Civil Wars

ISSN: 1369-8249 (Print) 1743-968X (Online) Journal homepage: http://www.tandfonline.com/loi/fciv20

Forecasting Civil Conflict with Zero-Inflated Count Models

Benjamin E. Bagozzi

To cite this article: Benjamin E. Bagozzi (2015) Forecasting Civil Conflict with Zero-Inflated Count Models, Civil Wars, 17:1, 1-24, DOI: 10.1080/13698249.2015.1059564

To link to this article: http://dx.doi.org/10.1080/13698249.2015.1059564

Published online: 11 Sep 2015.

Civil Wars, 2015
Vol. 17, No. 1, 1–24, http://dx.doi.org/10.1080/13698249.2015.1059564

Forecasting Civil Conflict with Zero-Inflated Count Models

BENJAMIN E. BAGOZZI
Department of Political Science, The University of Minnesota, 1472 Social Sciences, 267 19th Ave S, Minneapolis, MN 55455, USA

Advances in the study of civil war have led to the proliferation of event count data, and to a corresponding increase in the use of (zero-inflated) count models for the quantitative analysis of civil conflict events. Our ability to effectively use these techniques is met with two current limitations. First, researchers do not yet have a definitive answer as to whether zero-inflated count models are a verifiably better approach to civil conflict modeling than are 'less assuming' approaches such as negative binomial count models. Second, the accurate analysis of conflict-event counts with count models – zero-inflated or otherwise – is severely limited by the absence of an effective framework for the evaluation of predictive accuracy, which is an empirical approach that is of increasing importance to conflict modelers. This article rectifies both of these deficiencies. Specifically, this study presents count forecasting techniques for the evaluation and comparison of count models' predictive accuracies. Using these techniques alongside out-of-sample forecasts, it then definitively verifies – for the first time – that zero-inflated count models are superior to comparable non-inflated models for the study of intrastate conflict events.

The availability of fine-grained measures of (inter and intra-state) conflict-event counts has dramatically increased in recent years.[1] This, in turn, has allowed quantitative conflict scholars to use count data – in conjunction with empirical event count models[2] – to make novel inferences about the correlates of conflict frequency.[3] While such approaches have accordingly achieved a degree of prominence in the field of conflict research, they remain disadvantaged for two core reasons. First, and despite their theoretical appeal, the appropriateness of zero-inflated count models for the modeling of conflict events remains ambiguous. Conflict is a rare event,[4] and it is accordingly unclear as to whether the preponderance of peace cases typically observed in conflict data-sets is arising from this inherent rarity[5] or from a heterogeneous mixture of actual peace observations and inflated (i.e., 'irrelevant') peace observations. Moreover, methodological research into the nature of zero-inflated count models suggests that the potential for model misspecification is high, even in cases where zero inflation verifiably exists.[6] As a result, zero-inflated count models continue to be applied unevenly in conflict research, with many researchers still favoring non-inflated approaches,[7] others simultaneously reporting equivalent inflated and non-inflated models,[8] and still others exhibiting either (1) a degree of uncertainty over proper inflation stage specifications[9] or (2) wholesale reservations about the application of zero-inflated models to the study of conflict.[10]

With regard to the second problem alluded to above, conflict scholars are also currently unable to fully evaluate and compare (inflated and non-inflated) count models through the use of appropriate forecasting techniques. Political scientists have recently demonstrated the value of prediction to the advancement of the theoretical and practical understandings of civil conflict,[11] and the field of conflict research is placing increasing importance upon such techniques.[12] However, statistical forecasting tools are almost exclusively designed for either binary or continuous dependent variables.[13] This limits the applicability of existing forecasting methods to the evaluation of predicted conflict frequencies, and many conflict scholars interested in conflict prediction have accordingly been left to favor dichotomous dependent variables over these richer measures of conflict. When applied to graduated social-science variables (such as civil conflict), this practice of dichotomization discards relevant information, exacerbates existing measurement errors within one's variable of interest,[14] and has proven to be detrimental to both prediction and theory testing.[15] Hence, artificial dichotomization is a poor solution to the current limits of forecasting techniques, and the latter deficiency continues to undermine researchers' abilities to leverage advances in conflict forecasting for the study of conflict.

To address these problems, this study builds upon recent statistical advances in probabilistic count-data forecasting[16] to present the first comprehensive forecasting analysis of civil conflict frequency.[17] In doing so, this paper introduces a number of useful statistical tools for the evaluation, refinement, and presentation of conflict-event count forecasts. It then demonstrates with these tools that – when used correctly – zero-inflated count models can produce compelling levels of calibration and sharpness in civil conflict predictions. Specifically, 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 notably increase out-of-sample conflict-count forecasting accuracy. Herein, this project takes a novel approach, and includes past levels of (government and rebel initiated) material conflict in the inflation stage of the forecasting models used below. Doing so is argued to help account for the adverse effects of structural zeroes on one's abilities to forecast civil conflict processes. Hence, while the benefits of zero-inflated models are widely known among conflict scholars, this paper provides a suite of methods that allow conflict researchers to better assess the actual magnitude of these benefits with regard to conflict prediction, which in turn can help identify problems of model misspecification in zero-inflated conflict applications. These contributions are thus intended to serve as a useful starting point for future conflict forecasters that encounter a dependent variable that is limited in nature, or contaminated with structural zeroes.

The next section discusses the prevalence of zero inflation in conflict data, outlines the benefits of addressing this problem with zero-inflated models, and presents a rational for the inclusion of lagged conflict measures as inflation stage covariates. This is followed by the description of an event data-set of monthly civil-conflicts, a justification of the choices of count models, and an application of these models to a training set of monthly conflict-count-data. The heart of the analysis then demonstrates a range of novel forecasting tools and 'best-practices' that conflict scholars can use when evaluating models of conflict events. The conclusion discusses the broader implications of the study.

THEORETICAL MOTIVATION

Notwithstanding the concerns leveraged above, zero inflation is regarded as a pervasive problem by empirical conflict modelers. In particular, yearly, monthly, and weekly aggregations of militarized conflict – whether measured at the dyad, country, or sub-country level – are often claimed to be 'inflated' with structural zeroes.[18] These zeroes represent peace observations that would 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 been treated as structural zeroes[19] since such dyads likely could never experience war due to geographic distance and limited military capabilities. Treating these dyads 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.[20] On the other hand, truncating all potential structural zeroes from one's sample excludes a significant proportion of relevant conflict observations[21] and produces selection bias.[22] As an alternative, 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.[23] 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 down-weight structural zeroes' 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.[24] For example, in a study of human rights violations committed by the Revolutionary Armed Forces of Colombia (FARC), extant research[25] finds that there are many department-months in a sample of all Colombian departments wherein the FARC is wholly inactive. The authors 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 insights into the underlying dynamics of civil conflict. Extant findings suggest the presence of similar zero inflation processes for civil war events within highly disaggregated spatial grid-cells in Liberia.[26] 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.[27] As above, truncating all advanced industrialized countries from one's 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 against the state,[28] 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. This paper therefore follows recent research in proposing 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.

A variety of forecasting techniques are used below to explicitly evaluate claims of zero-inflated model superiority, relative to the claim that all zero cases were drawn from a single, homogeneous set of peace observations. Before doing so, however, note that 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 and count-stage zeroes. 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.[29] 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 conflict forecasting accuracy.[30] This deficiency – along with the (mis)specification concerns highlighted above – limits one's ability to use such covariates, where theoretically appropriate, in the inflation stages of zero-inflated count models of civil conflict.

One exception, however, in terms of both explanatory and forecasting power, is an observation's past levels of conflict. Lagged conflict values have been consistently identified as being among the largest and most robust predictors of inter- and intra-state conflict.[31] Indeed, conflict researchers have found that inter- and intra-state conflicts exhibit a strong temporal dependence[32] and that conflict forecasts can be vastly improved by the inclusion of a series of temporally lagged values of past conflict.[33] This paper thus argues that past levels of conflict (or lack thereof) not only directly affect subsequent levels of civil conflict in a reciprocal or inertial sense, but also help to inform scholars, ex-ante, as to which countries are currently able to experience any level of conflict. In this manner, one can improve conflict forecasting accuracy – and understandings of conflict processes – by including lagged conflict measures as inflation stage covariates.

Specifically, the contention is that zero-inflated peace observations not only arise cross-sectionally,[34] 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 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,[35] tit-for-tat dynamics,[36] or environmental and social pressures such as diplomatic mediators, religious observances, or seasonal harvests. These 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.[37] Scholars have identified comparable instances of secret truces or de-facto stalemates in civil conflict arenas as varied as the Russian Revolution,[38] the Irish Confederate Wars,[39] and the El Salvador civil war.[40] Similar to the aforementioned low conflict propensities within advanced industrialized states, self-enforcing truce periods 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 prevent a given observation from ever experiencing conflict.

ANALYSIS

This paper uses the Integrated Conflict Early Warning System (ICEWS) event data-set to forecast monthly frequencies of domestic civil conflict events within 29 Asian countries for the years 1997–2010.[41] The ICEWS data are part of a Defense Advanced Research Project Agency funded data-set encompassing over two million machine-coded daily events occurring between relevant actors within the Asia–Pacific region. To machine code these events, the ICEWS project coded news articles from over 75 electronic regional and international news sources using the TABARI (Text Analysis By Augmented Replacement Instructions) software program[42] and a Lockheed-Martin-developed Java variant known as JABARI. 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 on a categorical coding scheme developed by the

Conflict and Mediation Event Observation project.[43] The resultant ICEWS events – data-set has been characterized as being 'the most accurate event dataset currently available.'[44]

ICEWS-coded events data correspond to individual daily instances (i.e., actions) of conflict or cooperation between a pair of actors. Subsets of these events were aggregated to the country-month level (*it*) for two specific domestic actors of interest: government actors and rebel actors. In doing so, the paper created two dependent variables. The first is government conflict, which is a monthly count of government actor[45]-initiated, domestic material (i.e., physical, rather than verbal) conflict events[46] targeting violent rebel actors operating within a government's territory. The second dependent variable is rebel conflict, which aggregates monthly counts of violent rebel actor[47]-initiated material conflict events targeting government actors within a rebel's home-country.[48] To create these variables, 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 (monthly) count variables below. Each dependent variable has 5,040 observations across the 1997–2010 sample period. Histograms for government and rebel conflict are presented in the Supplemental Appendix, and indicate that the ranges of these variables are [0–98] and [0–126] conflicts per month, respectively.

Model Selection

Given the event-count nature of the dependent variables, the analysis sought to identify several suitable count models for the forecasting of the events at hand. To this end, one might first consider using a set of Poisson count models. However, the histograms presented above suggest that the government conflict and rebel conflict 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 and rebel conflict, 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, 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 through a parameterized relaxation of the mean–variance equality assumption. However, as discussed above, there is also strong reason to believe that many of the excess zeroes observed within the dependent variables under analysis are not true count-level zeroes, in the sense that they could ever take on values greater than zero.

Ignoring these structural zeroes by 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 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 data generating process (d.g.p.) – with an NB or Poisson likelihood equation that directly tests for the effect of one's covariates on the expected frequency of government or rebel conflict, conditional on the likelihood that a given observation was generated from the count-stage d.g.p. Accordingly, the expectation is that ZIP and ZINB models will be superior to Poisson and NB models for the modeling and forecasting of government and rebel conflict. Furthermore, due to the aforementioned presence of many extreme (high-count) values in government conflict and rebel conflict, conditional overdispersion was expected in these variables, even after accounting for the zero inflation described above. The analysis thus favored the ZINB model over the ZIP model for all zero-inflated forecasts.

Model selection statistics confirmed these suspicions. Vuong tests for non-nested models[49] are appropriate comparison tests for count models of interest, and these tests were used to compare ZIP, ZINB, NB, and Poisson models for all specifications presented below. Vuong tests indicated that across all specifications, the ZINB model outperforms the ZIP, Poisson, and NB models at the $p < 0.01$ level, while the NB model outperforms the ZIP and Poisson models at the $p < 0.01$ level. Likelihood ratio tests were also conducted where applicable, and similarly suggested both that the ZINB model is superior to the ZIP for the dependent variables, and that the NB model was superior to the Poisson model across all models compared. Information-based model selection criteria are also prominently featured in comparisons of count and zero-inflated models and therefore are applicable here.[50] Accordingly, Akaike information criterion (AIC) comparisons were calculated and compared for all models, with each comparison preferring the NB and ZINB models to comparable Poisson and ZIP models, as well as preferring ZINB models over NB models. To summarize, the zero-inflated, overdispersed nature of the dependent variables suggests that NB and ZINB count models should be used for modeling government and rebel conflict, and these are therefore the statistical forecasting models that are estimated and evaluated here.

Independent Variables

The primary independent variables used for forecasting government conflict and rebel conflict 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.[51] 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 the government conflict- and rebel conflict-dependent variables. The natural log of each lagged conflict variable (+0.05) was then taken prior to its inclusion on the right-hand side of these models in order to ensure that outliers were not disproportionately influencing the analysis.[52] For the ZINB models, all lagged conflict variables were then included within both the zero inflation stage and count stage estimating equations. As argued above, the justification for using lagged covariates within the inflation stage rests on the contention that recent levels of monthly conflict (or lack thereof) directly inform us, with ex-ante observability, as to which country-months are currently able to experience domestic conflict, and which cannot. If correct, this strategy will allow us to statistically partition the (potentially) 'inflated-zero' cases from the true 'count-zero' conflict cases, and to thereby improve the accuracy and precision of the count stage estimates and conflict forecasts discussed below.

Drawing on recent civil conflict studies, a number of control variables are also included in the models reported below. The count stages of the NB and ZINB models include yearly measures of the natural log of GDP per capita, GDP growth, and the natural log of a country's total population,[53] as – unlike many commonly studied correlates of intrastate conflict – these three variables have been found to enhance one's ability to predict civil war.[54] GDP per capita has also been found to be a strong predictor of a country's likelihood of ever experiencing domestic political violence,[55] and accordingly, the paper also includes ln GDPpc within the inflation stage of the ZINB models. Robustness tests re-ran all models discussed below (1) without ln GDPpc, GDP growth, or ln population (i.e., with only the lagged conflict measures included as covariates) and (2) with a range of additional controls added to each model.[56] All findings and conclusions remain unchanged under these alternative specifications. Finally, the analysis also explored the inclusion of additional monthly conflict lags as independent variables in the NB and ZINB models, and found that including lagged conflict measures beyond 3-month lags as independent variables generally does not improve the forecasting accuracy of these models, and in some cases reduces accuracy. Thus, conflict measures beyond 3-month lags are not included in the models reported directly below.

Results

All models are estimated on a training data-set encompassing the 1997–2004 country-month sample, with the aim of evaluating the forecasting accuracy of these model-estimates on a country-month validation data-set encompassing the years 2005–2010. The 1997–2004 training models thus serve as the primary models of reference here. Comparable ZINB and NB models were also estimated on the entire 1997–2010 data-set 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. Coefficient estimates, standard errors, and goodness-of-fit statistics for the training models of government conflict and rebel conflict are presented in Table A.1 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 the government conflict ZINB models. The findings are similar for rebel conflict, and again the NB model appears to overestimate count-stage coefficient estimates and standard errors, which is consistent with expectations of zero inflation. Here, for the rebel conflict models given in Table A.1, 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.

These findings contribute to substantive understandings of conflict processes in several manners. Consistent with past literature,[57] the 'outcome-stage' results for the rebel and government conflict models indicate that civil conflict between governments and domestic groups follows both an action–reaction pattern – wherein each side reciprocates each other's (increased or decreased) conflict behaviors – and (ii) a bureaucratic-inertial tendency – wherein each side perpetuates their own past policies irrespective of the other's actions. One key implication of the latter phenomenon, as discussed in extant research,[58] is that single-month reductions in civil conflict frequency by either side, although likely to be reciprocated by the other side, are unlikely to lead to lasting or significant reduction in overall conflict unless repeated signals are used. Lastly, the ZINB model estimates also indicate that the magnitude of these action–reaction and inertial relationships is overstated when zero inflation is ignored, implying that governments and rebel groups may not base their conflict decisions as consummately on past levels of conflict as previously thought.

Classification Matrices

To better evaluate the relative performances of the NB and ZINB models in terms of conflict forecasting, this section next presents a set of classification matrices for the government and rebel conflict-dependent variables. These matrices were created by first calculating in-sample and out-of-sample NB model predictions for the government conflict and rebel conflict counts using the NB expected value formula:

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

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

$$E(y_{it}|x_{it}, z_{it}) = \mathrm{e}^{x_{it}\hat{\beta}} - \pi_{it}\mathrm{e}^{x_{it}\hat{\beta}},$$

where, because the inflation equations used above follow a logistic cumulative distribution function,

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

and where here r_0 corresponds to the zero-only regime, $\hat{\gamma}$ corresponds to the 1997–2004 ZINB inflation stage coefficient estimates, z_{it} corresponds to the ZINB inflation stage covariates, $\hat{\beta}$ corresponds to the 1997–2004 ZINB outcome-stage coefficient estimates, x_{it} are the outcome stage covariates, and $E(y_{it}|x_{it})$ are the ZINB expected event counts. These in-sample and out-of-sample count forecasts were then used to derive classification matrix statistics, and corresponding 95 per cent CIs, for each set of models. Specifically, this paper calculated five classification matrix statistics for each model by dichotomizing the forecasted and observed counts in order to evaluate model prediction accuracy 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, the present analysis calculated and reported five relevant classification statistics for each threshold of interest: sensitivity, specificity, negative predicted values (NPVs), positive predicted values (PPVs), and the percentage correctly classified.[59] Sensitivity reports the proportion of actual conflict country-months that were correctly identified as conflict-months (for a given threshold) by the forecasting models. Specificity reports the proportion of peace-country-months that were correctly identified as such by the models. PPVs refer to the proportion of conflict-country-month forecasts that were actually observed to be conflict-country-months within the sample. NPVs refer to the proportion of peace-country-month forecasts that were actually observed to be peace-country-months within the sample. Finally, the 'correctly classified' statistic reports the percentage of cases within a given sample that were actually classified as either peace or conflict by a forecasting model.

The full set of classification statistics and corresponding 95 per cent confidence intervals – for the in-sample (1997–2004) and out-of-sample (2005–2010)

TABLE 1
CLASSIFICATION

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

Note: Values in parentheses are 95% CIs.

government and rebel conflict forecasts – are presented in Table 1. Beginning with government conflicts, Table 1 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 sensitivity statistics given in Table 1 indicate that the out-of-sample ZINB models are on average 8.4 percentage points better at accurately forecasting country-months that experience at least one conflict (sensitivity = 81.97 per cent) and at-least five conflicts (sensitivity = 82.72 per cent) than the NB models (sensitivity = 68.72 per cent and 79.58 per cent). At the same time, across all government conflict specifications reported in Table 1, the ZINB and NB models perform comparably well in terms of cases correctly classified (90.22 per cent ↔ 95.70 per cent), specificity (94.46 per cent ↔ 97.54 per cent), and NPV (91.35 per cent ↔ 97.99 per cent), which is unsurprising given the overabundance of zero observations within the samples of interest. Regarding the government conflict PPV statistics reported in Table 1, the NB model does do on average 4.7 percentage points 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 over-prediction of zeroes – and an under-prediction of government conflict – by the 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 predicting conflict-country-months, within both in-sample and out-of-sample settings.

One can draw similar conclusions from the classification statistics and confidence intervals that are reported for rebel conflict 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 percentage points better at accurately forecasting country-months that experience at least one conflict (sensitivity = 80.05 per cent) and at-least five conflicts (sensitivity = 80.41 per cent), relative to comparable NB models (sensitivity = 68.31 and 77.84 per cent). Across all rebel conflict specifications reported in Table 1 the ZINB and NB models again perform comparably well in terms of cases correctly classified (89.45 per cent ↔ 95.86 per cent), specificity (93.42 per cent ↔ 97.96 per cent), and NPV (90.04 per cent ↔ 97.71 per cent), which again is unsurprising given the overabundance of zero cases within the samples of interest. Finally, the PPVs reported in Table 1 suggest that the NB model does on average 5.9 percentage points better than ZINB models, which is likely the result of an over-prediction of zeroes, and an under-prediction of rebel conflict (for each conflict threshold) by the NB models. Thus, while both the NB and ZINB models do a comparable job of predicting peace-months, the ZINB model is again superior to the NB model in terms of predicting actual instances of rebel conflicts, within both an in-sample and out-of-sample setting.

Marginal Calibration Diagrams

For a more comprehensive evaluation of the government and rebel conflict forecasting models, one can instead compare the *marginal calibration* of the NB and ZINB count forecasts to the actual count values observed in the true (training and validation) data-sets. 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 when one's average observed count forecasts equal one's average probabilistic forecasts.[60] To calculate marginal calibrations for the count models discussed here, define P as a predictive probability distribution on the set of nonnegative integers resulting from the probabilistic forecasts derived from a series of count models. Assuming then that each observed count, $x^{(it)}$, is a random draw from its probabilistic forecast, a histogram of these observed counts is statistically comparable to the composite distributions of one's aggregated predictive distributions $P^{(it)}$.[61] One can thus represent these aggregations graphically via a marginal calibration diagram, which compares the predicted frequencies,

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

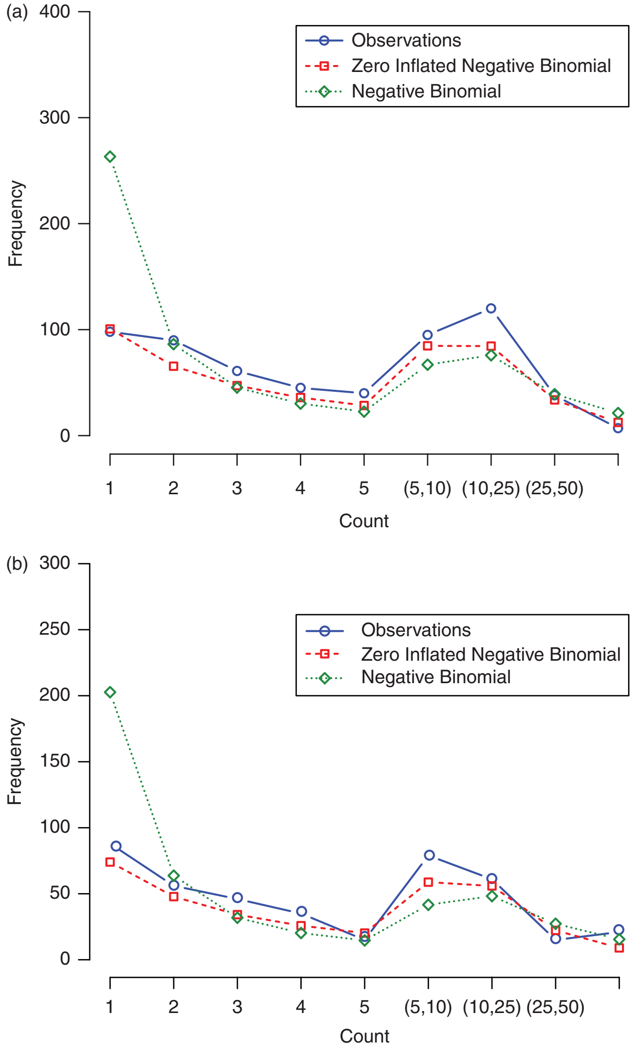
for specific x values or intervals $(x_a, x_b]$, to their empirical counterparts

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

in an extension of the marginal calibration diagram formulas presented in extant research.[62] 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 the 1997–2004 government and rebel conflict in-sample predictions, and for the 2005–2010 out-of-sample forecasts. These diagrams appear in Figures 1 and 2. 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.

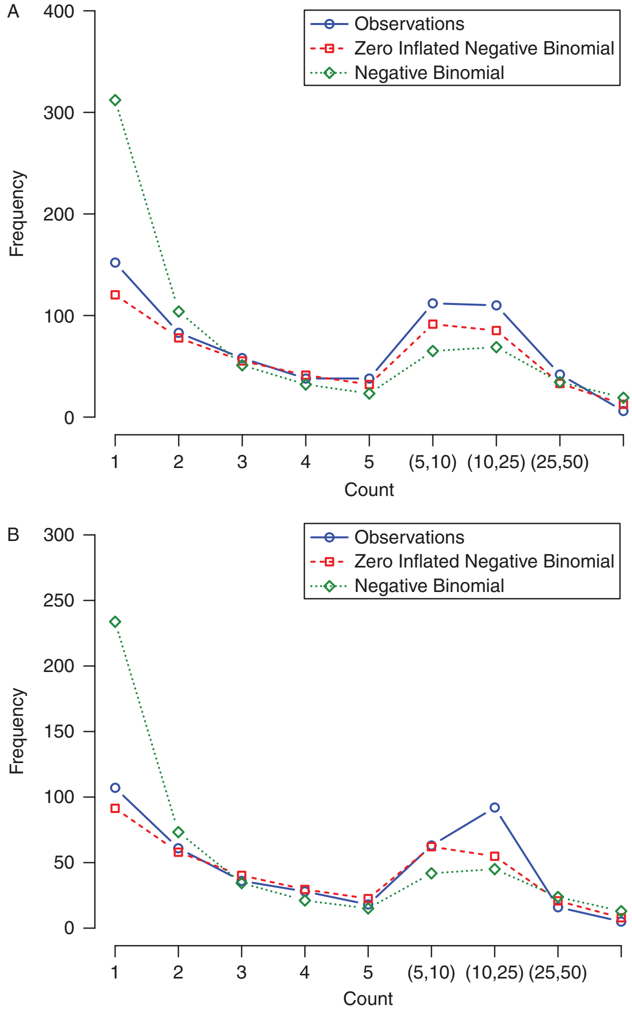
Figure 1 reports marginal calibration diagrams for the in-sample (1997–2004) and out-of-sample (2005–2010) forecasts of government conflict. It suggests that, although both models do a competent job of predicting government-initiated conflicts, the ZINB models are superior to the NB models in calibration. To see this,

FIGURE 1
MARGINAL CALIBRATION DIAGRAMS FOR GOVERNMENT-INITIATED CONFLICT. (A) IN-SAMPLE PREDICTIONS AND (B) OUT-OF-SAMPLE PREDICTIONS



the present section focuses on the out-of-sample predictions (Figure 1(b)). Turning to Figure 1(b), note first that the NB model substantially over-predicts the number of country-months experiencing a single instance of government-initiated civil conflict (by 130 per cent) while the ZINB model only slightly under-predicts the number of country-month instances of a single observed government-initiated conflict within the validation sample (by 16 per cent). The ZINB and NB models each do a fairly accurate job of forecasting observations with observed monthly conflict counts lying

FIGURE 2
MARGINAL CALIBRATION DIAGRAMS FOR CITIZEN-INITIATED CONFLICT. (A) IN-SAMPLE PREDICTIONS AND (B) OUT-OF-SAMPLE PREDICTIONS



between two and five (inclusive). However, as one begins to aggregate across higher levels of government-initiated conflict counts shown in Figure 1, one sees 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 are observed across bins (5;10], (10;25], and (25;50]. While the ZINB model under-predicts observations lying within these conflict thresholds by an average of 31 per cent, comparable NB predictions are off by an average of 54 per cent. The models perform comparably in predicting the (exceedingly rare) frequencies of monthly government-initiated conflicts lying within the final (50;100] interval. In sum, the NB model of government conflict over-predicts low-level country-month instances of government-initiated civil conflict and under-predicts higher levels of monthly conflict. By contrast, ZINB models of government conflict better predict monthly conflict across this variable's entire range.

Figure 2 presents marginal calibration diagrams of the ZINB and NB in-sample (1997–2004) and out-of-sample (2005–2010) predictions of rebel conflict. As above, this Figure suggests that the ZINB model is superior in calibration to the NB model, and a discussion of the out-of-sample predictions in Figure 2(b) elucidates this point. Here, the NB model over-predicts the number of country-months experiencing a single instance of rebel-initiated conflict by 119 per cent. By contrast, the 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 per cent. The ZINB and NB models each does 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, one finds in Figure 2 that as above, the 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 the ZINB model are off by an average 24 per cent, whereas comparable NB model predictions are off by an average 45 per cent. Lastly, although both models do a comparable job of predicting the frequencies of monthly rebel-initiated conflicts lying within the final (50;100] range, the ZINB model frequency forecasts are moderately closer to the actual count frequencies. Hence, the NB model over-predicts country-month instances of (rebel-initiated) conflict for low-level conflict counts and under-predicts higher levels of monthly conflict. By contrast, the ZINB model accurately predicts conflict across the entire range of monthly rebel-initiated conflicts.

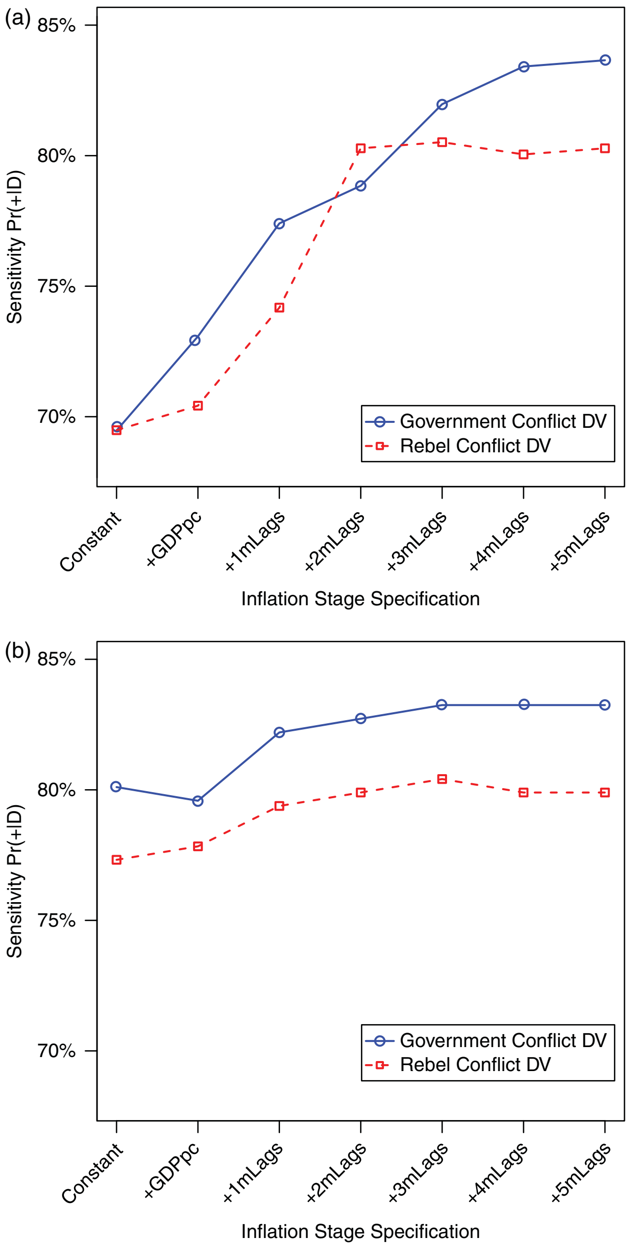
ZINB Comparisons

While the above analysis demonstrates the superiority of the ZINB conflict models over comparable NB models, it is only suggestive as to the forecasting advantages of including lagged conflict measures within the inflation stage of the ZINB models. To better assess the latter, the present section builds upon the ZINB analysis presented above by incrementally adding-in a number of lagged conflict variables to the inflation stages of the ZINB models. While doing so, the analysis holds 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 and rebel conflict. For the inflation stages, this exercise begins with a ZINB model reporting only an inflation stage constant, and then adds ln GDPpc to this stage, evaluating the results at both steps. The analysis next sequentially adds 1-to-5 month lagged values of government conflict and rebel conflict to the ZINB inflation stage, again evaluating the results at each step. For each dependent variable, the resultant seven (nested) ZINB specifications are then compared via a number of model fit statistics. Vuong tests indicate that for all ZINB models, the inclusion of each successive pair of lagged government conflict and rebel conflict inflation stage covariates produces a significant ($p < 0.01$) improvement in model fit and model performance. Likelihood ratio tests similarly suggest that the addition of 1–5-month lags of government and rebel conflict to the ZINB inflation stage produces a significant ($p < 0.01$) improvement in model fit. Finally, in comparing the AICs of the ZINB models, each pair-wise comparison preferred a more fully specified ZINB model to a ZINB model with fewer inflation stage (lagged conflict) covariates. Thus, a wide range of model fit statistics further confirm that past levels of conflict serve as significant, robust predictors of zero inflation.

To determine whether lagged inflation-stage covariates affect conflict forecasts, these additional models were evaluated with a series of sensitivity plots. These plots compare the sensitivity levels of the out-of-sample conflict predictions for the seven sets of (government conflict and rebel conflict) ZINB model variants described above. The sensitivity statistics used in these plots report the proportion of actual conflicts that one's models predicted as such, and thus are particularly useful in comparing the conflict-forecasting accuracy of ZINB models. Sensitivity plots are similar in motivation to predictive power plots,[63] with the exception that the present analysis plots the sensitivity levels of each model's forecasts, rather than the total-area under a receiver operating characteristic curve. The former are preferred not only because the (predicted) counts under examination encompass values greater than 1, but also because the extreme proportion of zeroes in the sample can obscure any differences in actual sensitivity levels when aggregating across all discrimination thresholds. As for the classification matrices, government conflict and rebel conflict sensitivity statistics were calculated for both the 'at least 1-monthly conflict' and 'at least 5-monthly conflicts' thresholds. This process was repeated iteratively as more lagged values of government and rebel conflict were incrementally added to the ZINB models' inflation stages (beginning with a model that includes only a constant in the inflation stage). The resultant sensitivity plots for 1-monthly conflict and 5-monthly conflict thresholds appear in Figure 3.

Beginning with Figure 3(a), observe that for government conflict, Figure 3(a) indicates that adding ln GDPpc to the inflation stage increases one's ability to accurately predict instances of government-initiated conflict by roughly 4 percentage points. By comparison, subsequently adding 1, 1–2, or 1–3-month lags of government and rebel conflict to the inflation stage of the government conflict ZINB models produces 8, 9, and 13 percentage points increases in sensitivity, again relative to the 'constant-only' inflation stage model. The rebel conflict ZINB model exhibits comparable increases in sensitivity (of 5, 11, and 11 percentage points) for these same three pairs of (1–3-month) lagged conflict models, although ln GDPpc 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

FIGURE 3
SENSITIVITY PLOTS FOR ZINB MODELS. (A) CONFLICT THRESHOLD = 1 AND (B) CONFLICT THRESHOLD = 5



conflict lags is negligible in either model. Indeed, in the case of rebel conflict, adding 4 or 4–5 month conflict lags to the inflation stage decreases sensitivity relative to the 1–3 lagged conflict specification. For the government conflict 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 one's ability to distinguish between inflated and non-inflated peace months appear 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 3(a) accordingly demonstrates that the addition of lagged conflict values to the inflation stage of the ZINB models produces a marked improvement in the accuracy of one's conflict forecasts.

Relative to Figure 3(a), the sensitivity plots in Figure 3(b) report higher initial levels of sensitivity, and hence one finds smaller additional gains in sensitivity across all covariate additions. Nevertheless, the trends shown in Figure 3(b) 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 the ZINB models increases the sensitivity levels of the government conflict and rebel conflict forecasts. By contrast, the addition of ln GDPpc to the inflation stages of these government conflict and rebel conflict models yields an average improvement in sensitivity of roughly 0 percentage points. 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 the government conflict or rebel conflict inflation stages. Hence, the benefit of including lagged levels of conflict within the inflation stages of these models dissipates after approximately 3 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, or de-facto truces appear to also preclude government and rebel actors from fighting for short periods of time. Modeling these dynamics allows us to better identify, and predict, the observations that are most likely to experience conflict in a given month.

CONCLUSION

While the study of conflict-event counts has grown markedly in recent years, ambiguities remain with regard to proper model choice and count forecasting techniques. This paper addresses these two deficiencies by refining and presenting tools that allow conflict modelers to evaluate and compare conflict count models for predictive accuracy and proper specification. In doing so, the present project verifies for the first time that zero-inflated count models improve one's ability to *forecast* monthly frequencies of rebel- and government-initiated conflicts, relative to non-inflated approaches. Notably, using zero-inflated models in conjunction with lagged conflict covariates gives the former models a decided advantage over comparable approaches. On average, the ZINB models discussed above were roughly 8 percentage points 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. Given the lasting uncertainty surrounding (1) whether zero-inflated count models are appropriate for conflict modeling to begin with and (2) the proper inflation and outcome stage specifications for such models, the forecasting techniques presented here thereby provide a verifiable framework for future researchers to use in theoretically advancing this research agenda.

This paper also makes several broader contributions to conflict-forecasting methodology. Event counts and related discrete (un)ordered outcomes are integral to the study of civil conflict. One 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.[64] This paper addresses these deficiencies by providing the first forecasting assessment of civil conflict frequency, while using a recently developed data-set that credibly measures monthly conflict frequency for both rebel and government initiators. Doing so provides examples of several techniques that one can use in assessing the accuracy, specificity, and sensitivity of one's count forecasts. To this end, the above analysis presents 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. Through these examples, this paper is intended to 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, this paper suggests that recent levels of rebel and government material conflict have a direct, positive effect on present levels of conflict, which is consistent with theories of conflict reciprocity and conflict inertia.[65] Identifying these patterns within the 29 Asian countries included in ICEWS is notable, as doing so demonstrates that the findings reported in previous country-specific studies of reciprocal and inertial conflict processes[66] are generalizable across a wider spectrum of countries, time periods, and political actors. However, the magnitudes of these reciprocal and inertial relationships are also found to be overestimated when zero-inflation is ignored, implying that the substantive effect of action–reaction and inertial patterns in conflict processes may be overstated. 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 researchers' 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, structurally inflated peace-periods.

Thus, for researchers interested in the direct effect of explanatory variables on civil conflict, the above findings together 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. Given the challenges associated with coding additional civil conflict covariates in forecasting models as one moves to smaller-and-smaller units of aggregation or real-time forecasting, lagged conflict variables will be especially useful in these contexts. In fact, the approach outlined here would likely yield even larger improvements in forecasting accuracy when applied to data-sets 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.

ACKNOWLEDGEMENTS

I thank the editors and reviewers at Civil Wars, as well as Phil Schrodt, John Freeman, Daniel Hill, Will Moore, Glenn Palmer, and Jon Pevehouse for their helpful comments and suggestions.

SUPPLEMENTAL DATA

Supplemental data for this article can be accessed at doi:10.1080/13698249.2015.1059564.

NOTES ON CONTRIBUTOR

Benjamin E. Bagozzi received his Ph.D. in 2013 from the Pennsylvania State University, and is an Assistant Professor of political science at the University of Delaware. His research centers on the development of methods for the study of international relations and political violence. Other research interests include international political economy, international environmental politics, and international organization. Current address: Department of Political Science and International Relations, University of Delaware, 347 Smith Hall, 18 Amstel Avenue, Newark, DE 19716 USA. Email: bagozzib@gmail.com.

NOTES

1. Philip A. Schrodt, 'Precedents, Progress, and Prospects in Political Event Data', *International Interactions* 38 (2012) pp.546–69. For prominent examples, Sean P. O'Brien, 'Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research', *International Studies Review* 12 (2010) pp.87–104; and Idean Salehyan, Cullen S. Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams, 'Social Conflict in Africa: A New Database', *International Interactions* 38 (2012), pp.503–11.
2. Most notably the Poisson or NB models and their zero-inflated counterparts.
3. E.g., Michael T. Koch and Skyler Cranmer, 'Testing the "Dick Cheney" Hypothesis: Do Governments of The Left Attract More Terrorism Than Governments of The Right?' *Conflict Management and Peace Science* 24 (2007) pp.311–26; James A. Piazza, 'Poverty, Minority Economic Discrimination, and Domestic Terrorism', *Journal of Peace Research* 48 (2011) pp.339–

53; and Cullen S. Hendrix and Idean Salehyan, 'Climate Change, Rainfall, and Social Conflict in Africa', *Journal of Peace Research* 49 (2012) pp.35–50.

4. Nathanial Beck, Gary King, and Langche Zeng, 'Improving Quantitative Studies of International Conflict: A Conjecture', *American Political Science Review* 94 (2000) pp.21–35.
5. Which alone is not sufficient justification for the use of zero-inflated models: see Dominique Lord, Simone P. Washington, and John N. Ivan, 'Poisson, Poisson-gamma and Zero-inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory', *Accident Analysis & Prevention* 37 (2005) pp.35–46 and Quan Li, 'Does Democracy Promote or Reduce Transnational Terrorist Incidents?' *Journal of Conflict Resolution* 49 (2005) pp.278–297, 293.
6. See Martin Ridout, John Hinde, and Clarice G.B. Demétrio, 'A Score Test for Testing a Zero-inflated Poisson Regression Model Against Zero-inflated Negative Binomial Alternatives', *Biometrics* 57 (2001) pp.219–23 and Daniel B. Hall and Zhengang Zhang, 'Marginal Models for Zero-inflated Clustered Data', *Statistical Modelling* 4 (2004), pp.160–80.
7. E.g., Koch and Cranmer (note 3) pp.311–26 and Hendrix and Salehyan (note 3) pp.35–50.
8. Charles C. Branas, Therese S. Richmond, and C. William Schwab, 'Firearm Homicide and Firearm Suicide: Opposite But Equal', *Public Health Reports* 119 (2004) pp.114–24 and James I Walsh and James A. Piazza, 'Why Respecting Physical Integrity Rights Reduces Terrorism', *Comparative Political Studies* 43 (2010) pp.551–77.
9. See Håvard Hegre, Gudrun Østby, and Clionadh Raleigh, 'Poverty and Civil War Events: A Disaggregated Study of Liberia', *Journal of Conflict Resolution* 53 (2009) pp.598–623 and Michael G. Findley, James A. Piazza, and Joseph K. Young, 'Games Rivals Play: Terrorism in International Rivalries', *Journal of Politics* 74 (2012) pp.235–48.
10. Li (note 5) pp.278–97 and Daniel Masters and Patricia Hoen, 'State Legitimacy and Terrorism', *Democracy and Terrorism* 8 (2012) pp.337–57.
11. Jack A. Goldstone, Robert H. Bates, David L. Epstein, Ted Robert Gurr, Michael B. Lustik, Monty G. Marshall, Jay Ulfelder, and Mark Woodward, 'A Global Model for Forecasting Political Instability', *American Journal of Political Science* 54 (2009) pp.190–208; Michael D. Ward, Brian D. Greenhill, and Kristen M. Bakke, 'The Perils of Policy by *P*-Value: Predicting Civil Conflicts', *Journal of Peace Research* 47 (2010) pp.363–75; and Patrick T. Brandt, John R. Freeman, and Philip A. Schrodt, 'Real Time, Time Series Forecasting of Inter and Intra-State Political Conflict', *Conflict Management and Peace Science* 28 (2011) pp.41–64.
12. Philip A. Schrodt, James Yonamine, and Benjamin E. Bagozzi, 'Data-based Computational Approaches to Forecasting Political Violence' in V. Subrahmanian (ed.) *Handbook of Computational Approaches to Counterterrorism* (New York, NY: Springer Press 2013) pp.129–62.
13. Claudia Czado, Tilmann Gneiting, and Leonhard Held, 'Predictive Model Assessment for Count Data', *Biometrics* 65 (2009) pp.1254–61.
14. Jacob Cohen, 'The Cost of Dichotomization', *Applied Psychological Measurement* 7 (1983) pp.249–53.
15. See Robert C. MacCallum, Shaobo Zhang, Kristopher J. Preacher, and Derek D. Rucker, 'On the Practice of Dichotomization of Quantitative Variables', *Psychological Methods* 7 (2002) pp.624–38 and Patrick Royston, Douglas G. Altman, and Willi Sauerbrei, 'Dichotomizing Continuous Predictors in Multiple Regression: A Bad Idea', *Statistics in Medicine* 25 (2006) pp.127–41.
16. Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E. Raftery, 'Probabilistic Forecasts, Calibration and Sharpness', *Journal of the Royal Statistical Society: Series B Statistical Methodology* 69 (2007) pp.243–68 and Czado, Gneiting, and Held (note 13) pp.1254–61.
17. Where, this paper's focus is on the frequency of *individual* civil conflict *events* arising during a larger period of conflict between government and rebel actors.
18. David H. Clark and Patrick M. Regan, 'Opportunities to Fight: A Statistical Technique for Modeling Unobservable Phenomena', *Journal of Conflict Resolution* 47 (2003) pp.94–115 and Benjamin E. Bagozzi, Daniel W. Hill, Will H. Moore, and Bumba Mukherjee, 'Modeling Two Types of Peace: The Zero-inflated Ordered Probit (ZiOP) Model in Conflict Research', *Journal of Conflict Resolution* 59 (2015) pp. 728–752.
19. See Zeev Maoz and Bruce M. Russett, 'Normative and Structural Causes of Democratic Peace, 1946–1986', *American Political Science Review* 87 (1993) pp.624–38 for a discussion of 'relevant dyads'.
20. Douglas Lemke and William Reed, 'The Relevance of Politically Relevant Dyads', *Journal of Conflict Resolution* 45 (2001) pp.125–44; Clark and Regan (note 18) pp.94–115; and Jun Xiang,

'Relevance as a Latent Variable in Dyadic Analysis of Conflict', *Journal of Politics* 72 (2010) pp.484–98.

21. D. Scott Bennett and Alan Stam, *The Behavioral Origins of War* (Ann Arbor, MI: University of Michigan Press 2006) p.61.
22. Lemke and Reed (note 20) pp.125–44 and Xiang (note 20) pp.484–98.
23. E.g., Clark and Regan (note 18) pp.94–115 and Xiang (note 20) p.484–98.
24. E.g., Lisa Hultman, 'Battle Losses and Rebel Violence: Raising the Costs for Fighting Terrorism and Political Violence', *Terrorism and Political Violence* 19 (2007) pp.205–22; Hegre, Østby, and Raleigh (note 9) pp.598–623 and Piazza (note 3) pp.339–53.
25. Jennifer S. Holmes, Sheila Amin Gutiérrez De Piñeres, and Kevin M. Curtina, 'A Subnational Study of Insurgency: FARC Violence in the 1990s', *Studies in Conflict & Terrorism* 30 (2007) pp.249–65.
26. Hegre, Østby and Raleigh (note 9) pp.598–623.
27. Bagozzi *et al.* (note 18).
28. Ted R. Gurr, 'Why Minorities Rebel: A Global Analysis of Communal Mobilization and Conflict Since 1945', *International Political Science Review* 14 (1994) pp.161–201.
29. Beck, King, and Zeng (note 4) pp.21–35; Bennett and Stam (note 21); and Ward, Greenhill, and Bakke (note 11) pp.363–75.
30. Ward, Greenhill, and Bakke (note 11) pp.363–75.
31. E.g., Ted R. Gurr and Mark I. Lichbach, 'Forecasting Internal Conflict: A Competitive Evaluation of Empirical Theories', *Comparative Empirical Studies* 19 (1986) pp.3–38 and O'Brien (note 1) pp. 87–104.
32. Nathanial Beck, Jonathan N. Katz, and Richard Tucker, 'Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable', *American Journal of Political Science* 42 (1998) pp.1260–88.
33. Jon C. Pevehouse and Joshua S. Goldstein, 'Serbian Compliance or Defiance in Kosovo? Statistical Analysis and Real-time Predictions', *Journal of Conflict Resolution* 43 (1999) pp.538–46 and Stephen M. Shellman, Claire Hatfield, and Maggie J. Mills, 'Disaggregating Actors in Intranational Conflict', *Journal of Peace Research* 47 (2010) pp.83–90.
34. E.g., as a result of geography or slow moving variables such as institutions and GDP.
35. Andrew H. Kydd and Barbara F. Walter, 'Sabotaging the Peace: The Politics of Extremist Violence', *International Organization* 56 (2002), pp.263–96 and Anthony Wanis-St. John, 'Back-Channel Negotiation: International Bargaining in the Shadows', *Negotiation Journal* 22 (2006) pp.119–44.
36. Robert Axelrod, *The Evolution of Cooperation* (New York, NY: Basic Books 1984).
37. 'British Troops in Secret Truce with Taliban', *The Sunday Times* 2006.
38. Piotr S. Wandycz, 'Secret Soviet-Polish Peace Talks in 1919', *Slavic Review* 24 (1965) pp.425–49.
39. Padraig Lehinan, *Confederate Catholics at War, 1641–49* (Cork, Ireland: Irish Committee of Historical Sciences 2002) p.73.
40. Elisabeth J. Wood, *Insurgent Collective action and Civil War in El Salvador* (Cambridge, UK: Cambridge University Press 2003).
41. O'Brien (note 1) pp.87–104. 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.
42. Philip A. Schrodt, 'Tabari: Textual Analysis by Augmented Replacement Instructions, version 7.0', 2009. http://eventdata.parusanalytics.com/tabari.dir/tabari.manual.0.7.6.pdf
43. Philip A. Schrodt and Omur Yilmaz, 'Cameo Conflict and Mediation Event Observation Codebook', 2007. http://eventdata.parusanalytics.com/data.dir/cameo.html
44. Vito D'Orazio, Jay Yonamine, and Philip A. Schrodt, *Predicting Intra-state Conflict Onset: An Event Data Approach Using Euclidean and Levenshtein Distance Measures*, Paper presented at the 69th annual Midwest Political Science Association meeting, 2011. p.4.
45. Government members, including members of governing parties and coalition partners; military troops, soldiers, state-military personnel; and police forces and officers were all considered to be 'government actors'.
46. Examples of material conflict events include fights with small arms and light weapons, attempts to assassinate actors, and car bombings; material conflict excludes verbal conflict behaviors such as verbal threats or accusations.
47. Domestic rebels (armed and violent groups or individuals), insurgents, and separatist groups were all treated as 'violent rebel actors'.

48. Conflict measures are separated into government and rebel-initiated conflicts as 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 (note 33) pp.83–90.
49. Quang H. Vuong, 'Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses', *Econometrica* 57 (1989) pp.307–33.
50. Czado, Gneiting, and Held (note 13) pp.1254–61.
51. D'Orazio, Yonamine, and Schrodt (note 44).
52. This did not dramatically affect the results, although it did moderately improve the calibration of the NB forecasts (the ZINB forecasts remained relatively unchanged).
53. These measures are from World Bank, 'World Development Indicators', http://data.worldbank.org/data-catalog/world-development-indicators 2011.
54. Ward, Greenhill, and Bakke (note 11) pp.363–75.
55. Bagozzi *et al.* (note 18).
56. 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.
57. Joshua S. Goldstein and John R. Freeman, *Three-Way Street: Strategic Reciprocity in World Politics* (Chicago, IL: University of Chicago Press 1990); Will H. Moore, 'Action-reaction or Rational Expectations? Reciprocity and the Domestic-international Conflict Nexus During the "Rhodesia Problem"', *Journal of Conflict Resolution* 39 (1995) pp.129–67; and Ronald A. Francisco, 'The Relationship Between Coercion and Protest: An Empirical Evaluation in Three Coercive States', *Journal of Conflict Resolution* 39 (1995) pp.263–82.
58. Goldstein and Freeman (note 57) and Moore (note 57) pp.129–67.
59. Classification formulas appear in the Supplemental Appendix.
60. As $T \rightarrow \infty$ and provided that all mass is placed on finite values. Gneiting, Balabdaoui, and Raftery (note 16) pp.243–68.
61. Czado, Gneiting, and Held (note 13) pp.1254–61.
62. Ibid.
63. Ward, Greenhill, and Bakke (note 11) pp.363–75.
64. Czado, Gneiting, and Held (note 13) pp.1254–61.
65. Goldstein and Freeman (note 57); Moore (note 57) pp.129–67; and Francisco (note 57) pp.263–82.
66. E.g., Moore (note 57) pp.129–67 and Francisco (note 57) pp.263–82.