Modeling Related Processes with an Excess of Zeros*

David Carlson

Washington University in St. Louis

carlson.david@wustl.edu

ABSTRACT

Political science research frequently models binary or ordered outcomes involving related processes, such as strategic interactions. However, traditional modeling of these outcomes ignores common data issues and cannot capture nuances. There is often an excess of zeros, the observed outcomes for different actors are inherently related, and competing actors may respond to the same factors differently. This paper develops a new model that addresses these issues simultaneously: a zero-inflated multivariate ordered probit. This model performs better than existing models at capturing the true parameters of interest, estimates the nature of the related processes, and captures the differences in actors' decision-making. I demonstrate these benefits through simulation exercises and two novel applications.

*The author would like to sincerely thank Siddhartha Chib, Jeff Gill, Jacob Montgomery, Guillermo Rosas, Margit Tavits, Michelle Torres, every member of the Comparative Politics Workshop, and every member of the Data Science Lab for their invaluable comments.

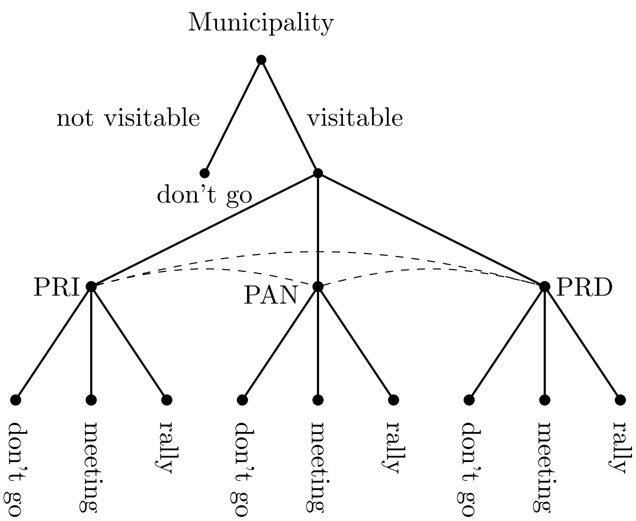
1. INTRODUCTION

Many questions in the literature concern binary or ordered outcomes. However, researchers regularly face two distinct problems: (1) there is often an excess of zeros in the outcome variable, and (2) outcomes may be related. For instance, consider the decision-making processes of competing parties regarding candidate visits during a presidential campaign. These campaigns can only visit a small proportion of localities within a country. The outcome therefore exhibits an excess of zeros. Which municipalities candidates choose to visit on their campaigns is a strategic choice, and municipal-level factors make certain municipalities never worth consideration by any party. When deciding among those municipalities that are considered, competitiveness and latent party support are important determinants in the parties' calculi. These decisions are also highly interdependent, however, with parties targeting municipalities that are also targeted by their rivals. Further, the manner by which parties react to different factors likely varies across parties.

In the first analysis of the paper, I test these claims on Mexican presidential campaigns in 2006 and 2012 for the three major parties – the PRI, the PAN, and the PRD (Langston and Rosas 2016). The outcome of interest is the level of visitation by each of the parties – no visit, hold a meeting, or hold a rally. Because there are three parties, the outcome is trivariate. In other words, each municipality has three outcomes, one for each party, that are inherently related. The vast majority of the municipalities were never visited by any party. Figure 1 depicts the parties' decisions. No current model can accurately capture the decision tree described. Zero-inflated models cannot measure the extent to which the parties are strategically interacting. Models allowing correlations would not capture the first stage. Both would lead to misleading and biased conclusions.

Ignoring these issues is problematic for accurate estimation. Furthermore, the added benefit of addressing them is that we are able to draw interesting substantive conclusions that current methods do not allow. We need to account for and explain the excess of zero observations. Additionally, we may want to model actors' decisions separately, allowing

Figure 1: Parties' decision trees.



Note: The decision of a party is split into two stages. The first stage determines whether or not a municipality is visitable. The second stage determines the type of visit, conditional on the municipality being visitable. This stage is likely related to the decisions of other parties.

different actors to respond to factors differently. Similarly, we may also want to account for the fact that the behavior of one actor is related to, or even shaped by, the (anticipated) behavior of other actors.[1]

I propose a novel model that can address these issues: a zero-inflated multivariate ordered probit (ZIMVOP). It consists of two stages. The first stage models the observation as a potential non-zero, splitting the population into "always zeros" and "potential non-zeros." The second stage is a multivariate ordered probit, allowing correlations of the disturbance terms across equations over dimensions.

While strategic interaction is an intuitive application of ZIMVOP, the applicability is much wider. For example, it allows us to control for unmodeled factors that may be driving related processes. The second application focuses on exactly this. Specifically, I test the hypothesis in the international relations literature that terrorism "spoils" interstate cooperation and negotiations, and this is precisely the goal of some terrorist activity (Kydd and

[1]Because these decisions are being made at an unobserved time, or simultaneously, standard strategic interaction models are inappropriate (Bas et al. 2008; Carson and Roberts 2005; Signorino 2002; Signorino 2003).

Walter 2002). However, it is unclear in the literature if attacks harm positive relations (e.g., negotiations), or result in condemnation (e.g., trade barriers). I demonstrate that terrorism affects both, by decreasing the former and increasing the latter. However, these processes are both driven by very similar observed and unobserved factors. Allowing the processes to be correlated accounts for some of the unobserved variability.

The outline of the paper is as follows. First, I discuss the issues of zero-inflated and correlated outcomes, and explain how ZIMVOP addresses both. As part of the discussion, I explain how it differs from previous models and also builds on them. Second, I specify the ZIMVOP model. Third, I demonstrate its effectiveness using simulated data and illustrate the usefulness through the two substantive applications discussed above. Finally, I conclude with a discussion of potential future applications and directions.

2. ZERO-INFLATED AND CORRELATED ERRORS: ISSUES AND SOLUTIONS

In this section, I discuss the issues associated with outcomes that exhibit an excess of zeros and current approaches to dealing with these issues. I then do the same for multivariate models that have correlated error terms, known as seemingly unrelated regressions (SUR). Neither of these families of models adequately addresses the problems of both zero-inflated outcomes and correlated error terms. Through the discussion, I highlight the advantages to current approaches and demonstrate that ZIMVOP, a synthesis of the two families of models, is an intuitive extension when dealing with data that raise both of these concerns.

2.1. *Models with Zero-Inflation*

King and Zeng (2001a and 2001b) introduce a unique approach to modeling rare outcomes, focusing primarily on international conflict. They argue that modeling conflict on all country dyads underestimates the effect of certain factors, producing biased estimates. This is due to the fact that the vast majority of dyads will never go to war, regardless of certain observed characteristics that may actually be deterministic in other dyads. The approach they suggest is to save data collection, maintain all non-zero observations in the data, randomly sample zero outcomes, and focus more time on the quality of the data than the quantity of data.

This recommendation saves data collection and may lead to less biased estimates. However, certain covariates may have different effects in a split-population approach (Harris and Zhao 2007). This split-population method models the outcome in two stages. The first stage models the likelihood that an observation is a potential non-zero, and the second stage models the outcome conditional on the observation being a potential non-zero. The split-population refers to splitting the population into "potential non-zeros" and "always zeros." An intuitive example relates to civil conflict. Bagozzi et al. (2015), using a zero-inflated ordered probit, find that a country's GDP has a reliable and negative effect on the potential for political violence, but on a potential non-zero, the effect is positive. That is, rich countries are less likely to experience political violence, but on a potential non-zero, income has a positive effect on the outcome, likely due to greater resources.

Similarly, Langston and Rosas (2016) model the likelihood of parties to visit municipalities in Mexican presidential campaigns with a zero-inflated ordered probit. They find that a party's previous vote share in a municipality is positively related to a municipality being visitable, a potential non-zero, but once visitable, the effect is negative. This could be due to the fact that beyond a certain point, high previous vote shares indicate a lack of necessity for parties to waste resources on a rally to gain votes in a municipality in which they already have a stronghold. It could also suggest that parties are reluctant to visit municipalities where they have very little support, but among those that are deemed visitable, they target the municipalities where they can gain the most ground. These examples highlight both the issues related to ignoring an excess of zeros and the benefits in addressing them. If the two stages were ignored, the nuanced effects of these covariates would be lost and the estimates of the effects would be biased, because a standard model with no inflation would lead to a correlation between the error terms and the explanatory variables (Bagozzi and Marchetti 2014; Dunne and Tian n.d.).

2.2. *Seemingly Unrelated Regressions*

In addition to the problems associated with an excess of zeros, outcomes also may share related data generating processes. The SUR class of models stacks regressions and allows the error terms across these stacked regressions to be correlated (Zellner 1962). Jointly estimating a set of equations improves asymptotic relative efficiency over the equation-by-equation case by combining information across equations (King 1989; Zellner and Huang 1962). In other words, in the limit, the estimators produce estimates with smaller mean squared errors and smaller variances. As an example, the processes generating observed multi-party vote shares are intuitively related. The errors of the predictions of one party's share are correlated with the errors of the predictions of other parties (Jackson 2002; Philips et al. 2015a and 2015b; Tomz et al. 2002; Tucker 2006). Similarly, positive political advertisement recall, negative ad recall, and turnout share a related data generating process (Ansolabehere et al. 1999). Failing to take the correlation into account would bias estimates and standard errors.

These correlations are also often substantively interesting, allowing inferences about the relationship between data generating processes. The process determining presidential vetoes of defense legislation is related to the process determining presidential vetoes of welfare legislation (King 1989). The error terms are positively correlated, indicating that if vetoes on welfare legislation is under (over) predicted, vetoes on defense legislation is also under (over) predicted. This implies that there is an underlying process shared by both that is not captured by the covariates, and, given covariates, a president who chooses to veto either type of legislation more (less) will also veto the other type more (less). This is an interesting finding that could not be uncovered with separate regressions for each type of veto. Further, shared covariates between the two types of vetoes are shown to affect the likelihood of the two types of vetoes in different ways. If the analysis were instead pooled this would be left uncovered. These examples demonstrate the usefulness of SUR models and suggest that many outcomes of interest may seem unrelated but in fact share some common data generating process.

2.3. *Building on Current Approaches*

When seemingly unrelated outcomes share a data generating process and exhibit an excess of zeros, combining zero-inflation and SUR will alleviate common problems analyses of these data often face. ZIMVOP is an effective tool for these types of data when the outcomes are ordered or binary. A zero-inflated bivariate ordered probit was developed by Gurmu and Dagne (2012), but it does not easily extend to the multivariate case (see also Kadel 2013).[2] Further, despite the development of the zero-inflated bivariate probit in the statistics literature, political science has yet to capitalize on the model. ZIMVOP generalizes the zero-inflated ordered probit to have a theoretically unbounded number of dimensions in an intuitive, straight-forward manner. It is an approachable method for political science questions that often face these problems in the outcome variable.

3. ZIMVOP SPECIFICATION

ZIMVOP has two major components that together set the model apart from current approaches. The first is a zero-inflation stage. This is simply a univariate standard probit stage that models the probability that an observation is a potential participator, or a potential non-zero. In the Mexico example, this would translate to whether or not any party will even consider visiting a given municipality.

The second component is a multivariate ordered probit stage. For each observation, there is a vector of outcomes, one for each dimension. In the Mexico example, the dimensions would be each party, one dimension for the PRI, one for the PAN, and one for the PRD. Each observation is a municipality in a given time period, and the outcome is a vector of length three, one outcome for each party, each component taking one of three values – 0

[2] The principles of ZIMVOP do not vary substantially from these bivariate models, but the implementation is much more straight-forward and these bivariate approaches only allow for one correlation parameter.

for no visit, 1 for a meeting, 2 for a rally. In this stage, conditional on being a potential non-zero, an ordered outcome is modeled separately for each dimension, but the error terms are allowed to correlate across dimensions (parties). Specifically, the linear models used to estimate the probit each have an error term. The error terms are not assumed independent, but instead the model allows these error terms to be correlated between dimensions at a given observation.[3] In other words, the decision processes at the second stage for each party are not assumed independent in a given municipality and time period. In combination, the zero-inflation stage and the correlated errors allow for more precise and efficient estimation of the second stage estimates, and we may have substantive interest both in the first stage and in the correlations.

3.1. *First Stage*

Let i denote an observation (e.g., a municipality in the Mexico example). Let s_i^* be a latent, unobserved variable capturing the likelihood of the observation being a potential non-zero. Both the first stage probit and the second stage multivariate ordered probit follow Albert and Chib's (1993) data augmentation approach. Sampling of latent variables leads to probability distributions for the observed outcomes. Potential non-zero status in the Mexico example would be whether or not the municipality is visitable. Being a potential non-zero is modeled with a matrix of covariates, Z, with each row, $\mathbf{z}_i$, a vector of observation-level covariates, including a constant. We let $s_i^* = \mathbf{z}_i'\gamma + \mu_i$, with γ an unknown vector to be estimated. Let μ_i be the random error.[4] Now, let s_i be defined as:

$$s_i = \begin{cases} 0 & \text{if } s_i^* \leq 0, \\ 1 & \text{if } s_i^* > 0. \end{cases}$$

[3]Harris and Zhao (2007) allow the errors from the first and second stage equations to be correlated. However, Gurmu and Dagne (2012) find that when moving from the zero-inflated univariate to the bivariate ordered probit allowing this correlation does not improve the model.

[4]I discuss all the prior distributions in Section 3.3.

Let $\Phi(\cdot)$ denote the normal CDF. Then,

$$\Pr(s_i = 1) = \Phi(s_i^*) = \Phi(\mathbf{z}_i'\gamma)$$

is the probability of the observation being a potential non-zero.

3.2. *Second Stage*

In the second stage, let $r = 1, \ldots, d$ denote the dimension, ranging from 1 to d, the total number of dimensions. If modeling a trivariate outcome as in the Mexico example, r would equal 1, 2, and 3, each number indicating a party. Let $\tilde{y}_{ri}^*$ be the unobserved, latent variable capturing the probability distribution of outcome level for observation i on dimension r conditional on observation i being a potential non-zero. Let X_r be a matrix of predictors for the level of participation on dimension r, with a constant. We let $\tilde{y}_{ri}^* = \mathbf{x}_{ri}'\beta_r + \epsilon_{ri}$, where ϵ_{ri} is the error term.

Let $\tilde{y}_{ri}$ be the level of participation for observation i on dimension r conditional on being a potential participant. Let j_r be the maximum possible outcome on dimension r. This second stage is an ordered probit, with cutoffs separating the levels of participation. Let a_{rk} be the cut-off points for levels $k = 1, \ldots, j_r - 1$. We can now define $\tilde{y}_{ri}$ as:

$$\tilde{y}_{ri} = \begin{cases} 0 & \text{if} \quad y_{ri}^* \leq a_{r1}, \\ k & \text{if} \quad a_{rk} < y_{ri}^* \leq a_{rk+1}, \quad k = 1, \ldots, j_r - 1, \\ j_r & \text{if} \quad y_i^* > a_{rj_r - 1}. \end{cases}$$

Finally, the observed vector of outcomes, $\mathbf{y}_i$, is equal to $s_i\tilde{\mathbf{y}}_i$. In the likelihood function that follows, allow i to index observations, p to index levels of observation on the first dimension, and q to index levels of observation on the second dimension. Let $m_{ijk} = 1$ if $\tilde{y}_{1i} = p$ and

$\tilde{y}_{2i} = q$, and $m_{ijk} = 0$ otherwise. The ensuing likelihood function for the bivariate case, which is easily generalized to higher dimensions, is:

$$\begin{aligned}\mathcal{L}(\mathbf{y},\mathbf{s}|X,Z,\beta,\gamma,\mathbf{a}) = &\prod_{i=1}^{N}\prod_{(p,q)=(0,0)}[\Pr(s_i=0)+(1-\Pr(s_i=0))\Pr(\tilde{y}_{1i}=0,\tilde{y}_{2i}=0)]^{m_{ipq}}\\ &\times\prod_{i=1}^{N}\prod_{(p,q)\neq(0,0)}[(1-\Pr(s_i=0))\Pr(\tilde{y}_{1i}=p,\tilde{y}_{2i}=q)]^{m_{ipq}}.\end{aligned}$$

To simplify the above, consider the probability of different outcomes. A zero outcome on three dimensions would be the probability that $s_i = 0$ added to the probability that $s_i = 1$ multiplied by the probability of all outcomes equaling zero: $\Pr(y_i = [0, 0, 0]) = \Pr(s_i = 0) + \Pr(s_i = 1) \times \Pr(\tilde{y}_{1i} = 0) \times \Pr(\tilde{y}_{2i} = 0) \times \Pr(\tilde{y}_{3i} = 0)$. An outcome of, for example, $[1, 0, 2]$, would be: $\Pr(s_i = 1) \times \Pr(\tilde{y}_{1i} = 1) \times \Pr(\tilde{y}_{2i} = 0) \times \Pr(\tilde{y}_{3i} = 2)$.

3.3. *Priors*

Let the first stage error terms, μ_i, follow the normal distribution with mean 0 and standard deviation 1: $\mu_i \sim \mathcal{N}(0, 1)$. In frequentist statistics, setting the standard deviation to 1 is necessary to identify the model (Cameron and Trivedi, 2005). Though in a Bayesian context we could put a hyperprior on the variance, I choose not to in order to make the model easier to interpret and to hasten convergence.

In the second stage we want to allow the error terms to be correlated. We therefore let the vector of error terms across dimensions, ϵ_i, be distributed multivariate normal with mean $\vec{0}_d$, a vector of zeros of length d, with variance-covariance matrix $\mathbf{\Sigma}_d$: $\epsilon_i \sim \mathcal{N}_d(\vec{0}_d, \mathbf{\Sigma}_d)$. The precision matrix $\mathbf{\Sigma}_d^{-1}$ is distributed inverse-Wishart with the $d \times d$ identity matrix as the mean and degrees of freedom ν: $\mathbf{\Sigma}_d^{-1} \sim \mathcal{IW}(\mathbf{I}_d, \nu)$.[5]

Because the variance is unconstrained in the specification, two cut-offs are set to identify the model.[6] Set a_{r1} to 0 and a_{r2} to some positive constant c_r. We can let all undefined a_{rk} follow a log-normal distribution with mean 0 and variance σ^2: $a_{rk>2} \sim \ln \mathcal{N}(0, \sigma^2)$. Note that no order is imposed on these cut-offs. Finally, we let our variables of interest, γ and β, have diffuse normal priors centered at zero. The model is written in `JAGS` and the code is provided in the Supplementary Information (SI) Section SI-1.

4. APPLYING ZIMVOP

This section first shows illustrative examples on simulated data to demonstrate the problems that can arise when researchers ignore the zero-inflation or the correlations in the underlying data-generating processes. I then apply the model to presidential campaigns in Mexico and I test the hypothesis that terrorism spoils interstate relations.

4.1. *Implementation on Simulated Data*

ZIMVOP synthesizes zero-inflation and SUR models. To isolate the gains of ZIMVOP in comparison to either models not accounting for zero-inflation or not accounting for correlated errors, I perform two sets of simulation exercises. The first set compares ZIMVOP to a multivariate ordered probit without zero-inflation, varies the degree of zero-inflation, and does not impose a correlation on the generated error terms. The second set compares ZIMVOP to an unpooled (i.e., separate equations for each dimension) zero-inflated probit, varies the correlation of the error terms, and does not vary the degree of zero-inflation.[7] By

[5] The inverse-Wishart is a conjugate prior for the multivariate normal distribution and it ensures generating positive-definite matrices. However, the inverse-Wishart has been criticized for the lack of independence between the variance and the correlations when sampling (Barnard et al. 2000). The best strategy to address this is to vary the degrees of freedom, ν, to ensure robustness of the results to different prior specifications. ν should always be equal to or greater than d to be uninformative. Note that the expected value of the precision matrix is a square matrix with diagonal elements equal to ν and off-diagonal elements equal to 0.

[6] Again, this is not strictly necessary, but aids in convergence and interpretability.

[7] All competing models are also run in `JAGS`.

performing these simulation exercises separately, as opposed to comparing all three models on the same sets of data, I can set up the data to make the competing model better able to capture the parameters of interest, allowing for a harder test of ZIMVOP.

Simulation Exercise I: Zero-Inflation The first set of simulations compare ZIMVOP to a model without the zero-inflation stage, a multivariate ordered probit (a SUR model). Data are generated through eight different zero-inflated processes. The generation of the data involves a zero-inflation stage with an intercept (γ_0) of -1.5 and a coefficient of interest (γ_1) changing from four to 11 by increments of one.[8] The first-stage equation is therefore;

$$
\begin{aligned}
&s_i^* = -1.5 + \mathbf{z}_{i1} \times \gamma_1 + \mu_i, \\
&\mu_i \sim \mathcal{N}(0,1), \quad \text{and} \\
&s_i = \begin{cases} 0 & \text{if } s_i^* \leq 0, \\ 1 & \text{otherwise.} \end{cases}
\end{aligned}
$$

The first stage values of predictors, Z, are all generated randomly and are not nested in the second-stage variables, which are independently generated. This is a harder test than nesting the values, because some of the variation of the zero-inflation stage should be accounted for in the second-stage intercept estimates, and much of it could be accounted for by the modeled correlation. In other words, if the zero-inflation stage is unmodeled, the second-stage estimates can in theory predict reasonable non-zero outcomes, and use the correlation and variance of the error terms to explain the excess zeros not following the pattern of the second stage.

The second-stage consists of three levels of outcome on three dimensions. The intercept term on each dimension is set to 0, and the three dimensions each have one predictor, set to 2, 2.5, and 2. The second-stage equation is therefore;

[8] SI-2 contains tables of the true parameters for both sets of simulations.

$$\tilde{y}_i^* = X_{i1} \begin{pmatrix} 2 \\ 2.5 \\ 2 \end{pmatrix} + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(\vec{0}_3, \mathbf{I}_3).$$

To demonstrate only the issues arising from not modeling the inflation, all correlations are set to zero, meaning the random errors used to generate the outcome are completely independent. This further allows the first stage to be captured by the correlation estimates, acting as an observation-level random effect. These simulations are repeated five times for a total of forty sets of data and eighty analyses, forty per model. Each unique data set is analyzed twice, once by both models, to ensure comparability.[9]

Despite the difficulty of the test, Table 1 shows that across specifications, the model accounting for the zero inflation performs better.[10] The root mean squared error (RMSE),[11] a measure of bias, of the second-stage estimates are smaller, while the standard deviations, a measure of precision and efficiency, of the posteriors are smaller. Further, the estimation of the correlation is much closer to the true values when modeling the zero-inflation. If we have a substantive interest in the correlations, we will get very biased results if we do not account for zero-inflation. Finally, across all simulations and specifications, the coverage probability in the model with a first stage is 95.2%, while the model without the first stage is 91.1%.[12] This suggests that the decrease in standard errors is not leading to overly restrictive posteriors.

Simulation Exercise II: Correlated Error Terms The second set of simulation exercises compare a zero-inflated multivariate ordered probit model with second-stage correlated errors

[9]The first of every simulation set-up, for both the first and second set, were run for 10,000 iterations and two chains. All $\hat{R}$'s were close to one and lack-of-convergence tests with the package `superdiag` indicated no problems. The remaining were run for 20,000 iterations to make convergence likely without having to test for convergence on all models.

[10]No patterns emerged as the true coefficient changed, so results are pooled. The graphs showing the statistics at different data generating processes are shown in Figure SI-1 in SI-2.

[11]RMSE is calculated by squaring the difference between the estimates and the true values and taking the mean.

[12]Coverage probability is the proportion of posterior distributions in which the true value falls within the 95% highest density region. Ideally this value would be 0.95.

Table 1: Comparing ZIMVOP to multivariate ordered probit (MVOP) and zero-inflated ordered probit (ZIOP)

	First set		Second set	
Test Statistic	ZIMVOP	MVOP	ZIMVOP	ZIOP
Coverage probability second stage	.952	.911	.963	.962
RMSE second stage	.059	.160	.075	.115
Standard deviation second stage	.340	.401	.309	.363
RMSE correlations	.008	.456	NA	NA

Note: Across simulations, ZIMVOP produces second-stage estimates with a smaller root mean squared error (RMSE) and tighter posteriors. The coverage probability is very close to .95, indicating that the increase in precision is not leading to poor coverage. The last row shows the RMSE of the correlation estimates. ZIMVOP significantly outperforms MVOP by this metric, indicating that failing to include the zero-inflation stage biases correlation estimates. The RMSE of correlations is suppressed for the second set, because the alternative model, ZIOP, does not produce these estimates and RMSE is a relative measure.

to the same model not allowing correlations. The simulations again generate data with a zero-inflation process, but vary the correlations of the second-stage error terms. The true data generating process sets the first-stage intercept, γ_0, to -1.5. The coefficient of interest in the first stage, γ_1, is set to nine. The first-stage equation is therefore;

$$
\begin{aligned}
s_i^* &= -1.5 + 9 \times \mathbf{z}_{i1} + \mu_i, \\
\mu_i &\sim \mathcal{N}(0,1), \quad \text{and} \\
s_i &= \begin{cases} 0 & \text{if } s_i^* \leq 0, \\ 1 & \text{otherwise.} \end{cases}
\end{aligned}
$$

The values of the predictors are again generated independently of the second stage values. They are not nested.

The second stage has three outcome levels on each of three dimensions. The coefficients used are the same as the earlier round of simulations. Intercepts are set to zero and the coefficients of interest are set to 2, 2.5, and 2. The outcomes generated however are determined using correlated error terms, with the correlation varying across simulations.[13] The second-stage equation is:

$$\tilde{y}_i^* = \begin{pmatrix} 2 \\ 2.5 \\ 2 \end{pmatrix} X_{i1} + \epsilon_i,$$

$$\epsilon_i \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}\right).$$

These simulations are repeated ten times each, resulting in two hundred seventy sets of data, analyzed once by each model.

Results of the exercise are shown in Table 1.[14] Again, we see a smaller RMSE in the second stage estimates with tighter posteriors as shown by the smaller standard deviations. Despite the increase in precision, the coverage probability is still close to 0.95. It is actually slightly conservative with a coverage of 0.96. Further, the ZIOP does not produce correlation estimates which are often substantively interesting.

When comparing the proposed model to both one not modeling the zero-inflation and one not modeling correlations, the proposed model outperforms these currently extant alternatives. Results hold across various specifications and different benchmarks. Overall, the RMSE of the second-stage estimates is reduced, and the posterior densities are more precise while still maintaining approximately 0.95 coverage. ZIMVOP is less biased, more precise and efficient, and produces substantively interesting results by modeling both the zero-inflation stage and the second-stage correlations.

[13] There are twenty-seven different data generating processes. The first correlation coefficient, ρ_{12}, varies from -0.8 to 0.8 by 0.2, and the other correlation coefficients are functions of ρ_{12}. See SI-2 for more detail.

[14] Again, no clear patterns emerged as the correlations changed, so results are pooled. Figure SI-2 in SI-2 shows this.

4.2. *Application I: Presidential Campaigns in Mexico*

Having demonstrated the benefits to our inferences using ZIMVOP, I will now apply it to substantive questions of interest, demonstrating its applicability and beneficial features. Langston and Rosas (2016) argue that municipal-level party support and competitiveness are significant determinants to party campaigns' calculi when deciding which municipalities to visit, and whether that visit is a meeting or a rally. Visits can help assess and signal local party strength and their mobilization networks, and can signal a party's interest in a locality. If they hold a rally and it is not well-attended, however, this can impose more costs than benefits, signaling a lack of strength in the area. Rallies are also expensive, and if the return is not great enough the cost is not worth it. To test the saliency of certain factors entering into this decision, they analyze the Mexican presidential campaigns of 2006 and 2012, focusing on the three major parties – the PRI, the PAN, and the PRD.

The current analysis builds on this with two main propositions: (1) the strategies of parties are not the same and will respond to local support differently, and (2) parties will engage in a Colonel Blotto-type interaction, targeting the municipalities their rivals are targeting.[15] ZIMVOP is uniquely suited to test these propositions because coefficient estimates vary between parties, and the correlation between the parties' decision processes captures the degree to which parties are basing their decisions on the observed or anticipated behavior of their competitors. This proposed strategic interaction between the parties should result in a positive estimate of the correlation. Further, the vast majority of municipalities are never visited. To obtain reliable estimates of the correlations between parties' calculi and the coefficients of interest, a zero-inflation stage is necessary.

The outcome is a vector of ordered party outcomes – 0 for no visit, 1 for a meeting, and 2 for a rally. There are three dimensions, one for the PRI, one for the PAN, and one for the PRD. For example, if the PRI holds a meeting in a municipality at a given time period,

[15] The Colonel Blotto game, first solved by Borel (1921), is a game based on the idea that battlefields will be won by whichever side sends the most troops, causing a pooling of resources at locations. This idea has been used as a metaphor for party competition in the political science literature (see Laslier and Picard 2002; Myerson 1993).

the PAN do not visit, and the PRD hold a rally, the outcome would be [1,0,2]. The first stage predictors consist of one matrix of municipal characteristics. I use three variables: *population*, *vote HHI*, and *2012 dummy*. The variable *population size* is a measure of how populated a municipality is. Sparsely populated municipalities are not worth a visit. The variable *2012 dummy* is a dummy variable indicating if the observation is in 2012 rather than 2006. The visits are, according to Langston and Rosas (2016), a common strategy in newer democracies that still have clientelistic networks. As a state's democracy grows and evolves, these visits should be less common. Finally, to capture compeitiveness, I include the HHI (Herfindahl-Hirschman) index of the previous vote as a general measure for party dominance in the municipality. This measure is a sum of the squared previous vote shares for each of the three parties. When it is high, it indicates one party has dominance over the others. These municipalities should be less appealing for parties to visit. Either it is the dominating party and there is little to be gained from a visit, or it is the weaker party and going would be both a waste of resources and potentially damaging.

The second stage includes all of the variables used in Langston and Rosas (2016), including the first stage variables. Variables included that are not in the first stage are *gira*, *concurrent*, *coparty mayor*, and *previous vote*. The variable *gira* is a dummy variable for whether or not the visit was part of a multi-stop tour. The variable *concurrent* is a dummy indicating whether or not the mayoral race is concurrent with the presidential race. The two main variables of interest are *coparty mayor*, an indicator for whether or not there is a mayor of the same party, and *previous vote*, the party's previous vote share. These capture the underlying political support for a party. In the original data, if a party visits a municipality more than once in a given time period it is included twice. For this analysis, to keep the number of observations the same and to not falsely give parties 0 outcomes or repeated outcomes, duplicates are dropped, keeping the highest outcome. This only affects about 150 out of over 3,000 observations. About half of the data are randomly selected municipalities that were not visited by any party.

Analysis The results of the analysis are presented graphically in Figure 2.[16] There is strong evidence for the two propositions. First, the correlation estimates are all positive and reliable. This suggests parties are engaging in Colonel Blotto dynamics and targeting municipalities their competitors also target. Second, the second stage estimates for the parties of the variables of interest vary considerably. The estimates for *coparty mayor* and *previous vote* vary across parties, suggesting different calculi for the parties.

For the PAN, having a mayor of the same party reliably increases the odds ratio of moving up a category in the visitation scheme. It is unreliable for the PRI, but indicates that the PRD may be responding to this in the opposite direction. The estimates for *previous vote* suggest something similar: parties are not responding in the same way to the same factors. Again, the PAN seem to go where they have support, but so do the PRD, while the PRI estimate, though (just barely) unreliable at a 95% level, suggests that the PRI is more likely to visit and hold rallies in municipalities with less support.

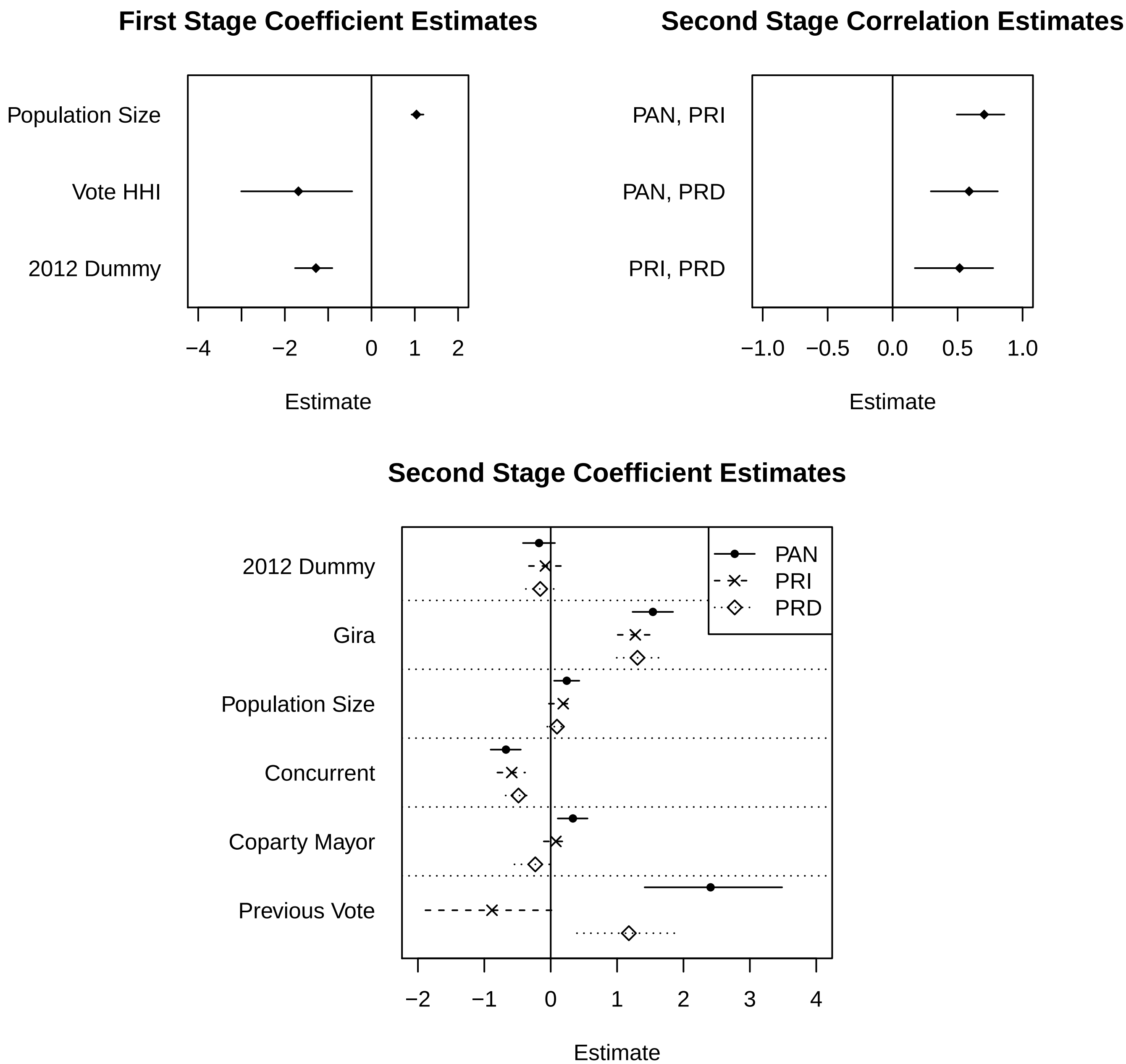
4.3. *Application II: The Effect of Terrorism on Diplomacy*

A common hypothesis in the international relations literature is that terrorism spoils interstate relations. Directed terrorist attacks (i.e., a foreign state or a national of that state claims responsibility for the attack) are often strategic and effective actions taken to disrupt state cooperation (Conrad and Walsh 2014; Findley and Young 2012; Kydd and Walter 2002). These arguments generally do not specify whether the impact or intent is to discourage active cooperation or encourage retaliation. Theoretically, I expect both to be affected. Terrorist attacks should increase negative gestures (e.g., economic sanctions) toward the state from which the attacks originated, and decrease positive gestures (e.g., treaty signing).

These gestures are rare. Some states would rarely interact, positively or negatively, because these interactions are costly and some dyads have little reason to disrupt the status quo or even communicate. Positive and negative gestures should also intuitively be analyzed

[16] Two chains of 150,000 iterations were run. The package `superdiag` indicated no evidence of lack-of-convergence, and all $\hat{R}$'s are close to one.

Figure 2: Results from the presidential campaign visits in Mexico.



Note: Estimates are shown with 95% credible intervals of the posterior distribution. The top-left panel shows the first stage estimates. All are reliable and in the predicted direction. The top right panel shows the correlations between parties in the second stage. All are reliable and positive, suggesting parties target municipalities that their competitors target. The bottom panel shows the second stage estimates. The variables of interest, coparty mayor and previous vote vary considerably in their estimates, indicating that parties strategize differently.

jointly. Similar observed and unobserved characteristics of country dyads will drive both the observed number of negative gestures and the observed number of positive gestures. For example, directed terrorist attacks are observed, whereas a latent cultural affinity may not be. These characteristics of the problem make ZIMVOP an appropriate modeling choice.

To test this proposition I utilize data from Conrad and Walsh (2014).[17] Data are at the country dyad level. Previous work on interstate interactions and terrorism typically analyze the directed dyad, arguing it is the most appropriate level of analysis. Further, only politically relevant dyads are included, with a basis in the literature for the exclusion of certain dyads (Conrad and Walsh 2014). It is monthly data ranging from 1990-2004.

The outcome variable is a bivariate, ordered indicator of positive and negative directed interstate gestures. It takes the value of zero if no actions are taken, 1 if one action is taken, and 2 if more than one are taken.[18] Following the literature, I include a lagged dependent variable, a construct summing the transformed dependent variable across the three previous months.[19]

The main explanatory variable is the number of directed terrorist attacks at a one, two, three, four, and five month lag, a common practice in similar analyses. Both the main explanatory variable and the summed lagged dependent variable are only in the second stage analysis. The first stage variables, which are also included in the second stage, are a set of controls that likely affect the possibility of interstate relations and the likelihood of directed terrorist attacks, with a basis in the literature. These are standard variables in both the conflict literature and the interstate cooperation literature (Conrad and Walsh 2014). They include *allies*, *contiguous*, *log distance*, *joint democracy*, *major dyad*, *dyadic trade*, *enduring rivalry*, and *GDP country 2*.

The variable *allies* is a dummy variable indicating if the states in the dyad belong to a

[17]Most of their data come from Bond et al. (2003).

[18]For example, if country A sends one positive gesture and 3 negative gestures to country B, the outcome for that dyad would be [1,2]. Beyond 2 actions, there is likely little difference in the substantive significance of these actions.

[19]There is little reason to include every month as its own variable if the reasoning is that histories of interstate actions can predict current actions. An action one month prior is not any more meaningful than using this constructed variable as a crude proxy for the dyad's history.

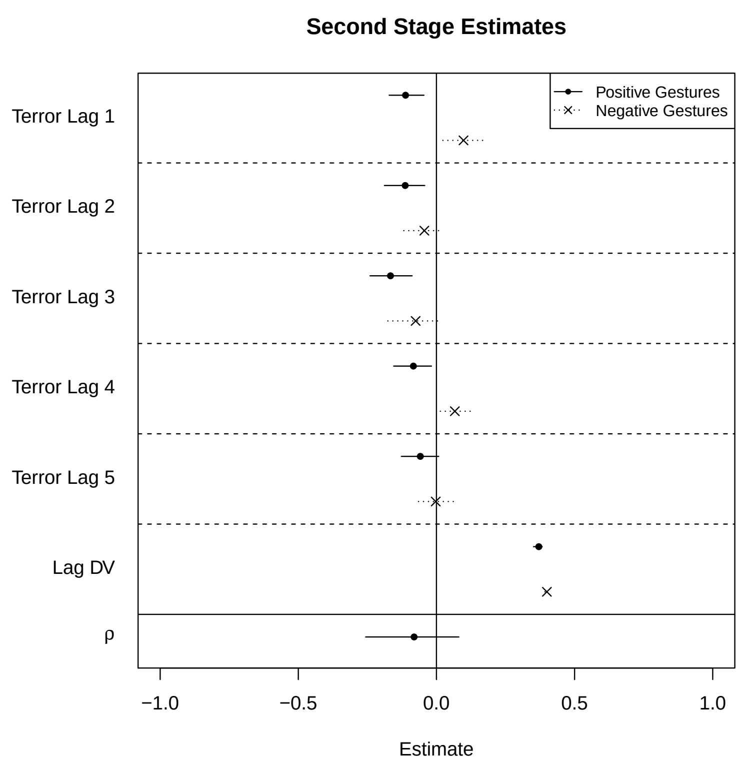
shared alliance. The variable *contiguous* indicates whether the states share borders. The logged distance between the states' capitals is *log distance*. The variable *joint democracy* indicates if both in the dyad are democracies. The variable *major dyad* indicates if the dyad consists of two important states in global affairs, which is based in the literature. The amount of trade between partners is captured by *dyadic trade*. If the states in the dyad have been in a rivalry for an extended period of time, *enduring rivalry* is coded as one, which is also based in the literature. Finally, *GDP country 2* is the GDP of the country receiving the signal.

Analysis Due to the large number of predictors in this model, Figure 3 only shows the second stage estimates for the variables of interest, the lagged dependent variable, and the correlation coefficient.[20] The results are generally consistent with expectations. There is strong evidence that terrorist attacks decrease the likelihood and degree of positive gestures. There is also suggestive evidence that terrorist attacks increase the likelihood and degree of negative gestures. The estimates for this latter proposition are only reliable at certain time lags. The most interesting result in terms of this latter proposition is that the strongest and most reliable of these estimates is the one-month lag. This suggests an immediate but not necessarily long-lasting effect on negative gestures. Though I did not hypothesize this nuance, it is, retrospectively, a sensible result.

The correlation coefficient is unreliable at the 95% level but its estimate is negative. This suggests that, given the covariates, states are not likely to send both a positive and negative gesture. They are instead more likely to resort to one or the other. This is conceptually more difficult to grasp than the correlations between parties in the first analysis. If a negative gesture and positive gesture were estimated to have the same probability, the negative correlation of the residuals indicates that the gesture sent, if one is sent, would be either a negative or a positive one. If one disturbance is positive, the other is likely to be negative.

[20] Other estimates are provided in SI-4.1. Two chains of 150,000 iterations were run. The package `superdiag` indicated no evidence of lack-of-convergence, and all $\hat{R}$'s are close to one.

Figure 3: Results from the analysis on the effect of terrorist attacks on positive and negative interstate gestures.



Note: Estimates are shown with 95% credible intervals of the posterior distribution. Only a selection of second-stage estimates are included. There is evidence that lagged terrorist attacks decrease the likelihood and degree of positive gestures. There is suggestive evidence that negative gestures are more likely following terrorist attacks. The correlation coefficient is not reliable at the 95% level, but it is estimated as negative. This suggests states are not gesturing both positively and negatively at the same time.

5. CONCLUSION

Binary and ordered outcomes are often of interest in political science, but analyses of these data can be problematic. There is often an excess of zeros, and the data generating processes of seemingly unrelated outcomes may in fact be related. These issues can lead to biased and inefficient estimators. This paper proposes a new model, ZIMVOP, that appropriately addresses these issues and in doing so opens the door to answering questions we have been unable to answer. ZIMVOP not only provides better estimates of our parameters of interest as shown in the simulation exercises, it also helps us recover useful information that otherwise would be lost. We can investigate the nature of the related processes causing observed outcomes and analyze the varying effects of observables at the zero-inflation stage and the outcome stage. I applied ZIMVOP to presidential campaign visits in Mexico and to the spoiling effect of international terrorism to illustrate the model's benefits.

Though ZIMVOP outperforms existing models in certain contexts, it is not as straightforward to interpret as its simpler alternatives. ZIMVOP is also fairly computationally intensive, with some models taking a very long time to converge. Further, it only applies to cases in which we believe processes are related and an excess of zeros suggests two stages of data generation. Nevertheless, the applicability of ZIMVOP is potentially wide.

For example, decision-making often results in unanimity. During U.S. Supreme Court agenda setting, most Justices vote to not hear the case. If we want to explain the likelihood of SCOTUS accepting a case, there is very likely a relationship between the processes of one Justice voting to hear the case and another Justice wanting to hear the case that is unexplained from observables. This would in fact be a very interesting question because some correlations, such as those of ideologically proximate Justices, may be positive, while those of ideologically distant Justices may be negative.

Survey questions are also a very well-suited application of the model. If, for example, we are interested in how party affiliation impacts opinions of high-level (high-knowledge) policies, the outcome, opinion towards the policy, will exhibit pooling at indifferent or do not know. We could allow a correlation between this response and other questions indicating political knowledge, common on most surveys. This correlation would help explain both the relationship between the respondents' knowledge and the response, and allow inferences about how others would respond if their knowledge was high enough to have an opinion.

Finally, ZIMVOP as proposed in this paper has a zero-inflation stage modeling the all-zero state. In other words, though the outcome is multivariate, the zero-inflation stage is univariate. This is computationally less demanding and in general theoretically sensible. There are municipal-level characteristics that make no municipality visitable, for example, and there are country dyad characteristics that make any signaling between states unlikely. With both of these examples, particularly when considering that the underlying processes are related, there would be no reason to deviate from this univariate zero-inflation. However, ZIMVOP could easily allow for inflation in each component, potentially opening up new avenues of application.

References

Albert, James H, and Siddhartha Chib. "Bayesian analysis of binary and polychotomous response data". *Journal of the American statistical Association* 88, no. 422 (1993): 669–679.

Ansolabehere, Stephen D, Shanto Iyengar, and Adam Simon. "Replicating experiments using aggregate and survey data: The case of negative advertising and turnout". *American Political Science Review* 93, no. 04 (1999): 901–909.

Bagozzi, Benjamin E, and Kathleen Marchetti. "Distinguishing occasional abstention from routine indifference in models of vote choice". *Political Science Research and Methods* (2014).

Bagozzi, Benjamin E, et al. "Modeling two types of peace: The zero-inflated ordered probit (ZiOP) model in conflict research". *Journal of Conflict Resolution* 59, no. 4 (2015): 728–752.

Barnard, John, Robert McCulloch, and Xiao-Li Meng. “Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage”. *Statistica Sinica* (2000): 1281–1311.

Bas, Muhammet Ali, Curtis S Signorino, and Robert W Walker. “Statistical backwards induction: A simple method for estimating recursive strategic models”. *Political Analysis* 16, no. 1 (2008): 21–40.

Bond, Doug, et al. “Integrated data for events analysis (IDEA): An event typology for automated events data development”. *Journal of Peace Research* 40, no. 6 (2003): 733–745.

Borel, Emile. “La théorie du jeu et les équations intégrales à noyau symétrique”. *Comptes Rendus de l’Académie des Sciences* 173, **numbers** 1304-1308 (1921): 58.

Cameron, A Colin, and Pravin K Trivedi. *Microeconometrics: Methods and applications.* Cambridge university press, 2005.

Carson, Jamie L, and Jason M Roberts. “Strategic politicians and US House elections, 1874–1914”. *Journal of Politics* 67, no. 2 (2005): 474–496.

Conrad, Justin, and James Igoe Walsh. “International cooperation, spoiling, and transnational terrorism”. *International Interactions* 40, no. 4 (2014): 453–476.

Dunne, John Paul, and Nan Tian. “The determinants of civil war and excess zeros”. Working Paper.

Findley, Michael G, and Joseph K Young. “Terrorism and civil war: A spatial and temporal approach to a conceptual problem”. *Perspectives on Politics* 10, no. 02 (2012): 285–305.

Gurmu, Shiferaw, and Getachew A Dagne. “Bayesian approach to zero-inflated bivariate ordered probit regression model, with an application to tobacco use”. *Journal of Probability and Statistics* 2012 (2011).

Harris, Mark N, and Xueyan Zhao. "A zero-inflated ordered probit model, with an application to modelling tobacco consumption". *Journal of Econometrics* 141, no. 2 (2007): 1073–1099.

Jackson, John E. "A seemingly unrelated regression model for analyzing multiparty elections". *Political Analysis* 10, no. 1 (2002): 49–65.

Kadel, Rajendra. "A latent mixture approach to modeling zero-inflated bivariate ordinal data". PhD thesis, University of South Florida, 2013.

King, Gary. "A seemingly unrelated Poisson regression model". *Sociological Methods & Research* 17, no. 3 (1989): 235–255.

King, Gary, and Langche Zeng. "Explaining rare events in international relations". *International Organization* 55, no. 03 (2001): 693–715.

— . "Logistic regression in rare events data". *Political analysis* 9, no. 2 (2001): 137–163.

Kydd, Andrew, and Barbara F Walter. "Sabotaging the peace: The politics of extremist violence". *International Organization* 56, no. 02 (2002): 263–296.

Langston, Joy, and Guillermo Rosas. "Presidential campaigns under single-district plurality: Visits, rallies, and the calculus of electoral mobilzation in Mexico". Submitted, 2016.

Laslier, Jean-Francois, and Nathalie Picard. "Distributive politics and electoral competition". *Journal of Economic Theory* 103, no. 1 (2002): 106–130.

Myerson, Roger B. "Incentives to cultivate favored minorities under alternative electoral systems." *American Political Science Review* 87, no. 04 (1993): 856–869.

Philips, Andrew Q, Amanda Rutherford, and Guy D Whitten. "Dynamic pie: A strategy for modeling trade-offs in compositional variables over time". *American Journal of Political Science* (2015).

— . "The dynamic battle for pieces of pie-Modeling party support in multi-party nations". *Electoral Studies* 39 (2015): 264–274.

Signorino, Curtis. “Strategy and selection in international relations”. *International Interactions* 28, no. 1 (2002): 93–115.

Signorino, Curtis S. “Structure and uncertainty in discrete choice models”. *Political Analysis* 11, no. 4 (2003): 316–344.

Tomz, Michael, Joshua A Tucker, and Jason Wittenberg. “An easy and accurate regression model for multiparty electoral data”. *Political Analysis* 10, no. 1 (2002): 66–83.

Tucker, Joshua Aaron. *Regional economic voting: Russia, Poland, Hungary, Slovakia, and the Czech Republic, 1990-1999.* Cambridge Univ Press, 2006.

Zellner, Arnold. “An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias”. *Journal of the American statistical Association* 57, no. 298 (1962): 348–368.

Zellner, Arnold, and David S Huang. “Further properties of efficient estimators for seemingly unrelated regression equations”. *International Economic Review* 3, no. 3 (1962): 300–313.