

## Modeling Efficiency of Foreign Aid Allocation in Malawi

Philip A. White, Candace Berrett, E. Shannon Neeley-Tass & Michael G. Findley

To cite this article: Philip A. White, Candace Berrett, E. Shannon Neeley-Tass & Michael G. Findley (2018): Modeling Efficiency of Foreign Aid Allocation in Malawi, *The American Statistician*, DOI: [10.1080/00031305.2018.1470032](https://doi.org/10.1080/00031305.2018.1470032)

To link to this article: <https://doi.org/10.1080/00031305.2018.1470032>



Accepted author version posted online: 17 May 2018.  
Published online: 17 Jul 2018.



Submit your article to this journal [↗](#)



Article views: 57



View Crossmark data [↗](#)



## Modeling Efficiency of Foreign Aid Allocation in Malawi

Philip A. White<sup>a</sup>, Candace Berrett<sup>b</sup>, E. Shannon Neeley-Tass<sup>b</sup>, and Michael G. Findley<sup>c</sup>

<sup>a</sup>Department of Statistics, Duke University, Durham, NC; <sup>b</sup>Department of Statistics, Brigham Young University, Provo, UT; <sup>c</sup>Department of Government, University of Texas at Austin, Austin, TX

### ABSTRACT

The Open Aid Malawi initiative has collected an unprecedented database that identifies as much location-specific information as possible for each of over 2500 individual foreign aid donations to Malawi since 2003. The efficient use and distribution of such aid is important to donors and to Malawi citizens. However, because of individual donor goals and difficulty in tracking donor coordination it is difficult to determine whether aid allocation is efficient. We compare several Bayesian spatial generalized linear mixed models to relate aid allocation to various economic indicators within seven donation sectors. We find that the spatial gamma regression model best predicts current aid allocation. While we are cautious about making strong claims based on this exploratory study, we provide a methodology by which one could (i) evaluate the efficiency of aid allocation via a study of the locations of current aid allocation as compared to the need at those locations and (ii) come up with a strategy for efficient allocation of resources in conditions where there exists an ideal relationship between aid allocation and economic sectors.

### ARTICLE HISTORY

Received April 2018  
Revised May 2016

### KEYWORDS

Bayesian; Donor collaboration; Gamma regression; SGLMM

## 1. Introduction

Each year, wealthy countries and international organizations send billions of dollars in foreign aid to developing countries across the world. The Organization for Economic Co-operation and Development (OECD) estimates that in 2015 the United States alone donated 31.1 billion dollars in development assistance and that other developed countries are donating anywhere between 700 million and 19 billion US dollars (OECD 2015). Despite the large dollar amounts and the potential impact of such aid, the allocation of these resources is not well studied. Aid giving is thus big business, and we need to understand better whether the business operates in the service of its so-called clients in recipient countries or rather in service of its own interests.

Donors' motives are varied, ranging from pursuing strategic interest (Alesina and Dollar 2000) to trying to facilitate positive social and economic outcomes. And yet, with so much spent on development assistance, donors, governments, and citizens may want these donations to be allocated efficiently—providing money to places and situations with the most need. Even where strategic interest concerns are high, and therefore aid may not be allocated with need in mind, foreign aid can affect peoples' lives in nontrivial ways, for better or for worse. Thus, it benefits stakeholders in both recipient and donor countries to know whether aid is in fact allocated in ways that respond to and meet the needs of recipient countries. For example, donations meant for improving education should ideally be sent to those places within the country where education is suffering the most.

Several initiatives have been introduced to improve aid targeting and coordination, including the Paris Declaration of 2005 which called for better alignment between government

objectives and donor aid allocation (Aced 2009; Findley, Marineau, and Powell 2016). However, the communication between recipients and donors, and between the various donor groups, that is necessary for tracking and efficiently allocating the vast array of projects, has proven nearly impossible. Moreover, there has been a dearth of scholarly attempts to track the complexity of aid allocation and so both academics and policy makers have a poor understanding of the patterns of foreign aid allocation. Because of the potentially substantial impact, answers to the question of efficient aid allocation are timely and important.

In 2011, the government of Malawi, together with the University of Texas at Austin's Climate Change and African Political Stability team, began efforts to collect and display donor information at the subnational level (Findley et al. 2011; Weaver et al. 2014). This initiative—Open Aid Malawi—resulted in an unprecedented and unique database containing as much location-specific information as possible for *each* individual donation to Malawi since 2003 (Peratsakis et al. 2012). This database contains over 2500 tracked donations all at the subnational level, to date the largest of its kind. Such a large and detailed resource makes it possible to examine donor aid across various project sectors.

Malawi is a particularly interesting country for examining aid allocation. Malawi's economy, education, and public health are some of the least developed anywhere in the world. Specifically, agricultural goods are their leading export, 50.7% of the population lives below the poverty line, the literacy rate is 68.2%, and the infant mortality rate is 69.3 deaths for every 1000 births. (National Statistical Office 2015). In contrast to most highly impoverished countries, Malawi has experienced almost

no violence, making it an ideal location to study aid allocation patterns, as the presence of violence makes stable aid allocation patterns difficult to determine. Put differently, given its political stability Malawi is a most likely case to observe donors responding to need, and we thus explore in this article the extent to which donors in fact allocate based on indicators of need.

In an effort to improve circumstances in Malawi, a large number of foreign donors support economic, health, and educational development plans in Malawi. We are specifically interested in quantifying how such assistance is being allocated within the country to determine whether sectoral aid matches sectoral need. In particular, we focus on seven project sectors or donation types: (1) Agriculture; (2) Education; (3) Governance; (4) Health; (5) Rural Development [RD]; (6) Roads, Public Works, and Transportation [RPT]; and (7) Water, Sanitation, and Irrigation [WSI].

Using district-specific economic, social, and demographic data (hereafter referred to as “economic indicators”), we model foreign aid allocation in all districts for these project sectors. While we do not expect the economic indicators to be the sole driving forces behind aid allocation, by the donors’ own admission, they should be key factors in determining whether or not aid is being allocated efficiently. Therefore, our goals are, first, to identify high-level economic indicators that are related to aid allocation while accounting for and making use of spatial dependence; second, to determine the ability of our model to capture a true relationship between aid distribution and the economic indicators; and third, to propose an efficient allocation scenario and its discrepancy from the actual allocation to inform future donations.

Analyses attempting to measure donor coordination in Malawi using this novel dataset have begun. For example, comparing donations before and after the Paris Declaration of 2005 with various economic and demographic indicators, Nunnenkamp, Sotirova, and Thiele (2015) found no evidence of increased donor coordination across sectors. De and Becker (2015) used instrumental variable techniques to show evidence of improved economic conditions at the local level from aid donations, emphasizing the need and ability for donors to coordinate their efforts using local economic indicators. Using a model to accurately capture the skewness present in aid data, we seek to significantly expand on these analyses by exploring the extent of the spatial dependence of the different sectors and to account for this spatial dependence to make inferences about relationships between donations and economic indicators at the local level.

Common analysis tools in the literature that examine aid allocation on this and other datasets include multiple regression (Feeny 2007; Nunnenkamp, Sotirova, and Thiele 2015), probit and Tobit regression (Collier and Dollar 2002; Dietrich 2013; De and Becker 2015), Poisson regression (Masaki 2017), structural equation models and two-stage least squares (Collier and Dollar 2002; Dietrich and Wright 2015), and mixed effects models (Hodler and Raschky 2014; Joablonski 2014; Briggs 2017). As geocoded aid data become more readily available, models that account for spatial dependence are becoming more common. For example, van Weezel (2015) used a Bayesian normally distributed mixed model with a spatial random effect to examine the relationship between foreign aid and civil

conflict. Nunnenkamp, Öhler, and Sosa Andrés (2017) studied the motivation (i.e., need, merit, or politics) for aid allocation in India using a spatial two-stage least-square model. Runfola et al. (2017) developed a methodology which they call geoSIMEX to account for different levels of spatial imprecision that arises when aid allocation data are reported at various spatial resolutions. Our model helps build on and expand the growing literature that accounts for spatial dependence in the modeling of foreign aid allocation data.

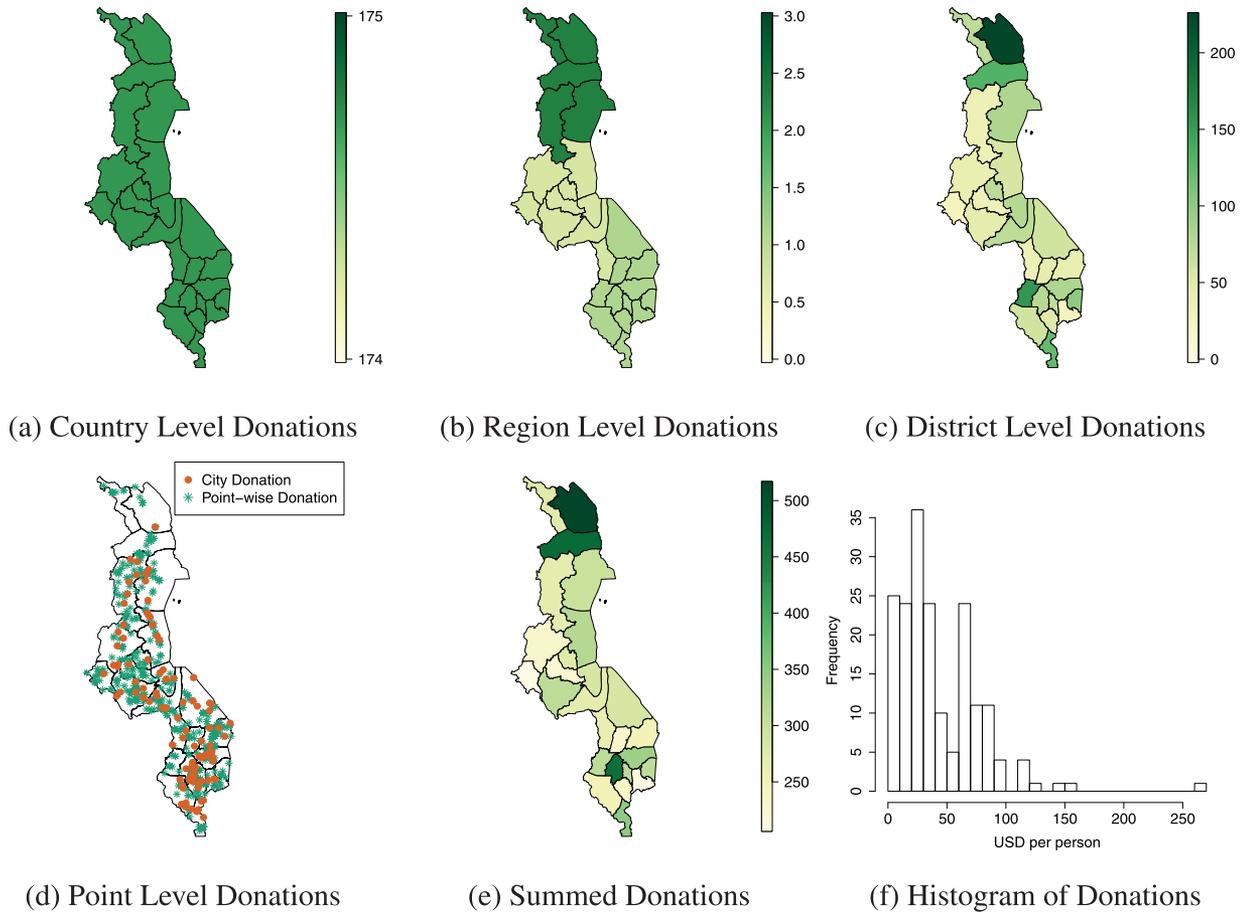
In this manuscript, we examine donation allocation within Malawi at the subnational level using a Bayesian spatial gamma regression model. We describe the data in more detail in Section 2. In Section 3, we explicitly define the model used for analysis as well as the comparison models. In Section 4.1, we provide an in-depth analysis of the various model parameters and their interpretations and implications for aid allocation in Malawi, contrasting these results with a more ideal allocation scenario in Section 4.2. Finally, we provide concluding remarks in Section 5.

## 2. Data

The Open Aid Malawi database (Peratsakis et al. 2012) contains over 2500 donations spanning the years 2003 to 2016. Each donation total (in US dollars; USD) corresponds to a specific project and its intended project sector (Agriculture; Education; Governance; Health; Rural Development; Roads, Public Works, and Transport; and Water, Sanitation, and Irrigation). The database also contains the spatial location and “precision” or level at which a project’s donation was designated (e.g., country, region, district, city). Figure 1(a)–1(d) illustrates the different precisions at which a project’s donation could be assigned. For example, Figure 1(b) shows the total money per person assigned to the three regions (North, Central, South) in Malawi, while Figure 1(d) shows the locations of project donations assigned to a city (orange circle points) or a more specific latitude–longitude (green stars).

In addition to the donation data, we also have population and economic indicator data for each district from the 2008 Malawi census (National Statistical Office 2008). These economic indicators include: proportion of people in poverty, average distance of people to schools (km), number of recently (within 2 weeks) injured people, food spending as a percentage of all expenses, mean area cultivated per household, and proportion of population with electric lighting. Descriptions of the variables are in Table 1; all covariates were standardized across districts to have a mean zero and variance one. However, we do not have these values for the three smallest districts—Likoma (a small island in Lake Malawi), Neno (split in 2003 from Mwanza), and Area under National Administration (various locations across the country). To accommodate this lack of information, we combined the donations to Neno and Mwanza (as they only recently split) and removed the other two districts from the analysis.

Because our economic indicator data are only available at the district level, we aggregated all the donation data to the district level. For the country- and region-level donations, we assumed these were dispersed proportionally among the districts according to district population. Additionally, each individual project



**Figure 1.** Money per person for donations at the various levels of available location information. Figures (a)–(c) show maps of money per person for (a) donations made to the entire country, (b) donations made to the individual northern, central, and southern regions, and (c) donations made to the individual districts. Figure (d) shows a map of the locations of city and point-wise donations. Figure (e) shows the total money per person within each district. Finally, (f) shows a histogram of the donations per person within each project sector and district.

could have several subprojects and each subproject could be assigned to a different location. However, only the total project donation amount is reported in the database and not the donation for the specific subproject. In this case, we divided the total project donation amount equally among the subprojects. Our final donation data are the total donation sum for each district in USD by project sector. Figure 1(e) shows the total donations per person summed across project sectors. We expect that more money should be sent to those districts with more people, thus we use dollars per person as the response variable by dividing the district donation total for each project sector by the district population.

For the dataset used in the analysis, the total sample size is  $N = 182$  since the data contain dollars donated per person in 26 districts for seven project sectors. Figure 1(f) shows a histogram of all 182 observed dollars per person for each sector and district, where the most extreme value is \$265.42 per person.

### 3. Model

#### 3.1. Model Definition

We expect some quantities, like agricultural output, public health outcomes, or nonurban areas, to be grouped spatially; therefore, it is plausible that accounting for spatial dependence could improve model performance in terms of prediction and

inferences. The normal linear regression model with spatial random effects is a standard model for spatially distributed quantities (Banerjee, Carlin, and Gelfand 2014). Because our data are strictly positive and right-skewed, we consider more general models. Following Diggle, Moyeed, and Tawn (1998), we make use of a spatial generalized linear mixed model (SGLMM).

Let  $Y_{ij}$  be the dollars per person (USD) for the  $i$ th district ( $i = 1, \dots, n$ , where  $n = 26$ ) and the  $j$ th project sector ( $j = 1, \dots, J$ , where  $J = 7$ ). We model

$$Y_{ij} | \mu_{ij}, \theta \sim f(y_{ij} | \mu_{ij}, \theta),$$

where  $\mu_{ij} = E(Y_{ij} | \beta_j, \phi_{ij}, \theta)$  and  $\theta$  is an applicable scale parameter (e.g., for the normal distribution, this is the common variance parameter). We model the mean using a link function,  $g(\cdot)$ ,

$$g(\mu_{ij}) = \mathbf{x}'_{ij} \beta_j + \phi_{ij}, \quad (1)$$

where  $\mathbf{x}_{ij}$  is a  $k \times 1$  vector of covariates (economic indicators) for the  $i$ th district and the  $j$ th project sector,  $\beta_j$  is a  $k \times 1$  vector of coefficients for each sector, and  $\phi_{ij}$  is a spatial random effect where each sector has its own spatial dependence structure. Because we expect a combination of the economic indicators to be related to aid allocation within each sector, we fit a sector-specific coefficient for each of the seven economic indicators (including population) as well as a sector-specific intercept to account for the different amounts of aid appointed to each

**Table 1.** Names and descriptions of the data variables used in the analysis.

Variable	Abbreviation	Description
<i>Sectors</i>		
Agriculture	Agriculture	Donations assigned to development of agriculture
Education	Education	Donations assigned to education development
Governance	Governance	Donations assigned to both economic and democratic governance
Health	Health	Donations assigned to for development of health initiatives
Integrated Rural Development	RD	Donations assigned for rural development
Roads, Public Works, and Transportation	RPT	Donations assigned to developing roads, transportation, and other public works
Water, Sanitation, and Irrigation	WSI	Donations assigned to water, sanitation, and irrigation development
<i>Economic Indicators</i>		
Proportion in Poverty	Poverty	Percent of population living below the national poverty line
Distance to School	Dist to School	Average distance (km) across households to the nearest school in each district
All Injured	All Injured	Number injured or ill in the previous 2 weeks (as of the 2008 census) in each district
Food as % of Expenditures	Food %	Average percent of household expenditures spent on food in each district
Mean Land Cultivated per House	Land Cultivated	Average area cultivated per household (in hectares) in each district
Electric Lighting Proportion	Electric %	Percent of households in the district with electric lighting
Population	Population	Population of each district

sector. Therefore, there are  $k = 8$  coefficients for each sector, or 56 total coefficients in our model.

Following Hughes and Haran (2013), we account for potential spatial confounding by using the Moran basis on the spatial random effect. Let  $\phi_j = (\phi_{1j}, \dots, \phi_{26,j})'$  be the vector of spatial random effects for sector  $j$  and model this with an intrinsic conditionally autoregressive (CAR; see, e.g., Besag 1974; Banerjee, Carlin, and Gelfand 2014) prior:

$$\pi(\phi_j | \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\text{rank}(\mathbf{Q})/2} \exp\left\{-\frac{1}{2\sigma^2} \phi_j' \mathbf{Q} \phi_j\right\},$$

where  $\sigma^2$  is a variance parameter and  $\mathbf{Q} = \text{diag}(\mathbf{A}\mathbf{1}) - \mathbf{A}$  is a precision matrix where  $\mathbf{A}$  is the  $n \times n$  binary neighborhood matrix such that  $A_{ik} = 1$  for  $i \neq k$  if districts  $i$  and  $k$  share a border and  $A_{ik} = 0$  otherwise, and  $\mathbf{1}$  is an  $n \times 1$  vector of ones. As detailed by Reich, Hodges, and Zadnik (2006), the spatial random effect may be correlated with spatially dependent covariates, creating collinearity between the model coefficients,  $\beta$ , and the spatial random effect,  $\phi$ , inhibiting model inferences and interpretability. Thus, Hughes and Haran (2013) proposed to remove confounding by using the Moran basis to approximate

the spatial random effect:

$$\phi_j \approx \mathbf{M} \delta_j,$$

where  $\mathbf{M}$  is fixed to be the first  $r$  eigenvectors of  $\mathbf{P}^\perp \mathbf{A} \mathbf{P}^\perp$ , where  $\mathbf{P}^\perp = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$  and  $\mathbf{X}$  is the  $n \times k$  matrix consisting of a column of ones and the economic indicators for each location, and  $\delta_j$  is the  $r \times 1$  vector of coefficients. Then,

$$\pi(\delta_j | \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{r/2} \exp\left\{-\frac{1}{2\sigma^2} \delta_j' (\mathbf{M}' \mathbf{Q} \mathbf{M}) \delta_j\right\},$$

making  $\mathbf{M}' \mathbf{Q} \mathbf{M}$  the precision matrix for  $\delta_j$ . Thus, rather than modeling the spatial random effect directly through  $\phi_j$ , we model  $\delta_j$ . One recommendation by Hughes and Haran (2013) is to set  $r$  to be the number of positive eigenvalues of  $\mathbf{P}^\perp \mathbf{A} \mathbf{P}^\perp$  to model positive spatial dependence and we follow their approach in our analysis. Thus, we set  $r = 7$  making the total number of parameters in each spatial model equal to  $(k \times J) + (r \times J) + 2 = 107$ .

Finally, for the remaining parameters we use relatively non-informative prior distributions. Specifically,

$$\theta \sim \text{Inverse-Gamma}(a_\theta, b_\theta)$$

$$\beta \sim \text{Normal}(\mathbf{0}, s^2 \mathbf{I})$$

$$\sigma^2 \sim \text{Inverse-Gamma}(a_\sigma, b_\sigma)$$

where  $a_\theta = 0.01$ ,  $b_\theta = 0.01$ ,  $s^2 = 10,000$ ,  $a_\sigma = 0.01$ ,  $b_\sigma = 0.01$ . We select these values for the hyperparameters so that our prior distributions are diffuse.

### 3.2. Model Selection

We consider eight Bayesian models, comparing both spatial and nonspatial models using gamma, log-normal, normal, and Weibull likelihood distributions (see Table 2; Appendix A shows the parameterizations used for each likelihood distribution). For the models using the Weibull likelihood distribution, instead of modeling the mean,  $\mu_{ij}$ , as a linear combination of the covariates, we model the parameter,  $\lambda_{ij} = \mu_{ij} / \Gamma(1 + 1/\theta)$ , as done in, for example, Kalbfleisch and Prentice (2002). For the nonspatial models, we assume  $\phi_{ij} \equiv 0$  for all  $i$  and  $j$  in Equation (1).

Table 2 also defines the link function used for each model. We chose these link functions because they are commonly used for each likelihood and they are often used due to computational convenience (McCullagh and Nelder 1989). For example, the canonical link for the Gamma distribution is the inverse link, but this link is not always positive and has vertical asymptotes.

**Table 2.** Models used for comparison. Note that for the Weibull likelihood, the link is on the scale parameter,  $\lambda$ , instead of the mean,  $\mu$ .

Model	Likelihood: $f(y_{ij}   \mu_{ij}, \theta)$	Link: $g(\mu_{ij})$	Spatial dependence
1	Gamma	$\log(\mu_{ij})$	No
2	Gamma	$\log(\mu_{ij})$	Yes
3	Log-normal	$\mu_{ij}$	No
4	Log-normal	$\mu_{ij}$	Yes
5	Normal	$\mu_{ij}$	No
6	Normal	$\mu_{ij}$	Yes
7	Weibull	$\log(\lambda_{ij})$	No
8	Weibull	$\log(\lambda_{ij})$	Yes

**Table 3.** Observed mean absolute error (MAE), mean squared error (MSE), and deviance information criteria (DIC) for the eight posed models. The smallest number for each criterion is bolded.

Model		MAE	MSE	DIC
1	Nonspatial Gamma	10.36	623.88	3574.69
2	Spatial Gamma	<b>8.93</b>	495.23	<b>3568.18</b>
3	Nonspatial Log-normal	10.09	513.63	3740.24
4	Spatial Log-normal	10.17	494.06	3910.73
5	Nonspatial Normal	11.06	436.83	4028.86
6	Spatial Normal	10.83	<b>420.02</b>	4194.91
7	Nonspatial Weibull	13.70	544.23	3814.99
8	Spatial Weibull	12.19	529.77	3860.22

We fit three chains (using different starting values) for each model using an adaptive Metropolis random walk sampler (Roberts and Rosenthal 2009) for 1,500,000 iterations using code written in R (R Core Team 2016). For memory, we saved only every tenth draw. Of the saved draws, we kept the final 30,000 as post-burn-in posterior draws. We checked for convergence by monitoring trace plots and computing Geweke's convergence diagnostic within chains (Geweke 1992) and Gelman–Rubin convergence diagnostic between chains (Gelman and Rubin 1992) using the R package `coda` (Plummer et al. 2006). For all models and all chains, less than 1% of Geweke diagnostics on parameters from all models and all chains exceeded 3 (or were below  $-3$ ) and none were above (below) 3.8 ( $-3.8$ ). No Gelman–Rubin multivariate diagnostic exceeded 1.1, and the maximum univariate diagnostic value was 1.3, but almost all were below 1.1. We also computed Monte Carlo standard errors using the R package `mcmcse` (Flegal et al. 2017) and these values are included with all parameter estimates in Appendix C.

We are interested in the model that performs the best in predicting foreign aid for each sector in each district; therefore, we use the posterior predictive distribution,

$$f(y_{\text{new}} | y) = \int_{\Psi} f(y_{\text{new}} | \psi) \pi(\psi | y) d\psi, \quad (2)$$

where  $\psi$  is the set of all model parameters and  $\Psi$  is its parameter space, for each model to determine which model most accurately predicts foreign aid in Malawi. We consider both mean

squared error (MSE) and mean absolute error (MAE) between the posterior predictive mean and observed donations. We also compute deviance information criteria (DIC; Spiegelhalter et al. 2002, 2014) for model comparison.

Table 3 shows the MAE and MSE for the in-sample predictions from each of the models. In most cases, the spatial models out predict their nonspatial counterparts; however, the improvement is modest. Additionally, other than the spatial gamma model, the DIC is larger for the spatial models than for their nonspatial counterparts. This indicates that while some spatial dependence is present, the additional parameters may not be worth adding for the small gains in model fit. In this case, however, we believe a spatial model is important especially as geocoded aid data and economic data become more sophisticated and available. That said, Model 2, or the spatial gamma model, provides the smallest MAE and the smallest DIC. Model 6, or the spatial normal model, has the smallest MSE. This is due to an anomalously high value in the RPT sector (see Figure B.2 in Appendix B) that no model captures well, but the normal model captures best and therefore its MSE is less affected by the outlier. This improvement in model fit is not seen in the MAE of the normal model because the outlier does not affect the MAE as much as the MSE. In our analysis, we focus on the spatial gamma model because this is the model with the smallest DIC and MAE.

## 4. Analysis

In this section, we examine the parameters and fitted values of the spatial gamma model described in Section 3. In Section 4.1, we discuss evidence for and against donor collaboration or efficient aid allocation by examining the posterior distributions of the model coefficients. We conduct an in-depth study in Section 4.2 to identify potential areas for improved aid allocation.

### 4.1. Parameter Estimates and Interpretation

Table 4 shows posterior means and standard deviations for all regression coefficients  $\beta$  as well as the marginal posterior probabilities that each regression coefficient is greater than 0. We

**Table 4.** Summary of the posterior distributions of the coefficients of the spatial gamma model. ' $P(> 0)$ ' means the posterior probability that the corresponding coefficient is greater than 0. Coefficients (other than intercepts) that have  $>90\%$  posterior probability of being either positive or negative are bolded. RD = Integrated Rural Development; RPT = Roads, Public Works, & Transportation; WSI = Water, Sanitation, Irrigation.

Sector covariate	Agriculture mean (sd) $P(> 0)$	Education mean (sd) $P(> 0)$	Governance mean (sd) $P(> 0)$	Health mean (sd) $P(> 0)$	RD mean (sd) $P(> 0)$	RPT mean (sd) $P(> 0)$	WSI mean (sd) $P(> 0)$
Intercept	3.55 (0.08) 1.00	3.5 (0.08) 1.00	4.21 (0.03) 1.00	4.46 (0.03) 1.00	2.69 (0.12) 1.00	2.54 (0.17) 1.00	2.76 (0.12) 1.00
Poverty	−0.08 (0.11) 0.22	0.09 (0.12) 0.80	−0.02 (0.06) 0.37	<b>0.07</b> (0.04) <b>0.97</b>	−0.29 (0.18) <b>0.04</b>	<b>0.34</b> (0.22) <b>0.94</b>	−0.07 (0.21) 0.37
Dist to School	0.00 (0.13) 0.49	−0.16 (0.14) 0.13	0.03 (0.07) 0.66	−0.02 (0.05) 0.31	<b>0.42</b> (0.2) <b>0.98</b>	−0.11 (0.24) 0.31	0.21 (0.23) 0.81
All Injured	−0.17 (0.11) <b>0.06</b>	<b>0.17</b> (0.09) <b>0.97</b>	0.00 (0.05) 0.52	0.01 (0.04) 0.65	−0.14 (0.15) 0.17	<b>0.59</b> (0.16) <b>1.00</b>	0.06 (0.15) 0.66
Food %	0.1 (0.13) 0.81	0.04 (0.17) 0.5	−0.01 (0.05) 0.42	0.00 (0.04) 0.43	0.06 (0.15) 0.64	−0.68 (0.17) <b>0</b>	−0.15 (0.17) 0.16
Land Cultivated	<b>0.13</b> (0.08) <b>0.95</b>	−0.02 (0.08) 0.43	−0.02 (0.04) 0.31	−0.11 (0.03) <b>0.00</b>	−0.08 (0.12) 0.23	−0.52 (0.12) <b>0.00</b>	−0.03 (0.12) 0.38
Electric %	<b>0.20</b> (0.07) <b>0.99</b>	−0.03 (0.08) 0.37	0.03 (0.04) 0.80	<b>0.04</b> (0.03) <b>0.93</b>	0.08 (0.09) 0.81	<b>0.94</b> (0.12) <b>1.00</b>	−0.15 (0.17) 0.19
Population	−0.05 (0.10) 0.34	−0.06 (0.09) 0.28	0.04 (0.04) 0.85	−0.02 (0.03) 0.25	−0.11 (0.10) 0.13	<b>0.39</b> (0.11) <b>1.00</b>	0.09 (0.13) 0.78

use these posterior probabilities to examine the relationship between the economic indicator and the donations within a sector. Bolded values correspond to coefficients with posterior probability greater than 0.9 of being either positive or negative. We emphasize that this analysis is exploratory and that relationships (or lack of relationships) should be interpreted with caution.

We use the relationships between aid and the economic indicators to determine whether or not there is evidence of efficient aid allocation. For example, for the Education sector, if aid is being allocated efficiently, we would expect education aid to be positively related to distance to school (greater distances to school on average indicate a greater need for education aid). In large part, there is no evidence that the economic indicators are related to districts' aid per person within a sector. Many of the coefficient posterior distributions envelop 0 (those nonbolded coefficients). In fact, two sectors—Governance and Water, Sanitation, and Irrigation (WSI)—appear to not be related to any of the economic indicators. For Governance, this is not surprising as more than 50% of the donations from that sector were given at the country or region level (see Figure 1(a) and 1(b)), and we assigned these projects equally per person to each of the districts within the country or region, respectively. For WSI, however, this is not the case and suggests that WSI aid allocation might not be related to the economic indicators.

There are, however, several coefficients that provide evidence of efficient aid allocation. For example, mean land cultivated per household has a 0.95 posterior probability of being positively related to agriculture aid. This makes sense since more land cultivated indicates more agricultural areas and thus more need for agricultural aid. Additionally, distance to school is positively related to RD aid (0.98 posterior probability), as expected, since districts where the people are farther from a school tend to be more rural and thus would benefit from RD aid.

In contrast, poverty has a posterior probability of 0.96 of being negatively related to RD aid. This means that as the proportion of the population living below the poverty line increases, the amount of RD aid decreases, contradicting what we would expect for efficient aid allocation. Also, number injured is positively related to education aid (posterior probability of 0.97), but there is not an intuitive reason for why this would make sense. A natural economic indicator for education would be distance to school, but this relationship is weaker (and negative—the opposite of what we would expect).

Several economic indicator variables are related to the RPT sector. This is likely due to a large outlier in this sector, visible in Figure 1(f).

The posterior mean for  $\sigma^2$ , the variance of the spatial dependence parameters, is 0.054 and a corresponding 90% credible interval is (0.005, 0.174). Additionally, we provide coefficient summaries for the regression coefficients of all models in Appendix C. As expected, the coefficients follow similar patterns across models and most of the time, models find the same variables to have high posterior probabilities of being different from zero.

Obviously, these interpretations of coefficients are conditional on the values of the other covariates and multicollinearity among the covariates could introduce changes in the values of the coefficients when some covariates are left out of the

**Table 5.** Variance inflation factors (VIF) and scaled generalized VIF's (GVIF) for economic indicator variables.

Covariate	VIF	GVIF <sup>1/14</sup>
Poverty	2.49	1.58
Dist to School	3.54	1.88
All Injured	2.36	1.54
Food %	1.94	1.39
Land Cultivated	1.40	1.18
Electric %	1.21	1.10
Population	1.11	1.05

model (thus changing interpretations). However, there is not much collinearity between the economic indicators, the highest absolute value correlation is 0.3, with most absolute value correlations less than 0.15. Table 5 provides variance inflation factors (VIF) for each economic indicator and scaled generalized variance inflation factors (GVIF; Fox and Monette 1992) which account for the interaction between sector and each economic indicator. All VIF's are below the commonly used rule of thumb of five to indicate concerning multicollinearity (see, e.g., Dobson and Barnett 2008). That said, interpretations of the coefficients must be made with care.

Figure 2 shows plots of observed, posterior predictive means, and the residuals for the total donations, as well as the same plots for Governance and RD donations. Plots for the remaining sectors are in Appendix B in Figures B.1 and B.2.

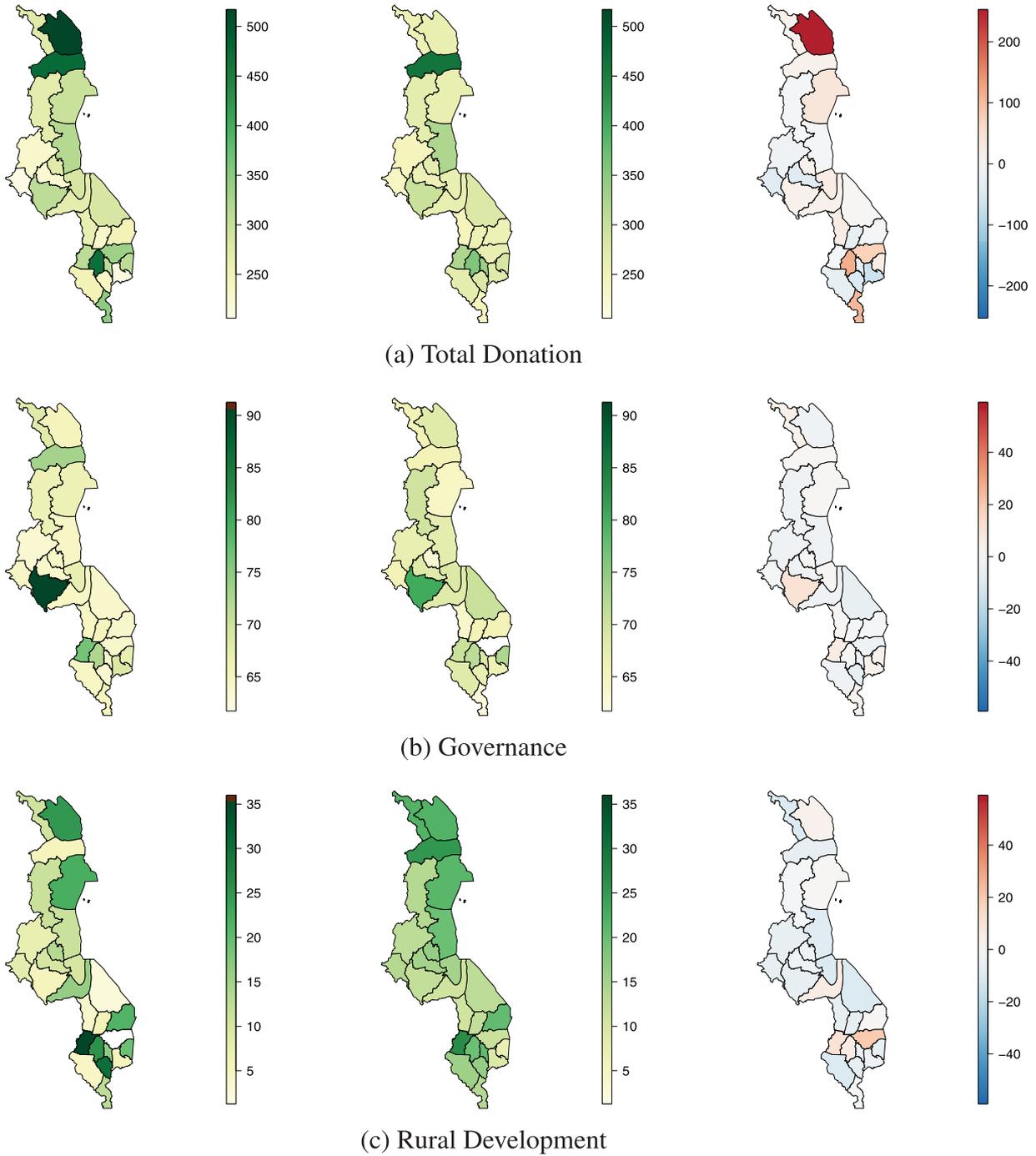
#### 4.2. More Efficient Aid Allocation

Although we do not expect the economic indicators to be perfect drivers of aid allocation, they can provide direction for better aid allocation. In this section, we consider the scenario where aid is allocated according to a more expected or ideal relationship with specific economic indicators. We call this scenario our “efficient scenario.” By doing this, we can first examine the potential for our model to capture a “true” relationship between aid and economic indicators. Second, and more important for donors as they identify locations to send aid, is to identify locations where more or less aid is needed within each sector.

To accomplish this task, we select a single economic indicator for each sector and identify its expected relationship to aid, shown in Table 6. We then use the posterior distributions from the fitted model of Section 4.1, but adjust the coefficients for these selected economic indicators. Specifically, we set the mean of the corresponding coefficient of each economic indicator of interest to be three posterior standard deviations above (or below for a negative relationship) zero. For example, the

**Table 6.** Selected economic indicator for each sector and the corresponding direction of the relationship.

Sector	Economic indicator	Expected relationship
Agriculture	Land Cultivated	Positive
Education	Dist to School	Positive
Governance	Poverty	Positive
Health	All Injured	Positive
RD	Electric %	Negative
RPT	Electric %	Negative
WSI	Land Cultivated	Positive



**Figure 2.** Observed (left), predicted (center), and residual (right) donations plots for total donations (a) and the Governance (b) and RD (c) sectors. All values are in USD.

adjusted mean of the coefficient corresponding to Land Cultivated for Agriculture is set to 0.24 (see Table 4), while the other coefficients for the Agriculture sector remain at the original posterior mean. Then, the adjusted posterior distribution on the coefficients is

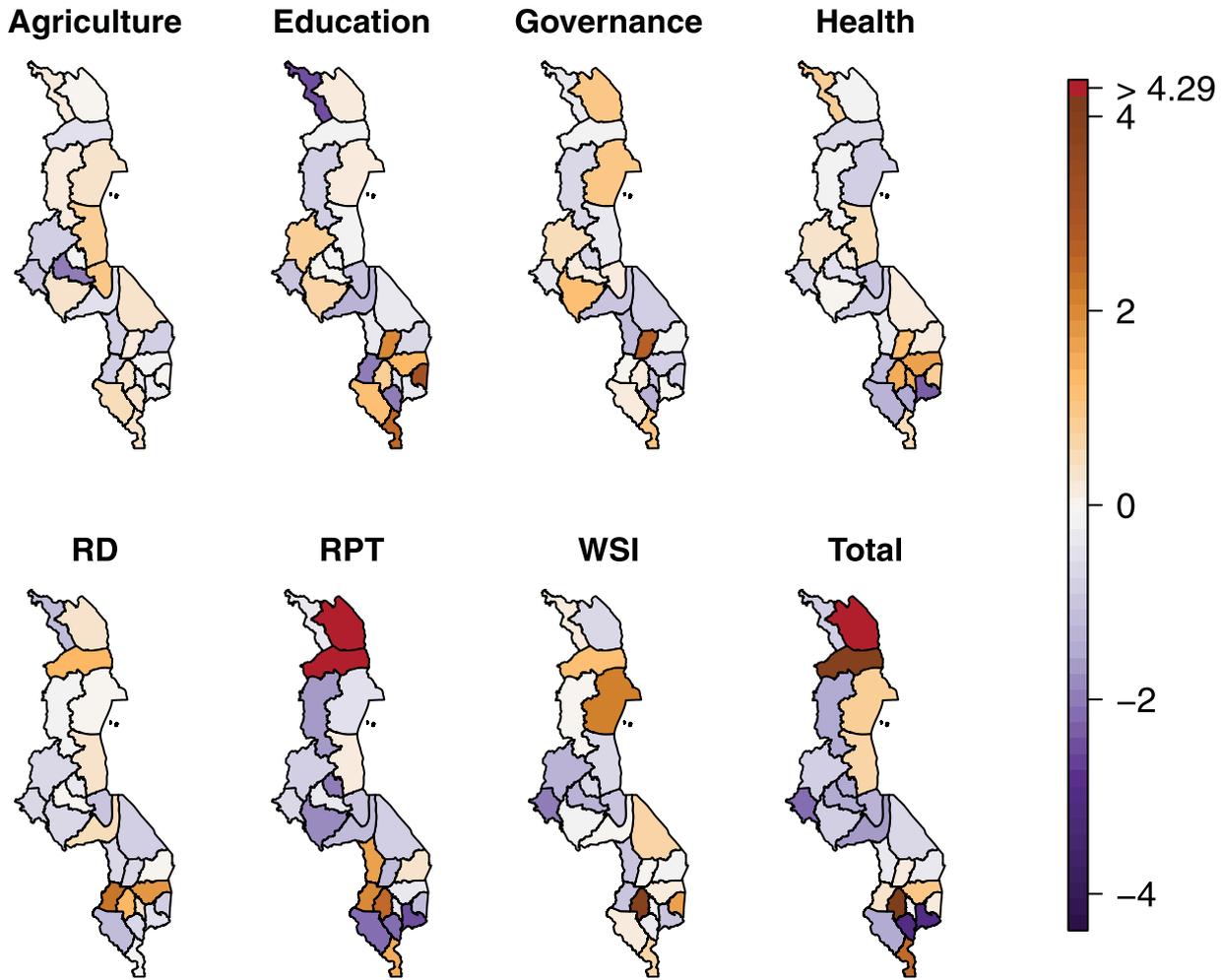
$$\beta^* \sim N(\bar{\beta}^*, \Sigma(\beta)), \quad (3)$$

where  $\bar{\beta}^*$  is the adjusted posterior mean of  $\beta$  as described above and  $\Sigma(\beta)$  is the posterior variance of  $\beta$  from the fitted spatial gamma model in Section 4.1. The posterior correlation among the  $\beta$ 's was quite small, and thus  $\Sigma(\beta)$  could also be a diagonal

matrix with the marginal variances of each coefficient on the diagonal.

We then obtain adjusted posterior predictive distributions on aid per person for each sector using Equation (2), but with the adjusted posterior distribution of  $\beta$ . The computational algorithm is as follows:

1. Draw a value of  $\beta^*$  from Equation (3).
2. Use a draw of  $\phi$  and  $\theta$  from the posterior draws obtained in Section 4.1.
3. Draw a value of the adjusted aid allocation,  $y_{ij}^*$ , for each  $i$  and  $j$  from the gamma likelihood, conditional on the values in steps 1 and 2.



**Figure 3.** Standardized residual maps between the observed and efficient scenario donations per person for each sector and for all sectors combined (bottom right).

4. Scale  $y_{ij}^*$  so that the new sector totals are the same as the data sector totals:  $\sum_{i=1}^{26} y_{ij}^* = \sum_{i=1}^{26} y_{ij}$ . (This requires that the total money within a sector does not change, only its distribution across the districts.)
5. Repeat for all post-burn-in posterior draws.

We can then use these draws from the adjusted posterior predictive distributions to make inferences on aid allocation.

To learn about the ability of our model to identify a “true” relationship, we refit the spatial gamma model to 26 randomly selected draws from the adjusted posterior predictive distribution. In this way, we are making use of frequentist sampling distribution ideas to compute a power-type diagnostic. For these 26 randomly selected draws, using a 90% posterior probability of being greater than zero (or less than zero for expected negative relationships) cut-off, we would have identified a relationship 85% of the time. This is a bit lower than 90%, but this may be due to rescaling the data to keep total donation amounts the same. Notably, however, we never found a relationship in the opposite direction, that is, a parameter that was truly positive (negative) never had a 90% posterior probability of being negative (positive).

Finally, we compare the adjusted posterior predictive distributions of donation amounts to the observed donation amounts.

**Figure 3** shows standardized residual maps; specifically, for district  $i$  and sector  $j$ ,

$$r_{ij} = \frac{y_{ij} - \bar{y}_{ij}^*}{SD(y_{ij}^*)},$$

where  $y_{ij}$  is the observed aid per person,  $\bar{y}_{ij}^*$  is the mean of the adjusted posterior predictive distribution of aid per person, and  $SD(y_{ij}^*)$  is the standard deviation of the adjusted posterior predictive distribution of aid per person. By standardizing this way, we can compare differences between the predictive means and the observed values relative to the variation of the predictive distributions, so that a large difference between the observed and predictive means is either dampened or enhanced by a large or small standard deviation, respectively. In **Figure 3**, darker orange values indicate that more aid is going to these locations than our efficient scenario suggests, while dark purple values indicate that less aid is going to these locations than our efficient scenario suggests. Looking at total donations per person within each district, many of the districts that indicate more need for aid are on the western side of Malawi. Note that as with interpretations in **Section 4.1**, these results are merely exploratory and we must be cautious in taking interpretations of these results too far.

## 5. Discussion

Foreign aid is important for many countries' economies. How the aid is allocated can be even more important to both the donors and citizens within the receiving country. Unfortunately, determining if aid is in fact efficiently allocated can be difficult to measure and assess. In this article, we made use of a unique dataset to examine how aid allocation within Malawi is related to various economic indicators within seven donation sectors. While this study is mainly exploratory and we are cautious about making strong claims based on the results, we found some evidence in support of efficient allocation as well as some evidence that aid is actually not being allocated according to economic indicators. Additionally, we also used the model to suggest a plan for efficient aid allocation according to an idealized relationship with a specific subset of the economic indicators. We provided maps that can give donors insights into adjustments of aid allocation to better facilitate more efficient allocation.

Although most countries do not have databases with similar location-specific information on donations, these methods can be applied to examine efficiency of cross-national aid allocation. And as more subnational geocoded aid data become available, these methods will become more relevant. On the whole, our findings suggest the importance of collecting good data and studying aid allocation with more rigor to better understand how foreign aid money is allocated.

### Appendix A: Likelihood Parameterizations

In this appendix, we include the parameterization used for each likelihood distribution in Models 1–8. For brevity, we drop subscript notation so that below,  $y_{ij} = y$ ,  $\mu_{ij} = \mu$ ,  $\mathbf{x}_{ij} = \mathbf{x}$ , and  $\boldsymbol{\beta}_j = \boldsymbol{\beta}$ .

*Gamma*

$$f(y|\mu, \theta) = \left(\frac{\mu}{\theta}\right)^{\mu^2/\theta} \frac{1}{\Gamma(\mu^2/\theta)} y^{(\mu^2/\theta)-1} \exp\{-y\mu/\theta\},$$

where  $E(y|\mu, \theta) = \mu$  and  $\text{var}(y|\mu, \theta) = \theta$ .

*Lognormal*

$$f(y|\mu, \theta) = \frac{1}{\sqrt{2\pi\theta}y} \exp\left\{-\frac{1}{2\theta}(\log y - \mu)^2\right\},$$

where  $E(\log y|\mu, \theta) = \mu$  and  $\text{var}(\log y|\mu, \theta) = \theta$ .

*Normal*

$$f(y|\mu, \theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left\{-\frac{1}{2\theta}(y - \mu)^2\right\},$$

where  $E(y|\mu, \theta) = \mu$  and  $\text{var}(y|\mu, \theta) = \theta$ .

*Weibull*

$$f(y|\lambda, \theta) = \frac{\theta}{\lambda} \left(\frac{y}{\lambda}\right)^{\theta-1} \exp\left\{-\left(\frac{y}{\lambda}\right)^\theta\right\},$$

where  $E(y|\lambda, \theta) = \lambda\Gamma(1 + 1/\theta)$  and  $\text{var}(y|\lambda, \theta) = \lambda^2(\Gamma(1 + 2/\theta) - \Gamma(1 + 1/\theta)^2)$ . Note that this parameterization corresponds to the Weibull regression model found in Kalbfleisch and Prentice (2002) as follows. As in our models, let  $\log \lambda = \mathbf{x}'\boldsymbol{\beta} = \beta_0 + \mathbf{x}'_1\boldsymbol{\beta}_1$ , where  $\beta_0$  corresponds to the intercept,  $\mathbf{x}_1$  corresponds to the vector of covariates without the intercept, and  $\boldsymbol{\beta}_1$  corresponds to the remaining vector of coefficients (all the coefficients without the intercept). Then, the pdf in (2.8) of Kalbfleisch and Prentice (2002) can be written by letting  $\theta = \gamma$ ,  $\exp\{-\beta_0\} = \lambda$ , and  $-\boldsymbol{\beta}_1\theta = \boldsymbol{\beta}$  where the parameters on the right correspond to the notation used in Kalbfleisch and Prentice (2002) and the parameters on the left correspond to the parameters used in our model.

### Appendix B: Predicted Values

Figures B.1 and B.2 provide additional plots of the observed, predicted, and residual values of donations per person for each sector.

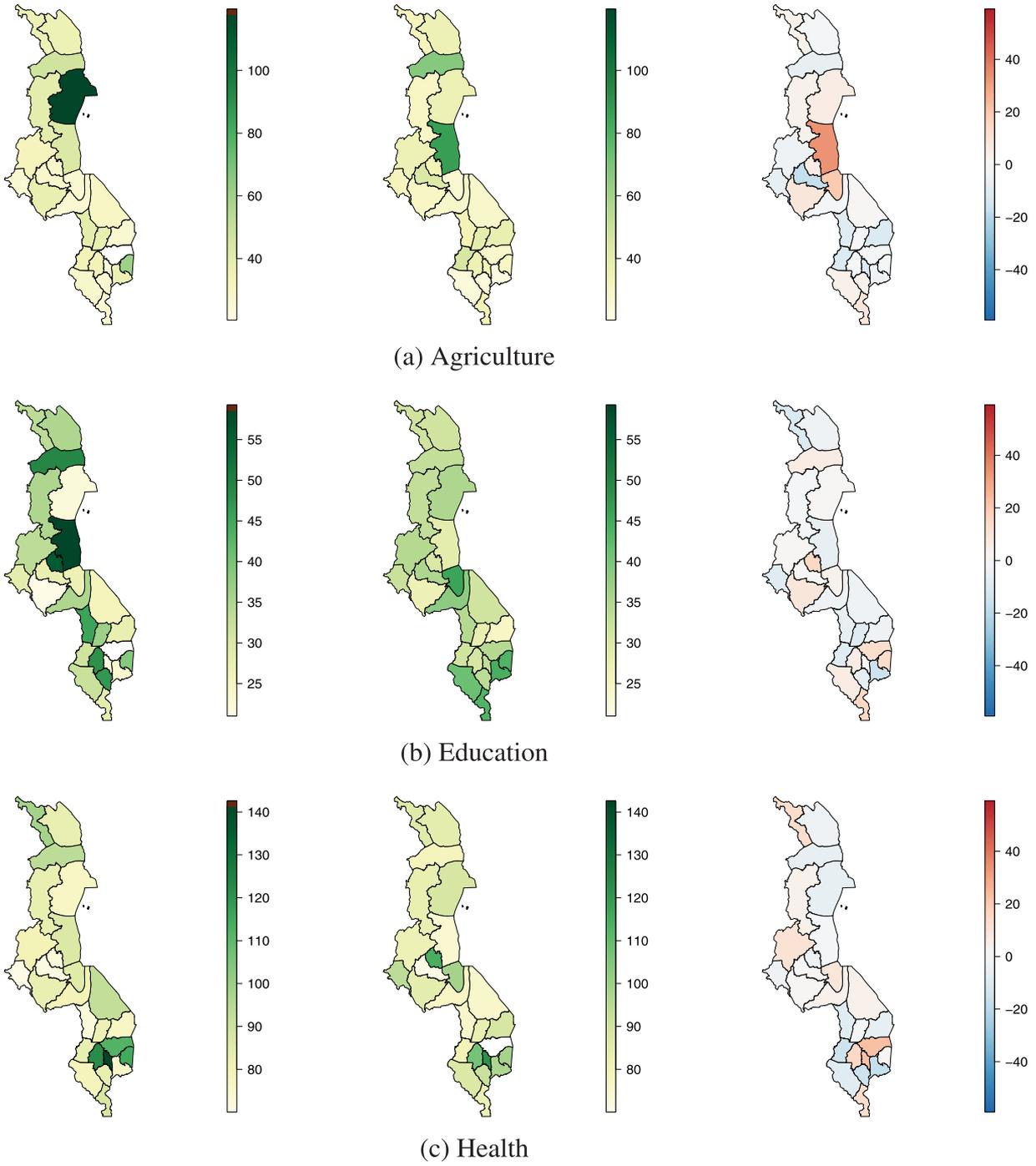
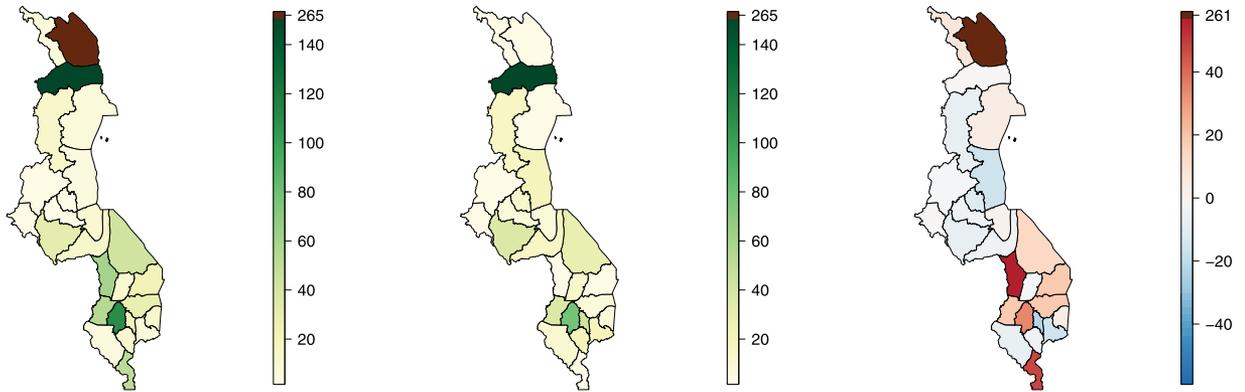
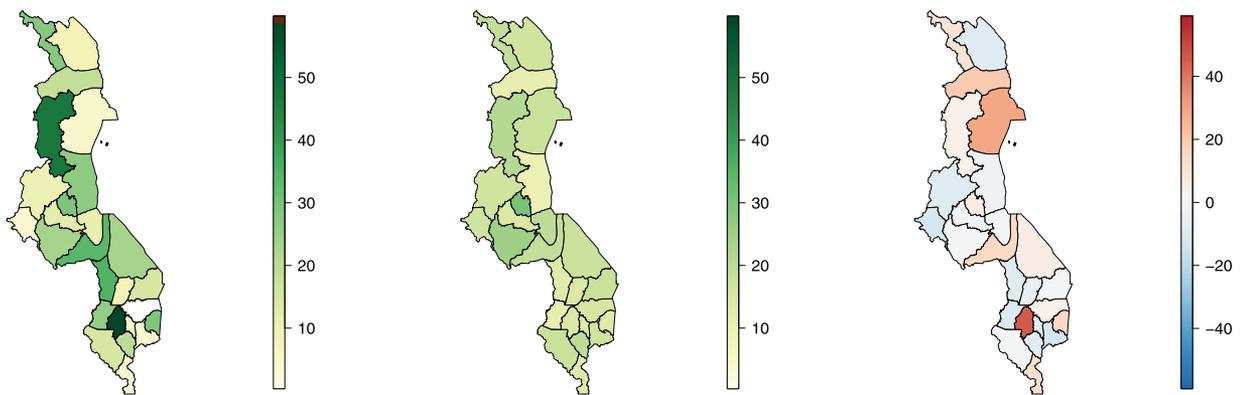


Figure B.1. Observed (left), predicted (center), and residual (right) donations plots for Agriculture, Education, and Health sectors. All values are in USD.



(a) Roads, Public Works and Transportation



(b) Water, Sanitation, and Irrigation

**Figure B.2.** Observed (left), predicted (center), and residual (right) donations plots for RPT and WSI sectors. Outliers (in observed and residual donations of the RPT sector) are colored brown with values marked on the color legends. All values are in USD.

## Appendix C: Coefficient Estimates

Tables C.1–C.7 provide the posterior mean of each coefficient along with the corresponding Monte Carlo standard errors and posterior probabilities that the coefficient is greater than 0 for all models and sectors.

**Table C.1.** Coefficient posterior means (Monte Carlo standard errors) for all models for the *Agriculture* sector.  $P(> 0)$  means the posterior probability that the corresponding coefficient is greater than 0. Coefficients (other than intercepts) that have  $> 90\%$  posterior probability of being either positive or negative are bolded.

Covariate	Intercept mean (mcse) $P(> 0)$	Poverty mean (mcse) $P(> 0)$	Dist to School mean (mcse) $P(> 0)$	All Injured mean (mcse) $P(> 0)$	Food % mean (mcse) $P(> 0)$	Land Cultivated mean (mcse) $P(> 0)$	Electric % mean (mcse) $P(> 0)$	Population mean (mcse) $P(> 0)$
1	3.58 (0.001) 1	<b>-0.1</b> (0.001) <b>0.07</b>	0.05 (0.002) 0.7	<b>-0.14</b> (0.001) <b>0.02</b>	0.06 (0.001) 0.82	<b>0.11</b> (0.001) <b>0.98</b>	<b>0.19</b> (0.001) <b>1</b>	-0.05 (0.001) 0.2
2	3.55 (0.002) 1	-0.08 (0.003) 0.22	0 (0.003) 0.49	<b>-0.17</b> (0.003) <b>0.06</b>	0.1 (0.003) 0.81	<b>0.13</b> (0.002) <b>0.95</b>	<b>0.2</b> (0.002) <b>0.99</b>	-0.05 (0.002) 0.34
3	3.57 (0.001) 1	-0.12 (0.003) 0.22	0.05 (0.004) 0.61	-0.1 (0.003) 0.26	0.01 (0.002) 0.54	0.1 (0.002) 0.82	<b>0.18</b> (0.002) <b>0.95</b>	-0.05 (0.001) 0.34
4	3.57 (0.001) 1	-0.12 (0.004) 0.26	0.05 (0.005) 0.58	-0.09 (0.004) 0.31	0.01 (0.003) 0.54	0.11 (0.002) 0.77	0.18 (0.002) 0.9	-0.05 (0.002) 0.35
5	37.77 (0.045) 1	-3.18 (0.141) 0.31	-0.51 (0.184) 0.47	-5.94 (0.132) 0.17	2.58 (0.104) 0.68	<b>6.64</b> (0.074) <b>0.92</b>	<b>10.24</b> (0.067) <b>0.99</b>	-1.9 (0.054) 0.33
6	37.66 (0.058) 1	-3.39 (0.176) 0.34	-0.19 (0.221) 0.49	-5.99 (0.157) 0.22	2.44 (0.128) 0.64	6.64 (0.092) 0.86	<b>10.25</b> (0.078) <b>0.97</b>	-1.9 (0.071) 0.36
7	3.72 (0.003) 1	-0.12 (0.006) 0.2	0.03 (0.008) 0.58	-0.07 (0.006) 0.3	-0.01 (0.006) 0.47	0.14 (0.005) 0.86	<b>0.2</b> (0.004) <b>0.98</b>	-0.07 (0.002) 0.21
8	3.7 (0.001) 1	-0.11 (0.004) 0.25	0.01 (0.005) 0.53	-0.11 (0.004) 0.25	0 (0.003) 0.52	0.15 (0.002) 0.89	<b>0.21</b> (0.002) <b>0.97</b>	-0.05 (0.001) 0.31

**Table C.2.** Coefficient posterior means (Monte Carlo standard errors) for all models for the *Education* sector.  $P(> 0)$  means the posterior probability that the corresponding coefficient is greater than 0. Coefficients (other than intercepts) that have  $> 90\%$  posterior probability of being either positive or negative are bolded.

Covariate	Intercept mean (mcse) $P(> 0)$	Poverty mean (mcse) $P(> 0)$	Dist to School mean (mcse) $P(> 0)$	All Injured mean (mcse) $P(> 0)$	Food % mean (mcse) $P(> 0)$	Land Cultivated mean (mcse) $P(> 0)$	Electric % mean (mcse) $P(> 0)$	Population mean (mcse) $P(> 0)$
1	3.51 (0.001) 1	<b>0.14</b> (0.002) <b>0.96</b>	<b>-0.17</b> (0.002) <b>0.03</b>	<b>0.18</b> (0.001) <b>0.99</b>	-0.04 (0.002) 0.28	0.01 (0.001) 0.6	-0.04 (0.001) 0.29	-0.06 (0.001) 0.19
2	3.5 (0.001) 1	0.09 (0.003) 0.8	-0.16 (0.004) 0.13	<b>0.17</b> (0.002) <b>0.97</b>	0.04 (0.004) 0.5	-0.02 (0.001) 0.43	-0.03 (0.001) 0.37	-0.06 (0.001) 0.28
3	3.5 (0.001) 1	0.18 (0.004) 0.87	-0.2 (0.005) 0.14	<b>0.23</b> (0.003) <b>0.93</b>	-0.09 (0.002) 0.26	0 (0.002) 0.49	-0.02 (0.002) 0.44	-0.05 (0.001) 0.33
4	3.5 (0.001) 1	0.19 (0.004) 0.83	-0.2 (0.005) 0.18	0.23 (0.004) 0.89	-0.09 (0.003) 0.29	0 (0.002) 0.5	-0.02 (0.002) 0.45	-0.04 (0.002) 0.36
5	34.61 (0.046) 1	6.44 (0.139) 0.85	-7.37 (0.184) 0.17	7.21 (0.129) 0.87	-2.61 (0.105) 0.32	0.16 (0.076) 0.51	-1.09 (0.066) 0.4	-2.22 (0.052) 0.3
6	34.46 (0.059) 1	6.07 (0.173) 0.78	-7.07 (0.227) 0.23	7.22 (0.163) 0.82	-2.59 (0.126) 0.35	0.07 (0.1) 0.5	-1.1 (0.08) 0.42	-2.25 (0.068) 0.34
7	3.61 (0.003) 1	<b>0.22</b> (0.007) <b>0.93</b>	<b>-0.25</b> (0.01) <b>0.1</b>	<b>0.22</b> (0.007) <b>0.93</b>	-0.11 (0.005) 0.19	0.01 (0.004) 0.55	0 (0.004) 0.46	-0.05 (0.003) 0.28
8	3.63 (0.001) 1	0.21 (0.003) 0.9	<b>-0.27</b> (0.004) <b>0.09</b>	<b>0.26</b> (0.004) <b>0.94</b>	-0.14 (0.003) 0.16	0 (0.002) 0.52	-0.01 (0.002) 0.45	-0.05 (0.001) 0.3

**Table C.3.** Coefficient posterior means (Monte Carlo standard errors) for all models for the *Governance* sector.  $P(> 0)$  means the posterior probability that the corresponding coefficient is greater than 0. Coefficients (other than intercepts) that have  $> 90\%$  posterior probability of being either positive or negative are bolded.

Covariate	Intercept mean (mcse)	Poverty mean (mcse)	Dist to School mean (mcse)	All Injured mean (mcse)	Food % mean (mcse)	Land Cultivated mean (mcse)	Electric % mean (mcse)	Population mean (mcse)
Model	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$
1	4.21 (0)	-0.02 (0.001)	0.03 (0.001)	0 (0.001)	-0.01 (0.001)	-0.02 (0)	0.03 (0)	<b>0.04</b> (0)
	1	0.34	0.7	0.5	0.4	0.28	0.87	<b>0.92</b>
2	4.21 (0)	-0.02 (0.001)	0.03 (0.002)	0 (0.001)	-0.01 (0.001)	-0.02 (0.001)	0.03 (0.001)	0.04 (0.001)
	1	0.37	0.66	0.52	0.42	0.31	0.8	0.85
3	4.21 (0.001)	-0.01 (0.003)	0.01 (0.004)	0 (0.003)	-0.01 (0.003)	-0.02 (0.002)	0.04 (0.002)	0.04 (0.001)
	1	0.48	0.53	0.51	0.46	0.44	0.63	0.64
4	4.21 (0.001)	-0.01 (0.004)	0.02 (0.005)	0 (0.004)	-0.02 (0.003)	-0.02 (0.002)	0.03 (0.002)	0.03 (0.002)
	1	0.48	0.53	0.5	0.47	0.45	0.6	0.6
5	67.6 (0.047)	-1.21 (0.142)	1.62 (0.182)	0.13 (0.126)	-1.02 (0.105)	-1.25 (0.077)	2.36 (0.064)	2.83 (0.053)
	1	0.43	0.58	0.51	0.43	0.4	0.7	0.74
6	67.58 (0.062)	-1.06 (0.174)	1.56 (0.231)	0.19 (0.16)	-1.17 (0.131)	-1.2 (0.098)	2.38 (0.08)	2.87 (0.067)
	1	0.44	0.56	0.52	0.43	0.42	0.67	0.7
7	4.54 (0.003)	0.04 (0.004)	-0.11 (0.007)	0.04 (0.005)	0.05 (0.006)	0.11 (0.006)	-0.06 (0.003)	0 (0.002)
	1	0.65	0.22	0.66	0.67	0.78	0.26	0.48
8	4.3 (0.001)	0 (0.003)	0 (0.004)	0 (0.003)	-0.05 (0.003)	-0.01 (0.002)	0.05 (0.002)	0.04 (0.001)
	1	0.5	0.49	0.5	0.38	0.48	0.65	0.63

**Table C.4.** Coefficient posterior means (Monte Carlo standard errors) for all models for the *Health* sector.  $P(> 0)$  means the posterior probability that the corresponding coefficient is greater than 0. Coefficients (other than intercepts) that have  $> 90\%$  posterior probability of being either positive or negative are bolded.

Covariate	Intercept mean (mcse)	Poverty mean (mcse)	Dist to School mean (mcse)	All Injured mean (mcse)	Food % mean (mcse)	Land Cultivated mean (mcse)	Electric % mean (mcse)	Population mean (mcse)
Model	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$
1	4.47 (0)	<b>0.07</b> (0.001)	-0.02 (0.001)	0.01 (0.001)	0 (0.001)	<b>-0.11</b> (0)	<b>0.04</b> (0)	-0.02 (0)
	1	<b>0.99</b>	0.31	0.6	0.46	<b>0</b>	<b>0.97</b>	0.18
2	4.46 (0)	<b>0.07</b> (0.001)	-0.02 (0.001)	0.01 (0.001)	0 (0.001)	<b>-0.11</b> (0.001)	<b>0.04</b> (0)	-0.02 (0)
	1	<b>0.97</b>	0.31	0.65	0.43	<b>0</b>	<b>0.93</b>	0.25
3	4.47 (0.001)	0.07 (0.003)	-0.03 (0.005)	0.02 (0.003)	0 (0.003)	-0.1 (0.002)	0.04 (0.002)	-0.02 (0.001)
	1	0.66	0.44	0.55	0.5	0.2	0.64	0.43
4	4.47 (0.001)	0.07 (0.004)	-0.04 (0.006)	0.03 (0.004)	-0.01 (0.003)	-0.09 (0.002)	0.04 (0.002)	-0.02 (0.002)
	1	0.65	0.43	0.56	0.48	0.26	0.63	0.43
5	88.67 (0.048)	7.43 (0.136)	-3.23 (0.173)	2.02 (0.125)	-0.65 (0.104)	<b>-9.57</b> (0.074)	3.69 (0.063)	-2.1 (0.054)
	1	0.88	0.33	0.63	0.45	<b>0.02</b>	0.79	0.31
6	88.62 (0.059)	7.42 (0.178)	-3.07 (0.232)	1.99 (0.155)	-0.66 (0.126)	<b>-9.57</b> (0.099)	3.59 (0.083)	-2.3 (0.071)
	1	0.82	0.37	0.61	0.47	<b>0.06</b>	0.74	0.33
7	4.58 (0.003)	0.07 (0.008)	-0.04 (0.01)	0 (0.007)	-0.01 (0.006)	-0.08 (0.004)	0.05 (0.003)	-0.03 (0.003)
	1	0.65	0.44	0.48	0.49	0.24	0.67	0.34
8	4.57 (0.001)	0.08 (0.004)	-0.04 (0.005)	0.02 (0.003)	-0.04 (0.003)	-0.1 (0.002)	0.05 (0.002)	-0.03 (0.001)
	1	0.69	0.43	0.54	0.41	0.2	0.65	0.38

**Table C.5.** Coefficient posterior means (Monte Carlo standard errors) for all models for the *RD* sector.  $P(> 0)$  means the posterior probability that the corresponding coefficient is greater than 0. Coefficients (other than intercepts) that have  $> 90\%$  posterior probability of being either positive or negative are bolded.

Covariate	Intercept mean (mcse)	Poverty mean (mcse)	Dist to School mean (mcse)	All Injured mean (mcse)	Food % mean (mcse)	Land Cultivated mean (mcse)	Electric % mean (mcse)	Population mean (mcse)
Model	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$	$P(> 0)$
1	2.6 (0.002)	<b>-0.27</b> (0.004)	<b>0.41</b> (0.005)	-0.12 (0.002)	0.05 (0.004)	-0.07 (0.002)	0.09 (0.002)	-0.1 (0.002)
	1	<b>0.06</b>	<b>0.99</b>	0.14	0.58	0.25	0.82	0.22
2	2.69 (0.002)	<b>-0.29</b> (0.004)	<b>0.42</b> (0.005)	-0.14 (0.003)	0.06 (0.003)	-0.08 (0.002)	0.08 (0.001)	-0.11 (0.001)
	1	<b>0.04</b>	<b>0.98</b>	0.17	0.64	0.23	0.81	0.13
3	2.36 (0.001)	<b>-0.52</b> (0.003)	<b>0.76</b> (0.004)	<b>-0.19</b> (0.003)	-0.01 (0.003)	-0.07 (0.002)	<b>0.15</b> (0.002)	<b>-0.18</b> (0.001)
	1	<b>0</b>	<b>1</b>	<b>0.09</b>	0.48	0.28	<b>0.92</b>	<b>0.04</b>
4	2.36 (0.001)	<b>-0.53</b> (0.004)	<b>0.76</b> (0.005)	-0.2 (0.004)	0 (0.003)	-0.06 (0.002)	0.15 (0.002)	<b>-0.18</b> (0.002)
	1	<b>0</b>	<b>1</b>	0.14	0.5	0.33	0.85	<b>0.09</b>
5	13.33 (0.047)	-4.03 (0.141)	6.95 (0.179)	-2.52 (0.127)	0.67 (0.102)	-1.82 (0.073)	1.73 (0.061)	-2.32 (0.054)
	1	0.27	0.82	0.34	0.55	0.35	0.65	0.3
6	13.39 (0.058)	-4.27 (0.18)	7.32 (0.234)	-2.71 (0.157)	0.68 (0.123)	-1.85 (0.096)	1.73 (0.079)	-2.46 (0.068)
	1	0.3	0.77	0.36	0.54	0.38	0.62	0.32
7	2.6 (0.003)	<b>-0.54</b> (0.005)	<b>0.72</b> (0.008)	-0.07 (0.006)	-0.04 (0.005)	-0.07 (0.004)	0.14 (0.004)	<b>-0.19</b> (0.003)
	1	<b>0</b>	<b>1</b>	0.31	0.39	0.27	0.9	<b>0.04</b>
8	2.71 (0.002)	<b>-0.29</b> (0.006)	<b>0.56</b> (0.007)	-0.21 (0.004)	0.03 (0.003)	-0.07 (0.003)	0.09 (0.002)	-0.12 (0.003)
	1	<b>0.1</b>	<b>0.98</b>	0.12	0.59	0.32	0.76	0.22

**Table C.6.** Coefficient posterior means (Monte Carlo standard errors) for all models for the *RPT* sector.  $P(> 0)$  means the posterior probability that the corresponding coefficient is greater than 0. Coefficients (other than intercepts) that have  $> 90\%$  posterior probability of being either positive or negative are bolded.

Covariate	Intercept mean (mcse) $P(> 0)$	Poverty mean (mcse) $P(> 0)$	Dist to School mean (mcse) $P(> 0)$	All Injured mean (mcse) $P(> 0)$	Food % mean (mcse) $P(> 0)$	Land Cultivated mean (mcse) $P(> 0)$	Electric % mean (mcse) $P(> 0)$	Population mean (mcse) $P(> 0)$
1	2.11 (0.004) 1	<b>0.89</b> (0.003) 1	-0.07 (0.006) 0.38	<b>-0.39</b> (0.004) <b>0.01</b>	0.21 (0.004) 0.9	0.11 (0.003) 0.8	<b>0.7</b> (0.002) 1	<b>0.43</b> (0.002) 1
2	2.54 (0.005) 1	<b>0.34</b> (0.006) <b>0.94</b>	-0.11 (0.006) 0.31	<b>0.59</b> (0.004) 1	<b>-0.68</b> (0.005) <b>0</b>	<b>-0.52</b> (0.003) <b>0</b>	<b>0.94</b> (0.004) 1	<b>0.39</b> (0.003) 1
3	2.98 (0.001) 1	-0.2 (0.003) 0.1	0.23 (0.004) 0.89	-0.17 (0.003) 0.12	<b>0.32</b> (0.002) <b>0.99</b>	<b>-0.47</b> (0.002) <b>0</b>	<b>0.32</b> (0.002) 1	-0.03 (0.001) 0.4
4	2.98 (0.001) 1	-0.19 (0.004) 0.16	0.22 (0.005) 0.83	-0.16 (0.004) 0.2	<b>0.31</b> (0.003) <b>0.96</b>	<b>-0.47</b> (0.002) <b>0</b>	<b>0.32</b> (0.002) <b>0.99</b>	-0.02 (0.002) 0.44
5	38.63 (0.045) 1	<b>-23.6</b> (0.14) <b>0</b>	<b>19.33</b> (0.179) 1	-7.41 (0.128) 0.12	<b>10.82</b> (0.104) <b>0.97</b>	<b>-11.87</b> (0.073) <b>0.01</b>	<b>13.93</b> (0.064) 1	<b>-8.57</b> (0.055) <b>0.02</b>
6	38.56 (0.057) 1	<b>-23.36</b> (0.176) <b>0</b>	<b>18.92</b> (0.222) <b>0.98</b>	-7.48 (0.161) 0.17	<b>10.92</b> (0.127) <b>0.94</b>	<b>-11.68</b> (0.094) <b>0.03</b>	<b>13.96</b> (0.081) <b>0.99</b>	<b>-8.69</b> (0.072) <b>0.06</b>
7	3.48 (0.003) 1	<b>-0.44</b> (0.006) <b>0</b>	<b>0.47</b> (0.009) <b>0.98</b>	<b>-0.25</b> (0.006) <b>0.03</b>	<b>0.48</b> (0.005) 1	<b>-0.52</b> (0.004) <b>0</b>	<b>0.19</b> (0.004) <b>0.94</b>	<b>-0.21</b> (0.003) <b>0.03</b>
8	3.71 (0.002) 1	<b>-0.66</b> (0.004) <b>0</b>	<b>0.48</b> (0.007) <b>0.93</b>	<b>-0.62</b> (0.004) <b>0</b>	<b>0.71</b> (0.003) 1	<b>-0.33</b> (0.004) <b>0.04</b>	<b>0.28</b> (0.003) <b>0.99</b>	<b>-0.18</b> (0.002) <b>0.06</b>

**Table C.7.** Coefficient posterior means (Monte Carlo standard errors) for all models for the *WSI* sector.  $P(> 0)$  means the posterior probability that the corresponding coefficient is greater than 0. Coefficients (other than intercepts) that have  $> 90\%$  posterior probability of being either positive or negative are bolded.

Covariate	Intercept mean (mcse) $P(> 0)$	Poverty mean (mcse) $P(> 0)$	Dist to School mean (mcse) $P(> 0)$	All Injured mean (mcse) $P(> 0)$	Food % mean (mcse) $P(> 0)$	Land Cultivated mean (mcse) $P(> 0)$	Electric % mean (mcse) $P(> 0)$	Population mean (mcse) $P(> 0)$
1	2.66 (0.002) 1	0.06 (0.003) 0.65	0 (0.004) 0.51	0.06 (0.004) 0.64	-0.1 (0.003) 0.2	<b>0.15</b> (0.002) <b>0.9</b>	-0.17 (0.003) 0.16	0.05 (0.002) 0.69
2	2.76 (0.002) 1	-0.07 (0.006) 0.37	0.21 (0.006) 0.81	0.06 (0.003) 0.66	-0.15 (0.004) 0.16	-0.03 (0.002) 0.38	-0.15 (0.004) 0.19	0.09 (0.002) 0.78
3	2.67 (0.001) 1	0.03 (0.003) 0.58	0.13 (0.004) 0.76	<b>0.26</b> (0.003) <b>0.95</b>	<b>-0.39</b> (0.003) <b>0</b>	-0.12 (0.002) 0.14	-0.12 (0.002) 0.14	0.08 (0.001) 0.78
4	2.66 (0.001) 1	0.04 (0.004) 0.58	0.12 (0.006) 0.7	<b>0.26</b> (0.004) <b>0.91</b>	<b>-0.39</b> (0.003) <b>0.01</b>	-0.12 (0.002) 0.21	-0.11 (0.002) 0.21	0.08 (0.002) 0.73
5	19.77 (0.048) 1	1.04 (0.137) 0.57	3.49 (0.183) 0.67	4.16 (0.126) 0.75	-7.18 (0.104) 0.1	-3.31 (0.078) 0.25	0.18 (0.067) 0.51	-0.63 (0.055) 0.44
6	19.75 (0.057) 1	1.01 (0.177) 0.55	3.5 (0.226) 0.63	4.26 (0.164) 0.71	-7.2 (0.129) 0.15	-3.47 (0.092) 0.29	0.11 (0.083) 0.5	-0.72 (0.068) 0.45
7	3.07 (0.003) 1	-0.09 (0.007) 0.26	<b>0.3</b> (0.009) <b>0.95</b>	<b>0.32</b> (0.007) <b>0.97</b>	<b>-0.31</b> (0.006) <b>0.01</b>	<b>-0.32</b> (0.005) <b>0</b>	-0.01 (0.003) 0.44	-0.08 (0.003) 0.19
8	3.16 (0.001) 1	0.04 (0.002) 0.59	0.16 (0.004) 0.81	<b>0.36</b> (0.004) <b>0.97</b>	<b>-0.39</b> (0.003) <b>0.01</b>	<b>-0.33</b> (0.003) <b>0.02</b>	-0.02 (0.001) 0.41	-0.07 (0.001) 0.21

## Acknowledgments

The authors thank the associate editors and referees for their constructive and helpful feedback. Additionally, C. Berrett thanks Dr. Amanda S. Hering for her insights during a discussion about the project.

## References

- Aced, C. (2009), "Aid Effectiveness in Education," *Prospects*, 39, 107–108. [1]
- Alesina, A., and Dollar, D. (2000), "Who Gives Foreign Aid to Whom and Why?" *Journal of Economic Growth*, 5, 33–63. [1]
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014), *Hierarchical Modeling and Analysis for Spatial Data* (2nd ed.), Boca Raton, FL: Chapman & Hall/CRC. [3,4]
- Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society, Series B*, 36, 192–236. [4]
- Briggs, R. C. (2017), "Does Foreign Aid Target the Poorest?" *International Organization*, 71, 187–206. [2]
- Collier, P., and Dollar, D. (2002), "Aid Allocation and Poverty Reduction," *European Economic Review*, 46, 1475–1500. [2]
- De, R., and Becker, C. (2015), "The Foreign Aid Effectiveness Debate: Evidence from Malawi," AidData Working Paper 6, Williamsburg, VA: AidData, available at <http://aiddata.org/aiddata-working-paper-series>. [2]
- Dietrich, S. (2013), "Bypass or Engage? Explaining Donor Delivery Tactics in Foreign Aid Allocation," *International Studies Quarterly*, 57, 698–712. [2]
- Dietrich, S., and Wright, J. (2015), "Foreign Aid Allocation Tactics and Democratic Change in Africa," *The Journal of Politics*, 77, 216–234. [2]
- Diggle, P. J., Moyeed, R. A., and Tawn, J. A. (1998), "Model-Based Geostatistics" (with discussion and rejoinder), *Applied Statistics*, 47, 299–350. [3]
- Dobson, A. J., and Barnett, G. (2008), *An Introduction to Generalized Linear Models* (3rd ed.), Boca Raton, FL: Chapman & Hall/CRC. [6]
- Feeny, S. (2007), "Impacts of Foreign Aid to Melanesia," *Journal of the Asia Pacific Economy*, 12, 34–60. [2]
- Findley, M. G., Marineau, J. F., and Powell, J. (2016), "The Swarm Principle: A Sub-National Spatial Analysis of Aid Targeting and Donor Coordination in Sub-Saharan Africa," unpublished manuscript, University of Texas at Austin. [1]
- Findley, M. G., Powell, J., Strandow, D., and Tanner, J. (2011), "The Localized Geography of Foreign Aid: A New Dataset and Application to Violent Armed Conflict," *World Development*, 39, 1995–2009. [1]
- Flegal, J. M., Hughes, J., Vats, D., and Dai, N. (2017), *mcmcse: Monte Carlo Standard Errors for MCMC, R Package Version 1.3-2*, Riverside, CA. [5]
- Fox, J., and Monette, G. (1992), "Generalized Collinearity Diagnostics," *Journal of the American Statistical Association*, 87, 178–183. [6]
- Gelman, A., and Rubin, D. B. (1992), "Inference from Iterative Simulation using Multiple Sequences," *Statistical Science*, 7, 457–511. [5]
- Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments," in *Bayesian Statistics 4*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., Oxford, UK: Clarendon Press, pp. 169–193. [5]
- Hodler, R., and Raschky, P. A. (2014), "Regional Favoritism," *The Quarterly Journal of Economics*, 129, 995–1033. [2]
- Hughes, J., and Haran, M. (2013), "Dimension Reduction and Alleviation of Confounding for Spatial Generalized Linear Mixed Models," *Journal of the Royal Statistical Society, Series B*, 75, 139–159. [4]
- Joablonski, R. S. (2014), "How Aid Targets Votes: The Impact of Electoral Incentives on Foreign Aid Distribution," *World Politics*, 66, 293–330. [2]
- Kalbfleisch, J. D., and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data* (2nd ed.), Hoboken, NJ: Wiley. [4,9]
- Masaki, T. (2017), "The Political Economy of Aid Allocation in Africa: Evidence from Zambia," *African Studies Review*, 61, 55–82. Available at [http://takaakimasaki.com/wp-content/uploads/2014/08/zambia\\_aid\\_allocatio\\_n\\_2016-9-23.pdf](http://takaakimasaki.com/wp-content/uploads/2014/08/zambia_aid_allocatio_n_2016-9-23.pdf). [2]
- McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), Boca Raton, FL: CRC Press. [4]
- National Statistical Office. (2008), *2008 Population and Housing Census*, Zomba, Malawi: Government of Malawi, available at [http://www.nsomalawi.mw/images/stories/data\\_on\\_line/demography/census\\_2008/Main%20Report/Census%20Main%20Report.pdf](http://www.nsomalawi.mw/images/stories/data_on_line/demography/census_2008/Main%20Report/Census%20Main%20Report.pdf) [2]
- (2015), *Malawi in Figures 2015*, Zomba, Malawi: Government of Malawi, available at [http://www.nsomalawi.mw/images/stories/data\\_on\\_line/general/malawi\\_in\\_figures/Malawi%20in%20Figures%202015.pdf](http://www.nsomalawi.mw/images/stories/data_on_line/general/malawi_in_figures/Malawi%20in%20Figures%202015.pdf) [1]
- Nunnenkamp, P., Öhler, H., and Sosa Andrés, M. (2017), "Need, Merit and Politics in Multilateral Aid Allocation: A District-level Analysis of World Bank Projects in India," *Review of Development Economics*, 21, 126–156. [2]
- Nunnenkamp, P., Sotirova, A., and Thiele, R. (2015), "Do Aid Donors Specialize and Coordinate within Recipient Countries? The Case of Malawi," AidData Working Paper 10, Williamsburg, VA: AidData, available at <http://aiddata.org/aiddata-working-paper-series>. [2]
- OECD (2015), *OECD Factbook 2015: Economic, Environmental and Social Statistics*, Paris, France: OECD Publishing. [1]
- Peratsakis, C., Powell, J., Findley, M., Baker, J., and Weaver, C. (2012), *Geocoded Activity-Level Data from the Government of Malawi's Aid Management Platform*, Washington, DC: AidData and the Robert S. Strauss Center for International Security and Law, available at <http://aiddata.org/subnational-geospatial-research-datasets> [1,2]
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006), "CODA: Convergence Diagnosis and Output Analysis for MCMC," *R News*, 6, 7–11. [5]
- R Core Team (2016), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [5]
- Reich, B. J., Hodges, J. S., and Zadnik, V. (2006), "Effects of Residual Smoothing on the Posterior of the Fixed Effects in Disease-Mapping Models," *Biometrics*, 62, 1197–1206. [4]
- Roberts, G. O., and Rosenthal, J. S. (2009), "Examples of Adaptive MCMC," *Journal of Computational and Graphical Statistics*, 18, 349–367. [5]
- Runfola, D., Marty, R., Goodman, S., Lefew, M., and B., Y. A. (2017), "geoSIMEX: A Generalized Approach to Modeling Spatial Imprecision," AidData Working Paper 38, Williamsburg, VA: AidData, available at <http://aiddata.org/aiddata-working-paper-series>. [2]
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society, Series B*, 64, 583–639. [5]
- (2014), "The Deviance Information Criterion: 12 Years On," *Journal of the Royal Statistical Society, Series B*, 76, 485–493. [5]
- van Weezel, S. (2015), "A Spatial Analysis of the Effect of Foreign Aid in Conflict Areas," AidData Working Paper 8, Williamsburg, VA: AidData, available at <http://aiddata.org/aiddata-working-paper-series>. [2]
- Weaver, C., Davenport, S., Baker, J., Findley, M., Peratsakis, C., and Powell, J. (2014), *Malawi's Open Aid Map*, Washington, DC: The World Bank, available at [http://wbi.worldbank.org/wbi/Data/wbi/wbicms/files/drupal-acquia/wbi/malawi\\_case\\_study2\\_0.pdf](http://wbi.worldbank.org/wbi/Data/wbi/wbicms/files/drupal-acquia/wbi/malawi_case_study2_0.pdf). [1]