The Problem of Missing Data in the Econometrics of Foreign Aid[1]

DRAFT. Do Not Cite. Results Subject to Change.

Robert L. Hicks[2] and Kurt E. Schnier[3]

September 2008

Abstract: In this paper we examine the impact of missing data for models of foreign aid allocation. We explore two types of missing data problems inherent with many models in the development literature. First, we examine the impact of estimating models over an incomplete set of donors by leveraging the many Non-DAC bilaterals and multilaterals in the PLAID database. Second, we examine how the extent of missing recipient characteristics impacts inference and policy conclusions. We do this by treating missingness using a variety of different techniques including formal imputation methods and other *ad hoc* approaches in use in the literature today. Our results show that......

JEL Classification: Need This

Keywords: Foreign Aid, Missing Data, Aid Allocation

[1]We thank x, y, and z and seminar participants at ZZZ University for their insightful comments. Usual disclaimers apply.

[2]Department of Economics, College of William and Mary, Williamsburg VA 23187. Ph: (757) 221-3707; E-mail: rob.hicks@wm.edu

[3]Department of Economics, Georgia State University, Atlanta GA ???. Ph: (404) 413-0159; E-mail: kschnier@gsu.edu

1 Introduction

In this paper we examine the impact of missing data for models of foreign aid allocation. We explore two types of missing data problems inherent with many models in the development literature. First, we examine the impact of estimating models over an incomplete set of donors by leveraging the many Non-DAC bilaterals and multilaterals in the PLAID database. Second, we examine how the extent of missing recipient characteristics impacts inference and policy conclusions. We do this by treating missingness using a variety of different techniques including formal imputation methods and other *ad hoc* approaches in use in the literature today. Our results show that......

The foreign aid literature has empirically examined the allocation of foreign aid from donor countries and organizations to recipient countries. Beginning with Dudley and Montmarguette (1977), the literature has focused on the choices made by a donor as it ostensibly seeks to maximize the impact of its aid portfolio. The literature has focused primarily on "rich" donor nations (cites here) whose official development assistance is found in the OECD DAC database, there has been some efforts to examine non-DAC financial flows (cite Neumeyer). In parallel, this same aid allocation literature has grappled with the difficult issue of missing data as it relates to recipient characteristics (and recipient-donor characteristics describing degrees of interrelatedness (e.g. trade between donor and recipient)). Taken together, the missing data associated with (1) missing flows from donors in the analysis and (2) missing recipient characteristics important for characterizing the allocation decisions of donors may lead to imprecise parameter estimates and poor policy guidance from estimated aid allocation models. Further we note that many of the issues arising here and estimation approaches we employ are directly applicable to models of aid effectiveness.

2 The Literature

There have been numerous models of aid allocation beginning with the pioneering work of Dudley and Montmarguette (1977), and continuing on to present [McGillivray and Oczkowski, 1991, McGillivray, 1992, McGillivray and Oczkowski, 1992, McGillivray, 2003, McGillivray, 2004]. Additionally, there has been some aid effectiveness work (e.g. Burnside and Dollar; ??, Easterly, Levine, Roodman) that has included aid allocation models in order to deal with endogeneity. When examining this literature, several points emerge

1. A subset of donors- the subset contained in the OECD DAC database- is examined. Other bilateral and multilateral donors, as well as private flows are largely ignored. **??**

2. When additional donors are included, they are often examined piecemeal and not by systematically expanding the coverage of donors in the econometric models (e.g. Neumeyer in Arab Aid).

3. When modeling donor allocation decisions, recipient and donor-recipient characteristics are included in the model. Oftentimes there are severe missing data problems for measures we might want in our models. To deal with these problems a variety of approaches have been employed including **??**

   (a) Listwise deletion: completely dropping from the analysis recipient countries with less than a full set of data

   (b) Multiple year moving averages: to smooth data and to eliminate holes in data series, three year moving averages are often employed. This effectively is like imputing missing values based on what in the near recent future has happened in the country. Countries having far fewer datapoints are assigned "operative" data without recognition that some of these imputations occur with more non-missing data than others.

   (c) Formal imputation methods. These methods formally seek to assign missing values by exploiting the underlying structure of the observed data. Hicks et al. used MCMC imputation methods in their models of aid allocation, yet the use of such methods in the presence of rather severe missing data problems has not been systematically examined in a development assistance setting.

In general there ought to be a concern about how we deal with missing values- whether from the standpoint of **??** or **??**

3 Discussion of Patterns of Missing Data

Include summary data on missing data, correlation between missing data and aid flows, show maps of where missing data is the biggest problem, explore correlation of missing data with observed data (e.g. GDP and missing data, institutions and missing data).

4 Methodology

Given the large number of missing observations within aid data, data imputation may improve our empirical estimates. To investigate this hypothesis we have developed five different data imputation methods to be utilized: imputed means, nearest neighbor propensity score matching, and weighted propensity score matching. Additionally, we run models using listwise deletion, that simply drops observations with *any* missing data. Each of these methods is described below and they are compared to two empirical models that do not utilize imputation methods: linear regression using full data and list-wise deletion of missing observations. The former regression model treats all missing values as zeros, whereas the later is the most commonly used method in aid data. These methods are compared using a simplified Monte Carlo experiment discussed in the Section ?? as well with an aid data set on —-.

4.1 Monte Carlo Anaysis

Consider a data set with N columns and J rows of independent variables denoted X_{ij}, $i = 1, ...N, j = 1toJ$ with the missing observations containing at least one missing element of X. In addition define the column vector P_{ij} that takes a value of 1 if the an observation is missing and zero otherwise. Given this notation the following sub-sections will explain the data imputation methods utilized.

4.1.1 Imputed Means

The imputed means model estimates the mean of $X_{ij|Obs}$ for each column in the data set and then imputes the mean value for all observations . Although this method is the simplest routine to utilize for data imputation, it does not allow of the other covariates to influence the imputed values. The other imputation methods allow for this correlation. We are planning to extend this to make it conditional on donor and recipient perhaps.

4.1.2 Nearest Neighbor Propensity Score Matching

Nearest neighbor propensity score matching conducts a probit regression of P_ij on the other non-missing covariates within the empirical model, as well as the indicator columns to control for the missing values of X_{kj} which are set equal to zero. Following estimation of the probit regression, the propensity score for each observation is calculated, denoted S_{ik} , is estimated as the probability an observation is missing conditional on the other covariates. Propensity scores are then divided depending on whether or not the observation is observed or not, $S_{ik|Miss}$ and $S_{ik|Obs}$ respectively. For each missing observation in $X_{ik|Miss}$ the distance between the missing values propensity score and all the propensity scores for the observed data is calculated as

$$Dist_e = \|S_{ij|miss} - S_{ej|Obs}\|, \forall e \in X_{ij|Obs} \tag{1}$$

Following the estimation of the distance between the propensity scores, the observed observation that possesses the lowest distance estimate is imputed for the missing value. This is conducted for all the missing observations within each column separately until all the missing values have been imputed.

4.1.3 Weighted Propensity Score Matching

The weighted propensity score matching model utilizes the propensity scores estimate to created a weighted average of the nearest neighbors to a missing observation. In the imputed means model all of the observed observations were assigned equal weight in their impact on the imputed missing values. With the weighted the propensity score matching model those observed data points that are closer missing data point are assigned more weight in the imputation. The imputed value for missing observation $X_{ik|Miss}$ is calculated as follows,

$$X_{ij|Miss} = \sum_{z=1}^{J_{Obs}} \frac{Dist_z}{\sum_{j=1}^{J_{Obs}} Dist_j} X_{ij|Obs} \tag{2}$$

where denotes the observed observations within column i. If the then the weighted propensity score will generate the same imputed values as though in the imputed means model.

4.1.4 Implementation

These results were generated from drawing from independent uniform variates to simulate missing data. We plan on changing this to a more plausible distribution (multivariate normal with nonzero off-diagonal elements in the covariance matrix), but haven't done it yet. We also will add linear regression interpolation, ordered propensity score matching, the EM algorithm for imputation, and MCMC multiple imputation.

We estimate everything using a Cragg Model for the results outlined in Hicks et al. This is a two-stage model, is inefficient, but is very robust in terms of convergence, allows for parameters for to have differences in signs for the gatekeeping and continuous parts of the aid allocation equation. It also does not have identification problem inherent with the Heckman model.

5 Results in the context of Foreign Aid

See these tables. No time to add a discussion.

We are running elasticities now, that captures both stages of the estimation.

6 Conclusion

This stuff is just to demonstrate that the citations are working. [Agosin and Machado, 2007, Ali and Isse, 2006, Anselin and Bera, 1998, Berthelemy and Tichit, 2004, Berthelemy, 2006, Feeny and McGilli Hagen, 2006b, Hagen, 2006a, Halonen-Akatwijuka, 2007, Irwin and Bockstael, 2004, Kenny, 2008, Mascarenhas and Sandler, 2006, **?**, Nath and Sobhee, 2007, Neumayer, 2003, Trumbull and Wall, 1994, White, 1992, Wood, 2008, Younas, 2008]

References

[Agosin and Machado, 2007] Agosin, M. R. and Machado, R. (2007). Openness and the international allocation of foreign direct investment. *Journal of Development Studies*, 43(7):1234–1247. Geographic: Selected Countries; Added: 200802.

[Ali and Isse, 2006] Ali, A. M. and Isse, H. S. (2006). An empirical analysis of the determinants of foreign aid: A panel approach. *International Advances in Economic Research*, 12(2):241–250. Geographic: Selected Countries; Added: 200702.

[Anselin and Bera, 1998] Anselin, L. and Bera, A. K. (1998). *Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics*. Source: In: Handbook of applied economic statistics; 237-89; Statistics: Textbooks and Monographs, vol. 155. New York; Basel and Hong Kong: Dekker,; Date Added: 200012.

[Berthelemy, 2006] Berthelemy, J.-C. (2006). Bilateral donors' interest vs. recipients' development motives in aid allocation: Do all donors behave the same? *Review of Development Economics*, 10(2):179–194. Geographic: Selected Countries; Added: 200609.

[Berthelemy and Tichit, 2004] Berthelemy, J.-C. and Tichit, A. (2004). Bilateral donors' aid allocation decisions–a three-dimensional panel analysis. *International Review of Economics and Finance*, 13(3):253–274. DOI: 10.1016/j.iref.2003.11.004; Geographic: Selected Countries; Added: 200407.

[Feeny and McGillivray, 2008] Feeny, S. and McGillivray, M. (2008). What determines bilateral aid allocations? evidence from time series data. *Review of Development Economics*, 12(3):515–529. Geographic: Selected Countries; Added: 200809.

[Hagen, 2006a] Hagen, R. J. (2006a). Buying influence: Aid fungibility in a strategic perspective. *Review of Development Economics*, 10(2):267–284. Added: 200609.

[Hagen, 2006b] Hagen, R. J. (2006b). Samaritan agents? on the strategic delegation of aid policy. *Journal of Development Economics*, 79(1):249–263. DOI: 10.1016/j.jdeveco.2004.11.009; Added: 200603.

[Halonen-Akatwijuka, 2007] Halonen-Akatwijuka, M. (2007). Coordination failure in foreign aid. *B.E.Journal of Economic Analysis and Policy: Topics in Economic Analysis and Policy*, 7(1):1–38. Added: 200712.

[Irwin and Bockstael, 2004] Irwin, E. G. and Bockstael, N. E. (2004). Land use externalities, open space preservation, and urban sprawl. *Regional Science and Urban Economics*, 34(6):705–725. DOI: 10.1016/j.regsciurbeco.2004.03.002; Geographic: U.S. Region: Northern America; Added: 200411.

[Kenny, 2008] Kenny, C. (2008). What is effective aid? how would donors allocate it? *European Journal of Development Research*, 20(2):330–346. Added: 200809.

[Mascarenhas and Sandler, 2006] Mascarenhas, R. and Sandler, T. (2006). Do donors cooperatively fund foreign aid? *Review of International Organizations*, 1(4):337–357. Geographic: Selected Countries; Added: 200810.

[McGillivray, 1992] McGillivray, M. (1992). The allocation of aid among developing countries: A multi-donor analysis using a per capita aid index. *World Development*, 20(11):1699–1702. Geographic: LDCs MDCs; Added: 199306.

[McGillivray, 2003] McGillivray, M. (2003). Modelling aid allocation: Issues, approaches and results. *Journal of Economic Development*, 28(1):171–188. Geographic: U.S. Region: Northern America; Added: 200309.

[McGillivray, 2004] McGillivray, M. (2004). Descriptive and prescriptive analyses of aid allocation: Approaches, issues, and consequences. *International Review of Economics and Finance*, 13(3):275–292. DOI: 10.1016/j.iref.2003.11.005; Geographic: Selected Countries; Added: 200407.

[McGillivray and Oczkowski, 1991] McGillivray, M. and Oczkowski, E. (1991). Modelling the allocation of australian bilateral aid: A two-part sample selection approach. *Economic Record*, 67(197):147–152. Geographic: Australia Region: Oceania; Added: 199203.

[McGillivray and Oczkowski, 1992] McGillivray, M. and Oczkowski, E. (1992). A two-part sample selection model of british bilateral foreign aid allocation. *Applied Economics*, 24(12):1311–1319. Geographic: U.K. Region: Europe; Added: 199306.

[Nath and Sobhee, 2007] Nath, S. and Sobhee, S. K. (2007). Aid motivation and donor behavior. *American Review of Political Economy*, 5(1):1–13. Added: 200805.

[Neumayer, 2003] Neumayer, E. (2003). What factors determine the allocation of aid by arab countries and multilateral agencies? *Journal of Development Studies*, 39(4):134–147. Geographic: Arab Countries; Added: 200308.

[Trumbull and Wall, 1994] Trumbull, W. N. and Wall, H. J. (1994). Estimating aid-allocation criteria with panel data. *Economic Journal*, 104(425):876–882. Geographic: Global; Added: 199503.

[White, 1992] White, H. (1992). The allocation of aid among developing countries: A comment. *World Development*, 20(11):1697–1698. Geographic: LDCs MDCs; Added: 199306.

[Wood, 2008] Wood, A. (2008). Looking ahead optimally in allocating aid. *World Development*, 36(7):1135–1151. DOI: 10.1016/j.worlddev.2007.08.002; Added: 200808.

[Younas, 2008] Younas, J. (2008). Motivation for bilateral aid allocation: Altruism or trade benefits. *European Journal of Political Economy*, 24(3):661–674. DOI: 10.1016/j.ejpoleco.2008.05.003; Geographic: OECD LDCs; Added: 200811.

Tables and Figures

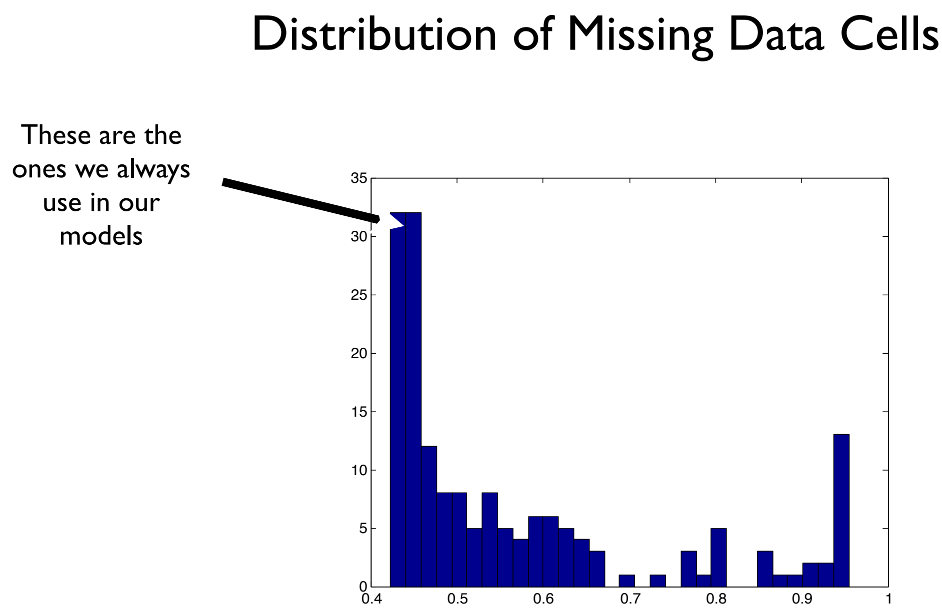

Figure 1

$N = 1000$; $M = 100$; $selection = 10\%$; $corr = 0$

Parameter	OLS	List-wise	Imputed	Linear	Nearest	Ordered	Weighted
Bias							
β_0	-0.0688	0.1005	0.0256	0.0249	0.0013	0.0003	0.0256
β_1	0.2351	-0.0186	-0.0075	-0.0072	0.0879	0.0898	-0.0075
β_2	-0.1240	-0.0120	-0.0151	-0.0137	-0.0621	-0.0567	-0.0152
β_3	0.1035	-0.0092	-0.0165	-0.0168	0.0341	0.0340	-0.0165
β_4	-0.1104	-0.1084	-0.0072	-0.0069	-0.0586	-0.0587	-0.0070
RMSE							
β_0	0.1147	0.1612	0.1254	0.1260	0.1084	0.1150	0.1254
β_1	0.2593	0.1352	0.1260	0.1266	0.1480	0.1506	0.1259
β_2	0.1660	0.1518	0.1340	0.1344	0.1383	0.1392	0.1340
β_3	0.1431	0.1230	0.1140	0.1143	0.1109	0.1190	0.1139
β_4	0.1499	0.1554	0.1158	0.1169	0.1264	0.1225	0.1158

$N = 1000$; $M = 100$; $selection = 25\%$; $corr = 0$

Parameter	OLS	List-wise	Imputed	Linear	Nearest	Ordered	Weighted
Bias							
β_0	-0.1301	0.1803	0.0401	0.0390	-0.0385	-0.0302	0.0406
β_1	0.4304	-0.0182	-0.0032	-0.0029	0.2495	0.2394	-0.0030
β_2	-0.2217	-0.0482	-0.0296	-0.0280	-0.1436	-0.1520	-0.0296
β_3	0.1907	-0.0142	-0.0297	-0.0315	0.1109	0.1080	-0.0297
β_4	-0.2161	-0.1931	-0.0145	-0.0121	-0.1333	-0.1264	-0.0154
RMSE							
β_0	0.1500	0.2231	0.1348	0.1356	0.1256	0.1207	0.1351
β_1	0.4450	0.1651	0.1338	0.1364	0.2724	0.2654	0.1333
β_2	0.2450	0.1913	0.1480	0.1502	0.1886	0.2046	0.1482
β_3	0.2171	0.1756	0.1327	0.1339	0.1589	0.1664	0.1322
β_4	0.2403	0.2413	0.1168	0.1192	0.1740	0.1617	0.1167

$N = 1000$; $M = 100$; $selection = 50\%$; $corr = 0$

Parameter	OLS	List-wise	Imputed	Linear	Nearest	Ordered	Weighted
Bias							
β_0	-0.1930	0.1957	0.0446	0.0465	-0.1690	-0.0960	0.0442
β_1	0.6188	0.0099	0.0031	0.0002	0.4970	0.4773	0.0042
β_2	-0.3030	-0.0237	-0.0245	-0.0292	-0.2111	-0.2669	-0.0237
β_3	0.2781	0.0102	-0.0399	-0.0402	0.2485	0.2233	-0.0387
β_4	-0.3106	-0.2600	-0.0270	-0.0232	-0.2048	-0.2459	-0.0279
RMSE							
β_0	0.2026	0.3827	0.1663	0.1737	0.1937	0.1565	0.1676
β_1	0.6281	0.3037	0.1641	0.1782	0.5094	0.4938	0.1636
β_2	0.3172	0.3460	0.1627	0.1724	0.2398	0.2920	0.1626
β_3	0.2979	0.3366	0.1715	0.1790	0.2805	0.2532	0.1714
β_4	0.3294	0.3809	0.1568	0.1761	0.2331	0.2729	0.1555

Figure 2: Results of Simple Monte Carlo Simulation

Table 1: Probit Equation Modeling Probability of Positive Aid Flow

	Probit (Prob(Aid>0))			
Variable	Listwise	Mean	Prop. Nearest	Prop. Weighted
Voting Affinity	1.1185	0.8801	-0.8457	1.5201
	0.0785	0.038	0.0215	0.0436
Colony	-1.2246	-0.5139	0.2159	-1.1683
	0.0493	0.0184	0.0077	0.0286
Bureacratic Quality	-0.0944	0.4783	0.5229	0.3898
	0.0466	0.0262	0.0261	0.0266
Democracy	0.1023	-0.3576	-0.1968	-0.4868
	0.0305	0.0184	0.0133	0.0196
Economic Risk	0.2103	0.0085	0.3927	-0.0047
	0.031	0.0185	0.0093	0.0189
FDI	0.2387	0.0793	0.3422	0.1125
	0.0579	0.0336	0.0218	0.0337
Disaster	0.0885	-0.0289	0.0078	-0.0122
	0.0091	0.0031	0.0014	0.0025
GDP	-0.0002	0.0031	0.0024	0.0038
	0.0026	0.0005	0.0004	0.0005
Population	-0.3558	-0.4086	-0.232	-0.2975
	0.0163	0.008	0.0052	0.0071
Intercept	0.0441	0.0333	0.0352	0.0255
	0.005	0.0016	0.0014	0.0017
donors	38	38	38	38
recipients	81	190	190	190
N	17348	83076	83076	83076

Table 2: General Linear Model of Aid Flow (Conditional on positive amount)

		GLM E(Aid |Aid>0)		
Variable	Listwise	Mean	Prop. Nearest	Prop. Weighted
Voting Affinity	0.8529	0.2968	0.3804	0.4506
	0.15	0.0885	0.0414	0.1089
Colony	0.7629	0.3017	0.2953	0.3391
	0.1754	0.1046	0.1052	0.1071
Bureacratic Quality	0.2682	-0.222	0.0275	-0.2712
	0.1378	0.0888	0.0629	0.0913
Democracy	-0.399	-0.2474	0.0998	-0.2738
	0.1304	0.082	0.0421	0.0827
Economic Risk	-0.1982	-0.6441	-0.3581	-0.647
	0.2534	0.1542	0.1123	0.1547
FDI	-0.0294	-0.0404	-0.0331	-0.0352
	0.0389	0.0152	0.011	0.0146
Disaster	-0.0328	0.0053	0.0057	0.0054
	0.0116	0.0013	0.0013	0.0013
GDP	-0.6004	-0.3428	-0.2705	-0.299
	0.0874	0.0469	0.0329	0.045
Population	0.3062	0.1426	0.1114	0.1613
	0.0148	0.0051	0.0041	0.0057
Intercept	1.2532	2.1346	1.1219	2.0278
	0.3043	0.1825	0.1234	0.1913
donors	38	38	38	38
recipients	81	190	190	190
N	17348	83076	83076	83076