Fiscal Effects of Aid

Emilija Timmis

Thesis submitted to the University of Nottingham for the

Degree of Doctor of Philosophy

March 2015

Abstract

This thesis analyses fiscal effects of aid, first of health aid on health spending for a sample of developing countries and then broadly for Ethiopia and Tanzania. Particular attention is paid to data quality and the severe difficulties in achieving a reliable disaggregation of aid into its on-budget and off-budget components. The first essay assesses the sensitivity of estimated health aid fungibility to how the missing data (often considerable) are treated and explores a novel (at least in economics) method of multiple imputation. The second essay provides a conceptual framework for the disaggregation of (sector) aid into its on-budget and off-budget components. Given that complete binary distinction is not feasible, the aid-spending relationship is explored from a broader fiscal effects angle. This yields new insights on assessing the effect of health aid on health spending.

Contributing to the growing body of evidence based on time-series methods, two essays adopt a case study approach to analyse distinct fiscal dynamics in Tanzania and Ethiopia, invoking detailed understanding of qualitative economic and political context to complement the quantitative data. Both essays employ current Cointegrated Vector Auto-Regressive (CVAR) techniques to distinguish long run equilibrium relationships from short term adjustment mechanisms, test for variable exogeneity and identify which variables adjust to disequilibrium. The fifth and final essay addresses the differences between donor and recipient data records for the two countries, demonstrating that the direction of the discrepancies is not necessarily predicted from the outset and affects the estimated fiscal effects of (and on) aid. These essays contribute to the growing literature using country case studies to assess the fiscal effects of aid.

Acknowledgements

First of all, I would like to thank my supervisors, Prof. Oliver Morrissey and Prof. David Harvey, whose advice and guidance have been invaluable throughout the PhD. In particular, I am forever grateful to Oliver for nudging me away from Masters in Law and supporting me through my application for MSc and PhD Economics.

I am thankful to the School of Economics for the rewarding opportunities to teach. I would especially like to thank Sir Prof. David Greenaway, Prof. Giovanni Facchini, and Robert Hine for trusting me to assist them in their modules, and Prof. Kevin Lee for having faith that I can deliver one myself. Undoubtedly, many words of gratitude go to the excellent School staff – both support and academic. The time spent with my PhD colleagues – whether sharing an office or a pint – and the friendships that resulted from this are particularly treasured.

I have thoroughly enjoyed my experiences away from the School, too. The internship at the Scottish Government was an exciting insight into policy work. The summer school and the subsequent exchange visit to the University of Copenhagen contributed an awful lot to my understanding of econometrics. I am especially grateful to Prof. Henrik Hansen for hosting me at their Department of Economics.

On a personal note, my families and friends, whose unquestionable support I have enjoyed before and throughout this venture, deserve a special mention. And finally, the last thing I expected to acquire during the PhD was a change in my marital status, but it turned out to be one of the most convincing results – for this I am obliged to Jonathan Timmis.

The generous financial support from Economic and Social Research Council (ESRC 1+3 Quota Award for Postgraduate Studies, and funding for Overseas Institutional Visit Scheme and ESRC–Scottish Government Internship Initiative) and the Nottingham School of Economics Graduate Teaching Fellowship are gratefully acknowledged.

Contents

List of Acronyms

General

CI: Co-Integrating, Co-Integration

CPI: Consumer Price Index

CT: Common Trend

CVAR: Cointegrated Vector Auto-Regressive (model)

DAC: Development Assistance Committee (OECD)

EU: European Union

FDI: Foreign Direct Investment

GAVI: Vaccine Alliance

GDP: Gross Domestic Product

GFATM: Global Fund to Fight AIDS, Tuberculosis, and Malaria

GNI: Gross National Income

IMF: International Monetary Fund

NGOs: Non-Governmental Organisations

ODA: Official Development Assistance

OECD: Organisation for Economic Cooperation and Development

TA: Technical Assistance

UN: United Nations

UNICEF: United Nations' Children's Fund

UNAIDS: United Nations Programme on HIV/AIDS

UNFPA: United Nations Population Fund

VECM: Vector Error-Correction Model

WB: World Bank

WDI: World Development Indicators

WHO: World Health Organisation

Chapter-Specific

Chapters 2 and 3

ABBB: Arellano-Bover/Blundel-Bond linear generalised method of moments estimator

DA: Data-Augmentation

DAH: Development Assistance for Health

DAH-G: Development Assistance for Health disbursed through Government

DAH-nG: Development Assistance for Health disbursed through non-Governmental Sectors

DR: Debt Relief

EM: Expectation Maximisation

GDPpc: Gross Domestic Product per capita

GHE-A: Government Health Expenditure as Agent

GHE-S: Government Health Expenditure as Source

GGE: General Government Spending

IP: Investment Projects (Health aid component)

MAR: Missing at Random

MCAR: Missing Completely at Random

MI: Multiple Imputation

NI: Non-Ignorable

ONM: other no mark (Health aid component)

SI: Single Imputation

SP: Sector Programme (Health aid component)

TC: Technical Cooperation (Health aid component)

Chapters 4, 5, and 6

AID: Foreign Aid

AR: Auto-Regressive

BORROW: Public Borrowing (non-concessional)

CAPEXP: Capital Expenditure (Ethiopia)

DEXP: Development Expenditure (Tanzania)

DOMREV: Domestic Revenue

EPRDF: Ethiopian People's Revolutionary Democratic Front

ERP: Economic Recovery Programme (Tanzania)

ESAP: Economic and Social Adjustment Programme (Tanzania)

ESAF: Enhanced Structural Enhancement Facility (Tanzania)

GRANTS: Aid Grants

GTP: Growth and Transformation Plan (Ethiopia)

HIPC: Highly Indebted Poor Countries

H-Q: Hannan-Quinn Information Criterion

IFS: International Financial Statistics (IMF)

JAST: Joint Assistance Strategy for Tanzania

LM: Lagrange Multiplier

LOANS: Foreign Aid Loans

LR: Likelihood Ratio

MA: Moving-Average

MoFED: Ethiopian Ministry of Finance and Economic Development

NTAX: Non-Tax Revenue

PASDEP: Plan for Accelerated and Sustained Development to End Poverty (Ethiopia)

PBS: Protection of Basic Services (Ethiopia)

RECEXP: Recurrent Expenditure (Ethiopia)

REXP: Recurrent Expenditure (Tanzania)

SAPs: Structural Adjustment Programmes

SC: Schwartz Information Criterion

TAX: Tax Revenue

TEXP: Total Expenditure

TShs: Tanzania Shillings

USSR: Union of Soviet Socialist Republics

UVAR: Unrestricted VAR

VAR: Vector Auto-Regressive (model)

List of Tables

List of Figures

List of Tables (Appendices)

List of Figures (Appendices)

Chapter 1

Introduction

The core declared purpose of foreign aid[1] is economic development and extreme poverty alleviation. The empirical evaluation of aid effectiveness is thus centred on the effects of aid on economic growth. Despite a considerable (and growing) body of theoretical and empirical literature, from simple models of the growth process, through to the literature on the direct impact of aid on growth, to sophisticated panel econometric studies, pioneered by the arguments for conditional aid effectiveness (Burnside and Dollar, 2000, subsequently challenged by Hansen and Tarp, 2001, among others), the aid effectiveness literature has not yielded a consensus over which, how, when, where aid works (i.e. delivers positive growth effects) or otherwise. Even the meta-studies arrive at conflicting conclusions (see Doucouliagos and Paldam, 2008, and Mekasha and Tarp, 2013).

The fragile conclusions vary with aid components considered (i.e. eliciting the directly growth-promoting component of aid, investment), growth model, panel dimensions (sample length and country coverage), empirical specification (explanatory and omitted variables, and conditional factors), econometric methodology, accounting for endogeneity of aid or otherwise, data sources, and researchers' priors; see Arndt et al. (2010) for a review of

[1] Aid, or foreign assistance, as defined by the Development Assistance Committee (DAC) of the Organisation for Economic Cooperation and Development (OECD), refers to concessional financing (grants and subsidised loans) delivered in form or financial flows, technical assistance, or commodities designed to promote economic development (see Radelet, 2006, for a comprehensive primer on foreign aid).

recent literature, with a more comprehensive account available in McGillivray et al. (2006). Herzer and Morrissey (2013:724) further argue that in addition to "using growth as the dependent variable but levels as independent variables; and the endogeneity problem of weak instruments", it is the unaccounted cross-country heterogeneity that compromises the robustness of the empirical aid effectiveness (although see Juselius et al., 2011).

A substantial proportion of aid is disbursed through the recipient's government and thus can be expected to affect its fiscal decisions. Given the centrality of the government to economic performance (both in terms of fostering and hindering effects), especially at the early stages of economic development, understanding the fiscal effects of aid is a prerequisite for understanding its broader macroeconomic effects, including aid's effect on growth or on human development (McGillivray and Morrissey, 2000; Morrissey, 2012). Aid should increase recipient's spending and can influence the composition (and potentially evolution) of spending: potentially promoting development spending over recurrent expenditures, or prioritising certain sector components, such as social-capital-enhancing spending on health or education (Gomanee et al., 2005). Aid can increase or decrease tax collection: it provides an alternative, arguably politically cheaper, source of government revenue and thus can provide perverse incentives to relax tax collection efforts or substitute for domestic (borrowing) deficit financing; or it can enlarge the tax base through growth, strengthen the revenue institutions and push for reforms. See McGillivray and Morrissey (2004) for a thematic review of the potential fiscal consequences of foreign aid and Morrissey (2014) for a review of recent empirical applications.

A considerable body of the fiscal effects literature attends to the aid-spending relationship, with particular focus on aid fungibility, i.e. whether aid is spent according to donor's intentions. Empirical findings here, too, give rise to conflicting views. Part of the problem arises from operational definitions: McGillivray and Morrissey (2004) distinguish between general fungibility (diversion of aid intended for public investment into government consumption spending), sector fungibility (whereby aid intended for a specific sector is (intentionally) spent under a different heading), and full additionality (assessing whether the total public spending in the sector increased by the amount of aid). A more substantial problem arises from the (donor) data records: for (earmarked) aid to be linked to a specific component of spending, it has to be disbursed through the recipient's government and be recorded on the budget as aid revenue. Donor aid figures, however, include aid that is not disbursed through the recipient's government (instead delivered through non-governmental organisations or in the form of donor projects), or may not even be spent in the recipient country (funding donor activities such as research, consulting, administration, and the like). Van de Sijpe (2013) demonstrates that unless the on-budget aid component is appropriately disaggregated from its off-budget component, the estimated extent of fungibility will be biased. Given the current state of (donor) aid records, a complete binary distinction between on-budget (potentially fungible) and off-budget (less likely to be fungible) sector aid is not feasible, thus preventing conclusive statements about the degree of sector aid fungibility.

Despite the efforts to standardise fiscal accounting across countries (e.g. the United Nations' System of National Accounts), there are reporting imperfections in recipients' records. Appropriate data collection, storage and processing is a costly process, and may not constitute a top priority or be infeasible for the recipient country agencies (see Jerven, 2013, for discussion). Lu et al. (2010) point at one of the bigger elephants in the room – the pervasiveness of missing data (whereby an observation is not recorded or reported). The way the missing data are handled (or ignored) may jeopardise the researcher's ability to draw valid inferences. Yet, the detail on how the missing data are treated is often itself missing: international organisations (the go-to source of development data) impute observations that are indistinguishable from the truly observed data, without an indication of what assumptions have been taken to impute the missing values, thus with measurement error exacerbating the problem.

Building on the recent (and influential) Lu et al. (2010) study that addresses the fungibility (or, more correctly, additionality) of health aid in low- and middle-income countries, Chapter 2 assesses the sensitivity of the estimated health aid fungibility results to how the missing data are 'treated'. In particular, we explore the novel (at least in economics) methodology of multiple imputation, proposed (but not appropriately applied) in the original study. Multiple imputation is a simulation-based approach used for analysing incomplete data. Multiple distinct datasets are created where the missing data are replaced by plausible values; these are individually analysed using standard econometric techniques; and the results are then combined to generate a single set of estimates, whose standard errors reflect the uncertainty associated with missing data (Little and Rubin, 1987). A "valuable addition to any data analyst's toolkit" (Schafer, 1999:4) as multiple imputation may be, we issue an important warning against imputing the outcome variable – a strategy undertaken by Lu et al. (2010) – and demonstrate that in simple applications it may generate a bias of ambiguous direction. Comparing two widely used data sources for public health spending aggregates, World Health Organisation and International Monetary Fund, we demonstrate that even the data that could be seen to be of best quality (as it is observed across several core data sources) can contain considerable measurement error. The discrepancy across measures of the same variable is on average 0.5 – 0.7 per cent of gross domestic product (GDP) in low- and middle-income countries over 1995-2006. Given that total public health spending averages 2 – 2.5 per cent of GDP, this represents a discrepancy of about 20-30 per cent across sources.

Retaining the focus on health aid and spending, Chapter 3 provides a conceptual framework for the disaggregation of (sector) aid into its on-budget and off-budget components, commenting on the degree of potential fungibility. After reviewing disaggregation strategies of health aid data available in the literature (namely, Lu et al., 2010, and Van de Sijpe, 2013), keeping identical modelling and estimation methods we demonstrate that the currently available estimates of the relationship between health aid and spending are not as conflicting as currently stated, despite the stark differences in disaggregated (and aggregate) sector aid data. However, if an inevitably *ad hoc* binary disaggregation of data is imposed in an attempt to produce fungibility estimates, the results can yield conflicting conclusions and policy implications. We show that this largely depends on whether donor projects, estimated to have the most robust positive association with recipient's total overall commitment to health spending, are attributed to aid's on- or off-budget component.

The most important limitation of aid fungibility studies is that they ignore broader fiscal effects of aid and the dynamics of fiscal relationships, especially on tax revenue and public (non-concessional) borrowing. In the past, the research on fiscal dynamics relied on the seminal Heller (1975) framework for fiscal response models, based on maximisation of government's utility function represented by deviations of actual fiscal aggregates from target levels. Criticisms (see Binh and McGillivray, 1993, among others) to this framework include both theoretical and empirical issues, such as equal treatment of overshooting and undershooting the government targets, and unavailability of the actual data on these government fiscal targets; McGillivray and Morrissey (2004) illustrate the sensitivity of results to the data.

With the passage of time (it has been fifty years since the first United Nations Conference on Trade and Development that launched "a debate about how much money rich countries should give to poor ones to reduce poverty and bolster growth", *The Economist*), the increasing availability of data has allowed for applications of time series methods.

Recognising the heterogeneity of fiscal mechanisms, a growing body of country case studies, stemming from Osei et al. (2005), employ the Cointegrated Vector Auto-Regressive (CVAR, Juselius, 2006) methods to analyse the fiscal (and growth) effects of aid. Although demanding of the data, CVAR is particularly useful in this context as it allows distinguishing long run equilibrium relationships from the short term adjustment mechanisms, and testing for variable exogeneity rather than imposing it from the outset. We employ the CVAR framework to explore the fiscal dynamics in two East African countries: Ethiopia[2] (Chapter 4) and Tanzania (Chapter 5). The advantage over existing similar CVAR applications stems from significantly longer time-series dimension (47 versus 26-39 yearly observations) and the fact that data are obtained from a single domestic source, thus reducing the extent of measurement error (through conversions and corrections) and, crucially, capturing the recipient's measure of aid – what is effectively disbursed through the budget and the government is aware of.

The atheoretical nature of the CVAR 'allows data to speak freely' to discriminate between competing hypotheses and theories, and does not impose *a priori* assumptions and restrictions, such as residual normality or variable exogeneity, instead allowing to test for these in the dynamic multiple equation setting. However, since the estimation of simultaneous long and short run equations involves a large number of parameters, the CVAR requires large samples. As 47 yearly observations do not constitute a statistically large sample, the CVAR analysis is complemented by a detailed qualitative understanding of the country-specific economic and political context, which ensures sound model specification and sensible interpretation of estimated results.

Modelling several variants of a model between total (central) government expenditure (possibly disaggregated into capital/development and recurrent components), domestic revenue (possibly disaggregated into tax and non-tax), aid (grants and loans), and, in the Tanzanian case, domestic public borrowing, we fail to identify any perverse or adverse effects of aid, while still illustrating differences in fiscal adjustment mechanisms between countries. We fail to identify any negative effect of aid on tax revenue. In Ethiopia, tax revenue is positively associated with both modalities of aid (grants and loans), with only grant aid adjusting to departures from long-run equilibrium (possibly reflecting donors rewarding reforms). In Tanzania (although the results are generally more fragile), too, only a positive association can be identified between aid and tax, again with only aid exhibiting any

[2] Chapter 4 on Ethiopia is a result of collaboration with Giulia Mascagni, at the time PhD Candidate in Economics at the University of Sussex, who provided the data.

adjusting behaviour (although aid and borrowing could be seen as substitutes). So while, rather plausibly, on-budget aid does not drive domestic revenue collection, it does not discourage (or substitute for) the collection of tax revenue. In terms of public spending, the aggregate expenditure is positively associated with aggregate aid flows, with aid adjusting to departures from equilibrium (financing the excess spending/deficit). Finally, disaggregation of spending components gives further insight into what aid (and its modalities) is funding: in both countries capital expenditure is strongly positively associated with aid, although the adjustment mechanism differs. In Ethiopia, aid (and particularly grants) adjusts to capital expenditures, whilst in Tanzania it is mainly the development expenditures that adjust to (shortfalls or windfalls) in aid.

With this thesis arguing strongly for the importance of disaggregation of aid into its on- and off-budget components and sensitivity of the estimated results to data and its sources, the final essay, Chapter 6, conducts an exercise assessing the sensitivity of the estimated fiscal effects of aid to its data source, comparing the standard (DAC) donor to recipient's aid data. In relation to the early chapters of the thesis, the DAC data represents a mixture of on- and off-budget aid disbursements, while the recipient's measure consists entirely of the on-budget aid flows only. However, eliciting the off-budget component is not possible due to the presence of non-traditional (i.e. non-DAC) donor disbursements in the recipient's data. It is shown that recipient and donor data differ, and not necessarily in a direction predictable from the outset: while for Tanzania the total DAC aid measure exceeds recipient's own records by more than three times (signalling a large proportion of DAC grants delivered through donor projects, or spent in donor countries), in Ethiopia the total amount of disbursed aid is higher in recipient's records (indicating the importance of non-traditional donor funds). Depending on which source is relied upon for aid data, the results can differ substantially. The comparison of the simple CVAR estimates of models with recipients versus DAC donor data reveal that the two aid measures do not even co-vary sufficiently to yield qualitatively consistent estimates. The estimated (long run) coefficients of aid can contrast in terms of sign (as in the Tanzanian case) or reflect different adjustment behaviour (the Ethiopian case).

The conclusion to this thesis (Chapter 7) provides a brief overview of the findings and the core lessons drawn from them, as well as their limitations, together with a brief indication of potential avenues of future research. Appendices A, B, C, D, and E (corresponding to

Chapters 2 through 6) contain the complementary materials that are referenced in the respective chapters and can be found at the end of the thesis.

Chapter 2

Data Imputation vs. Data Amputation:

An Application of Multiple Imputation

in the Context of Aid and Health Spending

1. Introduction

Aid donors tend to be interested whether the official development assistance funds are spent according to (donors') intentions. Since Heller's (1975) paper, fungibility literature[3] attempts to estimate whether aid earmarked for specific activity (such as health) is fully additional to recipient's planned spending or is it diverted elsewhere[4]. The study by Lu et al. (2010) has received substantial attention since its publication in *The Lancet*, both in economics (aid) and medical contexts, media and broader philanthropic community. Its key contribution to the aid fungibility[5] debate is the finding that each health aid dollar disbursed through the recipient government reduces government's domestically funded public

[3] The key references include Pack and Pack (1990, 1993, 1999), Feyzioglu et al. (1998), World Bank (1998), Franco-Rodriguez et al. (1998), Devarajan et al. (1999), McGillivray and Morrissey (2000, 2001, 2004).

[4] Following McGillivray and Morrissey (2000), aid's effect on recipient's revenue and borrowing is termed 'fiscal response', a broader concept than fungibility.

[5] Lu et al. (2010:1376) define fungibility as "aid substitut[ing] for domestic government spending", and thus in effect are concerned with additionality of aid funds rather than whether the aid funds are allocated to sectors intended by donors (see McGillivray and Morrissey, 2000, 2001). This is discussed in depth in Chapter 3.

expenditures on health in low and medium income countries by USD$ 0.46 in the short run and by $1.14 in the long run, therefore concluding that health aid is largely fungible.

The renewal of interest in aid fungibility has been partly based on increased availability of data (Van de Sijpe, 2013a). The quality of data presents a considerable challenge in development economics research. Together with inappropriate choice of econometric techniques and economic modelling, it may compromise the empirical conclusions. Lu et al. (2010) point at one of the bigger elephants in the room – the pervasiveness of missing data.[6] The way the missing data are handled (or ignored) may jeopardise the researcher's ability to draw valid inferences. Yet, the detail on how the missing data are treated is often itself missing: international databases impute observations that are indistinguishable from the truly observed data, without an indication of what assumptions have been taken to impute the missing values, so measurement error exacerbates the problem.

Building on Lu et al. (2010), this chapter explores the sensitivity of the estimated health aid fungibility results to how the missing data are 'treated'. In particular, we explore the novel (at least in economics) methodology of multiple imputation, proposed (but not appropriately applied) by the original study. The chapter is structured as follows: Section 2 discusses the data, with Section 3 discussing the missing data. Section 4 introduces the available tools to 'treat' the missing data and issues an important warning against imputing the outcome variable – a strategy undertaken by to Lu et al. (2010). The results are discussed in Section 5. Section 6 further assesses the appropriateness of imputing the dependent variable. Section 7 explores whether the supposedly better quality data are indeed any better. Section 8 concludes. Additional information is provided in Appendix A.

2. Data

For the purposes of this chapter and to provide directly comparable estimates, the original dataset of Lu et al. (2010) is used. The dataset contains 111[7] countries over the period 1995-2006. Following the original paper's variable definitions, we elaborate on the appropriateness of their construction and other relevant comments.

[6] It must be born in mind that appropriate data collection, storing and processing is a costly process, and such process, requiring compliance with multitude of certain statistical standards or consistency over time may not constitute a top priority or be altogether infeasible for the recipient country agencies (see Jerven, 2013, for discussion).

[7] Initial sample contains 113 countries; however, Lu et al. (2010) omit Angola and Eritrea for the analyses. For the purposes of comparison, we will follow their sample size of 111 countries.

Government health expenditures as agent (GHE-A) represent the total spending on health from government budgets, funded both domestically (e.g. through tax receipts) and externally (e.g. health aid disbursed through government agencies). Two organisations provide datasets containing information on government health expenditures as agent: World Health Organisation (WHO)[8] and International Monetary Fund (IMF)[9] (two separate datasets)[10]. The variable is expressed as percentage of GDP for country *i* in year *t*; the observations obtained from the WHO (IMF) are scaled by the GDP data obtained from WHO (IMF, respectively).

The Development Assistance for Health (DAH) is exceedingly difficult to trace to a recipient country or agency. Ravishankar et al. (2009) attempt this Sisyphean task. In the original study, Ravishankar et al. (2009:2114) define the DAH as "all flows for health from public and private institutions whose primary purpose is to provide development assistance to low-income and middle-income countries". More technically, it is the "sum of gross yearly disbursements on all health-sector grants and loans, and health-related programme expenditures from the channels of assistance, net of any transfers to other channels of assistance included in the study" (Ravishankar et al., 2009:2114). Health aid is defined as "all disease-specific support and general health-sector support, and excluded support for allied sectors such as water and sanitation, education, general budget support, and humanitarian assistance", and the measure includes "research funded by DAH channels of assistance", but excludes "research by institutions that do not meet our definition of a channel of assistance" (Ravishankar et al., 2009:2114). This DAH measure includes the OECD DAC data, "contributions from the Global Alliance for Vaccines and Immunization (GAVI) and core-funded activities of WHO", as well some private philanthropy flows (which are, however, limited to certain US donors) and (only) US-based NGOs' contributions. Aid flows from non-traditional (i.e. non-OECD) donors are not included. Further challenges are presented by blurred distinction between 'health' and 'health-related' aid disbursements; commitments, or disbursements[11]; as well as double counting, timing of fiscal years across countries and

[8] WHO. National Health Accounts. Geneva: World Health Organization, 2009. http://www.who.int/nha/country/en/ (accessed Dec 17, 2009).
[9] IMF Fiscal Affairs Department. Total Health Spending Database. Washington, DC: International Monetary Fund, 2009.
[10] The analyses on these two distinct datasets for the GHE-A variable are carried out separately and thus two sets of results are provided in the Chapter.
[11] Ravishankar et al. (2009:2113).

organisations, and identifying which institutions count as channels of assistance.[12] Thus it includes aid spent (delivered) in both recipient and donor countries, through both governmental and non-governmental organisations.[13]

Ravishankar et al. (2009) conducted a substantial amount of data "corrections", estimation, projection, and imputation.[14] The authors estimated the 'under-reported disbursement information to [OECD] CRS before 2002', as well as some other missing data, by predicting/projecting disbursements from the observed data, using basic yet untestable assumptions.[15] Furthermore, their DAH measure includes an estimate[16] of technical assistance and programme support component; this in-kind[17] component reportedly ranges from 9.2 to 13.7 per cent of the financial transfers for the study period of 1990 – 2007 (and from 9.2 to 11.4 per cent[18] in our sample period 1995-2006). Technical assistance and funds spent mostly in donor countries (i.e. programme support or research expenses, the exact relative magnitude of which is not reported by the authors) do represent aid, but they do not represent cash flows from donors to recipients (see GHE-S paragraph below)[19]. Table 2.1 below reports the total DAH estimates in 2007 US dollars by Ravishankar et al. (2009) for the period of 1995-2006, and illustrates the substantial increases in health aid.

Table 2.1: Total Development Assistance for Health Estimates 1995-2006 (Billions 2007 USD)

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
DAH*	8.0	8.1	8.4	8.7	9.8	10.7	10.9	12.4	13.6	15.6	17.9	19.0

**Total Estimated DAH (billions 2007 USD), Ravishankar et al. (2009:2116).*

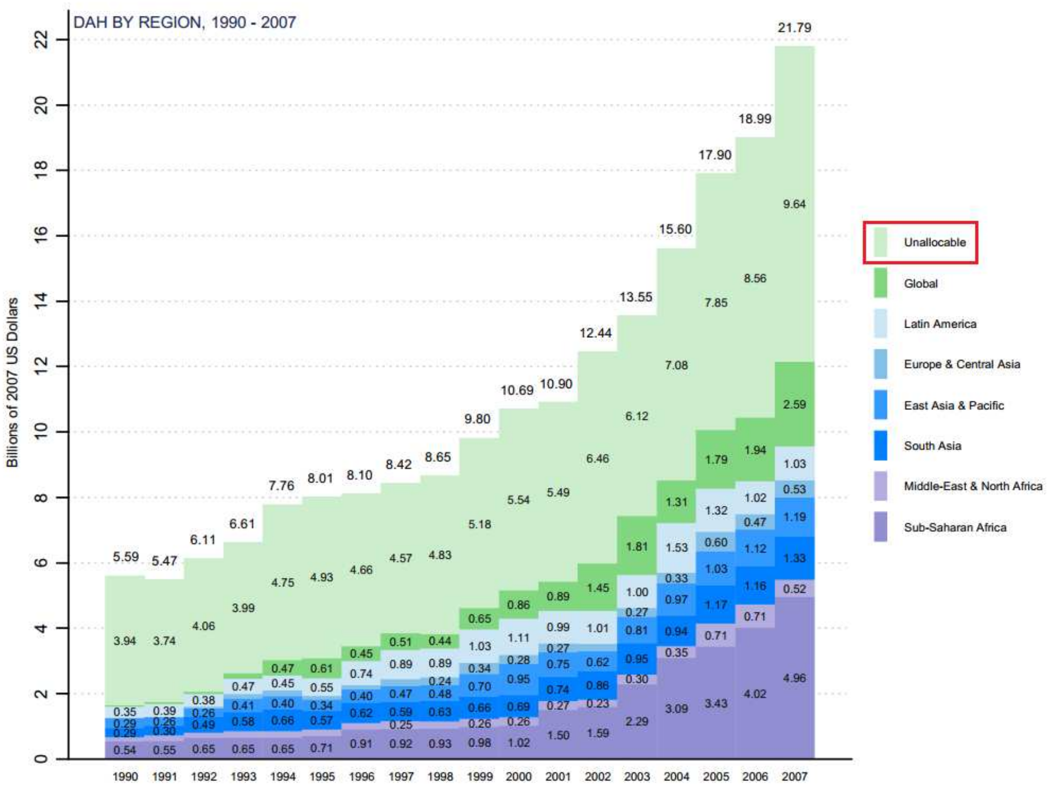
Defining health aid, accounting donor flows, dealing with missing data are not the only challenges facing the aid data. Figure 2.1, extracted from Ravishankar et al. (2009) webappendix[20] highlights the recipient-wise traceability issues with (total) DAH measure.

---

[12] And, of course, there are significant differences depending on whether the values in domestic currencies are first converted into dollars and then deflated, or vice versa (authors pick the former); see Ravishankar et al. (2009, Webappendix, p. 19).
[13] See also Van de Sijpe (2013a) and Dieleman et al. (2013) for further debate over this issue.
[14] And thus the DAH variable, too, suffers from missing data problem.
[15] See Ravishankar et al. (2009, Webappendix, pp. 18-24).
[16] See Ravishankar et al. (2009, Webappendix, pp. 53-55).
[17] In-kind component also includes commodities, such as drugs and medical supplies.
[18] ET's own calculations based on Ravishankar et al. (2009) data.
[19] This issue is particularly discussed in Van de Sijpe (2013:4). This will overstate the extent of estimated fungbility of aid.
[20] Available online at http://www.sciencedirect.com/science/article/pii/S0140673609608813#appd001

The geographically "unallocable" proportion of DAH ranges from 43.9 to 61.5 per cent of total estimated yearly DAH.

Figure 2.1: Traceability of estimated DAH flows to the Recipient Region



Source: Ravishankar et al. (2009, Webappendix p. 4). ***Original caption****: "Webfigure 4: DAH from 1990 to 2007 by focus region. All quantities are in real 2007 US$. DAH Funds for which we have no recipient country or region information are coded as "unallocable"."*

The key focus of Lu et al. (2010) study is to estimate the extent of health aid fungibility depending on whether the DAH is disbursed to recipient government sectors, or through non-governmental bodies. For this purpose, they use a final variation in DAH tracing discussed in Ravishankar et al. (2009): the decomposition of DAH by channel of assistance. This decomposition traces the DAH through: recipient governments; multilateral organisations (European Commission, UNICEF, UNAIDS, UNFPA, WHO); global health partnerships (GAVI and GFATM); International Development Association; NGOs, private-public partnerships, and other channels (excluding GAVI and GFATM); the residual is coded as 'unspecified'. The associated shortcomings of this further decomposition are discussed in the following paragraphs.

Development Assistance for Health disbursed through Government (DAH-G) is constructed[21] by Lu et al. (2010:1378) as the funds identified to be disbursed to the recipient government *plus* "disbursements that lacked any information about the channel of delivery". Figure 2.2 below illustrates why this may be a problem: as much as over a half of the (presumably geographically traceable) total DAH has an 'unspecified' channel of delivery. Treating these flows as going through the recipient government is likely to overestimate DAH-G and influence[22] the estimated results (see GHE-S paragraph).

Figure 2.2: DAH Composition by the Channel of Delivery

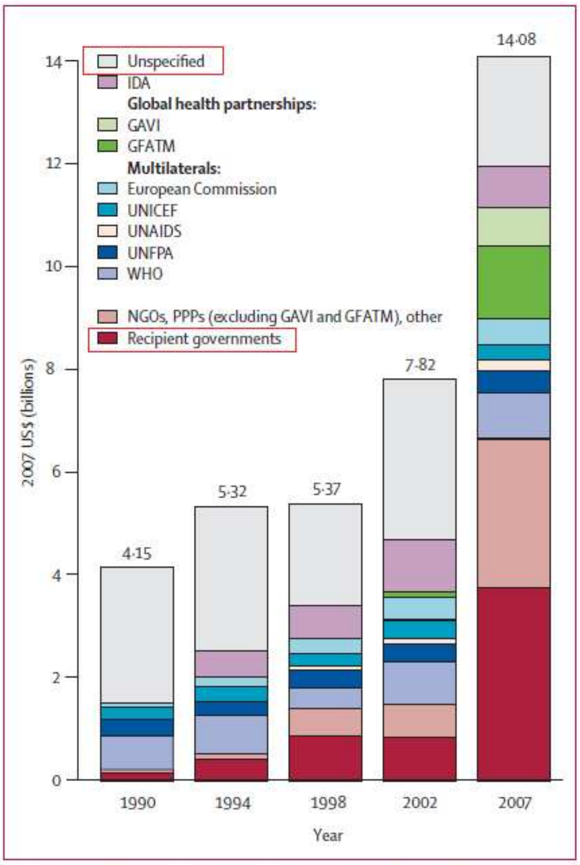


Figure 3: Publicly-financed development assistance for health (DAH) in 1990, 1994, 1998, 2002, and 2007 separated by channel
Non-governmental organisations (NGOs), public-private partnerships (PPPs) excluding the Global Alliance for Vaccines and Immunization (GAVI) and the Global Fund to Fight AIDS, Tuberculosis and Malaria (GFATM), and other miscellaneous channels are shown together. Disbursements from the creditor reporting system when the channel of delivery code was not specified are shown here as unspecified. IDA=International Development Association. UNICEF=UN Children's Fund. UNFPA=UN Population Fund. UNAIDS=Joint UN Programme on HIV/AIDS.

Source: Ravishankar et al. (2009:2117).

[21] Lu et al. (2010) exclude DAH provided in form of loans.
[22] Lu et al. (2010: 1378) reportedly test the sensitivity of their findings to this assumption, and claim them to be robust to final conclusions (see Lu et al., 2010, Webappendix p.6), although it is not clear whether they redefine/re-estimate GHE-S accordingly.

Development Assistance for Health disbursed through non-Governmental sectors (DAH-nG) represents health aid disbursed through non-Governmental organisations.

To elicit *domestically* funded public health expenditures, Lu et al. construct variable ***Government Spending as Source (GHE-S)***[23]. It is estimated by subtracting the health aid disbursed through the government (DAH-G) from government's total health spending *as agent* (GHE-A). Some key problems arise here in addition to the problems with constructing the components (especially potentially overestimated DAH-G), primarily because this way of constructing the (dependent) variable assumes that aid is (or is required or intended to be) fully spent in the fiscal year it was received. If this is not the case, aid will be estimated as fungible by construction.

Aid may not be spent in the year it was received for multiple reasons. Firstly, in cases where aid flows are volatile[24], it may be in the recipient government's rational interest *not* to spend the received (in cash) health aid in its entirety in the fiscal year it was received. In the presence of such aid 'smoothing', the assumption that aid is fully spent in the year it has been received inherent in the definition of GHE-S may lead to estimated coefficients associated with conclusion that it is fungible *by construction*.[25] Secondly, in the same issue of *The Lancet*, Ooms et al. (2010:1403) hypothesise further 'explicit policy choices' explaining why aid may have a crowding out effect, namely that "governments anticipate long-term unreliability of international health aid by stalling possible increases of recurrent health expenditure". Such behaviour would be consistent with IMF encouraging developing countries to build up financial reserves as a buffer against future adverse shocks, including aid shortfalls. In a way withholding (or smoothing) spending of the received health aid could *enliken*[26] rather than endanger the sustainability of health spending. Finally, if DAH-G is potentially overestimated (i.e. if at least a fraction of 'DAH-unspecified' assumed to go through the recipient government in fact does *not* flow through the recipient government, or a fraction of the aid is not in cash), the domestically funded component of government health spending (GHE-S) and therefore the estimated government's commitment to health will be underestimated by construction (reinforcing the downward bias from the assumption

[23] Lu et al. (2010) acknowledge that due to lack of reporting standards, it is not always certain whether GHE-A or GHE-S is reported in the original National Accounts.

[24] And/or not fully predictable.

[25] Altogether, it may severely disrupt recipient reporting as resources would be transferred across years, and it may not be clear whether spending is domestically or externally funded.

[26] Assuming the health aid saved is later indeed spent on health.

that all aid is spent in the year it is received). GHE-S is the dependent variable in Lu et al. (2010).

It is not difficult to agree with Lu et al. (2010:1376) argument that "[e]nhancement of public financing of health is important for the long-term financial sustainability of the health sector" and that if (or when) "the donor funding declines or stops, continuation of aid-funded health programmes would be difficult without the financial support of the domestic government" (unless, as discussed above, some of the aid is *consciously* saved up indeed to ensure such continuation). However, authors provide no clear discussion over what incentive structure is in place for the recipient government to adopt such 'long-term' view rather than manage the funds available to them given the short term urges; in practice, they may have little incentive not to *fungor*[27] - especially when donors do not provide any clear depiction of (health) aid exit strategy. As Ooms et al. (2010:1403) note, DAH is relatively more generous[28] than other aid, therefore recipient governments may "compensate for exceptional international generosity to the health sector by reallocating government funding to other sectors". This is, in fact, somewhat consistent with consistent with Paris Declaration on Aid Effectiveness' intention to align aid with developing countries' own priorities[29], when the aid is not distributed uniformly across sectors (see also McGillivray and Morrissey, 2000, 2001 on welfare-optimising response from a rational recipient government).

General Government Spending (GGE) captures the government size (data from the World Bank).[30] It is worth noting that in Lu et al. this measure includes the GHE-A.

Debt relief disbursement (DR), constructed from OECD CRS Action Relating to Debt database, is included in the model, potentially to control for the effects of the PRSP[31]-type

[27] *Fungor (latin)* –perform, enjoy.

[28] To put this in numbers, in addition to **Table 2.1** above, "health aid has grown relative to other foreign aid: development assistance for health (DAH) increased by 251 per cent between 1995 and 2010 (Institute for Health Metrics and Evaluation, 2013), while official development assistance (ODA, which includes health aid) increased by only 43 per cent (The World Bank, 2012)." (Dieleman et al., 2013:1755).

[29] Dieleman et al. (2013: 1755) point out that "recipient governments prioritise health in much the same manner as they always have. In 1995, the average low- and middle-income country spent 9.3 per cent of total general government expenditure (GGE, which includes some ODA) on health. In 2010, the share of GGE on health increased only to 9.4 per cent. In an environment characterised by an increasing share of aid for health but relatively constant government prioritisation, it is plausible that rational governments reallocate funds and displace domestic government health expenditure. This renders health aid fungible". Nevertheless, it must be noted that these statistics indicated that *total* amount spend on aid increased at least in line with total government spending.

[30] World Development Indicators Online Database. Washington, DC: The World Bank, 2009. http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/0,,contentMDK:20398986~menuPK:64133163~pagePK:64133150~piPK:64133175~theSitePK:239419,00.html (accessed Sep 11, 2009).

programmes, which are expected to route the released funds into social sector spending (including health). It excludes capitalised interest and is assumed to be evenly distributed (in dollar, rather than relative, terms) over 10 years.

As increments in the standard of living have been empirically established to be associated with increased health expenditures, ***Gross Domestic Product per capita (GDPpc)*** (in USD 2006 values) is also controlled for, constructed using IMF World Economic Outlook and UN Population Data[32]. It is noteworthy that Lu et al. (2010) use the non-logged version of this variable.

HIV prevalence rate (HIV), from UNAIDS database[33], is used to control for the spread of infectious diseases.[34] The positive correlation between aid and burden of disease is not necessarily informative about the direction of causality. Over[35] provides a brilliant summary: "If we believe that health assistance should reduce disease burden, then we might expect this correlation [between the total disease burden in a country and the total amount of health assistance it receives] to be negative. If we believe that health assistance should be directed to places with higher burdens, then the correlation should be positive. If we think the main criterion for allocating foreign assistance should be the cost-effectiveness of spending opportunities, not the size of the burden, and that many other factors influence health beside health assistance from abroad, then we should expect that the correlation will be weak. If we believe all of these things at once, as most of us do, then we have no prior belief whatsoever about the correlation between burden and assistance and will find it to be uninteresting".

All variables, except HIV prevalence and GDP per capita, are expressed as a percentage of GDP[36] for country *i* in time *t*. A general comment may be worth noting here. It is rather doubtful that government officials plan revenues and expenditures in terms of percentages

---

[31] Highly Indebted Poor Countries Poverty Reduction Strategy Paper.

[32] IMF. World Economic Outlook Database. Washington DC: International Monetary Fund, 2009. http://imf.org/external/data.htm (accessed Feb 1, 2009);
United Nations Department of Economics and Social Affairs, Population Division. World Population Prospects: The 2006 Revision. 2007. CD-ROM Edition-Extended Dataset in Excel and ASCII formats (United Nations publication, Sales No. E.07.XIII.7)

[33] UNAIDS, WHO. HIV data. Geneva: UNAIDS, 2009. http://www.unaids.org/en/KnowledgeCentre/HIVData/ (accessed Jan 20, 2010).

[34] HIV is correlated with GHE-A just about as much as with DAH-G.

[35] Over (2009, Blog Entry): http://www.cgdev.org/blog/%E2%80%9Cwho-what-where-when-how-and-how-much%E2%80%9D-international-health-aid-%E2%80%93-not-%E2%80%9Cwhy%E2%80%9D

[36] GDP figures from the World Bank are used for variables other that GHE-A/GHE-S. It is not clear why would the authors depart from the IMF/WHO statistics. We will stick to Lu et al.'s method.

of (hardly predictable or properly calculated) future GDP figures (and incorporating forecasted GDP growth rates) rather than in level terms, in domestic currency. [37] Clearly, expressing variables as percentage as GDP allow for international comparisons, but that inevitably introduce extra variation that is not at all in direct control of the recipient government.[38]

Table 2.2 below provides summary statistics for the explanatory variables (summary statistics for the dependent variable, GHE-A/GHE-S are provided in the Results section due to pervasive missing values). Appendix Figure A1 depicts the distributions of each of the variables, illustrating that these are skewed. The correlation coefficients among the independent variables are provided in Table 2.3.[39] The highest correlation coefficient in the sample is observed between the aid variables (firstly, between DAH-G and DAH-nG, but also with debt relief). This may (but does not necessarily) reflect common trends in aid funding (i.e. if donors increase their disbursements, they do so across delivery channels). Strong negative correlation between health aid disbursements (and debt relief) and GDP per capita may reflect decreasing aid funding with increments in standards of living. Disease burden, proxied by HIV prevalence, seems to bear positive (albeit not large) correlation with DAH-G, possibly favouring the argument that donors allocate more health aid to countries with higher identified disease burdens, or that health aid is at least partially contributing to disease diagnostics and recording.

Table 2.2: Summary Statistics (Explanatory Variables)

	DAH-G/GDP	DAH-nG/GDP	DR/GDP	GDPpc	GGE/GDP	HIV
Mean	.0027	.0007	.0048	2285	.1445	.0275
Std.d.	.0045	.0019	.0139	2825	.0635	.0516
Min	-.000	0	0	89	.0255	0
Max	.0386	.0198	.1517	21414	.5160	.2892
N	111	111	111	111	111	111
YO	1332	1332	1332	1332	1332	1332

[37] Unless knowingly assessed on exactly such measures by some institution that can pose any credible sanctions if some specific targets are not lived up to (e.g. donor holding-off next aid disbursement until some pre-set condition is fulfilled).

[38] Even the total government spending is not fully predictable at the time individual spending components are planned. Nevertheless, expressing key variables of interest (namely, GHE-A, GHE-S, DAH-G, and DAH-nG) as percentage of total government spending may more appropriately reflect the intended (or acknowledged) variation in and prioritisation of public health spending compared to other public spending components, still allowing for international comparisons. However, the possibly unrelated shocks to government spending may introduce variation just as bad as from GDP, just in other direction.

[39] A word of caution is worth mentioning: if the variables follow I(1) processes (and in such short sample variables expressed as proportion of GDP are likely to be), the correlation is spurious.

Table 2.3: Correlations across Explanatory Variables

	DAH-G /GDP	DAH-nG /GDP	DR/GDP	GDPpc	GGE /GDP	HIV
DAH-G/GDP	1					
DAH-nG/GDP	0.4859	1				
DR/GDP	0.3039	0.2546	1			
GDPpc	-0.3453	-0.1927	-0.2038	1		
GGE/GDP	-0.0004	0.0322	-0.0610	0.1359	1	
HIV	0.2323	0.1284	0.0385	-0.0809	0.2267	1

Note: Table reports correlation coefficients.

3. Missing Data

Lu et al. (2010:1376) raise an important issue of missing data in developing country records of government total health spending (GHE-A):

> "<…> WHO is forced to estimate missing data for a substantial proportion of the country years [of GHE-A variable]. The actual data and estimates are not always distinguished in the published tables, and detailed information about imputation methods and components is not available to the public. <…> WHO's imputation methods are not standardised, and <…>[d]etailed information about the components used to generate imputations was not provided by WHO".[40]

This overstates the precision of the data.

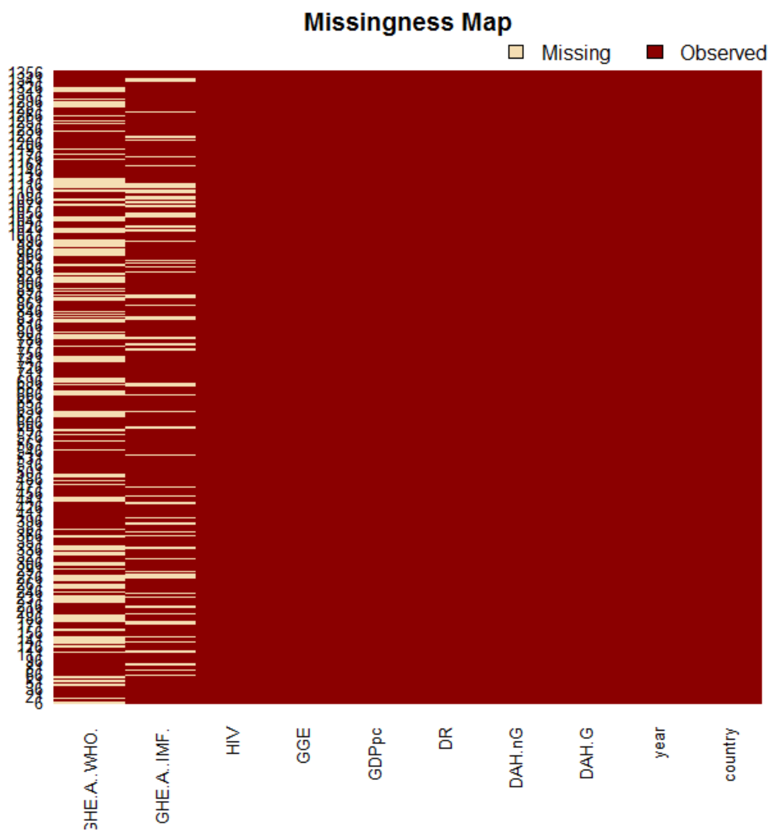
Unfortunately, after introducing the problem of missing data and proposing solving it using multiple imputation techniques, the authors average out the 100 imputed datasets and estimate the papers' results from what essentially is a singly imputed dataset. Although in this case the methods of 'recovering' the missing data are known (i.e. stated in the paper), this reintroduces the issue of overstating the certainty of observation.

With WHO's help in identifying which values were imputed and which were observed in the country financial reports, Lu et al. (2010) use a simple (though *ad hoc*) rule to code the government health expenditures (as agent) as missing or not. For a country year for which more than 10% information underlying the construction of the GHE-A observation is missing, the authors coded the observation as missing. For 1995-2006, this implies 33% of GHE-A observations missing in the WHO sample (442 out of 1332 country-year observations), and about 19% of GHE-A observations missing in the IMF sample (249/1332). The correlation

[40] "And imputations were often based on the assumption that the ratio of government health spending to general government spending was constant with time" (Lu et al., 2010:1377).

between the GHE-A samples reported by WHO and IMF is only 70.8 per cent, with only 84 observations coded as missing simultaneously in both samples. The remaining variables are implicitly *assumed* to be fully observed. This is reflected in the missingness map below, where red colour indicates an observed variable, whilst the observations coded as missing are coloured in yellow (Figure 2.3). Next section discusses analytical tools available to handle missing data statistically.[41]

Figure 2.3: The Missingness Map



4. Tools to Tackle Missing Data

Regarding the issue of missing data (i.e. when one or more variable is not recorded or observed for a particular observation), three types of concern are generally considered:

> "i) loss of information, efficiency or power due to loss of data;
>
> ii) complication in data handling, computation and analysis due to irregularities in the data patterns and non-applicability of standard software; and, most fundamentally,

[41] CSV file ImputationChapter113Amelia.csv

iii) very serious potential bias due to systematic differences between the observed data and the unobserved data" (Barnard and Men, 1999:17)[42].

The problem of missing data has been widely recognised – and statistical developments applied – in the context of medical statistics, as non-response in health surveys due to attrition or other problems may cause serious biases and inefficiency of coefficients in longitudinal analysis if not accounted for correctly. [43] Meanwhile, the majority of empirical studies within social sciences assume the absence of missing data (Honaker and King, 2010:3), or ignore the problem altogether, and analyse the data for which all variable values are observed.

Reliance on 'default' methods such as complete-case analysis (listwise deletion), where the observation is discarded if values are not observed for one or more variables for that datum, has been criticised on several grounds. Firstly, data are costly to collect; therefore discarding observations for which at least some information is available is regarded as wasteful. Secondly, in cases where a large fraction of data is missing, discarding this data may reduce statistical efficiency through sample reduction[44]. Thirdly, depending on the reasons why data are missing, the practice of listwise deletion may invoke severe biases.[45] Literature (see, for instance, Graham et al., 1994:15) also discusses missing data situations of omission, attrition, or planned missing data, among other topics related to *coarsened*[46] data; this, however, is more related to the survey data, and is of little direct relevance here.

Several alternative approaches to treating missing data have been suggested during last three decades, including single (SI) and multiple imputation (MI) techniques. The material below thus first provides a summary of the missingness mechanism, and how that may

[42] "Unfortunately, it is also the most difficult to handle because, typically, the reasons for not observing the full data (i.e. the so-called missing-data mechanism) are often at best partially understood (except for cases where missing data are induced by the design or latent-variable modelling)" (Barnard and Meng, 1999:18)

[43] "In fact, the problem is so universal that an unusually high response rate (e.g.95%) should make the investigator worry about possible design flaws in the survey, such as selection bias in the sample or a substantial amount of untrustworthy responses induced by too much monetary incentive" (Barnard and Meng, 1999:18)

[44] See Section 6.

[45] Note that the focus here is on missing data that does exist 'in a specific metaphysical sense' but is unobserved (e.g. government spending on health does actually *exist* in a given country, even if it is close to zero, provided there exists a country with an established government; whether that data are recorded or revealed by the government, may define whether that data is observed, or *missing*).

[46] Schafer and Graham (2002: 148) note that "Missing values are part of the more general concept of coarsened data, which includes numbers that have been grouped, aggregated, rounded, censored, or truncated, resulting in partial loss of information (Heitjan and Rubin, 1991)".

influence the choice of treatment applied to the missing data; several popular methods for treating missing data problem are briefly summarised; and the MI framework is introduced.

4.1 The Missingness Mechanism

"Knowledge, or the absence of knowledge, of the mechanisms that led to certain values being missing is a key element in choosing an appropriate analysis and in interpreting the results" (Little and Rubin, 1987:8). Three types of the underlying missingness mechanisms are distinguished in statistical literature (see Little and Rubin, 1987:8, Schaffer, 1997:11): Missing Completely at Random (MCAR), Missing at Random (MAR), and Non-Ignorable (NI). In the first (MCAR) case, the pattern of data missingness cannot be predicted by either the values of dependent or independent variable: missing values are a simple random sample of all values.[47] If the data are missing at random (MAR), the probability of missingness depends on the observed data, but not on the unobserved data.[48] The missingness process is non-ignorable when the probability that a cell is missing depends on the unobserved value of the missing response. For a more extended and technical discussion see Appendix Table A1.

The underlying missingness mechanism (or the respective assumption) may determine the validity of the methods employed in the empirical analysis. For instance, the default option of listwise deletion may bias the results unless the MCAR holds, whilst the inferences from analyses using MI are not biased under MCAR or MAR. All econometric estimates may be biased under NI. The methods applied in this chapter assume that data are MAR. Usually, MCAR can be rejected in favour of MAR. Unfortunately, "it is not possible to relax the MAR assumption in any meaningful way without replacing it with other equally untestable assumptions. <…> In the vast majority of studies, principled methods that assume MAR will tend to perform better than *ad hoc* procedures such as listwise deletion or imputation of means." (Schafer and Olsen, 1998:553).[49] Similarly, the presence or absence of NI can never be demonstrated using only the observed data. Thus, in most circumstances it is possible to verify (in statistical simulations) whether multiple imputation will outperform (or rather will

---

[47] An example of this could be if a proportion of observations are randomly deleted; or hard copies are kept in cellar in which some parts are exposed to intense humidity which over a period of time destroys a proportion of files arranged in disorderly fashion.

[48] For instance, war. In our sample, Rwanda exhibits missing data for several years after the 1994 genocide. Increments in government (military) spending (if recorded), or the very fact that other data are missing (i.e. the observed missingness matrix) could be used to predict the missingness of the data in other variables).

[49] Potthoff et al. (2006) discuss a "technique for assessing the degree to which MAR assumption is tenable". Kline and Santos (2013) assess the sensitivity of the estimated results to deviations from MAR.

be expected to perform at least as well as) listwise deletion, but it is not possible to verify absolutely the validity of any multiple imputation model King et al. (2001:50-51). Relating to Lu et al. (2010), the variable for which a significant proportion is missing is government spending (as agent) on health. In this case, it seems plausible to reject the MCAR assumption in favour of the MAR, whilst discussion between MAR and NI remains open.

4.2 Analysis Possibilities with Missing Data

As missing data are not specific to developing countries – essentially, data in any field are likely to have some observations missing – numerous solutions to tackle the issue have been be designed by statisticians (econometricians), increasingly so with developments in computing technologies. Three broad categories of available methodologies can be distinguished: non-imputation methods; single imputation methods; and multiple imputation methods. These are discussed in turn in the following paragraphs.

Non-imputation Methods

The standard (default) practice[50] of treating the observations containing missing data items is the *complete-case analysis* or *listwise deletion*, whereby an observation with at least one missing data point is discarded. If the rate of missingness is high, this results in a loss of statistical power. Where the question of missing data arises in context of a panel dataset, the waste of the data may be even higher, as, after discarding the country-year observations with missing data, the remaining scattered country years with fully observed information may, depending on panel method, be useless, and therefore eventually, too, be discarded. Unless the underlying missingness mechanism is MCAR, this method can yield biased results and underestimated variance[51]. Little and Rubin (1987:41) suggest that the most common strategy to deal with this bias in the selection of complete cases (at least in dealing with sample survey) is to assign the case weights in the subsequent analysis.

[50] For instance, King et al. (2001:49), reviewing 'recent' literature in political science, suggest that "approximately 94% use listwise deletion to eliminate entire observations (losing about one-third of their data, on average) when any one variable remains missing after filling in guesses for some", therefore losing valuable information, reducing standard errors and possibly causing selection bias.

[51] King et al. (2012:51): "Inferences from analyses using listwise deletion are relatively inefficient, no matter which assumption characterises the missingness, and they are also biased unless MCAR holds. Inferences based on multiple imputation are more efficient than listwise deletion (since no observed data are discarded), and they are not biased under MCAR or MAR (Little and Rubin, 1989; Little and Schenker, 1995). Both listwise deletion and basic multiple imputation approaches can be biased under NI, in which case additional steps must be taken, or different models must be chosen, to ensure valid inferences. Thus, multiple imputation will normally be better than, and almost always not worse than, listwise deletion".

The relative performance of listwise deletion compared to other methods such as multiple imputation under different missingness mechanism in theory is possible through comparison of the associated minimum square errors. Clearly, in practice such comparisons cannot be made due to the nature of the problem itself (the complete dataset is not available). Furthermore, related to the discussion above regarding the missingness mechanism, with a rare exception of highly controlled experiments, the missingness mechanism (i.e. MAR versus NI) cannot be verified, further reducing the validity of comparison of methods altogether.

An alternative strategy, which is somewhat less wasteful (at least in the case of univariate analysis), is the *available-case analysis*, which includes all the cases where the variable of interest is observed. However, the sample base changes from variable to variable, therefore yielding potential problems of comparability across different sample bases. The problem is again exacerbated when panel dataset is used. Little and Rubin (1987:55) also discuss some *weighting methods*, which are essentially based on probability sampling.[52]

Generally, unless the MCAR assumption holds, and the missingness rate is very small, these simple methods provide biased inference, result in loss of statistical power and too small standard errors, and, as discussed above, may contain other problems or restrictions.

Single Imputation Methods

Imputation refers to the class of methods that impute (i.e. fill in) the values of the items that are missing. Several examples of single imputation are summarised below (Little and Rubin, 1987:60-62):

i. *Mean imputation* refers to the popular case where missing values are simply substituted by the sample (or sub-sample) mean (estimated from the observed data). Clearly, this underestimates the magnitude of both variances and covariances.
ii. *Hot deck imputation* refers to the method where a missing value is replaced by a value from estimated distribution: in practice, the empirical distribution consists of values from observed units, such that hot deck imputation substitutes the missing values with values drawn from similar responding units.
iii. *Substitution* replaces non-responding units with units previously not selected into the sample. It is worth noting that the resulting sample should not be treated as

[52] "A unit selected with probability π_i is "representing" π_i^{-1} units in the population, and hence should be given the weight π_i^{-1} in estimates of population quantities". This method, however, requires the assumptions of underlying distribution of missing values or the entire population.

complete, as the substituted units are respondents and therefore may systematically differ from non-respondents.

iv. *Cold deck imputation:* a missing value is replaced by a constant value from an external source (e.g. a value from a previous realisation from the same survey).

v. *Regression imputation* "replaces missing values by predicted values from a regression of the missing item on items observed for the unit, usually calculated from units with both observed and missing variables present" (Little and Rubin, 1987:61).

vi. *Stochastic regression imputation* "replaces missing values by a value predicted by regression imputation plus a residual, drawn to reflect uncertainty in the predicted value" (Little and Rubin, 1987:61).

vii. *Composite methods,* as implied by the name, combine ideas from the different methods mentioned above (e.g. "hot deck and regression imputation can be combined by calculating predicted means from a regression but the adding a residual randomly chosen from the empirical residuals to the predicted value when forming the values for imputation" (Little and Rubin, 1987:61)).

However, even when the imputation model is correct, single imputation inference tends to overstate precision of imputed observation because it omits the *between*-imputation component of variability (Schafer, 1999:7) (see section 'Combination Step' below).

Multiple Imputation Methods

If one was to rely on the default method of complete-case analysis using the dataset in question and discard any country for which at least one yearly observation is missing, the sample size would be reduced by 62% in the case where IMF data were used, 65% in the case where the WHO data were used, and by 90% if a conservative researcher were only to trust data simultaneously fully observed in both samples. Single imputation methods would overstate the precision of the missing data. Thus we explore the third alternative.

Multiple imputation (MI) is a simulation-based approach used for analysing incomplete data. MAR assumption is often key to the validity of MI, although some frameworks have been developed to accommodate cases where the missingness of data is non-ignorable (NI). The methods used in this chapter are built on the MAR assumption.

Any MI analysis involves three steps: an imputation step, where imputation model is formulated and M distinct imputed datasets are created; an analysis step, where each of the imputed datasets is analysed separately using the usual econometric techniques; and a combination step, where the results of the separate analysis are combined to generate a single set of estimates, according to the rules proposed by Rubin (1987).

The Imputation Step

The imputation step refers to creation (simulation) of M>1 imputed datasets, where the observed values are fixed across the M datasets, but the missing values in each dataset are replaced by imputed values that vary across the datasets to reflect uncertainty associated with the missing value.

The validity of the method hinges on how the imputations were generated: different MI methods undertake different distributional assumptions of completed (that is, observed and unobserved) data; they also rely on different algorithms[53] and different computational techniques[54]. Schafer (1999:4) argues that "the imputations should, on average, give reasonable predictions for the missing data, and the variability among them must reflect an appropriate degree of uncertainty". For instance, in cases where a variable in question can only have non-negative values in reality (e.g. government expenditure on health), it is not plausible for the multiple imputation procedure to lead to negative imputed values. Thus distributional assumptions, computational techniques, and modelling are all to be carefully considered.

Imputation models account for other variables in the model to be analysed, and the quality of that data may define the levels of uncertainty associated with the imputed value: if a sole datum is missing for one variable, and all other variables – as well as the imputed one – demonstrate some 'stability' in the overall dataset (e.g. all observed values are growing at a constant rate), the uncertainty associated with the imputed datum (that is, the variability of the imputed datum across the M datasets) will be smaller than in cases where there is a lot of variability in the observed data itself).

In terms of modeling at the imputation step, the consensus in statistical literature is that the imputation model should contain at least all the variables (and their transformations) that are going to be used in the analysis model. The consideration regarding the inclusion of auxiliary variables (up to potentially all the information available in the dataset) is twofold. On one hand, building the imputation model that contains the information of the entire

[53] Expectation-maximisation (EM) versus data-augmentation (DA) algorithms.
[54] E.g. Markov Chain Monte Carlo.

dataset has the advantage that the resulting multiple-imputed dataset may be used for any analysis regarding that dataset. On the other hand, if one builds an imputation model to be used for a specific analysis, there is an advantage of being able to include various non-linear and interaction terms relevant for a specific research question, which, however, add to computational complexity. In the context of survey design, for instance, one may plan for auxiliary variables that may help predict the missing value of a related variable (e.g. it may be useful to include some questions about number of rooms in a house of individual or a type of a car they drive, if the researcher anticipates that some individuals may decline to indicate their income).[55]

The dependent variable containing missing values should not be imputed. However, Lu et al. (2010) effectively use the MI technique to construct the dependent variable. [56] The outline and the discussion regarding the MI paradigm are mainly related to the missing observations in the *explanatory* variables. It is arguably beneficial to include the dependent variable in the imputation model (whilst imputing the explanatory variables), so that the relationships between the dependent and independent variables are preserved and accounted for: "If values of X [independent variables] are missing as well as Y [the dependent variable], then cases with Y missing can provide a minor amount of information for the regression of interest, by improving prediction of missing X's [sic] for cases with Y present" (Little 1992:1227). However, in cases where the dependent variable itself has missing values, the imputation is of little value: "If the X's [sic] are complete and the missing values of Y are missing at random, then the incomplete cases contribute no information to the regression of Y on X_1, …, X_p" (Little 1992:1227, where *p* denotes the number of independent variables). As noted above, in Lu et al. (2010), the explanatory variables (the Xs) are assumed to be fully observed, and the outcome variable is assumed to be missing at random. Von Hippel (2007:83) demonstrates that "using imputed Ys can add needless noise to the estimates" and, complying with Little's (1992) argument, proposes an imputation strategy whereby all cases are used for imputation, but following imputation cases with imputed Y values are excluded from the analysis. This way, he argues, the relationships between dependent and explanatory variables are maintained during the imputation process (imputed explanatory variables (or Xs) that are later used) when Y is used to impute Xs. When the explanatory

[55] http://www.ats.ucla.edu/stat/stata/seminars/missing_data/mi_in_stata_pt1.htm

[56] The authors apply multiple imputation technique to the GHE-A/GDP variable; they then subtract the DAH-G/GDP measure from the completed (observed and imputed) GHE-A/GDP variable to arrive to an estimate of GHE-S/GDP, which is the dependent variable in their analysis. For further explanation refer to the *Data* section above.

variables are fully observed, there is no need for imputation, because maximum-likelihood estimates can be obtained by deleting cases with missing Y (note, however, that this is more true for cross-section analysis rather than panel structure). The author argues that such strategy is more efficient compared to an ordinary MI (i.e. the one retaining imputed Ys): "[it] tends to give less variable point estimates, more accurate standard-error estimates, and shorter confidence intervals with equal or higher coverage rates" (Von Hippel, 2007:85).

This argument issues an important warning for the practice undertaken by Lu et al. (2010), as the authors rely on multiple imputation to generate values *only* for the dependent variable (based on the view that explanatory variables are fully observed), which are all subsequently used in the analysis (and combination) step. The issue is aggravated by the fact that, despite slight alteration to the dependent variable, given the recommendations regarding the imputation step that at least all the variables to be used in the analysis step must be used in the imputation model, the resulting outcome is that imputation model is (nearly) (at least economically) identical to the analysis model in the Lu et al. (2010) study. The 'ultimate' dependent variable is constructed by subtracting DAH-G/GDP (used also in both in the imputation and analysis models as an explanatory variable), from the multiply imputed GHE-A/GDP variable. This is clear from equations (2.1)-(2.3).

Imputation model:

GHE-A/GDP = **DAH-Gov/GDP** + DAH-nG/GDP + GGE/GDP + GDPpc +DR + HIV + e (2.1)

Construction of the 'ultimate' dependent variable:

GHE-S/GDP = **GHE-A/GDP**[57]– **DAH-Gov/GDP** (2.2)

Analysis model

GHE-S/GDP = **DAH-Gov/GDP** + DAH-nG/GDP + GGE/GDP + GDPpc +DR + HIV + e (2.3)

Overall, Lu et al. (2010) imputation procedure goes rather radically against the proposition that the imputed values of the dependent should not be used in the analysis model (**equation 2.1**). Their imputation and analysis models are effectively (economically) identical and thus may be *forcing* the fit of the model used at the analysis step (**equations 2.1 and 2.3**). The 'alteration' of the dependent variable (**equation 2.2**) for the ultimate analysis

[57] Completed = observed *plus* imputed.

model by using the key variable of interest (health aid to government, DAH-G/GDP) on both right and left hand side may further disrupt sound estimates and their economic interpretation.

Analysis Step

Analysis is performed using each of *M* imputed datasets, treating each completed dataset as if it was complete (i.e. no data was missing) to obtain a set of completed-data estimates $\hat{b}_M = \hat{b}_{(1)}, \hat{b}_{(2)}, \dots, \hat{b}_{(m)}$. In this chapter, for comparability with Lu et al. (2010), the Arellano-Bover/Blundel-Bond (ABBB) linear generalised method of moments estimator, designed for panels with large cross-sectional dimension and few periods, is employed.[58]

The analysis model, following Lu et al. (2010) can be summarised as:

$$\left(\frac{GHE-S}{GDP}\right)_{it} = \beta_0\left(\frac{GHE-S}{GDP}\right)_{it-1} + \beta_1\left(\frac{DAH-G}{GDP}\right)_{it} + \beta_2\left(\frac{DAH-nG}{GDP}\right)_{it} + \beta_3\left(\frac{DR}{GDP}\right)_{it} + \beta_4 GDPpc_{it} + \beta_5\left(\frac{GGE}{GDP}\right)_{it} + \beta_6 HIV_{it} + \mu_i + \varepsilon_{it} \quad (2.4)$$

where variables are as described in Section 2.

Combination Step

The main difference and advantage of *multiple* imputation, as opposed to single imputation techniques, lies in the third step where the *M* sets of estimates from analysis step are combined in such manner that the standard errors reflect the uncertainty associated with the underlying missing data.

The parameter estimates are usually just a simple average across the results from all imputed datasets:

$$\bar{b} = M^{-1}\sum_{m=1}^{M}\hat{b}_{(m)} \quad (2.5)$$

where $\hat{b}_{(m)}$ denotes estimates from each complete (observed plus imputed) individual dataset *m*; and *M* denotes the total number of imputed datasets.

[58] As noted by Lu et al. (2010:1379), the methodology is suitable for "independent variables that are correlated with past and present realisations of the error; fixed effects; and heteroskedasticity and autocorrelation within individual panels".

The standard errors account not only for the ***within* imputation variance** (the average of variance across imputations) (Schafer, 1997)[59]:

$$U_{\hat{b}} = \frac{\sum_m^M SE^2_{\hat{b}(m)}}{M} \tag{2.6}$$

where $U_{\hat{b}}$ denotes the *within* variation of a particular regression coefficient, $\hat{b}$, and is simply an average of the squared standard error (SE) over *M* imputed datasets;

but also the ***between* imputation variance** (a function of the variance of the parameter estimate across the imputed datasets and the number of imputations, *M*):

$$B_{\hat{b}} = \frac{1}{(M-1)} \sum_{m=1}^{M} (\hat{b} - \bar{b})^2 \tag{2.7}$$

where $B_{\hat{b}}$ denotes the sample variance of the parameter estimate, $\hat{b}$, over *M* imputed datasets,

therefore accounting for the uncertainty related to the imputed values.[60] The final combination of these two variances is described by the following formula (Graham et al. 2007:207):

$$T_{\hat{b}} = U_{\hat{b}} + \left(1 + \frac{1}{M}\right) B_{\hat{b}} \tag{2.8}$$

The resulting standard error is a square root of the two variance components added together.

It has been previously argued that good statistical inference could be achieved with sufficiently small number of imputations, *M* (that is, *M*=3 or *M*=5). This is based on the argument that gains to *relative efficiency* (itself based on the concept of mean-square error (MSE)) (Graham et al. 2007:207), summarized as:

$$\left(1 + \frac{\gamma}{M}\right)^{-1} \tag{2.9}$$

where γ denotes the fraction of the data that is missing, diminish rapidly after the first few imputations (Rubin 1987:114). Graham et al. (2007:208) conduct a series of Monte Carlo

---

[59] Note, however, that the SE components related to the model itself are not reported separately from the SE component related to the imputation.

[60] http://www.ats.ucla.edu/stat/stata/seminars/missing_data/mi_in_stata_pt1.htm

simulations to demonstrate that, whilst the empirical estimates of efficiency are fairly close to the theoretical predictions given by Schafer and Olsen (1998), "other important quantities, such as standard errors of the estimate, *p*-values, and power all vary rather markedly with the number of imputations (M), [and] statistical power can vary rather more dramatically with M than is implied by the efficiency tables presented in the previous discussions of MI theory". Their resulting recommendations are, especially concentrating on preventing the statistical power falloff, that M should be 20, 30, 40, 100, and >100 for γ= 0.10, 0.30, 0.50, 0.70, and 0.90, respectively (Graham et al. 2007:212).[61] As, given the current state of available computing power, a larger M does not require much (if any) of additional time or resources of a researcher, the choice of M=100 is applied throughout this chapter.

Multiple Imputation Software

A number of software packages have been developed to perform the task of multiple imputation. Some of those would only perform the imputation step; some would be designed to handle all three steps. The early developments were mostly responding to the demand in the context of medical statistical analysis, followed by those designed towards the problems is social sciences, and have now been incorporated in the leading statistical software packages, such as Stata or SAS. Clearly, developments are eased by the increasing computing power, enabling the creators and users to both introduce more complicated methods, as well as speed up the process itself. Statistical packages for multiple imputation differ on the basis of "interface, features, and results" (Horton and Lipsitz, 2001: 244). For instance, they assume different distributions of completed (observed[62] plus imputed) data, rely on different algorithms, and differ in flexibility of which features of the vast array in multiple imputation they can accommodate. Some examples and comparison of software packages (such as SOLAS, NORM, or MICE) that implements multiple imputation may be found in Horton and Lipsitz (2001). The analyses of this chapter follow Lu et al. (2010) and thus employ Amelia II (King et al., 2001) MI software, developed specifically for use in social sciences.

[61] Similar recommendations are provided in the STATA 11 manual for multiple imputation (p. 5): it is argued that the actual M required depends not only on the rate of missingness, but also on the analysis model and the data itself, with some of analysis requiring M to be 50 or more; generally, the handbook advises "using at least 20 imputations to reduce the sampling error due to the imputations".

[62] The observed data contains the observed data and the missingness indicator matrix.

At the time of development of Amelia II, the methodology of multiple imputation was "largely unknown and unused" (Schafer and Olsen, 1998, as cited in Honaker et al., 2012:50), and used by a few applied statisticians or social scientists, although the concept of multiple imputation had been around for several decades (if one takes Rubin, 1987, as the departure point). Back in 2001, the lack of computing power (compared to today) was still an issue holding back a wider implementation of multiple imputation: "[...] although this method [multiple imputation] is easy to use in theory, in practice it requires computational algorithms that can take many hours or days to run and cannot be fully automated".[63]

Amelia II assumes that complete data is distributed as multivariate normal. Clearly, this is an approximation, but "many researchers have found that it works as well as more complicated alternatives specially designed for categorical or mixed data"[64]. This assumption is used in a large fraction of multiple imputation software. Amelia II also assumes that data is missing at random (MAR)[65], based on authors' argument that "MAR assumption can be made more plausible by including additional variables" (Honaker et al., 2012:4).

The attractiveness of Amelia II seems to lie in its algorithm. Most of the multiple imputation implementing software packages rely on expectation-maximisation (EM) or data-augmentation (DA) algorithms. Amelia II combines the "classic" EM algorithm (or, more specifically, expectation maximisation with importance sampling (EMis)) with bootstrap approach to take draws from its posterior. For each draw, Amelia II bootstraps the data to simulate estimation uncertainty and then runs the EM algorithm "to find the mode of the posterior for the bootstrapped data, which gives fundamental uncertainty too" (Honaker et al., 2010:5). Bootstrapping essentially replaces "the complicated process of drawing μ [mean] and Σ [variance] from their posterior density" (Honaker and King, 2010:564).

The representation of bootstrap-based EM, and the overall summary of MI analysis is provided in **Figure 2.4**. Note, however, that Amelia II only implements the first (imputation)

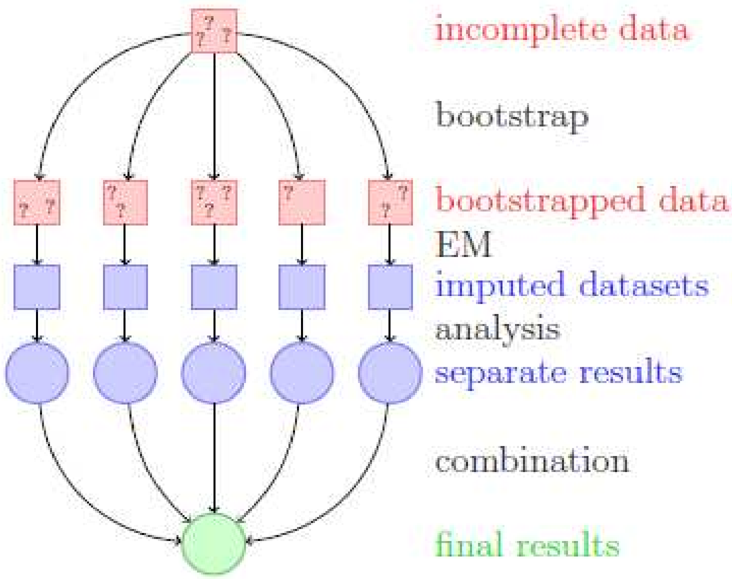
[63] There are further issues noted by King et al. (2001:50) that postpones the wider application of multiple imputation to data: "Because these algorithms rely on concepts of stochastic (rather than deterministic) convergence, knowing when the iterations are complete and the program should be stopped requires much expert judgment, but unfortunately there is little consensus about this even among the experts.

[64] Ezzati-Rice et al. (1995); Graham and Schafer (1999); Rubin and Schenker (1986); Schafer (1997); Schafer and Olsen (1998)) as cited in King et al. (2001:53).

[65] Although MAR is not directly testable, we can compare the basic variable descriptions (means, standard deviation, and correlation) between the fully observed data and observations containing some missing values. These are provided in the **Appendix Table A3**.

step of the MI, whilst the analysis and combination steps must be performed in other statistical software packages (e.g. STATA 13, using `mi estimate` commands).

Figure 2.4: The Process of Multiple Imputation Analysis in Amelia II



Source: Honaker et al. (2012:6.)

5. Results

Lu et al. (2010) claim to use multiple imputation techniques using Amelia II (version 1.2-13.0) to generate 100 imputations for each missing value of GHE-A. However, they then average the imputations into what becomes a singly imputed dataset (where each missing value is replaced by one average imputation), and conduct the analysis without exploiting the additional variation arising from the uncertainty assigned to missing values.

Having replicated the original Lu et al. (2010) results and confirmed that the dataset[66] is indeed one used in the original paper, we 'reinstate' the identified missing values in GHE-A variable in WHO and IMF datasets. Using Amelia II (version 1.6.4)[67], we carry out imputations based on varying underlying assumptions, to demonstrate the sensitivity of Lu et al. (2010)

[66] Downloaded from D. Roodman's blog: http://www.cgdev.org/blog/cross-post-aid-fungibility-debate-and-medical-journal-peer-review

[67] Latest available; no structural changes, mostly just bugs addressed.

findings.[68] Each variation of multiple imputed dataset contains 100 imputations (*M*=100). The imputed values are bounded to fall within the range of observed value. This rather *ad hoc* assumption has both positive aspects (i.e. avoiding irrational outlier imputed values) and negative aspects (such as imposing an upper bound that may not be correct). However, this is done following Lu et al. (2010), and is potentially as good as any other *ad hoc* bounds. As noted in Section 4, the dynamic panel estimator (ABBB) is used.

Figure 2.6 illustrates the observed and imputed values of the government health spending as agent as proportion of GDP (GHE-A/GDP) for WHO and IMF samples. The GHE-A summary statistics are provided in Appendix Table A4. Table 2.5 reports the estimation results for IMF samples for the short run (with full sets of short- and long-run estimates for WHO and IMF datasets reported in Appendix Tables A5a and A5b, respectively). Subsections 5.(1) – 5.(5) discuss multiple imputation under varying assumptions taken during imputation step. Subsection 5.(6) estimates the results only from (reportedly) completely observed cross-sectional units (countries). Subsection 5.(7) illustrates a widespread single imputation technique of sub-period averaging.

5.(1) MI: following Lu et al. (2010) assumptions

Lu et al. (2010) report generating 100 imputations for missing GHE-A/GDP simultaneously for WHO and IMF measures using all the right-hand-side variables of the analysis model, and including lags and leads[69] of the outcome variables.[70] Furthermore, although authors do not report this, it is obvious from their data they also impose imputation bounds equal to minima and maxima of the observed samples,[71] to avoid 'outlier' imputations (such as negative values,[72] or government health spending as proportion of GDP approaching one).[73]

Figure 2.6 illustrates the observed and imputed values of the government health spending as agent as proportion of GDP (GHE-A/GDP) for WHO and IMF. The imputed values tend to centre around the observed mean (0.0258 for WHO, and 0.0205 for IMF), and be less dispersed towards the tales, compared to the observed values. For individual countries (not depicted here), the uncertainty associated with the imputed missing value can be rather

[68] Missing values in WHO and IMF samples are imputed simultaneously, as suggested by Lu et al. (2010:1378).
[69] As the imputation models are predictive, and not cause (Amelia manual, Honaker et al. (2012:22).
[70] ET common seed 0128 is set across imputations for replication purposes.
[71] These are (0.002708; 0.09551) for WHO sample and (0.001712; 0.0867) for IMF sample.
[72] This imputation, nevertheless, returns some negative values for GHE-S/GDP.
[73] Dataset [**Amelia2014OneA**].

substantial. Summary statistics do not seem to change dramatically (see **Appendix Table A4**).

5.(2): Single imputation (averaged MI)

How much does estimating the multiple imputed dataset *appropriately* matter? Given that results of (1) are *fairly* different from Lu et al. (2010) despite following their claimed (and implied) assumptions, we average the imputed datasets of (1) into one, to construct a 'singly-imputed' averaged dataset[74] and estimate the results. The estimated coefficients are reported in column (2) in **Table 2.4**.

5.(3) MI: Assuming common time trend

As an alternative to inclusion of lags and leads of the missing variable, Amelia has an option to include time polynomials (up to third order) in the imputation model. This option allows for an assumption that observed values vary smoothly over time[75] and share a common trend.[76] Although an appealing option in theory, in our sample even the reportedly fully observed cases do not demonstrate 'smoothness' over time (see **Figure 2.5** for examples), let alone trying to fit a common trend. Partly, this may be due to scaling by GDP; partly, it may reflect an indeed varying health spending; partly, it could be driven by other reasons, or their combinations.

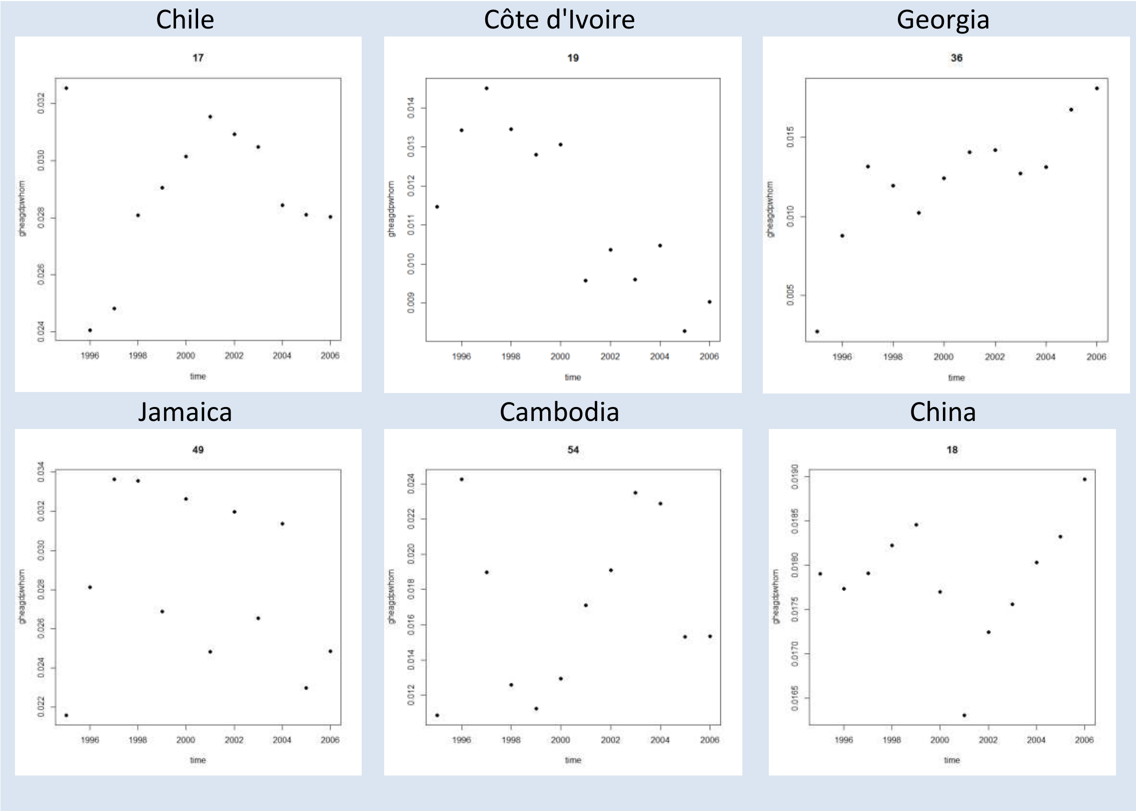
The comparisons between observed and imputed values of GHE-A are provided in **Figure 2.6**. Compared to multiple imputation assumptions discussed and reported in (1) (that includes lags and leads of variables) assuming a common time trend across countries during the imputation process (3) even more strongly centres the imputed values around the sample mean, deviating further from the observed distribution.

[74] Dataset [**Amelia2014OneA_Average**].

[75] If further interacted with the cross-section, this option would allow the pattern vary over time within the cross-sectional unit. However, given the small time series dimension this is not feasible, because of the amount of the extra parameters to be estimated.

[76] Dataset [**Amelia2014ThreeK3.dta**].The imputations with one (k=1) and two (k=2) time polynomials in the model were also carried out, but there were no seeming difference across them. Here only the results of the imputation model with three time polynomials are reported.

Figure 2.5: Example of GHE-A/GDP Country Patterns (WHO Data, selected countries)



5.(4) MI: Assuming fixed effects

Given the panel structure of our dataset (i.e. small time-series and large cross-sectional dimensions), cross-sectional observation fixed effects may be preferable to fitting a (common) trend over a short sample (especially given 'unsmoothness' exacerbated by expressing variables as proportion of GDP). Fixed effects simply imply that every cross-sectional unit (i.e. country) has a uniquely estimated constant term. This is a reasonable setting, unless one is strongly convinced that all cross-sectional units have the same patterns over time in all variables including the same constant term[77] (Honaker et al., 2012:21). The imputations are bounded as before.

Panel (4) of Figure 2.6 portrays the imputed distributions. Especially in the IMF sample (with lower fraction of missing data), modelling the imputation stage based on the fixed effects assumption seems to deliver the closest distribution of the imputed values to the observed values thus far, although not necessarily indicating of the most 'correct' imputation.

5.(5) MI: No (extra) assumptions

As a simple check we also impute a model with no assumptions regarding any of the determinants (country-specific constant or common trend). Although without bounding the range of imputed values the upper values comply with upper limits, between 7-8% of the 135600 observations are now imputed as negative. The imputed values are highly centred around the mean.[78]

5.(6) Complete-case analysis

Complete-case analysis discards observations for which any data are missing. Given our panel structure, we discard countries for which at least one value of GHE-A/GDP variable was identified as missing. GHE-A/GDP variable is reportedly fully observed (over the sample period) for 38 countries in the WHO sample and in 41 countries in the IMF sample. This corresponds to about a third (34-37 per cent) of the initial sample.

5.(7) Three-year sub-period averages

Given the pervasiveness of missing data in the context of developing economies, sub-period averages seem to be accepted as one of the standard solutions to compensate for some of

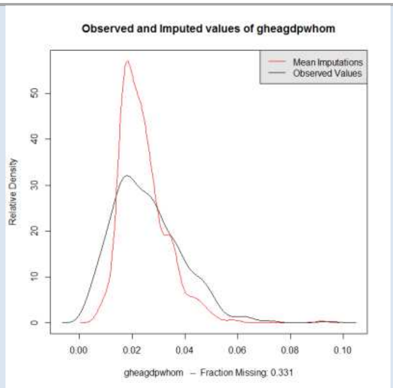
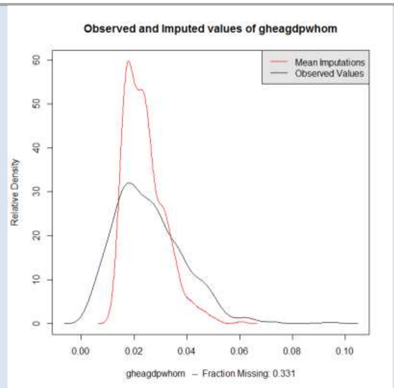
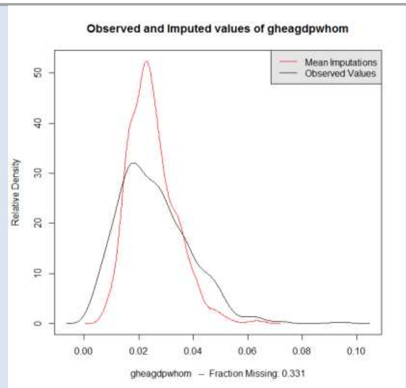
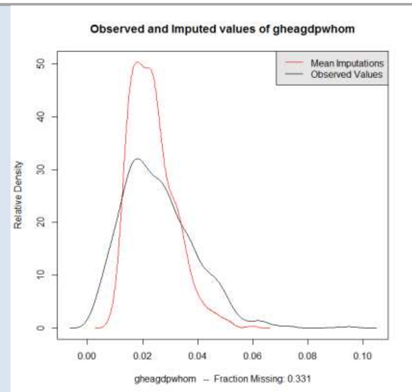
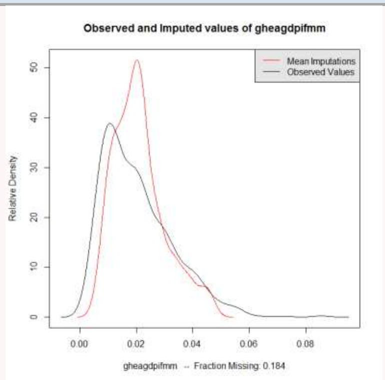
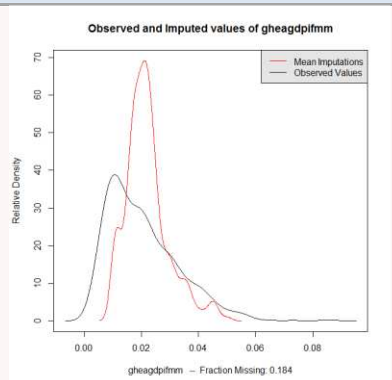
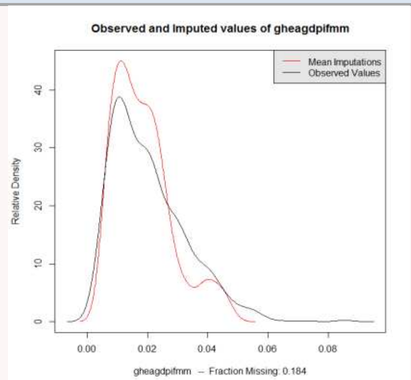
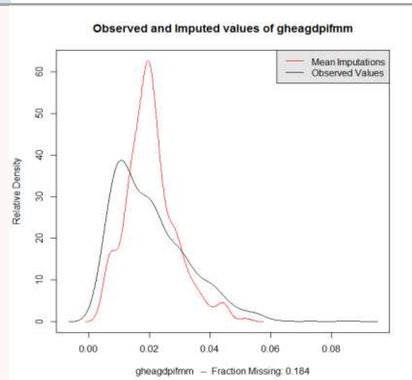
---

[77] Dataset [**Amelia2014FourFEA.dta**].
[78] Dataset [**Amelia2014FiveNoAs.dta**].

the missing data. In this approach, the sample period is divided into smaller subsamples (here, the 12 year sample period is divided into four three-year sub-periods). In each sub-period, the values across (potential maximum of) three observations are averaged. If at least one yearly observation in the sub-period is non-missing, the observation takes the average value derived from the non-missing yearly observations.[79] This is a variant of single imputation methodology, and thus overstates the precision of the imputed value. It decreases the time-series dimension, and reduces the variance of the variables across time. In our sample, following this approach leads to a sample of 64 (83) countries observed over the four periods in the WHO (IMF) sample.

[79] The remaining variables are also averaged into sub-period observations.

Figure 2.6: Observed and Imputed Values of GHE-A/GDP

	(1): 100 imputations according to Lu et al. (2010) assumptions (lags and leads)	**(3):** 100 imputations assuming three time polynomials/common trend	**(4):** 100 imputations assuming country fixed effects	**(5)**: No additional assumptions (only multivariate normality and MAR)
WHO (missingness rate 33%)				
IMF (missingness rate 19%)				

The table depicts observed and multiple imputed (M=100) values of the government health spending as agent as proportion of GDP (GHE-A/GDP) for WHO and IMF samples. The variation in the imputed values results from different assumptions taken during the multiple imputation process. The imputed values in (1)-(4) are bounded to the observed range; imputed values in (5) are unbounded.

Note that if imputed values in (1) were not bounded as described (not reported here), the imputed values would provide a much close correspondence between observed and imputed values, especially so for WHO, where the rate of data missingness is higher.

Table 2.4 provides the estimated results from variations of multiple and single imputation methods, and complete case analysis (for brevity, only results from IMF sample are discussed; WHO results are reported in the Appendix Table A5a and lead to comparable conclusions). The first column reproduces the Lu et al. (2010) estimates (significant coefficients are in bold, and the differences are italicised). The original paper (Column 'Lu' in Table 2.4) concluded that each dollar of health aid disbursed through the government (DAH-G/GDP) reduces government's domestically funded health spending (GHE-S/GDP) by $0.43 in the short run (and by $1.01 in the long-run, see Appendix Table A5). Meanwhile, each dollar of health aid disbursed through non-governmental organisations (DAH-nG/GDP) was found to *increase* government's domestically funded public expenditures on health by $0.58.[80] Government's commitment to domestically funded health spending was also increased as the size of the government grew. GDP per capita, disease burden, or debt relief with its associated pro-poor spending strategies were estimated to have no sizeable effect on government's domestically funded health spending.

The results from seven alternative strategies of handling missing data described in sub-sections 5.(1)-(7) are provided in the corresponding columns (1) through (7) in Table 2.4. As depicted in Figure 2.6, multiple imputed data following the Lu et al. (2010) assumptions (1) and those based on fixed effects (4) delivered the imputed distribution most comparable to observed data. Compared to the original Lu et al. (2010) estimates, there is little qualitative difference (although now government size does not seem to influence government's domestically funded health expenditure)[81]. Therefore, had Lu et al. (2010) used the MI techniques appropriately, their conclusions on aid fungibility would have been graver: each health aid dollar[82] assumed to be going through the recipient government[83] would be estimated to reduce government own health spending by $0.66 in the short run, and by $1.12 dollar in the long run[84]. Unless the estimates would have been met with higher

[80] Van de Sijpe (2013a: 1748) shows an interesting result: "if on- and off-budget health aid [here proxied (to some extent) by DAH-G and DAH-nG, respectively] are equally fungible, we see that $\hat{\beta}_{off} = \hat{\beta}_{on} - 1$". This is what we see in Lu et al.'s estimates. Following Van de Sijpe's interpretation, Lu et al. (2010) find that on and off budget aid are equally fungible. Van de Sijpe (2013a:1748) further notes that "[c]ontrary to Lu et al.'s interpretation, a marginal effect of DAH-G smaller than 0 does not necessarily mean aid is fungible; it could simply indicate that not all health aid is recorded on budget".

[81] In fact, this is a common conclusion across our estimations.

[82] Or, more correctly, each percentage point increase in health aid as percentage of GDP.

[83] Recall that DAH-G includes non-cash components, and health aid that cannot be traced to a particular channel.

[84] Note that the estimated long term effects are somewhat milder than (yet still comparable to) Lu et al. (2010) original estimates.

scepticism, this may have attracted even more attention and caused even graver policy implications.

Table 2.4: The Estimated Results (IMF, ABBB)

Dependent variable: Domestically funded public health spending (GHE-S)								
	(Lu)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Original Lu et al. (2010)	MI acc. to Lu et al. (lags and leads)	averaged (SI) (1)	MI common time polyno-mials	MI Fixed effects	MI with No (extra) assump-tions	Comple-te Case Analysis	3-year averages (SI)
Logged GHE-S/GDP	*.573*** (.055)*	.406*** (.084)	.603*** (.060)	.293*** (.084)	.414*** (.065)	.259*** (.081)	.582*** (.047)	.704*** (.184)
DAH-G/GDP	**-.433*** (.090)**	**-.663*** (.141)**	**-.597*** (.107)**	**-.716*** (.156)**	**-.603*** (.117)**	**-.729*** (.158)**	**-.560*** (.165)**	**-.536*** (.146)**
DAH-nG/GDP	**.580*** (.147)**	**.563*** (.215)**	**.571*** (.173)**	***.520* (.260)***	**.551*** (.190)**	***.497* (.264)***	**.428** (.179)**	*.320 (.293)*
DR/GDP	-.010 (.030)	.018 (.061)	.023 (.034)	.019 (.064)	.012 (.044)	.006 (.068)	-.002 (.026)	*-.071 (.042)*
GDPpc	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)
GGE/GDP	**.020** (.009)**	*.020 (.018)*	**.019* (.011)**	*.030 (.019)*	*.018 (.012)*	*.031 (.020)*	*.000 (.013)*	*.026 (.018)*
HIV	.028 (.026)	.026 (.041)	.027 (.024)	.060 (.046)	.048 (.033)	.060 (.048)	***.048** (.023)***	.003 (.041)
constant	**.005*** (.002)**	**.009** (.003)**	**.005** (.002)**	**.009** (.004)**	**.008*** (.002)**	**.009** (.004)**	**.008** (.003)**	*.005 (.005)*
N	**111**	**111**	**111**	**111**	**111**	**111**	**41**	**83**
T	**12**	**12**	**12**	**12**	**12**	**12**	**12**	**4**

Each cell reports the estimated coefficients from ABBB. Standard errors are reported in brackets. N denotes number of countries (not observations).

Simply averaging multiple imputed dataset into a singly imputed one (2)[85] (as Lu et al., 2010, have done) reduces the standard errors on multiple variables, in both short and long term. So, as predicted in theory, using single imputation for a missing value overstates precision associated with the unobserved value. The other single imputation solution of splitting the sample into three-year averages provides the most contrasting results: only the coefficient

[85] Note that these estimates are not identical to Lu et al.'s reported estimates, highlighting the possibility that some unreported assumptions were taken during the authors' multiple imputation process, i.e. complete replication was not possible.

of health aid disbursed through the government is estimated to be significant, rendering zero effect from all the remaining variables, including aid channelled through non-governmental organisations, (in fact, the significance of the DAH-nG coefficient appears to be sensitive across the treatments).

Overall, there appears to be little difference between the variations of assumptions taken during the multiple imputation step (though slightly more so in the WHO sample, where the fraction of the missing data is higher). The estimated effects of health aid variables are larger in absolute terms across the MI treatments than in the case of complete case analysis or mean imputation in three-year averaging. Applying multiple imputation methods in this context aggravates the estimated fungibility of health aid.

6. Check: Direction of Bias

The results section demonstrated that different approaches taken to tackle the missing data issue can alter the quantitative results. Two points are worth noting. Firstly, as we do not have the fully observed data, we cannot conclude which approach brings the estimates closest to the truth. Secondly, statistical literature (see Section 4) issues a warning against imputing the dependent variable (the case in Lu et al. and thus the focus of this chapter). To 'check' whether multiple imputation brings the estimates closer to the underlying parameters compared to alternative 'default' options of complete case analysis (listwise deletion) or expressing data in sub-period averages, we conduct a 'sensitivity check' of sorts. (Note that we keep the estimation methods as before to isolate the differences arising from the missing data issue. Fixed effects estimates, however, are also shown, to see whether the bias would move consistently across empirical strategies. The results are available in the Appendix Table A8).[86]

We take the largest subsample in which the variables are fully observed during the sample period of 1995-2006: the IMF sample of 41 countries (492 yearly observations); this corresponds to 37% of the Lu et al. (2010) sample. We completely randomly (complying with a more restrictive MCAR assumption) delete a fraction (about 19%, namely 90 out of 492 yearly observations, and then 33 % (161/492) for comparison purposes) of observations, to establish a missingness rate consistent with the initial IMF full sample (of 111 countries). We

[86] In terms of the bias, the FE results comply with the discussion of the subsection based on the ABBB estimates. Interestingly, however, FE estimates lead to conclusion that DAH-nG does not have a significant effect on government health spending (economically sound result, given that government is less likely to be aware of these flows; see Chapter 3 for discussion).

then impute the missing (discarded) data following the assumptions and structure of Section 5 (repeating the analysis 10 times[87] for robustness, keeping the 'location' of missing values identical, i.e. observation is coded as missing across repetitions). Although the sample is smaller, and the missingness assumptions are milder (and known, i.e. MCAR), we can now compare the estimated results using the imputation/amputation techniques to those of the fully observed sample. **Table 2.5** reports the estimated results (and the **Appendix Table A6** reports the range of estimates from the 10 imputations/estimations).[88]

Figure 2.7 contains the observed and imputed distributions for each variant of imputation assumptions. The upper panel reflects sample where the imposed missingness rate is 18.3% (as in the IMF sample). The lower panel conducts the equivalent exercise with the higher missingness rate (33%, as in the original WHO sample) imposed. As with the full Lu et al. (2010) sample of 111 countries (Section 5), the Lu et al. assumptions of including lags and leads of the variables and bounding the estimates to observed sample range (1), and fixed effects (also bounded to sample range) (4) assumptions return the imputed values closest to the observed distribution, on average. Imputations based on imposing a common trend across countries (3) perform poorly, potentially signalling substantial heterogeneity within the sample. Relying on the multivariate normality and MAR assumptions alone (5) tends to return imputations centred around a common observed mean, as depicted in column (3). The higher missingness rate in this sample does not seem to result in poorer multiple imputation performance, although modelling the imputation stage based on the fixed effects assumption seems to provide the best option across those considered here.

Table 2.5 lists the short run results from the ABBB estimations (the full set are reported in the **Appendix Table A7**). The first column reports the results from the fully observed sample of 41 countries over the period 1995-2006. Assuming that dynamic panel estimator (ABBB) is the correct econometric modelling choice, these estimates would represent the 'true' estimated effects: health aid channelled through the recipient government would reduce domestically funded public health spending, and the ODA health funds delivered through NGOs would significantly increase government's health spending. The disease burden, proxied by HIV prevalence rates would also be positively associated with GHE-S/GDP[89]. Changes in GDP per capita, government size (GGE/GDP) or debt relief would be estimated to have no significant effect on government's domestically funded health spending.

[87] Random seed.
[88] Dataset [**AmeliaEx2IMF.dta**].
[89] Note that HIV variable was insignificant in the original sample of 111 countries.

Unsurprisingly, none of the missing data handling approaches return results equivalent to those estimated from the fully observed data. However, some perform better than others: multiple imputation methods, based on Lu et al. (2010) assumptions (1) or fixed effects (4), deliver estimated coefficients that are qualitatively closest to the full sample estimates. Single imputation (sub-period averages) approach seems to perform particularly poorly, with the key coefficients of interest (health aid variables) are estimated as insignificant,[90] possibly because this particular approach wipes out the majority of year-on-year variation (and noise). Finally, complete case (listwise deletion) analysis could not be conducted at all, as only 2 out of 41 countries were fully observed for the 12 year period, impeding the use of panel estimation techniques, particularly illustrating the complications that can arise from the missing data in the limit.

Most importantly, imputing the dependent variable biases the results in unpredictable directions. Some assumptions at the imputation step result in higher (columns (3), (4), and (5)), some lower (column (1)) fungibility estimates in absolute terms. So whilst the multiple imputation seems to provide estimates most consistent to the full sample estimates, imputing dependent variables does bias the results, and does so in an ambiguous manner, even when the values are missing completely at random.

[90] As before, the significant of the DAH-nG/GDP coefficient is particularly sensitive. In fact, if the Fixed Effects model is used, DAH-nG/GDP is estimated to be insignificant consistently across treatments of missing data.

Figure 2.7: Observed and Imputed Values (Fully Observed Sample Only)

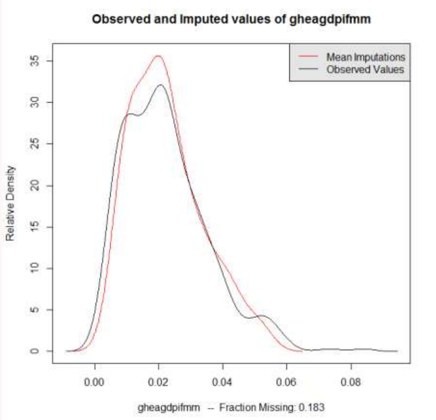
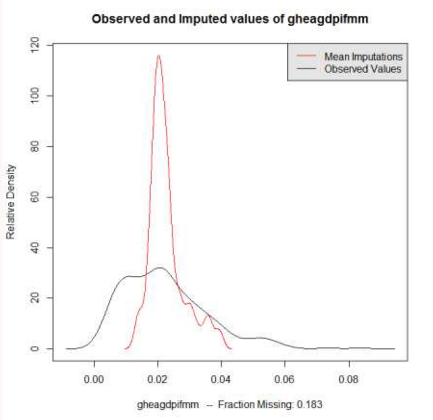
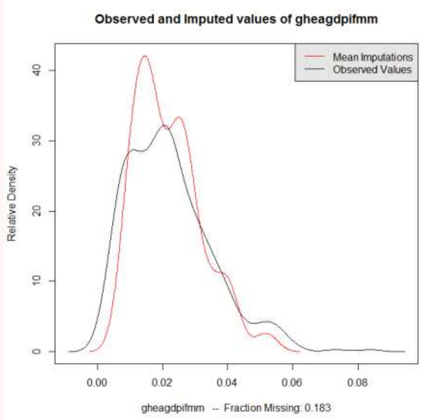
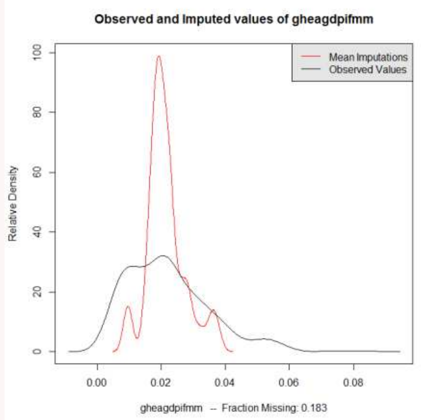
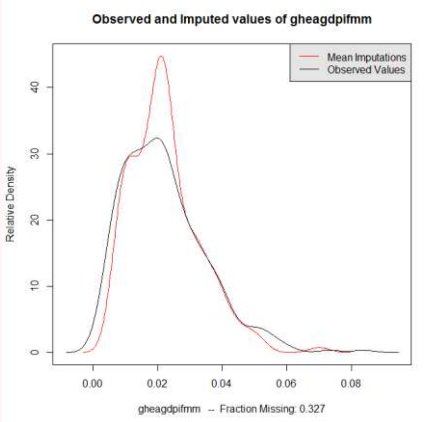
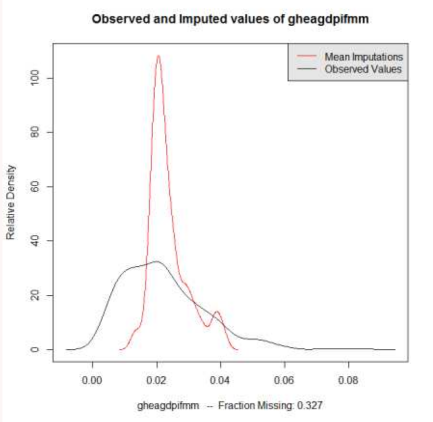
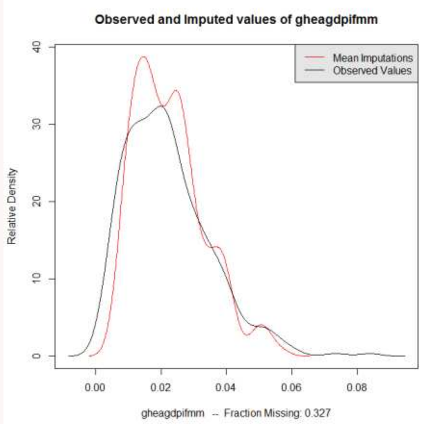
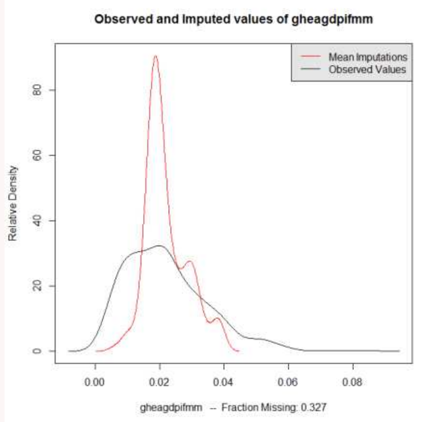
	(1)AmeliaEx2IMFone (As in Lu et al.)	(3)AmeliaEx2IMFthree (imposing common time trend)	(4)Fixed Effects	(5) No additional assumptions (no bounds, no time trends, or FE)
Missingness rate **19%**				
Missingness rate **33%**				

Table 2.5: Relative Performance across Missing Data Solutions (IMF, ABBB)

Dependent variable: Domestically funded public health spending (GHE-S)							
	Full sample	(1)	(3)	(4)	(5)	(6)	(7)
	Fully observed	**MI acc. to Lu et al. (lags and leads)**	**MI common time polyno-mials**	**MI Fixed effects**	**MI with No (extra) assump-tions**	**Complete Case Analysis**	**3-year averages (SI)**
Logged GHE-S/GDP	.582*** (.047)	.584*** (.066)	.188** (.092)	.440*** (.070)	.162* (.091)	-	.867*** (.321)
DAH-G/GDP	**-.560*** (.165)**	**-.531** (.215)**	**-.637*** (.242)**	**-.616*** (.224)**	**-.641*** (.236)**	-	-.540 (.328)
DAH-nG/GDP	**.428** (.179)**	**.312* (.189)**	.408 (.376)	**.457* (.254)**	.348 (.372)	-	.093 (.388)
DR/GDP	-.002 (.026)	-.0001 (.035)	.019 (.064)	-.005 (.043)	.001 (.076)	-	**-.100* (.055)**
GDPpc	-.000 (.000)	-.000 (.000)	.000 (0.000)	-.000 (.000)	.000 (.000)	-	.000 (.000)
GGE/GDP	.0002 (.013)	-.0001 (.018)	-.003 (.029)	-.011 (.025)	.001 (.033)	-	-.006 (.024)
HIV	**.048** (.023)**	**.058** (.026)**	.013 (.047)	.040 (.032)	.028 (.048)	-	**.131** (.062)**
constant	**.008** (.003)**	**.008* (.004)**	**.015*** (.006)**	**.013*** (.005)**	**.014** (.006)**	-	.0002 (.009)
N	**41**	**41**	41	41	**41**	2	40
T	**12**	**12**	12	12	**12**	12	4

Each cell reports the estimated coefficients from ABBB. Standard errors are reported in brackets.

7. Check: Better Observed – Better Quality?

Finally, we look at the quality of the data in the sub-sample in which both IMF and WHO health spending variables (GHE-A/GDP) are coded as fully observed in a country for the whole sample period simultaneously in both WHO and IMF samples. In theory, this sub-sample should reflect the best quality of the data. Specifically, we look at the discrepancy (absolute difference) between the GHE-A/GDP variable across the WHO and IMF samples, and try to elicit which explanatory variables may predict such discrepancies. Unfortunately, GHE-A/GDP measure is simultaneously fully observed over the entire sample period in only

11 countries[91] (132 country-year observations across 11, constituting 10% of the total sample). Therefore, we also report statistics for all the country-year observations for which the variable is simultaneously coded as observed (725 country-year observations across 110 countries). To see whether these form a representative sample (this is elaborated in Appendix Table A10), we also provide the sample statistics of explanatory variables for the total sample (1332 yearly observations across 111 countries). **Table 2.6** reports the (average) absolute differences between WHO and IMF records (as proportion of GDP); for a more contextualised measure, we also report these discrepancies expressed as percentage of the average value (IMF and WHO combined) of the GHE-A/GDP variable itself. The table also provides averages (and standard deviations) for all explanatory variables, for the fully simultaneously observed sample of 11 countries; pooled simultaneously observed yearly observations; total sample; and for each of the fully simultaneously observed countries.

Even in the sub-sample with the supposedly best quality of data, the data discrepancies can be substantial, signalling either substantial differences between reporting ministries (WHO data more likely to be based on Ministry of Health data, whilst IMF would tend to rely on Ministry of Finance), or even more underlying missing data that are not identified as such. In the 11 simultaneously fully observed countries, the average discrepancy between IMF and WHO measure of government total spending on health is 0.5% of GDP, or 20% of the average value of the GHE-A/GDP itself; in the sub-sample where all yearly observations for which both sources code GHE-A/GDP as observed, this increases to nearly 0.7% GDP, or 32% of the average value of GHE-A/GDP.

Country-wise, the discrepancies seem to be (but not always are) larger the poorer the country: for Cambodia, the discrepancies between the GHE-A/GDP measure across sources average to about 0.8% GDP (of 64% of the average GHE-A/GDP value); for Lesotho, the discrepancies reach nearly 2% GDP (42% of the average GHE-A/GDP value). However, some of the poorer countries have rather consistent data (e.g. Uzbekistan), and some of the richer ones (e.g.) Turkey exhibit above average discrepancies.

[91] Burundi, Cambodia, Côte d'Ivoire, Kazakhstan, Lesotho, Malaysia, Maldives, Namibia, South Africa, Turkey, and Uzbekistan. The full list of the differences of the observed country-years is reported in Appendix Table A9.

Table 2.6: Sub-sample Simultaneously Fully Observed in Both Sources (WHO, IMF)

	Discrepancies across sources		Health Spending		Health Aid		Other Explanatory			
	Absolute discrepancy (GHE-A WHO - IMF) (proportion of GDP)	Discrepancy as % of average* GHE-A/GDP	GHE-A/GDP (IMF)	GHE-A/GDP (WHO)	DAH-G/GDP	DAH-NG/GDP	DR	GGE/GDP	HIV	GDP pc
Fully observed (N=11, YO=132)	**.0051**[o] **(.0076)**	20%	.0260 (.0156)	.0270 (.0134)	.0023 (.0044)	.0004 (.0011)	**.0017** (.0044)	.1661 (.0745)	**.0582** (.0735)	**2181** **(<u>1858</u>)**
Pooled simulta-neously observed (N=110; YO=725)	**.0067** **(.0082)**	32%	.0219 (.0128)	.0258 (.0131)	.0028 (.0050)	.0006 (.0019)	.0051 (.0152)	.1458 (.0635)	.0273 (.0539)	**2250** **(2635)**
Full Sample (N=111; YO=132)	-	-	.0203 (.0122)†	.0259 (.0134) ‡	.0027 (0045)	.0007 (.0019)	.0048 (.0139)	.1445 (.0635)	.0275 (.0516)	**2285** **(2825)**
1: Burundi	.0008 (.0009)	11%	.0070 (.0010)	.0066 (.0010)	.0092 (.0098)	.0016 (.0025)	.0070 (.0064)	.2098 (.0407)	.0367 (.0105)	134 (39)
2: Cambodia	.0083 (.0038)	64%	.0087 (.0021)	.0170 (.0048)	.0077 (.0024)	.0012 (.0017)	.0004 (.0002)	.0501 (.0070)	.0167 (.0042)	379 (59)
3: Côte d'Ivoire	.0022 (.0016)	20%	.0099 (.0016)	.0113 (.0021)	.0009 (.0005)	.0005 (.0007)	.0116 (.0047)	.0806 (.0133)	.0572 (.0080)	865 (99)
4: Kazakhstan	.0021 (.0034)	8%	.0225 (.0035)	.0244 (.0042)	.0003 (.0002)	.0001 (.0001)	.0000 (.0000)	.1189 (.0104)	.00059 (.0005)	2266 (1216)
5: Lesotho	.0195 (.0122)	42%	.0534 (.0138)	.0339 (.0045)	.0033 (.0031)	.0001 (.0001)	.0001 (.0001)	.2800 (.0344)	.2212 (.0298)	626 (122)
6: Malaysia	.0010 (.0009)	6%	.0177 (.0042)	.0183 (.0034)	.0000 (.0000)	.0000 (.0000)	0 (0)	.1167 (.0108)	.0030 (.0014)	5082 (699)
7: Maldives	.0029 (.0054)	5%	.0407 (.0073)	.0433 (.0123)	.0002 (.0003)	.0000 (.0000)	0 (0)	.2116 (.0338)	.0001 (.0000)	2556 (266)
8: Namibia	.0109 (.0052)	26%	.0369 (.0056)	.0469 (.0058)	.0026 (.0011)	.0010 (.0016)	0 (0)	.2599 (.0406)	.1295 (.0292)	2513 (494)

9: South Africa	.0024	7%	.0317	.0330	.0002	.0001	0	.1885	.1461	4045
	(.0015)		(.0025)	(.0029)	(.0001)	(.0001)	(0)	(.0050)	(.0419)	(851)
10: Turkey	.0056	30%	.0300	.0340	.0000	.0000	0	.1185	.0284	4913
	(.0082)		(.0142)	(.0070)	(.0000)	(.0000)	(0)	(.0071)	(.0076)	(1159)
11: Uzbekistan	.0002	0.6%	.0280	.0282	.0009	.0001	0	.1934	.0004	612
	(.0002)		(.0050)	(.0051)	(.0007)	(.0002)	(0)	(.0175)	(.0003)	(136)

Table reports averages (proportion of GDP) over the sample period (1995-2006); standard errors are reported in parentheses. N denotes the number of countries. YO denotes the number of yearly observations

**The average value between IMF and WHO observations across time.*

º On average, the absolute value of difference between IMF and WHO values for government health spending as agent as proportion of GDP (GHE-A/GDP) differ about 20% of average value of GHE-A/GDP (average between IMF and WHO).

† YO=1083

‡ YO=890

Table 2.7: Sub-sample Simultaneously Fully Observed in Both Sources (WHO, IMF): What Determines the Discrepancies?

	Absolute discrepancy (GHE-A WHO - IMF) (proportion of GDP)	**Absolute discrepancy (GHE-A WHO - IMF) (proportion of GDP)**
DAH-G/GDP	0.1932	0.2180
DAH-NG/GDP	-0.0043	0.2029
DR	-0.1617	0.2029
GGE/GDP	0.3110	0.0686
HIV	0.5686	0.2401
GDP pc	-0.1648	0.0217
N	11	110
YO	132	725

Table reports correlation coefficients between the absolute values of the discrepancies, and explanatory variables.

Table 2.7 reports the correlation coefficients between the absolute value (as proportion of GDP) of discrepancies between IMF and WHO records of GHE-A/GDP. For the 11 simultaneously fully observed countries (left column), the differences tend to be (but not always are) larger the poorer the country, and the more aid flows through the government. The data are also less consistent the larger the government, and the higher the disease burden. (However, these correlations change rather dramatically in the pooled simultaneously observed sample, where more data seems to be associated with higher discrepancies, but the GDP is no longer a clear determinant of the size of such discrepancies).

Thus even the supposedly better quality data (in terms of being simultaneously observed across sources) exhibits substantial discrepancies. Nearly inevitably some fraction of data originating in developing countries is likely to be imputed.

8. Conclusions

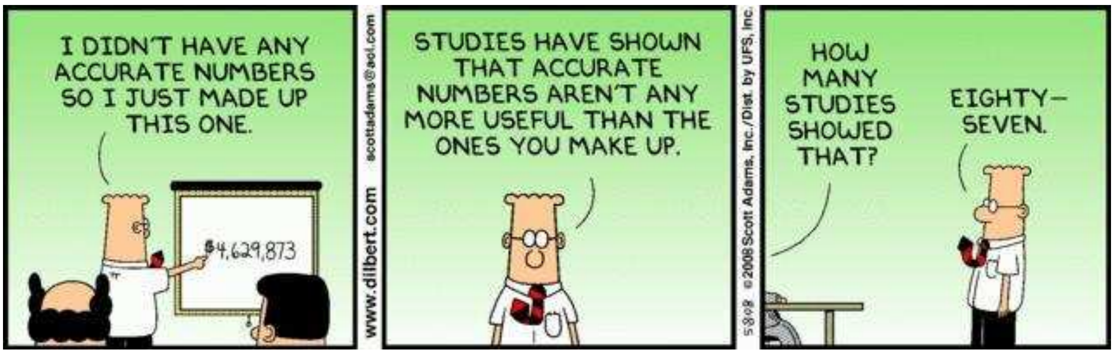
Three key issues have been raised in this chapter: missing data; geographical and institutional traceability of aid data; estimating fungibility of (health) aid.

Missing data is a pervasive problem in development economics. Data are costly to collect and to process, and this has historically not been a top priority from the recipients' perspective (which can be seen as rational given the array of pressing issues faced by developing countries). We have shown that, depending on statistical solutions applied to tackle the problem of missing data, the estimated results do vary. In the context of aid fungibility, and especially in such donor priority area as health, this may lead to substantial differences in the subsequent policy decisions. Furthermore, the missing data are often replaced by imputed estimates in a non-transparent way, and are typically not indicated as distinct from the truly observed data in the core international databases. As these international databases often are the source of data for research and evaluation, the resulting estimates may suffer from severe biases and overstated precision. When missing data afflicts the dependent variable and the (multiple) imputation methods are applied to 'recover' the missing observations, the estimates can be biased in an unknown (ambiguous) direction. We therefore reiterate the warning that dependent variable should not be imputed.[92]

[92] Multiple imputation is not a panacea for dealing with missing value in explanatory variables either. As Schafer (1999:4) suggests, whilst MI is a "valuable addition to any data analyst's toolkit" (especially

The missing (or poor quality) data is not solely a recipient's problem. Using Ravishankar et al. (2009) data, we have illustrated that aid flows, predominantly controlled and recorded by the donors and their organisations, lack information that would allow tracing the aid flows geographically or identifying the end organisation by which the aid is spent. And yet, to estimate the true extent of (health) aid fungibility, one needs to be able to distinguish between aid actually flowing to the recipient country and that spent elsewhere. Moreover, as shown by Van de Sijpe (2013), one also needs to distinguish between the cash aid flowing through the budget, and that going through NGOs operating in the recipient country or spent elsewhere. We have illustrated that health aid data used by Lu et al. (2010) do not allow for such fine decomposition. This leads to the third point.

We have argued that Lu et al. (2010) economic model biases the estimated effects of health aid fungibility towards the unfavourable conclusion. The likely overestimated health aid flows through the government, assumed to be spent in the year they were received, are subtracted from government's total aid spending to construct the dependent variable intended to describe the domestically funded component of the total public expenditures on health. This overlooks the fact that the receiving government would in practice implement forward-looking policy choices, which in turn implies that aid is fungible. This is the core topic in Chapter 3.



Source: http://dilbert.com/strips/comic/2008-05-08/

considering the simplicity and generality of the method), other – simpler, more conventional, non-simulation-based – methods may often provide a more suitable alternative to MI in the face of missing data problem.

Chapter 3

A Perspective on the Health Aid Fungibility Debate

1. Introduction[93]

Lu et al. (2010) argues that it is important to distinguish between aid flowing through the government, which is more likely to be fungible, and aid delivered through non-governmental organisations. The authors find that these flows have opposing effects on government's domestically funded health spending. Van de Sijpe (2013) formalises a similar binary distinction conceptually. He demonstrates than unless the 'off-budget' component of aid, defined as "aid not recorded on the recipient government's budget (e.g. donors building hospitals, training medical personnel, hiring consultants…)"[94], is included (and appropriately decomposed from the 'on-budget' component) in the fungibility estimations, the estimated fungibility of aid would be biased.

The fungibility debate interests many but is limited. Part of the problem arises from the differences in the (operational) definition of the issue itself. Lu et al. (2010:1376) define fungibility as "aid substitut[ing] for domestic government spending" and conduct their study along the lines of this definition to assess whether health aid disbursed through the recipient government reduces the (estimated) domestically funded public health expenditures (the

[93] I am grateful to Nicolas Van de Sijpe (Department of International Development and Centre for the Study of African Economies at the University of Oxford) for providing Van de Sijpe (2013) data.
[94] Van de Sijpe (2013:2)

limitations to their approach are discussed in Chapter 2). McGillivray and Morrissey (2004) distinguish between general fungibility (diversion of aid intended for public investment into government consumption spending), sector fungibility (whereby aid intended for a specific sector is (intentionally) spent under a different heading), and full additionality (assessing whether the total public spending in the sector increased by the amount of aid). See Morrissey (2012) for a review of recent literature.

Given the current state of (donor) data, a complete binary distinction between on-budget (potentially fungible) and off-budget (less likely to be fungible) sector aid is not feasible (see Van de Sijpe, 2013). Fungibility, if defined as aid spent according to the intentions of the donors, thus cannot be accurately estimated due to non-accessibility of necessary data, primarily, how much of aid actually flows through the recipient government and whether its intended sector can be identified. If Lu et al. (2010) were to be careful about their conclusion, it should have read that they demonstrated that health aid was not fully additional during the period of 1995-2006 (although there are issues with the estimations and construction of variables, discussed in Chapter 2), whilst the current data (even if carefully constructed, as in Van de Sijpe, 2013) do not allow for (empirically) conclusive statements about the (actual) fungibility of health aid.

Nonetheless, the current health aid data do allow for a certain degree of disaggregation (for instance, into sector support, technical cooperation, investment projects, and other, as in Van de Sijpe, 2013), allowing to explore the relationship between (sector) aid and spending, without invoking strict assumptions about the nature of such relationship. Approaching the question from the broader fiscal effects angle exerts less pressure on the data (though a clear distinction and correct disaggregation would indeed be useful): by denouncing the binary distinction underlying the aid fungibility studies, we can explore whether the (earmarked) health aid increases the total public spending on health, and, if so, which aid modalities do the best job in strengthening recipient's overall commitment to health spending.

This chapter is structured as follows: Section 2 introduces a simple (informal) conceptual framework of disaggregation of aid into its on-budget and off-budget components, commenting on the degree of potential fungibility. Section 3 discusses the different health aid disaggregation strategies given the available data. Section 4 briefly summarises the empirical estimation strategy, and provides two sets of results: firstly, isolating the differences arising from the alternative strategies of disaggregating health aid in Lu et al.

(2010) and Van de Sijpe (2013), using identical modelling and estimation methods, we demonstrate that differences may not be as pronounced as currently stated; and secondly conducting a simple sensitivity check to assess how would such results be altered if a binary distinction proposed in Section 2 was enforced on the data we illustrate the potentially conflicting conclusions and policy implications. Both sets of results return briefly to the missing data issues raised in Chapter 2, also testing for the hypothesis of aid smoothing. Section 5 concludes. Sensitivity checks are provided in the Appendix B.

2. Framework for Decomposition of On- and Off-budget Aid, and Data

Aid can either be ‘spent’ (in broad terms of end destination of cash or direct provision of goods/some services) in Recipient country, or Donor country. If aid is spent in the Recipient country, it can either be delivered through the Government[95], through Non-governmental organisations, or by the donor retaining control over the project. As Lu et al. (2010) did not explicitly consider the third possibility (delivery by the donor), our perspective reconciles the differences between the two existing studies: Lu et al.’s distinction between aid flowing through the recipient government as opposed to non-governmental channels and Van de Sijpe (2013) aid at the discretion of the recipient government versus control retained by the donor.

If aid delivered through the government is in cash, it (in theory) shows up on the recipient government accounts, and can be considered to be ‘on-budget’. If aid goes through the budget, it is recorded as revenue, and can potentially be linked to spending; in such cases fungibility could in principle be established: it is possible to assess/check whether aid was *spent* as intended by the donors. This component of (earmarked) aid is the most likely to be fungible, as the government (though not necessarily implementing ministry) is fully aware of these (liquid) flows. This is presented on the left side of **Figure 3.1**.[96]

Aid delivered through the recipient government but not in cash (e.g. medicines or technical assistants) would then represent a fraction of the ‘off-budget’ component. As the Recipient Government is (still) fully aware of such aid, it could in principle liquidate some of the goods

[95] For simplicity, it is assumed here the core aid flows are delivered through central government, which then decides whether to transfer any of the funds to the local government.
[96] No seeming disagreement between Lu et al. (2010) and Van de Sijpe (2013), as both assign it to DAH-G/on-budget category.

received and divert the resulting funds, or divert the services to other sectors.[97] Nevertheless, such off-budget aid is less likely to be fungible than (earmarked) aid received in cash.

Aid can be spent in the Recipient country through non-governmental organisations, and could be donated in cash (e.g. support to NGOs), or in direct provision of goods/services (say, books for a village school). Recipient Government may be at least partially, but probably not fully aware of these aid flows, as they are, too, 'off-budget'.[98] Such 'off-budget' aid may be considered partially fungible at most: to the extent that the government is aware of the non-government aid flows, (although unlikely to be able to 'capture' any of them) it would be able to reduce its own (domestically funded) component of spending (in aggregate, or within sectors); however, such actions would refer to fungibility of domestic revenues (tax, fungible by definition) rather than aid.[99]

A substantial fraction of aid is delivered through investment projects, where donors retain full (or a degree of) control of the project. In cases where donors retain full control of the project, aid would not flow through the recipient government as revenue, and thus could not be traced to expenditures. In such cases, aid would be spent as intended by the donor, and thus by definition be not fungible. More explicitly, if the government is aware of donor projects, it may reduce its own allocation, so aid through the donor may not be fully additional. The reason it is important to identify this component of aid is because it appears as part of (health) aid but cannot appear in public health spending; if not separated from on-budget health aid, it leads to a spurious appearance that health aid is fungible. In practice, however, donor projects could overlap (from the accounting perspective) with any of the components discussed above (and aid spent in the Donor country, too) and could induce a limited degree of fungibility if part of the project funds does flow through the government; this it is not depicted in the stylised **Figure 3.1**.

A considerable fraction of aid may be spent in the Donor country; for instance, paying for international organisations' administration costs (e.g. UN staff, research, or technical

[97] In Lu et al. (2010), such aid would be part of aid disbursed through the government(DAH-G), whilst in Van de Sijpe (2013:2) this would (conceptually) fall under the off-budget component (but would likely be unallocated to either on- or off-budget component assuming it is delivered through a donor project).
[98] See McGillivray and Morrissey (2000, 2001, 2004) for broader discussion of fungibility and related fiscal effects of aid.
[99] It is not exactly clear under which heading such aid would fall in either Van de Sijpe (2013) – though most likely would be recorded under donor projects in the OECD data, nor in Lu et al. (2010).

assistance[100]). The Recipient Government would be least likely to fund such activities itself; therefore the likelihood of averting such funds (including from own spending), assuming that it is even aware of it[101], is also least likely. Such aid could be considered to be least (if at all) fungible; however, in donor accounts it would still be considered 'aid'. This is visually presented in **Figure 3.1**.[102]

Note that in practice there will be some (conceptual or accounting) overlap between the components, such as Technical Assistance delivered as end services (e.g. training) in the recipient country; however, this would be unlikely to cross into the 'on-budget' component.

Finally, **Figure 3.1** also highlights the omission of financial flows from non-OECD donors (or 'collaborators'; see Walz and Ramachandran, 2011, for a summary of recent literature). Some of the non-DAC financial flows (although more likely to be in the form of FDI rather than fall under the OECD definition of ODA) to the Recipient country may also be delivered in cash to the Government, and thus be fungible. However, as the financial flows from non-OECD donors tend to be less conditional or earmarked (and less likely to meet the DAC definition of aid altogether), the resulting fiscal response may be more likely to take form in reduction of tax revenue or public borrowing, rather than diversion of spending from specific sectors. Such omission of potential non-DAC donor funds could bias the fungibility estimates upwards (i.e. the estimated extent of fungibility would be reduced). However, this currently cannot be tested, as non-traditional donor data are not consistently reported, and the financial flows are not clearly defined as aid.[103]

Overall, the aid fungibility question realistically applies only to the 'on-budget' component of (sector) aid. Following the 'strict' definition of aid fungibility, off-budget aid is not fungible: by the very fact that it is not going through the recipient government's budget, such aid is *spent* as donors intended. It is also less likely to be fungible in practice: though in principle goods received in kind may be resold and the resulting revenue diverted, or provided services may be diverted to benefit a different sector, this would not be likely with donor projects or aid spent in donor countries. Furthermore, using EU and WB budget support

[100] Technical assistance, as defined OECD, is, nevertheless, likely to deliver some end services to the Recipient country, but could also be regarded as intrusive if the donor-recipient views/priorities are not aligned.

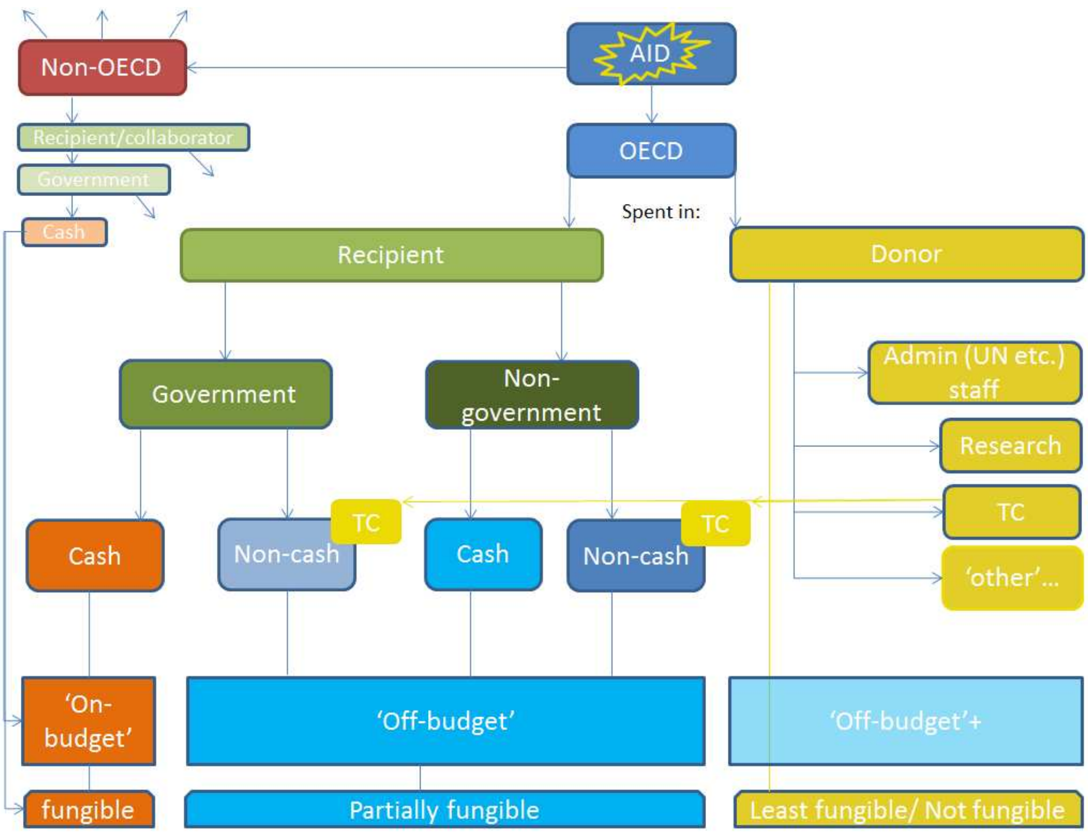
[101] Van de Sijpe (2013:13) notes that "In fact, Sundberg and Gelb (2006) argue that many aspects of TC, such as finance for training programs, analytical reports and expert advice involve resources that never even leave the donor country". This implies that the recipient is least aware of these flows.

[102] In Van de Sijpe (2013) this would fall under the off-budget aid; this is not explicitly discussed in Lu et al. (2010).

[103] Humanitarian aid is excluded from both measures.

data, Clist et al. (2012) show that donors actually choose the modality of aid delivery based on their assessment whether aid can be expected to be fungible (delivering aid largely in form of donor projects) or not (delivering most funds through general budget support) based on quality of recipient's institutions and the alignment of aid spending preferences, thus (to some extent) limiting the potentiality of fungibility. Nonetheless, off-budget can have broader fiscal effects than fungibility – aid can influence the level of spending through: complementarity; conditionality; aid illusion; strengthening of institutions resulting in subsequent changes in priorities; tax discretion (recipients reducing their own financing if what donors provide satisfy the level desired by the recipient) or improved collection; (see Morrissey 2012; McGillivray and Morrissey 2000, 2001, 2004).

Figure 3.1: Which Aid is Fungible? A Visual Representation

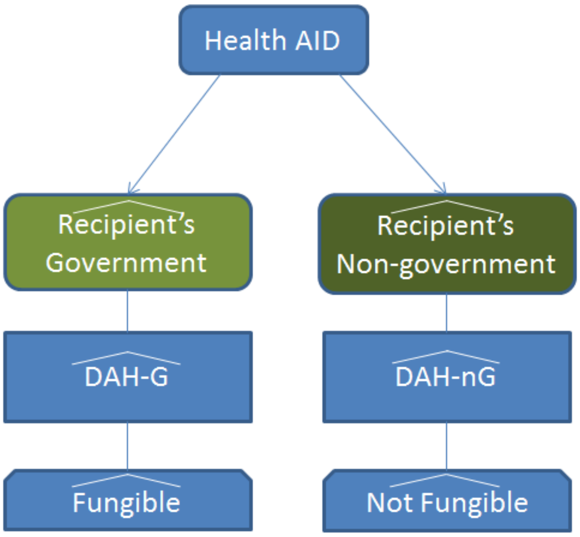


3. Existing Health Aid Disaggregation Strategies and Data

Two recent studies advocate the importance of disaggregation of health aid data: Lu et al. (2010) and Van de Sijpe (2013), with a few follow-up publications, such as Van de Sijpe (2013a) and Dieleman et al. (2013).

As discussed in Chapter 2, based on Ravishankar et al. (2009) dataset Lu et al. disaggregates health aid spending into the flows disbursed through the government (DAH-G) and those delivered through non-governmental organisations (DAH-nG)[104]. They evaluate the effect of these components on the (estimated) recipient's domestically funded health expenditures in 111 countries over 1995-2006, and conclude that health aid disbursed through the government is highly fungible (leading to less than one for one increments in total health spending), whilst health aid channelled through non-governmental organisations actually increases government's domestic resources devoted to health. This is sketched in Figure 3.2. Chapter 2 outlined the argument that DAH-G measure is overestimated due to inclusion of funds not channelled through government with certainty or not channelled in cash.

Figure 3.2: Lu et al.'s (2010) Health Aid Decomposition



Using a more precise definition of aid fungibility[105], Van de Sijpe (2013) formalises the health aid disaggregation into on-budget and off-budget components. He demonstrates that unless (health) aid is disaggregated, the resulting estimates of (total) health aid fungibility would be biased, with the size of bias depending on the variances of and covariance between the on- and off-budget components.

---

[104] In principle Lu et al.'s health aid measure includes funds spent in the donor country (unclearly allocated between DAH-G and DAH-nG) and potentially donor projects (see Chapter 2).

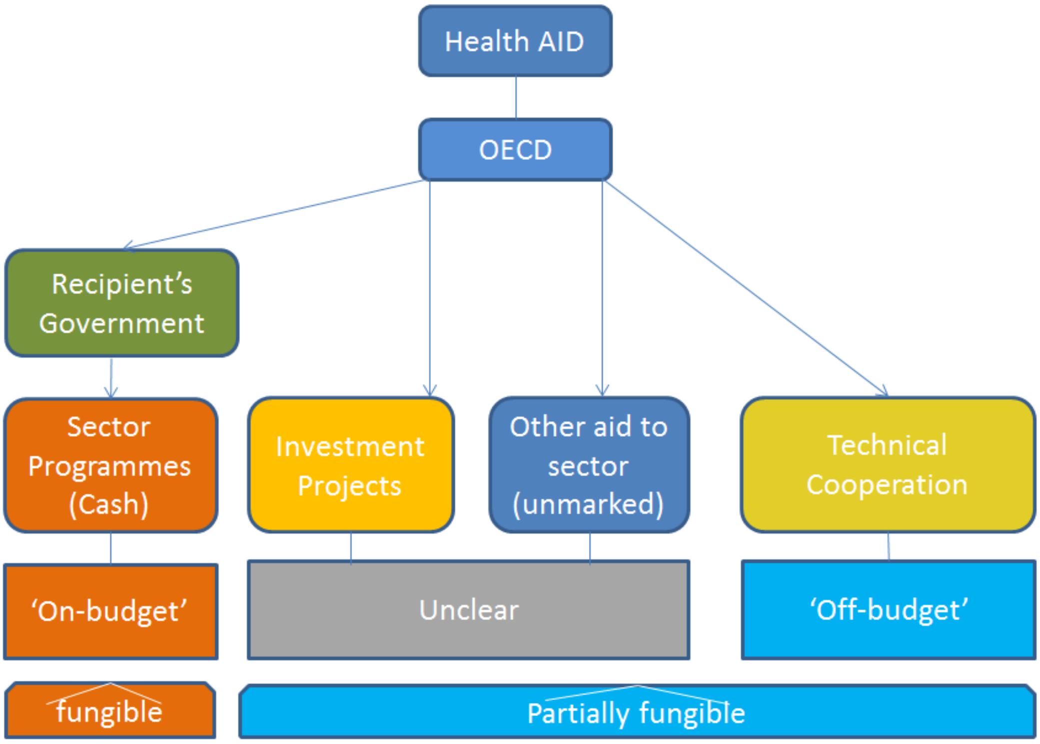
[105] "Fungibility occurs when aid is not used for the purpose intended by donors (McGillivray and Morrissey, 2004). More precisely, targeted aid is fungible if it is transformed into a pure revenue or income augmenting resource that can be spent whichever way the recipient government chooses (Khilji and Zampelli, 1994)" (Van de Sijpe, 2013:2).

Using OECD CRS and DAC data, Van de Sijpe (2013) distinguishes between four types of (earmarked) health aid components: sector programme (SP), technical cooperation (TC), investment projects (IP), and other (no mark) health aid (ONM). SP is a proxy for the on-budget health aid "as by definition programme aid involves a government to government transfer of resources" (Van de Sijpe, 2013:13). TC is used as a proxy for off-budget health aid, as this would either be spent in donor countries or would involve "direct payments from the donor government rather than a transfer of money to the recipient government" (Van de Sijpe 2013:13). The other two health aid components (IP, ONM) are not allocated either to on- or off-budget components but are instead included in the model as separate explanatory variables (together with general aid, a measure of support for NGOs, and other non-health sector aid). As donors tend to retain a certain degree of control over the donor projects (IP), it is much less likely to be disbursed through the recipient government than aid given as sector programme aid (see Makoro, 2008). Therefore, the "extent to which IP and ONM aid are reported in government budgets is more uncertain, so [their estimated coefficients] are less informative to gauge the degree of fungibility" (Van de Sijpe, 2013:14). Van de Sijpe's health aid disaggregation strategy is depicted in **Figure 3.3**.[106]

In effect, both studies (Lu et al., 2010, and Van de Sijpe, 2013) advocate a binary disaggregation of health aid, but draw the distinction along differing lines. Lu et al. end up doing the full approximation, attributing health aid flows to either category, without explicit statements which category IP or TC (or indeed any of the specific aid flows) is attributed. Alternatively, Van de Sijpe (2013) only allocates aid to a specific category when aid can be established as on- or off-budget from the OECD records (SP and TC, respectively). In effect, he demonstrates that, whilst important in avoiding biases of fungibility estimates, a binary distinction is currently not fully feasible in practice due to non-accessibility of necessary data (a fault in donor records). Even very carefully (de)constructed data (Van de Sijpe, 2013) does not allow for the full binary disaggregation of on- and off-budget data.

[106] Note that despite careful disaggregation, Van de Sijpe's data still involves a degree of estimation (where CRS data is not fully available) and scaling (to bring the overall numbers closer to DAC2 disbursements).

Figure 3.3: Van de Sijpe's (2013) Health Aid Decomposition

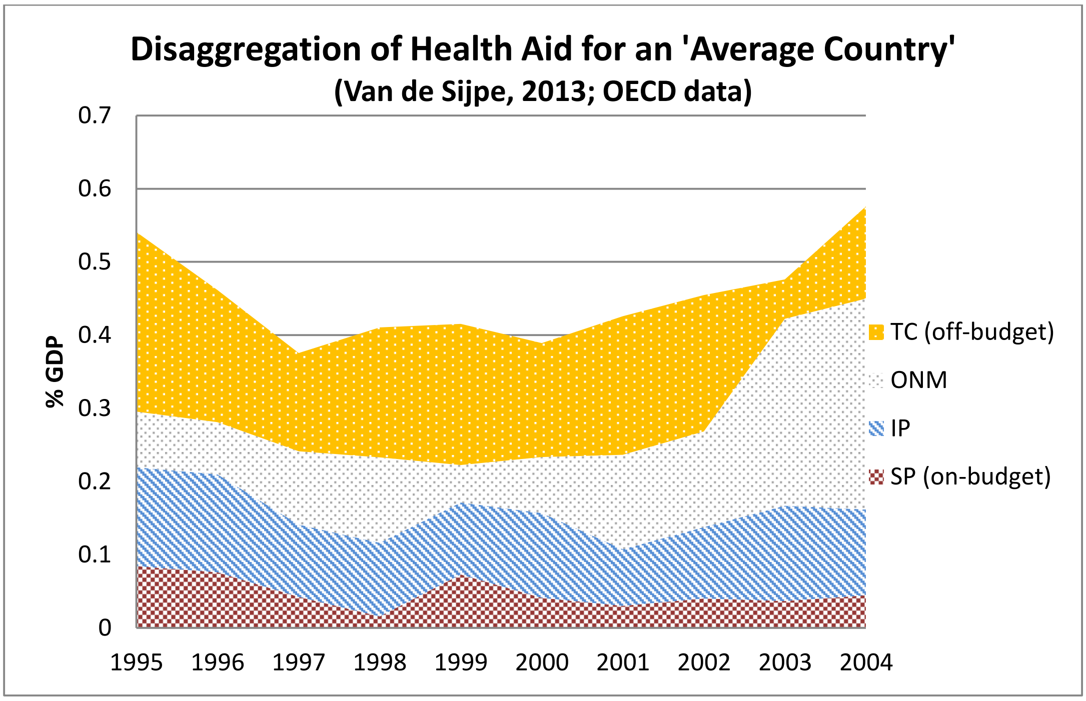


Data

Lu et al. (2010) sample contains 111 countries over 1995-2006; IMF sample is chosen as it has a lower missingness rate in the dependent variable compared to their WHO sample, and also because Van de Sijpe (2013:10) reports using IMF GHE-A data (though not publicly available). IMF dataset is likely to represent data from Ministry of Finance, whilst WHO samples are mostly based on records from the Ministries of Health. Van de Sijpe's dataset spans 108 countries over 1990-2004. We use the union of the two datasets to provide a directly comparable data. This resulting sample contains 108[107] low- and middle-income countries over 1995-2004. The resulting disaggregation of health aid in Van de Sijpe (2013, into SP, IP, TC, and ONM) and Lu et al. (2010, into DAH-G and DAH-nG) for an average country (or on aggregate) are depicted in Figures 3.4 and 3.5, respectively; the individual distributions of variables are depicted in Appendix Figure B1. Table 3.1 provides the summary statistics for the key health aid variables of both studies.

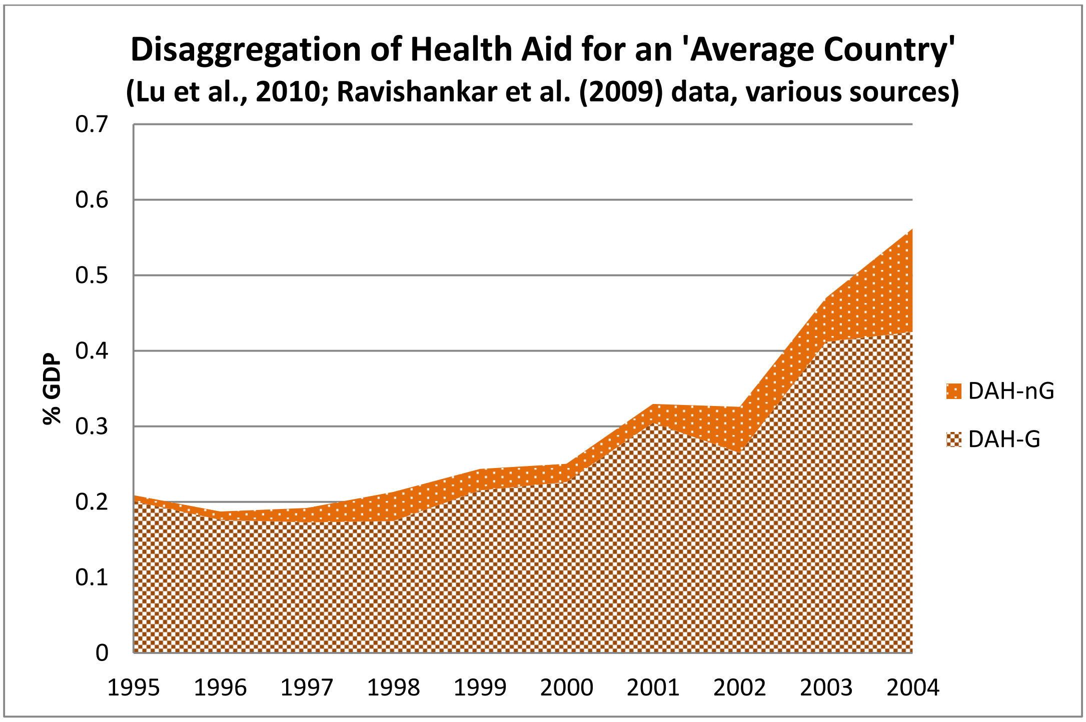
[107] Angola and Eritrea are dropped as in Lu et al. Libya is further excluded as contains 6 years of missing data for Van de Sijpe's health aid variables.

Figure 3.4: Disaggregation of Health Aid for an 'Average Country' (Van de Sijpe, 2013)



The figure depicts Van de Sijpe (2013) disaggregation of health aid (OECD CRS and DAC2) into its Sector Programme (SP, on-budget health aid), Technical Cooperation (TC, off-budget health aid), Investment Projects (IP) and Other (Not Marked) health aid (ONM) for 108 low- and middle-income countries over the period of 1995-2004. All variables are expressed as percentage of GDP.

Figure 3.5: Disaggregation of Health Aid for an 'Average Country' (Lu et al., 2010)



The figure depicts Lu et al. (2010) disaggregation of health aid (Ravishankar (2009) data compiled from various sources) into health aid disbursed through the government (DAH-G) and health aid channelled through non-governmental organisation (DAH-nG) for 108 low- and middle-income countries over the period of 1995-2004. All variables are expressed as percentage of GDP.

Table 3.1: Health Aid Summary Statistics

	On-budget		Off-budget		Other (unclassified)		Total	
Source	Lu	VDS	Lu	VDS	VDS	VDS	Lu	VDS
Name	DAH-G	SP	DAH-nG	TC	IP	ONM	DAH total	Health aid total
Mean	.0026	.0005	.0004	.0016	.0011	.0013	.0030	.0045
St.d.	.0041	.0011	.0013	.0025	.0025	.0024	.0048	.0060
Min	0	0	0	0	0	0	0	0
Max	.0386	.0175	.0140	.0269	.0398	.0205	.0392	.0736

Table reports health aid components' summary statistics (Lu et al., 2010, and Van de Sijpe, 2013) across the sample of 108 low- and middle-income countries during the period 1995-2004, expressed as proportion of GDP. The 'Total' health aid component is a sum of DAH-G and DAH-nG in Lu et al. data, and a sum of SP, IP, TC, and ONM in Van de Sijpe (2013) data.

Different disaggregation strategies result in stark differences. Lu et al.'s measure of 'on-budget'[108] health (DAH-G) is about five times higher than Van de Sijpe's (SP), on average across countries over the sample period (also with higher variance)[109]. This likely reflects overestimation of DAH-G and potential underestimation of 'on-budget' (SP) component in Van de Sijpe (2013). We concur with Van de Sijpe (2013a) that Lu et al.'s measure of on-budget aid (DAH-G) is likely to be overestimated (this is discussed in Section 2 of Chapter 2), as it contains a substantial amount that is not channelled in cash and/or not through recipient government (potentially including some of the investment projects(IP), TC, and/or ONM).[110] The likely underestimation of on-budget aid in Van de Sijpe is due to tracking sector programme funds only.

The opposite is true for the off-budget aid measure: Van de Sijpe's TC is about four times higher than Lu et al.'s DAH-nG (also with higher variance). This is likely due to the fact that

[108] Treating DAH-G as an 'on-budget' measure is a rough conceptual approximation: as argued in Chapter 2 and outlined in Van de Sijpe (2013), this includes non-cash components and aid likely not disbursed through the recipient government.
[109] Variability is important in Van de Sijpe's (2013) fungibility coefficient 'corrections', and clearly is very different using Lu et al.'s decomposition.
[110] We do not, however, agree with Van de Sijpe's (2013:1746) position that Lu et al.'s estimates "do not account for the fact that a lot of health aid is off-budget (that is, not recorded on recipient governments' budgets)". As noted in Dieleman et al. (2013), a conceptually similar distinction is partially recognised by decomposing the total DAH into DAH-G (roughly, on-budget) and DAH-nG (off-budget) components and demonstrating that they have different effects on government's domestically funded health expenditures. Nevertheless, as discussed in Van de Sijpe, Lu et al.'s disaggregation is severely limited.

DAH-nG may not include a substantial part of technical cooperation, estimated to be about 40% of total earmarked health aid in Van de Sijpe (2013:13), with its potential (partial) inclusion in DAH-G measure instead; or limited geographical traceability of funds spent/administered by donors without involvement of the recipient government or aid funds spent in donor countries.

The total health aid measure (sum over DAH-G and DAH-nG in Lu et al., and aggregate of SP, TC, IP and ONM in Van de Sijpe) is about 1.5 times higher (and slightly more variable) in Van de Sijpe's data, partly due to the inclusion of otherwise unclassified IP and ONM components (the sum of SP and TC components alone would instead be 30 % smaller than total DAH). The discrepancy between aggregate health aid measures illustrates further fragility of data, which stretches beyond disaggregation issues (although note that Lu et al. exclude health aid loans). Note that both measures exclude non-DAC donors' funds, and thus the total health aid may be equally underestimated in both studies.

Overestimation of the 'on-budget' aid component (by at least partly misclassifying off-budget aid as on-budget, as in Lu et al.) may lead to overestimated fungibility of aid (especially if the over-estimated aid figure is used to infer the domestically funded expenditure component). Underestimation of on-budget (misclassifying some of the on-budget aid as off-budget) component may, too, lead to distorted estimates, overestimating the positive effect of health aid. Such overestimation of the off-budget component may be seen as less harmful for the donor-recipient relations, and leading to less severe potential aid policy consequences (if the positive effect of aid on spending is overestimated, donors would not withdraw funds; although effectiveness of such funds may suffer). However, if such component is *substantially* overestimated (and the on-budget component severely underestimated), the resulting evaluation of relative performance of aid modalities may give misleading qualitative (rather than just quantitative) conclusions (attributing the positive effects on spending from on-budget aid to off-budget component), potentially discouraging donor from 'trusting' the recipient through the delivery of cash and reducing the health aid impact on spending.

In the fungibility context (where binary distinction is vital) the priority is measuring the on-budget aid component correctly, as only aid disbursed through the budget can appear as aid revenue and spending simultaneously. It matters whether the state of the data allows for binary disaggregation, or not. The residual un-allocable components in Van de Sijpe (2013), IP and ONM, amounting about a half (53%) of total (earmarked) health aid, lead to a complication with respect to fungibility estimates, as this leaves a considerable fraction of health aid outside the vital binary distinction, leading to less interpretable coefficients of unallocated aid (as the extent to which such aid is recorded on the recipient's budget is uncertain), and interdicting the corrections that the author proposes for correcting the fungibility estimates of total (disaggregated) aid given the remaining uncertainty in categorisation, and thus uncertainty in the relative variances of on- and off-budget aid.

In broader context (i.e. in measuring the effect of various aid modalities on total health spending, thus moving towards fiscal effects studies), the binary distinction is not necessary: the effect of more than two aid modalities on total public health expenditures can be estimated; and the resulting coefficients have a clear economic interpretation. Whether aid is fungible is not a concern under this approach; the question addressed is whether total health spending increases in proportion to increases in health aid.

4. Empirical Model and Results

Firstly, we compare Lu et al. and Van de Sijpe (2013) fungibility (rather, effect of components of health aid on total health spending) estimates arising from their differing decomposition of health aid. We use the same parsimonious model and Lu et al.'s dataset to isolate the differences in fungibility estimates arising solely from alternative ways of constructing the health aid data (i.e. keeping identical the estimator, the data for all (dependent and control) variables, the omitted variables, sample size). Secondly, using the same model, we conduct a sensitivity check by proposing an alternative binary decomposition of health aid. We also return to the missing data problem raised in Chapter 2 to inspect whether the estimated results differ depending on how the missing data are treated. The model is kept parsimonious and is based predominantly on Lu et al.'s specification to provide results that are comparable both to the Chapter 2 and Lu et al. (2010) and Van de Sijpe (2013) estimates. Lu et al. (IMF sample) and Van de Sijpe datasets overlap for 108 countries for 10 years (1995-2004), covering the majority of Lu et al. sample.[111] The estimated model can be summarised as:

[111] Lu et al. original findings are sensitive to this sample reduction (see **Appendix Table B10**); their core IMF results would be altered such that DAH-nG has no significant effect on GHE-S (the estimated domestically funded health spending), and GGE would no longer be significant.

$$GHE - A_{it} = \beta_1 AIDon_{it-k} + \beta_2 AIDoff_{it-k} \; (+\beta_3 AIDother_{it-k}) + \beta_4 DR_{it} + \beta_5 \ln(GDPpc)_{it} + \beta_6 GGEres_{it} + \beta_7 HIV_{it} + \mu_i \; (+\lambda_t) + \epsilon_{it} \tag{2.1}$$

for country *i = 1, …, N(=108)*, year *t = 1995,…,T(=2004)*; lag *k=0,1,2*. $GHE - A_{it}$ denotes recipient government spending on health (as agent, see below). $AIDon_{it-k}$ measures health aid disbursed through government's budget, whilst $AIDoff_{it-k}$ denotes health aid not recorded on recipient's government budget. $AIDother_{it-k}$ denotes any other aid modalities of aid potentially included in the model. DR_{it} refers to debt relief (assumed to be uniformly distributed across years for any given recipient) as in Lu et al. (2010). $\ln(GDPpc)_{it}$ is the natural logarithm of GDP per capita. $GGEres_{it}$ captures the total government expenditures less the $GHE - A_{it}$ itself to compare the increments in health spending to rest of the public spending. HIV_{it} measures HIV prevalence in the adult population, used as a proxy for disease burden. All variables, except HIV prevalence and GDP per capita are expressed as proportion of the GDP. μ_i captures time-invariant country effects, and λ_t denote a set of year dummies. ϵ_{it} is assumed to be a transient error. Data, except for the Van de Sijpe's (2013) measures of health aid variables, are as in Chapter 2 and Lu et al. (2010).

In Chapter 2 we argued that GHE-S (domestically funded health spending) construction contains some flaws, and thus GHE-A (total health aid spending) is a preferred choice for the dependent variable. First, it is not always clear whether it is GHE-A or GHE-S that is reported to the international databases (Lu et al., 2010); using GHE-A would avoid double deduction of aid flows in GHE-S (especially if aid is not spent in the year it was received, or was not recorded on the budget). Secondly, we argued that it is likely that DAH-G is likely to be overestimated, primarily due to assumption that health aid funds for which the channel of delivery is unclear are going through the budget. If this is the case, subtracting DAH-G from GHE-A to construct GHE-S would automatically underestimate recipient government's commitment to health (GHE-S) and overestimates the extent of fungibility (the opposite being true if Van de Sijpe's on-budget measure (SP) was used to construct GHE-A). Thirdly, if off-budget aid is to have any positive effect on government spending on health (increasing the efficiency of institutions, training of extra medical staff available to be employed, donors building the hospitals to be staffed by the government-paid staff, etc.), it would be more clearly reflected in the total measure (GHE-A) (a zero estimate would indicate no fungibility of off-budget aid).

We disagree with Dieleman et al. (2013) that GMM is the best estimator. Chapter 2 has demonstrated (see **Figure 2.5**) that the variables (expressed as a proportion of GDP) exhibit substantial year-on-year variation, the effect of which is exacerbated in GMM. This is in addition to Roodman's (2009b) criticism that system GMM is over-instrumented. We estimate a fixed effects (FE) model (with robust standard errors).

To check for the potential aid smoothing effects, we also estimate a variant of the model with lagged (and contemporaneous) aid included. If past health aid flows have a significant positive effect on total health spending in the model where contemporaneous health aid is also estimated to have a positive effect, we could conclude that we have found evidence of aid smoothing effects.

Finally, Chapter 2 has raised the issue of missing data that permeates developing country data. The chapter has also outlined the argument against imputing the dependent variable, and argued that defying this warning may bias the resulting estimates in unknown direction. Thus, for each model specification, we provide three sets of estimates:

a) a sample of countries for which all values are fully observed (in Lu et al.'s coding), corresponding to complete case analysis estimates: although it does reduce the sample considerably (to 50 countries[112]), at least the direction of bias is predictable, compared to multiple imputation;
b) a pooled (unbalanced) sample of country-years for which the dependent variable is coded as observed. This does not disrupt the fixed effects estimates, whilst allowing for a substantial increase in the sample size.
c) A full multiple-imputed and averaged Lu et al.'s sample from the overlapping sample of 108 countries. This is done in acknowledgement that data as supplied by the international organisations (where missing observations are imputed and coded as observed) are continued to be used. This further serves as a more direct[113] comparison to Lu et al. and Van de Sijpe's estimates (as the latter also uses IMF (though less accessible/not publicly provided) data).

[112] This sample excludes all countries for which at least one of the GHE-A observations is coded as missing in Lu et al. (2010) dataset (also exclude Angola and Eritrea). Libya is also dropped as it contains missing data for all health aid variables in Van de Sijpe (2013) data.
[113] Except for limited truncation of the sample (both Lu et al. and Van de Sijpe), alteration to the model (w.r.t. Van de Sijpe), and different dependent variable and estimator (w.r.t. Lu et al.)

The estimates reported in the main body refer to full multiple-imputed and averaged Lu et al.'s sample from the overlapping sample of 108 countries, with the results from other subsamples (a, b) reported in the Appendix B.

4.1 If modelled in the same way, do Lu et al. (2010) and Van de Sijpe (2013) yield such different results?

Table 3.2 compares the 'fungibility' results arising from Lu et al. and Van de Sijpe's propositions on disaggregating health aid. We isolate the differences from health aid variable constructions by using the same economic model (a version of Lu et al.'s parsimonious model, with all of their data, except for the health aid), fixed effects estimator (Lu et al. used ABBB dynamic system GMM), only varying decomposition of health aid. Column I reports results from Lu et al. disaggregation into DAH-G and DAH-nG; column II reports results from Van de Sijpe's disaggregation into on-budget (health SP) and off-budget (health TC) components, and including (as in author's original specification) the residual (uncategorised) health aid variables (health IP and health ONM), as well as other aid variables (general aid, support to NGOs, and other non-health aid).

Interestingly, results from Lu et al. and Van de Sijpe disaggregation strategies are very comparable. Both ways of health aid disaggregation imply that on-budget aid is somewhat fungible, but significantly associated with increments in government's overall commitment to health spending. Lu et al. disaggregation results in higher fungibility estimate. This is sensible given the argument that DAH-G is potentially overestimated, whilst Van de Sijpe's measure of on-budget health aid is potentially underestimated.[114] Van de Sijpe's measure is estimated less precisely. Both sets report narrowly defined off-budget (TC and DAH-nG) aid as having no significant effect on government health spending (not implausible). This effectively zero estimate indicates that either off-budget aid is not fungible (government chooses to maintain GHE-A), or at least fails to provide any evidence that it is fungible; alternatively, it could indicate that recipient government is not fully informed about these off-budget aid flows (spending does not respond). The results are consistent irrespective of whether we recognise the missing data problem or ignore it altogether (see Appendix Table B1 for full set of Lu et al. disaggregation estimates and Appendix Table B2 for Van de Sijpe's health aid measures). These findings effectively show that the key differences arise from modelling and estimator rather than basic disaggregation strategy.

[114] The estimate is much smaller than Van de Sijpe's own highly significant 0.84***(0.31) and only weakly significant. The results are not directly comparable to Lu et al. (2010) as they used GHE-S and not GHE-A as a dependent variable (and GMM instead of fixed effects).

Note that the most significant variables are unallocated health aid (IP, and ONM), signalling their potential importance in influencing/association with GHE-A, consistent with 'matched funding' (this is contrary to Van de Sijpe's results, where IP is estimated as insignificant).

Further note that results based on Van de Sijpe's health aid data are fragile depending on whether other aid variables are included. If only health aid variables (column III in **Table 3.2** for inclusion of all health variables; column IV for only explicitly on- and off-budget ones) are included[115], omitting other aid variables, the estimated effect of on-budget aid is insignificant, rendering only uncategorised health aid variables (health IP and ONM) bearing significant association (see **Appendix Table B4** for correlation coefficient across health aid measures). Thus, if only binary distinction was allowed for (e.g. in fungibility studies), Van de Sijpe's disaggregation could conclude no positive effects of aid and potentially full fungibility of on-budget aid (see section 4.2 for detailed discussion).

Potential aid smoothing effects are tested for by including lagged aid variables (see **Appendix Table B1** for Lu et al. disaggregation results, and **Appendix Table B2** for Van de Sijpe's). Columns 1, 4, 7 report the estimates where contemporaneous values of aid variables are used. Columns 2, 5, 8 use the lagged value of aid. Columns 3, 6, 9 report results where both contemporaneous and lagged values of aid are included; if both were estimated to be significant (and positive), we could conclude that government is potentially smoothing aid (i.e. hoarding back some of the health aid received to be spent over several years). Using Lu et al. disaggregation of health aid, we find evidence of some aid smoothing behaviour with respect to the health aid disbursed through the government (DAH-G), but only if the missing data problem is recognised (missing values of the dependent variable omitted rather than imputed as in Lu et al.). No such evidence is found if the missing data problem is effectively ignored. Using Van de Sijpe's health aid disaggregation, we find no evidence of aid smoothing behaviour in key explanatory variables (on-budget health aid, SP, and off-budget aid, TC).

Appendix Table B5 provides the (consistent) results with year dummies included in the estimation. Results are comparable, although ONM coefficient is no longer significant, strengthening the estimated differences of donor project funds from the rest of health aid.

[115] See **Appendix Table B3** for the full set of estimates across sample variations.

Table 3.2: Comparing Health Aid Disaggregation Strategies of Lu et al. and Van de Sijpe

	Dependent variable: GHE-A/GDP (Lu et al.)				
	FE, vce (R)	I	II	III	IV
		Lu et al. disaggr.	VDS disaggr.	VDS disaggr.	VDS disaggr.
		c	c	c	c
	N, YO	108,1080	108,1080	108,1080	108,1080
Health aid	**On-budget:** DAHG/Health SP	**.3570*****	**.5032***	.3074	.3182
		(.0707)	(.2825)	(.2970)	(.3225)
	Off-budget: DAHnG/Health TC	.1676	.0642	-.0541	-.0307
		(.1646)	(.1128)	(.1054)	(.1399)
	Uncategorised: Health IP		**.3928*****	**.3405*****	
			(.0525)	(.0535)	
	Uncategorised: Health ONM		**.3016*****	**.2368****	
			(.1098)	(.1103)	
Other aid	General aid		.0054		
			(.0204)		
	Support to NGOs		-.1308		
			(.1280)		
	Other non-health aid		**-.0114***		
			(.0066)		
Other controls	**Debt Relief**	-.0202	-.0094	-.0089	-.0089
		(.0439)	(.0493)	(.0466)	(.0509)
	Ln(GDPpc)	-.0011	-.0013	-.0012	-.0020
		(.0011)	(.0012)	(.0012)	(.0012)
	GGEres	-.0080	-.0086	-.0093	-.0082
		(.0105)	(.0109)	(.0110)	(.0111)
	HIV	.0406	.0360	.0406	.0392
		(.0314)	(.0304)	(.0323)	(.0334)
	Constant	**.0261*****	**.0284*****	**.0266*****	**.0330*****
		(.0078)	**(.0090)**	**(.0087)**	**(.0089)**
	R (w, b, o)	0.0624	0.0603	0.0518	0.0202
		0.0001	0.0009	0.0002	0.0114
		0.0014	0.0001	0.0016	0.0074

Table reports fixed effects (country-clustered robust standard errors) estimation results using full sample (108 countries, 1995-2004), and contemporaneous values of health aid (and other variables). Standard errors reported in the parentheses. Time dummies are not included. GGEres variable is constructing by deducting GHE-A/GDP from total government spending (GGE/GDP).

Appendix Tables B6, B7, and B8 report the results from first-difference estimations (one-, two-, and three-year differences respectively). Three-year differenced data broadly support our baseline results (it is not the residual autocorrelation driving the results), reporting IP as the most (and only) significant determinant of government total spending; in Lu et al., DAH-G coefficient is positive and significant at 5 per cent level (highlighting the potential misclassification of donor projects under the 'on-budget' heading). One-year differences provide less consistent results (estimating TC as the only significant variable). This is likely due to first-differencing on shorter time horizons amplifying a lot of noise and/or measurement error, which has been shown to be present in the data (see Chapter 2).

4.2 Alternative health aid disaggregation: sensitivity analysis

Van de Sijpe argued for effectively binary distinction of (health) aid into its on- and off-budget components, but the data did not allow him to do so completely, resulting in inclusion not only of his on- and off-budget variables (SP and TC), but also other health aid components (namely, IP and ONM). The upside of this is it avoids attributing other health aid components (IP and ONM) to either of the categories in an *ad hoc* fashion (given the uncertainty to which extent they appear on the recipient's budget). The downside is that it leaves the fungibility question largely unanswered: these unallocated health aid components (IP and ONM) amount to just over a half (53%) of total health aid (Van de Sijpe's measure); and thus his proposed coefficient 'corrections' are still fractionally *ad hoc*.

In contrast to Van de Sijpe's strict disaggregation, attributing uncertain components to the on-budget aid would be as unadvisable as Lu et al. assuming that health aid with unidentified channel of delivery (potentially including some of the donor project funding, among other components) is flowing through the recipient's budget (overestimating extent of fungibility). But as Van de Sijpe only proxies for the off-budget component using TC, it is highly likely to be underestimated, and, crucially, largely exclude off-budget aid that would be most likely to influence health spending (and only including the least fungible health aid component). Should a broader definition of the off-budget health aid be used (even if it contains some on-budget data), if there is no evidence that its coefficient is negative (i.e. is estimated to be zero or above), we could conclude that such, most likely off-budget, aid component is not detrimental (non-fungible) to health spending, whilst still estimating the effect of fungibility on what is most certainly an on-budget measure (SP).[116]

Effectively, this section provides a simple sensitivity check. Having a best-available disaggregated health aid data (Van de Sijpe, 2013), we can explore what are the effects of enforcing a binary distinction (as required by fungibility studies) of health aid data, or/and overestimating the off-budget component, and which components drive the differences in the results.

[116] If off-budget is found to be fungible, it may just show that donors spend a high proportion of what recipients term as 'enough' for a certain sector.

Six types of disaggregation of health aid are estimated:

1) Original disaggregation between sector programme (SP), Investment (donor) Projects (IP), Technical Cooperation (TC), and other unmarked (ONM) health aid (Van de Sijpe, 2013). As noted before, this disaggregation renders the IP and ONM coefficients uninterpretable in terms of aid fungibility.
2) Enforcing a binary distinction on Van de Sijpe's (2013) data: SP is treated as on-budget aid, and the sum of IP, TC and ONM is considered to be off-budget (denoted 'offbudgetVDS' below).
3) Sensitivity check on (2) to check whether it is the donor projects attributed to the off-budget component that are driving the results: SP is treated as on-budget aid, the sum of TC and ONM is considered to be off-budget (denoted 'offbudgetVDS2' below), and IP is estimated separately.
4) Descriptively, Lu et al. (2010) DAH measure (the sum of DAH-G and DAH-nG) is more likely to capture a larger extent of (geographically traceable) off-budget health aid, as it includes some broader headings, such as research funding or private (US-based) philanthropy funds, as Van de Sijpe's (2013, 2013a) proxy for off-budget aid, (DAC) Technical Cooperation, TC, is likely to omit substantial components of what donors classify (for accounting purposes) as (health) 'aid'. Van de Sijpe's (2013) carefully constructed measure of on-budget health aid (SP) should still provide a good proxy for the on-budget health aid. In this light, an alternative off-budget measure could be proposed by deducting SP from DAHtotal (a sum of DAH-G and DAH-nG minus SP; denoted 'offbudget' below). This measure results in some negative values (up to 1.3 per cent of GDP in value, see Table 3.3), which are primarily likely to be due to exclusion of health aid loans in Lu et al. (2010) DAH measure. Furthermore, Lu et al.'s DAHtotal is likely to underestimate the broader off-budget aid, as it would mostly consider funds spent in the recipient country, and non-governmentally given aid from US-based organisations only.
5) As in (4), but IP is estimated separately (and excluded from off-budget measure, 'offbudget2').
6) Lu et al. (2010) original disaggregation into health aid disbursed through the recipient government (DAH-G) and non-governmental organisations (DAH-nG).

Table 3.3 provides the summary statistics comparing the alternative health aid disaggregation strategies. The measures derived from mixing between Lu et al. (2010) and

Van de Sijpe (2013) health aid data yield some negative values (up to 1.3 – 1.8 per cent of GDP in value, see Table 3.3), which are primarily likely to be due to exclusion of health aid loans in Lu et al. (2010) DAH measure.

Table 3.3: Alternative Disaggregation of Health Aid: Summary Statistics

Variable	Construction	Estimation	Obs.	Mean	Std. Dev.	Min	Max
DAH-G		(6)	1080	0.0026	0.0041	0	0.0386
DAH-nG		(6)	1080	0.0004	0.0013	0	0.0140
DAHtotal	(DAHG + DAHnG)	-	1080	0.0030	0.0048	0	0.0392
Health IP		(1,3,5)	1080	0.0011	0.0025	0	0.0398
Health SP		(1-5)	1080	0.0005	0.0011	0	0.0175
Health TC		(1)	1080	0.0016	0.0025	0	0.0269
Health ONM		(1)	1080	0.0013	0.0024	0	0.0205
offbudgetVDS	(IP + TC + ONM)	(2)	1080	0.0040	0.0054	0	0.0688
offbudgetVDS2	(TC + ONM)	(3)	1080	0.0029	0.0040	0	0.0290
offbudget	(DAHtotal – SP)	(4)	1080	0.0025	0.0045	-0.0136	0.0386
offbudget2	(DAHtotal – SP – IP)	(5)	1080	0.0014	0.0038	-0.0182	0.0313

Table reports summary statistics (mean, standard deviation, minimum and maximum values) for alternative disaggregation of health aid variables, using Van de Sijpe (2013) and Lu et al. (2010) data. All variables are expressed as proportion of GDP.

We estimate the same country fixed effects model with country-clustered robust standard errors as in the previous section, for simplicity excluding non-health aid variables. Table 3.4 reports the results, with columns corresponding to six (1)–(6) alternative disaggregation strategies outlined above.

This primitive sensitivity check supports the postulated hypothesis about over- and under-estimation of health aid components, and highlights the sensitivity of the empirical results of alternative disaggregation strategies. Columns (2), (4) and (6) report the conflicting results from three alternative strategies of binary disaggregation of health aid data into on- and off-budget components, using best data available (Van de Sijpe, 2013, and Lu et al., 2010). A simplistic binary disaggregation of Van de Sijpe's health aid data (Column 2) reports that on-budget aid has no significant impact on total health spending despite flowing through the budget, whilst the off-budget component (a sum of donor projects, technical cooperation, and other health aid) has a significant positive effect on total health spending despite not flowing through the budget (in full, or even at large). Lu et al. (2010) disaggregation (Column 6) suggests the opposite: health aid channelled through the recipient government has a significant and positive (through less than one-or-one) effect on government public health expenditures; Lu et al. (2010) on-budget measure of aid (DAH-G) likely (though not explicitly) includes donor projects, potentially influencing the corresponding coefficient.

Unsurprisingly, combining the two data sources (Column 4) yields a conclusion that both on- and off-budget aid has a significant positive effect on GHE-A. This illustrates that conclusions (and thus policy recommendations) can depend drastically on the health aid disaggregation choices, especially if a binary distinction is enforced.

Table 3.4: Health Aid Disaggregation - Sensitivity Check

Dependent variable: GHE-A/GDP (Lu et al.)						
	(1) VDS original disaggr.	(2) VDS (binary)	(3) VDS	(4) Lu et al. +VDS (binary)	(5) Lu et al. +VDS	(6) Lu et al. original disaggr.
DAH-G						0.3570***
(on-budget)						(0.0707)
DAH-nG						0.1676
(off-budget)						(0.1646)
Health SP	0.3075	0.1835	0.1946	0.5842*	0.5190	
(on-budget)	(0.2970)	(0.3454)	(0.3333)	(0.3170)	(0.3160)	
Health IP	0.3405***		0.3127***		0.4103***	
(uncategorised)	(0.0535)		(0.0556)		(0.0478)	
Health TC	-0.0541					
(off-budget)	(0.1054)					
Health ONM	0.2368**					
(uncategorised)	(0.1103)					
offbudget				0.3172***		
(DAHt-SP)				(0.0567)		
offbudget2					0.2743***	
(DAHt-SP-IP)					(0.0657)	
offbudgetVDS		0.2142***				
(IP+TC+ONM)		(0.0538)				
offbudgetVDS2			0.1274			
(TC+ONM)			(0.0903)			
Debt Relief	-0.0089	-0.0038	-0.0027	-0.0179	-0.0145	-0.0202
	(0.0466)	(0.0458)	(0.0456)	(0.0400)	(0.0396)	(0.0439)
Ln(GDPpc)	-0.0012	-0.0011	-0.0010	-0.0011	-0.0009	-0.0011
	(0.0012)	(0.0012)	(0.0012)	(0.0011)	(0.0011)	(0.0011)
GGEres	-0.0093	-0.0082	-0.0081	-0.0084	-0.0084	-0.0080
	(0.0110)	(0.0109)	(0.0109)	(0.0105)	(0.0105)	(0.0105)
HIV	0.0406	0.0405	0.0407	0.0413	0.0416	0.0406
	(0.0323)	(0.0334)	(0.0335)	(0.0309)	(0.0312)	(0.0314)
N	1080	1080	1080	1080	1080	1080
r2_a	0.0447	0.0366	0.0384	0.0584	0.0602	0.0572

Table reports country fixed effects estimates (country-clustered robust standard errors) for various disaggregation strategies of health aid. Column (1) reports estimates from Van de Sijpe (2013) disaggregation of health aid (identical to column III in Table 2); Column (2) enforces a binary distinction on Van de Sijpe's data by aggregating IP, TC and ONM components into 'offbudgetVDS' variable; Column (3) estimates SP and IP separately, aggregating TC and ONM into 'offbudgetVDS2'. Column (4) uses 'offbudget' constructed by deducting SP from DAHtotal, therefore using both Lu et al. (2010) and Van de Sijpe's measures (SP). Column (5) further deducts IP from DAHtotal ('offbudget2') and including it separately in the estimation. Column (6) reports estimates from Lu et al. (2010) original disaggregation (and replicated Column I from Table 3.2). Standard errors are reported in parentheses.

Departing away from the binary distinction, primarily by disaggregating away the donor project component from the off-budget health aid (Columns 3 and 5 in Table 3.4), reiterates the previous result that IP has a robust significant positive effect on recipient health spending. This may be due to these flows potentially at least partially flowing through the budget, but is as likely (given the uncertainty to the extent of such flows appearing on the budget) to signal the positive effects of investment projects through complementarity (matched funding), conditionality, institutional and capacity building, and similar channels. The contrasting findings of the residual off-budget aid (after the removal of SP and IP components from the respective total measures of Van de Sijpe and Lu et al.) illustrates that the contents of Van de Sijpe (2013) and Lu et al. (2010) data differ substantially: the estimated effect of 'offbudget2' (Lu et al.'s DAH total less the sector program and donor projects) is positive and significant, whilst Van de Sijpe's 'offbudgetVDS2' (a sum of TC and ONM) portrays no sizeable effect.

Overall, using a (limited) parsimonious model and a fixed effects estimator, Van de Sijpe's on-budget component (sector programme, SP) is rather consistently estimated as having no (strongly) significant effect on total public health spending in low- and middle-income countries. This highlights potential underestimation of such on-budget measure (though may, too, be due to fungibility effects, or omission of important variables, such as tax revenue and other forms of aid). The proposed alternative off-budget measures based on binary disaggregation of health aid are estimated to have a highly significant positive effect on government's health spending (even if it is primarily driven by the contribution from donor projects, as the latter does constitute off-budget aid unless it flows through the government). Two key interpretations are worth noting. Either this off-budget aid contains a significant proportion of on-budget aid, or off-budget aid actually has a positive effect on total health spending (e.g. through requirement of some complementary funds, monitoring, etc.). Finally, as noted above, Lu et al. (2010) on-budget measure of aid (DAH-G) likely (though not explicitly) includes donor projects, rendering the corresponding coefficient positive and highly significant – contrasting findings from Van de Sijpe's data.

5. Conclusions

In evaluation of fiscal effects of aid (including fungibility), it is important to distinguish between on- and off-budget aid flows. We introduced a simple conceptual disaggregation of (health) aid into its on-budget and off-budget components, and argued that, if a careful definition of fungibility is adopted, off-budget aid is unlikely to be fungible. Given the current state of donor data records, however, a binary disaggregation of (sector) aid flows is not feasible in practice, preventing conclusive statements about fungibility of aid. Even aggregate figures reveal stark differences, with even larger relative discrepancies in the disaggregated data.

The existing health aid fungibility estimates (primarily those by Van de Sijpe, 2013, and Lu et al., 2010) yield less conflicting results than is currently stated. We showed that if the estimation strategy is the same, Lu et al. and Van de Sijpe's estimates lead to very comparable – qualitatively not conflicting – conclusions: on-budget aid increases health spending (even if partially fungible); and the narrow definition of off-budget aid (i.e. at least excluding donor projects) has no significant effect on GHE-A. Quantitatively, the estimates inevitably differ slightly due to different health aid disaggregation strategies, and do so in expected direction. However, the disaggregation strategy is important: if a simplistic (*ad hoc*) binary distinction was enforced on the data, policy recommendation using Van de Sijpe's data would delegate health aid to off-budget channels; Lu et al.'s – through the recipient government, and constructing a measure using both datasets would conclude that both channels are suitable for (effective) health aid delivery.

The lack of distinction between on- and off-budget aid may partly explain the mixed evidence of fungibility studies, especially if studies aiming to assess the relationship between fungibility and aid effectiveness are considered. Petterson (2007), who finds that despite high estimated sector fungibility, shows that aid effectiveness is not reduced in the face of such high aid fungibility. This is consistent with potentially over-estimated on-budget aid (leading to higher estimates of sector fungibility). The estimated effectiveness of aid would instead capture the effects of both on- and off-budget aid (e.g. by effectively taking into account impacts of technical cooperation), and thus the on- / off- budget disaggregation would be of lesser importance. The widely cited Feyzioglu et al. (1998), who assessed both general (or aggregate) fungbility (results sensitive to sample size) and sectoral fungibility (aid to agriculture, education, and energy found to be fungible, whilst aid to transport and communication sector were not), did not find any conclusive evidence of the fungibility of health aid. This may be partly due to their measure of aid: the authors used concessional loans to assess sector fungibility. Whilst concessional loans would likely be mostly on-budget (and most relevant to some of the other sectors such as energy), the health sector aid, especially during their sample period, is more likely to be disbursed in the form of grants, resulting in underestimating the on-budget component (and virtually omitting the off-budget component), and potentially resulting in inconclusive findings. Given the key argument of distinguishing between on- and off-budget aid, the findings are not directly comparable to other estimates of aid fungibility.

Departing from the limited fungibility question and approaching the problem from a broader (yet still limited) fiscal effects angle allows to analyse the issue and the data more plausibly. Relinquishing the binary disaggregation of aid required by the fungibility studies allows to explore the relationship between separate health aid components and total (domestically and externally funded) health spending, allowing to assess which health aid modality has the most sizeable effect on recipient's commitment to health sector. Donor (investment) projects are found to have the most robust strongly significantly positive effect. We postulated that this may be reflecting complementarity, conditionality, or institutional or capacity improvements possibly associated with such funds during the sample period. No health aid modality was found to have a significantly negative effect on total health spending. As Chapter 2 raised further aid recording issues and contested the assumption that received aid is ought to be spent in the current fiscal year, we also test for aid smoothing behaviour, and find little evidence to support such hypothesis. Acknowledging the incidence of missing data did not alter qualitative conclusions (although the size of estimated coefficients inevitably varies slightly).

To fully assess the sector fungibility of aid, one would need to be able to track aid funds available to other sectors (if the concern is over aid awarded to one sector being diverted onto spending in another sector; to some extent attempted by Van de Sijpe, 2013), and over time (due to potential spending lags). Furthermore, in all currently available health aid fungibility studies discussed in this chapter, tax (and other domestic) revenue constitutes an important omission as the government's domestically available funds are not controlled for. In this chapter, we maintained the estimated economic model as close as possible to the existing ones for comparability purposes. In the two following chapters we turn our attention to fiscal response analysis (cases studies), with the final Chapter 6 returning to the issue of omitting non-DAC aid flows.

Chapter 4

Fiscal Effects of Aid in Ethiopia:

Evidence from CVAR Applications[117]

1. Introduction

This chapter looks at fiscal dynamics in Ethiopia, particularly focusing on the fiscal effects of aid. Understanding of the fiscal effects of aid can be seen as a prerequisite to the analysis of the macroeconomics effects (and effectiveness) of aid. Since Osei et al. (2005), the Cointegrated Vector Auto – Regressive (CVAR) has been increasingly used for the analysis of fiscal effects of aid in individual country setting. We estimate a CVAR model including the following variables: government expenditure, disaggregated into recurrent and capital spending components, tax revenue, non-tax revenue, and aid, disaggregated into grants and loans. The VAR model is highly demanding of the data. Therefore we estimate two models with five variables at a time. Since our key interest is in whether aid has any adverse effects on tax revenue, the primary focus is on a model with disaggregated aid and revenue variables, and aggregated government spending. The alternative system then looks in more detail at the relationship between aid and public expenditure by disaggregating the latter into the capital and recurrent components, but aggregating government revenue, aiming to answer what is aid actually funding.

[117] This chapter is a result of collaboration with Giulia Mascagni, at the time PhD Candidate in Economics, University of Sussex, who provided the data.

The key advantage of this study over similar CVAR applications for Africa lies in our unique data. We use annual observations from 1960 to 2009 compiled by the Ethiopian Ministry of Finance and Economic Development (MoFED). Not only the series are longer than those used in most existing studies in the literature, but also they are obtained from a single domestic source. By using national data we are able to capture the recipient's measure of aid – what is effectively disbursed through the budget and the government is aware of. Therefore it is the component and the measure of aid most relevant for the analysis of its fiscal effects. The CVAR analysis is complemented by an in depth qualitative understanding of the Ethiopian context, which ensures sound model specification and sensible interpretation of estimated results.

Our findings are three-fold. Firstly, our results provide evidence for the existence of domestic budget equilibrium: government spending decisions in the long run are driven by domestic revenue, and this is continual across the three political regimes covered in our sample. Secondly, aid is positively associated with tax revenue[118], thus failing to provide evidence for a disincentive or substitution effects. Thirdly, aid is positively associated with public expenditure. The system with disaggregated spending components shows that aid grants exhibit a stronger positive association with capital expenditure than loans, despite the conventional expectation that the opposite should be the case because of the repayment requirement attached to aid loans. Aid components exhibit a more uniform effect on recurrent spending. In these equilibria, aid also exhibits adjustment behaviour, perhaps reflecting that donors' disbursement decisions are based on the recipient's fiscal behaviour, rather than aid driving fiscal trends. Interestingly, grants and loans do not seem to exhibit qualitatively differing effects on Ethiopian fiscal variables. Results are robust to alternative aggregation/disaggregation strategies and inclusions/exclusion of dummies.

The Chapter is structured as follows: Section 2 introduces the fiscal effects of aid and reviews the relevant literature to date; Section 3 discusses the quantitative data, the quantitative dataset and introduces some relevant aspects of Ethiopian fiscal history to provide some qualitative context. Section 4 describes the CVAR methodology, and summarises the misspecification tests and the determination of the cointegration rank. The long run structure is identified in subsections 4.3 and 4.4; subsection 4.5 briefly discusses the short run results, and Section 4.6 – the common driving trends in the model. The results for alternative system specification where total expenditure is disaggregated into recurrent

[118] Note that we use a measure of tax in (logged) levels rather than as a percentage of GDP, and therefore we cannot draw explicit conclusions on the effects of aid on tax *effort*.

and capital components are provided in section 5. Section 6 concludes. Additional information can be found in Appendix C.

2. Fiscal Effects of Aid: Framework, Hypotheses, and Applications

Last few decades of the research on fiscal effects of aid often relied on the seminal Heller (1975) framework of fiscal response models. The framework is based on the maximisation of the government's utility function, represented by deviations of actual fiscal aggregates from target levels. Criticisms (see Binh and McGillivray, 1993, among others) to this framework include both theoretical and empirical issues, such as equal treatment of overshooting and undershooting the government targets, or unavailability of the actual data on the government targets.

With the aim of overcoming the problems inherent to Heller's framework and single equation models, the Cointegrated Vector Auto – Regressive (CVAR) framework has attracted increased attention in the analysis of fiscal dynamics. CVAR offers several advantages: firstly, it does not require a strict theoretical economic structure but rather 'allows data speak freely' to discriminate between competing hypotheses or theories; secondly, it does not impose *a priori* assumptions and restrictions, such as residual normality or variable exogeneity, but allows to test for these in the dynamic multiple equation setting. However, since the estimation of simultaneous long– and short–run equations involves a large number of parameters, the CVAR ideally requires large samples, and this poses a challenge in the analysis of fiscal dynamics for developing countries. The framework is outlined in section 4 below.

Given the lack of robust economic theoretical framework, we use a simple government budget identity to equip ourselves with a set of hypotheses of fiscal effects of aid to be tested in the parsimonious CVAR model. The basic accounting identity of the budget simply states that all revenues plus borrowing must equal all expenditures:

$$TAX + NTAX + LOANS + GRANTS + BORROW = CAPEXP + RECEXP \tag{4.1}$$

where *TAX* denotes tax revenue, *NTAX* is non-tax revenue, *LOANS* are foreign aid loans, *GRANTS* denote foreign aid grants, *BORROW* is domestic (and, potentially, foreign non-concessional) borrowing, also possibly including any seignorage revenue, and *CAPEXP* and *RECEXP* are central government's capital and recurrent expenditure, respectively.

Together with the assumption of some government targets, previous literature of fiscal effects would often assume that aid is exogenous, putting some measure of tax (tax effort models) or expenditure variable (fungibility studies) on the left hand side. In this context, the CVAR framework has the clear advantage since it does not require any of these assumptions. The mechanism for the budget process does not have to be specified *a priori*, and therefore the CVAR can 'let the data speak' on the dynamics that drive the process, and discriminate against competing potential mechanisms.

Given the concessional nature of aid loans, (domestic) borrowing could be considered as the 'borrowing of last resort'[119]. Following such reasoning, the equation (4.1) can be rewritten as:

$$(TEXP) - (DOMREV + AID) = BORROW \tag{4.2}$$

where, for simplicity, we aggregated all variables into total domestic revenue (*DOMREV*), total government expenditure (*TEXP*) and aid (*AID*). Borrowing is then a function of interactions between domestically collected revenue, aid, and expenditure decisions[120]. In effect, borrowing is bounded in the long run at some sustainable level (for instance, one that ensures the feasibility of servicing the outstanding public debt). Viewing borrowing as a residual decision would allow regarding it as potentially a stationary process (see Figure 4.2), and therefore focus hypothesis testing on interactions between the remaining variables.

The equation (4.2) above suggests three main effects of aid. Firstly, we can expect a positive relation with expenditure – aid should be spent. Given the limited availability of data and contentious issue of what is a 'good' way to spend aid money, we will not delve into discussion of the fungibility of aid (see McGillivray and Morrissey (2000) for an overview of the debate). Our empirical focus is to simply test which – the aid or the government expenditure – adjust to the other if they form a long run equilibrium, and which spending component bears a stronger association with aid.

Secondly, aid can influence tax revenue. Several competing hypotheses can be formed about this potential relationship. Foreign aid may provide a politically cheaper source of revenue

119 This, of course, is debatable if one accepts that both commitment and disbursement of aid loans takes time.

120 Certainly, apart from the variables in the postulated system, each of the variables depends on other domestic (or foreign) factors; for instance, the revenue collection will depend on tax policy (e.g. tax base, rates, effectiveness of collection), whilst aid will in turn depend on the economic and political conditions in the donor country.

than taxation, and therefore discourage tax effort. This argument, in theory, is stronger for grants than for aid loans, as the latter – at least in theory – requires future repayments. On the other hand, aid may have a positive effect on tax revenue through its effect on income, expanding tax base, or strengthening tax administration or improving tax policies[121]. If the latter effect of aid on tax dominates the former, we would expect aid and tax to exhibit a positive long run association.

Finally, aid may not be all spent as additional public expenditure but also be used to decrease borrowing: since aid relaxes the domestic budget constraint (i.e. the budget identity excluding aid variables), the government could achieve the same level of expenditure with less borrowing. As we do not have the full series on domestic borrowing (see section 3), we are unable to test for this potential fiscal effect of aid.

Therefore, while the single equation 4.2, as accounting identity, would be expected to imply one cointegrating (or equilibrium) relationship between the variables, the economic perspective summarised in the paragraphs above would imply three cointegrating (equilibrium) relationships (see section 4 of this chapter). Firstly, a domestic budget relationship between expenditures and domestic revenue, where the government makes its spending decisions consistent with the planned revenue. Secondly, aid-spending relationship as described in the paragraph above. Finally, aid – tax relationship (if the spending and domestic revenue are cointegrated, and spending and aid cointegrated, so must aid and tax revenue). These three relationships would describe the inter-variable dynamics of equation 4.2, as such joint system of three cointegrating relationships would be expected to form a stationary system, with borrowing – the excluded variable – representing a stationary (adjusting) process. In a five variable system, this would imply two common trends driving the system (see section 4.6), with one potentially broadly representing the domestic agenda, and one – donor aid-driven processes.

In summary, the research questions we aim to explore are:

- Does the budget identity hold as an equilibrium relation in the long run?
- Is aid part of that long run equilibrium relation?
- Does aid discourage tax revenue?
- Does aid increase spending?

[121] Part of this effect will not be modelled directly as the recipient's measure of aid will exclude any non-cash aid components, and therefore ignore donors' staff's presence (expertise, consulting, etc.), for instance.

- Which components of spending are most affected by aid?
- Does aid heterogeneity (i.e. distinction between grants and loans) matter and what are the differences in the behaviour of the two aid components?

It is important to note that the CVAR methodology has also some caveats. First of all, the CVAR is very demanding on the data and therefore the number of variables should be as limited as possible to allow estimation and inference, particularly in small samples.[122] Secondly, the results are sensitive to specification choices (Lloyd et al. 2009). To address these concerns, we formulate two distinct models: one to focus on the tax – aid relationship; and one to examine which components of government spending are most affected by aid; and perform numerous robustness checks.

Similar empirical applications

Given the inconclusiveness of cross-country and/or panel evaluations of fiscal effects of aid, and as data improves, case-study approach is more often adopted in the literature. A number of authors have applied the CVAR analysis to investigate the fiscal effects of aid. Table 4.1 below outlines studies that applied the CVAR framework to developing country context (and one focusing on macroeconomic effects of aid), the length of their respective sample, variables employed and their data sources.

Since a large part of aid flows into the public budget, analysis of the fiscal effects of aid can be seen as a prerequisite to understanding the macroeconomic effectiveness of aid (McGillivray and Morrissey, 2000; Mavrotas, 2002). Juselius et al. (2011) look at the macroeconomic effects of aid in a sample of sub-Saharan African countries. They do not find any adverse macroeconomic effects of aid in Ethiopia, but do not consider the fiscal system.[123]

The common feature of the previous applications is that data come from various national and international sources and the time-series dimension is rather short. Also, given the aforementioned data requirements, the number of variables included in the models tends to be low, with a maximum of five. A notable exception is Martins (2010) who uses quarterly data (60 observations) to model a system of six fiscal variables. We favour of annual data for

[122] Also note that the CVAR lends itself to the "build-up of type I errors in a general-to-specific modelling strategy" (Lloyd et al., 2009:162).
[123] Several other studies also apply CVAR to developing countries, but their foci are beyond the fiscal system alone. These are, for instance, Mavrotas (2002), looking at the effect of aid on growth; M'Amanja et al. (2006) on aid, investment and growth in Kenya.

two reasons: firstly, quarterly data for Ethiopia are available but they are not as reliable as they are only compiled seriously after the introduction of Protection of Basic Services (PBS) project in 2005 when donors became more careful about monitoring and reporting; secondly, budget decisions are taken annually and intra-year dynamics do not necessarily add relevant information. Therefore, while we take Martins (2010) paper as a reference point, as it analyses Ethiopia, we depart from it both by using annual data and by exploiting deeper qualitative information about the country context.

Following from the submission of Osei et al. (2003), the set of studies by Overseas Development Institute (ODI) on Malawi, Uganda and Zambia all adopt the same approach of estimating a set of different models based on the CVAR methodology, including different sets of variables (although note that in Fagernas and Roberts (2004b) all variables were found to be stationary and only a simple VAR was implemented). For both Uganda and Malawi, Fagernas and Roberts (2004, 2004a) find that both grants and loans have the expected positive effect on total expenditure. They find no solid evidence that aid discourages tax effort in Malawi and that it may have reduced borrowing, and identify a positive long run effect of aid on domestic revenue in Uganda, with negligible effect on domestic borrowing. For Zambia, aid seems to be associated with weakened domestic revenue and increased borrowing.

Osei et al. (2003) focus on the impact of aid on fiscal policy in Ghana using two models: first using a measure of aggregate expenditure, and the second one further disaggregated into capital and recurrent expenditure. In both cases they provide support for strong exogeneity of foreign aid. Aid in Ghana is associated with beneficial policy responses: increased tax effort and decreased domestic borrowing, resulting into increased public spending. Results from the disaggregated system suggest that aid in Ghana is more strongly associated with current rather than capital expenditure, contrary to the evidence from Uganda and Malawi (Fagernas and Roberts, 2004a; Fagernas and Schurich, 2004).

Table 4.1: Summary of CVAR Fiscal Literature

Paper	Country	Obs	Variables	Data source
Fagernas and Schurich (2004)	Malawi	31	Fiscal	National, WDI, IMF
Fagernas and Roberts (2004a)	Uganda	26	Fiscal	National, IMF
Fagernas and Roberts (2004b)	Zambia	27	Fiscal	WDI, IMF-IFS, OECD-DAC
Osei et al. (2003)	Ghana	33	Fiscal	IMF, OECD-DAC
Bwire (2013)	Uganda	37	Fiscal	National, OECD-DAC
M'Amanja et al. (2005)	Kenya	39	Fiscal and growth	National
Lloyd et al. (2009)	19 LIC and MIC	30	Fiscal	WDI
Martins (2010)	Ethiopia	60*	Fiscal	National
Juselius et al. (2011)	36 SSA countries		Macro-economic	WDI, OECD-DAC, PWT

Note: '' denotes quarterly observations*

M'Amanja et al. (2005) relate fiscal variables to growth in Kenya using a measure of aid disaggregated into grants and loans. While grants are positively associated with growth in the long run, loans are required to finance fiscal deficits. Consequently, loans were found to have negative effects on growth in the long run. With weak significance of the effects of (low flows) of grants, authors find that "loans substitute for domestic tax effort to finance a fiscal deficit" and conclude that, in the case of Kenya is a potential obstacle to aid effectiveness and that grants seem to be a preferable aid modality than loans.

In Bwire's (2013) doctoral thesis on Uganda, aid is found to be associated with increased spending, increased tax revenue, and decreased domestic borrowing. Domestic revenue is found to be the main driver of spending plans, and the exogeneity of aid is not empirically supported.

Martins (2010) models fiscal dynamics in Ethiopia using disaggregated measures of expenditure and aid but leaving the domestic revenue aggregated (containing both tax and non-tax revenue); the system also includes domestic borrowing, but excludes a number of residual items.[124] For the period of 1993-2008, aid is found to be positively related to development expenditure, with aid adjusting to variations in expenditure and therefore suggesting that donors follow government's expenditure decisions by financing increased

[124] All papers exclude one or more variables from the analysis, most commonly domestic borrowing and/or non-tax revenue, to avoid estimating an identity.

expenditure. Domestic borrowing was found to be the most adjusting item, thus compensating for variations in both aid and other revenues. No evidence was found that aid may be discouraging tax revenue, and that government finances its expenditure in the order in line with Bwire's (2013) findings for Uganda, namely: domestic revenue, aid, and borrowing. While quarterly data allows increasing the sample size and conveniently considering only a period of relative political stability while preserving the number of observations, our approach is in favour of annual data instead: budget decisions are taken annually and intra-year dynamics do not necessarily add relevant information.

The discussed studies do not find much evidence for 'negative' fiscal effects of aid, but demonstrate that the underlying fiscal mechanisms differ across countries, and therefore justify a case-study approach. A general element emerging throughout CVAR applications is the importance of considering the country context, particularly when the number of observations is small. Knowledge of the historical and political context can help explain large residuals, and in designing the deterministic components of the CVAR, such as dummies and mean shifts related to country specific events. The next sections summarise our data, their advantages, and how the exploration of the qualitative context contribute to the quantitative set up.

3. Data and Qualitative Context

Data availability and reliability present a severe issue in African countries, and especially so its time-series dimension. Many African countries reached independence in the 1960s and only then did they start building national institutions, including statistical offices. Ethiopian case is different because the country has never experienced colonisation, but only a six-year invasion by Italian forces in 1935-1941. Upon his return, Emperor Haile Selassie embarked on a reform process with fiscal policy at its core, as public revenues were much needed for reconstruction and development. By the creation of the Central Statistical Office in 1961, Ethiopia had a well-established tradition in data collection with fiscal records dating back to 1949, and, in fact, today – together with South Africa – Ethiopia has the highest Statistical Capacity Rating in Africa (African Economic Outlook 2010)[125]. This is of particular importance for the CVAR analysis, as the approach is highly reliant on the data.

[125] http://www.africaneconomicoutlook.org/en/

Our dataset of 50 annual observations for the period of 1960 – 2009[126] was compiled in Ethiopia on the basis of MoFED data.[127] MoFED compiles National Accounts that may be then transferred to international institutions to apply the necessary modifications that make the data comparable across countries. Our choice to use national data has several advantages: firstly, the data series are consistent as they come from a single source, and thus avoid introducing any conversions or adjustments. Secondly, it is the one used for government decision making and is therefore relevant from a policy perspective. Finally – and crucially – it includes a measure of aid that represent the actual cash portion going through government's budget (other aid channels being through non-governmental organisations, or delivered in form of technical assistance or in-kind, to name a few).[128] Note that in the measure of budget aid we include not only budget support, but also other sources of aid that flow through the budget. In particular, budget support was withdrawn in 2005 due to the post-election tensions and has not been restored since; however, other types of aid were introduced, most notably a project called Protection of Basic Services (PBS). While PBS is a project, it flows through the budget and fully uses the country systems, thus exhibiting some similarities with general budget support.

While only budget aid is considered, we are still able to further disaggregate it into grants and loans. Such disaggregation is motivated by the expectation that these two types of aid would exhibit different effects because of the repayment requirement associated with loans. Furthermore, whilst grants are largely donor-determined, loans may be sought out by the recipient, especially when deficits are higher.

When disaggregating public expenditure into its capital and recurrent components, we maintain the government's original classification without any further manipulation. Whilst the distinction between development expenditure and pure government consumption along the lines of Martins (2010) is theoretically appealing, it is very difficult to credibly impute single expenditure items to either of the categories. An obvious example is the public sector salaries: some components may be considered a developmental expenditure (e.g. wages in health and education), but the wage bill is classified under recurrent.

[126] Data records follow Ethiopian fiscal years, following the Ethiopian Ge'ez calendar.
[127] While the data are of generally good quality, the quantitative data was reviewed during an extensive interview process (including with MoFED employees), and such qualitative information contributed to filling some prevailing data gaps, especially during the Derg period, were data were of poorer quality (see Mascagni, 2014).
[128] By omitting the off-budget aid, we ignore its potential indirect effects on government decisions.

Finally, we disaggregate domestic revenue into tax and non-tax revenue, and we show them to display different behaviour in Ethiopian fiscal dynamics. Unfortunately, we have to exclude domestic borrowing because full series for this variable is only available from 1974. Furthermore, given several negative values, its inclusion would be complicated given the log transformation applied to the data.

In the CVAR applications discussed above, variables are either analysed in levels (Bwire, 2013, Martins, 2010, Osei et al., 2003) or in logarithmic transformations (Juselius et al., 2011, M'Amanja et al., 2005). We explored both options and settled on data in logs because it was superior in terms of model fit, such as no autocorrelation and normality in the residuals (discussed in section 4). The log transformation requires all variables to be strictly positive. Since the first three years in the grants series are reported to be effectively zero, we discard first three years of observations, reducing the sample to the period of 1963 – 2009.[129]

The key variables are depicted in Figure 4.1 below. For presentation purposes, the data is expressed as proportion of GDP. During the sample period, Ethiopia experienced several important events, including two major political regime changes that are distinct in the data. In the beginning of our sample, Ethiopia was under the Imperial rule of Haile Selassie, who had ruled the country since 1930, with a six–year interruption due to the Italian invasion in 1935. After the 1974 revolution, the rule was assumed by the socialist military junta, known as Derg, which in turn was eventually replaced by the Ethiopian People's Revolutionary Democratic Front (EPRDF) that is still in power today. A brief description of each regime (in terms of key fiscal components) is provided in Table 4.2.[130] Table 4.3 provides regime averages for the key variables. Consulting fiscal history, as well as economic and political calendar, complements econometric results and ensure that the interpretation of quantitative results is realistic.

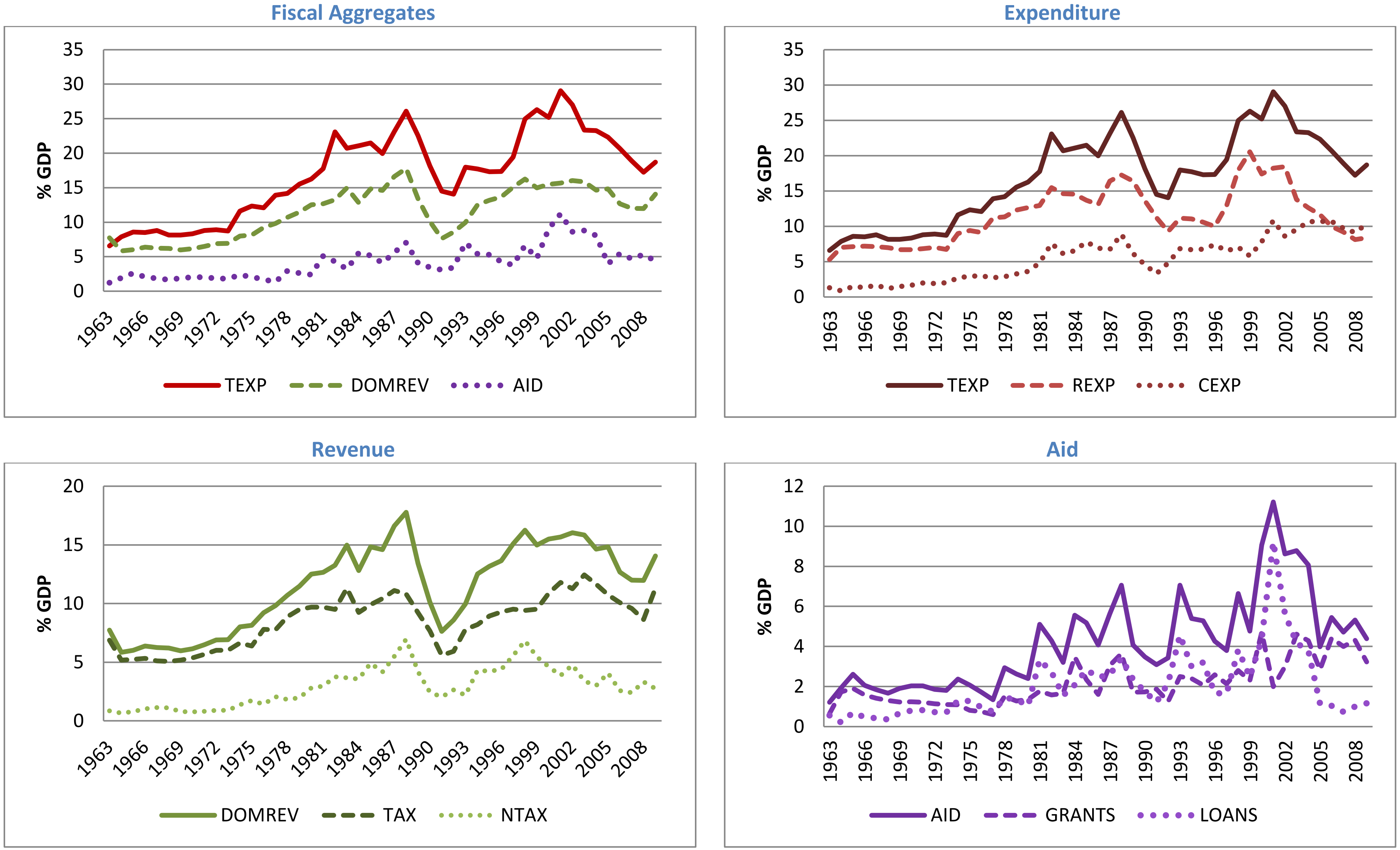
The first 11 years of data depict the feudal structure of the imperial rule. With underdeveloped "modern" sectors and little revenue obtained in rural areas through direct taxation (mainly due to widespread tax evasion), the imperial regime relied heavily on trade and indirect taxation. The sustained revenue mobilisation during this period was mainly

[129] We specifically do not express the variables as proportion of the GDP. Whilst it would allow to facilitate the interpretation (both statistic and economic), as well as international comparisons, it may introduce more measurement error, given the difficulties related to GDP accounting (see, for instance, Jerven, 2013).

[130] A more verbose description of the key events and policies is available on request; an even more extended version can be found in Mascagni (2014) PhD thesis at the University of Sussex.

driven by two elements of the government spending: expansion of the military and civilian bureaucracy. A large army (biggest military force in black Africa by 1960) was needed to address the tensions over borders, primarily with Eritrea (fully annexed by Ethiopia in 1962), but also with Somalia over the Ogaden region. The bureaucratic apparatus was expanding to meet increasing administrative and economic functions: the development planning – a central issue in the international debate in 1950s – included a succession five-year development plans. The importance of foreign investment was recognised by the government: fiscal incentives included exemptions from business income tax, duty free imports, and guarantees regarding the possibility of remitting a proportion of profits. Given Ethiopia's unique independence from colonial power, aid flows were comparatively low, with loans and grants contributing about a fifth of total expenditure on average. Although the US was the key ally and donor, Ethiopia could also turn for assistance to Federal Republic of Germany, Sweden, Czechoslovakia, Yugoslavia, and even the Soviet Union, as well as multilateral organisations such as the World Bank and UN missions. Finally, since mid-1960s, government systematically used domestic borrowing to finance its budget deficit, although still less than in subsequent regimes.

Figure 4.1: Data of Key Aggregated and Disaggregated Variables

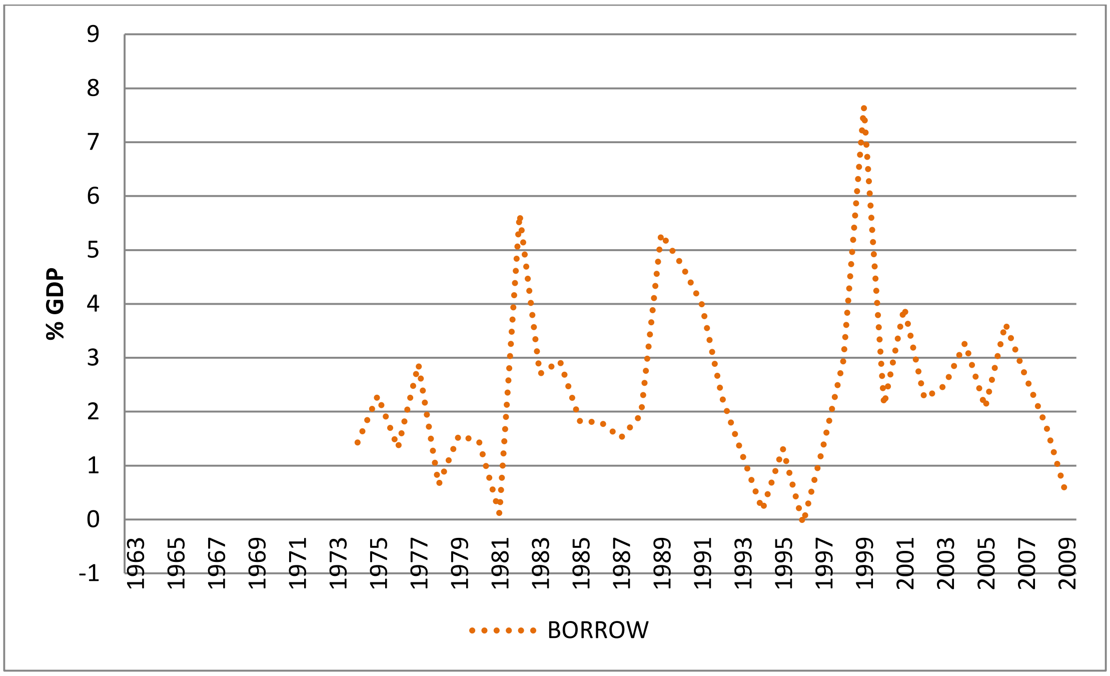


After the 1974 socialist revolution and the establishment of the socialist military junta (Derg), the growth in state expenditure was at increasing imbalance with the growth of the economy. The public expenditures consistently shifted away from development and services (capital expenditure) towards control functions, including military expenditure (recurrent spending). Thus, the continuity with the Imperial regime became increasingly clear, at least in terms of strong state, repressive political apparatus and lack of independent institutions. Domestic revenue mobilisation remained a priority. Tax policy during the Derg mainly relied on: direct taxation with high marginal rates[131] (personal incomes, and commercial profits, including those from state-owned enterprises); agricultural taxation (income and land use); and trade taxes (dominated by the revenues from taxes on exports). Although initially increasing, the tax revenue eventually declined, mainly due to shrinking tax base, and widespread avoidance and evasion. Non-tax revenues were also increased, initially due to expropriations and nationalisation, then through retaining the profits from state enterprises (that were already heavily taxed), and, towards the end of the regime, transfers from National Bank of Ethiopia as it had a large amount of accumulated reserves in domestic currency that were unused. Borrowing (**Figure 4.2**), both domestic and foreign, was also increasing to close the government resource gap; the situation deteriorated considerably towards the end of the regime (as the government was scaling up its military spending) and in 1990 payments of all debt obligations were frozen, except for those to international financial institutions and other critical ones. American aid (except its humanitarian component) was fully withdrawn from Ethiopia by 1977, mainly due to the uncompensated expropriation of American private assets. USSR became the largest foreign actor and the Western donors had little political leverage in the county. Nevertheless, trade mostly occurred with the western partners (Europe, the US, and Japan); Ethiopia was part of the Lomé agreement with European community, and thus a substantial fraction of aid was still Western. Foreign investment, however, decreased sharply, mainly due to non-compensatory expropriations and restrictions to private initiative. The end of the regime, 1989-1991 period, was described by deteriorating economic, military[132], and political situation in the country, as a decade of poor economic policies – increasing war effort, overextension of the state, the lack of investment, and deterioration of the terms of trade - resulted in economic crisis, accompanied by fiscal collapse (see **Figure 4.1**).

[131] Up to 89%.

[132] By 1990, reportedly, “the conflicts in the north were consuming more than two-thirds of Ethiopia’s annual budget” (Keller, 1992)

Figure 4.2: Public Borrowing



Coming to power in 1991 after 16 years of armed struggle, EPRDF embarked upon significant liberalisation and privatisation programmes, establishment of ethnic federalism and accompanied decentralisation, capacity building, and institution development, and the revenues and expenditure recovered rather quickly. As these efforts were supported by the Western donors (especially through Structural Adjustment Programmes (SAPs) in the 1990s)[133], the (budget) aid to Ethiopia increased substantially and consistently during the EPRDF period, with Ethiopia now being considered an 'aid darling' (although it still receives less official development assistance (ODA) per capita than most of other African countries). Both grants and loans each averaged to about 3% of GDP during the period, and aid dependency (expressed as total budget aid as a proportion of government expenditure) increased to about 28%. However, aid remains the most volatile source of revenue (Figure 4.1). The fiscal situation inherited from the Derg was disastrous: the revenue was at pre-revolution levels, and the level of debt exceeded 100% of GDP. However, EPRDF rapidly improved revenue performance, with the growth rate of revenue of 36% in 1992 already, and sustained at double rates until the Eritrean war. The deficit was also consistently decreased, with government budget producing surpluses in 1994-1996. The limited capacity for tax reform in the 1990s was further complicated by the 1993 secession of Eritrea. A

[133] Although these were interrupted by the 'peace conditionalities' during the armed conflict with Eritrea (1998-2001, the latter year also coinciding with a drought). Furthermore, the general budget support was withdrawn following the aftermath of 2005 election, but was soon substituted by the Protection of Basic Services project aid, flowing through the budget. The decrease in loans in the 2000s is due to debt relief under HIPC initiative.

major tax reform was eventually carried out in 2002 and represented a great effort of revenue mobilisation, which was falling short of the needs stemming from the administrative reforms, decentralisations and the re-militarisation of the late 1990s. IMF, along with other donors, played a crucial role in supporting the tax reform. Despite the expectations of annual tax growth rate of 24% on average (against the predicted GDP growth rate of 11%), the limitations to tax revenue mobilisation, such as low income and large share (40% GDP) of the agricultural sector, with further capacity and compliance constraints, remain. Supported by the emergence of developmental state, reflected especially in 2005 Plan for Accelerated and Sustained Development to End Poverty (PASDEP) and 2010 Growth and Transformation Plan (GTP), the (per capita) GDP growth rates also recovered following the stagnation and deterioration of the Derg regime. In fact, the post-2003 trends (Figure 4.1) are described by fast GDP growth rates rather than actually deteriorating fiscal effort. Aid grants have also been consistently increasing under EPRDF, except for the period of war with Eritrea.

Table 4.2: Key Qualitative 'Summary Statistics'[134]

	Imperial Period (1963-1974)	**Derg** (1974-1991)	**EPRDF** (1991-today)
Key description:	**Feudal system** *Oppressive, inequality, large state*	**Socialist military junta** *Nationalisation, repressive, large state*	**Ethnic federalism** *Liberalisation and privatisation (not land); 'free markets' and centralist state.*
Expenditure:	Expansion of military and civilian bureaucracy. Development plans' implementation limited.	Further expansion of state control. Development plans' implementation limited.	Development: commitment to poverty reduction (state-led). Demilitarisation – remilitarisation.
Tax (and non-tax):	Rely heavily on indirect and trade taxes. Avoidance and evasion.	Coercion. Extraction through non-tax revenue (expropriations; profits).	Actual reforms since 2002 (IMF); VAT; enforcement; rapid growth.
Borrowing:	Systematic but modest	Increasing.	'Relaxed' w.r.t. Infrastructure investment.
Aid:	No 'patron' (low flow); US, other bilateral and multilateral	USSR; other bilateral and multilateral.	Increasing but volatile; SAPs; peace conditionalities; PBS post-2005; HIPC. Strategic. 'Non-traditional donors'.
Foreign inv.:	Encouraged	Expropriated	Limited

[134] Based on Mascagni (2014) PhD thesis submitted to University of Sussex.

Table 4.3: Regime Averages of Selected Variables

Indicator	Description	Imperial	Derg	EPRDF
GDP pc	Deflated nominal	1001.2	995.8	1141.9
Agriculture	%GDP	71.3	59.3	49.9
Manufacturing	%GDP	3.7	5.5	5.1
Trade openness	%(Imports+Exports)/GDP)	10.5	13.4	29.4
Fiscal pressure on trade	%(Trade tax/(Imports+Exports))	21.5	23.1	14.8
Tax revenue	**%GDP**	**5.3**	**9.2**	**9.6**
Tax revenue growth	Annual change	11.5	8.7	**18.5**
Non tax revenue	%GDP	0.9	3.3	3.8
Fiscal Deficit	**%GDP**	**-**	**-2.1**	**-1.4**
Expenditure	%GDP	8.2	18.2	20.8
Grants[135]	**%GDP**	**1.0**	**1.8**	**3.0**
Loans	**%GDP**	**0.7**	**1.9**	**2.9**
Aid dependency	%(aid/expenditure)	21.1	19.7	28.2

4. Econometric Framework: CVAR[136] and Results

Model with Disaggregated Aid, Disaggregated Revenue, and Aggregated Spending

Using the data described in Section 3, we model a five-dimensional vector autoregressive model that includes central government's total expenditure (***texp***); domestic revenue, disaggregated into the tax and non-tax revenue components (***tax*** and ***non-tax***, respectively); and budget aid disaggregated into ***grants*** and ***loans***. Variables are transformed using natural logarithms.[137]

In the VAR framework, each variable is modelled as endogenous, and is expressed as a function of past own values, as well as past realisations of other variables (and deterministic components). The vector error-correction model (VECM) representation of the VAR includes both the stationary first differences of variables in x_t (Δx_t), and their value in levels (x_t), thus preserving both the long-run and short-run information in the data. In particular, the error correction form of the VAR (VECM) is represented by the following equation:

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi D_t + \varepsilon_t \quad (4.3)$$

[135] Note that aid figures reflect budget aid only.
[136] This section relies heavily on Juselius (2006). Analysis is conducted using CATS (Hansen and Juselius, 1995).
[137] In aggregate specification of Chapter 6, the three-dimensional VAR is used, where $x_t = texp_t, domrev_t, aid_t$.

$x_t = texp_t, tax_t, nontax_t, grants_t, loans_t$

where x_t is a $\mathrm{p} \times 1$ vector of endogenous variables described above, D_t is a vector of deterministic components (such as constant, deterministic trend, and dummy variables) with a vector of coefficients Φ; k denotes the selected lag length; ε_t is a $\mathrm{p} \times 1$ vector of unobservable error terms, that are assumed to be $\varepsilon_t \sim IN(0, \Omega)$. VECM allows a clear separation between the long-run coefficients in Π and the short-run coefficients in Γ_i.

The VECM representation illustrates that if variables are found to be I(1) – and macroeconomic variables usually are – stationary variables (Δx_t) are regressed on unit-root processes (x_{t-1}). In such case, the estimated coefficients would be spurious. However, if some variables in the system are driven by the same persistent shocks, there may exist linear combinations of these variables that are integrated of the lower order than the variables themselves (i.e. I(0)). These linear combinations would represent cointegrated relations, $\beta' x_t$, and could be interpreted as the long-run steady-state relationships. When cointegration exists, Π has reduced rank $r < p$ and is defined as follows:

$$\Pi = \alpha\beta' \tag{4.4}$$

where α and β are $p \times r$ matrices (with $r < p$); $\beta' x_t$ defines the stationary long-run cointegrating relationships $(r \times 1)$, and α denotes the adjustment coefficients to the equilibrium error. Intuitively, if all $x_t \sim I(1)$ and $\Delta x_t \sim I(0)$, then a full rank in Π would be logically inconsistent as it would imply that x_t must be stationary.[138] On the other hand, $r = 0$ implies that each variable in x_t is non-stationary and is driven by its own individual stochastic trend and therefore no cointegration exists. In this case, a simple VAR model with the variables in first differences would not imply any loss in long-run information.

The accompanying moving average (MA) representation of the VAR illustrates how the process can be described in terms of pulling and pushing forces. The steady state to which the process is pulled to is defined by the long run relations $\beta' x_t - \beta_0 = 0$. The forces α represent adjustment and they activate as soon as the process is out of steady state, i.e. when $\beta' x_t - \beta_0 \neq 0$.[139] The MA representation describes the non-stationary movement of

[138] The VECM representation of the VAR with full rank in Π and x_t~I(1) would imply that a stationary variable Δx_t equals a non-stationary variable x_{t-1}, lagged stationary variables Δx_{t-1} and a stationary error term. Since a stationary variable cannot equal a non-stationary variable, either Π=0 or it would have reduced rank.

[139] Juselius (2006:88-89).

the variables according to the common driving trends that represent the cumulated sum of the shocks to the system. "In this sense, the AR and MA representation are two sides of the same coin: the pulling and the pushing forces of the system" (Juselius, 2006:88). The inverted model can be summarised as:

$$x_t = C\sum_{i=1}^{t}(\varepsilon_i + \Phi D_i) + C^*(L)(\varepsilon_i + \Phi D_i) + X_o \qquad (4.5)$$

where $C = \beta_\perp(\alpha'_\perp(I-\Gamma_1)\beta_\perp)^{-1}\alpha'_\perp$ is the long-run impact matrix of rank *p-r*, with $\alpha'_\perp\varepsilon_t$ describing the common driving trends; $C^*(L)$ is a stationary lag polynomial, and X_o depends on the initial values.

4.1 Misspecification Tests[140]

This section discusses the formal misspecification tests and the corresponding results from our model. These tests aim to assess the validity of the assumptions underlying the statistical VAR model. The misspecification tests are also a helpful tool in guiding the analyst to correct model specification. Note that the residual autocorrelation tests and the ARCH tests are derived under the assumption of normally distributed errors and the normality tests are derived under the assumption of independent and homoscedastic errors and the lag length criteria are only valid under the assumption correctly specified model.[141] The misspecification test procedure, in search of the correctly specified model, therefore is rather iterative. The results discussed below refer to the final specified five-dimensional VAR(*k=2*) model with an unrestricted constant (allowed to cumulated to a drift in levels). We tested a model allowing for a trend in cointegrating relationships (which would cumulate to quadratic trends in data in (logged) levels), but the tests suggested it can be excluded from the cointegrating space. See Juselius (2006, Chapter 6) for a detailed discussion of deterministic components in the I(1) model.

The political regime changes are modelled as shift dummies in 1974 (from emperor to Derg) and 1991 (from Derg to EPRDF), taking form of $D_t = (\dots 0,0,0,1,1,1,\dots)$, allowing for mean changes in equilibrium relationships. Whilst the 1991 shift dummy is 'required' by the data as an outlier, the 1974 one is not. However, we model it for consistency with the qualitative data. Results do not hinge on the exclusion/inclusion of this dummy. The model estimated

[140] This section relies on Chapter 4 of Juselius (2006).
[141] Juselius (2006:77-71).

over the period of 1963-2009.[142] The testing results in the sub-sections below refer to the unrestricted VAR (UVAR).

4.1.1 Lag Length Determination

We employ the standard lag length determination procedure, which relies on three information criteria and the likelihood ratio (LR) lag reduction test. (Note that the criteria for the lag length selection are only valid under the assumption of correctly specified model).[143] The sequential LR tests for the lag length determination can be formulated as:

$$-2lnQ\left({}^{\mathrm{H_k}}/_{\mathrm{H_{k+1}}}\right) = T\left(ln|\widehat{\Omega}_{\mathrm{k}}| - ln|\widehat{\Omega}_{\mathrm{k+1}}|\right) \qquad (4.6)$$

where $\mathrm{H_k}$ denotes the null hypothesis of lag truncation at k against the alternative hypothesis of $k+1$; T defines the effective[144] sample size; $\widehat{\Omega}$ is the residual covariance matrix. The test is asymptotically distributed as χ^2 with p^2 degrees of freedom.

Results are summarised in the upper part of Table 4.4. With maximum lag length of five selected (and hence effective sample of 1968-2009) the lag length reduction tests suggest the lag length of five. However, note that no penalising factor is applied in LR tests; also, small samples often suffer from size issues. The two information criteria defined below include a (different) penalising factor related to the number of estimated parameters.

Schwartz information criterion (SC):

$$SC = ln|\widehat{\Omega}| + (p^2k)\frac{lnT}{T} \qquad (4.7)$$

Hannan-Quinn information criterion (H-Q):

$$H-Q = ln|\widehat{\Omega}| + (p^2k)\frac{2lnlnT}{T} \qquad (4.8)$$

As reported in the lower panel of Table 4.4, SC suggests $k=1$, whilst and H-Q favours $k=3$. However, the LM tests indicate some potential 'left-over' residual autocorrelation at lag length of one. Indeed, the residuals from an equivalent five-dimensional UVAR (k=1) indicate inferior model specification as the multivariate normality is rejected, as is the null of

[142] The notation here is shorthand for Ethiopian Ge'ez calendar years of 1963/64 through 2009/10.
[143] Juselius (2006:71).
[144] The effective number of observations must be identical when testing $\mathrm{H_k}$ against $\mathrm{H_{k+1}}$ and hence is defined by the longest lag length selected by the analyst. This also holds for the following discussion of the information criteria.

homoscedastic errors of order one and two. Since the lag-length of two seems to provide a better description of the data generating process, we select $k = 2$.[145]

Table 4.4: Lag Length Determination (Effective Sample: 1968 to 2009)

Lag Reduction Tests							
VAR(4)	<<	VAR(5)	:	ChiSqr(25)	=	85.214	[0.000]
VAR(3)	<<	VAR(5)	:	ChiSqr(50)	=	124.005	[0.000]
VAR(3)	<<	VAR(4)	:	ChiSqr(25)	=	38.792	[0.039]
VAR(2)	<<	VAR(5)	:	ChiSqr(75)	=	198.140	[0.000]
VAR(2)	<<	VAR(4)	:	ChiSqr(50)	=	112.927	[0.000]
VAR(2)	<<	VAR(3)	:	ChiSqr(25)	=	74.135	[0.000]
VAR(1)	<<	VAR(5)	:	ChiSqr(100)	=	259.290	[0.000]
VAR(1)	<<	VAR(4)	:	ChiSqr(75)	=	174.076	[0.000]
VAR(1)	<<	VAR(3)	:	ChiSqr(50)	=	135.284	[0.000]
VAR(1)	<<	VAR(2)	:	ChiSqr(25)	=	61.149	[0.000]

Model	k	T	Regr.	Log-lik	SC	H-Q	LM(1)	LM(k)
VAR(5)	5	42	30	513.591	-11.11	-15.04	0.149	0.579
VAR(4)	4	42	25	470.985	-11.30	-14.58	0.678	0.161
VAR(3)	3	42	20	451.589	-12.61	-15.23	0.315	0.483
VAR(2)	2	42	15	414.521	-13.07	-15.03	0.268	0.143
VAR(1)	1	42	10	383.947	-13.83	-15.14	0.055	0.055

Effective Sample: 1968:01 to 2019:01
SC : Schwarz Criterion; H-Q : Hannan-Quinn Criterion
LM(k): LM-Test for autocorrelation of order k

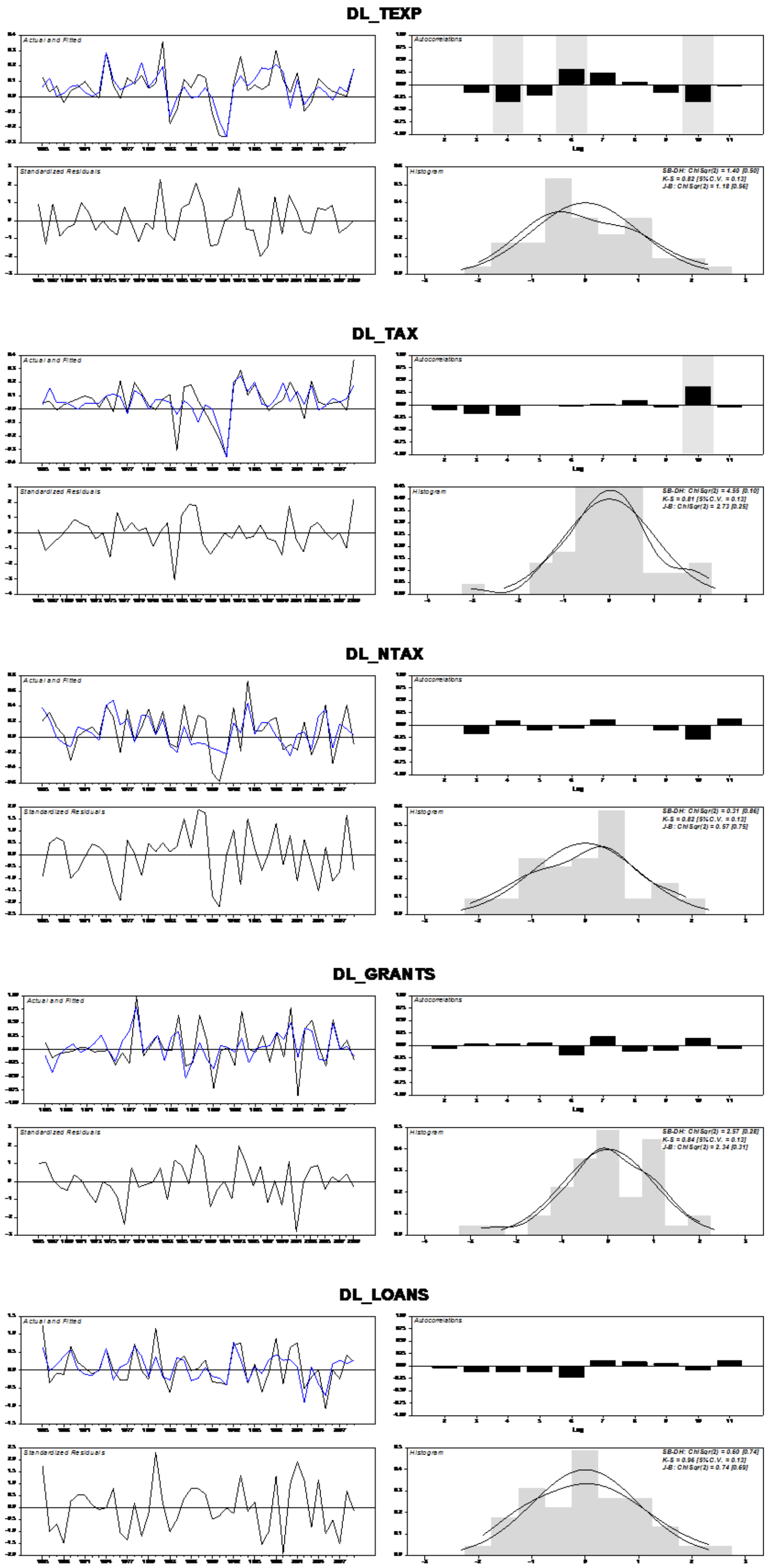
4.1.2 Residual Plots (Unrestricted VAR)

Whilst the output of the test results is itself informative, the graphical analysis available in the software packages (both RATS/CATS, and PcGive) often provide additional information, possibly revealing specification problems that test results fail to discover (Juselius, 2006:66). This is especially relevant in our study, as the sample is rather small from the time-series perspective.

For each equation, the Figure 4.3 below shows: (i) the plots of actual and fitted values of $\Delta x_{i,t}$, i=1,...,p (top left panel); (ii) the autocorrelogram of order 11 (top right panel); (iii) the standardised residuals (bottom left panel); and (iv) the empirical and normal distributions (bottom right panel). The graphs do not signal any particular issues.

[145] Note that Juselius (2006:72) argues that "a lag length of two is in most cases sufficient to describe a very rich dynamic structure even in a small-dimensional system".

Figure 4.3: Residual Plots (Unrestricted VAR)



4.1.3 Residual Autocorrelation, Heteroskedasticity, Normality, and Goodness of Fit

As the VAR methodology is based on the idea of decomposing the variation in the data into a systematic part describing the dynamics in the model and an unsystematic random part, the assumption of uncorrelated residuals (and hence this test) is an important one. The χ^2 and F test are derived under the assumption of independent errors; the violation of this assumption would result in the distribution of these tests deviating from χ^2 and F in unknown ways Juselius (2006:74).

The key test used to detect residual autocorrelation is the LM test of j[th] order autocorrelation, calculated using an auxiliary regression of estimated VAR residuals, $\widehat{\varepsilon_t}$, on the k lagged variables, $\mathbf{x_{t-1}}, \mathbf{x_{t-2}}, \ldots, \mathbf{x_{t-k}}$, and the j[th] lagged VAR residual, $\widehat{\varepsilon_{t-j}}$:

$$\widehat{\varepsilon_t} = \boldsymbol{A_1 x_{t-1}} + \boldsymbol{A_2 x_{t-2}} + \cdots + \boldsymbol{A_k x_{t-k}} + \boldsymbol{A_\varepsilon}\widehat{\varepsilon_{t-j}} + \widetilde{\varepsilon_t} \tag{4.9}$$

where the first *j* missing values $\widehat{\varepsilon_{-j}}, \ldots, \widehat{\varepsilon_{-1}}$ are set to 0. The LM test is calculated as a Wilks' ratio test with a small-sample correction:

$$LM(j) = -\left(T - p(k+1) - \frac{1}{2}\right) ln\left(\frac{|\widehat{\Omega}(j)|}{|\widehat{\Omega}|}\right) \tag{4.10}$$

The test is approximately distributed with χ^2 with p^2 degrees of freedom; H_o assumes no (left-over) autocorrelation in the residuals. In Table 4.5 we report the test p-values until order four. The null of no autocorrelation is not rejected at any order.

Table 4.5: Misspecification Tests

Residual normality *(p-values)*						
	Multivariate	**Univariate**				
		texp	***tax***	***non-tax***	***grants***	***loans***
	0.111	0.496	0.103	0.856	0.277	0.741
Residual autocorrelation and ARCH effects *(p-values)*						
		LM(1)	LM(2)	LM(3)	LM(4)	
Residual autocorrelation		0.190	0.225	0.156	0.864	
ARCH		0.606	0.131	0.340	1.000	
Trace correlation	0.518					

Note: All values are p-values.

To test[146] for residual heteroskedasticity is the m[th] order ARCH test, calculated as $(T + k - m)xR^2$, where T is the total sample zise, k is the lag length of the VAR, and R^2 is from auxiliary regression:

$$\varepsilon_{i,t}^2 = \gamma_0 + \sum_{j=1}^{m} \gamma_j \varepsilon_{i,t-j}^2 + error \tag{4.11}$$

The test is approximately distributed as χ^2 with m degrees of freedom, and the H_o assumes homoscedastic errors. The results in Table 4.5 demonstrate that no ARCH effects were detected at multivariate or univariate level, respectively. (The null of homoscedastic errors cannot be rejected)[147]. Note that the conditional heterogeneity testing is a standard procedure in CATS software, and no unconditional heteroskedasticity tests were performed. Also note that ARCH effects are more relevant when using financial data, rather than our small sample of annual fiscal data.

The normality (univariate and multivariate) tests discussed in this section are based on the Shenton-Bowman transformation. The reported multivariate normality test is that suggested in Hansen and Doornik (1996). The multivariate normality is not rejected at 10% level (see Table 4.5). Juselius (2006:76-77) notes that VAR estimates are more sensitive to deviations from normality due to skewness than to excess kurtosis. The results reported in the Appendix Table C1 do not indicate any particular departures from normality in terms of skewness (expected to be around 0) or kurtosis (expected to be around 3).

The measure for goodness of fit in the VAR model is the trace correlation, defined as:

$$Trace\ correlation = 1 - trace\left(\widehat{\Omega}[Cov(\Delta x_t)]^{-1}\right)/p \tag{4.12}$$

Roughly interpreted as R^2 in the linear regression model, the trace correlation for our data is 0.518 (Table 4.5). CATS also calculates an R^2 for each equation, $i = 1, \dots, p$, for the models in ECM form:

$$R_i^2 = 1 - \widehat{\Omega}_{ii}/Var\Delta x_{i,t} \tag{4.13}$$

where $\widehat{\Omega}_{ii}$ is the estimated residual variance of equation i. When the variables are integrated of order one, the R^2 is only meaningful when the dependent variable is given as

[146] Juselius (2006:74).

[147] Rahbek et al. (2002) simulation results have demonstrated that the cointegration rank tests are robust against moderate residual ARCH effects. (Juselius, 2006:75).

$\Delta x_{i,t}$, in which case the R^2 measures the explanatory power of the regressor variables as compared to the random walk model. [148] The results are reported in the Appendix Table C1.

4.1.4 Parameter Constancy Tests (UVAR)

Since parameter constancy is an important feature of the model[149], we report the results from a battery of tests (for both full and reduced model) in the appendix. The tests for R-form are more likely to accept the parameter constancy of the long-run as the effects of the non-constant parameters are averaged out. Meanwhile, the X-form[150] tests are more likely to be influenced by the instability of the parameters of the short-run structure. The two are also more likely to differ where the baseline sample is very short (Juselius, 2006:150). The parameter constancy tests, reported in the Appendix Table C2, do not show evidence of non-constant parameters.

4.2 Determination of Cointegration Rank

The determination of cointegration rank, *r,* is crucial in the CVAR analysis, as it influences all the subsequent econometric analysis by dividing the data into r pulling and $p-r$ pushing forces, corresponding to, respectively, equilibrium relations and common driving trends. In other words, the testing procedure aims to discriminate between the stationary (equilibrium) and the non-stationary relations.

The choice of cointegration rank is usually a difficult decision and in the context of developing countries it is aggravated by small samples. It is therefore preferable to consider additional information in addition to the formal testing procedure (Juselius, 2006:131). In the next paragraphs we consider all the available information for determining the cointegration rank.

The Johansen test, also called the trace test (Table 4.6), is the formal test procedure. It is based on the concentrated form of the VAR model (or R-form), where all short-run dynamics and deterministic components are concentrated out using the Frisch-Waugh theorem.[151] The

[148] Juselius (2006:73).
[149] "Simulation studies have shown that valid statistical inference is sensitive to violation of some of the assumptions, such as parameter non-constancy, autocorrelated residuals (the higher, the worse) and skewed residuals, while quite robust to others, such as excess kurtosis and residual heteroskedasticity" (Juselius, 2006:47).
[150] Also note that "because X-form version re-estimates all parameters, the degrees of freedom are fewer than for the R-form" (Juselius, 2006: 150). Where one is continuing with a model for which the recursive tests signal non-constancies (i.e. one did not choose to re-specify the model), the estimated parameters will measure average effects.
[151] For more details see Juselius (2006:116-117, 131-145).

procedure is to test the hypothesis $H_r: rank = r$, implying that there are at least $p - r$ unit roots and r cointegrating relations. If the test statistic exceeds the critical value, we reject the hypothesis of $p - r$ unit roots and r cointegrating relations, and conclude that there are fewer unit roots and more cointegrating relations in the model.

The distribution of this likelihood-ratio test is non-standard and it is influenced by the deterministic components of the VAR model. It therefore has to be simulated using our specified model (to account for the step dummies) in order to obtain critical values (reported in the Appendix Table C3). In addition, Juselius (2006:140-141) argues that in small samples the asymptotic distributions are generally a poor approximation to the true distributions and can therefore result in substantial size and power distortions. Therefore we apply the small sample Bartlett corrections to the trace statistic (see Johansen, 2002) that ensure a correct test size.

The uncorrected trace statistic allows accepting the hypothesis that there are two unit roots $(p - r)$ and three stationary relations (r), thus suggesting a rank of three $(r = 3)$. The Bartlett-corrected values may suggest three unit roots $(p - r)$ and two cointegrating relations (r), thus a rank of two $(r = 2)$. However it is only possible to accept this hypothesis with a borderline *p*-value of 0.062. Juselius (2006:145) suggests that in small samples it is better to avoid choosing the rank based on small p-values close to the 5% threshold and it therefore imposes some caution in accepting $r = 2$, whereas $r = 3$ would be a safest option.

Table 4.6: Rank Test

p-r	r	Eig. Value	Trace	Trace*	Frac95	P-value	P-value*
5	0	0.561	102.992	89.83	75.45	0.000	0.003
4	1	0.459	65.973	57.49	54.15	0.003	0.025
3	2	0.364	38.306	34.60	35.87	0.026	0.062
2	3	0.265	17.927	15.41	19.08	0.075	0.155
1	4	0.087	4.095	3.55	5.86	0.123	0.163

** denotes Bartlett corrections*

Juselius (2006:48-52, 131-145) suggests considering four additional pieces of information when deciding the cointegration rank: the characteristic roots of the model, the t-values of the α coefficients of unrestricted VAR, the recursive graphs of the trace statistic, and the graphs of the cointegrating relations (as well as economic interpretability of the results). Such information, reported in the Appendix Table C4, seems to support the choice of $r = 3$.

This choice is also confirmed by the parameter constancy tests of the model with $r = 3$, that do not signal any particular problem (Juselius, 2006:145).

Firstly, if the $(r+1)^{th}$ cointegrating vector is non-stationary and wrongly included in the model, then the largest characteristic root will be close to the unit circle). With $r = 3$, the modulus of largest characteristic root is 0.690. In such small sample it is difficult to make a sharp distinction between unit roots, near unit roots, and 'very stationary' roots (Juselius, 2006:145).[152] Secondly, if all of t-statistics of α coefficients of the $(r+1)^{th}$ cointegrating vector are small, say less than 2.6, then one would not gain a lot by including the $(r+1)^{th}$ vector as a cointegrating relation in the model" (Juselius, 2006:142). All first three alpha vectors from the unrestricted VAR have significant coefficients (with two significantly adjusting coefficients in the third alpha vector), whilst not really in the fourth. Thirdly, the recursively calculated components of the trace statistic should grow linearly for all i=1, ..., r, but stay constant form i=r+1, ..., p. The concentrated model illustrates that three components of the trace test statistic can be said to be growing linearly,[153] while the remaining two exhibit some volatile behaviour. Fourthly, the graphs of cointegrating relations should not reflect distinctly non-stationary behaviour; if they do, the choice of *r* should be reconsidered, as the model specification may be incorrect.[154] The final panel of **Appendix Table C4** illustrate that whilst three (first (*texp*), second (*tax*), and fourth (*grants*)) relationships do look rather stationary, two of them do less so.

Finally, in addition to the statistical tests to determine the rank of Π, it is crucial to ensure that the resulting equilibrium relations are economically interpretable. Following the discussion of Ethiopian qualitative data (Section 3) and broader literature (Section 2), we may expect to find the following three equilibrium relations:

1. A domestic budget equilibrium**,** where the government makes its spending decisions consistent with the planned domestic revenue. Whether aid is part of this equilibrium can and will be tested.
2. A relationship between government spending and aid, which we can expect to be positive. Formulating an equilibrium relation between these variables would also allow to test hypotheses about aid spending and to identify the adjusting variables.

[152] The modulus of largest characteristic root with r=2 is 0.584.
[153] Note that the unit root rejection line should be shifted from 1 to approximately 1.25 to account for Bartlett correction and the effect of a shift dummy (Juselius, 2006:145).
[154] Juselius (2006:142).

In particular, it is interesting to test whether it is government expenditure or aid that adjusts to deviations from such equilibrium relationship.

3. A relation between aid variables and tax revenue. If such a long-run relation exists, it would be possible to test whether a disincentive effect of aid on tax could be found. In addition, by disaggregating grants and loans we can test whether aid heterogeneity matters.[155]

The expected relations discussed here are only preliminary and they need to be tested empirically. In the next section, on identification of the long run structure, we assess their empirical validity and estimate the respective coefficients. .

4.3 Long Run Identification: Hypothesis Testing

We conduct a battery of long-run identification procedures to gain initial insight into the dynamics of the system. Namely, we test[156] whether the variables are long-run excludable, stationary, weakly exogenous, or purely adjusting. The Table 4.7 below summarises these results for the selected rank (r=3).

4.3.1 Long Run Exclusion[157]

The long-run exclusion tests test for a zero row restriction on β, i.e. whether the variable can be removed from the cointegration space without losing information. The tests of the same restriction in all cointegrating relations[158] do not impose identifying restrictions as they impose identical restrictions on all cointegrating relations. The likelihood ratio procedure tests the null of same restrictions on all r β vectors against the alternative of no restrictions on β. The test is approximately distributed as χ^2 with $r \times m$ degrees of freedom, where m denotes the number of restrictions.[159] A variable is said to be long-run excludable if its long-run coefficient can be accepted to be zero across all cointegrating vectors. For a system with

---

[155] Note that we do not include GDP in the specified model. GDP would capture the tax base effects on taxation, but would generate a lot of omitted variables (GDP determinants). Although one could argue that it would represent a reduced form, we maintain the focus on the fiscal system only.

[156] All tests in this section are likelihood ratio tests.

[157] Note we are formulating a system where we allow for a trend in levels, but not in CI relations. Since this is a testable restriction, we perform a test of long-run exclusion of a trend in the CI relationships to confirm that the trend is not required for our system.

[158] These would also include, for instance, tests of long-run homogeneity between variables for all long-run relationships.

[159] Note that we can only impose *p-m≥r* on the x_t endogenous variables without violating the rank condition (with no such constraint on the deterministic variables).

three cointegrating relationships, none of the variables of interest can be excluded (with a mild suggestion that the 1991 mean shift may be excludable).[160]

Table 4.7: Long Run Identification Tests

	texp	tax	non-tax	grants	loans	1991	1974
Long run exclusion	[0.000]	[0.002]	[0.008]	0.001]	[0.014]	[0.055]	[0.001]
Stationarity	[0.003]	[0.003]	[0.003]	[0.002]	[0.011]	**Excluded**	
Stationarity	[0.085]	[0.021]	[0.647]	[0.059]	[0.371]	**Included**	
Weak exogeneity	[0.007]	[0.096]	[0.002]	[0.020]	[0.073]		
Unit vectors in alpha	[0.194]	[0.061]	[0.040]	[0.007]	[0.054]		

Note: Table reports p-values for r=3.

4.3.2 Univariate Stationarity Tests

The univariate stationarity tests assert whether any variable is stationary (here, around the mean) by imposing zero restrictions on all other variables in one cointegrating vector, leaving other *r-1* vectors of long-run parameters unrestricted. For $r = 3$, this is implemented by imposing the restrictions on one cointegrating relation to include only the variable with deterministic components, and leaving the remaining two cointegrating vectors unrestricted. Note that the test results are sensitive to both the choice of rank, r, and the inclusion of the deterministic variables. We therefore report the results for all choices of rank (Appendix Table C5) and with shift dummies both included in and excluded from the cointegrating relations (Table 4.7). The LR test is asymptotically distributed as χ^2 with $(p1 - r) \times n_b$ degrees of freedom, where p1 contains p endogenous variables and the level shift dummies, and n_b denotes the number of restricted vectors (here $n_b = 1$). The null is stationarity.

For cointegration rank of our choice, $r = 3$, none of the variables can be accepted as stationary if the mean shift dummies are excluded. However, if the dummies are included, the stationarity of non-tax and loans cannot be rejected. This is likely to be due to 'slicing' of an already small sample. The DF-GLS tests largely support the hypothesis that most variables are I(1) processes with a drift, with the exception of loans (see Appendix Table C5).

[160] We consult the test statistics for the alternative (neighbouring) choices of rank: for *r=2*, none of the endogenous variables may be excluded either, although the exclusion of the 1991 shift dummy would be accepted. None of the variables could be excluded from a system with four cointegrating relations.

4.3.3 Weak Exogeneity Tests

Weak exogeneity tests identify which variables may not adjust to the long-run equilibrium by imposing a zero row in alpha vector (without imposing any restrictions on betas). If the null hypothesis is accepted, a variable with a zero row in alpha defines a common driving trend as the cumulated sum of the empirical shocks to the (weakly) exogenous variable.[161] A weakly exogenous variable therefore can be seen as having influenced the long-run stochastic path of the other variables without having been influenced by them itself. The LR test is asymptotically distributed as χ^2 with rm degrees of freedom, where m denotes the number of weakly exogenous variables.[162] For $r = 3$, tax revenue and loans are potentially weakly exogenous (but this will be re-tested once the long run structure is identified).

4.3.4 Tests for Unit Vectors in Alpha

Finally, mirroring the weak exogeneity test, unit vector in alpha test asserts whether a variable can be accepted as purely adjusting to the equilibrium error, with the remaining variables exclusively adjusting to the remaining $r - 1$ cointegrating relations.[163] Since a unit vector in alpha corresponds to a zero row in alpha orthogonal, the shocks to a variable that is purely adjusting to the cointegration relation would only have transitory (not permanent) effects on other variables, without any contribution to common stochastic trends. The LR test is asymptotically distributed as χ^2 with $v = m(p - r)$ degrees of freedom, where m denotes the number of known $\boldsymbol{\alpha}$ vectors. For $r = 3$, only government spending can be seen as purely adjusting, while such behaviour in tax and loans could only be borderline accepted.[164]

4.3.5 Individual Hypothesis Testing

In this section we test whether our hypothesised relationships are *individually*[165] stationary. Keeping the remaining $r - 1$ cointegrating relationships unrestricted (and thus unidentified),

[161] Therefore the test must comply with a condition that there at most can be (p-r) zero-row restrictions in alpha vector (Juselius, 2006:194).

[162] If more than one variable is found to be weakly exogenous, the joint test for weak exogeneity need to be performed. Should the individually weakly exogenous variables are found to be jointly weakly exogenous, then the cumulated shocks to these variables would completely define the autonomous driving trends (Juselius, 2006:202).

[163] Juselius (2006:201)

[164] In our identified system, a unit vector for government expenditure could be accepted as purely adjusting to beta 1 (internal budget equilibrium) with p-value of 0.19. The (unpredicted) shocks to government spending would be seen as having only transitory effects on other variables, which could be fairly plausible.

[165] As stated above, our main aim is to see whether the individually stationary cointegrating relationships hold together as an equilibrium system.

zero (or homogeneity) restrictions in a particular equilibrium relationship may be tested, allowing the remaining parameters to be estimated. The test is asymptotically distributed as χ^2 with $v = (m1 - r + 1) = (p1 - r) - (s1 - 1)$ degrees of freedom, where m1 is the number of restrictions, and s1 denotes the free parameters.[166]

The Table 4.8 below summarises the individually stationary (or not) relationships. Several variations of expected relationships are explored to check their 'stability' and reducibility.

Table 4.8: Stationarity (or otherwise) of Variable Combinations

	texp	*tax*	*nontax*	*grants*	*loans*	*Ds1991*	*Ds1974*	**p-value**
H1	1	-0.703	-0.275	-0.013	-0.05056	0	0	1 -
H2	1	-0.696	-0.338	0.001	0	0	0	0.391
H3	**1**	**-0.695**	**-0.338**	**0**	**0**	**0**	**0**	**0.692**
H4	1	-0.719	-0.281	0	-0.04538	0	0	0.765
H5	1	-0.731	-0.190	0	0	-0.177	-0.156	1 -
H6	1	-0.694	-0.330	0	0	-0.021	0	0.426
H7	1	-0.689	-0.347	0	0	0	0.016	0.409
H8	1	-0.785	0	0	0	-0.385	-0.361	0.630
H9	0	0	1	0	0	-1.358	-1.265	0.647
H10	**1**	**0**	**0**	**-0.761**	**-0.4520**	**0.785**	**-0.145**	**1 -**
H11	1	0	0	-0.807	-0.5583	0.993	0	0.869
H12	1	0	0	-0.607	-0.0117	0	-0.715	0.020
H13	1	0	0	-0.325	-0.8734	0	0	0.003
H14	1	0	0	-0.494	0	-0.258	-0.744	0.058
H15	1	0	0	0	1.3419	-2.794	-2.491	0.385
H16	0	0	0	1	2.0562	-4.337	-2.728	0.479
H17	0	0	0	1	-2.5671	0	0	0.005
H18	0	1	0	-0.9998	-0.7116	1.709	0.449	1 -
H19	**0**	**1**	**0**	**-0.884**	**-0.3932**	**1.135**	**0**	**0.628**
H20	0	1	0	-0.686	0.3080	0	-0.856	0.016
H21	0	1	0	-4.428	10.6359	0	0	0.002
H22	0	1	0	-0.559	0	0	0	0.002
H23	0	1	0	0	-2.1007	0	0	0.005

Note: Zeroes in the table are imposed and not estimated.
Also note that where r –1 = 2 conditions are imposed, the relationships are just-identified and therefore p-value is 1 by construction (i.e. restrictions are not testable) (Juselius, 2006:189).

The three relationships of interest are bolded in the table. H3 (later Beta 1) represents the internal budgeting. The relationship is stationary with p-value of 0.69. This (irreducible) relationship is very stationary irrespective of whether the shift dummies are included or excluded. The estimated coefficients clearly vary, but do so mildly. Also, a relationship with the excluded shift dummies would approximate a government's 'internal' decision making as

[166] Note that normalisation is not counted as a restriction as it is associated with an unrestricted α coefficient.

averaged over the whole sample. Importantly, such decision would release degrees of freedom, and test the remaining two long run equilibrium relationships. H10 (later Beta 2) summarises (positive) relationship between government total expenditure and aid variables. P-value is 1 by construction.[167] Such relationship needs at least 1991 to be stationary (but stationary regardless of inclusion of 1974). H19 (later Beta 3) represents a potential positive relationship between tax and aid variables. H1 reflects that expenditure bears a non-negative relationship with all the sources of funding.

4.4 Long Run Identification: Results

It is worthwhile to distinguish between a *just-identified* structure (with $r(r-1)$ identifying restrictions), where $r-1$ identifying (usually zero, or homogeneity) restrictions are imposed for each of r cointegrating relationships; and *over-identifying* restrictions, whereby more than $r-1$ identifying restrictions are imposed[168] for at least one of the cointegrating vectors. Whilst just-identifying restrictions do not change the value of the likelihood function as they do not constrain the parameter space, the over-identifying restrictions do, and therefore can be tested. Note that normalisation on one element in each vector does not change the likelihood as the corresponding α_i coefficient is normalised on the same β_i coefficient. However, that once we have identified a long run structure, the normalisation is an important choice, as we do not want to normalise on an insignificant variable)[169].

The LR test procedure for such hypothesis testing is asymptotically distributed as χ^2 with $v = \sum_{i=1}^{r}(m_i - r + 1) = \sum_{i=1}^{r}(p1 - r) - (s_1 - 1)$ degrees of freedom, where $p1$ is the number of parameters in β_i CI relationship, m_i are the restrictions imposed on the β_i vector, and s_1 are the parameters free to be estimated in each β_i.[170]

Imposing (a variant of)[171] just-identifying relations allows us to inspect whether the hypothesised (equilibrium) relationships represent economically sound system, i.e. whether the signs (and, possibly, magnitudes) of the coefficients are meaningful and significant.

[167] H11 however demonstrates that if 1974 dummy is excluded from this relationship when the other relationships are completely unconstrained, the relationship is stationary.

[168] I.e. the rank conditions are met, meaning no linear combination of other r-1 CI relations may produce a vector that resembles the first one (see Juselius, 2006:209-210, which further cites Johansen and Juselius, 1994 and Johansen, 1995). For further discussion of three aspects of identification (generic, empirical, and economic) see Juselius (2006:208).

[169] Juselius (2006:214).

[170] See also a summary of the process and also the discussion of the calculations of the degrees of freedom in Juselius (2006:212).

[171] Juselius (2006:218): “in general one can find many just-identified structures by rotating the cointegrating space”.

Results are provided in **Table 4.9**. As an over-identifying restriction, we impose a zero restriction on the 1974 dummy from the third cointegrating relationship, β_3, as its coefficient is reported as insignificant (**Table 4.9**). Furthermore, although the two mean-shift dummies appear significant in the first cointegrating relation, β_1, as it is reasonable to accept that the political regime changes would indeed affect the internal budget dynamics, further excluding the dummy variables from the first cointegrating vector we will free two degrees of freedom (reducing the first cointegrating relationship to an irreducible stationary long-run 'average' relationship between the fiscal variables) and will be able to test the system for stationarity. The results from the over-identified model are provided in **Table 4.10**, and sum up our key findings.

Long-run structure

To answer the economic questions, we impose the following restrictions on the long-run (β) coefficients: [172]

- To test whether there exists an internal budget equilibrium in the very long run, we exclude aid and dummy variables from β_1;
- To identify the relationship between aid and spending (with a special interest into which variables are adjusting), we exclude the domestic revenue variables from β_2;
- To explore the equilibrium between aid and tax, we exclude government spending and non-tax revenue from β_3.

β coefficients describe the stationary long-run equilibrium relations; the corresponding α coefficients describe the adjustment behaviour of the variables[173]. Normalisation of the β vectors is always done on a significant variable. In addition to this statistical criterion, normalisation is also decided to ease economic interpretability[174]. Note however that the results of the normalized beta should still be read as a vector and not as causal effects.

[172] Note that ordering of the β vectors does not affect the results.
[173] Note that α coefficient needs to be of opposite sign to its corresponding β coefficient to be equilibrium correcting.
[174] Contrary to a regression model, a change in the normalization will not change the ratio between the coefficients (Juselius, 2006:120).

Table 4.9: Just-identified Model with Disaggregated Grants and Domestic Revenue

	texp	tax	nontax	grants	loans	1991	1974	
Beta1a	**1**	**-0.73** (-18.19)	**-0.19** (-7.02)	-	-	**-0.18** (-5.04)	**-0.16** (-3.21)	~I(0)
Beta2a	**1**	-	-	**-0.76** (-7.39)	**-0.45** (-4.70)	**0.79** (4.93)	**-0.15** (-0.77)	~I(0)
Beta3a	-	**1**	-	**-1.000** (-4.83)	**-0.71** (-3.88)	**1.71** (5.14)	**0.45** (1.17)	~I(0)
Alpha1a	**-0.77** **(-3.21)**	**0.60** **(1.94)**	0.86 (1.32)	-1.16 (-1.26)	-1.36 (-1.08)	-	-	
Alpha2a	0.13 (0.68)	-0.13 (-0.54)	**-2.18** **(-4.15)**	**2.33** **(3.16)**	**2.30** **(2.26)**	-	-	
Alpha3a	-0.05 (-0.42)	0.19 (1.38)	**1.21** **(4.13)**	**-1.00** **(-2.42)**	-0.86 (-1.51)	-	-	
Normality	[p-value = 0.070]							
Stationarity	[p-value = '-']							
	Log-Likelihood = 426.899							

Note: t-statistics are reported in parentheses.

Table 4.10: Over-identified Model with Disaggregated Grants and Domestic Revenue

	texp	tax	nontax	grants	loans	1991	1974	
Beta1b	**1**	**-0.69** (-11.30)	**-0.34** (-8.56)	-	-	-	-	~I(0)
Beta2b	**1**	-	-	**-0.72** (-9.50)	**-0.29** (-5.15)	**0.53** (4.91)	**-0.36** (-8.72)	~I(0)
Beta3b	-	**1**	-	**-0.91** (-5.82)	**-0.38** (-3.35)	**1.16** (5.23)	-	~I(0)
Alpha1b	**-0.42** **(-3.06)**	**0.39** **(2.25)**	0.52 (1.41)	-0.45 (-0.87)	-1.07 (-1.45)	-	-	
Alpha2b	-0.22 (-1.52)	0.065 (-0.35)	**-1.89** **(-4.81)**	**1.718** **(3.15)**	**1.69** **(2.18)**	-	-	
Alpha3b	**0.24** **(3.20)**	0.08 (0.89)	**1.05** **(5.17)**	**-0.46** **(-1.63)**	-0.33 (-0.81)	-	-	
Normality	[p-value = 0.068]							
Stationarity	[p-value = 0.849]							
	Log-Likelihood = 426.498							

Note: t-statistics are reported in parentheses.

The first equilibrium relationship confirms the 'internal domestic equilibrium' hypothesis: government total expenditure is positively related to tax and non-tax revenue.[175] Although the logarithmic transformation infringes the interpretation of the coefficients as the

175 Note that this cannot be strictly seen as identity given non-compliance with logarithmic transformation.

homogeneity condition (i.e. expenditure is equal to the sum of domestic revenues[176]), the equilibrium could be interpreted as a very long run budgetary process equating public expenditure to domestic revenue (or continuity of statehood across the regimes), as the mean shift dummies are excluded and thus the relationship holds across the political regimes. As depicted in the α_1 coefficients, expenditure exhibits the strongest and most significant adjustment behaviour, adjusting to the equilibrium error in just over two years, although the α_1 coefficient on tax is of similar magnitude, implying that in the (very) long run tax is also adjusting to equilibrium error. This is in line with a sensible expectation that expenditure decisions are more sensitive to planned revenue.

The second identified equilibrium relationship reveals a positive association between both aid components and government spending.[177] Crucially, α_2 indicates that aid – and not expenditure – adjusts to departures from this equilibrium.[178] For instance, an increase in expenditure would result in increase in aid to restore the balance. Such behaviour suggests the hypothesis that donors may follow some 'disbursement rule' based on government spending decisions, whilst government spending behaviour does not seem to be conditional on disbursement of aid. Comparing between the aid modalities, the weaker relationship between expenditure and loans may reflect the fact that aid loans are often disbursed in lumps. The significance of the mean shift dummies reflects that the relationship has been changing across the political regimes.[179]

Finally, the identified long-run relationship between aid and taxation reveals no adverse effects of aid: both grants and loans are positively associated with the tax revenue.[180] Whilst expenditure and non-tax revenue exhibit some adjustment, grants can be seen as the most adjusting variable to departures from the equilibrium, consistent with the notion that donors support tax reforms by disbursing grants (potentially to relax some capacity constraints), with a significant change in 1991.[181] The lack of adjustment to the equilibrium

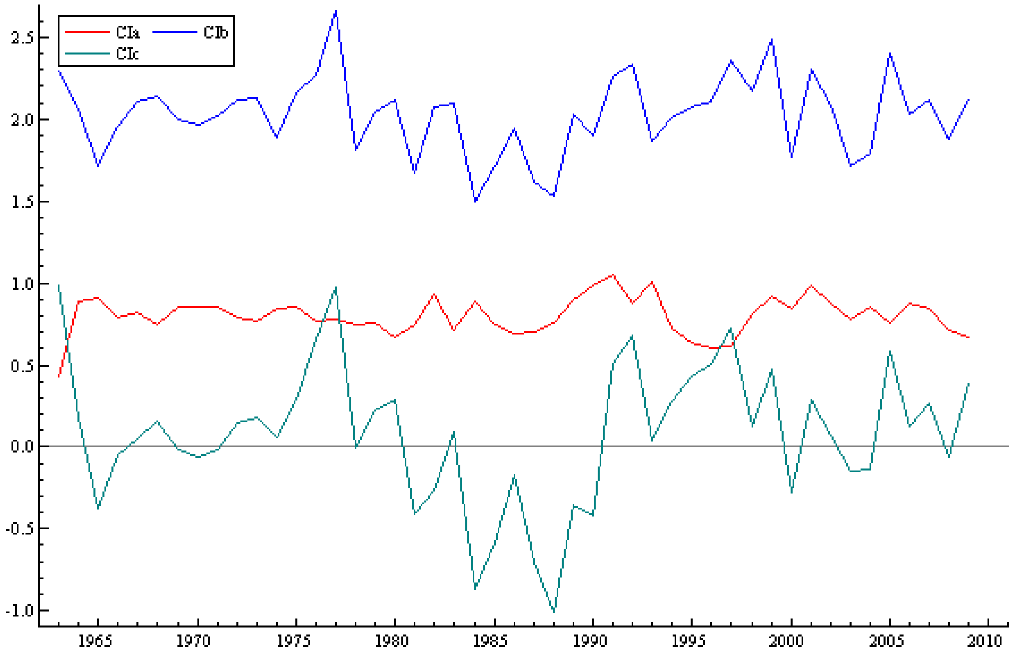
[176] And even then, assuming they are both fully measured.
[177] Note that the relationships that include non-zero aid coefficients require at least one mean shift dummy to be stationary, indicating important changes in aid behaviour and its relation to fiscal variables.
[178] In the light of this result, it might be more sensible to normalise on the (most adjusting) grants variable; however, normalising instead on government spending allows for the increased readability of the results.
[179] Note that non-tax revenue is also exhibiting some adjustment behaviour. This may be due to its 'stationary-like behaviour' when both dummies are included.
[180] Again, the vector is normalised on tax coefficient for the purposes of readability.
[181] Note that the 1974 dummy was insignificant and thus excluded from β_3.

error by aid loans indicates that shortfalls in taxation do not seem to be funded by concessional borrowing abroad.

The significance and the plausibility of the sign of dummy coefficients (Table 4.10) support their inclusion in the system. In the aid-spending relationship (Beta 2b) the 1974 dummy is negative and significant, and the 1991 dummy is positive and significant. This indicates that association between aid and spending was much stronger once EPRDF came to power. More interestingly, the 1991 dummy is positive, significant and large in the aid-tax relationship (1974 being insignificant and excluded altogether). This soundly reflects the changes in the variable relationships across regimes (justifying the inclusion of the dummies), stressing the convergence in donor and recipient government agenda for Ethiopia once EPRDF has come to power.

Keeping the identified long-run structure fixed, it is possible to test whether any of variables are weakly exogenous. Aid loans could be accepted as weakly exogenous (p-value=0.131), whilst weak exogeneity of tax is borderline rejected (p-value=0.049), likely due to the adjustment to the first equilibrium error. The remaining analysis in this chapter is conducted without imposing any weak exogeneity conditions. The joint stationarity of the over-identified system is accepted with a p-value of 0.849. Stationarity of individual vectors is depicted in the Figure 4.4 below. The Doornik-Hansen test suggest that the assumption of multivariate normality cannot be rejected (p-value=0.068).

Figure 4.4: Stationarity of Cointegrating Relations (Over-identifying Restrictions)



Note: (CIa= Beta1b, CIb= Beta2b CIc= Beta3b)

The figures in Appendix Table C6 report the parameter constancy tests for the over-identified model. The first thing to note is that the parameters, especially the alphas, were remarkably constant over time. In β_2 and β_3, the only indicated instability may be seen in the coefficients on grants and the 1991 dummy around the year 1995 (although the proportional change between those variables is rather constant).

4.5 Identification of the Short Run Structure[182]

While the CI relations in our model are $r = 3$ long run equilibrium relations between endogenous variables with the same time index, the short-run equations are $p = 5$ relations between p current variables (Δx_t); $\left(\mathrm{p} \times (\mathrm{k} - 1)\right)$ lagged variables Δx_{t-1} $(i = 1, \ldots, k-1)$; and r lagged equilibrium errors, $\beta(x_{t-1})$, from the identified long-run structure. Identification of the short-run structure requires $(p-1)$ restrictions on each of the simultaneous equations.

Two other important differences exist with respect to the long run identification. First of all, the distinction between endogenous and exogenous variables may change short run identification whereas it did not change the long run structure that is based on vectors. Secondly, identification of the short run structure requires uncorrelated residuals, whereas

[182] Modelled in PcGive.

no such requirement existed in the long run structure. Therefore the residual covariance matrix plays an important role here. In particular, uncorrelated residuals of a short-run structural model may be interpreted as estimated shocks, whilst large off-diagonal elements of covariance matrix can be a signal of significant current effects between the system variables (Juselius, 2006:230). Indeed, "the VAR model can be considered a reduced form model in the short run dynamics in the sense than potentially important current (simultaneous) effects are not explicitly modelled but are left in the residuals" (Juselius, 2006:230). The high correlation coefficients in the residual covariance matrix may also be due to the omission of relevant variables, but in our system it is most likely that it reflects contemporaneous effects between the fiscal variables.

As just-identified short-run structure is heavily over-parameterised, with many insignificant coefficients, in this section we report a parsimonious system, following Juselius (2006, Chapter 13), where the estimated coefficients with small t-statistics[183] were set to zero. Since there are some non-negligible correlation coefficients in the residual covariance matrix (see **Table 4.12**), the interpretation of the short-run equations as causal relationships (or reactions to structural shocks) should be taken with caution. The equation results are shown in the **Table 4.11**. The 30 over-identifying restrictions were accepted with a p-value of 0.5.

The government expenditure equation shows positive association with the past changes of foreign grants and loans, albeit with limited magnitude. This may reflect government's smoothing decisions in the face of volatile aid flows.[184] The tax equation indicates that even in the short run, aid is not inducing a reduction in tax revenue. This could well indicate a positive 'income effect' of aid on tax in the short run, as aid also seems to be positively associated with non-tax revenue. Grants do not seem to be reacting to any of the shocks in the short run, consistent with the qualitative suggestion that aid may be issued for strategic considerations; or that donors take time to react to Ethiopia's fiscal decisions. Finally, loans seem to be reduced in the face of higher tax (but not non-tax) revenues, which is a plausible prediction as the government need to borrow is reduced in the periods of growing revenues.

[183] P-value < 0.1.

[184] Or, since the data is in logs, just reflect that a percentage change in each aid component, which together amount to about a fifth of government's spending, corresponds to about a fifth of the percentage increase in the government spending, indicating that received aid is actually spent in one period.

Table 4.11: Short Run Equations (Over-identified Structure)

Dep. vrbl.	$\Delta texp_{t-1}$	Δtax_{t-1}	$\Delta non\text{-}tax_{t-1}$	$\Delta grants_{t-1}$	$\Delta loans_{t-1}$	$CI1b_{t-1}$	$CI2b_{t-1}$	$CI3b_{t-1}$	D91	D74
$\Delta texp_t$=				0.16	0.11	-0.58	0.25			0.16
				(4.3)	(4.5)	(-5.6)	(6.0)			(2.0)
Δtax_t=			0.2	0.08	0.46		-0.17	0.17	-0.26	
			(3.6)	(1.8)	(2.8)		(-2.5)	(3.3)	(-2.6)	
$\Delta nontax_t$=				0.24	0.16			0.19		0.42
				(2.4)	(2.2)			(2.2)		(1.7)
$\Delta grants_t$=								0.46		
								(4.0)		
$\Delta loans_t$=		-1.14	0.48			-1.74	0.74			
		(-2.6)	(2.8)			(-3.5)	(3.7)			
	LR test of over-identifying restrictions: Chi^2(30) = 29.325 [p-value=0.5006]									

Note: t-statistics are reported in the parentheses

Table 4.12: Residual Correlation Matrix (Parsimonious Structure)

	Δtexp	**Δtax**	**Δnon-tax**	**Δgrants**	**Δloans**
Δtexp	0.089				
Δtax	**0.356**	0.100			
Δnon-tax	**0.410**	-0.162	0.248		
Δgrants	-0.039	0.104	-0.010	0.321	
Δloans	**0.410**	0.171	-0.043	0.122	0.420

4.6 Identification of the Common Trends (MA)

Unlike the identification of the stationary long run relationships, the identification[185] of the MA is not invariant to the information set. Furthermore, given that some of the residual cross-correlations are non-negligible, the residuals cannot be strictly interpreted as structural shocks. Finally, we did not find enough evidence to substantiate the imposition of weak exogeneity restrictions, or identify variables that are purely adjusting to the identified long-run structure. Therefore the results in this section are to be taken with caution. The identified common trends and their loadings on other variables are provided in Table 4.13.

[185] (*p-r-1 = 1*) restrictions are required to just-identify each common trend.

Table 4.13: Composition and Loadings of the Common Trends

	texp	tax	non-tax	grants	loans
The composition of common trends (CT) $[\alpha'_{\perp}]$					
CT(1)	0.8	**1**	-0.46	**-0.44**	-
	(1.16)		(-1.16)	(-1.91)	
CT(2)	-1.29	-	0.197	**-0.93**	**1**
	(-0.83)		(0.30)	**(-1.85)**	
The effect of the common trends on other variables $[\beta_{\perp}(\alpha'_{\perp}(I-\Gamma_1)\beta_{\perp})^{-1}]$					
CT(1)	**0.54**	**0.69**	**0.20**	**0.62**	0.33
	(2.85)	(2.84)	(2.89)	(2.30)	(0.60)
CT(2)	-0.03	-0.04	**-0.02**	**-0.30**	**0.62**
	(-1.10)	(-0.98)	(-1.92)	(-6.8)	(6.88)

Note: t-statistics reported in the parentheses.

Broadly, the first common trend (CT1) seems to be mostly constructed from the unanticipated shocks to tax revenue, with a potential contribution from grants (other coefficients being insignificant). It seems to most strongly affect the domestic fiscal variables, notably, expenditure and tax revenue; and also grants, indicating support for potential donor response to tax mobilisation reforms. The second common trend seems to be composed from shocks to aid variables (loans), and loading to aid variables themselves, consistent to aid policy being fairly independent of recipient's fiscal dynamics.

The columns of C matrix more broadly illustrate how unanticipated 'shocks' to each variable ripple through the system, with a significant coefficient indicating a permanent effect; otherwise, the effect is transitory at most. Likewise, the rows indicate how each variable is affected by such 'shocks'. The results are summarised in Table 4.14 below. Unanticipated 'shocks' to government expenditure could be expected to have persistent positive effects on expenditure itself, tax (and non-tax) revenue, and, especially, grants. The effect on loans can be expected to be temporary at most, and negative. The unanticipated shocks to tax revenue would have positive permanent effects on all domestic fiscal variables and grant aid. Shocks to non-tax revenue[186] seem to affect all variables negatively, and, loans aside, permanently. This could be indicating of detrimental policies of expropriation or transfer of funds from the central bank. The effects of cumulated 'shocks' to grants are more difficult to interpret, as they seem negatively affect both expenditure and tax, whilst a permanent negative effect on loans could indicate that grants and loans are substitutes, perhaps from the donor perspective. Loans, on the other hand, do not seem to have permanent effects on the domestic fiscal variables, but they seem to negatively affect grants (again, consistent

[186] This may be driven by changes associated with nationalisation or privatisation programmes.

with the substitution between the aid modalities), and have a positive effect on loans themselves, possibly signalling repayment or servicing difficulties. Note again, that these results are indicative at most.

Table 4.14: Long Run Impact Matrix

	texp	tax	non-tax	grants	loans
texp	**0.49**	**0.54**	**-0.26**	**-0.21**	-0.03
	(1.85)	**(2.85)**	**(-2.00)**	**(-2.20)**	(-1.10)
tax	**0.61**	**0.69**	**-0.32**	**-0.26**	-0.04
	(1.83)	**(2.84)**	**(-2.00)**	**(-2.23)**	(-0.98)
non-tax	**0.19**	**0.20**	**-0.09**	**-0.07**	**-0.02**
	(2.00)	**(2.89)**	**(-2.07)**	**(-1.99)**	**(-1.92)**
grants	**0.89**	**0.62**	**-0.34**	0.01	**-0.30**
	(2.39)	**(2.30)**	**(-1.90)**	(0.05)	**(-6.76)**
loans	-0.53	0.33	-0.03	**-0.72**	**0.62**
	(-0.70)	(0.60)	(-0.08)	**(-2.68)**	**(6.88)**

Note: t-statistics reported in the parentheses.

5. Alternative Model with Disaggregated Aid, Aggregated Domestic Revenue, and Disaggregated Spending

To get more insight into what aid might be actually funding, or, alternatively, which spending decisions does it seem to be more adjusting to, we re-specify the system in a different way: keeping the aid flows disaggregated as above, we aggregate the domestic revenue (***domrev***) variables (given that they did not exhibit highly contrasting long-run behaviour) and disaggregate government expenditure into its capital (***cexp***) and recurrent components (***rexp***), keeping the total number of variables in the system to p=5. The structure of the deterministic terms is identical to that of the model above. The selected lag length is k=2 (Table 4.15). Although lag length selection tests support k=1, and such choice would not present residual autocorrelation issues, k=2 is superior in terms of normality. Although the Johansen trace test suggest r=4 (Table 4.16), such system lacks economic identification. The choice of cointegration rank is r=3, and it is supported by additional information (for instance the largest modulus of root is 0.8 for r=4, and for r=3 it is 0.648). For briefness, we focus on the long-run results only[187], as the fit of the system is slightly inferior[188] to the model above.

[187] Short-run structure is summarised in the Appendix Table C7.

Table 4.15: Lag Length Determination (Alternative System)

Lag Reduction Tests						
VAR(4)	<<	VAR(5)	:	ChiSqr(25)	=	105.487 [0.000]
VAR(3)	<<	VAR(5)	:	ChiSqr(50)	=	152.346 [0.000]
VAR(3)	<<	VAR(4)	:	ChiSqr(25)	=	46.860 [0.005]
VAR(2)	<<	VAR(5)	:	ChiSqr(75)	=	206.752 [0.000]
VAR(2)	<<	VAR(4)	:	ChiSqr(50)	=	101.265 [0.000]
VAR(2)	<<	VAR(3)	:	ChiSqr(25)	=	54.406 [0.001]
VAR(1)	<<	VAR(5)	:	ChiSqr(100)	=	253.858 [0.000]
VAR(1)	<<	VAR(4)	:	ChiSqr(75)	=	148.371 [0.000]
VAR(1)	<<	VAR(3)	:	ChiSqr(50)	=	101.512 [0.000]
VAR(1)	<<	VAR(2)	:	ChiSqr(25)	=	47.106 [0.005]

Model	k	T	Regr.	Log-lik	SC	H-Q	LM(1)	LM(k)
VAR(5)	5	42	30	532.201	-11.994	-15.925	0.346	0.637
VAR(4)	4	42	25	479.458	-11.707	-14.983	0.128	0.051
VAR(3)	3	42	20	456.028	-12.816	-15.437	0.042	0.649
VAR(2)	2	42	15	428.825	-13.746	-15.711	0.367	0.336
VAR(1)	1	42	10	405.272	-14.849	-16.159	0.288	0.288

Effective Sample: 1971:01 to 2012:01
SC : Schwarz Criterion; H-Q : Hannan-Quinn Criterion
LM(k): LM-Test for autocorrelation of order k

Table 4.16: Rank Determination (Alternative System)

p-r	r	Eig. Value	Trace	Trace*	Frac95	P-value	P-value*
5	0	0.536	105.570	90.497	76.896	0.000	0.003
4	1	0.439	70.972	61.131	54.245	0.001	0.011
3	2	0.362	44.973	39.571	35.296	0.004	0.017
2	3	0.350	24.734	21.395	19.817	0.011	0.032
1	4	0.112	5.328	5.144	5.805	0.067	0.074

Note: * *denotes Bartlett corrections*

Table 4.17: Misspecification Tests (Alternative System)

Residual normality *(p-values)*						
	Multivariate	**Univariate**				
		cexp	**rexp**	**domrev**	**grants**	**loans**
	0.012	0.291	0.002	0.120	0.597	0.956
Residual autocorrelation and ARCH effects *(p-values)*						
		LM(1)	LM(2)	LM(3)	LM(4)	
Residual autocorrelation		0.664	0.383	0.248	0.429	
ARCH		0.937	0.397	0.493	1.000	
Trace correlation	0.509					

Note: *all values are p-values.*

[188] Table 4.17 provides a summary of the long-run identification tests: loans are reported as long run excludable, but will be kept in the system as it is one of the key variables. For most of the variables (though not *cexp*), weak exogeneity and unit vector in alpha tests provide conflicting results). However, if the dummies are excluded, none of the variables are found to be stationary.

Table 4.18: Long run Identification Tests (Alternative System)

	cexp	rexp	domrev	grants	loans	1991	1974
Long run exclusion	0.139	0.012	0.020	0.035	0.480	0.140	0.192
Stationarity	0.001	0.001	0.001	0.001	0.005	**Excluded**	
Stationarity	0.019	0.212	0.051	0.010	0.883	**Included**	
Weak exogeneity [189]	0.548	0.014	0.711	0.164	0.896		
Unit vectors in alpha	0.009	0.796	0.806	0.153	0.142		

Table reports p-values for r=3.

Results from the over-identified system are summarised in Table 4.19 below. The first cointegrating vector (normalised on domestic revenue for readability purposes) mimics the previously identified domestic budget equilibrium: domestic revenue is positively associated with both components of government expenditure. Interestingly, the relationship is stronger with the capital expenditure, possibly reflecting that the periods with 'good' government policies targeting the collection of revenue tend to be reflected in more capital ('development') spending. The recurrent spending is the single adjusting variable to departures from this equilibrium. Vectors β_2 and β_3 roughly correspond to the second equilibrium in the previous model, with aid variables now related separately to capital and recurrent expenditure to identify any potentially differing effects. Both aid variables seem to be positively related to both components of government spending. Particularly, grants seem to be more strongly associated with capital expenditures than loans. With grants being the most adjusting variable to this equilibrium error, it potentially signals the accordance of donors and recipients on financing priorities and donors backing the commitment to increased domestic capital expenditures with more grants, rather than successful aid conditionality. Finally, the third equilibrium relationship indicates positive associations between aid and government recurrent expenditures. While some would argue this may point to 'general fungibility' issues, our view is that some aid is indeed intended to fund recurrent spending components (such as health, and education), although the positive relationship between loans and recurrent expenditure could in principle reflect war financing through loans both in Derg and EPRDF period.

[189] For r=4, the picture is different, with only capital expenditure reported as weakly exogenous.

Table 4.19: Long Run Results (Alternative System, Over-identified)

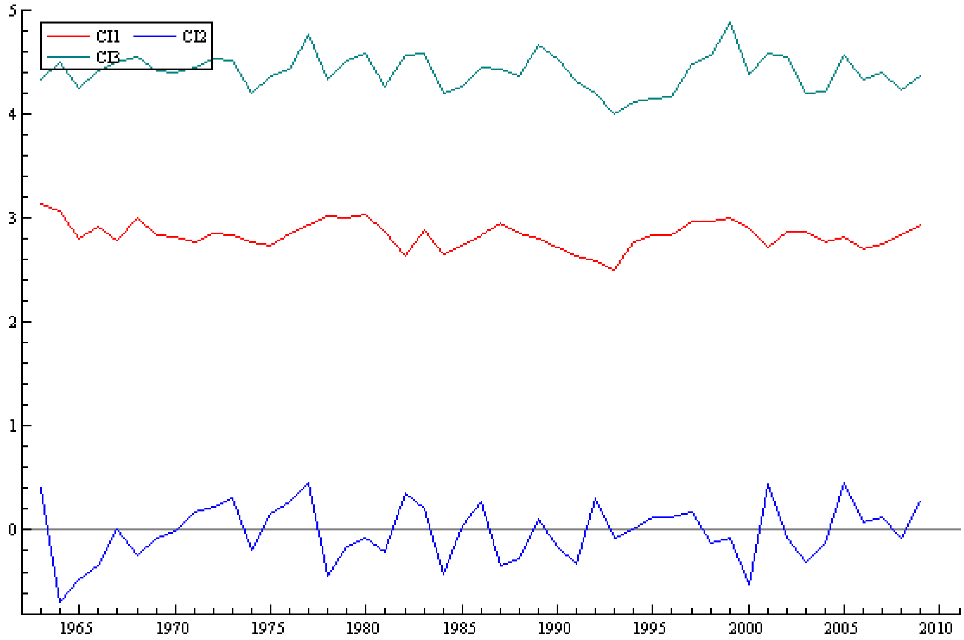
	cexp	rexp	domrev.	grants	loans	1991	1974	
Beta1c	**-0.60**	**-0.13**	**1**	-	-	-	-	~I(0)
	(-14.09)	(-1.84)						
Beta2c	**1**	-	-	**-0.89**	**-0.18**	**0.17**	**-0.71**	~I(0)
				(-15.03)	(-3.10)	(2.06)	(-6.88)	
Beta3c	-	**1**	-	**-0.36**	**-0.20**	-0.05	**-0.472**	~I(0)
				(-7.82)	(-4.62)	(-0.74)	(-6.13)	
Alpha1c	-0.21	**0.84**	-0.16	0.08	0.42	-	-	
	(0.53)	(4.98)	(-0.73)	(0.13)	(0.44)			
Alpha2c	-0.22	**0.33**	0.08	**0.75**	-0.21	-	-	
	(-1.17)	(4.14)	(0.77)	(2.64)	(-0.47)			
Alpha3c	-0.11	**-0.59**	*-0.25*	0.56	0.53	-	-	
	(-0.42)	(-5.13)	*(-1.64)*	(1.37)	(0.83)			
Normality	[0.068]							
Stationarity	[0.564]							

Note: t-statistics are reported in parentheses.

The positive and strong relation between grants and capital expenditure is consistent with other findings in the literature (Fagernas and Roberts, 2004a; Fagernas and Schurich, 2004; Martins, 2010). On the one hand, this result may seem counterintuitive: indeed it may be reasonable for loans to be more related to the more "productive" capital expenditure since they have to be repaid in the future. In practice, however, loans to Ethiopia are largely concessional, making repayment an issue rather distant in time that therefore might not have direct policy implications. On the other hand, grants may come with more conditionality, in the form of pressure to spend on 'productive' capital rather than recurrent ("consumption") expenditure, precisely because they do not require repayment. The idea that capital spending is preferable to recurrent expenditure may be slowly fading in the international debate. However, in historical perspective, this distinction may be behind the result of grants being mostly associated with capital expenditure. This suggests that donors back a commitment to increased domestic capital expenditure with grants.

Figure 4.5 displays the identified cointegrated relationships.

Figure 4.5: Cointegrating Relations (Alternative System)



Note: CI1=CI1c; CI2=CI2c; CI3=CI3c .

6. Conclusions

Using unique dataset of domestically collected Ethiopian fiscal data and paying particular attention to the quantitative context, this chapter shed light on the fiscal effects of aid in Ethiopia and it answered the questions raised in Section 2 on the hypothesised effects of aid.

We provide evidence for the existence of a domestic budget equilibrium that includes domestic revenues and government expenditure, but excludes aid. The domestic budget equilibrium between total expenditure and domestic revenue is confirmed in the two models estimated in this chapter as well as in the alternative systems used for checking robustness. This relation holds regardless of the political regime changes across the whole period considered. By looking at adjustment coefficients we also find that spending plans are mainly driven by tax revenue, while expenditure is the most adjusting variable.

Most crucially, we find no evidence of an adverse effect of aid on tax revenue, which implies that the government of Ethiopia is not substituting taxes with aid, nor has it discouraged in its tax revenue collection. On the contrary, we find a positive and robust relation between tax revenue and both grants and loans in the long run, which is also largely confirmed in the short run structure. This relation may be explained by a complementary beneficial effect of aid in improving tax administration and strengthening domestic institutions. Indeed, throughout the whole period the government of Ethiopia has received foreign advice on tax matters and this remains today one of the policy areas of highest agreement between the government and donors. Moreover, Ethiopia's history of independence from colonial powers has profoundly shaped the national character and pride, making financial independence a core priority of the current government. As a consequence, the case for a substitution or tax displacement effect of aid is particularly ill-grounded in Ethiopia as confirmed by this analysis.

Both aid variables are found to have a positive and robust relation with public expenditure. This relation is stronger between capital expenditure and grants, as shown in the alternative system using disaggregated expenditure data. This finding is consistent with the results in the literature and with the idea that donors may have a preference for grants to be spent on the more productive capital expenditure rather than on 'consumption' (recurrent) expenditure. We are also able to identify a 'donor disbursement rule' of sorts whereby donors back proven commitment to increased expenditure with additional funding, particularly grants.

The key differences between this study and Martins (2010) is the data frequency (and thus period covered, 1993-2008), and the (number of) variables used. Using quarterly data increases sample size. As the author notes, fiscal decisions are indeed taken throughout the year; however, the key aid, tax, and even spending decisions (e.g. large projects) would be decided on annual (or, at most, semi-annually during the official budget-revisions) rather than quarterly basis. The author uses the all the variables used in Chapter 4 of this thesis, plus public borrowing. In theory, adding another variable should not change the already identified cointegrating relationships. However, as the period covered differs between the studies, this can be expected to affect the results (as well as model specification), especially when both samples are very small. Two key results are identified in both studies: the absence of a negative association between domestic revenue and aid; and a positive association between aid grants and development expenditure.

All the results presented here are robust to different variations in the system, which is particularly valuable in the CVAR context where results are often very model-specific. We are able to test and confirm all the underlying statistical assumptions of the VAR model, more so in our main system than in the alternative one, thus supporting the validity of our results. Using exclusively national data sources we are able to avoid problems related to the different international measures of aid and capture exactly the component that is most relevant for the analysis of its fiscal effects. Our dataset also presents an advantage in terms of the length of the time series available, which is the longest in the CVAR fiscal literature for developing countries. Last but not least, the findings of this chapter are largely rooted and consistent with the Ethiopian context and with the qualitative evidence on the political economy of the country.

In efforts to reconcile the cross-country aid-tax estimates with country case studies conducted in this thesis, one should consider three aspects: the political economy arguments; the soundness of empirical estimates; and the appropriateness of the statistical techniques to available data.

The prevailing argument from the political economy perspective suggests that aid is expected to crowd-out tax collection because in the fledgling state it is less politically risky (or costly) to increase aid flows than to raise taxes[190] (Brautigam and Knack, 2004; Gupta et al., 2004). It is the main argument behind the empirical estimates of a negative aid coefficient in tax effort equations. The usual counter-argument is that aid-tax association could be positive if aid increases the domestic revenue mobilisation through support for reforms, and through increased tax base (Gupta, 2007; Brun et al., 2009; Clist and Morrissey, 2011). Returning to the first argument, Morrissey and Torrance (2015) consolidate the argument why may the political cost of aid dependence actually exceed the cost of tax collection, once the bureaucratic, accountability, and reduction-in-the-polity-autonomy costs are properly accounted for.[191] Overall, taken sufficiently broadly, political literature does not give robust prediction over the direction or sign of the tax-aid association.

From the empirical perspective, consistent with the political economy arguments, the studies find negative (Gupta et al., 2004; Benedek et al., 2012) or positive (Clist and Morrissey, 2011; Morrissey et al., 2014) association between tax effort and aid. Through replication, Morrissey and Torrance (2015) demonstrate that the empirical negative cross-country relationship between aid and tax is altogether not robust, being sensitive to specification, estimator, and introduction of lagged aid, or alteration of the sample or frequency of data. Crucially, the authors argue that it is not the amount of aid (the usual

[190] The argument implicitly assumes a significant level of political accountability and other aspects of democratic rule, which may not be exactly the case.

[191] Namely, they consider bureaucratic cost (multiple donors with distinct procedures and requirements each intervening in several sectors and thus imposing cost on line ministries versus cost of tax administration); accountability cost (to donors vs. the ‘electorate’); costs of (absence) of autonomy (with any aid conditionality reducing the autonomy of the recipient government by constraining policy action).

measure of the aid variable) that has influence on the level of tax revenue, but the nature or the donor-recipient relationship and policy dialogue.

From the statistical perspective, cross-country estimations ignore country heterogeneity; single equation tax effort models imply one-directional causality; existing estimations ignore the distinction between on- and off-budget aid, usually lumping the two together. On the other hand, whilst the multiple equation setting (and thus estimated feedback effects) and the respect for the country heterogeneity of the CVAR overcomes some of these criticisms, enforcing the CVAR method on too small of a sample, potentially combined with researcher's (latent) priors when invoking 'judgment' where statistical guidance is less that crystal clear may too produce less than robust results.

Taking that political economy arguments are not conclusive, empirical evidence is not robust, and the application of both panel and time-series statistical techniques each have their drawbacks, the task of reconciling the cross-country aid-tax effort evidence with those of country-case time-series application is both easier and less meaningful.

The key result demonstrated in Chapter 4 (and Chapter 5)[192] is that there is no *negative* association between aid and tax revenue. The key association is shown to be between the domestic variables (expenditures and tax revenue), and then aid and expenditures; only then do we identify a relationship between aid and tax. This is unsurprising. The primary purpose of on-budget aid is to finance expenditure; and one could consider the secondary aim to be doing so without reducing the domestic (tax) revenue.

This positive (or non-negative) association between aid and tax is stronger in the final subsample, during the EPRDF period (as indicated by the 1991 significant positive shift dummy[193]). Although this period saw higher levels of total (on-budget) aid, from the qualitative information (as off-budget fraction of aid is not captured in the quantitative data) it is clear that Ethiopia also received much higher levels of technical assistance (including support for the considerable tax revenue reform), and saw much more aligned donor-recipient dialogue with respect to fiscal and macroeconomic management, including tax reforms (see also Moore, 2014); this again resonates with Morrissey and Torrance (2015)

[192] Although data limitations complicated the analysis, no testing indicated existence of a negative aid-tax relationship in Tanzania, instead consistently estimating a significantly positive association in various versions of the estimated system. See Chapter 5 for more detail.

[193] The analysis of the sub-sample is not possible due to much too small (sub)sample size. 'Artificially' increasing the sample size by using quarterly data (as in Martins, 2010) would not provide much insight, as neither aid not tax collection are likely to move meaningfully on quarterly basis from a policy perspective.

that the positive associations in the data can be driven by other (structural) characteristics of the country (here the change of political winds and broader reforms not captured in by the variables in the estimated system).

Chapter 5

Fiscal Effects of Aid in Tanzania:

Evidence from CVAR Applications

1. Introduction

Tanzania provides an interesting case for comparison to Ethiopia. There are similarities in that both countries followed socialist regimes (for longer in Tanzania) before severe liberalisation policies in the 1980s/early 1990s. Their relationships with donors developed along rather different paths: with Ethiopia never colonised, its aid policy shifted radically during the 1980s from the US to USSR, and then back to Western powers after the end of the Cold war; meanwhile, after gaining independence in 1961, Tanzania enjoyed support from the socialist-oriented Scandinavian donors, as well as other Western donors (including the Netherlands, West Germany, and the World Bank).

This chapter thus aims to conduct a CVAR exercise to evaluate the fiscal relationships in Tanzania during the independence period, covering the years 1966-2012. The motivation is largely the same as for the Ethiopian chapter: regarding the understanding of the fiscal mechanisms as a prerequisite to understanding the broader macroeconomic effects of aid, the aim of this chapter is to identify the fiscal effects of (or on) aid in Tanzania. The key advantage, as before, is that recipient's measure of aid is available: the components of both cash grant and loan aid recorded as received by the Tanzanian government, and thus considered the most (directly) influential towards the fiscal policy. [194] As before, The CVAR analysis is complemented by a qualitative context, which ensures sound model specification and sensible interpretation of estimated results, especially in a small sample setting. The quality of the overall quantitative (even highly aggregated) data, however, is lower than that of Ethiopian data, which proves to be a severe challenge. In a way, this chapter demonstrates the limitations of the application of the CVAR method on the very small developing country sample.

The framework for the analyses of fiscal effects and the review of the relevant literature is already provided in Section 2 of Chapter 4 and will thus not be repeated here. The remainder of this chapter is as follows: Section 2 discusses the data and provides a description of the Tanzanian qualitative context. The modelling choices and estimated results are provided in section 3; given the nature of the data and the resulting delicacy of the results only long-run identification is discussed. Section 4 concludes by comparing the results from the two case studies. The complementary information is provided in Appendix D.

2. Data and Qualitative Context

The dataset used here is an extension of that used by Kweka and Morrissey (2000) and was provided by Josaphat Kweka (at the time in the World Bank office in Tanzania). Following variables are available: government expenditures, decomposed into development and recurrent components;[195] domestic revenue, decomposed into tax and non-tax revenue; government borrowing (budget financing); and aid, decomposed into grant and foreign financing components. The inclusion of GDP data in the study is complicated for two reasons. Firstly, with a system of fiscal variables alone, scaling by GDP would not add much information, whilst attempting to relate the fiscal system to GDP growth would imply (too) many omitted variables.

[194] Note that, as before, the indirect effects of off-budget aid are not modelled. To some extent, then, the trade-off of aid absorption (increase in net imports) and spending (widening in the fiscal deficit) is also ignored (see Killick and Foster (2007), for instance), implicitly assuming that all aid that goes through the government is expected to be spent or used to reduce domestic borrowing or building reserves. In fact, Killick and Foster (2007) find that increases in aid in Tanzania were not at all absorbed but nearly fully spent. Meanwhile, for Ethiopia, they found that during the given period, aid surges were not at all spent, and only partially absorbed, rebuilding the forex reserves and reducing government debt, to some extent justifying our key focus on aid's effect on taxation.

[195] We maintain the distinction as it is coded in the original dataset.

Secondly, and more importantly, Tanzanian GDP was rebased in 1987, and all data predating this transformation appears to have been erased from the majority of data sources, including those of the Tanzanian government. The data could be recovered from the National Accounts from the UN data website, but it is not clear how credible or comparable this would be to the post-1987 series, while the shift dummy included to account for the rebase in GDP would also capture the start of the liberalisation reforms (see below), although there is no reason to expect that these reforms would have had an immediate or even quick effect on GDP.

All variables are expressed in domestic currency (Tanzanian shillings), so no conversion is necessary – analysis is conducted using variables expressed in the domestic currency. However, the (urban) consumer price index (CPI) was used to convert the series into real values. Evidently, the GDP deflator was not available for the entire period studied due to aforementioned GDP data issues. Clearly, the CPI is severely distorted during the radical socialist regime (effectively at least until 1986), as nearly all nominal prices were set by the government bodies rather than market forces[196]. However, it is a measure available from IMF IFS database (with the base year in 2005) for the entire period in question.

Log values are again preferred as the raw values resemble processes with quadratic trends and thus complicate the estimation. This introduces slight further complications on the estimation, as aid loans and the borrowing variable have negative values for several years, and grants and non-tax domestic revenue contain zero values.[197] For the analysis where aid is disaggregated to grants and loans component, we scale the loans variable by adding a constant (80,000 TShs as the minimum value of loans observed is -73,365) so that all values are strictly positive, and shorten the sample to include only non-zero observations for grants (i.e. exclude 1966-67). Where the model includes domestic borrowing, this variable is also scaled by a constant (300,000 TShs). The non-tax variable is not used as a standalone variable as it is coded as 0 for the period 1968-1982.

The fiscal trends are summarised in **Figures 5.1-5.5**. **Figure 5.1** depicts all fiscal variables and aid as a fraction of GDP to provide a conventional comparison between these variables, whilst

[196] Clearly, this would severely distort the nominal (and real) fiscal variables, too, and thus the pro-market reforms need to be accounted for by adding a step dummy (although they essentially coincide with the GDP rebase, and thus perhaps could be entailed in the same dummy).

[197]Interestingly in 1989, both loans and domestic borrowing are recorded as negative – it could be that (some) additional aid was used to repay/reduce foreign and domestic borrowing.

Figures 5.2-5.5 provide information in levels. Figures in logs are reported in the **Appendix Table D1**.

Figure 5.1: Tanzanian Fiscal Aggregates Expressed as Proportion of GDP

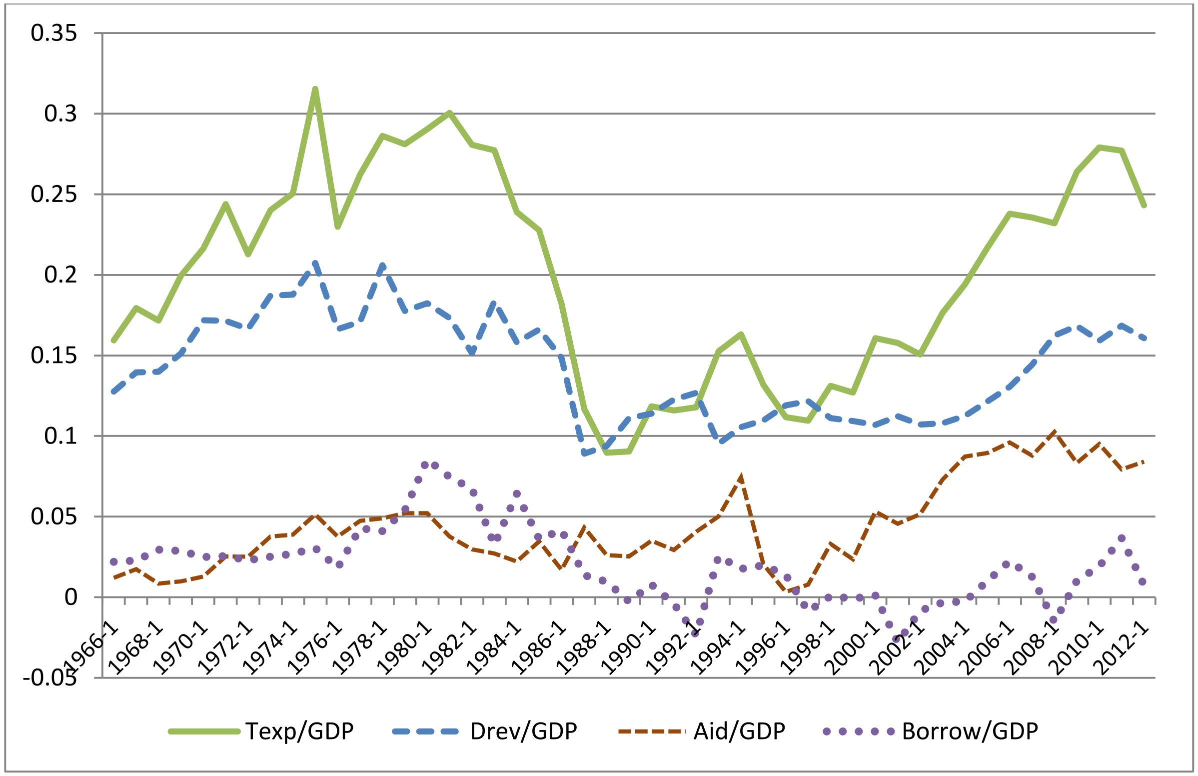


Figure 5.2: Expenditures (Levels, deflated)

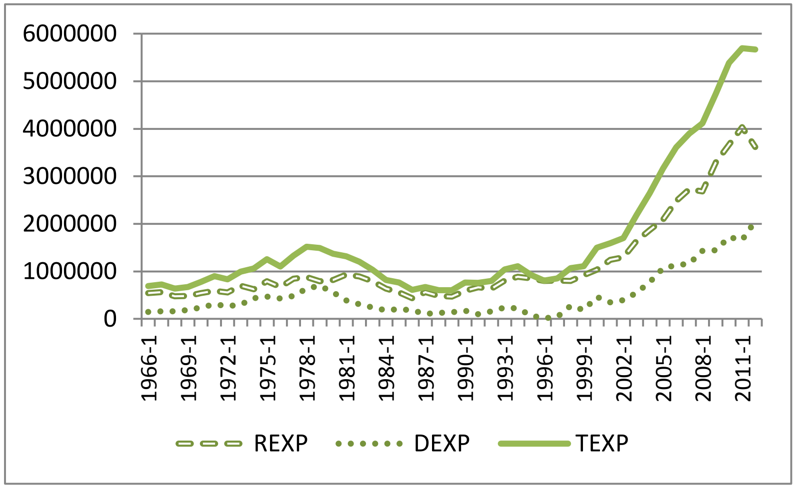


Figure 5.3: Domestic Revenue (Levels, deflated)

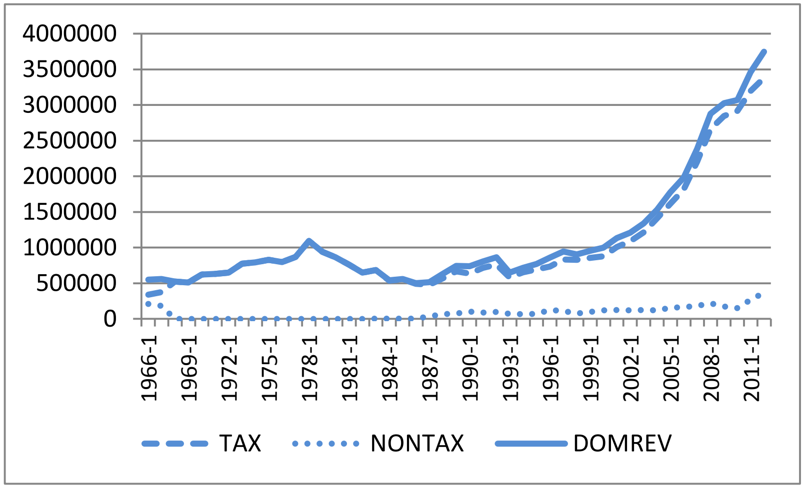


Figure 5.4: Aid (Levels, deflated)

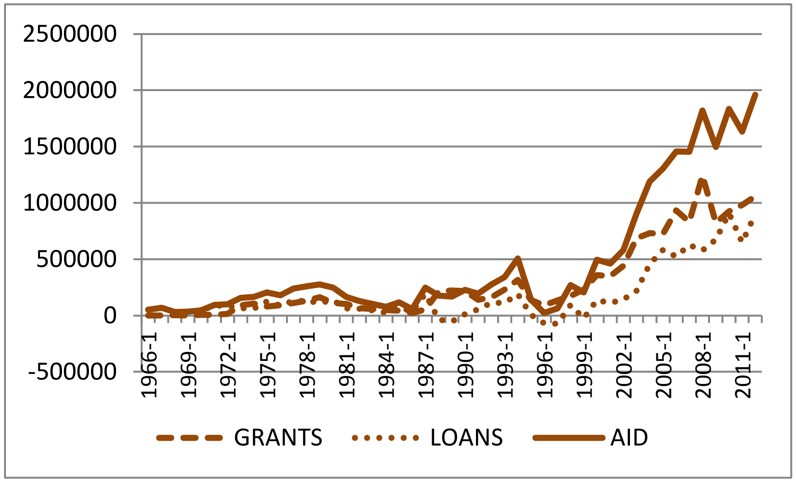
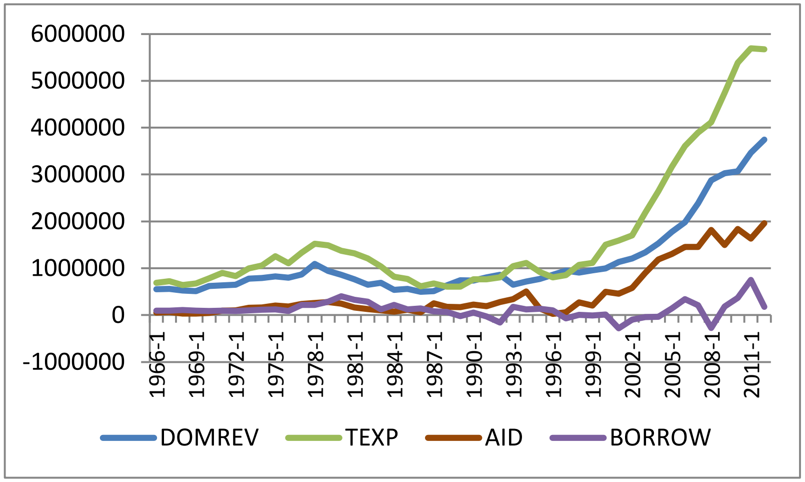


Figure 5.5: Aggregates (Levels, deflated)



The beginnings of independent Tanzania have been described as radically socialist, anti-market and anti-capitalist economy (Hyden and Karlstrom, 1993). President Nyerere's view was that prices had little-to-no role in allocation of economic resources, so the nominal prices, exchange rates and nominal interest rates were fixed and unchanging. The majority of the population remained in the (subsistence) agricultural sector (and thus outside the monetary economy), with enforced villagisation process. The government size was increasing rapidly (see Figure 5.2); however, Nyerere's 'ambitious and increasingly unrealistic' (Hyden and Karlstrom, 1993:1397) development plans soon had an adverse effect on the economy. Although some positive external shocks, such as the increases in the global coffee prices (following the frosts in Brazil), temporarily dampened the negative trends, Tanzania's terms of trade started declining, and debt burden increasing, by a surge of negative external shocks from 1979-onwards.

In the early 1970s aid flows to Tanzania were increasing (see Figure 5.4). Although not all aid was from countries with a close political ideology, the top six donors accounting for over 50 per cent of total cumulative aid flows during the period of 1970-84 were (in decreasing order) Sweden, the World Bank, West Germany, the Netherlands, Denmark and Norway (Hyden and Karlstrom, 1993). Aid flows peaked in around 1980, and were comparatively resilient, especially from Scandinavian donors, even with increasing pressure from the World Bank and the IMF to "modify the rigid domestic price system and exchange rate policy, the marketing policy in the agricultural sector and related issues" (Hyden and Karlstrom, 1993:1398). Domestic borrowing, which remained low at the beginning of the period partly due to increasing inflows of aid, started to increase rapidly in late seventies (Figures 5.1, 5.5). As the deficits were financed through the Central Bank, the inflationary pressures were rapidly increasing, and imports collapsing. The official statistics show a decline of 0.5 per cent in real GDP per capita during 1965-1985, although household surveys indicate a much more severe decline: "over 15 year period to 1984, real income per household fell by roughly 50 per cent" (Hyden and Karlstrom, 1993:1399). As the government started sanctioning those with 'above average private capital' toward the mid-1980s, even the most loyal donors like Sweden started realigning their views towards the IMF.

The reforms started in 1986, after Nyerere's resignation (and succession of Mwinyi), as the government launched a three year Economic Recovery Programme (ERP) and signed the agreement with the IMF. The programme entailed a mixture of short-term stabilisation measures (both fiscal and monetary), growth stimulating policies (with aid-funded rehabilitation of physical infrastructure), liberalisation of both domestic and foreign trade policies, gradual adjustments to the exchange rate and agricultural prices. As the programme expired, the government negotiated the Economic and Social Adjustment Programme (ESAP) with the World Bank with similar conditionalities attached, further accelerating the reform process in the very early 1990s with privatisation of the key sectors in the economy, establishing a private banking system and allowing foreign ownership enterprises and a multiparty system. The pro-market policies, however, had some adverse effects on the government capacity, as the civil servants increasingly turned to private sector for employment.

Although the programmes delivered economic growth and macroeconomic and structural reforms, the dialogue between the Tanzanian government and the donors deteriorated substantially in 1993-1994, due to poor fiscal performance in 1993/4, suspicions of corruption,[198] 'lack of will',[199] reached a critical decision (crisis) point in 1994, when IMF and World Bank programmes were put on hold, and non-project finance from the principle donors suspended (Helleiner et al., 1995:3). An independent committee, headed by Helleiner, was set up, with the 'minimum' issues to be assessed being (with the emphasis on the importance of 'ownership'):

> "1. The efficiency and relevance of the current dialogue between GOT [Government of Tanzania – E.T.] and donor community regularly taking place both inside and outside Tanzania.
> "2. The relevance and effectiveness of the totality of aid programmes, including the modes, composition and administration of cooperation (programme aid, project aid, technical assistance, etc.); conditionalities; donor cooperation; absorption capacity of the Tanzanian economy and the institutions through which the aid is channelled; problems of accountability" (Helleiner et al., 1995:2).

The severity of the crisis is thus evident.

[198] It is acknowledged that accompanying political transition may slow 'some elements of economic policy reform and institutional change', as well as highlight – or exaggerate – (for instance, though newly freer press) less attractive features of the transitional (i.e. from command to market) economic order (Helleiner et al., 1995:8).
[199] In return, the government of Tanzania regarded donors' demands unrealistic and impatient, intrusive with respect to domestic policy, as well as donors' 'lack of trust' and 'unwillingness to share information' (Helleiner et al., 1995:3).

The newly elected government, led by President Mkapa, addressed the macroeconomic situation without much delay, restored fiscal control mainly though introduction of "a cash management system which left no room for expenditures beyond the limits set by the revenue collections" (Helleiner, 2001:3), and responded to many recommendations of the Helleiner et al. (1995) report, and by the end of 1996 agreed on a three year Enhanced Structural Adjustment Facility (ESAF) with the IMF. The government's free market-oriented policies included privatisation of state-owned enterprises, among other liberalisation policies and structural and institutional reforms. The relationships with key Nordic donors were also restored by the beginning of 1997 (essentially rendering the composition of the key donors unchanged since the independence). The recipient's policy ownership was a clear characteristic of National Development Vision 2025, established in 1997, and National Poverty Eradication Strategy, issued in 1998, and way ahead of IMF's Poverty Reduction Strategy Papers (PRSPs) or World Bank's Comprehensive Development Frameworks (CDFs), although these were also later prepared in 2000 as part of the HIPC initiative[200] (Helleiner, 2001)[201]. Tanzanian Assistance Strategy (2002-2005), later developed into the Joint Assistance Strategy for Tanzania (JAST, 2006-2011) aimed to improve donor coordination and integrate more external resources under the government budget and exchequer system.[202]

Most notably (**Figure 5.1**), post-1995 reforms, the total spending has accelerated without a matched increase in the collection of the domestic revenue (although that, too, was growing). Partly, the widening gap between expenditure and domestic revenue can be filled by increasing aid (although this may not be sustainable in the very long run); inevitably, however, the shortfalls have to be covered by public borrowing (both domestic, and, increasingly, non-concessional borrowing abroad), which has been increasing (and increasingly volatile) since early 2000s. On the other hand, in National Debt Strategy (2002) Tanzanian government claimed that "[g]oing forward, and in the context of our move

---

[200] HIPC for Tanzania was completed in November 2001 (http://www.who.int/immunization_financing/analyses/debt_relief/country_data/Tanzania_datasheet_final.pdf).

[201] Note, interestingly, that technical assistance was regarded by the Government of Tanzania 'as unnecessarily wasteful use of scarce aid resources, contributing little either to local human resource use (employment) or to capacity-building' (Helleiner, 2001:7).

[202] Tanzanian Government Brochure https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=6&cad=rja&ved=0CFUQFjAF&url=http%3A%2F%2Fwww.tzdpg.or.tz%2Findex.php%3FeID%3Dtx_nawsecuredl%26u%3D0%26file%3Duploads%2Fmedia%2FBusanBrochure_GoTDPG.pdf%26t%3D1455886306%26hash%3D2b95f4a30a0e79cda11a9799656ac0783a2c4747&ei=QGl6Uo22MLPa4QSshID4DQ&usg=AFQjCNFpIKkE1dyhfLO9x-qOUv65XBwGdw&bvm=bv.55980276,d.bGE

towards self-reliance, the domestic debt market is expected to become an increasingly important funding source for the government", partly explaining increasing public borrowing since the HIPC Completion Point was reached in November 2001[203], [204],[205],[206].

Table 5.1 below provides the key summary statistics (means and standard deviations) for all variables used in the analyses. Variables are expressed as percentages of GDP to provide a conventional comparison, although one must note the GDP rebasing in 1987; and as percentage of total government expenditure to provide consistent information throughout the period (given the caveats of GDP data). Domestically collected government revenue – mainly tax – has been the key source of revenue throughout the whole period, constituting about three-quarters of government spending. Budget aid's contribution to total spending has increased post 1986 reforms from 12.2 to 30.95 per cent of total government expenditure. This was primarily driven by the increments in grants, as these average figures for contribution from loans remained relatively stable throughout the whole observed period.[207]

The composition of government spending itself, too, has changed over time. Although in absolute terms both development and recurrent spending has increased, the relative recurrent spending has increased compared to development spending. Government's domestic borrowing (public financing) has differed substantially between the two periods: whilst before the market-oriented reforms government on average borrowed (excluding aid loans) nearly 16 per cent of its total expenditures, since 1987 the public borrowing only constituted 2.4 per cent of total expenditures (and less than 1 per cent of GDP).[208]

Overall, the fiscal relationships seem to change post mid-1980s, and, from the qualitative perspective, it would be ideal to analyse each period separately. However, each of the subsamples would be far too short for the quantitative (CVAR) analysis. As the next best

[203] http://www.mof.go.tz/mofdocs/debt/nationaldebtstrategy.pdf
[204] http://www.imf.org/external/pubs/ft/dsa/pdf/2012/dsacr12185.pdf
[205] http://www.mof.go.tz/index.php?option=com_content&view=article&id=48:finance-a-debt-policy&Itemid=63
[206] The medium term report argues that increasing development spending needs motivate more borrowing (both domestic, and external); debt composition in 2010 was 0.7/0.3, domestic/foreign, respectively; domestic borrowing agreed with IMF not to exceed 1% GDP. Government projected revenue increments.
[207] Aid loans increased substantially during the 'transition' period in 1987.
[208] Note that the revenue and spending components are not equal as some spending elements (such as loan servicing) are omitted from the original data.

alternative, a shift dummy will be added to the cointegrating space aiming to capture the different relationships before and after the 1986 reforms.[209]

Table 5.1: Summary Statistics

	Variable name	Variables as proportion of GDP (%)			Variables as proportion of Total Government Expenditure (%)		
		Full sample (1966-2012)	Before GDP rebasing/ reforms (1966-1986)	Post GDP rebasing/ reforms (1987-2010)	Full sample (1966-2012)	Pre-1986 reforms	Post 1986 reforms
Total expenditure	*TEXP*	20.10 (6.41)	24.03 (4.52)	16.93 (6.00)	1	1	1
Development expenditure	*DEXP*	5.86 (3.27)	7.77 (2.83)	4.32 (2.79)	26.98 (9.28)	31.63 (7.41)	23.22 (9.04)
Recurrent expenditure	*REXP*	14.24 (3.61)	16.26 (2.75)	12.61 (3.43)	73.02 (9.28)	68.37 (7.42)	76.78 (9.04)
Total Domestic revenue	*DREV*	14.31 (3.18)	16.83 (2.08)	12.27 (2.34)	75.01 (16.37)	71.22 (7.97)	78.06 (20.51)
Tax revenue	*TAX*	13.46 (3.74)	16.36 (3.11)	11.12 (2.29)	69.47 (14.28)	68.45 (9.50)	70.30 (17.36)
Non-Tax revenue	*NTAX*	0.85 (1.01)	0.47 (1.41)	1.15 (0.26)	5.53 (6.61)	2.77 (8.41)	7.76 (3.49)
Total budget aid	*AID*	4.43 (2.72)	3.07 (1.46)	5.54 (3.01)	22.59 (12.62)	12.24 (4.47)	30.95 (10.69)
Grants	*GRANTS*	2.66 (1.79)	1.34 (1.03)	3.73 (1.55)	14.56 (10.30)	5.11 (3.73)	22.20 (6.97)
Loans	*LOANS*	1.77 (1.37)	1.73 (0.70)	1.81 (1.76)	8.03 (7.28)	7.13 (2.34)	8.76 (9.59)
Borrowing	*BORROW*	2.01 (2.39)	3.89 (1.92)	0.50 (1.50)	8.45 (10.52)	15.94 (6.24)	2.41 (9.35)

Note: standard errors are reported in parentheses.

3. Empirical Analyses

Using the data (logged values, denoted by lower case, deflated using the CPI) described in Section 2, the empirical analyses are conducted using the cointegrated vector autoregressive framework (CVAR, Juselius, 2006). In the VAR framework, each variable is modelled as

[209] For a similar discussion on the macroeconomic (rather than) fiscal indicators and the related CVAR modelling, see Juselius et al. (2013).

endogenous, and is expressed as a function of past own values, as well as past realisations of other variables (and deterministic components). The vector error-correction model (VECM) representation of the VAR includes both the stationary first differences of variables in x_t (Δx_t), and their value in levels (x_t), thus preserving both the long-run and short-run information in the data. In particular, the error correction form of the VAR (VECM) is represented by the following equation:

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi D_t + \varepsilon_t \tag{5.1}$$

$x_t = texp_t, tax_t, aid_t$

where x_t is a $p \times 1$ vector of endogenous variables described above, D_t is a vector of deterministic components (such as constant, deterministic trend, and dummy variables) with a vector of coefficients Φ; k denotes the selected lag length; ε_t is a $p \times 1$ vector of unobservable error terms, that are assumed to be $\varepsilon_t \sim IN(0, \Omega)$. VECM allows a clear separation between the long-run coefficients in Π and the short-run coefficients in Γ_i.

The VECM representation illustrates that if variables are found to be I(1) – and macroeconomic variables usually are – stationary variables (Δx_t) are regressed on unit-root processes (x_{t-1}). In such case, the estimated coefficients would be spurious. However, if some variables in the system are driven by the same persistent shocks, there may exist linear combinations of these variables that are integrated of the lower order than the variables themselves (i.e. I(0)). These linear combinations would represent cointegrated relations, $\beta' x_t$, and could be interpreted as the long-run steady-state relationships. When cointegration exists, Π has reduced rank $r < p$ and is defined as follows:

$$\Pi = \alpha \beta' \tag{5.2}$$

where α and β are $p \times r$ matrices (with $r < p$); $\beta' x_t$ defines the stationary long-run cointegrating relationships ($r \times 1$), and α denotes the adjustment coefficients to the equilibrium error. Intuitively, if all $x_t \sim I(1)$ and $\Delta x_t \sim I(0)$, then a full rank in Π would be logically inconsistent as it would imply that x_t must be stationary.[210] On the other hand, $r = 0$ implies that each variable in x_t is non-stationary and is driven by its own individual

[210] The VECM representation of the VAR with full rank in Π and $x_t \sim I(1)$ would imply that a stationary variable Δx_t equals a non stationary variable x_{t-1}, lagged stationary variables Δx_{t-1} and a stationary error term. Since a stationary variable cannot equal a non-stationary variable, either $\Pi = 0$ or it would have reduced rank.

stochastic trend and therefore no cointegration exists. In this case, a simple VAR model with the variables in first differences would not imply any loss in long-run information.

The accompanying moving average (MA) representation of the VAR illustrates how the process can be described in terms of pulling and pushing forces. The steady state to which the process is pulled to is defined by the long run relations $\beta' x_t - \beta_0 = 0.$ The forces α represent adjustment and they activate as soon as the process is out of steady state, i.e. when $\beta' x_t - \beta_0 \neq 0$ (Juselius 2006: 88-89). The MA representation describes the non-stationary movement of the variables according to the common driving trends that represent the cumulated sum of the shocks to the system. "In this sense, the AR and MA representation are two sides of the same coin: the pulling and the pushing forces of the system" (Juselius, 2006:88). The inverted model can be summarised as:

$$x_t = C \sum_{i=1}^{t} (\varepsilon_i + \Phi D_i) + C^*(L)(\varepsilon_i + \Phi D_i) + X_o \tag{5.3}$$

where $C = \beta_\perp(\alpha'_\perp(I - \Gamma_1)\beta_\perp)^{-1}\alpha'_\perp$ is the long-run impact matrix of rank *p-r*, with $\alpha'_\perp\varepsilon_t$ describing the common driving trends; $C^*(L)$ is a stationary lag polynomial, and X_o depends on the initial values.

The VAR model is highly demanding of the data. Therefore, we aim to estimate four distinct models:

- a three-dimensional VAR between total expenditure (***texp***), tax revenue (***tax***) and aggregated aid (***aid***):
  $x_t = texp_t, tax_t, aid_t$
- a four-dimensional model with total central government spending disaggregated into its development and recurrent components (***dexp*** and ***rexp***, respectively), tax revenue, and aid:
  $x_t = dexp_t, rexp_t, tax_t, aid_t$
- a four-dimensional model with aggregated government spending, tax revenue, and aid disaggregated into grants (***grants***) and loans (***loans***)[211]:
  $x_t = texp_t, tax_t, grants_t, loans_t$

[211] Transformed to allow for logarithmic transformation (See section 3.3).

- a four dimensional model with government spending, tax revenue, aid, and borrowing (***borrow***)[212],
  $x_t = texp_t, tax_t, aid_t, borrow_t$

The results of each model are discussed in the following subsections. As the relevant tests and other methodology are discussed in Section 4 of Chapter 4, we resist reproducing it at length here and discuss the results in parsimonious fashion, referring to the relevant sections of Chapter 4 where necessary.

3.1 Aggregated model[213]

As a point of departure, the simplest fiscal model is estimated with total expenditure, tax revenue and aggregated aid, all in logs. There are three reasons for such simplification. Firstly and most importantly, given the relatively small number of observations from a purely time-series perspective, a system with a minimal number of variables should yield the most robust results, as it would be least demanding on the limited data. Secondly, the disaggregation of domestic revenue is complicated as non-tax revenue is recorded as 0 for the period of 1968-1982, and thereafter is considerably smaller than tax revenue.[214] In the light of this, the analysis is restricted to tax variable only. Thirdly, disaggregation of aid into grants and loans is complicated by the five years of negative recorded loans. Whilst it is not unusual to observe some years of net repayment, this complicates the analysis conducted in logs. The budget financing also contains negative values, and is thus also excluded from the analysis here. (The disaggregation of total expenditure into recurrent and development components is possible, and the analysis is conducted in the next section).

3.1.1 Misspecification Tests

Lag length

Table 5.2 below provides the lag length testing results. Whilst both Schwarz and Hannah-Quinn information criteria indicate preference for lag length of one (*k=1*), such a model would exhibit second order autocorrelation issues. Therefore, lag length of two (*k=2*) is chosen for this model.

---

[212] Transformed to allow for logarithmic transformation (See Section 3.4).

[213] @cats(lags=2,det=drift,break=level,dum) 1966:1 2012:1
L_TEXP L_TAX L_AID
1986:1
dum96p

[214] *Non-tax* revenue is equal to on average 10% of *tax* revenue for the period from 1983, with a maximum of 16%.

Table 5.2: Lag-length Determination

Model	*k*	*T*	*Regr.*	*Log-lik*	*SC*	*H-Q*	*LM(1)*	*LM(k)*
VAR(5)	5	42	19	283.719	-8.438	-9.932	0.233	0.587
VAR(4)	4	42	16	273.227	-8.739	-9.997	0.060	0.300
VAR(3)	3	42	13	265.643	-9.179	-10.201	0.114	0.025
VAR(2)	2	42	10	261.010	-9.759	-10.546	0.367	0.018
VAR(1)	1	42	7	257.954	-10.415	-10.965	0.559	0.559

Effective Sample: 1971:01 to 2012:01
SC : Schwarz Criterion; H-Q : Hannan-Quinn Criterion
LM(k): LM-Test for autocorrelation of order k

Deterministics

An unrestricted constant is included to allow for a non-zero mean in the cointegrating relations and for non-quadratic trends in levels.[215] A permanent dummy in 1996, signalled by a large residual (over 3) in aid, captures the aid reforms following the Helleiner et al. (1995) report. Statistically, it ensures the normality of residuals in the aid equation. Finally, although not suggested by excessively large residuals, a shift dummy is included in 1986 to capture the introduction of market reforms to the economy. Whilst such inclusion is driven more by the economic rather than purely statistical reasoning and indeed affects the statistical results, it ensures economic interpretability of the estimated cointegrating relationships.[216]

Residuals

The residuals from the unrestricted VAR exhibit an excellent model fit (Table 5.3). The multivariate normality is not rejected (p-value = 0.766), nor is the univariate normality for each of the model variables. The trace correlation statistic of 0.480 indicates a good model fit. Table 5.3 also summarises the results for LM tests for autocorrelation up to fourth order. Whilst the second order test statistic is borderline, others indicated no residual autocorrelation. Whilst less relevant with annual data, the tests for residual heteroskedasticity are also supplied in Table 5.3 for completeness and indicate no significant ARCH effects. Appendix Figure D1 illustrates the residual fit visually.

[215] A model with a trend restricted to cointegration space was tested for, and the variable exclusion tests concluded that such trend can be excluded from the model (see Appendix Table D2). However, the estimated long run results are qualitatively comparable from both models.
[216] Without the inclusion of the shift dummy, it is virtually impossible to statistically justify the choice of r=2 vs. r=1; with r=1, the result is incredibly difficult to interpret, as the total expenditure ends up being positively associated with tax in the long run (sensible) but negatively associated with aid (difficult to interpret without purely external speculation).

Table 5.3: Residuals from Unrestricted VAR

Residual normality (p-values)				
	Multivariate	Univariate		
		texp	tax	aid
	0.766	0.671	0.260	0.918
Residual autocorrelation and ARCH effects (p-values)				
	LM(1)	LM(2)	LM(3)	LM(4)
Residual autocorrelation	0.477	0.045	0.106	0.424
ARCH	0.891	0.682	0.401	0.308
Trace correlation	0.480			

Note: All values are p-values

3.1.2 Determination of Cointegration Rank

Given the small sample and the model deterministics, we simulate the critical values for the Johansen test for the cointegration rank. Irrespective of whether we use the Bartlett-corrected values or not, the Johansen test suggests two cointegrating relationships (*r=2*) (Table 5.4). Juselius (2006:142) suggests consulting additional information for the critical choice of the cointegration rank (see **Appendix Table D4**). The characteristic roots of the model confirm this choice, indicating no large moduli once one common trend (*p-r=3-2=1*) is included in the model. The *t*-values of the alpha coefficients to the $(r+1)^{th}=3^{rd}$ cointegrating vector are all below 2.6, whilst at least one exceeds this value in the second cointegrating vector, again supporting the choice *r=2*. The recursive graphs illustrate the linear growth over time for the two out of the three components. The graphs of potential cointegrating relationships also illustrate potential stationarity of two out of three relationships. As the criteria of economic interpretability of the results is also satisfied (see next section), cointegration rank of two (*r=2*) can be selected with confidence.

Table 5.4: Trace Test

p-r	r	Eig. value	Trace	Trace*	Frac95	p-value	p-value*
3	0	0.526	57.363	50.889	28.295	0.000	0.000
2	1	0.365	23.790	21.068	15.035	0.002	0.006
1	2	0.072	3.346	3.136	3.870	0.063	0.072

** denotes Bartlett corrections*

3.1.3 Long-run identification: Hypothesis Testing

The formal discussion of the battery of long-run identification procedures is provided in Section 4.3 of Chapter 4. The tests of long-run exclusion (Table 5.5) demonstrate that none of the variables (nor the shift dummy) should be excluded from the cointegrating space.

Variable stationarity tests show that none of the variables are stationary around a (broken) mean or trend. Tests of weak exogeneity indicate that at the selected cointegration rank (*r=2*) none of the variables can be accepted as weakly exogenous (tax would be accepted as such if *r=1* was chosen). This is taken as a mild indication when identifying the common trends. Finally, the unit vector in alpha tests indicate that, irrespective of the choice of the cointegration rank, aid may be accepted as purely adjusting to the long-run equilibrium error.

Table 5.5: Long Run Identification Tests

	r	***texp***	***tax***	***aid***	Shift 1986
Long-run exclusion	*r=2*	0.000	0.000	0.000	0.001
	r=1	0.464	0.052	0.008	0.041
Stationarity	*r=2*	0.000	0.000	0.000	yes
	r=1	0.000	0.000	0.000	yes
Stationarity	*r=2*	0.000	0.000	0.000	no
	r=1	0.000	0.000	0.000	no
Stationarity (trend)	*r=2*	0.000	0.000	0.007	no
	r=1	0.000	0.000	0.007	no
Stationarity (trend)	*r=2*	0.000	0.000	0.002	yes
	r=1	0.000	0.000	0.004	yes
Weak exogeneity	*r=2*	0.003	0.027	0.000	-
	r=1	0.029	0.592	0.000	-
Purely adjusting	*r=2*	0.012	0.010	0.204	-
	r=1	0.000	0.000	0.092	-

The table reports p-values.

Individual Hypothesis Testing

In this section we test whether our hypothesised relationships are *individually*[217] stationary. Keeping the remaining $r-1$ cointegrating relationships unrestricted (and thus unidentified), zero (or homogeneity) restrictions in a particular equilibrium relationship may be tested, allowing the remaining parameters to be estimated. Table 5.6 below provides the results of testing for individually stationary or otherwise relationships among the variables for an unidentified long-run system with two cointegrating vectors. Such hypothesis testing indicates that while a strong positive very-long-run relationship between total government expenditure and tax revenue could be identified (although not one-for-one), any other variable combinations (a positive relationship between expenditure and aid, or tax and aid) would require the inclusion of the 1986 shift dummy to be stationary.

[217] As stated above, our main aim is to see whether the individually stationary cointegrating relationships hold together as an equilibrium system.

The Π matrix (Appendix Table D4) of unidentified system (with r=2) also contains interesting information for future long-run identification. The total expenditure may be significantly associated with tax but not necessarily aid. Tax in turn should bear a strong association with total government expenditure, and possibly aid. Aid exhibits a potential relationship with tax variable, but weaker relationship with the total expenditure. This confirms the unsurprising strong association between the domestically determined variables, and suggests somewhat counter-intuitively that the second equilibrium may be between aid and tax rather than aid and expenditure.

Table 5.6: Stationarity (or otherwise) of Variable Combinations

	texp	*tax*	*aid*	*Ds1986*	p-value
H1	1	-1.2505	0	0	0.48074
H2	1	-1.3555	0.0610	0	1
H3	1	-0	-0.8516	0.7511	1
H4	1	0	-08159	0	0.0005
H5	0	1	-0.6732	0.5541	1
H6	0	1	-0.6400	0	0.0003
H7	1	-1	0	0	0.0022

Note: Zeroes in the table are imposed and not estimated.
Also note that where r –1 = 1 conditions are imposed, the relationships are just-identified and therefore p-value is 1 by construction (i.e. restrictions are not testable) (Juselius, 2006:189).

3.1.4 Long Run Identification: Results

It is worthwhile to distinguish between a *just-identified* structure (with $r(r-1)$ identifying restrictions), where $r-1$ identifying (usually zero, or homogeneity) restrictions are imposed for each of r cointegrating relationships; and *over-identifying* restrictions, whereby more than $r-1$ identifying restrictions are imposed[218] for at least one of the cointegrating vectors. Whilst just-identifying restrictions do not change the value of the likelihood function as they do not constrain the parameter space, the over-identifying restrictions do, and therefore can be tested. Note that normalisation on one element in each vector does not change the likelihood as the corresponding α_i coefficient is normalised on the same β_i coefficient. However, once we have identified a long run structure, the normalisation is an important choice, as we do not want to normalise on an insignificant variable) (Juselius,

[218] I.e. the rank conditions are met, meaning no linear combination of other r-1 CI relations may produce a vector that resembles the first one (see Juselius, 2006:209-210, which further cites Johansen and Juselius, 1994, and Johansen, 1995). For further discussion of three aspects of identification (generic, empirical, and economic) see Juselius (2006:208).

2006:214). Results are provided in Table 5.7. As an over-identifying restriction, we exclude the shift dummy from the first cointegrating vector, as it is reported as insignificant. Results are provided in Table 5.8.

Table 5.7: Just-identified Model (Aggregated)

	texp	*tax*	*aid*	Shift 1986	
LR equilibrium relation (β_1)	**1.000**	**-1.27**	0	0.05	~I(0)
		(-20.74)		(0.84)	
LR equilibrium relation (β_2)	**-1.17**	0	**1.000**	-0.88	~I(0)
	(-9.15)			(-5.68)	
Adjustment coefficients (α_1)	**-0.37**	**0.30**	**-1.20**		
	(-3.59)	(2.91)	(-3.67)		
Adjustment coefficients (α_2)	**-0.02**	-0.06	**-0.51**		
	(-0.46)	(-1.56)	(-4.5)		
Multivariate normality	p-value = '-'				
Stationarity	0.481				
Trace correlation	0.463				
	Log-Likelihood = 272.046				

Note: t-statistics are reported in the parentheses.

Table 5.8: Over-identified Model (Aggregated)

	texp	*tax*	*aid*	Shift 1986	
LR equilibrium relation (β_1)	**1.000**	**-1.25**	0	0	~I(0)
		(-20.75)			
LR equilibrium relation (β_2)	**-1.18**	0	**1.000**	**-0.871**	~I(0)
	(-9.185)			(-5.625)	
Adjustment coefficients (α_1)	**-0.38**	**0.27**	**-1.183**		
	(-3.8)	(2.64)	(-3.74)		
Adjustment coefficients (α_2)	-0.01	-0.06	-0.5		
	(-0.289	(-1.54)	(-4.395)		
Multivariate normality	p-value = 0.862				
Stationarity	0.481				
Trace correlation	0.463				
	Log-Likelihood = 271.797				

Note: t-statistics are reported in the parentheses.

Guided by the previous indications, we identify the first cointegrating vector as a relationship between domestic variables (expenditure and tax) by imposing a zero restriction on aid (although coefficients are fairly similar, homogeneity restriction cannot be plausibly imposed). This relationship summarises the long run positive equilibrium relationship between government expenditure and tax revenue. Should there be a departure from this long-run equilibrium, both variables would adjust with similar speed (about 2.5 years), although the expenditure adjustment coefficient is estimated more precisely. Aid also adjusts, possibly quickly filling the deficit.

The second equilibrium is then identified as a positive relationship between aid and government total expenditure, by imposing a zero restriction on tax variable. Although the statistical results prioritise the association between aid and tax, the aid – expenditure relationship is of primary focus as theoretically this should entail a more direct effect (and more economically interpretable result).[219] The estimated result concurs with the postulated expectation: total government expenditure is positively related to aid in the long run, and deviations from this equilibrium would trigger aid to adjust: i.e. if there was an increase in government spending, aid would match the increase in about two years, and vice versa. The shift dummy's coefficient is significant, indicating that market reforms have altered the relationship between these two variables. The cointegrating relationships are depicted in Appendix Figure D2.

3.1.5 MA representation: common driving trends

The moving-average representation of the VAR allows one to inspect the driving forces in the model. The *p-r* common trends (CT) describe the non-stationarity in the process, originating from the cumulative sum of the unanticipated shocks. The tests above indicated that only tax could be potentially considered as weakly exogenous variable. However, from alpha orthogonal corresponding to the over-identified model (without imposing weak exogeneity of the tax variable) (see Table 5.9), it does not seem that cumulated residuals to tax variable alone could be considered a common stochastic trend – the total government expenditure residuals contribute, too, with a similar weight.

By reading the C-matrix (Table 5.10), we can elicit how the cumulated residuals from each VAR equation load into each of the variables (column-wise inspection) and the weights with which each variable in the system has been affected by any of the cumulated empirical shocks (row-wise inspection) (Juselius, 2006:259). Shocks (although note that residuals are correlated) to both government expenditure and tax revenue have comparable impact on the system variables, strongly, positively and permanently affecting all three variables. Aid, on the other hand, exhibits only a mild negative, transient at most effect.

[219] The results from alternative identification are available in Appendix Table D5.

Table 5.9: Common Trends (Over-identified Model)

		texp	*tax*	*aid*
Composition of common trends (α⊥)	CT1	**1.139**	**1.000**	-0.137
		(1.745)		(-1.110)
Loadings of common trends (β⊥)	CT1	**1.003**	**0.802**	**1.180**
		(2.732)	(2.732)	(2.732)

Note: t-statistics are reported in the parentheses.

Table 5.10: The Long Run Impulse Matrix C (Over-identified Model)

(Over-)identified model	$\hat{\varepsilon}_{texp}$	$\hat{\varepsilon}_{tax}$	$\hat{\varepsilon}_{aid}$
texp	**1.143**	**1.003**	-0.138
	(2.428)	**(2.732)**	(-1.192)
tax	**0.914**	**0.802**	-0.110
	(2.428)	**(2.732)**	(-1.192)
aid	**1.344**	**1.180**	-0.162
	(2.428)	**(2.732)**	(-1.192)

Note: t-statistics are reported in the parentheses.

Overall, the aggregated system provides sensible core results. There exists a domestic fiscal equilibrium which describes that over the long run domestic revenue and expenditure are closely related. There is also a positive association between aid and government spending, with aid adjusting to domestic fiscal decisions rather than driving them. Alternatively, a positive association between aid and tax could be identified, reassuring donors that aid does not have adverse effects on domestically collected tax revenue.

3.2 Model with Disaggregated expenditures[220]

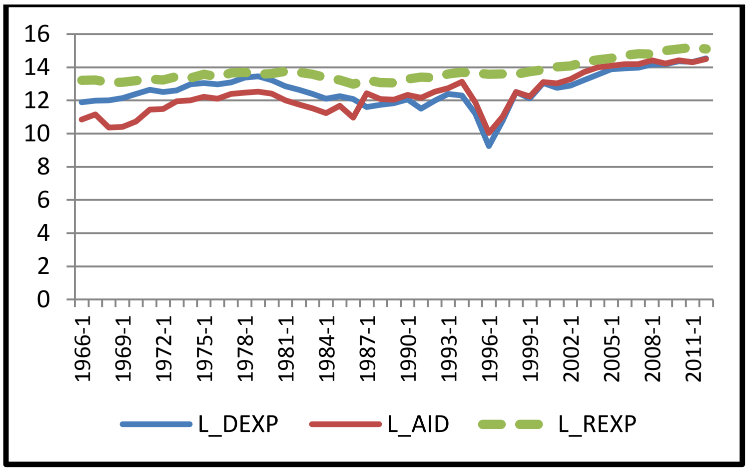
Ethiopian chapter asked whether aid has differing effects on the disaggregated government expenditure components, namely, development and recurrent spending. Whilst disaggregating aid into grants and loans poses some challenges, it is rather straightforward to formulate a model with disaggregated expenditure, and government tax revenue (or, alternatively, aggregated domestic revenue) and aid (grants plus loans).

One interesting thing to note from the data is the collapse in the development spending in 1996 coinciding with the collapse in the aid (Figure 5.6). Qualitatively, as there was no visible

[220] @cats(lags=2,det=cidrift,break=level,dum) 1966:1 2012:1
L_DEXP L_REXP L_TAX L_AID
1986:1
dum96p

dip in the recurrent spending, it is interesting as it suggests that donors' decision to suspend aid translated predominantly (if not solely) in reduction in the category of spending arguably more important for the long run. Development expenditure then recovered in line with the recovery of aid flows (grants and loans). Visually, development expenditures tend to move more in tandem with aid flows compared to recurrent expenditures, which appear to be more stable. This is sensible considering that public sector salaries constitute a considerable proportion recurrent expenditure, and consistent with tax being linked to recurrent spending and aid more likely to finance development/investment.

Figure 5.6: Selected Variables (Logs)



3.2.1 Misspecification Tests

Lag length

Lag length of two (*k=2*) is selected for this model (Table 5.11). As before, the information criteria indicate preference for lag length of one, and even the autocorrelation test statistics for *k=1* outperform those for *k=2*. However, for consistency, and improved economic interpretation of the results, *k=2* is accepted as superior specification.[221]

Table 5.11: Lag Length Selection (Model with Disaggregated Spending)

Model	*k*	*T*	*Regr.*	*Log-lik*	*SC*	*H-Q*	*LM(1)*	*LM(k)*
VAR(5)	5	42	25	400.270	-10.161	-12.782	0.109	0.108
VAR(4)	4	42	21	375.760	-10.418	-12.619	0.004	0.020
VAR(3)	3	42	17	361.396	-11.158	-12.940	0.081	0.321
VAR(2)	2	42	13	348.785	-11.981	-13.344	0.038	0.040
VAR(1)	1	42	9	334.377	-12.719	-13.662	0.251	0.251

Effective Sample: 1971:01 to 2012:01.

[221] k=1, although improves AC, completely disrupts the cointegration rank test results, and the estimated beta vectors are less sensible (no relationship between aid and development expenditure).

Deterministics

As above, a permanent dummy in 1996, signalled by a large residual (over 3) in aid (and now development expenditures, see Figure 5.6), captures the aid reforms following the Helleiner et al. (1995) report. Again, although not suggested by excessively large residuals, a shift dummy is included in 1986 to capture the introduction of market reforms to the economy. A trend restricted to cointegrating space cannot be excluded from the model. An interesting indication following from Figure 5.6 here is that the reduction in aid predating the 1995/6 reforms translated predominantly into reduction of development spending, having no visible effect on the recurrent expenditures.

Residuals

Compared to the aggregated model, the residuals (Table 5.12) illustrate a slightly inferior model fit, which is somewhat unsurprising given the extra demands on the data following from the increased number of variables. The multivariate normality is strongly rejected (p=0.001), primarily driven by the rejection of univariate normality in development expenditures variable, even with the aforementioned dummy structure. Model also exhibits some first order residual autocorrelation. The trace statistic of 0.575 suggests an acceptable model fit. The visual representation of the residuals is presented in Appendix Figure D3.

Table 5.12: Residuals from Unrestricted VAR (Model with Disaggregated Spending)

Residual normality (p-values)					
	Multivariate	Univariate			
		dexp	***rexp***	***tax***	***aid***
	0.001	0.020	0.833	0.104	0.449
Residual autocorrelation and ARCH effects (p-values)					
		LM(1)	LM(2)	LM(3)	LM(4)
Residual autocorrelation		0.010	0.462	0.354	0.079
ARCH		1.000	0.629	0.048	0.042
Trace correlation		0.575			

Note: The table reports p-values.

3.2.2 Determination of Cointegration Rank

Irrespective of whether we use the Bartlett-corrected values or not, the Johansen test (Table 5.13) suggests two cointegrating relationships (*r=2*). Additional information supports this choice (reported in the Appendix Table D6).

Table 5.13: Trace Test

p-r	r	Eig. value	Trace	Trace*	Frac95	p-value	p-value*
4	0	0.744	129.632	114.691	71.725	0.000	0.000
3	1	0.596	68.231	60.448	49.698	0.000	0.003
2	2	0.339	27.490	24.675	30.508	0.109	0.205
1	3	0.178	8.838	8.248	15.727	0.376	0.433

** denotes Bartlett corrections*

3.2.3 Long Run Identification: Hypothesis Testing

None of the variables are trend- or mean stationary, irrespective of whether the mean shift is accounted for, or not (see Table 5.14). None could be excluded at the selected cointegration rank. In terms of adjustment behaviour, tax is indicated as weakly exogenous (for *r=2*), and none of the variables can be accepted as purely adjusting.

Table 5.14: Long Run Identification Tests (Model with Disaggregated Spending)

p-values	*r*	***dexp***	***rexp***	***tax***	***aid***	Shift 1986	Trend
Long-run exclusion	*r=3*	0.000	0.000	0.000	0.000	0.002	0.073
	r=2	0.000	0.001	0.001	0.000	0.004	0.047
	r=1	0.000	0.041	0.881	0.006	0.001	0.419
Stationarity	*r=3*	0.001	0.001	0.001	0.006	no	no
	r=2	0.000	0.000	0.000	0.000	no	no
	r=1	0.000	0.000	0.000	0.000	no	no
Stationarity	*r=3*	0.000	0.002	0.002	0.003	yes	no
	r=2	0.000	0.000	0.000	0.000	yes	no
	r=1	0.000	0.000	0.000	0.000	yes	no
Stationarity	*r=3*	0.001	0.001	0.000	0.029	no	yes
	r=2	0.000	0.000	0.000	0.000	no	yes
	r=1	0.000	0.000	0.000	0.000	no	yes
Stationarity	*r=3*	0.002	0.002	0.002	0.048	yes	yes
	r=2	0.001	0.000	0.000	0.001	yes	yes
	r=1	0.000	0.000	0.000	0.000	yes	yes
Weak exogeneity	*r=3*	0.000	0.004	0.096	0.000	-	-
	r=2	0.000	0.004	0.316	0.000	-	-
	r=1	0.006	0.006	0.955	0.879	-	-
Purely adjusting	*r=3*	0.083	0.037	0.072	0.247		
	r=2	0.014	0.000	0.000	0.004		
	r=1	0.031	0.000	0.000	0.000		

Note: The table reports p-values.

3.2.4 Long Run Identification: Results (Model with Disaggregated Spending)

The key interest in this specification[222] is finding to which financing component – domestic (tax) or foreign (aid) – each of the government spending components are more strongly associated to. Therefore, one of the vectors looks at the development expenditures (excluding the recurrent spending), and the second at the recurrent expenditures (excluding the development spending). Individual hypothesis testing (reported in Appendix Table D7) report that two such (over-identified excluding insignificant variables once the system is estimated) relationships would be individually stationary. For readability purposes, each vector is normalised on each of the spending components rather than the most adjusting variable. The results summarised in Table 5.15 indicate that trend is not significant in the first cointegrating vector, and the shift dummy is insignificant in the second relationship. They are therefore excluded from the respective vectors, and the results of the over-identified model are provided in Table 5.16.

The results suggest that central government's development spending is strongly associated with domestic tax revenue, and, to a lesser extent, aid (which is sensible, as aid funded just over a fifth of total government spending over the full sample period). Should there be a departure from this equilibrium (e.g. a fall in tax revenue or aid), the development expenditure itself would be the most adjusting variable, which is consistent with Figure 5.6. The second cointegrating vector indicates that aid is positively associated with tax revenue and negatively related to the recurrent expenditure. This is consistent with a supposed donor behaviour whereby they would reward increasing tax collection efforts and be punitive towards recurrent expenditure excesses. Aid is the most adjusting variable, although the recurrent expenditures, too, adjust to equilibrium error.[223] The identified system is accepted as stationary with p-value=0.531.[224]

[222] It is clear that there is a strong 'trade-off' relationship between recurrent and development expenditure: if any one unit is spent on one item, it cannot be spent on another. This is a generic and uninteresting relationship, and it will not be directly identified.
[223] See Appendix Table D8 for and alternative identification.
[224] (Only) Tax could be accepted as weakly exogenous with p-value=0.532).

Table 5.15: Just-identified Model (Model with Disaggregated Spending)

	dexp	*rexp*	*tax*	*aid*	Shift 1986	Trend	
LR equilibrium relation (β_1)	**1.000**	0.000	**-1.105**	**-0.212**	**1.215**	-0.017	~I(0)
			(-4.703)	(-2.537)	(4.612)	(-1.241)	
LR equilibrium relation (β_2)	0.000	**1.000**	**-1.152**	**0.329**	0.282	**-0.034**	~I(0)
			(-7.177)	(5.757)	(1.568)	(-3.701)	
Adjustment coefficients (α_1)	**-0.587**	**0.186**	-0.016	**0.248**			
	(-4.339)	(3.741)	(-0.303)	(1.640)			
Adjustment coefficients (α_2)	**-0.528**	**-0.301**	0.147	**-1.647**			
	(-2.125)	(-3.301)	(1.532)	(-5.919)			
Multivariate normality	0.021						
Stationarity	p-value='-'						
Trace correlation	0.460						

Note: t-statistics are reported in the parentheses.

Table 5.16: Over-identified Model (Model with Disaggregated Spending)

	dexp	*rexp*	*tax*	*aid*	Shift 1986	Trend	
LR equilibrium relation (β_1)	**1.000**	0.000	**-1.384**	**-0.172**	**0.828**	0.000	~I(0)
			(-7.594)	(-1.965)	(8.711)		
LR equilibrium relation (β_2)	0.000	**1.000**	**-1.366**	**0.372**	0.000	**-0.023**	~I(0)
			(-10.029)	(5.695)		(-6.225)	
Adjustment coefficients (α_1)	-0.655	0.182	-0.011	0.191			
	(-4.830)	(3.608)	(-0.213)	(1.251)			
Adjustment coefficients (α_2)	-0.334	-0.298	0.106	-1.439			
	(-1.497)	(-3.593)	(1.197)	(-5.753)			
Multivariate normality	0.006						
Test of restricted model	0.531						
Trace correlation	0.456						

Note: t-statistics are reported in the parentheses.

3.2.5 MA Representation

As tax could be accepted as weakly exogenous variable, identification of one of the common driving trends is fairly straightforward: cumulated shocks to tax constitute a common trend (CT2), which positively (and permanently) loads into all of the system variables (Table 5.17). The other common trend is more complicated to identify, and seems to contain significant contributions from shocks to all remaining system variables.

The C matrix (Table 5.18) indicates that unanticipated shocks to aid may have permanent negative effects on recurrent spending, but no identifiable or permanent effect on development expenditure or tax revenue, consistent with aid funding specific projects rather than influencing the government investment levels in the long run. It would, nevertheless, have a positive and permanent effect on aid itself. An unanticipated shock to tax would have a strong, positive, and permanent effect on all variables. Similar shocks to development spending would have a small positive and lasting effect on recurrent expenditure, and – although perhaps transient only – effect on aid, with virtually no effect on other variables. Cumulated shocks to recurrent spending would be translated into permanently higher recurrent spending and lower aid, with only transient small effect on development expenditures and tax.

Table 5.17: Common Trends (Over-identified Model with Disaggregated Spending)

		dexp	*rexp*	*tax*	*aid*
Composition of common trends (α⊥)	CT1	**0.204**	**1.000**	0.000	**-0.254**
		(2.950)			(-3.465)
	CT2	0.004	0.000	**1.000**	0.073
		(0.050)			(0.946)
Loadings of common trends (β⊥)	CT1	0.125	0.766	0.237	-1.186
		(0.343)	(3.881)	(1.105)	(-2.422)
	CT2	1.722	0.502	0.970	2.213
		(4.014)	(2.163)	(3.838)	(3.844)

Note: t-statistics are reported in the parentheses.

Table 5.18: The Long Run Impulse Matrix C (Over-identified Model with Disaggregated Spending)

	$\hat{\varepsilon}_{dexp}$	$\hat{\varepsilon}_{rexp}$	$\hat{\varepsilon}_{tax}$	$\hat{\varepsilon}_{aid}$
dexp	0.032	0.125	**1.722**	0.094
	(0.245)	(0.343)	**(4.014)**	(0.922)
rexp	**0.158**	**0.766**	**0.502**	**-0.158**
	(2.253)	**(3.881)**	**(2.163)**	**(-2.881)**
tax	0.052	0.237	0.970	0.010
	(0.680)	(1.105)	(3.838)	(0.172)
aid	-0.234	**-1.186**	**2.213**	**0.463**
	(-1.343)	**(-2.422)**	**(3.844)**	**(3.396)**

Note: t-statistics are reported in the parentheses.

3.3 Model with Disaggregated aid (1968-2012) [225]

Disaggregating aid poses two data challenges. Firstly, loans are observed as negative for several years in the sample. Whilst it is not unusual to observe some years of net repayment, this complicates the analysis conducted in logs. To overcome this issue, a constant (80,000 local currency units) sufficiently large to render each observation positive is added to the

[225] Lutkepohl (2007:193): There are two possible interpretations of why some estimated VAR coefficients are not significantly different from zero. "First, some of the coefficients may actually be zero and this fact may be reflected in the estimation results. For instance, if some variable is not Granger-causal for the remaining variables, zero coefficients are encountered. Second, insignificant coefficient estimates are found if the information in the data is not rich enough to provide sufficiently precise estimates with confidence intervals that do not contain zero".

variable. Secondly, grants variable is recorded as zero for the first two years of the sample, again complicating analysis in logs. Therefore, the first two years are omitted from the sample. Furthermore, grants seem to be stationary around the mean (albeit with low p-value of 0.15). Thus unsurprisingly, nearly irrespective of model specification, it is estimated as the sole adjusting variable. No statistically acceptable or economically interpretable system was achieved, and thus model with both disaggregated expenditures and aid would not be estimated.

3.4 Model with Borrowing[226]

To model a system that includes borrowing (budget financing) without turning to a system in which variables are expressed as a proportion of GDP, the borrowing variable that contains some negative values (net repayment) is scaled by a constant (300,000 Tanzanian Shillings) to get rid of the negative values. The system further includes total government expenditures, total aid, and tax revenue.

3.4.1 Misspecification Tests

Lag length

The selected lag length is two (*k=2*). Again, the information criteria indicate preference for lag length of one (*k=1*); however, such choice is inferior in terms of residual autocorrelation.

Table 5.19: Lag Length Selection (Model with Borrowing)

Model	*k*	*T*	*Regr.*	*Log-lik*	*SC*	*H-Q*	*LM(1)*	*LM(k)*
VAR(5)	5	42	26	374.070	-8.558	-11.283	0.178	0.242
VAR(4)	4	42	22	357.882	-9.211	-11.517	0.501	0.116
VAR(3)	3	42	18	343.777	-9.963	-11.850	0.091	0.156
VAR(2)	2	42	14	335.931	-11.013	-12.481	0.120	0.003
VAR(1)	1	42	10	326.671	-11.996	-13.044	0.061	0.061

Effective Sample: 1971:01 to 2012:01.

Deterministics

Guided by large residuals, three permanent dummies are included in the model: aid reforms in 1995/6; and large residuals in borrowing in 2001 (HIPC debt relief) and 2008 (to cover for shortfalls in tax revenue partly due to the global financial crisis). A trend restricted to the

[226] @cats(lags=2,det=cidrift,break=level,dum) 1966:1 2012:1
L_TEXP L_TAX L_AID L_BORROWpos
1986:1
dum96p dum08p dum01p

cointegrating space and an unrestricted constant are included in the model, allowing for trends both in levels of variables and the cointegrating relationships, as above. Finally, in line with the previous, a mean shift in 1986 is included to capture the 1986 pro-market reforms.

Residuals

Model fit is acceptable, with trace correlation of 0.644 (Table 5.20). Multivariate normality cannot be rejected (p-value=0.088). The tests indicate no residual autocorrelation. Model residuals are depicted in Appendix Figure D4.

Table 5.20: Residuals from Unrestricted VAR

Residual normality (p-values)				
Multivariate	Univariate			
	texp	***tax***	***aid***	***borrow***
0.088	0.788	0.487	0.985	0.004[227]
Residual autocorrelation and ARCH effects (p-values)				
	LM(1)	LM(2)	LM(3)	LM(4)
Residual autocorrelation	0.280	0.054	0.536	0.446
ARCH	0.555	0.376	0.306	0.042
Trace correlation	0.644			

Note: Table reports p-values.

3.4.2 Cointegration rank

Johansen test indicates two cointegrating relationships (*r=2*) (Table 5.21). Such choice is supported by the additional information (Appendix Table D9).

Table 5.21: Trace Test

p-r	r	Eig. value	Trace	Trace*	Frac95	p-value	p-value*
4	0	0.852	151.091	131.172	71.436	0.000	0.000
3	1	0.551	65.218	55.090	49.065	0.001	0.011
2	2	0.426	29.198	24.922	29.821	0.061	0.179
1	3	0.090	4.221	3.712	15.862	0.857	0.899

* *denotes Bartlett corrections*

3.4.3 Long Run Identification: Hypothesis Testing

The long run exclusion test report that with the selected cointegration rank of two (r=2), none of the variables could be excluded from the cointegration space, except for a mild indication for the shift dummy. None of the variables are reported as mean- or trend-

[227] [Kurtosis of 5.765]

stationary, irrespective of whether the 1986 shift dummy is included in the test. For r=2, tax is reported as potentially weakly exogenous variable. No variables are found to be purely adjusting, although the borrowing variable could be borderline indicated as such. Results are provided in Table 5.22.

Table 5.22: Long Run Identification Tests

p-values	*r*	***texp***	***tax***	***aid***	***borrow***	Shift 1986	Trend
Long-run exclusion	*r=3*	0.000	0.000	0.000	0.000	0.006	0.006
	r=2	0.006	0.101	0.040	0.000	0.059	0.008
	r=1	0.002	0.411	0.187	0.000	0.019	0.002
Stationarity	*r=3*	0.000	0.000	0.000	0.000	no	no
	r=2	0.000	0.000	0.000	0.000	no	no
	r=1	0.000	0.000	0.000	0.000	no	no
Stationarity	*r=3*	0.000	0.000	0.001	0.000	yes	no
	r=2	0.000	0.000	0.001	0.000	yes	no
	r=1	0.000	0.000	0.000	0.000	yes	no
Stationarity	*r=3*	0.000	0.000	0.016	0.000	no	yes
	r=2	0.000	0.000	0.013	0.000	no	yes
	r=1	0.000	0.000	0.000	0.000	no	yes
Stationarity	*r=3*	0.000	0.000	0.004	0.000	yes	yes
	r=2	0.000	0.000	0.005	0.000	yes	yes
	r=1	0.000	0.000	0.000	0.000	yes	yes
Weak exogeneity	*r=3*	0.000	0.016	0.000	0.000	-	-
	r=2	0.005	0.670	0.003	0.000	-	-
	r=1	0.038	0.495	0.510	0.000	-	-
Purely adjusting	*r=3*	0.002	0.031	0.049	0.032		
	r=2	0.000	0.000	0.019	0.049		
	r=1	0.000	0.000	0.000	0.102		

Note: Table reports p-values.

3.4.4 Long Run Identification: Results (model with borrowing)

Life would be easy if cointegration tests indicated a sole cointegrating vector. In such case, the vector would represent an intuitive positive relationship between total government expenditure and all the revenue components, with borrowing being the most adjusting variable, and adjusting (albeit in overshooting way) spending (Table 5.23). Alas.

Table 5.23: Long Run Identification Tests (Example if r=1)

	texp	*tax*	*aid*	*borrow*	Shift 1986	Trend	
LR equilibrium relation (β_1)	**1.000**	**-0.237**	**-0.114**	**-0.718**	**0.546**	**-0.032**	~I(0)
		(-1.720)	(-2.428)	(-14.352)	(3.523)	(-4.001)	
Adjustment coefficients (α_1)	**0.106**	0.033	0.139	**1.173**			
	(2.292)	(0.714)	(0.744)	(12.389)			
Multivariate normality	0.001						
Trace correlation	0.446						
	Log-Likelihood = 328.928						

Note: t-statistics are reported in the parentheses.

With two cointegrating vectors, we posit a simple identification strategy by asking two simple questions. Firstly, what sort of equilibrium is formed among the variables over which the government has *direct* control (i.e. government spending, tax revenue, and budget financing). Secondly, what are the dynamics among *all* the revenue variables (except non-tax revenue) available to the government. The results for the just-identified system are provided in Table 5.24. With insignificant trend further excluded from the second cointegrating vector, the over-identified system is achieved (Table 5.25).

In the first cointegrating vector, the expenditure would positively depend on tax revenue and borrowing. This can be thought of as an 'extended' domestic fiscal equilibrium, as all the variables are under the direct control of the government. Following the theoretical (economic) postulated hypothesis, the borrowing is the most (and effectively only) adjusting variable: should the deficit (surplus) occur, the non-concessional borrowing would increase (decrease) quickly (in less than a year) to restore the budget.

The second cointegrating relationship indicates that tax is positively related to both aid (potential income effect) and borrowing (repayment/servicing requirements). Furthermore, aid and borrowing can be regarded as substitutes. Should a departure from this revenue equilibrium occur (for instance, a shortfall in the tax revenue combined with overtly constrained/prohibitively expensive further public borrowing), aid (increase) and spending (decrease) would adjust to equilibrium error. The identified cointegrating vectors are depicted in Appendix Figure D5.[228]

[228] A slightly differently identified model is summarised in Appendix Table D10.

Table 5.24: Just-identified Model (Model with Borrowing)

	texp	*tax*	*aid*	*borrow*	Shift 1986	Trend	
LR equilibrium relation (β_1)	**1.000**	**-0.506** (-4.422)	0.000	**-0.598** (-12.739)	**0.421** (2.896)	**-0.029** (-3.822)	~I(0)
LR equilibrium relation (β_2)	0.000	**1.000**	**-0.423** (-6.694)	**-0.448** (-5.911)	**0.465** (2.736)	-0.014 (-1.582)	~I(0)
Adjustment coefficients (α_1)	-0.011 (-0.200)	0.061 (1.053)	**-0.602** (-3.677)	**1.232** (10.186)			
Adjustment coefficients (α_2)	**0.215** (3.982)	-0.037 (-0.624)	**1.225** (7.283)	**0.222** (1.787)			
Multivariate normality	0.138						
Trace correlation	0.534						

Note: t-statistics are reported in the parentheses.

Table 5.25: Over-identified Model (Model with Borrowing)

	texp	*tax*	*aid*	*borrow*	Shift 1986	Trend	
LR equilibrium relation (β_1)	**1.000**	**-0.441** (-3.701)	0.000	**-0.654** (-13.078)	**0.454** (2.996)	**-0.031** (-4.017)	~I(0)
LR equilibrium relation (β_2)	0.000	**1.000**	**-0.616** (-11.901)	**-0.280** (-2.857)	**0.416** (3.526)	0.000	~I(0)
Adjustment coefficients (α_1)	0.067 (1.493)	0.039 (0.795)	-0.146 (-1.059)	**1.225** (12.297)			
Adjustment coefficients (α_2)	**0.152** (3.334)	-0.010 (-0.202)	**0.986** (7.062)	0.107 (1.064)			
Multivariate normality	0.099						
Test of restricted model	0.404						
Trace correlation	0.529						

Note: t-statistics are reported in the parentheses.

3.4.5 MA Representation

The moving-average representation of the VAR allows one to inspect the driving forces in the model. The *p-r* common trends (CT) describe the non-stationarity in the process, originating from the cumulative sum of the unanticipated shocks. One rather clear common trend (CT1) in this model is composed from cumulated unanticipated shocks to tax variable, which positively loads to expenditures, tax and aid (Table 5.26). The second one (CT2) seems to predominantly arise from the cumulated shocks to expenditures, with small but nevertheless significant contributions from shocks to both aid and borrowing. The second pushing force loads positively and significantly into expenditure and borrowing variables.

Table 5.26: Common Trends (Over-identified Model with Borrowing)

		texp	*tax*	*aid*	*borrow*
Composition of common trends (α⊥)	CT1	0.000	1.000	0.013	-0.030
				(0.258)	(-0.745)
	CT2	1.000	0.000	-0.146	-0.072
				(-4.234)	(-2.691)
Loadings of common trends (β⊥)	CT1	0.649	1.121	1.711	0.237
		(2.291)	(4.297)	(4.761)	(0.776)
	CT2	1.549	0.482	-0.144	2.042
		(4.100)	(1.387)	(-0.301)	(5.014)

Note: t-statistics are reported in the parentheses.

The C matrix (Table 5.27) illustrates how a shock[229] to each variable (each column) ripples through the system: a statistically significant coefficient would indicate that an unanticipated shock to the variable has a permanent effect on another variable; otherwise, the effect is transitory at most. An unanticipated shock to tax would positively and permanently affect expenditures, tax, and aid (latter perhaps indicating donors' reward policies or sustained tax revenue reform effort), and has no permanent effect on borrowing (if anything, it may temporarily increase domestic borrowing – but not reducing it). Shocks to borrowing permanently (although not much) reduce total government expenditure, and borrowing itself, and have very small, and - if at all - transitory negative effects on tax and aid. Shocks to aid permanently reduce borrowing, but also spending, without permanent or sizeable effects on tax or aid itself. Cumulated unanticipated shocks to expenditure seem to translate into permanent and large increases in government spending itself, as well as borrowing, with some positive transitory effect of tax and no permanent effect on aid (if anything, the latter may be temporarily reduced, again underlining donors' punitive behaviour towards recurrent spending excesses). It must be noted that these results need to be taken carefully, as, as in the other models discussed above, there is some residual correlation between the variables.

[229] One should again be wary of labelling them as empirical shocks given the highly correlated residuals.

Table 5.27: The Long Run impulse Matrix C (Over-identified Model with Borrowing)

	$\hat{\varepsilon}_{texp}$	$\hat{\varepsilon}_{tax}$	$\hat{\varepsilon}_{aid}$	$\hat{\varepsilon}_{borrow}$
texp	**1.549**	**0.649**	**-0.218**	**-0.132**
	(4.100)	**(2.291)**	**(-2.540)**	**(-2.038)**
tax	0.482	**1.121**	-0.056	-0.068
	(1.387)	**(4.297)**	(-0.705)	(-1.151)
aid	-0.144	**1.711**	0.044	-0.041
	(-0.301)	**(4.761)**	(0.403)	(-0.496)
borrow	**2.042**	0.237	**-0.296**	**-0.155**
	(5.014)	(0.776)	**(-3.195)**	**(-2.227)**

Note: t-statistics are reported in the parentheses.

4. Conclusion

Compared to the Ethiopian case study, the CVAR results on Tanzanian fiscal effects of aid are delicate, reflecting potential inferiority in terms of data quality, justifying focus on long run estimates only. Nevertheless, some reliable findings emerge.

The most statistically sound results seem to be from the most aggregated specification modelling total expenditure, tax, and aid. Although no variables are clearly found to be weakly exogenous, tax appears to be the 'most' exogenous variable (i.e. governments have limited ability to alter tax in short to medium term), and aid is found to be mostly adjusting. Aid does appear to be positively associated with tax and spending in the long run (although shocks to aid may have a mild transient negative effect – but the results are weak). Although on-budget aid, rather plausibly, does not drive the domestic revenue in either Ethiopia or Tanzania, it does not discourage or substitute for the domestically collected revenue. Unsurprisingly, tax is positively associated to expenditures in both countries.

As in Ethiopia, in Tanzania aid has a positive association with spending, and aid seems to be adjusting to funding the excessive deficits in both countries. However, the fiscal mechanism exhibits differences if the spending is disaggregated into development and recurrent components. In Ethiopia, aid (and especially grants) adjusts to capital expenditures (the two are positively related in the long run), positively indicating donors rewarding sound public investment decisions. In Tanzania it is mainly the development expenditures that adjust to (shortfalls or windfalls in) aid. This is especially pronounced during the period of re-assessing the aid disbursements in mid-1990s, where development expenditures dwindled following a sharp decrease in aid. In contrast to a positive relationship in Ethiopia, aid is negatively associated to recurrent spending in Tanzania. Although both variables adjust to departures from this long run equilibrium, the faster and stronger adjustment of aid indicates some potential punitive donor disbursement behaviour with respect to consumption spending excesses.

Finally, we find evidence that aid and borrowing could be considered substitutes. However, as it is found to be less (and less quickly) adjusting to equilibrium error, it is possible that public borrowing is not borrowing of last resort, rather signalling (potentially non-DAC) donors funding the at least a fraction of the deficit.

Chapter 6

Donor vs. Recipient Aid Records:

Different Tales

1. Introduction

The opening chapters of this thesis argued for the importance of distinguishing between on-budget and off-budget aid in analysing the fiscal effects of aid. Two country case studies in Chapters 4 and 5 demonstrated that using recipient's budget aid records the estimated fiscal effects of (and on) aid report much more sanguine results than is often estimated (or postulated) in the literature using the broader measures of aid. We argued that the omission of off-budget aid flows are less of a concern in the reduced form cointegrated VAR than in the conventional panel estimations.

Even with the recipient's measure of (on-budget) aid available we cannot disaggregate DAC aid flows into on-and off-budget components because recipient's data include non-DAC flows (comprehensive and accurate non-DAC data are not available). We can, nevertheless, illustrate the differences in total recipient and conventional (DAC) donor total aid flows (used in the majority of studies), and demonstrate the effect these differences have on the estimated fiscal effects. Chapter 2 illustrated that even data from respectable international databases (such as IMF and WHO) can provide substantially different versions of the same (even in terms of definition) variable, whose underlying data source could in principal be traced to the same developing country government. It would not be irrational to expect a certain degree of discrepancy arising from records originating from different sources.

This chapter is structured as follows. Section 2 compares the aid data (grants and loans) recorded by two East African recipients (Ethiopia and Tanzania) to the OECD DAC aid disbursements, and demonstrates that the direction of the discrepancies can vary. In section 3 simple cointegrated vector autoregressive models are estimated to expose the differences in the estimated fiscal effects of (and on) aid arising from the alternative sources of aid data. Section 4 concludes. Additional information is provided in Appendix E.

2. Data

Two countries' datasets are used in this chapter: Tanzania (as in Chapter 5) and Ethiopia (Chapter 4). These datasets contain the fiscal variables, such as central government expenditures and domestic revenues, and a recipient's measure of budget aid, disaggregated into grants and loans. The Tanzanian Central Bank's data for the period 1966-2012 are recorded in domestic currency (Tanzanian Shillings), and are deflated using the CPI measure (base year 2005, see Chapter 5 for more details). The Ethiopian data are available for the period 1963-2009 from Ethiopia's Ministry of Finance and Economic Development (MoFED), in domestic currency (millions of Ethiopian Birr), and are deflated using the GDP deflator (base year 1998, see Chapter 4).

OECD Development Assistance Committee (DAC) Official Development Assistance (ODA) disbursement data, used in a large fraction of studies estimating various effects of aid, are readily available from the OECD DAC Table 2a[230] for the whole period of interest. The DAC data are recorded in current US dollars[231]. To convert these data to domestic currencies (Tanzanian Shilling and Ethiopian Birr), IMF IFS' Official exchange rate (period (yearly) average) is used. This measure is available for the whole period of interest from a single source. In principle, the official – not an alternative measure of exchange rate – should be used to convert the official flows of money (this clearly ignores any secondary effects of available foreign exchange flowing through the budget, such as increased forex reserves, etc.) and for the purpose of this exercise it will be held that such conversion would deem the

[230] http://stats.oecd.org/Index.aspx?DataSetCode=Table2A

[231] A measure in constant (2012=100) USD is also available. The choice to use current values is driven by motivation to isolate data differences by using the same deflator on all series.

two (recipient's and DAC) datasets comparable.[232,233] The DAC measures of aid for Tanzania (Ethiopia) are deflated by the CPI (GDP) deflator to isolate data differences by using the same deflator on all country series.

It must be noted in the DAC measure of aid for Ethiopia, the values for aid loans are coded as missing for three years (1996, 1997, 1999). The only viable solution is to treat them as 0 (this is realistic, as a lot of aid was suspended around 1998 due to Ethiopia-Eritrea war, and it is loans that would practically be withheld first (whilst humanitarian aid (a grant component) would be expected to be ceased last). This, however, does not pose severe complications, as only the measure of total aid is used in estimations (and for the purposes of summary statistics these observations are treated as missing).[234]

The donor and recipient accounting of aid differs. Recipient's aid measure by definition only includes the on-budget aid, i.e. the aid (cash) receipts flowing through the (central) government (usually the ministry of finance). The donor measure of aid[235] would further include off-budget aid (transfers to non-governmental organisations, payments to donor agencies, research bodies, aid in-kind, technical cooperation component[236], etc.), but would exclude funds from non-traditional donors. Differences in the recorded aid would be

[232] "Ethiopian birr was pegged to the USD from its inception in 1945 until early 1990s. The Birr was valued at 2.50 per USD before the collapse of the Bretton Woods system in 1971, which forced an initial revaluation to 2.30, then in 1973 to 2.07 per USD". It was overvalued under Derg, and several devaluations were conducted when EPRDF came to power. "The current exchange rate system is classified as a (de facto) crawling peg to the USD, i.e. a managed (or dirty) float". (Martins, 2010b:25).
[233] "The gradual change in policy orientation from "controls" to "market" in Tanzania is associated with a change from a highly controlled exchange rate (until 1985) to a more liberalized regime from 1986 to the present (2002). The parallel exchange rate dominated price changes from the late 1970s to 1985; the parallel premium tapered off gradually from 1986, almost disappearing by 1992. The problem of inflation cuts across both regimes despite improvements in the past four to five years" (Rutasitara, 2002: Abstract).
[234] For the CVAR analysis, the variables are logged. For the graphs, the deflated levels are depicted.
[235] "Official Development Assistance (ODA) is defined as those flows to developing countries and multilateral institutions provided by official agencies, including state and local governments, or by their executive agencies, each transaction of which meets the following tests: i) it is administered with the promotion of the economic development and welfare of developing countries as its main objective; and ii) it is concessional in character and conveys a grant element of at least 25 per cent".(OECD, http://www.oecd.org/site/dacsmpd11/glossary.htm)
[236] "Technical Co-operation : This is defined as activities whose primary purpose is to augment the level of knowledge, skills, technical know-how or productive aptitudes of the population of developing countries, i.e., increasing their stock of human intellectual capital, or their capacity for more effective use of their existing factor endowment. Accordingly, the figures relate mainly to activities involving the supply of human resources (teachers, volunteers, experts in various sectors) and action targeted on human resources (education, training, advice). The supply of expertise designed primarily to support the implementation of capital projects ("Investment-Related Technical Co-operation'' - IRTC) is not included under this heading." (OECD, http://www.oecd.org/site/dacsmpd11/glossary.htm)

expected, but are often overlooked by researchers using the DAC data to estimate fiscal or growth effects of aid. Large discrepancies have indeed been confirmed for Uganda (up to 10% of GDP, Fagernas and Roberts, 2004a), Zambia (up to 20-40 % GDP, Fagernas and Roberts, 2004b), Senegal (DAC figures twice as high as aid reported by the Ministry of Finance, Ouattara, 2006).

Tables 6.1 and 6.3 record the ratio (period average, in percentages) of DAC aid (grants, loans, and the total of the two) observations to recipient aid data for Tanzania and Ethiopia, respectively. The statistics are further split into periods before and after pro-market reforms. Tables 6.2 and 6.4 report the correlation coefficients between donor (DAC) and recipient data for each country. Figures 6.1 and 6.2 depict the data (deflated levels) differences visually.

Tanzania

The OECD DAC records of aid grants for Tanzania consistently (for all years) exceed the recipient's data. While this follows intuition, the magnitude of this difference is alarming: the DAC grant measure is almost eight times the Tanzanian government's aid records on average. The correlation coefficient between these two measures is, nevertheless, large, 0.80.[237] Two interpretations of this discrepancy are available: what DAC records as grants may be treated by the recipient as loans due to misperception or misinformation associated with these flows. Alternatively, the large differential between recipient and donor grant flows could be explained by a large proportion of DAC grants delivered through donor projects or as technical cooperation, therefore constituting off-budget aid (see Chapter 3 for more detailed discussion).

Loans (net measure in both sources), on the other hand, provide opposite result: for the majority of years, the recipient's value exceeds that of DAC. For the years that both values are positive,[238] DAC loans average about 75% of the value recorded by the recipient. The correlation between the two series is virtually zero (0.034). This may reflect borrowing from non-DAC donors (in the past USSR, and more recently China and Gulf countries would be good examples, and although the reasons for *concessional* lending would be less clear, it could be expected to follow similar strategic motives of DAC donors). Whilst some non-DAC donors (Arab countries and EU members) report their aid flows to OECD-DAC, others (such

[237] Although it is acknowledged that correlation coefficient between I(1) variables is spurious, here the comparison is for the 'same' variable and is therefore seen as appropriate.
[238] The net value of DAC records is negative for 11 years (out of 47); only one value below zero is recorded in recipient's records, possibly reflecting borrowing elsewhere to repay DAC debt.

as BRICs) do not follow the DAC reporting standards, and accounting for their aid flows (and the motivation) is more complicated. Reviewing a body of literature, Walz and Ramachandran (2011) estimate that the aid flows from non-DAC countries range from \$11 billion to \$41.7 billion in 2009 (between 8 and 31% of global gross ODA).[239] We limit our analysis to the use of DAC donor data, because: i) this is what has been traditionally used in the aid literature; ii) non-DAC flows are not recorded on a consistent basis; iii) if included, non-DAC data would only strengthen our points on the size of discrepancies.

Table 6.1: Donor-to-Recipient Aid Measures Ratio (Tanzania)

Ratio (%), Tz	1966-2012	Pre-1986	Post-1986
Grants (DAC)-to-Grants (recipient)	**783 %**	1314 %	429 %
Loans (DAC)-to-Loans (recipient)	**75 %**	82 %	66 %
Total aid (DAC)-to-Total aid (recipient)	**320 %**	207 %	396 %

Statistics reported for Loans exclude negative values. Including the negative values, the DAC loans would constitute about a quarter (28%) of the recipient's values on average during the full sample period.

Table 6.2: Correlation Coefficients between Donor and Recipient Measures (Tanzania)

Correlation coefficient, Tz	Grants (DAC)	Loans (DAC)	Total aid (DAC)
Grants (recipient)	0.785		
Loans (recipient)		0.034	
Total aid (recipient)			0.778

Whilst differences in grants measures for Tanzania decrease post 1986 reforms, the discrepancies in the loan records increase, together with the increasing volume of loans (as recorded by the recipient). This supports the hypothesis that this is driven by non-traditional donors, but also poses a possibility that recipient's loan measure may include commercial foreign borrowing that is not concessional in nature.

Finally, the total aid, which is simply the sum of grants and (net) loans in both data sources, reflect astounding discrepancies in aid records. The donor measure of total disbursed aid

[239] DAC 2a Tables now include an attempted measure of total aid from 'all donors' rather than just DAC members (the latter's ODA contributions amounted to \$133.2 billion in 2009). However, the uncertainty associated with the estimation of non-DAC members' numbers, reported by Walz and Ramachandran, 2011, and especially considering the retrospective revisions, is not considered to be a 'robust' measure (for instance, China's recent aid estimates "range anywhere from \$1.5 to \$25 billion", p.1), as they may include FDI, military assistance, and other components that do not fall under the DAC's definition of ODA.

exceeds recipient's records by 3.2 times.[240] In studies attempting to evaluate aid's effect on growth, or the extent to which aid is spent in fungibility studies, using the OECD aid data would underestimate the beneficial effects of aid.

Ethiopia

For Ethiopia, the picture is rather different. Until about mid-to-late 1980s, DAC data for grants are *lower* than what is recorded by the Ethiopian Ministry of Finance. Though puzzling at first, the finding is explicable. As discussed in Chapter 4, during the Derg military junta regime (1974-1991), Ethiopia's major donor was USSR.[241] As the Russian Federation is not a DAC member even at present, these figures would not be included in the retrospective DAC tables. Since the late 1980s-early 1990s, that is, after the beginning of pro-market reforms and the fall of USSR, DAC grant measure are nearly twice *larger* than recipient's records, reflecting increasing Western donor presence. Overall, during the full sample period, DAC records exceed the recipient's grant measure by 36% on average, and the correlation coefficient is 0.78.

Table 6.3: Donor-to-Recipient Aid Measures Ratio (Ethiopia)

Ratio (%), Eth	1963-2009	Pre-1991	Post-1991
Grants (DAC)-to-Grants (recipient)	**136 %**	98 %	191 %
Loans (DAC)-to-Loans (recipient)	**24 %**	31 %	13 %
Total aid (DAC)-to-Total aid (recipient)	**79 %**	63 %	102 %

Note: Statistics reported for Loans exclude negative values. Including the negative values, the DAC loans would constitute about a quarter (28%) of the recipient's values on average during the full sample period.

Table 6.4: Correlation Coefficients between Donor and Recipient Measures (Ethiopia)

Correlation coefficient, Eth	Grants (DAC)	Loans (DAC)	Total aid (DAC)
Grants (recipient)	0.894		
Loans (recipient)		0.016	
Total aid (recipient)			0.866

[240] If recipient's loan measure indeed includes non-concessional borrowing abroad, the actual difference is even higher.

[241] While Tanzania was also socialist-oriented, its major inflows were from leftist Western donors, such as Scandinavian countries.

As in Tanzanian case, the donors' measure of (gross)[242] loans is lower than the values recorded by Ethiopian Ministry of Finance. However, in Ethiopia's case discrepancy is even higher: DAC loans amount to only about a quarter (24%) of recipient's records during the full sample average. Rather than decreasing, the discrepancy again increases in the latter years: since 1991, DAC loans amounted only to 13% of recipient's records, on average. This again signals the possibility of aid loans from non-traditional donors, as well as potential accounting of non-concessional loans under this heading.

Overall, the total aid (sum of grants and gross loans) measure in DAC ODA disbursement measure is *lower* than Ethiopian records (although the correlation coefficient is a respectable 0.866). This finding provides interesting evidence contradicting the expectation that DAC measures, which include both on- and off-budget aid, would generally exceed recipient's own on-budget aid records: the discrepancy is likely, but its direction is not certain. And whilst the two measures converge in the latter part of the Ethiopian sample (see **Figure 6.2**), this is shown to be by sheer coincidence. Overall, the data suggests that non-DAC aid is very important for Ethiopia, and not solely during the Derg.

Aid data is often scaled by Gross Domestic Product (GDP) or Gross National Income (GNI), especially in the cross-country studies. The final exercise in this section briefly compares three measures of aid: recipient's aid-to-GDP[243]; DAC-2a Total net aid-to-GNI; and WDI ODA-to-GNI[244]. The subsamples for which all measures are simultaneously available span 1988-2009 for Tanzania, and 1981-2009 for Ethiopia.

Although slightly dampened by the differences in GDP estimates, the differences between the DAC donor and recipient's data follow the discussion above. The key finding here is that WDI aid data substantially exceeds even DAC donor records. This is likely to be primarily due to the inclusion of non-DAC donor flows (although not always falling under the ODA

---

[242] Only gross loans are available in Ethiopian recipient data. For consistency, DAC gross loans data are used for Ethiopia (whilst for Tanzania loans referred to net measures in both samples).

[243] For Tanzania, GDP data is only available post-1987 GDP rebasing (see Chapter 5 for discussion). Indeed, the DAC 2a tables only provide the measure of total aid to GNI from 1988 onwards. To express the recipient's aid as a percentage of GDP IFS National Accounts data were used (note that IFS data records for GDP and GNI are highly similar; however, the data alterations to GDP measure in the DAC data are unknown). For Ethiopia, GDP data is available from government's own records (see Chapter 4).

[244] "Net official development assistance (ODA) consists of disbursements of loans made on concessional terms (net of repayments of principal) and grants by official agencies of the members of the Development Assistance Committee (DAC), by multilateral institutions, and by non-DAC countries to promote economic development and welfare in countries and territories in the DAC list of ODA recipients. It includes loand with a grant element of 25 per cent (calculated at rate of discount of 1-per cent)". World Bank, http:/data.worldbank.org/indicator/DT.ODA.ODAT.GN.ZS .

definition, as the latter may include FDI, military assistance, and other components that do not fall under the DAC's definition of ODA). The recipient measures, in theory, should include grants and loans from both DAC and non-DAC measures. The disparity between recipient's and WDI data is thus even further extending the difference between recipient's and donors' aid data records.[245]

Table 6.5: Aid-to-GDP(GNI) Ratios Across Three Sources of Data

	Aid/GDP (GNI)		
	Recipient's measure	**DAC-2a Disbursement Data**	**WDI**
Tanzania (1988-2009)			
Period Average	5.47 %	10.55 %	16.48 %
Ethiopia (1981-2009)			
Period Average	5.55 %	5.12 %	10.28 %

Note: recipient's measure refers to aid-to-GDP; DAC-2a Disbursement Data refers to DAC-2a Total net aid-to-GNI; and WDI refers to WDI ODA-to-GNI.

[245] If non-DAC flows were indeed considered aid, using only the DAC-donor data in aid estimations would overestimate the actual *cash* flows from DAC donors (underestimating positive effect of aid), but underestimating the extent of total (DAC and non-DAC donor) aid (overestimating the positive effect of aid).

Figure 6.1: Recipient-Donor Data Comparisons (Tanzania)

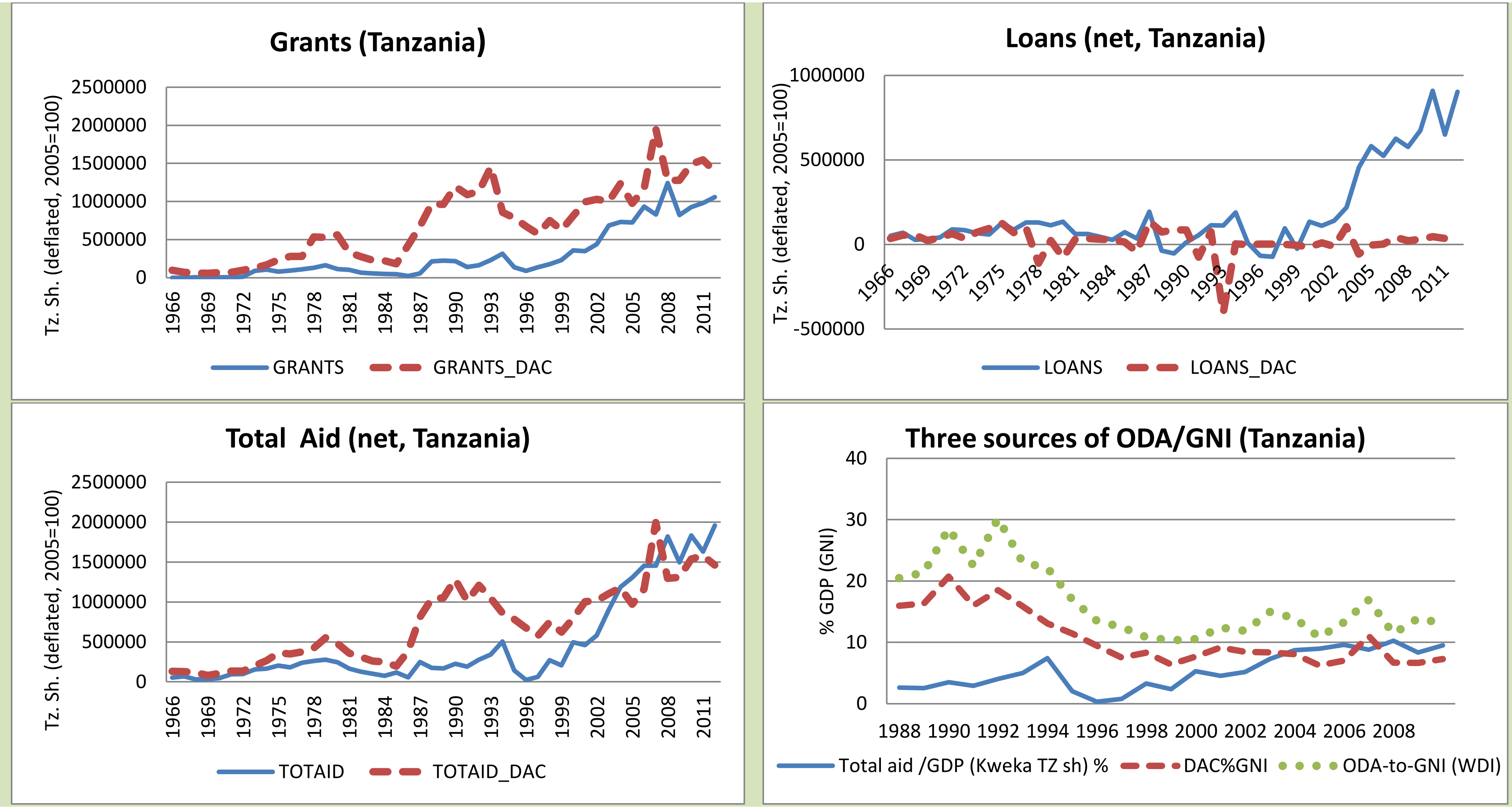
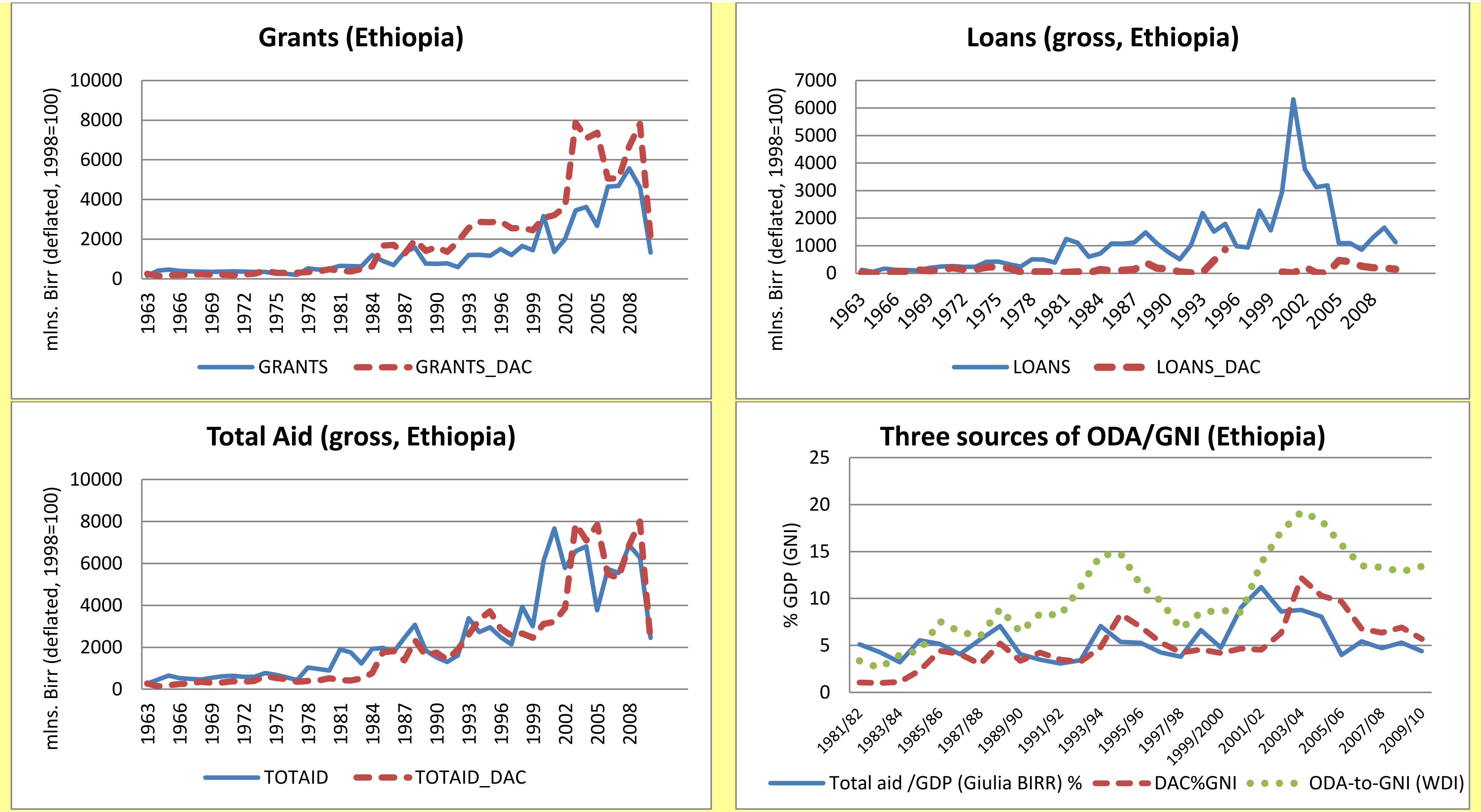


Figure 6.2: Recipient-Donor Data Comparisons (Ethiopia)



3. Empirical Results

We model a simple CVAR (VECM) (see Chapter 4 for methodology description) with three variables: (central) government total spending (***texp***), domestic revenue (sum of tax and non-tax revenues, ***domrev***), and total aid (grants plus loans), transformed using natural logarithm. To isolate the discrepancies between the aid measures across sources, the data for the domestic fiscal variables (***texp*** and ***domrev***) are the same (i.e. recipient's measure). For each country, two variants of the CVAR model are estimated: one with recipient's measure of aid (***aid***), and the second with DAC donors' ODA disbursement data (***aid_DAC***). To best isolate the effect of data differences, for each country the simplest statistically plausible specification across the two variants is estimated. Only long-run coefficients are reported.

For Tanzania, we model a CVAR with a lag length of one (*k=1*), an unrestricted constant to allow for a non-zero mean in the cointegrating relations and (non-quadratic) trends in levels[246] (model specification test results are available in the Appendix Table E1). Although the testing revealed a large (over 3) residual in the model with the recipient's measure of aid, the choice has been made not to include dummies into estimations to maintain the modelling choices consistent across the measures of aid. The model fit is not great in terms of low trace correlation and the rejection of multivariate normality, but crucially, there is no residual autocorrelation. Johansen (trace) test suggests cointegration rank of one (*r=1*), which would imply one long-run equilibrium relationship between fiscal aggregates and aid.

For Ethiopia, a CVAR with a lag length of two (*k=2*) is modelled, as a lower lag length would imply some residual autocorrelation of order one and two (model specification test results for Ethiopia are available in the Appendix Table E2). An unrestricted constant is included, but the model revealed no large residuals (over 3), thus no dummies were included. The model fit is acceptable (trace correlation of 0.334 (0.25) for recipient (donor) measure of aid), although the null hypothesis of multivariate normality is rejected with p-value of 0.025. Johansen (trace) test here too suggests cointegration rank of one (*r=1*), which would imply one long-run equilibrium relationship between fiscal aggregates and aid.[247]

The results are reported in Tables 6.6 and 6.7 (Tanzania) and Tables 6.8 and 6.9 for Ethiopia.

[246] A model with a restricted trend was tested, and it was shown that it can be excluded.
[247] The political changes and/or pro-market reforms are not modelled here to keep the models as simple as possible.

Table 6.6: CVAR Estimates: Recipient's Aid Data (Tanzania)

Recipient's data (TZ)	***texp***	***domrev***	***aid_DAC***	
LR equilibrium relation ($\beta_{5.1}$)	**1**	**-0.92**	**-0.28**	~I(0)
		(-4.43)	**(-3.07)**	
Adjustment coefficients ($\alpha_{5.1}$)	**-0.21**	-0.08	0.10	
	(-2.88)	(-1.27)	(0.32)	
Multivariate normality (p-value)	0.000			
Stationarity (p-value)	No over-identifying restrictions			
Trace correlation	0.104			
Log-likelihood	242.771			

Note: t-values are reported in parentheses.

Table 6.7: CVAR Estimates: DAC Donors' Aid Data (Tanzania)

DAC data (TZ)	***texp***	***domrev***	***aid_DAC***	
LR equilibrium relation ($\beta_{5.2}$)	**1**	**-1.42**	**0.113**	~I(0)
		(-13.19)	**(1.76)**	
Adjustment coefficients ($\alpha_{5.2}$)	**-0.37**	0.03	-0.17	
	(-4.21)	(0.36)	(-0.89)	
Multivariate normality (p-value)	0.065			
Stationarity (p-value)	No over-identifying restrictions			
Trace correlation	0.106			
Log-likelihood	268.382			

Note: t-values are reported in parentheses.

Table 6.8: CVAR estimates: Recipient's Aid Data (Ethiopia)

Recipient's data (ETH)	***texp***	***domrev***	***aid_DAC***	
LR equilibrium relation ($\beta_{5.3}$)	**1**	**-1.12**	0.04	~I(0)
		(-12.85)	(0.64)	
Adjustment coefficients ($\alpha_{5.3}$)	**-0.48**	0.12	-0.53	
	(-3.23)	(0.63)	(-1.09)	
Multivariate normality (p-value)	0.024			
Stationarity (p-value)	No over-identifying restrictions			
Trace correlation	0.275			
Log-likelihood	278.926			

Note: t-values are reported in parentheses.

Table 6.9: CVAR estimates: DAC Donors' Aid Data (Ethiopia)

DAC data (ETH)	***texp***	***domrev***	***aid_DAC***	
LR equilibrium relation ($\beta_{5.4}$)	**1**	**-0.99**	-0.05	~I(0)
		(-16.18)	(-1.42)	
Adjustment coefficients ($\alpha_{5.4}$)	**-0.52**	0.09	**0.96**	
	(-2.64)	(0.40)	**(2.19)**	
Multivariate normality (p-value)	0.007			
Stationarity (p-value)	No over-identifying restrictions			
Trace correlation	0.190			
Log-likelihood	280.33			

Note: t-values are reported in parentheses.

For Tanzania, the simple CVAR long run coefficients from models with alternative aid measures reveal contrasting results. The variant with recipient's measure of aid yield economically plausible results: total government expenditure in the long run is positively (and significantly) related to domestic revenue *and* aid ($\beta_{5.1}$), with domestic expenditures adjusting to equilibrium error ($\alpha_{5.1}$). Using the DAC donor aid data, the relationship is (significantly) different and more complicated to interpret: total government expenditure is still positively (and significantly) related to domestic revenue, but *negatively* related to aid ($\beta_{5.2}$). That is, if the DAC data rather than Tanzanian aid data are used, the sign of the estimated effect of aid changes, consistent with DAC overstating aid amount that can finance spending.

Ethiopian comparisons also reveal significant differences. In both variants of the model (with DAC donor data or recipient measure of aid), aid is estimated not to be significantly related to domestic fiscal aggregates in the long run ($\beta_{5.3}$, $\beta_{5.4}$), with government total spending positively associated to domestically collected revenue. The two data sources, however, suggest different adjustment mechanisms: recipient's data suggests that only spending would be adjusting to shortfalls or excesses of revenues; the DAC data suggests that in events where revenues fall short (or expenses exceed the equilibrium levels), donors step in with extra aid ($\alpha_{5.3}$, $\alpha_{5.4}$).

4. Conclusion

The recipient and donor data differ, and not necessarily in a direction predictable from the outset. Two key determinants of these differences were discussed. Firstly, recipient's aid records by definition solely account for on-budget aid; meanwhile, the DAC donors' total flows include both on-budget and off-budget components. Secondly, DAC donors' data only include flows from DAC member donors; meanwhile, the recipient's (on-budget) records will include financial flows from 'non-traditional' (or non-DAC) donors, which were increasingly present throughout the sample period (rendering the 'non-traditional' label somewhat faulty).[248]

[248] The latter flows do not always fall under the current definitions of ODA; there are also little (or no) incentive for some non-DAC donors (or, rather, co-operators) to report the destination, purpose, or magnitude of these 'aidic' flows, therefore although DAC now 'reports' aid data (also retrospectively) from 'all donors', the uncertainty associated with these records is considerable (the data are close to speculation) (see Walz and Ramachandran, 2011). Therefore a solution to the second factor may take a while to be realised.

Depending on which source is relied upon for aid data, the results can differ substantially. The comparison of the CVAR estimates of models with recipient versus DAC donor aid data revealed that the two aid measures do not even covariate sufficiently to yield qualitatively consistent estimates. The estimated effects of aid can contrast in terms of sign (as in the Tanzanian case) or reflect different adjustment behaviour (the Ethiopian case).

Chapter 7

Conclusion

There are many forms in which aid can be delivered. It could be given in monetary form, or as delivery of goods or services directly paid for by the donor. Aid can be disbursed in grants that require no repayment, or in subsidised loans. One can distinguish between aid delivered through the recipient's government, channelled through non-governmental organisations, or spent in the donor country. Aid could be earmarked for a specific heading or sector of donor's choice, or given as general budget support to be allocated at recipient's discretion. Conceptually and theoretically, we can postulate different fiscal (and, in turn, growth) effects of different modalities or components aid. Empirical evaluation, however, is in most cases infringed by the inaccessibility (or non-existence) of data necessary to fully and accurately disaggregate aid into its distinct components. Two running themes in this thesis therefore focus on data availability, quality, and consistency across sources, and to what extent it enables aid data disaggregation into its on-budget and off-budget components. We show that both contribute to inconclusiveness of the evidence.

Chapter 2 of this thesis provides the first exploration of the sensitivity of health aid additionality effects to treatment of missing data, reassessing findings of Lu et al. (2010). We demonstrate that multiple imputation of the outcome variable (health spending) leads to results being biased in an ambiguous direction, while the alternative of expressing variables as sub-period averages (a technique commonly applied in development contexts) wipes out most of the variation required for estimation. Furthermore, we bring into light the severe discrepancies in the health spending aggregates across core international data sources, namely WHO and IMF. This is in addition to the data deficiencies in health aid figures in terms of geographical and institutional traceability. These issues compromise the identification of the domestically funded health spending component, yielding conclusions that neither additionality nor fungibility of health aid can be accurately evaluated.

Consequently, Chapter 3 argues that whilst fungibility of health aid cannot be estimated, the broader health aid–spending relationship can be more successfully evaluated, as approaching the issue from the broader fiscal effects angle exerts less pressure on the data and produces interpretable coefficients. Using the best available disaggregated health aid data (Van de Sijpe, 2013), we show that none of the health aid components have a significantly negative effect on total (domestically and externally funded) health aid spending, and that donor projects have the most robust positive association with the recipient's commitment to public health. The size and/or significance of individual coefficients of the results are, nevertheless, sensitive to model specification. We do not identify any credible health aid smoothing effects. Using identical modelling and estimation strategies, we demonstrate that existing estimates of health aid fungibility depend largely on whether donor projects are counted as on- or off-budget.

The existing evaluations of the health aid effects do not include tax revenue in the model. This constitutes an important omission as while the foreign source of (earmarked) revenue is accounted for, the domestic funds available for the government are not. While it is impossible to identify domestic funds committed to a particular sector prior to (or even after) the aid receipts, inclusion of total tax receipts in future evaluations could potentially improve the model by controlling for a broader measure of government's revenue.

The thesis (Chapter 4 on Ethiopia and Chapter 5 on Tanzania) contributes to the growing body of evidence based on time-series methodology for evaluation of the broader fiscal effects of (and on) aid. The case study approach recognises the heterogeneity of developing countries in terms of fiscal dynamics, and allows pinning down the country-specific equilibrium and pushing forces and adjustment mechanisms. Contrary to many empirical applications, detailed understanding of the qualitative context is invoked in the thesis to complement the quantitative data. Not only does it offer guidance for sound statistical model specification and sensible economic interpretation of the estimated results, but it also provides a valuable check over the quality of recipient's quantitative fiscal data. The cumulated evidence from the spectrum of country case studies will eventually allow drawing more robust and reliable conclusions and identifying channels fostering or dampening the potential aid effectiveness. However, future applications should be careful to include the on-budget rather than aggregate (donor) measure of aid, and pay more attention to the qualitative context.

Using the recipient's measure of aid allows one to identify the most direct fiscal impact of aid. However, ignoring the off-budget component constitutes an important omission (even in the reduced form) as it neglects the less direct channels of aid's impact on fiscal aggregates (e.g. institutional or capacity building, facilitating reforms, providing locally unavailable expertise or goods, etc.). Even with both recipient and donor aid data available for Ethiopia and Tanzania, Chapter 6 argues that it is not possible to disaggregate the standard (DAC) aggregate aid flows into on-budget and off-budget components. This is due to the (unmeasurable) presence of disbursements from non-DAC donors in the recipient's data. Given that these flows do not need to comply with the DAC definition of aid, it is difficult to envisage a sizeable improvement to these records in the near future, even if the DAC has started imputing non-DAC flows in their tables. Relying on the recipient's aid records, where available, would nevertheless constitute an improvement over the donor aggregates in tracing the most direct fiscal effects of aid.

CVAR certainly provides an interesting tool to analyse the dynamic fiscal relationships. This thesis demonstrated both plausibility (Ethiopia) and limitations (Tanzania) of applying the method to developing country data: the samples are inevitably (very) small, especially if one accepts that the fiscal decisions follow yearly cycles. Flawed data limits any statistical analysis, but this is even more exposed in the CVAR, where the researcher is supposed to be led by careful testing of the data. As retrospective data is unlikely to be credibly improved, the application of the CVAR to the long-span fiscal and aid dynamics will remain limited. Where analysis is possible, the thesis reiterated the need to use the understanding of the qualitative data in complementing the qualitative analysis.

Inconclusive evidence may misguide policy responses and limit ability to evaluate the effects and effectiveness of aid. The use of less relevant or imprecise measures of aid has been shown to contribute to inconclusiveness of such evidence. In the final contribution, the recipient and donor aid data are directly compared. The anecdotal expectation is that DAC aid records would exceed recipient's own measures, particularly due to acknowledgement of at least a fraction of the off-budget aid. This thesis has shown that the direction of the donor-recipient record discrepancy is not necessarily predictable from the outset, and the explanation of such direction requires the knowledge of qualitative context.

Understanding which data one needs to answer the research question – or which questions can be answered with the data available to the researcher – is as important as it has ever been. It is still at times lacking in the aid literature: donor aid aggregates simply cannot conclusively contribute to aid fungibility or additionality debates, and have been shown in this thesis to garble the evidence of broader fiscal effects of aid. Future research should attend to this rudimentary point and adopt the (disaggregated) aid measure relevant for the question.

Attempts to disaggregate retrospective aid data, however, are likely to continue to be infringed by the lack of detailed data on how and where what aid was sent or spent. As demonstrated by Van de Sijpe (2013), (re)construction of disaggregated sector aid flows is feasible to an extent. Therefore the analysis of how different components of aid affect spending in other sectors – and especially donor priority sectors such as education – are also feasible. Clear description of such data disaggregation attempts is crucial, as then, at least in principle, it may be possible to ensure that the posed questions are feasible – even if that entails adjusting the research question. The same call for detailed description applies to broader data records in the core data sources, including the international databases: signposting which observations were indeed reported and which (and how) were constructed in house could tame the blissful ignorance of the missing data and enable the results to reflect the uncertainty associated to the imputed data.

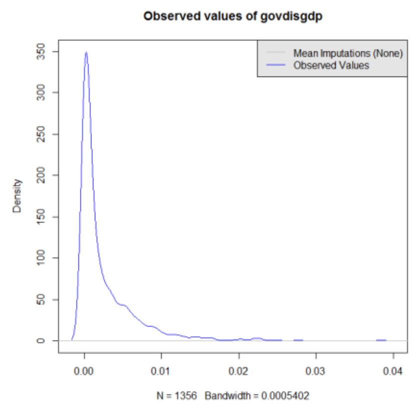
In conclusion:

“In God we trust, all others must bring data” – W. Edwards Deming

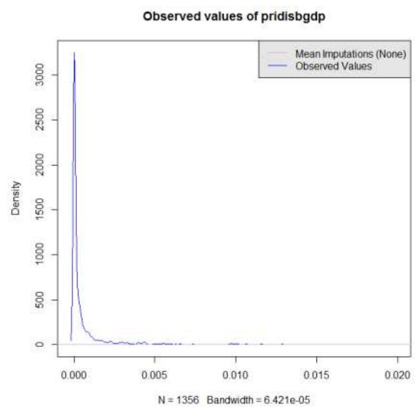
Appendix A

Appendix Figure A1: Distributions of Explanatory Variables

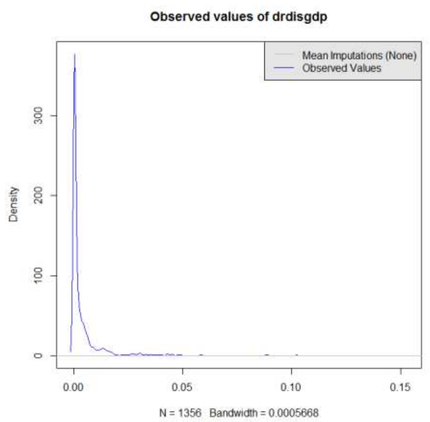
DAH-G/GDP



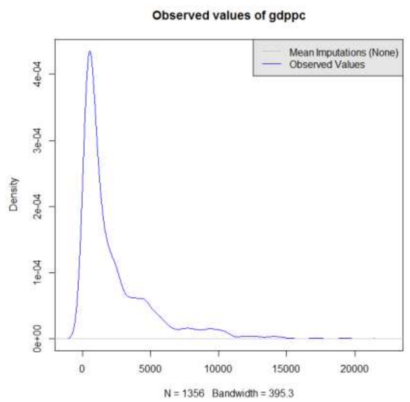
DAH-nG/GDP



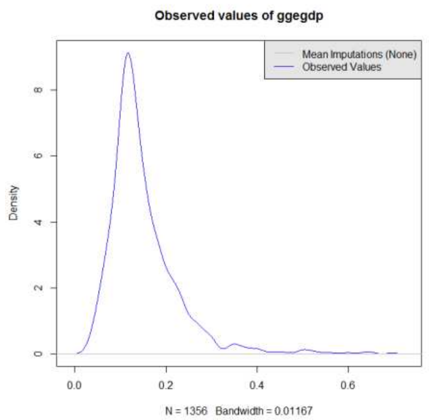
DR/GDP



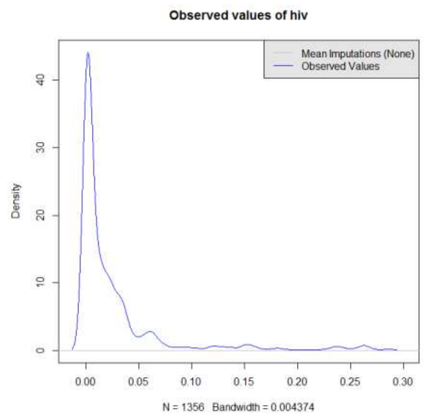
GDPpc



GGE/GDP



HIV

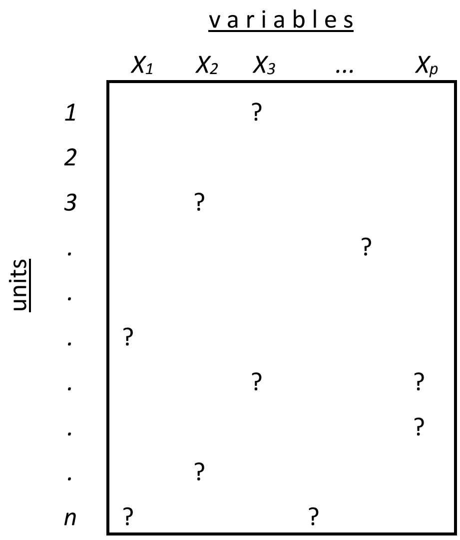


Appendix Table A1: Missingness Mechanism

Relying on Schafer (1997:9), one may consider complete-data (both observed and unobserved) as a "rectangular dataset whose rows can be modelled as independent, identically distributed (iid) draws from some multivariate probability distribution". The schematic representation of such dataset would resemble **Appendix Figure A2** below: the *n* rows represent observational units (e.g. country-years), whilst *p* columns represent the variables recorded for these units. Question marks denote the values that are missing, with remaining matrix entries representing the observed values.

Let *D* denote the complete-data matrix ($n \times p$), including both dependent, *Y*, and explanatory, *X,* variables: $D=\{Y, X\}$. Let I_m denote the missingness[249] indicator matrix, that has the same dimensions as *D* ($n \times p$), but for every observed corresponding cell entry in *D* has "1", and "0" if the corresponding observation is missing. The schematic representation can be depicted as in **Appendix Figure A3**.

Appendix Figure A2: Rectangular Dataset



Appendix Figure A3: (non)Missingness Matrix

	X_1	X_2	X_3	...		X_p
1	1	1	**0**	1	1	1
2	1	1	1	1	1	1
3	1	**0**	1	1	1	1
.	1	1	1	1	**0**	1
.	1	1	1	1	1	1
.	**0**	1	1	1	1	1
.	1	1	**0**	1	1	**0**
.	1	1	1	1	1	**0**
.	1	**0**	1	1	1	1
n	**0**	1	1	**0**	1	1

Three broad types of the underlying missingness mechanism can be distinguished. Following King et al. (2001:50), let D_{obs} and D_{mis} denote observed and missing portions of *D*, respectively, so that $D=\{D_{obs}; D_{mis}\}$. The authors use the following table (**Appendix Table A2**) to summarise the three alternative missingness assumptions, intending to "clarify the assumptions according to [one's] ability to predict the values of I_m", i.e. which values will be missing.

[249] Confusing name: Rather non-missingness/response indicator matrix.

Appendix Table A2: Missingness Mechanism Assumptions

Assumption	Acronym	One can predict I_m with:
Missing completely at random	MCAR	-
Missing at random	MAR	D_{obs}
Non-ignorable	NI	D_{obs} and D_{mis}

In the first case (MCAR), the pattern of missingness cannot be predicted neither by the values of the dependent or independent variable. That is I_m is independent of D – the missing data values are a simple random sample of all data values.[250] In the case of MAR[251], the probability of missingness depends on the observed data, but not on the unobserved data, that is, I_m is independent of D_{mis}.[252] "MAR is less restrictive than MCAR because it requires only that the missing values behave like a random sample of all values within subclasses defined by observed data. In other words, MAR allows the probability that a datum is missing to depend on the datum itself, but only indirectly through quantities that are observed" (Schafer, 1997:11). The missingness process is non-ignorable[253] when the probability that a cell is missing depends on the unobserved value of the missing response, that is, I_m is not independent of D.[254]

The underlying missingness mechanism[255] (or the respective assumption) may determine the validity of the methods employed in the empirical analysis. For instance, the default option of listwise deletion may contain bias in the results, unless the MCAR holds, whilst the

[250] P(M|D) = P(M)

[251] King et al. (2001: 51): "To an extent, the analyst, rather than the world that generates the data, controls the degree to which the MAR assumptions fit. It can be made to fit the data by including more variables in the imputation process to predict the pattern of missingness".

[252] P(M|D) = P(M|D_{obs})

[253] Schafer (1997:11): "To proceed further, we also need to assume that the parameters θ of the data model and the parameters ξ of the missingness mechanism are *distinct.* From a frequentist perspective, this means that the joint parameter space of (θ, ξ) must be the Cartesian cross-product of the individual parameter spaces for θ and ξ. From a Bayesian perspective, this means that any joint prior distribution applied to (θ, ξ) must factor into independent marginal priors for θ and ξ. In many situations this is intuitively reasonable, as knowing θ will provide little information about ξ and vice-versa. If both MAR and distinctness hold, then the missing-data mechanism is said to be ***ignorable*** (Little and Rubin, 1987; Rubin, 1987)."

[254] P(M|D) does not simplify; the observed data cannot alone predict whether a value is missing.

[255] Graham et al.(1994:15) also distinguish between "accessible and inaccessible" data mechanisms: a mechanism is said to be accessible if the cause of missingness has been measured and is available for use in the analysis; mechanism is termed 'inaccessible' if the cause of missingness has not been measured or otherwise is unavailable for analysis.

inferences from analyses using MI are not biased under MCAR or MAR. However, both may be biased under NI.

The methods applied in this paper assume that data are MAR. Usually, MCAR can be rejected in favour of MAR. Unfortunately, “it is not possible to relax the MAR assumption in any meaningful way without replacing it with other equally untestable assumptions” and “[i]n the vast majority of studies, principled methods that assume MAR will then perform better than *ad hoc* procedures such as listwise deletion of imputation of means.” (Schafer and Olsen, 1998:553).[256]

Similarly, “the presence or absence of NI can never be demonstrated using only the observed data. Thus, in most circumstances it is possible to verify whether multiple imputation will outperform (or rather will be expected to perform at least as well as) listwise deletion, but it is not possible to verify absolutely the validity of any multiple imputation model (or, of course, any statistical model)” (King et al. 2001: 50-51). Therefore the absolute validity of a multiple imputation model often cannot be proved in practice.

Relating to Lu et al. (2010), the variable for which a significant proportion is missing is government spending on health. In this case, it seems plausible to reject the MCAR assumption in favour of the MAR, whilst discussion between MAR and NI remains open.

Another issue to be considered in relation to missing data, mostly in the context of multiple imputation, is the pattern of missingness, as it may influence the simplicity of the method applied to the problem of missing data. Based on the data that is presented in a rectangular form (as above), consider an $(n{\times}p)$ data matrix X = $(X_1, X_2, \ldots, X_p)$, where the ordering of variables is not meaningful in economic sense (e.g., it does not make any difference whether a particular variable is labelled X_1 or X_2). Consider a permutation of column indices $(i_1, i_2, \ldots, i_p)$ such that X_{i1} is at least a s observed as X_{i2}, which in turn is at least as observed as X_{i3}. That is, X_{i3} has missing values in the same observations as X_{i2}, and possibly more, whilst X_{i2} has missing values in the same observations as X_{i1}, and possibly more. Provided such a permutation exists, the pattern of missingness in X is said to be *monotone*. This pattern is represented in **Appendix Figure A4**. Otherwise, the pattern of missingness is assumed to be *arbitrary*, as any set of variables may be missing for any unit (as in **Appendix Figure A3**):

[256] Potthoff et al. (2006) discuss a “technique for assessing the degree to which MAR assumption is tenable”.

Appendix Figure A4: Monotone Missingness Pattern

	X_{i1}	X_{i2}	X_{i3}	...		X_{ip}
1	1	1	1	1	1	1
2	1	1	1	1	1	0
3	1	1	1	1	1	0
.	1	1	1	1	1	0
.	1	1	1	0	0	0
.	1	1	1	0	0	0
.	1	1	1	0	0	0
.	1	1	0	1	0	0
.	1	0	0	0	0	0
n	1	0	0	0	0	0

This sort of information about the pattern of missingness may be important if it has a potential of suggesting any reasons for the values to be missing. Also, from a practical perspective, distinguishing between missingness patterns may simplify the imputation process: "[u]nder a monotone missing pattern, a multivariate imputation task can be formulated as a sequence of independent univariate (conditional) imputation tasks, which allows the creation of a flexible imputation model" (Stata 11 Handbook, Missing data: p. 7). In cases where the missingness pattern is arbitrary, one has to rely either on a multivariate normal model (see Little and Rubin, 2002), or use chained equations[257].

[257] http://www.ats.ucla.edu/stat/stata/seminars/missing_data/mi_in_stata_pt1.htm

Appendix Table A3: Inferring MAR

The table reports correlation coefficients. If not MAR, the summary statistics would differ substantially between fully observed data and observations with some missing values. They do not seem to vary substantially.

	DAH-G/GDP	DAH-nG/GDP	DR	GDPpc	GGE/GDP	HIV
DAH-G/GDP	1.0000					
DAH-nG/GDP	0.4805	1.0000				
DR	0.2947	0.2534	1.0000			
GDPpc	-0.3482	-0.1924	-0.2000	1.0000		
GGE/GDP	0.0544	0.0232	-0.0696	0.0618	1.0000	
HIV	0.2250	0.1278	0.0399	-0.0775	0.1580	1.0000

	GHEA/GDPwho	DAH-G/GDP	DAH-nG/GDP	DR	GDPpc	GGE/GDP	HIV
GHEA/GDPwho	1.0000						
DAH-G/GDP	0.1616	1.0000					
DAH-nG/GDP	0.1724	0.4680	1.0000				
DR	0.0339	0.3479	0.2853	1.0000			
GDPpc	0.2367	-0.3598	-0.1980	-0.2042	1.0000		
GGE/GDP	0.2513	0.0264	0.0273	-0.0432	0.0453	1.0000	
HIV	0.2252	0.2486	0.1439	0.0826	-0.0845	0.2115	1.0000

(obs=907)

	GHEA/GDPimf	GHEA/GDP	DAH-G/GDP	DAH-nG/GDP	DR	GDPpc	GGE/GDP	HIV
GHEA/GDPimf	1.0000							
DAH-G/GDP	0.1221	1.0000						
DAH-nG/GDP	0.0640	0.5111	1.0000					
DR	-0.0258	0.2958	0.2122	1.0000				
GDPpc	0.1849	-0.3392	-0.1946	-0.1916	1.0000			
GGE/GDP	0.3334	0.0438	-0.0253	-0.0935	0.0780	1.0000		
HIV	0.2411	0.2122	0.1403	0.0361	-0.0653	0.1653	1.0000	

(obs=1107)

	GHEA/GDPimf	GHEA/GDP	DAH-G/GDP	DAH-nG/GDP	DR
GHEA_gdp_Who	1.0000				
GHEA_gdp_Imf	0.6864	1.0000			
DAH-G/GDP	0.2038	0.1111	1.0000		
DAH-nG/GDP	0.1971	0.0364	0.4993	1.0000	
DR	0.0374	-0.0173	0.3443	0.2782	1.0000

(obs=742)

Appendix Table A4: GHE-A (Imputed variable) Summary Statistics across 'Treatments'

GHE-A/GDP (WHO)	Missing	Original paper	Imputed						
	Fully observed pooled sample	Lu et al. (imp ave, IHME)	(1) Amelia2014OneA as in Lu et al.	(2) Amelia2014OneA Averaged (SI)	(3) Amelia2014ThreeK3 Common trend (time poly k=3)	(4) Amelia2014FourFEA Fixed effects	(5) Amelia2014FiveNoAs No (extra) assumptions	(6) Fully observed	(7) 3-year averages
	WHO	WHO	WHO	WHO	WHO	WHO	WHO	WHO	WHO
Mean	.0257743	.0251097	.0252846	.0252846	.0252249	.0255207	.0250556	.0250816	.0254213
Std d	.0133041	.0121835	.0128206	.0120938	.0127951	.0124033	.0131121	.0126955	.0138871
Min	.0027078	.0027078	.0027078	.0027078	.0027078	-.0254634	-.0238577	.0027078	.0033117
Max	.0955123	.0955123	.0955123	.0955123	.0955123	.0955123	.0955123	.0731758	.0923938
N	907	1356	135600	1356	135600	135600	135600	456	260
m	0		100	100	100	100	100	0	0

GHE-A/GDP (IMF)	Missing	Original paper	Imputed						
	Fully observed pooled sample	Lu et al. (imp ave, IHME)	(1) Amelia2014OneA (100 imp)bounded to observed min/max	(2) Amelia2014OneA Averaged (SI)	(3) Amelia2014ThreeK3	(4) Amelia2014FourFEA Fixed effects	(5) Amelia2014FiveNoAs No (extra) assumptions	Fully observed	(7) 3-year averages
	IMF	IMF	IMF	IMF	IMF	IMF	IMF	IMF	IMF
Mean	.0204936	.0199258	.0206104	.0206104	.0207851	.0201403	.0205103	.0223602	.0208467
Std d	.0123881	.0120688	.0122574	.0118438	.0122372	.0121514	.0124485	.0130465	.0120211
Min	.0017122	.0017122	.0017122	.0017122	.0017122	.0017121	-.0246588	.0017122	.0021134
Max	.0867008	.0867008	.0867008	.0867008	.0867008	.0867008	.0867008	.0867008	.0709217
N	1107	1356	135600	1356	135600	135600	135600	516	340
m			100	100	100	100	100	0	

Appendix Table A5a: The Estimated Results (WHO)

Dependent variable: GHE-S/GDP								
WHO Variable/ Treatment	**Original Lu et al. (2010)**	**(1) MI acc. to Lu et al. descr. (lags and leads)**	**(2) averaged (SI) (1)**	**(3) Amelia2014ThreeK3 3 common time polynomials, bounded**	**(4) Amelia2014 FourFEA Fixed effects**	**(5) Amelia2014FiveNoAs No (extra) assumptions**	**(6) Fully observed (Subsample)**	**(7)** 3-year averages
N	***111***	**111**	**111**	**111**	**111**	111	38	64
Logged GHE-S/GDP	***.597****	**.321*****	**.659*****	**.191****	**.368*****	**.164***	**.626*****	**.990*****
	(.098)	**(.096)**	**(.061)**	**(.092)**	**(.123)**	**(.082)**	**(.074)**	**(.280)**
DAH-G/GDP	**-.457*****	**-.629*****	**-.542*****	**-.677*****	**-.519****	**-.649****	**-.766*****	**-.693*****
	(.107)	**(.190)**	**(.089)**	**(.253)**	**(.204)**	**(.255)**	**(.123)**	**(.121)**
DAH-nG/GDP	**.691*****	**1.166*****	**1.007*****	**1.098*****	**.818*****	**1.152*****	.213	.363
	(.155)	**(.308)**	**(.170)**	**(.322)**	**(.233)**	**(.328)**	(.254)	(.357)
DR/GDP	.053	.072	**.060***	.076	.083	.090	.028	.033
	(.038)	(.064)	**(.032)**	(.074)	(.057)	(.076)	(.096)	(.062)
GDPpc	-.000	-.000	-.000	-.000	-.000	.000	-.000	-.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
GGE/GDP	**.026*****	.022	**.022****	.027	.017	.030	.024	**.059***
	(.010)	(.019)	**(.009)**	(.023)	(.017)	(.023)	(.015)	**(.029)**
HIV	.042	.066	**.0717****	.060	.039	.052	.010	.092
	(.0319)	(.042)	**(.028)**	(.049)	(.033)	(.048)	(.019)	(.067)
constant	**.005****	**.011*****	**.003***	**.014*****	**.012*****	**.013*****	**.007****	-.006
	(.003)	**(.004)**	**(.002)**	**(.004)**	**(.003)**	**(.004)**	**(.003)**	(.006)

Each cell reports the estimated coefficient; Standard errors are reported in brackets; p-values are reported in square brackets.

LR effects (continued on next page)

LR effects								
		(1)	**(2)**	**(3)**	**(4)**	**(5)**	(6)	(7)
DAH-G/GDP	-1.135*** [p-value 0.000]	-.930*** (.251)	-1.592*** [p-value 0.000]	-.839*** (.285)	-.825*** (.254)	-.777*** (.282)	-2.046*** [p-value 0.000]	-66.873 [p-value 0.971]
DAH-nG/GDP	1.715** [p-value 0.013]	1.723*** (.525)	2.957*** [p-value 0.000]	1.361 *** (.436)	1.304** (.523)	1.379*** (.421)	.570 [p-value 0.427]	34.990 [p-value 0.971]
DR/GDP	.132 [p-value 0.250]	.107 (.100)	.177* [p-value 0.098]	.094 (.094)	.134 (.103)	.108 (.093)	.076 [0.771]	3.137 [0.971]
GDPpc	-.000 [N.A.]	N.A	-.000 [N.A.]	N.A.	N.A.	N.A.	-.000 [p-value 0.564]	-.000 [p-value 0.971]
GGE/GDP	.064*** [p-value 0.001]	.032 (.026)	.064** [p-value 0.014]	.034 (.027)	.028 (.025)	.036 (.027)	.0634* [p-value 0.0840]	5.684 [p-value 0.970]
HIV	.104 [p-value 0.273]	.098 (.064)	.211** [p-value 0.031]	.074 (.060)	.061 (.055)	.062 (.058)	.027 [p-value 0.610]	8.903 [p-value 0.970]

Appendix Table A5b: The Estimated Results (IMF)

Dependent variable: GHE-S/GDP								
IMF Variable/ Treatment	**Original Lu et al. (2010)**	**(1) MI acc. to Lu et al. descr.**	**(2) averaged (SI) (1)**	**(3) Amelia2014ThreeK3 3 common time polynomials**	**(4) Amelia2014FourFEA Fixed effects**	**(5) Amelia2014FiveNoAs No (extra) assumptions**	**(6) Fully observed**	(7) 3-year averages
N	**111**	**111**	**111**	**111**	**111**	**111**	**41**	83
Logged GHE-S/GDP	***.573****	**.406*****	**.603*****	**.293*****	**.414*****	**.259*****	**.582*****	**.704*****
	(.055)	**(.084)**	**(.060)**	**(.084)**	**(.065)**	**(.081)**	**(.047)**	**(.184)**
DAH-G/GDP	**-.433*****	**-.663*****	**-.597*****	**-.716*****	**-.603*****	**-.729*****	**-.560*****	**-.536*****
	(.090)	**(.141)**	**(.107)**	**(.1562842)**	**(.117)**	**(.158)**	**(.165)**	**(.146)**
DAH-nG/GDP	**.580*****	**.563*****	**.571*****	**.520***	**.551*****	**.497 ***	**.428****	.320
	(.147)	**(.215]**	**(.173)**	**(.260)**	**(.190)**	**(.264)**	**(.179)**	(.293)
DR/GDP	-.010	.018	.023	.019	.012	.006	-.002	-.071
	(.030)	(.061)	(.034)	(.064)	(.044)	(.068)	(.026)	(.042)
GDPpc	-.000	-.000	-.000	-.000	-.000	-.000	-.000	-.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
GGE/GDP	**.020****	.020	**.019***	.030	.018	.031	.000	.026
	(.009)	(.018)	**(.011)**	(.019)	(.012)	(.020)	(.013)	(.018)
HIV	.028	.026	.027	.060	.048	.060	.048**	.003
	(.026)	(.041)	(.024)	(.046)	(.033)	(.048)	(.023)	(.041)
constant	**.005*****	**.009****	**.005****	**.009****	**.008*****	**.009****	**.008****	.005
	(.002)	**(.003)**	**(.002)**	**(.004)**	**(.002)**	**(.004)**	**(.003)**	(.005)

Each cell reports the estimated coefficient; Standard errors are reported in brackets; p-values are reported in square brackets.

LR effects (continued on next page)

LR Effects								
		(1)	**(2)**	**(3)**	**(4)**	**(5)**	(6)	(7)
DAH-G/GDP	**-1.013*** [p-value 0.000]**	**-1.119*** (.260)**	**-1.506*** [p-value 0.000]**	**-1.017*** (.239)**	**-1.032*** (.233)**	**-.987*** (.2307)**	**-1.337** [p-value 0.001]**	-1.808 [p-value 0.102]
DAH-nG/GDP	**1.359*** [p-value 0.000]**	**.954** (.376)**	**1.440*** [p-value 0.001]**	**.740* (.380)**	**.944** (.327)**	**.673* (.365)**	**1.023** [p-value 0.011]**	1.080 [p-value 0.405]
DR/GDP	-.022 [p-value 0.749]	.031 (.103)	.060 [p-value 0.498]	.027 (.091)	.021 (.076)	.008 (.091)	-.004 [p-value 0.942]	-.238 [p-value 0.220]
GDPpc	-.000 [N.A.]	N.A.	-.000 [N.A.]	N.A.	N.A.	N.A.	-.000 [N.A.]	-.000 [p-value 0.266]
GGE/GDP	.047** [p-value 0.026]	.034 (.030)	.048** [0.078]	.042 (.026)	.031 (.021) [0.139]	.043 (.027)	.000 [p-value 0.988]	.086 [p-value 0.231]
HIV	.066 [p-value 0.282]	.043 (.068)	.068 [p-value 0.247]	.085 (.064)	.082 (.055)	.081 (.064)	**.116* [p-value 0.041]**	.012 [p-value 0.932]

Appendix Table A6: 10 Repetitions (Direction of Bias, IMF Sample)

Dependent variable: Domestically funded public health spending (GHE-S), IMF; ABBB

	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)
	MI acc. to Lu et al. (lags and leads)	**MI acc. to Lu et al. (lags and leads)**	**MI acc. to Lu et al. (lags and leads)**	**MI acc. to Lu et al. (lags and leads)**	**MI acc. to Lu et al. (lags and leads)**	**MI acc. to Lu et al. (lags and leads)**	**MI acc. to Lu et al. (lags and leads)**	**MI acc. to Lu et al. (lags and leads)**	**MI acc. to Lu et al. (lags and leads)**	**MI acc. to Lu et al. (lags and leads)**
Logged GHE-S/GDP	.584***	.590***	.589***	.588***	.588***	.589***	.589***	.589***	.591***	.580***
	(.066)	(.067)	(.066)	(.065)	(.067)	(.072)	(.067)	(.065)	(.070)	(.067)
DAH-G/GDP	-.531**	-.531**	-.534**	-.539**	-.537**	-.536**	-.537**	-.544**	-.531**	-.546**
	(.215)	(.216)	(.217)	(.215)	(.215)	(.217)	(.217)	(.215)	(.216)	(.217)
DAH-nG/GDP	.312*	.320*	.321*	.315	.322*	.310	.336*	.335*	.317*	.322
	(.189)	(.191)	(.190)	(.199)	(.193)	(.198)	(.191)	(.195)	(.192)	(.196)
DR/GDP	-.0001	.003	.001	-.000	.000	.003	.001	-.001	.001	.000
	(.035)	(.036)	(.036)	(.034)	(.035)	(.036)	(.035)	(.035)	(.034)	(.036)
GDPpc	-.000	-.000	-.000	-.000	-.000	-.000	-.000	-.000	-.000	-.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
GGE/GDP	-.0001	.001	.001	.001	.002	.001	-.000	.000	-.000	.000
	(.018)	(.017)	(.017)	(.017)	(.018)	(.018)	(.018)	(.018)	(.018)	(.018)
HIV	.058**	.057**	.059**	.059**	.058**	.059**	.058**	.059**	.059**	.059**
	(.026)	(.027)	(.026)	(.027)	(.027)	(.026)	(.026)	(.026)	(.026)	(.027)
constant	.008*	.007*	.007*	.007*	.007*	.007*	.007*	.007*	.007*	.008*
	(.004)	(.004)	(.004)	(.004)	(.004)	(.004)	(.004)	(.004)	(.004)	(.004)
N	41	41	41	41	41	41	41	41	41	41
T	12	12	12	12	12	12	12	12	12	12

Each cell reports the estimated coefficients from ABBB. Standard errors are reported in brackets. N denotes number of countries (not observations).

Dependent variable: Domestically funded public health spending (GHE-S); IMF, ABBB

	(3) MI common time polyno-mials	(3) MI common time polyno-mials	(3) MI common time polyno-mials	(3) MI common time polyno-mials	(3) MI common time polyno-mials	(3) MI common time polyno-mials	(3) MI common time polyno-mials	(3) MI common time polyno-mials	(3) MI common time polyno-mials	(3) MI common time polyno-mials
Logged GHE-S/GDP	.188**	.178**	.176*	.184*	.175*	.181*	.188**	.175**	.189**	.170*
	(.092)	(.090)	(.093)	(.095)	(.092)	(.094)	(.092)	(.092)	(.093)	(.091)
DAH-G/GDP	-.637***	-.636***	-.635***	-.644***	-.637***	-.631***	-.634***	-.636***	-.636***	-.624***
	(.242)	(.235)	(.230)	(.230)	(.238)	(.232)	(.236)	(.236)	(.229)	(.238)
DAH-nG/GDP	.408	.406	.378	.395	.367	.371	.394	.363	.385	.363
	(.376)	(.364)	(.363)	(.357)	(.369)	(.358)	(.352)	(.352)	(.361)	(.362)
DR/GDP	.019	.022	.015	.011	.017	.023	.020	.020	.019	.014
	(.064)	(.065)	(.063)	(.059)	(.064)	(.066)	(.065)	(.066)	(.065)	(.065)
GDPpc	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GGE/GDP	-.003	-.002	-.006	-.004	-.002	-.003	-.005	-.006	-.003	-.003
	(.029)	(.030)	(.030)	(.030)	(.029)	(.031)	(.030)	(.030)	(.029)	(.029)
HIV	.013	.022	.017	.019	.018	.016	.018	.022	.021	.015
	(.047)	(.050)	(.048)	(.048)	(.048)	(.048)	(.049)	(.048)	(.050)	(.048)
constant	.015***	.015***	.016***	.015***	.015***	.015***	.015***	.015***	.015***	.015***
	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)
N	41	41	41	41	41	41	41	41	41	41
T	12	12	12	12	12	12	12	12	12	12

Each cell reports the estimated coefficients from ABBB. Standard errors are reported in brackets.
N denotes number of countries (not observations).

Dependent variable: Domestically funded public health spending (GHE-S); IMF, ABBB

	(4) MI Fixed effects	(4) MI Fixed effects	(4) MI Fixed effects	(4) MI Fixed effects	(4) MI Fixed effects	(4) MI Fixed effects	(4) MI Fixed effects	(4) MI Fixed effects	(4) MI Fixed effects	(4) MI Fixed effects
Logged GHE-S/GDP	.440***	.445***	.446***	.438***	.439***	.441***	.441***	.438***	.441***	.439***
	(.070)	(.073)	(.073)	(.074)	(.069)	(.068)	(.070)	(.070)	(.075)	(.073)
DAH-G/GDP	-.616***	-.606***	-.617***	-.615***	-.619***	-.619***	-.612***	-.620***	-.619***	-.605***
	(.224)	(.228)	(.232)	(.231)	(.227)	(.231)	(.230)	(.228)	(.232)	(.227)
DAH-nG/GDP	.457*	.465*	.472*	.469*	.457*	.455*	.471*	.466*	.485**	.453*
	(.254)	(.242)	(.248)	(.254)	(.246)	(.242)	(.252)	(.245)	(.243)	(.253)
DR/GDP	-.005	-.001	.001	.001	-.002	-.004	-.006	-.005	-.006	-.003
	(.043)	(.043)	(.042)	(.041)	(.044)	(.042)	(.043)	(.042)	(.043)	(.043)
GDPpc	-.000	-.000	-.000	-.000	-.000	-.000	-.000	-.000	-.000	-.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
GGE/GDP	-.011	-.010	-.010	-.0010	-.011	-.012	-.011	-.013	-.013	-.012
	(.025)	(.024)	(.023)	(.024)	(.024)	(.025)	(.024)	(.024)	(.024)	(.024)
HIV	.040	.041	.040	.041	.044	.045	.041	.045	.044	.041
	(.032)	(.032)	(.032)	(.033)	(.033)	(.031)	(.032)	(.032)	(.033)	(.032)
constant	.013***	.013***	.013***	.013***	.013***	.013***	.013***	.013***	.013***	.013***
	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)
N	41	41	41	41	41	41	41	41	41	41
T	12	12	12	12	12	12	12	12	12	12

Each cell reports the estimated coefficients from ABBB. Standard errors are reported in brackets.
N denotes number of countries (not observations).

Dependent variable: Domestically funded public health spending (GHE-S); IMF, ABBB

	(5) MI with No (extra) assump-tions	(5) MI with No (extra) assump-tions	(5) MI with No (extra) assump-tions	(5) MI with No (extra) assump-tions	(5) MI with No (extra) assump-tions	(5) MI with No (extra) assump-tions	(5) MI with No (extra) assump-tions	(5) MI with No (extra) assump-tions	(5) MI with No (extra) assump-tions	(5) MI with No (extra) assump-tions
Logged GHE-S/GDP	.162* (.091)	.157 (.096)	.167* (.095)	.164* (.089)	.159* (.090)	.152 (.092)	.158* (.093)	.163* (.092)	.153* (.090)	.155* (.092)
DAH-G/GDP	-.641*** (.236)	-.633** (.245)	-.637*** (.227)	-.644*** (233)	-.656*** (.239)	-.643*** (.236)	-.663*** (.239)	-.639*** (.239)	-.645** (.251)	-.643*** (.240)
DAH-nG/GDP	.348 (.372)	.310 (.398)	.363 (.389)	.362 (.388)	.373 (.417)	.319 (.388)	.380 (.382)	.339 (.370)	.356 (.419)	.307 (.382)
DR/GDP	.001 (.076)	-.007 (.076)	-.004 (.074)	.005 (.076)	.003 (.073)	-.004 (.075)	-.000 (.076)	.001 (.075)	-.004 (.076)	-.011 (.079)
GDPpc	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)
GGE/GDP	.001 (.033)	-.001 (.033)	-.003 (.032)	-.000 (.032)	.001 (.031)	-.002 (.032)	-.002 (.032)	-.001 (.032)	-.003 (.032)	-.005 (.032)
HIV	.028 (.048)	.026 (.050)	.023 (.052)	.020 (.050)	.028 (.050)	.019 (.050)	.019 (.050)	.024 (.047)	.024 (.053)	.027 (.054)
constant	.014** (.006)	.015** (.006)	.015** (.006)	.015** (.006)	.015** (.006)	.015** (.006)	.015** (.006)	.015** (.006)	.015 (.006)	.016 (.006)
N	41	41	41	41	41	41	41	41	41	41
T	12	12	12	12	12	12	12	12	12	12

Each cell reports the estimated coefficients from ABBB. Standard errors are reported in brackets.
N denotes number of countries (not observations).

Appendix Table A7: Full Set of Estimates of Direction of Bias (Including Long Run)

	Multiple Imputation (N=41)					Complete Case[258] (N=2)	Single Imputation (N=40)
	Full sample (fully observed, N=41)	(1) Amelia2 ExIMFone	(3) Amelia2Ex IMFthree	(4) Amelia2Ex IMFfourFE	(5) Amelia2Ex IMFfive	(6)	(7) Sub-period averages (SI)
	ABBB	ABBB	ABBB	ABBB	ABBB		ABBB
Logged GHE-S/GDP	**.582*** (.047)**	**.584*** (.066)**	**.188** (.092)**	**.440*** (.070)**	**.162* (.091)**	-	**.867*** (.321)**
DAH-G/GDP	**-.560*** (.165)**	**-.531** (.215)**	**-.637*** (.242)**	**-.616*** (.224)**	**-.641*** (.236)**	-	-.540 (.328)
DAH-nG/GDP	**.428** (.179)**	**.312* (.189)**	.408 (.376)	**.457* (.254)**	.348 (.372)	-	.093 (.388)
DR/GDP	-.002 (.026)	-.0001 (.035)	.019 (.064)	-.005 (.043)	.001 (.076)	-	**-.100* (.055)**
GDPpc	-.000 (.000)	-.000 (.000)	.000 (0.000)	-.000 (.000)	.000 (.000)	-	.000 (.000)
GGE/GDP	.0002 (.013)	-.0001 (.018)	-.003 (.029)	-.011 (.025)	.001 (.033)	-	-.006 (.024)
HIV	**.048** (.023)**	**.058** (.026)**	.013 (.047)	.040 (.032)	.028 (.048)	-	**.131** (.062)**
constant	**.008** (.003)**	**.008* (.004)**	**.015*** (.006)**	**.013*** (.005)**	**.014** (.006)**		.0002 (.009)
				Long-Run Effects			
DAH-G/GDP	-1.337*** [0.0014]	-1.283 (.501)	-.783 (.287)	-1.102 (.398)	-.767 (.281)	-	-4.057 [0.637]
DAH-nG/GDP	1.023** [0.0114]	.754 (.473)	.499 (.458)	.816 (.453)	.417 (.445)	-	.702 [0.872]
DR/GDP	-.004 [0.942]	-.000 (.085)	.024 (.079)	-.008 (.077)	.002 (.091)	-	-.752 [0.677]
GDPpc	-.000 [N.A.]	N.A.	N.A.	N.A.	N.A.	-	.000 [N.A.]
GGE/GDP	.0005 [0.9879]	-.000 (.044)	-.004 (.036)	-.020 (.045)	.001 (.040)	-	-.043 [0.824]
HIV	.116** [0.0407]	.141 (.058)	.015 (.058)	.071 (.055)	.034 (.058)	-	.985 [0.668]

Standard errors reported in parentheses, p-values reported in the square brackets.

[258] Not done as only two countries 'fully observed'

Appendix Table A8: Fixed Effects Estimates of Direction of Bias

		Multiple Imputation (N=41)				Complete Case[259] (N=2)	Sub-period averages (SI) (N=40)
	Full sample (fully observed, N=41)	(1) Amelia2 ExIMFone	(3) Amelia2Ex IMFthree	(4) Amelia2Ex IMFfourFE	(5) Amelia2Ex IMFfive	(6)	(7)
	FE [260]	FE	FE	FE	FE		FE
DAH-G/GDP	**-.605**** **(.229)**	**-.589**** **(.262)**	**-.678***** **(.224)**	**-.618**** **(.236)**	**-.661***** **(.225)**	-	-.356 (.453)
DAH-nG/GDP	.169 (.303)	.016 (.314)	.190 (.259936)	.233 (.238)	.170 (.252)	-	-.105 (.468)
DR/GDP	.039 (.049)	.050 (.044)	.049 (.041)	**.064*** **(.037)**	.036 (.043)	-	.0338 (.050)
GDPpc	-.000 (.000)	-.000 (.000)	.000 (.000)	-.000 (.000)	.000 (.000)	-	.000 (.000)
GGE/GDP	.003 (.019)	.003 (.020)	.010 (.015)	.001 (.016)	.008 (.015)	-	-.011 (.031)
HIV	.054 (.040)	.047 (.044)	**.071*** **(.037)**	.057 (.036)	**.069*** **(.038)**	-	.051 (.049)
constant	**.018***** **(.004)**	**.018***** **(.004)**	**.015***** **(.003)**	**.017***** **(.003)**	**.016***** **(.003)**	**-**	**.019***** **(.006)**
	R2=0.17	R2=0.18	R2=0.23	R2= 0.14	R2= 0.22	-	R2=0.10[261]

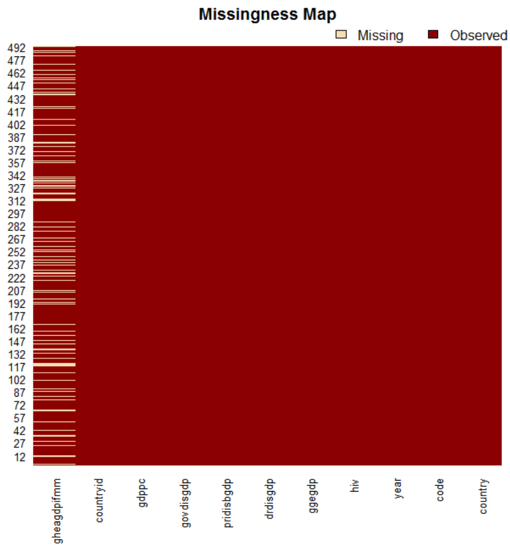
Standard errors reported in parentheses; p-values reported in the square brackets.

The fixed effects results differ substantially from the ABBB estimates. Firstly, considering only the full sample estimates (first column), the only variable (constant aside) estimated to have a significant effect on government's 'domestically funded' health expenditures is the health aid (assumed to be) flowing through the budget (however, this particularly may be affected by the fact that DAH-G/GDP is directly used in construction of the dependent variable). The insignificance of DAH-G/GDP is economically plausible: aid that does not flow through the government (DAH-nG/GDP) is less likely (if at all) to alter government's domestically funded spending as the government may not be fully aware of these flows. Debt relief, especially assumed to be uniformly distributed over 10 year period, constitutes small quantities of 'released' funds. The lack of estimated effect of GDP per capita is less explicable, as one would expect that public health bill would increase with increasingly wealthy population; Lu et al. do not log the variable. Lack of estimated association between

[259] Not done as only two countries 'fully observed'
[260] xtreg gheSgdpifmm govdisgdp pridisbgdp drdisgdp gdppc ggegdp hiv, fe vce(robust);
[261] Prob > F = 0.3691, others=0.

HIV and GHE-S/GDP is not particularly surprising, as it may go in either direction in the first place. In terms of comparability of results across alternative methodological choices to handle the missing data, again, none performs ideally, but multiple imputation estimates seem to be overall closer to the full sample estimates than those from single imputation (sub-period averages), confirming the statistical prediction even in our small sample.

Appendix Figure A5: The Resulting Distribution of Missing Values across Country-Year Observations



Appendix Table A9: The Pattern of Missingness across Countries over the Entire Year Sample Period

country	# WHO miss	# IMF miss	country	# WHO miss	# IMF miss	country	# WHO miss	# IMF miss
Algeria	10	4	Gabon	6	2	Nigeria	5	0
(Angola)	**0**	**0**	Gambia	2	0	Oman	11	0
Argentina	0	7	Georgia	0	4	Pakistan	0	5
Armenia	0	2	Ghana	11	0	Panama	1	0
Azerbaijan	11	0	Guatemala	0	2	Papua New Guinea	1	0
Bahrain	5	0	Guinea	11	0	Paraguay	0	2
Bangladesh	1	2	Guinea-Bissau	2	0	Peru	0	3
Barbados	0	2	Guyana	0	1	Philippines	0	5
Belize	4	1	Haiti	7	5	Rwanda	8	4
Benin	5	0	Honduras	8	2	Samoa	8	3
Bhutan	0	1	India	2	1	Saudi Arabia	2	0
Bolivia	4	0	Indonesia	2	2	Senegal	6	0

Botswana	6	0	Iran, Islamic Republic of	10	3	Sierra Leone	7	0
Brazil	0	4	Jamaica	0	3	Solomon Islands	2	1
Burkina Faso	8	0	Jordan	4	3	**South Africa**	**0**	**0**
Burundi	**0**	**0**	**Kazakhstan**	**0**	**0**	Sri Lanka	0	2
Cambodia	**0**	**0**	Kenya	9	0	Sudan	10	5
Cameroon	11	1	Kyrgyzstan	0	1	Suriname	11	8
Cape Verde	3	6	Lao People's Democratic Republic	1	1	Swaziland	9	0
Central African Republic	2	3	Lebanon	8	2	Syrian Arab Republic	7	0
Chad	8	2	**Lesotho**	**0**	**0**	Tajikistan	0	1
Chile	0	2	Liberia	3	5	Tanzania, United Republic of	9	3
China	0	2	Libyan Arab Jamahiriya	10	3	Thailand	1	5
Colombia	1	8	Madagascar	11	0	Togo	11	2
Comoros	7	4	Malawi	2	0	Trinidad and Tobago	4	8
Congo	7	11	**Malaysia**	**0**	**0**	Tunisia	7	0
Congo, the Democratic Republic of the	**12**	**12**	**Maldives**	**0**	**0**	**Turkey**	**0**	**0**
Costa Rica	0	8	Mali	7	0	Turkmenistan	12	0
Côte d'Ivoire	**0**	**0**	Mauritania	1	1	Uganda	8	2
Djibouti	11	5	Mauritius	1	1	Uruguay	2	7
Dominican Republic	0	6	Mexico	0	6	**Uzbekistan**	**0**	**0**
Ecuador	0	1	Mongolia	1	0	Vanuatu	0	5
Egypt	10	4	Morocco	1	2	Venezuela	1	4
El Salvador	0	7	Mozambique	4	0	Viet Nam	0	2
Equatorial Guinea	10	8	**Namibia**	**0**	**0**	Yemen	10	1
Eritrea	7	0	Nepal	0	1	Zambia	1	0
Ethiopia	9	0	Nicaragua	0	2	Zimbabwe	9	1
Fiji	1	4	Niger	9	0			

Appendix Table A10: Correlation Coefficients between Dependent and Explanatory Variables

corr	Fully simultaneously observed		Pooled simultaneously observed	
	GHE-A/GDP (IMF)	GHE-A/GDP (WHO)	GHE-A/GDP (IMF)	GHE-A/GDP (WHO)
DAH-G/GDP	-0.2207	-0.3043	0.0869	0.2086
DAH-NG/GDP	-0.2417	-0.2172	0.0305	0.1987
DR	-0.4318	-0.5252	-0.0142	0.0353
GGE/GDP	0.6604	0.5347	0.3969	0.3647
HIV	0.5244	0.3119	0.2592	0.2649
GDP pc	0.1769	0.3113	0.2350	0.2192
Corr (WHO IMF)	0.8144		0.7075	
	N=11; YO=132	N=11; YO=132	N=111; YO=725	N=111; YO=725

Table reports correlation coefficient between the government total health spending (GHE-A) and explanatory variables for subsamples where GHE-A data are coded as fully observed (11 countries, left part of the table), and at least partially observed (725 country-year observations, reported in the right part of the table.

For only 11 (eleven) countries are the GHE-A variable simultaneously fully observed in the WHO and IMF samples. These countries (Burundi, Cambodia, Cote d'Ivoire, Kazakhstan, Lesotho, Malaysia, Maldives, Namibia, South Africa, Turkey, and Uzbekistan (Angola dropped)) , although does not have a ring of a representative sample, have very comparable averages for key variables (and are geographically disbursed) to the full sample (but not in terms of covariates). Importantly, they are not on average richer than the full sample of 111 countries.

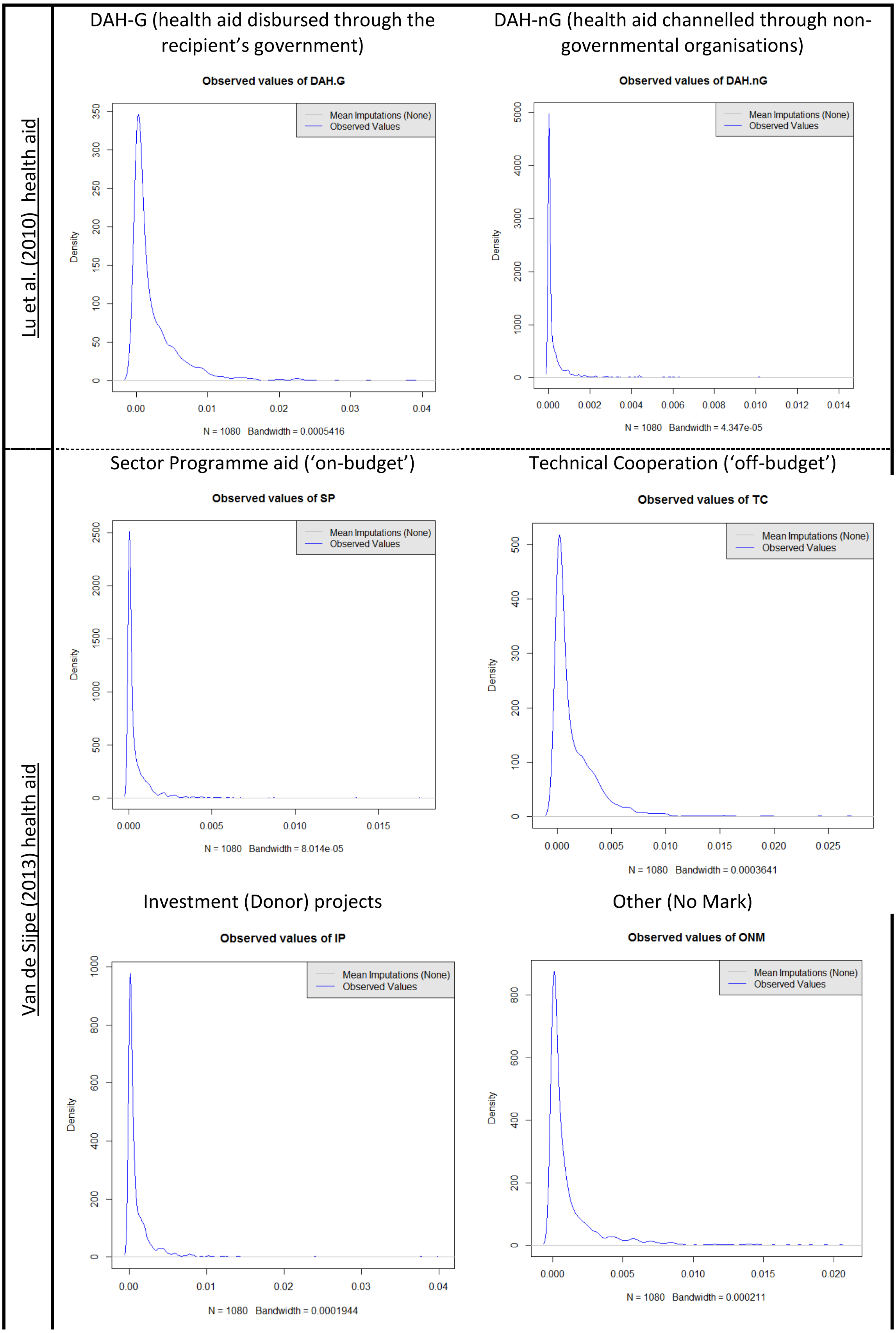
The cross-variable correlations, however, reveal important differences. The correlation between WHO and IMF GHE-A/GDP measures is higher for these 11 countries (0.81) compared to the pooled sample where these two measures overlap (0.69). The key difference is in correlation between government health spending (GHE-A/GDP) and health aid variables: whilst in the pooled fully observed sample (YO=725) the correlation between health aid and government health spending is small but positive (between 0.04 and 0.2), in the 11 fully simultaneously observed country sample it is negative and larger in magnitude (between -0.30 and -0.22).

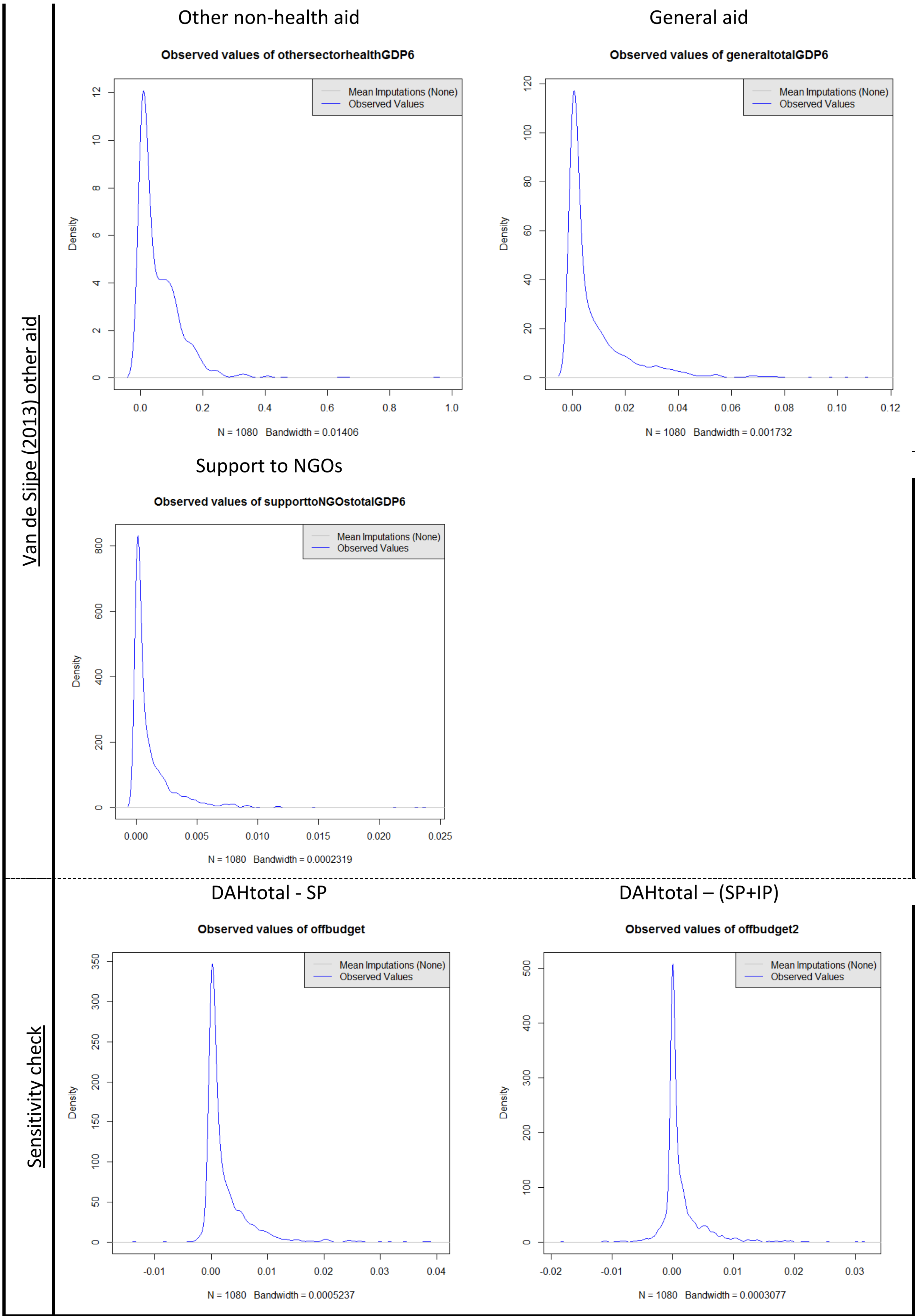
In the fully simultaneously observed sample of 11 countries over 12 year period[262], the GHE-A/GDP (for both IMF and WHO samples) seems to covariate relatively strongly with the independent variables. It bears considerable *negative* correlation with the aid variables (DAH-G/GDP, DAH-nG/GDP, and Debt Relief) (they received slightly less aid than the pooled sample, or full sample). Government total health spending is also strongly correlated with government total spending (GGE/GDP), and also recorded/estimated HIV prevalence, and indicating that in these countries health spending grows as government size increases an potentially with disease burden (although the latter may capture the reverse causality of higher spending leading to more attention to diseases). Potentially, this could reflect that *other variables* have different underlying missingness rate (although all recorded as fully observed): the countries outside the 11 fully observed country sample have poorer GHE-A data, and potentially poorer quality data across other variables.

[262] Note that a fraction of this correlation may spurious due to some I(1)ness in the short sample; however, this would be the case in the full sample as well.

Appendix B

Appendix Figure B1: Distributions of Health Aid Variables





The figure depicts disaggregated health aid variables (Lu et al., 2010; Van de Sijpe (2013); and the sensitivity check), for 108 low- and middle-income countries over 1995-2004.

Appendix Table B1: Lu et al. Distinction (FE Results) (Testing for Aid Smoothing and Missing Data Effects)

Dependent variable: GHE-A/GDP (Lu et al.) [Lu et al. health aid disaggregation]									
FE, vce (R)	a	a	a	b	b	b	c	c	c
N YO	50 500	50 450	50 450	107 896	107 808	107 808	108 1080	108 972	108 972
On-budget DAHG	**.3831*** (.1106)**		**.2419** (.1094)**	**.3727*** (.0635)**		**.2362*** (.0770)**	**.3570*** (.0707)**		**.2717*** (.0774)**
Lagged On-budget DAHG		**.3560** (.1541)**	**.2635* (.1553)**		**.3335*** (.0853)**	**.1834* (.0996)**		**.3170*** (.0883)**	.1326 (.0953)
Off-budget DAHnG	-.1544 (.3133)		-.2764 (.4101)	.0265 (.1830)		-.0799 (.2166)	.1676 (.1646)		.0669 (.1717)
Lagged Off-budget DAHnG		.1023 (.4921)	-.0525 (.3883)		.4341 (.3198)	.2823 (.2849)		**.3909* (.2183)**	.1641 (.1682)
DR	-.0438 (.0485)	.0102 (.0553)	.0061 (.0548)	-.0658 (.0495)	-.0152 (.0554)	-.0201 (.0544)	-.0202 (.0439)	.0213 (.0352)	.0166 (.0345)
Ln(GDPpc)	-.0021 (.0020)	-.0023 (.0021)	-.0016 (.0020)	-.0026 (.0016)	**.0030* (.0016)**	-.0024 (.0016)	-.0011 (.0011)	-.0016 (.0011)	-.0011 (.0010)
GGEGDPres	-.0102 (.0162)	-.0086 (.0168)	-.0085 (.0165)	-.0134 (.0144)	-.0137 (.0153)	-.0142 (.0153)	-.0080 (.0105)	-.0092 (.0116)	-.0096 (.0116)
HIV	.0447 (.0386)	.0356 (.0509)	.0392 (.0503)	.0403 (.0338)	.0247 (.0403)	.0257 (.0396)	.0406 (.0314)	.0257 (.0381)	.0271 (.0372)
constant	**.0336** (.0136)**	**.0354** (.0148)**	**.0301** (.0138)**	**.0377*** [.0114]**	**.0408*** (.0113)**	**.0363*** (.0111)**	**.0261*** (.0078)**	**.0297*** (.0084)**	**.0261*** (.0079)**
R (w, b, o)	0.0664 0.0073 0.0017	0.0457 0.0206 0.0115	0.0575 0.0045 0.0011	0.0752 0.0119 0.0057	0.0564 0.0320 0.0266	0.0662 0.0185 0.0125	0.0624 0.0001 0.0014	0.0463 0.0101 0.0046	-0.3588 0.0009 0.0000

The table reports fixed effects (with country-clustered robust standard errors) estimates for the model described in Section 2, using Lu et al. (2010, IMF) data for the period 1995-2004 for 108 countries (the overlapping sample with Van de Sijpe 2012) for all variables, including the health aid disaggregation. First three columns only use data where the dependent variable is fully observed for each country (50 countries, 450 yearly observations). The middle three columns report results using pooled sample where the dependent variable is observed for all the country-year observations used. Last three columns ignore the missing data problem altogether and report the estimates from the full sample, where the identified missing values are multiple-imputed and averaged as in Lu et al. (2010).
Potential aid smoothing effects are tested for by including lagged aid variables. Columns 1, 4, 7 report the estimates where contemporaneous values of aid variables are used. Columns 2, 5, 8 use the lagged value of aid. Columns 3, 6, 9 report results where both contemporaneous and lagged values of aid are included; if both were estimated to be significant (and positive), we could conclude that aid smoothing is taking place. We find evidence of some aid smoothing behaviour with respect to the health aid disbursed through the government (DAH-G), but only if the missing data problem is recognised. No such evidence is found if the missing data problem is effectively ignored.
Estimates are not directly comparable to Lu et al. as they use GHE-S/GDP as a dependent variable.

Appendix Table B2: Van de Sijpe's Distinction, Lu et al. Model; All VDS Aid Types Included (Testing for Aid Smoothing and Missing Data Effects)

Dependent variable: GHE-A/GDP (Lu et al.) [Van de Sijpe health aid disaggregation]									
FE, vce (R)	a	a	a	b	b	b	c	c	c
N	50	50	50	107	107	107	108	108	108
YO	500	450	450	896	808	808	1080	972	972
On-budget (VDS SP)	**.5899**** **(.2581)**		.4537 (.2779)	**.6620**** **(.2783)**		**.5569*** **(.2875)**	**.5032*** **(.2825)**		**.5981***** **(.2880)**
Lagged On-budget (VDS SP)		.4059 (.2795)	.3945 (.2739)		**.4334*** **(.2471)**	.3791 (.2370)		.2433 (.2911)	.1955 (.2967)
Off-budget (VDS TC)	.0873 (.2313)		-.0326 (.2387)	.0506 (.1270)		.0072 (.1393)	.0642 (.1128)		.0596 (.1023)
Lagged Off-budget (VDS TC)		.2478 (.2907)	.2145 (.2942)		.1148 (.1928)	.1431 (.1932)		.0257 (.1632)	.0634 (.1555)
Health IP	**.4019***** **(.1399)**		.4333*** (.1053)	**.3796***** **(.0629)**		**.3101***** **(.0680)**	**.3928***** **(.0525)**		**.3496***** **(.0486)**
Lagged Health IP		.2713 (.2073)	.2278 (.1861)		**.3609***** **(.0901)**	.1227 (.1264)		**(L).3406***** **(.0941)**	.0957 (.1059)
Health ONM	**.2836*** **(.1631)**		.0035 (.2071)	**.3157**** **(.1260)**		.0826 (.1383)	**.3016***** **(.1098)**		.1801 (.1098)
Lagged Health ONM		**.3985***** **(.1401)**	**.4078**** **(.1720)**		**.4277**** **(.1641)**	**.3779**** **(.1604)**		**(L) .3149**** **(.1194)**	**.2523**** **(.1127)**
General AID	.0197 (.0254)		-.0056 (.0239)	.0088 (.0214)		.0106 (.0228)	.0054 (.0204)		.0141 (.0184)
Lagged General AID		.0333 (.0246)	.0213 (.0239)		.0133 (.0189)	-.0009 (.0176)		(L) .0099 (.0171)	.0035 (.0154)
Support to NGOs	-.2631 (.1775)		-.1409 (.1397)	-.1791 (.1627)		-.1344 (.1341)	-.1308 (.1280)		-.0449 (.13654)
Lagged Support to NGOs		.0727 (.2531)	.1272 (.2814)		.1168 (.2040)	.1176 (.2155)		(L) .0589 (.1607)	.0173 (.1577)
Other non-health	**-.0204*** **(.0114)**		.0074 (.0151)	-.0116 (.0102)		.0111 (.0104)	**-.0114*** **(.0066)**	**(L)-.0150**** **(.0070)**	-.0014 (.0057)
Lagged Other non-health		**-.0269**** **(.0122)**	**-.0333***** **(.0112)**		**-.0189*** **(.0112)**	**-.0243**** **(.0110)**			**-.0158**** **(.0071)**
DR	-.0549 (.0642)	-.0017 (.0609)	.0189 (.0558)	-.0645 (.0532)	-.0037 (.0506)	.0168 (.0506)	-.0094 (.0493)	.0249 (.0320)	.0410 (.0304)
Ln(GDPpc)	-.0033 (.0020)	-.0023 (.0022)	-.0019 (.0023)	**-.0032*** **(.0017)**	**-.0031*** **(.0017)**	-.0024 (.0018)	-.0013 (.0012)	-.0016 (.0012)	-.0010 (.0012)
GGEGDPres	-.0089 (.0164)	-.0070 (.0166)	-.0061 (.0159)	-.0114 (.0142)	-.0118 (.0151)	-.0106 (.0144)	-.0086 (.0109)	-.0087 (.0119)	-.0083 (.0115)
HIV	.0364 (.0352)	.0249 (.0457)	.0263 (.0461)	.0364 (.0319)	.0185 (.0378)	.0205 (.0369)	.0360 (.0304)	.0184 (.0365)	.0199 (.0347)
constant	**.0434***** **(.0145)**	**.0364** **(.0154)**	**.0330*** **(.0164)**	**.0426***** **(.0122)**	**.0417***** **(.0121**	**.0360***** **(.0129)**	**.0284***** **(.0090)**	**.0310***** **(.0092)**	**.0263***** **(.0091)**
R (w, b, o)	0.0678 0.0322 0.0147	0.0535 0.0487 0.0257	0.0808 0.0132 0.0028	0.0755 0.0224 0.0133	0.0593 0.0537 0.0420	0.0881 0.0190 0.0102	0.0603 0.0009 0.0001	0.0478 0.0307 0.0154	0.0796 0.0029 0.0000

The table reports fixed effects (with country-clustered robust standard errors) estimates for the model described in Section 2, using Lu et al. (2010, IMF) data for the period 1995-2004 for 108 countries (the overlapping sample with Van de Sijpe 2012) for all variables, except the health aid disaggregation for

which data by Van de Sijpe (2013) is used. First three columns only use data where the dependent variable is fully observed for each country (50 countries, 450 yearly observations). The middle three columns report results using pooled sample where the dependent variable is observed for all the country-year observations used. Last three columns ignore the missing data problem altogether and report the estimates from the full sample, where the identified missing values are multiple- imputed and averaged as in Lu et al. (2010).
Potential aid smoothing effects are tested for by including lagged aid variables. Columns 1, 4, 7 report the estimates where contemporaneous values of aid variables are used. Columns 2, 5, 8 use the lagged value of aid. Columns 3, 6, 9 report results where both contemporaneous and lagged values of aid are included; if both were estimated to be significant (and positive), we could conclude that aid smoothing is taking place. Using Van de Sijpe's health aid disaggregation, we find no evidence of aid smoothing behaviour in key explanatory variables (on-budget health aid, SP, and off-budget aid, TC).

Appendix Table B3: Van de Sijpe's Distinction (Lu et al. Model); only Health Aid Included, other Aid Excluded (Testing for Missing Data Effects)

	Dependent variable: GHE-A/GDP (Lu et al.) [Van de Sijpe's disaggregation]						
		Health aid only			Only on-/off-budget health aid		
	FE, vce (R)	a	b	c	a	b	c
		Fully obs.	pooled	all	Fully obs.	pooled	all
	N	50	107	108	50	107	108
	YO	500	896	1080	500	896	1080
Health aid	**On-budget** SP	.3329 (.3563)	.5167 (.3256)	.3074 (.2970)	.2781 (.3861)	.5186 (.3612)	.3182 (.3225)
	Off-budget TC	-.0959 (.1713)	-.0352 (.1161)	-.0541 (.1054)	-.1473 (.1519)	-.0200 (.1450)	-.0307 (.1399)
	Health IP (unclassified)	**.2712*** (.1490)**	**.3296***** **(.0549)**	**.3405***** **(.0535)**			
	Health ONM (unclassified)	.1641 (.1602)	**.2352*** **(.1240)**	**.2368**** **(.1103)**			
	DR	-.0466 (.0518)	-.0574 (.0494)	-.0089 (.0466)	-.0540 (.0557)	-.0592 (.0535)	-.0089 (.0509)
	Ln(GDPpc)	-.0029 (.0020)	**-.0031*** **(.0017)**	-.0012 (.0012)	-.0034 (.0021)	**-.0041**** **(.0017)**	-.0020 (.0012)
	GGEGDPres	-.0102 (.0167)	-.0116 (.0142)	-.0093 (.0110)	-.0092 (.0166)	-.0114 (.0145)	-.0082 (.0111)
	HIV	.0456 (.0394)	.0422 (.0344)	.0406 (.0323)	.0420 (.0400)	.0400 (.0355)	.0392 (.0334)
	constant	**.0394***** **(.0141)**	**.0407***** **(.0119)**	**.0266***** **(.0087)**	**.0438***** **(.0147)**	**.0484***** **(.0122)**	**.0330***** **(.0089)**
	R (w, b, o)	0.0465 0.0117 0.0053	0.0691 0.0124 0.0071	0.0518 0.0002 0.0016	0.0350 0.0309 0.0191	0.0411 0.0354 0.0313	0.0202 0.0114 0.0074

Table reports estimates from the models and samples as described above (Van de Sijpe's disaggregation of health aid), omitting aid variables other than health aid (Columns 1,2,3), and then also omitting unclassified health aid variables (Columns 4,5,6).

Appendix Table B4: Correlations across Aid Variables (Van de Sijpe and Lu et al.)

	HealthSP	HealthTC	HealthIP	HealthON	GenAID	s.t.NGOs	otherAOD	DAH-Gov	DAH-nGov
Health SP	1.0000								
Health TC	0.4690	1.0000							
Health IP	0.2780	0.3657	1.0000						
Health ONM	0.2701	0.2740	0.2065	1.0000					
General Aid	0.4513	0.5327	0.2794	0.4304	1.0000				
Supp.to NGOs	0.5649	0.5958	0.3169	0.4706	0.5482	1.0000			
Othernohealth	0.5909	0.7313	0.4631	0.5116	0.6502	0.6726	1.0000		
DAH-Gov	0.3606	0.4498	0.6267	0.5975	0.4357	0.4569	0.5035	1.0000	
DAH-nonGov	0.2189	0.1838	0.1629	0.5599	0.2958	0.3159	0.2278	0.4541	1.0000

Appendix Table B5: Comparing Health Aid Disaggregation Strategies of Lu et al. and Van de Sijpe (Country and Time Fixed Effects)

Dependent variable: GHE-A/GDP (Lu et al.)				
FE, vce I	I Lu et al. disaggr.	II VDS disaggr.	III VDS disaggr.	IV VDS disaggr.
On-budget:	0.2773***	0.6194**	0.4472	0.4489
DAHG/Health SP	(0.0758)	(0.2583)	(0.2902)	(0.3315)
Off-budget:	0.0555	0.0384	-0.0588	0.0373
DAHnG/Health TC	(0.1749)	(0.1182)	(0.1064)	(0.1378)
Health IP (unclassified)		0.4050***	0.3718***	
		(0.0569)	(0.0552)	
Health ONM (unclassified)		0.1401	0.0922	
		(0.1189)	(0.1163)	
General aid		-0.0011		
		(0.0197)		
Support to NGOs		-0.2736**		
		(0.1206)		
Other non-health aid		-0.0054		
		(0.0064)		
Debt Relief	-0.0424	-0.0375	-0.0347	-0.0394
	(0.0499)	(0.0537)	(0.0491)	(0.0510)
Ln(GDPpc)	-0.0009	-0.0008	-0.0006	-0.0015
	(0.0012)	(0.0013)	(0.0012)	(0.0013)
GGEres	-0.0094	-0.0100	-0.0106	-0.0100
	(0.0096)	(0.0093)	(0.0094)	(0.0096)
HIV	0.0225	0.0149	0.0183	0.0171
	(0.0307)	(0.0302)	(0.0312)	(0.0320)
Country fixed effects	Y	Y	Y	Y
Time Fixed effects	Y	Y	Y	Y
N	1080	1080	1080	1080
R2	0.0901	0.1049	0.1000	0.0761

Table reports fixed effects (country-clustered robust standard errors) estimation results using full sample (108 countries, 1995-2004), and contemporaneous values of health aid (and other variables). Standard errors reported in the parentheses. Time dummies are included.

Appendix Table B6: Comparing Health Aid Disaggregation Strategies of Lu et al. and Van de Sijpe (First-Differenced Data, One-Year Differences)

Dependent variable: GHE-A/GDP (Lu et al.) [one-year differenced]				
	I	II	III	IV
	Lu et al. disaggr.	VDS disaggr.	VDS disaggr.	VDS disaggr.
D1_DAH-G	0.1093			
	(0.0667)			
D1_DAH-nG	-0.0397			
	(0.1455)			
D1_healthSP		0.1547	0.1267	0.1151
		(0.1236)	(0.1062)	(0.1047)
D1_healthTC		0.1155**	0.1268**	0.1168**
		(0.0503)	(0.0513)	(0.0470)
D1_healthIP		0.0903	0.0907	
		(0.0721)	(0.0677)	
D1_healthONM		0.0425	0.0571	
		(0.0701)	(0.0695)	
D1_generalaid		0.0044		
		(0.0113)		
D1_supportto NGOs		-0.0943		
		(0.0612)		
D1_othersectoraid		0.0018		
D1_drdisgdp	-0.0049	-0.0038	0.0024	-0.0013
	(0.0310)	(0.0394)	(0.0354)	(0.0329)
D1_l_gdppc	-0.0020*	-0.0019	-0.0020	-0.0021*
	(0.0011)	(0.0012)	(0.0012)	(0.0012)
D1_ggegdpRES	-0.0127	-0.0128	-0.0130	-0.0126
	(0.0086)	(0.0082)	(0.0083)	(0.0084)
D1_hiv	0.0348	0.0344	0.0347	0.0334
	(0.0370)	(0.0378)	(0.0375)	(0.0374)
		(0.0022)		
N	972	972	972	972
r2_a	0.0108	0.0117	0.0129	0.0124

Appendix Table B7: Comparing Health Aid Disaggregation Strategies of Lu et al. and Van de Sijpe (First-Differenced Data, Two-Year Differences)

Dependent variable: GHE-A/GDP (Lu et al.) [two year-differenced]				
	I	II	III	IV
	Lu et al. disaggr.	VDS disaggr.	VDS disaggr.	VDS disaggr.
D2_DAH-G	0.1693**			
	(0.0673)			
D2_DAH-nG	-0.0973			
	(0.1620)			
D2_healthSP		0.1888	0.1348	0.1319
		(0.3053)	(0.2796)	(0.2910)
D2_healthTC		0.0674	0.0973	0.1032
		(0.1293)	(0.1121)	(0.1086)
D2_healthIP		0.1513*	0.1427	
		(0.0894)	(0.0864)	
D2_healthONM		0.0363	0.0528	
		(0.1200)	(0.1134)	
D2_generalaid		0.0038		
		(0.0193)		
D2_supporttoNOGs		-0.1896		
		(0.1521)		
D2_othersectoraid		0.0028		
		(0.0042)		
D2_drdisgdp	0.0090	0.0065	0.0149	0.0109
	(0.0356)	(0.0441)	(0.0371)	(0.0372)
D2_l_gdppc	-0.0021*	-0.0019	-0.0019*	-0.0021*
	(0.0011)	(0.0012)	(0.0011)	(0.0011)
D2_ggegdpRES	-0.0148	-0.0139	-0.0146	-0.0144
	(0.0091)	(0.0088)	(0.0091)	(0.0091)
D2_hiv	0.0223	0.0208	0.0216	0.0200
	(0.0314)	(0.0325)	(0.0318)	(0.0319)
N	864.0000	864.0000	864.0000	864.0000
r2_a	0.0165	0.0146	0.0144	0.0120

Appendix Table B8: Comparing Health Aid Disaggregation Strategies of Lu et al. and Van de Sijpe (First-Differenced Data, Three-Year Differences)

Dependent variable: GHE-A/GDP (Lu et al.) [three year-differenced]				
	I	II	III	IV
	Lu et al. disaggr.	VDS disaggr.	VDS disaggr.	VDS disaggr.
D3_DAH-G	0.1718**			
	(0.0731)			
D3_DAH-nG	0.0313			
	(0.1496)			
D3_healthSP		0.3461	0.1965	0.2086
		(0.3759)	(0.3557)	(0.3693)
D3_healthTC		0.0562	-0.0164	0.0223
		(0.1303)	(0.1184)	(0.1197)
D3_healthIP		0.2645***	0.2366***	
		(0.0775)	(0.0742)	
D3_healthONM		0.1072	0.0565	
		(0.1049)	(0.1070)	
D3_generalaid		-0.0006		
		(0.0187)		
D3_supporttoNGOs		-0.0623		
		(0.1685)		
D3_othersectoraid		-0.0070		
		(0.0075)		
D3_drdisgdp	-0.0258	-0.0200	-0.0220	-0.0248
	(0.0516)	(0.0526)	(0.0504)	(0.0523)
D3_l_gdppc	-0.0014	-0.0014	-0.0013	-0.0017
	(0.0012)	(0.0012)	(0.0012)	(0.0012)
D3_ggegdpRES	-0.0125	-0.0132	-0.0132	-0.0131
	(0.0096)	(0.0097)	(0.0096)	(0.0097)
D3_hiv	0.0164	0.0117	0.0140	0.0131
	(0.0271)	(0.0272)	(0.0277)	(0.0280)
N	756.0000	756.0000	756.0000	756.0000
r2_a	0.0128	0.0137	0.0148	0.0059

Appendix Table B9: Sensitivity Check on GGE Alteration and Inclusion of Second Lag of Aid (Lu et al. Data only)

Dependent variable: GHE-A/GDP						
GHE-A/GDP (IMF)	FE, Lu (DAHG, DAHNG)	FE, Lu lagged	Fe, Lu L2 lagged	FE, Lu lagged and no lagged		
DAH-G	**.361*** (.076)**			**.264*** (.085)**		**.281*** (.081)**
DAH-G, Lagged		**.311*** (.099)**		.141 (.105)	**.280*** (.095)**	.112 (.089)
DAH-G, L2			.178 (.132)		.006 (.099)	-.045 (.109)
DAH-nG	.097 (.168)			-.016 (.1755)		-.052 (.177)
DAH-nG, Lagged		**.389* (.224)**		.209 (.173)	.225 (.188)	.075 (.156)
DAH-nG, L2			.474 (.309)		.218 (.228)	.188 (.200)
DR	-.038 (.043)	.004 (.032)	.028 (.018)	.000 (.032)	.013 (.022)	.010 (.022)
Log(GDPpc)	-.001 (.001)	-.001 (.001)	-.002 (.001)	-.001 (.001)	-.001 (.001)	-.001 (.001)
GGE/GDP	**.026*** (0.001)**	**.025*** (.009)**	**.025*** (.009)**	**.025*** (.008)**	**.025*** (.008)**	**.024*** (.008)**
HIV	.048 (.031)	.036 (.037)	.046 (.035)	.037 (.036)	.049 (.034)	.050 (.033)
constant	**.018** (.008)**	**.021** (.009)**	**.025*** (.009)**	**.018** (.008)**	**.022*** (.008)**	**.019** (.008)**
R2 (w, b, o)	W=0.0794 B=0.0671 O=0.0681	0.0620; 0.0371; 0.0392	0.0441; 0.0232; 0.0240	0.0750; 0.0659; 0.0668	0.0629; 0.0364; 0.0378	0.0802; 0.0660; 0.0668
YO; N=108	1080	972	864	972	864	864

Table demonstrates that is sensitivity of the results if GHE-A is not removed from GGE.

Appendix Table B10: Sensitivity of Lu et al. Findings to Minor Sample Change and Change in Estimator

Dependent variable: GHE-S/GDP

Dynamic panel-data estimation, one-step system GMM (xtabond2)

```
Group variable: countryid                         Number of obs      =       972
Time variable : year                              Number of groups   =       108
Number of instruments = 104                       Obs per group: min =         9
Wald chi2(6)  =     43.45                                        avg =      9.00
Prob > chi2   =     0.000                                        max =         9
             |              Robust
ghegdp_imf_s |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ghegdp_imf_s |
         L1. |  -.1014749   .0733773    -1.38   0.167    -.2452918    .0423421
             |
   DAH-G/GDP |  -.6492609   .1195594    -5.43   0.000    -.8835931   -.4149287
  DAH-nG/GDP |  -.0552009   .1827446    -0.30   0.763    -.4133737    .3029718
   drdisgdp |    -.06056    .057819    -1.05   0.295    -.1738831    .0527632
       gdppc |   3.33e-07   7.95e-07     0.42   0.676    -1.23e-06    1.89e-06
      ggegdp |   .0160934   .0132397     1.22   0.224     -.009856    .0420427
         hiv |   .0464201   .0467602     0.99   0.321    -.0452281    .1380684
       _cons |   .0163449   .0032589     5.02   0.000     .0099575    .0227323
```

Fixed effects

```
Fixed-effects (within) regression                 Number of obs      =      1080
Group variable: countryid                         Number of groups   =       108

R-sq:  within  = 0.1094                           Obs per group: min =        10
       between = 0.1983                                          avg =      10.0
       overall = 0.1819                                          max =        10

                                                  F(6,107)           =     10.34
corr(u_i, Xb)  = 0.1506                           Prob > F           =    0.0000

                            (Std. Err. adjusted for 108 clusters in countryid)
------------------------------------------------------------------------------
             |              Robust
ghegdp_imf_s |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   DAH-G/GDP |  -.5201887   .0717725    -7.25   0.000    -.6624693   -.3779082
  DAH-nG/GDP |    .145877   .1577036     0.93   0.357    -.1667521    .4585061
    drdisgdp |  -.0233904   .0357358    -0.65   0.514    -.0942325    .0474516
       gdppc |   1.09e-07   2.43e-07     0.45   0.655    -3.73e-07    5.91e-07
      ggegdp |   .0278818   .0075316     3.70   0.000     .0129512    .0428124
         hiv |    .037628    .027434     1.37   0.173    -.0167567    .0920128
       _cons |   .0129522   .0012911    10.03   0.000     .0103927    .0155117
-------------+----------------------------------------------------------------
     sigma_u |  .00987887
     sigma_e |  .00446321
         rho |  .83048387   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

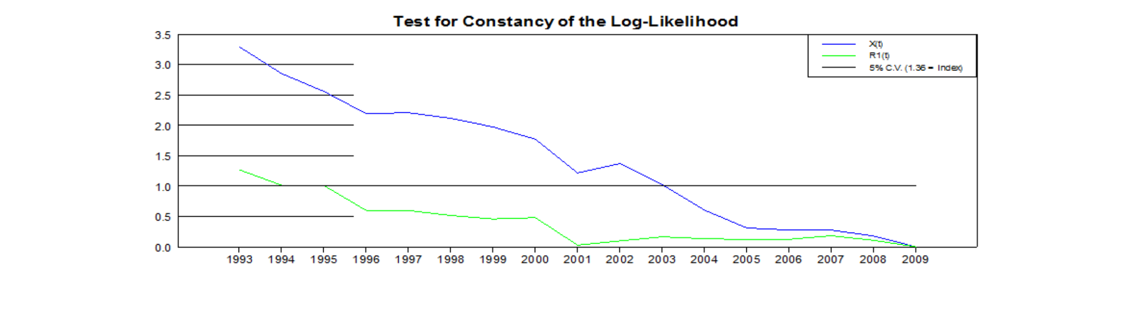
Sensitivity of Lu et al. finding to minor sample change (to 1995-2004, losing 2 years) and three countries removed. The upper section reports results from the original estimator, xtabond2. The lower section reports country-fixed effects estimates. Lu et al. (2010) findings are sensitive to this sample reduction; their core IMF results would be altered such that DAH-nG has no significant effect on GHE-S, and GGE would no longer be significant). Contrary to the original findings, GHE-A does not grow with the rest of GGE.

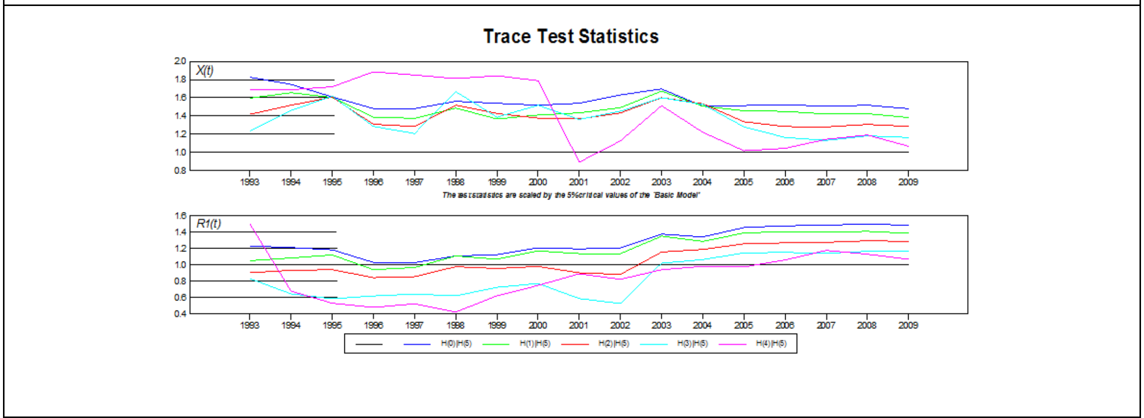
Appendix C

Appendix Table C1: Residual Tests from Unrestricted VAR (Disaggregated Aid Model)

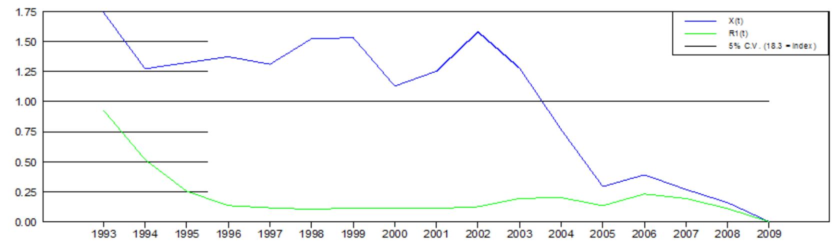
```
Tests for Autocorrelation
LM(1):                ChiSqr(25)  =  30.981 [0.190]
LM(2):                ChiSqr(25)  =  29.974 [0.225]
LM(3):                ChiSqr(25)  =  32.086 [0.156]
LM(4):                ChiSqr(25)  =  17.479 [0.864]

Test for ARCH:
LM(1):                ChiSqr(225) = 218.679 [0.606]
LM(2):                ChiSqr(450) = 483.740 [0.131]
LM(3):                ChiSqr(675) = 689.582 [0.340]
LM(4):                ChiSqr(900) = 675.000 [1.000]

Univariate Statistics

          Mean     Std.Dev Skewness  Kurtosis Maximum   Minimum
DL_TEXP    -0.000  0.069      0.323    2.491      0.161  -0.140
DL_TAX      0.000  0.086     -0.212    3.874      0.191  -0.264
DL_NTAX    -0.000  0.185     -0.125    2.438      0.354  -0.403
DL_GRANTS   0.000  0.250     -0.425    3.482      0.510  -0.702
DL_LOANS    0.000  0.339      0.193    2.437      0.795  -0.657

          ARCH(2)         Normality          R-Squared
DL_TEXP     0.078 [0.962]     1.402  [0.496]      0.677
DL_TAX      1.408 [0.495]     4.547  [0.103]      0.571
DL_NTAX     3.078 [0.215]     0.310  [0.856]      0.509
DL_GRANTS   0.114 [0.945]     2.566  [0.277]      0.511
DL_LOANS    2.578 [0.276]     0.599  [0.741]      0.518
```

Appendix Table C2: Parameter Constancy Tests of Unrestricted VAR (k=2)

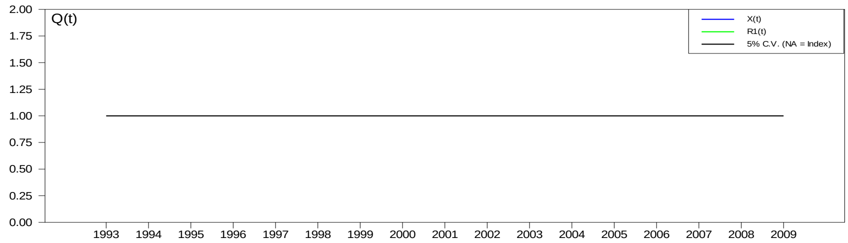




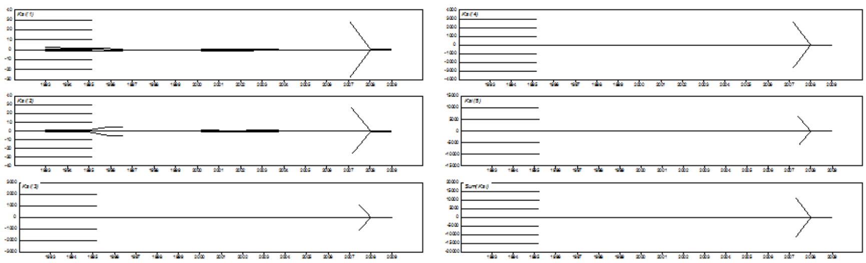
Test of Beta(t) = 'Known Beta'



Test of Beta Constancy

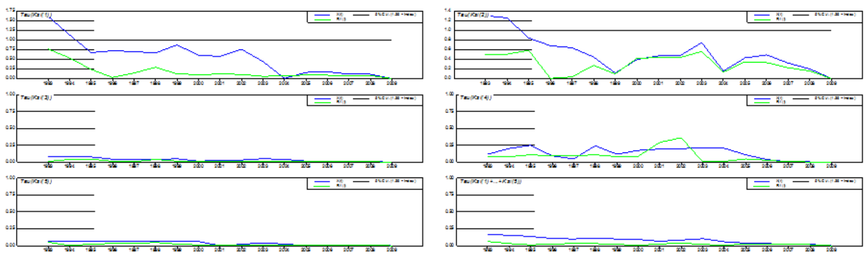


Transfomed Eigenvalues



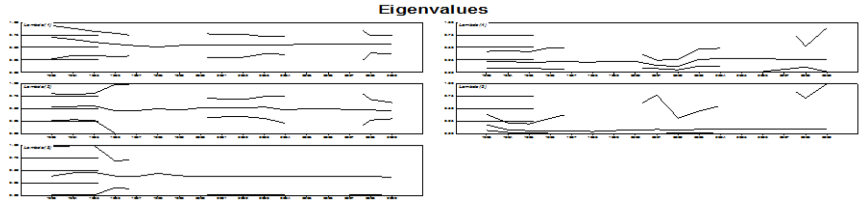
Ksi = log(Lambda/(1-Lambda)), Sum (Ksi) = Ksi(1) +...+ Ksi(r)

Eigenvalue Fluctuation Test



Tau(Ksi) = C(T)||Ksi(t)-Ksi(T)||

Eigenvalues



Appendix Table C3: Simulation of the Asymptotic Trace Test Distribution

```
Deterministic specification  : Unrestricted Constant (DRIFT)
Level Shifts (2)             : 1991:01 (0.600) 1974:01 (0.222)
Number of Replications (N)   : 2500
Length of Random Walks (T)   : 400

Quantiles of the Simulated Rank Test Distribution
 p-r r  Mean   S.E.    50%    75%    80%     85    90%    95%
  5  0 58.617 10.052 58.135 65.396 67.284 69.383 71.789 75.445
  4  1 39.213  8.459 38.395 44.636 46.095 48.066 50.244 54.154
  3  2 23.243  6.861 22.499 27.299 28.671 30.263 32.258 35.865
  2  3 10.956  4.491 10.305 13.485 14.408 15.503 16.910 19.076
  1  4  1.963  1.939  1.364  2.722  3.221  3.782  4.611  5.855

I(1)-ANALYSIS
 p-r r Eig.Value  Trace  Trace* Frac95 P-Value P-Value*
  5  0     0.561 102.992 89.825 75.445   0.000    0.003
  4  1     0.459  65.973 57.488 54.154   0.003    0.025
  3  2     0.364  38.306 34.598 35.865   0.026    0.062
  2  3     0.265  17.927 15.413 19.076   0.075    0.155
  1  4     0.087   4.095  3.549  5.855   0.123    0.163
```

Appendix Table C4: Additional Information for Rank Determination

UVAR estimates (Alpha coefficients)

```
@cats(lags=2,det=drift,break=level) 1963:1 2009:1
# L_TEXP L_TAX L_NTAX L_GRANTS L_LOANS
# 1991:1 1974:1

CATS for RATS version 2 - 03/01/2013 16:16

MODEL SUMMARY
Sample:                     1963:01 to 2009:01 (47 observations)
Effective Sample:           1965:01 to 2009:01 (45 observations)
Obs. - No. of variables:    30
System variables:           L_TEXP L_TAX L_NTAX L_GRANTS L_LOANS
Shift-dummy series:         C(1991:01) C(1974:01)
Constant/Trend:             Unrestricted Constant
Lags in VAR:                2

I(2) analysis not available for the specified model.

The unrestricted estimates:
BETA(transposed)
         L_TEXP L_TAX  L_NTAX L_GRANTS L_LOANS C(1991:01) C(1974:01)
Beta(1) 14.465 -7.072 -0.404   -3.874  -1.650     -0.124     -4.601
Beta(2)  8.717 -8.519 -4.150    2.553   0.633     -1.438      1.858
Beta(3) -2.545  0.733  1.434    0.973   1.143     -2.914     -1.702
Beta(4) -0.769  4.787 -1.412   -2.337   1.691      0.801     -2.027
Beta(5)  2.256  1.985 -1.981   -1.542  -0.203     -1.098      0.711

ALPHA
        Alpha(1) Alpha(2) Alpha(3) Alpha(4) Alpha(5)
DL_TEX   -0.023   -0.044   -0.033   -0.000   -0.004
        (-2.267) (-4.303) (-3.197) (-0.014) (-0.356)
DL_TAX    0.018    0.010   -0.046    0.016   -0.016
         (1.427)  (0.764) (-3.564)  (1.276) (-1.235)
DL_NTA   -0.119    0.024   -0.077    0.015    0.033
        (-4.340)  (0.877) (-2.812)  (0.536)  (1.185)
DL_GRA    0.122   -0.090   -0.073    0.075    0.039
         (3.282) (-2.416) (-1.966)  (2.004)  (1.060)
DL_LOA    0.114   -0.123   -0.144   -0.131    0.017
         (2.257) (-2.430) (-2.851) (-2.597)  (0.327)
```

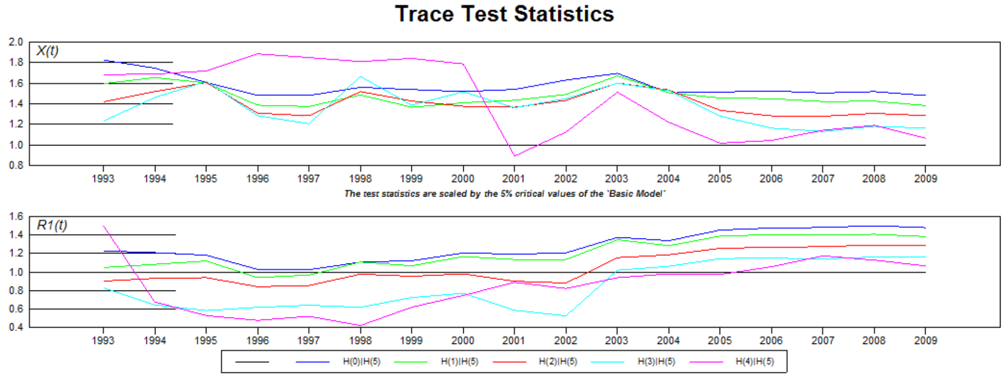
PI

	L_TEXP	L_TAX	L_NTAX	L_GRANTS	L_LOANS	C(1991:01)	C(1974:01)
DL_TEX	-0.650	0.512	0.154	-0.049	-0.027	0.167	0.079
	(-3.651)	(4.048)	(2.962)	(-0.859)	(-0.956)	(4.586)	(1.350)
DL_TAX	0.418	-0.199	-0.105	-0.104	-0.045	0.147	-0.033
	(1.890)	(-1.270)	(-1.631)	(-1.475)	(-1.307)	(3.260)	(-0.449)
DL_NTA	-1.258	0.718	-0.248	0.364	0.142	0.182	0.719
	(-2.651)	(2.129)	(-1.792)	(2.405)	(1.909)	(1.873)	(4.612)
DL_GRA	1.202	0.284	0.036	-1.010	-0.224	0.344	-0.729
	(1.871)	(0.623)	(0.190)	(-4.924)	(-2.223)	(2.621)	(-3.449)
DL_LOA	1.084	-0.461	0.410	-0.614	-0.656	0.459	-0.230
	(1.244)	(-0.745)	(1.610)	(-2.207)	(-4.794)	(2.576)	(-0.803)

Moduli of the roots of the companion matrix

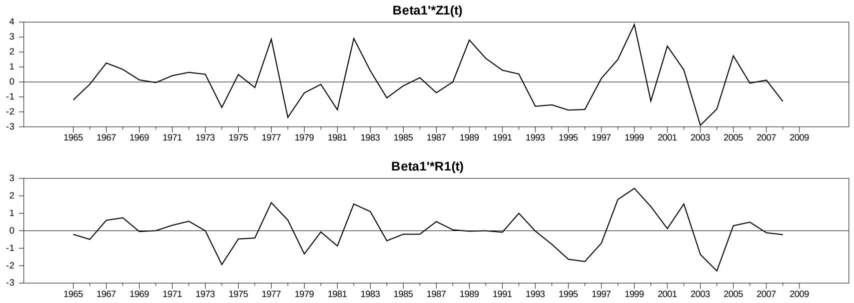
Rank	*Root 1*	*Root 2*	*Root 3*	*Root 4*	*Root 5*	*Root 6*	*Root 7*	*Root 8*	*Root 9*
Rank=5	0.971	0.633	0.633	0.605	0.605	0.593	0.593	0.300	0.300
Rank=4	1.000	0.672	0.672	0.591	0.591	0.587	0.587	0.289	0.289
Rank=3	1.000	1.000	0.690	0.562	0.562	0.525	0.525	0.395	0.376
Rank=2	1.000	1.000	1.000	0.584	0.584	0.554	0.554	0.401	0.401
Rank=1	1.000	1.000	1.000	1.000	0.505	0.505	0.323	0.307	0.307
Rank=0	1.000	1.000	1.000	1.000	1.000	0.454	0.454	0.289	0.289

Recursive Estimation Trade Test Statistic

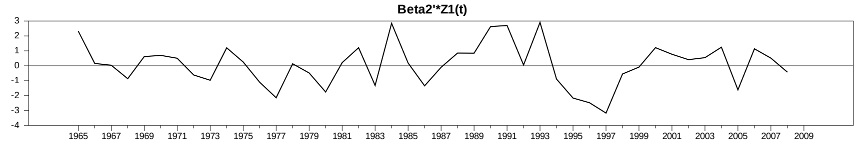


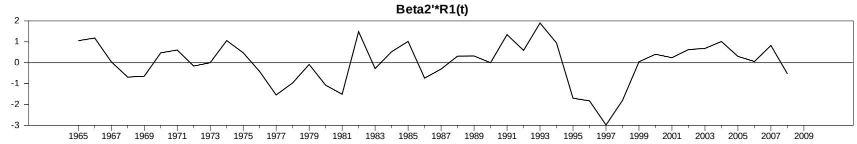
Graphs of Cointegrating Relations

Texp

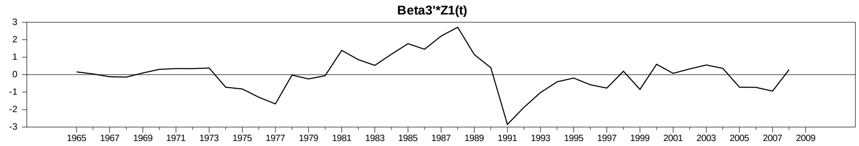


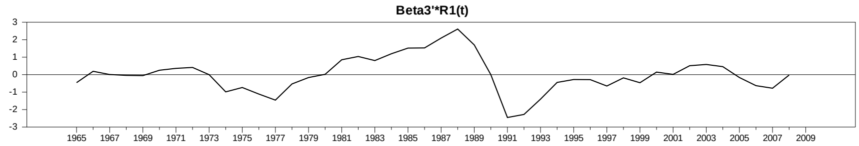
Tax



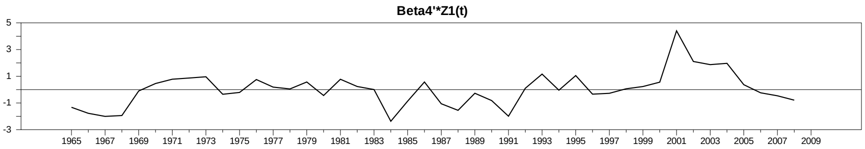


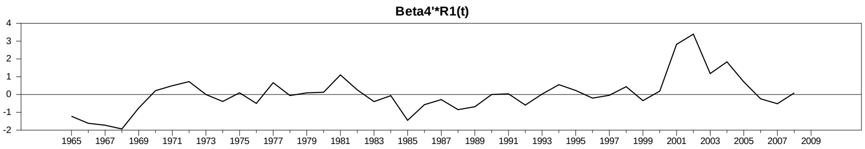
Non-tax



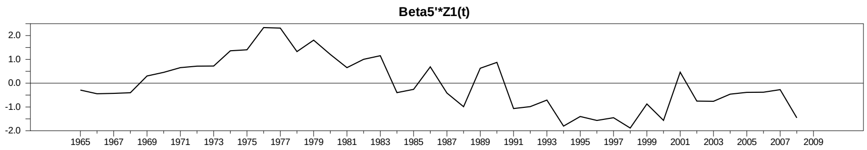


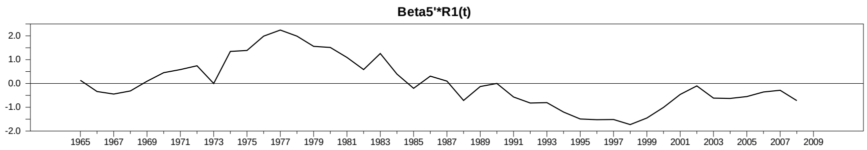
Grants





Loans





Appendix Table C5: Univariate Stationarity Tests

```
TEST OF STATIONARITY
LR-test, Chi-Square(5-r), P-values in brackets.

 r   DGF   5% C.V. L_TEXP    L_TAX  L_NTAX  L_GRANTS L_LOANS
 1    4     9.488  21.021   24.328  16.523   22.114  18.204
                   [0.000] [0.000] [0.002]  [0.000] [0.001]
 2    3     7.815  12.200   14.988   7.902   12.811   8.954
                   [0.007] [0.002] [0.048]  [0.005] [0.030]
 3    2     5.991   4.926    7.741   0.872    5.646   1.984
                   [0.085] [0.021] [0.647]  [0.059] [0.371]
 4    1     3.841   4.953    7.560   0.727    4.760   0.658
                   [0.026] [0.006] [0.394]  [0.029] [0.417]

Restricted Shift-Dummies included
in the cointegrating relation(s)

TEST OF STATIONARITY
LR-test, Chi-Square(7-r), P-values in brackets.

 r   DGF   5% C.V. L_TEXP    L_TAX  L_NTAX  L_GRANTS L_LOANS
 1    6    12.592  32.274   32.239  32.214   32.951  29.707
                   [0.000] [0.000] [0.000]  [0.000] [0.000]
 2    5    11.070  23.096   22.889  23.099   23.642  20.372
                   [0.000] [0.000] [0.000]  [0.000] [0.001]
 3    4     9.488  15.811   15.685  16.145   16.498  13.153
                   [0.003] [0.003] [0.003]  [0.002] [0.011]
 4    3     7.815   9.565    9.481  10.492   10.350   7.037
                   [0.023] [0.024] [0.015]  [0.016] [0.071]
```

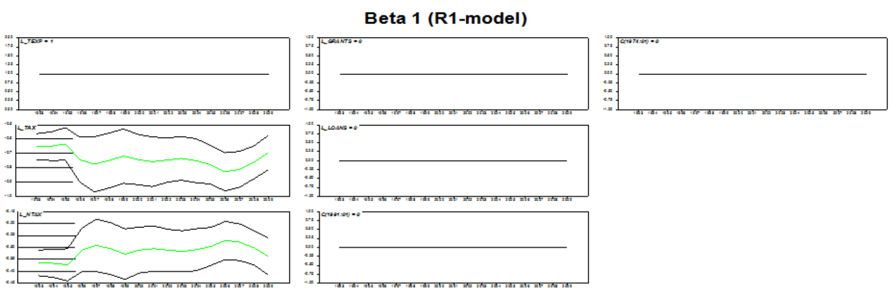
DF GLS

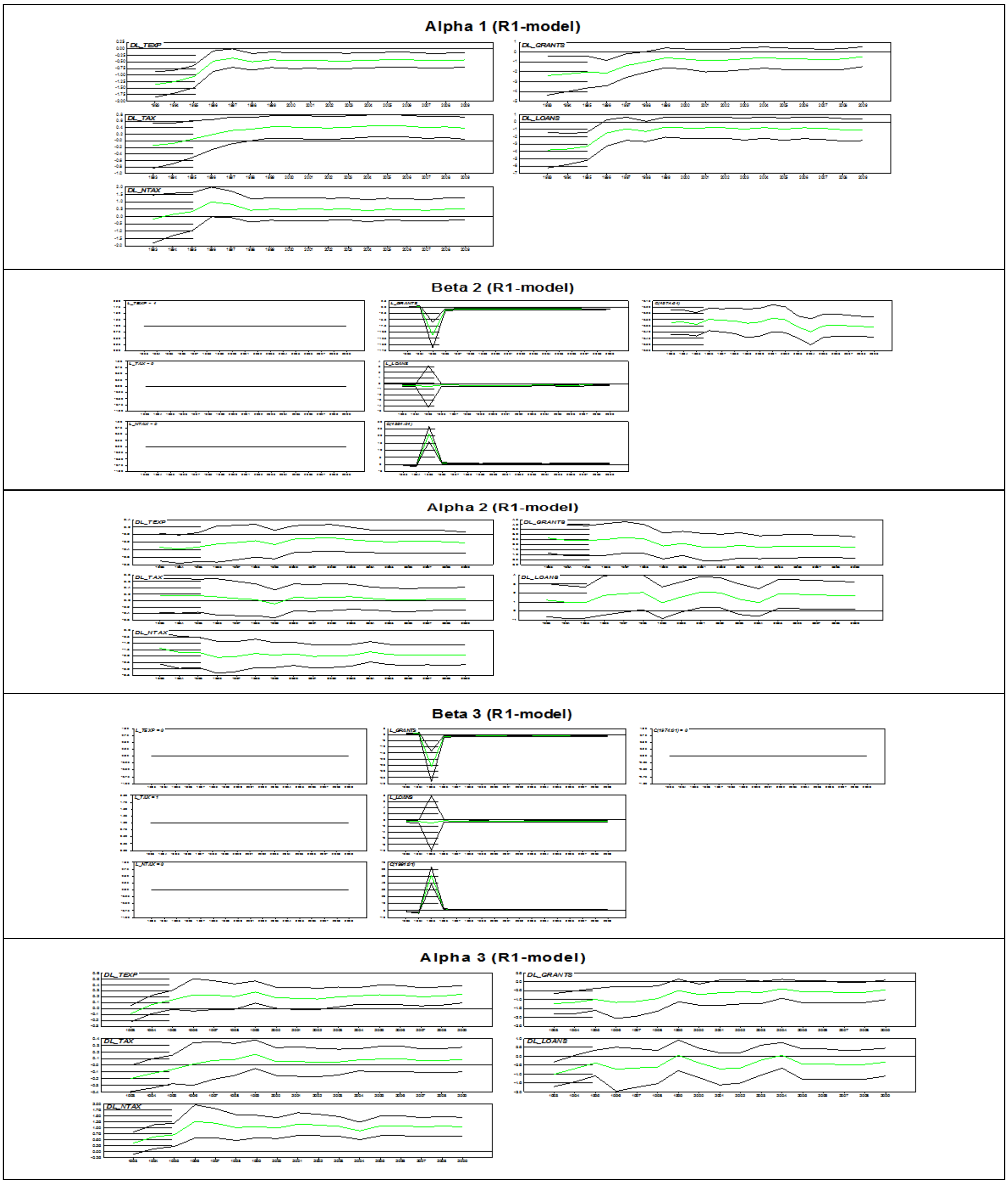
Results of Dickey-Fuller GLS test for trend-stationarity: logs

Variable	1 lag	2 lag	3 lag	4 lag	5 lag
Total expenditure	-2.05	-2.02	-2.16	-1.72	-2.26
Capital expenditure	-2.33	-2.20	-2.18	-2.33	-1.96
Recurrent expenditure	-2.79	-2.69	-2.91	-2.39	-3.24
Tax revenue	-1.02	-1.28	-1.13	-1.00	-0.89
Nontax revenue	-2.55	-2.65	-3.33	-2.87	-2.41
Loans	-3.91	-2.96	-3.65	-3.04	-2.56
Critical values (5%):	3.223	-3.176	-3.120	-3.059	-2.993

H0: non trend-stationarity

Appendix Table C6: Parameter Constancy Tests of Over-identified Model




Alpha 1 (R1-model)
DL_TEXP
DL_GRANTS
DL_TAX
DL_LOANS
DL_NTAX
Beta 2 (R1-model)
Alpha 2 (R1-model)
DL_TEXP
DL_GRANTS
DL_TAX
DL_LOANS
DL_NTAX
Beta 3 (R1-model)
Alpha 3 (R1-model)
DL_TEXP
DL_GRANTS
DL_TAX
DL_LOANS
DL_NTAX

Appendix Table C7: Alternative System: Short Run Results (Parsimonious Structure)

$$\Delta cexp_t = -0.31\,\Delta cexp_{t-1} + 0.04\,CI1_{t-1} - 0.27\,CI2_{t-1} - 0.60\,dum1991p + 0.21\,dum1974p$$

(-3.13) (4.02) (-3.68) (-3.62) (1.70)

$$\Delta rexp_t = 0.08\,\Delta \mathrm{loans}_{t-1} + 0.73\,\mathrm{CI1}_{t-1} + 0.25\,\mathrm{CI2}_{t-1} - 0.46\mathrm{CI3}_{t-1} + 0.13\,dum1974p$$

(3.90) (7.25) (5.38) (-7.11) (1.86)

$$\Delta domrev_t = 0.02\,CI1_{t-1} - 0.35\,dum1991p$$

(3.63) (-3.17)

$$\Delta grants_t = -0.60\,\Delta cexp_{t-1} + 0.78\,\Delta rexp_{t-1} + 0.77\,CI2_{t-1} + 0.02\,CI3_{t-1}$$

(-3.34) (2.75) (5.15) (1.94)

$$\Delta loans_t = -0.90\,\Delta cexp_{t-1} + 0.04\,CI3_{t-1} - 0.73\,dum1991p$$

(-3.31) (2.37) (-1.83)

```
LR test of over-identifying restrictions: Chi^2(31) =   22.956 [0.8507]
BFGS using analytical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
```

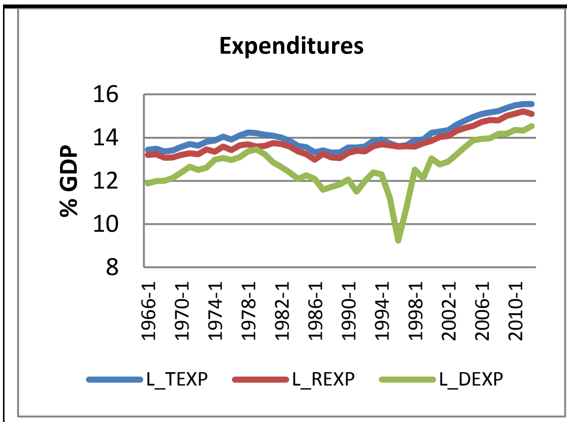
correlation of structural residuals (standard deviations on diagonal)

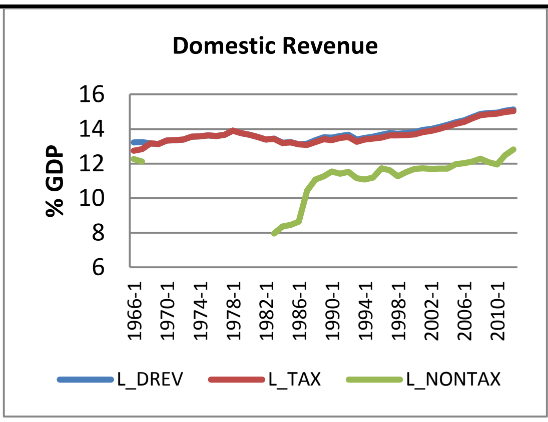
	D_L_CEXP	D_L_REXP	D_L_DOMREV	D_L_GRANTS	D_L_LOANS
D_L_CEXP	0.19032	0.24358	0.58971	0.50103	0.55183
D_L_REXP	0.24358	0.082380	0.39629	-0.057387	0.45842
D_L_DOMREV	0.58971	0.39629	0.12014	0.19295	0.27043
D_L_GRANTS	0.50103	-0.057387	0.19295	0.28888	0.31010
D_L_LOANS	0.55183	0.45842	0.27043	0.31010	0.45434

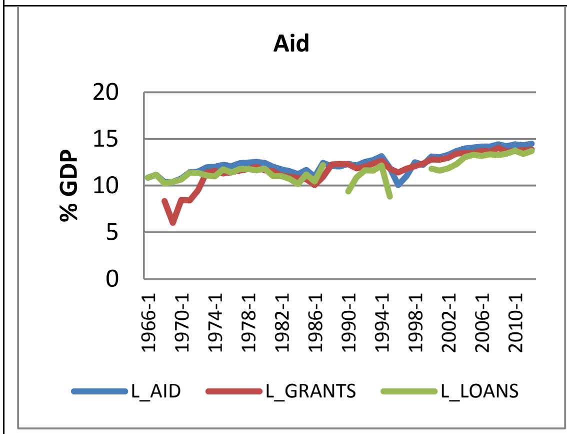
The generically (just-identified) short run structure is heavily over-parameterised. Here we report a parsimonious system, where the estimated coefficients with small (in absolute terms) t-statistics (p-value < 0.10) were set to zero (subject to passing a LR test). [30 restrictions]. Accepted with a p value 0.5. Since there are some non-negligible correlation coefficients in the residual covariance matrix, the interpretation of the short-run equations as causal relationships should be taken with caution.

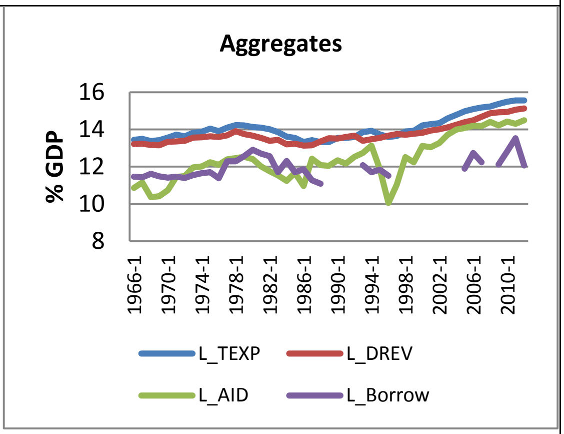
Appendix D

Appendix Table D1: Data Graphs (Tanzania)

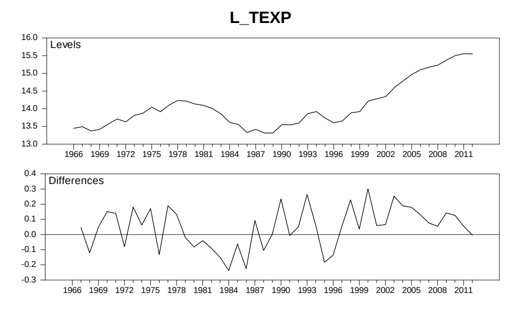


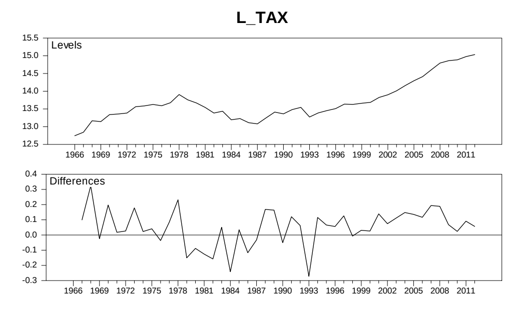


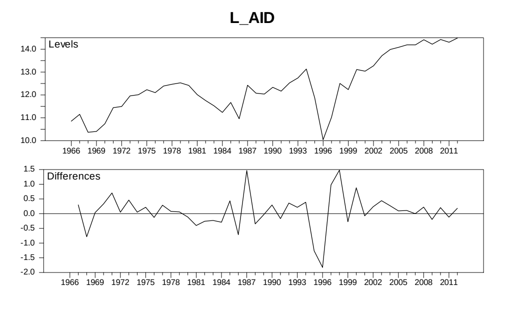


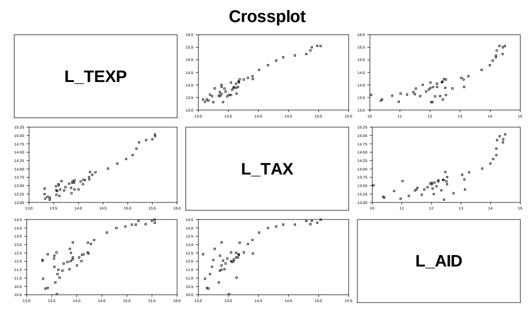


Variables in (logged) levels and their first differences









Appendix Table D2: Testing for Trend in Cointegrating Space

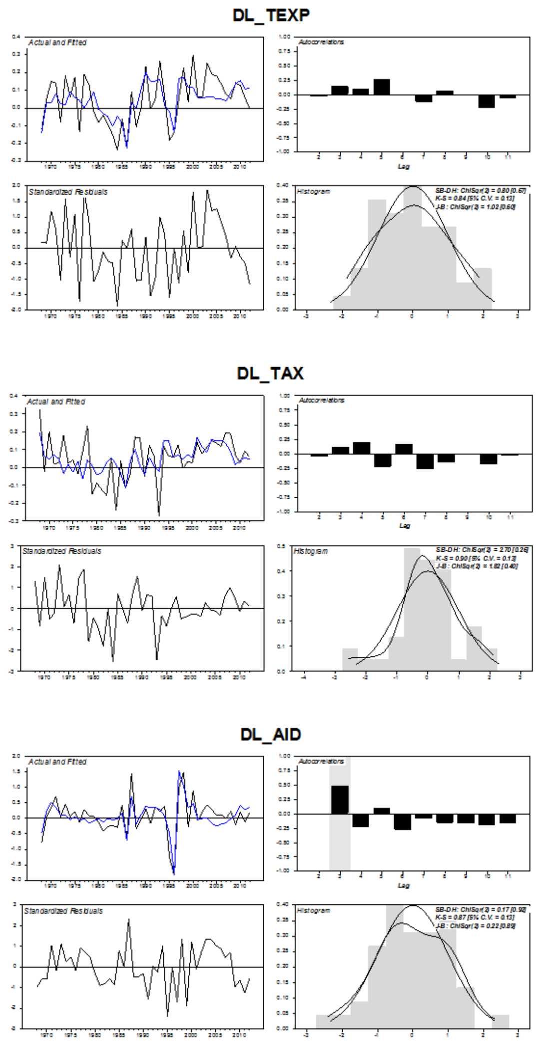
```
TEST OF EXCLUSION
 r  DGF   5% C.V. L_TEXP   L_TAX   L_AID  C(1986:01)  TREND
 1   1     3.841   0.633   3.630   4.714      0.169   0.135
                 [0.426] [0.057] [0.030]    [0.681] [0.714]
 2   2     5.991  18.939  22.134  22.305      6.600   2.981
                 [0.000] [0.000] [0.000]    [0.037] [0.225]
```

Appendix Table D3: Unrestricted VAR Estimates

```
@cats(lags=2,det=drift,break=level,dum) 1966:1 2012:1
# L_TEXP L_TAX L_AID
# 1986:1
# dum96p CATS for RATS version 2 - 07/18/2014 12:08

MODEL SUMMARY
Sample:                       1966:01 to 2012:01 (47 observations)
Effective Sample:             1968:01 to 2012:01 (45 observations)
Obs. - No. of variables:      35
System variables:             L_TEXP L_TAX L_AID
Shift-dummy series:           C(1986:01)
Dummy-series:                 DUM96P{0}
Constant/Trend:               Unrestricted Constant
Lags in VAR:                  2

I(2) analysis not available for the specified model.

The unrestricted estimates:
BETA(transposed)
        L_TEXP L_TAX  L_AID C(1986:01)
Beta(1)  1.512 -4.391 1.669     -1.297
Beta(2) -7.780  7.389 1.650     -1.749
Beta(3)  0.519 -1.745 0.501      1.988

ALPHA
       Alpha(1) Alpha(2) Alpha(3)
DL_TEX  -0.046    0.037    0.021
       (-3.071)  (2.442)  (1.420)
DL_TAX   0.011   -0.045    0.022
        (0.744) (-3.029)  (1.484)
DL_AID  -0.321    0.014    0.027
       (-6.708)  (0.292)  (0.574)

PI
        L_TEXP   L_TAX    L_AID   C(1986:01)
DL_TEX  -0.344    0.436   -0.006      0.038
       (-2.884)  (3.313) (-0.159)    (0.860)
DL_TAX   0.380   -0.421   -0.045      0.109
        (3.206) (-3.220) (-1.256)    (2.470)
DL_AID  -0.580    1.466   -0.499      0.447
       (-1.526)  (3.490) (-4.344)    (3.165)

Log-Likelihood = 273.719

RESIDUAL ANALYSIS

Residual S.E. and Cross-Correlations
         DL_TEXP     DL_TAX     DL_AID
        0.10067445 0.10002541 0.32126552
DL_TEXP   1.000
DL_TAX    0.275      1.000
DL_AID    0.675      0.111      1.000
```

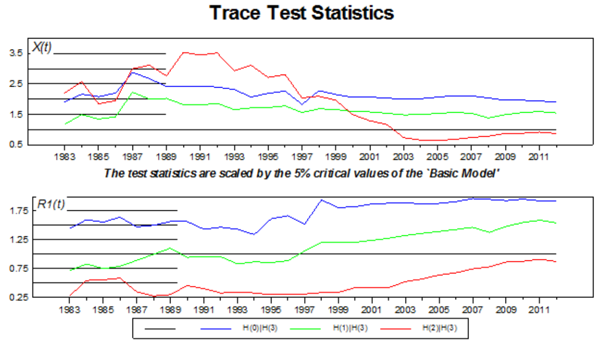
```
LOG(|Sigma|)                       = -12.165
Information Criteria: SC           =  -9.628
                      H-Q          = -10.383
Trace Correlation                  =   0.480
Tests for Autocorrelation
Ljung-Box(11):        ChiSqr(81)  = 146.090 [0.000]
LM(1):                ChiSqr(9)   =   8.583 [0.477]
LM(2):                ChiSqr(9)   =  17.212 [0.045]
LM(3):                ChiSqr(9)   =  14.499 [0.106]
LM(4):                ChiSqr(9)   =   9.147 [0.424]
Test for Normality:   ChiSqr(6)   =   3.335 [0.766]
Test for ARCH:
LM(1):                ChiSqr(36)  =  25.975 [0.891]
LM(2):                ChiSqr(72)  =  65.830 [0.682]
LM(3):                ChiSqr(108) = 111.034 [0.401]
LM(4):                ChiSqr(144) = 152.011 [0.308]
Univariate Statistics
        Mean    Std.Dev Skewness  Kurtosis Maximum   Minimum
DL_TEXP  -0.000  0.101      0.080   2.228      0.192  -0.192
DL_TAX   -0.000  0.100     -0.313   3.533      0.214  -0.259
DL_AID   -0.000  0.321     -0.130   2.677      0.753  -0.781
        ARCH(2)         Normality          R-Squared
DL_TEXP   0.763 [0.683]     0.798  [0.671]     0.440
DL_TAX    0.484 [0.785]     2.698  [0.260]     0.310
DL_AID    0.462 [0.794]     0.171  [0.918]     0.682
```

Appendix Figure D1: Residual Plots (UVAR)



Appendix Table D4: Additional Information for Determination of Cointegration Rank

Trace test statistics (recursive estimation)



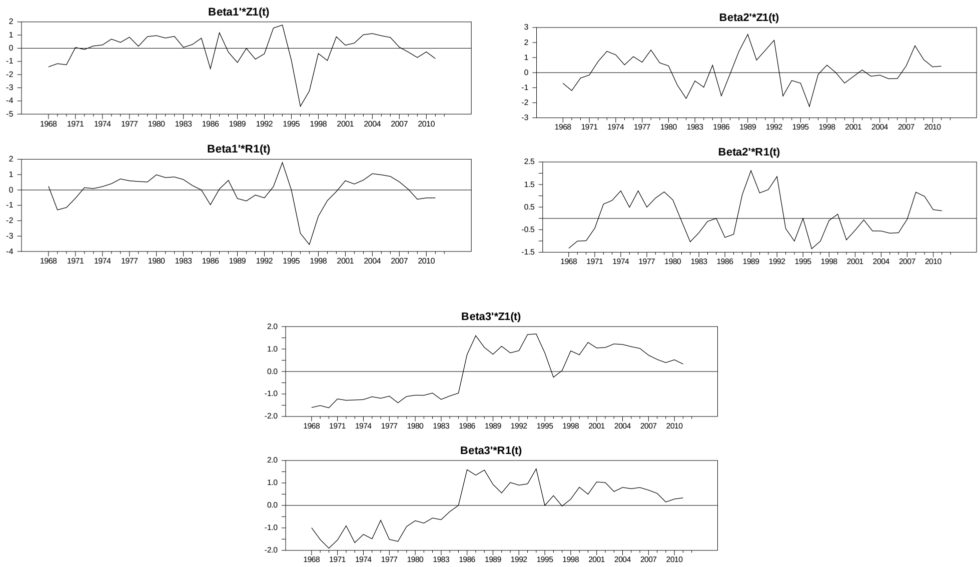
Moduli of the roots of the companion matrix

Rank	*Root 1*	*Root 2*	*Root 3*	*Root 4*	*Root 5*	*Root 6*
Rank=3	0.990	0.497	0.497	0.359	0.223	0.223
Rank=2	1.000	0.484	0.484	0.443	0.241	0.241
Rank=1	1.000	1.000	0.507	0.265	0.265	0.190
Rank=0	1.000	1.000	1.000	0.410	0.253	0.166

Alpha matrix coefficients (UVAR)

	L_TEXP	L_TAX	L_AID
Alfa matrix coefficients (α_1)	-0.046	0.011	-0.321
	(-3.071)	(0.744)	(-6.708)
Alfa matrix coefficients (α_2)	0.037	-0.045	0.014
	(2.442)	**(-3.029)**	(0.292)
Alfa matrix coefficients (α_3)	0.021	0.022	0.027
	(1.420)	(1.484)	(0.574)

Graphs of cointegrating relationships (UVAR)



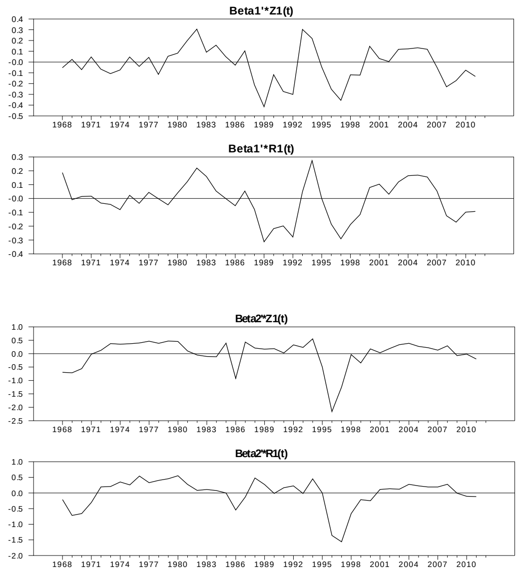
Pi matrix from UVAR with r=2

	L_TEXP	L_TAX	L_AID	C(1986:01)
DL_TEX	-0.355	0.473	-0.016	-0.004
	(-2.919)	(3.589)	(-0.456)	(-0.129)
DL_TAX	0.368	-0.383	-0.056	0.065
	(3.042)	(-2.914)	(-1.563)	(1.943)
DL_AID	-0.595	1.514	-0.513	0.392
	(-1.561)	(3.665)	(-4.548)	(3.748)

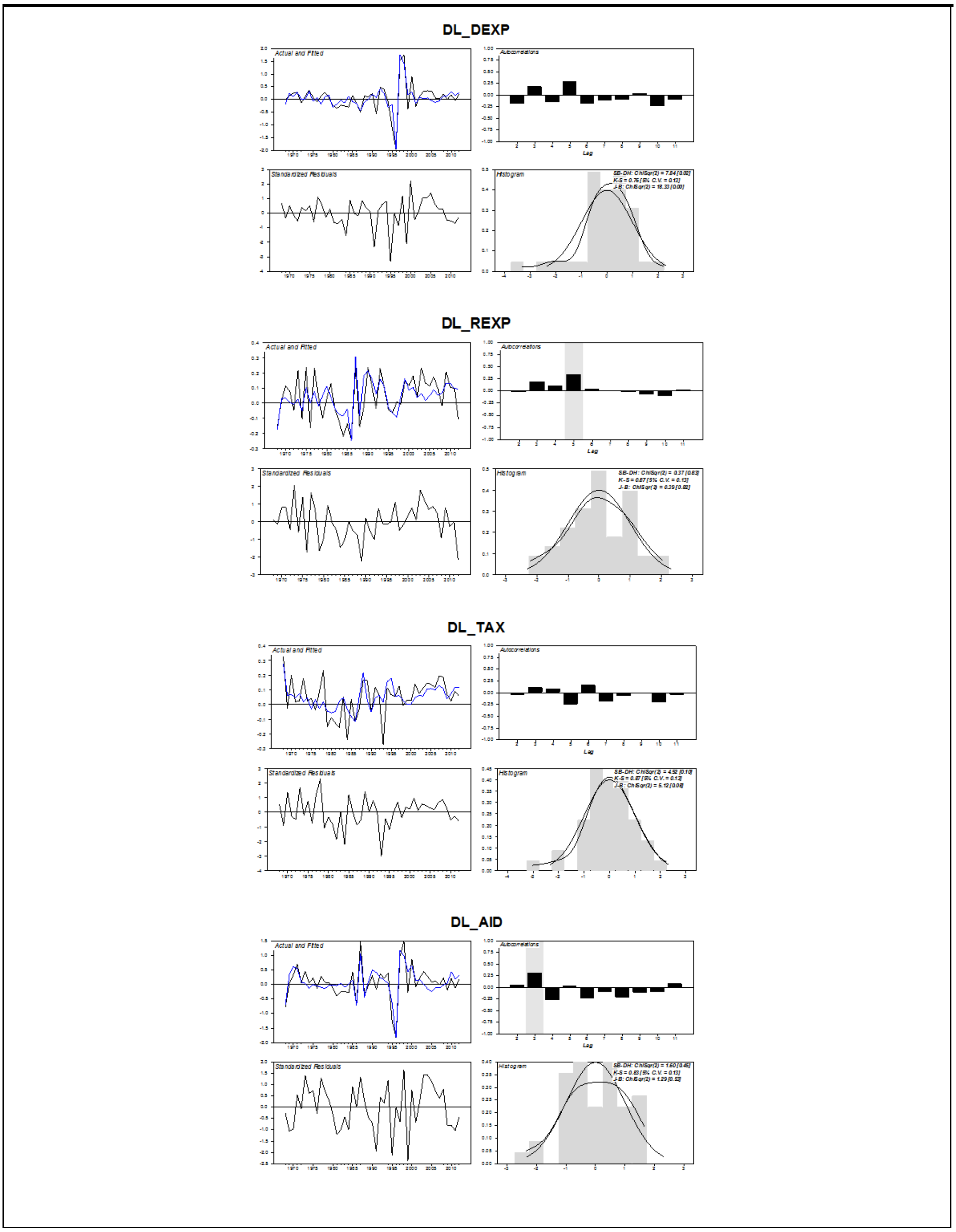
Appendix Table D5: Alternative Identification of the Aggregated Model

Aggregate model: not identified					
	texp	***tax***	***aid***	Shift 1986	
LR equilibrium relation (β_1)	1.512	-4.391	1.669	-1.297	~I(0)
LR equilibrium relation (β_2)	-7.780	7.389	1.650	-1.749	~I(0)
Adjustment coefficients (α_1)	0.202 (3.005)	-0.049 (-0.727)	1.411 (6.684)		
Adjustment coefficients (α_2)	-0.285 (-2.389)	0.351 (2.958)	-0.109 (-0.291)		
Multivariate normality	p-value = 0.862				
Trace correlation	0.463				
Aggregate model: just identified					
	texp	***tax***	***aid***	Shift 1986	
LR equilibrium relation (β_1)	1.000	-1.265 (-20.733)	0.000	0.050 (0.837)	~I(0)
LR equilibrium relation (β_2)	0.000	-1.485 (-8.594)	1.000	-0.823 (-4.842)	~I(0)
Adjustment coefficients (α_1)	-0.355 (-2.919)	0.368 (3.042)	-0.595 (-1.561)		
Adjustment coefficients (α_2)	-0.016 (-0.456)	-0.056 (-1.563)	-0.513 (-4.548)		
Multivariate normality	p-value = 0.862				
Trace correlation	0.463				
Aggregated model: (over-)identified					
	texp	***tax***	***aid***	Shift 1986	
LR equilibrium relation (β_1)	1.000	-1.251 (-20.744)	0.000	0.000	~I(0)
LR equilibrium relation (β_2)	0.000	-1.471 (-8.455)	1.000	-0.872 (-5.472)	~I(0)
Adjustment coefficients (α_1)	-0.368 (-3.102)	0.337 (2.780)	-0.596 (-1.590)		
Adjustment coefficients (α_2)	-0.010 (-0.289)	-0.057 (-1.543)	-0.499 (-4.394)		
Multivariate normality	p-value = 0.843				
Test of restricted model	p-value = 0.481				
Trace correlation	0.463				

Appendix Figure D2: Plots of Cointegrating Relationships (Over-identified Aggregated Model)



Appendix Figure D3: UVAR Residuals (Model with Disaggregated Expenditures)

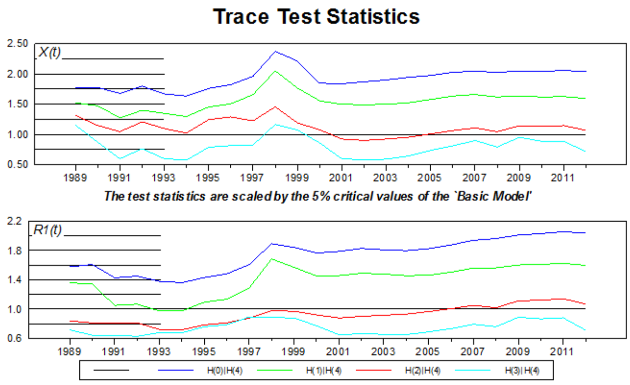


Appendix Table D6: Additional Information for Rank Determination (Model with Disaggregated Expenditures)

Moduli of the roots of the companion matrix

Rank	*Root 1*	*Root 2*	*Root 3*	*Root 4*	*Root 5*	*Root 6*	*Root 7*	*Root 8*
Rank=4	1.106	0.579	0.579	0.421	0.421	0.401	0.401	0.175
Rank=3	1.000	0.548	0.548	0.420	0.420	0.378	0.259	0.259
Rank=2	1.000	1.000	0.392	0.392	0.307	0.307	0.238	0.238
Rank=1	1.000	1.000	1.000	0.338	0.338	0.316	0.308	0.082
Rank=0	1.000	1.000	1.000	1.000	0.351	0.351	0.236	0.229

Trace test statistics (recursive estimation)



Alpha matrix coefficients (UVAR)

	L_DEXP	L_REXP	L_TAX	L_AID
Alfa matrix coefficients (α_1)	0.233 (5.793)	-0.052 (-3.802)	-0.001 (-0.074)	-0.014 (-0.309)
Alfa matrix coefficients (α_2)	0.230 (5.736)	0.029 (2.073)	-0.029 (-2.045)	0.308 (6.985)
Alfa matrix coefficients (α_3)	-0.050 (-1.249)	0.027 (1.955)	-0.040 (-2.856)	-0.095 (-2.168)
Alfa matrix coefficients (α_4)	0.053 (1.326)	0.035 (2.538)	0.034 (2.388)	0.034 (0.764)

Appendix Table D7: Individual Hypothesis Testing (Model with Disaggregated Expenditures)

	P-Value	L_DEXP	L_REXP	L_TAX	L_AID	C(1986:01)	TREND
Rel 1	0.00004	1.00000	-1.10817	0.00000	0.00000	0.00000	0.00000
Rel 2	0.00216	1.00000	0.00000	-1.45544	0.00000	0.00000	0.00000
Rel 3	0.00004	0.00000	1.00000	-1.17271	0.00000	0.00000	0.00000
Rel 4	0.00000	1.00000	0.00000	0.00000	-0.11461	0.00000	0.00000
Rel 5	0.00008	0.00000	1.00000	0.00000	-0.70741	0.00000	0.00000
Rel 6	0.02032	1.00000	2.06844	-3.82220	0.00000	0.00000	0.00000
Rel 7	0.00115	1.00000	-2.92933	-0.00000	1.19963	-0.00000	-0.00000
Rel 8	0.00010	-0.00000	-0.00000	1.00000	-0.60434	-0.00000	-0.00000
Rel 9	0.00003	-0.00000	1.00000	1.10831	-1.37409	-0.00000	-0.00000
Rel 10	0.11302	1.00000	-0.00000	-2.87030	0.80572	-0.00000	-0.00000
Rel 11	0.06025	1.00000	0.84977	-3.58950	0.66677	0.00000	0.00000
Rel 12	0.30037	1.00000	0.00000	-1.28301	-0.23868	0.89269	0.00000
Rel 13	0.03699	1.00000	0.00000	-2.71791	0.64967	0.00000	0.00669
Rel 14	1.00000	1.00000	-0.00000	-1.10490	-0.21250	1.21504	-0.01678
Rel 15	1.00000	0.00000	1.00000	-1.15160	0.32931	0.28214	-0.03417
Rel 16	0.02393	-0.00000	1.00000	-2.18107	0.71198	-0.72617	-0.00000
Rel 17	0.32023	0.00000	1.00000	-1.41560	0.42674	0.00000	-0.02503
Rel 18	0.00003	-0.00000	1.00000	1.10831	-1.37409	-0.00000	-0.00000

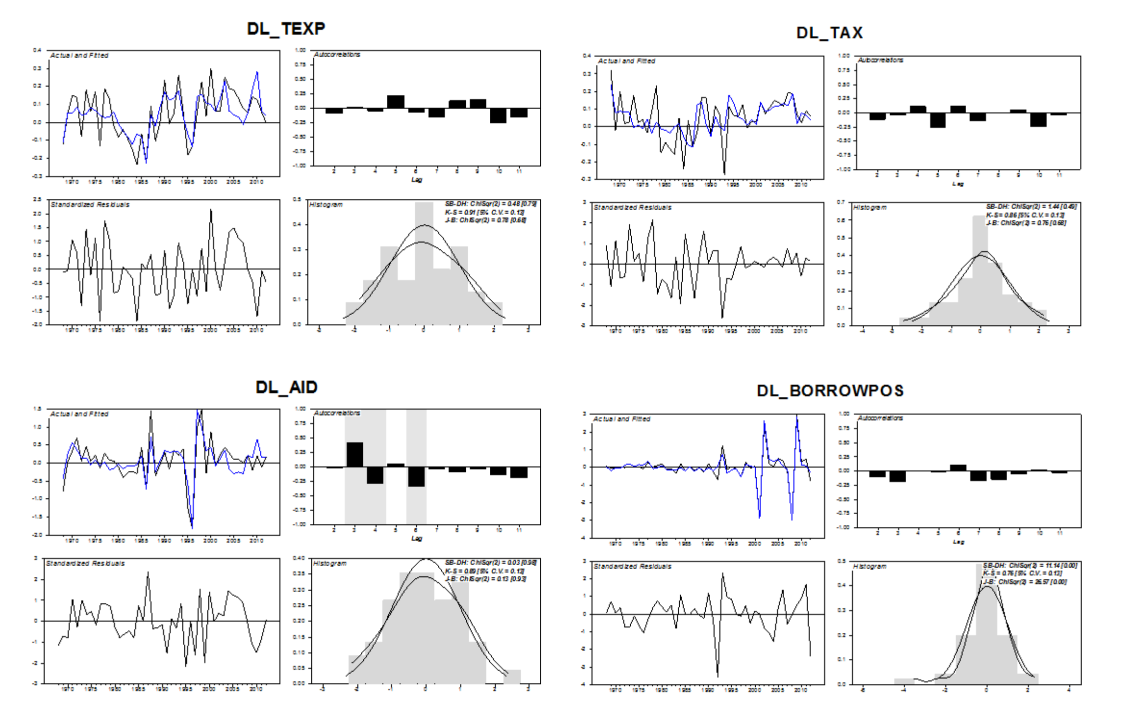
Appendix Table D8: Alternative Identification of the Model with Disaggregated Expenditures

Alternatively, this system could be identified along the domestic/foreign funded motivation. The first cointegrating relationship would summarise the ‘domestic’ equilibrium (aid excluded), where both development and recurrent expenditures would be positively related to tax revenue, with development expenditures adjusting more quickly (than recurrent revenue) to equilibrium error (although aid, excluded from the beta, indicates a very strong adjustment behaviour). The second equilibrium could be identified as expenditure-aid relationship, with aid positively related to development expenditures in the long run, and negatively associated to the recurrent spending.

```
THE MATRICES BASED ON 2 COINTEGRATING VECTORS:

BETA(transposed)
            L_DEXP    L_REXP    L_TAX    L_AID C(1986:01)  TREND
Beta(1)      1.000     0.645    -1.848   0.000      1.397   -0.039
             (.NA)   (1.535)  (-5.758)  (.NA)     (3.143)  (-1.595)
Beta(2)     -1.892     1.816     0.000   1.000     -1.787   -0.030
          (-20.802)  (4.847)    (.NA)   (.NA)    (-4.144)  (-1.244)
ALPHA
        Alpha(1) Alpha(2)
DL_DEX   -0.680    -0.049
        (-5.720) (-0.478)
DL_REX   -0.077    -0.139
        (-1.754) (-3.668)
DL_TAX    0.082     0.052
         (1.789)   (1.302)
DL_AID   -0.878    -0.595
```

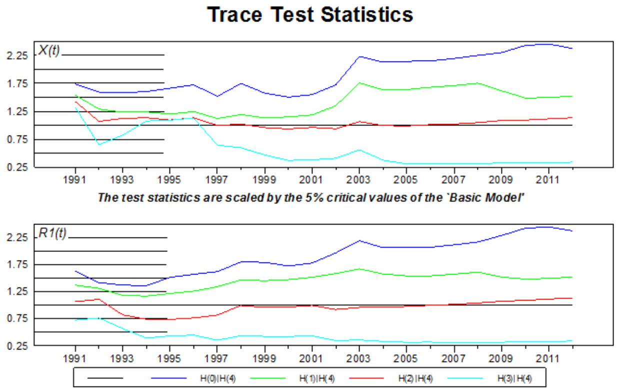
	(-6.595)	(-5.162)				
PI						
	L_DEXP	L_REXP	L_TAX	L_AID	C(1986:01)	TREND
DL_DEX	-0.587	-0.528	1.257	-0.049	-0.863	0.028
	(-4.339)	(-2.125)	(5.720)	(-0.478)	(-6.641)	(3.867)
DL_REX	0.186	-0.301	0.141	-0.139	0.141	0.007
	(3.741)	(-3.301)	(1.754)	(-3.668)	(2.954)	(2.708)
DL_TAX	-0.016	0.147	-0.152	0.052	0.022	-0.005
	(-0.303)	(1.532)	(-1.789)	(1.302)	(0.444)	(-1.708)
DL_AID	0.248	-1.647	1.623	-0.595	-0.163	0.052
	(1.640)	(-5.919)	(6.595)	(-5.162)	(-1.121)	(6.452)

Appendix Figure D4: Residuals from UVAR (Model with Borrowing)



Appendix Table D9: Additional Information for Rank Determination (Model with Borrowing)

Trace test statistics (recursive estimation)



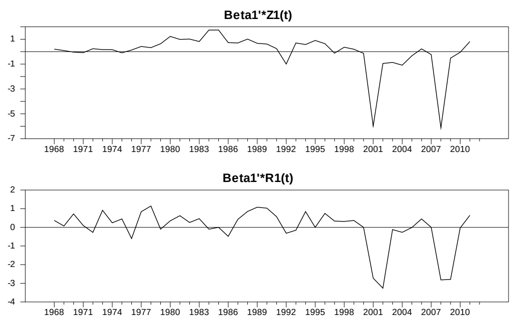
Moduli of the roots of the companion matrix

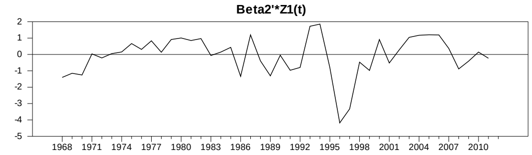
Rank	*Root 1*	*Root 2*	*Root 3*	*Root 4*	*Root 5*	*Root 6*	*Root 7*	*Root 8*
Rank=4	1.047	0.604	0.496	0.496	0.290	0.290	0.242	0.242
Rank=3	1.000	0.710	0.518	0.518	0.325	0.325	0.242	0.242
Rank=2	1.000	1.000	0.595	0.375	0.317	0.317	0.246	0.154
Rank=1	1.000	1.000	1.000	0.629	0.327	0.258	0.258	0.170
Rank=0	1.000	1.000	1.000	1.000	0.481	0.408	0.235	0.172

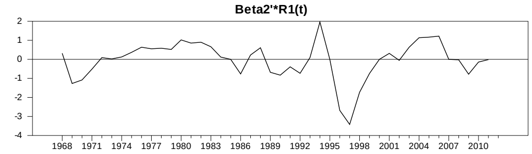
Alpha matrix coefficients (UVAR)

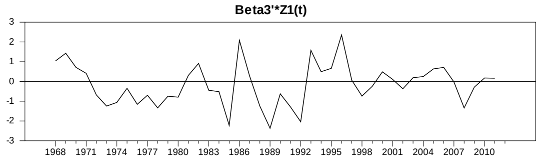
	L_TEXP	L_TAX	L_AID	L_BORROWpos
Alfa matrix coefficients (α_1)	-0.038 (-2.794)	-0.012 (-0.835)	-0.050 (-1.096)	-0.422 (-13.536)
Alfa matrix coefficients (α_2)	-0.052 (-3.806)	0.013 (0.914)	-0.331 (-7.233)	0.026 (0.831)
Alfa matrix coefficients (α_3)	-0.030 (-2.210)	0.051 (3.623)	0.012 (0.253)	-0.087 (-2.798)
Alfa matrix coefficients (α_4)	0.022 (1.576)	0.023 (1.615)	0.021 (0.450)	-0.014 (-0.440)

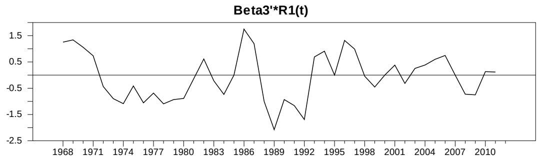
(Potential) CI relations

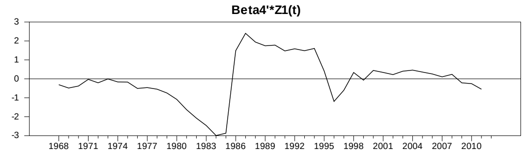


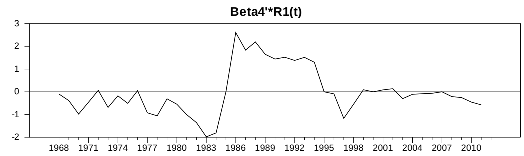

Beta2'*Z1(t)


Beta2'*R1(t)

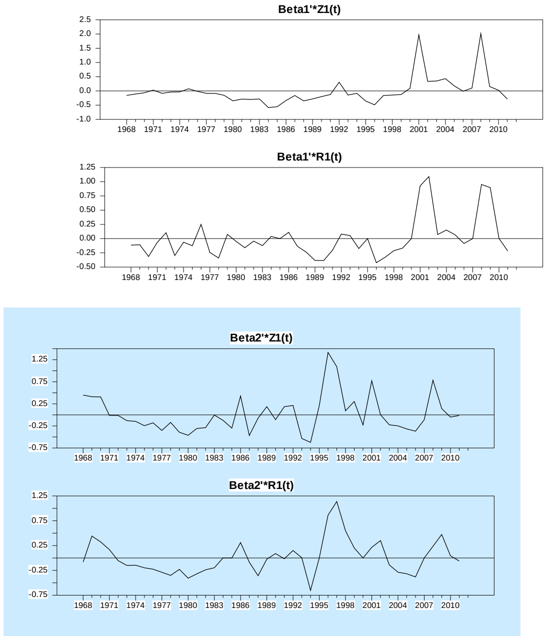

Beta3'*Z1(t)


Beta3'*R1(t)


Beta4'*Z1(t)

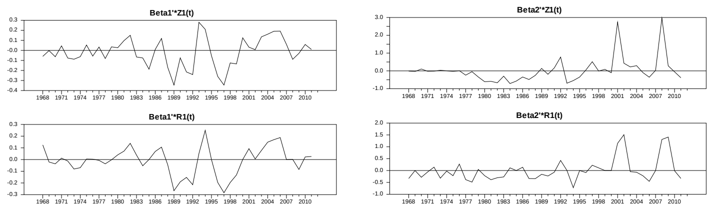

Beta4'*R1(t)

Appendix Figure D5: Cointegrating Vectors (Over-identified Model with Borrowing)



Appendix Table D10: Alternative Over-identified Model with Borrowing

Alternatively, this model could be identified as a system where the first cointegrating relationship describes a very long run equilibrium between expenditure and tax revenue, with all variables adjusting to departures from such equilibrium; and the second one describes long-run interactions between the revenue variables: tax is positively associated with both aid and borrowing in the long run, and aid and borrowing can be regarded as substitutes. The key difference form the identification in the main text is that now the most adjusting variable to the second cointegrating vector is borrowing rather than aid.

	texp	***tax***	***aid***	***borrow***	Shift 1986	Trend	
LR equilibrium relation (β_1)	1.000	-1.000	0.000	0.000	0.320 (3.899)	-0.015 (-4.647)	~I(0)
LR equilibrium relation (β_2)	0.000	1.000	-0.154 (-2.886)	-0.929 (-14.537)	0.300 (1.708)	-0.023 (-2.847)	~I(0)
Adjustment coefficients (α_1)	-0.328 (-3.013)	0.322 (2.717)	-1.715 (-3.993)	0.962 (3.654)			
Adjustment coefficients (α_2)	0.098 (3.234)	0.015 (0.448)	0.179 (1.496)	0.914 (12.458)			
Multivariate normality	0.017						
Test of restricted model	0.100						
Trace correlation	0.550						
							

Appendix E

Appendix Table E1: Model Specification Testing (Tanzania) – Recipient Aid Data

```
@cats(lags=1,det=drift) 1966:1 2012:1
# L_TEXP L_DOMREV L_TOTAID
```

Lag length determination (Tanzania)

Model	*k*	*T*	*Regr.*	*Log-lik*	*SC*	*H-Q*	*LM(1)*	*LM(k)*
VAR(5)	5	42	16	256.619	-7.948	-9.206	0.379	0.665
VAR(4)	4	42	13	245.995	-8.243	-9.265	0.153	0.206
VAR(3)	3	42	10	236.715	-8.602	-9.389	0.205	0.347
VAR(2)	2	42	7	234.146	-9.281	-9.831	0.986	0.670
VAR(1)	1	42	4	229.896	-9.880	-10.194	0.735	0.735

Residuals from Unrestricted VAR

Residual normality (p-values)

	Multivariate	Univariate		
		texp	***domrev***	***aid***
	0.000	0.319	0.069	0.000

Residual autocorrelation and ARCH effects (p-values)

	LM(1)	LM(2)	LM(3)	LM(4)
Residual autocorrelation	0.617	0.703	0.381	0.046
ARCH	0.416	0.232	0.074	0.028
Trace correlation	0.198			

Determination of Cointegration Rank (Trace Test)

p-r	r	Eig. value	Trace	Trace*	Frac95	p-value	p-value*
3	0	0.311	32.102	31.066	29.804	0.026	0.035
2	1	0.266	14.942	14.657	15.408	0.059	0.065
1	2	0.015	0.688	0.683	3.841	0.407	0.408

Long-Run Identification Tests (p-values)

	r	***texp***	***domrev***	***aid***
Long-run exclusion	*r=2*	0.004	0.001	0.002
	r=1	0.170	0.342	0.231
Stationarity	*r=2*	0.001	0.001	0.029
	r=1	0.001	0.005	0.062
Weak exogeneity	*r=2*	0.001	0.275	0.004
	r=1	0.223	0.319	0.880
Purely adjusting	*r=2*	0.307	0.001	0.108
	r=1	0.554	0.001	0.155

```
TEST OF RESTRICTED MODEL:    CHISQR(1) = 1.437 [0.231]
BARTLETT CORRECTION:         CHISQR(1) = 0.960 [0.327] (Correction Factor: 1.496)
        L_TEXP L_DOMREV  L_TOTAID
Beta(1)  1.000   -1.369     0.000
         (.NA) (-14.169)    (.NA)
ALPHA
       Alpha(1)
DL_TEX  -0.324
       (-4.138)
DL_DOM  -0.017
       (-0.222)
DL_TOT  -0.766
       (-2.060)
        L_TEXP  L_DOMREV L_TOTAID
DL_TEX  -0.324    0.443     0.000
       (-4.138)  (4.138)    (.NA)
DL_DOM  -0.017    0.024     0.000
       (-0.222)  (0.222)    (.NA)
DL_TOT  -0.766    1.049     0.000
       (-2.060)  (2.060)    (.NA)
Log-Likelihood = 242.053
```

Appendix Table E2: Model Specification Testing (Tanzania) – DAC Donors Aid Data

```
@cats(lags=1,det=drift) 1966:1 2012:1
# L_TEXP L_DOMREV L_TOTAID_DAC
```

Lag length determination (Tanzania)

Model	*k*	*T*	*Regr.*	*Log-lik*	*SC*	*H-Q*	*LM(1)*	*LM(k)*
VAR(5)	5	42	16	266.308	-8.410	-9.668	0.836	0.623
VAR(4)	4	42	13	262.592	-9.034	-10.056	0.695	0.847
VAR(3)	3	42	10	259.595	-9.692	-10.478	0.765	0.999
VAR(2)	2	42	7	255.720	-10.308	-10.859	0.613	0.756
VAR(1)	1	42	4	249.856	-10.830	-11.144	0.282	0.282

Residuals from Unrestricted VAR

Residual normality (p-values)

	Multivariate	Univariate		
		texp	***domrev***	***Aid_Dac***
	0.046	0.454	0.074	0.120

Residual autocorrelation and ARCH effects (p-values)

	LM(1)	LM(2)	LM(3)	LM(4)
Residual autocorrelation	0.336	0.576	0.886	0.730
ARCH	0.341	0.174	0.150	0.185
Trace correlation	0.154			

Determination of Cointegration Rank (Trace Test)

p-r	r	Eig. value	Trace	Trace*	Frac95	p-value	p-value*
3	0	0.318	24.681	23.885	20.731	0.015	0.019
2	1	0.139	7.079	6.943	9.751	0.144	0.151
1	2	0.004	0.171	0.170	0.000	.NA	.NA

Long-Run Identification Tests

	r	***texp***	***domrev***	***aid***
Long-run exclusion	*r=2*	0.000	0.000	0.057
	r=1	0.002	0.001	0.171
Stationarity	*r=2*	0.063	0.108	0.622
	r=1	0.001	0.003	0.005
Weak exogeneity	*r=2*	0.000	0.194	0.211
	r=1	0.002	0.749	0.421
Purely adjusting	*r=2*	0.307	0.001	0.108
	r=1	0.554	0.001	0.155

```
TEST OF RESTRICTED MODEL:    CHISQR(1) = 1.870 [0.171]
BARTLETT CORRECTION:         CHISQR(1) = 1.289 [0.256] (Correction Factor: 1.451)
THE MATRICES BASED ON 1 COINTEGRATING VECTOR:
BETA(transposed)
        L_TEXP L_DOMREV  L_TOTAID_DAC
Beta(1)  1.000   -1.364         0.000
         (.NA) (-14.211)        (.NA)
ALPHA
       Alpha(1)
DL_TEX  -0.326
       (-4.135)
DL_DOM  -0.016
       (-0.206)
DL_TOT  -0.045
       (-0.261)
PI
        L_TEXP  L_DOMREV L_TOTAID_DAC
DL_TEX  -0.326    0.444         0.000
       (-4.135)  (4.135)        (.NA)
DL_DOM  -0.016    0.022         0.000
       (-0.206)  (0.206)        (.NA)
DL_TOT  -0.045    0.062         0.000
       (-0.261)  (0.261)        (.NA)
Log-Likelihood = 267.447
Trace Correlation                =   0.097
```

Appendix Table E3: Model Specification Testing (Ethiopia) – Recipient Aid Data

```
@cats(lags=2,det=drift) 1963:1 2009:1
# L_TEXP L_DOMREV L_TOTAID
```

Lag length determination (Tanzania)

Model	*k*	*T*	*Regr.*	*Log-lik*	*SC*	*H-Q*	*LM(1)*	*LM(k)*
VAR(5)	5	42	16	288.862	-9.484	-10.742	0.894	0.659
VAR(4)	4	42	13	282.198	-9.967	-10.989	0.475	0.121
VAR(3)	3	42	10	277.287	-10.534	-11.321	0.787	0.973
VAR(2)	2	42	7	266.423	-10.818	-11.368	0.108	0.401
VAR(1)	1	42	4	253.750	-11.015	-11.330	0.013	0.013

Residuals from Unrestricted VAR

Residual normality (p-values)

	Multivariate	Univariate		
		texp	***domrev***	***aid***
	0.025	0.484	0.123	0.781

Residual autocorrelation and ARCH effects (p-values)

	LM(1)	LM(2)	LM(3)	LM(4)
Residual autocorrelation	0.153	0.444	0.151	0.214
ARCH	0.172	0.146	0.217	0.401
Trace correlation	0.334			

Determination of Cointegration Rank (Trace Test)

p-r	r	Eig. value	Trace	Trace*	Frac95	p-value	p-value*
3	0	0.396	34.298	31.183	29.804	0.013	0.034
2	1	0.221	11.595	10.373	15.408	0.180	0.258
1	2	0.008	0.377	0.351	3.841	0.539	0.554

Long-Run Identification Tests (p-values)

	r	***texp***	***domrev***	***aid***
Long-run exclusion	*r=2*	0.000	0.000	0.004
	r=1	0.007	0.001	0.723
Stationarity	*r=2*	0.001	0.001	0.001
	r=1	0.000	0.000	0.000
Weak exogeneity	*r=2*	0.010	0.728	0.011
	r=1	0.002	0.536	0.392
Purely adjusting	*r=2*	0.717	0.004	0.688
	r=1	0.535	0.009	0.004

```
TEST OF RESTRICTED MODEL:    CHISQR(1) = 0.126 [0.723]
BARTLETT CORRECTION:         CHISQR(1) = 0.086 [0.769] (Correction Factor: 1.460)
        L_TEXP L_DOMREV  L_TOTAID
Beta(1)  1.000   -1.063     0.000
         (.NA) (-44.946)    (.NA)
ALPHA
       Alpha(1)
DL_TEX  -0.509
       (-3.195)
DL_DOM   0.115
        (0.578)
DL_TOT  -0.428
       (-0.814)
        L_TEXP  L_DOMREV L_TOTAID
DL_TEX  -0.509    0.541     0.000
       (-3.195)  (3.195)    (.NA)
DL_DOM   0.115   -0.122     0.000
        (0.578) (-0.578)    (.NA)
DL_TOT  -0.428    0.455     0.000
       (-0.814)  (0.814)    (.NA)
Log-Likelihood = 278.863
```

Appendix Table E4: Model Specification Testing (Ethiopia) – DAC Donors Aid Data

```
@cats(lags=2,det=drift) 1963:1 2009:1
# L_TEXP L_DOMREV L_TOTAID_DAC
```

Lag length determination (Tanzania)

Model	*k*	*T*	*Regr.*	*Log-lik*	*SC*	*H-Q*	*LM(1)*	*LM(k)*
VAR(5)	5	42	16	291.423	-9.606	-10.864	0.750	0.999
VAR(4)	4	42	13	282.327	-9.973	-10.996	0.374	0.144
VAR(3)	3	42	10	273.469	-10.353	-11.139	0.182	0.693
VAR(2)	2	42	7	263.715	-10.689	-11.239	0.102	0.199
VAR(1)	1	42	4	257.063	-11.173	-11.488	0.260	0.260

Residuals from Unrestricted VAR

Residual normality (p-values)				
	Multivariate	Univariate		
		texp	***domrev***	***Aid_dac***
	0.025	0.221	0.022	0.464
Residual autocorrelation and ARCH effects (p-values)				
	LM(1)	LM(2)	LM(3)	LM(4)
Residual autocorrelation	0.236	0.372	0.192	0.028
ARCH	0.049	0.111	0.170	0.159
Trace correlation	0.251			

Determination of Cointegration Rank (Trace Test)

p-r	r	Eig. value	Trace	Trace*	Frac95	p-value	p-value*
3	0	0.337	29.084	25.711	20.696	0.004	0.011
2	1	0.203	10.606	9.426	9.215	0.032	0.051
1	2	0.009	0.414	0.391	0.000	.NA	.NA

Long-Run Identification Tests (p-values)

	r	***texp***	***domrev***	***aid***
Long-run exclusion	*r=2*	0.000	0.000	0.005
	r=1	0.004	0.006	0.388
Stationarity	*r=2*	0.002	0.002	0.004
	r=1	0.000	0.000	0.000
Weak exogeneity	*r=2*	0.040	0.292	0.040
	r=1	0.013	0.721	0.069
Purely adjusting	*r=2*	0.202	0.451	0.186
	r=1	0.186	0.026	0.016

```
TEST OF RESTRICTED MODEL:    CHISQR(1) = 0.747 [0.388]
BARTLETT CORRECTION:         CHISQR(1) = 0.478 [0.489] (Correction Factor: 1.562)
THE MATRICES BASED ON 1 COINTEGRATING VECTOR:
BETA(transposed)
        L_TEXP L_DOMREV  L_TOTAID_DAC
Beta(1)  1.000   -1.068         0.000
         (.NA) (-40.983)        (.NA)
ALPHA
       Alpha(1)
DL_TEX  -0.468
       (-2.432)
DL_DOM   0.180
        (0.828)
DL_TOT   0.716
        (1.650)
PI
        L_TEXP  L_DOMREV L_TOTAID_DAC
DL_TEX  -0.468    0.500         0.000
       (-2.432)  (2.432)        (.NA)
DL_DOM   0.180   -0.192         0.000
        (0.828) (-0.828)        (.NA)
DL_TOT   0.716   -0.764         0.000
        (1.650) (-1.650)        (.NA)
Log-Likelihood = 279.959
Test for Normality:   ChiSqr(6)   =  15.475 [0.017]
Trace Correlation                 =   0.191
```

Bibliography

All websites last accessed September 2014, unless otherwise stated.

Arndt, C., S. Jones, F. Tarp, (2010), 'Aid, Growth and Development: Have We Come Full Circle', *Journal of Globalization and Development,* 1(2), Article 5.

Baldacci, E., B. Clements, S. Gupta and Q. Cui (2008), 'Social Spending, Human Capital and Growth in Developing Countries', *World Development,* 36, 1317-41.

Barnard, J. and X. L. Meng (1999), 'Applications of Multiple Imputation in Medical Studies: from AIDS to NHANES, *Statistical Methods in Medical Research*, 8, 17-36.

Batniji, R. and E. Bendavid (2012), 'Does Development Assistance for Health Really Displace Government Health Spending? Reassessing the Evidence', *PLoS Medicine*, 9(5): e1001214. Note: Retracted: http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001214#pmed-1001214-t004 (Last accessed on 7th August, 2014)

Binh, T. and M. McGillivray (1993) 'Foreign Aid, Taxes and Public Investment: A Comment', *Journal of Development Economics*, 41, 173-176.

Bougheas, S, I. Dasgupta, I. and O. Morrissey (2007), 'Tough Love or Unconditional Charity?', Oxford *Economic Papers*, 59(4), 561-582.

Burnside, C. and D. Dollar, (2000), 'Aid, Policies and Growth', *American Economic Review* 90(4), 847-868.

Bwire, T. (2013), 'Aid, Fiscal Policy and Macroeconomy of Uganda: CVAR approach', Ph. D. Thesis, University of Nottingham.

CATS 2.0 (Cointegration Analysis of Time Series). Illinois: Estima.

Clist., P., A. Isopi and O. Morrissey (2012), 'Selectivity on Aid Modality: Determinants of Budget Support from Multilateral Donors', *Review of International Organisations,* 7(3), 267-284.

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

Devarajan, S., A. S. Rajkumar and V. Swaroop (1999), 'What Does Aid to Africa Finance?', *World Bank Policy Research Paper* No. 2092.

Dieleman, J. L., M. C. Graves and M. Hanlon (2013), 'The Fungibility of Health Aid: Reconsidering the Reconsidered', *Journal of Development Studies*, 49(12), 1755-1762.

Doucouliagos, H. and M. Paldam (2008), 'Aid Effectiveness on Growth. A Meta Study', *European Journal of Political Economy,* 24, 1-24.

Eberhardt, M. and F. Teal (2013), 'Structural Change and Cross-country Growth Empirics', *World Bank Economic Review*, 27(2), 229-271.

The Economist, (2014), 'Aid to the Rescue', *The Economist*, Print Edition, 16th August, London. Available online at: http://www.economist.com/news/finance-and-economics/21612183-new-research-suggests-development-aid-does-foster-growthbut-what-cost-aid

Ezzati-Rice, T. M, W. Johnson, M. Khare, R. J. A. Little, D. B. Rubin and J. L. Schafer (1995), 'A Simulation Study to Evaluate the Performance of Model-Based Multiple Imputations in NCHS Health Examination Surveys', *Proceedings of the Annual Research Conference*. Washington, DC: Bureau of the Census, 257-266.

Fagernäs, S. and C. Schurich (2004). 'The Fiscal Effects of Aid in Malawi', *Economics and Statistics Analysis Unit, Working Paper* 7. London: Overseas Development Institute.

Fagernäs, S., J. and Roberts (2004a), 'The Fiscal Effects of Aid in Uganda', *Economics and Statistics Analysis Unit, Working Paper* 9. London: Overseas Development Institute.

Fagernäs, S. and J. Roberts (2004b), 'The Fiscal Effects of Aid in Zambia', *Economics and Statistics Analysis Unit, Working Paper* 10. London: Overseas Development Institute.

Fagernas, S. and J. Roberts (2004c), 'Fiscal Impact of Aid: A Survey of issues and synthesis of country studies of Malawi, Uganda and Zambia'. *Economics and Statistics Analysis Unit, Working Paper* 11. London: Overseas Development Institute.

Farag, M., A. K. Nandakumar, S. S. Wallack, G. Gaumer and D. Hodgkin, (2009), 'Does Funding from Donors Displace Government Spending for Health in Developing Countries', *Health Affairs,* 28, 1045-55.

Feyzioglu, T., V. Swaroop and M. Zhu (1988), 'A Panel Data Analysis of the Fungibility of Foreign Aid', *World Bank Economic Review*, 12(1), 29-58.

Franco-Rodriguez, S., O. Morrissey and M. McGillivray (1998), 'Aid and the Public Sector in Pakistan: Evidence with Endogenous aid', *World Development*, 26(7), 1241-1250.

Gomanee, K., O. Morrissey, P. Mosley and A. Verschoor (2005), 'Aid, Government Expenditure and Aggregate Welfare', *World Development,* 33(3), 355-370.

Graham, J. W., S. M. Hofer and A. M. Piccinin (1994), 'Analysis with Missing Data in Drug Prevention Research' in L. M. Collins and L. Seitz (Eds.), *Advances in Data Analysis for Prevention Intervention Research*, 13-63. Washington, DC: National Institute on Drug Abuse.

Graham, J. W., A. E. Olchowski and T. D. Gilreath (2007), 'How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory', *Prevention Science*, 8, 206-213.

Graham, J. W. and J. L. Schafer (1999), 'On the Performance of Multiple Imputation for Multivariate Data with Small Sample Size', in R. Hoyle (Ed.), *Statistical Strategies for Small Sample Research*, 1-29. Thousand Oaks, CA: Sage

Gottret, P. And G. Schieber (2006), *Health Financing Revisited: A Practitioner's Guide*. Herndon, VA: World Bank Publications.

Gupta, S., B. Clements, A. Pivovarsky and E. Tiongson (2004) 'Foreign Aid and Revenue Response: Does the Composition of Aid Matter?', in *Helping Countries Develop: The Role of Fiscal Policy*. S. Gupta, B. Clements and G. Inchauste ed. Washington DC: International Monetary Fund, 385-406.

Hansen, H. and K. Juselius, (1995), *Cats in Rats: Cointegration Analysis of Time Series*. Illinois: Estima.

Hansen, H. and J. A. Doornik (1996), 'An Omnibus Test for Univariate and Multivariate Normality', Discussion paper, Nuffield College, Oxford University.

Hansen, H. and F. Tarp (2000), 'Aid Effectiveness Disputed', *Journal of International Development* 12(3), 375-398.

Hansen, H. and F. Tarp (2001), 'Aid and Growth Regressions', *Journal of Development Economics* 64(2), 547-570.

Heitjan, D. F. and D. B. Rubin (1991), 'Ignorability and Coarse Data', *Annals of Statistics*, 19, 2244-2253.

Helleiner, G., T. Killick, N. Lipumba, N. Ndulu and K. E. Svendsen, (1995), 'Report of the Group of Independent Advisers on Development Cooperation Issues between Tanzania and Its Aid Donors'. Copenhagen: Royal Danish Ministry of Foreign Affairs.

Helleiner, G. (2001), 'Local Ownership and Donor Performance Monitoring: New Aid Relationships in Tanzania?', *Journal of Human Development*, 3(2), 2002, 251-261.

Heller, P. S. (1975), 'A Model of Public Fiscal Behavior in Developing Countries: Aid, Investment and Taxation', *American Economic Review*, 65, 429–445.

Hendry, D. F. and J. A. Doornik (1996), *Empirical Econometric Modelling using PcGive for Windows*. London: Timberlake Consultants Press.

Herzer, D. and O. Morrissey (2013), 'Foreign Aid and Domestic Output in the Long Run', *Review of World Economics,* 149(4), 723-748.

von Hippel, P. T. (2007), 'Regression with Missing Y's: An Improved Method for Analyzing Multiply-Imputed Data', *Social Methodology*, 37, edited by Yu Xie, MA: Blackwell Publishing.

Honaker, J. and G. King (2010), 'What to Do about Missing Values in Time-Series Cross-Section Data', *American Journal of Political Science,* 54(2), 561-581.

Honaker, J., G. King and M. Blackwell (2012), *AMELIA II: A Program for Missing Data, Unpublished User's Guide*. Cambridge, MA: Harvard University, http://gking.harvard.edu/amelia/

Horton, N. J. and S. R. Lipsitz (2001), 'Multiple Imputation in Practice: Comparison of Software Packages for Regression Models with Missing Variables, *American Statistician*, 55, 244-254.

Hyden, G. and B. Karlstrom (1993), ‘Structural Adjustment as a Policy Process: The Case of Tanzania’, *World Development*, 21, 1395-1404.

IMF. World Economic Outlook Database. Washington DC: International Monetary Fund, 2009. http://imf.org/external/data.htm;

IMF Fiscal Affairs Department. Total Health Spending Database. Washington, DC: International Monetary Fund, 2009.

Jerven, M. (2013), *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It*. Ithaca and London: Cornell University Press.

Johansen, S. (1995), ‘Identifying Restrictions of Linear Equations. With Applications to Simultaneous Equations and Cointegration’, *Journal of Econometrics*, 69(1), 111-132.

Johansen, S. (2002), ‘A Small Sample Correction of the Test for Cointegrating Rank in the Vector autoregressive Model’, *Econometrica*, 70, 1929-1961.

Johansen, S. and K. Juselius (1994), ‘Identification of the Long-run and Short-run Structure: An Application to the ISLM Model’, *Journal of Econometrics*, 63, 7-36.

Juselius, K., (2006), *The Cointegrated VAR Model: Methodology and Applications*. Oxford University Press, Oxford.

Juselius, K., N. F. Moller and F. Tarp (2011), ‘The Long-run Impact of Foreign Aid in 36 African Countries’, *UNU-WIDER Working Paper* 2011/51.

Juselius, K., A. Reshid and F. Tarp (2013), ‘The Real Exchange Rate, Foreign Aid and Macroeconomic Transmission Mechanisms in Tanzania and Ghana’, *UNU-WIDER Working Paper* 2013/090.

Keller, E. J. (1992), ‘Drought, War and the Politics of Famine in Ethiopia and Eritrea’, *The Journal of Modern African* Studies, 30(4), 609-24.

Kenward, M. G. and J. Carpenter (2007), ‘Multiple Imputation: Current Perspectives’, *Statistical Methods in Medical Research,* 16, 199-218.

Khilji, N. M. and E. M. Zampelli (1994), 'The Fungibility of US Military and Non-Military Assistance and the Impacts on Expenditures of Major Aid Recipients', *Journal of Development Economics*, 43(2), 345-362.,

King, G., J. Honaker, A. Joseph and K. Scheve (2001), 'Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation', *American Political Science Review*, 95(1), 49-69.

Killick, T. and M. Foster (2007), 'The Macroeconomics of Doubling Aid to Africa and the Centrality of the Supply Side', *Development Policy Review,* 25, 167-92.

Kline, P. and A. Santos (2013), 'Sensitivity to Missing Data Assumptions: Theory and an Evaluation of the U.S. Wage Structure', *Quantitative Economics,* 4, 231-267.

Kweka, J. P. and O. Morrissey (2000), 'Government Spending and Economic Growth in Tanzania, 1965-1996', *CREDIT Research Paper* No. 00/6, University of Nottingham.

Little, R. J. A. (1992), 'Regression with Missing X's: A Review', *Journal of the American Statistical Association*, 87(420), 1227-1237.

Little, R. J. A. and D. B. Rubin (1987), *Statistical Analysis with Missing Data.* New York: John Wiley.

Little, J. R. A. and D. B. Rubin (1989), 'The Analysis of Social Science Data with Missing Values', *Sociological Methods and Research,* 18, 292-326.

Little, R. J. A. and D. B. Rubin (2002), *Statistical Analysis with Missing Data (2nd Ed).* New York: Wiley.

Little, R. J. A. and N. Schenker (1995), 'Missing Data'. In G. Arminger, C. C. Clogg and M. E. Sobel (Eds.), *Handbook of Statistical Modelling for The Social and Behavioral Sciences*. New York: Plenum.

Lloyd, T., M. McGillivray, O. Morrissey and M. Opoku-Afari (2009). 'The Fiscal Effects of Aid in Developing Countries: A Comparative Dynamic Analysis'. In G. Mavrotas and M. McGillivray (eds.), *Development Aid: A Fresh Look*. Basingstoke: Palgrave Macmillan UNU-WIDER Studies in Development Economics and Policy, chapter 7, 158–79.

Lu, C. H., M. T. Schneider, P. Gubbins, K. Leach-Kemon, D. Jamison, C. J. L. Murray, (2010), 'Public Financing of Health in Developing Countries: A Cross-National Systematic Analysis', *The Lancet,* 435, 1375-87. Webappendix available at: http://www.sciencedirect.com/science/article/pii/S0140673610602334

Lütkepohl, H. (2007), *New Introduction to Multiple Time Series Analysis*. Berlin and Heidelberg: Springer.

Makoro (2008, April), 'Putting Aid on Budget: Synthesis Report', A Study for the Collaborative Africa Budget Reform Initiative (CABRI) and the Strategic Partnership with Africa (SPA). Retrieved from http://www.makoro.co.uk/files/13/publication/cabri_synthesis_eng_proof_6.pdf

Martins, P. M. G. (2010), 'Fiscal Dynamics in Ethiopia: The Cointegrated VAR model with Quarterly Data', *Economics Departments Working Paper Series* 2009-10, University of Sussex.

Martins, P. (2010b), 'Do Capital Inflows Hinder Competitiveness? The Real Exchange Rate in Ethiopia', *CREDIT Working Paper* No. 10/07.

Mascagni, G. (2014), PhD Thesis, University of Sussex.

Mavrotas, G. (2002), 'Aid and Growth in India: Some Evidence from Disaggregated Aid Data', *South Asia Economic Journa*l, 3(1), 19-49.

McGillivray, M., S. Feeby, N. Hermes and R. Lensink, (2006), 'Controversies over the Impact of Development aid: It Works; It Doesn't; It Can, But That Depends…', *Journal of International Development*, 18, 1031-1050.

McGillivray, M. and O. Morrissey (2000), 'Aid Fungibility in *Assessing Aid*: Red Herring or True Concern?', *Journal of International Development*, 12(3), 413-428.

McGillivray, M and O. Morrissey (2001), 'A Review of Evidence on the Fiscal Effects of Aid', *CREDIT Working Paper* 01/13.

McGillivray, M., and O. Morrissey (2001a), 'Aid Illusion and Public Sector Fiscal Behaviour', *Journal of Development Studies*, 37(6), 118–136.

McGillivray, M. and O. Morrissey, (2004), ‘Fiscal Effects of Aid’, in T. Addison and A. Roe (eds), *Fiscal Policy for Development.* Basingtoke: Palgrave/WIDER, 72-96.

Mekasha, T. and F. Tarp, (2013), ‘Aid and Growth: What Meta-Analysis Reveals’, *Journal of Development Studies,* 49, 564-583.

Mishra P. and D. Newhouse (2007), *Health Aid and Infant Mortality*. Washington DC: International Monetary Fund.

Morrissey, O. (2012), ‘Aid and Government Fiscal Behaviour: What Does the Evidence Say’, *UNU-Wider Working Paper No.* 2012/01.

Morrissey, O. (2014), ‘Aid and Government Fiscal Behaviour: Assessing Recent Evidence’, *forthcoming in World Development.*

Murray, C. J. L. (2012), ‘Doubts About the Evidence that Foreign Aid for Health Is Displaced into Non-Health uses – Author’s Reply’, *The Lancet,* 380(9846), 973. Available at: http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(12)61530-X/fulltext

Musgrove, P., R. Zeramdini and G. Carrin, ‘Basic Patterns in National Health Expenditure’ (2002), *Bulletin of World Health Organization,* 80, 134-46.

M’Amanja, D., T. Lloyd and O. Morrissey (2005), ‘Fiscal Aggregates, Aid and Growth in Kenya: A Vector Autoregressive (VAR) Analysis’, *CREDIT Research Paper* 05/07.

M’Amanja, D. and O. Morrissey (2006), ‘Foreign Aid, Investment and Economic Growth in Kenya: A Time Series Approach’, *CREDIT Research Paper* 05/06.

OECD (2007), ‘Reporting Directives for the Creditor Reporting System’. Paris: OECD. Retrieved from http://www.oecd.org/development/stats/1948102.pdf.

OECD (2008), ‘2008 Survey on Monitoring the Paris Declaration: Making Aid More Effective by 2010’. Paris: OECD. Retrieved from http://www.oecd.org/development/effectiveness/41202121.pdf.

OECD (2010), ’Reporting directives for the Creditor Reporting System: Corrigendum on the channels of delivery’. Paris: OECD. Retrieved from http://www.oecd.org/dac/aidstatistics/45917818.pdf.

Ooms, G., K. Decoster, K. Miti, S. Rens, L. Can Leemput, P. Vermeiren and W. Van Damme (2010), 'Crowding Out: Are Relations between International Health Aid and Government Health Funding Too Complex to be Captured in Averages Only?', *The Lancet,* 375 (9723), 1403-1405.

Ooms, G., D. Stuckler, A. Basu and M. McKee (2010a), 'Financing the Millennium Development Goals for Health and Beyond: Sustaining the 'Big Push'', *Globalisation and Health,* 6(17), 1-8.

Osei, R., O. Morrissey and T. Lloyd (2003), 'Modelling the Fiscal Effects of Aid: An Impulse Response Analysis for Ghana', *CREDIT Research Paper* 03/10.

Osei, R., O. Morrissey and T. Lloyd (2005), 'The Fiscal Effects of Aid in Ghana', *Journal of International Development*, 17, 1037-1054.

Ouattara, B. (2006), 'Aid, Debt and Fiscal Policies in Senegal', *Journal of International Development*, 18, 1105–1122.

Over, M. (2009, Blog Entry), 'The "Who, What, Where, When, How and How Much" of the International Health Aid – But Not the "Why"', Centre for Global Development Blog Entry. Available at http://www.cgdev.org/blog/%E2%80%9Cwho-what-where-when-how-and-how-much%E2%80%9D-international-health-aid-%E2%80%93-not-%E2%80%9Cwhy%E2%80%9D

Pack, H. and J. R. Pack (1990), 'Is Foreign Aid Fungible? The Case of Indonesia', *Economic Journal*, 100(399), 188-194.

Pack, H. and J. R. Pack (1993), 'Foreign Aid and the Question of Fungibility', *Review of Economics and Statistics,* 75(2), 258-265.

Pack, H. and J. R. Pack (1999), 'Foreign Aid and Fiscal Stress'. In G. R. Saxonhouse and T. Srinivasan (Eds.) *Development, Duality, and the International Economic Regime: Essays in Honors of Gustav Ranis,* 452-476. Ann Arbor Michigan: Michigan Press.

Potthoff, R. F., G.E. Tudor, K. S. Pieper and V. Hasselblad (2006), 'Can One Assess Whether Missing Data are Missing at Random in Medical Studies?', *Statistical Methods in Medical Research,* 15(3), 213-234.

Radelet, S. (2006), ‘A Primer of Foreign Aid’, *Center for Global Development Working Paper* No. 92.

Rahbek, A., E. Hansen and J. G. Dennis (2002), ‘ARCH Innovations and their Impact on Cointegration Rank Testing’, *Preprint no. 12, Department of Theoretical Statistics, Working Paper* no. 22, Centre for Analytical Finance.

Rajan, R. G. and A. Subramaniam (2008), ‘Aid and Growth: What Does the Cross-Country Evidence Really Show?’, *The Review of Economics and Statistics,* 90(4), 643-665.

Ravishankar, N., Gubbins, P., Cooley, R. J., Leach-Kemon, K., Michaud, C. M., Jamison, D. T., & Murray, C. J. (2009), ‘Financing of Global Health: Tracking Development Assistance for Health from 1990 to 2007’, *The Lancet*, 373, 2113–2124. Webappendix available at: http://www.sciencedirect.com/science/article/pii/S0140673609608813#appd001

Roodman, D. (2006), ‘How to Do xtabond2: An Introduction to “Difference” and “System” GMM in Stata’, Centre for Global Development Working Paper No. 103. Available at: http://www.cgdev.org/sites/default/files/11619_file_HowtoDoxtabond8_with_foreword_0.pdf

Roodman, D. (2007), ‘The Anarchy of Numbers: Aid, Development and Cross-Country Empirics’, *World Bank Economic Review*, 21(2), 255-277.

Roodman, D. (2009a), ‘How to Do xtabond2: An Introduction to Difference and System GMM in Stata’, *Stata Journal*, 9(1), 86-136.

Roodman, D. (2009b), ‘A Note on the Theme of Too Many Instruments’, *Oxford Bulletin of Economics and Statistics*, 71(1), 135-158.

Roodman, D. (2010, blog entry), ‘Squishy Findings on Ad Fungibility’, Centre for Global Development Blog Entry. Available at: http://www.cgdev.org/blog/squishy-findings-aid-fungibility

Roodman, D. (2012), ‘Doubts about the Evidence that Foreign Aid for Health is Displaced into Non-health Uses’. *The Lancet,* 380, 972-973.

Roodman, D. (2012, blog entry), ‘Cross-post: The Aid Fungibility Debate and Medical Journal Peer Review’, Centre for Global Development Blog Entry. Available at: http://www.cgdev.org/blog/cross-post-aid-fungibility-debate-and-medical-journal-peer-review

Royston, P. (2007), ‘Multiple Imputation of Missing Values: Further Update of Ice, with Emphasis on Interval Censoring’, *Stata Journal*, 7(4), 445-464.

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley.

Rubin, D. B. (1996), ‘Multiple Imputation After 18+ Years’, *Journal of the American Statistical Association,* 91(434), 473-489.

Rubin, D. B. and N. Schenker (1986), ‘Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse’, *Journal of the American Statistical Association*, 81(394), 366-374.

Rutasitara, L. (2002), ‘Exchange Rate Regimes and Inflation in Tanzania’, *African Economic Research Consortium Research Paper* 138.

Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data.* London: Chapman and Hall.

Schafer, J. L. (1999), ‘Multiple Imputation: A Primer’, *Statistical Methods in Medical Research,* 8, 2-15.

Schafer, J. L. and J. W. Graham (2002), ‘Missing Data: Our View of the State of the Art’, *Psychological Methods,* 7, 147-177.

Schafer, J. L. and M. K. Olsen (1998), ‘Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst’s Perspective’, *Multivariate Behavioural Research,* 33, 545-571.

STATA (2009), *Stata Longitudinal Data/Panel-Data Reference Manual: Release 11.* College Station, TX: Stata Press.

Sundberg, M. and A. Gelb (2006), ‘Making Aid Work’, *Finance and Development*, 43(4), 14-17.

Temple, J. (2010), 'Aid and Conditionality', in D. Rodrik and M. Rosenzweig (Eds.), *Handbook of Development Economics,* Chapter 67(5), 4415-4523. Amsterdam: North Holland.

United Nations Department of Economics and Social Affairs, Population Division. World Population Prospects: The 2006 Revision. 2007. CD-ROM Edition-Extended Dataset in Excel and ASCII formats (United Nations publication, Sales No. E.07.XIII.7)

UNAIDS, WHO. HIV data. Geneva: UNAIDS, 2009. http://www.unaids.org/en/KnowledgeCentre/HIVData/.

Van de Sijpe, N. (2013), 'Is Foreign Aid Fungible? Evidence from the Education and Health Sectors', *World Bank Economic Review,* 27(2), 320-356.

Van de Sijpe, N. (2013a), 'The Fungibility of Health Aid', *Journal of Development Studies*, 49(12), 1746-1754.

Walz, J. and V. Ramachandran (2011), 'Brave New World: A Literature Review of Emerging Donors and Changing Nature of Foreign Assistance', *Center for Global Development Working* Paper No. 273.

WHO. National Health Accounts. Geneva: World Health Organization, 2009. http://www.who.int/nha/country/en/.

World Bank, (1998), *Assessing Aid: What Works, What Doesn't and Why*. New York: Oxford University Press.

World Development Indicators Online Database. Washington, DC: The World Bank, 2009. http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/0,,contentMDK:20398986~menuPK:64133163~pagePK:64133150~piPK:64133175~theSitePK:239419,00.html.

Yontcheva, B. and N. Masud (2005), *Does Foreign Aid Reduce Poverty? Empirical Evidence from Nongovernmental and Bilateral Aid.* Washington DC: International Monetary Fund.

Zimmermann, F. and K. Smith (2011), 'More Actors, More Money, More Ideas for International Development Cooperation, *Journal of International Development,* 23, 722-738.