WORKING PAPERS

What We Miss By Missing Data: Aid Effectiveness Revisited

Anja Breitwieser
Katharina Wick

March 2013

Working Paper No: 1302

DEPARTMENT OF ECONOMICS

UNIVERSITY OF VIENNA

What We Miss By Missing Data: Aid Effectiveness Revisited

Anja Breitwieser* Katharina Wick†

March 2013

Abstract

Missing data is a major problem in empirical development economics, as it may entail efficiency losses as well as biased results. This is an issue within the literature that investigates the effect of foreign aid on welfare. Using multiple imputation techniques, we address these problems and find lower aid effectiveness than previous studies suggest. In addition, imputation allows for comparison of different welfare indicators within the same framework. We find that if aid effectiveness is evaluated based on such indicators, the respective indicator choice can matter for the results.

JEL Classification: F35, H41, I15, I25, O10, O11.
Keywords: Aid Effectiveness, Welfare Indicators, Sample Selection, Missing Data, Multiple Imputation Techniques

*University of Vienna, Department of Development Studies, Sensengasse 3/2/2, 1090 Vienna, Austria. E-mail:anja.breitwieser@univie.ac.at

†University of Vienna, Department of Economics, Brünnerstrasse 72, 1210 Vienna, Austria. E-mail:katharina.wick@univie.ac.at.

1 Introduction

Despite increasingly large aid flows of the international community to developing countries over the last couple of years (IMF (2007)), observed effectiveness of these flows has been notoriously low. While there was - and still is - a strong focus on economic growth as indicator of effectiveness (Burnside and Dollar (2000), Hansen and Tarp (2001), Clemens et al. (2004) and Dalgaard et al. (2004)), lately a shift towards alternative indicators has set in. Arguably, the link between economic growth and foreign aid is rather remote, resting on several links in between, as for instance the effect of aid on education, health and infrastructure. Education, health and infrastructure, however, do not only matter for a country's productivity but are direct indicators of its residents' well-being.

This broader notion of aid effectiveness is supported by the definition of Official Development Aid (ODA) - which identifies "the promotion of economic development and welfare of developing countries" (OECD (2003)) as the main objective of foreign aid - and is reinforced by the UN Millennium Development Goals (MDGs) which specify a wide array of socioeconomic goals of development policy. Therefore an investigation into the effect of foreign aid on welfare indicators other than growth is essential.

A limited number of studies have investigated the effect of foreign aid on welfare indicators. Dietrich (2011), Mishra and Newhouse (2009), and Gauri and Khaleghian (2002) analyze the effect of aid on infant mortality and immunization rates (DPT, measles and hepatitis B) respectively. With respect to the education sector, Michaelowa (2004), Dreher et al. (2006), Michaelowa and Weber (2007a), Michaelowa and Weber (2007b), and Christensen et al. (2010) investigate the effect of foreign aid on net and gross enrollment rates (primary, secondary, and tertiary), and primary completion rates. While these studies are a most welcome diversion from the focus on economic growth, the evidence they provide is scattered, as these analyses vary in many respects, including time and country coverage, choice of indicator, aid measure, and estimation procedures. To contextualize these findings and derive more general results, we analyze the effect of foreign aid on a large variety of welfare indicators - measuring progress in different areas such as health, education and infrastructure - within the same framework.

One of the main problems in empirical development economics is missing data. The most common approach to address this issue is listwise deletion, i.e. in a panel setting dropping all country-year pairs for which any variable included in the estimated equation is missing. Using listwise deletion, a relatively small fraction of missing data, however, can results in considerable losses of observations (Ross (2006))[1]. If the observed data is a random sub-sample of the unobservable complete data, this reduction in sample size causes efficiency losses in the analysis. If the missingness in one variable, can be explained by other variables, the consequences are even more profound. For instance, if countries with higher income are more likely to report data, or authoritarian regimes are less likely to do so (Bueno de Mesquita et al. (2003), Hollyer et al. (2011)), the observed sample cannot generally be considered a random sub-sample of the unobservable complete sample and observed case analysis[2] can be biased. In this case, ignoring the missing data pattern entails ignoring sample selection bias.

An alternative way to deal with missing data is to use multiple imputation techniques. Multiple imputation allows to base our analysis on all potential observations and hence can handle efficiency losses and biased results.

Although the method of multiple imputation has not been employed much in economics, other sciences (such as political science (Honaker and King (2010), Gelman et al. (1999)) or medicine (Lee and Carlin (2010), Raghunathan (2004)) widely resort to this method. We argue that this approach warrants some attention also in empirical (development) economics. Since both methods, listwise deletion as well as imputation, have their downside and may be legitimately criticized, it is a decision between two second-best options. We argue that - depending on the research question - multiple imputation may be an expedient alternative to listwise deletion. In the course of our analysis, we carefully inspect the imputation results to convince the reader and ourselves that our imputations yield reasonable complete datasets.

Our analysis contributes to the existing literature in two ways. First, by taking account of the missing data pattern and applying multiple imputation techniques, we are able to address the considerable reduction in sample size and potential bias, which can be associated with missing data.

Second, since within a certain sector various different indicators exist, the choice of

[1]Ross (2006) for example reports for his study a fraction of 18 percent of the data missing, which subsequently results in the loss of three quarters of observations.

[2]The missing data literature uses the term 'complete case analysis' (Graham (2009)) instead of 'observed case analysis'. We refer to the potential unobservable dataset without any missing data as complete. To avoid confusion, we thus refer to an analysis based only on the observed data as observed case analysis. Observed case analysis thus refers to deleting any observations from the dataset which have missing values on any of the variables (listwise deletion) and then proceeding with conventional statistical methods of analysis (Allison (2001)).

the respective indicator is non-trivial. Empirical studies, for instance, on aid effectiveness in the health sector so far considered inter alia DPT immunization (Dietrich (2011)), child mortality rates (Mishra and Newhouse (2009)) or life expectancy (Wilson (2011)) as indicators of effectiveness. Observed outcomes however, may not be independent from indicator choices. Differing results could be attributed to differences in the particular samples. Typically, this issue is addressed by using a common sample for the analysis. However, analyzing numerous welfare indicators for developing countries in a common sample can reduce the sample size considerably, such that the analysis becomes highly inefficient or infeasible. Resorting to our imputed (and thus 'complete') data allows comparisons between indicators.

Our results indicate that great care has to be taken when analyzing welfare indicators to establish results on aid effectiveness. We find that results of such an analysis (1) may be biased due to sample selection and (2) cannot be used to draw general conclusions about aid's effectiveness. We show that even indicators that arguably capture welfare in the same broad category (such as enrollment rates and the pupil-teacher ratio that may serve as a proxy for the education sector), may suggest qualitatively different results.

The paper is structured as follows. In section 2 we analyze the missing data pattern and argue for the expedience of imputation techniques in the context of our research question. Section 3 describes the multiple imputation approach and its results. In addition we present our estimation method. Section 4 presents the data. The results are presented in section 5. Section 6 concludes.

2 Missing Data

The maximum number of observations (N), that is, the number of country-year pairs for which aggregate foreign aid inflows have been reported to the OECD between 1970 and 2009, is 5274.[3] However, an analysis of the effect of these aid flows on welfare indicators would be based on a considerably smaller sample due to missing observations.

Rubin (1976) develops a framework for the different types of missing data patterns. In this framework, missingness is classified into missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). For any data set a ma-

[3]Since previous studies have predominantly used sector aid measures, their potential sample size is mostly lower.

trix R can be defined such that it identifies the observed and the missing observations within the data. That is, R contains the value zero for missing and one for observed values. R can be considered as a combination of random variables with a joint probability distribution. MCAR implies that the probability of missingness neither depends on the observed data nor on the unobserved parts of the data. That is, $P(R|Y^c) = P(R)$, where Y^c denotes the complete data. MAR is less restrictive and allows the distribution to depend on the observed data, such that: $P(R|Y^c) = P(R|Y^o)$, where Y^o denotes observed parts of the data[4]. In other words, conditional on the observed data the probability of missingness does not depend on the missing data. The missing data is considered missing not at random (MNAR) if the condition $P(R|Y^c) = P(R|Y^o)$ does not hold. That is, if conditional on Y^o, the probability of missingness does depend on the missing data (Y^m) (Schafer and Graham (2002)).

If the data is MCAR, that is the observed sample is a random sub-sample of the unobserved complete sample, observed case analysis is subject to efficiency losses but unbiased. In the context of our analysis this assumption, however, is rather strong. Countries with higher income have more monetary and human resources available for data collection and are thus less likely to have missing data (Bueno de Mesquita et al. (2003)). Dictatorships on the other hand may be less willing to collect and report data on certain variables (Bueno de Mesquita et al. (2003), Hollyer et al. (2011)). In the context of foreign assistance, outside agencies (bilateral or multilateral donors) may compel governments to regularly report certain data. Since the deterrent effect of non-compliance increases with the flow of foreign aid potentially withdrawn, missingness would in that case be a function of the amount of aid received. These examples support the notion that the observed data (eg. on income, political regime, etc.) influences missingness and that the MCAR assumption is too strong. Testing our data we indeed find that income, form of government, civil liberties and other variables significantly predict missing observations. In this case, however, results from complete case analysis, cannot generally be assumed to be unbiased.

Eliminating MNAR from the set of possible missing data mechanisms is less straight forward since testing is not possible in this case. MNAR implies that the pattern of missingness depends on the values of the variable under consideration itself. One could argue that for example school enrollment rates are more likely to be missing for lower enrollment rates as governments might be more reluctant to collect or report these numbers. Yet, governments are usually not exempt from internal or external public pressure, which limits their autonomy of decision. In a democratic environment for

[4] That is, the complete data Y^c is partitioned, such that Y^c=(Y^o,Y^m), where Y^m denotes the missing parts of the data.

instance, citizens could pressure their governments into collecting the respective data and hence make missingness a function of a country's regime type. Similarly, in particular when considering aid recipient countries, external pressure for data collection might be exerted by bilateral and multilateral donors. Inclusion of auxiliary variables (eg. civil liberties, freedom of press, etc.) in the imputation model can thus support the MAR assumption.

Table 1 reports the median, the number of observations and the number of countries for each dependent variable for two different samples[5]. For each welfare indicator used in the subsequent analysis, countries are split into two samples with respect to their degree of missing data. Table 1 reports on the left side summary statistics for countries with many missing values[6]. On the right hand side summary statistics for countries with few missing observations are reported.

Consider for example the variable 'gross enrollment in primary and secondary education'. Aid recipients with more missing observations than the average country have a sample median for gross primary and secondary enrollment of 64.82%. Yet, for the countries with high data coverage, the enrollment rate is considerably higher (76.99 %). In the following row, summary statistics for our aid measure are reported for the two gross primary and secondary enrollment samples. Since the aid measure is not subject to missing observations, the number of observations increases in both samples. While in the low-observation sample a recipient country received on average 91.41 US$ in PPP per capita per year, a country in the high-observation sample received on average only 68.64 US$ in PPP per capita per year. Similar patterns are found for HIV prevalence, the tuberculosis detection rate, the female to male primary enrollment ratio, the number of internet users, and the electricity generating capacity.

In other words, for some of our dependent variables countries with high rates of missing observations show worse welfare outcomes and at the same time large per capita aid flows. This could bias the aid coefficient upwards. Assuming that a country's missing data is - if it were observable - similar in magnitude to the observed data, the missing data would have both: low values for the welfare indicators and high values for the amount of foreign aid received. Multiple imputation - which relies on considerably

[5]Note that Table 1 does not indicate, that the pattern of missingness depends on the value of the variable itself (i.e. that it is MNAR), since the summary statistics do not account for the conditional effects of the observed data (eg. income, political regime, etc.).

[6]That is, countries with a higher rate of missingness than the median rate of missingness for the respective variable.

Table 1: Missing data pattern

	few observations			many observations		
	median	no. of obs.	countries	median	no. of obs.	countries
DPT	68.319	3195	114	88.34	741	38
Aid	76.435	4461		106.60	813	
Infant mortality*	59.230	453	16	63.12	4669	136
Aid	158.668	605		71.03	4669	
Tuberc.	61.238	2152	118	76.99	583	34
Aid	84.733	4603		56.06	671	
HIV	2.492	1791	121	1.15	533	31
Aid	84.546	4655		55.06	619	
Life expect. fem*	75.809	33	4	62.62	5144	148
Aid	314.839	130		75.18	5144	
Female/Male enrol	86.224	1661	76	88.91	2332	76
Aid	94.936	2770		65.76	2504	
Enrol. prim.	94.712	1798	77	91.31	2474	75
Aid	100.749	2682		60.74	2592	
Enrol. prim. & sec.	64.816	1469	78	76.99	2117	74
Aid	91.416	2901		68.46	2373	
Completion prim.	75.384	907	77	68.29	1971	75
Aid	90.063	2804		70.89	2470	
Pupil/Teacher	32.039	1162	76	33.55	2347	76
Aid	107.093	2603		55.74	2671	
Phone	6.440	3897	116	14.86	717	36
Aid	80.267	4544		86.18	730	
Internet	4.320	1110	78	6.81	1313	74
Aid	97.557	3039		58.69	2235	
EGC	4847.176	3612	134	12659.92	405	18
Aid	79.679	4833		96.50	441	

Samples split by the median amount of missingness in the respective welfare indicator.
* Due to the low percentage of missing observations the samples is split by whether or not the respective welfare indicator has any missing observations for the respective country.

more information than this rather stylized argument - re-introduces these observation into the analysis. Therefore one would expect the effect of foreign aid to decrease.

3 Methodology

3.1 Multiple Imputation

Summary statistics of all variables included in the imputation model are reported in the Appendix (table 5 and table 4). Data on the total population, the share of the urban population, and whether a recipient country ever was a colony or shares a common language with important donors are available for all 5274 observations. All other variables have missing values. The amount of missing observations varies considerably by variable. While for the variables real GDP per capita, infant mortality rate, and life expectancy the rate of missingness is less than 0.01, for the bulk of our welfare indicators data coverage is lower. Among our variables HIV prevalence has the lowest data coverage, with almost 56% of its observations missing, predominantly due to early observations missing[7].

Since the variables included in the imputation model consist of continuous, categorical, and binary variables and many of them are subject to logical constraints (e.g. ratios may not be greater then one) we choose a variable-by-variable imputation approach, namely the chained equation approach (also called fully conditional specification, FCS van Buuren (2007)). To reflect the uncertainty about the imputed values multiple complete data sets are generated. Single imputation methods dilute the uncertainty of the imputed values by treating the imputed values as if they were observed. Multiple imputation methods use the variance between the imputed datasets to account for this uncertainty. To take recent indications into account that imputing three to five complete data sets (a number suggested by earlier works) may not suffice (Graham et al. (2007)), while at the same time keeping computational costs at reasonable levels we generate twenty complete data sets. The analysis is then performed individually on the imputed datasets and results are obtained by combining them into a single set of estimates according to Rubin's rule (Rubin (1987)). That is, for each imputed dataset estimated coefficients and respective standard errors are obtained. Let $\hat{Q}_j$ be the estimated (aid) coefficient form dataset j, where $j = 1, ..., m$ and U_j the respective

[7]While imputing values for HIV prevalence in the 1970ties might seem unusual, note that the aim of missing data procedures is not to accurately predict the missing values but rather to preserve the relations in the data (Schafer and Graham, 2002).

standard error. The overall point estimate (eg. for foreign aid) from the imputed datasets is then calculated as the arithmetic mean of the individual estimates. That is (Schafer,1997):

$$\bar{Q} = \frac{1}{m}\sum_{j=1}^{m}\hat{Q}_j \tag{1}$$

For the overall standard error two sources of uncertainty have to be considered. First, the standard errors of the m=20 individual imputations U_j, where $j = 1, ..., m$ (within-imputation variance) and second the uncertainty associated with imputed values, captured in the variation in the individual estimated coefficients $\hat{Q}_j$ (between-imputation variance; B).

$$\bar{U} = \frac{1}{m}\sum_{j=1}^{m}\hat{U}_j \tag{2}$$

$$B = \frac{1}{m-1}\sum_{j=1}^{m}(\hat{Q}_j - \bar{Q})^2 \tag{3}$$

The overall standard error is then obtained by:

$$\sqrt{T} = \bar{U} + (1 + \frac{1}{m})B \tag{4}$$

Thus, while multiple imputation can increase efficiency by basing the analysis on a larger sample, high uncertainty in the imputed values - that is, large differences in the estimated coefficients from the individual imputations - affects the overall standard errors in the opposite direction.

Multiple imputation using a fully conditional specification can be described as follows: Let $X = (X_1, ..., X_l)$ be the set of completely observed variables, $Y = (Y_1, ..., Y_k)$ be the set of incomplete variables, where Y_k^m denotes missing observations and Y_k^o observed observations, and $R = (R_1, ..., R_k)$ be a set of dummy variables indicating missing (R=0) and observed values (R=1). Instead of assuming the multivariate distribution to be multivariate normal - as in the multivariate normal imputation approach - the distribution $P(Y, X, R|\theta)$ is obtained from the conditional densities $P(Y_i|X, Y_{-i}, R, \theta)$ for each Y_i from the individual variables, where θ depicts the model parameters and Y_{-i} is the set of all variables with missing observations except for Y_i. After starting from simple guesses, imputed values are then obtained from iteration over these individually specified imputation models. In particular, for each variable an individual regression is fit to draw random values from the respective predictive distribution. Since this approach uses iterations over a sequence of univariate models to impute for each Y_i^m, different univariate models can be used for each variable (e.g. logistic regression for binary variables)(van Buuren (2007); van Buuren et al. (2006); Gelman (2004)).

Variables that do not contain any missing observations are included as predictor variables in the imputation model. Missing observations for non-permanent United Nations Security Council membership[8] are imputed using a logistic regression. Missing observations for the variables regime type from Cheibub et al. (2010), civil liberties, political rights, and the polity index from the Correlates of War Project (COW (2003)) are imputed using an ordered logistic regression. The rest of the variables with missing data is imputed by predictive mean matching.

Predictive mean matching is closely related to ordinary linear regression with the only exception that in the final step, the imputations for the missing values are sampled from the set of observed values of the respective variable. Predictive mean matching is of avail for continuous variables which do not meet the normality assumption or in order to control for potential nonlinear relationships (White et al. (2011)). While deviation from the normality assumption is an issue, we chose predictive mean matching for the latter reason in particular.

To account for the panel data structure in the imputation model, Graham (2009)) suggests to include indicator variables for each cluster or separate imputation by clusters. In order to avoid overburdening the model with more than 100 indicator variables, the imputations are run separately by regions, where for each region the respective recipient dummies are included.[9]

3.1.1 Imputation results

The aim of missing data procedures like multiple imputation is not to accurately predict the missing values but rather to preserve the relations in the data. Hence, the

[8]Missing observations for non-permanent United Nations Security Council Membership data mainly stem from political transition (e.g. independence or secession) processes. Countries with missing values for at least one time period include: Angola, Antigua and Barbuda, Bahamas, Bahrain, Belize, Bhutan, Botswana, China, Comoros, Cyprus, Djibouti, Ethiopia, Fiji, Gambia, Guyana, Lesotho, Maldives, Mauritius, Mozambique, Namibia, Papua New Guinea, Russia, Samoa, Sao Tome & Principe, Serbia, Seychelles, Singapore, Solomon Islands, Suriname, Swaziland, United Arab Emirates, Viet Nam, and Yemen.

[9]The regions in our analysis are Europe and Central Asia; Latin America and the Caribbean; sub-Saharan Africa; East Asia, the Pacific and South Asia; Middle East, North and West Africa.

performance of the imputation model can only be evaluated in the context of the respective analysis model. If the occurrence of missing data is beyond the control of the researcher - which is the case in most empirical applications - assumptions about the pattern of missingness have to be made. These assumptions have to be made irrespective of the choice of missing data procedures. That is, it applies to multiple imputations techniques as well as to listwise deletion. Some of these assumptions, however, are inherently non-testable (eg. distinguishing MAR from MNAR).

Diagnostics for the imputed data are available, since multiple imputation models are fitted to observed data. In particular, with a chained equation imputation approach, diagnostics can check the respective model fit with conventional methods. Fitting the individual imputation equations (for each imputed variable) to our observed data and plotting the fitted values, we generally do not observe any alarming patterns. In addition to examining the model fit, we argue for the plausibility of the imputed values by comparing densities as well as mean and standard deviation of the observed, imputed and complete data. While this obviously is not a formal hypothesis test, inspection of the distributions and summary statistics can flag potential problems in the imputation model. It should be noted though that under a MAR assumption, differences with respect to the distribution of the imputed and observed values are not necessarily problematic. In fact, some differences are expected[10] (Abayomi et al. (2008), Raghunathan and Bondarenko (2007)). Consider for example the distribution of the observed, imputed and complete data for the 'gross primary and secondary enrollment ratios' (Figure 1). As discussed in section 2 we expect predominantly bad outcomes to be missing. This notion is supported by the fact that the imputed values lie to the left of the observed values. Observed, imputed and complete data distributions for the other welfare indicators are shown in Figure 2 in the appendix.

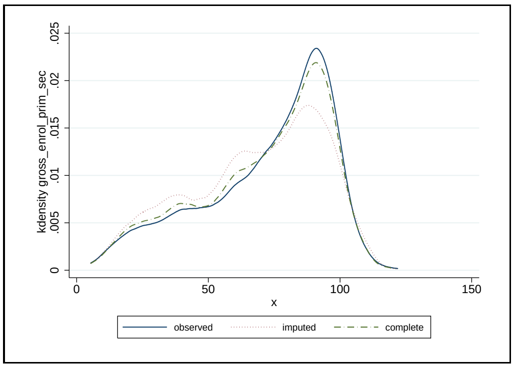

Figure 1: Kernel densitiy

[10]This would not be the case under a MCAR assumption.

Imputations by a chained equation approach are obtained by iteration over individually specified imputation models. Since this is an iterative approach, convergence should be monitored. We examine convergence via plots of the iteration numbers against the mean and the standard deviation of our imputed values (StataCorp (2011)). We do not find any evident long-term trends. That is, we find no indication against convergence.

3.2 Endogeneity and IV estimation

The possibility of endogeneity in aid-growth regressions is a prominent issue in the aid effectiveness literature since aid flows can hardly be regarded as exogenous to the income level of the recipient country and hence to growth. Similarly, if aid is given for philanthropic and developmental motives, countries that score low in terms of our welfare indicators will receive more aid. This negative relationship can bias the estimated coefficient for foreign aid downwards. In addition, some of the ambiguity of the direction of causation might be driven by third variables like natural disasters. Earthquakes for example involve both a sudden negative shock to the welfare indicators as well as a spike in (humanitarian) aid donations. In order to mitigate this source of endogeneity we exclude humanitarian aid and food aid from our foreign aid measure. A bias in the opposite direction would also be possible. A country with insufficient institutional capacity may receive less aid flows (due to its inability to conform with donors' formal requirements) and experience difficulties in public service delivery and hence would not perform well with respect to our welfare indicators. Including country-specific fixed effects reduces this bias as institutions arguably change very slowly over time.

To address endogeneity we use instrumental variable estimation. Several instruments for foreign aid have been suggested in the aid-growth literature so far. Assuming that donor countries give more aid to former colonies than to any other random recipient country, 'former colonial ties' (Djankov et al. (2008)) and common language (Rajan and Subramanian (2008)) have been suggested as exogenous instruments. These instruments however have in common, that they are time invariant and thus are infeasible in the presence of country fixed-effects. Alesina and Dollar (2000) suggest using data on political contiguousness between recipient and donor countries. In particular, they propose that correlation of votes in the UN general assembly serves as a good proxy for common political interest of any two countries. In this vein, we use data on the correlation of votes[11] between recipient countries and important donor countries and

[11] We use two types of variables here. The first one is calculated on the basis of all votes (ALL) in the UN general assembly. The second one recognizes the differing degree of importance with respect to the topic of the poll and is only based on keyvotes (KEY). Keyvotes are classified according to

groups (Canada, France, Germany, Italy, Japan, UK, USA, and the G7 group) from Dreher and Sturm (2012). Since a recipient's strategic importance is particularly high when it is a member of the UN security council (UNSC) we in addition use UNSC membership (Dreher et al. (2009)) to instrument for aid.

Estimates are obtained from a 2SLS estimator with country fixed effects. That is, the estimated results only take the variation within a country into account. However, since taking all country characteristics explicitly into account is impossible, we use fixed effects to capture this heterogeneity. To control for global shocks and trends (i.e. changes in the cost and effectiveness of health interventions for example with respect to immunization between 1960 and 2009) time dummies are included.

4 Data

4.1 Dependent Variables

Since we are interested in the effect of sample selection, which arguably affects different indicators in different ways as well as the effect of indicator choice, we use a wide array of different welfare indicators as dependent variables. Unless stated otherwise, the data is obtained from the World Development Indicators (IBRD (2011)).

For the effect of foreign aid on health, data on immunization rates, infant mortality rate, prevalence of diseases (HIV and tuberculosis) and female life expectancy are included in the analysis. The variable DPT immunization (DPT Immun.) refers to the percentage of children aged between 12 and 23 months that have been adequately vaccinated (i.e. have received the doses necessary to achieve full immunization) against DPT. Data is available from the World Health Organization and UNICEF (WHO (2011)). HIV prevalence (HIV) captures the share of a country's population between 15 and 49 years of age who is infected with HIV. Tuberculosis detection rate (Tuberc.) refers to the percentage of new tuberculosis cases and relapses to estimated incident cases. The dependent variable female life expectancy (Life expect. fem.) indicates the number of years a female newborn would live given the patterns of mortality at the time of the newborn's birth prevail throughout her life. The infant mortality rate (infant mortality) is the number of infants dying under the age of one, in a given year per 1,000 live births.

Kilby (2009).

For the education sector we consider gross school enrollment rates, completion rates, the pupil-teacher ratio, and female to male enrollment ratios. The variables gross school enrollment primary (Enrol. prim.) and gross school enrollment primary and secondary (Enrol. prim. & sec.) are the total enrollment rates in primary, and primary and secondary education irrespective of the students age. Hence, the enrollment rates can exceed 100% since over-aged and under-aged children are included. The primary completion (Completion prim.) rate is the gross intake rate to the last grade of primary, that is the number of new students in the last grade of primary school to the number of students who are theoretically at the entrance age of the last grade of primary education. The pupil to teacher ratio (Pupil/Teacher) refers to the number of students enrolled in primary education to the respective number of teachers. The female to male primary enrollment (Female/Male enrol) ratio is calculated from the female primary enrollment ratios over the male primary enrollment ratios. Hence, an indicator equal to one denotes parity between the two groups with respect to primary enrollment rates. A value above one indicates disparity in favor of girls and a value below one indicates disparity in favor of boys.

The data on the stock of physical infrastructure and communication includes electricity generating capacity (EGC) measured as the number of kWh available per capita, the fixed telephone lines per 100 people (Phone), and the number of internet users per 100 people (Internet). Whereas the last two are obtained from the World Development Indicators (IBRD (2011)), the former is from Canning (1998). Though the original data from Canning ends in 1998, the data series has been extended substantially by using information from the World Development Indicators, so that the observed data series covers the time span between 1970 and 2005. 'Electricity generating' capacity and 'telephone lines' are estimated in first-differences to account for high autocorrelation in these variables.

4.2 Independent Variables

International organizations like the OECD and the World Bank[12] collect data on aid flows. Net official development assistance (net ODA), a frequently used measure of aid flows provided by the Development Assistance Committee (DAC), includes total grants as well as concessional loans[13] minus principal repayments. While most empir-

[12]The World Bank Debt Reporting System (DRS) comprises data of official loans from bilateral and multilateral donors to developing countries. However, not all of the data is publicly available.

[13]A loan is booked as ODA if its grant element is at least 25%.

ical studies use sector aid (Dietrich (2011), Michaelowa and Weber (2007b), Mishra and Newhouse (2009)) in their analysis, we prefer an aggregate aid measure. Data on sector aid *commitments* are generally available, but *disbursed* sector aid data is available only from the 1990 onwards. Since commitments and disbursements often differ substantially and it is actual *disbursements* we are interested in, using sector aid measures would considerably reduce our sample size.
In addition we control for food aid and humanitarian aid, which is mainly intended for short term relief. Hence, foreign aid flows (Aid) are measured as: net ODA minus food and humanitarian aid. The aid measure is available between 1960 and 2009, and calculated on per capita basis.

Following the empirical literature (Deacon (2009), Alesina et al. (1999), Saiz (2002)), we control for the income level of a country in logs (GDP), its regime type (DEMOC), total population in logs (POP) and the percentage share of the urban population (Urban). Data on real GDP in constant 2009 US Dollars is originally taken from the Penn World Table (Heston et al. (2011)) and has been deflated using the US GDP deflator (NIPA (2011)), to align its base year to the base year of our aid measure. In addition, GDP is corrected for purchasing power parity. The indicator variable for regime type takes the value one if the country has been classified as a democracy by Cheibub et al. (2010) and zero otherwise. Data on total population and urban population has been obtained from the World Development Indicators (IBRD (2011)). Even though most variables are reported on a yearly basis we use five year averages in our analysis to avoid influences from business cycle fluctuations.

5 Empirical Results

Our results from instrumental variable estimation for the observed as well as complete data in 5-year averages are summarized in Table 2. In the first column from the left, estimated aid coefficients and standard errors based on the observed data set are reported. The respective test statistics for the relevance (underidentification test) and validity (Hansen test) of our instruments for foreign aid as well as income are reported at the far right. Except for the female to male primary enrollment ratio, the test statistics support the choice of instruments. Due to the low p-value for the Hansen test, the results for the female to male enrollment ratio should not be interpreted. In the third column from the left, estimated aid coefficients and respective p-values using the (multiple imputed) complete data are reported.

First, consider the results using the observed data (column 1). While for the majority of indicators the estimated aid coefficient is insignificant, for some indicators in the health sector a significant positive effect of foreign aid is found. In particular, we find that foreign aid has a significant positive effect on DPT immunization rates and a significant negative effect on HIV prevalence.

Table 2: Fixed Effects IV Estimates, 5-year averages

	Aid, observed	Aid, complete	N, obs.	N, comp.	Groups, obs.	Groups, comp.	Underid, p-val	Hansen, p-val
DPT^A	9.088** (3.723)	-9.124 (6.257)	743	1029	144	152	0.009	0.096
Infant mortality^A	1.012 (4.455)	1.772 (5.028)	936	1029	145	152	0.019	0.469
Tuberc.^{B}	6.655 (9.684)	0.313 (10.302)	531	1029	144	152	0.038	0.279
HIV^B	-2.137** (0.888)	1.022 (0.729)	438	1029	118	152	0.009	0.338
Life expect. fem^B	0.074 (1.062)	-1.270 (1.423)	948	1029	145	152	0.003	0.356
Completion prim.^{A}	1.925 (3.411)	0.700 (2.563)	597	1029	133	152	0.016	0.200
Pupil/Teacher^B	3.901 (2.507)	4.504* (2.606)	816	1029	141	152	0.082	0.776
Female/Male enrol.^{B}	6.217** (3.122)	5.722 (3.717)	874	1029	141	152	0.017	0.027
Enrol. prim.^{B}	5.415 (4.679)	5.416 (4.750)	890	1029	141	152	0.015	0.938
Enrol. prim. & sec.^{B}	0.853 (2.878)	3.196 (2.938)	831	1029	140	152	0.019	0.633
Internet^B	-0.119 (2.663)	-5.304* (2.876)	558	1029	144	152	0.186	0.104
D.Phone^A	0.119 (0.115)	0.059 (0.117)	754	1029	144	152	0.091	0.344
D.EGC^B	-6.8e-06 (7.7e-06)	-1.9e-05 (2.4e-05)	802	1029	142	152	0.009	0.449

* $p<0.10$, ** $p<0.05$, *** $p<0.01$
All estimates include the control variables GDP, Democ, POP and Urban as well as time dummies. The estimates are obtained from the user-written Stata command xtivreg2 (Schaffer (2010)). Aid as well as GDP per capita are instrumented using (A) UN general assembly key votes (KEY) and UN security council membership (UNSC) or (B) all UN general assembly votes (ALL) and UN security council membership (UNSC) as exogenous instruments.

Next, consider the results from the multiple imputed data set (column 2). While we found a positive significant effect of aid on DPT immunization in the observed sample, the effect turns insignificant using the complete data. The same applies to the effect of aid on HIV prevalence. In the case of the pupil-teacher ratio, where we found an insignificant effect in the observed sample the estimated aid coefficient from the multiple imputed data set is positive significant[14]. Similarly, for the number of internet users per 100 people, where no effect of foreign aid is found in the observed sample, the estimated coefficient in the complete sample is significant and negative.

For the welfare indicators DPT immunization and HIV prevalence switching from the observed to the complete sample results in a loss of significance. Since the overall standard error comprises the within- and between-imputation variance, this result could be due to considerable differences with respect to the estimated coefficients for the individual imputations (between-imputation variance) or a higher within-imputation variance (see section 3.1). A result based on a high between-imputation variance would be undesirable, in the sense that the loss of significance would be driven by the large uncertainty associated with the imputations. To address this issue, table 9 and 10 in the appendix report the estimation results for the observed, the overall imputed (complete) and the individual imputation datasets for HIV prevalence and DPT immunization respectively. First, consider the variable DPT immunization. The overall complete standard error is larger than the standard error based on the observed data. However, since the majority of the individual standard errors are higher as well the increase in the variance can be explained by an increase in the within-imputation variance. With respect to HIV prevalence, the overall complete standard error is lower than the observed standard error and so are most of the individual imputation standard errors. In this case the loss in significance seems to be associated with the decrease in the size of the estimated coefficients and not an increase in the between-imputation variance.

Summarizing, in all those instances where the estimated coefficients differ in the two data sets, the change implies a decrease in aid effectiveness when using the complete dataset. Thus, we find evidence for a sample selection bias, which - in all cases where present - suggests a more favorable result for the effectivity of aid in improving welfare outcomes.

In addition, the results in column 2 suggest that (independent of sample selection problems) it does matter which particular indicator is chosen for analysis. Considering the pupil-teacher ratio as an indicator of aid effectiveness implies a different result than the gross enrollment rate.

[14] Note that a high-pupil teacher ratio is undesirable.

5.1 Robustness

First, we assess the robustness of our findings with respect to different specifications and estimators. The results are reported[15] in table 3. To facilitate comparison the first column reproduce our main results. In the next column the results from rerunning the estimation on 3 year averages are reported. In column 3 additional controls (government spending and trade to GDP) are added to our basic set of controls. Column 4 reports the results using an alternative measure for regime type. We replace the regime type variable obtained from Cheibub et al. (2010) by an indicator variable ranging from minus ten (autocracy) to plus ten (democracy) obtained from the Correlates of War dataset (2003). To elude that our results are influenced by instrumenting for GDP per capita, we next lag GDP per capita one period and treat it as exogenous. The estimates are reported in column 5. Finally the last column (6) report the results for pooled OLS estimation including time and region dummies.

Our results are robust for the different fixed effects estimations. Focusing on those four dependent variables where we find a different effect for the observed and the complete dataset, we find that, for DPT immunization and HIV prevalence the results are the same in all estimates based on a fixed effects estimator[16]. The results with respect to internet users hold in two out of three FEIV estimates and the results for the pupil/teacher ratio hold in two out of four FEIV estimates and for the pooled OLS estimates [17]. Except for the pupil/teacher ratio in neither of the above variables we find any evidence contrary to our main findings. If the effect of foreign aid is insignificant in the observed and the complete sample (i.e. infant mortality, tuberculosis detection rate, female life expectancy, primary completion rate, female/male enrollment ratio, primary enrollment ratio, primary and secondary enrollment ratio, telephone lines per capita, and electricity generating capacity)), this pattern prevails for the robustness checks, with occasional occurrences of changes in significances supporting our findings. Only in four cases we find evidence contrary to our argument. In neither case these results are robust.

In general, the robustness checks support the notion that observed effectiveness of foreign aid decreases when the analysis is based on the complete data.

[15]For reasons of presentability we only report the estimated coefficients and standard errors for foreign aid in the tables of this section.

[16]Only those results, where the test statistics support the validity and relevance of the instruments are reported in table 3 and considered to assess robustness.

[17]For the other dependent variables the results are less robust to discarding the assumption of country fixed effects. Pooled OLS results, however, are more likely to be subject to omitted variable bias.

Table 3: Robustness

	(1) FEIV, 5-year avg.		(2) FEIV, 3-year avg.		(3) FEIV, add. controls		(4) FEIV, alt. controls		(5) FEIV, lagged GDP, exg.		(6) pooled OLS	
	Aid, observed	Aid, complete	Aid, observed	Aid, complete	Aid, observed	Aid, complete	Aid, observed	Aid, complete	Aid, observed	Aid, complete	Aid, observed	Aid, complete
DPT^A	**9.088****	**-9.124**	-	-	**11.502****	**-9.882***	**7.598***	**-8.896**	**9.445****	**0.020**	-4.794	-3.611
	(3.723)	(6.257)			(5.174)	(5.901)	(4.299)	(6.092)	(3.824)	(3.395)	(4.571)	(3.992)
Infant mortality^A	1.012	1.772	2.895	4.869	5.881	2.126	2.954	1.330	-5.146	-5.164	-2.567	0.442
	(4.455)	(5.028)	(3.591)	(6.025)	(6.563)	(4.696)	(4.690)	(4.842)	(3.700)	(4.136)	(3.675)	(4.393)
Tuberc.^{B}	6.655	0.313	-1.335	8.764	11.978	0.422	1.158	0.622	9.279	8.112	-0.693	-3.324
	(9.684)	(10.302)	(6.421)	(12.461)	(7.859)	(9.839)	(7.260)	(10.342)	(10.780)	(7.853)	(3.464)	(5.167)
HIV^B	**-2.137****	**1.022**	**-1.570*****	**0.780**	**-2.264****	**1.050**	**-2.579****	**1.039**	**-1.657****	**0.023**	0.855	-0.748
	(0.888)	(0.729)	(0.580)	(0.874)	(0.921)	(0.696)	(1.122)	(0.729)	(0.840)	(0.506)	(3.992)	(0.662)
Life expect. fem^B	0.074	-1.270	0.155	-1.196	-1.380	-1.272	-0.286	-1.263	**2.520***	**1.060**	0.620	0.265
	(1.062)	(1.423)	(0.992)	(1.611)	(1.489)	(1.325)	(1.300)	(1.414)	(1.337)	(1.102)	(1.035)	(1.019)
Completion prim.^{A}	1.925	0.700	2.282	0.697	5.164	0.730	4.878	0.485	3.509	1.542	*-13.434***	*-3.268*
	(3.411)	(2.563)	(3.116)	(2.490)	(3.995)	(2.585)	(4.333)	(2.647)	(3.555)	(2.714)	(6.753)	(3.380)
Pupil/Teacher^B	**3.901**	**4.504***	-2.054	3.076	**5.008**	**4.686***	*4.554**	*4.320**	-	-	**0.852**	**5.759****
	(2.507)	(2.606)	(1.856)	(2.987)	(3.124)	(2.473)	(2.435)	(2.506)			(4.523)	(2.669)
Female/Male enrol.^{B}	-	-	-0.158	-1.695	-	-	-	-	-	-	7.571**	0.656
			(1.906)	(3.453)							(3.526)	(2.171)
Enrol. prim.^{B}	5.415	5.416	4.623	4.631	7.246	4.936	4.344	4.910	6.020	4.885	-	-
	(4.679)	(4.750)	(4.168)	(5.678)	(6.577)	(4.500)	(4.927)	(4.624)	(5.230)	(4.077)		
Enrol. prim. & sec.^{B}	0.853	3.196	2.302	2.891	3.758	2.612	1.139	2.923	*-0.502*	*4.064**	2.726	-0.964
	(2.878)	(2.938)	(2.538)	(3.810)	(4.071)	(2.767)	(3.159)	(2.832)	(3.062)	(2.454)	(4.421)	(3.048)
Internet^B	**-0.119**	**-5.304***	0.699	-5.281	-	-	**2.733**	**-4.955***	-	-	-1.854	0.973
	(2.663)	(2.876)	(1.830)	(4.335)			(4.855)	(2.714)			(1.129)	(0.743)
D.Phone^A	0.119	0.059	**0.292*****	**0.152**	-0.025	-0.025	0.117	0.054	**0.328****	**0.210***	-	-
	(0.115)	(0.117)	(0.110)	(0.104)	(0.180)	(0.167)	(0.142)	(0.120)	(0.129)	(0.118)		
D.EGC^B	-6.8e-06	-1.9e-05	8.3e-06	8.8e-06	*-2.1e-05**	*-3.0e-05*	-4.7e-06	-1.8e-05	5.8e-08	-6.8e-06	-4.2e-02	3.3e-06
	(7.7e-06)	(2.4e-05)	(1.2e-05)	(1.9e-05)	(1.2e-05)	(2.8e-05)	(7.4e-06)	(2.3e-05)	(6.8e-06)	(1.5e-05)	(7.4e-02)	(8.2e-06)

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Estimates in column 1, 2, and 5 include the control variables GDP, Democ, POP and Urban as well as time dummies. Estimates in column 3 in addition include trade to GDP and government spending as controls. Estimates reported in column 4 proxies regime type by *polity2* instead of *democ*. The estimates are obtained from the user-written Stata command xtivreg2 (Schaffer (2010)). Aid as well as GDP per capita (except for column 5, where GDP per capita is assumed exogenous) are instrumented using (*A*) UN general assembly key votes (KEY) and UN security council membership(UNSC) or (*B*) all UN general assembly votes (ALL) and UN security council membership (UNSC) as exogenous instruments. Estimates are not reported if the underidentification or Hansen test do not support the relevance or validity of the respective set of instruments. Estimation results, in support of (bold) or contrary to (emphasized) our main findings are highlighted.

Second, to evaluate robustness of the complete sample results with respect to the underlying multiple imputation datasets, we use m=5 and m=10 imputations to rerun the main regression. Table 6 and table 7 in the appendix report the results for the complete data analysis based only on the first 5 imputations and the first 10 imputations respectively. The results of the paper are robust to changes in the number of underlying multiple imputations.

Third, since for some of our dependent variables data collection starts after 1970 (see table 4 in the appendix) we re-estimate our main regression using different starting dates for the dependent variables. For instance, data collection for HIV prevalence started in the 1990s. Hence, estimates for HIV prevalence from the observed and the complete datasets in table 8 are based on the time period 1990 to 2009. In other words, we only include in-sample imputations in the complete data estimations. Using in-sample imputations only is an instructive robustness check, however, multiple imputation techniques are intended to replicate distributions. Hence, in-sample imputations are not necessarily preferable to out-of-sample imputations. In contrary, if the missing data mechanism is MAR - in particular if missingness is not independent from the time periods covered - in-sample imputations can be subject to the same bias as the observed data. While for DPT immunization - where after the start of data collection in 1980 only 13 observations are missing - a significant positiv effect of aid is now found in the observed as well as in the complete dataset, all other results are equipollent[18].

6 Conclusion

Access to schooling, health care and transportation facilities are important constituents of human well-being and lack thereof can aggravate income poverty. Increased emphasize on these aspects is thus a valuable turn in the empirical aid effectiveness literature. Yet, data availability is an issue when using effectiveness indicators other than growth, such as schooling rates, health indicators or measures of physical infrastructure.

Not only can relatively low fractions of missing data in individual variables reduce the number of observations considerably and thus cause efficiency losses, but missing data can also entail biased results. Since it is mostly countries with low income, where data availability is poor (due to a general lack of human and financial resources), coun-

[18]While for HIV prevalence the estimated coefficient for foreign aid is now negative significant in both samples, the size of the effect decreases. That is, - supporting our main finding - the effect is weaker in the complete sample.

tries which arguably should be the focus of analysis, simply 'drop out'. In addition, since missing observations alternate between indicators, comparisons between them (i.e. estimating several indicators in a common sample) can be highly inefficient or infeasible.

We use multiple imputation techniques to address these consequences of missing data. This technique reintroduces the missing observations into the analysis, while taking the inherent uncertainty of the imputed values into account. As a result, it increases the amount of observations in the analysis, as well as minimizes potential bias.

Our findings illustrate the importance to deliberate the choice of welfare indicators in aid effectiveness studies for two main reasons. First, sample selection bias due to the pattern of missing data in the respective indicator may alter regression results. Second, even once a potential selection bias is properly accounted for, not all indicators within a sector yield the same conclusion with respect to aid effectiveness. For instance, our findings indicate that aid differently affects the gross-enrollment rates and the pupil-teacher ratio - both indicators commonly used to evaluate advancements in the education sector. The analyzed indicator therefore has to be chosen very consciously to match the research question at hand.

References

K. Abayomi, A. Gelman, and M. Levy. Diagnostics for multivariate imputations. *Applied Statistics 57(3):273-291*, 2008.

A. Alesina and D. Dollar. Who gives foreign aid to whom and why? *Journal of Economic Growth 5(1):33-63*, 2000.

A. Alesina, R. Baqir, and W. Easterly. Public goods and ethnic divisions. *The Quarterly Journal of Economics 114(4):1243-1284*, 1999.

P.D. Allison. Missing data. *Thousand Oaks, CA: Sage Publications*, 2001.

B. Bueno de Mesquita, a. Smith, R.M. Siverson, and J.D. Morrow. The logic of political survival. *Cambridge: MIT Press*, 2003.

C. Burnside and D. Dollar. Aid, policies, and growth. *American Economic Review 90(4):847 868*, 2000.

D. Canning. A database of world stocks of infrastructure: 1950-1995. *The World Bank Economic Review 12: 529-548*, 1998.

J.A. Cheibub, J. Gandhi, and J.R. Vreeland. Democracy and dictatorship revisited. *Public Choice 143(1-2):67-101*, 2010.

Z. Christensen, D. Homer, and D. Nielson. We dont need no education: The effects of education-specific foreign aid school enrollment in low-income countries. *Paper prepared for presentation at the conference on Aid Transparency and Development Finance: Lessons from AidData, Oxford University.*, 2010.

M. Clemens, S. Radelet, and R. Bhavnani. Counting chickens when they hatch: The short term effect of aid on growth. *Center for Global Development, Working Paper 44*, 2004.

COW. Correlates of War Project. *Version 3.03*, 2003.

C.J. Dalgaard, H. Hansen, and F. Tarp. On the empirics of foreign aid and growth. *The Economic Journal 114:F191-F216*, 2004.

R. Deacon. Public good provision under dictatorship and democracy. *Public Choice (139): 241-262*, 2009.

S. Dietrich. The politics of public health aid: Why corrupt governments have incentives to implement aid effectively. *World Development 39(1):55-63*, 2011.

S. Djankov, J.G. Montalvo, and M. Reynal-Querol. The curse of aid. *Journal of Economic Growth 13(3): 169-194*, 2008.

A. Dreher and J. Sturm. Do imf and world bank influence voting in the un general assembly? *Public Choice 151(1):363-397*, 2012.

A. Dreher, P. Nunnenkamp, and R. Thiele. Does aid for education educate children? evidence from panel data. *KOF Working Papers 146*, 2006.

A. Dreher, J. Sturm, and J. Vreeland. Development aid and international politics: Does membership on the un security council influence world bank decisions? *Journal of Development Economics 88:1-18*, 2009.

V. Gauri and P. Khaleghian. Immunization in developing countries: Its political and organizational determinants. *World Development 30(12):2109-2132*, 2002.

A. Gelman. Parameterization and bayesian modeling. *Journal of the American Statistical Association 99(466): 53745*, 2004.

A. Gelman, G. King, and C. Liu. Not asked and not answered: Multiple imputation for multiple surveys. *Journal of the American Statistical Association 93: 846-857*, 1999.

J.W. Graham. Missing data analysis: Making it work in the real world. *Annual Review of Psychology 60: 549-576*, 2009.

J.W. Graham, A.E. Olchowski, and T.D. Gilreath. How many imputations are really needed? some practical clarifications of multiple imputation theory. *Prevention Science 8(3): 206-213*, 2007.

H. Hansen and F. Tarp. Aid and growth regressions. *Journal of Development Economics 64:547-570*, 2001.

A. Heston, R. Summers, and B. Aten. Penn World Table Version 7.0. *Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania*, 2011.

J. Hollyer, B.D. Rosendorff, and J.R. Vreeland. Democracy and transparency. *Journal of Politics 73(4): 1191-1205*, 2011.

J. Honaker and G. King. What to do about missing values in time series cross-section data. *American Journal of Political Science 54(3): 561-581*, 2010.

IBRD. World development indicators. *Washington, DC: The World Bank*, 2011.

IMF. Fiscal policy response to scaled-up aid. *International Monetary Fund, Fiscal Affairs Department*, 2007.

C. Kilby. Donor influence in international financial institutions: Deciphering what alignment measures measure. $https : //ncgg.princeton.edu/IPES/2009/papers/S1045_paper3.pdf$, 2009.

K.J. Lee and B. Carlin. Multiple imputation for missing data: Fully conditional specification versus multivariate normal imputation. *American Journal of Epidemiology 171(5): 624-632*, 2010.

K. Michaelowa. Aid effectiveness reconsidered: Panel data evidence for the education sector. *Hamburgisches Welt-Wirtschafts-Archiv Discussion Paper 264*, 2004.

K. Michaelowa and A. Weber. Aid effectiveness in the education sector: A dynamic panel analysis. *Theory and Practice of Foreign Aid 357-386*, 2007a.

K. Michaelowa and A. Weber. Aid effectivness in primary, secondary and tertiary education. *Background Paper All Global Monitoring Report 2008/ED/EFA/MRT/PI/51 UNESCO*, 2007b.

P. Mishra and D. Newhouse. Does health aid matter? *Journal of Health Economics 28:855-872*, 2009.

NIPA. Implicit price deflators for GDP. *US Department of Commerce, Bureau of Economic Analysis*, 2011.

OECD. Dac glossary of key terms and concepts: Official development assistance (oda). *http://www.oecd.org*, 2003.

T. Raghunathan and I. Bondarenko. Diagnostics for multiple imputations. $http : //ssrn.com/abstract = 1031750$, 2007.

T.E. Raghunathan. What to do with missing data? some options for analysis of incomplete data. *Annual Review of Public Health 25: 99-117*, 2004.

R. Rajan and A. Subramanian. Aid and growth: What does the cross-country evidence really show? *The Review of Economics and Statistics 90(4):643-665*, 2008.

M. Ross. Is democracy good for the poor? *American Journal of Political Science 50(4):860-874*, 2006.

D.B. Rubin. Inference and missing data. *Biometrica 63: 581-592*, 1976.

D.B. Rubin. Multiple imputation for nonresponse in surveys. *New York: John Wiley & Sons*, 1987.

A. Saiz. Democracy to the road: the political economy of potholes. *US Department of Commerce, Bureau of Economic Analysis*, 2002.

J.L. Schafer and J.W. Graham. Missing data: Our view of the state of the art. *Psychological Methods 7(2):147-177*, 2002.

M.E. Schaffer. xtivreg2: Stata module to perform extended iv/2sls, gmm and ac/hac, liml and k-class regression for panel data models. *http://ideas.repec.org/c/boc/bocode/s456501.html*, 2010.

StataCorp. Stata 12 multiple imputation reference manual. *College Station, TX: Stata Press*, 2011.

S. van Buuren. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research 16: 219242*, 2007.

S. van Buuren, J.P.L. Brand, and C.G.M. Groothuis-Oudshoorn. Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation 76(12): 1049-1046*, 2006.

I.R. White, P. Royston, and A.M. Wood. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicin 30: 377-399*, 2011.

WHO. Immunization surveillance, assessment and monitoring. *http://www.who.int/immunization_monitoring/data/en/index.html*, 2011.

S.E. Wilson. Chasing success: Health sector aid and mortality. *World Development 39(11): 20322043*, 2011.

7 Appendix A

Table 4

	(1) starting date	(2) missing before start	(3) missing after start	(4) overall missing	(5) number of observations
Life. expect. fem.	1970	0	6	6	1033
DPT	1980	229	13	242	797
Infant mortality	1970	0	24	24	1015
Completion prim.	1970	0	256	256	783
Pupil/Teacher	1970	0	157	157	882
Female/Male Enrol.	1970	0	92	92	947
Enrol. prim	1970	0	76	76	963
Enrol. prim & sec.	1970	0	138	138	901
Tuberc.	1990	472	4	476	563
HIV	1990	472	112	584	455
Internet	1970	0	447	447	592
EGC	1970	0	168	168	871
Phone	1970	0	43	43	996
Key (inst.)	1980	229	7	236	803
All (inst.)	1970	0	25	25	1014

Column 1 reports the first year for which we have any data for the respective indicator. Column 2 reports the number of (missing) observations between 1970 and the year reported in column 1. Column 3 reports the number of missing observations after the year reported in column 1. Column 4 reports the overall number of missing observations between 1970 and 2009 and column 5 reports the overall number of non-missing observations between 1970 and 2009 for the respective indicator.

Table 5: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Urban	43.727	23.101	2.4	100	5274
Aid	81.085	127.537	-118.14	1894.737	5274
Aid^2	22837.322	104813.212	0	3590028	5274
Pop	29755995.159	118793876.51	53600	1331380000	5274
colony Spain	0.158	0.418	0	3	5274
colony France	0.576	1.156	0	3	5274
colony UK	1.04	1.415	0	3	5274
language span.	0.453	1.075	0	3	5274
language french	0.558	1.123	0	3	5274
language eng.	0.922	1.33	0	3	5274
GDP	5483.895	6582.616	143.987	80090.352	5077
GDP growth	131.393	929.595	-9988.248	19412.227	4921
Gov. consumption	12.729	8.709	0.665	58.588	5076
Inflation	57.651	593.083	-33.585	26762.018	4661
Pop. density	131.072	401.914	1.008	6232.836	5234
Polity2	9.282	6.937	0	20	4665
Democ	2.709	1.492	0	5	5038
Political rights	4.359	2.04	1	7	4826
Civil liberties	4.312	1.689	1	7	4826
Measles	71.516	24.403	0	99	3868
Infant mortality	62.772	42.599	2.4	210.2	5122
tradeofgdp	78.484	47.038	0.309	413.455	4631
DPT	72.088	25.403	0	99	3936
Female/Male enrol.	87.793	17.179	0	159.244	3993
Female/Male Enrol. sec.	83.646	27.874	0	208.141	3295
Enrol. prim & sec.	72.003	24.466	5.313	121.901	3586
Completion prim.	70.525	28.135	0	151.718	2878
Enrol. prim.	92.744	27.079	8.004	232.841	4272
Enrol. sec.	48.552	30.272	0	117.854	3720
Rail	3625.531	9637.196	0	63506	2832
Pupil/Teacher	33.048	13.136	2.925	100.236	3509
Enrol. tert.	13.106	15.372	0	118.102	3196
Life expect.	60.607	10.694	26.819	80.146	5177
Life expect. fem.	62.707	11.36	28.532	82.400	5177
Phone	7.749	10.541	0.006	55.111	4614
EGC	5634.869	21582.504	0	356090	4017
Tuberc.	64.596	39.156	0	860	2735
Internet	5.672	10.335	0	75.03	2423
HIV	2.185	4.532	0.1	26.5	2324
All Can.	0.473	0.154	0	0.968	4963
All France	0.402	0.148	0	0.895	4963
All UK	0.388	0.152	0	0.903	4963
All Ger.	0.465	0.18	0	1	4681
All Italy	0.48	0.165	0	1	4963
All Jpn.	0.521	0.154	0	0.932	4963
All USA	0.189	0.117	0	0.808	4963
All G7	0.393	0.155	0	0.912	4963
Key Can.	0.462	0.246	0	1	3577
Key France	0.467	0.24	0	1	3577
Key UK	0.458	0.25	0	1	3577
Key Ger.	0.459	0.252	0	1	3577
Key Italy	0.487	0.252	0	1	3577
Key Jpn.	0.462	0.252	0	1	3577
Key USA	0.312	0.229	0	1	3577
Key G7	0.456	0.251	0	1	3577

Maximum number of observations: 5274.

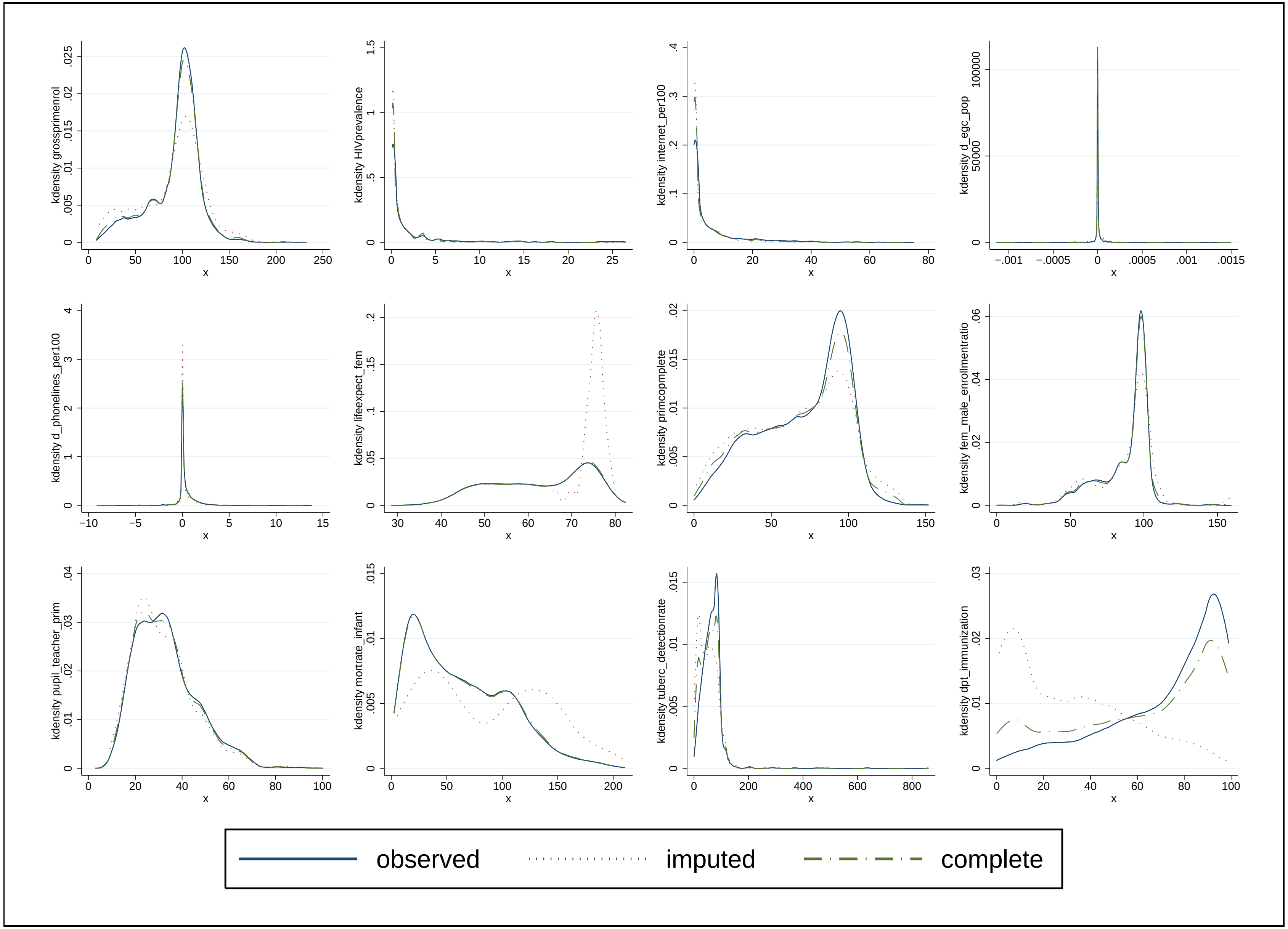

Figure 2: Kernel densities

8 Appendix B

Table 6: Fixed Effects IV Estimates, 5-year averages, 10 imputed datasets

	Aid, observed	Aid, complete	N, observed	Groups, observed	Underid, p-val	Hansen, p-val		
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
DPT^{A}	9.088**	-8.945	743	1029	144	152	0.009	0.096
	(3.723)	(6.331)						
Infant mortality^{A}	1.012	1.582	936	1029	145	152	0.019	0.469
	(4.455)	(5.018)						
Tuberc.^{B}	6.655	1.518	531	1029	144	152	0.038	0.279
	(9.684)	(10.681)						
HIV^{B}	-2.137**	1.035	438	1029	118	152	0.009	0.338
	(0.888)	(0.779)						
Life expect. fem^{B}	0.074	-1.286	948	1029	145	152	0.003	0.356
	(1.062)	(1.487)						
Completion prim.^{A}	1.925	0.124	597	1029	133	152	0.016	0.200
	(3.411)	(2.414)						
Pupil/Teacher^{B}	3.901	4.368	816	1029	141	152	0.082	0.776
	(2.507)	(2.750)						
Female/Male enrol.^{B}	6.217**	5.812	874	1029	141	152	0.017	0.027
	(3.122)	(4.229)						
Enrol. prim.^{B}	5.415	5.176	890	1029	141	152	0.015	0.938
	(4.679)	(4.865)						
Enrol. prim. & sec.^{B}	0.853	3.130	831	1029	140	152	0.019	0.633
	(2.878)	(2.969)						
Internet^{B}	-0.119	-5.455*	558	1029	144	152	0.186	0.104
	(2.663)	(2.981)						
D.Phone^{A}	0.119	0.054	754	1029	144	152	0.091	0.344
	(0.115)	(0.120)						
D.EGC^{B}	-6.8e-06	-2.2e-05	802	1029	142	152	0.009	0.449
	(7.7e-06)	(2.6e-05)						

* p<0.10, ** p<0.05, *** p<0.01
All estimates include the control variables GDP, Democ, POP and Urban as well as time dummies. The estimates are obtained from the user-written Stata command xtivreg2 (Schaffer (2010)). Aid as well as GDP per capita are instrumented using (*A*) UN general assembly key votes (KEY) and UN security council membership (UNSC) or (*B*) all UN general assembly votes (ALL) and UN security council membership (UNSC) as exogenous instruments.

Table 7: Fixed Effects IV Estimates, 5-year averages, 5 imputed datasets

	Aid, observed	Aid, complete	N, observed	Groups, observed	Underid, p-val	Hansen, p-val		
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
DPT[A]	9.088** (3.723)	-8.005 (5.618)	743	1029	144	152	0.009	0.096
Infant mortality[A]	1.012 (4.455)	1.268 (4.274)	936	1029	145	152	0.019	0.469
Tuberc.[B]	6.655 (9.684)	2.067 (8.549)	531	1029	144	152	0.038	0.279
HIV[B]	-2.137** (0.888)	0.830 (0.561)	438	1029	118	152	0.009	0.338
Life expect. fem[B]	0.074 (1.062)	-0.880 (1.145)	948	1029	145	152	0.003	0.356
Completion prim.[A]	1.925 (3.411)	-0.177 (2.257)	597	1029	133	152	0.016	0.200
Pupil/Teacher[B]	3.901 (2.507)	3.330* (1.971)	816	1029	141	152	0.082	0.776
Female/Male enrol.[B]	6.217** (3.122)	5.600 (3.768)	874	1029	141	152	0.017	0.027
Enrol. prim.[B]	5.415 (4.679)	5.171 (4.069)	890	1029	141	152	0.015	0.938
Enrol. prim. & sec.[B]	0.853 (2.878)	3.003 (2.494)	831	1029	140	152	0.019	0.633
Internet[B]	-0.119 (2.663)	-4.381** (2.130)	558	1029	144	152	0.186	0.104
D.Phone[A]	0.119 (0.115)	0.061 (0.118)	754	1029	144	152	0.091	0.344
D.EGC[B]	-6.8e-06 (7.7e-06)	-2.0e-05 (2.5e-05)	802	1029	142	152	0.009	0.449

* p<0.10, ** p<0.05, *** p<0.01
All estimates include the control variables GDP, Democ, POP and Urban as well as time dummies. The estimates are obtained from the user-written Stata command xtivreg2 (Schaffer (2010)). Aid as well as GDP per capita are instrumented using (*A*) UN general assembly key votes (KEY) and UN security council membership (UNSC) or (*B*) all UN general assembly votes (ALL) and UN security council membership (UNSC) as exogenous instruments.

Table 8: Fixed Effects IV Estimates, 5-year averages, indiv. starting time

	Aid, observed	Aid, complete	N, observed	Groups, observed	Underid, p-val	Hansen, p-val		
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
DPT^{A}	9.088**	9.724**	743	801	144	152	0.009	0.096
	(3.723)	(4.572)						
Infant mortality^{A}	1.012	1.772	936	1029	145	152	0.019	0.469
	(4.455)	(5.028)						
Tuberc.^{B}	6.655	6.250	531	559	144	151	0.038	0.279
	(9.684)	(9.497)						
HIV^{B}	-2.137**	-1.667*	438	559	118	151	0.009	0.338
	(0.888)	(0.853)						
Life expect. fem^{B}	0.074	-1.270	948	1029	145	152	0.003	0.356
	(1.062)	(1.423)						
Completion prim.^{A}	1.925	0.700	597	1029	133	152	0.016	0.200
	(3.411)	(2.563)						
Pupil/Teacher^{B}	3.901	4.504*	816	1029	141	152	0.082	0.776
	(2.507)	(2.606)						
Female/Male enrol.^{B}	6.217**	5.722	874	1029	141	152	0.017	0.027
	(3.122)	(3.717)						
Enrol. prim.^{B}	5.415	5.416	890	1029	141	152	0.015	0.938
	(4.679)	(4.750)						
Enrol. prim. & sec.^{B}	0.853	3.196	831	1029	140	152	0.019	0.633
	(2.878)	(2.938)						
Internet^{B}	1.320	-1.075	530	559	144	151	0.049	0.343
	(2.913)	(2.112)						
D.Phone^{A}	0.119	0.059	754	1029	144	152	0.091	0.344
	(0.115)	(0.117)						
D.EGC^{B}	-6.8e-06	-1.9e-05	802	1029	142	152	0.009	0.449
	(7.7e-06)	(2.4e-05)						

* p<0.10, ** p<0.05, *** p<0.01
All estimates include the control variables GDP, Democ, POP and Urban as well as time dummies. The estimates are obtained from the user-written Stata command xtivreg2 (Schaffer (2010)). Aid as well as GDP per capita are instrumented using (A) UN general assembly key votes (KEY) and UN security council membership (UNSC) or (B) all UN general assembly votes (ALL) and UN security council membership (UNSC) as exogenous instruments.

Table 9: FEIV, 5year, HIV

	Aid	N	Groups	Underid, p-val	Hansen, p-val
HIV observed	-2.137** (0.888)	438	118	0.009	0.338
HIV complete m=1-20	1.022 (0.729)	1029	152		
HIV imput. m=1	0.701 (0.634)	1029	152	0.080	0.227
HIV imput. m=2	0.882* (0.514)	1029	152	0.043	0.309
HIV imput. m=3	0.643 (0.489)	1029	152	0.034	0.231
HIV imput. m=4	1.099** (0.506)	1029	152	0.064	0.084
HIV imput. m=5	0.824* (0.469)	1029	152	0.104	0.180
HIV imput. m=6	0.660 (0.481)	1029	152	0.038	0.302
HIV imput. m=7	1.604* (0.962)	1029	152	0.310	0.568
HIV imput. m=8	1.688* (0.964)	1029	152	0.219	0.541
HIV imput. m=9	0.890 (0.578)	1029	152	0.059	0.226
HIV imput. m=10	1.355* (0.786)	1029	152	0.515	0.428
HIV imput. m=11	1.059** (0.529)	1029	152	0.026	0.807
HIV imput. m=12	0.767 (0.596)	1029	152	0.038	0.222
HIV imput. m=13	1.238* (0.648)	1029	152	0.147	0.940
HIV imput. m=14	0.656 (0.706)	1029	152	0.078	0.114
HIV imput. m=15	1.473** (0.674)	1029	152	0.099	0.682
HIV imput. m=16	0.779 (0.595)	1029	152	0.372	0.760
HIV imput. m=17	1.082 (0.685)	1029	152	0.109	0.685
HIV imput. m=18	0.907 (0.701)	1029	152	0.237	0.401
HIV imput. m=19	0.807* (0.450)	1029	152	0.005	0.250
HIV imput. m=20	1.322* (0.679)	1029	152	0.036	0.479

Table 10: FEIV, 5year, DPT

	Aid	N	Groups	Underid, p-val	Hansen, p-val
DPT observed	9.088**	743	144	0.009	0.096
	(3.723)				
DPT complete m=1-20	-9.124	1029	152		
	(6.257)				
DPT imput. m=1	-4.398	1029	152	0.080	0.024
	(4.013)				
DPT imput. m=2	-10.565**	1029	152	0.043	0.183
	(5.172)				
DPT imput. m=3	-11.723**	1029	152	0.034	0.730
	(5.568)				
DPT imput. m=4	-7.297*	1029	152	0.064	0.177
	(3.925)				
DPT imput. m=5	-6.041*	1029	152	0.104	0.057
	(3.452)				
DPT imput. m=6	-8.868*	1029	152	0.038	0.292
	(4.797)				
DPT imput. m=7	-7.734	1029	152	0.310	0.217
	(5.656)				
DPT imput. m=8	-9.301	1029	152	0.219	0.071
	(6.263)				
DPT imput. m=9	-8.687	1029	152	0.059	0.239
	(5.382)				
DPT imput. m=10	-14.833*	1029	152	0.515	0.810
	(8.969)				
DPT imput. m=11	-8.457*	1029	152	0.026	0.273
	(4.586)				
DPT imput. m=12	-9.273*	1029	152	0.038	0.191
	(5.194)				
DPT imput. m=13	-8.415	1029	152	0.147	0.151
	(5.122)				
DPT imput. m=14	-6.836	1029	152	0.078	0.118
	(5.298)				
DPT imput. m=15	-11.871*	1029	152	0.099	0.660
	(6.445)				
DPT imput. m=16	-12.079	1029	152	0.372	0.487
	(8.432)				
DPT imput. m=17	-10.311*	1029	152	0.109	0.262
	(5.276)				
DPT imput. m=18	-11.759*	1029	152	0.237	0.618
	(6.814)				
DPT imput. m=19	-3.425	1029	152	0.005	0.093
	(2.735)				
DPT imput. m=20	-10.601**	1029	152	0.036	0.750
	(4.684)				