Online Appendix for (Under What Conditions) Do Politicians Reward Their Supporters? Evidence From Kenya's Constituencies Development Fund

A Technical Details of the Point Process Model

The discussion here closely follows Baddeley (2010). The observed data in our analysis are the locations of n CDF projects, $x = \{x_1, ..., x_n\}$, whose spatial distribution is a realization of the point process X in a given constituency R; $x \in R$. The Poisson process model estimates parameters of the intensity function for all locations $u \in R$. The intensity function is:

$$E\left[N(X \cap B)\right] = \int_B \lambda(u)du$$

where $E\left[N(X \cap B)\right]$ is the expected number of points in B, a region within R. For R we can estimate the intensity as the count of points in x divided by the area of R. This is the intensity in the entire constituency. Point patterns may not occur with uniform intensity, since some areas of a constituency likely receive more projects than others.

We define $\lambda(u)$ is the intensity of a local Poisson process at location u. Note that covariates Z are measured at every point in R. The stochastic component of the model is defined as:

$$X \sim \text{Poisson}(\lambda(u))$$

The systematic component of the model is defined as:

$$\lambda(u) = e^{Z(u)\beta}$$

The assumptions for the point process model are familiar to regular users of standard generalized linear models. First, the observations (project locations and dummy points) are independent of one another. While this is rarely strictly true in any kind of data, we constructed our data in a way to better fit this assumption. We counted only unique project locations, rather than treating each individual project in a given year as a separate project. For instance, if CDF funds went to projects at Huduma Primary School in several years (e.g., to build several new classrooms across several years or if a single project had a funding allocation recorded over multiple years in the CDF database), we represent this as a single project in our dataset. Second, the intensity function (reporting the propensity for an area u to receive projects) is log-linear in the spatial covariates, as is standard in the Poisson generalized linear model and given the non-negative nature of count-type data. Renner et al. (2015) discusses these modeling assumptions in more detail.

$Z(u)$ are the values of spatial covariates at location u; these are defined at every point in the study area (in this case, in each of the 196 constituencies), and stored as high-resolution raster data. Our definition of units of analysis for estimation in this framework follow from the point nature of the data. Two kinds of points are used to estimate the intensity λ as a function of

spatial covariates: points representing actual project locations and "dummy" points representing "pseudo-absences," or places without a project. Modeling continuous space is not computationally feasible, so we break up continuous space using the dummy point scheme. This combined set of points form a quadrature scheme that breaks up the area of analysis R into disjoint spatial units ("tiles") that can be analyzed using familiar Poisson log-linear regression.

We make two choices regarding the model defaults in our analysis. Although these choices do not affect the substantive results, we report them here for transparency (we also report robustness to alternative dummy point choices below). First, we face a choice regarding the number of dummy points to include in each constituency-level point process model. A higher number of dummy points leads to a more stable estimate, but at significant computational cost. Ideally, we would set the density of dummy points identically for all constituencies. However, this approach would lead to a computationally impractical number of dummy points for large constituencies (e.g., virtually anywhere in North Eastern Province). As a result, we vary the number of dummy points used as a flexible function of constituency area. To do so, we calculate the bounding box of the constituency (in meters), and set a quantity Q equal to the longest dimension of that bounding box divided by 100. Then we set the spacing of dummy points equal to max(Q, 250). This ensures that, for large constituencies, we retain a relatively fine grid of dummy points (ensuring high approximation of two-dimensional space). For small urban constituencies, this ensures that the dummy points are spaced 250 meters apart.

The second choice regards the methods for estimating the parameters of interest. Options include maximum pseudolikelihood, logistic likelihood, variational Bayes likelihood, and the Huang-Ogata method. We use the maximum pseudolikelihood method, as it is equivalent to the maximum likelihood in the case of Poisson regression and is unbiased in the presence of a large number of dummy points (such as the number we specify). See Baddeley and Turner (2000) and Baddeley and Turner (2005) for further details.

We estimate all models using the `spatstat` package in `R` (Baddeley et al. 2015).

A.1 Dummy Point Specifications

Our choice to use a fine grid of points follows the recommendations in the literature, which suggests that finer grids of dummy points provide a better approximation of the likelihood function, reduce bias, and improve RMSE (Warton and Shepherd 2010; Baddeley and Turner 2000, p. 312).

We show that this modeling choice has little effect on our substantive findings. Tables A1 and A2 replicate Table 3 from the paper, using dummy point grids of 1 and 2.5 kilometer resolution respectively. In some cases, the statistical significance of the coefficients degrade, but the sign and overall magnitude are quite similar to the better specified dummy point configuration included in

the paper. An alternative approach would be to randomly allocate dummy points, rather than use a fine grid. Renner et al. (2015) and Warton and Shepherd (2010) discuss the benefits of a grid over the random approach.

Table A1: Explaining the Relationship Between Project Placement and Residual Support: 1 kilometer dummy point spacing + normal linear model

	Dependent variable:								
	Coefficient of Project Placement and Residual Support								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	−0.11	−0.11					−0.11	−0.12*	−0.16**
	(0.07)	(0.07)					(0.07)	(0.07)	(0.07)
Member of Ruling Coalition	−0.05*	−0.07***					−0.06**	−0.07**	−0.08***
	(0.02)	(0.03)					(0.03)	(0.03)	(0.03)
Incumbent	0.01	0.003					0.01	0.01	0.01
	(0.03)	(0.03)					(0.03)	(0.03)	(0.03)
Margin of Victory (2002)			0.01	0.004			0.02	0.02	0.02
			(0.03)	(0.03)			(0.03)	(0.03)	(0.03)
Segregation of Supporters			0.08	0.12			0.07	0.12	0.28***
			(0.07)	(0.08)			(0.07)	(0.08)	(0.11)
Ethnic Heterogeneity					−0.15	0.06	0.06	0.16	0.11
					(0.38)	(0.50)	(0.39)	(0.49)	(0.49)
Population Clustering					−0.04*	−0.04	−0.06**	−0.06*	−0.06*
					(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Margin of Victory × Segregation									−0.35**
									(0.14)
Constant	0.03	0.06	−0.002	0.01	0.002	−0.003	0.03	0.05	0.05
	(0.02)	(0.04)	(0.02)	(0.04)	(0.01)	(0.03)	(0.02)	(0.04)	(0.04)
Province FE	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	196	196	196	196	196	196	196	196	196
R^2	0.03	0.07	0.01	0.03	0.02	0.03	0.07	0.09	0.12
Adjusted R^2	0.02	0.02	−0.003	−0.02	0.01	−0.02	0.04	0.02	0.05

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are estimated using OLS, with province fixed effects as noted. Outcome variable is the estimated coefficient of the supporters variable from the constituency-level point-process model examining the determinants of project placement, where dummy points are spaced at one kilometer intervals.

Table A2: Explaining the Relationship Between Project Placement and Residual Support: 2.5 kilometer dummy point spacing + normal linear model

	Dependent variable:								
	Coefficient of Project Placement and Residual Support								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	−0.11	−0.11					−0.10	−0.12	−0.15**
	(0.07)	(0.07)					(0.07)	(0.07)	(0.07)
Member of Ruling Coalition	−0.04*	−0.07**					−0.05**	−0.07**	−0.07**
	(0.03)	(0.03)					(0.03)	(0.03)	(0.03)
Incumbent	0.01	0.003					0.02	0.01	0.01
	(0.03)	(0.03)					(0.03)	(0.03)	(0.03)
Margin of Victory (2002)			0.01	0.01			0.02	0.03	0.03
			(0.03)	(0.03)			(0.03)	(0.03)	(0.03)
Segregation of Supporters			0.10	0.13			0.08	0.13	0.28***
			(0.07)	(0.08)			(0.08)	(0.08)	(0.11)
Ethnic Heterogeneity					−0.30	−0.13	−0.10	−0.03	−0.07
					(0.38)	(0.51)	(0.39)	(0.50)	(0.50)
Population Clustering					−0.03	−0.03	−0.05**	−0.05	−0.05
					(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Margin of Victory × Segregation									−0.33**
									(0.15)
Constant	0.03	0.05	−0.01	0.01	0.001	−0.01	0.02	0.04	0.04
	(0.02)	(0.04)	(0.02)	(0.04)	(0.01)	(0.03)	(0.02)	(0.04)	(0.04)
Province FE	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	196	196	196	196	196	196	196	196	196
R^2	0.03	0.06	0.01	0.03	0.02	0.03	0.06	0.08	0.11
Adjusted R^2	0.01	0.01	−0.001	−0.02	0.01	−0.02	0.03	0.01	0.03

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are estimated using OLS, with province fixed effects as noted. Outcome variable is the estimated coefficient of the supporters variable from the constituency-level point-process model examining the determinants of project placement, where dummy points are spaced at 2.5 kilometer intervals.

B Description of Data Construction and Coding

This appendix describes data sources, construction, and coding for both the intra-constituency analyses generating the coefficients in Figures 4 and 5 and the cross-constituency analyses in Table 3. Table B1 provides summary statistics for the untransformed convariates used in the cross-constituency regressions.

B.1 Intra-constituency Analyses

Number of Supporters: Construction of the spatial distribution of the number of supporters proceeds in two steps. First, we generate a Voronoi diagram using the polling stations in a given constituency, which partitions the constituency into tiles partitioning the constituency. Then, we multiply the number of supporters of the MP by the spatial distribution of population in that tile. The main difference here (relative to the calculation of coethnics described below) is that prior to multiplying the number of supporters by population, we divide the respective tile of the diagram by its total population, providing a raster tile layer containing the proportion of the population in each tile of the grid square. By multiplying this "weighting" layer by the number of supporters in the polling station catchment area (represented by the tile), we approximate the spatial distribution of the number of supporters in the vicinity of the polling station.

Population Density: We use population density raster data available from `http://www.worldpop.org.uk`. This data provides data on the spatial distribution of population at a high degree of geographic resolution (approximately $100m^2$), providing the estimate number of people within the raster grid square. See Linard et al. (2012) for more information.

Number of Coethnics: We calculate the number of coethnics in two steps. First, using voter registration information and methods described in Harris (2015), we estimate the percentage of each ethnic group at the polling station level. Then, we generate a Voronoi diagram using the polling station points to partition each constituency into tiles defining areas closest to each polling station. We overlay these catchment areas over the population density raster described above, and then multiply the percentage of the MP's coethnics estimated at the polling station corresponding to a given tile with the population density within that tile. This provides us with an approximation of the number of coethnics and their spatial distribution within a given constituency.

Distance to MP's Home Polling Station (squared): We proxy the MP's home area by matching the MP to the polling station at which he or she is registered to vote. Then, we create

a raster reporting, for each grid cell, the distance between the polling station and the MP's home polling station.

Number Living in Poverty: To create a raster representing the spatial distribution of those in poverty in a given constituency, we combined information from the population density raster with information from the Kenya 1km poverty raster, which reports the proportion of individuals in a grid square below the poverty line and is available at `http://www.worldpop.org.uk`. To arrive at a count of those falling below the poverty line for each grid square, we reprojected the poverty data to match that of the population density raster and then multiplied the poverty raster times the population density raster. See Tatem et al. (2015) for more information on the poverty data

Distance to Road: We utilize the World Bank/Kenya Ministry of Roads and Public Works dataset described in Government of Kenya (2006) to create a raster identifying the square of the distance from each point in each constituency to a paved road.

B.2 Cross-constituency Analysis

Female: Coded as zero if the MP is male, one if the MP is female.

Member of Ruling Coalition: Coded as one if the MP was a member of the National Rainbow Coalition, the winning presidential candidate's party, zero otherwise.

Incumbent: Coded as zero if the MP is in his or her first term, one if the MP has previously held an MP position.

Margin of Victory – 2002: Coded as a dummy variable taking a value of 1 if the margin of victory, defined as the difference in percentage of vote share for the first and second place candidates in the 2002 parliamentary election, is above the median, zero otherwise.

Segregation of Supporters: This variable is constructed using the `seg` package in `R` available at `https://cran.r-project.org/web/packages/seg/`. Specifically, we use the spatial information theory index, H, described in Reardon and O'Sullivan (2004, p. 139):

$$H = 1 - \frac{1}{TE} \int_{p \in R} \tau_p E_p dp$$

where E is the overall entropy of the total population (see equation 8 in Reardon and O'Sullivan

(2004); τ_p is the population density at point p in region R; E_p is the entropy at point p; and T is the total population in R. This measure is increasing in segregation, equal to 1 when all individuals in the constituency are surrounded by people with identical political preferences.

Ethnic Heterogeneity: To calculate ethnic heterogeneity, we follow the method for estimation of ethnic identity outlined in Harris (2015), and calculate a traditional Theil-type index of ethnic heterogeneity at the constituency-level.

Population Clustering: We measure population clustering in terms of relative entropy, which we operationalize by comparing the grid square-level populations we observe in each constituency with the hypothetical situation in which the population was evenly distributed across all grid squares. Higher values represent more clustering. The measure is calculated using the function `RelativeEntropy` in the `RelValAnalysis` package (Wong 2014).

Table B1: Summary statistics for untransformed covariates for cross-constituency regressions.

Statistic	N	Mean	St. Dev.	Min	Max
Female	210	0.029	0.167	0	1
Member of national ruling coalition	210	0.595	0.492	0	1
Incumbent	210	0.400	0.491	0	1
Ethnic heterogeneity	210	0.338	0.262	0.000	0.867
Margin of victory (2002)	205	0.368	0.253	0.001	0.964
PPM coefficient	196	0.006	0.168	−0.695	0.462
PPM SE	196	0.127	0.081	0.009	0.521
Population clustering	210	0.629	0.658	0.024	3.389
Segregation of supporters	205	0.068	0.039	0.007	0.213
Above median vote margin	205	0.498	0.501	0	1

C Imputation of Missing Spatial Data

Our procedures for geo-referencing projects leads to some uncertainty about location. This uncertainty has two sources. The first is the uncertainty about the true locations of the features (schools, villages, health centers) to which the majority of our projects are exactly matched. This uncertainty may be due to human choices (e.g., where a GIS technician chooses to capture a point representing a school – for example, near the principal's office or in the middle of the school yard), or due to fundamental uncertainty based on technological or positional factors beyond the control of the data collector. For instance, if we have two measurements of the location of a school taken at two different times, it is not uncommon for these measurements to vary by 50 meters (understandable given that the area of a school easily encompasses such a span). Instead of simply picking one point, we impute the actual location of the school along the line connecting the two reports of its location.

The second source of uncertainty comes from projects that could not be exactly matched to a specific point and were instead matched to the smallest administrative unit (EA, sub-location, location) in which they could be located. Such project locations were imputed multiple times within the polygon representing that administrative unit with the likelihood of being placed at each point proportional to the estimated population density at that point. For example, suppose that a CDF Project provided Huruma Primary School with an additional classroom. If we were unable to link Huruma Primary School to a specific point in our database, but if the CDF records noted that Huruma was located in Makutano location (an administrative area smaller than a constituency), then we would randomly impute a point representing Huruma Primary School N times within Makutano location, running N separate analyses with each spatial imputation. We then combine the N results into one result using the formulas for combining point and uncertainty estimates from N analyses outlined in King et al. (2001). In this way, we reflect the fundamental uncertainty about the location of georeferenced projects in our data. We know of no other work that explicitly incorporates such effect-attenuating spatial uncertainty into an analysis.

C.1 Main results without imputation

In Table C1, we replicate the main results from Table 3, including only the highest quality point matches. That is, we re-estimate the constituency-level associations between project placement and residual support, excluding those projects whose locations were imputed within an administrative polygon, and then use these new estimates to reproduce Table 3. The results are substantively similar to the results included in the paper, suggesting that our imputation efforts are not driving the main results.

Table C1: Explaining the Relationship Between Project Placement and Residual Support: Highest quality matches only

	Dependent variable:								
	Coefficient of Project Placement and Residual Support								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.27^{***}	-0.27^{***}					-0.27^{***}	-0.28^{***}	-0.32^{***}
	(0.09)	(0.09)					(0.09)	(0.09)	(0.09)
Member of Ruling Coalition	-0.04	-0.07^{*}					-0.06	-0.07^{*}	-0.08^{**}
	(0.03)	(0.04)					(0.03)	(0.04)	(0.04)
Incumbent	0.04	0.03					0.04	0.03	0.04
	(0.03)	(0.03)					(0.03)	(0.03)	(0.03)
Margin of Victory (2002)			0.05	0.03			0.06^{*}	0.05	0.05
			(0.03)	(0.04)			(0.03)	(0.04)	(0.04)
Segregation of Supporters			0.13	0.14			0.11	0.14	0.33^{**}
			(0.09)	(0.11)			(0.10)	(0.11)	(0.14)
Ethnic Heterogeneity					-0.24	-0.01	0.03	0.15	0.09
					(0.50)	(0.65)	(0.51)	(0.64)	(0.64)
Population Clustering					-0.02	0.01	-0.04	-0.02	-0.02
					(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
Margin of Victory × Segregation									-0.42^{**}
									(0.19)
Constant	0.03	0.03	-0.01	-0.03	0.01	-0.03	0.01	0.02	0.01
	(0.03)	(0.05)	(0.02)	(0.05)	(0.02)	(0.04)	(0.03)	(0.05)	(0.05)
Province FE	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	196	196	196	196	196	196	196	196	196
R^2	0.06	0.10	0.02	0.05	0.01	0.04	0.09	0.12	0.14
Adjusted R^2	0.05	0.05	0.01	0.0004	-0.01	-0.01	0.05	0.05	0.07

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are estimated using OLS, with province fixed effects as noted. Outcome variable is the estimated coefficient of the supporters variable from the constituency-level point-process models, using only the highest quality matches for projects.

D Aggregated Ward-Level Results

A novel aspect of our analysis is the decision to model the spatial variation in project placement directly without relying on aggregated units. Figure D1 illustrates the different possible levels of aggregation that one might adopt to study the allocation of development projects in Kenya: the district (top left panel; $N = 71$), as in Burgess et al. (2015); the parliamentary constituency (top right panel; $N = 210$), as in Jablonski (2014); or the local government ward (bottom left panel; roughly 10-12 per constituency). This contrasts with our approach, which is to avoid aggregation altogether and to study the distribution of points in continuous space within each constituency (as depicted in the bottom right panel).

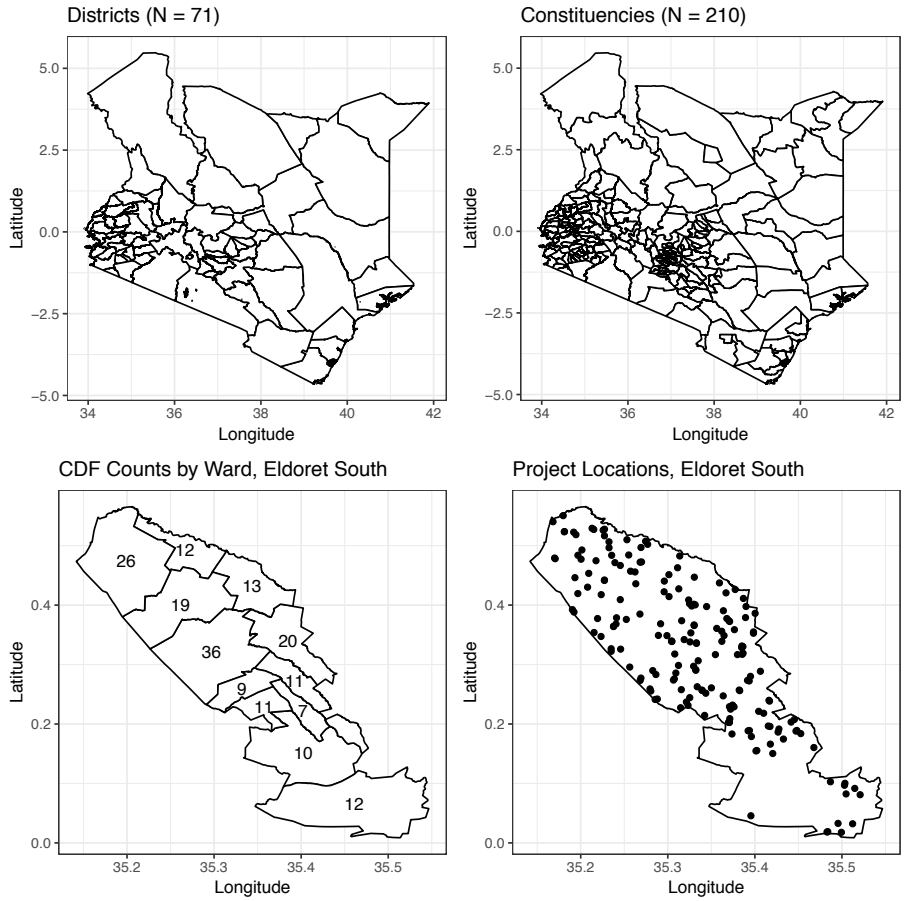

Figure D1: Comparison of Different Possible Levels of Aggregation of the Kenya Data. District- and constituency-level approaches are illustrated with respect to the country as a whole; ward- and point-level approaches are illustrated with respect to a single constituency, Eldoret South. District, constituency, and ward level analyses rely on aggregation to pre-existing polygons. Our approach, represented by the lower right panel, examines the placement of projects in continuous space.

An analysis at the level of the district, constituency, or ward would count the number of projects in each of these administrative units and estimate the association between that number and the other covariates of interest measured at the unit level. An analysis at the point-level ignores the

administrative boundaries altogether and estimates the association between the likelihood that a project is placed at each point and the characteristics of the population and other relevant factors (which may be related to the administrative unit) located at that point.

Figure 2 in the main text illustrates the loss of information that occurs when outcomes and explanatory variables are aggregated to the level of the ward. The situation is clarified further if one considers Figures D2 and D3, which provide maps of where CDF projects were actually placed superimposed over the distribution of political support in both Langas and Timboroa, the two wards highlighted in the text. As an inspection of the two maps makes clear, the spatial distribution and number of supporters in each ward varies widely. Langas is relatively highly populated and clustered, while Timboroa is much less densely populated and has lower levels of clustering. Aggregation to the ward level washes away all of that salient spatial information about where voters are, where they are not, and how their location relates to where projects are put. Aggregation changes the question that we would be answering, forcing us to focus on allocation to geographic units, rather than to local voters (which, as the maps illustrate, are rarely uniformly distributed within a given unit).

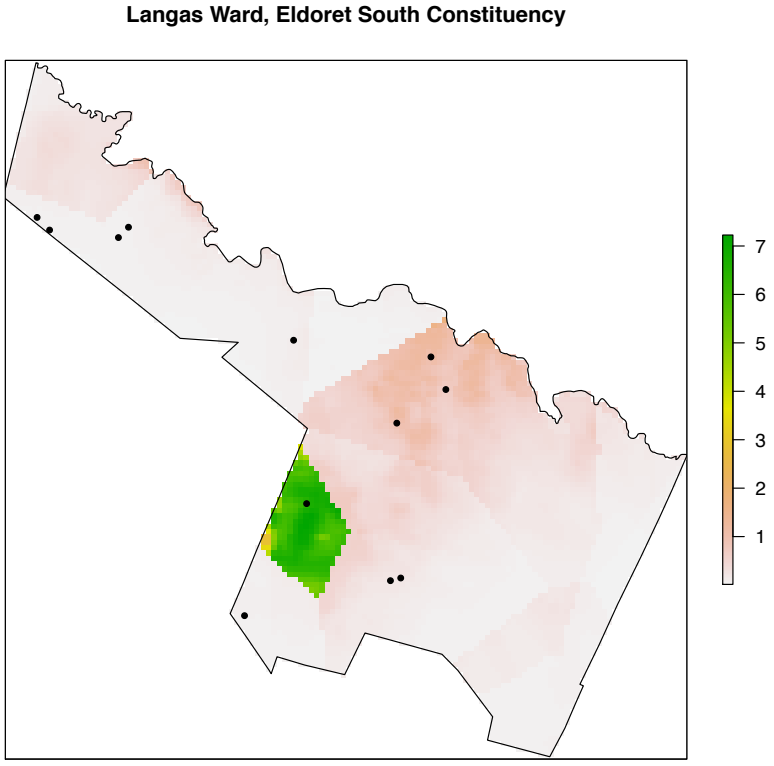

Figure D2: Spatial Distribution of Supporters within Langas Ward: The scale shows the estimated number of supporters in each grid cell in Langas Ward.

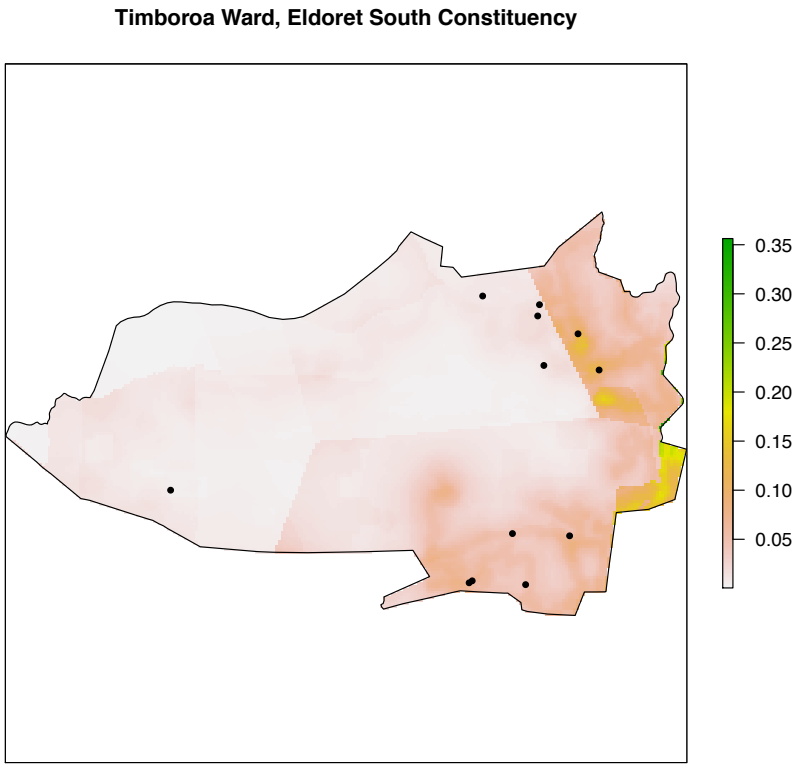

Figure D3: Spatial Distribution of Supporters within Timboroa Ward: The scale shows the estimated number of supporters in each grid cell in Timboroa Ward.

Notwithstanding our reasons for running our analysis at the point-level, we present a parallel version of Table 3, with project counts and covariates aggregated to the ward level, in Table D1. As expected, these results bear little resemblance to the higher-resolution results presented in the main paper precisely because the units capture completely different kinds of variation.

This exercise, and these arguments, have broader implications. Researchers who use aggregated data to study distributive politics usually do so because the available data is itself aggregated, not because aggregated data is optimal for studying the allocation decisions they are interested in understanding. This is also more generally true of social scientists who use existing GIS data in their analyses. Outside of the relatively narrow purview of studies of ecological inference in voting behavior, researchers rarely even raise the issue of whether the level of aggregation they employ in their analyses is appropriate for the question they are studying (Amrhein 1995; Fotheringham and Wong 1991). The materials presented in this appendix suggest that perhaps they should.

Table D1: Explaining the Relationship Between Project Placement and Residual Support: Data aggregated to ward level.

	Dependent variable:								
	Coefficient of Project Placement and Residual Support								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	6.91**	7.75**					7.67**	7.89**	7.72**
	(3.01)	(3.07)					(3.01)	(3.09)	(3.16)
Member of Ruling Coalition	−0.95	−1.43					−0.90	−1.15	−1.18
	(1.06)	(1.17)					(1.14)	(1.21)	(1.22)
Incumbent	1.11	1.05					1.53	1.34	1.35
	(1.07)	(1.09)					(1.07)	(1.10)	(1.11)
Margin of Victory (2002)			0.05	−0.68			−0.03	−0.42	−0.43
			(1.11)	(1.18)			(1.11)	(1.19)	(1.20)
Segregation of Supporters			3.56	3.87			4.46	4.11	4.91
			(3.02)	(3.52)			(3.18)	(3.57)	(4.59)
Ethnic Heterogeneity					−18.58	−16.57	−17.93	−18.04	−18.29
					(16.13)	(21.23)	(16.52)	(21.17)	(21.24)
Population Clustering					−1.14	−0.89	−1.55	−1.01	−1.01
					(0.97)	(1.41)	(1.01)	(1.43)	(1.43)
Margin of Victory × Segregation									−1.75
									(6.29)
Constant	0.29	0.45	0.34	1.13	0.34	0.05	0.05	0.38	0.36
	(0.89)	(1.67)	(0.76)	(1.50)	(0.52)	(1.42)	(1.02)	(1.74)	(1.75)
Province FE	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	196	196	196	196	196	196	196	196	196
R^2	0.03	0.07	0.01	0.04	0.02	0.03	0.07	0.09	0.09
Adjusted R^2	0.02	0.02	−0.002	−0.01	0.01	−0.01	0.04	0.02	0.01

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are estimated using OLS, with province fixed effects as noted. Data in this analysis are aggregated to the ward level.

E Model Specification

Our specification for the intra-constituency models is theory driven: we include covariates that in practical and political terms should determine where development projects are placed. However, we are acutely aware that model specification is a subjective process. In this appendix, we examine one important issue that may affect our model specification: collinearity. This examination leads us to consider dropping one variable (coethnicity) from some versions of our intra-constituency models due to its collinearity with the supporters variable. This, in turn, raises the possibility of omitted variable bias. Using model selection and model averaging techniques, we show that the fully specified model (which includes the coethnicity covariate) is usually preferred over the truncated model (which omits it), and that the substantive results still hold when information from the two models are combined.

Given the spatial nature of our data, many of the covariates are highly correlated. As discussed in the main text, our central approach to dealing with this is the residualization of the independent variables as a function of population distribution. However, a second source of collinearity affects the coethnicity and supporters variables because both are constructed using the same underlying data source: polling station-level election data. The coethnicity measure is constructed using information from the voter register; the supporters measure is constructed using actual electoral returns. As a result, these two variable retain some degree of collinearity. Figure E1 displays the distribution of constituency-level correlations between all of the pairs of residualized independent variables used in the intra-constituency regressions. While most pairs show relatively low levels of correlation in most constituencies, residualized coethnicity and support show a consistent positive correlation relative to other pairs.

We proceed in several steps to explore how this issue may affect our estimates. First, we rerun our main regressions, truncating our sample to those observations where the absolute value of the correlations between coethnicity and support are less than 80% and 60%, to see if the results change substantially for the subset of constituencies where collinearity is high. If the main results are driven by collinearity between support and coethnicity, then excluding those constituencies from analysis should significantly change the main results. The results for the truncation at 80% (with 127 observations remaining) is shown in Table E1 and 60% (with 51 observations remaining) in Table E2. In both cases, the substantive results mainly hold, though statistical significance is reduced due to the diminished sample size. One clear exception to this is the result for female MP's. In the 80% tables, the coefficient is statistically indistinguishable from zero; in the 60% table, the coefficient is dropped from the model as there are no constituencies with female MPs

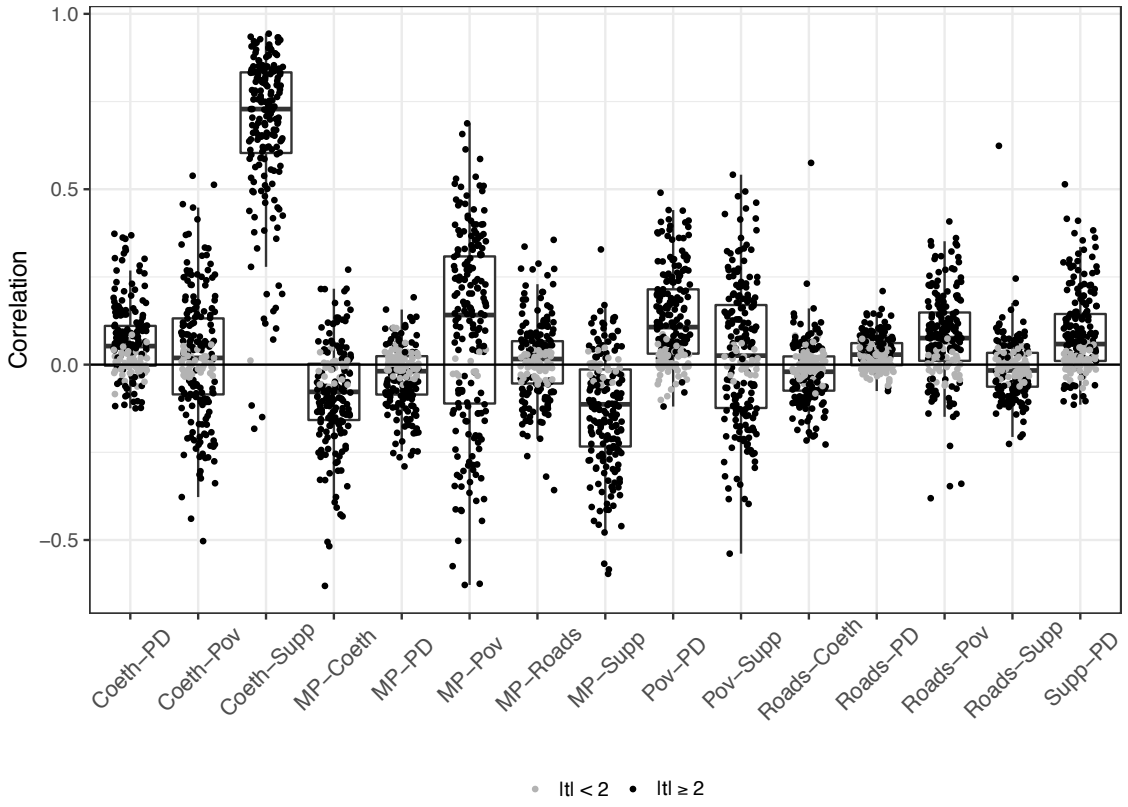

Figure E1: Distribution of Constituency-level Correlations between All Pairs of Independent Variables Used in the Intra-constituency Regressions

remaining in the sample.[36] Otherwise, these findings provide evidence of the robustness of the results.

Second, we can use the relative Aikike Information Criterion (AIC) of the two intra-constituency models under consideration ("full model" versus the "truncated model" without coethnicity) for each constituency to help select the better model, where lower AIC is better (Burnham and Anderson 2004). Figure E2 shows the density of $AIC_{full} - AIC_{trunc}$. The density shows that, in 105 of 196 constituencies, the full model is strictly a better fit. In the remaining 91 cases, the truncated model is a better fit. Interestingly, there is significant asymmetry in the difference between the AICs. When the full model is a better fit, it tends to be better by a larger margin than when the truncated model is a better fit. This can be seen by the thick left tail of the distribution in Figure E2. When the truncated model is better, it tends to be by a small amount, as can be seen in the right tail. Simply put, if we had to choose one specification for all of the models, then the full specification would be better.

We do not, however, have to choose one model to fit all constituencies. Next, we try two different ways of combining information from the different specifications and show that our results

[36] Recall that the initial result was predicated on very few women in the entire sample.

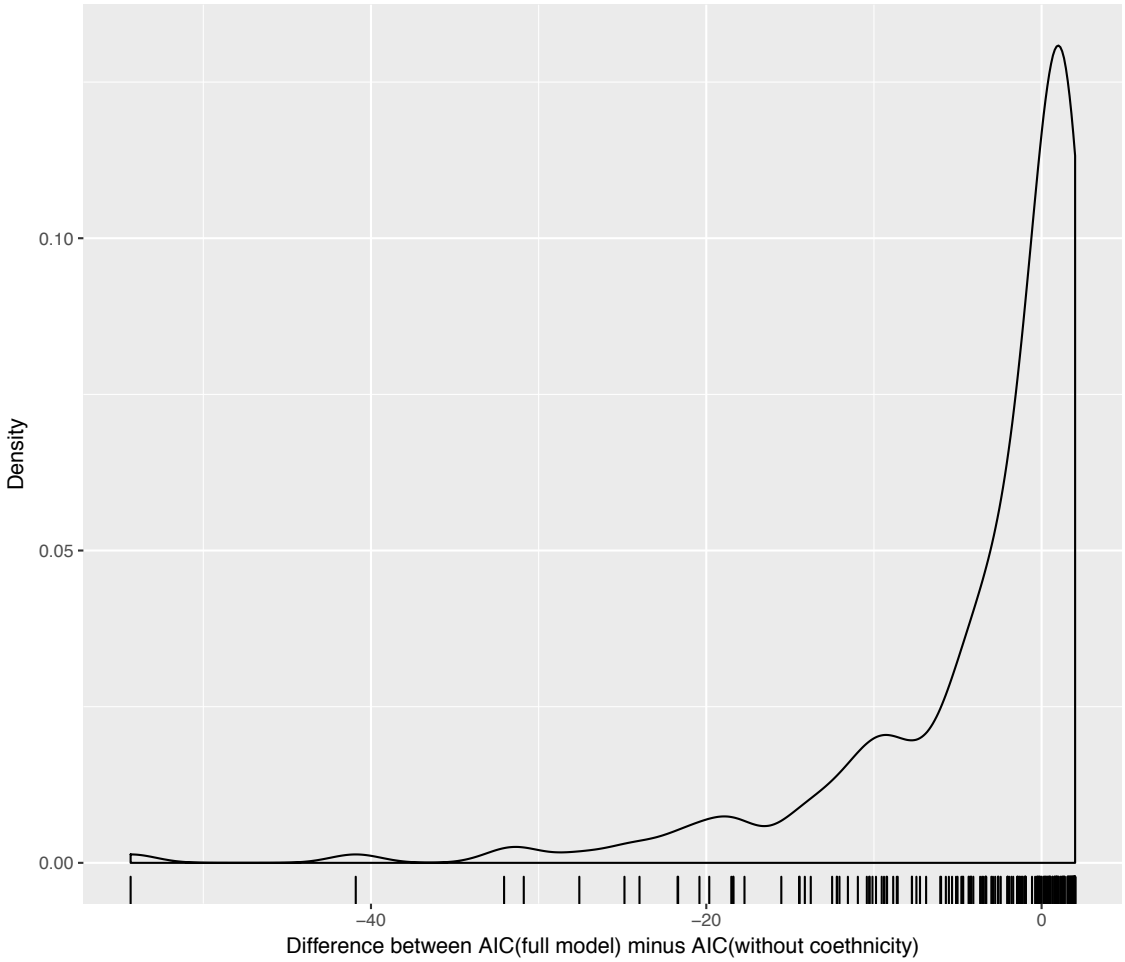

Figure E2: Density of Difference between $AIC_{full} - AIC_{trunc}$. Values less than zero suggest that the full model fits better than the truncated model. The density suggests that when the full model fits better than the truncated model, it does so by a large margin. When the full model fits better than the truncated model, it does so by a relatively small margin.

are largely preserved (in substance, if not significance). The first approach generates weights from each model's AIC, following Burnham and Anderson (2004). Using those weights, the coefficients and standard errors from the two models are combined and then used in the main cross-constituency results. In this way, information from the two models are directly combined for each constituency. The results for the model using coefficients weighted to contain information from both models is in Table E3. The substance of our results is unchanged.

Another approach is to use the AIC as a strict decision metric in deciding which model is more appropriate. In table E4, we replicate the main results, this time replacing the outcome coefficient with the one dictated by the lowest AIC. That is, for the 91 cases where the truncated model is a better fit, we replace the outcome with the coefficient from that model. Again, we find the substance of our results retained. Taken together, these efforts demonstrate that our results are robust to issues related to model fit and choice.

Table E1: Explaining the Relationship Between Project Placement and Residual Support: Corr. less than 0.8.

	Dependent variable:								
	Coefficient of Project Placement and Residual Support								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	0.09	0.07					0.06	0.03	0.03
	(0.10)	(0.10)					(0.10)	(0.10)	(0.10)
Member of Ruling Coalition	−0.02	−0.05*					−0.03	−0.05*	−0.05*
	(0.02)	(0.03)					(0.03)	(0.03)	(0.03)
Incumbent	0.01	0.01					0.02	0.03	0.03
	(0.03)	(0.03)					(0.03)	(0.03)	(0.03)
Margin of Victory (2002)			0.02	0.002			0.01	0.01	0.01
			(0.03)	(0.03)			(0.03)	(0.03)	(0.03)
Segregation of Supporters			0.09	0.19**			0.11	0.24**	0.26**
			(0.08)	(0.09)			(0.09)	(0.10)	(0.11)
Ethnic Heterogeneity					−0.08	−0.28	0.07	−0.06	−0.07
					(0.37)	(0.49)	(0.40)	(0.49)	(0.49)
Population Clustering					−0.03	−0.02	−0.05*	−0.07*	−0.06*
					(0.02)	(0.03)	(0.03)	(0.04)	(0.04)
Margin of Victory × Segregation									−0.07
									(0.17)
Constant	0.02	0.04	0.01	−0.001	0.02	−0.01	0.01	0.01	0.01
	(0.02)	(0.05)	(0.02)	(0.04)	(0.01)	(0.05)	(0.03)	(0.05)	(0.05)
Province FE	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	127	127	127	127	127	127	127	127	127
R^2	0.02	0.11	0.01	0.12	0.02	0.09	0.06	0.17	0.17
Adjusted R^2	−0.01	0.04	−0.01	0.05	0.01	0.02	0.001	0.07	0.06

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are estimated using OLS, with province fixed effects as noted. This analysis excludes constituencies in which the correlation between coethnicity and supporters is less than 0.8.

Table E2: Explaining the Relationship Between Project Placement and Residual Support: Corr. less than 0.6.

	Dependent variable:								
	Coefficient of Project Placement and Residual Support								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female									
Member of Ruling Coalition	-0.09^{***}	-0.09^{**}					-0.08^{**}	-0.08^{**}	-0.08^{**}
	(0.03)	(0.04)					(0.03)	(0.04)	(0.04)
Incumbent	0.02	0.02					0.05	0.05	0.05
	(0.04)	(0.04)					(0.03)	(0.04)	(0.04)
Margin of Victory (2002)			0.04	0.02			0.03	0.04	0.04
			(0.04)	(0.04)			(0.04)	(0.04)	(0.05)
Segregation of Supporters			0.29^{***}	0.26^{**}			0.25^{**}	0.31^{**}	0.32^{*}
			(0.10)	(0.13)			(0.11)	(0.14)	(0.16)
Ethnic Heterogeneity					-0.55	-0.62	0.12	-0.23	-0.22
					(0.51)	(0.68)	(0.52)	(0.66)	(0.67)
Population Clustering					-0.03	0.004	-0.07^{*}	-0.08	-0.08
					(0.04)	(0.05)	(0.03)	(0.05)	(0.06)
Margin of Victory × Segregation									-0.03
									(0.22)
Constant	0.06^{**}	0.04	-0.01	-0.06	0.04^{**}	-0.03	0.02	-0.07	-0.07
	(0.03)	(0.07)	(0.03)	(0.07)	(0.02)	(0.08)	(0.04)	(0.09)	(0.09)
Province FE	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	51	51	51	51	51	51	51	51	51
R^2	0.14	0.23	0.14	0.20	0.06	0.13	0.31	0.34	0.34
Adjusted R^2	0.10	0.10	0.11	0.06	0.03	-0.02	0.21	0.15	0.13

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are estimated using OLS, with province fixed effects as noted. This analysis excludes constituencies in which the correlation between coethnicity and supporters is less than 0.6.

Table E3: Explaining the Relationship Between Project Placement and Residual Support

	Dependent variable:								
	Coefficient of Project Placement and Residual Support								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.25^{***}	-0.24^{***}					-0.24^{***}	-0.25^{***}	-0.28^{***}
	(0.08)	(0.08)					(0.08)	(0.08)	(0.08)
Member of Ruling Coalition	-0.03	-0.05^{*}					-0.05	-0.06^{*}	-0.06^{*}
	(0.03)	(0.03)					(0.03)	(0.03)	(0.03)
Incumbent	0.01	0.001					0.01	0.01	0.01
	(0.03)	(0.03)					(0.03)	(0.03)	(0.03)
Margin of Victory (2002)			0.02	0.01			0.03	0.03	0.03
			(0.03)	(0.03)			(0.03)	(0.03)	(0.03)
Segregation of Supporters			0.11	0.12			0.09	0.13	0.29^{**}
			(0.08)	(0.09)			(0.08)	(0.09)	(0.11)
Ethnic Heterogeneity					-0.33	-0.28	-0.04	-0.10	-0.15
					(0.41)	(0.55)	(0.42)	(0.54)	(0.53)
Population Clustering					-0.05^{*}	-0.04	-0.07^{***}	-0.07^{*}	-0.07^{*}
					(0.02)	(0.04)	(0.03)	(0.04)	(0.04)
Margin of Victory × Segregation									-0.36^{**}
									(0.16)
Constant	0.04^{*}	0.06	0.01	0.01	0.02	-0.004	0.04	0.05	0.04
	(0.02)	(0.04)	(0.02)	(0.04)	(0.01)	(0.04)	(0.03)	(0.04)	(0.04)
Province FE	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	196	196	196	196	196	196	196	196	196
R^2	0.06	0.09	0.01	0.04	0.03	0.04	0.11	0.12	0.14
Adjusted R^2	0.05	0.04	0.0001	-0.01	0.02	-0.01	0.07	0.05	0.07

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are estimated using OLS, with province fixed effects as noted. The outcome in this analysis is built by combining the full model and the truncated model using AIC to generate weighted averages of the coefficients from the two models.

Table E4: Explaining the Relationship Between Project Placement and Residual Support

	Dependent variable:								
	Coefficient of Project Placement and Residual Support								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.27^{***}	-0.27^{***}					-0.26^{***}	-0.28^{***}	-0.31^{***}
	(0.08)	(0.08)					(0.08)	(0.08)	(0.08)
Member of Ruling Coalition	-0.03	-0.05^{*}					-0.04	-0.05^{*}	-0.06^{*}
	(0.03)	(0.03)					(0.03)	(0.03)	(0.03)
Incumbent	0.01	0.01					0.02	0.02	0.02
	(0.03)	(0.03)					(0.03)	(0.03)	(0.03)
Margin of Victory (2002)			0.03	0.01			0.03	0.04	0.03
			(0.03)	(0.03)			(0.03)	(0.03)	(0.03)
Segregation of Supporters			0.14^{*}	0.16^{*}			0.12	0.17^{*}	0.33^{***}
			(0.08)	(0.09)			(0.08)	(0.09)	(0.12)
Ethnic Heterogeneity					-0.32	-0.26	0.02	-0.05	-0.10
					(0.42)	(0.56)	(0.43)	(0.55)	(0.54)
Population Clustering					-0.04	-0.04	-0.06^{**}	-0.06^{*}	-0.06^{*}
					(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
Margin of Victory × Segregation									-0.34^{**}
									(0.16)
Constant	0.04	0.05	0.002	0.01	0.01	-0.01	0.02	0.04	0.04
	(0.02)	(0.04)	(0.02)	(0.04)	(0.01)	(0.04)	(0.03)	(0.04)	(0.04)
Province FE	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	196	196	196	196	196	196	196	196	196
R^2	0.07	0.10	0.02	0.04	0.02	0.03	0.11	0.13	0.15
Adjusted R^2	0.06	0.05	0.01	-0.01	0.01	-0.02	0.08	0.06	0.08

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are estimated using OLS, with province fixed effects as noted. The outcome in this analysis uses AIC to replace full model coefficients with the truncated model coefficient when the AIC suggests that the truncated model is better.

F Robustness to Periodization

A drawback of the 2003-2007 period we study is that a constitutional referendum in 2005 led to a political crisis that precipitated significant realignments in the Kenyan party system. After 2005, some MPs who had previously been in the governing coalition (and, in some cases, in the Cabinet) shifted to the opposition, which may have changed their access to non-CDF government resources and, potentially, the allocative strategies they pursued with respect to CDF resources. If this is true, then our results may be driven by these shifts rather than representing stable patterns across time.

To address this concern, we break our CDF project observations up into two periods, 2003-2005 and 2006-2007. We re-estimate the constituency-level associations between project placement and residual support for each period and then use the estimated coefficients in each time period to re-run the main results reported in Table 3. Table F1 presents the fully saturated model results for both periods. Although there are some differences in the statistical significance of the results across the two periods (in large part because we are estimating the constituency-level associations with fewer projects, and thus with greater uncertainty), the signs and general magnitudes of the coefficients are broadly similar.

Table F1: Explaining the Relationship Between Project Placement and Residual Support: Projects started from 2003 - 2005 (column 1) and 2006 - 2007 (column 2)

	Dependent variable:	
	2003 - 2005	2006 - 2007
	(1)	(2)
Female	-0.28^{***}	-0.31
	(0.09)	(0.40)
Member of Ruling Coalition	-0.09^{***}	-0.26^{*}
	(0.03)	(0.16)
Incumbent	0.02	0.21
	(0.03)	(0.14)
Margin of Victory (2002)	0.02	-0.08
	(0.03)	(0.15)
Segregation of Supporters	0.13	1.24^{**}
	(0.13)	(0.58)
Ethnic Heterogeneity	0.10	-1.60
	(0.58)	(2.70)
Population Clustering	-0.08^{**}	-0.28
	(0.04)	(0.18)
Margin of Victory × Segregation	-0.32^{*}	-1.31
	(0.17)	(0.80)
Constant	0.05	0.11
	(0.05)	(0.22)
Province FE	Yes	Yes
Observations	196	196
R^2	0.12	0.10
Adjusted R^2	0.05	0.02
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Models are estimated using OLS, with province fixed effects as noted. This analysis splits the project sample into two periods (2003-2005 and 2006-2007) to see if the same relationships hold across periods.	

Appendix G: Additional Robustness Checks

Tables G1 and G2 present the main results, adding robust standard errors and WLS estimation to demonstrate robustness. The main results on segregation of supporters and population clustering remain substantively similar to the original results. The result on the interaction between margin of victory and segregation retains the same sign in both sets of results, though statistical significance diminishes.

The analyses presented in the paper are limited to 196 of Kenya's 210 constituencies because of missing data on one or more covariates of interest. Table G3 presents the results from a saturated model based on 50 imputations of the constituency-level dataset, which allows us to incorporate information from all 210 constituencies, as well as to impute uncertainty via multiple imputation. The imputation and combination of results was carried out in `Zelig` (Choirat et al. 2017; Imai et al. 2008). As the results show, the substantive and statistical significance of the results findings reported in the paper are unchanged.

Table G4 contains the regressions mentioned in footnote 33 of the paper. These regressions only contain observations where segregation was above the sample mean. Even after dropping 93 observations, our results remain substantively similar, though, with only two women in the truncated dataset, we lose significance on that coefficient (though the sign remains substantively similar). Similarly, the coefficient on the interaction is not significant, though the sign remains negative. The coefficient on segregation, meanwhile, becomes larger, as expected.

Table G1: Explaining the Relationship Between Project Placement and Residual Support: Robust (HC3) standard errors

	Dependent variable:								
	Coefficient of Project Placement and Residual Support								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	−0.16	−0.15					−0.16	−0.17	−0.19
	(0.15)	(0.16)					(0.14)	(0.15)	(0.16)
Member of Ruling Coalition	−0.05**	−0.08***					−0.07***	−0.08***	−0.08***
	(0.02)	(0.02)					(0.02)	(0.03)	(0.03)
Incumbent	0.02	0.01					0.02	0.02	0.02
	(0.02)	(0.02)					(0.02)	(0.02)	(0.02)
Margin of Victory (2002)			0.03	0.02			0.04	0.04	0.04
			(0.03)	(0.03)			(0.03)	(0.03)	(0.03)
Segregation of Supporters			0.14*	0.16*			0.12*	0.16*	0.28***
			(0.08)	(0.08)			(0.07)	(0.08)	(0.10)
Ethnic Heterogeneity					−0.22	0.08	0.06	0.20	0.17
					(0.37)	(0.48)	(0.33)	(0.48)	(0.48)
Population Clustering					−0.03*	−0.03	−0.06**	−0.06*	−0.06*
					(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Margin of Victory × Segregation									−0.26
									(0.17)
Constant	0.04**	0.05	−0.01	−0.002	0.01	−0.01	0.02	0.04	0.04
	(0.02)	(0.04)	(0.02)	(0.03)	(0.01)	(0.03)	(0.02)	(0.04)	(0.04)
Province FE	No	Yes	No	Yes	No	Yes	No	Yes	Yes

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are estimated using OLS with HC3 standard errors, with province fixed effects as noted. Outcome variable is the estimated coefficient of the supporters variable from the constituency-level point-process model examining the determinants of project placement.

Table G2: Explaining the Relationship Between Project Placement and Residual Support: Weighted Least Squares

	Dependent variable:								
	Coefficient of Project Placement and Residual Support								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	−0.08	−0.08					−0.08	−0.10	−0.12
	(0.07)	(0.08)					(0.07)	(0.07)	(0.08)
Member of Ruling Coalition	−0.05**	−0.06***					−0.05**	−0.06***	−0.07***
	(0.02)	(0.02)					(0.02)	(0.02)	(0.02)
Incumbent	0.02	0.01					0.02	0.03	0.02
	(0.02)	(0.02)					(0.02)	(0.02)	(0.02)
Margin of Victory (2002)			0.01	0.01			0.02	0.03	0.04*
			(0.02)	(0.02)			(0.02)	(0.02)	(0.02)
Segregation of Supporters			0.11*	0.15**			0.13**	0.18**	0.25***
			(0.06)	(0.07)			(0.06)	(0.07)	(0.08)
Ethnic Heterogeneity					−0.07	−0.004	0.07	−0.01	0.04
					(0.29)	(0.38)	(0.29)	(0.37)	(0.37)
Population Clustering					−0.01	−0.02	−0.04**	−0.05*	−0.05*
					(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Margin of Victory × Segregation									−0.22*
									(0.13)
Constant	0.03**	0.05	0.003	0.01	0.02*	0.01	0.02	0.04	0.04
	(0.01)	(0.04)	(0.01)	(0.03)	(0.01)	(0.03)	(0.02)	(0.04)	(0.04)
Province FE	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	196	196	196	196	196	196	196	196	196
R^2	0.04	0.05	0.02	0.03	0.01	0.01	0.07	0.10	0.11
Adjusted R^2	0.02	0.002	0.01	−0.02	−0.01	−0.04	0.04	0.03	0.04

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are estimated using weighted least squares, with province fixed effects as noted. Outcome variable is the estimated coefficient of the supporters variable from the constituency-level point-process model examining the determinants of project placement.

Variable	Coefficient	Standard Error	t-statistic	p-value
Female	-0.19	0.07	-2.77	0.01
Member of Ruling Coalition	-0.08	0.03	-3.03	0.00
Incumbent	0.02	0.02	0.90	0.37
Margin of Victory (2002)	0.04	0.03	1.60	0.11
Segregation of Supporters	0.28	0.10	2.79	0.01
Ethnic Heterogeneity	0.18	0.46	0.40	0.69
Population Clustering	-0.06	0.03	-1.85	0.07
Margin of Victory $\times$ Segregation	-0.25	0.14	-1.84	0.07

Table G3: Combined results from 50 imputed datasets, accounting for missing constituency data.

Table G4: Explaining the Relationship Between Project Placement and Residual Support: High Segregation Only

	Dependent variable:								
	Coefficient of Project Placement and Residual Support								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	−0.16	−0.14					−0.16	−0.18	−0.18
	(0.11)	(0.11)					(0.11)	(0.11)	(0.11)
Member of Ruling Coalition	−0.08***	−0.10***					−0.10***	−0.09***	−0.09***
	(0.03)	(0.03)					(0.03)	(0.03)	(0.03)
Incumbent	0.02	0.02					0.04	0.05	0.05
	(0.03)	(0.03)					(0.03)	(0.03)	(0.03)
Margin of Victory (2002)			0.01	0.004			0.01	0.03	0.03
			(0.03)	(0.04)			(0.03)	(0.03)	(0.06)
Segregation of Supporters			0.36**	0.35**			0.42**	0.41**	0.41**
			(0.17)	(0.17)			(0.16)	(0.17)	(0.18)
Ethnic Heterogeneity					−0.17	−0.01	−0.13	−0.10	−0.10
					(0.48)	(0.61)	(0.45)	(0.57)	(0.57)
Population Clustering					−0.04	−0.08	−0.07**	−0.11**	−0.11**
					(0.03)	(0.05)	(0.03)	(0.05)	(0.05)
Margin of Victory × Segregation									−0.01
									(0.45)
Constant	0.05**	0.04	−0.04	−0.06	0.02	−0.04	−0.004	−0.05	−0.05
	(0.02)	(0.06)	(0.03)	(0.06)	(0.02)	(0.06)	(0.04)	(0.07)	(0.07)
Province FE	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	104	104	104	104	104	104	104	104	104
R^2	0.10	0.13	0.04	0.06	0.02	0.05	0.20	0.22	0.22
Adjusted R^2	0.07	0.04	0.03	−0.02	0.003	−0.03	0.14	0.11	0.10

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are estimated using OLS, with province fixed effects as noted. This analysis includes only constituencies with above average segregation of supporters.