Foreign Aid and Domestic Taxation: Multiple Sources, One Conclusion

Influential research has argued that foreign aid displaces domestic tax revenue when it is given in the form of grants. These claims are based on data that are deeply problematic: several different sources are amalgamated into one dataset, with no apparent checks on compatibility. In this article, a variety of econometric strategies are used to overcome these issues of data quality. The weight of evidence points to a modest but positive effect from all foreign aid on domestic tax revenue. Fears over a negative effect for aid grants appear unwarranted, and are accounted for by the inappropriate use of data or endogeneity concerns.

**Key Words**: Foreign Aid, Tax Revenue, Data Quality, MIMIC model

# Introduction

Should development aid be given in the form of loans or grants? This question has received attention recently, with the House of Commons International Development Committee (2014) arguing that the UK should move away from providing aid as grants towards loans being the overriding default for middle income countries (and commonly used with low income countries). The question is also of some vintage, with Schmidt (1964, p.387) stating that the relative merits of loans and grants had “long been in dispute” over five decades ago. This article discusses a critical element in choosing between them: their reported effect on domestic tax revenue. This perceived difference can be traced back to Gupta et al. (2004), who presented empirical evidence that while aid loans have no negative effects on tax revenue, aid grants do. The obvious policy implication is that donors should be more cautious when providing aid grants than when providing aid loans.

Any negative effect from international aid to domestic tax revenue is a concern in three distinct ways. First, the immediate effect is to reduce the effective value of aid flows for the recipient government, as they would not be additional resources but rather crowd-out tax revenue. In essence aid would finance tax cuts in recipient countries rather than government spending or investment. Second, over the medium term this may have a knock-on effect on the public support for aid. A quote from the House of Commons International Development Committee (2013, p.4) illustrates well how such a resource transfer is perceived: “[w]e cannot expect the people in the UK to pay taxes to improve education and health in Pakistan if the Pakistan elite is not paying income tax.” Third, an even greater fear relates to a potential pernicious effect of aid over a longer time horizon, where it could conceivably undermine governance through fracturing the social contract. Deaton (2013, p.295) argues that “[o]ne of the strongest arguments against large aid flows is that they undermine these constraints, removing the need to raise money with consent and in the limit turning what should be beneficial political institutions into toxic ones.”

In recent times, the evidence base for a differential effect from grants and loans consists of two papers by IMF researchers. Gupta et al. (2004) make a strong claim: aid grants depress domestic tax revenue but aid loans do not. The magnitude of the reported effects were such that for some countries the value of an aid grant was completely offset by the fall in tax revenue. The basis for Gupta et al.’s (2004) claim was disputed by Clist and Morrissey (2011) and Carter (2010, 2013). The former extended the dataset and found that the negative relationship is not present if a) the period 1985-2005 is examined or b) aid is included in specifications with a more reasonable lag. The latter found the results to be fragile to sample, and suggested that endogeneity may be driving the negative relationship between aid grants and tax revenue. Endogeneity is a concern as the composition and volume of aid may be influenced by domestic tax revenue itself. Benedek et al. (2014) can be seen as a follow up to Gupta et al. (2004)[[1]](#footnote-1), and responds to criticism with new methods and data but ultimately presents very similar findings. Here, I examine the evidence base for these claims which influence important decisions about the correct composition of aid, and feeds into the wider debate regarding its effect on governance.

The article proceeds as follows. Section 2 attempts, and fails, to replicate Benedek et al.’s (2014) results using the provided and described data. An advanced discussion of the suitability of GMM techniques is found elsewhere (Carter, 2010, 2013) but the data are found to have more basic weaknesses. These weaknesses, stemming mainly from the use of multiple different sources, are explored in section 3. A variety of solutions are examined in section 4, and section 5 concludes.

# Inability to Replicate Results

I have attempted to replicate Benedek et al.’s (2014) results using the provided dependent variable and listed sources. There are several discrepancies. Limiting the sample to the countries in appendix 1 of Benedek et al. (2014), there are six countries which are apparently included in the regression results but for whom no data exists for the dependent variable (four countries are not in the dataset, two exist but have no data)[[2]](#footnote-2). Comparisons between summary statistics in Benedek et al. (2014) and those from the reported sources reveal several other discrepancies. For example, the paper reports a mean value of 18.75 for trade openness, (exports+imports)/GDP, but a newly constructed dataset using the reported data sources gives a mean of 54.

Table : Determinants of Tax Revenue 1980-2009

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model 1 | Model 2 | Model 1 | Model 2 |
| With Source dummies: | No | Yes | No | Yes | From Benedek et al. |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Aid | 0.005 | 0.006 |  |  | -0.007\* |  |
|  | (1.19) | (1.49) |  |  | (0.004) |  |
| Aid Squared | -0.0001 | -0.0001 |  |  | 0.000 |  |
|  | (-1.27) | (-1.49) |  |  | (0.000) |  |
| Loans |  |  | 0.011\*\* | 0.010\*\* |  | 0.000 |
|  |  |  | (2.33) | (2.16) |  | (0.004) |
| Loans Squared |  |  | -0.0003\* | -0.0003 |  | -0.000 |
|  |  |  | (-1.72) | (-1.64) |  | (0.000) |
| Grants |  |  | 0.003 | 0.005 |  | -0.006\* |
|  |  |  | (0.50) | (0.82) |  | (0.003) |
| Grants Squared |  |  | -0.0001 | -0.0002 |  | 0.000 |
|  |  |  | (-1.03) | (-1.28) |  | (0.000) |
| Agriculture | -0.006\* | -0.006 | -0.007\* | -0.007\* | -0.008\*\*\* | -0.003 |
|  | (-1.67) | (-1.63) | (-1.89) | (-1.88) | (0.002) | (0.003) |
| Industry | -0.007 | -0.006 | -0.002 | -0.003 | 0.000 | 0.000 |
|  | (-1.19) | (-1.29) | (-0.64) | (-0.76) | (0.003) | (0.000) |
| GDP pc (Logged) | 0.248\*\* | 0.230\* | 0.209 | 0.145 | 0.304\*\*\* | 0.305\*\*\* |
|  | (2.09) | (1.88) | (1.62) | (1.23) | (0.122) | (0.122) |
| Trade Openness | 0.003\*\*\* | 0.003\*\*\* | 0.003\*\* | 0.003\*\* | -0.002\*\* | -0.002 |
|  | (2.90) | (3.11) | (2.30) | (2.61) | (0.000) | (0.001) |
| WEO Gen |  | -0.350 |  | 0.081 |  |  |
|  |  | (-1.26) |  | (0.71) |  |  |
| GFS 2001 Cen |  | . |  | 0.155 |  |  |
|  |  | . |  | (1.60) |  |  |
| GFS 2001 Cen |  | -0.275 |  | 0.088 |  |  |
|  |  | (-1.32) |  | (1.07) |  |  |
| GFS 1986 Cen |  | -0.377\* |  | . |  |  |
|  |  | (-1.71) |  | . |  |  |
| GFS 2001 Bud |  | -0.561\*\* |  | -0.192 |  |  |
|  |  | (-2.27) |  | (-1.58) |  |  |
| Afri R M |  | -0.265 |  | 0.183 |  |  |
|  |  | (-0.88) |  | (1.28) |  |  |
| Overall-R-Squared | 0.28 | 0.29 | 0.32 | 0.36 | Not Reported |
| Between-R-Squared | 0.28 | 0.27 | 0.39 | 0.41 | Not Reported |
| Observations | 2174 | 2174 | 1968 | 1968 | 2589 | 2589 |
| Countries | 99 | 99 | 97 | 97 | 118 | 118 |

Note: The dependent variable is total tax revenue divided by GDP, logged. Year dummies and a constant are included but not reported. Estimation is by OLS using country fixed effects, with clustered standard errors. Columns 1-4 provide t-statistics in parentheses, columns 5-6 provide standard errors in parentheses. The source dummies are relative to OECD Central government data, where Cen stands for Central, Gen for General and Bud for Budgetary. ‘African R M’ stands for African Revenue Mobilisation. GFS denotes the IMF Government Finance Statistics, and WEO the World Economic Outlook. 2001 and 1986 refer to different vintages of IMF data. Columns (5) and (6) are taken from Benedek et al. (2014) table 1, columns (1) and (4), where the estimator is also OLS with country fixed effects.

While an accurate replication has proved impossible, the regression results from the attempted replication are found in columns (1) and (3) of Table 1. For ease of reference, Benedek et al. (2014) results are reproduced in columns (5) and (6). These new results concur that tax revenue has a negative association with agriculture and a positive association with income per capita. However, the coefficients reported in (1) and (3) differ in their sign on several key variables: Aid, Aid Squared, Industry and Trade Openness. The least important of these differences relates to industry: the coefficient is insignificant in all specifications. The coefficient for trade openness is positive in Table 1, but was found to be negative by Benedek et al. (2014). Despite apparently using the same sources, there seems to be quite a large difference in the two variables, as shown by comparing summary statistics. Logically, we would expect higher ‘trade openness’ to mean higher tax revenue as a percentage of GDP as import taxes are relatively easy to collect, and so the results here are more intuitively plausible. In later analysis, I follow Clist and Morrissey (2011) who disaggregate ‘trade’ into imports and exports as they logically have different effects.

## A negative effect from aid?

Turning to the coefficients of interest, there is a large disparity between the coefficients found using the new dataset and those reported by Benedek et al. (2014). They find that aid has a negative effect on tax revenue but has *increasing* returns. They find a similar pattern for aid grants, but a positive effect for loans with *decreasing* returns. By contrast, I find the pattern of a positive effect with diminishing returns for all types of aid: total, grants and loans. Benedek et al. (2014) conclude that aid (especially in the form of grants) has a negative association with tax revenue. They acknowledge but do not emphasise that their results imply that higher levels of grants actually have a positive effect on domestic tax revenue collection.

The results here are more in line with expectation (at least regarding non-linearity): aid's positive effect on tax revenue diminishes as it becomes increasingly large relative to GDP. Table 2 reports the implied turning points, taken from Table 1. The turning point for total aid is calculated to be 65% of GDP, meaning that aid can be said to have a negative effect on 0.3% of the sample and a positive effect for the other 99.7%. Grants are found to be more negative than loans when using this metric, but still can only be said to have a negative effect on domestic tax revenue in 1.8% of the sample. The calculation of the turning point has a high degree of uncertainty around it and should be treated with caution. However, it is useful illustratively as it is clear from table 2 that the estimated turning points imply aid has an aggregate benefit on domestic tax revenue for the vast majority of the sample.

Table : Turning Point

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Coefficient | Coefficient Squared | Turning Point | Mean | % of sample above turning point |
| Aid/GDP | 0.005 | -0.00008 | 65% | 8.2% | 0.3% |
| Loans/GDP | 0.011 | -0.0003 | 39% | 2.6% | 0.1% |
| Grants/GDP | 0.003 | -0.0001 | 26% | 5.6% | 1.8% |

Note: Turing points are calculated using the coefficients in columns 1 and 3 in Table 1. The final two columns use the sample years and recipients.

The results from these regressions do not support the conclusion that aid, in whatever form, is negatively associated with domestic tax revenue. It is not clear how the sample or data used here varies from Benedek et al. (2014) as I use their reported sources for independent variables and their provided dataset for the dependent variable, but the difference is meaningful. Perhaps the most important point to notice from the reported results are that the estimated coefficients for total aid and aid grants are not significant. Of the 6 relevant coefficients in (5) and (6) only two are significant, and only then at the 10% level. Indeed, it appears that aid is a relatively small factor in tax revenue. This is consistent with other results such as Gupta et al. (2007); he focuses on determinants of tax revenue and finds aid has a small but significant positive effect.

# Multiple Sources

Clist and Morrissey (2011) reported a system break when extending the dataset of Gupta et al. (2004) from 1970-2000 to 1970-2005. This time effect implied that while Gupta et al. (2004) were right to report a negative relationship, this effect was only relevant in the period 1970-1985, with a positive relationship a better characterisation of the period 1986-2005. Using the reported data sources I cannot replicate the weak negative association found by Benedek et al. (2014) nor the system break of Clist and Morrissey (2011) with the new data and time period. Regressions that augment the independent variables with an interaction term (where aid variables are multiplied by a dummy for post-1984) lack extra explanatory power as the augmented variables are insignificant, as shown in the appendix in table A1. A key element of the time trend may actually have been in using two different data sources as pre-1990 data (termed ‘historical’) was not comparable with post-1990 data[[3]](#footnote-3), but was treated as such. While there appears to be no systematic time trend effect in the data, there are several different sources used.

While Benedek et al. (2014) treat the dependent variable as a single variable, they acknowledge that there are actually several sources. Each of these has their own data definition and methodology and so the data are not strictly comparable. Three different types of data are used: General, Central and Budgetary definitions are treated as the same. The difference in source is compounded by differences in coverage between datasets. The OECD data covers higher-income countries, and so the two samples overlap for only three countries: Chile, Mexico and Turkey. In the majority of cases sources are inconsistent by recipient: on average countries are represented by about 1.8 sources over the 30-year period.

Does the difference in sources matter? Columns (2) and (4) of Table 1 report the regression results from the two models proposed by Benedek et al. (2014) augmented by source dummies. There is clearly some effect of the source: the GFS 2001 Central Government figures appear to be relatively high, and GFS 2001 Budgetary figures are relatively low. These are significant effects, with the latter coefficient much larger than the coefficient on aid loans, for example. However, this table does not represent a particularly elaborate test. If different sources were consistent in their differences, then the problem is easily rectified using source dummies, and the coefficients would be stable.

Table : Estimating the Effects on Tax Revenue by Source, 1980-2009

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Source of the | OECD | WEO | GFS 2001 | GFS 2001 | GFS 1986 | GFS 2001 | African |
| Dependent Variable: | Cen | Gen | Gen | Cen | Cen | Bud | R M |
| Aid | 0.095 | 0.001 | -0.032\*\* | -0.007 | 0.009 | -0.010 | 0.018\* |
|  | (0.28) | (0.30) | (-3.71) | (-1.65) | (0.97) | (-0.94) | (2.69) |
| Aid Squared | -0.012 | -0.000 | 0.001\*\* | 0.000 | -0.000 | 0.000 | -0.000\* |
|  | (-0.06) | (-0.22) | (3.11) | (1.60) | (-1.13) | (0.97) | (-2.71) |
| Agriculture | -0.038 | -0.000 | 0.021\* | -0.004 | -0.018\* | 0.007 | -0.011\*\* |
|  | (-3.19) | (-0.11) | (2.16) | (-0.56) | (-3.13) | (1.04) | (-2.92) |
| Industry | 0.010 | 0.001 | -0.008 | -0.010\* | -0.002 | 0.000 | 0.010\* |
|  | (2.32) | (0.19) | (-1.40) | (-2.41) | (-0.43) | (0.01) | (2.56) |
| GDP pc (Logged) | -0.091 | 0.242\* | -0.922\*\*\* | 0.228 | 0.087 | 0.631 | 0.320\* |
|  | (-0.28) | (2.65) | (-7.68) | (1.08) | (0.39) | (1.89) | (2.28) |
| Trade Openness | -0.006 | 0.003\*\* | 0.004 | 0.001 | 0.001 | 0.004\*\* | 0.003 |
|  | (-0.74) | (3.45) | (1.87) | (0.28) | (0.49) | (3.46) | (1.22) |
| Overall-R-Squared | 0.45 | 0.33 | 0.09 | 0.12 | 0.02 | 0.08 | 0.59 |
| Between-R-Squared | 0.21 | 0.15 | 0.01 | 0.09 | 0.03 | 0.05 | 0.57 |
| Observations | 80 | 783 | 151 | 349 | 124 | 150 | 537 |
| Countries | 3 | 55 | 19 | 32 | 11 | 24 | 31 |

Note: see table 1.

Table 3 presents results from a more elaborate test, which allows the estimated relationship between independent variables and dependent variable to differ by source. If the relationship is robust and the source does not differ systematically we would expect stable parameter estimates. Of the seven sources, we might expect variation for those with few observations or countries, but the results are quite different across the sources. Benedek et al. (2014) report that aid has a negative effect with increasing returns to scale. This pattern is only found when using GFS 2001 data, indeed the two largest datasets find opposite coefficient signs. Most variables are found to be significant in at least one regression or other, but this is not a sign of consistency. Trade Openness stands out as the only variable not to be found significant with both positive and negative signs. To summarise, Table 3 shows that coefficient estimates are not robust to the various sources used and questions the robustness of results in Benedek et al. (2014).

More elaborate techniques, such as GMM, would seem to compound the problem rather than solve it. For example, in the 2,458 cases where the same source is available for two consecutive years for a country, there is an average difference of 1.3% in tax revenue as a percentage of GDP. In the 96 cases where different sources are used for consecutive years, the mean difference is 4.9%. Difference GMM estimations would see large random fluctuations that result from using different sources. These are not controlled for and represent outliers that may erroneously drive the results. While Carter (2013) dealt with many of the more technical aspects of empirical research in this area (concluding that there is little evidence of a negative effect from aid on tax), it appears there are much more fundamental problems with the data that are only compounded by the use of methods such as GMM.

# Four options

Ideally, we would simply have better data. In lieu of this, the second best solution is to take seriously both the data problems which exist and the urgent need for policy advice, which will not always wait for better data to emerge. There are four approaches: to treat the candidate dependent variables as interchangeable, to treat them completely separately, to model their relationship or to use completely new data. In order to explore the validity of the first three options, I have constructed a new dataset which includes all of the independent *and dependent* variables described by Benedek et al. (2014) for 1980-2011 for their chosen sample, from the original sources. These four strategies can be applied more generally in analogous situations, where a variety of overlapping datasets exist.

## Treat variables as interchangeable

Benedek et al. (2014) treat the different variables as essentially interchangeable and replace any missing observations from one dataset with those of any other. The benefit of this approach is that it minimises selection issues by maximising the sample size. However, it is potentially very problematic and at very least introduces abnormally high measurement error. The preceding section showed that sample-specific dummies are significant when added to the main specification used in Benedek et al. (2014), and that running seven separate regressions leads to very different coefficient estimates for the variables of interest. A further complication that is noted here, but not explored further, is that the order in which missing data was replaced may have large effects. This may seem trivial, but with 7 sources there are 5040 possible orderings[[4]](#footnote-4), which is problematic unless the different sources are completely compatible.

As a further test of whether the various sources are measuring the same thing, I use the newly constructed dataset to investigate cases where more than one source exists for any given observation. If the variables are to be treated as interchangeable they must be more than correlated: two variables could be perfectly correlated if one is always double the other but they would not be interchangeable. The requirement for treating the variables as exchangeable is that they must be consistent with each other. As such, I run 20 regressions of the form

$$y\_{i}=α\_{ij}+β\_{ij}y\_{j}+ε$$

where $y\_{i}$ is a candidate dependent variable from source $i$ and $y\_{j}$ is a candidate dependent variable from source $j$ being used as an independent variable. The tests are F-tests of two assumptions, where the null hypotheses represent variables which could be substituted without introducing bias or noise.

$$H\_{0}:α\_{ij}=0 $$

$$H\_{0}:β\_{ij}=1 $$

The results of these tests are displayed as $p$ statistics in Table 4. Columns in Table 4 refer to the dependent variable (i.e. $i$) and rows to the independent variable (i.e. $j$). Of the 40 comparisons shown, Table 4 shows that the null hypothesis is rejected in 33 cases at the 1% level: the different sources do not measure the same thing in the same way.

Table : P Statistics to Test for Agreement Between Sources

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Budgetary GFS | Central GFS | General GFS | WEO | OECD |
| Budgetary GFS | Alpha  | . | 0.00 | 0.00 | 0.00 | 0.00 |
|  | Beta | . | 0.14 | 0.91 | 0.75 | 0.00 |
|  | N | 1691 | 1018 | 497 | 1385 | 23 |
|  |  |  |  |  |  |  |
| Central GFS | Alpha | 0.00 | . | 0.00 | 0.00 | 0.00 |
|  | Beta | 0.01 | . | 0.34 | 0.00 | 0.28 |
|  | N | 1018 | 1797 | 843 | 1161 | 52 |
|  |  |  |  |  |  |  |
| General GFS | Alpha | 0.00 | 0.00 | . | 0.00 | 0.00 |
|  | Beta | 0.00 | 0.00 | . | 0.00 | 0.01 |
|  | N | 497 | 843 | 868 | 676 | 27 |
|  |  |  |  |  |  |  |
| WEO | Alpha | 0.00 | 0.00 | 0.00 | . | 0.64 |
|  | Beta | 0.00 | 0.00 | 0.00 | . | 0.52 |
|  | N | 1385 | 1161 | 676 | 3602 | 57 |
|  |  |  |  |  |  |  |
| OECD | Alpha | 0.00 | 0.00 | 0.00 | 0.00 | . |
|  | Beta | 0.00 | 0.00 | 0.00 | 0.00 | . |
|  | N | 23 | 52 | 27 | 57 | 89 |

Note: The dependent variable is given in the first column, with the independent variable from columns 3-7. P statistics are testing the rejection of the null hypothesis, where p<0.05 indicates rejection at the 5% level. These regressions do not limit the sample to that of Benedek et al. (2014), and so may state a larger number of sources per observation with regard their sample. The results are not symmetrical and are influenced by which variable is designated the dependent.

The differences between sources have been shown in section 3 to make a difference when there is only one source per observation.

Table 4 reports tests regarding whether the various sources measure the same variable, and demonstrates very strongly that they do not meet simple tests of their agreement. In only one case do the tests fail to reject the null hypothesis: when the OECD variable is used as the dependent variable and the WEO variable is used as the independent variable. In the regression where those roles are reversed, the null hypotheses are rejected.

## Different regressions

The second option, if treating the various dependent variables as interchangeable is not sensible, is to treat all variables as completely independent. This avoids needlessly introducing noise into the regressions, though it does potentially introduce problems of sample selection bias that Benedek et al. (2014) avoid. The newly created dataset means that the maximum number of sources for the maximum number of variables was created. Unlike the results in Table 3, any individual country can appear multiple times if the relevant data exist. I make three other changes to the preferred specification of Benedek et al. (2014). First, the specification introduces a one-year lag for aid variables. This is more realistic, as it is difficult to imagine aid having a contemporary effect on tax revenue. It also reduces fears of endogeneity (while not removing them), as tax short-falls could easily lead to short-term contemporary increases in aid, say in the case of a natural disaster. Second, I disaggregate trade into imports and exports as they are likely to have different effects (see Clist and Morrissey, 2011). Specifically, it is to be expected that imports increase tax revenue, as they provide a relatively easy avenue for tax collection. By contrast, exports are usually taxed less than domestic consumption, and so we could expect a negative coefficient given the opportunity cost of high exports includes foregone tax revenue. Third, I move away from estimating nonlinearities in the aid-tax relationship. Given data constrains and quality, it does not seem prudent to attempt to recover precise estimates of turning points. As can be seen from Table 3 estimates of turning points are very unstable, and so the preferred estimation is less ambitious.

Table : Preferred Specification 1980-2011, by data source

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source of Dependent Variable | Budgetary GFS | Central GFS | General GFS | WEO | OECD |
| Loans/GDP,  | -0.000 | 0.004 | -0.020\*\*\* | -0.000 | 0.123 |
| Lagged | (-0.06) | (0.47) | (-3.95) | (-0.21) | (1.07) |
| Grants/GDP | 0.006\* | -0.005 | 0.006 | 0.004\*\* | -0.017 |
| Lagged | (1.76) | (-1.25) | (0.75) | (2.12) | (-0.07) |
| Agriculture | -0.006 | -0.010 | -0.013 | -0.009\*\* | -0.054\* |
|  | (-1.49) | (-1.38) | (-1.35) | (-2.52) | (-3.28) |
| Industry | -0.004 | -0.002 | 0.000 | 0.000 | 0.014\*\* |
|  | (-1.53) | (-0.36) | (0.03) | (0.14) | (6.04) |
| Ln(GDP pc) | 0.349 | 0.237 | 0.173 | -0.134 | -0.268 |
|  | (1.57) | (1.33) | (0.89) | (-1.26) | (-0.89) |
| Imports | 0.008\*\*\* | 0.004 | 0.003 | 0.004\*\*\* | 0.000 |
|  | (4.16) | (1.48) | (1.20) | (3.17) | (0.03) |
| Exports | -0.003 | -0.003 | -0.005 | 0.002 | -0.013 |
|  | (-1.41) | (-0.94) | (-1.42) | (1.34) | (-2.76) |
| Overall-R-Squared | 0.00 | 0.11 | 0.00 | 0.13 | 0.55 |
| Between-R-Squared | 0.03 | 0.11 | 0.03 | 0.06 | 0.40 |
| Observations | 908 | 756 | 254 | 1540 | 77 |
| Countries | 81 | 63 | 35 | 99 | 3 |

Note: See Table 1 notes for details. Here, the dependent variable is given in the first row, using all available data from the reconstructed dataset. I do not include IMF data using the 1986 definition or data from the African Revenue Mobilisation dataset due to availability constraints. Loans and Grants are both measured as a percentage of GDP, logged and lagged.

What can be learnt from the results reported in Table 5? It is heartening that despite problems with the dependent variable, coefficients on the variables agriculture, industry, exports and imports tend to agree. The most consistent coefficient estimate is that for agriculture, which is always found to be negative. The last column, relating to the OECD source, has fewer observations and is the least similar to others. As in Clist and Morrissey (2011), the decision to split trade into imports and exports is justified by the coefficients on the two variables being of opposite signs. The signs are in line with theoretical predictions: imports are relatively easily to tax and exports lead to foregone domestic consumption, which would have been taxed. The log level of per capita income is insignificant in every regression.

Turning to the variables of interest, aid grants and loans are significant in only 3 of a possible 12 cases. Both are found to be positive and negative in difference specifications. However, it is remarkable that in no specification is the coefficient on aid grants negative and significant. In fact, the only significant effects for aid variables are *positive* for grants and *negative* for loans.

## MIMIC model

The third option represents a middle way between treating the different possible dependent variables as interchangeable and treating them separately. A multiple indicators multiple cause (MIMIC) model can be applied in this case. The model was introduced by Jöreskog (1975), and an example of its use can be found in Giles (1999). To describe briefly, the model states that there is an unobserved latent variable $y^{\*}$ (tax revenue), but multiple indicators $y\_{i},…,y\_{m}$ are observed (the various candidates for the dependent variable). The MIMIC model then considers two sorts of relationships. For the relationship between the causes and the latent variable, an equation of the following form is estimated:

$y^{\*}=γ\_{1}x\_{1},… , γ\_{k}x\_{k}+ϵ$ (1)

The relationship between the latent variable and the indicators is also modelled, using equations of the form:

$y\_{1}=a\_{1}+ β\_{1}y^{\*}+u\_{1}$

$…,$ (2)

$y\_{m}=a\_{m}+ β\_{m}y^{\*}+u\_{m}$

These are then jointly estimated using maximum likelihood[[5]](#footnote-5). Figure 1 is perhaps useful to visualise the model. The benefit of this approach should be seen predominately as a robustness check on the results presented in Table 5. The MIMIC model allows multiple candidate dependent variables to be used simultaneously, and can provide a larger sample than those using one candidate at a time.

Figure : The MIMIC model



One complication of using MIMIC models when there is not absolute overlap in sources it may be that models do not converge. Only one combination of sources works, and so Table 6 reports the coefficients from that regression. A further complication is that it is not feasible to use fixed effects in this case.

Table : MIMIC Model, 1980-2011

|  |  |  |
| --- | --- | --- |
| Equation (2) |  | Equations of the From (1) |
| Aid Loans | 0.002 |  | Dependent Variable:  | Bud. Tax Revenue (GFS) |
| Lagged | (0.40) |  | Latent Variable | 1.000 |
| Aid Grants | 0.029\*\*\* |  |  | (constrained) |
| Lagged | (8.68) |  | Constant | -3.028\*\*\* |
| Agriculture | -0.011\*\*\* |  |  | (-15.71) |
|  | (-5.25) |  | R Squared | 0.22 |
| Industry | 0.007\*\*\* |  | Dependent Variable:  | Revenue (WEO) |
|  | (4.19) |  | Latent Variable | 0.806\*\*\* |
| Ln(GDP pc) | 0.104\*\*\* |  |  | (9.87) |
|  | (5.56) |  | Constant | 2.205\*\*\* |
| Imports | 0.005\*\*\* |  |  | (16.19) |
|  | (6.19) |  | R Squared | 0.81 |
| Exports | 0.004\*\*\* |  | Log Likelihood | 69426.71 |
|  | (3.39) |  | Observations | 3281 |
| R Squared | 0.47 |  | Countries | 114 |

Note: In Equation (2) on the left panel, the dependent variable is the estimated latent variable. This is constrained in the first equation in the form (1), such that the coefficient on the slope of GFS tax revenue is equal to one. Thus the latent variable can be thought of as the underlying concept of tax revenue, with two indicators. Z statistics are shown below parameter estimated in parentheses. R squared is calculated on an equation basis, and refers to the equation above it.

Bearing in mind the various caveats regarding the difficulty in model convergence, I present only one converged specification which uses two candidate dependent variables: the GFS's Budgetary Tax Revenue and the WEO's Revenue variables. Table 6 contains the full results for all equations. One estimated coefficient must be constrained in the model, and so $\hat{β}$ for the latent variable's effect on BA Tax Revenue is constrained to equal 1. The benefit of this approach is evident in the sample size, as the sample is much larger than those presented in table 5. Dealing first with the equations of the form (1) on the right panel, we can see that both indicators are positively correlated with the latent variable. The WEO variable is 5 units (where the variables are expressed as a percentage of GDP, logged) higher. The coefficient on the latent variable in the WEO equation is less than one, meaning that revenue is estimated to be less affected by tax revenue than the GFS measure. Moving to the left panel, we see familiar and reassuring parameter estimates in most cases. For the variables of interest, we see that both types of aid are consistent with higher tax receipts in the following year, but the result is only significant (at the 1% level) for aid grants. The control variables are almost all as expected and all significant: higher tax receipts are associated with less agriculture, more industry, being richer and importing more. The only variable which is not in line with expectation is that related to exports, where a positive effect is found.

## New data

A fourth option in attempting to replicate the results of Benedek et al. (2014) has recently become available due to the newly created International Center for Tax and Development Government Revenue Dataset (ICTD GRD). This new dataset is the culmination of a three-year effort to provide a consistent and accurate estimate of tax revenue in developing countries. Crucially, there was an emphasis on compatible sources in the construction of this dataset to allow comparisons across countries and time periods. Table 7 provides the results of applying my preferred specification (used in Table 5) to the new data. This exercise provides several benefits. First, it provides a further test of previous research in this area, by using what appears to be the most accurate dataset available. In maintaining a comparable specification with minimal adjustments, the new data can be used to evaluate previous findings in this area. Second, as the new dataset is used to tackle new issues, it is useful to benchmark it against older datasets using a familiar econometric specification.

Table : ICTD Dataset, 1980-2009

|  |  |  |  |
| --- | --- | --- | --- |
| Aid Variables: | Concurrent | Once Lagged | Twice Lagged |
|  | (1) | (2) | (3) |
| Loans/GDP,  | 0.002 | 0.004 | 0.002 |
| Lagged | (0.76) | (1.54) | (0.59) |
| Grants/GDP | -0.004\* | -0.001 | -0.001 |
| Lagged | (-1.78) | (-0.76) | (-0.49) |
| Agriculture | -0.007\*\* | -0.006\*\* | -0.005\* |
|  | (-2.39) | (-2.24) | (-1.94) |
| Industry | -0.005\* | -0.005\*\* | -0.006\*\* |
|  | (-1.84) | (-1.99) | (-2.25) |
| Ln(GDP pc) | 0.043 | 0.081 | 0.064 |
|  | (0.52) | (0.93) | (0.75) |
| Imports | 0.008\*\*\* | 0.006\*\*\* | 0.006\*\*\* |
|  | (5.40) | (3.81) | (3.82) |
| Exports | -0.003\* | -0.003\* | -0.003\* |
|  | (-1.91) | (-1.69) | (-1.77) |
| Overall-R-Squared | 0.39 | 0.38 | 0.38 |
| Between-R-Squared | 0.41 | 0.42 | 0.42 |
| Observations | 2173 | 2138 | 2102 |
| Countries | 113 | 112 | 112 |
|  |  |  |  |

Note: As for table 1, but with the dependent variable from the International Center for Tax and Development Government Revenue Dataset (ICTD GRD).The aid variables are either concurrent, or lagged one or two years.

Table 7 has the now familiar pattern for the majority of coefficients: tax revenue is positively associated with having higher imports, a smaller agriculture sector, lower exports and a smaller industrial sector. Of the six coefficients on the aid variables, only the coefficient for aid grants in column (1) is significant, with a negative sign. When a lag of one or two years is used however, the coefficient for aid grants is both smaller and insignificant. This specification and data provide the only negative and significant effect of aid grants on domestic tax revenue I have found. However, a negative contemporary association is not necessarily evidence that aid grants cause lower domestic tax revenue. In any given year a contemporary association is more likely to be the result of aid donors compensating recipients in adverse conditions (for example an economic downturn). If aid does corrode tax revenue, it is more likely to do so over a period of time, and yet the effect becomes insignificant over this time period.

# Discussion and Conclusion

Replication has long been recognised as performing a vital role in assessing the credibility of empirical results (see Camfield and Palmer-Jones, 2013, for a recent review of the issues). For this reason I have attempted to replicate the results presented by Benedek et al. (2014) as it is the most high profile recent empirical research paper that reports differential effects from aid grants and loans on tax revenue. This empirical result underpins one of the main perceived differences between aid grants and aid loans, and so has real consequences for the delivery of development finance. However like Carter (2013), I have failed to replicate the specific results. In only one case have I found a reported negative effect for aid grants: when contemporary aid grants are used. The specific replication failed even when using the exact dataset for the dependent variable reportedly used by Benedek et al. (2014).

While Carter (2013) provides an extensive critique of the methods used by Benedek et al. (2014) to combat endogeneity (mainly the use of GMM), I argue that there appear to be much more fundamental issues with the original data used. Specifically, several datasets are apparently used to construct a single dependent variable despite different data definitions. Tests show that previous specifications are not robust to the use of source dummies or running separate regressions by the dependent variable's source. Furthermore, the different candidate dependent variables are statistically and conceptually measuring different things. The use of system GMM in this setting is particularly worrying, as large yearly differences may simply be due to the use of different sources in consecutive years. Jerven (2013) has drawn needed attention to the issue of data quality, but this is not a counsel of despair. Occasionally, careful econometrics can be used to confront data quality issues head on. In this case, the number of overlapping datasets is the fundamental issue, and econometric techniques can be employed to deal with this problem.

Given the data quality, it is heartening that the various empirical approaches employed here point in a single direction with clear policy implications. In short, there is no convincing evidence that aid grants undermine domestic tax revenue, so aid donors making a decision between grants and loans do not need to factor in a differential effect on tax revenue. The only significant and negative coefficient relating to aid in Table 5 and Table 6 actually relates to aid loans. The only significant and negative coefficient for aid grants relates to contemporaneous aid, and is thus dogged by endogeneity concerns. The weight of the empirical evidence suggests that aid has a relatively small, possibly positive, influence on domestic tax revenue. There is not sufficient evidence to suggest that the composition of aid should be influenced by concerns relating to differential effects on tax revenue. Wider concerns relating to a longer-term corrosive effect on government cannot be dismissed on the basis of these results (even if aid leads to higher tax revenue, it may be that tax is a lower proportion of total government revenue than it would be without aid), but neither can be used to support them.

# Bibliography

Benedek, D., Crivelli, E., Gupta, S., & Muthoora, P. (2014) Foreign Aid and Revenue: Still a Crowding-Out Effect?, *FinanzArchiv*, 70: 67-96.

Camfield, L. and Palmer-Jones, R. (2013) Three `Rs' of Econometrics: Repetition, Reproduction and Replication, *Journal of Development Studies*, 49(12): 1607-1614.

Carter, P. (2010) Foreign aid and taxation, revisited. CSAE Conference Paper, 180.

Carter, P. (2013) Does Foreign Aid Displace Domestic Taxation? *Journal of Globalization and Development*, 4(1):1-47.

Clist, P. and Morrissey, O. (2011) Aid and tax revenue: Signs of a positive effect since the 1980s. *Journal of International Development*, 23(2):165-180.

Deaton, A. (2013) *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton: Princeton University Press.

Giles, D. E. (1999) Measuring the hidden economy: Implications for econometric modelling. The Economic Journal, 109(456): 370-380.

Gupta, A. S. (2007) Determinants of Tax Revenue Efforts in Developing Countries. IMF Working Paper, 184.

Gupta, S., Clements, B., Pivovarsky, A., and Tiongson, E. (2004) *Helping Countries Develop: The Role of Fiscal Policy, chapter Foreign Aid and Revenue Response: Does the Composition of Aid Matter?*, Mimeo, Washington: International Monetary Fund.

House of Commons International Development Committee (2013) *Pakistan,* Tenth Report of Session 2012-13, Volume 1, London: The Stationery Office Limited.

House of Commons International Development Committee (2014) *The Future of UK Development Cooperation: Phase 1: Development Finance*, Eighth Report of Session 2013–14, London: The Stationery Office Limited.

IMF (2001) Government Finance Statistics Manual 2001. IMF: Washington, DC.

Jerven, M. (2013) *Poor numbers: how we are misled by African development statistics and what to do about it*, New York : Cornell University Press.

Jöreskog, K. G. and Goldberger, A. S. (1975) Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351): 631-639.

Knack, S. (2001) Aid dependence and the quality of governance: cross-country empirical tests, *Southern Economic Journal*, 310-329.

Schmidt, W. (1964) The Economics of Charity: Loans Versus Grants, *The Journal of Political Economy*, 72(4): 387-395.

# Appendix

Table A1 provides evidence that, as suspected, there is no system break in the data provided by Benedek et al. (2014).

Table A1: System Break

|  |  |  |
| --- | --- | --- |
| Variable | Model 1 | Model 2 |
| Aid | 0.000 |  |
|  | (0.03) |  |
| \*dummy | 0.005 |  |
|  | (0.64) |  |
| Aid Squared | -0.000 |  |
|  | (-0.28) |  |
| \*dummy | -0.000 |  |
|  | (-0.30) |  |
| Loans |  | -0.018\* |
|  |  | (1.73) |
| \*dummy |  | -0.009 |
|  |  | (-0.80) |
| Loans Squared |  | -0.000\*\* |
|  |  | (-2.10) |
| \*dummy |  | 0.000 |
|  |  | (0.99) |
| Grants |  | -0.005 |
|  |  | (-0.43) |
| \*dummy |  | 0.010 |
|  |  | (0.84) |
| Grants Squared |  | 0.000 |
|  |  | (0.06) |
| \*dummy |  | -0.000 |
|  |  | (-0.74) |
| Agriculture | -0.006\* | -0.007\* |
|  | (-1.80) | (-1.95) |
| Industry | -0.007 | -0.003 |
|  | (-1.19) | (-0.67) |
| Ln(GDP pc) | 0.251\*\* | 0.212 |
|  | (2.12) | (1.64) |
| Trade Openness | 0.003\*\*\* | 0.003\*\* |
|  | (3.03) | (2.38) |
| Overall R-Squared | 0.28 | 0.32 |
| Between R-Squared | 0.28 | 0.39 |
| Observations | 2174 | 1968 |
| Countries | 99 | 97 |

*Note: `\*dummy' denotes the preceding variable multiplied by a dummy which takes the value 1 if the year is 1985 or later. For all other details, see Table 1.*

# Data Sources

All data and Stata code can be accessed from <https://sites.google.com/site/paulclist/data>

*Tax Revenue, Organisation of Economic Cooperation and Development (OECD)* total tax revenue as a percentage of GDP.

*Tax Revenue, Government Finance Statistics (GFS)* Two vintages are used by Benedek et al. (2014): 1986 and 2001. These refer to two different data definitions. Furthermore, three types of variable are used: Budgetary, Central Government or General Government.

*Tax Revenue, World Economic Outlook (WEO)* is a non-public variable with unclear definition that is expressly not designed for cross-country comparison.

*Aid: Total, Grants and Loans* Net ODA, ODA Grants, and ODA Loans, relative to GDP, are from the OECD database.

*Agriculture and Industry* These are taken from the World Bank's World Development Indicators (WDI), and describe the added value as a percentage of GDP.

*Trade Openness, Exports and Imports* The numerator is taken from the IMF's International Finance Statistics (IFS) database. The denominator (GDP) is taken from the World Economic Outlook.

*GDP per capita* is taken from the World Bank's World Development Indicators (WDI) database, and calculated in constant (2000) U.S. dollars.

1. Both are written by IMF-based researchers, and they share one author. [↑](#footnote-ref-1)
2. In later correspondence, another dataset *was* made available which included data for these six countries. However, that dataset gave no information on the underlying sources, and so the original dataset is used here. Regressions using the latter dataset also fail to replicate the reported results. [↑](#footnote-ref-2)
3. The IMF (2001, pp.3-4) states that “The methodology for compiling government finance statistics described in this [2001] manual differs substantially from the methodology of the 1986 GFS Manual.” [↑](#footnote-ref-3)
4. This is due to 7!=5040 capturing all possible permutations. [↑](#footnote-ref-4)
5. In Stata, the relevant command is sem. All Stata code and data are freely available at https://sites.google.com/site/paulclist/data [↑](#footnote-ref-5)