Appendix S1 Construction of the sectoral aid data

Creditor Reporting System

As already discussed in the main text, the OECD's (2002) Creditor Reporting System (CRS) disaggregates development assistance along a number of dimensions, including the sector or purpose of aid and the aid type or prefix code. Unfortunately, because CRS disbursements reported by most donors are incomplete in at least some years they need to be supplemented with additional information. This appendix describes in detail a data construction method that further makes use of two OECD Development Assistance Committee (DAC) data tables to construct more complete disaggregated aid disbursements.[1]

Starting from the CRS database, I download annual gross disbursements in millions of US$ for the period 1990-2004 for the following sectors:[2] education (DAC5 sector code 110), health (120), commodity aid/general programme assistance (500), action relating to debt (600), donor administrative costs (910), support to NGOs (920) and other sectors (the sum of all remaining sector codes).[3] These data are obtained in a recipient-donor-year (RDY) format, i.e. for each year showing the amount of foreign aid transferred from each donor to each recipient. Education and health disbursements are further partitioned into four aid types or prefix codes: investment projects (IP), sector programme (SP) aid, technical cooperation (TC), and other (no mark) (ONM).

The prefix codes are useful because, to some extent, they allow the separation of on- and off-budget aid flows (see the main body of the paper for details). Definitions of the prefix codes can be found in OECD (2002, p. 22) (also see OECD, 2000, pp. 47-48): "free-standing technical cooperation is defined as financing of activities whose primary purpose is to augment the level of knowledge, skills, technical know-how or productive aptitudes of the population of aid recipient countries. ... It includes the cost of personnel, training and research, as well as associated equipment and administrative costs" and mainly comes in the form of "supply of human resources (teachers, volunteers and experts) or action targeted on human resources (education, training, advice)" (OECD, 2000, p. 47).[4] Sector programme aid "comprises contributions to carry out wide-ranging development plans in a defined sector such as agriculture, education, transportation, etc. Assistance is made available 'in cash' or 'in kind', with or without restriction on the specific use of the funds, but on the condition the recipient executes a development plan in favour of the sector concerned. Investment projects comprise schemes to increase and/or improve the recipient's stock of physical capital and financing the supply of goods and services in support of such schemes" (OECD, 2002, p. 22). This includes investment-related technical cooperation, which is "the financing of services by a donor country with the primary purpose of contributing to the design and/or implementation of a project or programme aiming to increase the physical capital stock of the recipient country. These services include consulting services, technical support, the provision of know-how linked to the execution of an investment project, and the contribution of the donor's own personnel to the actual implementation of the project (managers, technicians, skilled labour etc.)" (OECD, 2002, p. 22). Other (no mark) is the residual category.

A very small part of education and health aid in CRS is listed under a combination of prefix codes (e.g. IP & TC). In these cases, I allocate an equal part of the aid amount to each of the prefix codes that make up the combination.

At this stage it is important to note that CRS does not record zeros. If no aid is given in a sector the observation is simply missing, so, in general, it is difficult to tell whether an observation is missing because no aid is disbursed or because existing aid flows are not reported. Whenever total education or health disbursements are available, which is the case when at least one of the four prefix codes is available, I set missing values for the other prefix codes to zero. Similarly, whenever aggregate disbursements are available, missing observations for sectoral disbursements, as well as education and health prefix codes, are changed to zero. The prefix codes always sum to total education and health disbursements. Similarly, aggregate CRS disbursements equal the sum of the underlying sectors, apart from tiny discrepancies.[5] I also download CRS data on aggregate grants and loans, which will become useful later. Again, missing observations for these two variables are turned to zero whenever aggregate CRS disbursements are available. CRS grants and loans always sum to aggregate CRS disbursements.

The aggregate and sectoral disbursements thus obtained from CRS in a recipient-donor-year format form the backbone of the data construction. From here on I refer to these variables as CRS^{agg}_{RDY} and CRS^{s}_{RDY} (for $s = 1, \ldots, S$), respectively. CRS disbursements at the prefix code level are labelled $CRS^{s,p}_{RDY}$, where s now refers to the education or health sector and $p = IP, SP, TC, ONM$. Because these aid measures are incomplete I attempt to improve on them, which first of all requires data from DAC table 2a.

Development Assistance Committee table 2a

The data in DAC2a should be complete but do not allow a full disaggregation of aid according to sector or prefix code. I download data on grants and loans extended, again in a RDY format. Missing values for loans are set to zero when grants are observed, and vice versa. Total disbursements, $DAC2a^{agg}_{RDY}$, are then calculated as the sum of grants and loans. The OECD makes a distinction between Official Development Assistance (ODA) and Official Assistance (OA), where OA is simply ODA directed to countries on part II of the DAC list of aid recipients, comprised of transition countries and more advanced developing countries (OECD, 2000, p. 11 and p. 64). Whether aid transferred to a given recipient is classified as ODA or OA may vary over time. While OECD (2002, p. 4) states that the CRS database contains both ODA and OA, in the CRS data I downloaded no observations are available for recipient-years that are listed on part II of the recipient list in DAC2a. As a result, I focus only on ODA in DAC2a and exclude part II recipient-years. Conversely, for Serbia CRS data is available but DAC2a data is not so Serbia is dropped from the sample.[6] In addition, I only select donors that are also available in DAC table 5, for reasons that will become clear shortly. Two donors are excluded from DAC2a because of this: GFATM (Global Fund to Fight Aids, Tuberculosis and Malaria) and UNFPA (United Nations Population Fund).

Calculating the amount of aid missing from CRS

I now have data on (supposedly) complete aggregate DAC2a disbursements and incomplete aggregate and sectoral CRS disbursements, both in a recipient-donor-year format. By subtracting CRS disbursements from DAC2a disbursements I obtain a residual for aggregate disbursements (RES^{agg}_{RDY}). For each RDY observation this residual captures the amount of disbursed aid that is missing from the CRS database:

$$RES^{agg}_{RDY} = DAC2a^{agg}_{RDY} - CRS^{agg}_{RDY} \tag{S1.1}$$

The aim is to allocate this residual across sectors.

RES^{agg}_{RDY} is negative for quite a few observations. In the majority of such cases CRS disbursements exceed DAC2a disbursements by only a very small margin but there are also a number of observations where the difference is larger. I replace DAC2a grants (loans) by the CRS amount in all cases where CRS grants (loans) exceed DAC2a grants (loans). I then recalculate $DAC2a^{agg}_{RDY}$ as the sum of DAC2a grants and loans, and recalculate RES^{agg}_{RDY}. If the DAC2a value is negative and the CRS value is zero, however, no replacement is carried out, whereas if the DAC2a value is negative and the CRS value is non-zero the former *is* replaced by the latter.

The rationale for these adjustments is that it is very unlikely that aid is reported if it never actually took place. It is far more likely actual aid is underreported, i.e. it is more likely DAC2a figures are missing something when they are exceeded by CRS figures, even though they are supposed to be complete. It might also be the case that negative amounts of aid go unreported in CRS and this is what causes the CRS figures to exceed the DAC2a figures. This is less probable, however, since negative amounts of aid, which presumably capture the repayment of unused aid money or resources, are quite rare in the data.[7]

Applying this rationale consistently is also what leads me not to replace the DAC2a value by the CRS value if the former is negative and the latter is zero. A zero CRS value means no aid is reported to CRS, while the negative value for DAC2a implies there was some aid, albeit negative. The situation where DAC2a aid is negative and CRS aid is non-zero is more tricky. On the one hand, the DAC2a database is supposed to be complete so its value is more likely to be the true one but, on the other hand, negative amounts of aid are rare and it is difficult to interpret them, which tilts the balance of favour of the CRS figure. Hence, in this case I replace the negative DAC2a amount by the non-zero (and always positive) CRS amount. Because there are only a few such observations (9 for grants and 17 for loans, out of a total of 43216 RDY observations) this choice should not have a substantial impact on the data.

For some RDY observations (1230 in total) CRS data is available but DAC2a data is not.[8] For these observations no residual can be calculated. Even so, I do not delete these observations from the CRS database, they are simply treated as having a zero residual. Conversely, if observations are available in DAC2a but missing from CRS, all CRS variables are changed to zero so that the complete DAC2a value is recorded as a residual.

Having calculated a total residual for each RDY observation, I collapse the dataset by summing over recipients, yielding a residual for aggregate disbursements in a donor-year format ($RES_{DY}^{agg,C}$, where DY stands for donor-year and C makes clear this residual is formed by *collapsing* RES_{RDY}^{agg} over all recipients):

$$RES_{DY}^{agg,C} = \sum_{R} RES_{RDY}^{agg} \tag{S1.2}$$

While the RDY data contains 113 negative residuals, $RES_{DY}^{agg,C}$ is always positive. The reason for collapsing the dataset is that now, with data from one more DAC table, it becomes possible to allocate $RES_{DY}^{agg,C}$ across sectors for each donor-year.

Development Assistance Committee table 5

To do this, one more piece of information, which comes from DAC table 5, is needed. DAC5 comprises a sectoral disaggregation of total ODA but only in a donor-year format. While this means the data are not available from a recipient perspective, the advantage of DAC5 is that it should contain more complete information than CRS. I label total and sectoral aid from this table as $DAC5_{DY}^{agg}$ and $DAC5_{DY}^{s}$, respectively. As in CRS, the sectors of interest are: education (DAC5 sector code 110), health (120), commodity aid/general programme assistance (500), action relating to debt (600), donor administrative costs (910), support to NGOs (920) and other sectors (the sum of all remaining sector codes). Missing observations for sectoral aid are set to zero whenever $DAC5_{DY}^{agg}$ is available. A problem is that $DAC5_{DY}^{agg}$ is not always equal to the sum of the sectoral aid variables. Four observations show up with large discrepancies: AsDF (Asian Development Fund) 1996, AsDF 2002, France 1997, and IDB (Inter-American Development Bank) Special Fund 1996.

For AsDF 2002 and France 1997 $DAC5_{DY}^{agg}$ exceeds the sum of the sectoral aid variables. In both cases this is because the entry for total sector allocable aid exceeds the sum of its underlying series.[9] Hence, for both observations I scale up all sector allocable series so that their sum matches total sector allocable aid. This means education and health aid are scaled up but also the other sector allocable series, which make up part of other sector aid. Therefore, after scaling up, other sector aid is recalculated as the sum of the underlying sectors. For all other observations discrepancies are extremely small, most likely due to rounding errors. To get rid of these small discrepancies $DAC5^{agg}_{DY}$ is recalculated as the sum of the sectoral aid variables. For AsDF 1996 and IDB Special Fund 1996 the sectoral sum exceeds $DAC5^{agg}_{DY}$ so these observations are also taken care of in this way.

Lastly, from DAC5 I also download data that partition health and education aid into the four prefix codes, again in a donor-year format ($DAC5^{s,p}_{DY}$).[10] Because, for AsDF 2002 and France 1997, education and health aid have been scaled up (see previous paragraph) I also scale up the prefix codes for these observations so that they still sum to total education and health aid. As before, missing observations for the prefix codes are set to zero whenever at least one of the other prefix codes within the sector is observed.

Unfortunately, the prefix codes in DAC5 do not always sum to total education and health aid. There is one observation for which the education total exceeds the sum of the prefix codes, while for health there are three such observations. For these observations I scale up the prefix codes so that their sum matches the sector total. I then recalculate education and health totals as the sum of their prefix codes for all other observations. This takes care of the one observation in both sectors for which the sum of the prefix codes exceeds the sector total. It also sorts out the many observations for which there are extremely small discrepancies. As this leads to changes in the values of education and health aid I recalculate $DAC5^{agg}_{DY}$ as the sum of the underlying sectors to ensure consistency.

This means I now have, in donor-year format, (supposedly) complete aid data disaggregated by sector from DAC5 and incomplete aid data disaggregated by sector from the collapsed CRS dataset ($CRS^{agg}_{DY} = \sum_R CRS^{agg}_{RDY}$, $CRS^{s}_{DY} = \sum_R CRS^{s}_{RDY}$, $CRS^{s,p}_{DY} = \sum_R CRS^{s,p}_{RDY}$). The plan is to calculate sectoral residuals for each donor-year and to use these to allocate each donor's total residual across sectors in each year. Going back to the data in recipient-donor-year format (RES^{agg}_{RDY}) this donor- and year-specific sectoral allocation of the total residual is then applied to all recipients that receive aid from the relevant donor in a given year that is not accounted for in CRS.

There is, however, one problem that needs to be solved before proceeding. The sectoral residuals must be calculated from DAC5 data, whereas the total residual is based on DAC2a data (as DAC5 is not available in RDY format). Apart from the possibility of reporting inconsistencies between the two tables, a bigger problem arises because donors have a choice in DAC5 to report either commitments or disbursements. I received some information from the DAC for the years 2001-2004 as to who reports what. Out of the 127 DY observations with data in DAC5 for which I have this information 72 refer to disbursements and 55 to commitments. However, these 55 observations include many of the larger donors, such as the United States, Japan, the European Commission, Germany and France.

As a consequence I scale all DAC5 aid variables, including the education and health prefix codes, by the ratio of aggregate DAC2a disbursements to total DAC5 ODA so that the sectoral aid variables from DAC5 sum to DAC2a aggregate disbursements:

$$\widehat{DAC5}^{s}_{DY} = DAC2a^{agg}_{DY}\left(\frac{DAC5^{s}_{DY}}{DAC5^{agg}_{DY}}\right) \tag{S1.3}$$

for $s = 1, \ldots, S$, and:

$$\widehat{DAC5}^{s,p}_{DY} = DAC2a^{agg}_{DY}\left(\frac{DAC5^{s,p}_{DY}}{DAC5^{agg}_{DY}}\right) \tag{S1.4}$$

for the education and health sectors and $p = IP, SP, TC, ONM$. This amounts to assuming that the sectoral allocation in DAC5 (of commitments or disbursements) is an accurate guide to the sectoral allocation of DAC2a disbursements. The correlation between $DAC5^{agg}_{DY}$ and $DAC2a^{agg}_{DY}$, at least, is very high (0.90). A few positive $DAC5^{agg}_{DY}$ values are scaled to zero because $DAC2a^{agg}_{DY}$ is zero but since these observations have no aggregate disbursements residual that needs to be allocated anyway this is not a problem. Scaling the data in this way ensures that the sectoral aid variables from DAC5 sum to DAC2a aggregate disbursements. This allows for a calculation of sectoral residuals that is more consistent with the calculation of the total residual, which is based on $DAC2a^{agg}_{RDY}$.

If, after the scaling, sectoral values in CRS exceed those in DAC5, I replace the latter by the former. I first carry out this replacement at the level of the prefix codes and recalculate total education and health aid as the sum of their prefix codes. I then repeat this strategy for all sectoral aid variables. At this stage the only changes for education and health aid occur for observations for which there is no prefix code disaggregation. So, after these changes the prefix codes still sum to total education and health aid for all observations that have data on both. As before, the DAC5 value is not replaced by the CRS value if the DAC5 value is negative and the CRS value is zero. However, if the DAC5 value is negative and the CRS value is non-zero then the former is replaced by the latter. The adjustments are limited in number and size, which is brought out by the high correlation (0.99) between the sum of the DAC5 sectoral aid variables (after scaling and replacement: $\sum_{s=1}^{S} \widehat{DAC5}^{s}_{DY}$) and $DAC2a^{agg}_{DY}$.

Allocating the residual across sectors

The total residual in donor-year format, RES_{DY}, is now calculated as the sum of the DAC5 sectoral aid variables minus aggregate CRS disbursements:

$$RES^{agg}_{DY} = \sum_{s=1}^{S} \widehat{DAC5}^{s}_{DY} - CRS^{agg}_{DY} \tag{S1.5}$$

The correlation with the collapsed residual that was computed earlier from the recipient-donor-year dataset ($RES^{agg,C}_{DY}$) is 0.97. Sectoral (prefix) residuals in this DY format are calculated as the difference between

DAC5 sectoral (prefix) aid variables and sectoral (prefix) CRS disbursements:

$$RES^{s}_{DY} = \widehat{DAC5}^{s}_{DY} - CRS^{s}_{DY} \tag{S1.6}$$

$$RES^{s,p}_{DY} = \widehat{DAC5}^{s,p}_{DY} - CRS^{s,p}_{DY} \tag{S1.7}$$

The sectoral residuals sum to RES^{agg}_{DY} and residuals for the prefix codes sum to the total residuals for education and health. Whenever the CRS value is missing, the full DAC5 value is recorded as residual, as before.

Two sectoral residuals are negative. Finland 1991 has a negative residual for health IP (the DAC5 value is negative, while the CRS value is zero). For this observation I turn the health prefix code residuals to missing. UK 1996 has a negative residual for action relating to debt. Because this observation has a large total residual it would be a shame to lose it. Moreover, the absolute value of the negative action relating to debt residual is less than 0.1% of the total residual. Therefore, I set the action relating to debt residual to zero for this observation and recalculate RES^{agg}_{DY} as the sum of the sectoral residuals.

Now it is possible to calculate the shares of the sector residuals in the total residual ($SHRES^{s}_{DY}$), as well as the share of the prefix code residuals in the total education and health residuals ($SHRES^{s,p}_{DY}$):

$$SHRES^{s}_{DY} = \frac{RES^{s}_{DY}}{\sum_{s=1}^{S} RES^{s}_{DY}} \tag{S1.8}$$

$$SHRES^{s,p}_{DY} = \frac{RES^{s,p}_{DY}}{\sum_{p} RES^{s,p}_{DY}} \tag{S1.9}$$

This donor- and year-specific allocation of RES^{agg}_{DY} across sectors is then applied to the total residual calculated in the original RDY format (RES^{agg}_{RDY}):

$$\widehat{RES}^{s}_{RDY} = SHRES^{s}_{DY} RES^{agg}_{RDY} \tag{S1.10}$$

That is, I apply the sectoral residual shares of a given donor-year to the total residuals of all recipients to which the donor gives aid in that year that is not fully accounted for in CRS. In other words, I assume the sectoral allocation of a donor's total residual is the same for all recipients with which this donor has a residual. For instance, if Botswana and Tanzania receive an unallocated residual from the US in 2004, and (S1.8) shows that half of the total residual of the US in 2004 consists of education aid and half consists of health aid, then for both Botswana and Tanzania half of the total residual with the US in 2004 is classified as education aid and half as health aid. Total education and health residuals are allocated across prefix codes in the same way:

$$\widehat{RES}^{s,p}_{RDY} = SHRES^{s,p}_{DY} \widehat{RES}^{s}_{RDY} \tag{S1.11}$$

Creating more complete sectoral aid disbursements

I add the sectoral residuals to the CRS disbursements in the RDY database, and likewise for the education and health prefix codes:

$$\widetilde{CRS}^{s}_{RDY} = CRS^{s}_{RDY} + \widehat{RES}^{s}_{RDY} \tag{S1.12}$$

$$\widetilde{CRS}^{s,p}_{RDY} = CRS^{s,p}_{RDY} + \widehat{RES}^{s,p}_{RDY} \tag{S1.13}$$

For some observations insufficient information is available in DAC5 to allocate RES^{agg}_{RDY} across sectors.[11] As a result, the sum of the newly calculated sectoral variables does not necessarily equal $DAC2a^{agg}_{RDY}$.[12] Similarly, education and health prefix codes do not always sum to the education and health total because for some donors insufficient information is available to allocate the education and health residuals across prefix codes.

Therefore, as a final step in the data construction, after collapsing the data to a recipient-year (RY) format, I scale the sectoral disbursements so that their sum equals a plausible measure of aggregate disbursements received. Before collapsing the data I replace missing $DAC2a^{agg}_{RDY}$ by CRS^{agg}_{RDY} for the 1230 RDY observations that have CRS data but are missing from DAC2a.[13]

I collapse the RDY dataset by summing over donors:

$$\widetilde{CRS}^{s}_{RY} = \sum_{D} \widetilde{CRS}^{s}_{RDY} \tag{S1.14}$$

$$\widetilde{CRS}^{s,p}_{RY} = \sum_{D} \widetilde{CRS}^{s,p}_{RDY} \tag{S1.15}$$

In this final recipient-year (RY) dataset there are observations for which both aggregate DAC2a and CRS disbursements are zero. The reason why these observations are zero rather than missing (as one would expect) is that Stata turns missing values to zero when collapsing data. I turn all aid variables to missing for these observations. In addition, there are seven observations with non-zero aggregate DAC2a disbursements but zeros for all sectoral aid variables. Since, for these observations, there is no information at all about the allocation of aggregate disbursements across sectors, all variables are turned to missing. Similarly, there is one observation with zeros for all health prefix codes, but a non-zero health total. For this observation health prefix codes are changed to missing.

As before the collapse, when I sum the sectoral disbursements I do not always get a number that equals aggregate DAC2a disbursements ($DAC2a^{agg}_{RY} = \sum_D DAC2a^{agg}_{RDY}$), and, similarly, the sum of the prefix codes does not always equal total education and health aid. I first scale the prefix codes so that their sum equals total education and health aid. This is done by multiplying each prefix code with the ratio of total sectoral (education or health) aid to the sum of the prefix codes:

$$\overline{CRS}^{s,p}_{RY} = \widetilde{CRS}^{s}_{RY} \left(\frac{\widetilde{CRS}^{s,p}_{RY}}{\sum_p \widetilde{CRS}^{s,p}_{RY}} \right) \tag{S1.16}$$

For Chinese Taipei (more commonly known as Taiwan) several years have negative values for total health aid while the sum of the health prefix codes is positive. In addition, in the remaining observed years (except 1990) the sum of the health prefix codes always exceeds total health aid and these are the only observations in the dataset for which this is the case. Similarly, in all observed years except 1990 Chinese Taipei has a value for total education aid that is smaller than the sum of the prefix codes (the latter is also the case for Somalia 1997). This seems to suggest data for Chinese Taipei contains a great deal of measurement error. Given that Chinese Taipei has no data after 1996 in any case, it is dropped from the dataset in its entirety. For both sectors Cayman islands 1991 has a negative prefix sum. However, because total education and health aid are also negative, scaling should not be a problem for this observation. For now, I keep this observation and simply apply the scaling, as it will be dropped at a later stage for other reasons in any case.

I now apply the same strategy to the sectoral aid variables to make sure their sum matches an aggregate measure of disbursements received. Recall that aggregate DAC2a disbursements in this RY format are calculated by summing DAC2a disbursements in the RDY format over all donors, and that donors that are missing from DAC5 or CRS were not selected when downloading data for $DAC2a^{agg}_{RDY}$. Consequently, aid from these donors is not included in $DAC2a^{agg}_{RY}$. Therefore, in addition to $DAC2a^{agg}_{RY}$, I download grants and loans from DAC2a in a RY format, selecting 'all donors (total)' in the donor dimension. Missing grants are set to zero when loans are observed, and vice versa. Total disbursements, $DAC2a^{agg}_{RY,AD}$ (AD stands for all donors), are calculated as the sum of grants and loans extended. The correlation between this measure and $DAC2a^{agg}_{RY}$ is extremely high (0.99). The sum of the sectoral variables has a similarly high correlation with both measures.

I scale the sectoral variables so that their sum equals the maximum of $DAC2a^{agg}_{RY}$ and $DAC2a^{agg}_{RY,AD}$. Again, this follows the rationale that it is unlikely non-existing aid is reported, so the higher figure should be the most accurate one. While $DAC2a^{agg}_{RY,AD}$ should include aid from more donors, $DAC2a^{agg}_{RDY}$ (on which $DAC2a^{agg}_{RY}$ is based) has been adjusted upwards for those observations where it is exceeded by aggregate CRS disbursements (see above).

For 4 observations (Costa Rica 1992, Mexico 1992, Panama 1992, Saudi Arabia 1991) the sum of the sectoral aid variables ($\sum_{s=1}^{S} \widetilde{CRS}^{s}_{RY}$) slightly exceeds $DAC2a^{agg}_{RY}$ (for some other observations the difference is negligibly small and due to the way Stata stores data). This may arise if a recipient receives a negative total residual from a donor for which no sectoral allocation can be calculated. Since $DAC2a^{agg}_{RY}$ incorporates this negative amount of aid while the sectoral aid variables do not, the sectoral sum may exceed $DAC2a^{agg}_{RY}$ if the negative residual is not offset by positive residuals from other donors for which the sectoral allocation is also lacking. For these observations $\sum_{s=1}^{S} \widetilde{CRS}^{s}_{RY}$ may also exceed $DAC2a^{agg}_{RY,AD}$, which here is only the case for Panama 1992. Since $\sum_{s=1}^{S} \widetilde{CRS}^{s}_{RY}$ only exceeds $DAC2a^{agg}_{RY,AD}$ and $DAC2a^{agg}_{RY}$ if it does not incorporate negative amounts of aid that are known to have taken place but that I was not able to allocate across sectors, it is likely to exaggerate aid disbursements for the observations where this is the case. As a result, I scale only to the maximum of $DAC2a^{agg}_{RY,AD}$ and $DAC2a^{agg}_{RY}$. This maximum value is labelled $DISB_{RY}$. Consequently, the final measures of sectoral and prefix code aid disbursements are:

$$\widehat{CRS}^{s}_{RY} = DISB_{RY}\left(\frac{\widetilde{CRS}^{s}_{RY}}{\sum_s \widetilde{CRS}^{s}_{RY}}\right) \tag{S1.17}$$

$$\widehat{CRS}^{s,p}_{RY} = DISB_{RY}\left(\frac{\overline{CRS}^{s,p}_{RY}}{\sum_s \widetilde{CRS}^{s}_{RY}}\right) \tag{S1.18}$$

One observation (Cayman islands 1991) has a negative sectoral sum. For this observation the only residual that can be allocated across sectors is negative, whereas for the two donors with a positive residual no sectoral allocation is available. Hence, each sectoral aid variable, and their sum, is negative, whereas $DAC2a^{agg}_{RY}$ is positive. I turn all variables to missing for this observation.

There are ten recipient-year observations with sectoral CRS data but missing DAC2a data. When examining the time series around these observations in more detail, for all but one (Slovenia 1992) it is evident that aggregate CRS disbursements are a lot lower than aggregate DAC2a disbursements in subsequent years. Hence, I choose not to rely solely on the CRS data, which could seriously underestimate the total amount of aid, and instead turn all variables to missing when $DAC2a^{agg}_{RY,AD}$ is missing.

Finally, I drop high-income countries, defined as countries with a 2005 GNI per capita of 10726 US$ or more (following World Bank, 2006). Many of the high-income countries are small islands (e.g. Antigua and Barbuda, Aruba, Netherlands Antilles) or oil exporters (e.g. Kuwait, Qatar, United Arab Emirates). Two remaining observations (Turkey 2000 for education SP aid and Barbados 2001 for health SP aid) are smaller than zero. Since in both cases it concerns extremely small negative values (less than 0.0001% of GDP in absolute value) and since negative aid values are difficult to interpret, I set these observations to zero.

Table S1.1 shows summary statistics for the scaling that takes place in the final step of the data construction (see equations (S1.17) and (S1.18)). $scaling$ is computed as the ratio of the sum of the constructed sectoral disbursements (before scaling) to $DISB_{RY}$:[14]

$$scaling = \frac{\sum_{s=1}^{S} \widetilde{CRS}^{s}_{RY}}{DISB_{RY}} \tag{S1.19}$$

This is compared to the scaling that would take place if I simply scale sectoral CRS disbursements so that their sum matches a measure of total aggregate disbursements, following the logic behind equation (5) in the main text:

$$scaling_{CRS} = \frac{\sum_{s=1}^{S} CRS_{RY}^{s}}{DISB_{RY}} \tag{S1.20}$$

As can be seen from table S1.1, the difference between $scaling$ and $scaling_{CRS}$ is large. On average, the constructed disbursements before scaling make up more than 76% of aggregate, complete disbursements, whereas for CRS disbursements this is only 31.9%. This difference reflects the information added to the sectoral CRS disbursements by the data construction method described in this appendix. For the majority of observations the scaling performed in the final step of the data construction is limited in magnitude and a lot smaller than if CRS sectoral disbursements are scaled without any adjustment. For instance, for more than three quarters of observations CRS disbursements constitute less than half of aggregate aid. For the constructed sectoral disbursements this is the case for less than 10% of observations. This makes it more likely that the sectoral allocation of the aid data before scaling is a reasonable reflection of the actual sectoral allocation one would find if data were complete. This is again the best that can be done with the available data, and not scaling the sectoral disbursements runs the risk of underestimating the amount of aid received.

Table S1.1: Scaling variables

	$scaling$	$scaling_{CRS}$
Observations	2192	2192
Mean	0.768	0.319
Standard deviation	0.191	0.264
Minimum	0.016	0
1st percentile	0.174	0
5th percentile	0.391	0
10th percentile	0.515	0.015
25th percentile	0.656	0.097
Median	0.804	0.258
75th percentile	0.925	0.494
90th percentile	0.981	0.726
95th percentile	0.996	0.843
99th percentile	1	0.981
Maximum	1.128	1

Source: author's analysis based on data described in the text.

Notes

[1]All data used in this appendix can be accessed via the OECD's International Development Statistics (IDS) online databases on aid and other resource flows at www.oecd.org/dac/stats/idsonline.

[2]In CRS, the sector is recorded using a 5-digit purpose code, the first 3 digits of which refer to the corresponding sector in DAC table 5 (see OECD, 2002, Annex 5, pp. 87-106). It is these 3 digits I focus on to demarcate sectors. DAC5 contains a disaggregation of total official development assistance along the same sectors and aid types as CRS, but in a donor-year format, not by recipient (see below for more information).

[3]Other sector aid consists of: population programmes (130), water supply and sanitation (140), government and civil society (150), other social infrastructure and services (160), economic infrastructure and services (200), production sectors (300), multisector/crosscutting (400), emergency assistance (700) and unallocated/unspecified (998).

[4]In addition to the supply of experts, teachers and volunteers, and expenditure on research, equipment and materials, the DAC directive lists the cost of students and trainees, and the financing of development-oriented social and cultural programmes as part of TC (OECD, 2000, pp. 59-62).

[5]Throughout the data construction, tiny discrepancies between totals and their underlying components may arise, even if the former is (re)calculated explicitly as the sum of the latter. This is because Stata stores numbers as binary and many decimal numbers have no exact binary representation, which may lead to small calculation 'errors' (Cox, 2006; Gould, 2006). It would be possible to deal with this by transforming all variables into integers and then transforming them back after the data construction (Gould, 2006). I forego this option, because it adds another layer of complexity and because the discrepancies that arise are negligibly small. I do consistently store variables as 'double' in Stata, so as to keep discrepancies as small as possible.

[6]The dataset still contains 'Serbia & Montenegro, FRY' as a recipient for 1994-2004.

[7]Recall I am working with gross disbursements so these negative amounts of aid do not reflect loan repayments. In the RDY CRS data there is not a single negative observation for the aid variables I distinguish. In the DAC2a dataset 185 out of a total of 43216 RDY combinations are negative for grants and/or loans.

[8]Some of these observations arise because I have excluded donors GFATM and UNFPA from DAC2a, due to the fact that they are absent from DAC5.

[9]Sector allocable aid includes aid for social infrastructure and services (including education and health), economic infrastructure and services, production sectors, and multisector/crosscutting aid. What remains is aid that cannot be allocated across sectors: commodity aid/general programme assistance, action relating to debt, emergency assistance, administrative costs of donors, support to NGOs and unallocated/unspecified aid.

[10]In contrast with the CRS database, DAC5 classifies combinations of prefix codes as ONM (OECD, 2000, p. 118). My decision to instead allocate an equal part of the aid amount to each of the prefix codes that make up the combination in the CRS data should have little effect, though, since only a very small part of education and health aid is listed under a combination of prefix codes.

[11]While bilateral donors' ODA is typically available for all years in DAC5, data for multilateral donors is more patchy. Data for IBRD (International Bank for Reconstruction and Development) and IDA (International Development Association), for instance, is only available for 4 and 5 years in the beginning of the 90s, respectively (IBRD is also missing from DAC2a). In the years with data, the magnitude of aggregate ODA in DAC5 and aggregate disbursements in DAC2a is relatively similar for both bilateral and multilateral donors. Generally speaking, multilateral donors' coverage is worse in CRS as well.

[12]Conversely, there are also observations where aggregate CRS disbursements are zero but a DAC2a total is available that has been allocated across the different sectors. For these recipients with no sectoral CRS data the sectoral disbursements I end up with are based entirely on how the residuals of the donors that deal with this recipient are allocated across sectors.

[13]Some of these 1230 observations involve the two donors (GFATM and UNFPA) that are available in DAC2a but missing from DAC5. With hindsight, I should not have excluded these donors from DAC2a. In fact, I could have included all available donors in DAC2a even if they are absent from DAC5 or CRS and then sum over all donors to obtain aggregate disbursements in a RY format. Before I scale the constructed sectoral disbursements to a plausible measure of aggregate disbursements, however, I also download RY data from DAC2a with 'all donors' as donors, and use this variable as a candidate measure of aggregate disbursements received (see p. 9 below), so the effect of this omission – if any – should be extremely small.

[14]Note the maximum value exceeds one. This is the observation for Panama 1992.

References

Cox, Nicholas J.. 2006. "Stata tip 33: sweet sixteen: hexadecimal formats and precision problems." *Stata Journal* 6(2): 282–283.

Gould, William. 2006. "Mata matters: precision." *Stata Journal* 6(4): 550–560.

OECD. 2000. "DAC Statistical Reporting Directives."

OECD. 2002. "Reporting Directives for the Creditor Reporting System."

World Bank. 2006. "World Development Indicators."