Forecasting Civil Conflict with Zero-Inflated Count Models

Benjamin E. Bagozzi
The Pennsylvania State University
Department of Political Science

June 12, 2012

Abstract

This article presents a number of useful forecasting techniques for the evaluation of political violence *counts*. To demonstrate these techniques, negative binomial (NB) and zero-inflated negative binomial (ZINB) count models are applied to a recently developed country-month event-count dataset of rebel and government-initiated violent conflicts. Using these count-forecasting methods, we compare model predictions of the conflicts occurring within these data, and find that moving from a NB model to a zero-inflated count model can produce up-to an 13% improvement in civil conflict forecasting accuracy. We also find that including (1-3 month) lagged values of monthly conflict frequency in the inflation stage of our zero-inflated conflict models can lead to as much as a 12% improvement in conflict forecasting accuracy. Substantively our findings suggest that, while past values of government and rebel initiated conflict are indeed positively related to present values, the magnitude of this positive relationship tends to be overstated when zero-inflation is not accounted for.

Introduction

Political scientists have recently demonstrated the importance of *prediction* to the advancement of our theoretical and practical understandings of civil conflict (Ward, Greenhill and Bakke 2010; Weidmann and Toft 2010; Brandt, Freeman and Schrodt 2011). However, across the sciences, statistical forecasting tools are almost exclusively designed for either binary or continuous dependent variables (Czado, Gneiting and Held 2009). This limits our ability to forecast intrastate conflicts, since across most levels of temporal and spatial aggregation, "civil conflict" is best operationalized through intermediary levels of measurement such as counts, durations, and discrete (un)ordered outcomes with more than two categories.[1] Nevertheless, scholars interested in predicting conflict have favored dichotomous dependent variables over these richer measures of civil conflict, due (in part) to the forecasting limitations mentioned above. When applied to graduated social-science variables (such as civil conflict), this practice of dichotomization discards relevant information (i.e. variance) and may exacerbate existing measurement errors within one's variable of interest (Cohen 1983). As a consequence, scholars report artificial dichotomization to be detrimental to both prediction and theory testing (MacCallum et al. 2002; Royston, Altman and Sauerbrei 2006).

To address these problems, our study builds upon recent statistical advances in probabilistic count-data forecasting (Gneiting, Balabdaoui and Raftery 2007; Czado, Gneiting and Held 2009) to present the first comprehensive forecasting analysis of civil conflict *frequency*. In doing so, this article introduces a number of useful statistical tools for the evaluation, refinement, and presentation of conflict-event count forecasts. We then demonstrate with these tools that—when used correctly—count models can produce compelling levels of calibration and sharpness in civil conflict predictions. Specifically, we find that by leveraging count models' split-population modeling capabilities in a manner that statistically accounts for the presence of excess (i.e. structural) zeroes within civil conflict data, one can further increase conflict-count forecasting accuracy by as much as 13%.[2] Herein, we take a novel approach, and include past levels of (government and

[1] See, e.g., Hegre, Ostby and Raleigh (2009) for a count measure of civil conflict, Fearon (2004) for a duration measure of civil conflict, Hill et al. (2011) for a discrete ordered measure of civil conflict, and Buhaug (2006) for a discrete unordered measure of civil conflict.

[2] On average, for both government and rebel initiated conflicts, when using an "at-least one monthly conflict"

rebel initiated) material conflict in the inflation stage of our forecasting count-models. It is argued and shown below that doing so helps to account for the adverse effects of structural zeroes on our abilities to forecast civil conflict processes. Specifically, we find that the addition of 1-3-month lagged measures of civil conflict to the *inflation stages* of our forecasting models increases our ability to predict actual instances of out-of-sample monthly conflict by as much as 12%.[3] While the benefits of zero inflated models are widely known among conflict scholars, this paper thereby provides—for the first time— collection of methods that allow conflict researchers to assess the actual magnitude of these benefits with regards to conflict *prediction*, which in turn can help to identify problems of model misspecification in (especially low sample-size) zero inflated conflict applications.

This article proceeds as follows. In the next section we briefly discuss the prevalence of zero-inflation in conflict data, outline the benefits of addressing this problem with zero-inflated models, and present a rational for the inclusion of lagged conflict measures as inflation stage covariates. We then introduce a newly developed event dataset of monthly civil-conflicts, justify our choice of zero inflated count model, and apply this model to a training set of these monthly conflict-count-data. The heart of our analysis section then uses our training dataset, a validation dataset, in-sample and out-of-sample predictions, classification matrices, marginal calibration diagrams, and sensitivity plots to demonstrate that accounting for zero inflation with past levels of conflict can substantially increase the accuracy and precision of one's (already commensurate) civil conflict forecasts. Finally, we conclude by discussing the implications of our findings for those interested in the modeling and forecasting of civil conflict events.

Theoretical Motivation

Yearly, monthly, and weekly aggregations of militarized conflict—whether measured at the dyad, country, or sub-country level—are often "inflated" with structural zeroes (Clark and Regan 2003; Pevehouse 2004; Hill et al. 2011). These zeroes represent peace-observations that would

threshold and our out-of-sample forecasts.

[3]On average, for government and rebel initiated conflicts, when using an "at least one monthly conflict" threshold.

likely *never* experience conflict under any realistic circumstances. For instance, within dyad-year studies of interstate war, pairs of countries such as Switzerland and Costa Rica (i.e. "irrelevant dyads") have consistently been considered to be structural zeroes[4] since such dyads could *never* go to war with one another due to their geographic distance and limited military capabilities. Treating these cases as "peace-zeroes" within a statistical model of conflict can lead to biased inferences because such cases effectively have zero probability of *ever* experiencing an event of interest (Lemke and Reed 2001; Clark and Regan 2003; Xiang 2010). On the other hand, truncating all potential structural (peace-year) zeroes from one's sample excludes a significant proportion of relevant-conflict observations (Bennett and Stam 2004, 61) and produces selection bias (Lemke and Reed 2001; Xiang 2010). As an alternative to these to approaches, scholars have begun to recognize that—by (1) including *all* observations in one's analysis and (2) then accounting for the likelihood of zero-inflation among peace observations probabilistically—one can address the challenges created by structural zeroes in an unbiased fashion (e.g., Clark and Regan 2003; Benini and Moulton 2004; Pevehouse 2004; Xiang 2010). In essence, this approach allows one to use ex-ante observable and theoretically informed covariates to account for the probability that a given zero observation is structural, and to then probabilistically discount these structural zeroes' leverage within one's primary analysis; without dropping these observations entirely.

The zero-inflated technique described above has proven to be especially useful to studies of terrorism and civil conflict (e.g., Hultman 2007; Hegre, Ostby and Raleigh 2009; Piazza 2011). For example, in a department-month study of human rights violations committed by the Revolutionary Armed Forces of Colombia (FARC), Holmes, Pineres and Curtina (2007) find that there were many department-months in their sample of all Colombian departments wherein the FARC was wholly inactive. The authors accordingly account for these structural zero-observations with a zero inflated count model, since the FARC was likely incapable of committing *any* number of human rights violations greater than zero in departments where it was not active, and find that doing so yields valuable theoretical and statistical insights into the underlying dynamics of civil conflict onset and

[4] See, e.g., Maoz and Russett (1993) for a discussion of "relevant dyads".

intensity. At the country-year level, advanced industrialized polities have similarly been shown to engender a non-negligible quantity of structural-zeroes within ordinal variables of government repression and civil war (Hill et al. 2011). As above, truncating all such advanced-industrialized countries from one's civil conflict analysis is likely to produce selection bias and exclude a non-negligible number of actual (and potential) instances of civil conflict. Indeed, even within advanced industrial democracies, minority groups occasionally rebel gainst the state (Gurr 1993), and home-grown terrorist attacks can occur. For those interested in producing accurate and comprehensive civil conflict *forecasts*, these are very costly cases to miss. We therefore propose that, when faced with the potential of excess zeroes in a civil conflict forecasting application, analysts can make the most of conflict forecasts by (1) including all zero-observations in the forecasting model and (2) accounting for any resultant zero-inflation econometrically:

- *Hypothesis 1: Modeling conflict frequency with zero inflated count models will improve the accuracy of civil conflict count forecasts.*

However, the advantages of zero-inflated models are dependent upon one's inflation stage specification. That is, given the existence of structural zeroes, a zero inflated model's ability to reduce the adverse effects of zero inflation on one's outcome stage analysis is contingent upon the degree to which these models' inflation stages accurately distinguish between structural zeroes (in our case the "always-zero" observations) and count-stage zeroes (e.g., peace-years that could potentially experience conflict under different circumstances). Due to the inherent rarity of conflict in space and time, this prerequisite represents an especially acute challenge for applications of zero-inflated models to studies of conflict. Indeed, while a great many covariates do have statistically significant relationships with inter and intra-state conflict, each variable therein tends to explain only a small amount of the actual variation in conflict onset and escalation (Beck, King and Zeng 2000; Bennett and Stam 2004; Ward, Greenhill and Bakke 2010). As consequence, many of the most well-known correlates of civil conflict have been shown to offer only a negligible—and at times negative—level of improvement in actual conflict forecasting accuracy (Ward, Greenhill and

Bakke 2010). This deficiency limits our ability to effectively use such covariates, where theoretically appropriate, in the inflation stages of zero inflated, civil conflict count models.

One exception, however, in terms of both explanatory and forecasting power, is an observation's past levels of conflict. Such lagged conflict values have been consistently identified as being among the largest and most robust predictors of inter and intra-state conflict (e.g., Lichbach and Gurr 1981; Gurr and Lichbach 1986; O'Brien 2002, 2010). Indeed, conflict researchers have found that inter and intra-state conflicts exhibit a strong temporal dependence (Beck, Katz and Tucker 1998; Weidmann and Ward 2010) and that conflict forecasts can be vastly improved by the inclusion of a series of temporally lagged values of past conflict (Pevehouse and Goldstein 1999; Shellman, Hatfield and Mills 2010). We argue here that these past levels of conflict (or lack thereof) not only directly affect subsequent levels of civil conflict in a reciprocal or inertial sense,[5] but also help to inform us, ex-ante, as to which countries are currently *able* to experience *any* level of domestic conflict. In this manner, we can improve both our conflict forecasting accuracy—and our understanding of conflict processes—by including lagged conflict measures as *inflation stage covariates*.

Specifically, we contend that zero-inflated peace-observations not only arise cross-sectionally,[6] but also evolve (and devolve) temporally, even within conflict-prone states. As the above discussion of zero-inflated conflict studies elucidated, it is likely that many civil-conflict "peace-observations" are *cross-sectional* structural-zeroes, representing (for example) advanced developed democracies whose probability of experiencing *any* rebel or government initiated domestic material conflict under reasonable circumstances is effectively zero for all time periods. However, even among conflict-prone countries, un-observed, secret, or informal truces frequently arise between government and rebel forces due to (for example) concerns over extremist factions sabotaging peace negotiations (Kydd and Walter 2002; Wanis-St. John 2006), tit-for-tat dynamics (Axelrod 1984), or environmental and social pressures such as mediators, religious observances, or seasonal harvests.

[5]See, for example, Gurr (1970); Francisco (1995) for theories of reciprocal hostility; and Goldstein and Freeman (1990) for a 'policy-inertia' theory of conflict.

[6]E.g., as a result of geography or slow moving variables such as institutions and GDP.

Such temporary stalemates may be unobservable to any actors other than the two sides involved, or may be common knowledge that—due to resource constraints—cannot be archived and coded for all observations. A recent example of such a phenomenon can be found in 2006 media reports of a "secret truce" between British troops and Taliban forces in southern Afghanistan, where after months of heavy fighting, both sides agreed to pull out of the town of Musa Qula; resulting in a temporary peaceful stalemate in the area.[7] Scholars have identified comparable instances of secret truces or defacto stalemates in civil conflict arenas as varied as the Russian Revolution (Wandycz 1965), the Irish Confederate Wars (Lehinan 2002, 73), and the El Salvador civil war (Wood 2003). Similar to the aforementioned (and time-invariant) low conflict propensities within advanced industrialized states, self-enforcing truce-periods of this sort are marked by unobserved characteristics that disproportionately preclude domestic actors from initiating any level of conflict greater than zero. In this sense, the recent occurrence (and levels) of civil conflict serve as an informative, time varying proxy for the broader-array of unobservable factors that often preclude a given observation from ever experiencing conflict. Hence,:

- *Hypothesis 2: Past levels of civil conflict serve as significant and robust predictors of zero inflation within zero inflated conflict-count models.*

Analysis

To evaluate these hypotheses, this paper uses a newly developed, Integrated Conflict Early Warning System (ICEWS) event-dataset to forecast the monthly frequencies of domestic civil conflict events within 29 Asian countries for the years 1997-2010 (O'Brien 2010).[8] These data are part of a Defense Advanced Research Project Agency (DARPA) funded project which has recently created a dataset of over 2-million machine-coded daily events occurring between relevant actors within the Asia-Pacific region. To machine code these events, the ICEWS project utilized news articles from over 75 electronic regional and international news sources. The coding of these news

[7] "British Troops in Secret Truce with Taliban," *The Sunday Times* 2006.

[8] The 29 countries included in the analysis are listed in the Supplemental Appendix and encompass all Asian and Oceanic polities with a population above one million.

stories was then undertaken by the Penn State Event Data Project's TABARI (Text Analysis By Augmented Replacement Instructions) software program (Schrodt 2009) and a Lockheed-Martin-developed java variant of TABARI known as JABARI. Specifically, TABARI and JABARI used sparse parsing and pattern recognition techniques to machine-coded millions of news stories from the aforementioned news sources for daily political events based primarily on a categorical coding scheme developed by the Conflict and Mediation Event Observation (CAMEO) project (Schrodt and Yilmaz 2007; Schrodt, Gerner and Yilmaz 2009). The resultant ICEWS events-dataset has recently been characterized as being "the most accurate event dataset currently available" (D'Orazio, Yonamine and Schrodt 2011, 4).

For our analysis, these raw ICEWS-coded events data were aggregated to the country-month level (*it*) for two specific domestic-actors of interest: government-actors and violent-rebel-actors. In doing so, we created two specific dependent conflict-variables. The first dependent variable is $government\ conflict_{it}$, which is a monthly count of government-actor[9] initiated, domestic material (i.e. physical, rather than verbal) conflicts targeting violent rebel-actors operating within a government's own country. The second dependent count variable is $rebel\ conflict_{it}$, which aggregates monthly counts of violent rebel-actor[10] initiated material conflicts targeting government actors within a rebel's country of origin. We choose to disaggregate conflict measures separately into government *and* rebel-actor initiated conflicts because recent findings suggest that a failure to do so increases the risk of Type I and II errors in studies of intrastate conflict (Shellman, Hatfield and Mills 2010). To create these two variables, daily ICEWS-coded events were first collapsed into daily counts of *government-actor→rebel-actor* and *rebel-actor→government-actor* country-level material conflicts. The resulting country-day event-counts for government material conflict and rebel material conflict were then aggregated to the monthly-count level for use as independent and dependent (monthly) count-variables below. Each of our dependent variables have 5,040 observations across our entire 1997-2010 sample period. Frequency histograms for for *government*

[9]Government members, including members of governing parties and coalition partners; military troops, soldiers, all state-military personnel; and police forces and officers were all considered to be "government actors".

[10]Domestic rebels (armed and violent groups and individuals), insurgents, and separatist groups were all considered to be "violent rebel actors".

and *rebel* $conflict_{it}$ are presented in Figure 1, and indicate that the ranges of these variables are $[0-98]$ and $[0-126]$ conflicts per-month, respectively.

[*Insert Figure 1 about here*]

Model Selection

Given the event-count nature of our two dependent variables, we identify several suitable count models (and associated distributions) for the forecasting of our events of interest. To this end, we first considered using a set of Poisson count models. However, the histograms presented above suggest that our *government* $conflict_{it}$ and *rebel* $conflict_{it}$ count distributions contain both an excess number of zero counts (i.e. "peace-country-months") and a right-skewed series of relatively high count values. Together these traits suggest that each dependent count variable exhibits high degrees of overdispersion and positive contagion. This is confirmed by examining the standard deviations of *government* $conflict_{it}$ and *rebel* $conflict_{it}$, which with values of 6.82 and 6.10, are significantly larger than these variables' respective means of 1.89 and 1.81. Conditional overdispersion,[11] if present, would violate a Poisson model's mean-variance equality assumption, which would thereby undermine the Poisson model's applicability in estimating and forecasting the event counts described above. Accordingly, the negative binomial (NB) model is favored as a baseline forecasting model below, as it accounts for conditional overdispersion by through a parameterized relaxation of the mean-variance equality assumption. However, as argued above, there is also strong reason to believe that many of the excess zeroes observed within our dependent variables are not true count-level zeroes, in the sense that they could ever take on values greater than zero. Rather, it is likely that many of these "peace-months" are structural-zeroes, representing cases such as Japan or New Zealand, whose probability of experiencing *any* rebel or government initiated domestic material conflict, under reasonable circumstances, is effectively zero. Even within traditionally conflict prone countries, un-observed, secret, or informal truces could arise between government and rebel forces. Similar to the time-invariant peace-qualities of Japan and New Zealand, unobserved truce-months—that disproportionately preclude domestic actors from

[11] That is, the persistence of count-overdispersion once one has conditioned on all covariates.

initiating any level of conflict greater than zero—likely arise within many of the (more conflict-prone) countries under analysis. If such phenomena do exist, then these cases would engender a second, *time varying* form of structural zeroes within our sample of interest.

Ignoring either form of structural zeroes, and treating such observations as count stage zeroes in an NB count model, can bias one's coefficient estimates and standard errors, whereas an ad-hoc removal of all potentially-inflated zeroes from one's sample likely discards relevant conflict-observations and produces selection bias. In order to avoid these biases, structural zeroes must be accounted for statistically through the use of a zero inflated Poisson (ZIP) or zero inflated negative binomial (ZINB) model. The ZIP and ZINB models specifically allow one to explicitly model and test for the presence of inflated zeroes through likelihood functions which combine the results from a binary equation—estimating whether a zero observation is more likely to have come from the zero-only or count-stage d.g.p—with the results of a NB or Poisson likelihood equation that directly tests for the effect of one's covariates on the expected frequency of $government\ conflict_{it}$ or $rebel\ conflict_{it}$, conditional on the likelihood that a given observation was generated from the count-stage d.g.p. Accordingly, we expect that ZIP and ZINB models will be superior to Poisson and NB models for the modeling and forecasting of $government\ conflict_{it}$ and $rebel\ conflict_{it}$. Furthermore, due to the aforementioned presence of many extreme (high-count) values in $government\ conflict_{it}$ and $rebel\ conflict_{it}$, conditional overdispersion was believed to be persistent in our dependent variables and models, even after accounting for the zero-inflation described above. We therefore favored the ZINB model over the ZIP model for all zero-inflated forecasts discussed below.

A variety of model selection statistics confirmed our suspicions. Vuong comparison tests for non-nested models (Vuong 1989) are the most appropriate comparison tests for our models of interest, and these tests were accordingly used to compare ZIP, ZINB, NB, and Poisson models for all model specifications presented below. Vuong tests indicated that across all specifications, the ZINB model outperforms the ZIP, Poisson, and NB models at the $p < .01$ level, while the NB model outperforms the ZIP and Poisson models at the $p < .01$ level. Likelihood ratio tests were also conducted where applicable, and similarly suggested both that the ZINB model is superior to the ZIP for our dependent variables of interest, and that the NB model was superior to the Poisson model across all models compared. Standard information based model selection criteria are also prominently featured in comparisons of count and zero inflated models and therefore are applicable here (Harris and Zhao 2007; Czado, Gneiting and Held 2009). Accordingly, Akaike information criterion (AIC) comparisons were calculated and compared for all models used below, with each comparison therein preferring the NB and ZINB models to comparable Poisson and ZIP models, as well as preferring our ZINB models over our NB models. To summarize, the zero-inflated, overdispersed nature of our dependent variables suggests that count models of the NB and ZINB variety should used for the modeling and forecasting of $government\ conflict_{it}$ and $rebel\ conflict_{it}$, and NB and ZINB models are therefore the statistical forecasting models that we estimate and evaluate fin our analyses.

Independent Variables

The primary independent variables used for forecasting $government\ conflict_{it}$ and $rebel\ conflict_{it}$ are past monthly counts of material domestic conflict. The use of lagged conflict-count measures as predictors within conflict forecasting models has become common in the field, in part due to the challenges associated with the scaling of conflict-cooperation scored events.[12] For the study at hand; one, two, and three month lags of rebel initiated material conflict and government initiated material conflict were included in the forecasting models of both our $government\ conflict_{it}$ and $rebel\ conflict_{it}$ dependent variables. The natural log of each lagged conflict variable (+0.5) was then taken prior to its inclusion on the right hand side of our models in order to ensure that outliers were not disproportionately influencing the analysis.[13] For our ZINB models, all independent lagged-conflict variables were then included within both the zero inflation stage and count stage estimating equations. As argued above, the justification for using these lagged covariates within our inflation stage rests on the contention that recent levels of monthly conflict (or lack thereof)

[12] See (e.g., D'Orazio, Yonamine and Schrodt 2011).

[13] Logging the independent variables did not dramatically affect our results, although it did moderately improve the calibration of the NB forecasts for all time periods examined (the ZINB forecasts remained relatively unchanged).

directly inform us, with ex-ante observability, as to which country-months are currently *able* to experience domestic conflict, and which are not. If correct, this strategy will allow us to statistically partition our (potentially) 'inflated zero' cases from the true 'count-zero' conflict cases, and to thereby improve the accuracy and precision of our count stage estimates and our conflict forecasts.

Drawing on recent civil conflict studies, a limited number of control variables are also included in the models reported below. In the count stages of our NB and ZINB models, we include yearly measures of the natural log of GDP per capita, GDP growth, and the natural log of a country's total population,[14] as—unlike many commonly studied correlates of intrastate conflict—these three variables have been found to enhance our ability to predict civil war (Ward, Greenhill and Bakke 2010). GDP per capita has also been found to be a strong predictor of a country's likelihood of *ever* experiencing domestic political violence (Hill et al. 2011), and accordingly, we also include $ln\ GDP_{pc}$ within the inflation stage of our ZINB models. As robustness tests, we re-ran all models discussed below (i) without $ln\ GDP_{pc}$, *GDP growth*, or *ln population* (i.e. with only our lagged conflict measures included as covariates) and (ii) with a range of additional controls added to each model.[15] Our findings and conclusions remain unchanged under these alternative specifications. Finally, we also explored the inclusion of additional monthly conflict lags as independent variables in our NB and ZINB models. We find that including lagged conflict measures beyond 3-month lags as independent variables generally does not improve the forecasting accuracy of our models, and in some cases reduces accuracy. Thus, we choose not to include conflict measures beyond 3-month lags in the models reported directly below.

Estimation Model Results

ZINB and NB models of $government\ conflict_{it}$ and $rebel\ conflict_{it}$ are estimated with 1-3 month lags of $ln\ government\ conflict_{it}$ and $ln\ rebel\ conflict_{it}$ included as our key predictors. All models are estimated on a training dataset encompassing the 1997-2004 country-month sample, with the aim of evaluating the forecasting accuracy of these model-estimates on a country-month

[14]These measures are taken from the World Bank's World Development Indicators (World Bank 2011).

[15]Additional controls included monthly counts of verbal (government and rebel) conflict events, monthly counts of verbal (government and rebel) cooperative events, the natural log of GDP, and the natural log of unemployment.

validation dataset encompassing the years 2005-2010. The 1997-2004 training models thus serve as our primary models of reference here. Comparable ZINB and NB models were also estimated on the entire 1997-2010 dataset in order to evaluate whether any discrepancies existed in probability distributions across the 1997-2004 training sample and 2005-2010 validation sample, and no major discrepancies were found.[16] Coefficient estimates, standard errors, and goodness-of-fit statistics for the training models of $government\ conflict_{it}$ and $rebel\ conflict_{it}$ are presented in Table A of the Supplemental Appendix.

Several initial conclusions can be drawn from these model estimates and test statistics. The ZINB and NB government-conflict models suggest that past values of government and rebel initiated material conflict are positively associated with current monthly frequencies of government initiated conflicts, although the NB model tends to overestimate the magnitude and precision of these relationships. Lagged measures of rebel and government conflict are generally significant (negative) determinants of inflated observations in our $government\ conflict_{it}$ ZINB models. The findings are similar for $rebel\ conflict_{it}$, and again the NB model appears to overestimate our count-stage coefficient estimates and standard errors, which is consistent with our expectations of zero inflation, as well as with the results reported for the Vuong tests and AICs above. Here, for the rebel-conflict models in Table A, higher (lower) past levels of government and rebel initiated conflict are again generally associated with higher (lower) current levels of rebel initiated conflict at statistically significant levels, while the ZINB inflation stages suggest that increases in past levels of government and rebel initiated civil conflict generally decrease the probability that a peace observation is from the "zero-only" d.g.p. and increase that observation's likelihood of coming from the conflict-count d.g.p.

Classification Matrices

To better evaluate the relative performances of our NB and ZINB models in terms of conflict *forecasting*, we next present a set of classification matrices for our government and rebel conflict dependent variables. To create these matrices, we began by calculating in-sample and out-of-

[16] These full models are available upon request.

sample NB model predictions for our $government\ conflict_{it}$ and $rebel\ conflict_{it}$ counts using the the NB expected value formula:

$$E(y_{it}|x_{it}) = e^{x_{it}\hat{\beta}} \tag{1}$$

where $\hat{\beta}$ corresponds to our (1997-2004) NB coefficient estimates, x_{it} corresponds to our covariates, and $E(y_{it}|x_{it})$ corresponds to our expected number of event counts. We then predicted in-sample and out-of-sample ZINB model $government\ conflict_{it}$ and $rebel\ conflict_{it}$ count frequencies using the ZINB expected value formula

$$E(y_{it}|x_{it}, z_{it}) = e^{x_{it}\hat{\beta}} - \pi_i e^{x_{it}\hat{\beta}} \tag{2}$$

where, because our specific inflation equations follow a logistic probability distribution,

$$\pi_{it} = Pr(i \in r_0|z_i) = \frac{1}{1+e^{-z_{it}\hat{\gamma}}} \tag{3}$$

and where here, r_0 corresponds to the zero-only regime, $\hat{\gamma}$ corresponds to our 1997-2004 ZINB inflation stage coefficient estimates, z_{it} corresponds to our ZINB inflation stage covariates, $\hat{\beta}$ corresponds to our 1997-2004 ZINB outcome-stage coefficient estimates, x_{it} are our outcome stage covariates, and $E(y_{it}|x_{it})$ are our ZINB predicted expected event counts. These in-sample and out-of-sample count forecasts were then used to derive classification matrix statistics, and corresponding 95% confidence intervals, for each set of models. Specifically, we calculated five classification matrix statistics for each model by first dichotomizing our forecasted and observed counts in order to evaluate the accuracy of our model predictions across two intuitive conflict thresholds:

1. Rebel and government initiated conflicts $\geq$ one conflict per country-month

2. Rebel and government initiated conflicts $\geq$ five conflicts per country-month

In addition to reporting the true "peace-conflict" proportions for each of these dichotomized thresholds, we calculate and report five relevant classification statistics for each threshold of interest:

sensitivity, specificity, negative predicted values, positive predicted values, and the percent correctly classified.[17] Sensitivity reports the proportion of actual conflict country-months that were correctly identified as conflict-months (for a given threshold) by our forecasting models. Specificity reports to the proportion of peace-country-months that were correctly identified as such by our models. Positive predictive values (PPVs) refer to the proportion of our conflict-country-month forecasts that were actually observed to be conflict-country-months within our sample. Negative predictive values (NPVs) refer to the proportion of peace-country-month forecasts that were actually observed to be peace-country-months within the sample. Finally, our 'correctly classified' statistic reports the percentage of cases within a given sample that were actually classified as either peace or conflict by our forecasting model.

Table 1 presents classification statistics–and corresponding 95% confidence intervals—for our in-sample (1997-2004) and out-of-sample (2005-2010) *government* and $rebel\ conflict_{it}$ forecasts. Beginning with $government\ conflicts_{it}$, this Table demonstrates that across both conflict thresholds the ZINB model is superior to the NB model in predicting country-months that actually experience a given threshold of government initiated conflict greater than zero (i.e. sensitivity). Specifically, the $government\ conflict_{it}$ sensitivity statistics in Table 1 indicate that our out-of-sample ZINB models are on average 8.4% better at accurately forecasting country-months that experience at least one conflict (sensitivity= 81.97%) and at-least five conflicts (sensitivity= 82.72%) than our NB models (sensitivity= 68.72% & 79.58%). At the same time, across all $government\ conflict_{it}$ specifications reported in Table 1, the ZINB and NB models perform comparably well in terms of cases correctly classified (90.22% − 95.70%), specificity (94.46% − 97.54%), and NPV (91.35% − 97.99%), which is unsurprising given the overabundance of zero, "peace year" observations within the samples of interest. Regarding the $government\ conflict_{it}$ PPV statistics reported in Table 1, the NB model does do on average 4.7% better than ZINB models. However, the sensitivity scores discussed above, as well as the slightly lower NPVs reported in Table 1, together suggest that these relatively higher NB PPVs are the result of an overprediction of zeroes—and an underprediction

[17] The formulas for these classification statistics appear in the Supplemental Appendix.

of government conflict (for each conflict threshold)—by our NB models; rather than any superior ability in *conflict* forecasting. In sum, while both models do a comparable job of predicting peace-months, the ZINB model is superior to the NB model in terms of predicting conflict-country-months, within both in-sample and out-of-sample settings, which further corroborates hypotheses 1 and 2.

[*Insert Table 1 about here*]

We draw similar conclusions from the classification statistics and confidence intervals that are reported for $rebel\ conflicts_{it}$ in Table 1. Across both conflict thresholds the ZINB model is superior in sensitivity to the NB model in predicting actual instances of rebel-initiated material violence. Specifically, Table 1 indicates that the out-of-sample ZINB model is on average 7.2% better at accurately forecasting country-months that experience at least one conflict (sensitivity= 80.05%) and at-least five conflicts (sensitivity= 80.41%), relative to comparable NB models (sensitivity= 68.31% & 77.84%). Across all $rebel\ conflicts_{it}$ specifications reported in Table 1 the ZINB and NB models again perform comparably well in terms of cases correctly classified (89.45% – 95.86%), specificity (93.42% – 97.96%), and NPV (90.04% – 97.71%), which again is unsurprising given the overabundance of zero, "peace year" cases within the samples of interest. Finally, the PPVs reported in Table 1 suggest that the NB model does on average 5.9% better than ZINB models, which is likely the result of an overprediction of zeroes, and an underprediction of rebel conflict (for each conflict threshold) by our NB models. Thus, while both the NB and ZINB models do a comparable job of predicting peace-months, the ZINB model is superior to the NB model in terms of predicting actual instances of $rebel\ conflicts_{it}$, within both an in-sample and out-of-sample setting, which is strong support for hypothesis one, and indirect support for hypothesis 2.

Marginal Calibration Diagrams

For a more comprehensive evaluation of our government and rebel conflict forecasting-models, we next compare the marginal calibration of our NB and zero ZINB count forecasts to the actual count values observed in our true (training and validation) datasets. In contrast to the classification matrices discussed above, marginal calibration diagrams offer a comprehensive view of count-model forecasting accuracy *across the entire range of possible event counts*. Specifically, marginal calibration comparisons evaluate the calibration of probabilistic count forecasts against a set of observed counts, where marginal calibration is fully achieved if one's average observed count forecasts equal one's average probabilistic forecasts as $T \to \infty$, provided that all mass is placed on finite values (Gneiting, Balabdaoui and Raftery 2007). To calculate marginal calibrations for our models of interest, we first define P as a predictive probability distribution on the set of nonnegative integers resulting from the probabilistic forecasts derived from our count models. Assuming then that each observed count, $x^{(it)}$, is a random draw from its respective probabilistic forecast, a histogram of these observed counts is statistically comparable to the composite distributions of our aggregated predictive distributions $P^{(it)}$ (Czado, Gneiting and Held 2009). When can then represent these aggregations graphically via a marginal calibration diagram, which compares the predicted frequencies,

$$\hat{p}_x = \sum_{i=1}^{n} (P_x^{(it)} - P_{x-1}^{(it)}) \quad or \quad \hat{p}_{(x_a,x_b]} = \sum_{i=1}^{n} (P_{x_b}^{(it)} - P_{x_a}^{(it)}) \tag{4}$$

for specific x values or intervals $(x_a, x_b]$, to their empirical counterparts,

$$f_x = \sum_{i=1}^{n} 1(x^{(it)} = x) \quad or \quad f_{(x_a,x_b]} = \sum_{i=1}^{n} 1(x_a < x^{(it)} \leq x_b), \tag{5}$$

in an extension of the marginal calibration diagram formulas presented in Czado, Gneiting and Held (2009). This diagnostic tool thereby allows one to evaluate the performance of count forecasts across the *entire range* of observed counts, rather than for a single dichotomous threshold at a time, as was the case for the classification tables presented above. Marginal calibration diagrams comparing observed count values to ZINB and NB model forecasts were calculated for our 1997-2004 government and rebel conflict in-sample predictions, and for our 2005-2010 out-of-sample forecasts.[18] These marginal calibration diagrams appear in Figures 2 and 3. Importantly, the zero-category (peace-country-month) values and predictions are omitted from these figures so as not to visually distort the variation that exists across the (NB and ZINB model) predicted frequencies and their empirical counterparts for the monthly counts of government and rebel initiated conflict greater than zero (i.e. conflict country-months), which are of the most interest to the study at hand.

Figure 2 reports marginal calibration diagrams for our in-sample (1997-2004) out-of-sample (2005-2010) forecasts of $government\ conflicts_{it}$. This Figure suggests that, although both models do a competent job of predicting government initiated conflicts, the ZINB models are superior to the NB models in calibration. To see this, we focus our discussion here on the out-of-sample predictions (Figure 2b). Turning to Figure Figure 2b, note first that our NB model substantially over predicts the number of country-months experiencing a single instance of government initiated civil conflict (by 130%) while the ZINB model only slightly under predicts the number of country-month instances of a single observed government initiated conflict within our validation sample (by 16%). The ZINB and NB models then each do a fairly accurate job of forecasting observations with observed monthly conflict counts lying between two and five (inclusive). However, as we begin to aggregate across higher levels of government initiated conflict counts in Figure 2, we see an increased divergence in NB-to-ZINB forecasting accuracy that again favors the ZINB model. In particular, the ZINB model does a much better job of predicting the spike in government conflict frequencies that we observe across bins $(5,10]$, $(10,25]$, and $(25,50]$. Specifically, while the ZINB model under predicts observations lying within these conflict thresholds by an average of 31%, our comparable NB predictions are off by an average of 54%. Lastly, both models do a comparable job of predicting the (exceedingly rare) frequencies of monthly government initiated conflicts lying within the final $(50,100]$ interval.[19] Overall, Figure 2 indicates that the NB model of $government\ conflicts_{it}$ tends to over-predict low-level country-month instances of government initiated civil conflict and under-predict higher levels of monthly conflict. By contrast, the ZINB

[18] Marginal calibration diagrams calculated over the entire 1997-2010 period are comparable to those described here, and are available upon request.

[19] Although the NB model forecasts were in this case slightly closer to the actual count frequencies.

model of *government* $conflicts_{it}$ does a much better job of accurately predicting monthly conflict frequencies across this variable's entire range, although the ZINB model has similar difficulties in accurately predicting country-months experiencing very high levels of government-initiated civil conflicts. Thus, the marginal calibration diagrams in Figure 2 suggest that ZINB models do a superior job of forecasting *government* $conflicts_{it}$, relative to a comparable NB models, which is strong support for hypothesis 1.

[*Insert Figures 2 and 3 about here*]

Figure 3 presents marginal calibration diagrams of our ZINB and NB model in-sample and (1997-2004) out-of-sample (2005-2010) predictions of *rebel* $conflicts_{it}$. As above, this Figure suggests that the ZINB model is superior in calibration to the NB model, and to elucidate this we focus our discussion heretofore on our out-of-sample predictions (Figure 3b). Here, our NB model over-predicts the number of country-months experiencing a single instance of rebel initiated conflict by 119%. By contrast, our ZINB model does a much better job of prediction within this range of monthly conflicts, with ZINB out-of-sample forecasts under-predicting the frequency of single-rebel-conflict country-months by only 15%. The ZINB and NB models each do a commensurate job of forecasting countries experiencing between two and five conflicts per month (inclusive). Aggregating across higher levels of monthly rebel initiated conflict counts, we find in Figure 3 that as above, our ZINB model better predicts the increased number of country-months experiencing conflicts across these heightened conflict intervals. For example, within the $(5,10]$, $(10,25]$, $(25,50]$ monthly conflict intervals, the out-of-sample country-month predictions made by our ZINB model are off by an average 24%, whereas comparable NB model predictions are off by an average 45%. Lastly, although both models do a comparable job of predicting the frequencies of monthly rebel initiated conflicts lying within the final $(50,100]$ range, we can note here that the ZINB model frequency forecasts are slightly closer to the actual count frequencies. Hence, the NB model discussed here over-predicts country-month instances of (rebel initiated) material conflict for low-level country-month conflict counts (i.e. values of *rebel* $conflicts_{it}$ ranging from zero to approximately three) and under-predicts higher levels of monthly conflict (i.e. values of *rebel*

$conflicts_{it}$ between five and 50). By contrast, the ZINB model accurately predicts conflict across the entire range of monthly rebel-initiated conflicts, although it also occasionally under predicts the number of countries experiencing low-level rebel-initiated conflicts. Therefore, and in support of our hypotheses, the marginal calibration diagrams in Figure 3 suggest that our ZINB models provide more accurate forecasts of $rebel\ conflicts_{it}$ than do our NB models.

ZINB Comparisons

While the above analysis demonstrates the superiority of the ZINB conflict models over comparable NB models (hypothesis 1), it is only suggestive as to the forecasting-advantages of including lagged conflict measures within the inflation stage of the ZINB models (hypothesis 2). To better assess the latter, we build upon the ZINB analysis presented above by incrementally adding-in an ever-expanding number of lagged conflict variables to the inflation stages of our ZINB models. While doing so, we hold these models' outcome (i.e. count) stage covariate specifications fixed to the count-stage specifications reported above, with additions of 3 and 4 month lagged values of $government\ conflict_{it}$ and $rebel\ conflict_{it}$. In our inflation stages, we begin with a ZINB model reporting only an inflation stage constant, and then add $ln\ GDP_{pc}$ to this stage, evaluating the results at both steps. We next sequentially add 1-to-5 month lagged values of $government\ conflict_{it}$ and $rebel\ conflict_{it}$ to the inflation stage of our ZINB model, and again evaluate the results at each step. For each of our two dependent variables, the resultant seven (nested) ZINB specifications are then compared via a number of model-fit statistics. Vuong tests indicate that for all ZINB models of both $government\ conflict_{it}$ and $rebel\ conflict_{it}$, the inclusion of each successive pair of lagged $government\ conflict_{it}$ and $rebel\ conflict_{it}$ inflation stage covariates produces a significant ($p < .01$) improvement in model fit and model performance. Likelihood ratio tests similarly suggest that the addition of 1-5 month lags of $government\ conflict_{it}$ and $rebel\ conflict_{it}$ to the inflation stage of our ZINB models produces a significant ($p < .01$) improvement in model fit. Finally, in comparing the AICs of our ZINB models, we find that each pair-wise comparison preferred a more-fully specified ZINB model to a given ZINB model with fewer inflation stage (lagged conflict) covariates. Hence, a wide range of model fit statistics further corroborate our initial findings (in Table A) that past levels of civil conflict serve as significant and robust predictors of zero inflation, which is in support of hypothesis 2.

To determine whether lagged inflation-stage covariates affect our actual conflict *forecasts*, we evaluate our new models using a series of sensitivity plots. These plots compare the sensitivity levels of our out-of-sample conflict predictions for the seven sets of ($government\ conflict_{it}$ and $rebel\ conflict_{it}$) ZINB model variants described above.[20] The sensitivity statistics used in these plots report the proportion of actual conflicts that our models predicted as such, and thus are particularly useful in comparing the conflict-forecasting accuracy of our ZINB models. Our sensitivity plots are very similar in motivation to the predictive power plots used by Ward, Greenhill and Bakke (2010, 369), with the exception that we plot the fixed *sensitivity* levels of our model's forecasts, rather than the total-area under a receiver operating characteristic (ROC) curve. We favor the former not only because our dependent variables and predictions encompass values greater than one, but also because the extreme proportion of zeroes in our sample—in conjunction with comparable specificity levels across all ZINB models—together tend to obscure the differences in actual sensitivity levels across these models, when aggregating across all discrimination thresholds. As in our classification matrices, we calculated $government\ conflict_{it}$ and $rebel\ conflict_{it}$ sensitivity statistics for both the "at-least 1-monthly-conflict" and "at least 5-monthly conflicts" thresholds. We then repeated this process iteratively for our ZINB models as more lagged values of $government\ conflict_{it}$ and $rebel\ conflict_{it}$ were incrementally added to the models' inflation stages (beginning with a ZINB model that includes only a constant in the inflation stage). The resultant sensitivity plots for our 1-monthly conflict and 5-monthly conflict thresholds appear in Figure 4.

[*Insert Figure 4 about here*]

Beginning with Figure 4a, we can note that these plots strongly support hypothesis 2, for both our $government\ conflict_{it}$ and $rebel\ conflict_{it}$ ZINB models. For $government\ conflict_{it}$, Figure 4a first indicates that adding $ln\ GDP_{pc}$ to our inflation stage increases our ability to accurately predict instances of government initiated conflict by roughly 4%. By comparison, subsequently

[20] In-sample sensitivity plots are comparable, and are available upon request.

adding 1, 1-2, or 1-3 month lags of *government* and $rebel\ conflict_{it}$ to the inflation stage of our $government\ conflict_{it}$ ZINB models produces 8%, 9%, and 13% increases in sensitivity, again relative to the "constant-only" inflation stage model. Our $rebel\ conflict_{it}$ ZINB model exhibits comparable increases in sensitivity (of 5%, 11% and 11%) for these same three pairs of (1-3 month) lagged conflict models, although $ln\ GDP_{pc}$ contributes little to sensitivity in this case. Next, note that the marginal increase in sensitivity provided for by the addition of 4, and 4-5 month conflict lags is negligible in either model. Indeed, in the case of $rebel\ conflict_{it}$, adding 4 or 4-5 month conflict lags to our inflation stage actually decreases sensitivity relative to our 1-3 lagged conflict specification. For the $government\ conflict_{it}$ model, additions of 4 (or 4-5) month conflict lags do slightly improve sensitivity, but do so at a decreasing rate, relative to the gains made by earlier inflation-stage covariate-additions. Thus, the contributions of past conflict-levels to our ability to distinguish between inflated and non-inflated peace-months appears to diminish after 2-3 months, suggesting that these inflation-covariates are accounting for a temporally varying—rather than fixed—form of zero inflation. Figure 4a is therefore strong support for hypothesis 2, as it demonstrates that the addition of lagged conflict values to the inflation stage of our ZINB models produces a marked improvement in the accuracy of our conflict forecasts.

Relative to Figure 4a, the sensitivity plots in Figure 4b report higher initial levels of sensitivity, and hence we find smaller additional gains in sensitivity across all covariate additions. Nevertheless, the trends in Figure 4b are comparable to those discussed above. Relative to a ZINB model with a "constant-only" inflation stage, the inclusion of 1, 1-2, and 1-3 month lagged conflicts in the inflation stage of our ZINB models increases the sensitivity levels of our $government\ conflict_{it}$ (of 2%, 3%, and 3%, repetitively) and $rebel\ conflict_{it}$ (of 2%, 3%, and 3%) forecasts. By contrast, the addition of $ln\ GDP_{pc}$ to the inflation stages of these $government\ conflict_{it}$ and $rebel\ conflict_{it}$ models yields an average improvement in sensitivity of roughly 0%. As above, any additional gains in sensitivity are negligible and in many cases negative when 4-5 month lagged conflict measures are added to either our $government\ conflict_{it}$ or $rebel\ conflict_{it}$ inflation stages. Hence, the benefit of including lagged levels of conflict within the inflation stages of our models dissipates after approximately three months. Substantively, this suggests that—in addition to the time-invariant structural factors that may predispose some countries from ever experiencing civil conflicts—unobserved short-term (i.e. 1-3 month) temporal dynamics such as mutually reinforcing stalemates, tit-for-tat strategies, de-facto truces or exogenous conditions (e.g. seasonal weather) appear to also preclude government and rebel actors from fighting for short periods of time. Modeling these temporal dynamics in the manner presented above allows us to better identify, and hence predict, the observations which are most likely to experience conflict in any given month. Therefore overall, the sensitivity results discussed here are further evidence in support hypothesis 2, which contends that the inclusion of lagged conflict variables within the inflation stage of our ZINB models will significantly improve our conflict forecasting accuracy.

Conclusion

In this paper, we present a variety of useful tools that allow forecasters to evaluate and compare count models of social conflict for predictive accuracy and proper specification. We find that zero inflated count models improve our ability to forecast monthly frequencies of rebel and government-initiated conflicts. Using zero-inflated models in conjunction with lagged conflict covariates gives these models a decided advantage over comparable approaches. On average, the ZINB models discussed above were roughly 8% better than comparable NB models at accurately forecasting out-of-sample country-months that experienced at least one civil conflict and at-least five civil conflicts. Likewise, marginal calibration diagrams suggest that ZINB models do a much better job of forecasting monthly conflict *frequency* across the entire range of possible monthly conflict counts, again relative to comparable NB models. While the benefits of zero-inflated models are well known among conflict researchers, we also demonstrate for the first time that the inclusion of lagged values of rebel and government initiated conflicts *within the inflation stages* of zero inflated count models yields a considerable improvement in forecasting accuracy, relative to ZINB models which do not include such covariates in the inflation stage. Specifically, the addition of 1-3 month lagged measures of (logged) civil conflict frequency to the inflation stage of our ZINB models improved our ability to accurately forecast countries experiencing at least one, and at least five, monthly civil conflicts by 12% and 3% respectively . Averaging across these two thresholds, as well as across our government *and* rebel conflict models, our final ZINB models accurately predicted 81% of all monthly-conflict events.

This article makes several key contributions to conflict-forecasting methodology. Event counts, selection models, and discrete (un)ordered outcomes are integral to the study of civil conflict. However, a current limitation to civil conflict forecasting, and to conflict studies in general, has been the underdevelopment—across the sciences—of forecasting models (and assessment techniques) for discrete dependent variables of the count, ordered, or unordered varieties (Czado, Gneiting and Held 2009). We address these deficiencies above by providing the first forecasting assessment of civil conflict *frequency*; while using of a newly developed dataset that uniquely measures monthly conflict frequency for both rebel and government initiators. In doing so, we provide examples of several robust techniques that one can use in assessing the accuracy, specificity, and sensitivity of one's count forecasts. To this end, we present marginal calibration diagrams, comparative fit statistics, and classification statistics that together allow the researcher to begin to gain a sense of count model forecasting precision. It is thereby hoped that, through these examples, this article will serve as a useful starting point for future conflict-event forecasting researchers faced with a dependent variable that is limited in nature, or contaminated with structural zeroes.

Substantively, our results indicate that recent levels of rebel and government initiated material conflict have a direct, positive effect on present levels of each type of conflict, which is in line with theories of conflict reciprocity and conflict inertia (Gurr 1970; Hibbs 1973; Francisco 1995; Goldstein and Freeman 1990). However, we also find that the magnitude of this positive relationship tends to be overstated when the presence of zero-inflation is ignored within one's statistical model. The results discussed above also suggest that time-varying peace-inducing dynamics—such as secret or de-facto truces—do occur, and that modeling such phenomena can enhance our abilities to predict and understand civil conflict. Specifically, past levels of monthly government and rebel initiated conflicts serve as excellent ex-ante observable indicators of time-varying, structurallyinflated peace-periods. However, sensitivity analyses indicate that the contribution of these lagged conflict measures to the modeling of such temporal stalemates dissipates markedly for any conflict measures beyond 3-month lags. Therefore, any gains to be had from the modeling of temporary (structural) peace-spells with lagged conflict measures appear to be temporally limited to the 1-3 month period immediately prior to a civil conflict period of interest. Finally, in line with past scholarship (Hill et al. 2011), we find that $ln\ GDP_{pc}$ has a positive and significant effect on the likelihood of a country-month *ever* experiencing government initiated conflicts targeting rebels, but a *null* effect on the likelihood of domestic rebel groups initiating such conflicts against government actors. This adds another layer of nuance to past $ln\ GDP_{pc}$-peace-inflation findings, and suggests that $ln\ GDP_{pc}$ serves as more of a constraint against government-initiated conflict than against citizen-initiated violence.

For researchers interested in the direct effect of *any variable* on civil conflict, these findings suggest that one can substantially reduce the bias imposed by excess zeroes by (1) using a zero-inflated model and (2) including appropriate lagged values of conflict within the inflation stage of zero inflated models. A key advantage of this approach is that—no matter the temporal aggregation or cross-sectional unit of observation—lagged dependent (conflict) variables will be available to the researcher for a majority of the sample of interest. Given the challenges associated with coding additional (time varying) civil conflict covariates in forecasting models as one moves to (1) smaller-and-smaller units of temporal (or cross-sectional) aggregation or (2) real-time forecasting, lagged conflict variables will be especially useful in these contexts. In fact, the approach outlined here is likely to yield even larger improvements in forecasting accuracy when applied to datasets aggregating over *smaller* temporal or geographic units of observation, such as days or districts, since under these circumstances the level of zero-inflation will in most cases become more severe. Finally, while zero-inflated *count* models are used here, the methods and forecasting tools described above are applicable to the entire range of zero-inflated, limited dependent variable models currently available.[21]

[21] Such as mixture models for duration data (Svolik 2008), binary dependent variables (Xiang 2010), and discrete ordered outcomes (Harris and Zhao 2007; Hill et al. 2011).

There are several promising directions for future research. The use of lagged civil conflict values in the inflation stage of our ZINB models, while an improvement over simpler NB and ZINB models, falls short of an ideal inflation stage specification. There is a breadth of available time-varying and structural factors that could likely be used to better model the inflation stage of such models, and a comprehensive assessment of which set of covariates best predicts the propensity for a peace observation to be a structural zero would greatly advance conflict research and conflict forecasting. Accomplishing this task will bring us closer to developing a systematic approach to the accurate identification of the (proportionally small) subset of (non-structural) observations within our samples whose ex-ante probabilities of conflict is indeed high (Beck, King and Zeng 2000). Drawing on recent environmetric advances in *spatial* zero-inflated count models (Agarwal, Gelfand and Citron-Pousty 2002; Ver Hoef and Jansen 2007), a second avenue by which the above study could be directly improved upon would be through the development of spatial, or space-time, ZINB forecasting models of civil conflict. Accounting for space and time in studies of civil conflict has been shown to be critical for both theory testing and the development of accurate conflict forecasts (Ward and Gleditsch 2002; Weidmann and Toft 2010; Weidmann and Ward 2010), and the models presented in Agarwal, Gelfand and Citron-Pousty (2002) and Ver Hoef and Jansen (2007) therefore serve as excellent templates for the future refinement of zero-inflated conflict forecasting models.

References

Agarwal, Deepak K., Alan E. Gelfand and Steven Citron-Pousty. 2002. "Zero-inflated Models with Application to Spatial Count Data." *Environmental and Ecological Statistics* 9(4):341–355.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York, NY: Basic Books.

Beck, Nathaniel, Gary King and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94(1):21–35.

Beck, Nathaniel, Jonathan N. Katz and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(4):1260–1288.

Benini, Aldo A. and Lawrence H. Moulton. 2004. "Civilian Victims in an Asymmetrical Conflict: Operation Enduring Freedom, Afghanistan." *Journal of Peace Research* 41(4):403–422.

Bennett, D. Scott and Alan Stam. 2004. *The Behavioral Orgins of War*. Ann Arbor, NJ: University of Michigan Press.

Brandt, Patrick T., John R. Freeman and Philip A. Schrodt. 2011. "Real Time, Time Series Forecasting of Inter- and Intra-state Political Conflict." *Conflict Management and Peace Science* 28(1):41–64.

Buhaug, Halvard. 2006. "Relative Capability and Rebel Objective in Civil War." *Journal of Peace Research* 43(6):691–708.

Clark, David H. and Patrick M. Regan. 2003. "Opportunities to Fight: A Statistical Technique for Modeling Unobservable Phenomena." *The Journal of Conflict Resolution* 47(1):94–115.

Cohen, Jacob. 1983. "The Cost of Dichotomization." *Applied Psychological Measurement* 7(3):249–253.

Czado, Claudia, Tilmann Gneiting and Leonard Held. 2009. "Predictive Model Assessment for Count Data." *Biometrics* 65(4):1254–1261.

D'Orazio, Vito, James E. Yonamine and Philip A. Schrodt. 2011. "Predicting Intra-State Conflict Onset: An Event Data Approach Using Euclidean and Levenshtein Distance Measures." Presented at the 69th annual Midwest Political Science Association meeting.

Fearon, James D. 2004. "Why Do Some Civil Wars Last so Much Longer than Others?" *Journal of Peace Research* 41(3):275–301.

Francisco, Ronald A. 1995. "The Relationship Between Coercion and Protest: An Empirical Evaluation in Three Coercive States." *Journal of Conflict Resolution* 39(2):263–282.

Gneiting, Tilmann, Fadoua Balabdaoui and Adrian E. Raftery. 2007. "Probabilistic Forecasts, Calibration and Sharpness." *Journal of the Royal Statistical Society: Series B Statistical Methodology* 69(2):243–268.

Goldstein, Joshua S. and John R. Freeman. 1990. *Three-Way Street: Strategic Reciprocity in World Politics*. Chicago, IL: University of Chicago Press.

Gurr, Ted R. 1970. *Why Men Rebel*. Princeton, NJ: Princeton University Press.

Gurr, Ted Robert. 1993. "Why Minorities Rebel: A Global Analysis of Communal Mobilization and Conflict since 1945." *International Political Science Review* 14(2):161–201.

Gurr, Ted Robert and Mark Irving Lichbach. 1986. "Forecasting Internal Conflict : A Competitive Evaluation of Empirical Theories." *Comparative Empirical Studies* 19(3):3–38.

Harris, Mark N. and Xueyan Zhao. 2007. "A Zero-Inflated Ordered Probit Model, with an Application to Modelling Tobacco Consumption." *Journal of Econometrics* 141(2):1073–1099.

Hegre, Haavard, Gudrun Ostby and Clionadh Raleigh. 2009. "Poverty and Civil War Events: A Disaggregated Study of Liberia." *Journal of Conflict Resolution* 53(4):598–623.

Hibbs, Douglas A. 1973. *Mass Political Violence: A Cross-National Causal Analysis*. New York, NY: John Wiley and Sons.

Hill, Daniel W., Benjamin E. Bagozzi, Will H. Moore and Bumba Mukherjee. 2011. "Strategic Incentives and Modeling Bias in Ordinal Data:The Zero-inflated Ordered Probit (ZiOP) Model in Political Science." http://qssi.psu.edu/files/NF4Hill.pdf. Paper presented at the New Faces in Political Methodology meeting, Penn State, 30 April 2011.

Holmes, Jennifer S., Sheila Amin Gutierrez De Pineres and Kevin M. Curtina. 2007. "A Subnational Study of Insurgency: FARC Violence in the 1990s." *Studies in Conflict & Terrorism* 30(3):249–265.

Hultman, Lisa. 2007. "Battle Losses and Rebel Violence: Raising the Costs for Fighting Terrorism and Political Violence." *Terrorism and Political Violence* 19(2):205–222.

Kydd, Andrew H. and Barbara F. Walter. 2002. "Sabotaging the Peace: The Politics of Extremist Violence." *International Organization* 56(2):263–296.

Lehinan, Pádraig. 2002. *Confederate Catholics at War, 1641-49*. Cork, Ireland: Irish Committee of Historical Sciences.

Lemke, Douglas and William Reed. 2001. "The Relevance of Politically Relevant Dyads." *Journal of Conflict Resolution* 45(1):125–144.

Lichbach, Mark Irving and Ted Robert Gurr. 1981. "The Conflict Process : A Formal Model." *Journal of Conflict Resolution* 25(3):3–29.

MacCallum, Robert C., Shaobo Zhang, Kristopher J. Preacher and Derek D. Rucker. 2002. "On the Practice of Dichotomization of Quantitative Variables." *Psychological Methods* 7(1):624–638.

Maoz, Zeev and Bruce M. Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946-1986." *American Political Science Review* 87(3):624–638.

O'Brien, Sean P. 2002. "Anticipating the Good, the Bad, and the Ugly: An Early Warning Approach to Conflict and Instability Analysis." *Journal of Conflcit Resolution* 46(6):791–811.

O'Brien, Sean P. 2010. "Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research." *International Studies Review* 12(1):87–104.

Pevehouse, Jon. 2004. "Interdependence Theory and the Measurement of International Conflict." *Journal of Politics* 66(1):247–266.

Pevehouse, Jon C. and Joshua S. Goldstein. 1999. "Serbian Compliance or Defiance in Kosovo? Statistical Analysis and Real-Time Predictions." *Journal of Conflict Resolution* 43(4):538–546.

Piazza, James A. 2011. "Poverty, Minority Economic Discrimination, and Domestic Terrorism." *Journal of Peace Research* 48(3):339–353.

Royston, Patrick, Douglas G. Altman and Willi Sauerbrei. 2006. "Dichotomizing Continuous Predictors in Multiple Regression: A Bad Idea." *Statistics in Medicine* 25(1):127Ű141.

Schrodt, Philip A. 2009. "TABARI: Textual Analysis by Augmented Replacement Instructions, Version 7.0." http://eventdata.psu.edu/tabari.dir/tabari.manual.0.7.3b3.pdf.

Schrodt, Philip A., Deborah J. Gerner and Omur Yilmaz. 2009. *International Conflict Mediation: New Approaches and Findings*. New York: Routledge chapter Conflict and Mediation Event Observations (CAMEO): An Event Data Framework for a Post Cold War World.

Schrodt, Philip A. and Omur Yilmaz. 2007. "CAMEO Conflict and Mediation Event Observation Codebook." http://eventdata.psu.edu/cameo.dir/CAMEO.CDB.09b5.pdf.

Shellman, Stephen M., Calre Hatfield and Maggie J. Mills. 2010. "Disaggregating Actors in Intranational Conflict." *Journal of Peace Research* 47(1):83–90.

Svolik, Milan W. 2008. "Authoritarian Reversals and Democratic Consolidation." *American Political Science Review* 102(2):153–168.

Ver Hoef, Jay M. and John K. Jansen. 2007. "Space-Time Zero-Inflated Count Models of Harbor Seals." *Environmetrics* 18(7):697–712.

Vuong, Quang H. 1989. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica* 57(2):307–333.

Wandycz, P.S. 1965. "Secret Soviet-Polish Peace Talks in 1919." *Slavic Review* 24(3):425–449.

Wanis-St. John, Anthony. 2006. "Back-Channel Negotiation: International Bargaining in the Shadows." *Negotiation Journal* 22(2):119–144.

Ward, Michael D., Brian D. Greenhill and Kristin M. Bakke. 2010. "The Perils of Policy by P-Value: Predicting Civil Conflicts." *Journal of Peace Research* 47(4):363–375.

Ward, Michael D. and Kristian Skrede Gleditsch. 2002. "Location, Location, Location: An MCMC Approach to Modeling the Spatial Context of War and Peace." *Political Analysis* 10(3):244–260.

Weidmann, Nils B. and Michael D. Ward. 2010. "Predicting Conflict in Space and Time." *Journal of Conflict Resolution* 54(6):883–901.

Weidmann, Nils B. and Monica Duffy Toft. 2010. "Promises and Pitfalls in the Spatial Prediction of Ethnic Violence: A Comment." *Conflict Management and Peace Science* 27(2):159–176.

Wood, Elisabeth Jean. 2003. *Insurgent Collective action and Civil War in El Salvador*. Cambridge, UK: Cambridge University Press.

World Bank. 2011. "World Development Indicators." http://data.worldbank.org/data-catalog/world-development-indicators.

Xiang, Jun. 2010. "Relevance as a Latent Variable in Dyadic Analysis of Conflict." *Journal of Politics* 72(2):484–498.

Figure 1: *Monthly Frequencies of Rebel and Government Initiated Domestic Material Conflicts, 1997-2010*

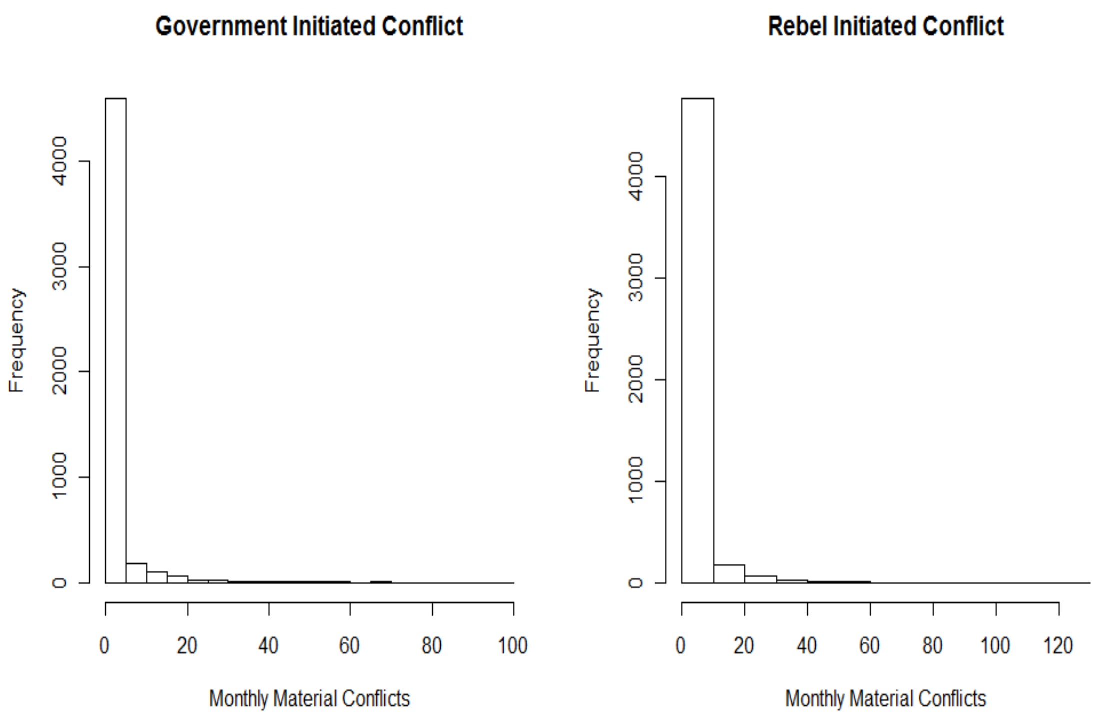


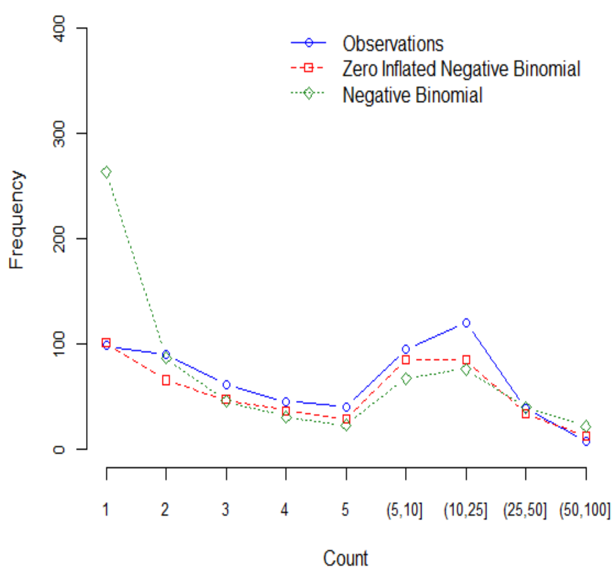
Table 1: Classification

		NB: Threshold 1	ZINB: Threshold 1	NB: Threshold 2	ZINB: Threshold 2
		Monthly Conflicts ≥ 1	*Monthly Conflicts* ≥ 1	*Monthly Conflicts* ≥ 5	*Monthly Conflicts* ≥ 5
Monthly Government Initiated Material Conflict, 1997-2004 (In Sample)					
Sensitivity	$Pr(+\|D)$	75.08%	81.82%	78.00%	83.67%
		(70.71%-79.46%)	(77.77%-85.52%)	(71.33%-81.33%)	(77.00%-87.33%)
Specificity	$Pr(-\|\sim D)$	96.44%	94.63%	96.18%	95.04%
		(95.07%-97.75%)	(92.87%-96.11%)	(95.61%-96.93%)	(93.58%-96.70%)
Positive PV	$Pr(D\|+)$	87.28%	83.22%	74.29%	70.51%
		(83.99%-91.11%)	(79.62%-86.68%)	(72.40%-76.70%)	(65.83%-76.74%)
Negative PV	$Pr(\sim D\|-)$	92.24%	94.11%	96.86%	97.62%
		(91.11%-93.43%)	(93.00%-95.17%)	(95.98%-97.31%)	(96.74%-98.12%)
Correctly Classified		91.29%	91.19%	93.92%	93.63%
Number of Cases (conflict/peace)		594/1,824	594/1,824	300/2,118	300/2,118
Number of Obs.		2,418	2,418	2,418	2,418
Monthly Government Initiated Material Conflict, 2005-2010 (Out of Sample)					
Sensitivity	$Pr(+\|D)$	68.27%	81.97%	79.58%	82.72%
		(63.22%-75.00%)	(75.48%-85.10%)	(74.35%-82.20%	(77.49%-85.86%)
Specificity	$Pr(-\|\sim D)$	96.54%	94.46%	97.54%	96.29%
		(95.15%-98.06%)	(92.52%-95.98%)	(96.88%-97.90%)	(94.79%-97.60%)
Positive PV	$Pr(D\|+)$	85.03%	81.00%	78.76%	71.82%
		(81.68%-90.38%)	(76.62%-84.41%)	(75.12%-80.23%)	(65.34%-78.72%)
Negative PV	$Pr(\sim D\|-)$	91.35%	94.79%	97.66%	97.99%
		(90.25%-92.96%)	(93.15%-95.57%)	(97.09%-97.94%)	(97.43%-98.32%)
Correctly Classified		90.22%	91.67%	95.70%	94.89%
Number of Cases (conflict/peace)		416/1,444	416/1,444	191/1,669	191/1,669
Number of Obs.		1,860	1,860	1,860	1,860
Monthly Citizen Initiated Material Conflict, 1997-2004 (In Sample)					
Sensitivity	$Pr(+\|D)$	70.63%	78.44%	74.76%	77.35%
		(64.69%-74.06%)	(73.59%-81.25%)	(68.39%-78.32%)	(69.26%-85.44%)
Specificity	$Pr(-\|\sim D)$	95.61%	93.42%	97.06%	95.50%
		(93.31%-97.69%)	(90.44%-95.50%)	(95.92%-97.77%)	(94.31%-96.97%)
Positive PV	$Pr(D\|+)$	85.28%	81.10%	78.84%	71.56%
		(79.93%-90.99%)	(75.36%-85.48%)	(73.78%-81.78%)	(68.75%-76.98%)
Negative PV	$Pr(\sim D\|-)$	90.04%	92.33%	96.33%	96.64%
		(88.49%-90.90%)	(90.95%-93.06%)	(96.79%-95.46%)	(95.56%-97.79%)
Correctly Classified		90.00%	89.45%	94.21%	93.18%
Number of Cases (conflict/peace)		640/1,778	640/1,778	309/2,109	309/2,109
Number of Obs.		2,418	2,418	2,418	2,418
Monthly Citizen Initiated Material Conflict, 2005-2010 (Out of Sample)					
Sensitivity	$Pr(+\|D)$	68.31%	80.05%	77.84%-	80.41%
		(62.91%-74.41%)	(75.82%-83.33%)	(74.23%-82.47%)	(72.68%-87.63%)
Specificity	$Pr(-\|\sim D)$	96.72%	93.72%	97.96%	97.12%
		(95.47%-97.91%)	(92.19%-95.68%)	(97.24%-98.68%)	(95.74%-98.20%)
Positive PV	$Pr(D\|+)$	86.10%	79.12%	81.62%	76.47%
		(82.98%-89.93%)	(76.02%-83.90%)	(77.67%-86.75%)	(70.54%-82.46%)
Negative PV	$Pr(\sim D\|-)$	91.13%	94.05%	97.43%	97.71%
		(89.88%-92.63%)	(93.02%-94.90%)	(97.05%-97.94%)	(96.86%-98.52%)
Correctly Classified		89.90%	90.59%	95.86%	95.37%
Number of Cases (conflict/peace)		426/1,434	426/1,434	194/1,666	194/1,666
Number of Obs.		1,860	1,860	1,860	1,860

Values in parentheses: 95% CIs. $+$ =predict conflict; $-$ =predict peace; D =actually conflict; $\sim D$ =actually peace.

Figure 2: Marginal Calibration Diagrams for Government Initiated Conflict

(a) In-Sample Predictions



(b) Out-of-Sample Predictions

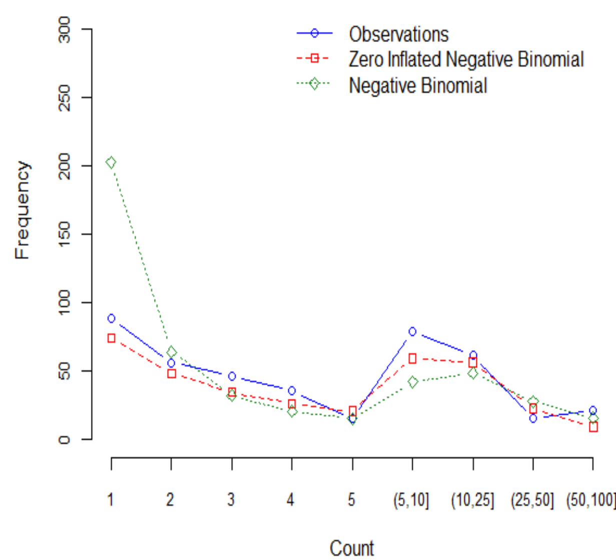
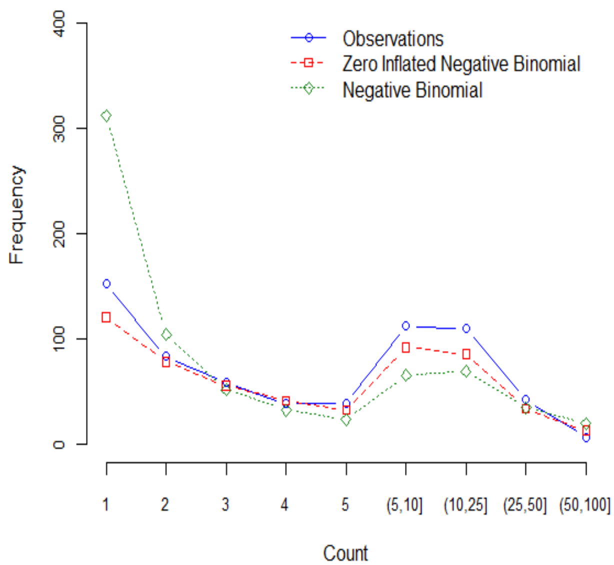


Figure 3: Marginal Calibration Diagrams for Rebel Initiated Conflict

(a) In-Sample Predictions



(b) Out-of-Sample Predictions

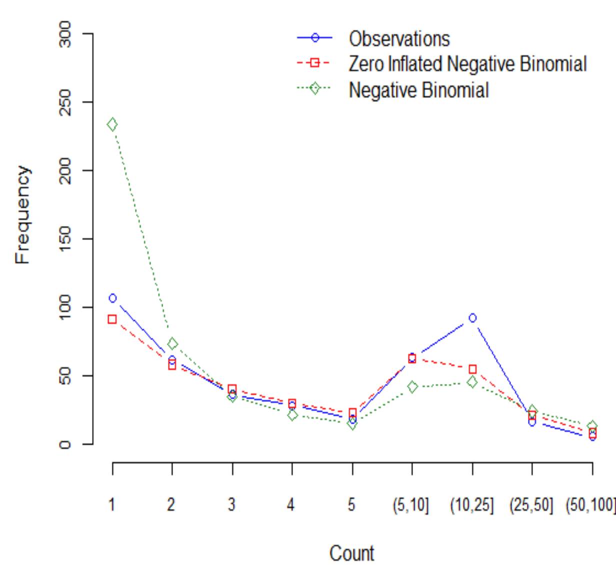
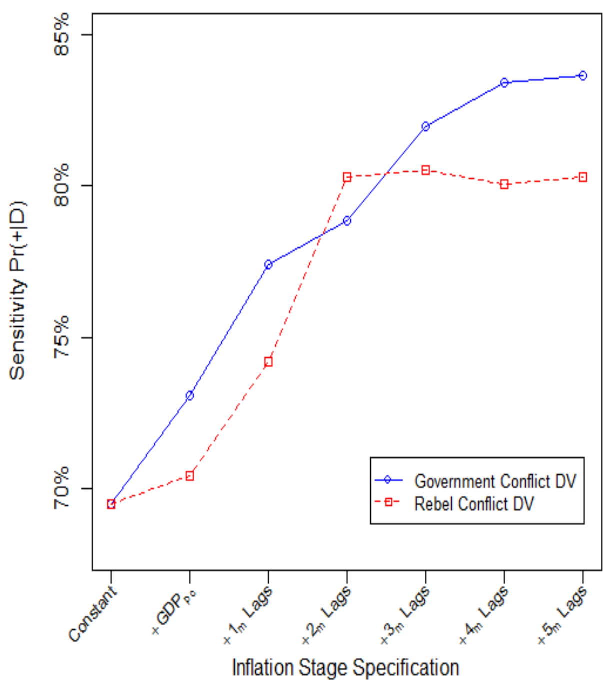


Figure 4: Sensitivity Comparisons for ZINB Out of Sample Predictions

(a) Conflict Threshold of 1



(b) Conflict Threshold of 5