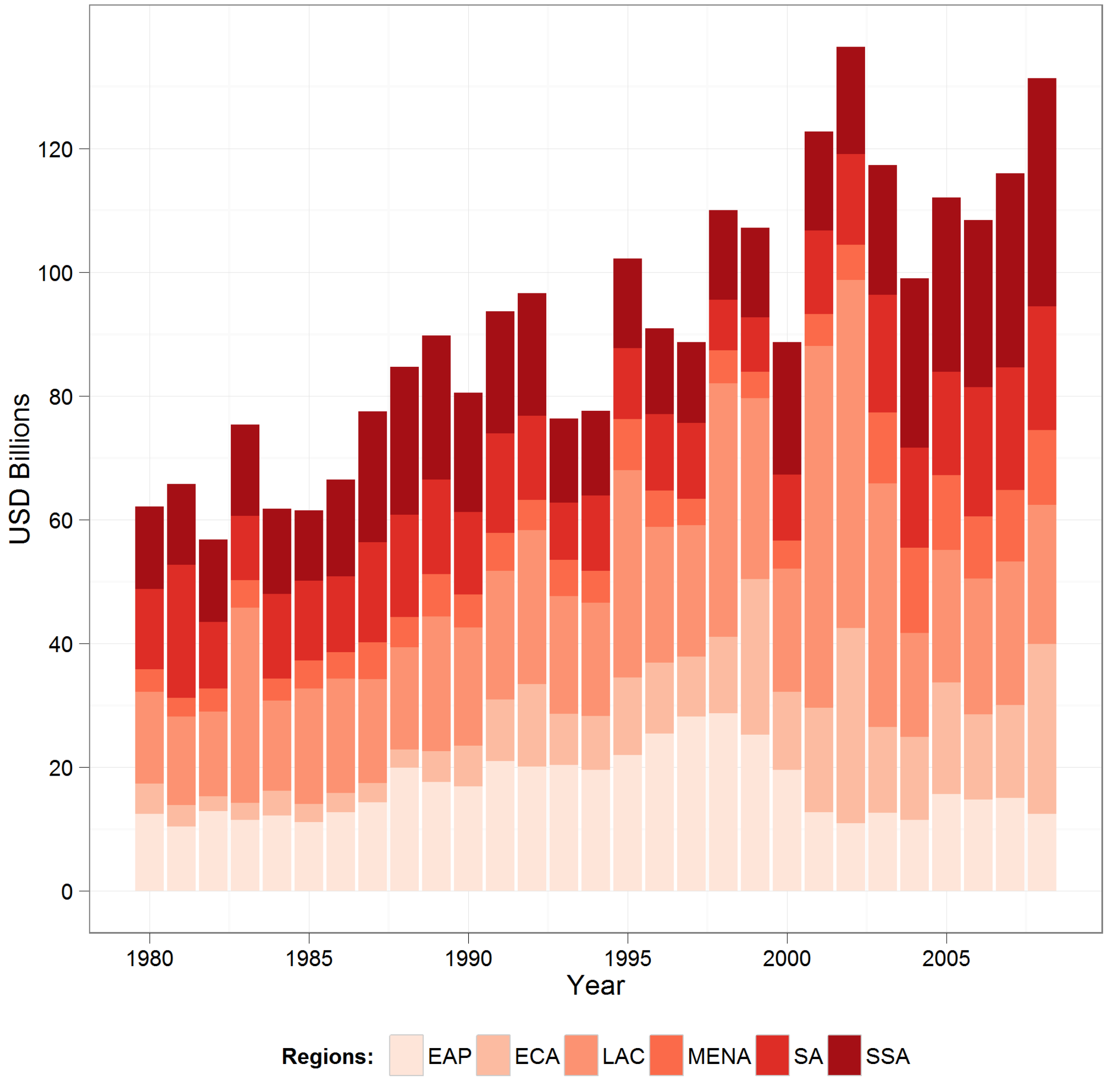
The fifth stylized fact is that aid often has been volatile and unpredictable at the country-level. A sense

[2]Commitments rather than disbursements are used as the latter is missing for a large number of entries (projects). For further details on the AidData database see www.aiddata.org. This source is used as it claims to be somewhat more comprehensive than standard OECD-DAC data on official aid flows (Nielson et al., 2009). That is, it covers OECD-DAC bilateral donors, non-DAC bilateral donors, multilateral donors (World Bank, United Nations agencies etc.) and some vertical funds. However, the database remains a 'work in progress', meaning that new historical information is regularly introduced and there is greater coverage of recent flows. Thus, the estimates should be seen as indicative, rather than accurate or comprehensive.

[3]For detailed FDI statistics see http://unctadstat.unctad.org/. In 2000 the ratio of aid to FDI inflows to developing countries was around 30%; since then FDI has grown at a faster rate than foreign aid.

[4]Estimates based on official census population estimates and AidData values illustrated in Figure 1.2.

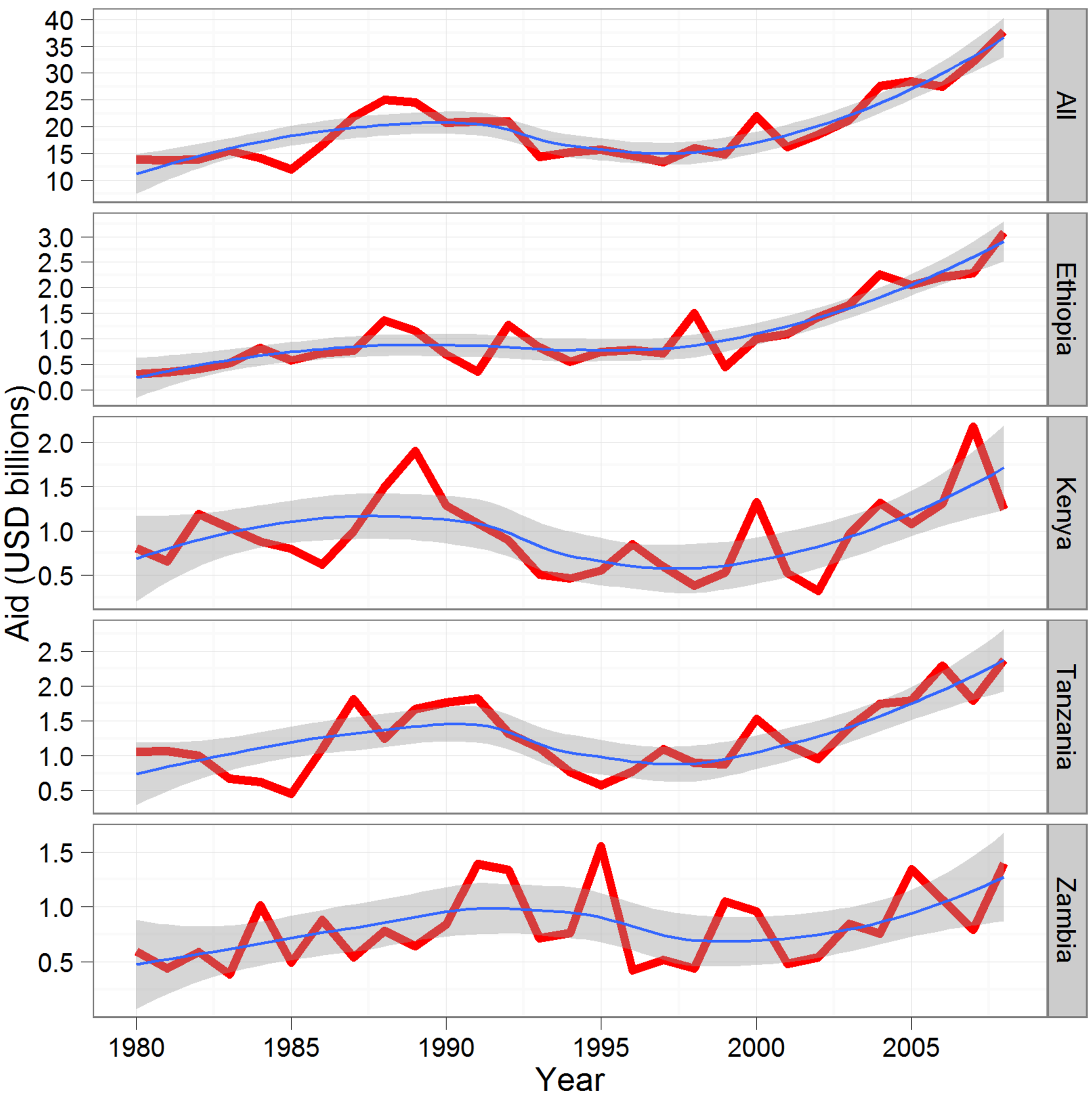
Figure 1.1: Trends in total aid flows to developing countries, by region, 1980-2008



Notes: values in constant 2000 prices; regions are as follows – "EAP" is East Asia & Pacific; "ECA" is Europe & Central Asia; "LAC" is Latin America & Caribbean; "MENA" is Middle East & North Africa; "SA" is South Asia; and "SSA" is Sub-Saharan Africa.

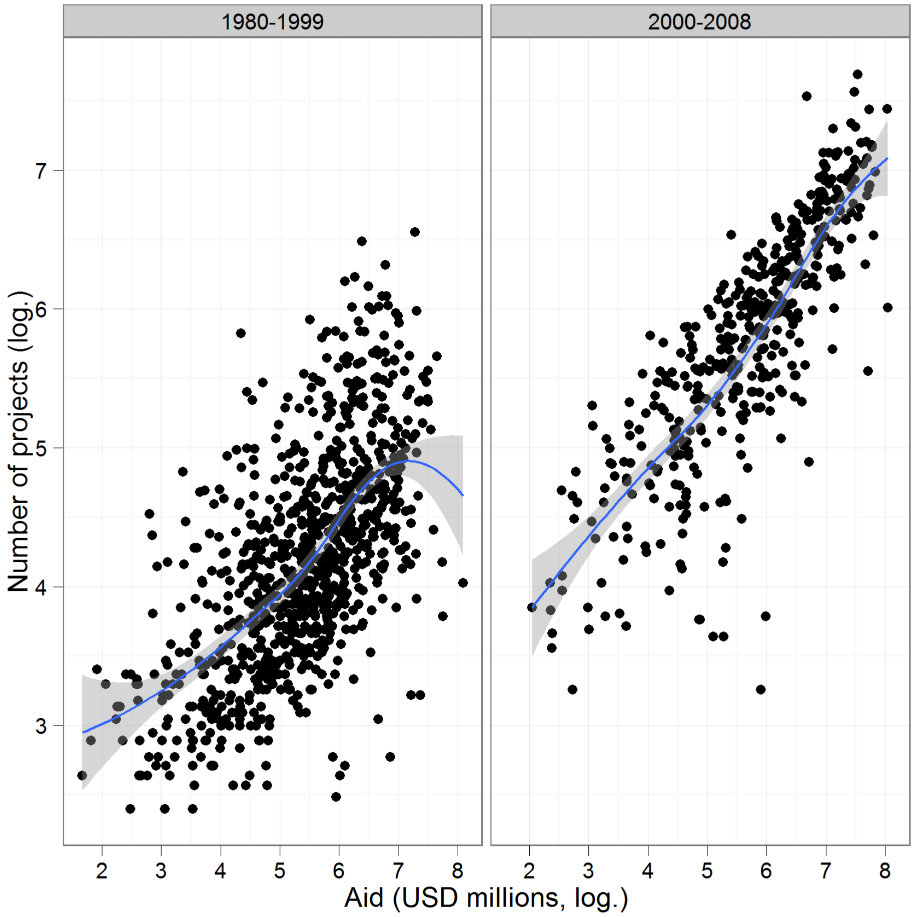
Source: author's calculations from AidData database snapshot as at 08-07-2011.

Figure 1.2: Trends in aid flows to sub-Saharan Africa, 1980-2008



Notes: blue line represents a smoothed conditional mean, with 95% confidence interval in grey shading; values in constant 2000 prices.
Source: author's calculations from AidData database snapshot as at 08-07-2011.

Figure 1.3: Scatter plot of numbers of projects and aid volumes for countries in sub-Saharan Africa, 1980-2008



Notes: x-axis plots total committed value of projects per year for a given country; y-axis plots number of projects per year and country; both axis scaled by the natural logarithm transform; blue line represents a smoothed conditional mean, with 95% confidence interval in grey shading; values in constant 2000 prices.

Source: author's calculations from AidData database snapshot as at 08-07-2011.

of this is given in Figure 1.2 which plots, using the same data, trends in aid to sub-Saharan Africa as a whole (labelled "All") as well as to selected countries within that region. The figure indicates both substantial changes over time and large differences across countries in terms of aid levels and its volatility. Indeed, year-on-year changes in aid can be very dramatic at the country level (see Bulir and Hamann, 2008), an example being the sharp fall in aid to Zambia from 1995-1996. (Further implications of these characteristics of aid are discussed in Chapter 2).

Finally, foreign aid is fragmented. The sheer number of donors and projects with which recipient countries have to deal is often very large. Using the AidData database, and continuing to focus only on countries in sub-Saharan Africa, Figure 1.3 plots the number of projects (including multi-sector funding lines such as general budget support) received by a given country in a given year, against the total committed value of these projects (both axis are scaled exponentially). The broad linear pattern shown in the panels of the figure indicates that countries which receive more aid by value do not tend to show larger average project values; thus, there are no apparent economies of scale from a recipient point of view. Moreover, contrary to the principle of harmonisation espoused in the 2005 Paris Declaration on Aid Effectiveness, there has been a relative proliferation of projects in recent years. This is, indicated by the difference in height of the conditional mean of the two panels of the figure.[5] Thus, the median project value for the sub-Saharan region fell to $0.95 million in 2000-2008, compared to $3.6 million in 1980-1999. Also, of the 35 country/year observations with more than 1000 projects, all of these occur since 2000, including a high of 2188 projects in Mozambique in 2007.

1.2 Evaluation challenges

The review of the previous section indicates something of the nature and complexity of the aid landscape. In this light, the current section turns to the challenges involved in assessing whether or not aid is effective. Five main challenges are highlighted, and can be referred to as problems of: scope, confounding, simultaneity, supply-side idiosyncrasies, and data availability.

A trivial starting point, but one which has profound consequences, is that since aid has been given for multiple objectives, it must be valid to assess aid's effectiveness across multiple dimensions. In light of the numerous types and sources of aid, it might be more appropriate to pose specific questions about aid effectiveness along the lines of: "has flow X of aid been effective in achieving Y?", where X is a particular instance of aid (possibly in a specific time and place) and Y a particular outcome of interest (to which the aid was targeted). The specificity of such questions makes them appealing. However, if one chooses to focus on narrow instances of aid, there is a risk of neglecting relevant side-effects (positive or negative). To give an example, aid given to improve education in an under-performing region may not only boost education in that region but also may generate health benefits. Nonetheless, it may undermine public services in other regions to the extent that scarce administrative resources congregate around the aid flow in detriment to other areas. Assessment of these wider effects may lead an analyst to quite different conclusions about the impact of that aid flow on social welfare, compared to remaining within the confines of one or two specific outcomes.[6]

A fundamental challenge, therefore, is how to balance and connect evidence about aid at different levels of

[5] Of course, this could be an artefact of the data – i.e., there may be greater coverage of smaller projects in recent years.

[6] See Ravallion (2012) for not dissimilar comments about the limitations of randomized controlled trials, particularly to inform policy.

aggregation (impact). This is the problem of scope. Evidence of aid's effectiveness at the micro-level (i.e., from specific projects on specific outcomes) is one part of the story but is not likely to provide a sufficient basis to inform policy-makers (donors or recipients). Rather, it must be complemented and linked to evidence at higher levels of aggregation, which can better account for spill-overs, longer-run adjustment effects and changes in aggregate welfare, which are an essential concern. Indeed, a prevailing research theme over generations has been a lack of clear correspondence between the macro- and micro-evidence about foreign aid, a challenge dubbed early on as the 'micro-macro paradox' (see Mosley, 1986, 1987).

Movement from narrower to broader outcomes and/or levels of aggregation both exacerbates and generates new technical problems in impact evaluation. These are highly pertinent due to aid's inherent complexity; but they are not specific to foreign aid *per se*. Such issues are discussed in a very large literature on impact evaluation (see Chapters 2 to 6; Blundell and Dias, 2000; Banerjee and Duflo, 2009; Imbens and Wooldridge, 2009; White, 2009b; Woolcock, 2009), which employs the concept of a counterfactual as the cornerstone of a formal and rigorous definition of causality – as is now widely used in the economics profession and throughout this thesis.[7] To say that an intervention has had a causal impact on a given outcome requires a measure of what would have happened to that outcome in the absence of the intervention, the counterfactual. In other words, the causal impact is defined as the difference between the actual outcome and the outcome (for the same unit) in the counterfactual state of the world.

A variety of problems plague construction of the counterfactual. In the context of measuring aid's effectiveness, dealing with the presence of confounders is critical. To be valid, the counterfactual must fully incorporate the effects of *all* variables that determine both the unit's exposure to the intervention (e.g., the amount or type of aid received) and the outcome, uniquely excluding the effect of the intervention itself. Since we study aid outside of the scientific laboratory, we neither cannot nor do not observe the human world in two parallel states (e.g., an identical unit observed at the exact same time, 'with' and 'without' aid). Thus, the counterfactual must be constructed artificially, typically from past observations or observations on similar units. The risk is that in so doing, the analyst fails to control for all variables that are common causes of both the intervention (or treatment) and outcome(s), such that the causal effect of the intervention is confounded.

To return to the example of assessing the impact of aid towards education, outcomes in the education sector (e.g., school enrolment rates) are likely to be driven by a variety of factors including household incomes, health and social norms which also may be correlated with receipt of the intervention. In order to isolate the unique effect of an intervention (supported by aid), the effects of changes in any one of these multiple factors must be incorporated in estimating the counterfactual. This is challenging even at the micro-level, where the variety of influences on specific outcomes can be modelled and observed more easily (albeit at some cost). At an aggregate level, the determinants of development outcomes are less well-understood, more open, and more prone to measurement error (for a discussion of this problem in relation to economic growth see Temple, 2000); thus, a wider range of possible confounders might be at play. The point is that to assess aid's impact, at either the macro- or micro-levels, problems of confounding are likely to be fundamental.

An additional difficulty, which is closely related to the problem of confounding but conceptually distinct,

[7] It is worth flagging that the use of counterfactual language to describe and analyse causality is not without its critics (e.g., Dawid, 2000). However, following proponents such as Rubin (1974, 1976) (also Morgan and Winship, 2007), the counterfactual approach has become a predominant general paradigm in the empirical social sciences, but nonetheless is associated with a variety of methodologies (see Chapter 4).

is that of simultaneity bias.[8] As noted in Section 1.1, aid typically has at least some developmental objectives. This means that countries (or sub-regional units) with lower developmental indicators would be expected to receive higher amounts of aid (on average). Thus, aid can be specified as a function of the very developmental outcomes it seeks to influence. In turn, this means that aid and developmental outcomes may be simultaneously determined (e.g., see Brückner, 2011). This problem is relevant in both dynamic and static contexts, particularly when measures of aid and outcomes are taken as averages over multiple time periods. As in the case of confounding, one way to deal with this is to find some external source of variation that explains differences in aid receipts but is not directly related to the outcome of interest – technically-speaking, a valid external instrument for aid can be used (see Chapters 2 and 3 for further discussion).

The fourth problem in assessing aid effectiveness, noted in Section 1.1, is that aid supplies have changed over time, often in an unpredictable fashion. Put more simply, the quality of aid has not been constant either across countries or over time. On the one hand this means that empirical analysis based on aggregate measures of aid may miss critical nuances in how aid may affect development outcomes. For example, aid-financed investments in new roads may have very different implications to investments in reducing infant mortality. On the other hand, a corollary of variations in aid supplies over time is that in order to analyse aid effectiveness from historical data, it may be misleading to focus exclusively on conditions and actions by recipients (on the demand-side). Rather, and as is clearly expressed in commitments intended to enhance aid effectiveness (such as made in Paris and Accra in 2005 and 2008, see OECD, 2008b), issues of how aid is allocated and administered by the donors themselves is vital to understand when, where and to what extent aid actually 'works'. This represents an added complication to the overall analytical challenge.[9]

Finally, a major challenge is that of data. The evaluation challenges already identified imply that detailed and high quality data over time is needed, at various levels of aggregation, covering not only aid flows (disbursements) but also the characteristics of both donors and recipients, and a range of developmental outcomes. The better the range, granularity and quality of the data, the better progress the we should be able to make to address the technical problems of evaluating aid's impact. Sadly, we do not have this luxury. Data on aid flows principally comes from the records of major bilateral and multilateral institutions submitted and compiled by the OECD-DAC.[10] This information has improved substantially over recent years, but historical data remains particularly patchy. For instance, prior to 2002, data on actual aid disbursements to individual countries by sector ('purpose-disaggregated aid') is frequently missing (c.f., Clemens et al., 2011); rather, for sector-specific information the analyst typically must rely on aid commitments. Historical data also typically does not cover aid from the former Eastern Bloc countries, which was significant in some instances. In recent years, 'non-DAC' donors have gained renewed importance. Donors such as China, India and Brazil are becoming increasingly important, particularly in infrastructure, agriculture and business-support sectors. These aid flows (grants or concessional lending) typically do not show up consistently in publicly available, official sources.

Data on donor/recipient characteristics, as well as outcomes of interest such as GDP growth, is generally more widely available from standard public sources such as the World Bank's World Development

[8]In econometric terms, both confounding and simultaneity can be described as specific instances or sources of an endogeneity problem.

[9]One approach to dealing with this issue has been to distinguish between different types of aid. Clemens et al. (2011), for example, focus on 'early-impact' aid defined as that portion of aid which is expected to impact on growth over the time period studied (short panels).

[10]See Section 1.1 for a discussion of the AidData initiative to improve coverage of aid data to a wider range of donors.

Indicators and the Penn World Tables. These are not without their own problems, demonstrated by missing observations (especially in the 1970s and 1980s), large jumps in certain series such as GDP due to revisions by official government agencies, changes between different data set releases (e.g., Johnson et al., 2009), and concerns that the commonly-used aggregate data series represent poor proxies for actual development outcomes. At the micro- and meso-levels, detailed data on developmental outcomes typically only comes from low frequency survey data. Indeed, evaluations of aid at these levels often requires specific survey data (see Chapters 5 and 6), meaning that only a small share of interventions can be (or have been) rigorously assessed.

To conclude this section, it is important to note that challenges of the kind mentioned above are far from unique to assessing the effectiveness of foreign aid. Many other important empirical questions in economics face similar issues and are beset by on-going controversies. A close comparison is the debate regarding the causal link between education and welfare. At the micro-level, huge effort has been expended to precisely estimate the causal effect of schooling on income (Card, 2001). At the macro-level, there is continued disagreement about the degree to which education stimulates growth, and a topical line of discussion focuses on the quality of aggregate education outcome measures (Hanushek and Kimko, 2000). Despite these debates, it is noteworthy that one does not tend to hear economists calling for funding of public education to be stopped. However, foreign aid is often lambasted as a hopeless waste.[11]

1.3 Chapter overview

The previous sections outlined some stylized facts about foreign aid and the main challenges involved in rigorously evaluating its effectiveness. Chapter 2 takes this a step further and provides a brief, but necessarily selective, review of previous economics studies of the impact of aid on a range of outcomes and at different levels of aggregation (i.e., macro-, meso- and micro- levels). The objective of the chapter is to bring the reader up to date with the state of knowledge cumulated in existing scholarship, and thereby pinpoint specific and relevant points for constructive engagement within the broad terrain encompassed by aid effectiveness. The chapter highlights that from a simple theoretical macroeconomic framework, the effect of aid on economic growth is expected to be positive and directly proportionate to the size of the aid inflow. Specifically, an inflow of aid equal to 10% of GDP is expected to increase GDP growth by around 1 percentage point. However, both the sign and magniture of this impact becomes more ambiguous as increasing complexity and general equilibrium effects are admitted, making empirical research essential. At the aggregate level, where most of the controversy about aid has focussed, the literature review indicates there has been a movement away from use of dynamic panel (GMM) approaches. Rather, there remains scope for further methodological improvements which take due account of the cumulative, long-run contribution of aid and carefully address the technical problems associated with aid's likely endogeneity.

This challenge is the subject of Chapter 3 (co-authored with Finn Tarp and Channing Arndt), which builds on the long-run cross-section approach developed in Rajan and Subramanian (2008) to investigate the effect of aid on economic growth. Based on a detailed analysis of their method, three key modifications are introduced – a better instrumentation strategy for aid; an improved specification of the aid-growth

[11] As Jeffrey Sachs stated in a recent interview, one of the biggest obstacles blocking foreign aid from effectively supporting development is cynicism amongst donor-nations (Guardian Open Weekend, posted 12 April 2012, www.guardian.co.uk/business/video/2012/apr/12/jeffrey-sachs-interview-video).

relationship; and a new econometric estimator. On its own, each modification leads to a rejection of the null hypothesis that aid has no effect on growth. Taken together, the three modifications suggest that, on average over the period 1970-2000, the impact of aid on GDP growth has been positive. The estimated magnitude of this effect is closely in line with the theoretical prior mentioned above. Finally, the chapter includes a variety of robustness and validity checks, including of the underlying instrument. These provide further support to the main conclusions.

Returning to the literature review, Chapter 2 notes that less attention has been given to the effect of aid on a wider range of meso-level outcomes, such as education and health. These effects are no less important, particularly as they are often the explicit targets to which aid is directed, and can provide an important means to sense-check findings at the macroeconomic level. At the same time, existing contributions to this literature are subject to many of the same critiques as that of earlier aid-growth studies, meaning there is scope for methodological improvements. This provides the motivation for Chapter 4 (also co-authored with Finn Tarp and Channing Arndt), which applies the same long-run methods developed in Chapter 3 to both an extended dataset (covering the period 1970-2007) and a broad range of final and intermediate outcomes. The chapter also explicitly attempts to connect meso- to macro-outcomes via a decomposition of the aid-growth effect into constituent transmission channels.

The results of Chapter 4 point to a robust positive impact of aid on growth for the 1970-2007 period, confirming the findings of Chapter 3. The aggregate effects of aid are also coherent across other outcomes. On average and over the long-run, foreign aid reduces poverty with no significant impacts on inequality. Aid also contributes to more rapid expansion of 'modern' sectors (industry), a relative decline of agriculture's share in GDP, higher investment, higher levels of average schooling and life expectancy, as well as lower infant mortality. The decomposition of aggregate aid effectiveness focuses on the effect of aid on growth through changes in physical capital, education and health, as per a standard production function. Based on systems estimators, the results confirm the strong direct effect of aid on each of these channels. It also indicates that investments in physical capital and improvements in health are the most robust transmission channels through which aid promotes growth. However, as has been found elsewhere (e.g., Pritchett, 2001), the effect of education on growth remains ambiguous, possibly due to problems associated with the underlying measure of education that has been employed.

Chapters 5 and 6 take a narrower focus, investigating specific microeconomic interventions. At this level, the link to foreign aid becomes weak and less direct. The main focus is on the impact of concrete interventions rather than the unique contributions of different financing instruments to changes in outcomes. As is noted in Chapter 2, the latter issue remains relatively unresearched; however, it is not considered here. Instead, Chapters 5 and 6 address a different gap in the literature. Specifically, they consider the impact on the welfare of smallholder farmers of interventions to simulate export farming in rural Africa by supporting organic out-grower contract farming schemes. Seed finance for these interventions was provided by the Swedish development agency, but in both chapters the actual intervention has been implemented and financially sustained by private sector operators (buying intermediaries).

These kinds of interventions are of interest for two main reasons. First, contract farming for tropical commodities in the context of liberalised agricultural markets has received little academic attention in sub-Saharan Africa. This is despite the fact that, as the role of state-run marketing boards has been heavily curtailed, competitive (and often weakly regulated) markets are now the norm in many sub-Saharan African countries. Second, very little attention has been given to the role and potential for organic farming in low income contexts. One of the main attractions of organic farming in Africa is that conversion to organic status often involves only small changes to farm processes and does not depend on repeated application of purchased inputs (for which markets are often missing). Nonetheless, certified organic products can gain access to niche export markets and, hence, command substantial price advantages. Among existing studies of this area, none rely on comprehensive farm budget-related survey data and, as such, fail to evaluate the impact of such schemes in a rigorous manner. Indeed, looking more widely, there are very few rigorous evaluations of interventions in African tropical commodity value chains. It is therefore highly relevant to deepen our understanding of these markets and the potential contribution of foreign aid to support them.

Chapter 5 (co-authored with Simon Bolwig and Peter Gibbon) considers the case of the Sipi organic arabica coffee scheme, situated on the northern slopes of Mount Elgon in Kapchorwa District in eastern Uganda. In 2005 the scheme encompassed 3,870 organic farmers. To evaluate its impact on household welfare, a small random survey of 112 participants and 48 non-participant farmers (from the same area) was administered. Using the survey data, two simple hypotheses are tested, namely: (i) whether there is a significant difference in revenue between certified organic and non-certified farmers, controlling for other relevant determinants; and (ii) whether there is a significant revenue effect from the application of organic farming practices, controlling for the participation in the organic contract farming scheme and other relevant factors. In light of the discussion in Section 1.2, identifying these causal effects is far from straightforward. Among other things, participation in the scheme was not assigned on a random basis and pre-intervention data is unavailable for the same households. To address these concerns, a Heckman selection model is employed, which is often preferred in small samples and has the advantage of being consistent in the presence of heterogeneous treatment effects. Estimation of this model suggests that the average expected effect of participation in the scheme, relative to a counterfactual of no participation, is equal to a 75% increase in net coffee revenue or between 10% and 15% of total household revenue. Evidence for the yield-benefits of applying organic farming techniques are slightly more ambiguous, but also appear to be positive.

Chapter 6 (co-authored with Peter Gibbon) extends the analysis of organic schemes in two ways. First, it analyses a different scheme for a different commodity and with bespoke survey data collected at two different periods of time (but not on the same households), providing a total of 222 observations split between scheme participants and non-participants. Second, the analysis focuses more explicitly on some of the mechanisms through which benefits are likely to accrue (a point discussed briefly in the preceding chapter). The scheme of interest is for organic cocoa, located in the remote Bundibugyo region of Uganda on the border with the Democratic Republic of Congo. Again, quantifying the causal impact of the scheme on household welfare is hampered by various potential sources of bias. However, due to a better understanding of the nature of the scheme and how the participants were selected, different analytical techniques are employed. In its most simple form, scheme eligibility is used as an excluded instrument for the share of organic grade cocoa sold by farmers to the buying intermediary (Esco). The results point to an average effect of scheme eligibility on net household cocoa income of around 100%, primarily driven by increased on-farm post-harvest cocoa processing which improves product quality and ensures farmers gain access to a price premium. Changes in farm methods, however, have not been restricted to participants in the scheme. Evidence points to a general pattern of market deepening and demand-induced technology adoption. While this cannot be attributed solely to the Esco scheme, the latter has played a leading role in providing a consistent and credible source of demand for high(er) quality cocoa. However, the chapter concludes with some cautious reflections on the extent to which these successes might be generalised.

The final two chapters return to a more encompassing view of foreign aid, but take different perspectives to earlier chapters. In part motivated by the financial crisis of 2008/09, Chapter 7 considers the determinants of aid from the supply-side. This is important with respect to enhancing the overall quality of aid, particularly as regards its predictability – flagged as a concern in previous studies (see Chapter 2). Using the 50 years of data from the full period 1960-2009, the chapter specifically examines how the supply of aid from major OECD-DAC donors has been related to domestic (home) conditions such as their level of income, share of government spending in the economy and rate of unemployment. Based on a simple empirical framework, a distinction is made between long-run supply trends and short-run dynamics, which motivates use of time series error correction methods. Panel-based econometric techniques are employed that are consistent in the presence of parameter heterogeneity and dependence between countries in cross-section. The results support the use of the error correction framework, but point to very substantial heterogeneity between countries. There is also good evidence that donor behaviour is not immutable. Rather, the importance of different domestic determinants of aid supplies changes over time. Today, aid supplies appear somewhat less responsive (but not immune) to domestic conditions in donor-countries than in the past. In turn, this means we should not be over-confident in our ability to predict future aid volumes from historical patterns.

Finally, in recognition that foreign aid continues to evolve, Chapter 8 considers emerging tendencies in the composition and nature of aid. Attention is focussed on what may be described as innovative development financing models, namely specialised global partnership funds (e.g., The Global Fund) and market-based approaches (e.g., Global-Giving). Empirical evidence about these financing models is scarce, in part due to their novelty. Nonetheless, the chapter aims to provide a balanced assessment of the strengths and limitations of these innovative approaches. To do so, the chapter presents a typology of different development financing models, which helps discern the distinctive features and scope of innovative aid models. This is supported by an overview of some of the leading examples of these models and some headline information regarding their achievements. Second, a synthesis of the strengths, limitations and weaknesses of these models is presented, based on emerging evidence from the field and relevant literature. The argument is that these innovative models are genuinely distinctive and have been particularly successful in unbundling the task of raising development funds from the task of project design, selection and implementation. Nevertheless, they do not adequately address some of the most complex problems in aid financing, which includes how to achieve long run financial sustainability and how to support lasting institutional progress (genuine capacity-building). Moreover, they often replicate established problems with traditional models of development financing – namely, fragmentation, a bias toward the interests of donors, and an extreme focus on quick results. Thus, these innovative instruments must not be seen as a panacea for aid's deficiences.

1.4 Reflections

The present thesis considers the effectiveness of foreign aid at various levels of aggregation, on a range of outcomes and from a number of different angles. To conclude this introduction, it is helpful to reflect on the main contributions of the thesis, as well as to consider potential weaknesses and gaps that merit future investigation.

Taken as a whole, a principal contribution of the thesis is that it reflects – and in so doing, makes the case for – the diversity and complexity of the aid landscape. Foreign aid cannot be taken as a single 'thing' which can be adequately assessed either at any one level or exclusively from the demand-side. As such, the challenge of rigorously evaluating the impact of aid in order to draw policy-relevant lessons is particularly acute. This thesis has addressed the complexity of aid by looking at effects at different levels and from different perspectives. Even so, in part due to the paucity of high quality data, it has not been possible to fully account for changes in the quality or objectives of aid over time.

Notwithstanding this shortcoming, a recurrent finding is that foreign aid has had a positive developmental impact. Consistent with other recent literature, the results indicate a positive and statistically significant impact of aid on growth which, when viewed as an investment, has a respectable economic return over the long-term. At the meso-level, a coherent story emerges of positive effects on a range of social indicators as well as on government finances. These effects can be linked through to the macroeconomic level, where the channels of physical capital investment and health appear to be the most robust. At the microeconomic level, a case for aid is also apparent, based here on two specific experiences in support of organic farming in Uganda. Simply put, there is no basis for the pessimistic and cynical attitude toward foreign aid that is frequently found in a range of popular and academic literature.

A second overall contribution of the thesis is that it combines some of the latest econometric techniques with the best available data to shed light on both topical and under-researched areas within the varied terrain of aid effectiveness. This is where some more specific contributions are made. At the macroeconomic level, the thesis contributes with a new cross-section instrument for aid, an improved econometric specification and a new 'doubly robust' estimator for use in the instrumental variables context (see Chapter 3). These improved methods are, for the first time, extended to a wide range of final and intermediate (meso-level) outcomes, thereby providing a consistent and robust treatment of aid effectiveness at this level (see Chapter 4). Additionally, the main transmission channels from aid to growth are identified, which complements and extends previous work, assessed using panel data, on the link between aid and savings and/or investment (e.g., see Boone, 1996; Hansen and Tarp, 2001; Clemens et al., 2011). Chapter 7 also makes a methodological contribution, in this case to the analysis of the dynamics of aid supplies. Specifically, the chapter carefully takes account of the time series properties of the data, as well as the heterogeneity of donor-specific aid supply functions. This leads to the application of more suitable econometric techniques than have been used previously.

In addition to the connection to foreign aid, the microeconomic evaluations in Chapters 5 and 6 contribute to the economics of the development of agricultural markets in sub-Saharan Africa. Both chapters explore the under-researched topic of the performance of smallholder outgrower schemes in tropical commodities operating under competitive (i.e., non-monopsonistic) conditions. The chapters contribute new data, rigorous methods and interpretation of why and how the schemes appear to be functioning as they do. The direct link to foreign aid is not the prevailing focus of these chapters; nonetheless, these evaluations confirm the potential for aid to operate as a catalyst of private sector initiatives that benefit rural households.

It should be remarked that this thesis does not view foreign aid as developmental silver-bullet. The main findings of Chapters 3 and 4, that aid has had a positive causal impact on growth and other outcomes, are only valid 'on average' and 'over the long run'. These cross-section averages are likely to hide numerous instances of statistically negligible and possibly even harmful effects of aid, particularly when viewed at lower levels of aggregation. Indeed, none of the chapters of this thesis take the view that aid either is efficient or achieves optimal results. Rather, there is ample room to enhance aid effectiveness. In this light, a further contribution of the two final chapters is that they help nuance our judgment. The chapters show that the supply of aid remains strongly influenced by short-run and long-run factors in donor countries, implying that the volume of aid is not entirely sensitive to the needs of actual recipients, nor is it likely to be efficiently allocated to meet such needs. Also, while aid is evolving, a number of innovative aid instruments continue to suffer from some of the established weaknesses of more traditional models. Thus, further and more vigorous reform of the aid architecture is necessary.

Some words of caution also might be sounded as regards how far the present findings can go to inform policy. The evidential basis for aid's effectiveness and behaviour (at all levels) has been developed from historical, non-experimental data. As such, and despite best efforts to address potential sources of bias, parameter estimates will remain open to question. This concern is particularly germane to results derived from the aggregate (cross-country) aid instrument developed in Chapter 3, as well as from the additional instruments used in the aid-growth decomposition of Chapter 4. In part this is because the process through which growth (and other aggregate outcomes) is generated is not known with certainty, meaning it is hard to exhaustively justify or test for instrument exclusion restrictions. This point has already been made (e.g., Clemens and Bazzi, 2009), and additional research is merited to explore this and other solutions to the macroeconomic endogeneity problem.

A further concern, intimated in the previous paragraph, is that the present results do not adequately attend to the heterogeneity of aid's effects – e.g., across countries, across different aid instrument (e.g., budget support) and across aid allocated to different objectives. As better and longer time series of data become available, it will be increasingly feasible to undertake robust country-specific analyses of aid effectiveness (at various levels).[12] In principle, it should also be possible to extend the analysis of Chapter 7 to simultaneously estimate the supply- and demand-side determinants of aid for individual countries over time. Last, there can be no doubt that rigorous microeconomic evaluations of aid interventions remain far too scarce (Ravallion, 2009). Evidence of this nature can only be built in a cumulative fashion to understand what works, where and why. In this vein, additional attention must be given to understanding the efficacy of different financing modes, as well as to pinpointing the underlying mechanisms at play (Chapter 6 is one attempt; for further discussion see Deaton, 2010). Also, where possible, evaluations of national-scale interventions (possibly co-financed by aid) would be valuable in order to connect project information with outcomes at the meso-level.

To conclude, this thesis has sought to push forward our understanding of aid effectiveness on a number of fronts. It has contributed consistent evidence of aid effectiveness at the macro-, meso- and micro-levels. It has examined the supply behaviour of donors, highlighting the substantial heterogeneity in how conditions in donor-countries affect aid supplies as well as changes in this behaviour over time. It also considered innovative development financing models. These have clear strengths, but they do not effectively address some of most persistent weaknesses of traditional foreign aid instruments. Notwithstanding these contributions, there is no doubt that foreign aid will continue to be a source of controversy. This being the case, aid's effectiveness remains an important and fruitful area of empirical research, not least to counter mere speculation.

[12] See Juselius et al. (2011) for a recent contribution in this regard.

Chapter 2

What do we know about the impact of foreign aid?†

2.1 Introduction

As a complement to the introduction of Chapter 1, the present chapter provides a tour of the literature concerned with the impact of foreign aid on development outcomes. The intention is not to provide a comprehensive literature review, but rather to summarise how research on aid has evolved, to identify some key findings, and to suggest gaps that merit attention. Whilst the majority of past research has been empirical in nature, a useful starting point is to develop some theoretical priors about aid's effectiveness. Indeed, a key aim of empirical analysis is to discriminate between competing hypotheses. Theoretical guidance is useful in practical terms because it helps clarify what might constitute reasonable *ex ante* expectations about the impact of aid – often a missing element in the debate.[1] It is also necessary due to the dangers of pursuing 'measurement without theory'. As Koopmans (1947) recognised long ago, without an explicit theoretical framework, the choice of what to observe may be arbitrary, the interpretation of results can be misleading, and robust policy-relevant conclusions cannot be drawn.

In that light, without seeking to provide any comprehensive treatment, Section 2.2 introduces some basic theoretical issues regarding the link between aid and aggregate economic growth. This helps to frame a plausible prior for the magnitude of its expected empirical effect. Due to their sheer variety, specific theoretical issues regarding the impact of aid on other (meso- and micro-) outcomes are not reviewed here. However, at a very generic level the simple macro-theoretical framework that is presented sheds some light at lower levels of aggregation – i.e., where development outcomes can be expressed as a function of inputs and where aid augments one or more of these inputs (e.g., in an education production function). The point is that the final impact of aid on selected outcomes crucially depends on the productivity of the specific inputs affected by aid, as well as their interaction with other variables in equilibrium.

Turning to the empirical evidence, detailed literature reviews of foreign aid are readily available elsewhere (e.g., Tsikata, 1998; Hansen and Tarp, 2000; Hjertholm and White, 2000; Tarp, 2000, 2006; Riddell,

†This chapter loosely draws on, but substantially extends, the background article published as Arndt et al. (2012).

[1] Inflated expectations about what aid can achieve are suggested by the phenomenon of persistent public over-estimates of the share of the budget devoted to foreign aid (for the USA see Destler, 1996).

2007; Temple, 2010). A limitation of these, however, is that they tend to exclusively address the impact of aid on a single aggregate outcome – GDP growth. This is no doubt important; but, aid is often intended to have impacts in other domains. Indeed, there is a small but growing literature on aid effectiveness at the meso-level, which refers to outcomes at the sector- rather than the economy-wide level (e.g., on education or health). There is also long history of assessments – produced both by academics and donors – of individual aid projects. These dimensions deserve treatment precisely because of concerns that aid may be less than the sum of its parts. In other words, to understand aid's impacts we must look both at its parts (micro- and meso-outcomes) as well as the sum of these (macro-outcomes). Section 2.3 thus reviews the accumulated evidence from previous studies, distinguishing between insights at different levels of aggregation (i.e., macro-, meso- and micro-levels). Section 2.4 concludes with some remarks on the current state of the literature and suggestions about future lines of research.

Before proceeding, a few caveats are required. In order to narrow down the vast range of studies concerned with foreign aid, the focus will be limited to academic economics studies. In large measure, and most prominent in the recent literature, these take a counterfactual perspective and, at least implicitly, define the impact of foreign aid as the difference in developmental outcomes in the 'with aid' versus 'without aid' scenarios (see Chapter 1). Also, this review is backward looking. As such, it is open to the critique that it refers to outdated models of 'doing' aid and does not reflect recent progress, such as the emergence of vertical and philanthropic funding mechanisms, changes following international initiatives such as the Paris Declaration on Aid Effectiveness and the Accra Agenda for Action, and the shift toward sector and general budget support. Be this as it may, these developments are comparatively new and effective change on the ground remains slow in many areas (for discussion see Chapter 8; also OECD, 2008a; Armytage, 2011), meaning there is scarce rigorous empirical evidence for their effectiveness (positive or negative).

2.2 Aid in theory

A primitive model of the impact of aid on aggregate income can be derived from a generic production function and national accounts identities.[2] Thus, define national income (Y), as a function of capital (K) and labour (L):

$$Y = f(K, L) \Leftrightarrow dY = f_K dK + f_L dL \tag{2.1}$$

where dY is thre total derivative of income and f_K,f_L are partial deerivatives of income with respect to capital and labour respectively. Second, recall the standard expenditure identity:

$$Y = C + I + X \tag{2.2}$$

which states that income is the sum of domestic consumption (C), investment (I) and net exports (X) and where no distinction is made between the public and private sectors. Lastly, we define two behavioural equations and one identity, namely:

$$C = (1-\rho)Y + (1-\beta)F, \ \rho \in (0,1), \beta \in [0,1] \tag{2.3}$$

$$K_t - K_{t-1} = dK = I \tag{2.4}$$

$$X = -F \tag{2.5}$$

[2]This exposition draws on Obstfeld (1999). Please refer to Appendix 2A for a summary of variables and parameters used hereafter.

where ρ is the marginal propensity to save in income, F represents the balance on the capital account (assuming reserves are held constant), which for our purposes can be considered an inflow of foreign aid, and β captures the share of aid that does not leak into consumption.

Combining equations (2.1) to (2.5) gives:

$$Y = (1-\rho)Y + (1-\beta)F + \frac{dY - f_L dL}{f_K} - F \tag{2.6}$$

which, rearranging, provides a general formula for proportional changes in income:

$$g_Y = \frac{dY}{Y} = \frac{f_K}{Y}(\rho Y + \beta F) + \frac{f_L dL}{Y} \tag{2.7}$$

and, in turn, yields:

$$\frac{\delta g_Y}{\delta(F/Y)} = f_K \beta \tag{2.8}$$

The above implies that the expected increment to income growth from a positive aid inflow depends, *ceteris paribus*, positively on the product of the share of aid devoted to investment and the marginal product of capital. Notably, assuming a constant-returns to scale production function and perfectly competitive domestic capital markets (see Caselli and Feyer, 2007), the latter is equal to the rental rate on capital; i.e., $r = f_K \Rightarrow \delta g_Y / \delta({}^F\!/\!_Y) = r\beta$.

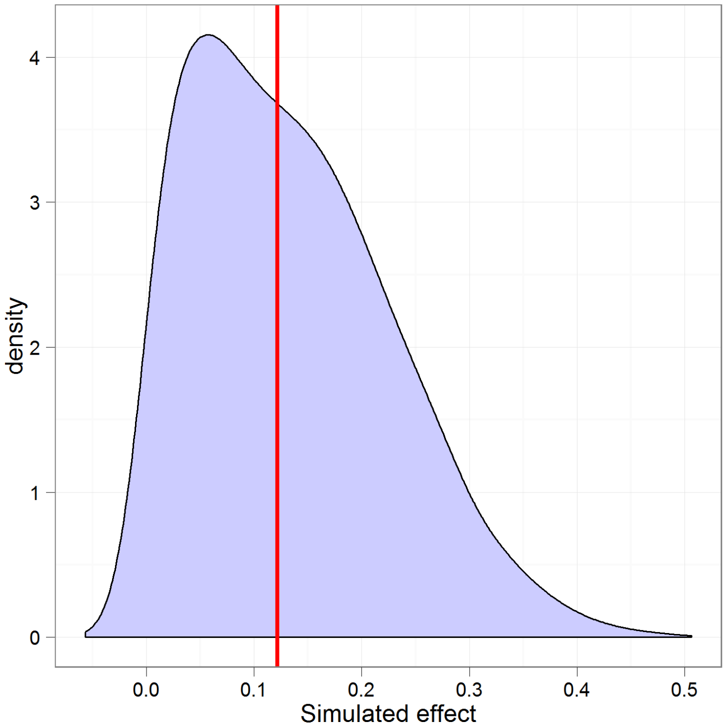
Three important points fall out of this framework. First, the model assumes that the only mechanism through which aid affects income growth is via investment in physical capital. This is a crude simplification and neglects direct and indirect effects of aid on human capital (e.g., education) and productivity (e.g., governance).[3] Nonetheless, it makes the point that the aggregate effects of aid occur through mediating variables, such as proximate sources of growth (for discussion of types of growth determinants see Sachs and Warner, 1997). Second, the exposition highlights that aid effectiveness is dependent on local conditions, including domestic returns on capital. Thus, aid impacts are best understood in specific local contexts. Third, the framework indicates that expectations about what aid can achieve also should be relatively modest. This is illustrated, in a purely suggestive sense, in Figure 2.1, which plots an estimated (mixture) distribution for $\delta g_Y / \delta({}^F\!/\!_Y) = r\beta$ based on 10,000 random draws for r and β. The figure indicates a median 'growth return on aid' of around 0.12, meaning that an Aid/GDP ratio of 10% (which would be high in comparative historical terms; see Chapter 1), would be expected to augment the growth rate by just over 1 percentage point. Moreover, the vast majority of the distribution lies in the positive domain.[4]

The need for modest expectations accords with similar results in Rajan and Subramanian (2008) and Dalgaard and Erickson (2009), where specific functional forms for Y are employed. This view is reinforced when exchange rate effects associated with capital inflows are considered (à la Dutch Disease), which raises the spectre of capital appreciation, weakening of the current account balance and, possibly, lower growth. These effects are neglected in the previous framework but become explicit from a modified Mundell-Fleming model (see Lane et al., 1999; Boughton, 2006). Together with equation (2.5), comparative static analysis of the national accounts identity:

$$Y = C(Y, E) + I(F, r) + X(Y, E) \tag{2.9}$$

[3] As Temple (2010) and others have argued, this simplification begs the question as to why aid is required – that is, if aid generates a return equal to that of private capital, some explanation for the absence of sufficient private capital is required.
[4] See Rajan and Subramanian (2008) for similar results based on a necoclassical growth model.

Figure 2.1: Simulated distribution of the effect of aid on growth



Source: authors' calculations.
Notes: effect is estimated as $r\beta$, where β is drawn randomly from a uniform distribution on the unit simplex. For illustrative purposes, r is drawn randomly from a normal distribution with mean of 27% and standard deviation of 9% as per the naïve cross-country empirical estimates provided by Caselli and Feyer (2007); 10,000 observations are used; Gaussian kernel applied; median = 0.12, indicated by the vertical line.

where E is the real exchange rate, gives:

$$\left.\frac{dY}{dF}\right|_{dr=0} = -\frac{1 + C_E/X_E - I_F}{1 - C_Y + (C_E X_Y)/X_E} \tag{2.10}$$

Assuming Cobb-Douglas technology such that $r = f_K = \alpha Y/K$, and setting $F = \gamma K, (\gamma > 0)$, we have the (partial) point elasticity:

$$\left.\frac{dY/Y}{dF/F}\right|_{dr=0} = \varepsilon_{YF} = -\left(\frac{1 + C_E/X_E - I_F}{1 - C_Y + (C_E X_Y)/X_E}\right)\frac{\alpha\gamma}{r} \tag{2.11}$$

Interpretation of this result is more involved, but two main insights stand out. First, under usual assumptions regarding the signs of the partial derivatives, ε_{YF} will be positive only if $I_F > 1 + C_E/X_E$, and thereby depends on the relative sizes of the consumption versus net export exchange rate effects. Second, the magnitude of the point elasticity is dependent on the ratio of the capital share (α) to the return on capital. This underlines that as greater complexity is included in our theory framework, the nature of the expected impact of aid on growth becomes increasingly ambiguous. In similar vein, Temple (2010) notes that the number of theoretical models regarding foreign aid exceeds the number of available units (observations), meaning it is impossible to discriminate between competing theories. Simulation models of the aggregate impact of aid confirm the possibility of complex general equilibrium effects. For instance, Adam and Bevan (2006) find that when there is an existing shortage of physical capital, there can be potentially large medium-run gains from aid-financed (public) investments, despite short-run Dutch Disease effects. However, these effects are sensitive to domestic conditions, particularly the degree of domestic bias in the aggregate supply effects. In short, empirical analysis is essential to progress.

2.3 Aid in practice

2.3.1 Macro-level studies

Academic economic studies seeking to quantify the effectiveness of aid have most often focussed on the relationship between aid and aggregate outcomes. There are two main reasons to concentrate on this level. First, aggregate or macroeconomic outcomes are of especial interest from a social welfare and overall policy perspective. To improve living standards significantly, poor countries must produce more (per capita). To do so requires countries (and regions) to initiate and sustain long run processes of building physical capital and human capital, acquiring technology, and nurturing institutions that facilitate growth. The role of foreign aid for development, broadly conceived, is to support these long run cumulative processes. Evidence that such process are taking place across an economy is most frequently available from aggregate indicators such as investment (savings) and GDP growth.[5] Second, aggregate-level evaluations not only complement the variety of project (micro-level) evaluations undertaken by donors (see Section 2.3.3) but also are better able to take account of possible spillovers from individual aid-financed interventions. Indeed, explaining what has been described as the ‘micro-macro paradox’ (Mosley, 1986, 1987), which is the apparent inconsistency between evidence for highly satisfactory *ex post* rates of return to foreign aid at the micro-level but lower and more ambiguous returns at the macro-level, has remained a guiding concern of research.

Following Tarp (2006), among others, research regarding the aggregate impact of aid can be classified into different generations. Inspired by the Harrod-Domar and two-gap models, the first generation considered the link between aid and savings. As the detailed review of Hansen and Tarp (2000) notes, these studies generally found that aid tends to be correlated with higher total savings; nonetheless, they also suggested a (smaller) proportion of aid is likely to be consumed. The second generation of literature, associated with numerous studies of the 1980s and 1990s, turned attention to the linkages from aid→investment and from investment→growth. These studies, also reviewed in Hansen and Tarp (2000), consistently found a positive link between aid and investment. However, the relationship from investment to growth was not found to be robust, echoing a wider debate over the cross-country determinants of growth (e.g., Sala-i-Martin, 1997).

In part due to the small number of observations available, these first two generations largely relied on static cross-section OLS regressions, with each observation representing the time average for a given variable and country. However, studies in the second generation showed an increasing awareness that OLS estimates of the impact of aid may be biased. In particular, scholars recognised that economic growth may itself be a determinant of aid – i.e., countries with low incomes and slow growth may receive

[5]This is not to claim that such standard (national accounts) indicators are unproblematic. In particular, they tend to neglect (changes in) the distribution of income, which also is an essential component of development.

more aid precisely because of their poor economic performance. This would introduce simultaneity bias into OLS estimates of the aid-growth relation. To address this, a small number of second generation studies employed instrumental variables (IV) estimators, often as robustness checks on OLS results (e.g., Boone, 1996). In retrospect, however, these applications were often rudimentary and problematic, reflecting a nascent understanding of the properties/use of IV estimators within the profession at the time, particularly as applied to macroeconomic phenomena. For example, in one of the more sophisticated studies, Mosley et al. (1987) set up a three stage least squares (3SLS) system incorporating separate equations for growth, aid and mortality; and where aid is specified as both a determinant and a function of the other two endogenous variables. However, the study includes limited discussion or analysis of whether the excluded instruments are appropriate (valid and relevant). As such, it is difficult to assess whether their results are consistent.

The third generation of literature, which emerged in the later 1990s, took advantage of newly available panel data and, frequently, dynamic panel estimators (e.g., Arellano-Bond and Blundell-Bond GMM methods). This methodological shift held distinct promises – a stepwise increase in the number of observations; methods to deal with country-specific fixed effects; and the possibility of using internal instruments to deal with endogenous variables such as aid. Also, this literature came to focus almost exclusively on the reduced form aid-growth relation, as opposed to structural models (as per Mosley et al., 1987) or the wider range of outcomes investigated in early generations. Broadly speaking, a message of guarded optimism emerged from many such studies. One of the most cited of these, due to Burnside and Dollar (2000), concluded that: "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, p. 847). However, Hansen and Tarp (2001) (also Easterly et al., 2004) found these results to be fragile, a view echoed by Roodman (2007) who concludes that the findings of this generation are extremely sensitive to methodological choices.

More recently, the outlines of a fourth generation of literature has emerged. Again, this marks a departure from previous generations on the basis of method. Specifically, the widespread use of dynamic panel methods to deal with endogenous variables has received scrutiny. This comes from four main angles. First, 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 systems approach. Second, 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 used as instruments for the level equation. This assumption cannot be tested but may be suspect given the highly complex nature of the growth process and 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) indeed conclude that the principle issue for system GMM is not one of strong or weak instruments, but rather the validity of these moment conditions. Third, Roodman (2009) warns that the Blundell-Bond estimator may yield a false sense of certainty as use of a large number of internal instruments can over-fit the endogenous variables and may weaken the power of Hansen/Sargan tests. Fourth, internal instruments do not avoid bias arising from systematic measurement error in the endogenous regressors, which is an important limitation in the context of aid-growth regressions (see below).

Cognizant of these concerns, a range of recent studies has moved away from this combination of panel data and dynamic panel GMM estimators. One of most influential is Rajan and Subramanian (2008), who return to static cross-section methods. However, in contrast to the first and second generations of literature, they do so with the benefit of longer time series (e.g., 1970-2000). Using the rich data on donor-recipient flows at the country-level, they carefully develop an external instrument for aid based on its supply-side determinants. Their results indicate no systematic effect of aid on growth regardless of the time period or measure of aid used. The latter study provides the starting point for the contribution in Chapter 3 (published as Arndt et al., 2010a), which makes a number of methodological enhancements remaining within a long-run cross-section framework. Other studies, applying different methods, include Clemens et al. (2011) and Brückner (2011). Not only do these studies all confirm a positive impact of aid on growth, but they also suggest a parameter estimate for this effect which is well within the domain identified theoretically in Section 2.2.

Finally, an alternative perspective on the aid-growth literature comes from meta-studies. The professed benefit of this approach is that it is less subjective than a generic literature review, but remains agnostic as to the 'correct' method or dataset. That is, while it is assumed there is a true effect to be found – be it a point estimate or distribution – no one approach is given emphasis over others. A series of papers in this vein is due to Doucouliagos and Paldam (2006; 2008; 2010), who undertake a systematic analysis of the large number of aid-growth regressions contained in over 68 papers produced since the 1980s, together covering the period 1970-2004. Their main finding, also summarised in Doucouliagos and Paldam (2009), is that the literature has failed to establish a significant positive impact of aid on growth. However, this conclusion is strongly contested. Mekasha and Tarp (2011) review the same papers as Doucouliagos and Paldam (2008), but correct a number of the latter's methodological and coding errors, and also take due account of effect heterogeneity over countries and time. In doing so, they find that the null hypothesis of 'no effect' of aid on growth is consistently rejected at conventional significance levels, and that the average effect of aid on growth is located squarely in the domain anticipated by theory (see Section 2.2). With the caveat that individual studies are potentially fragile (as per the critique due to Roodman, 2007), this seems a reasonable summary of the current state of knowledge regarding the macro-impact of aid.

2.3.2 Meso-level studies

Recent macro-level studies focus on one outcome – growth. However, as indicated in Section 2.2, and explicit in earlier generations, aid is expected to influence growth through various intermediary channels. These channels, which generally refer to meso-level outcomes, are also of direct empirical interest. First, responding to a critique of the aid-growth literature, they can point to the mechanisms that might link aid to growth. Indeed, a recent wave of scepticism about aid (c.f., Moyo, 2009; Rajan and Subramanian, 2008), has stimulated a search for mechanisms that might undermine its effectiveness. Second, it is widely recognised that growth is not the only objective of development. Ensuring an equitable distribution of wealth and improving (in terms of access and quality) key public services such as education and health can be considered at least equally vital to progress in human welfare. Third, given the methodological challenges associated with estimating causal effects at the macro-level, perhaps sharper insight may be gained from investigating narrower outcomes.

Plausible explanations for nefarious impacts of aid coalesce around three main themes. The first is political economy effects. Djankov et al. (2008) argue that the impact of aid is analogous to a natural resource curse because it stimulates rent-seeking behaviour and engenders a fall in overall governance quality (also Svensson, 2000, 2006; Rajan and Subramanian, 2007).[6] In similar vein, Brautigam and Knack (2004) and Knack and Rahman (2008) point to the costs of dealing with multiple and fragmented donors, often pursuing inconsistent and/or overlapping objectives. Empirically these outcomes are hard to verify, largely because they are not directly measurable. Nonetheless, based on various researcher-constructed indexes (e.g., the International Country Risk Guide), the same studies point to evidence of lower quality of governance, bureaucratic inefficiency and weaker public service delivery in the presence of larger aid flows.

The second theme refers to the fiscal costs of aid due to distorted government budgets. On the revenue side, aid may undermine domestic revenue mobilisation, in the extreme being consistent with a fall in tax revenues equal to the aid inflow (see Azam et al., 1999). Such concerns have been voiced frequently (Moss et al., 2006), but supporting empirical evidence is ambiguous. Pivovarsky et al. (2003) find a positive revenue impact from concessional loans but a small negative effect from grants. Teera and Hudson (2004) find that foreign aid is not a significant determinant of tax revenues across a range of developing country subsamples, a finding largely echoed by Baunsgaard and Keen (2010). On the expenditure side, the fungibility of aid may mean that real spending does not increase by the full amount of the aid flow, such that aid supports fiscal consolidation. Equally, aid may be fungible between sectors, for example, meaning that aid earmarked for education, does not boost education spending in a counterfactual sense. Quantitative evidence for these effects also is mixed, but broadly points to the existence of fungibility at lower levels of aggregation. Pack and Pack (1993) find that aid significantly undermined developmental public expenditures in the Dominican Republic over the period 1968-1986. Based on cross-country evidence, Feyzioglu et al. (1998) conclude aid is most fungible between sectors within a given country. In contrast, Pack and Pack (1990) find a positive effect of aid on expenditures both on aggregate and for intended sectors in Indonesia – i.e., there is a 'flypaper effect' (see also van de Walle and Mu, 2007).

The third theme refers to the challenges of dealing with large and lumpy capital inflows. Bulir and Hamann (2008) and Hudson and Mosley (2008), among others, point to the macroeconomic and fiscal costs associated with volatile and unpredictable aid inflows, especially for low income countries with limited access to external private capital markets. From the perspective of undermining income growth, concerns also persist that foreign aid can lead to (excess) exchange rate overvaluation, thereby slowing the growth of export and manufactured sectors which, in turn, may be considered important sources of technological learning and productivity enhancement (Rodrik, 2009). Empirically, Rajan and Subramanian (2005, 2011) find this effect to be important – e.g., that the rate of growth of value added of manufactured exports in developing countries has been undermined by upward pressure on exchange rates induced by aid. Nonetheless, in a recent review essay, Magud and Sosa (2010) point out that there is very little evidence that Dutch Disease effects undermine economic growth at the economy-wide level. Thus, the issue is one of effectively managing resource inflows rather than avoiding them *per se*.

With respect to the effectiveness of aid in improving aggregate social outcomes, evidence is generally more positive. For example, using a range of econometric methods and differing measures of aid, various authors conclude that aid helps lower infant mortality (e.g., Gomanee et al., 2005; Masud and Yontcheva, 2007; Mishra and Newhouse, 2009) and has enhanced education access and outcomes (e.g., Michaelowa, 2004; Dreher et al., 2008).[7] Foreign aid also has supported the expansion of specific public goods. For

[6]As one author explains: "In Tanzania ... large and rising aid levels in the 1970s and 80s helped sustain large government subsidies to state-owned enterprises and parastatals. Larger public sectors create more opportunities for corruption. If public firms displace private investment, a weakened private sector produces less pressure on government to establish accountable and transparent procedures and institutions" (Knack, 2001, p. 313).

[7]With some exceptions and variations, the preferred econometric method among these studies has been in keeping with the third

example, the abolition of user fees in low income countries, often financed with donor support (such as the UNESCO-led Education for All initiative), appears to have generated large and rapid improvements in access to primary services, especially amongst the poorest (e.g., Deininger and Mpuga, 2005). Also, many commentators point to the huge increase in access to HIV/AIDS treatment due to the US President's Emergency Plan For AIDS Relief (PEPFAR/Emergency Plan). According to Gerson (2008), for example, since 2005 the number of people in sub-Saharan Africa with access to antiretroviral medicines has increased from 50,000 to 1.4 million.

2.3.3 Micro-level studies

Evidence about aid effectiveness at the micro-level typically concerns the performance of individual donor-financed projects. In principle, there is a comparative wealth of information here as (one assumes) all donors track the status and performance of their project portfolio in some manner. Taken at face value, this evidence is generally favourable. For example, since 1996, over 65% of the World Bank's projects received a rating of 'moderately satisfactory or better' (Independent Evaluation Group, 2010). Similarly, a review of the performance of the UK's Department for International Development over the period 2005-2008 found that 75% of projects were 'expected to achieve or largely achieve their objectives' in 2008, up from 65% in 2000/01 (Conlin and Beauean, 2009), leading the authors to conclude 'there is credible evidence that, overall, [project] efficiency and effectiveness are improving'. This kind of result is not new. As noted previously, a favourable trend at the micro-level compared to more ambiguous results at higher levels of aggregation was identified in the 1980s by Mosley (1987).

Despite the above, and as Temple (2010) suggests, at least since the mid-1990s, donor-portfolio performance information generally has not been employed by the research community as raw data for investigating aid effectiveness (for an exception see Cassen and Associates, 1994). A number of reasons might explain this. The first is a lack of access to the relevant information at a reasonable level of detail or sufficient scale (e.g., across multiple donors), a trend which appears to be changing as demands for greater transparency mount.[8] Second, the quality and credibility of internal project information remains subject to doubt. Different donors record different information and apply alternative definitions of success. This is true even for sister organizations; e.g., as the Independent Evaluation Group note:

> "Project ratings across the Bank Group are not comparable, as they refer to distinct frameworks and methods. The World Bank uses an objectives-based system ... [while the] IFC [International Finance Corporation] and MIGA's [Multilateral Investment Guarantee Agency] project rating systems are based on quantitative and qualitative benchmarks rather than on achievement of specific development objectives." (2010, p. 13).

Furthermore, success is often defined in terms of process outputs (e.g., number of mosquito nets delivered) rather than development outcomes (e.g., reduction in cases of malaria). Even when the latter are employed, such outcomes typically are not defined in counterfactual terms. Thus, as White (2009a) emphasises, it is not possible to attribute causal effects to the project as opposed to other determinants. Third, a persistent critique of project information systems is that they generally do not track outputs/outcomes

generation of the aid-growth literature – i.e., use of dynamic panel GMM techniques. As already noted, these have been the subject of criticism.

[8] See, for example, the International Aid Transparency Initiative: www.aidtransparency.net.

once the project has been completed. Thus, we do not gain insight into longer run effects including their sustainability.

In addition to portfolio information, there are formal evaluations of donor activities at the project, sector, country and regional levels. These too have been subject to criticism. Most prominently, the vast majority of such evaluations have focussed on process rather than counterfactual impacts. As one review makes clear: "Documentation shows that UN agencies, multilateral development banks, and developing country governments spend substantial sums on evaluations that are useful for monitoring and operational assessments, but do not put sufficient resources into the kinds of studies needed to judge which interventions work under given conditions, what difference they make, and at what cost." (Savedoff et al. 2006: 2). To address these knowledge gaps, randomized control trials (RCTs) have been employed to investigate the efficacy of developmental interventions over recent years (for discussion see Banerjee and Duflo, 2009; Barrett and Carter, 2010; Card et al., 2011). Results from such trials, as well as those of a relatively small number of rigorous *ex post* impact evaluations, point to strong welfare gains in some, but not all, instances. By way of examples, Kremer and Miguel (2004) found that deworming treatment given to children from randomly selected schools in a Kenyan district improved health and increased school participation, but also had negative peer effects; Lokshin and Yemtsov (2005) found that rehabilitation of school infrastructure yielded large gains for the poorest households relative to other infrastructure projects in rural Georgia; Jalan and Ravallion (2003) found interventions providing piped water reduce the prevalence and duration of diarrhoea among young children, although this effect is weaker for children from poor families; and van de Walle and Mu (2007) report positive economic effects from support to rural road infrastructure for the kilometres of roads rehabilitated in Vietnam.

In sum, the *potential* for aid to be effective at the micro-level has been demonstrated across a range of interventions and settings. However, it is important to note that evidence from RCTs typically informs about the effectiveness of different policy interventions and does not address the effectiveness of the financing method (e.g., foreign aid vs. domestic government funding). Also, as Ravallion (2012) opines, the usefulness of RCTs to inform government policy is open to question, particularly where the government's implementation capacity is limited.

2.4 Conclusion

This chapter has reviewed the economics literature concerned with aid effectiveness. Looking across a wide-range range of studies, including micro-, meso- and macro-level outcomes, three main conclusions can be highlighted. First, the weight of evidence points to the existence of positive causal effects of aid on average and over a range of development outcomes. Such estimates are in line with a simple theoretical framework, particularly when longer time frames are considered. Second, a major drawback of previous literature is that the stock of internally valid evidence remains less complete and less convincing than is desired (see Ravallion, 2009). In part this is because the effectiveness of aid has frequently been evaluated in a manner that fails to pay due attention to the problem of how to distinguish its impact from other factors. Also, despite the explosion of RCTs, these have most often been informative about interventions in the social sectors (health and education) as opposed to more complex interventions associated with governance and enterprise-growth.

A third conclusion refers to the appropriate time-frame over which any effects accruing from aid can be expected to materialize. It stands to reason that at least some of the channels through which aid may affect more aggregate outcomes are likely to exert a cumulative but not immediate impact on the rate of income growth. For example, changes in education and health move only slowly at the aggregate (population) level and thus can have a positive influence on economic growth with a substantial lag. This follows from simple demographics whereby improvements in schooling indicators 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.[9] Existing literature has not given due weight to this challenge. Indeed, at the macro- and meso-levels, the predominant method for assessing aid's impact has been to use (dynamic) panel data. By construction, this limits the estimated impact of aid to the periodicity of the panels (typically periods of 5 years).[10] At the micro-level, a problem is that impact evaluations typically occur shortly after the conclusion of a project (or during its implementation). Thus, the longer-run effectiveness, spill-overs and sustainability of such interventions are unlikely to captured fully.

In sum, our knowledge about the effectiveness of foreign aid has progressed but many questions remain open. Four areas for future research can be suggested. First, there remains scope for methodological improvements in how aid's impact is quantified at the aggregate level, in particularly giving due importance to the cumulative, long-run effects of aid and dealing (as best possible) with problems of endogeneity. Second such methodological improvements can be extended to the assessment of aid over a wide range of meso-level outcomes, the aim being to build a more consistent picture of how aid works in practice at this level. Third, to enhance the policy relevance of research, we need to pay greater attention to pinpointing the principal mechanisms through which aid is and is not effective. Insight of this nature has been scarce in previous literature but is relevant at all levels of aggregation. Finally, rigorous micro-level evaluations of aid should consider neglected areas. These include the role of aid in promoting economic activities such as smallholder agricultural exports, the relative importance of *how* project interventions are financed, and the long-term impact of aid-financed interventions.

[9] Ashraf et al. (2009) and Acemoglu and Johnson (2007), for example, find that the initial economic impact of gains in life expectancy from the health interventions introduced from the 1940s 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-intervention 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. (2009) 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.

[10] Clemens et al. (2011) go some way to address this issue by using lagged aid and restricting their measure of aid to be "early impact" in nature.

2A Appendix: summary of variables and parameters

Variables	Definition
Y	National income
K	Physical capital stock
L	Labour
C	Domestic consumption
I	Investment
X	Net exports
F	Balance on the capital account (foreign aid)
E	Real exchange rate
r	Real rental rate on K
Parameters	**Definition**
ρ	Marginal propensity to save in income
β	Share of aid devoted to investment
α	Share of income spent on capital (Cobb-Douglas production function parameter)
γ	Aid to physical capital ratio

Chapter 3

Aid, growth, and development: have we come full circle?†

3.1 Introduction

The discussion of the previouis chapter suggested that there remain concerns that aid is not effective at the aggregate level and, at worse, may be counter-productive. This view is not infrequently found in the popular and academic press (e.g., Moyo, 2009; Djankov et al., 2008). At the same time, the fundamental empirical challenge of identifying an appropriate and valid counterfactual, which can be used to determine what would have happened in the absence (presence) of aid, has not been settled. Using observational data, there is no way of establishing a plausible counterfactual without making assumptions that are bound to be debatable both in theory and practice. Nonetheless, methodological advances in the program evaluation literature have enhanced our understanding of causal attribution and suggest a range of appropriate empirical methods. Whilst these are never foolproof, their careful application provides a rigorous basis on which to address causal questions.

The objective of this chapter is to apply insights from the program evaluation literature to the age-old question of whether foreign aid promotes economic growth. In doing so we follow the recommendation of Temple (2010) to build explicitly on existing empirical work. Thus, our starting point for developing an appropriate empirical strategy is Rajan and Subramanian (2008) (henceforth RS08), which is a highly influential contribution that finds aid has no discernible impact on growth. On close analysis, however, a number of concerns question this conclusion and motivate methodological improvements. Once such modifications are incorporated, either individually or in conjunction, the empirical aid-growth relationship is shown to conform to the theoretical prior that foreign aid exerts a respectable long-run positive effect on growth in developing countries (see Chapter 2). Consequently we conclude there is no paradox between the micro- and macro-evidence for aid effectiveness (*pace* Mosley, 1987).

The rest of the chapter is structured as follows: Section 3.2 discusses the empirical strategy of RS08, including a brief summary of their approach and main results (§3.2.1), as well as a more detailed analysis

†This chapter combines and edits Arndt et al. (2010a) and Arndt et al. (2010b). It is co-authored with Channing Arndt and Finn Tarp.

of the validity of their instrumentation strategy (§3.2.2). This motivates various modifications, developed in Section 3.4. They consist of an improved instrument (§3.3.1), alterations to the specification (§3.3.2) and alternative regression estimators (§3.3.3). The results of applying these improvements are given in Section 3.4. Section 3.5 presents a battery of robustness and sensitivity tests. Section 3.6 concludes.

3.2 Empirical strategy of RS08

3.2.1 Description

As a first step, it is helpful to examine simple OLS estimates of the relationship between aid and growth. These are reported by RS08 and, using the same dataset, are replicated in column I of Table 3.1. As expected, they show a negative estimated coefficient on Aid/GDP. These estimates are expected to be biased downward principally due to a simultaneity problem. That is, precisely because of their weak growth, slower growing countries typically received comparatively more aid during the period (1970-2000). Thus, the challenge is to find valid and relevant external instruments that explain variation in aid receipts across developing countries, but which are unrelated to their growth performance.

Cognizant of the endogeneity of aid, RS08 investigate the aid-growth relationship using instrumental variables (IV) methods. Their preferred approach focuses on long-run averages, rather than dynamic panel methods. This is sensible. As discussed in Chapter 2, dynamic panel methods are subject to doubt given the expected cumulative effect of aid and corresponding concerns regarding the validity of internal instruments in GMM estimators. Their long-run approach echoes the average OLS estimator proposed by Mankiw et al. (1992) as well as the long-difference approach used by Acemoglu and Johnson (2007).

RS08 consider four periods separately: 1960-2000; 1970-2000; 1980-2000; and 1990-2000. In each period, their instrument for aid is generated from a "zero" 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 is then used as a single excluded instrument in a 2SLS estimation, where average growth over the period is the dependent variable.

Core results from RS08 for the 1970-2000 period are replicated in column II of Table 3.1. 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, and move on to show this holds for 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). The same basic result also emerges when the question is considered in a dynamic panel setting (RS08 Table 10).

Table 3.1: Alternative models for aid and growth, 1970-2000

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
	OLS	2SLS	LIML	LIML	LIML	LIML	LIML	LIML
Aid / GDP	-0.08*	0.10	0.10	0.07	0.08	0.12	0.21*	0.19*
	(0.04)	(0.07)	(0.08)	(0.06)	(0.06)	(0.10)	(0.11)	(0.10)
Initial per cap. GDP	-1.67***	-1.41***	-1.40***	-1.44***	-1.44***	-1.44***	-1.34***	-1.36***
	(0.32)	(0.43)	(0.39)	(0.37)	(0.36)	(0.30)	(0.33)	(0.33)
Initial level of policy	2.28***	2.14***	2.13***	2.16***	2.15***	2.28***	2.29***	2.28***
	(0.47)	(0.62)	(0.56)	(0.52)	(0.53)	(0.46)	(0.52)	(0.51)
Initial life expectancy	0.02	0.08*	0.08**	0.07**	0.07**	0.04	0.05	0.05
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Geography	0.39**	0.61**	0.62**	0.58***	0.58***	0.25	0.29	0.28
	(0.18)	(0.26)	(0.24)	(0.22)	(0.22)	(0.22)	(0.24)	(0.23)
Specification	RS08	RS08	RS08	RS08	AJT	AJT	AJT	AJT
No. of excluded instruments	1	1	7	3	1	7	3	1
Regional dummies	SSA, EA	SSA, EA	SSA, EA	SSA, EA	SSA, EA	SSA, A, LA	SSA, A, LA	SSA, A, LA
Obs.	78	78	78	78	78	78	78	78
R-squared	0.70	0.59	0.58	0.62	0.61	0.69	0.65	0.66
Weak identification stat.	-	31.60	5.50	10.40	33.41	4.32	5.06	14.32
Stock-Wright LM S stat.	-	-	11.85	3.05	2.12	18.17	5.40	5.20
(probability)			0.158	0.384	0.145	0.020	0.145	0.023
Hansen J stat.	-	-	7.72	0.29	-	11.88	0.31	-
(probability)			0.358	0.865		0.104	0.857	
Difference-in-Hansen C stat.	-	-	0.40	0.01	-	0.23	0.00	-
(probability)			0.529	0.936		0.628	0.980	

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

Notes: only selected variables reported; columns (I) and (II) replicate results in Rajan and Subramanian (2008); column (III) employs a full set of aggregate instruments (see Table 3.2) in place of a single generated instrument; column (IV) restricts the aggregate instrument set to the mean population ratio, the colony dummy and their interaction; column (V) uses only log. initial (recipient) population as the instrument; columns (VI) to (VIII) replicate columns (III) to (V) but use a preferred set of conditioning variables (not shown); regional dummies are included as indicated, SSA = Sub-Saharan Africa, EA = East Asia, A = Asia, LA = Latin America & Caribbean; weak identification statistic is the first stage partial-F statistic in col. (II), and the Kleibergen-Paap Wald F statistic elsewhere; initial policy refers to the Sachs-Warner trade policy index; geography refers to the average of the number of frost days and tropical land area; standard errors, given in parentheses, are robust to arbitrary heteroskedasticity; dependent variable is mean real growth rate; Aid/GDP is treated as endogenous in all models except column (I); in columns (I) to (V) Aid /GDP is taken from RS08, in columns (VI) to (VIII) it is re-estimated from OECD-DAC (2008) data treating possible missing values as zeroes.

Source: authors' estimates; see Appendix 3C.

3.2.2 Instrument validity

The supply-side approach to instrumentation developed by RS08 represents the state of the art in the aid-growth literature.[1] Nevertheless, it has been subject to criticism. Clemens and Bazzi (2009) note that different authors have used the same variables as exogenous instruments for a wide range of endogenous variables. This raises the possibility that these exogenous instruments are correlated with other omitted variables, thereby invalidating the exclusion restriction on which valid causal inference depends. They direct specific attention to the reliance of the RS08 (fitted) instrument on the natural logarithm of the aid recipient's population size. They find that log population has a "statistically significant partial relationship with several variables that are plausible growth determinants" (2009, p. 11) and that are omitted from RS08's specification.

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 of bias is, ultimately, 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, straightforward validity checks of the RS08 (generated) instrument based on Sargan or Hansen tests are not possible because their IV model is just identified – i.e., the number of excluded instruments equals the number of endogenous variables. Nonetheless, recalling that the zero stage of the RS08 approach generates a single instrument as a linear combination of variables, it is possible to use modified versions of these same variables as excluded instruments directly in the aggregate aid-growth regressions. This provides for a large number of potential instruments, and therefore permits over-identification tests to be run either on the full set or on specific subsets of instruments.

Following this logic, we collapse the bilateral aid dataset along the donor dimension, thereby transforming the explanatory variables used in the bilateral zero stage regressions for use at a more aggregate level. For continuous zero 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 the specific colonizer, it is more appropriate to take the maximum value of the dummy for a given recipient (again, across all donors). Ignoring relatively minor variables such as currently being a colony and the population-colony interaction terms employed in RS08, this yields a set of eight possible instruments as per the rows of Table 3.2.

Column I of Table 3.2 verifies whether these aggregate instruments are adequate proxies for the fitted instruments generated from the zero stage regressions. As expected, the explanatory power is high. Moreover, underlining the contention of Clemens and Bazzi (2009), a driving force behind the fitted aid instruments appears to be the population ratio term. 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 from column I of Table 3.2 indicate that other variables make some (albeit smaller) contribution to the overall fitted instrument. Thus, to further test instrument validity, we re-estimate the RS08 model employing the full set of eight aggregate instruments. The results, reported in column III of Table 3.1, closely replicate column II; and the Hansen J test reports a probability of 0.358, which fails to reject the validity of the exclusion restriction assumption.

Nevertheless, we take one further step to investigate the exclusion restriction. Following the intuition

[1] Note that in a recent contribution, Clemens et al. (2011) avoid the use of an external instrument for aid and rather address the endogeneity problem by lagging the aid term and differencing their specifying equation.

Table 3.2: 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	0.10**	0.01	-0.19	-0.25	1.11	0.29	1.96	0.16
Pop. ratio x 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	-	-	-	-

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

Notes: columns (I) and (II) report standardized OLS regression coefficients in which the dependent variable is the fitted aid instrument taken from Table 3.3 columns (I) and (V) respectively; columns (III) and (IV) report standardized OLS regression coefficients from regressions of residuals saved from columns (IV) and (VI) of Table 3.1 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; see Appendix 3C.

of Sargan-type tests, we save the residuals and regress them against the set of excluded instruments. This provides initial insight as to *which* variables in the instrument set may be suspect. Standardized coefficients from these regressions are given in column III of Table 3.2. 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, the Portuguese and French colonizer effects are significant. In the alternative AJT-specification shown in column IV of Table 3.2, the French and UK colonizer terms are both significant.

Following this simple but intuitive OLS approach, columns V to VIII of Table 3.2 report formal tests of the orthogonality of each of the individual aggregate instruments to the growth regression errors. Specifically, they 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. The most suspicion falls on the colonizer terms; however, the population ratio variables do not give cause for concern, providing comfort as to their suitability as exogenous instruments in these models.

3.3 Methodological improvements

3.3.1 Instrumentation strategy

The results of Section 3.2.2 indicate that the RS08 instrumentation approach is broadly convincing but is weakened by inclusion of suspect variables in the zero stage. From a theoretical point of view, the validity of using colonizer dummies and their interactions as instruments is questionable. The institutional transplants and broader colonizing strategies pursued by 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 (e.g. La Porta et al., 2008), historical events (e.g. Nunn, 2008, 2009), and other institutional forms on contemporary economic outcomes. Put simply, the colonial relations variables are not orthogonal to growth and therefore should not be included in the zero stage regression explaining aid.

As a first step towards improving the RS08 instrument, we re-run their aid-growth model using a smaller and 'less suspect' subset of the aggregate instruments used in Section 3.2.2. These are the population ratio, a dummy for ever-having been a colony and their interaction. Results are given in column IV of Table 3.1, showing that the Hansen J test is now passed with a high level of confidence. Nevertheless, compared against column II, the results also suggest a trade-off between efficiency and transparency in instrument selection. While the use of multiple aggregate instruments is more transparent, it does not exploit the full information about bilateral aid flows contained in the zero stage. This may be one reason why the weak identification statistics are considerably lower in column IV versus column II of Table 3.1. In fact, as shown in column V of Table 3.1, even if only one aggregate instrument is employed, namely the log of the recipient's initial population, the strength of the instrument returns to similar values to those in column II and all coefficients are essentially unchanged.[2] Consequently, using a single instrument is likely to be more efficient but there are also potential information gains from employing a zero stage, especially in small aggregate samples such as those used in (static) cross-country regressions.

Thus, to strengthen the instrumentation approach, we return to the zero stage regressions. Aside from removal of suspect terms, additional concerns motivate further modifications. First, there are errors in the calculation of average Aid/GDP in all stages of RS08's regressions. 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.[3] 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 variables and calculate period averages for Aid/GDP and aid per capita, setting missing entries to zero. This affects the dependent variable employed in the zero stage regression as well as the endogenous aid variable used in the IV estimations.

Second, in the RS08 strategy, recipient GDP occurs in the denominator of the dependent variable in the zero 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 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

[2] This result further underlines the reliance of the RS08 instrumentation strategy on population size (as per Clemens and Bazzi, 2009).

[3] Confirmed in correspondence with the OECD DAC Secretariat.

ratio, and/or if there is a direct association between recipient GDP levels and population size or past colonial experiences.

Third, it is apparent that individual donor countries exhibit distinct attitudes to giving foreign aid (Alesina and Dollar, 2000), which reflect cultural and historical factors. These time-invariant influences can be understood as fixed effects and may be included as RHS variables in the zero 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.

To address these concerns, we modify the RS08 specification of the zero stage regression. In place of Aid/GDP, we use aid per capita (Aid/POP) as the dependent variable which accords closely with the explicit aid allocation rules used by donors, such as the World Bank (see Annex 1 of IDA15, 2008).[4] We drop the colonizer-specific variables (and interactions) and only include a dummy for whether a country was ever a colony (COLONY) or is presently a colony (CURCOL). Finally, adding donor-specific fixed effects (DONOR), our zero stage regression emerges as follows:

$$\begin{aligned}\text{Aid}_{dr}/\text{POP}_r = \beta_0 + \beta_1\text{CURCOL}_{dr} + \beta_2\text{COLONY}_r + \beta_3\text{COMLANG}_{dr} + \beta_4\log(\text{POP}_d/\text{POP}_r) \\ + \beta_5\text{COLONY}_r \times \log(\text{POP}_d/\text{POP}_r) + \theta_d\text{DONOR}_d + \epsilon_{dr} \quad (3.1)\end{aligned}$$

where the subscripts d and r represent donors and recipients respectively.

Results from these modifications are given in Table 3.3. 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. This change has a moderate impact and the pair-wise correlation between the fitted values from these two models is 0.83. Column III also 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, but there is 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 (3.1). 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 (Tarp et al., 1999): (i) which recipients should receive aid; and (ii) how much to supply – i.e., 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 3.3 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. 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 zero stage regression.

[4] Note that in all subsequent regression stages the endogenous variable of interest remains Aid/GDP.

Table 3.3: Alternative zero 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	-

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

Notes: column (I) replicates Rajan and Subramanian’s zero 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 metropole (colony-specific) fixed effects & 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 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; see Appendix 3C

3.3.2 Specification

Before presenting the results of the aid-growth IV regressions using the improved instrument, it is appropriate to discuss additional areas where the RS08 approach can be strengthened. The first of these is the choice of covariates. Given the relatively small sample available in the aggregate regressions (78 countries), inclusion of redundant variables may lead to a loss of efficiency and/or contribute to undesirable multicollinearity. In the case of RS08, we note that the three macroeconomic initial conditions (inflation, money supply, and budget balance) as well as ethnic fractionalization are insignificant in RS08’s cross-section outcome regressions for all periods. In addition, and 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 (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 forced changes in the top government elite. 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 may also introduce unwanted reverse causality.

It is also 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'. 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. Finally, it is appropriate to include additional variables that reflect initial socio-economic conditions such as education and health indicators, as well as additional geographic characteristics such as 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 aid.

Consequently, we propose a revised covariates specification (denoted AJT). This involves dropping contemporaneous outcome covariates and redundant variables, adding an alternative set of regional dummies and including additional controls. These are selected following Sala-i-Martin et al. (2004) who undertake comprehensive Bayesian averaging of long-run growth estimates. We include variables identified by these authors that are among those with the highest posterior probability of inclusion and refer to initial conditions. To this, we add civil liberties in 1972 and distance to major ports. The first of these captures additional dimensions of initial institutional quality, including the ability of citizens to bring the government to account which is 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 (2012) as a robust correlate of growth.

3.3.3 Estimators

Another area that can be strengthened refers to the choice of IV estimator. In light of the expected complexity of the growth process, as well as the different properties of alternative estimators, it is useful to investigate whether or not empirical results hold across different estimators. While RS08 employ a 2SLS estimator, this is not the only option. Other suitable IV estimators, which offer moderate differences, include LIML (limited information maximum likelihood), Fuller's modified LIML (with alpha = 1) and the continuously updated GMM estimator (GMM-CUE).

Taking motivation from the program evaluation literature, the 'doubly robust' estimators of Robins and Rotnitzky (1995) have attractive properties – their use having been described as "best practice" by Imbens and Wooldridge (2009, p. 25). A range of such estimators have been proposed in the literature. One of the more straightforward of these, presented in Imbens (2004), 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. More formally, propensity scores can be derived from estimates on a logistic form:

$$\pi(x_i) = \Pr(W_i = 1 \mid X_i = x_i) = \exp(X_i'\lambda)/[1 + \exp(X_i'\lambda)] \tag{3.2}$$

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:

$$\min_{\delta,\zeta} \sum_{1=1}^{N} \frac{[Y_i - (\delta W_i + X_i'\zeta)]^2}{W_i\hat{\pi}_i + (1 - W_i)(1 - \hat{\pi}_i)} \tag{3.3}$$

where the estimate of δ represents the treatment effect of interest; and inference is made with robust standard errors. As long as one of either the propensity score or the linear regression is correctly specified, consistent estimates are generated.

Unfortunately, existing doubly robust estimators cannot be applied straightforwardly to the aid-growth problem. As evident above, they assume a binary treatment/control framework in which receipt of the treatment is conditionally independent of potential outcomes. That is, they presume selection into treatment is random, or at least a situation of 'selection on observables' applies. Both the endogeneity and continuous scale of foreign aid violate these assumptions. Even so, some simple extensions enable doubly robust estimators to be applied to the instrumental variables context. To do so, we begin with a standard 2SLS set-up containing an endogenous treatment variable (W), a single continuous instrumental variable (Z^*; presumed valid), an outcome of interest (Y), and a vector of additional controls (X), i.e.:

$$\begin{aligned} W_i &= \rho Z_i^* + X_i'\theta + u_i \\ Y_i &= \alpha W_i + X_i'\beta + v_i \end{aligned} \tag{3.4}$$

where u and v are correlated error terms. The critical step required to extend the IPWLS (inverse probability weighted least squares) 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 the 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 'larger' aid recipients. Assuming the parameter for α in equation (3.4) 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 (v_i) in the outcome equation of interest. Thus, it follows that:

$$\mathrm{E}[Z_i^* v_i \mid X_i] = \mathrm{E}[g(Z_i^*) v_i \mid X_i] = \mathrm{E}[Z_i v_i \mid X_i] = 0$$

Following this logic, equations (3.2) and (3.3) can be modified by replacing the treatment variable W with the binary instrument Z. Essentially, this is 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 first stage of equation (3.4) looks like $W_i = \rho Z_i + X_i'\theta + u_i$ when the dichotomous instrument is used in place of Z_i^*. In turn, this yields the modified reduced form of the problem:

$$Y_i = \alpha\rho Z_i + X_i'(\beta + \alpha\theta) + (v_i + \alpha u_i) \tag{3.5}$$

Ignoring weights, these coefficients are directly comparable to those in equation (3.1), where W_i has been replaced by Z_i. Specifically we now have: $\delta = \alpha\rho$; and $\zeta = (\beta + \alpha\theta)$. Of course, the reduced form is of interest *per se* because failure to find a relationship here indicate the absence of a significant treatment effect. Nevertheless, to compare against (unweighted) standard 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 (or other linear IV estimator), employing a dichotomous instrument and weights estimated from a propensity score procedure applied to the binary instrument.

The new estimator developed here, denoted henceforth as IPWLS, presents a useful addition to the range of (IV) estimators suitable to explore the relationship running from aid to growth. The principal advantage of this new estimator, versus standard IV estimators, is that it addresses (non-linear) differences in the distribution of the explanatory variables between the treatment and control groups. 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. Additionally, dichotomization of the instrument represents a useful robustness check. If results arising from the binary instrument are not comparable to those from its continuous counterpart, this might indicate that the latter findings are driven by peculiarities in the distribution of the instrument. It also relaxes the assumption of a constant linear relationship between treatment and outcomes (aid and growth), instead 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, such as diminishing returns to aid are addressed by the dichotomization. Finally, because the instrument is derived from a zero stage regression, dichotomization provides a check against measurement error or misspecification in the zero stage.

3.4 Empirical results

Section 3.2 motivated and proposed three main improvements to RS08's empirical approach. These refer to their instrument, their specification and their chosen estimator. Tables 3.4 and 3.5 present results for different combinations of these modifications. We focus on the 1970-2000 period, allowing the effects of aid to be considered over a generation of elapsed time. 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. 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.

Column I of Table 3.4 reports regression results using the RS08 specification, the new preferred instrument (based on the zero stage regression in column V of Table 3.3) and the LIML estimator. Column II introduces the doubly robust IPWLS estimator, while columns III and IV replicate columns I and II with our new specification. Columns V and VI continue with the same model, but respectively use the GMM-CUE and Fuller estimators. Thus, in all columns the modified instrument from Section 3.3.1 is employed.

Table 3.4: Regressions with modified aid instrument, 1970-2000

	(I) LIML	(II) IPWLS	(III) LIML	(IV) IPWLS	(V) GMM-CUE	(VI) Fuller
Aid / GDP	0.22*	0.21*	0.25**	0.13***	0.25**	0.24**
	(0.12)	(0.13)	(0.12)	(0.05)	(0.12)	(0.12)
Initial per cap. GDP	-1.34***	-1.92***	-1.03***	-1.33***	-1.03***	-1.05***
	(0.40)	(0.39)	(0.38)	(0.27)	(0.38)	(0.37)
Initial level of policy	2.14***	2.58***	2.12***	2.44***	2.12***	2.12***
	(0.60)	(0.62)	(0.54)	(0.46)	(0.54)	(0.53)
Initial life expectancy	0.09**	0.05	0.04	0.03	0.04	0.03
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Geography	0.63**	0.48**	0.29	0.25	0.29	0.29
	(0.25)	(0.24)	(0.26)	(0.21)	(0.26)	(0.25)
Coastal pop. density, ’65			0.00**	0.00***	0.00**	0.00**
			(0.00)	(0.00)	(0.00)	(0.00)
Primary schooling, 1960			2.58**	2.26**	2.58**	2.56**
			(1.15)	(0.88)	(1.15)	(1.13)
Malaria risk, 1966			-1.50*	-1.06*	-1.50*	-1.49*
			(0.85)	(0.58)	(0.85)	(0.83)
Invest. goods price, 1960-1964			-0.01	-0.01	-0.01	-0.01
			(0.00)	(0.00)	(0.00)	(0.00)
Civil liberties, 1972			-1.28*	-0.98*	-1.28*	-1.24*
			(0.70)	(0.50)	(0.70)	(0.68)
Air distance (log.)			0.09	-0.03	0.09	0.09
			(0.38)	(0.33)	(0.38)	(0.38)
Specification	RS08	RS08	AJT	AJT	AJT	AJT
Scale of excluded instrument	Continuous	Binary	Continuous	Binary	Continuous	Continuous
Regional dummies	SSA, EA	SSA, EA	SSA, A, LA	SSA, A, LA	SSA, A, LA	SSA, A, LA
R-squared	0.57	0.70	0.59	0.77	0.59	0.60
Kleibergen-Paap Wald F stat.	29.48	24.42	17.28	39.78	17.28	17.28
Stock-Wright LM S stat.	4.33	3.53	5.77	6.49	5.77	5.77
(probability)	0.037	0.060	0.016	0.011	0.016	0.016

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

Notes: dependent variable is average real growth rate; endogenous variable is Aid/GDP, re-estimated from OECD-DAC (2008) data treating possible missing values as zeroes; specification of columns (I) and (II) follows Rajan and Subramanian (2008) (selected covariates shown); remaining columns use a modified specification, removing contemporaneous and redundant covariates and adding additional initial conditions; estimator given in column title; standard errors in parentheses are robust to arbitrary heteroskedasticity; number of observations = 78.

Source: authors’ estimates; see Appendix 3C.

Table 3.5: Summary of results from model modifications, 1970-2000

Instrument	Specification	Estimator	
		2SLS/LIML	IPWLS
RS08	RS08	0.10	0.15*
	AJT	0.10	0.10**
AJT	RS08	0.22*	0.21*
	AJT	0.25**	0.13***

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

Notes: AJT refers to our preferred instrument (Table 3.3, Column V) and/or specification (Table 3.4 Column III). Cells show the estimated coefficient on Aid/GDP from IV regressions involving different combinations of instruments (rows), specifications (rows) and estimators (columns); standard errors on which inference is based are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Source: authors' estimates; see Appendix 3C.

Not that with respect to the implementation of the IPWLS estimator (columns III and IV), a binary instrument is required. This is derived by taking the fitted instrument from RS08's zero stage regression, sorting countries in ascending order (from lowest to highest predicted aid shares), and then selecting the first 30 for the 'control' and the rest for the 'treatment' group. 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. Thus, in practice, the control group approximately corresponds to all countries falling below the 40^{th} percentile.

Turning to results, the range of test statistics reported in Table 3.4 indicates that the new instrument continues to perform strongly across different specifications and estimators. Under-identification tests (not shown), which can be interpreted as testing the null hypothesis of a zero correlation between the instruments and the endogenous regressors, are all rejected. The weak identification test (the Kleibergen-Paap Wald F statistic, 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 comparable to the levels achieved using RS08's original approach (Table 3.1, column II). Perhaps more importantly, 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), finds a significant (partial) correlation between the instrument and dependent variable in all cases.

Moving across the columns of Table 3.4, we note that the treatment effect – i.e., the coefficient on the endogenous aid variable – is consistently positive, significant and in a domain that is consistent with the RS08 prior (see Chapter 2). The main effect of using the new and strengthened instrument (column I) is that the treatment effect estimate edges upwards (from 0.l0 to 0.22). The doubly-robust estimator leaves this result almost unchanged, but enhances the overall explanatory power of the model. According to the Kleibergen-Paap Wald F statistic, switching to the modified specification (column III onwards) slightly reduces the strength of the instrument in the LIML first stage. However, by placing greater emphasis on the most informative observations, the strength of the instrument is considerably improved for the IPWLS estimator. Finally, the alternative IV estimators (columns V and VI) are virtually identical to the results of column III.

The preferred estimate presented in Column IV of Table 3.4 represents a new estimator, a new specification, and a new instrumentation strategy.[5] To get a better sense of the individual and joint impact of alternative combinations of our three main modifications, Table 3.5 provides a summary of the various models. Each cell reports the estimated impact of aid on growth over the 1970-2000 period for some combination of our modifications; the base case of no modifications (the RS08 result) is also shown in the top left cell. Movements horizontally thus entail a shift from an unweighted to a weighted instrumental variables estimator (e.g., LIML to IPWLS), while movements vertically entail either a switch between specifications (e.g., row one to row two) or between instruments (e.g., row one to row three). This helps identify what is driving the differences. A first point to note is that, with the exception of employing the AJT specification alone, all other modifications (taken either individually or jointly) yield a significant aid-growth relationship. Thus, our rejection of the RS08 result is not dependent on taking all modifications together; nor is it driven by any one modification alone. However, the probability of falsely rejecting the null hypothesis falls with the number of modifications employed. Thus, when all three modifications are incorporated, the estimated coefficient on Aid/GDP of 0.13 becomes significant at the 1% level. Second, the size of the estimated coefficient of interest tends to increase when the new instrument is employed. However, this effect appears to be modified (reduced) when the IPWLS estimator is also used. Thus, there is a somewhat complex pattern of interactions between the modifications. Nonetheless, broadly speaking, our modification of the generated instrument and introduction of a new estimator appear to be most crucial to our results.

Given our preferred approach employs a single generated instrument, it remains to be established whether the underlying instrumentation strategy remains valid in the context of the new set of conditioning variables. Thus, maintaining the improved specification and LIML estimator, we replace the generated instrument with different sets of aggregate instruments as per Sections 3.2.2 and 3.3.1. These results are reported in columns VI to VIII of Table 3.1, which respectively employ a full set of aggregate instruments (used in the RS08 zero stage), a preferred sub-set of three 'least suspect' instruments, and initial log population size only. The most important result is that the Hansen J statistic is considerably strengthened when three as opposed to eight aggregate instruments are used, passing all conventional test thresholds. Moreover, in the last two columns, the coefficient on Aid/GDP remains positive and significant, once again supporting our principal results.

At this point, it is helpful to reflect on what exactly the estimated regression parameters represent. As has been established in the literature, instrumental variables estimators typically cannot be interpreted as average treatment effects. This is only appropriate under strong additional assumptions – in this case, homogeneity across countries in their response to aid. Rather, what is actually recovered depends on the instruments chosen as well as the extent of heterogeneity in responses to changes in these instruments. In the case of a single binary instrument and endogenous response variable, IV estimators often recover a local average treatment effect (LATE), defined as the average treatment effect for the subpopulation that switches from the control to the treatment group on account of the switch in the instrument. This can be understood as a weighted average of the marginal treatment effects for the sub-population that is

[5]Note that adding the three initial macroeconomic conditions employed by RS08, which had been excluded for redundancy, to the models estimated in columns III and IV of Table 3.4 leaves all results essentially unchanged; moreover, these three variables continue to be redundant.

responsive to the instrument (Heckman, 2001).

In the present case, the complex mapping from multiple instruments to a continuous endogenous variable significantly complicates a LATE-type interpretation of the estimates. Furthermore, given the small sample sizes involved, distinguishing between key sub-populations and important observations is exceedingly difficult operationally. Nevertheless, in light of the instruments used, a possible interpretation is that the estimates reflect the growth response to aid for countries whose total aid inflows have been most influenced by differences in relative population sizes or ever having been a colony. These drivers typically would be associated with political rather than altruistic motivations. If politically motivated aid is less effective, this may bias downward the estimated aid coefficient. Finally, one notes that moving from the continuous instrument to the binary instrument (used in the doubly robust estimations) has a relatively small effect on the estimated parameter of interest.

Lastly, we consider how this evidence stacks up against theory. In all the regression specifications reported so far, tests have been made against a null hypothesis of a zero relationship between the explanatory and dependent variables. Nevertheless, it is straightforward to calculate the t-statistic for testing whether the estimates differ from the theoretical prior that the point estimate for the long-run effect of foreign aid on growth is around 0.1 (see Chapter 2). With the exception of the OLS regression reported in column I of Table 3.1, all point estimates for the aid growth relationship across all regressions reported in Tables 3.1, 3.4 and 3.5 are not significantly different from 0.1.[6] To put it differently, there is no basis on which to reject the theoretical prior that aid has a positive long-run effect on growth.

3.5 Sensitivity tests

The estimates presented in Column IV of Table 3.4 represent a new estimator, a new specification, and a new instrumentation strategy. The previous section 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 IPWLS results we employ a more flexible doubly robust estimator. This relaxes the assumption that the coefficients on the covariates are the same for treatment and control groups. Following Imbens and Wooldridge (2009), one can estimate versions of equation (3.3) separately for the treatment and control groups, this time with the covariates stated as deviations from overall sample means, such that the treatment effect is given by the difference in the estimated intercept terms.[7] Appendix Table 3A.1 summarizes the results from this estimator (denoted FDR) for different combinations of instruments and specifications. These are compared to the estimates from standard (unweighted) LIML estimators summarised in Table 3.5. 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. For both

[6]For these regressions, the null hypothesis that the coefficient of the estimated aid-growth parameter is equal to 0.1 is, in fact, never close to being rejected, including in RS08's original IV specification (Table 3.1, Column II).

[7]Evidently, in estimating equation (3.3) 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.

estimators, the vast mass of this interval lies in the positive domain. Moreover, the point estimates in the bottom right cell are highly consistent with the preferred results in Column IV of Table 3.4.

Both the IPWLS and the 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, Appendix Figure 3A.1 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).[8] Panel (a) refers to the RS08 specification and instrument (from which weights are then derived) replicated in column II of Table 3.1; panel (b) uses the specification and instrument from column IV of Table 3.4, 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 3A.2 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 $12 \times 78 = 936$ 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 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 3A.3. 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

[8] Appendix 3B 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 3A.3. They reinforce the conclusion that although the 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.

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 Appendix Table 3A.3 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.[9] 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 3.3), 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 LIML and IPWLS estimators, results for each alternative period are presented in Appendix Table 3A.4. 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 3.4 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 in Chapter 2, a meaningful and robust average short run effect of aid on growth may be very difficult to discern from the available empirical data.

3.6 Conclusion

To conclude, we respond to the question posed in the title to this chapter: has the aid, growth and development literature gone full circle? Our answer is "no". While in the most recent literature the

[9] 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.

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 program 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. Based on a detailed analysis of their approach, we subsequently developed a better instrumentation strategy, an improved specification and a preferred estimator. The improved 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 program evaluation literature, we excluded covariates, such as revolutions and institutional performance, which represent potential channels through which aid affects growth. With respect to the zero stage instrumentation, we (i) excluded suspect variables; (ii) corrected errors in the implementation of the RS08 instrumentation strategy; (iii) employed aid per capita in place of Aid/GDP to preclude spurious correlation with the chosen instruments; (iv) introduced donor-specific fixed effects; and (v) 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.

The results of this chapter provide solid support for the view that the effect of aid on growth is positive over the long-run, confirmed here in both the 1970-2000 and 1960-2000 periods. The preferred doubly robust estimator places the point estimate of the long run semi-elasticity of income with respect to the share of aid in recipient GDP at 0.13 (IPWLS). This suggests that an aid inflow on the order of 10 percent of GDP spurs the 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% confidence interval around our 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, when the longer run macro evidence is combined with the evidence at the micro- and meso-levels, a consistent case for aid effectiveness emerges. There is no micro-macro paradox.

Two specific caveats to the analysis can be highlighted. First, we focus exclusively on aggregate measures of aid, which are likely to mask substantial heterogeneity in terms of its quality and underlying objectives. As noted by Clemens et al. (2011) (also Roodman, 2007), there is no reason to expect that all aid will contribute to economic growth either *per se* or over the same period of time. Our focus on effects over a long horizon (1970-2000) goes some way to address this critique, but it does not deal with (unobserved) differences in aid quality. Second, our external instrument for aid cannot be validated exhaustively and therefore will remain open to criticism. This underlines the importance of using different methods and research strategies to explore aid effectiveness. As a consequence, we find ourselves in a similar position to Winters (2004) in his review of the implications of trade liberalization for growth. While he concludes that trade liberalization stimulates growth over the long-run and on average, he adds that: “For a variety of reasons, the level of proof remains a little less than one might wish but the preponderance of evidence certainly favours that conclusion.” (2004, p. 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, 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.

3A Appendix: additional tables and figures

Table 3A.1: Summary of Flexible Doubly Robust (FDR) results, 1970-2000

Instrument	Specification	Estimator 2SLS/LIML	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 (from Table 3.3 Column V) and specification (see Table 3.4 Column III). Cells show 90 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 the middle; column (I) provides estimates from standard IV estimators (2SLS or LIML); column (II) employs the flexible doubly robust estimator (FDR) as described in the text; estimates in column (I) come directly from results in Tables 3.1 and 3.3; standard errors used to calculate confidence intervals are robust to arbitrary heteroskedasticity; dependent variable is the average real growth rate.

Source: authors' estimates; see Appendix 3C.

Table 3A.2: Summary of sequential exclusion procedure, 1970-2000

Specification →		RS08		AJT	
	Instrument →	RS08	AJT	RS08	AJT
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
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
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 the modified instrument and/or specification (c.f., Table 3.5); '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; see Appendix 3C.

Table 3A.3: Robustness and sensitivity tests, 1970-2000

	(I) IPWLS	(II) IPWLS	(III) IPWLS	(IV) LIML	(V) LIML	(VI) LIML
Aid measure	0.15	0.15***	0.12**	0.24**	0.42**	0.03**
	(0.24)	(0.04)	(0.05)	(0.12)	(0.19)	(0.01)
Initial per cap. 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

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

Notes: unless otherwise indicated, the endogenous variable is Aid/GDP, measured as per the models in Table 3.4; 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; see Appendix 3C.

Table 3A.4: Aid-growth regressions, alternative periods

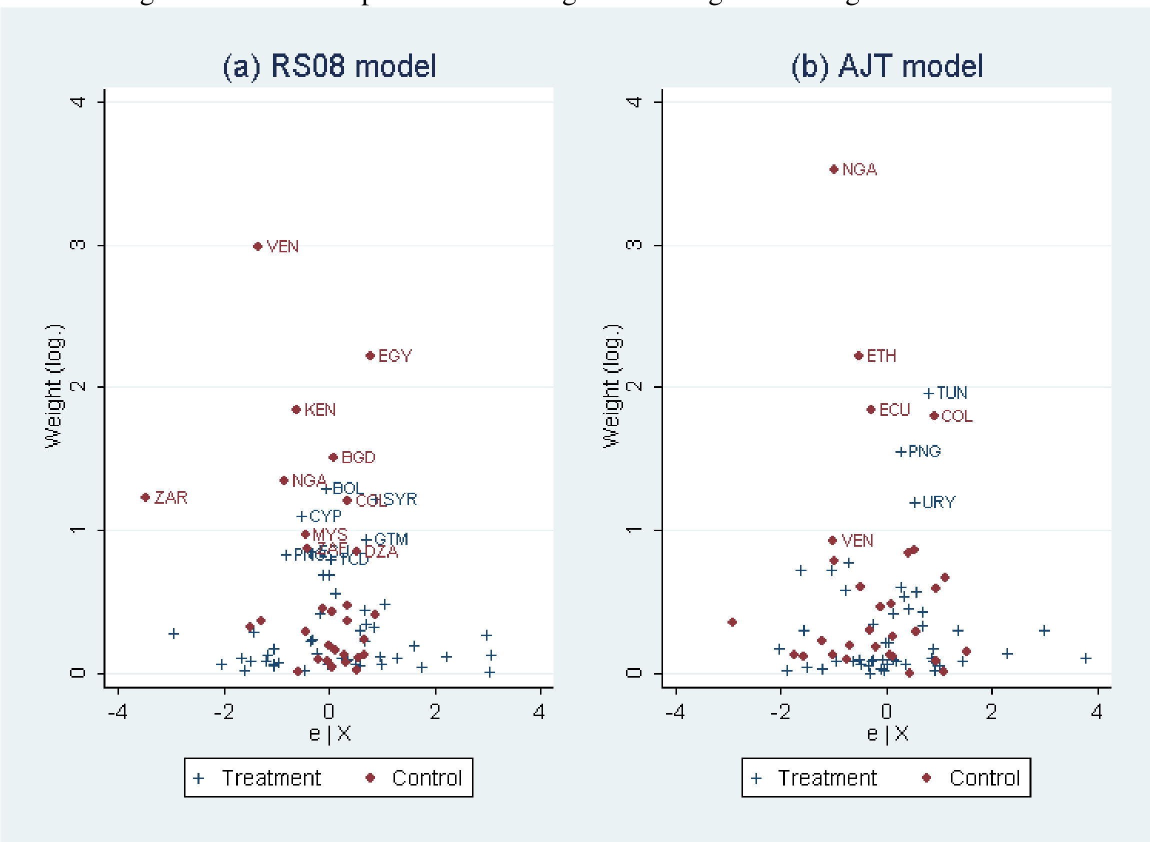
	1960-2000		1980-2000		1990-2000	
	(I) LIML	(II) IPWLS	(III) LIML	(IV) IPWLS	(V) LIML	(VI) IPWLS
Aid / GDP	0.16*	0.09**	0.02	0.05	-0.11	0.11
	(0.08)	(0.04)	(0.14)	(0.10)	(0.19)	(0.14)
Initial per cap. GDP	-0.67**	-0.83***	-1.41***	-1.36***	-0.72	-0.12
	(0.31)	(0.24)	(0.42)	(0.35)	(0.74)	(0.57)
Initial 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 in 1965	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 in 1960	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 in 1966	-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 1960-64	-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

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

Notes: the endogenous variable is Aid/GDP, measured as per the models in Table 3.4; 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; see Appendix 3C.

Figure 3A.1: Scatter plot of IPWLS regression weights versus growth residuals



Notes: panel (a) refers to the model summarised by the top right cell of Table 3.5; panel (b) refers to the model summarised by the bottom right cell of Table 3.5; y-axis plots the log. of (inverse propensity score) estimated weights from these models (transformed by natural logarithms for clarity); x-axis plots the residual from an OLS regression of the growth rate against core control variables (excluding Aid/GDP).

Source: authors' estimates; see Appendix 3C.

3B Appendix: summary statistics by country

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 Republic	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	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

Country	Growth rate	Endogenous Aid/GDP		Fitted aid measure		Estimated weight	
		RS08	AJT	RS08	AJT	RS08	AJT
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
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 Republic	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

3C Appendix: data sources and definitions

The base data, including explanatory variables used in the preliminary stage regressions, is from Rajan and Subramanian (2008) – kindly supplied by the authors. Other variables and their respective sources are described below.

Table 3C.1: Variables taken from Rajan and Subramanian (2008)

Variable	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	Average of five different indices of ethno-linguistic fractionalization which is the probability of two random people in a country not speaking the same language.

Table 3C.2: Variables taken from Sala-i-Martin et al. (2004)

Variable	Description
Coastal population density	Density of population 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.

Note: all variables represent ‘fixed’ initial conditions

Table 3C.3: Variables used to (re)calculate the aid instrument

Variable	Description	Source
Aid	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 www.oecd.org/dac/stats (accessed 2009)
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)

Chapter 4

Unpacking the effectiveness of foreign aid over the long run†

4.1 Introduction

Significant volumes of foreign aid have been channelled to developing countries for more than four decades; nevertheless, an answer to the question "What has aid accomplished?" remains complex, controversial and incomplete. This is not for lack of effort. A large literature considers aid effectiveness particularly from the perspective of the impact of aid on aggregate economic growth. A recent and influential contribution by Rajan and Subramanian (2008) finds little evidence that aid has contributed to economic growth. This was disputed in Chapter 3, which finds a positive and significant effect of aid on aggregate economic growth at levels consistent with simple growth theory.

Building on the analysis of the previous chapter, the present chapter seeks to provide a broader assessment of aid effectiveness, also from the perspective of development economics. We consider a range of development outcomes potentially affected by aid as well as the transmission channels from key intermediate outcomes to economic growth. This is important for at least two reasons. First, many outcomes are valued independently of their contribution to growth. Access to 'merit goods', such as basic health care and primary education are often considered to be core human rights. Accordingly, these outcomes are fundamental elements when considering the accomplishments of aid. Second, quantifying transmission channels from key intermediate outcomes to growth provides a coherence test for the aid-growth relationship. If no robust evidence of a relationship can be found between aid and important growth determinants such as investment and human capital, then the impact of foreign aid on growth is likely to be negligible, or at least becomes much harder to explain.

Like many empirical questions in the economics literatures, studying aid effectiveness is beset by difficulties in determining causality. In order to address these challenges, we follow the Structural Causal Model (SCM) due to Pearl (2009). This has been employed in numerous fields but is relatively new to the economics profession. We set out a general SCM, consistent with the empirical growth literature, define the target effects of interest, and suggest how these effects can be identified from observational data.

†This chapter is a substantially revised version of Arndt et al. (2011). It is co-authored with Channing Arndt and Finn Tarp.

Using the SCM, we also show that the transmission effects of interest are not likely to be identifiable even if aid were randomized. The empirical analysis is then pursued in three steps: we (i) calculate reduced form estimates of the impact of aid on a range of final economic outcomes (growth, poverty, inequality and structural change); (ii) apply the same reduced form approach to a set of intermediate economic outcomes (such as investment, consumption and tax take) and a range of social outcomes (such as health and education); and (iii) quantify a simplified representation of the full structural form, where aid impacts economic growth through a key subset of outcomes identified in step two.

In performing this analysis, the chapter contributes in three areas. First, although a number of previous studies have investigated the effect of aid on outcomes other than growth, this is the first to formally link effects at the meso- and macro-levels using a systematic methodology. This responds, at least in part, to the challenge set forth by Bourguignon and Sundberg (2007) to unpack the causal chain from aid to final outcomes. Second, practical application of the SCM framework in the context of a long-run cross-section approach and relying on external supply-side instruments for aid represent important distinctions and contributions of the present exercise. Third, we respond to the question posed at the outset. We find no evidence that nearly 40 years of development assistance has had an overall detrimental effect on development outcomes. Rather, a coherent and favourable picture emerges. Aid has promoted structural change, reduced poverty, and stimulated growth. Growth effects are attributable to the channels of health and physical capital investments. We also find that aid has had a positive effect on education outcomes. These findings are consistent with significant strands of the existing literature and add further weight to the conclusion that, while perhaps less potent than hoped, aid has registered significant accomplishments in helping to achieve development goals.

The rest of this chapter is structured as follows: Section 4.2 sets out the methodology and data used to guide causal inference; Section 4.3 presents the results; Section 4.4 discusses their economic significance and their relationship to the extant literature; and Section 4.5 details our conclusions.

4.2 Methodology

4.2.1 Causal framework

A variety of approaches have been developed, often in separate literatures, to address questions of causality, which are at the core of assessing aid's impact. For example, the potential outcomes approach associated with the contributions of Donald Rubin (e.g., Rubin, 1974), is rooted in concepts of (as if) randomization. This is useful in certain situations, particularly when an experimental element determines units' treatment status. However, in observational settings, the potential outcomes approach does not provide clear guidance as to channels of impact and/or exactly what effects will be estimated under different sets of conditioning variables (see Heckman, 2008; Pearl, 2010). To address causality in a clear and rigorous fashion, we adopt the SCM (Pearl, 2009). Two aspects of this approach deserve mention. First, it presumes the causal relations of interest can be depicted in directed acyclic graph (DAG), in which variables are connected by a network of directed edges (paths). This means that the joint probability distribution of the network can be factorised into the product of conditional probability distributions.[1] In turn, the correlation between any two variables, X and Y, can be stated as a function of individual

[1]The factorization is given by: $P(x_1, ..., x_n) = \prod_i P(x_i \mid PA_i)$ where the set PA_i refers to the Markovian parents (direct antecedent variables) of x_i in the network (Pearl, 2009).

effects (parameters or coefficients) that constitute the paths connecting X and Y. Following Brito and Pearl (2006), this is given as:

$$\rho_{X,Y} = \sum_{\text{paths } p_j} T(p_j) \tag{4.1}$$

where $T(p_j)$ represents the product of all parameters along path p_j, and the summation ranges over all unblocked paths (see below) between X and Y. Thus, different causal effects, including conditional/unconditional and direct/indirect effects, can be decomposed as functions of path coefficients.

The second aspect of the SCM is a graphical condition for causal identification in the presence of unmeasured confounders or correlated errors, known as the 'backdoor criterion'. This points to an admissible set of variables which, if adjusted upon (e.g., by stratification or conditioning), removes all sources of potential bias in the estimated causal effect. With respect to the causal effect of X on Y, this set is defined by (Pearl 2010: 17-18), as:

Definition 1. *A set S is admissible or 'sufficient' for adjustment if: (i) no element of S is a descendant of X; and (ii) the elements of S 'block' all paths that end with an arrow pointing to X – i.e., all "back-door" paths from X to Y are blocked.*

and where the notion of blocking is defined in (Pearl 2010: 8), as:

Definition 2. *A set S of nodes is said to block a path p if: either (i) p contains at least one arrow-emitting node that is in S; or (ii) p contains at least one collision node that is outside S and has no descendant in S.*

In the SCM framework, the causal effect of X on Y can be estimated if an admissible set is observed such that all backdoor paths from X to Y can be blocked by adjusting on this set.[2]

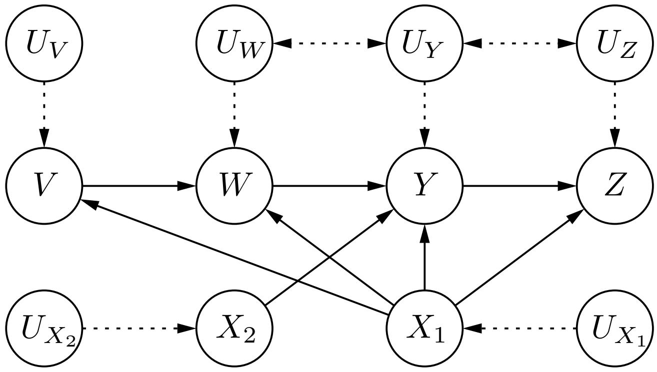
4.2.2 Application to foreign aid

Application of the SCM to assess aid effectiveness is useful for two main reasons. First, given the present interest in identifying aid's transmission channels, it provides a transparent method that puts a clear focus on the underlying causal mechanisms and the range of (empirical) assumptions required to identify individual channels of effect. This is particularly helpful to untangle the role and/or endogeneity of intermediate variables and their interactions in the system. Indeed, the method reveals that even if aid were to be applied randomly across countries, its transmission channels would remain unidentified. Second, due to its graphical approach, the method also helps pinpoint potential falsification tests for the hypotheses of interest. Nonetheless, it should be noted that the appropriate empirical methods to which the SCM points in this instance are relatively straightforward instrumental variables estimators.

In applying the SCM to assess aid effectiveness, it is necessary to encompass sets of both final and intermediate outcomes. In what follows, the former are defined by the set $\mathbf{Z}$, with elements $Z_1, Z_2, ..., Z_N$. In similar fashion, $\mathbf{Y}$ denotes the set of intermediate outcomes with elements Y_m and its subset $\mathbf{Y^{m'}}$,

[2] This is a special case of the general theorem due to Tian and Pearl (2002). It holds that a sufficient condition for identifying the causal effect of X on Y is that every path in a DAG between X and any of its children traces at least one arrow emanating from a measured variable.

Figure 4.1: General causal diagram summarising the linkages between aid and final outcomes



Notes: this figure is a causal directed acyclic graph (DAG) of the relationship between aid (W) and aggregate outcomes (Z), via intermediate outcomes (Y); V is an exogenous determinant of aid; U terms are unobserved, possibly errors; solid lines represent directed relationships between observed variables; broken lines represent directed relations due to unobserved variables (errors).

which includes all elements of $\mathbf{Y}$ excluding the single element Y_m; i.e., $\mathbf{Y^{m'}} \subset \mathbf{Y} = \{y : y \in \mathbf{Y}, y \neq Y_m\}$.

With this notation, Figure 4.1 sets out the presumed DAG, representing a primal model of the aid-final outcome relationship. Solid lines represent directed relationships between observed variables, depicted by the nodes (circles); broken lines represent effects emanating from unobserved variables, which can be thought of as error terms. We also presume there exists a variable V, which is a parent (ancestor node) of aid (W) and has an error structure that is unrelated to the error structure of any other variables, indicated by the absence of arcs to (unobserved) error terms.

The next step is to explicate the effects to be estimated. The aggregate aid-growth debate focuses on a single target effect – the causal effect of aid on GDP growth. Employing the notation of Pearl (2009), this can be expressed as $\partial/\partial w E[G \mid do(w), \mathbf{x}]$, where G is the outcome of interest, and $\mathbf{X}' = [\mathbf{X_1} \quad \mathbf{X_2}]$ is a set of exogenous background variables. The $do(\cdot)$ operator is used to denote the effect of a physical intervention or manipulation. This is conceptually distinct from statistical conditioning, such as stratification or inclusion on the RHS of a regression, but may be empirically equivalent under certain conditions (Pearl, 2010).

Taking guidance from Figure 4.1, we can distinguish between four relevant types of effects for chosen elements of $\mathbf{Y}$ and $\mathbf{Z}$:

(E1)	Total effect of W on Z_n	=	$\partial/\partial w E[Z_n \mid do(w), \mathbf{x}]$
(E2)	Total effect of W on Y_m	=	$\partial/\partial w E[Y_m \mid do(w), \mathbf{x}]$
(E3)	Direct effect of W on Y_m	=	$\partial/\partial w E[Y_m \mid do(w), do(\mathbf{Y^{m'}}), \mathbf{x}]$
(E4)	Total effect of Y_m on Z_n	=	$\partial/\partial Y_m E[Z_n \mid do(Y_m), \mathbf{x}]$

Effects E1 and E2 are straightforward replications of the aid-growth effect, employing alternative final or intermediate outcomes. These effects are also 'total' as they capture all open paths running from W

to the chosen outcome, regardless of the mediating variables involved. While of general interest, such estimates are not sufficient to quantify the individual effect of a chosen intermediate variable on a given final outcome, which is required to decompose aid effectiveness into its constituent channels. Rather, following the path decomposition property of DAGs, we first need an estimate of the direct effect of aid on the chosen intermediate outcome, excluding effects transmitted via other intermediate variables that causally precede the chosen Y_m. This is given by effect E3, which includes the additional term $do(\mathbf{Y^{m'}})$. Next, we require the total effect of Y_m on Z_n, given by effect E4. The product of E3 and E4 yields the decomposition of interest.

The DAGs have a corresponding (non-parametric) structural equation representation. Assuming the functions are autonomous, Figure 4.1 corresponds to equations (4.2) to (4.5) below. These are general expressions from which specifications used in the empirical growth regression literature can be derived as special cases (e.g., using additive errors). For instance, by equation (4.5), aggregate GDP growth ($G \in \mathbf{Z}$) can be represented by the function: $g = f_G(\mathbf{x}, \mathbf{y}, u_G)$. Ignoring the error term, this maps directly to the standard equation Mankiw et al. (1992), which defines growth as a function of initial and steady state income Mankiw et al. (1992). As a result, $\mathbf{X}$ contains initial income as well as various fixed factors that affect long-run productivity, while $\mathbf{Y}$ contains proximate time-varying factors, such as the rate of accumulation of human and physical capital that also affect steady state income.

$$v = f_V(\mathbf{x}, u_V) \tag{4.2}$$
$$w = f_W(v, \mathbf{x}, u_W) \tag{4.3}$$
$$y_m = f_Y(w, \mathbf{x}, u_{Y_m}) \tag{4.4}$$
$$z_n = f_Z(\mathbf{y}, \mathbf{x}, u_{Z_n}) \tag{4.5}$$
$$E[u_V u_j] = 0 \; \forall \; j \in J = \{W, Y_1, Z_1, ..., Y_M, Z_N\}$$
$$E[u_j u_k] \neq 0 \; \forall \; j, k \in J.$$

4.2.3 Identification

Effects E1 and E2

The DAG of Figure 4.1 provides a sufficient and general basis to determine whether the previously defined target effects can be identified. The immediate issue is that neither E1 nor E2 can be identified directly. Following Section 4.2.1, we note the existence of backdoor paths such as $W \dashleftarrow U_W \dashrightarrow U_Y \dashrightarrow U_Z \dashrightarrow Z_n$, which could reflect omitted variables bias, or some form of simultaneity. As these paths contain unobserved variables, the set of observed controls ($\mathbf{X}$) is not sufficient for adjustment. This means that in the absence of randomization of aid, simple estimates of either E1 or E2 (controlling for $\mathbf{X}$) will be biased – i.e., aid must be considered endogenous.

Nonetheless, given the assumptions encoded in the DAG, we see that $\mathbf{X}$ *is* admissible to identify the causal effect of V on any one of W, Y_m or Z_n; that is $U_V \perp\!\!\!\perp U_W, U_Y, U_Z \mid \mathbf{X}$. Again employing the path decomposition property of the DAG (see equation 4.1), effects E1 and E2 thus can be recovered indirectly as the ratio of causal effects due to V. However, as noted by Balke and Pearl (1997), this is only feasible with the additional assumption that the underlying functional forms are linear, such that error terms are additive. This can be seen algebraically by taking the reduced form associated with equations

(4.4) and (4.5):

$$
\begin{aligned}
z_n &= f_z(\mathbf{y}, \mathbf{x}, u_Z) \\
&= f_Z([f_Y(w, \mathbf{x}, u_{Y_1}), ..., f_Y(w, \mathbf{x}, u_{Y_M})], \mathbf{x}, u_Z) \\
&= \sum_{m=1}^{M} \alpha[\beta_{1m} w + \mathbf{x}'\beta_{2m} + u_{Y_m}] + \mathbf{x}'\gamma + u_Z \\
&= \tilde{\lambda} w + \mathbf{x}'\tilde{\mu} + \tilde{u} \qquad (4.6)
\end{aligned}
$$

where the tilde superscripts denote aggregated parameters. Multiplying (4.6) through by V, taking expectations and rearranging, yields an instrumental variables estimand: E1 $= Cov(V, Z_n \mid X)/Cov(V, W \mid X)$. Effect E2 can be estimated analogously: E2 $= Cov(V, Y_m \mid X)/Cov(V, W \mid X)$.

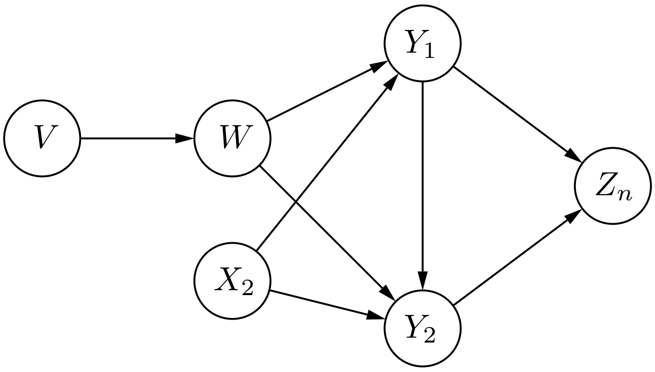
Effects E3 and E4

In contrast to E1 and E2, neither E3 nor E4 are easily identifiable even if aid were to be randomized. This is driven by the (likely) existence of interaction between intermediate variables, illustrated in Figure 4.2 for the case of two intermediaries and one final outcome. As can be seen from the definitions in Section 4.2.2, effect E3 is a restriction on E2, implying the former can be estimated by blocking (closing) paths running through other intermediate variables. Importantly, this cannot always be achieved by conditioning on the set $\mathbf{Y^{m'}}$. For instance, to estimate E3(Y_2) it is appropriate to include Y_1 in the conditioning set thereby closing the 'frontdoor' path $W \rightarrow Y_1 \rightarrow Y_2$. In contrast, it is incorrect to condition on Y_2 if we wish to estimate E3(Y_1). This is not only because Y_2 is a collider ($W \rightarrow Y_2 \leftarrow Y_1$), making it redundant, but principally because the act of conditioning on a collider can introduce additional bias; that is, it violates the backdoor criterion (Definition 2, condition (ii); Cole et al., 2010). The implication is that we must either be able to externally manipulate (e.g., randomize) all intermediate variables in the set $\mathbf{Y^{m'}}$, thereby severing their connection to aid; or, the structure of interaction between the elements of $\mathbf{Y}$ must be known, allowing an admissible set to be selected.[3]

Recovery of effect E4 faces similar difficulties. As indicated by Figure 4.1, correlation between the error terms (U_Y, U_Z) means that estimation must proceed indirectly as with effects E1 and E2. However, use of V (or randomized aid) as a suitable instrument is problematic. This can be seen from Figure 4.2, where the total effect of Y_1 on Z_n is given by the set of frontdoor paths $\{Y_1 \rightarrow Z_n, Y_1 \rightarrow Y_2 \rightarrow Z_n\}$. To employ V as a valid instrument for Y_1, we must include Y_2 in our adjustment set – i.e., it is necessary to control for all other intermediate variables for which there exists a frontdoor path connecting V and elements of $\mathbf{Y^{m'}}$. Following Definition 2, this blocks the causal path $V \rightarrow W \rightarrow Y_1 \rightarrow Y_2 \rightarrow Z_n$, meaning that only the direct effect of Y_1 on Z_n would be identified. According to the figure, and aside from externally manipulating $\mathbf{Y}$, an appropriate solution would be to use $\mathbf{X_2}$ as a set of exogenous instruments for Y_1 and Y_2. This introduces two additional requirements: (i) order conditions must be met – i.e., we need at least as many elements in $\mathbf{X_2}$ (instruments) as there are elements of $\mathbf{Y}$, and (ii) unless all elements of $\mathbf{X_2}$ are pairwise orthogonal (and relevant), estimation of effect E4 for multiple intermediate variables must be undertaken in one equation, rather than separately. In the case of Figure 4.2, a single instrumental variables (IV) regression equation, in which both Y_1 and Y_2 are treated as endogenous conditional on $\mathbf{X_1}$, would yield consistent estimates of E4 for these two intermediate variables.

[3] Of course, as the set $\mathbf{Y}$ refers to intermediate outcome variables (e.g., infant mortality), it is difficult to conceive how such outcomes could be randomised in any feasible or ethical way.

Figure 4.2: Simplified causal diagram summarising the linkages between aid and a single final outcome



Notes: this figure is a simplified version of Figure 4.1 including a presumed interaction structure between intermediate variables; X_1 and error terms (U) follow Figure 4.1 but are omitted for clarity.

Finally, Figures 4.1 and 4.2 suggest a falsification test. Given the Markovian properties of the assumed DAG (see Sections 4.2.1 and 4.2.2), it follows that $V \perp\!\!\!\perp Z \mid Y, X_1$. In practical terms, we should expect the instrument for aid to be redundant if included as a control variable in estimates of effect E4. If not, the model is likely to be misspecified, which might arise if material elements of $\mathbf{Y}$ have been neglected.

4.2.4 Estimation strategy

The previous sub-sections suggest that estimation should proceed in two steps. The first is to estimate the reduced forms E1 and E2 for a relatively wide range of variables. While interesting in itself, this is also a form of exploratory analysis, which informs the second step – estimating effects E3 and E4. Here, due to the need to identify multiple endogenous variables, a more manageable subset of $\mathbf{Y}$ and $\mathbf{Z}$ is chosen, constituting a simplified causal system. The latter step proceeds by selecting additional instrumental variables for the chosen intermediate variables, and applying both single-equation and systems IV estimators.

Choice of relevant and valid instruments in both stages is far from trivial. Regarding the aid instrument, given the set of conditioning variables is fixed across different outcomes, the relation between the instrument and aid remains the same (as per a first stage regression) – i.e., instrument strength is unchanged. Nonetheless, because the instrument is derived from observational data, and the 'true' set of exogenous background variables (such as initial conditions) that block all paths between the instrument and outcomes is unknown, there is no guarantee that the instrument is equally valid in all cases. Put differently, as the instrument is not randomized there may be some outcome for which U_V is not independent of U_Y or U_Z. Consequently, a metric of instrument validity would be valuable. Foremost, however, is the problem of instrument selection for which the methods developed in Chapter 3 (c.f., Arndt et al., 2010a, denoted hereafter as AJT10) are pertinent and, thus, provide the point of departure. Specifically, AJT10 generate an external instrument for aid (per capita) from a model of its supply-side determinants at the donor-recipient level, representing a modification of the instrument proposed by Rajan and Subramanian (2008), which itself was inspired in the earlier contribution of (inter alia) Tavares (2003). Predicted aid receipts from this model are then aggregated upwards to give a total predicted aid inflow for each country, averaged over a specified period. We redeploy the exact same approach here, but now use an extended dataset covering the period 1970-2007 contrary to the 1970-2000 period covered in Chapter 3.

Generation of a single instrument for aid means that over-identification tests cannot be employed. To address this, we also replicate the procedure in Chapter 3 whereby aggregated versions of the supply-side variables used to estimate the aid instrument are employed directly as instruments in (aggregate-level) aid-growth regressions. Thus, we employ relative population size, a dummy for whether the (recipient) country was ever a colony and their interaction (product). Subsequent Hansen/Sargan tests deriving from the same IV regressions as above, but now using the disaggregated instrument set, thus provide some insight as to instrument validity.

Choice of instruments for the intermediate variables in the systems estimates is not straightforward. Moreover, guidance from the literature is less clear. Following related exercises concerned with identifying macroeconomic transmission channels (e.g., Tavares and Wacziarg, 2001; Baldacci et al., 2004), a first approach to instrument selection is to impose (valid) exclusion restrictions on the aggregate conditioning set $\mathbf{X}$ (also Greene, 2002). This is suggested in Figures 4.1 and 4.2, where $\mathbf{X_2} \subset \mathbf{X}$ only affects final outcomes indirectly, via the intermediate variables. Thus, for the system estimates, a set $\mathbf{X_2}$ is employed as excluded instruments for the chosen intermediate variables, while $\mathbf{X_1} \subset \mathbf{X}$ is used as a standard conditioning set across all equations. The latter approach presumes that $\mathbf{X}$ can be appropriately partitioned and that $\dim(\mathbf{X_2}) \geq \dim(\mathbf{Y^{m'}})$. As discussed in Section 4.3.2, in the present exercise the latter requirement only holds with equality, meaning that over-identification tests cannot be employed. Thus, in addition to partioning $\mathbf{X}$, we introduce a small number of other variables as additional excluded instruments for $\mathbf{Y^{m'}}$, which are chosen in light of theoretical relevance and previous literature (see Section 4.3.2 and Appendix 4A for further details; for a similar approach see Tavares and Wacziarg, 2001). This yields the expanded instrument set $\mathbf{R} = [\mathbf{R_1} \quad \mathbf{X_2}]$.

With respect to E3 and E4, further complications derive from our interest in structural parameters. Although single-equation instrumental variables (IV) estimates of each equation of the system are expected to be consistent, they may be inefficient as they ignore correlation of cross-equation disturbances. Thus, gains could be realised from combining an appropriate single-equation estimator with the feasible GLS approach to estimating stacked equations due to Zellner (1962). In the case of OLS, this gives the seemingly unrelated regressions (SUR) estimator; in the case of 2SLS, this gives the 3SLS estimator. Systems methods are likely be especially important for estimating effect E3 as we do not include other intermediate variables on the RHS of these regressions – i.e., no direct interaction between the elements of $\mathbf{Y}$ is assumed. However, by proceeding with a simultaneous estimation of the system, residual cross-equation correlation is taken into account. In the absence of *a priori* knowledge of the interaction structure, this constitutes an imperfect but practical approach to recovering E3.

Finally, there is the question of the time period over which causal effects are to be estimated. A large part of the (modern) aid-growth literature has employed panel data, focussing on relatively short term effects of up to five years. In contrast, as discussed in Chapter 3 (also Rajan and Subramanian, 2008; Arndt et al., 2010a), there are good reasons to dispense with dynamic specifications because the impact of much developmental aid is likely to be cumulative and (very) long-term in nature, captured in Woolcock's (2011; 2009) metaphorical distinction between growing sunflowers versus oak trees (also Clemens et al., 2004; Temple, 2010). Thus, as in Chapter 3, the present focus remains on effects that cumulate over long time frames (1970-2007), given by the simple average of variables over this period. Further issues of variable selection and measurement are taken up below.

4.2.5 Variable selection and measurement

With respect to the measures of aid to be employed, as our intention is to investigate a range of intermediate outcomes, it might seem appropriate to employ sector-specific or narrower measures of aid (e.g., Clemens et al., 2004, 2011; Roodman, 2007). Such approaches are appealing in principle, but in practice entail substantial drawbacks. First, it is not clear that aid given to a given sector (objective) should only be expected to affect outcomes within the same sector (objective). For instance aid to education may well bring health-related benefits. Moreover, sector-specific measures of aid are problematic due to difficulties in attributing multi-sector funds to individual sectors, thereby adding to measurement error concerns. Second, OECD-DAC data regarding aid disbursements at the sector level are only available for a small number of recent years. This means that over long time horizons, it is necessary to impute sector-specific disbursement data from data on aid committments, which are known to diverge significantly even on aggregate (Odedokun, 2003). In light of these practical limitations, we employ aggregate measures of aid throughout; nonetheless, we recognise that such measures are imperfect and mask substantial differences in both quality and development intentions.

The measurement scale of aid is also an important empirical choice. On the one hand, it may be appropriate to scale the total aid received by a given country (over time) by its population size, leading to aid per capita as the 'treatment' variable of interest. This is not only an intuitive measure, but also is technically appealing as population size should not be confounded with GDP (or GDP growth) and many of the intermediate outcomes of interest are expressed in population terms (e.g., average years of schooling, life expectancy; see below). On the other hand, employment of aid per capita has specific limitations relative to the aid to GDP ratio (Aid/GDP), which has been more commonly used in the literature to date (see Chapter 3). First, it is hard to give a sensible and clear interpretation to any estimate effect of aid per capita on key macroeconomic outcomes, where variables are often measured in terms of or scaled by GDP. For instance, suppose we find that an inflow of US$10 of aid per capita causes the GDP growth rate to rise by 1 percentage point. Although this may be of interest *per se*, the problem is that the implied benefit-cost ratio of such aid fundamentally depends on the initial size of the economy.[4] Second, and relatedly, it is reasonable to assume that the real cost of providing a given flow of public services, such as education, tends to increase with GDP. Thus, especially over long time frames, the relative purchasing power of aid over a wide range of outcomes is best considered in economic terms and not in population terms. For these reasons, we employ Aid/GDP as the measure of aid throughout (unless noted otherwise).

Turning to the final and intermediate outcomes of interest, a vast range of candidate variables might be considered. However, data availability and computational limitations mean that exclusions must be imposed *ex ante*. With respect to final outcomes, we focus on growth, poverty, inequality and the sectoral composition of value added. The first three of these variables are intimately connected (see Bourguignon, 2003); therefore, we should expect to see a consistent pattern of effects across them. The remaining variables capture the extent of changes across different macroeconomic sectors (agriculture, industry and services). Historical experiences indicate that sustained growth transitions are normally associated with a declining share of agriculture and a rising share of industry in value added. At the same time, there are concerns that aid may provoke Dutch Disease, which is often associated with faster growth in service sectors than manufactures (e.g., Rajan and Subramanian, 2011). By including these variables we hope to gain insight into whether aid is associated with specific growth syndromes.

[4] Based on these hypothetical estimates, an economy of size US$700 per capita would exhibit a benefit-cost ratio of 0.7, while an economy of size US$7,000 per capita would exhibit a benefit-cost ratio of 7.

For intermediate outcomes, a number of 'usual suspects' emerge from previous literature. These fall into the following groups: (i) sub-components of GDP (investment, private consumption, government consumption); (ii) components of government revenue and spending; (iii) aggregate education and health outcomes (e.g., average years of schooling, life expectancy); and (iv) monetary and financial sector effects. A number of variables from each category is employed in the reduced form analysis, thus providing coverage over a wide range of meso-level aid effects. Details of the specific variables and sources of data are given in Appendix 4C.

The majority of outcome variables are measured as the average over the period 1970-2007 (as per aid and its generated instrument). However for a small number of outcomes (e.g., for education and health) the number of observations are scarce in the early years of this period, but increases over time. Thus, to avoid the long period average being dominated by more recent observations, a simple average of the earliest and latest observations is used in these cases, thereby assuming a linear trend over time. In a few other cases (e.g., for poverty rates), data is unavailable in the 1970s and early 1980s. In these cases, we define the dependent variable as the endpoint level (see Appendix 4C for the variables to which this applies).[5] It should be noted that this long-run averaging procedure applies only to the outcome and 'treatment' variables (whose impact we are interested in measuring; e.g., aid). So as not to be contemporaneous with the impact of aid or other intermediate growth drivers over the period of interest, and to avoid any possible confounding, background control variables are measured as their starting values (i.e., for 1970), or the closest available data point. Finally, to assure direct comparability with previous studies, we employ the same sample of 78 developing countries and the same set of control variables as in Chapter 3 (AJT10).[6]

4.3 Results

4.3.1 Reduced form

Table 4.1 summarises results from separate regressions for the selected set of final outcomes, using OLS, LIML, and inverse probability weighted least squares (IPWLS) estimators. The latter constitutes an extension of doubly robust methods to the instrumental variables context and serves as a robustness check on the LIML results (see Chapter 3; Arndt et al., 2010a) . To assist interpretation, all variables are standardised. Each cell of Table 4.1 gives the standardised coefficient on the aid to GDP ratio and, in the adjacent cell, the estimated probability that the true parameter estimate is zero. These correspond to results from individual regressions in which the row variable is the regressand. The table also reports the number of observations in each regression, and our prior regarding the expected direction of the partial correlation between aid and each outcome.

Appendix Table 4B.1, gives the full regression estimates corresponding to the growth outcome summarised in Table 4.1. Specifically, the first row of the latter extracts results from columns I, II and III of the former. The full regression results also report tests of instrument strength, which are relevant to *all* the reduced form IV regressions reported in this section as they employ the same first stage. These statistics give no cause for concern. Additionally, the table shows the results hold-up when aid per capita is used instead of

[5]The practice of using the endpoint level is encountered in the cross-country growth regression literature where final income can be used in place of the growth rate, and initial income is dropped from the RHS (e.g., Mankiw et al., 1992).

[6]We add a dummy for being an oil producer in 1960 due to the extension of the dataset from 2000-2007, which covers a period of rapid economic growth in oil-producing countries driven by rising oil prices.

the aid to GDP ratio, providing further credence to our results. Indeed, when applied to the full range of intermediate and final outcomes, as well as the structural form (see Section 4.3.2), use of aid per capita yields highly consistent results with those presented here (for full details see Arndt et al., 2011).

The results reported in Table 4.1 broadly conform to priors. The finding of a positive (causal) relationship running from aid to growth, established in Chapter 3, is replicated for the extended dataset used herein. We also find that aid reduces poverty but leaves inequality unaffected on average, which is consistent with the theoretical relation between growth, poverty and inequality (Bourguignon, 2003). Aid also leads to a drop in the weight of agriculture in GDP, implying the aid stimulates the growth of non-agricultural sectors relatively more. Indeed, the IPWLS estimates (and OLS) indicate a corresponding increase in industry's GDP share; however the impact on services is more ambiguous.

The final two columns of the table report additional test statistics. First are results from Durbin-Wu-Hausman χ^2 tests of the null hypothesis that the aid variable can be treated as exogenous (i.e., that the OLS results are consistent).[7] This is rejected in only some cases, suggesting that concerns surrounding aid's endogeneity may not be material for all outcomes, but certainly cannot be ignored in relation to growth. The final column gives the probability from Hansen-J over-identification tests, based on the same regression specification in the LIML column, but employing three aggregated instruments for aid instead of the single generated aid instrument (see Section 4.2.4). A significant result ($< 10\%$) is grounds to reject the (joint) null hypothesis that the instruments are valid (are uncorrelated with the regression error). This test is passed comfortably in all cases except for the Gini coefficient, implying that both the OLS and IV point estimates for the Gini may be biased. On the other hand, assuming the tests are independent, the probability that the Hansen-J test is passed at the 10% level in all seven cases is as low as one in two, even if the null is in fact always true.

Table 4.2 reports reduced form results for the effect of Aid/GDP on the chosen set of intermediate outcomes, adopting the same format as Table 4.1. Again, estimates are broadly consistent across the different estimators and broadly conform to priors. Investment, government consumption and revenues are all positively affected by aid inflows. Sub-components of government spending indicate that aid also boosts expenditure in social sectors. Although the OLS estimates indicate a positive and borderline-significant effect of aid on military spending, this does not hold up in the IV estimates. The impact of aid on key social outcomes corroborates positive results of a number of previous studies which employ panel techniques (see Section 4.4). Aid has a positive causal effect on average years of schooling, and secondary schooling in particular. Health outcomes also conform to expectations concerning the sign of the coefficients on aid; while the LIML estimates slightly exceed conventional significance levels, the IPWLS estimates for both infant mortality and life expectancy are significant. Results for monetary and financial sector indicators are ambiguous, suggesting there is no evidence of a systematic effect of aid on inflation, real interest rates, or credit to the private sector.

Lastly, the test statistics in the final two columns of Table 4.2 broadly follow the pattern of Table 4.1. For nearly two thirds of the intermediate outcomes we must reject the null hypothesis that aid is exogenous (at conventional significance levels of $< 10\%$). Thus, although aid may not be endogenous for all possible outcomes, such endogeneity needs to be taken seriously for a wide range of intermediate outcomes – i.e., *ex ante*, aid cannot be assumed to be exogenous. With respect to the validity of the generated aid instrument, however, for the large majority of outcomes we cannot reject the null of the Hansen-J test, the two exceptions being government size and fertility rates. Thus, again, the instrument appears to perform

[7] Implemented in Stata via the `endog()` option of the `ivreg2` command used to estimate the LIML results.

Table 4.1: Summary of reduced form results for relationships between Aid/GDP and final outcomes

	Prior	N	OLS	Pr.	LIML	Pr.	IPWLS	Pr.	Endog. test	Hansen-J
GDP per capita growth	+	78	-0.115	0.12	0.639	0.09*	0.612	0.03**	0.01**	0.84
Agriculture, value added (% GDP)	-	76	0.035	0.69	-0.652	0.08*	-1.062	0.01***	0.03**	0.89
Industry, value added (% GDP)	+	76	0.220	0.10	0.648	0.24	0.822	0.04**	0.42	0.68
Services, etc., value added (% GDP)	?	76	-0.273	0.04**	0.242	0.56	0.634	0.23	0.19	0.65
Poverty headcount at $2 a day	-	64	0.018	0.84	-0.471	0.05*	-0.438	0.19	0.01**	0.42
Poverty headcount at $1.25 a day	-	64	0.084	0.57	-0.487	0.09*	-0.285	0.42	0.01**	0.78
Gini index	?	65	-0.186	0.14	-0.035	0.91	-0.328	0.57	0.56	0.07*

significance: * 0.1, ** 0.05, *** 0.01

Notes: each cell of columns OLS, LIML and IPWLS reports the standardised coefficient on Aid/GDP from individual reduced form regressions in which the row variable enters as the dependent variable; adjacent columns report the corresponding probability that the true parameter is equal to zero; estimation method is indicated by the column headings; aid treated as endogenous in LIML and IPWLS only; ‘Prior’ gives the *a priori* expected direction of the aid-outcome relationship; ‘Endog. test’ reports the probability from a Durbin-Wu-Hausman test that aid can be treated as exogenous; ‘Hansen-J’ gives the probability associated with the Hansen-J statistic from a LIML regression using three aggregated instruments; all regressions include the same set of control variables (see text) and employ robust standard errors.

Source: authors’ calculations; see Appendix 4C for variable definitions and sources.

Table 4.2: Summary of reduced form results for relationships between Aid/GDP and intermediate outcomes

	Prior	N	OLS	Pr.	LIML	Pr.	IPWLS	Pr.	Endog. test	Hansen-J
Investment (% GDP)	+	78	0.319	0.00***	0.795	0.03**	0.357	0.16	0.15	0.73
Consumption (% GDP)	-	78	0.174	0.24	-0.515	0.25	-0.779	0.05**	0.07*	0.70
Government (% GDP)	+	78	0.513	0.00***	0.758	0.06*	0.873	0.05*	0.51	0.03**
Revenue, excluding grants (% GDP)	?	69	0.470	0.03**	2.362	0.00***	1.188	0.00***	0.00***	0.39
Health expend., public (% GDP)	+	78	0.403	0.10	0.363	0.36	-0.610	0.23	0.91	0.21
Education expend., public (% GDP)	+	76	0.485	0.00***	1.423	0.00***	1.644	0.00***	0.01**	0.18
Military expenditure (% GDP)	+	77	0.387	0.11	0.361	0.32	0.255	0.65	0.94	0.37
Av. years total schooling, 15+	+	72	0.206	0.10	1.010	0.04**	0.511	0.08*	0.03**	0.96
Av. years primary schooling, 15+	+	72	0.267	0.04**	0.673	0.13	0.246	0.28	0.33	0.97
Av. years secondary schooling, 15+	+	72	-0.006	0.97	1.476	0.03**	0.818	0.04**	0.00***	0.68
Life expectancy at birth, total (years)	+	78	-0.087	0.09*	0.187	0.16	0.329	0.06*	0.02**	0.33
Infant mortality rate	-	75	0.055	0.52	-0.306	0.17	-0.434	0.06*	0.09*	0.46
Mortality rate, under-5 (per 1,000)	-	75	0.101	0.23	-0.320	0.15	-0.297	0.14	0.04**	0.44
Death rate, crude (per 1,000 people)	-	78	0.216	0.01***	-0.162	0.40	-0.116	0.57	0.02**	0.97
Fertility rate (births / woman)	?	77	-0.053	0.47	-0.344	0.09*	0.102	0.49	0.12	0.05*
Consumer price inflation (%)	?	77	0.114	0.34	-0.677	0.18	-0.793	0.21	0.07*	0.70
Real interest rate (%)	?	77	-0.017	0.95	-0.488	0.37	0.199	0.70	0.36	0.27
Domestic credit to private sector (% GDP)	?	78	-0.053	0.61	-0.009	0.98	0.360	0.51	0.90	0.69

significance: * 0.1, ** 0.05, *** 0.01

Notes: each cell of columns OLS, LIML and IPWLS reports the standardised coefficient on Aid/GDP from individual reduced form regressions in which the row variable enters as the dependent variable; adjacent columns report the corresponding probability that the true parameter is equal to zero; estimation method is indicated by the column headings; aid treated as endogenous in LIML and IPWLS only; 'Prior' gives the a priori expected direction of the aid-outcome relationship; 'Endog. test' reports the probability from a Durbin-Wu-Hausman test that aid can be treated as exogenous; 'Hansen-J' gives the probability associated with the Hansen-J statistic from a LIML regression using three aggregated instruments; all regressions include the same set of control variables (see text) and employ robust standard errors.

Source: authors' calculations; see Appendix 4C for variable definitions and sources.

well, providing further support to is overall suitability.

4.3.2 Structural form

The reduced form results identify some of the key channels through which aid affects various intermediate and final outcomes. To estimate the structural form, however, it is not feasible to encompass all these variables. A pared-down system is necessary. To do so, we restrict the set of final outcome variables to growth. With respect to intermediate outcomes (transmission channels), we focus on standard inputs into an aggregate production function, namely investment and human capital – captured by education (average school years) and life expectancy.

The simplified model takes the form of Figure 4.2 but includes three intermediate outcomes. No specific assumptions are made as regards the interaction structure between these elements of $\mathbf{Y}$. Algebraically, this gives a triangular system in five equations: per capita growth; three aggregate inputs to growth (investment, education and health) in which aid features as a determinant; and a supply-side model for aid. The full system of equations and corresponding exogenous variables employed in both the baseline and augmented models (see below) is given in Appendix 4A. Simply put, the baseline model specifies growth as a function of the full set of control variables used previously and the three (proximate) growth drivers – investment, education (average years of schooling), and health (life expectancy). These three growth drivers are also specified as functions of the same control variables to which Aid/GDP is added. Finally, the equation for Aid/GDP replicates the first stage equation used in the reduced form models.

We begin by ignoring endogeneity concerns and separately estimate the individual equations of the baseline model by OLS, employing the same RHS specifications as in Section 4.3.1. Summary results are reported in column I of Appendix Table 4B.2. Next, we estimate the same equations by the iterated SUR (iSUR) method which, as noted by Pagan (1979), can be interpreted as an instrumental variables estimator as it numerically produces LIML parameter estimates on convergence (also Gao and Lahiri, 2000).[8] Thus, applied to the baseline model, which includes a separate equation for aid, the iSUR estimates in column II of Appendix Table 4B.2 address the endogeneity of aid, but not that of the intermediate outcomes.

The previous estimates exhibit possible confounding of investment and human capital with growth; however, as discussed in Section 4.2.3, we cannot rely on the aid instrument for identification. Thus, we define an augmented model which has the same structure as the baseline model but with modified specifications. Specifically, the set of control variables used in the growth equation is restricted due to the partitioning procedure. The control variables removed, namely $\mathbf{X_2} =$ [initial education, air distance, investment prices], are added to the RHS of the equations for each of the proximate growth drivers as well as the aid equation. This is due to their essential redundancy in previous estimates of growth (see Appendix Table 4B.1) and their plausible relevance to the chosen transmission channels. Since this gives a just-identified system, we also add three supplementary exogenous variables – namely, a dummy for being landlocked, ethnic fractionalization and (current) HIV prevalence. These are chosen in light of theoretical relevance and previous literature. HIV prevalence is a material determinant of average life expectancy in a number of developing countries. It is also plausibly exoegenous, controlling for initial income and life expectancy, as despite the epidemic being most significant in sub-Saharan Africa (for

[8]In the systems estimates (only) we correct for a small number missing values, using predicted values from an OLS regression of the variable of interest against the full set of exogenous variables used in steps one and two. Results are not sensitive to the choice of interpolation method.

which a regional dummy is included in all regressions), differences in prevalence rates within this region are largely determined by sexual behaviour and biological factors, rather than differences in incomes (for discussion see Buvé et al., 2002). The disadvantages of being landlocked are understood to operate largely through lower returns to investment (Faye et al., 2004), while evidence suggests that higher levels of ethnic fractionalization can create particular challenges for public education provision (Buchmann and Hannum, 2001), and also tends to increase political risks thereby lowering investment levels (Mauro, 1995; Montalvo and Reynal-Querol, 2005). The point is that these variables are plausible determinants of proximate growth drivers, only affecting growth indirectly through these channels.

Instrumental variables estimators are appropriate to estimate this augmented model. Column III of Appendix Table 4B.2 estimates each equation separately using a standard 2SLS estimator; column IV applies a 3SLS estimator; and column V reapplies the iSUR method to the augmented system. It is relevant to investigate the strength and validity of the expanded instrument set, particularly in light of the use of systems estimators. This can be gauged directly from single-equation IV estimates of the growth equation, where we apply the full set of excluded instruments, including the generated instrument for aid, to each of the three endogenous intermediate outcomes. The results (not reported) indicate the new instrument set does not violate over-identification tests (Hansen J statistic = 0.90; probability = 0.92).[9] Nonetheless, instrument strength varies across the three endogenous variables. Shea's (1997) partial R^2 measure from the individual first stage regressions ranges from 0.16 (education), 0.20 (investment) to 0.34 (health). Overall, the instrument set appears weak, with a Kleibergen-Paap Wald F statistic of only 1.38, suggesting there is likely to be substantial bias in the system from weak instruments. However, Anderson-Rubin χ^2 Wald tests, which are robust to the presence of weak instruments, confirm a statistically significant relationship between the set of endogenous variables and growth. Further regressions confirm the merit of including the additional instrument set, rather than simply relying on the partition of $\mathbf{X}$ alone. Indeed, in a similar model employing only $\{\mathbf{V}, \mathbf{X_2}\}$ as instruments for the intermediate variables, the specification fails rank-based under-identification tests and gives a Kleibergen-Paap Wald F statistic of only 0.35.

Despite these issues, Appendix Table 4B.2 reveals that the estimated parameters are broadly stable across all estimators and specifications. There are two main exceptions. First, as might be expected due to aid's endogeneity, the OLS estimates (column I) differ most and particularly in the growth equation. Second, the estimated parameter on education in the growth equation is sensitive to the estimation method and specification. The parameter is positive in the baseline model and highly significant under iSUR but is insignificant in all IV estimates of the augmented model. Overall, the principal transmission channels linking aid to growth are investment and health (proxied by life expectancy). In all estimates, aid appears to have a significant and positive impact on each one of the three intermediate outcomes. In support of earlier results, the estimated coefficients on aid from the structural estimates of these intermediate equations typically fall within sampling variation of the reduced form estimates (see Tables 4.1 and 4.2).

To decompose the effects of aid on growth, Table 4.3 quantifies the contribution of each channel based on the iSUR results in columns II and V of Appendix Table 4B.2. The first column gives the direct effect of aid on the specified channel (effect E3), while the second column gives the effect of that channel on growth, controlling for the other channels (effect E4). The final column reports their product, thereby giving an estimate of the individual effect of aid on growth due to the specified channel. The sum of these individual effects gives the estimated aggregate contribution of aid to growth due to the three channels.

[9]These results derive from continuously updated (CUE) GMM estimates with robust standard errors, which yield highly comparable results to those of the estimates of the augmented model reported in Appendix Table 4B.2; the same goes for 3SLS GMM systems estimates – available on request from the authors. Note that the advantage of these GMM approaches, relative to the present simple 3SLS estimates, is they do not assume errors are independent and identically distributed (i.i.d.).

Table 4.3: Structural estimates of transmission channels linking aid to growth via selected intermediate outcomes (Y_m)

Y_m	Aid $\rightarrow Y_m$ [A]	$Y_m \rightarrow$ growth [B]	Aid $\rightarrow$ growth [A $\times$ B]
	(a) iSUR - baseline		
Investment	0.795	0.326	0.259
	(7.19)	(3.73)	(3.33)
Education	0.543	0.406	0.220
	(7.79)	(2.90)	(2.73)
Health	0.187	0.852	0.159
	(4.23)	(3.85)	(2.89)
Overall		1.584	0.639
		(5.73)	(5.12)
	(b) iSUR - augmented		
Investment	0.774	0.523	0.404
	(8.11)	(6.30)	(5.00)
Education	0.503	-0.066	-0.033
	(6.84)	(-0.51)	(-0.51)
Health	0.211	0.555	0.117
	(5.63)	(2.44)	(2.27)
Overall		1.011	0.488
		(3.68)	(4.22)

Notes: all results are taken from the system estimates summarised in Appendix Table 4B.2 – panel (a) refers to column II of Table 4B.2 and panel (b) to column V of Table 4B.2; first column gives the direct effect of Aid/GDP on the channel of interest; second column gives the effect of the channel on growth; final column gives their product; 'overall' channel gives the column sum of the individual channels; all coefficients are expressed in standardised form; estimated t-statistics are given in parentheses.
Source: authors' calculations; see Appendix 4C for variable definitions and sources.

Corresponding t-ratios are also reported in parentheses, where the variance of the total effect of aid on growth via each channel given in the final column is approximated by the unbiased variance estimator due to Goodman (1960), which assumes the component point estimates are independent.[10]

The pattern of results from the decomposition confirms that the total effect of aid on growth that is due to the chosen channels is positive, significant and highly consistent with the reduced form results. In the baseline model, the sum of the aid→growth effect over each of the three channels equals 0.639 which is almost identical to the LIML and IPWLS estimates from Table 4.1 (first row). The individual coefficients and relative contributions of the individual channels varies somewhat between the baseline and augmented models, reflecting the fact that the latter employs instrumental variables to address the potential endogeneity of the proximate drivers. Nonetheless, the investment channel is the most important channel in both models (accounting for 80% of the total contribution under the augmented model), which arises mainly from the larger direct effect of aid on investment compared to on education or health. In both models, Aid/GDP makes a positive and significant contribution to growth through the health channel, but with a slightly smaller relative contribution.

[10]Alternative variance estimates, which assume some positive correlation between the point estimates, yield similar results.

An outstanding puzzle is the absence of a robust effect running from education to growth in the augmented model (which is despite finding that aid appears to enhance educational outcomes). There are various possible rationales for this finding. One is the absence of dynamics in our model – by focussing on a long period cross-section, we neglect the time profile of different impacts. It may be that aid has significantly financed education relatively recently, but insufficient time has passed for this to cumulate into detectable labour market and subsequent growth effects. The empirical weakness of the education-growth relationship also is not new. Pritchett (2001) suggests various reasons for the absence of a detectable growth-impact of education in macro-data, including low quality of schooling and weak growth in the demand for skilled labour. Attenuation bias due to measurement error may be another explanation. As Cohen and Soto (2007) show, standard cross-country measures of education outcomes are noisy and potentially misleading in empirical applications (also Hanushek and Kimko, 2000). Moreover, the discussion in Section 4.3 flagged the problem of weak instruments, particularly with respect to education.

Lastly, the systems model passes the falsification test proposed in Section 4.2.3. When the generated aid instrument is employed directly (as an included instrument) in a single-equation IV estimate of the growth equation, alongside the three intermediate growth drivers (appropriately instrumented), the estimated coefficient on aid is insignificant. Corroborating the results of Table 4.3, this implies that the transmission channels included in our simplified structural model account for the majority of the causal effect of aid on growth.

4.4 Interpretation

Thus far, discussion of results has concentrated on the sign and domain of parameter estimates. It is also helpful to reflect on whether these ranges are economically plausible. This cannot be directly ascertained from previous tables as results are given in standardized form. Consequently, for a selected number of final and intermediate outcomes, Table 4.4 presents the reduced form point estimates and 95% confidence intervals for the expected return to an average annual aid inflow equal to 5% of GDP over the period 1970–2007 (which is slightly greater than double the observed median Aid/GDP for all countries in the sample; see Appendix 4C). We find the long-run impacts of aid are both plausible and material. According to the IPWLS point estimates, such an aid inflow is expected to increase the average rate of economic growth by around 1.5 percentage points, reduce poverty by around 9 percentage points, raise investment by around 5 percentage points in GDP, augment average schooling by 1.4 years, boost life expectancy at birth by 4 years and reduce infant mortality by twenty in every 1000 births.[11]

The internal consistency of the reduced and structural form results corroborate the present findings. Our reduced form results also are externally consistent. Investment is frequently identified as a principal growth determinant (Mankiw et al., 1992; Sala-i-Martin et al., 2004), and evidence points to (very) long-run growth effects from improvements in aggregate health (Jack and Lewis, 2009). The present estimates of the impact of aid on growth imply a reasonable aggregate ‘return on aid’ over the entire period. That is, for a counterfactual per capita growth rate of 1.5% per annum, the relative income gain due to aid minus its cost (a constant 5% share of GDP) yields a modified internal rate of return around 19% over a 37 year horizon (1970-2007).[12] This is very close to the estimate of an approximate 20%

[11] These effects refer to the expected change in the average of the outcome variable over the full period – i.e., the difference in the average for that variable versus its counterfactual average.

[12] The modified internal rate of return (MIRR) is an improvement on the simple internal rate of return (IRR) as it ensures a

Table 4.4: Estimated effect on various outcomes of receiving an aid inflow equal to 5% of GDP (annual average for 1970-2007)

	OLS			IPWLS		
Variable	Lower	Point	Upper	Lower	Point	Upper
GDP per capita growth	-0.62	-0.27	0.08	0.11	1.45	2.79
Agriculture, value added (% GDP)	-2.49	0.62	3.72	-33.06	-18.90	-4.74
Poverty headcount at $1.25 a day	-6.76	2.67	12.10	-31.36	-9.09	13.18
Investment (% GDP)	2.09	4.35	6.62	-1.84	4.87	11.59
Government (% GDP)	2.73	5.72	8.70	-0.15	9.73	19.61
Revenue, excluding grants (% GDP)	0.52	5.49	10.46	4.94	13.88	22.83
Av. years total schooling, 15+	-0.12	0.57	1.27	-0.15	1.42	2.99
Life expectancy at birth, total (years)	-2.39	-1.10	0.19	-0.21	4.14	8.49
Infant mortality rate	-5.36	2.56	10.47	-41.58	-20.33	0.92

Notes: the table reports the raw estimated effect of a 5% Aid/GDP inflow on selected outcome variables based on the reduced form regressions summarised in Tables 4.1 and 4.2; 'lower' and 'upper' refer to 95% confidence limits; estimators indicated by column headings.
Source: authors' calculations; see Appendix 4C for variable definitions and sources.

return to foreign aid due to Dalgaard and Hansen (2005). Also, our replication of AJT10 with an extended dataset yields highly consistent point estimates for the aid-growth coefficient.[13] Thus, for the periods 1960-2000, 1970-2000 (estimated in Chapter 3) and 1970-2007 (here), the estimated impact of aid on growth lies in a comparable, positive domain.

The reduced form results for the intermediate outcomes are also consistent with previous studies (see Chapter 2). Gomanee et al. (2005), Masud and Yontcheva (2007) and Mishra and Newhouse (2009) all find positive effects of certain kinds of aid on health outcomes; while Michaelowa (2004) and Dreher et al. (2008) report positive effects of aid on education enrolment rates. Similar to our falsification test, Hansen and Tarp (2001) find that aid is not significant in a growth regression which controls for investment and human capital, but that aid remains a significant determinant of investment. Furthermore, our results provide a basis to reject the (largely) theoretical concerns that aid undermines domestic revenue mobilization (e.g., Moss et al., 2006). Rather, our results are closer to those of Pivovarsky et al. (2003), who find a positive revenue impact from concessional loans (but a small negative effect from grants). Similarly, and contrary to concerns that aid's positive developmental impact is muted due to its fungibility (Pack and Pack, 1993), our results corroborate van de Walle and Mu (2007) and show that some aid 'sticks' to the social sectors and, thus, is not entirely fungible.

With respect to the link between poverty and growth, the reduced form results enable us to back-out an estimate of the aid-induced growth semi-elasticity of poverty (GSEP). This is given by the estimated absolute change in the poverty (headcount) rate divided by the estimated percentage change in mean

unique return rate where cashflows are both positive and negative and also takes a more conservative view of the return on reinvested interim cash flows (see Lin, 1976).

[13]Specifically, for the equivalent specification and estimator, Arndt et al. (2010b) report a coefficient of 0.42 on aid and a standard error of 0.19 for the period 1970-2000. The comparable (unstandardized) coefficient corresponding to column III of Table 4B.1 is 0.30, with a standard error of 0.18.

income due to aid over the period.[14] For both the US$1.25 and US$2 poverty measures, we find that the aid-induced GSEP is around 0.30 (or 0.26 and 0.31 respectively), meaning that a 1% increase in mean income tends to lead to a 0.30 percentage point fall in the headcount poverty rate. This is situated just below the average of the range of GSEP estimates calculated by Klasen and Misselhorn (2008, Table 7), suggesting there is no reason to conclude that aid is any less effective in reducing poverty than other growth drivers over the long-run.

Finally, we note that the relationship between aid and productivity is left largely unresolved here. In the structural form estimates, the sum of the effects of aid via the three selected channels appears to capture the vast majority of the effects of aid on growth (around 100% in the baseline model and 75% in the augmented model). One interpretation is that any productivity-enhancing effects due to aid are thus small, or that any such effects (whether positive or negative) are cancelled-out by channels excluded from the simplified our structural equations. It is also plausible that productivity effects may already be captured by the three existing channels. For example, technical change is often embodied in new capital goods, making it difficult to separate from investment; and health improvements would typically register as productivity improvements in a standard productivity analysis. The implications of aid for productivity therefore remains an important area for future analysis.

4.5 Conclusions

This chapter aimed to answer the question: "What has aid accomplished over the past four decades?". Evidence of this kind speaks to the first order policy problem facing donors (recipients) – namely, whether they should continue to provide (accept) foreign aid. Applying the SCM framework, we began by presenting a general structural model of the relationship between aid and aggregate outcomes, which is consistent with the framework employed in the literature on growth empirics. To estimate this model, we first calculated reduced form estimates of the relationship between aid and final outcomes. The results confirm a robust positive impact of aid on growth for the 1970-2007 period, thereby replicating the findings of Chapter 3 using an extended dataset. The aggregate effects of aid are also coherent. On average and over the long-run, foreign aid reduces poverty with no significant impacts on inequality. Aid also contributes to more rapid expansion of 'modern' sectors (industry) and a relative decline of agriculture's share in GDP.

To gain insight into relevant transmission channels, we applied the same reduced form approach to a set of intermediate outcomes. These revealed a range of positive and significant effects due to aid – e.g., on investment, government revenue and spending, and on social outcomes. Lastly, we estimated a simplified representation of the full structural form, allowing aid to impact on growth through inputs into an aggregate production function. The results were consistent with those of the reduced form, confirming a strong direct effect of aid on investment, education and health outcomes. It also indicated that investments in physical capital and improvements in health are the most robust transmission channels through which aid promotes growth. The effect of education on growth is ambiguous.

[14]Calculated as GSEP = $-\beta_p/[(1+\beta_g)^{37}-1]$, where β_p is the estimated coefficient on Aid/GDP in the reduced form aid-poverty regression and β_g is the estimated coefficient on Aid/GDP in the reduced form aid-growth regression (appropriately scaled). Note that the latter coefficient estimates the expected increase in the average annual growth rate over the period 1970-2007, while the former estimates the expected overall change in the poverty rate due to aid over the same timeframe. Consequently, to compare like with like, we need to calculate the expected overall percentage change in mean income, which is given by the denominator of the GSEP equation.

In summary, based on results covering a wide range of outcomes, aid can point to a series of accomplishments with a positive impact on the growth and development process. There is no evidence that aid is detrimental. Aid contributed to economic growth by stimulating physical capital investments and improving human capital, particularly health. Overall, the experience of the past four decades or so provide no support to the argument that aid flows should cease. Moreover, the present analysis provides some guidance on the form of assistance by highlighting the importance of physical and human capital accumulation. Finally, while positive, the magnitude of the estimated effects of aid tend are generally moderate but cumulate in to material effects over the long-run. It follows that aid should not be considered a panacea or silver-bullet to stimulating growth and development.

4A Appendix: model summary

The following equations summarise the simplified system estimated in the structural form analysis of Section 4.3.2.

Eq.	Baseline model	Augmented model
(i)	$growth = f(ki, edu, health, x)$	$growth = f(ki, edu, health, x_1)$
(ii)	$ki = f(aid, x)$	$ki = f(aid, r, x)$
(iii)	$edu = f(aid, x)$	$edu = f(aid, r, x)$
(iv)	$health = f(aid, x)$	$health = f(aid, r, x)$
(v)	$aid = f(v, x)$	$aid = f(v, r, x)$

Variable definitions: $growth$ = Real GDP growth per capita; ki = Investment / real GDP; edu = Av. years total schooling; $health$ = Life expectancy at birth; aid = Aid per capita; v = Generated aid instrument.

Vector definitions: with the exception of HIV prevalence, all elements of the following vectors refer to initial conditions only (i.e., values do not refer to period averages); $x' = [x_1\ x_2]$; $r' = [r_1\ x_2]$; x'_1 = (Income per capita, Sachs-Warner trade policy index, Life expectancy, Geography, Coastal population density, Malaria prevalence index, Civil liberties, Oil producer); x'_2 = (Primary education enrolment rate, Price of investment goods, Air distance to major cities); r'_1 = (Prevalence of HIV, Ethnic Fractionalization, Landlocked).

4B Appendix: additional tables

Table 4B.1: Estimates of reduced form relation between aid and growth, 1970-2007

	Aid / GDP				Aid per capita			
	I	II	III	IV	V	VI	VII	VIII
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
Aid measure	-0.11	0.64*	0.61**	0.43*	0.06	0.30**	0.23**	0.24**
	(0.07)	(0.38)	(0.29)	(0.25)	(0.07)	(0.14)	(0.11)	(0.12)
GDP per capita (PPP)	-0.71***	-0.32	-0.35	-0.43**	-0.66***	-0.70***	-0.70***	-0.69***
	(0.15)	(0.24)	(0.23)	(0.19)	(0.14)	(0.13)	(0.13)	(0.13)
Primary schooling	0.15	0.16	0.24	0.15	0.14	0.11	0.24	0.12
	(0.47)	(0.52)	(0.45)	(0.47)	(0.46)	(0.41)	(0.40)	(0.40)
Trade policy index	0.77***	0.69**	0.70***	0.71***	0.77***	0.81***	0.80***	0.80***
	(0.22)	(0.28)	(0.26)	(0.24)	(0.23)	(0.21)	(0.19)	(0.21)
Life expectancy	0.03*	0.06**	0.06***	0.05***	0.04**	0.04**	0.04**	0.04**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Geography	0.08	0.10	0.12	0.10	0.10	0.17*	0.17*	0.15
	(0.11)	(0.12)	(0.10)	(0.11)	(0.11)	(0.10)	(0.09)	(0.10)
Coastal pop. dens.	0.00***	0.00**	0.00***	0.00**	0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Malaria prevalence	-0.96***	-1.19***	-1.08***	-1.13***	-1.01***	-1.07***	-0.92***	-1.05***
	(0.26)	(0.39)	(0.32)	(0.33)	(0.27)	(0.28)	(0.24)	(0.27)
Investment prices	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Civil liberties	-0.05	-0.34	-0.39	-0.26	-0.16	-0.41	-0.39	-0.35
	(0.27)	(0.39)	(0.32)	(0.32)	(0.28)	(0.33)	(0.26)	(0.32)
Air distance	-0.26	-0.34*	-0.39**	-0.32*	-0.28	-0.31*	-0.33**	-0.30*
	(0.18)	(0.21)	(0.20)	(0.19)	(0.18)	(0.17)	(0.17)	(0.17)
Oil producer	0.44**	0.91***	1.01***	0.78***	0.53***	0.59***	0.68***	0.58***
	(0.18)	(0.28)	(0.21)	(0.22)	(0.17)	(0.16)	(0.13)	(0.15)
N	78	78	78	78	78	78	78	78
R2	0.74	0.54	0.55	0.64	0.74	0.69	0.69	0.71
Weak id. stat.		9.16	12.13	4.26		18.44	11.70	9.56
Anderson-Rubin test		0.02	0.01	0.06		0.02	0.01	0.06
Endogeneity test		0.012	0.018	0.005		0.034	0.071	0.029

significance: * 0.1, ** 0.05, *** 0.01

Notes: columns report full results for growth regressions with alternative measures of aid (indicated); intercept and region dummies included but not shown; growth and aid measures enter in standardized form; columns I and V estimated by OLS; columns II and VI estimated by LIML, using a single generated aid instrument; columns III and VII replicate the latter with IPWLS; columns IV and VIII use three instruments for aid taken from the zero stage regression employed to generate the single aid instrument; final two rows report probabilities; endogeneity test is Durbin-Wu-Hausman χ^2; standard errors (in parentheses) are robust.

Source: authors' calculations; see Appendix 4C for variable definitions and sources.

Table 4B.2: Summary of regression estimates of simplified structural form

	Baseline model		Augmented model		
	I OLS b	II iSUR b	III 2SLS b	IV 3SLS b	V iSUR b
a. Growth					
Investment	0.11	0.33***	0.49**	0.49**	0.52***
Education	0.06	0.41***	-0.02	-0.02	-0.07
Life expectancy	0.53**	0.85***	0.55	0.55	0.55**
Primary schooling	0.04	-0.47			
Investment prices	-0.00	-0.00			
Air distance	-0.15	0.05			
b. Investment					
Aid / GDP	0.32**	0.80***	0.73**	0.77***	0.77***
Primary schooling	0.68	0.68	0.07	-0.00	0.01
Investment prices	-0.00**	-0.00*	-0.00	-0.00	-0.00
Air distance	-0.69***	-0.74***	-0.98***	-0.96***	-0.95***
HIV prevalence			0.07***	0.08***	0.08***
Landlocked			-0.38	-0.39*	-0.39*
Ethnic fraction.			0.96**	0.90***	0.90***
c. Education					
Aid / GDP	0.12	0.54***	0.52**	0.50**	0.50***
Primary schooling	1.11***	1.12***	1.01**	1.04***	1.04***
Investment prices	-0.00*	-0.00	-0.00	-0.00	-0.00
Air distance	-0.13	-0.18	-0.27	-0.28*	-0.28*
HIV prevalence			0.02	0.02	0.02
Landlocked			-0.04	-0.04	-0.04
Ethnic fraction.			0.55*	0.58**	0.58**

continued overleaf ...

	Baseline model		Augmented model		
	I	II	III	IV	V
	OLS	iSUR	2SLS	3SLS	iSUR
	b	b	b	b	b
d. Health					
Aid / GDP	-0.09*	0.19***	0.22	0.21	0.21***
Primary schooling	-0.06	-0.06	0.03	0.05	0.04
Investment prices	0.00	0.00	-0.00	-0.00	-0.00
Air distance	-0.06	-0.09	0.04	0.04	0.03
HIV prevalence			-0.02	-0.02*	-0.02*
Landlocked			-0.36***	-0.36***	-0.36***
Ethnic fraction.			-0.11	-0.10	-0.10
e. Aid per capita					
Aid instrument	0.03***	0.03***	0.03***	0.03***	0.03***
Primary schooling	-0.34	-0.34	-0.17	-0.26	-0.25
Investment prices	-0.00	-0.00	-0.00	-0.00	-0.00
Air distance	0.14	0.14	0.20	0.22	0.23
HIV prevalence			-0.02	-0.02	-0.02
Landlocked			0.30	0.30	0.30*
Ethnic fraction.			-0.55	-0.63**	-0.63**
N	78	78	78	78	78
R2 equation a	0.76	0.71	0.70	0.70	0.69
R2 equation b	0.64	0.56	0.69	0.68	0.68
R2 equation c	0.85	0.79	0.81	0.82	0.82
R2 equation d	0.94	0.92	0.93	0.93	0.93
R2 equation e	0.70	0.70	0.73	0.73	0.73

significance: * 0.1, ** 0.05, *** 0.01

Notes: columns report point estimates (beta coefficients) of the respective equations of the simplified full system – see Appendix 4A; estimator and model specification indicated by the column headings; all endogenous variables enter in standardized form; exogenous variables common to all equations are excluded; standard errors are robust for single-equation estimators (OLS and 2SLS) only; 'education' refers to total years of schooling; 'life expectancy' is at birth.

Source: authors' calculations; see Appendix 4C for variable definitions and sources.

4C Appendix: summary statistics and variable sources

The table below summarises the variables used in the analysis, the measurement scale employed and the original data sources (with source-variable reference code where available). Please see the notes at the end of the table for further details.

	N	Median	Mean	St. dev	Scale	Source	Reference code
Aid variables							
Aid / GDP	78	2.28	3.49	3.77	[A]	[1]	-
Aid per capita	78	26.64	36.78	44.74	[A]	[1]	-
Generated aid instrument	78	62.52	62.26	12.13	[A]	[2]	-
Final outcomes							
Real GDP growth per capita	78	1.68	1.73	1.79	[A]	[3]	rgdpchg
Agriculture, value added (% GDP)	76	20.65	22.59	13.42	[A]	[4]	NV.AGR.TOTL.ZS
Industry, value added (% GDP)	76	29.63	29.71	9.90	[A]	[4]	NV.IND.TOTL.ZS
Services, etc., value added (% GDP)	76	48.72	47.71	9.66	[A]	[4]	NV.SRV.TETC.ZS
Poverty headcount at $2 a day	64	43.30	45.39	29.72	[D]	[4]	SI.POV.2DAY
Poverty headcount at $1.25 a day	64	21.65	28.35	24.03	[D]	[4]	SI.POV.DDAY
Gini index	65	44.19	44.46	7.87	[D]	[4]	SI.POV.GINI
Intermediate outcomes							
Investment in real GDP	78	17.18	18.53	10.30	[A]	[3]	ki
Private consumption in real GDP	78	68.36	70.02	19.95	[A]	[3]	kc
Government consumption in real GDP	78	16.60	18.04	8.41	[A]	[3]	kg
Revenue, excluding grants (% GDP)	69	19.95	21.66	8.82	[A]	[4]	GC.REV.XGRT.GD.ZS
Health expend., public (% GDP)	78	2.67	2.89	1.39	[A]	[4]	SH.XPD.PUBL.ZS
Education expend., public (% GDP)	76	3.76	3.97	1.50	[A]	[4]	SE.XPD.TOTL.GD.ZS
Military expenditure (% GDP)	77	1.91	2.34	1.70	[A]	[4]	MS.MIL.XPND.GD.ZS
Life expectancy at birth, total (years)	78	60.96	59.00	9.50	[B]	[4]	SP.DYN.LE00.IN
Infant mortality rate	75	67.48	71.78	35.38	[B]	[4]	SP.DYN.IMRT.IN
Death rate, crude (per 1,000 people)	78	11.05	11.92	4.62	[B]	[4]	SP.DYN.CDRT.IN

continued overleaf ...

	N	Median	Mean	St. dev	Scale	Source	Reference code
Intermediate outcomes (contd.)							
Fertility rate (births / woman)	77	4.66	4.68	1.38	[B]	[4]	SP.DYN.TFRT.IN
Consumer price inflation (%)	77	10.31	52.69	150.30	[A]	[4]	FP.CPI.TOTL.ZG
Real interest rate (%)	77	6.58	7.00	8.78	[A]	[4]	FR.INR.RINR
Domestic credit to private sector (% GDP)	78	23.74	29.37	21.40	[A]	[4]	FS.AST.PRVT.GD.ZS
Av. years total schooling, 15+	72	4.89	5.01	2.10	[B]	[5]	BAR.SCHL.15UP
Av. years primary schooling, 15+	72	3.64	3.49	1.42	[B]	[5]	BAR.PRM.SCHL.15UP
Av. years secondary schooling, 15+	72	1.35	1.37	0.72	[B]	[5]	BAR.SEC.SCHL.15UP
Control variables							
Income per capita	78	7.88	7.84	0.79	[C]	[6]	-
Sachs-Warner trade policy index	78	0.32	0.32	0.29	[C]	[6]	-
Life expectancy	78	51.99	52.88	9.77	[C]	[6]	-
Geography	78	-1.00	-0.55	0.77	[C]	[6]	-
Ethnic fractionalization	78	0.54	0.47	0.29	[C]	[6]	-
Primary eduction enrolment rate	78	0.67	0.65	0.29	[C]	[7]	p60
Coastal population density	78	30.36	101.25	358.82	[C]	[7]	dens65c
Malaria prevalence index	78	0.54	0.51	0.43	[C]	[7]	sa_mr
Price of investment goods	78	85.83	93.68	62.20	[C]	[7]	iprice1
Civil liberties	78	0.33	0.41	0.27	[C]	[7]	civ72
Air distance to major cities	78	8.47	8.41	0.50	[C]	[7]	airdist
Oil producer	78	0.00	0.36	0.48	[C]	[7]	oildummy
Landlocked	78	0.00	0.21	0.41	[C]	[7]	landlock
Prevalence of HIV (% of pop 15-49)	68	0.80	3.12	5.70	[D]	[4]	SH.DYN.AIDS.ZS

Scales: [A] full period mean (1970-2007); [B] average of earliest start and latest end values; [C] initial value (1960s or early 1970s if the former unavailable); [D] latest end value only;

Sources: [1] authors' calculations from OECD-DAC (www.oecd.org/dac/stats/idsonline; downloaded May 2009); [2] authors' estimates based on the method set out in Arndt et al. (2010b,a), using updated and cleaned OECD-DAC dataset; [3] Penn World Tables v6.3 (http://pwt.econ.upenn.edu) [4] World Bank, World Development Indicators and Global Development Finance (http://data.worldbank.org/data-catalog; downloaded 20-12-2010); [5] World Bank, Education Statistics (http://data.worldbank.org/data-catalog; downloaded 20-12-2010); [6] Arndt et al. (2010b,a); [7] Sala-i-Martin et al. (2004).

Chapter 5

The economics of smallholder organic contract farming in tropical Africa: the case of Ugandan coffee[†]

5.1 Introduction

As the market for certified organic agricultural products has grown over recent years, organic activists, NGOs and some donors have promoted certified organic export production in a number of tropical African countries. This chapter is a preliminary assessment of the resulting schemes, focusing broadly on the revenue effects of participation relative to conventional farming. While this subject is well covered in the literature on organic farming in Northern countries, where conclusions converge on a finding of similar levels of profitability for the two farming systems (since price premiums and lower non-labour input costs compensate for organic agriculture's normally lower yields)[1], none of the handful of existing economic studies of organic farming in the tropics (Damiani, 2002; Lyngbaek et al., 2001; Bray et al., 2002; Carpenter, 2003; Bacon, 2005; van der Vossen, 2005) report comprehensive farm budget-related survey data.

Assessment of the revenue effects of certified organic relative to conventional (non-organic) farming in tropical Africa has to take into account two differences between farming systems there and in developed countries. Firstly, conventional agriculture in developed countries is industrial in character while that in tropical Africa is generally semi-industrial or non-industrial. For example, fertilizer consumption levels in tropical Africa are a fraction of those in other developing regions, and are falling.[2] This has implications both for changes in farmers' outlays on synthetic inputs, and for changes in yields, when conversion takes place. Also it has implications for the extent to which farmers in tropical Africa who are certified to organic standards have to adopt a new set of farming practices in order to maintain soil fertility and thus remain economically viable, as they have to in developed countries.

[†]This chapter is an edited reproduction of Bolwig et al. (2009). It is co-authored with Simon Bolwig and Peter Gibbon.

[1]For recent overviews see Nieberg and Offerman (2003) for Europe, and Dimitri et al. (2007) for the US.

[2]The World Bank's World Development Indicators for 2006 give Sub-Saharan African fertilizer consumption of 12.3 kg./hectare for 2002-03, as against 106.6 kg. for South Asia and 89.5 kg. for Latin America. Sub-Saharan Africa fertilizer consumption in 1989-91 was 14.2 kg./hectare.

Secondly, the institutional context for both conventional but organic agriculture in developed countries is deeper and more extensive than in tropical Africa. This means that in Africa little or no public assistance is available for conversion, while private credit and domestic savings are generally too low to support independent conversion – implying that organic farming in tropical Africa is a realistic option only for very large-scale operators or in the context of privately financed and coordinated contract farming schemes.[3] However, participants in such schemes may be selected by scheme owners rather than self-recruited. In other words, transposing a participation effects focus to tropical Africa requires close attention to variables confounding the independent effects of adoption of organic agricultural systems, including the prevalence of non-(certified) organic farming systems that are 'organic by default', as well as organization of smallholder certified organic agriculture in contract farming schemes.

Recent years have also seen an increase in interest in all types of contract farming arrangements for smallholders in tropical Africa, in a context of rising concern that African smallholders are being excluded from remunerative value chains, whether these are for exports or for higher-value products sold on domestic markets (Hazell et al., 2007; Reardon and Berdegué, 2007). Contract farming is seen as a solution to problems such as declining public investment and private market failure said to underlay exclusion, since it increases economies of scale and thereby reduces private traders' transaction costs (Simmons, 2003; Warning and Key, 2002; Poulton et al., 2004; Dorward et al., 2004). On the other hand, Little and Watts (1994) and Havnevik et al. (2007) challenge whether contract farming schemes generate sustainable income benefits for participants,[4] while others claim that they increase rural inequalities since – also in pursuit of lower transaction costs – it is typically only better-off smallholders that are recruited to them (Key and Runsten, 1999).

In assessing the extent to which certified organic contract farming schemes have positive revenue effects for smallholders, a question pertinent to the evaluation of both organic agriculture and contract farming as possible routes out of Africa's well-advertised problem of agricultural stagnation and decline,[5] we explicitly take into account the problem of non-random selection into schemes. This entails controlling for the possibility that observed positive revenue differences between participants and non-participants will reflect differences in farmers' factor endowments or abilities, rather than the unique impact of participation itself.[6] A second research question concerns the unique contribution of organic farming methods, as opposed to scheme participation as such, to any positive revenue benefits that are found.

These questions are examined with survey data collected in Uganda in 2006 from participants in Kawacom (U) Ltd.'s Sipi organic coffee contract farming scheme, as well as from a control group of non-organic coffee smallholders in the same area, using a standard OLS regression and a full information maximum likelihood estimate of the Heckman selection model. A limiting factor for the validity of findings on the second research question is the relatively short period since the scheme's certification (in 2000-01). This means that smallholder rates of adoption, and experience in using recommended organic farming methods, is likely to be restricted. Therefore, their full potential benefits arguably remain to be seen.

[3] In tropical Africa, organically certified large-scale commercial farming is found in Kenya (Gibbon and Bolwig, 2007, p. 25-38), Zambia (Parrott et al., 2003, p. 110), and Gambia. In all cases production is for fresh vegetables for the UK market. In 2007 the total number of farms involved was not more than five or six.

[4] Little and Watts (1994) argue that as smallholders alter their cropping patterns and invest in specific assets to optimize their benefits from schemes, their negotiating power declines relative to scheme owners.

[5] For recent discussions of the extent and basis of African agriculture's crisis – and solutions to it – see *inter alia* World Bank (2007), Havnevik et al. (2007) and Koning (2002).

[6] See Warning and Key (2002) and Benfica et al. (2006) for a parallel research question in relation to contract farming schemes *per se*.

The remainder of this chapter is organized in four sections. Section 5.2 describes the Sipi scheme and its context in greater detail, and provides descriptive statistics on its participants relative to the control group. Section 5.3 describes the data collection and analysis methods used. Section 5.4 presents the empirical analysis of the two hypotheses and discusses the results. Section 5.5 concludes.

5.2 The Sipi organic arabica scheme and its context

Uganda is one of the two leading exporters of certified organic produce by value in tropical Africa (the other being Kenya). At the time of the survey there were between 20 and 25 certified organic exporters, while total organic exports were worth just under $7 million annually (Gibbon, 2006). Organic exports were dominated by the traditional cash crops, led by coffee, and were overwhelmingly to the European market. With a few exceptions, all organic export operations were organized as contract farming schemes. Most such schemes were supported to different degrees by one or more donor.[7]

Coffee is central to Uganda's rural economy, with an estimated 350,000 smallholder producers. Coffee also has been Uganda's single most important export good since the late 1960s, but its relative importance has declined over the last decade due to falling prices, wilt disease affecting robusta coffee, and the rise of non-traditional exports.[8] Production peaked at 254,000 tons in 1996-97, but has since 2004 been oscillating around 150,000 tons (equivalent to 3% of global output). Uganda produces robusta and mild arabica coffees, and a little Hard arabica. Arabica accounts for about 20% of coffee exports. Against the overall trend, arabica production increased from 15,000 tons in 1993 when the sector was liberalized to 35,000 tons in 2007, with troughs in 1998-1999 and 2000-02. Growth was stimulated by rising prices during 1994-95, 1997-98 and 2004-08, while the major price slump of 1999-2003 (the global coffee crisis) led to a temporary decline in output.

Quality is a key competitive factor in the international market for Mild arabica. The most important coffee bean quality attributes are physical defects and cup defects (undesired taste characteristics), which are affected mainly by processing and handling. The quality of Ugandan coffee deteriorated rapidly in the first few years after liberalization in the early 1990s as exporters rushed to establish market share through aggressive procurement practices, *inter alia* through buying unripe or poorly processed beans (Ponte, 2002). Similar practices were observed in the 2005 and 2006 seasons when rising export prices following a period of production decline caused a new scramble for coffee.[9]

Since the late 1990s, some Ugandan exporters have tried alternative modes of coffee procurement to the predominant open market one that relies on several layers of middlemen. The resulting schemes often involve certification to various sustainability standards (organic, Utz Kapeh, Fair Trade and proprietary) as well as allowing other forms of product differentiation such as bean quality and geographical origin. The central motive for their establishment was to protect trading margins during the coffee crisis. A facilitating factor has been the availability of donor support. The schemes often resemble contract farming in their design, although with sometimes low levels of commitment on the part of both buyer and farmers.

[7] The Export Promotion of Organic Products from Africa (EPOPA) programme, funded by Sida, is the most important contributor to the development of the sector. It was supporting 18 organic exporters in 2007, including Kawacom.

[8] In 2006, coffee made up 19.5% of total export value. During the coffee boom in 1995 this figure was 67% (Bank of Uganda, 2006; Uganda Coffee Development Authority, 2003).

[9] Coffee procurement in Uganda is quite competitive. There were 24 exporters buying coffee in the 2006/07 season, with the top five accounting for 65% of total purchases (Uganda Coffee Development Authority, 2007).

One of the earliest and largest is the Kawacom Sipi Organic arabica scheme.

The Sipi scheme, is operated by Kawacom (U) Ltd, a subsidiary of the international commodity trading house Ecom Agroindustrial Corporation.[10] Kawacom is the third largest exporter of conventional coffee from Uganda (Uganda Coffee Development Authority, 2007) and the biggest exporter of organic coffee. The scheme is situated on the northern slopes of Mount Elgon in Kapchorwa District in eastern Uganda. Farms are situated in a contiguous area at 1650-2150 m.a.s.l. The area was chosen due to its favorable agro-climatic conditions and because the then dominant buyer in the region, Bugisu Cooperative Union, had only a weak presence. Mobile phone network coverage was established in 2000 and in 2003 a new tarmac road significantly improved accessibility. This eased procurement but also intensified competition for organic coffee from other traders.

The project encompassed 3,870 organic farmers in 2005, most of who were registered and certified in 2000-01. Except for location in the scheme area, there were no barriers to entry. Registration is free, and as a result it encompasses 62% of all households in the area. Organic certification is to both the EU and US standards and is paid for by Kawacom.[11] A group certification system is used, based on an elaborate internal control system (ICS). The central component of the ICS is an annual or semi-annual farm inspection performed by locally-recruited company field officers. The latter have been trained in organic farming methods, they run demonstration farms and they conduct occasional training. The field officers give technical advice during the farm inspections and monitor the performance of each farmer in terms of his/her compliance to the organic standards and other project requirements. Very few farmers have been evicted from the project on account of non compliance, however. The annual third party certification consists of reviewing ICS documentation as well as visits to selected farms and collection points.

Project farmers are required to follow certain production and on-farm processing practices, most of which are specified in a contract issued to each farmer by Kawacom at the time of registration. The practices are those necessary to conform to organic standards and others known to improve the physical quality of coffee beans in terms of size, moisture content, appearance and aroma. In addition, the technical advice disseminated emphasizes farm practices – mainly but not exclusively organic – that should enhance yield per area unit. Kawacom purchases only dry parchment coffee from scheme farmers, i.e. beans whose pulp has been removed through wet-processing (hand pulping) and subsequent fermentation and sun drying. The most common reasons for rejecting coffee are excessive moisture and foreign matter content. In such cases the farmer can reprocess the coffee or sell it off-scheme. In rare cases is coffee rejected on suspicion that it was harvested on non-certified farms.

In 2005 Kawacom procured 715 tons of organic coffee from the scheme, or 198 kg. per farmer on average. The coffee is purchased at designated collection points and stored for later transportation to a factory in Kampala where it is further processed and graded for export. The farmer is paid cash on delivery. Kawacom buys all the coffee offered for sale by its organic farmers during the main buying season, irrespective of the size of its organic orders. Any surplus is sold as conventional. Prices are communicated daily by mobile phone through the network of field staff and contact farmers. The contract obliges Kawacom to pay an organic premium if the coffee is 'of suitable quality'. The size of the premium is not specified and there has been no direct price negotiation between Kawacom and the farmers. In 2005 Kawacom paid a price premium of about USH 300, or 15% above the prevailing price in the Mount

[10]Historically, Ecom traded coffee internationally under the name Esteve.

[11]In 2003 the scheme was also certified to the Utz Kapeh standard, but this did not entail significant changes in grower requirements.

Elgon area. This premium reflects both an organic premium realized at the export level, the higher quality of organic coffee, and price competition from other traders operating in the scheme area.[12]

In summary, Kawacom employs various means to enable and induce growers to comply with its organic and quality standards: regular farm inspections, group training and individual advice, input provision (on a very limited scale), a policy of rejection of sub-standard and suspected off-scheme coffee, a price premium, and a procedure for de-registering farmers who consistently or grossly violate project standards and rules.

Table 5.1 compares the mean values of selected variables, for a sample of 112 scheme participants and a control group of 48 non-participant farmers in the same area (see Section 5.3 below). The two groups differ in their endowments of key production factors. Scheme participants operate larger farms, cultivate more coffee trees, and have larger family labor endowments (proxied by household size). A higher proportion of scheme participants also have farming as their primary occupation (in terms of time spent) than is the case for the control group.

As may be expected, a significantly larger proportion of scheme participants use organic practices for coffee farming, although one-fifth of them do not apply any. While no significant difference is found in the total revenue earned by certified and non-certified households, certified farmers earn higher revenue from the sale of coffee and from all crop sales, and their net coffee revenue is about three times greater than for the non-certified farmers. All data refers to 2005.

5.3 Methodology

5.3.1 Data collection and methods

The household survey of certified organic farmers and non-certified farmers used a questionnaire administered to heads of households by trained enumerators. It covered information on household demographics, farm area, number of coffee trees, farm equipment, expenditure over the previous two seasons on labor and other inputs and assets and on processing and marketing, as well as production, sales, and farm and non-farm income. As is common in farm budget-related surveys in Africa, no attempt was made to collect data on family labor inputs. This is because subjects typically find it more difficult to recall such inputs relative to hired labor ones, because of the difficulty in attributing accurate time values to some family labor tasks such as supervision, and because of difficulty in applying a common metric to labor by children and by adults.

In order to assess the extent to which organic and other farm practices were adopted and/or enforced as a result of contracts, data also was collected on farmers' use of a range of farm practices recommended during inspections and training, in most cases through physical observation by the enumerator. Organic practices were operationalized in terms of a range of positive farming interventions. Non-use of synthetic inputs, the central regulatory requirement for organic certification, was treated as a condition qualifying such positive interventions to be recorded, rather than as an organic practice in itself. In other words,

[12]Non-organic traders in the project area are willing to pay a premium for organic coffee due its superior physical quality. This also means that the farm-gate price within the project area is a little higher than in other parts of the area, to the benefit of local non-organic farmers and organic farmers selling off-scheme.

Table 5.1: Descriptive statistics for certified organic farmers and non-organic farmers

Variable	Unit	Certified organic	Non-organic	χ^2 / z stat.
(a) Sample characteristics:				
Respondents	Count	112	48	-
No use of organic practices	% group	20.5	39.6	6.3*
Use of >2 organic practices	% group	33.9	12.5	7.74**
Farming as primary occupation	% group	83.0	75.0	1.39**
(b) Household characteristics (means):				
Whole farm size	1000 m^2	10.8	7.9	-2.5*
Productive coffee trees	Count	650.1	308.1	-4.0**
Coffee farm altitude	1000 m.a.s.l.	1.9	1.8	-3.6**
Age of household head	Years	46.3	47.1	0.0
Education of household head	Years	6.9	6.6	-0.3
Household size	Count	7.2	6.2	-2.3*
Dependents (< 6 years)	% household	20.0	20.0	0.1
(c) Household revenue (means):				
Total household revenue	1000 USH	1424.6	1235.5	-1.6
Total crop revenue	1000 USH	679.6	374.1	-2.4*
Total non-crop revenue	1000 USH	655.7	852.2	1.0
Total coffee revenue	1000 USH	566.0	176.7	-6.0**
Net coffee revenue	1000 USH	518.8	154.5	-6.2**
(d) Variable coffee production & processing costs (means):				
Sales expenses (transport)	1000 USH	0.9	0.1	-3.6**
Hired labor (food, wages)	1000 USH	32.6	18.4	-2.4*
Equipment and inputs	1000 USH	13.7	7.2	-3.3**

significance level: * 5%, ** 1%

Notes: 'organic practices' excludes the non-use of synthetic inputs (see section 5.3); total household revenue calculated as the sum of gross crop revenue, revenue from the sale of livestock less livestock purchases, and income earned in off-farm activities; non-crop revenue is the sum of the latter two; net coffee revenue is total coffee revenue minus all costs given under group (d); family labor inputs and land purchases are not included in the calculation of costs; m.a.s.l. are metres above sea level; USH are Ugandan Shillings (US$ 1 = USH 1777 as at 2005); for group (a), significance tests report χ^2 statistic from cross-tabulation; for groups (b) to (d) significance tests report the z-stat. from a Wilcoxon rank-sum test.

Source: own analysis based on Sipi scheme survey (§5.3.1).

the handful of farmers found to have used synthetic inputs were excluded from consideration in terms of having followed organic practices, whether they in fact followed such practices or not. The positive interventions considered were: use of organic pesticides, mulching, animal manure, and composting (including mixing leguminous residues into the soil). This group of interventions, rather than others that may be counted as organic such as intercropping, use of trap crops etc., were selected for consideration for four reasons. Firstly, they had been promoted by the scheme from its inception, unlike e.g., planting agroforestry trees, which was a recent addition. Secondly, their use was not contingent on physical or similar factors, unlike the use of soil erosion measures that depend on slope. Thirdly, they were not applied for reasons unrelated to organic coffee production, unlike intercropping with food crops including leguminous cover crops, which is a traditional food security strategy. Fourthly, there had to be enough observations of them to render analysis meaningful.

A two-stage random sampling method was used for selection of both scheme participants and the control group. Scheme participants were randomly sampled in a number of parishes chosen purposively to reflect the range of agro-ecological conditions in the scheme area, using a list of registered farmers provided by Kawacom. Sampling of the control group population was performed randomly, from lists prepared by village leaders in nearby parishes chosen to match the (range of) agro-ecological conditions represented in the sampling frame for scheme participants.

5.3.2 Analytical methods

For empirical analysis, two specific null hypotheses can be formalized. These are: *Hypothesis I* – there is no significant difference in revenue between certified organic and non-certified farmers, controlling for other relevant determinants; and *Hypothesis II* – there is no significant revenue effect from application of organic farming practices, controlling for participation in the organic contract farming scheme and other relevant factors. Together these indicate we are concerned with evaluating the effects of different farming activities on household revenue. If we conceive of these activities as forms of intervention (analogous to, say, a labor training programme), it is evident we face a treatment evaluation problem.

The literature dealing with how treatments can be rigorously estimated is vast and cannot be reviewed here (e.g., Blundell and Dias, 2002; Heckman et al., 1999; Vella and Verbeek, 1999). To provide an organizing framework for discussion, however, the evaluation problem can be stated as a system of equations involving an outcome of interest (y) and a selection equation for treatment (t) over observations i. In general form these are:

$$y_{1i} = \alpha_1 + x_i'\beta_1 + u_{1i} \tag{5.1}$$

$$y_{0i} = \alpha_0 + x_i'\beta_0 + u_{0i} \tag{5.2}$$

$$t_i = \mathcal{I}(x_i'\theta + z_i'\gamma + v_i > 0) \tag{5.3}$$

where y_{1i}refers to the outcome for treated respondents ($t_i = 1$) and y_{0i} for the control group ($t_i = 0$); u_{ki} and v_i are the error terms. Note that the participation equation (5.3), which is an indicator function, invokes a latent variable framework in which the vector of selection variables (x, z) capture the propensity to participate in the scheme above a threshold. This can be summarized in the following general switching model:

$$y_i = \alpha_0 + x_i'\beta_0 + t_i(\alpha_1 - \alpha_0) + t_i x_i'(\beta_1 - \beta_0) + u_{0i} + t_i(u_{1i} - u_{0i}) \tag{5.4}$$

Differences in regime between participants and non-participants refer to the extent to which the treatment has an effect only through the intercept of the joint outcome equation. The assumption of non-distinct regimes is made frequently (Ravallion, 2005), and is reasonable in this case given that the scheme is relatively recent and there is substantial similarity between the control and treatment groups as regards demographic characteristics and location (see Table 5.1). Consequently, in equation (5.4) we restrict $\beta_1 = \beta_2 = \beta$, thus giving the familiar reduced form common coefficient model for outcomes over a single treatment:

$$y_i = \alpha_0 + x_i'\beta + t_i\delta + u_{0i} + t_i(u_{1i} - u_{0i}) \tag{5.5}$$

where $\delta = \alpha_1 - \alpha_0$ captures the treatment effect given by the difference in intercepts of equations (5.1) and (5.2).

As the present setting is non-experimental we cannot assume that the choice to participate in the scheme is purely random. Given the available data, which does not include repeated measures for each household, three main types of estimator can be employed to deal with endogenous selection. The first of these is matching estimators, which require (*inter alia*) that selection into the programme occurs only on observed variables – i.e., $\mathrm{E}\left[u_{ki} \mid x_i, t_i\right] = 0$. In such cases, propensity score matching can be used; or, where parametric assumptions apply, a standard OLS regression of the form given by equation (5.5) is consistent as long as all relevant selection variables are included as regressors.

If the assumption of participation on observables is doubted, then matching methods will be biased and either instrumental variable (IV) or Heckman selection models (Heckman, 1979) are appropriate. It is generally recognized that the latter are more robust, particularly for small samples. Nonetheless, they are sensitive to model specification and distributional assumptions (Blundell and Dias, 2000; Heckman et al., 1999). As a result, it is recommended to include in equation (5.3) variables that do not enter the outcome equation, denoted by the vector z. Effectively, this amounts to placing exclusion restrictions on x, analogous to an IV identification strategy. Tests for heteroscedasticity and collinearity between the selection and outcome equations are advised in order to check for deviations from the underlying assumptions required for consistency and robust inference. Following Puhani (2000), the collinearity test applied here uses the inverse Mills ratio (Heckman's lambda) estimated from the FIML model as the dependent variable in an OLS regression against the structural variables in the outcome equation. A large F-statistic (a high R^2) would then indicate the presence of significant collinearity.

The question of heterogeneous treatment concerns how the treatment effect is specified. A standard approach is to assume homogeneous effects, implicitly treating individual deviations from the average effect as white noise not correlated with the participation decision (conditional on x). Allowing for correlation between the participation decision and the individual treatment effect adds substantial complexity to the analysis and interpretation of results. Given this chapter's modest aims, simplicity is paramount and a homogeneous framework is assumed. Even so, it is important to note that Heckman selection estimators remain consistent under the assumption of heterogeneous effects, while IV estimators are invalidated (Blundell and Dias, 2000).

The above discussion indicates that unless selection on observables can be guaranteed, a Heckman model would be most appropriate. As there is no prior reason to discard the possibility of unobserved selection factors, the hypotheses are investigated via a simple OLS specification as well as a full information maximum likelihood (FIML) estimation of the Heckman model. The latter incorporates a test for sample selection bias, indicating whether OLS results may be biased. The FIML method differs from the (original) two-step estimation approach as the selection and outcome equations are estimated jointly, thereby enhancing asymptotic efficiency (Puhani, 2000). To check for robustness we also report results from alternative estimators, as well as relevant misspecification tests.

Before describing the empirical specification, it is necessary to reflect on the nature of the 'treatments'. As noted previously, two effects need to be distinguished: (i) participation in the organic scheme; and (ii) use of organic farming techniques. Although Table 5.1 confirms that farmers in the control group also employ some organic techniques, it also suggests that certification is a strong predictor of the number of techniques used. For example, approximately 34% of certified farmers use two or more organic techniques compared to only 13% of the control group. As a consequence, we consider participation in the scheme to be the main potential source of selection bias. The number of organic practices used is taken to be a second-order decision that is not subject to selectivity bias once we have controlled for scheme participation and other household characteristics. The validity of this approach is investigated empirically in Section 5.4 through an analysis of the determinants of the two treatments.

In terms of empirical implementation, the OLS estimates are based on equation (5.5) where the vector x defines a parsimonious set of structural regressors affecting both the outcome and the participation decision. These are: whole farm size (log.), number of productive coffee trees (log.), altitude above sea level, age of the head of household, his/her education (in number of years) and the size of the household. Both hypotheses are tested simultaneously by including the two treatments as additional regressors; these are: a dummy variable for participation in the scheme (C) and the number of organic practices used (P). In all specifications the dependent variables (y) refer to the logarithm of components of household revenue. For ease of exposition we focus only on gross crop revenue and net coffee revenue as described in Section 5.4; however, the results from alternative dependent variables are comparable and support the overall analysis.[13]

For the FIML selection model, equations (5.3) and (5.5) are estimated simultaneously (for details see Greene, 2002). For the exogenous predictors of participation (vector z), which do not enter the outcome equation, we use two dummy variables that proxy for the orientation of the household towards agriculture as well as its long-term welfare status. The former (non-crop) is constructed from the ratio of non-farm revenue to total revenue, taking the value of one for those falling in the top tercile and zero otherwise. This indicator is strongly associated with comparable indicators such as households stating their primary occupation as agriculture and those in receipt of a salary income. Thus, we interpret it to be capturing the 'deep' structure of household revenue generation. The welfare indicator (walls) takes a value of one if the walls of the household are made of brick and zero otherwise. Once again, it is assumed that this variable changes only slowly and therefore is exogenous to the outcome variable(s) over the measurement period.

5.4 Empirical findings

5.4.1 Hypothesis tests

Moving to the results, it is useful to review whether concerns regarding endogenous selection are warranted. Table 5.2 reveals the extent to which the observed levels of the two treatment variables can be attributed to the structural regressors. Scheme participation (certification, C) is modeled using a binomial probit estimator. The results show that specific household endowments relating to coffee production, farm altitude and the 'instruments' (z) are significant predictors of participation. At a minimum, this suggests

[13] These are available on request from the authors.

that certification is non-random and underlines the relevance of techniques that account for endogenous selection. The significance of the two exclusion restriction variables (non-crop, walls) also supports the feasibility of employing IV and Heckman selection methods. Use of organic practices (P) is a count variable and therefore is modelled by a Poisson regression. In addition to the variables entering vector x (see above), the specification also conditions on scheme participation and the number of inspections received from scheme managers (due to a potential training effect). The results show that only scheme participation (C) is a material partial correlate of the number of techniques used. Moreover, the overall model has weak explanatory power as indicated by the insignificance of the χ^2 and pseudo-R^2 measures. In sum, these results support the chosen empirical strategy in which scheme participation is considered to be the primary potential source of selection bias, while use of organic practices is conditionally random.

Table 5.2: Probit model for scheme participation and Poisson model for use of organic practices

Dep. variable	Certified organic (C)		Organic practices (P)	
	beta	s.e.	beta	s.e.
Whole farm size	0.06	(0.19)	0.02	(0.09)
Trees (no.)	0.39*	(0.16)	-0.08	(0.08)
Altitude	5.79**	(1.73)	-0.10	-0.68
Age	-0.02$^+$	(0.01)	0.01	(0.00)
Education	0.0	(0.04)	0.01	(0.02)
Household size	0.07	(0.06)	0.01	(0.03)
Non-crop	-0.63*	(0.27)	-0.16	(0.15)
Walls	0.47$^+$	(0.26)	0.01	(0.12)
Certified (C)	-		0.43*	(0.19)
Inspections	-		-0.02	(0.07)
Constant	-12.42**	(3.53)	0.44	(1.42)
N	147		147	
Log-like.	-66.7		-231.6	
R-sq. (pseudo)	0.23		0.02	
Chi-sq.	40.7**		9.9	

significance level: ** 1%, * 5%, $^+$ 10%

Notes: variables and models are as discussed in the text; standard errors (s.e.) are robust (Huber/White/Sandwich); samples exclude missing observations and outliers, defined as households with net coffee revenue ±3.5 standard deviations from the mean.

Source: own analysis based on Sipi scheme survey (§5.3.1).

Results for the models encompassing the two main hypotheses are set out in Table 5.3. Estimates for both the OLS and FIML estimators are reported for the dependent variables of interest. Four main results can be highlighted. The first is the strong goodness-of-fit of all models, given by the relatively high R^2 and χ^2 statistics, which are significant at the 1% level. Secondly, the estimated coefficients are highly comparable across the different models and estimators, also running in the expected directions. For example, size of the farm and number of productive trees are both significantly and positively associated with gross crop revenue; however, whole farm size is not associated with net coffee revenue once more specific characteristics of the household's coffee endowment are controlled for. The results of the selection equations (part (ii) of the table) are also consistent with the results from the individual probit model

Table 5.3: Regression results for effect of certification and organic practices on agricultural revenue

Dep. variable →	Gross crop revenue (log.)				Net coffee revenue (log.)			
Estimator →	(I) OLS		(II) FIML		(III) OLS		(IV) FIML	
Model ↓	beta	s.e.	beta	s.e.	beta	s.e.	beta	s.e.
(i) Outcome eq.:								
Whole farm size	0.35**	(0.09)	0.34**	(0.08)	0.12	(0.09)	0.11	(0.09)
Trees (no.)	0.52**	(0.06)	0.47**	(0.07)	0.71**	(0.08)	0.65**	(0.08)
Altitude	-1.63*	(0.66)	-2.21**	(0.76)	-2.27**	(0.70)	-3.02**	(0.80)
Age	-0.00	(0.00)	-0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Education	0.03	(0.02)	0.03+	(0.02)	0.02	(0.02)	0.02	(0.02)
Household size	-0.01	(0.02)	-0.02	(0.02)	-0.02	(0.02)	-0.02	(0.02)
Practices (P)	0.05	(0.05)	0.05	(0.05)	0.09+	(0.05)	0.09+	(0.05)
Certified (C)	0.38*	(0.15)	0.78**	(0.30)	0.78**	(0.16)	1.31**	(0.27)
Constant	9.50**	(1.36)	10.64**	(1.56)	10.85**	(1.46)	12.30**	(1.69)
(ii) Selection eq.:								
Whole farm size			0.04	(0.21)			0.05	(0.20)
Trees (no.)			0.44*	(0.17)			0.41*	(0.16)
Altitude			5.58**	(1.49)			5.58**	(1.39)
Age			-0.02+	(0.01)			-0.02*	(0.01)
Education			-0.01	(0.04)			0.00	(0.04)
Household size			0.04	(0.06)			0.05	(0.06)
Non-crop			-0.67**	(0.26)			-0.77**	(0.24)
Walls			0.50*	(0.25)			0.49*	(0.25)
Constant			-11.89**	(3.27)			-11.72**	(3.15)
rho (adj.)			-0.4	(0.24)			-0.49*	(0.19)
N	132		132		147		147	
Log-like.	-130.4		-194.3		-160.4		-225.4	
R^2	0.60		-		0.61		-	
F / χ^2-stat.	32.71**		249.8**		24.66**		205.8**	
Collinearity test	-		0.85		-		0.82	
Hetero. test	2.86		2.35		9.30		10.77+	

significance level: ** 1%, * 5%, + 10%

Notes: variables and models are as described in the text; collinearity test reports the F-statistic from a regression of Heckman's lambda against structural regressors in (i); heteroscedasticity test reports the χ^2 statistic from a Breusch-Pagan (LM) test also against structural regressors in (i); robust (Huber/White/Sandwich) standard errors are given; samples exclude missing observations and outliers, defined as households ±3.5 standard deviations from the mean of the dependent variable.

Source: own analysis based on Sipi scheme survey (§5.3.1).

for scheme participation. Together these findings indicate the models are well specified and are able to explain a substantial proportion of variation in the dependent variables.

The third issue refers to the problem of endogenous selection. This is captured by the adjusted *rho* statistic, which is the hyperbolic arctangent of the correlation (*rho*) between the residuals in the selection and outcome equations. For gross crop revenue, selection bias is not significant and therefore the OLS results are likely to be consistent. In contrast, for net coffee revenue the outcome and selection equations cannot be considered independent (at the 5% level). Thus, moderate selection bias exists and the OLS results may not be reliable. The difference between these findings relates to the fact that the organic scheme does not embrace non-coffee crops.[14] Thus, issues of selection have a narrow domain and may be marginal in the context of each household's overall crop production. With respect to the FIML model for net coffee revenue (column D), the collinearity and heteroscedasticity tests are insignificant at the 5% level. This indicates that the assumptions underpinning the FIML approach are not substantially violated and, therefore, these estimates can be preferred. In any case, robust standard errors are used to address any remaining heteroscedasticity (which cannot be rejected at the 10% level).

The final observation refers to the coefficients on the treatment variables (certification and use of organic practices). The most striking result is a consistent positive significant effect from participation in the scheme. This is observed in both the gross crop revenue and net coffee models. Given the empirical strategy applied, these estimates control for other observed determinants of revenue, the use of organic techniques and any unobserved (latent) selection bias. In other words, we can reject the null of Hypothesis I and conclude there is a positive treatment effect *ceteris paribus*. As to be expected given the semi-log specification, the relative magnitude of the participation effect is larger for net coffee revenue, simply reflecting the point that coffee revenue is only one component of gross crop revenue. The results for the use of organic practices also are relatively clear-cut. With respect to gross crop revenue no significant effect can be found. This implies there is no measurable gain at the level of gross crop revenue from augmenting the use of organic practices holding all other variables constant, including scheme participation. Once again, this is not the case for the specific net coffee revenue component. Here we find a modest positive effect from the use of organic practices, approximating a 9% increase in net coffee revenue for a one unit increase in the number of practices applied. However, this result is only significant at the 10% level, suggesting some additional caution is warranted in interpretation. We discuss the economic significance of these results in subsection (b) below.

Finally, to confirm the robustness of the results, Table 5.4 compares results for alternative estimators using net coffee revenue as the dependent variable. A consistent story emerges across all the estimators. The expectation in Hypothesis I of no effect from participation is rejected strongly; similarly, Hypothesis II can be rejected, but more cautiously. That is, organic practices appear to generate a small positive revenue effect, but this is only observable when net coffee revenue is the dependent variable.

5.4.2 Economic significance

It is all very well finding statistically significant results. But are they plausible from an economic perspective? Economic significance can be evaluated by calculating the size of the estimated treatment effects relative to the counterfactual of no treatment. As per standard practice, we estimate expected

[14] Technically, certification is for farms and therefore covers all crops. But there are no certified organic traders for the non-coffee crops produced in the area.

Table 5.4: Summary of alternative regression results for net coffee revenue (log.)

	(A) OLS		(B) two-stage IV		(C) LIML		(D) FIML	
	beta	s.e.	beta	s.e.	beta	s.e.	beta	s.e.
Whole farm size	0.12	(0.09)	0.11	(0.10)	0.09	(0.11)	0.11	(0.09)
Trees	0.71**	(0.08)	0.59**	(0.10)	0.58**	(0.11)	0.65**	(0.08)
Altitude	-2.27**	(0.70)	-3.87**	(1.10)	-3.84**	(1.13)	-3.02**	(0.80)
Age	0.00	(0.00)	0.00	(0.01)	0.00	(0.01)	0.00	(0.00)
Education	0.02	(0.02)	0.02	(0.02)	0.02	(0.02)	0.02	(0.02)
Household size	-0.02	(0.02)	-0.03	(0.03)	-0.03	(0.03)	-0.02	(0.02)
Practice (P)	0.09+	(0.05)	0.13*	(0.05)	0.09+	(0.05)	0.09+	(0.05)
Certified (C)	0.78**	(0.16)	1.81**	(0.54)	1.89**	(0.54)	1.31**	(0.27)
Constant	10.85**	(1.46)	13.77**	(2.18)	13.91**	(2.28)	12.30**	(1.69)
Selection	-		-		-0.71*	(0.32)	-0.49*	(0.19)
N	147		147		147		147	
Log-like.	-160.4		-167.0		-		-225.4	
R^2	0.61		0.58		-		-	
F/χ^2-stat.	24.66**		23.35**		182.0**		205.8**	

significance level: ** 1%, * 5%, + 10%

Notes: variables and models are as discussed in the text; robust (Huber/White/Sandwich) standard errors are given; LIML refers to limited information maximum likelihood, also known as the Heckman two-step estimator; FIML is the full information version of the same Heckman selection estimator; 'Selection' refers to the inverse Mills ratio for LIML and rho for FIML; in model (B), 'Certified' is treated as endogenous given by fitted probabilities from a first stage probit as per Table 5.2; selection equations not reported; samples exclude missing observations and outliers, defined as households with net coffee revenue $\pm$ 3.5 standard deviations from the mean.

Source: own analysis based on Sipi scheme survey (§5.3.1).

revenue for each household conditional on participation and no participation. The average treatment effect (ATE) is simply the mean difference between these two estimates over all individuals, or formally: ATE $= \mathrm{E}[y_i \mid x_i, t_i = 1] - \mathrm{E}[y_i \mid x_i, t_i = 0]$. Using the FIML results for net coffee revenue, Table 5.5 presents the estimated gain from scheme participation by different sub-groups, whether or not they participated in actuality. The average effect is a revenue increase of USH 170,430 per household, equivalent to a gain of 75% in net coffee revenue relative to the counterfactual of no participation. For those households that actually participated in the scheme, the increase is slightly lower at 67% (versus 96% for the control group), reflecting the higher likelihood that households with larger coffee farming assets become involved.

Table 5.5: Average effect of scheme participation (organic certification) on net coffee revenue, by group

		Actually organic certified?		
		No	Yes	All
No. of organic practices in use	0	153.72 (92.32)	203.16 (61.33)	182.34 (74.38)
	1	65.18 (103.86)	130.21 (63.62)	121.73 (68.87)
	2	153.02 (94.97)	174.08 (71.05)	166.59 (79.56)
	≥3	77.18 (104.03)	207.77 (69.2)	191.45 (73.55)
	All	137.23 (95.71)	182.96 (67.02)	170.43 (74.88)
Median test, over organic (C):			0.14 (pr = 0.71)	
Median test, over practice (P):			6.56 (pr = 0.26)	

Notes: for each group, figures give (mean) expected revenue increment in 1000 USH arising from participation in the scheme versus the counterfactual of no participation holding all other factors constant, including number of organic practices; figures in parentheses report the latter increment as a % of estimated net coffee revenue in the counterfactual scenario; median test reports the relevant chi-sq. statistic.

Source: own analysis based on Sipi scheme survey (§5.3.1).

Obviously, these are substantial effects; however, a focus on point estimates can be misleading. The 95% confidence interval around the coefficient on scheme participation ranges from 0.78 to 1.82 for the FIML model; the comparable OLS interval is 0.46 to 1.10 which (given these coefficients can be interpreted directly) translate into a relative revenue gain from participation ranging from around 60% to 200%. Thus, although the effect of participation is economically significant, undue stress should not be placed on the precision of the results. In any case, the revenue impact is smaller when viewed in terms of gross crop revenue or total household revenue. For example, the estimated (overall) ATE is equivalent to 12% of observed total household revenue for certified farmers or 14% for non-certified farmers. These are credible orders of magnitude and give credence to the overall direction of the results.

With regard to the effect of using organic techniques, the nature of the specification suggests a constant proportional gain in net coffee revenues from each additional technique applied. The estimated 90% confidence interval indicates these effects range from around 1% to 18% of net coffee revenues. This is not unreasonable as an average effect, especially given the relatively crude way in which organic practices have been operationalized. The insignificance of organic practices at the level of gross crop revenue also is comprehensible once we recall net coffee revenue is approximately 66% of gross crop revenue (for all farmers). Thus, given the modest effect of organic practices at the specific level of coffee revenue, one might only expect to see a significant positive effect at a broader level (gross crop revenue) if the benefits of these techniques applied to crops other than coffee alone. However, both the recent establishment of the scheme and the difficulty that farmers face in generalizing the application of certain organic farming techniques to all the crops they cultivate suggests that more general spillovers are yet to be realized.[15] As such, the absence of a more general effect from organic practices is in line with reasonable expectations.

5.4.3 Economic interpretation

Finally, how do these results cohere with what we know about the economics of certified organic smallholder farming in tropical Africa? The existence of a considerable treatment effect accruing purely from participation in the scheme may be explained with reference to the price premiums offered to certified farmers in the context of the workings of the coffee market. As a scheme member, a price premium from selling organic coffee is only available for produce that has been processed. While in the conventional market processed coffee beans also command a price premium, this is subject to the vagaries of the market and usually is lower. Moreover, processing is costly in terms of time, labor and equipment, suggesting that in the conventional market the decision to process represents an investment with uncertain returns. The existence of a price premium for scheme members may act to offset the risks associated with processing and, therefore, is likely to increase the extent to which farmers engage in these value-added activities.

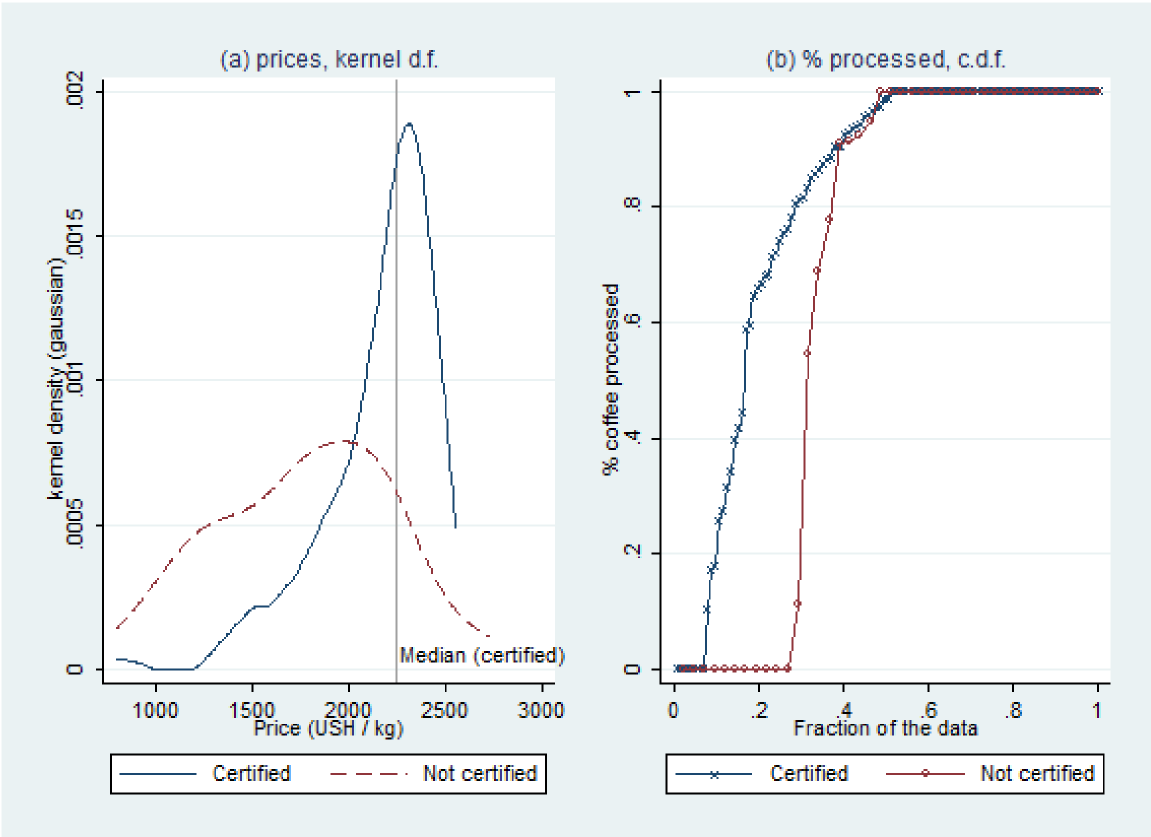
This perspective is substantiated from a review of the distribution of average prices received by scheme participants and the control group as well as the proportion of their coffee crop (fully) processed. Plotted in panel (a) of Figure 5.1, average prices received by farmers who are not certified organic tend to be lower than those received by certified farmers and show much larger variance (dispersion). Approximately only 10% of farmers who are not certified organic receive at least the median price received by certified farmers. The tighter distribution of average prices received by certified famers supports the existence of premium prices that are realized through processing.[16] Panel (b) of Figure 5.1 plots the cumulative distribution of the proportion of the coffee crop fully processed. As can be seen, there appear to be stronger incentives to engage in processing for certified farmers – less than 10% of certified organic farmers process none of their crop compared to over 30% of the control group. One also notes that the same distribution for the control group is extremely disjointed, suggesting a distinct regime shift between engaging in processing and not doing so. Clearly, this is not the case for certified organic farmers as the distribution is much smoother. In addition to the price premium rationale, an additional explanation for this pattern recognizes

[15] Farmers in the scheme area also cultivated plantain bananas, maize and legumes (in order of importance) in addition to coffee. The organic techniques referred to in this chapter are applicable to these crops too, but the volume of organic material available to most farmers meant that its application was restricted to coffee.

[16] Note that due to the complexities of the coffee market, prices available to farmers are not given but rather reflect a number of endogenous choices including the decision to engage in processing. Consequently, we have not used price as an explanatory variable in this analysis.

that the scheme introduced clearer quality criteria and more transparent measurement of quality and volume than in the non-organic market. Once again, this might act to reduce the risks of engaging in processing, thereby increasing the proportion of farmers gaining access to higher prices. Considered in this light, the effect of scheme participation supports arguments to the effect that contract out-grower schemes can help correct for classic market failures in developing country agricultural markets and thus yield positive welfare effects (see Sections 5.1 and 5.5).

Figure 5.1: Distributions of (a) average prices received, and (b) proportion of coffee crop processed for certified organic farmers and non-organic farmers



Notes: panel (a) is a Gaussian kernel probability density estimate of the average prices received by farmers, split between certified and non-certified; panel (b) is the empirical cumulative distribution of the proportion of coffee crop fully processed.

Source: own analysis based on Sipi scheme survey (§5.3.1).

In contexts where alternative farming systems are of an ‘organic by default’ character, any relationship between ‘genuine’ organic practices and revenue can be expected to operate through improved yields (see Section 5.2). This can be examined by employing yield per productive tree rather than revenue as the dependent variable in the same FIML model as before. The results are given in Table 5.6; they show that, controlling for selection bias, scheme membership and other plausible structural determinants of yields, there is a positive marginal effect from the use of organic practices. However, and as can be seen from the standard error for the latter coefficient, it is only significant at the 15% level reflecting both the small sample size and the way in which organic practices have been operationalized. Even so, the magnitude of the effect is analogous to that found when using net coffee revenue as the dependent variable – in this case, each additional organic practice generates a 7% increase in yield per tree on average. This

Table 5.6: Regression results for determinants of coffee yield per tree (log.)

	Selection eq.		Outcome eq.	
	beta	s.e.	beta	s.e.
No. organic practices (P)	-		0.07^{++}	(0.04)
Certified (C)	-		1.01**	(0.23)
Constant	-12.11**	(3.23)	4.22**	(1.46)
Whole farm size	0.03	(0.20)	0.12^{++}	(0.08)
Trees (no.)	0.50**	(0.17)	-0.40**	(0.06)
Altitude	5.55**	(1.41)	-2.51**	(0.69)
Age	-0.02*	(0.01)	0.00	(0.00)
Education	0.00	(0.04)	0.02	(0.01)
Household size	0.05	(0.06)	-0.03	(0.02)
Walls	0.56*	(0.25)	-	
Non-crop income	-0.69**	(0.24)	-	
rho (adj.)	-		-0.53**	(0.19)
N	146			
Log-like.	-200.6			
Chi-sq.	57.5**			

significance level: ** 1%, * 5%, $^{++}$ 15%

Notes: variables and models are as discussed in the text; mode estimated by FIML; standard errors (s.e.) are robust (Huber/White/Sandwich); samples exclude missing observations and outliers, defined as households with net coffee revenue ±3.5 standard deviations from the mean.

Source: own analysis based on Sipi scheme survey (§5.3.1).

result illustrates that the impact of organic techniques on net coffee revenue is likely to occur through improvements in yields and confirms the specificity of the effect for coffee as opposed to other crops.

5.5 Conclusion

We have analyzed the revenue effects of both participation in an organic coffee smallholder contract farming scheme and the application of recognized organic farming techniques. Controlling for a range of factors, including household endowments and non-random selection into the scheme, we find a positive individual effect from both of these activities. Scheme participation (organic certification) is associated with an increase in net coffee revenue of around 75% on average, equivalent to 12.5% of mean (total) household revenue. This is accounted for by the enhanced incentives provided by the scheme to engage in processing of the coffee crop, thereby enabling farmers to access guaranteed price premiums. The effect of applying organic techniques is more modest. We estimate that each additional organic technique used generates a gain equal to around 9% of net coffee revenue, explained by a positive association between these practices and yield per tree. This provides evidence of positive revenue effects arising not only from the scheme itself, but also from the specific application of organic techniques.

Returning to the issues raised in the introduction to this chapter, evidence has been generated in favor of the superior profitability of certified organic farming for smallholders in tropical Africa, relative to the dominant alternative scenario of farming systems that are ‘organic by default’. However, this superiority is bound up with the organization of certified organic production in contract farming schemes. At the same time, evidence has been generated on the conditions under which smallholder contract farming schemes allow for superior farmer profitability. One condition is that they succeed in disseminating low-cost farming techniques that result in higher yields than those obtainable in the default scenario. On the other hand, currently more important in the Kawacom Sipi case, is that they provide product marketing guarantees in relation to receiving a price premium for meeting given quality requirements. This appears to reduce smallholders’ uncertainty about the net returns to processing of the coffee crop. In other words, the evidence presented here also supports the case for contract farming schemes with specific design features, rather than for contract farming schemes as such, as a route out of African agriculture’s stagnation and decline. Of course, the order of importance between these contributing factors may change in the future, as low-cost and effective farming techniques such as organic ones become adopted more widely and deeply.

More generally, the results found here suggest the usefulness of further research in two main areas. The first concerns more detailed work on the economics of organic farming techniques in tropical Africa. Which techniques are most readily adopted, and why? Which generate the highest returns, and why? The other concerns a comparison of the design features of the plethora of new types of smallholder contract farming schemes that are emerging in tropical Africa in response to increased market differentiation in developed countries, particularly in terms of their incentive effects for smallholders.

Chapter 6

Developing agricultural markets in sub-Saharan Africa: organic cocoa in rural Uganda[†]

6.1 Introduction

The development of deep and efficient agricultural markets remains a key challenge across sub-Saharan Africa (hereafter, Africa) (World Bank, 2008). This chapter examines the effects of a specific market intervention, namely an organic cocoa scheme introduced and operated by an exporter. The analysis is based on surveys of cocoa smallholders conducted in the remote Bundibugyo region of western Uganda in 2005 and 2009. The survey design incorporates 'treatment' and 'control' households, corresponding to smallholders from locations eligible and not eligible for organic certification. This enables us to evaluate the welfare impact of the scheme and the corresponding economic drivers. Although we do not have repeated observations on the same households, the repeat dimension of the design allows changes over time to be considered. These include market developments common to all farmers, as well as the persistence of scheme effects.

The contribution of this chapter is threefold. First, we address areas where past research efforts have been thin. With some exceptions (e.g., Minten et al., 2009; Maertens and Swinnen, 2009), neither contract farming in liberalised markets nor organic farming have received much academic attention in Africa. This is despite that conversion to organic status often involves only small changes to farm processes but can enhance access to niche export markets and, hence, to substantial price benefits. Meanwhile, with respect to technology adoption, existing studies mainly focus on the diffusion of specific yield-enhancing as opposed to quality-improving technologies (Doss, 2006). These tend to treat technology adoption decisions as static and binary. Scant attention has been given to incremental improvements in farm practices which, as in the present case, can enhance product quality thereby improving access to export markets. Second, we provide a comprehensive and practical review of the empirical challenges involved in evaluating the welfare effects of "modern" contract farming schemes. Third, this study is unique in going beyond aggregate effects; rather, we provide a simple and intuitive decomposition of the scheme's

[†]This chapter is an edited reproduction of Jones and Gibbon (2011). It is co-authored with Peter Gibbon.

impacts.

The rest of the chapter is structured as follows: Section 6.2 provides an overview of organic cocoa contract farming in Africa. Section 6.3 introduces the scheme (§6.3.1) and data collection methods (§6.3.2). Section 6.4 discusses the economic mechanisms associated with the scheme and relevant empirical methods. Section 6.5 reports the results, including descriptive statistics from the surveys. Section 6.6 summaries the findings and reflects on general lessons from this case.

6.2 Organic contract farming for cocoa in Africa

6.2.1 Overview

Smallholder contract farming is the main source of certified organic exports from Africa. Assessment of its impacts requires an understanding of the specific characteristics that differentiate it from other types of contract farming on the sub-continent, as well as from organic farming outside the continent. Glover's (1984; 1987) classic definition of contract farming refers to annual contracts typically specifying production calendars, minimum/maximum delivery volumes, inputs and services provided by the buyer, quality requirements and prices. Producers are obliged to sell all output of a designated crop to the buyer, who in turn pledges to purchase it all, subject to quality. This definition implicitly refers to large, state-sponsored schemes for bulk export products (e.g. tea, sugar, tobacco, groundnuts), usually where smallholders were resettled on land cleared for this specific purpose (Little and Watts, 1994). While some schemes of this form persist, more modern forms involve private companies making arrangements with established farmers either for non-traditional exports, such as fresh produce (c.f., Dolan and Humphrey, 2000; Gogoe, 2003; Maertens et al., 2007), or for traditional export crops with new 'sustainable' qualities. While fresh produce schemes inherit several characteristics from traditional contract farming, newer schemes of the second type tend not to, reflecting the nature of both end and local markets.

In the case of traditional export crops, many local markets are now highly competitive. Thus, contractual monopsony is difficult to enforce and buyers are reluctant to supply inputs on credit. At the same time, end markets for traditional products with certified 'sustainable' qualities remain thin. Buyers adopt risk minimization strategies that emphasise strict quality control and price incentives, sometimes backed by farmer training to attain required quality dimensions. Hence, buyers normally only commit to provide smallholders with certification and to pay a premium for product of the required quality. In turn, smallholders promise to deliver produce and observe production rules.

Organic farming prohibits the use of synthetic inputs and promotes reliance on local inputs. Its distinctive feature is building soil fertility and controlling weeds, diseases and pests through rotations and using naturally occurring organisms and materials. Attaining organic certification mainly involves demonstrating non-use of synthetic inputs rather than following prescribed techniques. This is because organic standards emerged in developed countries with widespread use of synthetic inputs. In these cases, if synthetics are withdrawn, yields collapse unless alternative methods are adopted. Thus, there is no need to explicitly require specific crop management techniques. Use of synthetics, however, has been low and stagnant across much of Africa (Crawford et al., 2003). Most smallholder agriculture is 'organic by default' and certified farmers can, theoretically, earn price premiums without major changes to farming processes. Even so, because most smallholders are poor, certification costs typically have to be met by the buyer –

although financial assistance has been available from the donor community.

The literature evaluating organic farming in the tropics is small (Chapter 5, also Bray et al., 2002; Damiani, 2002; van der Vossen, 2005; Lyngbaek et al., 2001; Bacon, 2005; Eyhorn, 2007).[1] Virtually all studies report results from Latin America where 'organic by default' agriculture is uncommon. Most are based on small samples and only two report comprehensive farm budget data (Chapter 5, also Eyhorn, 2007). Whilst two studies employ econometric techniques (Chapter 5 and Bacon, 2005), the range of issues investigated is limited. For example, Chapter 5 analysed survey data from an organic coffee contract farming scheme in Uganda. Controlling for a range of factors, it finds positive net revenue effects from both participation in the scheme and, more modestly, from applying organic farming techniques. Nonetheless, it does not formally investigate the economic mechanisms through which the observed revenue effect is produced. Moreover, to date no study has considered changes over time or wider spillovers that organic contract schemes may generate.

6.2.2 The international cocoa market

Global cocoa prices have risen since 2000 and remain resilient despite the 2008/09 financial crisis. The upward trend, although erratic, has become more consistent during the last three years. This relates to repeated global supply deficits and a growing consensus that production in Côte d'Ivoire (historically the leading supplier) faces long-term problems. By the 2008-09 season, prices had reached levels not seen since the mid-1980s. Recent years also have witnessed a growing emphasis on product quality and value-chain sustainability. Two of the three major global players have made explicit commitments to support sustainable production and this remains an area of expansion. In 2009, global cocoa output certified to 'sustainable' standards reached ca. 40,000 tons or 1.2% of world production. Organic cocoa production was even lower, at ca. 20,000 tons (Tropical Commodity Coalition, 2009). The price premium for organic cocoa ranges from US$100 to US$300 per ton (ICCO, 2006). However, due to its niche and 'luxury' status, demand for organic cocoa is discontinuous and production capacity exceeds demand. Hence, exporters must ensure that organic cocoa also has quality attributes that command premiums in the conventional market.

Various attributes are captured under the rubric 'cocoa quality', including moisture content at shipment, flavour, and acid contents. Critical to attaining them is to optimise ripeness by harvesting pods every 2-3 weeks during the season and opening them within 3-5 days of harvesting. Extracted beans should be fully fermented and then sun dried (ICCO, 2007), requiring a minimum mass of raw beans (ca. 50 kg.) and a moderate commitment of labour time.[2] Scientific research (e.g., Hii et al., 2009) and market opinion indicate that the highest quality beans depend on careful and timely natural fermentation and drying by smallholders, rather than this being delayed and then carried out mechanically.

[1]The economic literature on smallholder schemes in the tropics certified to other sustainability standards, including Fairtrade, is even smaller (c.f., Pariente, 2002; Becchetti and Costantino, 2006; Giovannucci et al., 2008).

[2]A mass of beans is necessary to achieve an optimal fermentation temperature. Fermentation and sun drying are not labour intensive, but require ongoing care and monitoring.

6.3 The Bundibugyo scheme

6.3.1 Description

Cocoa production in Uganda dates from the 1950s but had minor importance until recently. Exports stood at 2,130 tons in 2001, reaching 5,386 tons in 2005 and 10,090 tons in 2009.[3] Production today is by 15-18,000 smallholders, overwhelmingly in Bundibugyo District bordering the Democratic Republic of the Congo. The cocoa area lies at an altitude of 700-1050 m. with average rainfall of 1400 mm. per year and average temperatures of 28-35°C. The District is remote and has neither mains electricity nor any tarmac roads. To date, few of the plant health problems plaguing cocoa in other regions of Africa have been experienced.[4]

In late 2005 there were four companies buying cocoa in Bundibugyo, one of which operated a contract farming scheme. By early 2009 there were six, three of which operated such schemes. All schemes were either certified organic or technically ‘in conversion’. The scheme operated by Esco (U) Ltd is the oldest. Farmer (re-)registration and certification was undertaken with support from a Sida project and the first exports occurred in 2002.[5] In addition to cocoa, vanilla production is certified although Esco has made little attempt to encourage it since international prices collapsed in 2003-04.

Esco has used the characteristics of different cocoa-growing areas to determine eligibility to join their scheme. At the outset, the company selected a number of parishes for inclusion based on an informal assessment of numbers of cocoa farmers and their specialization in cocoa production.[6] All households in these parishes were permitted to register as scheme members. Given zero entry costs, the vast majority of households in selected parishes initially did so. With respect to organic certification of these households, an ‘internal control system’ (ICS) has been used. This entails farm inspections by company field officers trained in organic farming methods. Inspections are also used to provide technical advice. In 2005, there were only four field officers and formal training was confined to 30-40 ‘contact farmers’ with demonstration plots in each village. Training has emphasized farm practices – partly but not exclusively organic – that should enhance yields. Annual third party certification involves reviewing records of cocoa purchases from individual farmers against ICS documentation on potential output, as well as visits to selected farms.

By 2005, the scheme comprised 1,721 farmers in two adjacent parishes (Ngamba and Burondo). Farmers were required by contract to follow organic standards and sell to Esco, who in turn provided some subsidized inputs (including cocoa seedlings), but only to ‘contact farmers’ and only in 2001-02.[7]

[3]These figures are based on EU import data (Market Access Database), which is considered to be more reliable than official Ugandan export data.

[4]The commonest cocoa plant health problems are black pod, witches broom and swollen shoot diseases. According to Bowers et al. (2001), these are commonest where production is in large mono-cropped plantations. They are more common where it is grown in small stands in more bio-diverse contexts as in Bundibugyo. Furthermore, the commonest of these diseases (black pod) is spread by windborne rain. Thus the remoteness of Bundibugyo from other centres of cocoa production also plays a benign role.

[5]The scheme was originally set up in 1998 by a Sudanese company but quickly abandoned due to an insurgency (1999-2001), during which the population of the scheme area were evacuated to IDP camps. Esco is a subsidiary of the international trading house Schluter SA. It received technical assistance worth about $100,000 from Sida during 2001-05 to set up the scheme.

[6]District administration in Uganda operates at the county, sub-county, parish and village levels.

[7]Most farmers in the area nevertheless obtained drying tarpaulins free, as a by-product of their period in the IDP camps, where these were provided for shelter. Later, the Sida project provided shade tree seedlings free to all farmers for a time.

Meanwhile, Esco buying posts in the scheme area only accepted cocoa that had been fully fermented and properly dried. In 2005 Esco offered scheme members a fermented cocoa price (at Ush 1,900/kg = US$1.07), 30% higher than that for fermented 'conventional' cocoa outside the area, and 100% higher than for unfermented cocoa. Nonetheless, in 2005 Esco procured only 269 tons of organic cocoa. Farmers complained that sometimes they could not sell to Esco as its buying posts ran short of cash.

Between 2005 and 2009 the scheme underwent important changes. Despite support from Sida elapsing, it was physically extended and the number of certified farmers grew to over 5,000.[8] Farmers not selling to the scheme were expelled – growers' lists examined for 16 villages in the original scheme area showed that around 26% of farmers certified in 2001-02 were struck off between 2005 and 2009.[9] Certification now covered US as well as EU standards. Instead of being managed from a rented store in Burondo parish, its base was a well-equipped town office. The field staff now numbered 30-35, training had been stepped up – principally in tree management and crop processing – and Esco had initiated a savings society.[10]

Esco now bought cocoa on a continuous basis. Backed by a new bye-law, farmers were strongly encouraged to harvest cocoa every two weeks. Buying posts always had cash and now employed technical instruments to read moisture levels, rather than relying on subjective assessment. An 'organic premium' was paid only to certified farmers presenting cocoa with $< 8\%$ moisture content, with a further premium for moisture $< 7.5\%$. Discounts at roughly Ush 200 intervals were applied for each degree of moisture content above 8%, until cocoa was determined (at 13%) to be unacceptable. The same system (except for the 'organic premium') was applied by Esco to conventional farmers outside the scheme and appears to have been copied by other buyers. Importantly, conventional farmers have been able to sell to Esco on a spot basis only, and therefore have not benefited from any contractual certainty. In 2008-09 the organic premium price was around 16-18% higher than what a 'conventional' farmer would have received for the same crop. All prices had moved upwards since 2005, with good quality organic commanding Ush 3,300 ($1.86)/kg. As a result of these changes, Esco's organic purchases dramatically increased to 2,593 tons in 2008.

6.3.2 Data collection

Surveys of households eligible to participate in the scheme and non-eligible households were undertaken in late 2005 and early 2009 using a questionnaire administered to household heads. This covered household demographics, factor endowments, agricultural revenue and expenditure, marketing behaviour and selected aspects of consumption. Data also was collected on farmers' use of a range of farm technologies, in most cases through physical observation.

Two-stage random sampling was used to select eligible and non-eligible households in both surveys. In 2005 three of the 38 'organic' (eligible) villages in Ngamba and Burondo parishes were purposively selected to reflect the range of local agro-ecological conditions. 30 farmers were then randomly sampled

[8] 6,950 were certified at the time of the second survey (January-February 2009) but not all of these had been certified at the start of 2008/09 season, whose results are covered.

[9] An analysis of survey data from 2005 comparing expelled farmers with those remaining in the scheme did not indicate any bias in terms in factor endowments or even total cocoa output in favour of those remaining in the scheme (details available on request from authors).

[10] Farmers depositing money when making sales to Esco received a additional premium of Ush 100/- per kg. A further premium of Ush. 100/- per kg. was paid to farmers delivering over 2 tons a year.

from a total of 199 farmers registered by Esco in these villages. The 2005 control (non-eligible) group was randomly sampled from a list 163 cocoa farmers prepared by local leaders in three villages in Busaru parish, a nearby ‘conventional’ area. These villages were chosen for their similarity to the agro-ecological conditions represented by scheme participants, in terms of soils, mix of altitudes and rainfall patterns.

In 2009, 16 of the original 38 villages in the scheme were selected for the ‘organic’ sampling frame, including all three sampled in 2005. 697 farmers living in these villages had been certified continuously since 2001-02, of which 90 were sampled at random from growers’ lists. It was not possible in 2009 to select a control group from Busaru parish again, since by this time most farmers there also had become certified organic. Although these might represent a useful sample for pre- and post-treatment analysis, these villages were only certified organic in 2008. Hence, while they were no longer ‘conventional’, they had not been certified organic for long enough for any scheme participation effects to be evident. Instead, eight villages in Bundinyama parish and three villages in Kaghema parish were selected to represent the District’s remaining conventional area. A control group of 78 non-eligible households were randomly selected from a list of 825 cocoa farmers prepared by local leaders in these villages. Thus, over the two survey rounds, a total of 222 households from 30 separate villages were interviewed.

6.4 Analytical framework

6.4.1 Economic aspects

Esco’s Bundibugyo scheme is structured so as to provide incentives to scheme members to process their cocoa crop to a high grade, following recommended organic techniques, and sell the processed crop to Esco. In turn, Esco provides a guaranteed price premium, a commitment to buy all high grade cocoa offered by farmers, and transparent measures of quality. Nevertheless, scheme members freely choose whether to process cocoa and how much to sell to Esco. Thus, the impact of participation in the scheme is of interest as opposed to formal membership.

These characteristics suggest four mechanisms through which scheme participation may affect household welfare. First, certified farmers may choose to sell properly processed cocoa to Esco, rather than to another intermediary, in order to benefit from an organic price premium. For farmers that already produced all cocoa to a high grade, farm practices would be largely unchanged. Second, for other farmers the scheme may induce greater specialisation in production of high grade cocoa, entailing a shift away from production of a more diverse range of crops or different standards of cocoa. If farmers previously sold only raw beans, they may also adopt cocoa processing technologies for the first time. Third, organic certified farmers may adopt recommended organic practices. As noted in Section 6.2, however, these specific practices are not required for ongoing certification.

Finally, various indirect benefits may accrue to scheme participants and other cocoa farmers. Perceptions of participation risk and coordination costs may fall as the benefits of adopting different technologies are revealed. Such information and social network externalities have been shown to be important drivers of change in smallholder agriculture (Conley and Udry, 2001; Besley and Case, 1993). Additionally, intensification of local economic activity, initially driven by growth in one product market, can generate productivity gains in other markets through *inter alia* private investment in local infrastructure and enhanced access to credit (Ravallion, 2005). Previous studies of commercialization and technology adoption highlight the potential for such household and regional spillovers from crop-specific interventions. Govereh and Jayne (2003) study cotton commercialization in Zimbabwe, for example, and suggest that higher incomes and improved access to inputs from participation in cash crop schemes can relax constraints to other household production activities (also Von Braun and Kennedy, 1994).

However, in the presence of multiple market failures, a superficial reading of incentives often provides a poor guide to behaviour. Farmers may resist full commercialization or improved technologies when these jeopardize a minimum level of food crop production or expose the household to large income fluctuations (Ellis, 1993). In contrast to selling raw cocoa, the decision to process beans to a high grade is not risk-free. Farmers must defer sale for at least two weeks, during which prices can change or buyers disappear from the market. Smallholders also may reject increased specialization in cocoa production and/or processing if, for example, farmgate cocoa prices or local food markets are unreliable. Indeed, producers often respond to price volatility by reserving a share of land for food production, despite higher expected returns from other crops (Fafchamps, 2003; Byerlee et al., 2006).

The structure of intermediation between smallholders and the world market is also of direct importance to quality-related technology decisions. Longitudinal research among cocoa producers in Ghana has identified sources of increased productivity in higher levels of effective competition among local buying companies in conjunction with a specific institutional complex (Teal et al., 2006). In contrast, the experience of other cocoa-producing countries (e.g., Cameroon and Côte d'Ivoire) suggests that unregulated competition in the context of a fully liberalized cocoa market may correspond to a decline in cocoa quality and, thus, to discounting against the world price (Losch, 2002). In the absence of institutions that can credibly regulate farmgate quality and/or entry into the intermediary market, a decline in quality can arise from free-riding on the public goods nature of implicit or explicit quality standards (c.f., Poulton et al., 2004).

Bringing these ideas together, two specific questions merit attention. First is whether the Esco scheme generates welfare benefits for those who participate in it. In theory, because individuals are free *not* to participate, and Esco does not have a monopsony position, we do not expect negative effects. Nevertheless, the scheme may be ineffective if its price incentives are negligible or if it fails to encourage farmers to (further) adopt quality-enhancing cocoa processing technologies and/or organic practices. Thus, the second empirical question focuses on the economic mechanisms at play – namely, whether the scheme has induced the adoption of post-harvest processing and/or organic practices.

6.4.2 Empirical strategy

With respect to the empirical challenges of the analysis, three potential sources of bias must be addressed. The first is endogenous programme placement. As discussed in Section 6.3, location-based characteristics were used by Esco managers to decide which households were eligible; thus, placement of the scheme was non-random. Second, there is self-selection bias, which refers to the choice to participate in the scheme. Formal registration as a scheme member was cost-free and does not guarantee actual participation, which refers to selling organic grade cocoa to Esco. It may be the case that more entrepreneurial or risk-taking households choose to participate, and that these are better-off regardless. Finally, there is geographic heterogeneity due to, *inter alia*, differences in soil productivity or access to infrastructure, which may be correlated with the other two sources of bias.

As has been established in the impact evaluation literature (see Chapter 5; also e.g., Angrist and Krueger, 1999), one way to address these problems is to fully and directly control for all sources of bias. Specifically, consider a general formula for the impact of actual scheme participation ("esco") on a welfare outcome (y) for household i in location j at time t:

$$y_{ijt} = \alpha_{1i} + \mu_j + \theta t + \delta \text{esco}_{it} + X'_{it}\beta + \epsilon_{it} \tag{6.1}$$

where X is a vector of time-varying household characteristics, representing elements of a household production function and ϵ is a white noise error term. Causal identification through 'selection on observables' requires that all elements of the equation are observed without error, including fixed differences in motivation or productivity across individual households α_{1i}, and geographic fixed effects μ_j.

The available data does not permit direct estimation of equation (6.1) as the household-specific effects are not observed. Similar difficulties have been confronted in the extensive micro-finance evaluation literature, from which further methodological inspiration is taken (for overview and references see Tedeschi, 2008). One strategy is to find proxy variable(s) for the household-specific effects, and include them alongside the other elements of equation (6.1) in a simple OLS regression. Such an approach is found in Coleman's (1999) quasi-experimental study. To evaluate the impact of a group lending programme in Thailand, he constructs a variable for scheme membership potential which includes both current borrowers as well as non-borrowers located in control villages who wish to join the programme. This variable is included to control for selection bias, allowing a distinct variable for participation to capture scheme impact.

A similar approach can be implemented here. Actual participation in the scheme is measured as the volume of organic grade cocoa sold to Esco as a share of the total volume of cocoa sold.[11] This is zero for all non-eligible households; and for eligible households ranges from zero (eight cases) to one (24 cases) with a mean of 0.35 and standard deviation of 0.42. *Potential* participation is measured as the share of fully processed cocoa sold to all buyers out of all cocoa sales by the household, constructed in the same way for eligible and non-eligible households. This reflects a revealed preference to engage in cocoa processing regardless of scheme eligibility. Thus, using previous notation and denoting the proxy for unobserved household-specific effects by "potential", a feasible estimating equation is:

$$y_{ijt} = \alpha_2 + \mu_j + \theta t + \delta \text{esco}_{it} + \pi \text{potential}_{it} + X'_{it}\beta + \epsilon_{it} \tag{6.2}$$

Despite the viability of this approach in theory, in practice it is not fail-safe. First, mismeasurement of either the actual or potential participation variables would generate attenuation bias. This is relevant in this study as both these variables depend on recall. Second, correlation between the same variables, as well as with the geographical fixed effects, may generate multicollinearity, leading to imprecise estimates for the main coefficient of interest (δ). Third, there is no reliable way to verify whether the proxy for household-specific effects is adequate. It also may be confounded with scheme impacts. For instance, the decision to process cocoa by a non-member household could reflect positive spillover effects from the Esco scheme, rather than innate household characteristics. Thus, by including the proxy on the RHS of equation (6.2) we run the risk of over-controlling and, thus, underestimating δ (see Wooldridge, 2005). Similarly, inclusion of a full set of location fixed effects may over-control for spillovers arising from participation within each location (e.g., Morduch, 1998).

An alternative estimation strategy is to find a valid instrument for scheme participation, enabling us to omit the proxy for household-specific effects. In a wide range of studies, exogenous aspects of scheme

[11] The main results are robust to alternative definitions of actual participation, such as the percentage (by number) of sales of high grade cocoa to Esco. Results available on request.

eligibility or availability have been used in this way. For example, Pitt and Khandker (1998) construct instruments for participation in microfinance programmes in Bangladesh from asset ownership criteria employed to exclude participants. Such an approach is also possible here. By virtue of the scheme's design, geographic rather than household-specific factors were used to determine scheme eligibility. As the latter is conditionally independent of the former, it represents a potential instrument for scheme participation. However, this only deals with selection bias, meaning that location characteristics must be included to address placement bias and geographic heterogeneity. This is problematic as all households in each chosen location were eligible to join the scheme. Consequently, a dummy variable for scheme eligibility, the potential instrument, is collinear with the vector of location-specific fixed effects, regardless of whether we specify geographic fixed effects at the sub-county, parish- or village-levels.

To address these concerns, the instrumental variable (IV) procedure can be adapted by defining a set of (continuous) proxies to substitute for location fixed effects. Assuming these are not collinear with the eligibility instrument, an IV approach is feasible. This method of using a vector of location characteristics in place of fixed effects is not new and also has been employed in the micro-finance literature (e.g., Pitt and Khandker, 1998). It also carries some advantages relative to the previous method – the adequacy of the location proxies can be directly tested; there is a lower risk of over-controlling; there are potential efficiency gains; and the IV approach can address any bias from mis-measurement of the participation variable. More formally, the second approach looks like:

$$\text{esco}_{ijt} = \alpha_3 + \eta t + \gamma \text{eligible}_{jt} + X'_{it}\varphi + Z'_{jt}\omega + v_{it} \tag{6.3}$$

$$y_{ijt} = \alpha_4 + \theta t + \delta \text{esco}_{ijt} + X'_{it}\beta + Z'_{jt}\lambda + \epsilon_{it} \tag{6.4}$$

where Z is a vector of proxies for the omitted location fixed effects. Equation (6.3) is the first-stage participation prediction equation, in which "eligible" is a dummy variable for scheme eligibility and varies only at the location level. Under the assumption $\text{E}[v_{it}\epsilon_{it}] = 0$, consistent estimates of equation (6.4) can be acquired by replacing the "esco" variable by its fitted values from the first stage (as per a two stage least squares procedure). Further practical aspects of this approach are described below.

6.5 Results

6.5.1 Descriptive statistics

Following the discussion of Sections 6.3 and 6.4, the distinction between eligible and non-eligible households is not the principal focus of analysis. More pertinent are differences in the extent of households' actual participation in the scheme, as well as their potential participation, used to proxy for unobserved household effects. Before presenting econometric results, it is helpful to investigate whether observed household and location characteristics vary with these two measures. As they are continuous, however, Table 6.1 presents descriptive statistics using a set of 2×2 categories created by splitting the potential participation variable (share of cocoa processed) at is mean and the measure of actual participation at 50%. In each case a value of one is given to observations above the split point, and zero otherwise.

With respect to the rows of Table 6.1, all monetary values are expressed in 2005 local prices, using a common deflator based on official information.[12] Net revenues are calculated as gross sales minus

[12] Specifically, we use the 'all items' consumer price index for Mbarara (the nearest town in Western Uganda for which an index

variable costs of equipment and labour, including exchange labour.[13] All estimates are calculated using sample weights. Sample design weights were calculated directly from the sampling frame for each village. However, to ensure that no specific time period or group of eligible/non-eligible households is over-represented, design weights were subsequently recalibrated to ensure an equal balance across the eligible/non-eligible households and over time periods. If nothing else, this is important due to the smaller absolute sample in the 2005 round. The final two columns of the table report whether there are significant differences in the conditional expectation of the row variable according to actual ("A") or potential participation categories ("P"), controlling for any common changes over time. This derives from separate regressions of each row variable against dummy variables representing the actual/potential participation categories of the table, plus a time period dummy equal to one in 2009 and zero otherwise. The column stars denote the significance of individual Wald tests that the actual/potential participation dummy variables are equal to zero.

The results indicate that the econometric concerns raised in Section 6.4.2 are pertinent. Four points merit comment. First, scheme eligibility guarantees neither actual nor potential participation in the scheme. For instance, around 11.2% of households who process a below-average level of cocoa (low potential) and sell less than 50% of their cocoa to Esco as organic are scheme eligible. Second, there are systematic differences between households according to their actual *and* potential participation rates. For example, high potential participants appear to own more cocoa trees and have larger farms than those who process a below average share of their cocoa. They also appear to have higher net cocoa revenues when compared to households within the same category of actual participation. Consequently, to avoid biased estimates, any assessment of scheme impacts must control for differences in (pre-existing) household characteristics. Third, location characteristics are important. Controlling for potential participation, the most active actual participants appear to live further away from Bundibugyo town. Finally, not shown in the table, one also notes distinct changes over time across all households, eligible and non-eligible. For example, use of hired labour, processing intensity and total number of sales have all increased substantially, pointing to possible dynamic processes at the regional level. We discuss these briefly in subsection 6.5.4.

6.5.2 Income effects

The results of Table 6.1 motivate the use of multivariate techniques to separate out the complex determinants of welfare differences between households. Thus, Table 6.2 presents results from a range of models for the welfare impact of the Esco scheme, focussing on the first of the two empirical strategies presented in Section 6.4.2. In all cases the dependent variable is the logarithm of net cocoa income. The null hypothesis of interest is that actual participation in the scheme has no causal effect on net cocoa income on average. Columns (I) to (V) correspond to different versions of equation (6.2). Column (I) represents a naïve model, which excludes household and location fixed effects; column (II) adds the potential participation measures to proxy for unobserved household effects; and columns (III) to (V) include sub-county, parish and village fixed effects respectively. The core explanatory variables include standard elements of a cocoa production function (e.g., Teal et al., 2006), as well as the number of times the household makes cocoa sales (of all types to all buyers). The latter is included to control for changes in

is available) taken from Uganda Bureau of Statistics (2009).

[13] We do not impute estimates of household labour costs. Besides difficulties of recall and of arriving at a common metric for adults and non-adults, family labour is frequently employed for tasks where paid labour is never or almost never used (e.g., supervision, monitoring the drying of beans and fermentation). Therefore there is an absence of reference material on which cost estimates could be based. Further details about variable construction are available from the authors on request

Table 6.1: Descriptive statistics by potential and actual participation

	Potential (P) ⟶	Low	Low	High	High		Signif.	
	Actual (A) ⟶	Low	High	Low	High	*Overall*	P	A
Household characteristics								
a.	Scheme eligible	11.2%	100.0%	29.2%	100.0%	*60.1%*		***
b.	Farm size (log. acres)	1.3	1.2	1.7	1.7	*1.5*	***	
c.	Cocoa trees (log)	6.6	6.4	7.0	7.1	*6.8*	***	
d.	Age of household head	40.2	43.2	44.9	49.2	*44.4*	**	
e.	Household farm workers	1.5	1.9	1.7	1.7	*1.7*		
f.	Use of hired labour	54.4%	30.2%	65.4%	72.4%	*55.6%*		
Location characteristics								
g.	Village altitude (log m.a.s.l)	6.7	6.6	6.7	6.7	*6.7*		
h.	Village distance to town (log)	2.3	3.0	2.3	2.9	*2.7*		***
i.	No. primary schools in parish	5.9	0.0	4.6	1.0	*2.9*		***
j.	No. secondary schools in parish	1.0	0.0	0.6	0.0	*0.4*		***
Revenue variables (10^3 USH)								
k.	Total farm revenue (gross)	701.7	1252.4	1169.7	2050.3	*1293.5*	***	***
l.	Cocoa revenue (gross)	660.9	1090.6	1073.1	1917.8	*1185.6*	***	***
m.	Cocoa revenue (net)	507.2	960.2	840.3	1779.3	*1021.7*	***	***
n.	Cocoa / total revenue (gross)	95.0%	87.6%	91.8%	90.8%	*91.3%*		
Variable costs								
o.	Labour costs (USH/tree)	62.2	48.6	81.8	70.0	*65.6*		
p.	Equip. costs (USH/acre)	28.0	15.9	21.6	23.8	*22.3*		
Intermediate outcome vars.								
q.	Cocoa volume (log. kg of FDE)	5.7	6.3	6.0	6.6	*6.2*	*	**
r.	Price received (USH / kg)	1240.7	1312.5	1630.2	1901.8	*1521.3*	***	***
s.	Sales to Esco (FFC+SFC / all)	11.8%	45.0%	53.9%	89.6%	*50.1%*	***	***
Technology indicators								
t.	Cocoa processing (%)	37.8%	63.5%	98.1%	97.7%	*74.3%*	***	*
u.	Organic grade cocoa (% vol.)	6.2%	63.7%	23.8%	81.1%	*43.7%*	**	***
v.	No. of sales	3.0	3.2	3.3	5.7	*3.8*		***
w.	No. of organic practices	0.4	1.3	0.4	0.8	*0.7*		***
Observations (unweighted)		78	49	9	86	*222*		
Sum of inverse probability weights		87.7	16.7	46.8	70.8	*222.0*		

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

Notes: actual (A) and potential (P) participation categories are binary dichotomisations of their continuous counterparts (see text); high (low) A is defined as observations above (below) 50%; high (low) P is defined as observations above (below) its mean; all statistics are calculated using sampling weights (calculated as the inverse probability of selection, adjusted to balance the sample between eligible and non-eligible households); for all non-revenue items, descriptive statistics are group means; all revenue variables are expressed in 2005 prices; final two columns reports results of tests for whether the actual and potential participation dummies are equal to zero in simple OLS regressions of the row variable, controlling for period effects; MASL refers to metres above sea level (variable g); FDE refers to fermented dry equivalent (variable q); variable s gives the number of sales to Esco of fully- and semi-fermented cocoa (FFC and SFC) as % number of all cocoa sales; total number of sales (variable v) is used as a proxy for harvesting frequency.

Source: own analysis based on Bundibugyo cocoa surveys (2005, 2009).

harvesting frequency over time (see above); it may also reflect a range of other household characteristics, such as households' commitment to cocoa farming, market orientation and/or commercial acumen. Due to space limitations, however, results for this vector of control variables are not discussed at length.[14]

A principal result is that the point estimate for the coefficient on actual participation (δ) is positive but declines and loses significance as we include a more complete set of controls. The introduction of the proxy for household-specific effects (column II) has a negligible effect vis-à-vis the naïve model. Nonetheless, location fixed effects are highly significant in all relevant specifications. For instance, when defined and included at the parish or village levels (columns IV and V), they yield an estimate for δ that is insignificant and is around half the size of the naive estimate. This supports the need to control for endogenous program placement and suggests that once location effects are included, the null hypothesis remains intact.

Nevertheless, the suitability of this approach remains in doubt (see Section 6.4.2). Aside from possible defects with our chosen proxy for household-specific effects, a relevant concern is multicollinearity between actual participation and other explanatory variables, especially location fixed effects and participation potential, making it difficult to identify the effects of actual participation precisely. The variance inflation factor (VIF) for actual participation, reported in the table, increases sharply from less than two in column (I), to around six in column (V). This is directly reflected in the magnitude of the standard errors on the point estimates for actual participation, which roughly doubles once the location fixed effects are included.

These concerns motivate application of the second empirical strategy. If nothing else, this provides a robustness check on the previous results. We focus on finding proxies for spatial fixed effects at the village level as this exploits the maximum amount of information available in the data. Despite the fact that Esco managers selected entire parishes as eligible, village heterogeneity may correlate with relevant unobserved (household-specific) variables. Focusing on villages also corresponds to the least restrictive assumptions regarding the amount of information used by Esco in its selection of eligible zones. Proxies for village fixed effects (VFEs) are generated from observations of specific village characteristics, as well as village means of household-level variables. With respect to the first type, we use the altitude of the village above sea level, its distance from Bundibugyo town, and the number of primary and secondary schools in the parish. With respect to the second type, we calculate village means of the household characteristics used in the model (taken from vector X), as well as the share of cocoa in total household revenue, the share of land devoted to cocoa production, the number of non-farm labourers in the household and a dummy for recent acquisition of land. These variables are intended to capture specific factors that Esco considered in its selection of eligible areas.

Column (I) of Table 6.3 gives results for a modified version of equation (6.2), where a full list of fifteen proxies is used in place of the individual VFEs. Column (II) repeats this model, but employs a reduced number of proxies based on a stepwise exclusion procedure.[15] Following Coleman (1999) (also Pitt and Khandker, 1998), various tests can be applied to examine the adequacy of these proxies. The results of a

[14] None of the results for control variables are unexpected. Note also that the two survey rounds are pooled. The appropriateness of pooling is confirmed by running an extended version of the fixed effects model given in Column (V) of Table 6.2, including period-specific variants of each time-varying covariate. A joint F-test of these period-covariate interaction terms is insignificant at the 10% level.

[15] We employ a general-to-specific (backward selection) approach. The general model is given in column (I) of Table 6.3; the probability threshold for retaining the VFE proxies is 25%. None of the core independent variables are affected by this selection procedure.

Table 6.2: Estimates of net cocoa income effect, household-specific effects proxy method

	(I) OLS	(II) OLS	(III) OLS	(IV) OLS	(V) OLS
Actual participation	0.998***	0.974***	0.608**	0.463	0.524
	(0.10)	(0.14)	(0.29)	(0.30)	(0.33)
Potential participation	-	0.059	-0.010	0.062	0.074
		(0.30)	(0.30)	(0.29)	(0.26)
No. of sales (log.)	0.267***	0.269***	0.265***	0.270***	0.245**
	(0.07)	(0.07)	(0.07)	(0.07)	(0.10)
Period dummy	-0.098	-0.109	0.012	0.028	0.134
	(0.15)	(0.16)	(0.14)	(0.15)	(0.21)
Cocoa trees (log)	0.737***	0.736***	0.718***	0.673***	0.684***
	(0.14)	(0.15)	(0.16)	(0.16)	(0.15)
Farm size (log)	0.105	0.100	0.129	0.203	0.229
	(0.14)	(0.13)	(0.14)	(0.13)	(0.17)
HH farm labourers	0.186*	0.186*	0.169	0.154	0.132
	(0.10)	(0.10)	(0.10)	(0.10)	(0.12)
Hired labour (1=yes)	0.186	0.185	0.200	0.187	0.182
	(0.21)	(0.22)	(0.23)	(0.23)	(0.16)
Age of hhld head	-0.004	-0.004	-0.005	-0.004	-0.004
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Hhld head male	0.363*	0.362*	0.362*	0.347*	0.334*
	(0.21)	(0.20)	(0.20)	(0.19)	(0.20)
Fixed effects	None	None	3 subcounties	5 parishes	30 villages
Participation VIF	1.17	1.67	4.40	4.71	5.96
Number of obs.	222	222	222	222	222
R-sq.	0.63	0.63	0.64	0.65	0.64
RMSE	0.72	0.72	0.71	0.71	0.72
F-stat.	45.23	40.66	59.75	37.70	34.69

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

Notes: dependent variable is the log. of real net cocoa revenue; specification is as per equation (6.2) in the text; 'Potential participation' is a proxy for unobserved household characteristics; fixed effects at different geographic levels are indicated below the coefficients but are not reported; intercept also not shown; all estimates are by OLS; standard errors (in parentheses) are robust to arbitrary heteroscedasticity and clustering at the village level.

Source: own analysis based on Bundibugyo cocoa surveys (2005, 2009).

Table 6.3: Estimates of net cocoa income effect, location fixed effects proxy method

	(I) OLS	(II) OLS	(III) GMM	(IV) GMM	(V) GMM	(VI) OLS
Actual participation	0.529**	0.656***	0.989***	0.986***	0.899***	-
	(0.22)	(0.15)	(0.19)	(0.28)	(0.25)	
Scheme eligibility	-	-	-	-	-	0.700***
						(0.11)
Potential participation	0.328*	0.328	-	0.004	0.043	-
	(0.18)	(0.24)		(0.29)	(0.22)	
No. of sales (log.)	0.222***	0.283***	0.260***	0.260***	0.283***	0.301***
	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)
Period dummy	0.036	-0.053	-0.008	-0.009	-0.041	0.120
	(0.20)	(0.10)	(0.12)	(0.14)	(0.16)	(0.13)
Cocoa trees (log)	0.574***	0.584***	0.694***	0.694***	0.669***	0.679***
	(0.10)	(0.10)	(0.14)	(0.14)	(0.12)	(0.15)
Farm size (log)	0.221*	0.176	0.190	0.189	0.187	0.214
	(0.13)	(0.12)	(0.12)	(0.12)	(0.14)	(0.14)
HH farm labourers	0.091	0.157*	0.162*	0.162*	0.118	0.142
	(0.10)	(0.09)	(0.09)	(0.09)	(0.10)	(0.10)
Hired labour (1=yes)	0.072	0.063	0.154	0.154	0.143	0.199
	(0.12)	(0.14)	(0.21)	(0.21)	(0.14)	(0.24)
Age of hhld head	-0.003	-0.003	-0.005	-0.005	-0.006	-0.004
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Hhld head male	0.260	0.293*	0.369*	0.368*	0.357*	0.401**
	(0.22)	(0.16)	(0.20)	(0.20)	(0.21)	(0.19)
Village land acquisition	-0.384	-0.354	-0.555	-0.556	-0.654*	-0.605
	(0.41)	(0.41)	(0.37)	(0.37)	(0.37)	(0.37)
Parish primary schools	-0.123**	-0.144***	-0.130***	-0.130***	-0.141***	-0.174***
	(0.06)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)
Parish secondary schools	0.615**	0.581***	0.534***	0.535***	0.574**	0.789***
	(0.29)	(0.18)	(0.17)	(0.18)	(0.24)	(0.15)
Participation / member VIF	4.51	2.39	2.96	4.06	3.60	2.17
Number of obs.	222	222	222	222	222	222
R-sq.	0.64	0.65	0.64	0.64	0.64	0.64
RMSE	0.67	0.66	0.69	0.69	0.69	0.71
F-stat.	16.75	30.23	45.74	42.36	34.99	70.19

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

Notes: dependent variable is the log. of real net cocoa revenue; column (I) includes a set of fifteen proxy village fixed effects (11 not shown); column (II) onwards includes a sub-set of three of these (reported) selected by stepwise exclusion; column (III) employs a two-step GMM estimator in which scheme eligibility is used as an excluded instrument; column (IV) runs the same model adding the potential participation measure; column (V) includes four additional excluded instruments (see text); column (VI) is the reduced form associated with the model estimated in column (III), derived from equations (6.3) and (6.4) in the text and estimated by OLS; intercept not shown; standard errors (in parentheses) are robust to arbitrary heteroscedasticity and clustering at the village level.

Source: own analysis based on Bundibugyo cocoa surveys (2005, 2009).

Hausman test comparing the model in Table 6.2 column (V) against that of Table 6.3 column (II) cannot reject the null hypothesis that there are no systematic differences between the estimates for the common variables (probability = 63.8%). Auxiliary regressions, which regress the predicted VFEs taken from estimates of Table 6.2 column (V) on the full set of regressors in Table 6.3 columns (I) and (II) are also highly supportive. They indicate that once either the full or reduced set of location proxies are added, there is no remaining correlation between the household regressors and the estimated VFEs, which would have signalled the existence of omitted variables bias.

The estimates in columns (I) and (II) of Table 6.3 suggest a positive and significant welfare impact from scheme participation. The point estimates are highly comparable to those in the final three columns of Table 6.2, but are now significant. This reflects a reduction in multicollinearity (i.e., smaller standard errors) as well as an increase in degrees of freedom. From these estimates we expect that a 10% increase in the volume of sales of organic cocoa to Esco (as a share of total) generates around a 6% increase in net cocoa revenue. Using the selected subset of VFE proxies, column (III) of Table 6.3 presents the results for the IV approach. The table reports results for the second-stage, corresponding to equation (6.4), based on the GMM two-step procedure. Various auxiliary under- and weak-identification tests are passed comfortably, attesting to the overall strength of the excluded instrument (Kleibergen-Paap Wald F-statistic = 142.06). The point estimate for δ is significant and within (99%) sampling error of previous estimates, at 0.99 log points. Other coefficients are consistent with previous results.

Two robustness checks are run on these IV results. Column (IV) adds the measure of potential participation, which in principle is orthogonal to the instrument and, thus, should not alter the estimates for δ. Column (V) employs additional instruments – namely, four dummy variables corresponding to quintiles of mean village cocoa revenue share.[16] This permits a test of over-identification (Hansen J-Test), which gives no cause for concern (probability = 54%). In both sets of results the estimates for δ remain around one and all other estimates are broadly unchanged. This provides strong confirmatory evidence against the null hypothesis. It suggests that the first empirical strategy, which delivers lower point estimates for δ, may suffer from over-controlling bias.

The reduced form associated with the latter IV approach is of independent interest. This is reported in the final column of Table 6.3, derived by substituting equation (6.3) into equation (6.4). The positive and significant coefficient on scheme eligibility confirms its strength as an instrument (see Baum et al., 2007). Moreover, given that the effect of eligibility on scheme participation is expected to be greater than zero and less than one (as the scheme continues to operate and we assume rational agents), then the reduced form gives direct and unbiased information about the sign and magnitude of the welfare effect from scheme participation (δ).[17] There are additional reasons to focus on these estimates. As Dave and Kaestner (2002) note in a different context, reduced form results often hold greater policy relevance than structural estimates because they capture the overall impact of the underlying policy tool, taking into account the efficacy of this tool in altering behaviour. This is pertinent here as the reduced form captures the average welfare impact associated with scheme eligibility (the policy tool), which is the joint (multiplicative) effect of eligibility on participation (γ) and participation on welfare (δ). Interpretation of the reduced form also is intuitive in this case as the 'treatment' variable is binary. The results in column (VI) tell us that, on average, households that have been eligible to participate in the scheme are now 0.70

[16] In other words, villages are placed into quintiles according to mean household revenue share. Choice of this variable accords with Esco's location eligibility decision rule (see Section 6.3). Each quintile includes both member and non-member villages.

[17] See Angrist and Krueger (1999) also Baum et al. (2007) for further discussion of the properties of the reduced form. The 95% confidence interval for γ in equation (6.3) taken from the first stage corresponding to the estimates reported in column (IV) of Table 6.3 ranges from 0.48 to 0.82 with a point estimate of 0.65.

log points wealthier than non-eligible households, controlling for endogenous program placement, village heterogeneity and (observed) household characteristics.

As a further check on the robustness of the results in Table 6.3, it is helpful to run a set of 'placebo regressions' (c.f., Card, 1990). The idea is to investigate whether pre-treatment or highly persistent variables are systematically related to the treatment variable. If so, then it is likely that relevant variables have been omitted from the specification, meaning that estimated treatment effects will be biased. As no pre-treatment variables are available, five variables that are expected to be uncorrelated with scheme participation or eligibility are examined. Each row of Appendix Table 6A.1 reports selected results from two separate OLS regressions in which the dependent variable is one of these 'predetermined' row variables, and the RHS variables are based on the reduced form model from column (VI) of Table 6.3. The first two columns report the coefficient on scheme eligibility and associated probability that this is equal to zero for a regression without any proxies for village fixed effects (VFEs); the third and fourth columns report results when the selected subset of VFE proxies employed in Table 6.3 (column II onwards) are included.

The placebo regressions confirm the necessity of including VFE proxies to address endogenous placement bias. Non-cocoa revenue is presumed to be independent of scheme eligibility as the majority of sampled households strongly specialise in cocoa production; nevertheless, it may be correlated with certain unobserved (village) characteristics. This is supported by the significant result when the VFE proxies are not present; once included, however, the partial correlation coefficient on scheme eligibility is not significantly different from zero. The same pattern is repeated for village altitude – once the VFE proxies are included as controls, the partial correlation coefficient on scheme membership is insignificant. All other placebo regressions support the preferred specification.

6.5.3 Economic mechanisms

The previous sub-section established a positive and significant welfare impact from the scheme. The remaining question is whether this is attributable to a price premium alone, or to changes in productivity associated with farm practices (technology). Together these two channels are expected to be exhaustive. Controlling for common changes over time, as well as household and geographic characteristics, no other economic mechanism should generate changes in real net cocoa income. Their effects are also algebraically additive, because both final and intermediate outcome variables are expressed in natural logarithms. Thus, the individual effects of prices and technology are linear in logs. Consequently, by focussing on these two intermediate outcomes, one is able to disaggregate the aggregate net income effect. Empirically this is implemented by separately regressing: (i) the log of the average price received by each household; and (ii) a measure of real cocoa output (specifically, the log of the quality-adjusted cocoa weight) on the regressors taken from the reduced form specification (column VI, Table 6.3). This specification is chosen for its direct policy relevance (see Section 6.4.2) and for its simplicity, being estimated by OLS. However, in employing the same approach as before, any remaining bias in the reduced form estimates is likely to transfer to estimates for these intermediate outcomes.

The results are given in Columns (I) and (II) of Table 6.4. Scheme eligibility has a 0.15 log. point effect on prices, which is highly consistent with the magnitude of the price premium discussed in Section 6.3. It also has a 0.47 effect on quality-adjusted output (column II). Adding the estimates for these two channels gives a combined effect of 0.61, which is closely in line with the aggregate estimate of 0.70 taken from column (VI) Table 6.3. This confirms the exhaustive nature of the price and technology mechanisms, which respectively account for around 23% and 77% of the measured welfare effect. Of further interest is the extent to which productivity improvements derive from changes in post-harvest processing or use of pre-harvest organic practices. Thus, Table 6.4 examines the impact of scheme eligibility on different aspects of technology uptake, measured here as the share of all cocoa processed to a high grade (column III) and the number of organic practices employed (column IV). These estimates show a strong positive effect running from scheme eligibility to post-harvest processing. However, the effect on pre-harvest technologies is indistinguishable from zero and there appears to be a negative change in the average of number of practices employed over time. Consequently, the principal driver behind the observed income effects has been adoption of quality-enhancing post-harvest practices.

6.5.4 Spillover effects

As a final element in the analysis, we consider feedback effects from the scheme. An initial insight is gained from Table 6.4, which shows strong positive period effects on real prices (column I) as well as on the use of post-harvesting technologies (column III). These reflect positive changes for all farmers – e.g., average prices have risen by over 40% and *all* households have increased their use of post-harvesting technologies by over 20%. Appendix Figure 6A.1 provides further evidence of positive general trends. The figure plots the cumulative distribution of the intensity of cocoa processing for farmers, differentiated by scheme eligibility and survey round. It shows a marked increase in use of processing technologies across all farmers over time, strongly led by eligible households. In 2005 around 50% of non-eligible farmers processed none of their cocoa to at least a semi-fermented standard; in 2009 this had fallen to 10%. Similarly, in 2005 around 20% of eligible farmers processed all of their cocoa; in 2009 this had increased to 90%. Thus, it appears that low-cost post-harvest processing has diffused throughout the cocoa farming community, backed by a general increase in demand for high grade cocoa as well as more accurate measurement of cocoa quality. This trend is supported by the establishment of other organic cocoa schemes in the locality (see Section 6.3).[18]

6.6 Conclusion

We have undertaken a detailed study of an organic contract cocoa scheme in rural Uganda. The scheme is of broad interest because, in contrast to older models of contract farming in Africa, it operates only on the basis of a pared-down contract between the scheme operator and members. Existing literature provides limited guidance regarding the economic dimensions of such a scheme. To fill this gap, an analytical framework was presented. This drew attention to a range of pre- and post-harvest technology choices and their relation to output quality. Following the micro-finance literature, two empirical approaches were developed to deal with possible sources of bias. These yield broadly consistent results, although the statistical significance of the main coefficient of interest remains sensitive to specific methodological choices. In addition, a simple decomposition of the scheme effects was provided. Together, this represents the most comprehensive and careful analysis of organic contract farming in Africa to date. Nevertheless, we recognise that in observational studies such as these there is no guarantee that all sources of bias have been eliminated.

[18] Other spillover effects include a general increase in membership of savings associations, also led by scheme members. Results available from authors on request.

Table 6.4: Decomposition of income effects associated with scheme membership

	(I) Price recvd OLS	(II) Cocoa vol. OLS	(III) Processed OLS	(IV) Practices Poisson
Scheme eligibility	0.146***	0.467***	0.388***	-0.134
	(0.04)	(0.11)	(0.06)	(0.32)
No. of sales (log.)	-0.012	0.265***	-0.015	0.076
	(0.02)	(0.07)	(0.02)	(0.18)
Period dummy	0.439***	-0.237**	0.211***	-0.774***
	(0.02)	(0.10)	(0.05)	(0.23)
Cocoa trees (log)	0.071**	0.627***	0.037	-0.070
	(0.03)	(0.13)	(0.04)	(0.17)
Farm size (log)	0.018	0.191	0.102**	0.423**
	(0.03)	(0.14)	(0.05)	(0.17)
HH farm labourers	-0.054	0.185*	-0.004	0.118
	(0.06)	(0.09)	(0.06)	(0.19)
Hired labour (1 = yes)	0.071***	0.361*	0.029	-0.122
	(0.02)	(0.18)	(0.05)	(0.23)
Age of hhld head	-0.001	-0.004	-0.001	0.003
	(0.00)	(0.00)	(0.00)	(0.01)
Hhld head male	0.093**	0.220*	0.039	0.288
	(0.04)	(0.13)	(0.08)	(0.48)
Number of obs.	222	222	222	222
R-sq.	0.63	0.66	0.30	-
RMSE	0.22	0.59	0.31	-
F-stat. / Chi-sq.	104.41	43.26	91.38	161.53

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

Notes: the dependent variable in column (I) is the natural log. of the average price received by each household from cocoa sales (in 2005 prices); the dependent variable in column (II) is the log. of the volume of cocoa output, measured in kilograms of fermented dry equivalent (FDE); the dependent variable in column (III) is the share of cocoa processed to a high grade; in column (IV) the dependent variable is the number of organic practices adopted; all specifications estimated using sampling weights; the intercept and four proxy village fixed effects (see Table 6.3, column V) are included but not shown; columns (I) to (III) estimated by OLS; column (IV) estimated via a Poisson regression; standard errors (in parentheses) are robust to arbitrary heteroscedasticity and clustering at the village level.

Source: own analysis based on Bundibugyo cocoa surveys (2005, 2009).

Based on the reduced form results, which are of direct policy interest, we find that the average effect of scheme eligibility on net household cocoa income is around 100% (0.70 log points). This is primarily driven by increased post-harvest cocoa processing, which improves product quality and enables farmers to access a price premium. Changes in farm methods, however, have not been restricted to participants in the scheme. Evidence points to a general pattern of market deepening and demand-induced technology adoption. While it would be rash to attribute this solely to Esco, the latter has played a leading role in providing a consistent and credible source of demand for high quality cocoa. It also has stimulated widespread adoption (by other buyers) of tools to measure cocoa quality.

We end by reflecting on the general lessons from this case. On the one hand, institutional, temporal and market specificities make generalisation problematic. The surveys occurred over a period of rapid increases in international cocoa prices and growing demand for high quality 'sustainable' cocoa. While the prospects for organic cocoa are reasonable, this does not extend to many other markets where international price trends remain uncertain. Moreover, (organic) contract farming schemes are not all alike; much depends on the details of the effective incentive structure as well as management quality.

Even so, a first lesson is that the benefits of the Esco scheme derive primarily from incentives to adopt quality-enhancing techniques rather than from its specifically 'organic' aspects. Ongoing success has been due to a credible commitment to purchase high grade cocoa, transparent quality measurement and improving scheme management. Nevertheless, certification as organic has been essential for the scheme to access a (more stable) premium niche export market, thereby enabling Esco to maintain attractive purchase commitments to members. Such market access also was critical for Esco to establish the scheme in the first place.

A second lesson, *pace* Poulton et al. (2004), is that there are likely to be trade-offs between market power and competition in niche agricultural markets. The ability to establish some market power through farming contracts provides incentives for intermediaries to invest in supplier relations (e.g., certification) and make buying commitments based on upstream export expectations. Competition, on the other hand, provides outside options that can protect sellers against aggressive pricing. It is notable in this case that many eligible and non-eligible farmers continue to sell some of their cocoa crop to other buyers. Thirdly, technology adoption need not be considered a binary step-change. Technologies such as quality-enhancing farm practices can change gradually over time, stimulated by social learning as well as direct dissemination. However, basic incentive compatibility constraints always apply, and (expected) market conditions remain fundamental drivers of household behaviour.

Similar considerations apply to whether the experience of the scheme is replicable. International market conditions are likely to continue to make it profitable to invest in smallholder contract farming for high quality cocoa (and similar versions of other traditional export crops) in the medium-term, even if the markets for specific qualities such as organic or Fairtrade remain limited. Thus, whether 'aid for trade' of the kind provided to Esco by Sida continues to be available should not greatly affect replicability. Nevertheless, the benefits of such schemes can be eroded by market saturation.

Finally, a pertinent question is whether other quality-related contract farming schemes can function as effectively as has Esco's since 2005. This depends on local and institutional success factors, of which two are worth underlining. The first is corporate. Esco is part of a small international trading house focused on the Great Lakes Region, and increasingly on cocoa as a commodity. Its international status meant that its resources were adequate to upgrade the scheme even after Sida support elapsed – something that would not be possible for most Ugandan-owned beneficiaries of the same Sida programme. Meanwhile, the company's narrow geographical and crop focus meant the scheme's success was critical for Esco's overall profitability in a way that would not be the case for companies on the scale of Cargill, ADM or Barry Callebaut (who may therefore have approached it in a different way). The second group of factors is local. Cocoa in Bundibugyo is free from serious plant health problems and Bundibugyo is a remote district with few other income streams. Hence, cocoa production can be a full-time activity and competent field staff can be recruited at low cost. While not all of these conditions are likely to be reproducible, some replication should be possible given foreseeable market conditions.

6A Appendix: additional tables and figures

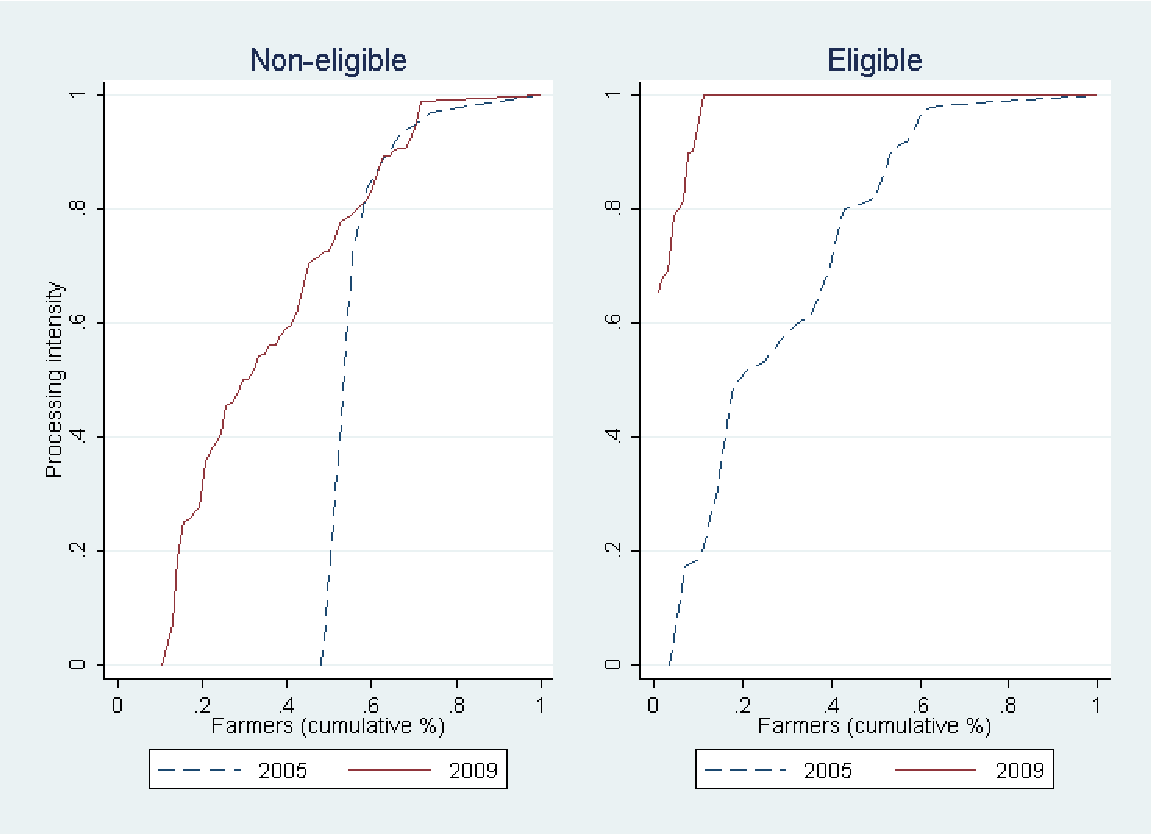
Table 6A.1: Summary results of placebo regressions

	Without VFE proxies		With VFE proxies	
Dependent variable	Coefficient	Probability	Coefficient	Probability
Non-cocoa revenue	2.78	0.00	1.28	0.28
Dependency rate	0.01	0.59	-0.04	0.42
Household non-farm labourers	0.06	0.42	0.16	0.34
Exclusively agricultural hhld	-0.02	0.72	-0.02	0.81
Village altitude	-0.07	0.02	0.00	0.95

Notes: each row reports summary results from two separate OLS regressions where the dependent variable is the indicated row variable and the explanatory variables follow the specification in column (VI) of Table 6.3, respectively without and with the selected subset of VFE proxies; columns give the estimated coefficient on scheme eligibility from each of these regressions and the corresponding probability that this is equal to zero (from a Wald test).

Source: own analysis based on Bundibugyo cocoa surveys (2005, 2009).

Figure 6A.1: Cumulative distribution of farmers by cocoa processing intensity (in %, vertical axis) with farmers distinguished by scheme eligibility and survey period



Source: own analysis based on Bundibugyo cocoa surveys (2005, 2009).

Chapter 7

Aid supplies over time: accounting for heterogeneity, trends and dynamics†

7.1 Introduction

The supply of foreign aid is most often considered from the perspective of recipients. Who receives aid, how much and for what kinds of activities has generated a huge literature, much of it critical. Substantially less analysis has been devoted to quantifying supply-side factors that influence either the long-run level or shorter-run dynamics of aid disbursements made by individual donors. Nevertheless, this question has received renewed attention. The financial crisis of 2008/09 stimulated numerous warnings that aid flows were likely to fall as macroeconomic conditions in donor countries worsened and fiscal costs mounted. For example, Roodman (2008) calculated that after the Nordic financial crisis of 1991, aid from Norway, Sweden and Finland fell by 10 per cent, 17 per cent and 62 per cent respectively (measured from peak to trough and adjusted for inflation). Similar concerns prompted Laurie Garrett of the US Council on Foreign Relations to warn: "As the global marketplace retrenches, there is great danger that the poorest billion people will be abandoned, their hopes for escaping poverty and disease forgotten by all but themselves." (Garrett 2009: 2).

The objective of this chapter is to take a careful look at the past relationship between economic conditions in donor countries and their supply of foreign aid. Building on the contributions of previous studies, which show significant relationships between domestic economic factors and aid disbursements, this study addresses some of the main gaps in the literature. In particular, descriptive evidence suggests that aid supply relationships have dynamic properties and are heterogeneous – both between countries and over time. Previously, lack of adequate data has limited our ability to study these characteristics. However, moderately long time series data is now available for at least twenty of the largest OECD donors, covering a maximum span of 1960 to 2009 (yearly). As such, econometric techniques can be used which are consistent in the face of slope heterogeneity and non-stationarity. These are derived from country-specific estimates of aid supply behaviour, which are of interest in themselves.

The rest of the chapter is structured as follows: Section 7.2 motivates the analysis of aid supply behaviour,

†This chapter is an edited reproduction of Jones (2011).

drawing inspiration from a descriptive analysis of the historical data. It also provides a review of the existing literature on the determinants of aid supplies, indicating gaps that have not been addressed. In response, Section 7.3 proposes a simple conceptual framework from which an empirical model is derived. Section 7.4 discusses appropriate panel data methods to test this model, giving particular attention to challenges of parameter heterogeneity, non-stationarity and cross-section dependence.

Section 7.5 presents the results – encompassing evidence on the long-run trends in aid supplies (§7.5.1), their dynamics (§7.5.2) and their interpretation (§7.5.3). The main findings each represent new contributions to the literature. They include evidence for substantial heterogeneity between countries in their aid supply behaviour, the evolution of donor behaviour over time (e.g., due to the end of the Cold War) and the merit of distinguishing between long- and short-run determinants of aid supplies. These insights suggest that aid supplies are more complex and ambiguous than previously appreciated, thereby limiting the confidence we can have in our ability to predict future trends from those of the past. Section 7.6 concludes.

7.2 Background and motivation

A small number of previous studies explore the determinants of either the level of or changes in the aggregate supply of aid. These broadly fall into one of two categories. The first group of studies explore the large and persistent differences in donors' aid effort, typically defined as the foreign aid to domestic income ratio (alternatively, aid supply per capita). These show that much of these differences are attributable to fixed factors including historical relations and cultural preferences. Macroeconomic variables, such as the domestic income level or the government's fiscal position also appear to be important long-run influences on aid supply levels. The findings of these studies are broadly coherent and appear to fit the data well. For example, Round and Odedokun (2004) estimate a panel fixed effects model of the Aid/GDP ratio which is able to account for over 80 per cent of the variation in the ratio around its global mean. Similarly, Chong and Gradstein (2008) employ data from the World Values Survey, as well as standard macroeconomic indicators, to identify factors which affect support for foreign aid within the electorate of donor countries. Based on the their full specification, 65 per cent of the variation in aid disbursements is accounted for by donor fixed effects and a further 31 per cent is explained by their other chosen explanatory variables – donor income, government behaviour and political leaning.

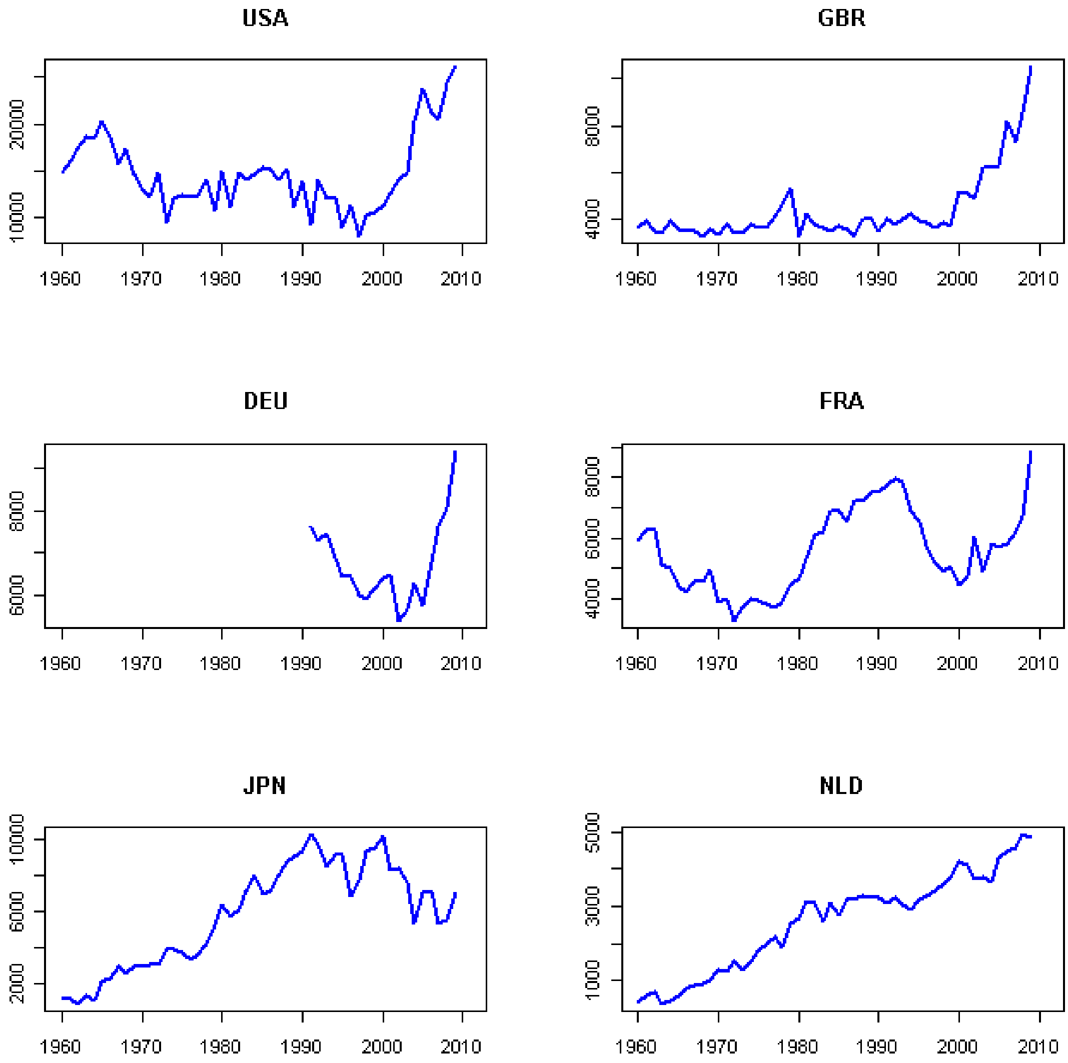
The second category of studies seeks to account for aid dynamics, particularly the response of aid supply to domestic macroeconomic shocks. A leading contribution in this regard is Pallage and Robe (2001), who investigate the cyclicality of aid from both a donor and recipient perspective. This line of enquiry has received renewed attention in the wake of the recent financial crisis, focussing on what tends to happen to aid budgets when donors face financial sector problems. Following Roodman (2008), various studies have quantified both country-specific and cross-country responses to such shocks. Dang et al. (2009) estimate that real aid disbursements tend to fall by up to 25 per cent in response to systemic banking crises relative to a 'no crisis' counter-factual. Similarly, Frot (2009) finds that aid tends to decline by 13 per cent in response to a financial crisis, or 5 per cent per annum. Thus, a proximate motivation for the present study is the passage of time. Disbursement data is now available up to and including 2009, which covers a period of substantive macroeconomic shocks in donor countries. This represents an informative window on aid supply behaviour, providing an opportunity to update past estimates of the (domestic) determinants of aid and examine the accuracy of predictions about what would happen to aid in the wake of the crisis.

Despite the above findings, there are remaining gaps in the literature. These directly inform the three main areas where this chapter makes a contribution. The first of these is to differentiate between long- and short-run determinants of aid supplies. Both determinants are relevant but for different reasons, meaning that there is no reason to expect them to be the same *a priori*. As indicated by the first category of previous studies, long-run determinants are likely to be rooted in a small number of slowly moving domestic factors. For instance since the early post-war period, when foreign aid became established as an instrument for promoting economic development, donors have made various commitments to donate a fixed share of their national income in aid. For example, as Clemens et al. (2007) document, the World Bank's Pearson Commission, concluded in 1969, recommended that bilateral donors increase net aid disbursements to 0.7 per cent of gross national product by at least 1980. Exhortations to achieve this figure have been repeated in various multilateral settings such as General Assembly resolutions and the UN's 2002 International Conference on Financing for Development held in Monterrey, Mexico. Recently, aid targets have gained renewed force from public pledges to increase foreign aid (or meet existing aid targets), such as those made at the 2005 Gleneagles G8 and Millennium +5 summits. Admittedly, despite such good intentions, aid targets seem to have operated more as a lobbying tool. Clemens et al. (2007) argue that, at best, rich countries have agreed to gradually move towards the 0.7 per cent goal, rather than attain it within a specific time frame. Even so, the persistence of the concept of aid supply targets suggest there are deep, long-run domestic influences on aid supplies.

At the same time, short-run aid dynamics are likely to be determined by a host of other factors – domestic or global. Country-specific studies indicate the importance of idiosyncratic local (political) factors in explaining changes in aid supplies. For example, among other things, Fleck and Kilby (2010) find that the War on Terror has driven a large increase in US foreign aid disbursements, but also has reduced the weight given to needs in the allocation of aid to core recipients. This echoes the cross-country work of Boschini and Olofsgand (2007), who find that the end of the Cold War led to a significant global reduction in bilateral aid supplies, but also that donors were not equally affected by these geopolitical events. Recent global trends may also weaken the impact of short-run domestic factors on aid supplies. The Paris Declaration on Aid Effectiveness, for example, identifies improvements in the delivery of aid as critical for enhancing aid effectiveness, thereby suggesting that recipient- as opposed to donor-country circumstances should drive aid disbursements.

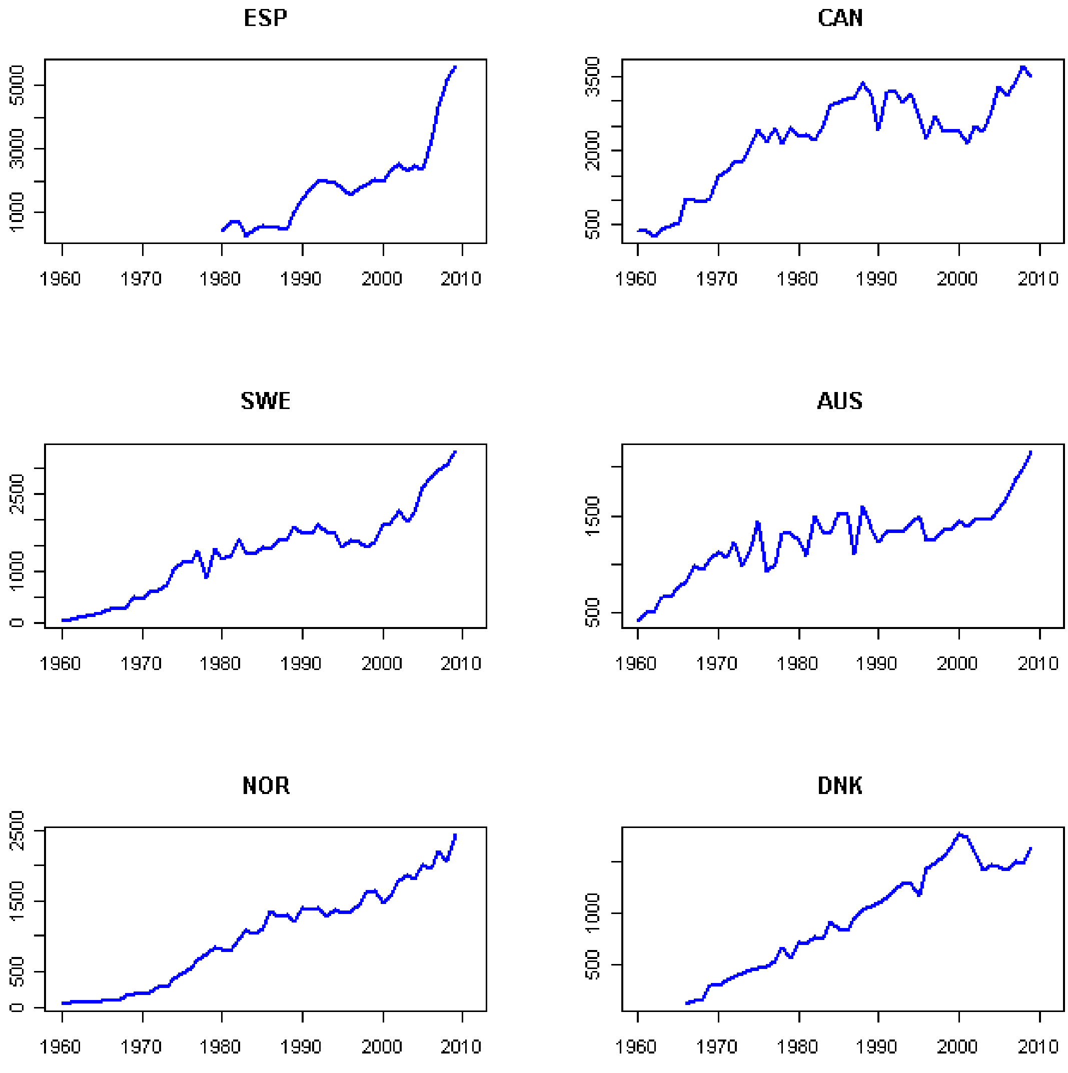
The previous point signals the second contribution – taking heterogeneity between countries and over time seriously. Past studies studies typically employ (static) fixed effects panel estimators, imposing a restriction of homogeneous slope parameters. A brief review of the historical behaviour of bilateral aid flows suggests this assumption may be strong which has implications for the consistency of econometric results (see Section 7.4). Figures 7.1 to 7.4 plot real bilateral aid flows by country in billions of USD, grouping donors by size (see Appendix 7C for details on variables, data sources and country abbreviations used throughout). They indicate very substantial differences in aid behaviour between countries, in terms of both absolute levels and trends. For example, while some donors have consistently increased their aid budgets (e.g., Norway), others display prolonged episodes of either cuts or surges in real disbursements (e.g., Italy; Finland). Country-specific studies also remark on the 'individuality' of specific donors. For instance, Bertoli et al. (2008) highlight the distinct behaviour of Italian aid disbursements compared to its peers, as well as the heterogeneous time series behaviour of aid disbursements (to GDP) across donors. Similarly, Hallet (2009) identifies large differences between donors in the extent (if any) to which they cut aid in response to a domestic recession, in turn suggesting that results for the 'average' donor should be interpreted with care.

Figure 7.1: Bilateral aid flows from larger donors (USD millions), 1960-2009



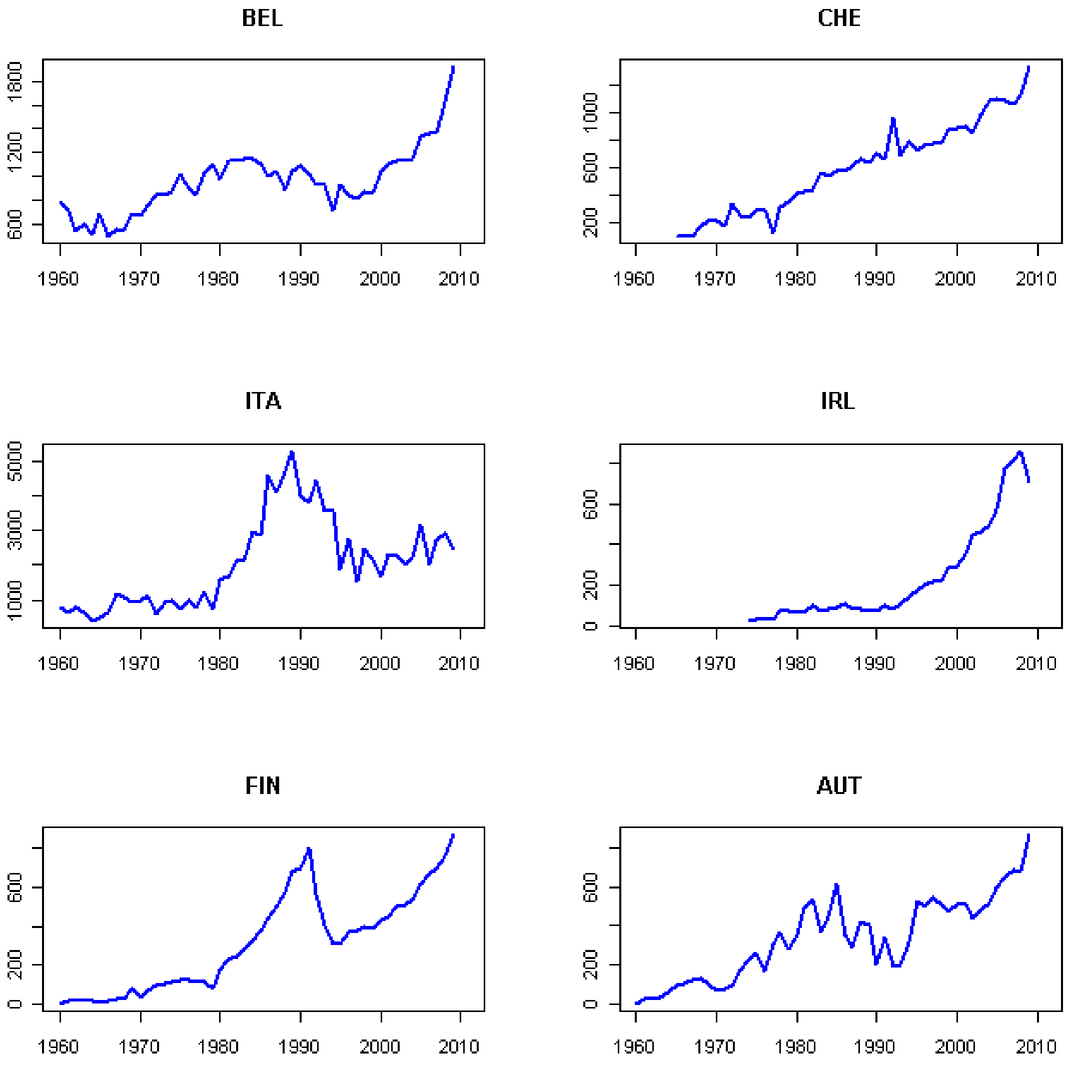
Notes: country codes are as per Appendix Table 7C.1.
Source: author's calculations, see Appendix 7C for variable definitions and sources.

Figure 7.2: Bilateral aid flows from medium-size donors (USD millions), 1960-2009



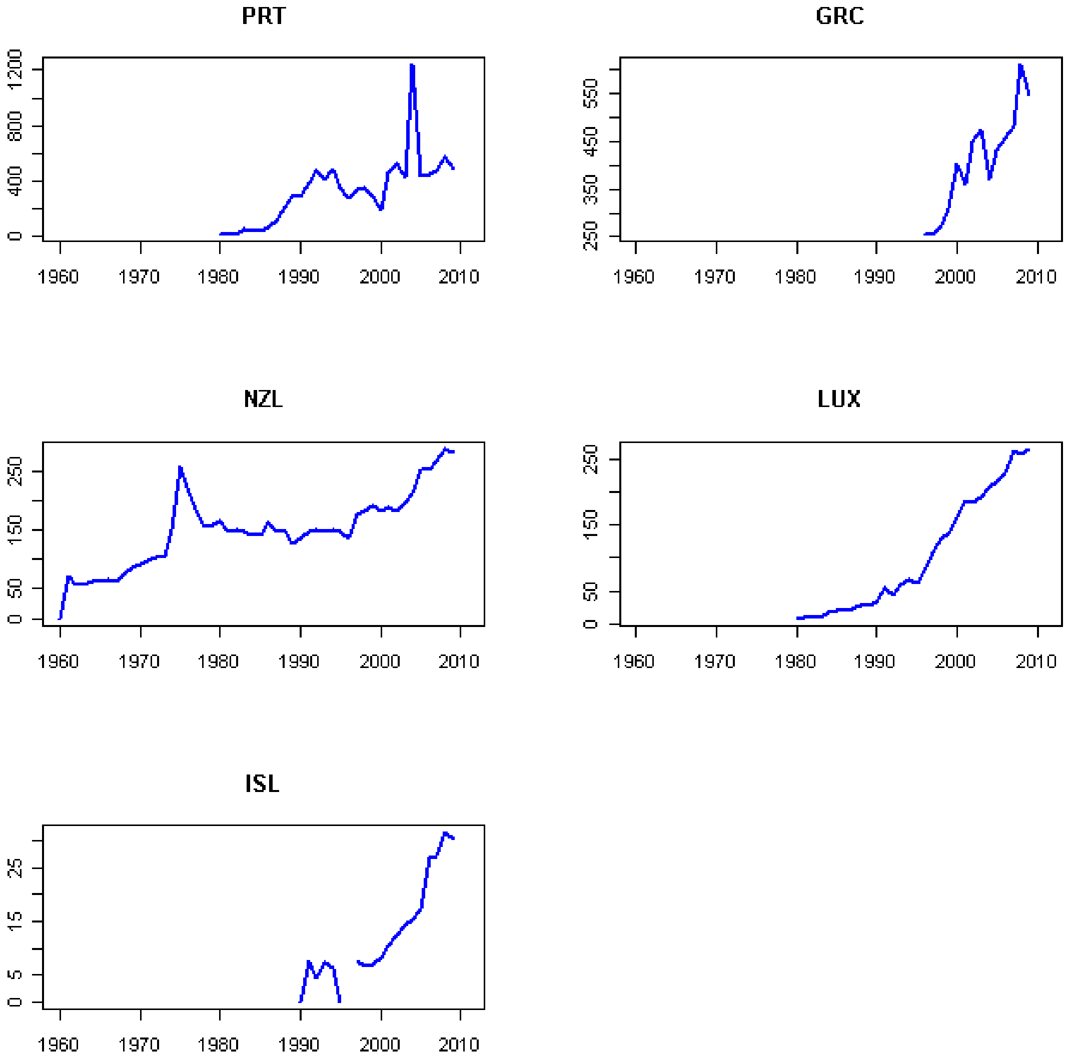
Notes: country codes are as per Appendix Table 7C.1.
Source: author's calculations, see Appendix 7C for variable definitions and sources.

Figure 7.3: Bilateral aid flows from smaller donors (USD millions), 1960-2009



Notes: country codes are as per Appendix Table 7C.1.
Source: author's calculations, see Appendix 7C for variable definitions and sources.

Figure 7.4: Bilateral aid flows from the smallest donors (USD millions), 1960-2009



Notes: country codes are as per Appendix Table 7C.1.
Source: author's calculations, see Appendix 7C for variable definitions and sources.

The figures confirm substantial short-run heterogeneity in terms of responses to macroeconomic shocks. Corresponding to the financial crisis, virtually all countries in the sample faced a severe economic slowdown in 2008 and 2009. However, while some of the smaller donors cut aid disbursements, this was not the case for most of the larger donors.[1] In fact, many of these countries continued to increase disbursements in real terms. As the five largest donors account for around 65 per cent of total bilateral aid (compared to less than 2 per cent represented by the smallest five), 2009 saw a real increase in total bilateral aid of around 14 per cent.

Additionally, the figures indicate that aid is unlikely to be mean reverting for most countries; or, in the minimum, it is highly path-dependent. If aid is not a stationary process, then it is unclear whether standard fixed effects (within) estimators capture anything meaningful as they rely on variation around a unit-specific mean for identification (Smith, 2001). Also, if the variables exhibit cointegration, then results must be interpreted with care in order to differentiate between long-run and dynamic parameters. Consequently, the time series properties of the data must be investigated in order to establish the suitability of chosen empirical techniques.

To underline this point, the dynamic properties of aid are expected to be material due to the way in which aid decisions are made. As noted early on by Mosley (1985), aid outflows are dependent on government budget processes which are both path dependent and temporally lagged – i.e., aid disbursements in year t largely reflect the information set and budget allocation at $t-1$. As a result, aid flows are likely to exhibit state dependence, meaning that past realizations of aid have a direct and independent effect on current realizations. Multi-year aid commitments and long-term diplomatic relationships, in which aid often plays a role, also suggest state dependence may be material, meaning that serial correlation of (estimated) error terms cannot be dismissed as only a problem for the standard errors. Despite the relevance of including dynamics, many previous studies propose static models for aid supply behaviour. Where studies include a lag of the dependent variable in the explanatory model (e.g., Boschini and Olofsgand, 2007; Mold et al., 2010), a significant coefficient bounded between zero and one typically is found, supporting the state dependence thesis. Thus, the third contribution of this chapter is to address the time series properties of the data in a rigorous and consistent way.

Before proceeding, some disclaimers are necessary. This chapter does not address questions regarding where aid is allocated or whether it is effective.[2] Thus, no assumptions are made about how changes in aggregate aid supplies are likely to affect recipient countries or sectors. Additionally, no attempt is made to analyse trends in aid supplied by non-DAC donors (such as China and India), philanthropic organizations, vertical funds (e.g., The Global Fund) or multilateral institutions. These are only excluded for reasons of clarity and length, not because of irrelevance.

7.3 Conceptual framework

Taking guidance from the previous section, it is helpful to start with a clear conceptual framework for donor aid supply decisions. The core assumption of the proposed model is that, over the long-run, donors

[1] Although some of the recent resilience in aid flows may be due to increases in aid to fragile states such as Iraq and Afghanistan, this does not appear to be a single determining factor. For example, aggregate net aid figures for 2008 excluding disbursements to to Iraq, Egypt, Israel increased by 5.6 per cent versus 9 per cent before these exclusions. (Country-specific disbursement data is not available for 2009).

[2] For a recent empirical analysis of the aggregate long-run evidence see Arndt et al. (2010b).

seek to meet a target level of real aid. This target largely reflects various long-run or fixed factors, as described in the literature (see Section 7.2). Nevertheless, the aid target is allowed to vary over time, for example being subject to a time trend as well as certain macroeconomic factors. Although the chosen target can be understood as a stabilizing force or attractor in aid supply decisions, there is considerable scope for fluctuation around the target due to random error as well as unanticipated macroeconomic events. These lead countries to deviate temporarily from the target supply ratio. However, in the event of a deviation from the aid target at time $t-1$, adjustment toward the target is expected to occur in time t but is likely to be incomplete and is potentially subject to new shocks. These assumptions are not chosen for their convenience. A somewhat similar approach is found in Mosley (1985), who postulates that while donor countries may have a "desired" level of foreign aid giving, which reflects domestic incomes and perceptions of aid quality, actual disbursements will deviate from this level due to local budgetary pressures and the behaviour of other donors.

The heterogeneity and time-varying nature of long-run aid targets is explicitly embedded in the proposed model, as is the possibility for short-run variation around these trends.[3] Aid targets are set on a country-specific basis and include time-varying components, denoted by the vector z, and a unit-specific trend. For now, the aid target will remain in general form, denoted by the function $\tilde{\theta}_i(z_t, t)$. Consequently, a very simple model for actual real net disbursements, denoted by $\tilde{a}$, for country i at time t is:

$$\tilde{a}_{it} = \tilde{a}_{i,t-1} \left(\frac{\tilde{\theta}_i(z_t, t)}{\tilde{a}_{i,t-1}} \right)^{\alpha_i} \cdot \tilde{\epsilon}_{it} \tag{7.1}$$

where it is assumed that $\alpha \geq 0$, and $\tilde{\epsilon}$ represents an unknown general error term, about which no assumptions currently are made. Note that the model encapsulates two special cases of aid supply behaviour. If $\alpha = 0$ then aid supplies follow a random walk and the proposition that they are driven by a long-run target does not hold. If $\alpha = 1$ then aid supplies are always equal to the target plus error.

In order to take the model to the data, some transformations and elaborations are necessary. Taking logs, such that $a \equiv \ln(\tilde{a})$, yields a dynamic linear specification:

$$a_{it} = (1 - \alpha_i)a_{i,t-1} + \alpha_i\theta_i(z_t, t) + \epsilon_{it} \tag{7.2}$$

which, in turn, can be stated in error correction form:

$$\Delta a_{it} = \alpha_i \Delta\theta_i(z_t, t) - \alpha_i[a_{i,t-1} - \theta_i(z_{t-1}, t-1)] + \epsilon_{it} \tag{7.3}$$

Next, it is necessary to make some assumptions about the nature of the aid target $\theta_i(z_t, t)$ and the error term. With respect to the former, a simple linear specification is chosen that incorporates country-specific fixed effects and a quadratic time trend. Namely:

$$\theta_i(z_t, t) = \mu_i + \gamma_i y_{it} + \delta_{i1}t + \delta_{i2}t^2 \tag{7.4}$$

which, in the case of $\gamma_i = 1$, translates equation (7.1) into a time-varying target aid-to-income ratio. With respect to the error term, much of the recent literature concerning the impact of the financial crisis on aid supplies can be understood as attempts to isolate factors which lead donors to vary aid disbursements over the short-term. In equation (7.3) such factors are hidden in the error term. Making them explicit, one can assume that the (log) error term is a linear function of additional variables, denoted by the vector X

[3]Please refer to Appendix 7B for a summary of the variables and parameters used in this model.

plus mean zero random error. Giving this a dynamic structure yields: $\epsilon_{it} = \sum_{j=0}^{J} X'_{i,t-j}\beta_j + \eta_{it}$. Thus, re-specifying equation (7.3) to incorporate these two extensions we have a more general error correction model:

$$\begin{aligned} \Delta a_{it} &= -\alpha_i \left[a_{i,t-1} - \{\mu_i + \gamma_i y_{i,t-1} + \delta_{i1}(t-1) + \delta_{i2}(t-1)^2\}\right] \\ &\quad + \lambda_{1i} + \lambda_{2i}t + \lambda_{3i}\Delta y_{it} + \sum_{j=0}^{J} X'_{i,t-j}\beta_j + \eta_{it} \end{aligned} \tag{7.5}$$

where the first three parameters of the second line (RHS) are composites from previous equations, i.e.: $\lambda_{1i} = [\mu_i + \alpha_i(\delta_{i1} - \delta_{i2})]$, $\lambda_{2i} = 2\alpha_i\delta_{2i}$ and $\lambda_{3i} = \alpha_i\gamma_i$.

A few aspects of equation (7.5) are worth highlighting. Filling a gap in the literature (see Section 7.2), there is now an explicit distinction between long-run and short-run determinants of aid supplies. Long-run determinants are given by the levels terms inside the square brackets, which represent the deviation of actual aid from its target at $t-1$. Aside from error correction, short-run movements around this relation are driven by the difference term, the vector X and random noise. This distinction between the timing of effects is important because unconstrained linear estimates of equation (7.5) or its autoregressive distributed lag equivalent would yield coefficients that are composites of the long-run and short-run parameters. State dependence of aid is incorporated via a lag of real aid on the RHS. Also, changes in GDP potentially exert an immediate impact on aid supplies, via the first difference term, as well as a lagged impact via changes to the target level of aid at $t-1$. Thus, if a financial shock is included as an element of X, then this specification allows one to distinguish between direct effects and indirect effects via income or other elements of X.

7.4 Empirical methods

Before turning to the results, it is necessary to reflect on appropriate methods. As already indicated, there are at least three material challenges – (i) heterogeneity in relationships between countries and over time; (ii) the time series nature of the data; and (iii) covariance between units over time. These are discussed in turn.

The potential existence of heterogeneity in economic relations is often discussed in cross-country work, leading to warnings against using pooled estimation results to inform country-specific policy. While this is correct, heterogeneity is not merely a secondary concern. It has implications for the choice of estimation strategy. Lee et al. (1997) show that standard panel estimators incorporating fixed effects and/or instrumental variables will be inconsistent under slope heterogeneity. Similarly, Haque et al. (1999) show that neglect of heterogeneity and dynamics generates misleading inferences about the determinants of savings behaviour across countries.[4] For moderately large panels in both dimensions N and T, a useful approach is the mean group estimator of Pesaran and Smith (1995), which is the simple average of unit-specific OLS coefficients and is an unweighted version of the random coefficients approach due to Swamy (1970). Although this approach is consistent under a wide range of conditions, it may be

[4] As Pesaran and Smith (1995) argue: "aggregating or pooling dynamic heterogenous panels can produce very misleading estimates ... [meaning] that the common assumption of homogeneity in dynamic models is far from innocuous." (p.102). For further discussion of appropriate estimators in heterogeneous (dynamic) panel settings also see Coakley et al. (2006); Pesaran (2006).

inefficient and biased in small samples. Indeed, Mark and Sul (2003) warn against the small sample fragility of single equation estimators, arguing in favour of aggregate (pooled) panel approaches. These concerns are relevant here as the number of countries is modest (N=20), meaning that cross-section averages may be vulnerable to outliers. Consequently, rather than making a prior decision, the empirical approach will be to compare results from a range of panel estimators that make differing homogeneity assumptions (to be tested) and which take identification from either the within- or between- dimensions of the data. Additionally, the stability of coefficients over time will be examined by running estimates on different temporal subsets of the data.

With respect to the second challenge, in order to avoid spurious results it must be established whether the levels relations (equation 7.4) are stationary. This is necessary due to the nature of the real aid and GDP series. Table 7.1 explores this further, giving results of country-specific ADF unit root tests (including a constant and trend term) for these two variables. The cells of the respective columns report the probability associated with the null hypothesis that the variable contains a unit root. The vast majority of these are greater than 0.10, meaning that the null cannot be rejected. The final two rows of theses columns implement the simple panel unit root tests proposed by Maddala and Wu (1999), where the null hypothesis remains that *all* panels contain a unit root. These remain insignificant at the 5 per cent level, indicating that these series are not (individually) trend stationary.[5]

Nevertheless, following equation (7.4), it is plausible that a linear combination of the variables on the form of $a_{it} = \theta_i(z_t, t)$ is stationary. This hypothesis of cointegration can be tested in various ways, depending on the degree of homogeneity imposed on the long-run relations between countries. In the spirit of Engle and Granger (1987), a two-step approach is adopted. First, the long-run relations are estimated, testing for cointegration at the panel- and unit-specific levels. Second, the dynamic model (equation 7.5) is run, inserting lagged values of the estimated residuals from the first step as the error correction term. The potential existence of cointegtation, however, raises the additional problem of the appropriate number of cointegrating relations. *A priori* one expects that national income is (strictly) exogenous to aid. Nevertheless, to correct for any potential bias from the endogeneity of GDP, a Dynamic OLS (DOLS) estimator is used (Stock and Watson, 1993), which is asymptotically unbiased and normally distributed even in the presence of endogenous regressors. Here, this amounts to adding $k = 1$ leads and lags of the differenced RHS variable (real GDP) to the specification, estimated either on a panel basis (see Mark and Sul, 2003) or separately for individual countries.

The third challenge refers to correlation between units in cross-section, which would violate the classical regression assumption of unit independence. Although time dummies are often used to address such effects, a simple and preferable alternative is the common correlated effects (CCE) approach of Pesaran (2006). For the mean group estimators this involves augmenting the model specification with (weighted) averages of the dependent and independent variables and is consistent under both heterogeneity and cross-sectional dependence. This approach is advantageous as it economizes on degrees of freedom and avoids the problem of missing time dummies that can arise in the unbalanced panel case (as here). Moreover, these additional regressors can be added to aggregate panel specifications, possibly allowing for different degrees of homogeneity in the factor loadings.

[5] This finding is not unique. Bertoli et al. (2008) report that unit root tests applied to various measures of the Aid/GDP ratio are unable to reject the null of of a unit root. Note that a variety of (more sophisticated) panel unit root and cointegration tests are also available; however, these are generally not applicable as the panels are unbalanced and the number of panels is modest.

Table 7.1: Unit root and cointegration tests, by country

	Obs	Aid	GDP	CEQs	t_R	t_E	Rejects
AUS	50	0.560	0.056	1	-4.21	-3.40	2
AUT	49	0.267	0.611	1	-3.37	-3.16	1
BEL	50	0.970	0.433	1	-3.65	-3.69	1
CAN	50	0.014	0.355	1	-5.52	-3.95	3
CHE	44	0.822	0.000	.	-5.29	-3.98	2
DNK	44	0.450	0.902	1	-4.79	-3.66	2
ESP	30	0.066	0.627	0	-3.56	-3.52	0
FIN	48	0.822	0.530	.	-6.16	-2.95	1
FRA	50	0.499	0.385	1	-4.88	-3.28	2
GBR	50	1.000	0.245	1	-6.23	-5.34	3
IRL	36	0.129	0.337	1	-3.29	-4.40	2
ITA	50	0.729	0.969	1	-5.18	0.89	2
JPN	50	0.885	0.622	1	-4.35	-4.25	3
LUX	26	0.898	0.690	1	-5.81	-4.97	3
NLD	50	0.014	0.151	1	-4.88	-4.62	3
NOR	50	0.191	0.939	1	-5.71	-2.16	2
NZL	49	0.362	0.145	1	-3.59	-4.60	2
PRT	30	0.540	0.850	0	-5.00	-1.50	1
SWE	50	0.005	0.136	1	-3.94	-3.33	2
USA	50	0.960	0.262	1	-3.75	-2.10	1
χ^2 / mean	45	52.700	52.000	1	-4.70	-3.40	1.9
(prob.)		0.085	0.097				

Notes: Aid is the log. of net bilateral aid excluding debt relief at 2005 prices; GDP also in logs and 2005 constant prices; for Aid and GDP, each country-variable cell reports the probability associated with the null hypothesis that the column variable contains a unit root, calculated via an Augmented Dickey Fuller test (with 3 lags and a trend term); column CEQs reports the number of cointegrating equations between Aid and GDP (on the form of equation 7.4) from a Johansen vector error-correction procedure, chosen by an information criterion; column t_R reports the t-statistic from an Augmented Dickey Fuller test on the residuals from country-specific estimates of the long-run cointegrating relation; t_E reports the t-statistic on the (lagged) error correction term in country-specific dynamic regressions; the final column reports the number of instances the null hypothesis of no cointegration is rejected for the tests in the preceding three columns; final two rows report the test statistic means or, for Aid and GDP, the χ^2 test statistic and associated probability from a Fisher meta-test of the combined column probabilities (see Maddala and Wu, 1999).
Source: author's estimates, see Appendix 7C for variable definitions.

7.5 Results

7.5.1 Long-run trends

Table 7.2 presents results for estimates of the long-run relationship between aid and the proposed aid target equation. Columns (1) to (4) take identification from within-group variation, based on standard fixed effects panel estimators. Column (1) is the basic specification with two-way fixed effects (time and country), column (2) adds the panel dynamic OLS (PDOLS) regressors, column (3) includes additional CCE regressors, and column (4) incorporates both the PDOLS and CCE terms. The results are broadly similar, especially for the coefficients on the log. level of real GDP which capture the income elasticity of aid. However, introduction of the CCE terms, which are inverse population-weighted cross-section averages of aid and GDP at time t, alters the direction and magnitude of the time coefficients both with and without the PDOLS terms. In the former case, these dynamic terms are jointly insignificant whilst all other terms remain essentially unchanged. Hausman tests to compare these different estimates (not reported) reject the null that there are no significant differences between column (4), assumed to be consistent but potentially inefficient, and each of the estimates in columns (1) to (3). However, this is largely due to differences in sample size. When the same estimates are restricted to the smaller sub-sample as per column (4), one cannot reject the Hausman test null for the panel CCE estimates of column (3), implying they are consistent. Therefore, this is the preferred pooled estimator.

Moving to the mean group (MG) estimators, the DOLS terms are also redundant once the CCE terms are already added (not reported). Thus, only results for the MG estimators including the CCE terms are reported, the unweighted version of which is the preferred estimator from the Monte Carlo simulations of heterogeneous panel data described in Coakley et al. (2006).[6] Both the weighted and unweighted results (columns 5 and 6), broadly support the pattern of the within estimators. For the 'average' country one finds a strictly positive income elasticity of aid. The mean aid term remains highly significant, suggesting that cross-section dependence is material in the sense that the contemporaneous aid efforts of other donors affects individual aid decisions. An important difference however, is that the quadratic time trend has shifted to an inverted-U shape, indicating the possibility of downward bias from the pooled estimates due to slope heterogeneity (as per Lee et al., 1997).

The MG estimates confirm very substantial heterogeneity between countries, shown visually in Figures 7.1 to 7.4. The Swamy estimator (column 5) incorporates a natural Hausman-type test of a null hypothesis of slope coefficient homogeneity between countries, which is easily rejected ($\chi^2_{114} = 10840$). Table 7A.1 summarises the country-specific OLS estimates (each incorporating the CCE terms). Three main points stand out. First is the variability in the overall goodness-of-fit of the specification between countries, given by the R-squared statistic, ranging from under 60 per cent in the cases of Belgium and Italy to over 95 per cent for the Nordic donors, among others. This is revealing in itself. A stronger goodness-of-fit is indicative of more stable and, thus, predictable aggregate aid disbursements viewed retrospectively over the long-run, which is likely to have a first order relation to the variability of disbursements at the country-level.[7] Second, the estimated income elasticities of aid span a very wide range, being negative

[6] Following these authors and Pesaran (2006), a consistent non-parametric estimator of the asymptotic variance of the unweighted MG estimator is used here as follows: $Var(\hat{\beta}_i) = \sum_{i=1}^{N}(\hat{\beta}_i - \bar{\beta})^2/[N(N-1)]$

[7] The poor performance of Italy as a donor has been remarked elsewhere (e.g., Bertoli et al., 2008). There are also interesting correspondences between the ranking of countries according to the regression R-squareds and the Center for Global Development's Commitment to Development Index see http://www.cgdev.org/section/initiatives/_active/cdi/.

or insignificantly different from zero in numerous cases (e.g., UK and Italy) to positive and significant in others (e.g., Ireland, Sweden). Third, the quadratic time trends also differ substantially, not only in size but also in their overall direction. For some of the larger donors the shape of the trend is U-shaped, reflecting a declining long-run trend to real aid (especially as a share of GDP) which has only been reversed very recently.

The above results point to extremely heterogeneous long-run tendencies in aid supplies, which in itself validates application of the Pesaran and Smith (1995) MG estimator. This also can be seen visually in Figure 7.5 which compares actual disbursements to the long-run target derived from country-specific empirical estimates of equation (7.4). Broadly speaking for the selected countries, there is a clear co-evolution between the actual and target levels of aid over time which supports the error-correction framework. However, differences between the countries are substantial in terms of both the trend in actual/target aid and the degree to which the actual aid follows its target. For instance, Norway ("NOR") shows a positive upward trend and very stable aid supplies over time. In contrast, Italy ("ITA") shows sharp annual variations in aid as well as changes in the average trend of disbrusements from negative (1960-1980) to positive (1980s) to negative (1990 onwards). The parsimony of the long-run target in equation (7.4) does not fully capture these variations, leading to a much wider gap between targets and actuals in this case, and hence the lower R-squared statistic (Table 7A.1). Overall, this demonstrates the advantages of moving from a (pooled) panel approach to country-specific estimates of aid supply dynamics.

It remains to determine whether these estimates are spurious. Considering the non-stationarity of both aid and GDP at the country-level, tests for cointegration must be applied. Although debate persists regarding the most appropriate of these, at the panel-level the computationally straightforward Dickey-Fuller (DF) tests set out in Kao (1999) are employed and reported in Table 7.2.[8] The preferred DF_γ statistics reject the null hypothesis of no cointegration for all the estimators excluding those which do not adjust for common effects (i.e., columns 1 and 2). Nevertheless, considering the finding of material cross-country heterogeneity, it is meaningful to examine cointegration on a country-by-country basis, based on the underlying equations used to derive the MG results (column 5, Table 7.2). Three different tests are summarised in the final four columns of Table 7.1. The first is based on Johansen's vector error-correction model and identifies the number of cointegrating equations (CEQs) by minimizing the Hannan and Quinn information criterion (Gonzalo and Pitarakis, 1998).[9] Second, the stationarity of the residuals estimated from the country-specific long-run equations are tested using individual Dickey-Fuller tests; relevant t-statistics are reported, t_R. Third, cointegration is verified from an error correction representation where the table reports the t-statistic, t_E, on the lagged error correction term in a simple form of equation (7.5). The final column of the table reports the number of times that the null of no cointegration is rejected from these three tests.[10] The vast majority of countries reject at least two of the tests (and only one country, Spain, none of them), confirming the overall message that the long-run equation estimates are

[8]For more advanced treatment of cointegration tests in the panel context see Persyn and Westerlund (2008); Banerjee (1999). Due to the unbalanced nature of the present dataset, however, tests proposed by these authors cannot by implemented in a straight-forward manner. For instance, the group-mean tests developed by Westerlund (2007) generally reject the null hypothesis of no cointegration for a non-zero fraction of the panel when the dataset is restricted to a balanced subset of panels. Results available on request.

[9]This approach is used as it is least sensitive to the relatively small number of observations available for certain countries; thus fewer countries have 'missing' results compared to alternative tests based on the Johansen procedure.

[10]Asymptotic critical values for the two t-statistics are $t_R < -3.8738$ and $t_E < -3.7782$ based on the methods set out in MacKinnon (1996) and Ericsson and MacKinnon (2002) respectively, calculated using software available on MacKinnon's personal webpage `http://econ.queensu.ca/faculty/mackinnon` that takes into account the number of cointegrating variables, deterministic terms and the quadratic time specification.

Table 7.2: Model for long run bilateral aid supplies, 1960-2009

	(1) FEs	(2) PDOLS	(3) CCE	(4) PDOLS-CCE	(5) Swamy	(6) MG
	b/se	b/se	b/se	b/se	b/se	b/se
GDP (log)	1.84**	2.17**	2.06**	2.36**	2.30**	2.76***
	(0.84)	(0.85)	(0.95)	(0.91)	(0.93)	(0.87)
Time	-0.01	-0.02	0.02	0.02	0.08*	0.09**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Time^2	-0.00	0.00	-0.00	-0.00	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
FD.GDP (log)		1.24		0.77		
		(1.54)		(1.28)		
D.GDP (log)		-1.89**		-1.04		
		(0.68)		(0.66)		
LD.GDP (log)		-0.18		0.58		
		(1.41)		(1.38)		
Aid mean			0.76**	0.80**	0.76**	0.83**
			(0.30)	(0.36)	(0.35)	(0.34)
GDP mean			-1.85	-2.33	-1.86	-2.88*
			(1.73)	(1.63)	(1.56)	(1.49)
Constant	-16.71	-20.48*	1.38	2.61	-1.28	3.29
	(9.92)	(10.03)	(14.19)	(14.56)	(12.72)	(12.21)
N	906	877	906	877	906	906
R2	0.69	0.69	0.69	0.69		0.83
Kao-DF$_\gamma$	-0.11	-0.65	-9.12	-10.87	-8.61	-8.65
prob.	0.46	0.26	0.00	0.00	0.00	0.00
Kao-DF$_t$	-6.70	-7.24	-7.09	-7.73	-5.50	-5.53
prob.	0.00	0.00	0.00	0.00	0.00	0.00

significance: * 0.1, ** 0.05, *** 0.01

Notes: dependent variable is the log of net bilateral aid; columns (1) to (4) apply pooled panel estimators, which are respectively a simple fixed effects (FEs) estimator, FEs augmented by panel dynamic OLS (PDOLS) terms, FEs augmented by common correlated effects (CCE) terms, and FEs with both PDOLS and CCE terms; column (5) is the Swamy (1970) random coefficients estimator; column (6) is the unweighted mean-group estimator of Pesaran and Smith (1995); all models use robust estimates of the variance-covariance matrix.

Source: author's estimates, see Appendix 7C for variable definitions.

Table 7.3: Error correction model for bilateral aid flows, 1960-2009

	(1) PCSE b/se	(2) IV b/se	(3) GMM b/se	(4) Swamy b/se	(5) MG b/se	(6) MG b/se
Error correction (lag)	-0.12***	-0.13***	-0.75***	-0.25***	-0.30***	-0.48***
	(0.01)	(0.02)	(0.22)	(0.06)	(0.05)	(0.07)
Δ Real GDP	1.14***	2.25	2.07	0.41	0.01	0.41
	(0.37)	(3.14)	(6.84)	(1.12)	(0.96)	(0.96)
Δ Gov. spend in GDP	0.70***	0.56	1.28	0.56	0.64	0.65
	(0.17)	(0.47)	(4.24)	(0.51)	(0.42)	(0.40)
Δ Govt. saving in GDP	-0.51	-0.96	7.52	-0.87	-0.22	-0.39
	(0.40)	(1.26)	(12.36)	(0.94)	(0.68)	(0.69)
Δ Unemployment	-0.77	0.06	11.17	0.87	2.24	1.56
	(0.70)	(3.13)	(19.54)	(3.40)	(3.07)	(3.21)
Δ Upper income GDP	-0.37	-1.37	-5.68	0.26	0.55	0.17
	(0.42)	(1.40)	(7.28)	(1.18)	(1.02)	(0.94)
Δ Low income GDP	0.09	-0.04	0.49	-0.04	-0.05	-0.12
	(0.13)	(0.25)	(1.66)	(0.27)	(0.21)	(0.19)
Mean Δ aid	1.00***	0.97***	0.69*	0.83***	0.95***	0.83***
	(0.10)	(0.19)	(0.37)	(0.25)	(0.20)	(0.19)
Δ Democracies (%)	-0.02	0.12	0.58	-0.18	-0.12	0.18
	(0.29)	(0.54)	(2.85)	(0.71)	(0.60)	(0.63)
Bank crisis (dummy)	0.02	0.01	0.04	0.01	0.02	0.02
	(0.02)	(0.04)	(0.09)	(0.05)	(0.04)	(0.03)
Time	0.00	0.00	-0.00	0.00	0.00	-0.00
	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Constant	-0.07	-0.05	0.10	-0.07	-0.08	-0.04
	(0.05)	(0.13)	(0.58)	(0.09)	(0.08)	(0.07)
N	735	725	682	735	735	735
R2	0.25				0.49	0.54
Chi2	1199.5	189.7	7608.2	70.9	234.0	251.1

significance: * 0.1, ** 0.05, *** 0.01

Notes: dependent variable is the first difference of the log. of net bilateral aid; columns (1) to (3) are pooled models estimated respectively by a Prais-Winsten regression including 5 year period and country dummies, a 2SLS panel fixed effects instrumental variables estimator (including period dummies), and the Arellano-Bond GMM procedure; all pooled models define the lagged error correction term (L.EC) from residuals estimated from Table 7.2 column (4); column (4) uses the Swamy (1970) random coefficients estimator and defines L.EC from residuals estimated from Table 7.2 column (5); columns (5) and (6) use the Pesaran and Smith (1995) mean-group estimator; column (5) defines L.EC from the residuals estimated from the averaged coefficients in Table 7.2 column (6) ; column (6) defines L.EC from the country-specific residuals associated with Table 7.2 column (6); column (1) reports panel-corrected standard errors (which adjust for panel-specific autocorrelation and heteroskedasticity as well as cross-unit correlation); all other models report robust (sandwich) estimates of the variance-covariance matrix.

Source: author's estimates, see Appendix 7C for variable definitions.

Table 7.4: Error correction model of bilateral aid flows, various data subsets

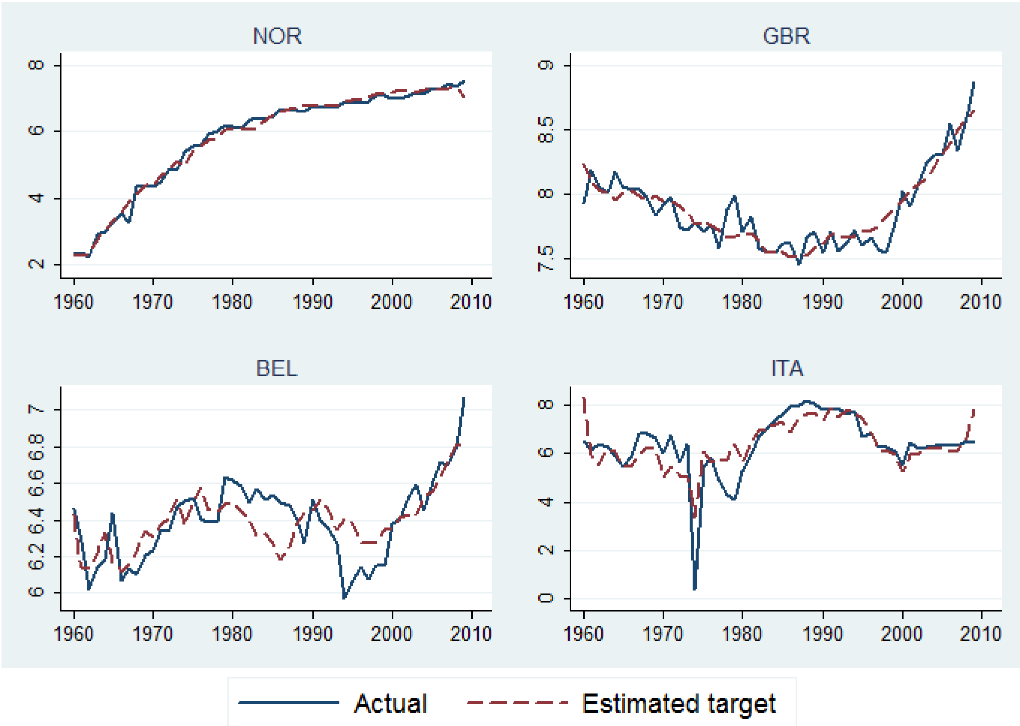
	(1) 1960-91	(2) 1992-09	(3) 1990s	(4) 2000s	(5) Donors-I	(6) Donors-II
	b/se	b/se	b/se	b/se	b/se	b/se
Error correction (lag)	-0.25***	-0.29***	-0.35***	-0.25***	-0.19***	-0.35***
	(0.03)	(0.04)	(0.06)	(0.06)	(0.05)	(0.04)
Δ Real GDP	0.00	1.55***	2.19**	1.34**	1.10*	-0.48
	(0.55)	(0.55)	(0.88)	(0.56)	(0.58)	(0.55)
Δ Gov. spend in GDP	0.85***	0.73***	0.61	1.00***	0.96***	0.54**
	(0.29)	(0.27)	(0.39)	(0.33)	(0.29)	(0.26)
Δ Govt. saving in GDP	0.30	-1.07**	-2.24***	-0.11	-0.70	-0.48
	(0.92)	(0.54)	(0.87)	(0.63)	(0.86)	(0.54)
Δ Unemployment	-1.79	-0.44	-0.19	-1.06	-0.15	-4.21***
	(1.13)	(0.88)	(1.36)	(1.27)	(1.39)	(1.02)
Δ Upper income GDP	0.54**	-0.76	0.03	-1.88**	-0.34	0.01
	(0.27)	(0.71)	(1.10)	(0.95)	(0.62)	(0.70)
Δ Low income GDP	0.15	-0.17	-0.32*	0.05	0.17	0.08
	(0.14)	(0.17)	(0.17)	(0.27)	(0.19)	(0.21)
Mean Δ aid	1.06***	0.79***	0.69***	0.80***	0.84***	1.14***
	(0.12)	(0.17)	(0.12)	(0.23)	(0.16)	(0.16)
Δ Democracies (%)	-0.19	0.70**	0.52	-0.65	-0.47	0.77*
	(0.36)	(0.33)	(0.37)	(1.09)	(0.45)	(0.43)
Bank crisis (dummy)	0.04	0.02	-0.05**	0.02	0.05	-0.05*
	(0.12)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)
Time	0.00	0.00**	0.00	0.01*	0.01*	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-0.00	-0.24***	-0.12	-0.30**	-0.13*	0.02
	(0.05)	(0.08)	(0.13)	(0.15)	(0.07)	(0.12)
N	382	353	194	196	413	322
R2	0.18	0.28	0.43	0.24	0.19	0.32
Chi2	1000.6	213.7	243.3	97.4	216.0	239.6

significance: * 0.1, ** 0.05, *** 0.01

Notes: underlying model is as per Table 7.3; columns (1) to (4) restrict the sample to the time periods indicated and include dummies for large, medium and small groups of donors (not reported); column (5) restricts the sample to donors whose average net real ODA is above US$1 billion (see Table Appendix Table 7C.1); column (6) is restricted to all other donors; both columns (5) and (6) include individual donor and 5 year time period dummies; all models estimated by a Prais-Winsten regression using panel-corrected standard errors, which adjust for panel-specific autocorrelation and heteroskedasticity as well as cross-unit correlation.

Source: author's estimates, see Appendix 7C for variable definitions.

Figure 7.5: Actual aid supplies versus estimated aid targets (selected countries)



Notes: all aid variables expressed in log. form; estimated long-run targets are the fitted values from country-specific estimates of equation (7.4) including CCE terms for aid and GDP, as per the MG estimator described in the text; actual aid is observed disbursements.

Source: author's calculations.

not spurious, at least once the CCE terms are included.

7.5.2 Short-run dynamics

Table 7.3 reports alternative estimates of the dynamic model, based on the error correction representation given by equation (7.5). Following the two-step procedure, estimated residuals from alternative long-run estimates are employed as the error correction terms (lagged). Results from a variety of different estimators are reported in order to verify the degree of sensitivity to pooled versus MG estimation, as well as to the use of residuals (error correction terms) derived from the range of results reported in Table 7.2. Variables employed to represent short-term 'shock' influences on aid disbursements follow the existing literature (see Section 7.2). For instance, Dang et al. (2009) find a significant relationship between aid disbursements and the government fiscal balance, unemployment rates and banking crises. Various CCE-type terms are also added – namely, the average (log.) growth rates of upper and lower income countries, the contemporaneous cross-section average change in aid disbursements and the percentage point change in the share of countries in the globe with competitively elected legislatures ('democracies').

The chosen specification includes only one variable in lagged form, the error correction term. Thus, all shock terms enter at time t only. This is primarily to conserve degrees of freedom, particularly for the estimators derived from country-specific estimates. To account for potential serial correlation in errors at the panel level, however, column (1) of Table 7.3 employs a Prais-Winsten panel estimator which adjusts for panel-specific autocorrelation and contemporaneous cross-panel correlation.[11] This specification employs the residuals estimated from the preferred long-run pooled estimates (Table 7.2, column 3) to define the error correction term and incorporates both 5-year period and country dummies. Objections may be raised, however, that this will be inconsistent if changes in GDP (ΔGDP) are endogenous. Consequently, panel instrumental variables techniques are employed – column (2) runs a two-stage least-squares within estimator, where ΔGDP is instrumented by its second and third lagged levels. To address both omitted dynamics and endogeneity concerns, column (3) runs the Arellano-Bond GMM estimator (Arellano and Bond, 1991), which incorporates one lag of the dependent variable (not reported; highly insignificant) and treats ΔGDP as endogenous. Columns (4) to (6) apply the MG estimators described previously – column (4) is the Swamy weighted estimator and defines the error correction term from the residuals derived from the estimates in column (5) of Table 7.2; column (5) is the unweighted MG estimator, where the error correction term is generated from the residuals from the averaged estimates in Table 7.2 column (6); column (6) repeats this estimator but uses residuals derived from the country-specific long-run equations, thereby more fully accounting for heterogeneity in these relations.

The results are broadly comparable. For the majority of coefficients, excluding the error correction terms, the estimates are well within sampling variation (95 per cent level). The error correction terms are highly significant in all specifications, with a t-statistic of over ten in many cases. This confirms the merit of the conceptual framework developed herein. Nevertheless, point estimates of the speed of error correction are sensitive to the empirical approach employed. For example, where allowances are made for greater heterogeneity between countries (e.g., column VI), the 'average' speed of error correction appears to be larger than in the pooled estimates of columns (1) and (2). Although the majority of estimates for the shock variables retain the same sign and approximate magnitude, they are often insignificant, implying there is no unambiguous relationship between changes in aid and these contemporaneous macroeconomic events (including the GDP growth rate). In some columns, aid appears to be positively related to the share of non-aid government spending in GDP (elasticity of around 0.7), but this is not consistently significant across estimators. The cross-section average change in aid is highly significant and close to one in all estimates. This points to material peer aid effects, as found elsewhere (e.g., Mosley, 1985; Round and Odedokun, 2004), further supporting the relevance of directly addressing cross-unit covariance in the estimation strategy. Lastly, the two instrumental panel estimators (columns 2 and 3) do little to alter the overall insight, but rather carry very large standard errors.

Most importantly, the results point to substantial heterogeneity between countries in aid supply decisions. Indeed, the test for coefficient stability across units is rejected (based on the Swamy estimator in column IV; $\chi^2_{209} = 348.03$). Table 7A.2, which summarises the country-specific estimates of the short-run model, shows large differences across the estimated coefficients in terms of both their magnitude and sign. Critically amongst these, the speed of error correction differs by a factor of around three (e.g., USA = 0.25; Denmark = 0.83) and is insignificantly different from zero in three cases (Portugal, Italy and Switzerland). The estimated partial dynamic effect of a banking crisis on aid disbursements also exemplifies the extent of slope heterogeneity. It is insignificant in most cases (at the 5% level), but significantly negative in only two cases – Norway and Finland – for cutting aid in response to domestic banking problems in the 1990s.

[11]Implemented in Stata v11.1 via the `xtpcse` command, with options `corr(psar1) np1`.

To get a further sense of the extent of this heterogeneity, Table 7.4 estimates the same dynamic model for different data subsets – namely, the pre- and post-Cold War period (in columns 1 and 2); specific decades (columns 3 and 4); larger donors (column 5, consisting of the 'large' and 'medium' donor groups in Appendix Table 7C.1); and smaller donors (column 6, comprising the remaining donors in Appendix Table 7C.1). Due to the smaller number of observations available for each subset, the MG estimators cannot be used. Thus, the Prais-Winsten panel estimator is employed (as per column 1, Table 7.3), but in light of the heterogeneity in dynamic relations, the error correction term derived from the country-specific long-run estimates (column 6, Table 7.3) is employed. Despite use of this pooled estimator, the results give further credence to the extent of heterogeneity in aid supply dynamics. This applies between groups of countries and between different time periods, and is evident in the variation across columns in the coefficient estimates and regression summary statistics. In keeping with Boschini and Olofsgand (2007), the end of the Cold War is reflected in substantive changes in the determinants of foreign aid. Before 1991 the model performs relatively poorly, with few coefficients significantly different from zero. After 1991, however, aid appears to be more systematically affected by macroeconomic events and there is a significant positive relationship between changes in aid and changes in the global share of democracies. One interpretation of this shift is that during the Cold War, foreign aid was used by some donors as a geo-political tool to support pro-Western regimes. Following the fall of the Berlin Wall this motive no longer held; rather, donors became more selective in their choice of recipients, often using aid as a tool to promote democratic rule. Even so, the idea of two historical regimes in aid may be overly simplistic. Columns (3) and (4) of the table also suggest material differences between the 1990s and 2000s, with the most recent years showing a moderate trend increase in aid but weaker error correction and a lower overall model fit (R-squared).

Comparing across groups of larger and smaller donors, further differences are apparent. For example, smaller donors appear to be more sensitive to average changes in global governance than are larger donors. The average larger donor shows a somewhat slower pace of error correction and greater sensitivity to changes in domestic GDP. Also, only the disbursements of smaller donors tend to be cut during domestic banking crises, controlling for contemporaneous changes in macroeconomic indicators. However, this appears to be driven principally by events in the 1990s (see column 3), but is not apparent in any other time period. Drilling deeper, this result is likely to reflect the specific effects of the Nordic banking crisis, discussed in Roodman (2008), as both Norway and Finland report significant negative coefficients on the bank crisis dummy variable from the underlying country-specific estimates (Table 7A.2).

7.5.3 Interpretation

What are we to make of these findings? As found in other settings (e.g., Haque et al., 1999; Lee et al., 1997), the results confirm that a neglect of dynamics and heterogeneity can lead to inconsistent and misleading conclusions about cross-country behaviour. This is made plain in Table 7A.3, which reports alternative estimates of the determinants of the level of aid on the form of equation (7.2). Columns (1) to (3) are static regressions, allowing for differing degrees of heterogeneity. Note that estimates in column (2) broadly replicate findings from previous studies (e.g., Dang et al., 2009), being a static fixed effects regression. Columns (4) to (6) include a dynamic term, again allowing for differing degrees of heterogeneity. Inferences about aid supply behaviour are highly dependent on the estimator chosen. Many variables that appear significant in the static regressions with slope homogeneity imposed (columns 1 and 2), lose significance when one allows for slope heterogeneity (column 3) and/or dynamics (columns 4 to 6). In the latter estimates, consistent estimates of the speed of dynamic adjustment also requires that heterogeneity is addressed. The findings of this chapter demonstrate that dynamics and heterogeneity are both material, meaning that only the MG results (column 6) are likely to be consistent.[12] Also, despite the moderate sample size (in dimension N), the MG results appear not unduly affected by outliers, as is evident from a comparison of the mean and median coefficient estimates from Tables 7A.1 and 7A.2

An additional finding, not directly apparent from levels regressions, is the merit of differentiating between long- versus short-term influences on aid supplies. This distinction reflects the non-stationarity of aid and is naturally incorporated in the conceptual framework proposed in Section 7.3. In all but a few cases, the error correction representation finds strong empirical support. Nevertheless, the specific parameters of this model vary substantially, pointing to fundamental heterogeneity in aid supply behaviours in terms of both long- and short-run parameters. Indeed, compared with previous studies, the present results are substantially more ambiguous precisely due to the dominance of heterogeneity in both N and T dimensions. This is a tangible example of where the use of 'average' results to inform about expected behaviour at the country-level would be misleading. However, this is not to say that 'average' parameter estimates are redundant. The first point is that they must be estimated consistently, which means that standard panel estimators are unlikely to be reliable. Secondly, 'average' results must be interpreted carefully – i.e., as convenient ways to describe (true) parameter distributions that span a wide range.

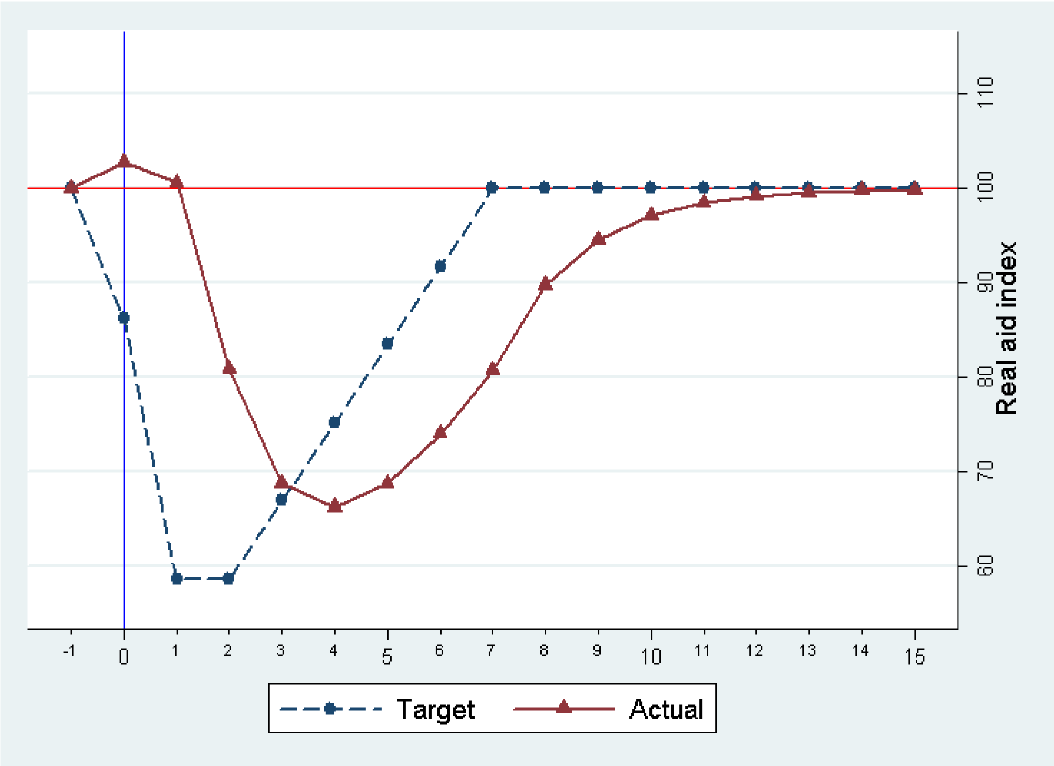
With these points in mind, it is useful to reflect on how our preferred results alter our understanding of aid supply behaviour, particularly the response to domestic shocks. Taking the main coefficients from column (6) of Table 7.3 as the preferred estimates of the average past response to domestic economic conditions, an interesting pattern emerges. Figure 7.6 plots the expected dynamic path of aid supply occurring after a simulated set of shocks to macroeconomic aggregates at time t=0 and t=1, followed by a 5 year recovery. The timing and magnitude of the shocks are chosen to loosely replicate some of the larger shocks to real variables observed in response to the crisis (e.g., an initial shock in 2008, worsening in 2009). Critically, the figure shows a large divergence between the path for actual aid and that of the target, which refers to the long-run relation. The former is both smoother and more moderate than the latter, capturing the fact that the estimated long-run target immediately 'absorbs' the full magnitude of shocks. Short-term influences on aid, including the error correction parameter, modify the extent of immediate cuts to aid. In the present case, they temporarily push real disbursements upward as the share of spending in GDP rises (e.g., due to automatic stabilizers and the contraction of GDP). Thus, it takes two periods after the shock for actual aid to be materially lower than its base value and the direction of error correction remains consistently downward, taking around 4 periods for actual aid to bottom out after the initial shock. In sum, this indicates that although initial aid supply responses to negative shocks may be moderate, 'full' adjustment can be large and occurs over a much longer period.

7.6 Conclusion

This chapter has taken advantage of increasingly rich time series data about bilateral supplies of foreign aid. Building on the existing literature, a principal contribution was to take into account heterogeneity between donors, differences between long and short term response factors and the non-stationary nature of the data. In doing so, it has been established that an error correction model for aid supplies is well supported in the data for most countries. This means that bilateral aid supplies have broadly followed

[12]The results in column (6) of Table 7A.3 are comparable to those of columns (5) and (6) of Table 7.3. The latter results, however, distinguish between long- and short-run parameters, are based on the two-step procedure and include CCE terms.

Figure 7.6: Simulation of aid supply response to macroeconomic shocks



Notes: simulation is based on coefficients from column (6) of Table 7.3. Assumed real shocks at time [x axis] t=0 are ΔGDP = -5%, ΔGovt. saving = -3.5%, ΔGovt. spending = +3%; at t=1 the same shocks are repeated, at double their previous magnitude; at t=2 there are no external changes to the system; recovery occurs in a linear fashion from t=3 to t=7, after which there are no further external changes; both random error and trend changes in aid are assumed to be zero throughout.

Source: author's calculations.

long-run trends, rooted in fixed and slow-moving factors, but have also shown substantial short-run variation. At the same time, heterogeneity both between countries and over time is substantial. Thus, where feasible, country-specific estimates of aid supply behaviour are to be relied upon in place of pooled panel estimators.

Two more specific lessons also stand out. First, there is evidence that aid supply behaviours continue to evolve over time (on average). During the Cold War, bilateral supplies of aid were comparatively insensitive to domestic macroeconomic events or global governance conditions. Since 1991, short-run changes in aid appear to be more responsive both to improvements in global democracy and to changes in domestic macroeconomic conditions. The first decade of this century also marks differences with the 1990s (on average), including a trend increase in aid and the absence of any systematic direct supply response to domestic banking crises. This leads to the second lesson. Previous studies have been too confident in their ability to predict aid supplies, in many cases forecasting a drastic and rapid reduction in aid in response to the present financial crises. To date, neither the average nor aggregate aid volumes have fallen as expected. To a certain extent this is broadly consistent with the present results, which would predict a lagged but more persistent response (fall in aid) operating through macroeconomic aggregates and the error correction mechanism. However, given the fundamental heterogeneity in the historical data, one must be sceptical of our ability to use the past to predict the future.

On a more positive note, the cautious nature of these findings does not inevitably lead to pessimistic conclusions. Distinct differences between donors is evident in both long-run trends and short-run dynamics. Thus, the kind of analysis undertaken here can be used to make comparisons between donors and advocate against poor performance or unpredictable behaviour (as shown by Italy). Also, the capacity of donors to change their aid supply behaviour over time can be seen in a positive light. The moderate and declining success in explaining short-term aid dynamics, particularly in recent years, may signal that aid is becoming less responsive to domestic factors than in the past. If so, this could have positive implications for aid predictability at the aggregate level. Consequently, the research agenda remains open as new and better data becomes available. In particular, as the number of observations for individual countries increases, country-specific models that combine donor and recipient behaviours will need to be elaborated.

7A Appendix: additional tables

Table 7A.1: Country-specific long-run estimates, summary

Country	Obs.	R^2	GDP log	Time	$Time^2$	Aid mean	GDP mean
AUS	50	0.81	1.58	-0.09***	0.00*	0.24	1.38
AUT	49	0.66	-0.98	0.14	-0.00**	0.27	1.84
BEL	50	0.49	5.81***	-0.14***	0.00***	0.61***	-3.53***
CAN	50	0.92	-0.94	-0.01	-0.00***	1.08***	4.03***
CHE	44	0.98	1.24*	0.12***	-0.00***	0.45***	-1.45*
DNK	44	0.97	-0.74	0.08**	-0.00***	-0.18	3.17***
ESP	30	0.76	13.56	0.17	-0.00	-0.86	-19.27*
FIN	48	0.94	4.35***	0.20***	-0.00***	1.72***	-8.16***
FRA	50	0.71	3.54***	-0.00	0.00	0.55***	-5.35***
GBR	50	0.83	-2.10*	-0.12***	0.00***	0.26	2.86***
IRL	36	0.96	4.46***	0.28***	-0.00***	-0.04	-11.54***
ITA	50	0.57	-0.29	0.51	-0.00	4.82***	-17.09***
JPN	50	0.89	0.72	0.10**	-0.00**	0.02	-1.31
LUX	26	0.99	1.65	0.45***	-0.01***	-0.09	0.96
NLD	50	0.93	0.52	0.05**	-0.00***	0.27	1.03
NOR	50	0.99	3.60***	-0.01	-0.00***	0.82**	1.81**
NZL	49	0.78	2.00**	-0.02	0.00	0.37*	-0.01
PRT	30	0.85	9.94***	0.22	-0.01*	5.20***	-7.59
SWE	50	0.96	6.38***	0.01	-0.00***	0.40	1.06
USA	50	0.62	0.87	-0.10***	0.00***	0.60*	-0.47
Mean			2.76	0.09	-0.00	0.83	-2.88
Median			1.61	0.06	-0.00	0.38	-0.24
St.dev			3.90	0.18	0.00	1.52	6.67

significance: * 0.1, ** 0.05, *** 0.01

Notes: columns report the underlying country-specific (single equation) parameter estimates from which the mean-group estimator given in Table 7.2 column (6) is derived; R^2 refers to the R-squared of each regression; summary statistics refer to the vector of estimated parameters for each variable.

Source: author's estimates, see Appendix 7C for variable definitions.

Table 7A.2: Country-specific short-run estimates, summary

Country	R^2	L.EC	ΔGDP	ΔSpend	ΔUnemp.	Av. ΔAid	ΔDemoc.	Bank crisis
AUS	0.42	-0.54***	3.39**	1.15	2.35	0.37	0.36	0.03
AUT	0.49	-0.21**	-9.35**	0.43	-29.06	2.67***	6.64*	-0.27
BEL	0.53	-0.25**	-2.94***	0.46	-6.09*	0.16	-1.49**	0.07
CAN	0.49	-0.25*	0.88	-3.60*	-1.76	1.72***	-0.96	-0.07
CHE	0.58	-0.34	-1.12	-1.21	5.82	0.21	-1.68	0.01
DNK	0.50	-0.82***	-1.33	-0.05	1.76	-0.32	-0.31	0.07
ESP	0.49	-0.51**	0.05	1.83	-5.60	0.65	3.36	0.41
FIN	0.53	-0.41***	4.38	2.74**	5.21**	1.54***	-0.83	-0.14*
FRA	0.54	-0.93***	7.79**	0.16	8.20	0.61	2.97	0.25**
GBR	0.58	-0.76***	-1.78	-1.48**	0.57	0.44	-1.10	-0.10
IRL	0.59	-0.51***	-1.47	0.56	-2.82	-0.20	-0.13	-0.16
ITA	0.36	0.18	4.68	4.99	28.28**	2.32	-7.15**	0.09
JPN	0.55	-0.62***	1.92	-0.38	41.26***	1.39***	0.67	0.08
LUX	0.88	-1.35***	4.17**	0.30	4.96	0.11	-2.15	0.03
NLD	0.39	-0.47***	-2.44	-0.20	-4.95	-0.06	-1.90*	0.11
NOR	0.57	-0.39***	2.33	-0.13	-2.37	0.37	0.48	-0.18***
NZL	0.67	-0.55**	-0.95	1.15	-0.92	0.70	1.44	-0.05
PRT	0.68	-0.05	-8.32*	2.11	-17.84**	1.55	1.27	0.12
SWE	0.62	-0.56***	4.42**	1.87***	5.29*	1.88**	3.78*	0.05
USA	0.39	-0.26***	3.91*	2.27	-1.18	0.57	0.43	-0.03
Mean		-0.48	0.41	0.65	1.56	0.83	0.18	0.02
Median		-0.49	0.46	0.44	-0.17	0.59	0.12	0.03
St.dev		0.33	4.31	1.79	14.36	0.86	2.81	0.15

significance: * 0.1, ** 0.05, *** 0.01

Notes: columns report the underlying country-specific (single equation) parameter estimates from which the mean-group estimator given in Table 7.3 column (6) is derived; selected variables shown; R^2 refers to the R-squared of each regression; summary statistics refer to the vector of estimated parameters for each variable.

Source: author's estimates, see Appendix 7C for variable definitions.

Table 7A.3: Static and dynamic levels regressions, 1960-2009

	(1) RE b/se	(2) FE b/se	(3) MG b/se	(4) FE b/se	(5) GMM b/se	(6) MG b/se
L.Aid				0.72***	0.43***	0.18***
				(0.07)	(0.16)	(0.06)
GDP (log)	1.14***	2.00	1.89	0.49	1.87***	1.64*
	(0.14)	(1.41)	(1.48)	(0.34)	(0.61)	(0.87)
Spend / GDP	6.12***	6.74**	3.85*	1.57*	4.89*	3.65*
	(2.14)	(2.39)	(2.04)	(0.87)	(2.53)	(2.09)
Save / GDP	3.71*	3.47	-1.98	0.39	2.21	-1.72
	(2.16)	(2.46)	(1.99)	(0.65)	(1.89)	(1.88)
Unemployment rate	-0.04*	-0.03	-0.01	-0.02*	-0.00	-0.00
	(0.03)	(0.04)	(0.06)	(0.01)	(0.02)	(0.04)
Democracies (%)	-0.22	-0.05	-1.62**	-0.18	-0.29	-1.15
	(0.75)	(0.88)	(0.71)	(0.23)	(0.37)	(0.84)
Bank crisis (dummy)	0.15*	0.12	0.09	0.03	0.03	0.04
	(0.08)	(0.08)	(0.08)	(0.04)	(0.05)	(0.05)
Time	-0.02	-0.07	0.05	-0.02	-0.09**	0.03
	(0.04)	(0.09)	(0.08)	(0.02)	(0.04)	(0.09)
$Time^2$	0.00	0.00	-0.00	0.00	0.00**	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-10.16***	-20.62	-18.86	-4.59	-20.48***	-17.38
	(1.87)	(16.55)	(17.94)	(3.98)	(7.29)	(10.95)
N	773	773	773	763	742	763
R2		0.46	0.88	0.77		0.89
Chi2	168.5		39.7		404.4	117.4

significance: * 0.1, ** 0.05, *** 0.01

Notes: dependent variable is the log. of net bilateral aid; columns (1) to (3) are static panel regressions estimated respectively by random effects, fixed effects, and the mean-group estimator (Pesaran and Smith, 1995); columns (4) to (6) add a lag of the dependent variable and are estimated respectively by fixed effects, Arellano-Bond GMM, and the mean-group estimator; all standard errors are based on robust estimates of the variance-covariance matrix.

Source: author's estimates, see Appendix 7C for variable definitions.

7B Appendix: summary of variables and parameters

Variables	Definition
a	Log. of real net disbursements of aid
t	Time (year)
z	Vector of variables determining long-run aid target
X	Vector of variables determining short-run variations in aid supply
y	Log. GDP
μ	Country-specific fixed effect (intercept) determing long-run aid supply
Parameters	**Definition**
α	Error correction parameter
γ	Elasticity of aid with respect to income
δ_1	Change in (log). of long run aid supplies with respect to time
δ_2	Change in (log). of long run aid supplies with respect to time squared

7C Appendix: data description & sources

The majority of data is taken from OECD statistical series for the 1960-2009 period – see summary table below. For consistency between countries and over time, all money-valued variables are stated in constant 2005 USD with a PPP adjustment. The principal unit of observation is the members of the DAC, which covers all major donors in advanced countries. Newer members such as South Korea, Turkey and ex-Soviet bloc Eastern European countries are excluded due to the comparatively low value and predominantly local focus of their aid disbursements, as well as the small number of available observations. Of the remaining N=23 countries, a further 3 countries are excluded (Germany, Iceland and Greece) from the regression analysis due to the smaller number of valid observations (T<20). Aid supplied by individual countries is defined as total net bilateral aid disbursements minus debt relief, which excludes disbursements to multilateral organizations but includes support to NGOs and international private organizations. All definitions follow those of the Development Assistance Committee (DAC) of the OECD.

Table 7C.1: Sample coverage and summary statistics

Country	Code	Group	Obs.	Net bilateral aid (1960-2009) Mean	Median	Change
United States	USA	Large	50	14,677	14,066	1.080
Germany	DEU	n/a	19	6,733	6,426	0.670
Japan	JPN	Large	50	5812	6,236	3.290
France	FRA	Large	50	5575	5,480	-0.215
United Kingdom	GBR	Large	50	4,369	3,802	1.922
Netherlands	NLD	Medium	50	2,600	3,086	5.192
Canada	CAN	Medium	50	2,224	2,393	4.703
Italy	ITA	Medium	50	2,055	1,929	4.358
Spain	ESP	Medium	30	1,866	1,831	12.39
Sweden	SWE	Medium	50	1,348	1,451	12.05
Australia	AUS	Medium	50	1,246	1,317	3.422
Norway	NOR	Small	50	996.6	1,095	8.664
Denmark	DNK	Small	44	969.5	992.9	6.627
Belgium	BEL	Small	50	952.2	929.0	1.237
Switzerland	CHE	Small	45	600.0	613.6	6.674
Greece	GRC	n/a	14	405.5	417.3	14.95
Austria	AUT	V. small	50	335.4	344.9	10.21
Portugal	PRT	V. small	30	326.5	342.3	-2.563
Finland	FIN	V. small	50	313.8	311.5	11.80
Ireland	IRL	V. small	36	236.0	105.0	13.41
New Zealand	NZL	V. small	50	149.5	148.1	2.221
Luxembourg	LUX	V. small	30	103.6	64.33	14.32
Iceland	ISL	n/a	19	12.71	8.305	7.608

Notes: ‘Group’ refers to the categorization of donors by aid volume as used in the empirical analysis (Table 7.4); ‘Obs’ gives the number of valid (annual) observations on aid for each country; all other columns report summary statistics for aid over the period, with ‘Change’ being the annual average percentage change in outflow.

Source: author’s calculations; see next table for variable definitions and sources.

Table 7C.2: Variable description and sources

Variable	Source	Comment
Net bilateral aid	OECD.Stat DAC dataset, by donor	Net disbursements of bilateral aid excluding debt relief. Converted from current USD to constant international dollars (at 2005 prices) by application of GDP deflator index and PPP conversion factor.
Real GDP	OECD Economic Outlook 86, December 2009	Converted from local currency units to international dollars (at 2005 prices) by application of GDP deflater index and PPP conversion factor.
GDP deflater	ibid.	(Used to convert to constant values).
Govt. saving in GDP	ibid.	Net government saving divided by GDP (local currency units).
Govt. spend in GDP	ibid.	General government total disbursements minus net bilateral aid (above), divided by GDP (local currency units).
Unemployment	ibid.	Unemployed as share of labour force.
Population	OECD.Stat National Accounts, Table 3	Working age population
Exchange rates	OECD.Stat National Accounts, Table 4	(Used to convert to US dollars)
PPP conversion	ibid.	(Used to convert to international dollars)
Upper income GDP	Author, using Penn World Tables, v6.3	Average GDP growth of all countries with per capita real incomes above US$ 12000
Low income GDP	ibid.	Average GDP growth of all countries with per capita real incomes below US$ 1000
Bank crisis (dummy)	Laeven and Valencia (2010)	Systemic banking crisis; variable takes a value of one for all years up to and including reported start and end dates; 2009 is assumed to be a crisis year for all financial crises starting 2008; for countries with no recorded financial crisis, one year is randomly given a value. This is to avoid the variable being dropped in the mean group estimates.
Democracies (%)	Beck et al. (2001)	Calculated by the author from the `checks_lax` variable as the share of all countries (with data available) who have legislatures that are competitively elected. Data is not available before 1975, therefore values for changes between 1960-1974 are imputed as a random normal variable with mean equal to the average change between 1975-1984 and standard deviation equal to the standard error of that estimate.

Note: all data accessed July 2010; OECD.Stat found at http://stats.oecd.org/Index.aspx; Penn World Tables found at http://pwt.econ.upenn.edu/php_site/pwt_index.php.

Chapter 8

Innovating foreign aid – progress and problems†

8.1 Introduction

The effectiveness of official foreign aid remains controversial. Although there have always been aid sceptics, the aid debate has been intensified by recent publication of a range of pessimistic studies about the capacity of official development assistance (ODA) to stimulate economic growth (see Chapter 2, also Rajan and Subramanian, 2008; Moyo, 2009). Frustration over foreign aid also stems from the perceived high costs of aid bureaucracies and the limited capacity of donors to mobilize funds for high priority long-term interventions (Birdsall, 2008). At the same time, other models of development financing have expanded rapidly. Private capital flows to developing countries have surged (at least before the financial crisis of 2008/09), numerous specialised global funds have been established, and market-based approaches to supporting social causes have gained acclaim.

The present chapter critically examines two innovative models for 'doing' aid – (a) multi-country specialized partnership funds, which include but are not limited to vertical funds; and (b) market-oriented approaches such as web-based philanthropic marketplaces. These are chosen because they have introduced substantial innovations in the structure of the donor-recipient relationship and its underlying dynamics. In contrast to interest in innovative aid modalities and risk-sharing devices (see Girishankar, 2009), these models frequently employ traditional financing instruments, but do so in new ways. Despite important differences, these models also share common features, namely an emphasis on partnerships, selectivity and tangible results. As a result, they are considered together.

Expectations surrounding the potential of these innovative models are high. Hilts (2005) claims that smart aid in the health sector, provided through specialised independent funds, avoids the major problems of ODA. The Global Fund explicitly describes itself as providing a 'new approach to international health financing' which puts the Paris Principles of enhanced aid effectiveness into practice. Adelman and Eberstadt (2008) call for a new 'business model' for (US) foreign aid based on greater use of private sector approaches (also Whittle and Kuraishi, 2008; Hoffman, 2008). GlobalGiving and Kiva, prominent

†This chapter is an edited reproduction of Jones (2012).

market-oriented models, have attracted praise from celebrated aid sceptics – William Easterly (2006, 2008) and Dambisa Moyo (2009). Media comment has been effusive (e.g., Hubbard-Preston, 2008).

The contribution of this chapter is twofold. First, a typology of development financing models is presented, which helps discern the distinctive features and scope of innovative aid models. This is supported by an overview of some of the leading examples of these models and their achievements to date. Second, a synthesis of the strengths, limitations and weaknesses of these models is presented, based on emerging evidence from the field and relevant literature. The objective is to provide a balanced assessment of the strengths and limitations of these innovative models. The thesis is that innovative models are genuinely distinctive and have been successful in raising finance for specific goals. Nevertheless, they do not address some of the most complex problems in aid financing – achieving long run sustainability and supporting lasting institutional progress. Moreover, they often replicate established problems with traditional aid models – fragmentation, a supply-side bias and an extreme focus on quick results. Some of these arguments have been made elsewhere before (e.g., Delph, 2008; Kirby-Zaki et al., 2008); however, this chapter brings together and deepens these scattered contributions.

Before proceeding some caveats are in order. The present chapter avoids discussion of *why* innovative aid models have expanded rapidly. This chapter also does not seek to provide a comprehensive critique of 'old' aid models, typified by ODA. Among others, Birdsall (2008) provides a clear summary of such critiques (also Easterly, 2008; Barder, 2009). At the same time it is important to recognise that ODA is not static. Recent reforms have encouraged alignment with country-led poverty reduction plans and wider use of programme aid modalities. These changes are not in focus here and for the present purposes orthodox foreign aid is defined as bilateral and multilateral flows between public sectors in developing and advanced countries according to traditional modalities – i.e., project loans and grants. Admittedly this is something of a caricature; but scholars note that the extent to which ODA has undergone genuine reform remains limited (see Section 8.3; also Wood et al., 2008; OECD, 2008a).

The rest of the chapter is structured as follows: Section 8.2 outlines a typology of development financing models. Section 8.3 provides an overview of two specific innovative models, followed by discussion of their strengths, weaknesses and limitations. Section 8.4 summarises the analysis and concludes with reflections on its implications for donor practice.

8.2 A typology of development financing models

Development financing is extremely multifaceted. As a starting point to organise thinking, it is helpful to distinguish different approaches to development financing according to: (i) what kind of good is being funded; (ii) the information challenges involved in designing, implementing and monitoring intervention; and (iii) the actors involved in intermediating the donor-beneficiary relationship.[1] While these different dimensions are highly interlinked, they merit individual discussion.

Table 8.1 organises developmental goods into six domains based on the economic characteristics of the goods (outcomes) being financed. They range from strictly private goods, through public goods, to merit goods. The latter are distinctive because their provision need not be justified on economic grounds. Rather, access to goods such as basic health care, primary education and potable water can be seen

[1] Note this typology is primarily employed as an heuristic tool and, therefore, is neither intended to be precise nor comprehensive. Alternative frameworks can be found in (Girishankar, 2009) and Barder (2009).

Table 8.1: Typology of developmental domains

Domain		Good type	Information costs: Design	Information costs: Implementation & Monitoring	Principal intermediaries
1	Governance institutions	Public	high	very high	Public sectors
2	Humanitarian	Merit	low	high	Public sectors, NGOs
3	Basic needs	Merit	low	medium	Public sectors, NGOs
4	Social goods & services	Public, some private	medium	medium	Public sectors, Public-private partnerships
5	Economic infrastructure	Public and private	medium	medium-low	Public-private partnerships, private sector agents, public sectors
6	Productive sector	Private	variable	low	None. Typically direct relationship between funder and beneficiary

Note: design and monitoring costs refer to information challenges only, not absolute monetary costs.
Source: author.

as inalienable human rights (e.g., Piron, 2002), the clearest example being humanitarian assistance.[2] Importantly, while a core dimension of development is the expansion of private economic production, many challenges facing developing countries involve at least some public goods aspects.

As also indicated in the table, each of these broad domains poses distinctive information challenges in terms of design, implementation and ongoing monitoring. For example, the design costs associated with many merit goods are low, as we largely know how to provide clean water and build low-cost housing (see Sachs, 2006). Nevertheless, there may be extreme monitoring problems where the intended impact of a development intervention is diffuse or is not easily observable. Again, the contrast with pure private goods is helpful. The Austrian School emphasises that the price signal can provide sufficient and reliable information to guide economic behaviour. However, as Barder (2009) cautions, there is no clear analogue to the price signal to guide developmental or public goods production (i.e., domains 1 to 5, Table 8.1).

Indeed, it has been noted extensively with respect to foreign aid that the scope for agency problems is severe due to the existence of multiple intermediaries, uncoordinated beneficiaries and large information gaps (e.g., Svensson, 2006; Ostrom et al., 2002; Martens, 2005; Azam and Laffont, 2003; Paul, 2006; Barder, 2009). To resolve these agency problems, commitment devices can be employed. While these come in various forms – e.g., *ex ante* or *ex post* conditionality, and *interim* monitoring – a broad distinction between devices is whether they are hard or soft. The difference is that the former incorporate pre-specified thresholds that define eligibility for a potential financing relationship or trigger its termination. In contrast,

[2]As Nelson Mandela put it: "... overcoming poverty is not a gesture of charity. It is an act of justice. It is the protection of a fundamental human right, the right to dignity and a decent life." (speech at The Campaign to Make Poverty History, 2005, Trafalgar Square, London, UK).

soft devices are more open and give greater scope for ongoing negotiation. Critically, there is a direct link between the information challenges of design and monitoring and the kinds of commitment devices that are suitable. Hard commitment devices tend to be effective where reliable, sufficient and timely information is available, such as prices in thick markets. These devices also reduce the need for extensive intermediation between donors and beneficiaries, allowing for arms-length transactions. On the other hand, where high quality and relevant information is not easily available, softer approaches tend to be used. These typically require more extensive interaction between donors and beneficiaries, often through intermediaries such as public aid agencies.

This leads naturally to the last element of the typology which refers to how a given development financing relationship is intermediated and, thus, the different types of actors involved. Although all forms of development financing ultimately involve a relationship between persons (citizens or corporations) in advanced and developing countries, this can be intermediated in very different ways. An intermediary is understood here as an active agent with a non-negligible degree of influence over development financing outcomes, for example via selection and/or implementation of projects. At the one extreme there may be (virtually) no intermediation where two private agents undertake a direct transaction such as in foreign direct investment or cash remittances. At the other extreme, public agencies in both source and recipient destinations can actively intermediate the relationship as in orthodox development assistance.

8.3 Innovative aid models

8.3.1 Overview

The typology elaborated in the previous section can be used to isolate the distinctive features of different development financing models. The first of these is specialized partnership funds (SPFs) which have expanded rapidly over recent years.[3] Table 8.2 provides a summary of some of the main examples of these vehicles. SPFs typically perform two main functions. They raise and pool funds for highly specific causes, (normally in the area of merit goods) and they disburse grants to local implementing agents (from both the public and private sectors). Thus, in contrast to orthodox foreign aid, SPFs establish some separation between fundraising for specific causes and the design/implementation of interventions. SPFs also tend to rely on hard commitment devices, which form an integral part of an emphasis on measurable results. In particular, alongside formal evaluation of outcomes, explicit eligibility criteria are used to regulate access to funds. For example, the Global Fund is open to all low income countries but requires lower and upper middle income countries to pass differentiated targeting and cost-sharing tests. Other vehicles, such as the Education for All Fast Track Initiative (EFA-FTI) and the Millennium Challenge Corporation (MCC), also incorporate specific selection criteria on governance and policy soundness.

A second type of innovative aid model refers to a range of market-oriented approaches. These go under various names such as social venture funds/banks, market-based grants, and peer-to-peer loans. Leading examples of such approaches are summarized in Table 8.3. Again, although diverse, two distinctive features are shared. First, there is an emphasis on using market-like mechanisms to allocate and monitor developmental funds. This ranges from a reliance on dispersed information (as in GlobalGiving and Kiva)

[3] Such funds are often referred to as vertical funds but they are not only of this form. Another label is 'Global Program Funds' which are large multi-country funds that contain a significant element of earmarked funding for specific objectives with thematic, sectoral or subsectoral breadth (Kirby-Zaki et al., 2008).

to the direct use of commercial debt and equity capital instruments by non-profit making organizations (as opposed to grants). In the sense of Hirschman (1970) these approaches enhance the credibility and likelihood of 'exit' in the face of poor performance by either party. Second, these approaches tend to shorten the aid chain by making one of either source or host country intermediaries redundant. In doing so, they aim to provide greater and more direct feedback between donor and recipients. A tag line of GlobalGiving, for example, is that individual donors can monitor the direct impact of their support, a concept which reflects the established practice whereby individuals sponsor disadvantaged children through (international) NGOs.

These features parallel those of SPFs. For example, there is a strong emphasis on selectivity, through vetting and monitoring of potential applicants, as well as achieving direct and tangible impacts based on partnerships. In the case of market-based grants and peer-to-peer loans, selectivity is exercised both by the financier (through selection of beneficiaries) and through requirements that local partners (who propose projects) meet certain eligibility criteria and provide adequate ongoing reports. For example, among other things, GlobalGiving requires local partners to be: (a) registered non-profit organizations in good standing, attested by documentary evidence including independent letters of reference; (b) non-discriminatory in its selection of staff, board, and beneficiaries; and (c) able to communicate in English.[4] Similar requirements are made by Kiva of its local microfinance partners. In contrast, the Acumen Fund and the Charity Bank undertake their own due diligence of individual projects in the same way as commercial financial institutions. A more specific innovation of these social funds/banks is that their use of commercial debt and equity instruments to finance ventures that may not otherwise be able to access funding. Thus, there is a greater reliance on competition and selectivity in these models, as well as a separation of responsibility for fundraising and project design/implementation.

Before considering the strengths, weaknesses and limitations of these innovative models, it is helpful to get a sense of their size relative to more traditional instruments. On the one hand, it is well established that although some scaling-up of ODA has taken place since 2000, this expansion has been modest in historical perspective (OECD, 2008a). In constant 2000 prices, total net official assistance to developing countries in 2007 (USD 80.4 billion) was below that of 1995 (USD 92.5 billion).[5] In contrast, private flows to developing countries have seen extremely rapid growth (see UNCTAD, 2008). Comprehensive or consistent data on disbursements of innovative financial approaches, however, is scarce.[6] Nevertheless, three conclusions emerge from existing figures. As is evident from Table 8.3, market-based approaches remain minor players in absolute levels terms, despite rapid growth rates. More generally, however, private grants to fund development interventions have seen very substantial growth over recent years. These are important sources of 'new' funding for both SPFs and market-like mechanisms (see Tables 8.2 and 8.3). Figures in OECD (2008a) indicate that, in 2007, private grants were equal in value to around 20% of net official flows to developing countries (after repayments to the IMF). Thus, given that SPFs combine both substantial flows of official assistance and private grants (see further below), there is no doubt that many of these are now very significant players in the development financing landscape.

---

[4]In addition, for organizations that do not meet these criteria or for projects that are not sponsored by eligible local partners, GlobalGiving organises "Open Access Challenges". It is not clear, however, what proportion of total grants are channelled through the latter.

[5]Based on figures in OECD (2008a).

[6]See Girishankar (2009) for a recent attempt to take stock of the contribution of a wide range of innovative approaches.

Table 8.2: Examples of specialized partnership funds

Vehicle	Active period	Total spend, 10^6 US$	Objectives	Main funding sources	Organization / approach
Global Fund	2000 onwards	17,100 (projected)	Combat of HIV/AIDS, tuberculosis & malaria esp. in poor countries	Official grants (96%) + other private sources (e.g., Bill & Melinda Gates Foundation); existing financing pledges extend until 2010.	Independent public-private partnership; administratively autonomous since Jan. 2009; implementation via grants to local partners; World Bank acting trustee.
GAVI	2006-2015	4,000 (projected)	Reduce child mortality by increasing access to vaccination and immunization	IFFIm (legally-binding long-term official grant commitments to 2026 used to back international bond issues); Advanced Market Commitments; direct official funding.	Independent alliance based on a public-private partnership model; hosted by UNICEF; provides grant support to national governments in response to country proposals.
Education For All - Fast Track Initiative	2002-2015	1391 (projected)	Accelerate progress towards universal primary school completion by 2015	Official grants to core trust funds (Catalytic Fund, Education Program Development Fund).	Partnership of bilateral donors, international agencies and development banks; World Bank managed and administered; quick-disbursing grants made to (mainly) national governments.
Global Partnership on Output Based Aid (GPOBA)	2003 onwards	154.7 (spent to January 2009)	Provide performance-based subsidies to basic services delivery where costs cannot be fully met via user fees alone	Official donor pledges.	Partnership of donors and international organizations (IFC, SIDA, AusAid, DFID). Beneficiaries are local private service companies.
Millennium Challenge Corporation (MCC)	2004 onwards	6,700 (approved as at Sept. 2008)	Reduce poverty and stimulate economic growth	US Government (Congress).	Independent government corporation; highly selective (based on governance criteria); 5-year grant commitments made to selected eligible countries.

Source: author's collation from respective institution websites, retrieved between December 2008 and August 2009.

Table 8.3: Examples of market-oriented aid approaches

Example	Founded	Summary	Progress
GlobalGiving	1997	A philanthropic marketplace where individual donors select to support specific projects implemented and/or managed by grassroots organizations. The latter must meet stipulated eligibility criteria and provide regular updates. 10% of all donations are taken to meet running expenses.	Since 2002 over US$20.7 million in donations have been channeled to over 1,300 projects worldwide, principally in education, health and economic development. Around 50% of all funding is to projects in USA (37%) and India (12%) [in 2008]
Acumen Fund	2001	Global venture fund based on 'entrepreneurial approaches'. Investments are in entrepreneurial social businesses that deliver affordable, essential goods and services through innovative, market-oriented approaches.	As at end December 2008, US$35 million of funds committed across 5 portfolios, namely: health (60%), housing (14%), water (11%), agriculture (8%), energy (7%)
Charity Bank	2002	Registered bank that takes individual/corporate deposits and uses these to make loans to charitable causes e.g., registered charities, community associations, social enterprises etc.. Mainly focussed on UK but with small overseas portfolio.	Balance sheet of approx. $100 million (end July 2009); since launch, loans of $140 million have been made, unlocking $250 million of additional funding for borrowers. Average loan size in 2009 = $256,000
Kiva	2005	Person-to-person micro-lending marketplace. Lenders choose beneficiaries which have been pre-selected by local microfinance partners who then manage the transaction. The latter must meet stipulated eligibility criteria. Running costs met by seed capital, grants and optional lender fees.	As at August 5, 2009 the cumulative value of loans (since 2005) was approx. US$85 million across 49 countries, reaching over 200,000 beneficiaries. Average loan size is about US$400.

Sources: material collated in August 2009 from www.charitynavigator.org, www.globalgiving.com, www.acumenfund.org, www.charitybank.org and www.kiva.org

8.3.2 Strengths

The above discussion indicates that innovative models incorporate substantial departures from orthodox aid. A number of strengths are common to SPFs and market-based mechansisms – they remove public sector involvement in the source country (e.g., via official aid agencies) in the design/implementation of specific interventions, they often shorten the aid chain, and they tend to rely on harder commitment devices. For example, certain market-based approaches have allowed charitable giving to operate along market-like or 'crowd-sourced' lines (Brabham, 2008), exploiting advances in global interconnectivity, thereby directly connecting ultimate donors and final beneficiaries. Similarly the political independence of SPFs is assured legally, as private-public parternships, as well as by their focus on specific objectives. Additionally, the wider use of hard conditionality in both models points toward an improved capacity to make time consistent commitments and choose well-performing agents to implement projects over time.

These strengths have not gone unnoticed. Backed by astute campaigning, SPFs have been able to rapidly mobilize and disburse relatively large volumes of funds. To cite some examples, the International Finance Facility for Immunisation has secured legally binding funding commitments from numerous advanced countries until 2026 (see Table 8.2). As a consequence, the GAVI Alliance reports that over the period 2000-08 it has protected a total of 213 million children with new and underused vaccines and prevented 3.4 million future deaths. Similarly, the Global Fund boasts funding commitments of over USD 17 billion and already has become a major player in global health, providing an estimated 20% of donor funding to HIV/AIDS, 64% for malaria, and 70% for tuberculosis (McCarthy, 2007, p. 307). The U.S. President's Emergency Plan for AIDS Relief (PEPFAR), also a major contributor to the Global Fund, reports that anti-retroviral has increased dramatically in its focus countries and attributes gains of over three million cumulative life years to its financial support (PEPFAR, 2008, p.12). As such, SPFs indicate definite benefits of unbundling fundraising from aid project design/implementation, as recommended by Barder (2009) (among others).

As Table 8.3 makes clear, market-based models are significantly smaller, reflecting the more private (direct) nature of the transactions. Nonetheless, and despite its youth, Kiva has placed over $85 million of microfinance loans across 49 countries. Market-based financing to social businesses also fill a key gap between micro-finance and private capital markets. While the former typically has focussed on the poorest self-employed (women), the latter generally target large transactions. Thus, scholars have highlighted a lack of financing for small and medium-sized businesses in low income countries (e.g., Beck and Demirguc–Kunt, 2006). This is even more acute for higher-risk ventures such as those with a social development purpose. At the same time, where commercial lending or investment instruments are used, there is scope for individuals in advanced countries to earn a (modest) return on social financing. This may expand the volume of financing available and is consistent with the Product Red™ concept of making charity beneficial for all parties (Ponte et al., 2009). Thus, a key strength of market-base mechanisms is they fill a niche in the development financing landscape.

Considered together the achievements of these models are noteworthy. In contrast to the relatively fragmented and protracted mechanisms of orthodox aid, there is a broad consensus that the innovative aspects of these vehicles have been crucial to their success (Radelet, 2004; Lu et al., 2006; McCarthy, 2007; The Global Fund, 2007).

8.3.3 Limitations and weaknesses

Innovative aid models are not a panacea for the problems of orthodox aid. A straightforward limitation of both the innovative models considered above is that they are suited to only certain developmental domains (as described in Table 8.1). Innovative models are conspicuously absent from both humanitarian and larger-scale infrastructure financing. These are cases where orthodox approaches, based on public sector fundraising, design and implementation are likely to remain predominant. They are also areas where additional funds are needed. For example, the World Bank's recent Africa Country Infrastructure Diagnostic identifies that an annual USD75 billion in additional financing is required to address the region's critical infrastructure gap (Foster, 2008).[7]

In addition, it is doubtful that these innovative models are suited to financing complex, long term challenges such as public institution- and capacity-building. While a focus on tangible results can be an advantage, it also limits the scope of interventions that are likely to be considered. Indeed, it is no surprise that the vast bulk of funding associated with these innovative aid models are associated with health, education and productive activities (see Tables 8.2 and 8.3). Supporting institutional reforms and/or activities where the technology of intervention is not known with certainty are unlikely to be addressed effectively by innovative models that rely on easily observed outcomes and hard commitment devices.[8] This is a long-standing critique of ODA, but remains germane to innovative approaches (Dodd et al., 2007; Reich et al., 2008). Admittedly, some of the larger SPFs are cognizant of this challenge. A recent independent evaluation of the Global Fund, for example, identifies an unresolved conflict between retaining a strict focus on channelling funds to the most efficient/effective local agents and becoming more closely involved in implementation processes through local capacity-building (The Global Fund, 2007). However, it remains unclear how these conflicts can be resolved.

A related limitation is the dependence of both SPFs and market-based models on 'eligible' local partners. In the case of market-based approaches these local partners are relied on to source the majority of potential beneficiaries (projects). Hence, it is unsurprising that market-based approaches retain a very distinct and uneven geographic incidence. For example, over 14,000 loans have been extended to entrepreneurs in Peru via Kiva, compared to less than 2,000 to Bolivia and only 1 to Costa Rica.[9] Similarly, in 2008 around 50% of all funding channelled through GlobalGiving went to projects in the USA (37%) and India (12%) (GlobalGiving, 2008). SPFs, on the other hand, are dependent on local implementation capacity and the ability of local partners to deal with very large volumes of funds. Various cases suggest this generates tensions, for example between funds' adherence to selectivity- and performance-criteria, versus building flexible and sustainable long-term partnerships (The Global Fund, 2007; Birdsall, 2008; Herrling et al., 2009). In Mozambique, the Health Ministry has had difficulties keeping up with the range of specific demands placed on it by the Global Fund, in turn leading to disbursement delays and failure to meet targets. According to the Health Minister, no funds from the Global Fund were disbursed to the country in the first half of 2009.[10]

These are more than just limitations of scope. They touch on deeper concerns regarding how development

[7] This does not ignore the potential for wider use of public private partnerships and private sector investment facilitation funds, such as the Emerging Africa Infrastructure Fund; however such funds operate on commercial principles and, thus, are distinct from SPFs.

[8] Delph (2008) notes that recommendations to establish a global fund for agriculture have been largely dismissed by major donors for precisely these reasons; also see Foster et al. (2001).

[9] Information correct as at 5 August 2009, taken from www.kivalytics.org/loans/location.php

[10] News item reported at: http://allafrica.com/stories/200907160884.html

interventions are prioritized. Holistic approaches to development recognize its complex and highly interdependent character. As such, the long-term effectiveness and sustainability of any single aid intervention, even the provision of core merit goods, ultimately depends on the strengthening of broader systems and local capacities. This is a particular challenge for SPFs as they often focus on the provision of (quasi-)public goods.[11] To take the example of primary health, various commentators suggest that access to funding is no longer a binding constraint in many developing countries. Rather, the critical challenge is to build national health systems with adequate personnel, high quality information and competent planning processes (Sridhar and Batniji, 2008; Bertozzi et al., 2008; AbouZahr et al., 2010). In their absence, health funding can be distorted by donor priorities rather than country needs and, thus, may be only weakly effective on aggregate. Garrett (2007) cites examples from across Africa where health system capacity indicators (e.g., doctors per capita) have stagnated or declined despite huge volumes of funding via SPFs for specific diseases such as HIV/AIDS. Sridhar and Batniji (2008) show extremely large discrepancies in the allocation of funds to different diseases, as well as the excessive focus of major global health funds on treatment as opposed to prevention. Case studies of Uganda and Ethiopia, documented in Bernstein and Sessions (2007), suggest that where local government capacity bottlenecks exist, pressure to achieve disbursement targets has led SPFs to push for 'external' solutions, such as outsourcing of key functions, as opposed to internal strengthening.

A more general concern refers to governance and accountability dynamics. SPFs and market-based models emphasize a partnership approach. Nonetheless, the agenda is squarely set by advanced country donors in the sense that the latter actors determine what might be considered 'appropriate' interventions. Thus, and as in most other development financing models, supply-side considerations take the lead. To take a concrete example, doubts have been raised that political interests can skew the content or application of the MCC's country eligibility criteria.[12] There is also the concern that rigid selection criteria may exclude the poorest beneficiaries, for whom marginal returns to aid may be highest. This is a variant of the long-standing argument that application of generic aid allocation criteria may not always enhance aid effectiveness (see Hansen and Tarp, 2001; Guillaumont, 2008).

Similarly, market-based approaches are driven by the developmental concerns of individual donors in advanced countries. This gives little space for independent analysis or for ultimate beneficiaries to voice needs or sanction poor implementation. In addition, these approaches can subject beneficiaries to vicissitudes in the tastes and interests of donors. This is not only relevant because fashions change, but it is also material. Preliminary evidence indicates large fluctuations in the availability of funding over time, at least at the country level. Although aggregate level statistics are generally hard to come by, and these market-based initiatives are at an early stage of development, figures from GlobalGiving (2008) indicate numerous countries faced a 50% or larger drop in funding in 2008 compared to 2007 (e.g., Indonesia, Burundi, Cameroon, Niger, Sierra Leone). Of course this may be for good reasons, but it is hard to imagine there were no valid projects or needs in these "losing" countries.

This issue connects to an extensive literature regarding the depth and validity of processes used by external actors to select and prioritize developmental interventions (e.g., White, 1996; Leal, 2007).[13] It is curious

[11] Even so, market-based approaches are not immune from this critique. For example, to the extent that such approaches become significant in any one country/location they might unbalance the provision of financial services, possibly weakening the growth of domestic finance providers.

[12] Stubbs (2009) notes that the inclusion of Georgia in the list of eligible countries, and its subsequent conclusion of an MCA Compact, is indicative of "covert" political bias. Georgia did not pass all indicators (e.g., 2/6 Ruling Justly indicators and 3/6 Economic Freedom indicators) and various countries with better scores were not selected.

[13] White's discussion of 'participation', for which 'partnership' could easily be substituted, expresses the point: "The status of

that the nature of the partnership claims made by many innovative models have received little scrutiny. This is in contrast to the widespread criticism of how the elaboration of Poverty Reduction Strategies often have not given space for genuine country ownership (e.g., Wang and Stewart, 2003; Whitfield, 2005). One exception is Mawdsley (2007), who notes that despite the MCC's stated commitment to country-ownership, the vast majority of all signed country compacts have 'independently' selected the *same* priority action areas (agribusiness, rural entrepreneurial development, and transport infrastructure).

There is also the question of sustainability. Although some innovative aid models have been able to mobilize short or medium run funding, this is both time limited and in no way guarantees financial sustainability for recipients, particularly if external funding dwindles. In the case of the Global Fund, the largest of all initiatives on an annualised spending basis, 96% of projected funding derives from official grants which in turn mainly come from the larger DAC donors. Similarly, the MCC is a US government corporation with exclusive funding through Congressional appropriations. This dependence on official public funding is the rule rather than the exception (see Table 8.2), also leading some to question the real extent of additionality (e.g., Godal, 2005). As Poore (2004) notes, while the Global Alliance on Vaccines and Immunization (GAVI) has been concerned with ensuring financial sustainability, the same organization recognises that the transition of financial responsibility from the Fund to national governments and local partners is in no way assured. This is reminiscent of established concerns with orthodox aid where operating and maintenance costs are often neglected, to the detriment of long run effectiveness. Similarly, for market-based approaches there is scant evidence of any real emphasis on ensuring a smooth transition for recipients away from dependence on external aid flows.[14]

Due to their larger size, a particular concern associated with SPFs is that they often operate as separate aid channels with their own systems and processes. This creates duplication and places distortionary burdens on local administrative resources via use of government systems and employment of scare personnel (Godal, 2005; Dodd et al., 2007; Garrett, 2007; Oomman et al., 2007, 2008). Referring specifically to global health partnerships, Caines (2005) argues they have exacerbated overlaps between donors and have aligned poorly with country systems. Empirically these concerns are little different to critiques of orthodox aid models which have burdened countries with hundreds of donors, projects and their ensuing paraphernalia (Knack and Rahman, 2008; Brautigam and Knack, 2004). In addition, and as suggested by Martens (2005), the shift to hard commitment devices, with its emphasis on selectivity and *ex post* conditionality, appears to have increased *ex ante* transaction costs during project negotiation and initiation. With respect to the MCC, for example, Herrling et al. (2009) note that start-up processes and procedures have been much slower than originally envisaged. The US Government Accountability Office (Gootnick, 2007) estimates it has taken an average of 633 days (about 21 months) for African countries to progress from being selected as eligible to signing a compact; and only 23% of disbursements took place as planned for compacts to African countries operational through to end March 2007. These problems are explained by the MCC's highly optimistic initial assessments of partner country capacity, including the availability of qualified personnel (Gootnick, 2007).

Finally, evidence for the comparative efficiency and effectiveness of either SPFs or market-based models is lacking. This should be distinguished from their ability to raise or disburse funds, which cannot be disputed. The question is whether 'orthodox' aid delivery mechanisms of the same funding volumes may have achieved similar or even better impacts. Certainly the high upfront (*ex ante)* costs and coordination

participation as a 'Hurrah' word, bringing a warm glow to its users and hearers, blocks its detailed examination. Its seeming transparency ... masks the fact that participation can take on multiple forms and serve many different interests." (1996, p. 6)

[14]This is most problematic for grant-based models. Where loans or capital investments are in play, there are built-in incentives to achieve independence due to the repayment costs for beneficiaries.

challenges involved in dealing with innovative mechanisms suggest this possibility is material. There is also a concern that fragmentation and selectivity in delivery may undermine scale economies in service provision and exclude the most needy (where marginal benefits may be highest). Rigorous evaluations of the long run and comparative effectiveness of these innovative mechanisms are scarce. For example, with respect to results-based financing schemes such as those supported by the GPOBA (see Table 8.2), Oxman and Fretheim (2008) conclude that most existing evaluations are not rigorous and there is almost no evidence of their relative cost-effectiveness. They also note that unanticipated adverse effects have occurred in some instances, such as a reduction in provision of services to the severely ill, which may be associated with the specific ways in which performance is monitored.

8.4 Concluding remarks

This chapter has surveyed two innovative aid models- specialised partnership funds and market-based approaches. A new typology of development financing models was sketched and applied to highlight the distinctive features of these innovative models. These were found to be the different ways in which aid relationships are intermediated, the shift towards harder commitment devices and a results-based focus. As such, these models have genuine strengths that merit recognition. Specialised partnership funds have been effective in mobilising and disbursing large volumes of funds, especially where the technology of intervention is known. Market-based approaches have created innovative ways to connect individual donors and recipients, both widening and decentralising the reach of aid, increasing the potential scale of external funding for social ventures in low income countries.

These innovative models also have made progress in unbundling the funding of development interventions from their design and implementation. This has a number of potential advantages. It can reduce political interference, enhance transparency, promote specialisation and allow greater competition among implementing actors, thereby improving performance. Nevertheless, lessons must be learnt from downsides associated with the way in which unbundling has been instigated by these models to date. The evidence suggests that development priorities are often set in a top-down manner, thereby giving a weak voice to beneficiaries, and insufficient support has been given to capacity building of key domestic actors (e.g., the recipient country public sector). The creation of new structures and processes has added to the proliferation of donors and has increased transaction costs associated with receiving aid funds. Also, information regarding the comparative effectiveness of these innovative vehicles remains scarce.

There are four main implications of this analysis for development practice. First, greater attention should be given to understanding which financing models are best suited to which kinds of development challenges. Delineating the appropriate conditions for employing innovative approaches, such as market-like mechanisms, would be helpful. In this respect, innovative models are best understood as only one part of the financing toolkit rather than all-purpose tools. Second, a better understanding of the appropriate domain of innovative approaches is not likely to come from theory. Investment in and learning from rigorous evaluations of innovative approaches must be given priority. Third, as part of this learning process, the comparative benefits of establishing new structures and processes for delivering aid, in contrast to using existing channels and relations (but in different ways), must be analysed. Indeed, it is already evident that both multilateral and bilateral donors are beginning to employ certain aspects of innovative approaches within 'orthodox' structures. Examples include the inclusion of challenge funds, hard selection criteria and output-based approaches in standard (sector) development programmes.

Thus, it may be the case that greater gains can be made from innovating within existing aid structures as opposed to creating new ones. This adds to Girishankar's (2009) conclusion that a careful distinction should be made between the benefits of using innovative models/structures to support global priorities versus country programmes. Finally, as aid models evolve, due attention should be given to a wide range of impacts rather than just a narrow set of short-term results. Among others, capacity building and long run sustainability must be emphasised. Indeed, there are ample lessons from developed countries that innovative approaches to public service delivery have yielded both positive and perverse consequences. These lessons can be usefully applied in the search for development financing models that enhance aid effectiveness, which is the ultimate goal.

Chapter 9

Conclusion

This thesis has brought together a collection of essays which address the relation between foreign aid and development. At first sight, the essays might seem somewhat disparate in nature. However, a unifying question – addressed from different angles and employing different empirical methods – is whether aid is effective in enhancing social welfare in developing countries. The introduction (Chapter 1) made the case for the heterogeneity of aid and, in turn, the particular evaluation challenges she presents. Chapter 2 reflected on a selection of previous theoretical and empirical literature on aid effectiveness (broadly conceived as spanning macro- through to micro-level analyses), which helped identify some of the specific topics that motivated the remaining chapters. Among these, the resilience of an apparent inconsistency between robust returns to aid at the micro-level and ambiguous returns at the macro-level (the 'micro-macro paradox') was highlighted.

Chapters 3 and 4 addressed this paradox head on. Taking the contribution of Rajan and Subramanian (2008) (RS08) as its starting point, the former chapter focused on quantifying the impact of aid on economic growth in a cross-section of developing countries over a single (long) period of 30 years. To do so, three modifications to the RS08 methodology were employed: (i) strengthening the generated aid instrument; (ii) improving the regression specification; and (iii) developing a (new) doubly robust estimator for the instrumental variables context. The results indicated a highly significant positive average impact of aid on growth that is consistent with the simple prior suggested in Chapter 2 (median point estimate = 0.12). This finding was further validated in Chapter 4, which not only extended the same dataset from 1970-2000 to 1970-2007 but also decomposed the aid-growth effect into constituent channels. It found that investment and health are the principal proximate means by which aid stimulates growth (on average). It also found that positive effects of aid are echoed across a range of other developmental outcomes, such as poverty and structural change.

Chapters 5 and 6 took a different route to addressing the theme of aid effectiveness. They focused on two organic farming out-grower schemes in Uganda that had received seed financing from the Swedish International Development Agency. The principal research questions were whether these schemes had improved household welfare and, if so, how. Evidently, the connection to aid effectiveness is only indirect as the survey design did not admit consideration of the specific contribution (or return to) the initial aid finance vis-à-vis other possible sources of capital. Nonetheless, the general issue is whether projects of this nature can be effective in raising rural household incomes. This is a fundamental challenge in tropical Africa, and is a domain where aid interventions have been limited (at least since the 1990s) and often problematic. Although based on two very specific cases, the chapters revealed strong welfare effects from the schemes. These were found to be driven by establishing credible incentives for smallholders to produce to a higher quality standard, as well as the price premia associated with gaining access to the organic market.

Chapters 7 and 8 shifted attention away from direct questions about impacts to questions about the overall aid architecture. This reflects the point that aid's developmental impacts are fundamentally determined by the behaviour of donors (e.g., the volume of aid that is supplied) as well as how aid interventions are designed and delivered. Partly motivated by the recent (and continuing) global financial crisis, Chapter 7 reviewed the determinants of the aggregate supply of aid by the major OECD-DAC donors. A crucial methodological innovation was the application of country-specific cointegration methods to distinguish between factors affecting long-run (revealed) aid targets and short-run variation around these targets. This framework was found to be useful, indicating very substantial differences in aid supply behaviours between countries and over time. Thus, the past is not necessarily a good guide to the future. This point is underlined in Chapter 8 which reflected on a number of innovative models for 'doing' aid – namely, specialized partnership funds and market-based approaches. These were found to be distinctive to orthodox models of aid, and have certain advantages such as insulation from political interference and greater selectivity. However, they were also seen as replicating established weaknesses in aid, such as retaining a supply-side bias and being poorly suited to long-term complex challenges. Thus, they are no panacea.

Bringing these chapters together, what have we learnt? A key lesson is that the aid landscape is fundamentally complex, heterogeneous and evolving. On the one hand, this means that any credible attempt to evaluate aid's effectiveness must confront these challenges explicitly. Indeed, recognition of the sheer complexity of aid warns against accepting the kind of glib statements about aid that are too often found in popular discussion, such as: "... aid is not benign – it's malignant" (Moyo, 2009). On the other hand, it also means that we must have the humility to appreciate that a comprehensive or definitive view of aid effectiveness is neither likely to be meaningful nor possible. Different research methods are essential, as is a constant questioning of existing research findings using new data and new techniques (c.f., Chapter 3).

These caveats do not mean that we are destined to remain in a 'no man's land' of aid doubt. Rather, and in keeping with a number of recent contributions to the literature (e.g., Clemens et al., 2011), at least three consistent messages can be drawn from the analysis. The first, as noted in the chapter summaries given above, is the finding of a coherent set of positive effects running from aid to a range of development outcomes measured at the micro, meso and macro levels. Simply put, there is no micro-macro paradox. In turn, this means that calls for foreign aid to be totally abolished have no basis in solid evidence. Second, it is also the case that aid is far from optimal and it is not a developmental panacea. Amongst other things, the effectiveness of foreign aid is marred by vicissitudes in its overall supply as well as failures in institution building and genuine country ownership, which are replicated by innovative aid models. Thus, there is ample scope for aid to be better allocated, designed and implemented. Third, and relatedly, more reasonable expectations of what aid can achieved are in order. Chapters 3 and 4 emphasized that returns to aid must be understood in the context of long-run development process, such as changes to a country's stock of human capital. Indeed, it is only when we focus on these long-run processes do we find that average economic returns to aid become significant, positive and respectable (e.g., Chapter 4 suggests a 19% internal rate of return on aid over a 37 year period). The corollary is that we should not expect aid to deliver quick, transformative changes at the economy-wide level.

To conclude, what are the gaps and remaining challenges in aid effectiveness research (broadly conceived)? Above all, it is critical to recognize the importance of accumulating evidence over time, based on rigorous and transparent evaluation methods. Average cross-country estimated effects are likely to mask numerous instances of negligible and even harmful effects. These need to be better understood. High quality evidence at the meso-level is lacking and can be usefully enlarged. Also, robust project-level evaluations are surprisingly scarce and, even where they are available, they cannot be generalized in any straightforward manner (see the discussion in Chapter 6). Moreover, as both the aid architecture and the nature of development challenges evolve, the effectiveness of aid is also likely to alter (for discussion in the case of Mozambique see Arndt et al., 2007).

Additionally, greater attention needs to be given to the heterogeneity of aid and, hence, differences in aid quality. Chapters 3 and 4 employed relatively crude aggregate measures of aid, in part due to weaknesses in the raw data. This is slowly changing as new databases are produced and donors become more transparent. Also, with the passage of time, it is becoming increasingly feasible to undertake robust time series analysis of aid effectiveness at the country-level (e.g., Juselius et al., 2011). This is likely to be an important area of fruitful engagement. Finally, we have a limited empirical understanding of the effectiveness of aid at the micro-level in comparison to other financing instruments. Research in this area may provide important insights as to when and where negative side-effects from aid emerge, such as effort disincentives due to lack of time consistency. Indeed, it is here that there is much speculation, but little reliable evidence. Consequently, the aid effectiveness research agenda remains open.

Bibliography

AbouZahr, C. et al. (2010). Better data needed: everyone agrees, but no one wants to pay. *The Lancet*, 375(9715):619.

Acemoglu, D. and Johnson, S. (2007). Disease and Development: The Effect of Life Expectancy on Economic Growth. *Journal of Political Economy*, 115(6):925–985.

Adam, C.S. and Bevan, D.L. (2006). Aid and the Supply Side: Public Investment, Export Performance, and Dutch Disease in low-Income Countries. *World Bank Economic Review*, 20(2):261–290.

Adelman, I. (2001). Fallacies in development theory and their implications for policy. In G.M. Meier and J.E. Stiglitz (Eds.), *Frontiers of development economics: the future in perspective*, pp. 103–149. New York: Oxford University Press.

Adelman, C.C. and Eberstadt, N. (2008). Foreign Aid: What Works and What Doesn't. Development Policy Outlook 3, American Enterprise Institute for Public Policy Research.

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

Angrist, J. and Krueger, A. (1999). Empirical strategies in labor economics. *Handbook of Labor Economics*, 3:1277–1366.

Angrist, J.D. and Pischke, J.S. (2008). *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press. ISBN 0691120358.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 58(2):277–97.

Armytage, L. (2011). Evaluating aid: An adolescent domain of practice. *Evaluation*, 17(3):261.

Arndt, C., Jones, S. and Tarp, F. (2007). Aid and Development: The Mozambican Case. In S. Lahiri (Ed.), *Theory and practice of foreign aid*, volume 1, pp. 235–288. Amsterdam: Amsterdam: Elsevier B.V.

—— (2010a). Aid, Growth, and Development: Have We Come Full Circle? *Journal of Globalization and Development*, 1(2):Article 5.

—— (2010b). Aid, Growth, and Development: Have We Come Full Circle? WIDER Working Paper 2010–96, World Institute for Development Economic Research (UNU–WIDER).

—— (2011). Aid Effectiveness: Opening the Black Box. WIDER Working Paper 2011–44, World Institute for Development Economic Research (UNU–WIDER).

—— (2012). Aid Effectiveness. In M. Ndulu (Ed.), *Problems, Promises, and Paradoxes of Aid: Africa's Experience*. Ohio University Press (forthcoming).

Ashraf, Q.H., Lester, A. and Weil, D.N. (2009). When Does Improving Health Raise GDP? In D. Acemoglu, K. Rogoff and M. Woodford (Eds.), *NBER Macroeconomics Annual 2008*, volume 23, pp. 157–204. University of Chicago Press. URL http://www.nber.org/chapters/c7278.

Azam, J.P., Devarajan, S. and O'Connell, S.A. (1999). Aid dependence reconsidered. CSAE Working Paper Series 1999–05, Centre for the Study of African Economies, University of Oxford.

Azam, J.P. and Laffont, J.J. (2003). Contracting for aid. *Journal of Development Economics*, 70(1):25–58.

Bacon, C. (2005). Confronting the coffee crisis: can fair trade, organic, and specialty coffees reduce small–scale farmer vulnerability in northern Nicaragua? *World Development*, 33(3):497–511.

Baldacci, E., Hillman, A.L. and Kojo, N.C. (2004). Growth, governance, and fiscal policy transmission channels in low-income countries. *European Journal of Political Economy*, 20(3):517–549.

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176.

Banerjee, A. (1999). Panel Data Unit Roots and Cointegration: An Overview. *Oxford Bulletin of Economics and Statistics*, 61(S1):607–29.

—— (2007). *Making aid work*. The MIT Press. ISBN 0262026155.

Banerjee, A.V. and Duflo, E. (2009). The Experimental Approach to Development Economics. *Annual Review of Economics*, 1(1):151–78. doi:10.1146/annurev.economics.050708.143235.

Bank of Uganda (2006). State of the economy: June 2006. Technical report, Bank of Uganda, Kampala.

Barder, O. (2009). Beyond Planning: Markets and Networks for Better Aid. Working Papers 185, Center for Global Development. URL http://ideas.repec.org/p/cgd/wpaper/185.html.

Barrett, C. and Carter, M. (2010). The power and pitfalls of experiments in development economics: some non–random reflections. *Applied Economic Perspectives and Policy*, 32(4):515.

Baum, C.F., Schaffer, M.E. and Stillman, S. (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal*, 7(4):465–506.

Baunsgaard, T. and Keen, M. (2010). Tax Revenue and (or?) Trade Liberalization. *Journal of Public Economics*, 94:563–577.

Becchetti, L. and Costantino, M. (2006). The effects of fair trade on marginalised producers: an impact analysis on Kenyan farmers. Working Paper 2006–41, Ecineq: Society for the Study of Economic Inequality.

Beck, T., Clarke, G., Groff, A., Keefer, P. and Walsh, P. (2001). New tools in comparative political economy: The database of political institutions. *World Bank Economic Review*, 15(1):165–176. Database update December 2009.

Beck, T. and Demirguc–Kunt, A. (2006). Small and medium-size enterprises: Access to finance as a growth constraint. *Journal of Banking & Finance*, 30(11):2931–2943.

Benfica, R., Tschirley, D. and Boughton, D. (2006). Interlinked transactions in cash cropping economies: the determinants of farmer participation and performance in the Zambezi river valley of Mozambique. In *Prcoceedings of the International Association of Agricultural Economics, (12-14 August, Gold Coast, Australia)*. URL http://ageconsearch.umn.edu/bitstream/25244/1/cp060855.pdf.

Bernstein, M. and Sessions, M. (2007). A trickle or a flood: Commitments and disbursements for HIV/AIDS from the Global Fund, PEPFAR and the World Bank's Multi-Country AIDS Program (MAP). Technical report, Centre for Global Development.

Bertoli, S., Cornia, G.A. and Manaresi, F. (2008). Aid Effort and Its Determinants: A Comparison of the Italian Performance with other OECD Donors. Working Papers Series WP2008–11, Universita' degli Studi di Firenze Dipartimento di Scienze Economiche.

Bertozzi, S.M., Laga, M., Bautista-Arredondo, S. and Coutinho, A. (2008). Making HIV prevention programmes work. *The Lancet*, 372(9641):831–844.

Besley, T. and Case, A. (1993). Modeling technology adoption in developing countries. *The American Economic Review*, 83(2):396–402.

Birdsall, N. (2008). Seven deadly sins: reflections on donor failings. In Easterly (2008), chapter 20.

Blundell, R. and Dias, M.C. (2000). Evaluation methods for non-experimental data. *Fiscal studies*, 21(4):427–468.

—— (2002). Alternative approaches to evaluation in empirical microeconomics. *Portuguese Economic Journal*, 1(2):91–115.

Bolwig, S., Gibbon, P. and Jones, S. (2009). The economics of smallholder organic contract farming in tropical Africa. *World Development*, 37(6):1094–1104.

Boone, P. (1996). Politics and the effectiveness of foreign aid. *European Economic Review*, 40(2):289–329.

Boschini, A. and Olofsgand, A. (2007). Foreign aid: An instrument for fighting communism? *The Journal of Development Studies*, 43(4):622–648.

Boughton, J.M. (2006). Different strokes? Common and uncommon responses to financial crises. In A. Mody and A. Rebucci (Eds.), *IMF-supported programs: Recent staff research*, chapter 8, pp. 131–139. Washington, D.C.: International Monetary Fund.

Bourguignon, F. (2003). The growth elasticity of poverty reduction: explaining heterogeneity across countries and time periods. In T.S. Eicher and S.J. Turnovsky (Eds.), *Inequality and Growth: Theory and Policy Implications*, pp. 3–27. MIT Press.

Bourguignon, F. and Sundberg, M. (2007). Aid Effectiveness – Opening the Black Box. *American Economic Review – Papers and Proceedings*, 97(2):316–321.

Bowers, J., Bailey, B., Hebbar, P., Sanogo, S. and Lumsden, R. (2001). The Impact of Plant Diseases on World Chocolate Production. *Plant Health Progress*, online. doi:10.1094/PHP--2001--0709--01--RV.

Brabham, D.C. (2008). Crowdsourcing as a model for problem solving. *Convergence: The International Journal of Research into New Media Technologies*, 14(1):75.

Brautigam, D.A. and Knack, S. (2004). Foreign aid, institutions, and governance in sub-Saharan Africa. *Economic Development and Cultural Change*, 52(2):255–285.

Bray, D., Sanchez, J. and Murphy, E. (2002). Social dimensions of organic coffee production in Mexico: lessons for eco-labeling initiatives. *Society & Natural Resources*, 15(5):429–446.

Brito, C. and Pearl, J. (2006). Graphical Models for Identification in Structural Equation Models. In R. Dechter and T.S. Richardson (Eds.), *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, pp. 47–54. Corvallis, OR: AUAI Press. URL escholarship.org/uc/item/0jh2v829.pdf.

Brückner, M. (2011). On the simultaneity problem in the aid and growth debate. *Journal of Applied Econometrics*, forthcoming. doi:10.1002/jae.1259.

Buchmann, C. and Hannum, E. (2001). Education and stratification in developing countries: a review of theories and research. *Annual Review of Sociology*, 27:77–102.

Bulir, A. and Hamann, A.J. (2008). Volatility of Development Aid: From the Frying Pan into the Fire? *World Development*, 36(10):2048–2066.

Bun, M.J. and Windmeijer, F. (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *The Econometrics Journal*, 13(1):95–126.

Burnside, C. and Dollar, D. (2000). Aid, policies, and growth. *The American Economic Review*, 90(4):847–868.

Buvé, A., Bishikwabo-Nsarhaza, K. and Mutangadura, G. (2002). The spread and effect of HIV-1 infection in sub-Saharan Africa. *The Lancet*, 359(9322):2011–2017.

Byerlee, D., Jayne, T. and Myers, R. (2006). Managing food price risks and instability in a liberalizing market environment: Overview and policy options. *Food Policy*, 31(4):275–287.

Caines, K. (2005). Key evidence from major studies of selected global health partnerships. Background paper prepared for the High–Level forum on the health MDGs, working group on global health initiatives and partnerships, UK, DfID Health Resource Centre. URL www.hlfhealthmdgs.org/Documents/GHPBackgroundPaperFinal.pdf.

Card, D. (1990). The impact of the Mariel boatlift on the Miami labor market. *Industrial and Labor Relations Review*, 43(2):245–257.

—— (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, 69(5):1127–1160.

Card, D., DellaVigna, S. and Malmendier, U. (2011). The Role of Theory in Field Experiments. NBER Working Paper 17047, National Bureau of Economic Research.

Carpenter, D. (2003). An investigation into the transition from technological to ecological rice farming among resource poor farmers from the Philippine island of Bohol. *Agriculture and Human Values*, 20(2):165–176.

Caselli, F. and Feyer, J. (2007). The Marginal Productivity of Capital. *Quarterly Journal of Economics*, 72(122):535–568.

Cassen, R. and Associates (1994). *Does Aid Work?: Report to an Intergovernmental Task Force*. ISBN-13 9780198773863. Oxford: Oxford University Press.

Chong, A. and Gradstein, M. (2008). What determines foreign aid? The donors' perspective. *Journal of Development Economics*, 87(1):1–13.

Clemens, M. and Bazzi, S. (2009). Blunt instruments: on establishing the causes of economic growth. Working paper 171, Center for Global Development.

Clemens, M.A., Kenny, C.J. and Moss, T.J. (2007). The trouble with the MDGs: confronting expectations of aid and development success. *World Development*, 35(5):735–751.

Clemens, M., Radelet, S. and Bhavnani, R. (2004). Counting Chickens When They Hatch: The Short–term Effect of Aid on Growth. Working Papers 44, Center for Global Development.

Clemens, M., Radelet, S., Bhavnani, R. and Bazzi, S. (2011). Counting Chickens When They Hatch: Timing And The Effects Of Aid On Growth. *The Economic Journal*, forthcoming. doi:10.1111/j.1468-0297.2011.02482.x.

Coakley, J., Fuertes, A.M. and Smith, R. (2006). Unobserved heterogeneity in panel time series models. *Computational Statistics & Data Analysis*, 50(9):2361–2380.

Cohen, D. and Soto, M. (2007). Growth and human capital: good data, good results. *Journal of Economic Growth*, 12(1):51–76.

Cole, S., Platt, R., Schisterman, E., Chu, H., Westreich, D., Richardson, D. and Poole, C. (2010). Illustrating bias due to conditioning on a collider. *International Journal of Epidemiology*, 39(2):417.

Coleman, B. (1999). The impact of group lending in Northeast Thailand. *Journal of Development Economics*, 60(1):105–141.

Conley, T. and Udry, C. (2001). Social learning through networks: The adoption of new agricultural technologies in Ghana. *American Journal of Agricultural Economics*, 83(3):668–673.

Conlin, S. and Beauean, S. (2009). PCR Synthesis Report. Evaluation Report Ev705, Department for International Development, UK. URL www.dfid.gov.uk/Documents/publications1/evaluation/proj-synth-rpt-ev705.pdf.

Crawford, E., Kelly, V., Jayne, T. and Howard, J. (2003). Input use and market development in sub-Saharan Africa: an overview. *Food Policy*, 28(4):277–292.

Dalgaard, C.J. and Erickson, L. (2009). Reasonable expectations and the first Millennium Development Goal: how much can aid achieve? *World Development*, 37(7):1170–1181.

Dalgaard, C.J. and Hansen, H. (2005). The Return to Foreign Aid. Discussion Papers 05–04, University of Copenhagen. Department of Economics.

Damiani, O. (2002). Organic agriculture in Costa Rica: the case of cocoa and banana production in Talamanca. Technical report, Office of Evaluation and Studies, IFAD, Rome.

Dang, H.A., Knack, S. and Rogers, H. (2009). International aid and financial crises in donor countries. Policy Research Working Paper Series 5162, The World Bank.

Dave, D. and Kaestner, R. (2002). Alcohol taxes and labor market outcomes. *Journal of Health Economics*, 21(3):357–371.

Dawid, A. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95(450):407–424.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48(2):424–455.

Deininger, K. and Mpuga, P. (2005). Economic and Welfare Impact of the Abolition of Health User Fees: Evidence from Uganda. *Journal of African Economies*, 14(1):55–91.

Delph, E. (2008). Global Vertical Programmes: The Future of Aid or a Quick Fix? Reality check, Eurodad. URL www.eurodad.org/whatsnew/reports.aspx?id=2736.

Destler, I. (1996). Foreign Policy and the Public: Will Leaders Catch the Full Message. *Brown Journal of World Affairs*, III(I):265.

Dimitri, C., Greene, C. and Wellson, A.J. (2007). Recent growth patterns in the US organic foods market. In A. Wellson (Ed.), *Organic agriculture in the US*, pp. 129–190. Hauppage, NY: Nova.

Djankov, S., Montalvo, J.G. and Reynal-Querol, M. (2008). The curse of aid. *Journal of Economic Growth*, 13(3):169–194.

Dodd, R., Schieber, G., Cassels, A., Fleisher, L. and Gottret, P. (2007). Making health systems work. WHO/HSS/healthsystems Working Paper 9, Department for Health Policy, Development and Services, WHO, Geneva.

Dolan, C. and Humphrey, J. (2000). Governance and trade in fresh vegetables: the impact of UK supermarkets on the African horticulture industry. *Journal of Development Studies*, 37(2):147–176.

Dorward, A., Fan, S., Kydd, J., Lofgren, H., Morrison, J., Poulton, C., Rao, N., Smith, L., Tchale, H., Thorat, S. et al. (2004). Institutions and Policies for Pro–poor Agricultural Growth. *Development Policy Review*, 22(6):611–622.

Doss, C. (2006). Analysing technology adoption using micro-studies: limitations, challenges and opportunities for improvement. *Agricultural Economics*, 34:207–219.

Doucouliagos, H. and Paldam, M. (2006). Aid effectiveness on accumulation: A meta study. *Kyklos*, 59(2):227–254.

—— (2008). Aid effectiveness on growth: A meta study. *European Journal of Political Economy*, 24(1):1–24.

—— (2009). The Aid Effectiveness Literature: The sad results of 40 years of research. *Journal of Economic Surveys*, 23(3):433–461.

—— (2010). Conditional aid effectiveness: A meta-study. *Journal of International Development*, 22(4):391–410.

Dreher, A., Nunnenkamp, P. and Thiele, R. (2008). Does Aid for Education Educate Children? Evidence from Panel Data. *The World Bank Economic Review*, 22(2):291–314.

Easterly, W. (2003). Can foreign aid buy growth? *The Journal of Economic Perspectives*, 17(3):23–48.

—— (2005). How to assess the need for aid? The answer: Don't ask. In *Financing Development: What are the Challenges in Expanding Aid Flows?*, Proceedings of the 3rd AFD-EUDN Conference. Research Department, Agence Française de Développement.

Easterly, W.R. (2006). *The white man's burden: why the West's efforts to aid the rest have done so much ill and so little good*. Penguin Group USA. ISBN 1594200378.

Easterly, W. (Ed.) (2008). *Reinventing foreign aid*, volume 1. The MIT Press. ISBN 0262550660.

Easterly, W., Levine, R. and Roodman, D. (2004). Aid, policies, and growth: comment. *The American Economic Review*, 94(3):774–780.

Ellis, F. (1993). *Peasant economics: farm households and agrarian development*. Cambridge: Cambridge Univ Press. ISBN 0521457114.

Engle, R.F. and Granger, C.W.J. (1987). Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2):251–76.

Ericsson, N.R. and MacKinnon, J.G. (2002). Distributions of error correction tests for cointegration. *Econometrics Journal*, 5(2):285–318.

Eyhorn, F. (2007). *Organic Farming for Sustainable Livelihoods in Developing Countries? The Case of Cotton in India*. Zurich: vdf Hochschulverlag AG. ISBN 3728131113.

Fafchamps, M. (2003). *Rural poverty, risk and development*. Edward Elgar Publishing. ISBN 1843764369.

Faye, M., McArthur, J., Sachs, J. and Snow, T. (2004). The challenges facing landlocked developing countries. *Journal of Human Development*, 5(1):31–68.

Feyzioglu, T., Swaroop, V. and Zhu, M. (1998). A Panel Data Analysis of the Fungibility of Foreign Aid. *World Bank Economic Review*, 12(1):29–58.

Fleck, R.K. and Kilby, C. (2010). Changing aid regimes? U.S. foreign aid from the Cold War to the War on Terror. *Journal of Development Economics*, 91(2):185–197.

Foster, V. (2008). Africa infrastructure country diagnostic: overhauling the engine of growth. Unpublished report, World Bank. URL http://siteresources.worldbank.org/INTAFRICA/Resources/AICD_exec_summ_9-30-08a.pdf.

Foster, M., Brown, A. and Naschold, F. (2001). Sector programme approaches: will they work in agriculture? *Development Policy Review*, 19(3):321–338.

Frot, E. (2009). Aid and the Financial Crisis: Shall we expect development aid to fall? Technical report, Stockholm: Stockholm School of Economics. URL http://dx.doi.org/10.2139/ssrn.1402788.

Gao, C. and Lahiri, K. (2000). Further consequences of viewing LIML as an iterated Aitken estimator. *Journal of Econometrics*, 98(2):187–202.

Garrett, L. (2007). The Challenge of Global Health. *Foreign Affairs*, 86(15):14–38.

Garrett, L.A. (2009). *The Future of Foreign Assistance Amid Global Economic and Financial Crisis: Advancing Global Health in the U.S. Development Agenda.* Council on Foreign Relations Press.

Gerson, M. (2008). The AIDS Relief Miracle, (Op-Ed March 12). *The Washington Post*. URL www.washingtonpost.com/wp-dyn/content/article/2008/03/11/AR2008031102460.html.

Gibbon, P. (2006). An overview of the certified organic export sector in Uganda. DIIS Working Paper 2006/13, Danish Institute for International Studies.

Gibbon, P. and Bolwig, S. (2007). The Economic Impact of a Ban on Imports of Air freighted Organic Products to the UK. DIIS Working Paper 2007/23, Danish Institute for International Studies.

Giovannucci, D., Potts, J., Killian, B., Wunderlich, C., Schuller, S., Soto, G., Schroeder, K., Vagneron, I. and Pinard, F. (2008). *Seeking sustainability: COSA preliminary analysis of sustainability initiatives in the coffee sector*. International Institute for Sustainable Development.

Girishankar, N. (2009). Innovating Development Finance. Policy Research Working Paper 5111, Washington D.C.: World Bank.

Glennie, J. and Prizzon, A. (2012). From high to low aid: a proposal to classify countries by aid receipt. ODI Background Note, Overseas Development Institute. URL www.odi.org.uk/resources/docs/7621.pdf.

GlobalGiving (2008). *2006-2008 Report: A two-year report on the good things going on at GlobalGiving, one earth-changing project after another*. GlobalGiving. URL http://www.globalgiving.com/aboutus/media/GGF_2007_AR.pdf.

Glover, D. (1984). Contract farming and smallholder outgrower schemes in less-developed countries. *World Development*, 12(11–12):1143–1157.

—— (1987). Increasing the benefits to smallholders from contract farming: Problems for farmers' organizations and policy makers. *World Development*, 15(4):441–448.

Godal, T. (2005). Do we have the architecture for health aid right? Increasing global aid effectiveness. *Nature Reviews Microbiology*, 3(11):899–903.

Gogoe, S. (2003). *Costs and Benefits of Small-Holder Compliance with EurepGAP Protocol in Ghana*. Master's thesis, Natural Resources Institute, Univsersity of Greenwich.

Gomanee, K., Girma, S. and Morrissey, O. (2005). Aid, public spending and human welfare: evidence from quantile regressions. *Journal of International Development*, 17(3):299–309.

Gonzalo, J. and Pitarakis, J.Y. (1998). Specification via model selection in vector error correction models. *Economics Letters*, 60(3):321–328.

Goodman, L.A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, 55:708–713.

Gootnick, D.B. (2007). Millennium Challenge Corporation: Progress and Challenges with Compacts in Africa, GAO Testimony Before the Subcommittee on Africa and Global Health. GAO Testimony Before the Subcommittee on Africa and Global Health, Committee on Foreign Affairs, House of Representatives, USA. URL www.gao.gov/new.items/d071049t.pdf.

Govereh, J. and Jayne, T. (2003). Cash cropping and food crop productivity: synergies or trade-offs? *Agricultural Economics*, 28(1):39–50.

Greene, W.H. (2002). *Econometric Analysis*. Prentice Hall, 5th edition.

Guillaumont, P. (2008). Adapting aid allocation criteria to development goals. Working Papers 2008/10, Paris: CERDI.

Hallet, M. (2009). Economic cycles and development aid: What is the evidence from the past? ECFIN Economic Brief 5, European Commission's Directorate-General for Economic and Financial Affairs.

Hansen, H. and Tarp, F. (2000). Aid effectiveness disputed. *Journal of International Development*, 12(3):375–398.

—— (2001). Aid and growth regressions. *Journal of Development Economics*, 64(2):547–570.

Hanushek, E. and Kimko, D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review*, 90(5):1184–1208.

Haque, N.U., Sharma, S. and Pesaran, M.H. (1999). Neglected Heterogeneity and Dynamics in Cross-Country Savings Regressions. IMF Working Papers 99/128, International Monetary Fund.

Hauk, W.R. and Wacziarg, R. (2009). A Monte Carlo study of growth regressions. *Journal of Economic Growth*, 14(2):103–147.

Havnevik, K.J., Bryceson, D., Birgegaard, L.E., Matondi, P. and Beyene, A. (2007). *African agriculture and the World Bank: development or impoverishment?* Uppsala: Nordic Africa Institute.

Hazell, P., Dorward, A., Poulton, C. and Wiggins, S. (2007). The future of small farms for poverty reduction and growth. 2020 Discussion Paper 42, International Food Policy Research Institute.

Heckman, J.J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Heckman, J. (2001). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy*, 109(4):673–748.

Heckman, J.J. (2008). Econometric Causality. *International Statistical Review*, 76(1):1–27.

Heckman, J.J., LaLonde, R.J. and Smith, J.A. (1999). The economics and econometrics of active labor market programs. *Handbook of Labor Economics*, 3:1865–2097.

Herrling, S., Kinder, M. and Radelet, S. (2009). From innovation to impact: Next steps for the Millennium Challenge Corporation. MCA Monitor Analysis, Washington, DC: Center for Global Development.

Hii, C., Law, C., Cloke, M. and Suzannah, S. (2009). Thin layer drying kinetics of cocoa and dried product quality. *Biosystems engineering*, 102(2):153–161.

Hilts, P.J. (2005). *Rx for Survival: why we must rise to the global health challenge*. Penguin Press. ISBN 159420070X.

Hjertholm, P. and White, H. (2000). Survey of Foreign Aid: History, Trends and Allocation. Discussion Papers 00–04, University of Copenhagen. Department of Economics.

Hoffman, K. (2008). Placing Enterprise and Business Thinking at the Heart of the War on Poverty. In Easterly (2008), chapter 18.

Hubbard-Preston, H. (2008). The new face of private philanthropy (26 Sept.). *The New York Times*. URL www.nytimes.com/2008/09/26/your-money/26iht-mphil.html.

Hudson, J. and Mosley, P. (2008). The macroeconomic impact of aid volatility. *Economics Letters*, 99(3):486–489.

ICCO (2006). A study on the market for organic cocoa. Technical Report EX/130/10, International Cocoa Organization.

—— (2007). Supply chain management for total quality in Africa. Technical Report EX/133/7, International Cocoa Organization.

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

Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1):4–29.

Imbens, G.W. and Wooldridge, J.M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1):5–86. doi:10.1257/jel.47.1.5.

Independent Evaluation Group (2010). *Results and Performance of the World Bank Group: IEG Annual Report 2010*. Washington D.C. : World Bank Group. URL http://go.worldbank.org/EWM6Z9T2E0.

Jack, W. and Lewis, M. (2009). Health Investments and Economic Growth: Macroeconomic Evidence and Microeconomic Foundations. In M. Spence and M. Lewis (Eds.), *Health and Growth*, pp. 1–41. World Bank.

Jalan, J. and Ravallion, M. (2003). Does piped water reduce diarrhea for children in rural India? *Journal of Econometrics*, 112(1):153–173.

Johnson, S., Larson, W.D., Papageorgiou, C. and Subramanian, A. (2009). Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates. NBER Working Papers 15455, National Bureau of Economic Research, Inc.

Jones, S. (2011). Aid Supplies Over Time: Accounting for Heterogeneity, Trends, and Dynamics. UNU-WIDER Working Paper 2011/04, World Institute for Development Economic Research (UNU-WIDER).

—— (2012). Innovating foreign aid – progress and problems. *Journal of International Development*, 24:1–16. doi:10.1002/jid.1758.

Jones, S. and Gibbon, P. (2011). Developing Agricultural Markets in Sub-Saharan Africa: Organic Cocoa in Rural Uganda. *Journal of Development Studies*, 47(10):1595–1618.

Juselius, K., Framroze, N. and Tarp, F. (2011). The Long-Run Impact of Foreign Aid in 36 African Countries: Insights from Multivariate Time Series Analysis. UNU-WIDER Working Paper 2011/51, World Institute for Development Economics Research (UNU-WIDER). URL http://ideas.repec.org/p/unu/wpaper/wp2011-51.html.

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics*, 90(1):1–44.

Key, N. and Runsten, D. (1999). Contract farming, smallholders, and rural development in Latin America: the organization of agroprocessing firms and the scale of outgrower production. *World Development*, 27(2):381–401.

Kirby-Zaki, J., Merghoub, B. and Selowsky, M. (2008). Global Program Funds at Country Level: What have we learned? Unpublished report, The Global Programs and Partnership Group, Concessional Finance and Global Partnerships Vice Presidency, The World Bank.

Klasen, S. and Misselhorn, M. (2008). Determinants of the Growth Semi-Elasticity of Poverty Reduction. IAI Discussion Papers 176, Ibero-America Institute (IAI) for Economic Research.

Knack, S. (2001). Aid dependence and the quality of governance: cross-country empirical tests. *Southern Economic Journal*, 68(2):310–329.

Knack, S. and Rahman, A. (2008). Donor Fragmentation. In Easterly (2008), chapter 11.

Koning, N. (2002). Should Africa protect its farmers to revitalize its economy? Gatekeeper's Series 105, International Institute for Environment and Development, UK.

Koopmans, T. (1947). Measurement without theory. *The Review of Economics and Statistics*, 29(3):161–172.

Kremer, M. and Miguel, E. (2004). Networks, social learning, and technology adoption: the case of deworming drugs in Kenya. *Econometrica*, 72(1):159–217.

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):379–392.

La Porta, R., F., L. and Shleifer, A. (2008). The economic consequences of legal origins. *Journal of Economic Literature*, 46(2):285–332.

Laeven, L. and Valencia, F. (2010). Resolution of Banking Crises: The Good, the Bad, and the Ugly. IMF Working Papers 10/146, International Monetary Fund.

Lane, T., Ghosh, A., Hamann, J., Phillips, S., Schulze-Ghattas, M. and Tsikata, T. (1999). IMF-supported programs in Indonesia, Korea, and Thailand: a preliminary assessment. Occaisonal Paper 178, International Monetary Fund.

Leal, P.A. (2007). Participation: the ascendancy of a buzzword in the neo-liberal era. *Development in Practice*, 17(4):539–548.

Lee, K., Pesaran, M.H. and Smith, R. (1997). Growth and Convergence in Multi-country Empirical Stochastic Solow Model. *Journal of Applied Econometrics*, 12(4):357–92.

Lin, S. (1976). The modified internal rate of return and investment criterion. *The Engineering Economist*, 21(4):237–247.

Little, P. and Watts, M. (1994). *Living under contract: contract farming and agrarian transformation in sub-Saharan Africa*. University of Wisconsin Press. ISBN 0299140644.

Lokshin, M. and Yemtsov, R. (2005). Has Rural Infrastructure Rehabilitation in Georgia Helped the Poor? *World Bank Economic Review*, 19(2):311–333.

Losch, B. (2002). Global Restructuring and Liberalization: Côte d'Ivoire and the End of the International Cocoa Market? *Journal of Agrarian Change*, 2(2):206–227.

Lu, C., Michaud, C.M., Gakidou, E., Khan, K. and Murray, C.J. (2006). Effect of the Global Alliance for Vaccines and Immunisation on diphtheria, tetanus, and pertussis vaccine coverage: an independent assessment. *The Lancet*, 368(9541):1088–1095.

Lunceford, J.K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*, 23(19):2937–2960.

Lyngbaek, A., Muschler, R. and Sinclair, F. (2001). Productivity and profitability of multi-strata organic versus conventional coffee farms in Costa Rica. *Agroforestry Systems*, 53(2):205–213.

MacKinnon, J.G. (1996). Numerical Distribution Functions for Unit Root and Cointegration Tests. *Journal of Applied Econometrics*, 11(6):601–18.

Maddala, G.S. and Wu, S. (1999). A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test. *Oxford Bulletin of Economics and Statistics*, 61(0):631–52.

Maertens, M., Dries, L., Dedehouanou, F. and Swinnen, J. (2007). High-value supply chains, food standards and rural households in Senegal. In J. Swinnen (Ed.), *Global supply chains, standards and the poor. How the globalization of food systems and standards affects rural development and poverty*, pp. 159–172. Wallingford: CABI International.

Maertens, M. and Swinnen, J. (2009). Trade, standards, and poverty: Evidence from Senegal. *World Development*, 37(1):161–178.

Magud, N. and Sosa, S. (2010). When and Why Worry About Real Exchange Rate Appreciation? The Missing Link between Dutch Disease and Growth. IMF Working Paper 10/271, International Monetary Fund.

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

Mark, N.C. and Sul, D. (2003). Cointegration Vector Estimation by Panel DOLS and Long-run Money Demand. *Oxford Bulletin of Economics and Statistics*, 65(5):655–680.

Martens, B. (2005). Why do aid agencies exist? *Development Policy Review*, 23(6):643–663.

Masud, N. and Yontcheva, B. (2007). Does foreign aid reduce poverty? Empirical evidence from nongovernmental and bilateral aid. IMF Working Papers 07/100, International Monetary Fund.

Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, 110(3):681–712.

Mawdsley, E. (2007). The Millennium Challenge Account: Neo-liberalism, poverty and security. *Review of International Political Economy*, 14(3):487–509.

McCarthy, M. (2007). The Global Fund: 5 years on. *Lancet*, 370(9584):307–308.

McGillivray, M. (2003). Modelling aid allocation: issues, approaches and results. *Journal of Economic Development*, 28(1):171–188.

Mekasha, T.J. and Tarp, F. (2011). Aid and Growth: What Meta-Analysis Reveals. UNU-WIDER Working Paper 2011/22, World Institute for Development Economics Research (UNU-WIDER).

Michaelowa, K. (2004). Aid Effectiveness Reconsidered – Panel Data Evidence for the Education Sector. Discussion Paper Series 26374, Hamburg Institute of International Economics.

Minten, B., Randrianarison, L. and Swinnen, J. (2009). Global retail chains and poor farmers: Evidence from Madagascar. *World Development*, 37(11):1728–1741.

Mishra, P. and Newhouse, D. (2009). Does health aid matter? *Journal of Health Economics*, 28(4):855–872.

Mold, A., Prizzon, A., Frot, E. and Santiso, J. (2010). Aid Flows in Times of Crises. In *Conference on development cooperation in times of crisis and on achieving the MDGs (IFEMA Convention Centre, Madrid)*.

Montalvo, J. and Reynal-Querol, M. (2005). Ethnic diversity and economic development. *Journal of Development Economics*, 76(2):293–323.

Moral-Benito, E. (2012). Determinants of Economic Growth: A Bayesian Panel Data Approach. *The Review of Economics and Statistics*, 94(2):566–579.

Morduch, J. (1998). Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh. Working paper 198, Princeton University, Woodrow Wilson School of Public and International Affairs, Research Program in Development Studies.

Morgan, S. and Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge University Press.

Mosley, P. (1985). The Political Economy of Foreign Aid: A Model of the Market for a Public Good. *Economic Development and Cultural Change*, 33(2):373–93.

—— (1986). Aid-effectiveness: The Micro-Macro Paradox. *IDS Bulletin*, 17(2):22–27.

—— (1987). *Foreign aid, its defense and reform*. University Press of Kentucky. ISBN 0813116082.

Mosley, P., Hudson, J. and Horrell, S. (1987). Aid, the Public Sector and the Market in Developing countries. *The Economic Journal*, 97(387):616–641.

Moss, T., Pettersson, G. and van de Walle, N. (2006). An Aid-Institutions Paradox? A Review Essay on Aid Dependency and State Building in sub-Saharan Africa. Working Papers 74, Center for Global Development.

Moyo, D. (2009). *Dead Aid: Why aid is not working and how there is a better way for Africa*. London: Penguin books.

Nieberg, H. and Offerman, F. (2003). The profitability of organic farming in Europe. In D. Jones (Ed.), *Organic agriculture: Sustainability, markets and policies*, pp. 140–150. Paris: OECD.

Nielson, D., Powers, R. and Tierney, M. (2009). Broad Trends in Foreign Aid: Insights from PLAID 1.6. URL http://irtheoryandpractice.wm.edu/projects/plaid/workshop_papers/nielsonpowerstierney_14sep09.pdf.

Nunn, N. (2008). The long-term effects of Africa's slave trades. *Quarterly Journal of Economics*, 123(1):139–176.

—— (2009). The Importance of History for Economic Development. *Annual Review of Economics*, 1:65–92.

Obstfeld, M. (1999). Foreign resource inflows, saving and growth. In K. Schmidt-Hebbel and L. Serven (Eds.), *The Economics of Saving and Growth: theory, evidence, and implications for policy*, chapter 5, pp. 107–146. Cambridge, UK: Cambridge University Press.

Odedokun, M. (2003). Analysis of deviations and delays in aid disbursements. *Journal of Economic Development*, 28(1):137–169.

OECD (2008a). Aid targets slipping out of reach? Report of the development finance working group, Organisation for Economic Co-operation and Development. URL `http://www.oecd.org/dataoecd/47/25/41724314.pdf`.

—— (2008b). The Paris Declaration on Aid Effectiveness and the Accra Agenda for Action. Technical report, Organisation for Economic Co-operation and Development. URL `www.oecd.org/dataoecd/11/41/34428351.pdf`.

Oomman, N., Bernstein, M. and Rosenzweig, S. (2007). Following the Funding for HIV/AIDS: a comparative analysis of the funding practices of PEPFAR, the Global Fund and World Bank MAP in Mozambique, Uganda and Zambia. HIV/AIDS Monitor, Center for Global Development.

—— (2008). *Seizing the opportunity on AIDS and health systems*. Washington DC: Center for Global Development.

Ostrom, E., Clark, G., Shivakumar, S. and Andersson, K. (2002). Aid, incentives, and sustainability: an institutional analysis of development cooperation. SIDA Studies in Evaluation 02/01, Swedish International Development Cooperation Agency (SIDA).

Oxman, A.D. and Fretheim, A. (2008). An overview of research on the effects of results-based financing. Technical Report 16–2008, Nasjonalt kunnskapssenter for helsetjenesten, Oslo.

Pack, H. and Pack, J.R. (1990). Is Foreign Aid Fungible? The Case of Indonesia. *Economic Journal*, 100(399):188–94.

—— (1993). Foreign Aid and the Question of Fungibility. *The Review of Economics and Statistics*, 75(2):258–65.

Pagan, A. (1979). Some consequences of viewing LIML as an iterated Aitken estimator. *Economics Letters*, 3(4):369–372.

Pallage, S. and Robe, M.A. (2001). Foreign Aid and the Business Cycle. *Review of International Economics*, 9(4):641–72.

Pariente, W. (2002). The impact of fair trade on a coffee cooperative in Costa Rica: A producers' behaviour approach. Technical report, Universite Paris I, Pantheon Sorbonne.

Parrott, N., Elzakker, B.V. and Eco, A. (2003). *Organic and like-minded movements in Africa*. Tholey-Theley: IFOAM.

Paul, E. (2006). A survey of the theoretical economic literature on foreign aid. *Asian Pacific Economic Literature*, 20(1):1–17.

Pearl, J. (2009). *Causality: Models, reasoning and inference*. Cambridge University Press, 2nd edition.

—— (2010). An Introduction to Causal Inference. *The International Journal of Biostatistics*, 6(2):Article 7.

PEPFAR (2008). The Power of Partnerships: The U.S. President's Emergency Plan for AIDS Relief. Annual report to congress, Office of the United States Global AIDS Coordinator.

Persyn, D. and Westerlund, J. (2008). Error-correction-based cointegration tests for panel data. *Stata Journal*, 8(2):232–241.

Pesaran, M.H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica*, 74(4):967–1012.

Pesaran, M.H. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68(1):79–113.

Piron, L.H. (2002). The right to development: a review of the current state of the debate for the department for international development. Unpublished report, Overseas Development Institute, UK. URL www.odi.org.uk/resources/download/1562.pdf.

Pitt, M. and Khandker, S. (1998). The impact of group-based credit programs on poor households in Bangladesh: Does the gender of participants matter? *Journal of Political Economy*, pp. 958–996.

Pivovarsky, A., Clements, B.J., Gupta, S. and Tiongson, E. (2003). Foreign Aid and Revenue Response: Does the Composition of Aid Matter? IMF Working Papers 03/176, International Monetary Fund.

Ponte, S. (2002). Brewing a Bitter Cup? Deregulation, Quality and the Re-organization of Coffee Marketing in East Africa. *Journal of Agrarian Change*, 2(2):248–272.

Ponte, S., Richey, L.A. and Baab, M. (2009). Bono's Product (RED) Initiative: corporate social responsibility that solves the problems of 'distant others'. *Third World Quarterly*, 30(2):301–317.

Poore, P. (2004). The Global Fund to fight AIDS, tuberculosis and malaria (GFATM). *Health Policy and Planning*, 19(1):52–53.

Poulton, C., Gibbon, P., Hanyani-Mlambo, B., Kydd, J., Maro, W., Larsen, M., Osorio, A., Tschirley, D. and Zulu, B. (2004). Competition and coordination in liberalized African cotton market systems. *World Development*, 32(3):519–536.

Pritchett, L. (2001). Where Has All the Education Gone? *The World Bank Economic Review*, 15(3):367–391. doi:10.1093/wber/15.3.367.

—— (2002). It pays to be ignorant: a simple political economy of rigorous program evaluation. *The Journal of Policy Reform*, 5(4):251–269.

Puhani, P. (2000). The Heckman correction for sample selection and its critique. *Journal of Economic Surveys*, 14(1):53–68.

Radelet, S. (2004). The Global Fund to Fight AIDS, Tuberculosis and Malaria: Progress, potential, and challenges for the future. Unpublished report, Washington DC: Center for Global Development.

Rajan, R. and Subramanian, A. (2005). What undermines aid's impact on growth? NBER Working Paper 11657, National Bureau of Economic Research, Cambridge, MA.

—— (2007). Does Aid Affect Governance? *The American Economic Review*, 97(2):322–327.

Rajan, R.G. and Subramanian, A. (2008). Aid and Growth: What Does the Cross-Country Evidence Really Show? *The Review of Economics and Statistics*, 90(4):643–665.

—— (2011). Aid, Dutch disease, and manufacturing growth. *Journal of Development Economics*, 94(1):106–118.

Ravallion, M. (2005). Externalities in rural development: Evidence for China. In R. Kanbur and A. Venables (Eds.), *Spatial Inequality and Development*, pp. 137–162. Oxford University Press.

—— (2009). Evaluation in the Practice of Development. *World Bank Research Observer*, 24(1):29–53.

—— (2012). Fighting Poverty One Experiment at a Time:'Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty': Review Essay. *Journal of Economic Literature*, 50(1):103–114.

Reardon, T. and Berdegué, J. (2007). The retail-led transformation of agrofood systems and its implications for development policies. Background paper written for World Development Report 2008, World Bank. URL www.rimisp.org/getdoc.php?docid=6432.

Reich, M.R., Takemi, K., Roberts, M.J. and Hsiao, W.C. (2008). Global action on health systems: a proposal for the Toyako G8 summit. *Lancet*, 371(9615):865–869.

Riddell, R. (2007). *Does foreign aid really work?* Oxford University Press, USA. ISBN 0199295654.

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

Rodrik, D. (2009). The real exchange rate and economic growth. *Brookings Papers on Economic Activity*, 2008(2):365–412.

Roodman, D. (2007). The anarchy of numbers: aid, development, and cross-country empirics. *The World Bank Economic Review*, 21(2):255–277.

—— (2008). History Says Financial Crisis Will Suppress Aid. Global Development: Views from the Center. URL http://blogs.cgdev.org/globaldevelopment/2008/10/history-says-financial-crisis.php.

—— (2009). A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics*, 71(1):135–158.

Round, J.I. and Odedokun, M. (2004). Aid effort and its determinants. *International Review of Economics & Finance*, 13(3):293–309.

Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

—— (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Sachs, J. (2006). *The end of poverty: economic possibilities for our time*. Penguin Group USA. ISBN 0143036580.

Sachs, J. and Warner, A. (1997). Fundamental sources of long-run growth. *The American Economic Review*, 87(2):184–188.

Sahn, D., Dorosh, P. and Younger, S. (1999). *Structural adjustment reconsidered: economic policy and poverty in Africa*. Cambridge University Press.

Sala-i-Martin, X. (1997). I just ran two million regressions. *The American Economic Review*, 87(2):178–183.

Sala-i-Martin, X., Doppelhofer, G. and Miller, R.I. (2004). Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *The American Economic Review*, 94(4):813–835.

Savedoff, W., Levine, R. and Birdsall, N. (2006). When will we ever learn? Improving lives through impact evaluation. Report of the evaluation gap working group, Center for Global Development, Washington D.C.

Shea, J. (1997). Instrument relevance in multivariate linear models: a simple measure. *Review of Economics and Statistics*, 49:348–352.

Simmons, P. (2003). Overview of smallholder contract farming in developing countries. FAO-ESA Working Papers 02/04, Agricultural and Development Economics Division of the Food and Agriculture Organization of the United Nations.

Singer, P. (2009). *The life you can save: Acting now to end world poverty*. New York: Random House, Inc.

Smith, R.P. (2001). Estimation and inference with non-stationary panel time-series data. Paper prepared for the RC33 Conference, Cologne October 2000, Department of Economics, Birkbeck College, London.

Sridhar, D. and Batniji, R. (2008). Misfinancing global health: a case for transparency in disbursements and decision making. *The Lancet*, 372(9644):1185–1191.

Stiglitz, J.E. (2002). Information and the change in the paradigm in economics. In T. Frangsmyr (Ed.), *Les Prix Nobel. The Nobel Prizes 2001*. Stockholm: Nobel Foundation.

Stiglitz, J. (2003). *Globalization and its Discontents*. WW Norton & Company.

Stock, J.H. and Watson, M.W. (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica*, 61(4):783–820.

Stubbs, R. (2009). The millennium challenge account: influencing governance in developing countries through performance-based foreign aid. *Vanderbilt Journal of Transnational Law*, 42:621–1029.

Svensson, J. (2000). Foreign aid and rent-seeking. *Journal of International Economics*, 51(2):437–461. doi:DOI:10.1016/S0022-1996(99)00014-8.

—— (2006). The institutional economics of foreign aid. *Swedish Economic Policy Review*, 13(2):115–137.

Swamy, P.A.V.B. (1970). Efficient Inference in a Random Coefficient Regression Model. *Econometrica*, 38(2):311–23.

Tarp, F. (2000). *Foreign Aid and Development: Lessons Learnt and Guidelines for the Future*. Routledge: London and New York.

—— (2006). Aid and development. *Swedish Economic Policy Review*, 13:9–61.

Tarp, F., Bach, C., Hansen, H. and Baunsgaard, S. (1999). Danish Aid Policy: Theory and Empirical Evidence. In K.L. Gupta (Ed.), *Foreign aid: new perspectives*. Kluwer Academic Publishers. ISBN 0792384040.

Tavares, J. (2003). Does foreign aid corrupt? *Economics Letters*, 79(1):99–106.

Tavares, J. and Wacziarg, R. (2001). How democracy affects growth. *European Economic Review*, 45(8):1341–1378.

Teal, F., Zeitlin, A. and Maamah, H. (2006). Ghana Cocoa Farmers Survey 2004: Report to Ghana Cocoa Board. Technical report, Centre for the Study of African Economies, University of Oxford and Haruna Maamah ECAM Consultancy Ltd., Accra, Ghana. URL www.gprg.org/pubs/reports/pdfs/2006-04-teal-zeitlin-maamah.pdf.

Tedeschi, G. (2008). Overcoming selection bias in microcredit impact assessments: a case study in Peru. *Journal of Development Studies*, 44(4):504–518.

Teera, J.M. and Hudson, J. (2004). Tax performance: a comparative study. *Journal of International Development*, 16(6):785–802.

Temple, J. (2000). Growth regressions and what the textbooks don't tell you. *Bulletin of Economic Research*, 52(3):181–205.

—— (2010). Aid and Conditionality. *Handbook of Development Economics*, 5:4416–4522.

The Global Fund (2007). Summary Paper on Study Area 1: Organizational Efficiency & Effectiveness. Technical report, Technical Evaluation Reference Group, The Global Fund. URL http://www.theglobalfund.org/documents/board/16/GF-BM16-04-ReportTERG5YearSA1.pdf.

Tian, J. and Pearl, J. (2002). A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pp. 567–573. Menlo Park, CA: AAAI Press/The MIT Press.

Tropical Commodity Coalition (2009). Cocoa Barometer 2009. Technical report, Tropical Commodity Coalition, The Hague.

Tsikata, T.M. (1998). Aid Effectiveness – A Survey of the Recent Empirical Literature. IMF Policy Discussion Papers 98/1, International Monetary Fund.

Uganda Bureau of Statistics (2009). Consumer Price Index: September 2009. Monthly bulletin, Uganda Bureau of Statistics. URL www.ubos.org/onlinefiles/uploads/ubos/cpi/septcpi2009.pdf.

Uganda Coffee Development Authority (2003). Annual report 2003. Technical report, Uganda Coffee Development Authority, Kampala.

—— (2007). Monthly report for September 2007. Technical report, Uganda Coffee Development Authority, Kampala.

UNCTAD (2008). *World Investment Report 2008: TNCs, FDI and Infrastructure*. Geneva: UNCTAD.

Vella, F. and Verbeek, M. (1999). Estimating and interpreting models with endogenous treatment effects. *Journal of Business & Economic Statistics*, 17(4):473–478.

Von Braun, J. and Kennedy, E. (1994). *Agricultural commercialization, economic development, and nutrition*. Johns Hopkins University Press.

van der Vossen, H. (2005). A critical analysis of the agronomic and economic sustainability of organic coffee production. *Experimental Agriculture*, 41(04):449–473.

van de Walle, D. and Mu, R. (2007). Fungibility and the flypaper effect of project aid: Micro-evidence for Vietnam. *Journal of Development Economics*, 84(2):667–685.

Wang, M. and Stewart, F. (2003). Do PRSPs Empower Poor Countries And Disempower The World Bank, or is it the Other Way Round? QEH Working Papers 108, Queen Elizabeth House, University of Oxford.

Warning, M. and Key, N. (2002). The social performance and distributional consequences of contract farming: An equilibrium analysis of the Arachide de Bouche program in Senegal. *World Development*, 30(2):255–263.

Westerlund, J. (2007). Testing for Error Correction in Panel Data. *Oxford Bulletin of Economics and Statistics*, 69(6):709–748.

White, S.C. (1996). Depoliticising development: the uses and abuses of participation. *Development in practice*, 6(1):6–15.

White, H. (2009a). Some Reflections On Current Debates In Impact Evaluation. 3ie Working papers 2009–1, International Initiative for Impact Evaluation.

—— (2009b). Theory-based impact evaluation: principles and practice. *Journal of Development Effectiveness*, 1(3):271–284.

Whitfield, L. (2005). Trustees of development from conditionality to governance: poverty reduction strategy papers in Ghana. *The Journal of Modern African Studies*, 43(04):641–664.

Whittle, D. and Kuraishi, M. (2008). Competing with Central Planning: Marketplaces for International Aid. In Easterly (2008), chapter 17.

Winters, L.A. (2004). Trade liberalisation and economic performance: an overview. *The Economic Journal*, 114(493):4–21.

Wood, B., Kabell, D., Sagasti, F. and Muwanga, N. (2008). Synthesis Report on the First Phase of the Evaluation of the Implementation of the Paris Declaration. Unpublished report, Copenhagen: Kabell Konsulting ApS.

Woolcock, M. (2009). Toward a plurality of methods in project evaluation: a contextualized approach to understanding impact trajectories and efficacy. *Journal of Development Effectiveness*, 1(1):1–14.

—— (2011). The Importance of Time and Trajectories in Understanding Project Effectiveness. Development Impact blog. URL http://www.samea.org.za/News-52.phtml.

Wooldridge, J.M. (2005). Violating ignorability of treatment by controlling for too many factors. *Econometric Theory*, 21(5):1026–1028.

World Bank (2007). *World Development Report 2008: Agriculture for development*. Washington DC: The World Bank. ISBN 0821368087.

—— (2008). *Annual Review of Development Effectiveness 2008: Shared Global Challenges*. World Bank. URL http://go.worldbank.org/U2T30HQKG0.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *Journal of the American Statistical Association*, 52(298):348–368.