Bridging Levels of Analysis:
Selection as a Multi-level Process

Mark David Nieman,[*] Olga Chyzh,[†] and Douglas M. Gibler[‡]

ABSTRACT

Many dyadic processes are nested within broader system-level structures, yet such dependency is often ignored in empirical research. As a result, empirical estimates of observable events data are prone to inferential problems associated with non-random sample selection. We demonstrate that this type of multi-level data generating process (DGP) is equivalent to a selection DGP and can be modeled using known selection models. Using Monte Carlo simulations from a multi-level DGP, we show that a selection model outperforms other commonly employed estimators, including random effects models; the multi-level structure of the data also helps meet the exclusion restriction. To further illustrate the importance of the proposed modeling approach, we replicate two prominent empirical studies of government-opposition behavior—a model of civilian protest outcomes and estimates of civilian killing by insurgent groups—and demonstrate that structural selection affects many of the inferences we draw from the observable data.

[*]Assistant Professor of Political Science, Iowa State University.
[†]Assistant Professor of Political Science and Statistics, Iowa State University.
[‡]Professor of Political Science, University of Alabama.

1. INTRODUCTION

Issues associated with sample selection remain ubiquitous within social science research. While we know much about the process of selection, we still ignore its general principles when collecting and analyzing events data. As a result, many existing data collection efforts (1) lack clear theoretical guidelines as to what cases constitute the target population, (2) often inadvertently select on observable factors that correlate with dependent variables of interest, and (3) span different levels of aggregation (individual, dyadic pairs, system/network). These issues, in turn, lead to substantial problems in subsequent analyses, including the production of biased, inefficient, and/or incorrect model estimates.

These problems can be ameliorated by using a *multi-level/selection* framework that *crosses* levels of analysis, where estimates of the outcome variable are informed and enhanced using data from higher levels of aggregation. We argue that selection can be treated as a special case of a multi-level model, in which structural conditions influence whether common types of micro-level events data are observed. In what follows, we demonstrate, theoretically and using simulations, that our approach allows researchers to combine disconnected datasets in theoretically-informed ways and enhances statistical inferences by leveraging existing data.

As an example of our approach, consider the ample research on the development and evolution of insurgent groups. These studies present a useful focus because they can readily be grouped in accordance to the level of aggregation into research that uses aggregated state-level data (e.g., Collier and Hoeffler 2002; Fearon and Laitin 2003; Gibler and Miller 2014) and the events-type data that relies on dyadic, rebel-government interactions (e.g., Buhaug et al. 2009; Cunningham et al. 2009; Cunningham 2011, 2013). Aggregated state-level data has proven to be an immensely powerful and robust tool for analyzing civil war onset, duration, and outcome. Fundamentally, however, civil war is an interaction between the government and an insurgent group (Cunningham et al. 2009). Many disaggregated events-type datasets account for this interactive aspect of conflict and serve as valuable tools for a number of different research agendas and permit testing of causal mechanisms. However, absent from these studies is the recognition that the observation of any given rebel group is non-random (Nieman 2015). Insurgent group data are mostly limited to countries that have experienced a civil war which is, of course, a non-random sample.

Rebel-government interactions are nested within the state itself, so the factors that weaken the government are also likely to be correlated with those that provide relative strength to rebel groups. More problematic is that data on rebel-government interactions suffer from a high degree of censoring. We can only collect data on the rebels that are sufficiently strong to challenge government forces. Weaker, latent rebel groups are forced to remain inconspicuous lest they provoke government retaliation. These groups will be unobserved and, therefore, will be absent from our datasets; their absence will then bias the estimates we derive from the observed data.

We begin in the next section with a review of problems related to non-ignorable missingness, multi-level data structures, and traditional techniques of modeling sample selection. We then demonstrate that sample selection can be shown as a special case of a multi-level data structure. Next, we present the results of Monte Carlo simulations that demonstrate the effects of ignoring structural factors when estimating micro-level outcomes using common estimation techniques. Finally, we provide two empirical applications to illustrate the implications of our argument. First, we replicate Chenoweth and Stephan (2011) and the argument that non-violent protest campaigns are more likely than violent campaigns to achieve successful outcomes. Second, we re-examine Wood (2010) and the finding that civilian targeting is inversely related to rebel group strength. For both studies we find that prior conclusions are largely determined by the structural selection processes that make events data observable.

2. STRUCTURE AND SELECTION

Political scientists have collected and analyzed an impressive array of events data. Comparative and international relations scholars have, for example, collected micro-level events data on protests, riots, coups, and insurgent groups.[1] Much of this data collection effort, however, only allows for meaningful inference if we are able to understand the underlying processes that would result in us observing those events. A failure to model events within the broader context is akin to omitted variable bias. Consider, for example, the anti-government protests in Tahir Square, Cairo, that took place in 2011. These protests were both similar *and* different to the 2005 youth protests in Paris. Though both are examples of youth protests, one must account for substantial differences in the social context in which they occur to understand why the former helped shape the long-term fate of its entire region, while the latter had only small short term-effects on the domestic policies of a particular government. In these cases as well as more generally, knowing the underlying context is important to correctly identify points of comparisons between types of protests, as well as the points of departure in making such comparisons meaningful.

The problem analysts and data scientists face is that most social events are non-random. The underlying processes behind the occurrence of observable events are frequently correlated with the dependent variables of interest—one of the most serious impediments to deriving meaningful inferences regarding causal relationships. Scientists can only collect data on the observable events and often struggle to identify relevant "control groups" or non-events to use for the baseline comparison. While most data collectors may be able to agree on what constitutes an insurgency or a protest, defining and identifying an appropriate "non-protest" or a "non-insurgency" seems like an impossible task.

Thus, social science data are frequently characterized by non-ignorable missingness that is correlated (1) with the outcome variable and (2) the multi-level structures in the data.

[1]See, for example, Chenoweth and Stephan (2011); Cunningham et al. (2009); Cunningham (2011); LaFree and Dugan (2007); Powell and Thyne (2011); Salehyan et al. (2012).

The first issue—that of the correlation between the outcome variable and the missingness in the data—is known as the selection problem and may be accounted for through the use of bivariate estimation techniques (e.g., Heckman selection models).[2] The second issue—that the missing data and the outcome variable is also often correlated with the multi-level structures in the data—has so far received little scholarly attention.

Traditional statistical approaches to modeling selection involve multi-stage estimation (e.g., Heckman 1979) of the same unit of observation, which improves parameter estimates associated with one stage by using the information from other stages. Application of such multi-stage approaches, however, imposes rather strict demands on data availability. The "selection" equation requires a random sample of both uncensored *and* censored observations that most micro-level events datasets do not provide. In addition, two-step approaches, where the selection stage is estimated and information regarding censoring is then included in the outcome stage, require that the outcome stage dependent variable be normally distributed (Freedman and Sekhon 2010; Terza 1998), a condition that is often not met with events data. Finally, analyses using multi-level data typically use listwise deletion and exclude cases in which observations on one level are unobserved—a common occurrence in most micro-level event data analyses. This causes potentially relevant counter-factual cases to be ignored, thus biasing causal inferences derived from the analysis.

We argue that understanding and properly modeling multi-level structures can help alleviate the selection bias introduced by the correlation between the missingness in the data and the outcome variable. Bridging bivariate estimation techniques with multi-level modeling does introduce several important technical issues, mostly related to differences in the levels of analysis, but these issues are not insurmountable and we outline several techniques that bridge data in theoretically informed and methodologically sounds ways. In what follows, we provide a brief overview of the types of data structures common to social sciences, common ways to model missingness/sample selection issues, and our proposed approach for

[2]While selection models are prevalent in political science and economics applications, to the best of our knowledge they have not been used in this context (for an exception, see Gibler 2016).

treating these data types.

2.1. *Non-Ignorable Missingness*

Social science data collection is frequently constrained by two related issues: (1) a lack of theoretical guidelines on what observations make up the target population for a particular research question, and (2) data availability/missingness being correlated with the outcome of interest. The first issue is exemplified in the research on conflict among international states. Conflict is often identified as the presence of a threat, display, or use of force by officials of one state against another state in the international system (Palmer et al. 2015). If one were to use all pairs of states as a target population, then approximately 99.5% of the resulting sample would never have a positive dependent variable since most states are at peace with most other states most of the time (Lemke and Reed 2001). A common remedy for this is to limit the target population dyads of states that have the opportunity (e.g., physical capabilities) to engage in conflict. These are the "politically relevant dyads" outlined in the literature (see, for example, Siverson and Starr 1991; Weede 1976).[3] However, while intuitively appealing, narrowing the sample to politically relevant dyads does not fully address the issue, since 10-26% of conflict occurs between "irrelevant" dyads during any given time period (Lemke and Reed 2001; Maoz and Russett 1993). This type of selection criterion also still provides samples that include many dyads that lack any face validity for the opportunity of conflict (e.g., Austria-Hungary—Ethiopia) (Lemke and Reed 2001, 130). Thus, selecting the target population by imposing inflexible assumptions such as political relevance may cause as many problems as it solves.[4]

The second issue—the correlation between the patterns in data missingness and the outcome variable—is frequently encountered in the study of sub-state or non-state actors,

[3] Empirically, politically relevant dyads are usually defined as either geographically contiguous states or dyads that include at least one major state (Russett and Oneal 2001).

[4] Xiang (2010) offers one alternative approach where, in the absence of data on the dependent variable in the selection equation, relevance is treated probabilistically.

such as terrorist groups, insurgencies, and civil wars. Thus, data collection on civil war is (and has to be) necessarily conditioned by some minimal level of noticeable subnational activity. This implies that data collection on terrorist or insurgent groups will remain missing until the group's activity reaches a recordable level, and likewise will remain missing even though the group may substantially affect the conflict. We also see this correlation when identifying ethnic groups, social movements, refugee flows, and other types of political behaviors that require some tipping point in behavior in order to be recordable.

2.2. *Multi-Level Structures*

Notably, both of these data collection problems are in part caused and exacerbated by the multi-level structure that is characteristic to social science data (Steenbergen and Jones 2002). For example, international states are nested within regions and time-periods, and the states themselves have varying regions of conflict within their territories. These multi-level structures may occur naturally or by design, and the multi-level data structures may actually result from a lack of availability of certain types of data as well. So, the resources available to an insurgent or terrorist group must often be modeled as nested with its entire host state in order to model the effects of GDP per capita, geographical location, and terrain, even though most theories suggest regional variation and regionally-specific effects for each of these variables. This suggests that misspecificaiton of the multi-level, conditional structure of the data can lead to both type I and II errors in multiple contexts.

Statistically, variation at different levels of data aggregation is modeled as:

$$\boldsymbol{Y} = \boldsymbol{\alpha}_j + \boldsymbol{\beta}\boldsymbol{X} + \boldsymbol{\epsilon} \tag{1}$$

$$\boldsymbol{\alpha}_j = \boldsymbol{\gamma}\boldsymbol{Z} + \boldsymbol{\eta} \tag{2}$$

where $\mathbf{Y}$ is a vector of the dependent variable, $\mathbf{X}$ is a matrix of individual-level covariates, $\boldsymbol{\beta}$ is a vector of individual-level coefficients, $\boldsymbol{\epsilon}$ is a vector of error terms at the individual-level,

$\boldsymbol{\alpha}$ is a vector containing a random intercept common to all individual observations nested within a common group j, $\mathbf{Z}$ is a matrix of group-level covariates, $\boldsymbol{\gamma}$ is vector of group-level coefficients, and $\boldsymbol{\eta}$ is a vector of the group-level error terms. The values for the outcome variable for individual observations thus have multiple sources of variation that stem from both individual and group-level characteristics.

The most common application of multi-level models examines individual observations that are "nested" within an encompassing group structure, such as counties within states, classrooms within a school, or individuals within countries. A simple multi-level approach models common sources of variation across states and years as random effects in the data.[5] This approach captures unobserved heterogeneity across groups while simultaneously keeping information and efficiency gains from the partially pooled data (Gelman and Hill 2006, 254-275). Naturally, in the presence of multi-level structures in the data, "ignoring the aggregate information excludes potentially important effects and treating the aggregate information as individual level effects confuses covariance in the model" (Gill 2009, 395).

The presence of multi-level structures and the need to appropriately model them has been recognized for some time. However, missing from these treatments is the fact that the multi-level structures are frequently correlated with both the dependent variable and the patterns of missingness in the data, and traditional tools for modeling sample selection or multi-level structures do not properly account for this correlation. Indeed, the most common method for controlling for selection bias—Heckman and Heckman-type models—simply ignore the multi-level or nesting structures in the data. We change this and demonstrate the appropriateness of our approach.

[5] Multi-level models can also accommodate more complex structures, such as group-level variation in slopes, or n-number of levels.

Per Heckman (1976, 1978, 1979), the sample selection problem takes on the form:

$$\boldsymbol{s}^* = \boldsymbol{\gamma Z} + \boldsymbol{\eta}, \tag{3}$$

so that

$$\boldsymbol{s} = \begin{cases} 1 \text{ if } \boldsymbol{s}^* > 0 \\ 0 \text{ if } \boldsymbol{s}^* \leq 0 \end{cases}$$

and

$$\boldsymbol{y}^* = \begin{cases} \boldsymbol{\beta X} + \boldsymbol{\epsilon} & \text{if } s = 1 \\ 0 & \text{if } s = 0 \end{cases} \tag{4}$$

where $\boldsymbol{s}^*$ is a vector of the latent dependent variable in the "sample selection" equation, $\boldsymbol{s}$ is a vector of the latent variable's observed realization, $\boldsymbol{y}^*$ is the dependent variable in the "outcome" equation, $\boldsymbol{Z}$ and $\boldsymbol{X}$ are matrices of exogenous covariates, $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are vectors of the regression coefficients, and $\boldsymbol{\eta}$ and $\boldsymbol{\epsilon}$ are vectors of the error terms, correlated at ρ.

When the dependent variable in the outcome equation, $\boldsymbol{y}^*$, is measured on a continuous scale (i.e. $\boldsymbol{y}^* = \boldsymbol{y}$), Equations 3 and 4 may be estimated in one of two ways. First, the two equations may be estimated simultaneously using a full information maximum likelihood (FIML) approach. Or, second, the two equations may be estimated sequentially as a two-step procedure. In the second approach, the error correlation between equations is modeled by calculating the inverse Mill's ratio from Equation 3 and using it as an additional regressor in Equation 4.[6]

When the dependent variable in the outcome equation $\boldsymbol{y^*}$ is measured on a discrete scale (i.e. $\boldsymbol{y}^* \neq \boldsymbol{y}$), however, the two-step estimation (e.g., estimating two probit models,

[6]The inverse Mill's ratio, λ, is calculated as a ratio of the probability density function and the cumulative density function for the individual observations in the selection equation, or $\lambda = \frac{\phi(\gamma Z)}{\Phi(\gamma Z)}$.

sample selection with count data, etc) *no longer produces unbiased estimates.* Use of the inverse Mill's ratio assumes (1) a bivariate normal distribution of the error terms in the selection and the outcome equations and (2) and that the omitted variable η, modeled via the inverse Mill's ratio, has a linear effect on the dependent variable in the outcome equation. If either of the assumptions is not met, the inclusion of the inverse Mill's ratio will lead to model misspecification and may induce bias, rather than help alleviate it (Greene 2006, 2010; Freedman and Sekhon 2010; Terza 1998; Winship and Mare 1992).[7] Thus, the FIML approach is actually preferred over the two-stage approach when there is a limited dependent variable in the outcome equation, as is often the case with events data.

3. MODELING SELECTION AS A MULTI-LEVEL STRUCTURE

For the reasons discussed above, modeling events data must account for any underlying selection process from the structural level and should do so using a FIML rather than a two-step approach. Applications using selection models have been most commonly used to model selection at the same level of analysis. We argue that these models may also provide leverage in dealing with multi-level data structures.

We start by demonstrating that the selection data generating process (DGP) is a special-case of the multi-level DGP. Consider the case where the multi-level structure is

$$\boldsymbol{y}^* = \boldsymbol{\alpha}_j^* (\boldsymbol{\beta X} + \boldsymbol{\epsilon}) \tag{5}$$

$$\boldsymbol{\alpha}_j^* = \boldsymbol{\gamma Z} + \boldsymbol{\eta} \tag{6}$$

where $\text{Corr}(\epsilon, \eta) = \rho$.[8] In this case, effects from the individual-level are moderated by structural aspects. In practice, this may imply that measurement error at one level of

[7] The use of a two-step method with a limited dependent variable in the outcome equation may, in some cases, increase, rather than reduce, the amount of bias in the estimates, compared to a naïve probit model (Freedman and Sekhon 2010).

[8] If $\rho = 0$, then the interaction of the group- and individual-level factors is equivalent to separating the data into subsamples, i.e. selecting on the independent variable.

analysis is related to measurement error at another level of analysis or covariates at one level are related to those at another level. Accuracy of records of economic output at a province level, for example, is likely to be correlated with data at the national level. Or, a state's military strength relative to other international states is likely to be correlated with a government's military strength relative to domestic opposition groups as well.

If $\boldsymbol{\alpha}_j^*$ is observed as a dichotomous variable (e.g., violence must reach a certain threshold to be included in a dataset), then $f(\alpha_j^*)$ from Equation 6 is analogous to $f(s^*)$ from Equation 3, where $f(\cdot)$ is a cumulative density function. Similarly, the effect of $\boldsymbol{\alpha}_j^*$ in Equation 5 is equivalent to that in Equation 4. In other words, the effects of individual-level covariates are correlated with the group-level process that determines whether the individual-level outcomes are observed in the first place. For example, the effect of rebel characteristics on securing concessions from a central government will be related to the state-level characteristics that make civil conflict more likely. Thus, ignoring the structural characteristics and estimating only the effect of individual-level covariates may result in omitted variable bias.[9]

Concerns over the exclusion restriction—that at least one exogenous variable is excluded from one of the equations—common to Heckman-type selection models (Sartori 2003; Winship and Mare 1992), are overcome due to different aggregation levels in the data for the selection and the outcome equations. Thus, model identification does not depend on restrictive distributional assumptions on the errors. In fact, the distributional assumption of the error terms are easily relaxed or modified (e.g., Greene 2006, 2010; Xiang 2010). We demonstrate the utility of this approach using both simulations and empirical applications.

3.1. *Monte Carlo Analysis*

As an initial proof of our approach, we provide the results of a Monte Carlo simulations that demonstrate that estimates of predictors of political event data—such as protest behavior,

[9]Estimating a multi-level model on cases with an observed individual-level outcome does not alleviate the problem since doing so ignores information from the group-level where individual-level data is unobserved.

coups, and political violence—may be biased. Civil war studies present a useful focus because they are conducted at either the state-level of analysis or at the dyadic-level (e.g., rebel-government interactions). We know, however, that rebel-government interactions are nested within the state itself, so factors that weaken the government are likely to be correlated with those that provide relative strength to rebel groups. Even more problematic is that data on rebel-government interactions are highly censored because only rebels who are sufficiently strong to challenge government forces will actually do so.

We generate 100 groups (j) with 50 individual-level (i) observations per group. Y adheres to the following multi-level DGP:

$$\begin{aligned} y_i^* &= s_j^* \left(-0.5 + 1X_i + \epsilon_i\right) \\ s_j^* &= 0.5 + 1Z_j + \eta_j \end{aligned}$$

where X and Z are each distributed $U[-2,2]$, ϵ and η are distributed $N\left(0,1\right)$ where $\text{corr}(\epsilon,\ \eta) \ = \ \rho$, $y = 1$ if $y^* > 0$ and 0 otherwise, and $s_j = 1$ if $s_j^* > 0$ and 0 otherwise. We set $\rho \ \in \ \{0.7,\ 0.4,\ -0.4,\ -0.7\}$. We run 100 simulations at each level of ρ.

We estimate several model specifications on the generated data. Following the traditional treatment of binary outcomes in much of the social science literature, we estimate two probit models, one with just the individual-level variables, and one where the group-level variable is included in the same equation:

$$y_i^* = \beta_0 + \beta_1 X_i + \epsilon_i, \tag{7}$$

$$y_i^* = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \epsilon_i \tag{8}$$

We also estimate two random effect models, allowing the intercept to vary. This is a common estimation procedure designed to capture group effects. We estimate one model with just the individual-level variables and one model with the group-level variable:

$$y_i^* = \beta_{0j} + \beta_1 X_i + \epsilon_i, \tag{9}$$

$$y_i^* = \beta_{0j} + \beta_1 X_i + \beta_2 Z_i + \epsilon_i \tag{10}$$

Finally, we estimate a Heckman probit where the group-level variables are in the selection equation and the individual-level variables are in the outcome equation. As we argue above, we expect that cases where individual-level data are observed are not random but instead occur where specific structural, group-level characteristics are present. Unobserved but related state-level factors are also likely to be correlated with unobserved individual-level characteristics. Measurement error of conceptually related variables, for example, is likely to be correlated across levels of aggregation. Thus, group-level variables can be treated as similar to a selection stage for the individual-level.

$$y_i^* = \beta_0 + \beta_1 X_i + \epsilon_i \text{ if } s_j = 1 \text{ where } s_j = 1 \text{ if } s_j^* > 0, \text{ and 0 otherwise,} \tag{11}$$

$$s_j^* = \gamma_0 + \gamma_1 Z_j + \eta_j \tag{12}$$

We present the results of the Monte Carlo analysis in Figure 1 and Table 1.[10] As our estimates suggest, each of the probit models exaggerates the effect of X as ρ moves away from zero. Moreover, the probit models exert a high degree of certainty in their biased estimates. Estimates of the group variable in probit models that include Z have the wrong sign on the parameter when ρ is positive and are biased towards zero when ρ is negative. Finally, the probit models overestimate the effect of the constant when ρ is positive and underestimate the effect (with the wrong sign) of the constant when ρ is negative. Bias on the constant is problematic given that estimators of discrete data generating processes are hyper-conditional—the effects of a variable on the probability of an outcome are conditional on their position on the cumulative density function (Berry et al. 2010; Williams 2016).

[10]The coefficient reported on the *constant* in the RMSE tables are of the outcome equation for the Heckman probit specification.

Thus, the substantive effects of a quantity of interest, as well as the predictive power of the model, will be incorrect due to bias in the constant.

[Figure 1 about here.]

[Table 1 about here.]

Next, we examine parameter estimates from the probit models with random effects. Like the initial probit models, random effects probit models exaggerate the effect of X the farther ρ is from zero—actually, exaggeration of the effect of X is much more pronounced here than in the traditional probit models. Random effects probit also performs poorly when a group-level variable is included, recovering estimates with the incorrect sign when ρ is positive and estimates that are biased towards zero when ρ is negative. The model also overestimates the constant when ρ is positive and underestimates the constant when ρ is negative. In sum, these estimates suggest that the random effects probit actually exerts the greatest degree of bias when compared to other models, which is especially problematic given how often this approach is used to address group level heterogeneity in structured data.

Finally, we turn to the probit model that accounts for potential structural selection effects. Figure 1 and Table 1 highlight, in particular, that a Heckman selection model is the only model that produces unbiased estimates of the true effects of the state-level covariates, Z. Moreover, accounting for selection, the Heckman probit model also recovers unbiased estimates for each parameter regardless of the value of ρ. A Heckman selection model also performs best in terms of estimating the true value of the model's intercept, β_0. So, if we assume that state-level covariates have an effect on rebels' decisions to attack—a common finding in the empirical literature (Chatagnier and Castelli 2016; Gibler 2016; Nieman 2015)—then ignoring or mistreating state-level structures will lead to substantially biased estimates.

Ultimately, this is a relatively simple demonstration of our argument, but it also holds importance for many recent data collection efforts, including the identification of terror group dynamics, ethnic group affiliation, and other conflict-related group behavior. Beyond the simple bivariate outcome case, the approach generalizes to other types of outcome variables, including limited dependent variables (see Greene 2006; Terza 1998). This is important because a two-stage approach to account for structural selection effects is not appropriate with non-normally distributed data (Freedman and Sekhon 2010; Terza 1998). In the next section, we apply our approach to modeling the selection effect of structural state-level factors to two recent studies of civil strife.

4. EMPIRICAL APPLICATIONS

We replicate two recent studies to demonstrate the impact of ignoring multi-level selection issues. First, Chenoweth and Stephan (2011) is a widely-cited work that suggests non-violent protests are an effective means of eliciting government change. We argue, however, that the likelihood of protest, and non-violent protest in particular, will be conditioned by the *ex ante* likelihood of success given the structural conditions of the state. Thus, it may be the case that nonviolent protests only occur when conditions promoting change are more likely.

Our second replication examines whether the structural conditions in the state also affect the strategies of conflict. Wood (2010) argues that rebel groups that lack the capacity to garner popular support are also the groups most likely to target civilians during civil conflicts. Of course, the observation of rebel groups is non-random and heavily conditioned by the relative strength of the government and the likelihood of success. As we demonstrate above with Monte Carlo estimates, this should lead to biased estimates of covariate effects if the structural conditions promoting success are not properly treated.

4.1. *Structure, Protest Occurrence, and Protest Outcomes*

Chenoweth and Stephan (2011) treat the protest campaign as the unit-of-analysis, with data that explores government–protest movement interactions. However, observation of a protest campaign will likely be (negatively) correlated with the expectation of government repression (Chyzh and Labzina 2015; Pierskalla 2010; Ritter and Conrad 2016). Further, since the use of repression differs systematically across states (Davenport 2007; Hill and Jones 2014; Regan and Norton 2005), we expect the likelihood of both protest and success will covary similarly. Complicating this relationship even more is the fact that protest strategy will covary with these same structural conditions as well. Few nonviolent campaigns are likely to start in under the watchful eyes of regimes that are likely to shoot protesters on sight. So, to account for the possibility of a non-random sample of protest movements, we examine the probability of success by non-violent or violent movements in achieving government concessions in the context of a selection model that crosses levels of analysis. We model state-level data as a selection equation and use those estimates to inform campaign-level data in the outcome equation.

We focus our replication on Table 3.1 from Chenoweth and Stephan (2011). They measure a protest movement's *success* as a binary outcome coded 1 if they achieve their stated goals, 0 otherwise. *Non-violent resistance* is measured as 1 if the movement is primarily non-violent, 0 otherwise. They also control for level of democracy, the number of participants in the movement, and the state's population.[11]

We account for structural factors using the model of state conflict from Fearon and Laitin (2003, Table 1, Model 1).[12] We employ data from Gibler and Miller (2014), who extend and expand Fearon and Laitin's dataset following the original authors' coding rules. The structural model includes common predictors of domestic strife, such as political and government strength factors like *democracy*, political *instability*, *GDP/capita*, and whether a state has territory that is *non-contiguous*.[13] The model also estimates conditions that favor challenges to government authority, such as the size of the *population*, the amount of

[11]See Chenoweth and Stephan (2011) for a discussion of how control variables are measured.
[12]See Regan and Norton (2005) for a similar structural/state-level approach to modeling protest behavior.
[13]Following Gibler and Miller (2014), we include a *democracy squared* term to account for possible non-linear effects.

mountainous terrain, *oil exports*, and *ethnic* and *religious fractionalization*.[14]

Table 2 reports the results of our analyses using probit and Heckman probit selection models, where the selection equation is the structural or macro-level and the outcome equation is the individual or micro-level event data. The first column in Table 2 displays the replication of Chenoweth and Stephan (2011, Table 3.1) using a probit model. The second column displays the subset of the data for which the structural and campaign data overlap. This ensures that the models in columns 2 and 3 include the same set of observations in order to provide a proper comparison. The third column displays the results when the multi-level selection process is also modeled.

[Table 2 about here.]

Comparing the models in Table 2 demonstrates that structural factors appear to influence the likelihood of whether protests occur. First, several factors associated with government weakness—*population size*, *non-contiguous territory*, and *political instability*—encourage the likelihood of observing protest movements, while high levels of *democracy* are associated with fewer protests. The negative constant implies the likelihood of protest at any given time is small. Second, the negative coefficient on *rho*, which measures the correlation of the error terms between the two equations, suggests that the unobserved factors (e.g., protesters' strength relative to the government) that increase the probability of protests' success also decrease the likelihood of protest at any given time: e.g., a strong opposition is more likely to obtain concessions from the government and, hence, are less likely to take to the streets in protest in the first place (Pierskalla 2010). Taken together, these results indicate there is a selection effect in the data.

Once structural conditions are modeled, the type of protest does not exert a statistically significant influence in our model. The coefficient on *non-violent* protests is roughly the same as its standard error. However, the coefficient for *democracy* is now statistically significant

[14] See Fearon and Laitin (2003) or Gibler and Miller (2014) for a discussion of how the variables are measured.

at conventional levels, as including estimates for the structural factors influencing protests reduces the degree of uncertainty associated with the coefficient. It is also worth noting that the constant is positive and statistically significant once selection based on structure is modeled. This suggests that when protests are observed, regardless of other factors, they are likely to succeed. This finding has important substantive implications and is consistent with formal models that predict that protesters behave strategically (Chyzh and Labzina 2015; Pierskalla 2010; Ritter 2014).

4.2. *Structure, Civil Conflict, and Civilian Targeting*

Our second application explores the effect of structure-based selection on civilian targeting by rebel forces. Wood (2010) argues that rebel groups with stronger capabilities vis-à-vis the government can use a mix of selective incentives and repression to garner support and resources from the population. Weaker rebel groups, on the other hand, often lack the capacity to offer incentives to the population to garner support and instead rely to a greater degree on civilian targeting. The unit-of-analysis is the dyad-year, where the dyad consists of an insurgent group and the government.

We previously argued that outbreaks of civil conflict are non-random, and data on rebel groups can only be collected if civil conflicts are observed. These two points imply that observed rebel groups are likely to be more capable than the population of potential rebel groups (Chatagnier and Castelli 2016; Nieman 2015). The onset and continuation of civil conflict, moreover, is more likely to occur in weak states, anocracies, states with lootable resources, economic shocks, as well as a number of other factors (Cederman et al. 2011; Cunningham et al. 2009; Fearon and Laitin 2003; Nieman 2011; Ross 2004). Taken together, state factors affect the likelihood of civil conflict, which in turn likely affects the type of rebel groups that are observed and their interactions with the government. Thus, we expect that this structural selection effect influences rebel group behaviors, including the tactic of civil targeting.

We focus on Wood (2010, Table 2, Model 1) in our replication. Wood (2010) measures the count of *rebel-civilian one-sided killing* as the direct, intentional killing of civilians in non-combat situations by rebel forces (Eck and Hultman 2007).[15] *Rebel capability* is the ratio of troops to the scaled number of government troops (Uppsala Conflict Data Program 2009).[16] He also controls for *government violence* against the population, *identity conflicts*, *territorial conflict*, the overall degree of *conflict severity*, the *age* of the conflict, *democracy*, *GDP/capita*, and whether the conflict takes place during the *Cold War*.[17] We measure structural factors related to conflict using the same model as above but also add a lagged dependent variable to account for conflict duration (Fearon and Laitin 2003).

Table 3 reports the results of the analyses using a Poisson model[18] and a Poisson that accounts for selection (Miranda et al. 2004; Miranda and Rabe-Hesketh 2006). The first column reports the exact replication of Wood (2010) using a Poisson model. The second column reports the results of a Poisson model conditioned by a multi-level selection process.

[Table 3 about here.]

The estimates in Table 3 demonstrate that structural factors influence the micro-level interactions that take place within them. As was the case before, the *rho* parameter in column 2 is negative and statistically significant.[19] This indicates that the unobservable factors from the macro-level are negatively correlated with the unobservable micro-level factors that affect one-sided rebel-civilian killing. Thus, the same factors that lead an opposition to arm and fight the government also make them less likely to engage in one-sided civilian killing.

[15] The measure does not include indirect civilian deaths resulting from sieges, disease, collateral damage, or extrajudicial executions (Wood 2010, 606).

[16] The scaling of the measure accounts for the potential presence of multiple insurgencies in one state.

[17] See Wood (2010) for a discussion of how the control variables are measured.

[18] Wood (2010) uses a negative binomial model, rather than a Poisson model, in his analysis. We focus on a Poisson model for ease of simultaneously estimating the selection and outcome equations (Miranda et al. 2004; Miranda and Rabe-Hesketh 2006). Estimated coefficients and the significance levels associated with all theoretically relevant explanatory variables are unaffected by this change in model choice.

[19] The negative correlation of the macro- and micro-level errors is consistent with Gibler (2016), who found that structural conditions affect reporting of crisis events in narratives compiled by the International Conflict Group.

The structural factors also affect the substantive results in the dyadic analysis of rebel tactics. The coefficient on *rebel capability*, for instance, is significantly smaller when conditioned by structural selection. Figure 2 compares the substantive effects of the two models using predicted values (90% confidence intervals) from Monte Carlo simulations based on estimates from Table 3. The solid line displays predicted values using model 1 and the dashed line displays predicted values accounting for structural selection. The model that ignores selection identifies a steep, declining slope in civilian casualties as rebel capabilities increase, while the model that accounts for structural selection factors shows almost no decline in civilian casualties at all. Moving from a *rebel capacity* of 0.2 to 0.8 in the replication results in a decrease of civilian casualties of 94.9 to 65.33. Comparatively, moving from a *rebel capability* of 0.2 to 0.8 results in a change in civilian casualties from 29.4 to 27.5 once structural selection is accounted for. Substantively, this means that ignoring macro-level factors would lead one to significantly overestimate the degree to which rebel capacity reduces civilian killings by insurgent groups.

[Figure 2 about here.]

Finally, it is also worth noting that the sign on the coefficient for *identity conflict* changes from negative to positive, and is statistically significant in both models. Ignoring structural selection could lead one to incorrectly infer that identity conflicts are more likely to result in civilian targeting than non-identity conflicts.

5. CONCLUSION

Most of social science suggests there is a trade-off between thick detail of relatively few cases and thinly-informative operationalizations of important variables across a large number of cases. This need not be the case. Rich data on selected sets of cases provides important information for understanding agency, causal mechanisms, and making inferences, so long as potential selection effects are well understood. However, different levels of analysis and aggregation often prevent rich data usage.

We argue that selection models are a special type of multi-level model. We use both Monte Carlo analyses and empirical replications to demonstrate that model estimates are improved by accounting for the the non-random processes at the structural or macro-level that makes such events occur in the first place. Our empirical applications demonstrate that inferences from recent work on protest movements and civilian targeting during civil conflicts are likely to be incorrect. Non-violent protests are not more effective once the structural environment that influences the likelihood of protest is considered. Those same environmental factors also heavily influence the observation of rebel groups and their capacity for killing civilians. As we argue, accounting for structural selection issues improves estimates and associated inferences of causal variables and relationships which, in turn, enhances our theoretical understanding and increases the quality of policy prescriptions based on these theories.

REFERENCES

Berry, W. D., J. H. DeMeritt, and J. Esarey (2010). Testing for interactions in binary logit and probit models: Is a product term essential? *American Journal of Political Science 54*(1), 248–266.

Buhaug, H., S. Gates, and P. Lujala (2009). Geography, rebel capability, and the duration of civil conflict. *Journal of Conflict Resolution 53*(4), 544–569.

Cederman, L.-E., N. B. Weidmann, and K. S. Gleditsch (2011). Horizontal inequalities and ethnonationalist civil war: A global comparison. *American Political Science Review 105*(03), 478–495.

Chatagnier, J. T. and E. Castelli (2016). The arc of modernization: Economic structure, materialism, and the onset of civil conflict. *Political Science Research and Methods*.

Chenoweth, E. and M. J. Stephan (2011). *Why civil resistance works: The strategic logic of nonviolent conflict*. Columbia University Press.

Chyzh, O. and E. Labzina (2015). Bankrolling repression? modeling third-party influence on protests and repression. Presented at the 2015 meeting of the Peace Science Society (International) in Oxford, MS.

Collier, P. and A. Hoeffler (2002). On the incidence of civil war in africa. *Journal of Conflict Resolution 46*(1), 13–28.

Cunningham, D. E., K. S. Gleditsch, and I. Salehyan (2009). It takes two: A dyadic analsys of civil war duration and outcome. *Journal of Conflict Resolution 53*(4), 570–597.

Cunningham, K. G. (2011). Divide and conquer or divide and concede: How do states respond to internally divided separatists? *American Political Science Review 105*(2), 275–297.

Cunningham, K. G. (2013). Actor fragmentation and civil war bargaining: How internal divisions generate civil conflict. *American Journal of Political Science 57*(3), 659–672.

Davenport, C. (2007). State repression and the tyrannical peace. *Journal of Peace Research 44*(4), 485–504.

Eck, K. and L. Hultman (2007). One-sided violence against civilians in war insights from new fatality data. *Journal of Peace Research 44*(2), 233–246.

Fearon, J. D. and D. D. Laitin (2003). Ethnicity, insurgency, and civil war. *American Political Science Review 97*(01), 75–90.

Freedman, D. A. and J. S. Sekhon (2010). Endogeneity in probit response models. *Political Analysis 18*(2), 138–150.

Gelman, A. and J. Hill (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gibler, D. M. (2016). Combining behavioral and structural predictors of violent civil conflict: Getting scholars and policymakers to talk. *International Studies Quarterly*.

Gibler, D. M. and S. V. Miller (2014). External territorial threat, state capacity, and civil war. *Journal of Peace Research 51*(5), 634–646.

Gill, J. (2009). *Bayesian Methods: A Social and Behavior Sciences Approach* (Second ed.). Boca Raton, FL: Chapman and Hall/CRC Press.

Greene, W. (2006). A general approach to incorporating selectivity in a model. Working paper. url: https://archive.nyu.edu/bitstream/2451/26072/2/6-10.pdf.

Greene, W. (2010). A stochastic frontier model with correction for sample selection. *Journal of Productivity Analysis 34*(1), 15–24.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement 5*(4), 475–492.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica 46*(4), 931–959.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Hill, D. W. and Z. M. Jones (2014). An empirical evaluation of explanations for state repression. *American Political Science Review 108*(03), 661–687.

LaFree, G. and L. Dugan (2007). Introducing the global terrorism database. *Terrorism and Political Violence 19*(2), 181–204.

Lemke, D. and W. Reed (2001). The relevance of politically relevant dyads. *Journal of Conflict Resolution 45*(1), 126–144.

Maoz, Z. and B. Russett (1993). Normative and structural causes of the democractic peace, 1946-1986. *American Political Science Review 87*(3), 624–638.

Miranda, A. et al. (2004). Fiml estimation of an endogenous switching model for count data. *Stata Journal 4*(1), 40–49.

Miranda, A. and S. Rabe-Hesketh (2006). Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. *Stata Journal 6*(3).

Nieman, M. D. (2011). Shocks and turbulence: Globalization and the occurrence of civil war. *International Interactions 37*(3), 263–292.

Nieman, M. D. (2015). Statistical analysis of strategic interaction with unobserved player actions: Introducing a strategic probit with partial observability. *Political Analysis 23*(3), 429–448.

Palmer, G., V. d'Orazio, M. Kenwick, and M. Lane (2015). The mid4 dataset, 2002–2010: Procedures, coding rules and description. *Conflict Management and Peace Science 32*(2), 222–242.

Pierskalla, J. H. (2010). Protest, deterrence, and escalation: The strategic calculus of government repression. *Journal of Conflict Resolution 54*(1), 117–145.

Powell, J. M. and C. L. Thyne (2011). Global instances of coups from 1950 to 2010 a new dataset. *Journal of Peace Research 48*(2), 249–259.

Regan, P. M. and D. Norton (2005). Greed, grievance, and mobilization in civil wars. *Journal of Conflict Resolution 49*(3), 319–336.

Ritter, E. H. (2014). Policy disputes, political survival, and the onset and severity of state repression. *Journal of Conflict Resolution 58*(1), 143–168.

Ritter, E. H. and C. R. Conrad (2016). Preventing and responding to dissent: The observational challenges of explaining strategic repression. *American Political Science Review 110*(1), 85–99.

Ross, M. L. (2004). How do natural resources influence civil war? evidence from thirteen cases. *International Organization 58*(1), 35–67.

Russett, B. and J. Oneal (2001). *Triangulating Peace: Democracy, Interdependence, and International Organization*. New York: W.W. Norton.

Salehyan, I., C. S. Hendrix, J. Hamner, C. Case, C. Linebarger, E. Stull, and J. Williams (2012). Social conflict in africa: A new database. *International Interactions 38*(4), 503–511.

Sartori, A. E. (2003). An estimator for some binary-outcome selection models without exclusion restrictions. *Political Analysis 11*(2), 111–138.

Siverson, R. M. and H. Starr (1991). *The Diffusion of War: A Study of Opportunity and Willingness*. University of Michigan Press.

Steenbergen, M. R. and B. S. Jones (2002). Modeling multilevel data structures. *American Journal of Political Science 46*(1), 218–237.

Terza, J. V. (1998). Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics 84*(1), 129–154.

Uppsala Conflict Data Program (2009). www.ucdp.uu.se/database.

Weede, E. (1976). Overwhelming preponderance as a pacifying condition among contiguous asian dyads, 1950–1969. *Journal of Conflict Resolution 20*(3), 395–411.

Williams, L. K. (2016). Long-term effects in models with temporal dependence. *Political Analysis 24*(2), 243–262.

Winship, C. and R. D. Mare (1992). Models for sample selection bias. *Annual Review of Sociology 18*, 327–350.

Wood, R. M. (2010). Rebel capability and strategic violence against civilians. *Journal of Peace Research 47*(5), 601–614.

Xiang, J. (2010). Relevance as a latent variable in dyadic analysis of conflict. *Journal of Politics 72*(2), 484–498.

Table 1: Root Mean Squared Errors with Varying Degrees of Error Correlation Between Levels.

	Rho = −0.7					Rho = −0.4					
Var.	Probit	Probit w/ Group	RE	RE Group	Heckman Probit	Var.	Probit	Probit w/ Group	RE	RE Group	Heckman Probit
X	0.259	0.284	0.475	0.474	0.273	X	0.201	0.205	0.226	0.225	0.209
Z	—	0.591	—	0.494	0.470	Z	—	0.788	—	0.777	0.462
Const.	0.494	0.938	0.497	1.056	0.375	Const.	0.331	0.545	0.324	0.566	0.313

	Rho = 0.4					Rho = 0.7					
Var.	Probit	Probit w/ Group	RE	RE Group	Heckman Probit	Var.	Probit	Probit w/ Group	RE	RE Group	Heckman Probit
X	0.208	0.212	0.234	0.234	0.220	X	0.266	0.294	0.463	0.463	0.285
Z	—	1.256	—	1.268	0.456	Z	—	1.498	—	1.580	0.439
Const.	0.355	0.592	0.396	0.651	0.350	Const.	0.574	1.101	0.817	1.401	0.429

Table 2: Probit Estimation of Protest Movement Outcomes and State Structure.

Variable	Replication	Subsample	Structure-Selection
Protest Success			
Non-violent	0.548*	0.463†	0.189
	(0.290)	(0.321)	(0.168)
Democracy	0.031†	0.027†	0.022**
	(0.019)	(0.020)	(0.009)
Members	0.229**	0.221**	0.118**
	(0.076)	(0.084)	(0.053)
Population	-0.262**	-0.295**	-0.250**
	(0.104)	(0.115)	(0.071)
Constant	-0.102	0.426	3.384**
	(0.952)	(1.052)	(0.537)
Domestic Protest			
GDP/capita			0.016
			(0.043)
Population			0.151**
			(0.021)
Mountains			0.007
			(0.026)
Non-contiguous			0.249*
			(0.135)
Oil exporter			-0.170†
			(0.113)
Democracy			-0.227**
			(0.070)
Democracy^2			-0.112**
			(0.056)
Instability			0.308**
			(0.095)
Ethnic Frac			0.079
			(0.145)
Religious Frac			0.060
			(0.132)
Constant			-3.813**
			(0.465)
Rho			-0.930**
			(0.078)
Log-likelihood	-79.88	-66.11	-616.36
Observations	141	115	7883 (115)

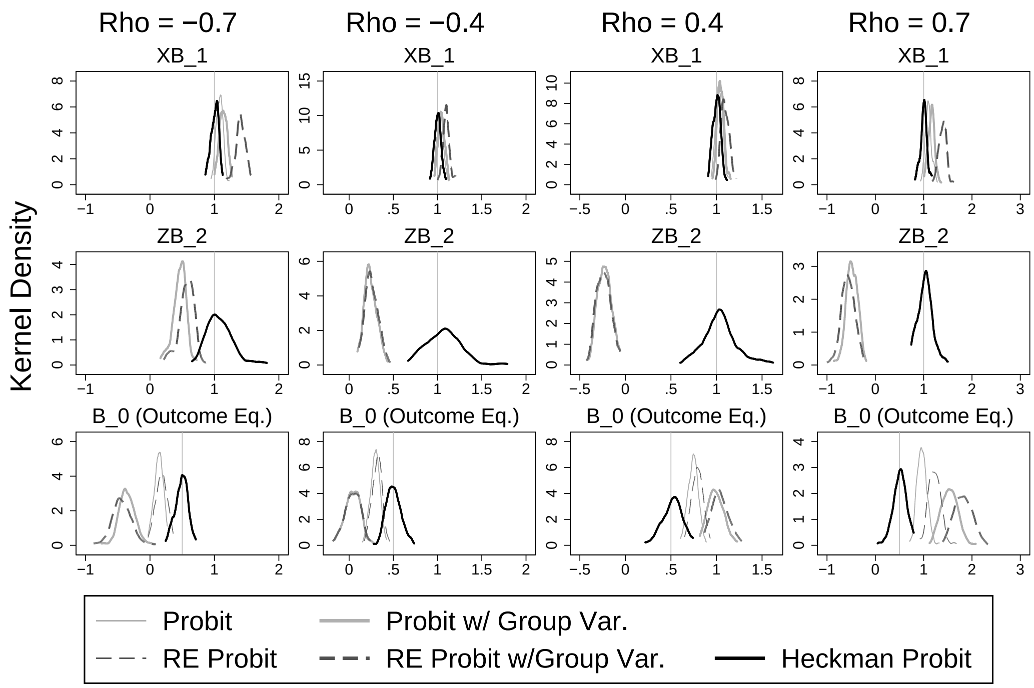
$^{**}p < 0.05$, $^{*}p < 0.10$ two-tailed, $^{\dagger}p < 0.10$ one-tailed. Robust standard errors in parentheses. The number under observations parentheses in the structure-selection model are uncensored cases.

Table 3: Poisson Estimation of Rebel One-sided Civilian Killing and State Structure.

Variable	Replication	Structure-Selection
Rebel Civilian Killing		
Rebel capacity	-0.585**	-0.075**
	(0.230)	(0.035)
Government violence	0.001**	0.001**
	(0.001)	(0.001)
Identity conflict	0.866*	-0.756**
	(0.286)	(0.115)
Territorial conflict	-0.938*	-0.565**
	(0.274)	(0.080)
Conflict severity	0.303**	0.521**
	(0.076)	(0.022)
Age	-0.183	-0.510**
	(0.285)	(0.030)
Democracy	0.046**	0.031**
	(0.021)	(0.008)
GDP/capita	-0.206	-0.131**
	(0.347)	(0.036)
Cold War	-0.762	-0.963**
	(0.522)	(0.079)
Constant	3.830	3.011**
	(3.126)	(0.337)
Civil Conflict		
GDP/capita		-0.142**
		(0.054)
Population		0.026
		(0.029)
Mountains		0.080**
		(0.033)
Non-contiguous		-0.258†
		(0.175)
Oil exporter		0.153
		(0.149)
Democracy		-0.616**
		(0.149)
Democracy^2		-0.587**
		(0.139)
Instability		0.175†
		(0.132)
Ethnic Frac		0.374*
		(0.220)
Religious Frac		-0.421**
		(0.196)
Prior Conflict		2.340**
		(0.101)
Constant		-1.129*
		(0.593)
Rho		-0.305**
		(0.037)
Log-likelihood	-66091.33	-6190.31
Observations	609	3293(609)

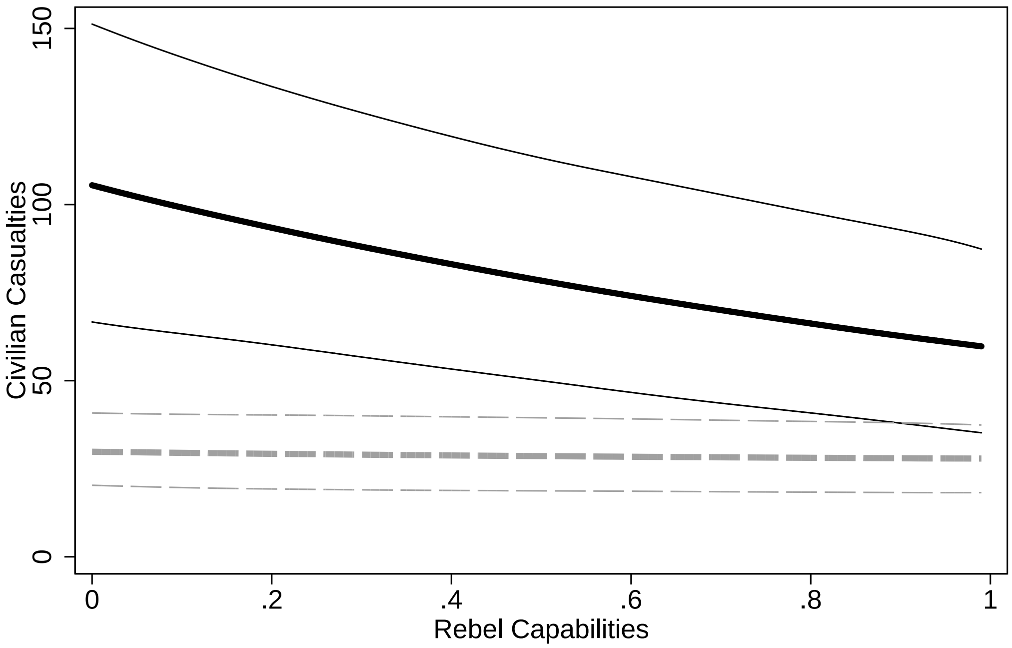
$^{**}p < 0.05$, $^{*}p < 0.10$ two-tailed, $^{\dagger}p < 0.10$ one-tailed. Robust standard errors in parentheses. The number under observations parentheses in the structure-selection model are uncensored cases.

Figure 1: Monte Carlos with Varying Degrees of Error Correlation Between Levels.



Note: Vertical line represents the true value of the coefficient. Results of 100 simulations with 5000 observations (100 group−level observations with 50 individual−level observations each.

Figure 2: Substantive Effects of Changing Rebel Capacity on Civilian Targeting.



Note: The solid line displays predicted values using the replication of Wood (2010). The dashed line displays predicted values accounting for structural selection. Predicted values (90% confidence intervals) are generated using Monte Carlosimulations, based on the estimates from Table 3.