International Studies Quarterly (2024) **68**, sqae137

Inference with Extremes: Accounting for Extreme Values in Count Regression Models

RESEARCH NOTE

David Randahl and Johan Vegelius
Uppsala University, Sweden

Processes that occasionally, but not always, produce extreme values are notoriously difficult to model, as a small number of extreme observations may have a large impact on the results. Existing methods for handling extreme values are often arbitrary and leave researchers without guidance regarding this problem. In this paper, we propose an extreme value and zero-inflated negative binomial (EVZINB) regression model, which allows for separate modeling of extreme and nonextreme observations to solve this problem. The EVZINB model offers an elegant solution to modeling data with extreme values and allows researchers to draw additional inferences about both extreme and nonextreme observations. We illustrate the usefulness of the EVZINB model by replicating a study on the effects of the deployment of UN peacekeepers on one-sided violence against civilians.

Los procesos que producen, de manera ocasional, pero no siempre, valores de carácter extremo son notoriamente difíciles de modelar, debido a que un pequeño número de observaciones con carácter extremo puede tener un gran impacto en los resultados. Los métodos existentes para manejar estos valores extremos son, con frecuencia, arbitrarios y dejan a los investigadores sin orientación con respecto a este problema. En este artículo, proponemos, con el fin de resolver este problema, un modelo de regresión de Valor Extremo y de Binomio Negativo Inflado Cero (EVZINB, por sus siglas en inglés), que permite modelar por separado las observaciones extremas y las observaciones no extremas. El modelo EVZINB ofrece una solución elegante para modelar aquellos conjuntos de datos que tienen valores extremos y permite a los investigadores hacer inferencias adicionales sobre las observaciones extremas y las no extremas. Ilustramos la utilidad del modelo EVZINB replicando un estudio sobre los efectos del despliegue de las fuerzas de paz de la ONU sobre la violencia unilateral contra los civiles.

Nous savons bien que les processus qui tendent parfois, mais pas tout le temps, à produire des valeurs extrêmes sont difficiles à modéliser. En effet, un petit nombre d'observations extrêmes peut entraîner des conséquences importantes sur les résultats. Les méthodes existantes de traitement des valeurs extrêmes sont souvent arbitraires et ne fournissent aucun conseil aux chercheurs concernant ce problème. Dans cet article, nous proposons le modèle de régression extreme value and zero-inflated negative binomial (EVZINB ou Valeur extrême et binomial négatif avec excès de zéros), qui permet de modéliser séparément les observations extrêmes et non extrêmes, pour résoudre ce problème. Le modèle EVZINB propose une solution élégante de modélisation des données avec des valeurs extrêmes et permet aux chercheurs d'effectuer d'autres déductions quant aux observations extrêmes et non extrêmes. Nous illustrons l'utilité du modèle EVZINB en répliquant une étude des effets du déploiement des forces de maintien de la paix de l'ONU sur les violences unilatérales à l'encontre de civils.

Introduction

As the availability and resolution of data increases in the political and social sciences, researchers are increasingly faced with difficult modeling decisions when the data they are trying to model do not conform to well-behaved distributions. For example, when modeling the number of fatalities from one-sided violence (OSV) against civilians (Eck and Hultman 2007) on a country-month or country-year level, it can be assumed that certain countries will never experience any fatalities from OSV against civilians simply because they are not at risk of political violence and thus produce "structural" zeroes (see, for instance, Bagozzi 2015). Conversely, in other countries there may be periods of very large counts of OSV against civilians during campaigns of ethnic cleansing or genocide, which can be thought of as a separate process from the vast majority of cases. Fitting a regular count regression model in these circumstances may produce inaccurate inferences if the factors that cause the structural zeros or the extreme counts are different from the factors that cause the nonextreme positive counts. The former of these problems, excessive structural zeroes, has received much attention in the literature, and it has been shown that zero-inflated count models are appropriate from both an empirical and theoretical level (for a longer discussion on zero-inflated models, see Hilbe 2011, ch. 11, and Bagozzi 2015).

The latter problem, that of the influence of extreme values on the estimation of count regression model, has, however, received little or no attention in the literature, and it is this issue that this paper aims to address. In this paper, we propose an *extreme value and zero-inflated negative*

David Randahl (1990) holds a PhD in peace and conflict research at Uppsala University (2022). He is a Postdoctoral Researcher at the Department of Peace and Conflict Research, Uppsala University (2022–present).

Johan Vegelius (1981) holds a PhD in statistics (2022) and PhD in physics (2011) from Uppsala University. He is a Postdoctoral Researcher at the Department of Medical Sciences, Uppsala University (2022–present).

Authors' notes: We would like to thank Lovisa Vildö, Håvard Hegre, Espen Geelmuyden Rød, Shaobo Jin, Fan Yang Wallentin, and all the members of the Monday Group on Wednesdays for thoughtful comments and invaluable support throughout the writing process. Equal authorship is implied.

Randahl, David, and Johan Vegelius. (2024) Inference with Extremes: Accounting for Extreme Values in Count Regression Models. *International Studies Quarterly*, https://doi.org/10.1093/isq/sqae137

binomial (EVZINB) regression model to allow for modeling of count data with extreme values. This EVZINB model allows researchers to model data that contains extreme values without arbitrary inclusion or exclusion criteria, and allows researchers to draw inferences about which factors influence the likelihood of extreme values and which factors influence the "extremeness" of the extreme values. We show the empirical utility of the EVZINB model by replicating a study on the effects of the deployment of UN peacekeeping forces on OSV against civilians by rebel groups and governments (Hultman, Kathman, and Shannon 2013). The results of the replication study show that while the conclusion of the original study, that an increase in UN peacekeeping troops leads to a lower level of OSV against civilians, holds true in the median case, the EVZINB model allows us to show that this relationship is more complex for more extreme cases and does not hold for the total overall effect, where no statistically significant effect is seen. In addition, we show that the EVZINB model outperforms the negative binomial (NB) and zero-inflated negative binomial (ZINB) regression models on a number of crucial evaluation metrics.

This paper proceeds by motivating the EVZINB model empirically and statistically. This is followed by the replication study on the effects of UN peacekeeping forces on OSV against civilians by rebel groups and governments. In the final sections we discuss the wider applicability of the EVZINB model. In the online appendix, we also provide a small supplementary study to further highlight the utility of the model.

Motivating the Extreme Value Inflated Model

Extreme values tend to arise in a multitude of different disciplines where the phenomena studied have a self-reinforcing component. However, extreme values need not only appear from processes that often or always follow an extreme value distribution. Rather, extreme values can also appear in phenomena, which in the majority of cases produce values from nonextreme distributions. Examples of such phenomena include different types of organized political violence (for instance, Lacina and Gleditsch 2005; Eck and Hultman 2007; Hultman, Kathman, and Shannon 2013), crime rates (Disha 2019), and mass protests (Weidmann and Rød 2019). Extreme-valued distributions are also present outside the social sciences in widely different fields, such as the sizes of solar flares (Litvinenko 1996), the use of word distribution in languages (Ferrer i Cancho and Solé 2003), earthquake sizes in California (Gutenberg and Richter 1944), rainfall distributions (Myhre et al. 2019), etc.

However, the fact that extreme values exist but are rare causes a number of different problems for researchers who are aiming to model the phenomena. On the theoretical level, it may be problematic to include the most extreme observations of the phenomenon of interest in the analysis, as it may well be argued that the observation arises from a different process. For instance, when studying OSV against civilians, a researcher could argue that cases where an active genocide is ongoing should be excluded, as genocide arises from a different process than other forms of OSV against civilians. On the other hand, excluding the most prominent cases of OSV against civilians may also seem like a strange choice for this researcher. Worse yet, since the extreme values by their nature are very large compared to the vast majority of cases, the extreme values tend to have a large impact on the results of any type of modeling. This means that the decision to include or exclude certain cases may severely affect the results of the modeling. To not be forced to make an arbitrary inclusion or exclusion decision for a single or handful of case(s), the researcher may decide to use an "objective" solution, such as "trimming" (excluding) or "winsorizing" (censoring observations to a threshold value), for a certain number or percentage of cases (Dixon and Yuen 1974). Yet, these techniques are in most cases neither statistically nor theoretically sound, as they simply mask the arbitrary nature of the inclusion or exclusion criteria and may severely bias the results by either removing important observations or artificially changing values on the dependent variable.

To show the effects of inclusion, exclusion, or censoring of extreme values, we created a simple example regression model, where we modeled the country-month counts of OSV against civilians in Africa between 1989 and 2019 (Pettersson and Öberg 2020) against population and two dummy variables indicating democracy and autocracy, with hybrid-regimes as the residual category (Hegre et al. 2019). We modeled this using both a regular NB and a ZINB regression model. We focus on the count model and therefore only present this part of the ZINB model. In the *original* models, we used all available country-month observations; in the *trimmed* models, we removed the ten largest counts; and in the *winsorized* models, we censored the ten largest observations to the eleventh largest value in the data set. The results of these regressions are found in table 1.

The results in the table show that by trimming or winsorizing the ten largest counts, i.e., the 0.05 percent most extreme values, in our data set, the coefficient for autocracy changes from being negative and statistically significant to being positive and statistically significant. Similarly, the coefficient for democracy, while staying in the same direction and level of significance, is more than halved for both the NB and ZINB specifications. This example may be simplistic in terms of the covariates included in the model, but it highlights the effects of the researchers' choice of including or excluding certain observations in the analysis.

The problems with extreme values are not limited to their effect on the observed results; they may also affect the estimation method and the possibility to make diagnostic tests or alternative specifications of the models. For instance, when rerunning the NB regression model from table 1 using bootstrapping, the algorithm failed to run in approximately 28 percent of the bootstraps. This shows that failing to properly deal with the extreme values may not only cause biased results, but it may also make certain tools of analysis unavailable to the researcher.

The EVZINB Model

To alleviate the problems associated with modeling processes that sometimes but not always exhibit extreme values, we propose the EVZINB regression model.[1] The EVZINB model extends the NB and zero-inflated generalized linear models, developed for count data in the 1980s and 1990s (Hilbe 2011), with elements from extreme value modeling developed in the late 2000s (Clauset, Rohilla Shalizi, and Newman 2009). NB regression models, zero-inflated or not, have been used for a wide variety of applications since the early 1990s, including analyzing power outages following hurricanes (Liu et al. 2005), safety measures for highways (Hadi et al. 1995), and analyzing societal determinants of

[1] This model can also be used when zero-inflation is not present, in which case it would reduce to an extreme value-inflated negative binomial regression model, EVINB.

Table 1. Regression results for OSV against civilians counts with different methods of handling extreme values. Only coefficients from the count model are shown for the ZINB.

	NB original	NB trimmed	NB winsorized	ZINB original	ZINB trimmed	ZINB winsorized
Constant	4.018***	1.558***	1.882***	4.311***	2.919***	3.172***
	(0.075)	(0.065)	(0.066)	(0.064)	(0.069)	(0.069)
log(pop)	0.020***	0.025***	0.023***	0.011***	0.009***	0.008***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
autocracy	−1.430***	0.522***	0.336**	−1.101***	0.362***	0.245*
	(0.133)	(0.116)	(0.118)	(0.139)	(0.099)	(0.102)
democracy	−3.818***	−1.484***	−1.777***	−3.363***	−0.792***	−1.046***
	(0.102)	(0.089)	(0.090)	(0.120)	(0.103)	(0.108)
Observations	19,499	19,489	19,499	19,499	19,489	19,499
NB-α	40.86	30.83	32.10	30.25	8.33	9.11
AIC	42,727.1	41,113.2	41,546.7	41,527.9	39,827.4	40,286.1

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

displacement following floods (Vestby et al. 2024). Extreme value modeling, on the other hand, has been used within political and conflict science to analyze and discuss grand questions such as the overall decline in large deadly wars, the distribution of mass atrocities, or terrorist attacks (see, for instance, Cirillo and Taleb 2016; Clauset 2017; Cunen, Lid Hjort, and Mokleiv Nygård 2020).

The EVZINB model builds on the ZINB model, which assumes that the data originates from two separate subprocesses: one generating zeroes and one generating counts (which may also produce zeroes). In ZINB models, these two subprocesses are modeled separately and can be thought of as two different components of the models, which may include different covariates. The first component aims to model structural excess zeroes using one set of covariates, while the second component models the count process separately from these excess zeroes (Hilbe 2011). Our proposal is to extend this framework to a three-component regression model, where both excess zeroes and extreme values are modeled separately. This three-component regression model can be seen as a regression model with latent states, where the latent states represent the subprocesses from which the data are generated.

The benefits of this approach are manifold. First, by allowing for three separate states in the model, each of the states can be estimated while filtering out the effects of the other two. This will lead to more stable and less biased parameter estimates for each of the components and to fewer convergence issues in large data. Second, it allows the researcher to specify different covariates on each of the components, allowing for a more nuanced analysis. Third, filtering out the effects of extreme observations should give more stable out-of-sample predicted values from the model, enhancing the predictive performance of the model. Fourth, the regression model allows for the estimation of the probability that any given observation is part of one of the three states, widening the possibilities for further analysis. The ability to model and draw inferences from all three components of the model also differentiates the EVZINB model from earlier attempts to model extreme values, since in these cases the researchers are primarily interested in drawing inferences about the tail risks and the shapes of the extreme value distributions.

The theoretical motivation for the model is perhaps easiest to see with regards to political violence; we also argue that this type of data, which originate from a data-generating process with different latent states, should be relatively common. In fact, there are many processes that could become self-reinforcing after a certain threshold, from the number of people participating in demonstrations to the number of people infected by a certain disease and the number of consumers using a specific product.[2]

Statistical Methodology

In the presence of count data, a common approach is to employ the Poisson regression models. It has, however, been recognized that count data commonly display overdispersion, i.e., the variance is larger than the mean. One possibility to account for overdispersion is the NB regression, which relaxes the mean-variance equality restriction. It is, however, also common with data exhibiting a proportion of excess zeroes, which cannot be appropriately modeled using the NB regression model. The ZINB regression model accounts for excess zeroes by introducing a proportion of zeroes, which can be modeled using logistic regression (Hilbe 2011). The latent states associated with only zeros and moderate count data will be denoted Z and NB, respectively.

We propose to introduce another latent state, generating extreme values, alongside the extra latent states generating zeroes (Z) and moderate count data (NB). This state will be referred to as the EV state. Define the unobserved random variable $\boldsymbol{W}$, which is $\boldsymbol{W} = (1, 0, 0)^T$, $\boldsymbol{W} = (0, 1, 0)^T$, or $\boldsymbol{W} = (0, 0, 1)^T$ if the latent state is Z, NB, or EV, respectively. Then, we introduce the observed random variable Y with the following conditional properties:

$$\begin{aligned} Y|\boldsymbol{W} &= (1, 0, 0)^T \quad &&= 0 \\ Y|\boldsymbol{W} &= (0, 1, 0)^T &&\sim \text{NegBin}\left(\mu_{NB}, \alpha_{NB}\right) \\ Y|\boldsymbol{W} &= (0, 0, 1)^T &&\sim \text{Pareto}\left(c_{EV}, \alpha_{EV}\right), \end{aligned}$$

where μ_{NB} and α_{NB} are the mean and dispersion parameters of an NB distribution, respectively, α_{EV} and c_{EV} are the shape and cut-off parameters of a Pareto distribution and the prior probabilities $\pi_Z = \Pr\left(W = (1, 0, 0)^T\right)$, $\pi_{EV} = \Pr\left(W = (0, 0, 1)^T\right)$ and $\pi_{NB} = 1 - \pi_Z - \pi_{EV}$. The exact expressions for the probability mass functions of the NB and

[2]That the process contains a self-reinforcing component is, however, not a requirement for a process to produce extreme values nor to model the process using the EVZINB model. Rainfall distributions and earthquake magnitudes are examples of processes where there is no (explicit) self-reinforcement.

Pareto distributions are given in the online appendix. The distribution of the random variable Y defining the data-generating process in this study can be summarized as

$$Y \sim \text{EVZINB}\left(\pi_Z, \pi_{EV}, \mu_{NB}, \alpha_{NB}, \alpha_{EV}, c_{EV}\right). \tag{1}$$

The prior probabilities (π_Z and π_{EV}) of the latent states are modeled using multinomial logistic regression, and the extreme-value state is assumed to follow a Pareto distribution with shape parameter α_{EV} modeled as a function of the covariates. α_{EV} is a real, positive number governing the tail of the Pareto distribution. Whereas the Pareto distribution is proportional to $y^{-\alpha_{EV}}$, the exponential and Gaussian distributions are proportional to e^{-ky} and e^{-cy^2}, respectively, for positive numbers c and k. Hence, the Pareto distribution is associated with substantially more probability in its tail, and is therefore suitable to model extremely large observations. In fact, when α_{EV} approaches 1 from above, the expected value approaches infinity. Another property is that the conditional probability of yielding an observation twice the magnitude of some number $c_1 > c_{EV}$ is the same for any c_1. In this sense, the Pareto distribution is scale invariant (Mandelbrot 1983) and, thus, suitable for modeling self-enhancing phenomena like those discussed in the introduction. As described below, each observation will be associated with a shape-parameter α_{EV}, which indicates the potential extremeness of a distribution with corresponding covariates. The full model will be referred to as an EVZINB regression model. The model-implied latent-state probabilities, or the prior probabilities $\pi_{Z,i}$, $\pi_{NB,i}$, and $\pi_{EV,i}$ of observation i, are modeled as

$$\pi_{Z,i} = \frac{\exp\left\{\boldsymbol{\gamma}_Z^T \boldsymbol{x}_{\pi,i}\right\}}{1 + \exp\left\{\boldsymbol{\gamma}_Z^T \boldsymbol{x}_{\pi,i}\right\} + \exp\left\{\boldsymbol{\gamma}_{EV}^T \boldsymbol{x}_{\pi,i}\right\}} \tag{2}$$

$$\pi_{EV,i} = \frac{\exp\left\{\boldsymbol{\gamma}_{EV}^T \boldsymbol{x}_{\pi,i}\right\}}{1 + \exp\left\{\boldsymbol{\gamma}_Z^T \boldsymbol{x}_{\pi,i}\right\} + \exp\left\{\boldsymbol{\gamma}_{EV}^T \boldsymbol{x}_{\pi,i}\right\}},$$

where $\pi_{NB,i} = 1 - \pi_{Z,i} - \pi_{EV,i}$ and $\boldsymbol{x}_{\pi,i}$ is a column vector of covariates (starting with 1 for intercept),[3] and $\boldsymbol{\gamma}_Z$ and $\boldsymbol{\gamma}_{EV}$ are zero-inflation and extreme value-inflation parameter vectors, respectively. The conditional mean $\mu_{NB,i}$ of the NB latent state of observation i, and the observation-specific shape $\alpha_{EV,i}$ of the EV latent state are modeled as

$$\mu_{NB,i} = \exp\left\{\boldsymbol{\beta}_{NB}^T \boldsymbol{x}_{NB,i}\right\} \tag{3}$$

$$\alpha_{EV,i} = \exp\left\{\boldsymbol{\beta}_{EV}^T \boldsymbol{x}_{EV,i}\right\},$$

where $\boldsymbol{x}_{NB,i}$ and $\boldsymbol{x}_{EV,i}$ are covariates (starting with 1 for intercept) and $\boldsymbol{\beta}_{NB}$ and $\boldsymbol{\beta}_{EV}$ are column vectors of parameters. Additional model parameters are the dispersion parameter of the NB latent state α_{NB} and c_{EV} which is the lower bound of observations from the Pareto distribution (EV latent state).[4] The model parameters, hence, include $\boldsymbol{\gamma}_Z$, $\boldsymbol{\gamma}_{EV}$, $\boldsymbol{\beta}_{NB}$, $\boldsymbol{\beta}_{EV}$, α_{NB}, and c_{EV}. Those are estimated with maximum likelihood using a version of the EM algorithm of Dempster, Laird, and Rubin (1977), combining a generalized expectation maximization (GEM) algorithm, e.g., Wu (1983), Lange (1995), and an expectation conditional maximization either (ECME) algorithm of Liu and Rubin (1994). The model estimation provides the model-implied (ex-ante) latent state probabilities $\pi_{Z,i}$, $\pi_{NB,i}$, and $\pi_{EV,i}$ and the ex-post latent state probabilities after observing the dependent variable (commonly referred to as the responsibilities) of all observations. If the EV state is not included, the model is equivalent of the ZINB model. If the Z state is not included, the model reduces to an extreme value-inflated negative binomial model (EVINB).

For prediction, it is common to use the expected value of the dependent variable, given the covariates and the estimated parameters. However, the EV state lacks expected value if the shape parameter $\alpha_{EV,i} \leq 1$. Instead, we use the harmonic mean for the EV state. The harmonic mean of a random variable X is defined as $E\left[X^{-1}\right]^{-1}$ and exists for any $\alpha_{EV,i} > 0$ for the Pareto distribution. When using the EVZINB model for prediction, it should be noted that the harmonic mean provides a more conservative estimate than $E[X]$. For more details on the harmonic mean and prediction, see online appendix A1 and A2.

In order to investigate whether a covariate has a significant contribution in any part of the model, a likelihood ratio (LR) test is developed. Define the vector of parameters to be $\boldsymbol{\theta}$ and the log likelihood to be $\ell(\boldsymbol{\theta})$ with the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ which maximizes $\ell(\boldsymbol{\theta})$. Define the parameter vector $\tilde{\boldsymbol{\theta}}$, which includes the restricted elements $\gamma_{Z,p} = \gamma_{EV,p} = \mu_{NB,p} = \alpha_{EV,p} = 0$ where, for example, $\gamma_{Z,p}$ represents the element of $\boldsymbol{\gamma}_Z$ corresponding to the p^{th} covariate in $\boldsymbol{\gamma}_Z$. Then the test statistic

$$\chi^2 = -2\left(\ell\left(\tilde{\boldsymbol{\theta}}\right) - \ell\left(\hat{\boldsymbol{\theta}}\right)\right)$$

asymptotically follows a χ^2 distribution with the degrees of freedom equal to the number of restrictions, i.e., the number of components in which the covariate is present.

Simulation Study

To investigate the performance of the proposed method, a simulation study is conducted. Data of three covariates were generated, and the parameters $\boldsymbol{\gamma}_Z$, $\boldsymbol{\gamma}_{EV}$, $\boldsymbol{\beta}_{NB}$, α_{NB}, $\boldsymbol{\beta}_{EV}$, and c_{EV} specified. A total of 1,000 replications of sample size $n = 1,000$ are generated. The results of the simulation study show that the EVZINB model provides unbiased parameter results under the proposed process, while both the NB and ZINB models are substantially biased, even when the proportion of extreme values is as low as 0.7 percent. Full details on the simulation design as well as the full simulation results are provided in online appendix A2.

Empirical Study

To illustrate the empirical usefulness of the EVZINB regression model, we replicate Hultman, Kathman, and Shannon's (2013) study on the effect of UN peacekeeping on *one-sided violence against civilians*. The dependent variable of the study is the count of OSV against civilians per conflict month in all conflicts coded by the Uppsala Conflict Data Program (UCDP) between 1991 and 2008 (Eck and Hultman 2007; Sundberg, Eck, and Kreutz 2012). Hultman et al. use NB regression models and a number of different model specifications to gauge the effects of UN Peacekeeping troops, police, and military observers on the amount of OSV against civilians conducted by rebel groups, governments, or both, and their main findings are that UN peacekeeping troops and police have a negative effect on OSV against civilians, i.e., that the more peacekeeping troops and police are present in a conflict zone, the less violence is perpetrated against civilians. We have chosen to replicate this study as it is one of relatively few high-impact studies on OSV against civilians, which is modeled using the raw count of fatalities from OSV as the dependent variable and which exhibits

[3] It is possible to use different covariates for the Z and EV latent states, but the same are used in the presentation for notational convenience.

[4] See online appendix A1 for details on the Pareto distribution.

both a high degree of zero-inflation as well as a number of large outliers of extreme values.[5]

To limit the number of comparisons, we only replicate Model 1, focusing on both government and rebel OSV against civilians, from Hultman et al. This model contains eight covariates, with the three theoretically relevant being the number of UN military troops, police, and observers, measured in 1,000's of deployed peacekeepers. In addition, five control variables are included: conflict duration in months, two dummy variables measuring whether the conflict is over government and whether any OSV against civilians was used in the previous month, the natural logarithm of the population of the country the conflict is taking place in, and the number of battle-related deaths in the conflict in the previous month.[6]

The data on the dependent variable is highly zero-inflated, with approximately 76 percent of observations being zeroes. In addition, the data contains a number of extreme values and is severely overdispersed with a rate of unconditional overdispersion (variance/mean) of approximately 83,000.[7] To test the utility of our proposed model, we estimate the model using both the original NB regression model as well as the ZINB and the EVZINB models. In online appendix B4, we also report the results from trimmed and winsorized versions of the NB and ZINB models for comparison.

For both the ZINB and EVZINB models, we need to specify covariates for the additional components in the model. In the ZINB model, we need to specify which covariates govern the zero-inflation, i.e., the covariates that separate the zero process from the count process. In the EVZINB model, we also need to specify the covariates that govern extreme value-inflation, i.e., the covariates that separate the extreme values from the zero- and count processes, and the covariates that govern the extreme value (Pareto) component, i.e., the covariates that govern "how extreme" the extreme values get. For the sake of simplicity, we have chosen to keep all covariates from the original model for the estimation of both the zero-inflation and extreme value-inflation. In the extreme value component, we have chosen to only retain UN military troops (lag), conflict duration, all battle deaths (lag), and population covariates, as we believe that these covariates are likely to have an effect on the more extreme values. In this component we have also chosen to log-transform the UN military troops (lag) and conflict duration covariates.[8]

We estimate the three models using bootstrapping with 1,000 bootstrapped samples, and we report the coefficients as the median across all bootstrapped samples, and the standard error as the standard deviation across the bootstrapped samples. Bootstrapping is needed for the estimation of standard errors in the EVZINB model, and we decided to use bootstrapped estimates for all models as this allows for a comparison of the models on equal terms. Since the data are panel data, we use a cluster bootstrap on the conflict level, where all observations of each conflict are either included or excluded. Cluster bootstrapping is robust to the complex autocorrelation and heteroskedasticity, which may be assumed to be present in the data (Cameron, Gelbach, and Miller 2008).[9] Lastly, bootstrapping allows us to assess the statistical significance through bootstrapped *p*-values without relying on any distributional assumptions, which may not be fulfilled when the data is sparse (e.g., in the extreme value component) or when the distribution of the residuals from the regression is highly skewed (a likely consequence of highly overdispersed data). Of the 1,000 bootstrapped samples, the NB was inestimable for 215, highlighting one of the issues with the NB model for this type of data. None of the bootstrapped samples were inestimable for the ZINB or EVZINB models.

Results

The results for the three regression models can be found in table 2. Focusing on the effects of the UN military troop presence, we can see that all three models agree that the effect of UN military troops is negative in the NB component, indicating that an increased presence of UN military troops leads to fewer civilian killings. In the ZINB model, this effect is not statistically significant. However, there is a statistically significant positive effect on the zero-inflation for the variable, indicating that an increased presence of UN military troops increases the likelihood of observing a zero count of civilian killings. This can be interpreted as the UN military troops increasing the likelihood that *no civilian killing occur*, but given that we are in the count process, the negative effect is no longer statistically significant. The EVZINB model agrees with both the NB and ZINB models that an increased presence of UN military troops is associated with a lower count of civilian killings in the count process and an increased likelihood of observing no civilian killings. The EVZINB model also offers insight into how UN military troops affect the likelihood of the process entering into the extreme value component and how extreme values will be if the process enters into this. While these results are not statistically significant, the point estimates indicate that an increase in UN military troops slightly decreases the likelihood of the process entering into the extreme value component, while the negative coefficient for the Pareto component indicates that if the process enters into this component, the values will tend to be *more* extreme.

Both the ZINB and EVZINB models offer more nuanced pictures of how the presence of UN military troops affect OSV but are also more difficult to interpret as the effects of a variable may differ across the different components of the model. This is especially true with regards to statistical significance since the statistical significance is only measured in terms of the effect in the individual components rather than the total effect across all components. To further investigate how the presence of UN military troops affect OSV, we in figure 1 reproduce figure 4 in Hultman, Kathman, and Shannon (2013), which shows the predicted number of civilian killings for different levels of UN military troop

[5]We also believe that one of the reasons that there are few published articles on this type of data are the problems of nonconvergence and/or model misspecification that the EVZINB model is aiming to solve, i.e., that many of the potential research questions that could be asked of this type of data are not possible to answer without solving the problem of extreme value-inflation.

[6]In Hultman et al., this variable is the count of all battle-related deaths in the conflict in the previous year. We have log-transformed this variable because this allows for more stable estimation across all models. Log-transforming this control variable does not affect the conclusions with regards to the theoretically relevant variables.

[7]More details on the distribution of the dependent variable is given in online appendix B1.

[8]In general, we recommend researchers to be parsimonious when choosing covariates as only data points above C_{EV} enter into the estimation for the extreme value (Pareto) component, and we also generally recommend either log-transforming the covariates or using covariates that operate on a relatively narrow range, such as dummy variables. These issues are further elaborated on in online appendix A1. As the covariates need to vary above C_{EV}, we also chose to exclude any OSV dummy variable for the Pareto component since near all observations above C_{EV} are positive for this variable.

[9]Hultman, Kathman, and Shannon (2013) use clustered standard errors to correct for this.

Table 2. Regression results across all components for the three different regression models.

	Negative binomial	Zero-inflated negative binomial		Extreme value and zero-inflated negative binomial			
	NB β	NB β	ZI γ	NB β	ZI γ	EVI γ	Pareto β
UN military troops$_{t-1}$	−0.522**	−0.299	0.140*	−0.127	0.157*	0.060	−1.003
	(0.192)	(0.296)	(0.265)	(0.213)	(0.172)	(0.304)	(8.881)
UN police$_{t-1}$	−9.770*	−7.265+	3.599	−3.020	−0.421	−6.080	
	(8.925)	(34.389)	(22.393)	(21.151)	(9.506)	(9.655)	
UN observers$_{t-1}$	21.311***	12.638	−4.426+	6.154	−3.657+	−4.114	
	(9.261)	(7.569)	(5.910)	(5.256)	(4.506)	(9.457)	
Conflict duration	−0.002	−0.010	−0.008**	−0.002	−0.006*	−0.005	0.295
	(0.005)	(0.006)	(0.713)	(0.003)	(0.002)	(0.004)	(0.266)
log(population)	0.741*	0.605	−0.568**	0.033	−0.473**	0.405	−0.178
	(0.284)	(0.381)	(1.263)	(0.251)	(0.155)	(1.094)	(0.491)
log(all battle deaths)$_{t-1}$	0.176*	0.064	−0.416**	0.057	−0.336***	0.003	−0.050
	(0.098)	(0.058)	(2.170)	(0.063)	(0.080)	(0.104)	(0.084)
Any OSV dummy$_{t-1}$	2.044***	−0.094	−13.030***	0.298	−2.993***	0.153	
	(0.321)	(0.383)	(12.937)	(0.242)	(0.809)	(0.638)	
Government conflict	2.255***	1.914+	−1.483**	0.783	−0.780**	2.186+	
	(0.637)	(0.743)	(1.897)	(0.484)	(0.435)	(1.847)	
Constant	−9.394*	−4.525	10.512***	1.073	8.623***	−9.485	0.652
	(3.320)	(4.331)	(2.701)	(3.192)	(2.057)	(14.510)	(4.904)
$C_E V$				96			
α_{nb}	16.75	5.00		1.96			
Observations	3,746	3,746		3,746			
Log likelihood	−6,254.19	−5,689.65		−5,574.38			
Akaike inf. crit.	12,528.4	11,417.3		11,216.8			

Note: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Coefficients reported as the median of the bootstrapped coefficients. Standard errors in parenthesis.

presence when the other continuous variables are set to their means and the dummy variables are set to 1. In addition to the predicted number of civilian killings, we also show the median 50th and 95th percentiles for each of the models, as well as component probabilities for the ZINB and EVZINB models. For the median 50th and 95th percentile plots, we interpret the lines as the values that the model implies that we would expect 50 percent and 5 percent of observations to exceed, respectively. Including these plots allows us to show how the covariate affects both the most common case (the 50th percentile) as well as the more extreme cases (the 95th percentile).

The results highlight a number of important and interesting differences between the three models. First, it is clear that the conclusion that an increased presence of UN military troops leads to a decrease in the number of civilian killings is less pronounced, and not statistically significant, in the EVZINB model. Looking at the panes for the 50th and 95th percentiles, it can be seen that UN military troops have divergent effects on the most common case, i.e., the 50th percentile, and the more extreme cases, i.e., the 95th percentile, as the median 50th percentile of civilian killings decreases as UN military troops increase, while the median 95th percentile of civilian killings increases as UN military troops increase from about 0 to about 5,000 after which they start decreasing. The decrease in predicted killings at the 50th percentile is statistically significant, while the increase and decrease in predicted killings at the 95th percentile are not. The curved relationship seen for the 95th percentile in the EVZINB model appears as an increase in UN military troops both decrease the likelihood of extreme values appearing *and* tend to make them more extreme if they happen. While these results are not statistically significant, and should not be interpreted causally, they highlight an important feature of the EVZINB model: that covariates may have divergent effects on the median and extreme cases.

Model Performance

While the section above highlights that the EVZINB model allows for a more nuanced analysis of effects than the NB and ZINB models, it is also important to test the performance of the models on a range of different metrics in order to determine the usefulness of the EVZINB model.

Model Fit

To compare the fit of the NB, ZINB, and EVZINB models, we use a bootstrapped comparison of the Aikaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). The results, which can be seen in the online appendix, figure B2, show that with AIC correction, the EVZINB model outperforms the NB model in 100 percent and the ZINB model in 99 percent of the bootstrapped samples. Using a BIC correction, which penalizes the additional parameters in the EVZINB model more severely, the EVZINB model still outperforms the NB model in 100 percent of the bootstrapped cases, and outperforms the ZINB model in 73 percent of of the bootstrapped samples. These results indicate that the EVZINB model fits the data substantially better than the competing models.

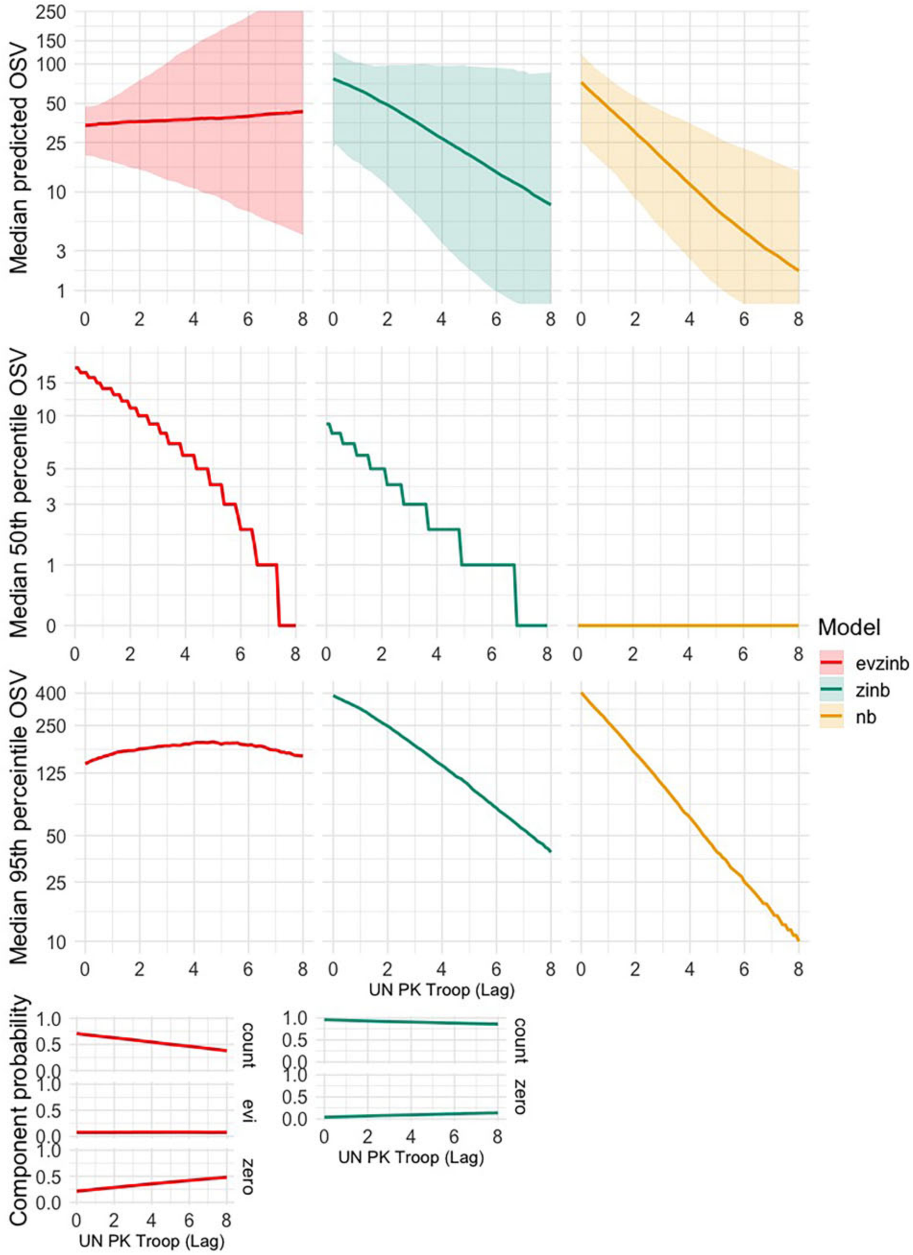

Figure 1. Median predicted fatalities, median predicted 50th percentile of fatalities, and median predicted 95th percentile of fatalities for different values of *UN Military Troops (lag)* across all bootstrapped samples with the remaining continuous covariates set to their means, and the any OSV against civilians$_{(t-1)}$ and Government Conflict dummies set to 1. Highlighted areas indicate 90 percent bootstrapped intervals.

Predictive Performance

Apart from model fit, we also test the predictive performance of the three models by making predictions on the observations left out in each bootstrapped sample, the so-called out-of-bag (OOB) observations. We then calculate the *root mean squared error* (RMSE) and the *root mean squared log error* (RMSLE) across all bootstraps and compare the mean and medians across these metrics. The results of the OOB predictive performance can be seen in table 3. This table shows that the EVZINB model outperforms both the NB and ZINB models on OOB predictive performance. A pairwise comparison of the RMSE and RMSLE in each of the bootstrapped samples also showed that the EVZINB model outperforms both the NB and ZINB models in over 92 percent and 96 percent of bootstrapped samples on the RMSLE and in 82 percent and 80 percent of bootstrapped samples on the RMSE.

Table 3. Out-of-bag RMSE and RMSLE for all models across all bootstraps

	RMSE	RMSLE	
Model	Median	Mean	Median
EVZINB	536	1.89	1.77
ZINB	3,912	2.42	2.13
NB	4,001	2.11	2.05

Discussion

The replication of Hultman, Kathman, and Shannon (2013), has shown that the EVZINB model can be empirically and theoretically useful. By employing the EVZINB model instead of the more conventional NB or ZINB models, we were able to disentangle what the effect is of UN military troops on civilian killings not only for the predicted values of civilian killings but also how the effects of these covariates differ between the effect on the most common cases and on the more extreme cases.

We were able to show that the main finding of Hultman, Kathman, and Shannon (2013), that an increased presence of UN military troops leads to fewer civilian casualties, indeed holds in the most common cases, but that this conclusion does not hold for the overall total effect of UN military troops. Instead, we were able to show that while UN military troops still lower the expected civilian killings in the median case, this is not true for the more extreme cases of civilian killings. These findings do not mean that we argue that UN military troops are ineffective in preventing violence against civilians. Rather, this conclusion helps us better understand under what conditions, when, and where UN peacekeeping may be most effective in preventing civilian killings. The EVZINB model could also identify which cases are at risk of entering into the extreme value state by producing predicted probabilities for each state. Such findings would be important for the ongoing academic debate on the effects of peacekeeping on different types of violence and under different circumstances (see, for instance, Gromes 2019; Bara 2020; Bara and Hultman 2020).

The additional inferences that can be drawn from the EVZINB model compared to a NB or ZINB model also allow us to more precisely analyze a phenomenon and distinguish the effects on the more common and more extreme cases. Additionally, researchers can test more fine-grained hypotheses not only about which covariates affect which outcomes but also test hypotheses relating to the extreme and less extreme cases separately. There may, for instance, be covariates that both decrease the likelihood of observing a zero, but also decrease the likelihood of entering the extreme value domain. With conventional models such as the NB and ZINB models, hypotheses relating to these covariates would be difficult to test as the overall effects may be zero, even if the effects are pronounced in both directions.

In addition to more fine-grained analysis, the EVZINB model also opens up for new avenues of research. In particular, the convergence and misspecification issues that arise when trying to model data containing extreme values using the NB and ZINB models may have prevented researchers from approaching certain research questions as the modeling problems have been too severe, thus causing a file-drawer problem. We believe that this is one of the reasons for the low number of high-impact research articles published on such data in the last decade. Using the EVZINB model, researchers using such data may be encouraged to restart research on these types of data. The EVZINB model also allows for new types of research questions as it is possible to focus the analysis on, for instance, the covariates for the multinomial process, which differentiates between the zero, count, and extreme value latent states, allowing researchers to better understand which conditions need to be present for a process to start producing extreme values. Focusing on the transitions between states may be especially useful when communicating results with policymakers as the (prior) state probabilities may be used to evaluate the risk of a low-intensity armed conflict escalating into a high-intensity armed conflict, while the ex-post (posterior) state probabilities of the same observation may be used to assess the likelihood that the armed conflict, given a certain number of fatalities, has entered into the specific state.

A straightforward extension of the EVZINB model would be to also allow regression parameters to be estimated for the Pareto threshold, $\hat{c}_{EV}$, i.e., the point after which observations may enter into the latent Pareto process. While perhaps not immediately evident, estimating regression parameters specifically for $\hat{c}_{EV}$ may create new avenues for research, and open up new types of research questions as $\hat{c}_{EV}$ is a measure for *when* a process have a chance of progressing from a well-behaved count process to an extreme value process. This means that by investigating covariates effect on $\hat{c}_{EV}$, we can ask questions such as: *What factors affect the threshold for when low-intensity armed conflicts may progress into large-scale armed conflicts?* The EVZINB model could also be constructed as a hidden Markov process where the latent states correspond to hidden Markov states.

Conclusion

In this paper, we have introduced the extreme value and zero-inflated regression model for count data, which contains both an inflated number of zeroes and extreme values. The extreme value and zero-inflated regression model can be thought of as a latent states regression model, where we can estimate both which covariates affect the likelihood of different states of the process and how these covariates affect the behavior of the process given its state.

We have shown that this model is both empirically and theoretically motivated and that the model can retrieve correct parameter estimates from simulated data. We have also shown the empirical usefulness of the model through replication of Hultman, Kathman, and Shannon's (2013) paper on the effect of UN peacekeeping troops on civilian killings. In the replication study, the extreme value and zero-inflated regression model allowed us to draw additional inferences about when UN military troops decrease civilian killings. Additionally, the EVZINB model outperformed both the NB and ZINB models with regards to efficiency of the estimated parameters, the AIC and BIC corrected LRs, and the predictive performance of the model.

In the discussion, we also presented a number of different empirical lenses through which the extreme value and zero-inflated regression model can be viewed, and how this model can allow researchers to ask novel questions about the nature of their data and to ask questions previously not possible to answer. The extreme value and zero-inflated regression model can also easily be extended to a non-zero-inflated version for count data which do not suffer from zero-inflation but still contains extreme values. With future development, a unified framework for analysis of the effect of covariates across states of the model could be developed, allowing for a more specific analysis of the marginal effects of certain covariates in different conditions.

The EVZINB model and tools related to analyzing this model are available in the `R` package `evinf` on CRAN.

Supplementary Information

Supplementary information is available in the *International Studies Quarterly* data archive.

Funder Information

The research was funded by the European Research Council, project H2020-ERC-2015-AdG 694640 (ViEWS), Riksbankens Jubileumsfond, grant M21-0002 (Societies at Risk), and Swedish Research Council, grant number 2017-01175.

References

Bagozzi, Benjamin E. 2015. "Forecasting Civil Conflict with Zero-Inflated Count Models." *Civil Wars* 17 (1): 1–24.

Bara, Corinne. 2020. "Shifting Targets: The Effect of Peacekeeping on Postwar Violence." *European Journal of International Relations* 26 (4): 979–1003.

Bara, Corinne, and Lisa Hultman. 2020. "Just Different Hats? Comparing UN and Non-UN Peacekeeping." *International Peacekeeping* 27 (3): 341–68.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics* 90 (3): 414–27.

Cirillo, Pasquale, and Nassim Nicholas Taleb. 2016. "On the Statistical Properties and Tail Risk of Violent Conflicts." *Physica A: Statistical Mechanics and its Applications* 452: 29–45.

Clauset, Aaron. 2017. "The Enduring Threat of a Large Interstate War." *Technical report*, One Earth Future Foundation.

Clauset, Aaron, Cosma Rohilla Shalizi, and M.E.J. Newman. 2009. "Power-Law Distributions in Empirical Data." *SIAM Review* 51 (4): 661–703.

Cunen, Céline, Nils Lid Hjort, and Håvard Mokleiv Nygård. 2020. "Statistical Sightings of Better Angels: Analysing the Distribution of Battle-Deaths in Interstate Conflict over Time." *Journal of Peace Research* 57 (2): 221–34.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)* 39: 1–38.

Disha, Ilir. 2019. "Different Paths: The Role of Immigrant Assimilation on Neighborhood Crime." *Social Science Quarterly* 100 (4): 1129–53.

Dixon, W.J., and K.K. Yuen. 1974. "Trimming and Winsorization: A Review." *Statistische Hefte* 15 (2): 157–70.

Eck, K., and L. Hultman. 2007. "One-Sided Violence against Civilians in War: Insights from New Fatality Data." *Journal of Peace Research* 44 (2): 233–46.

Ferrer i Cancho, Ramon, and Ricard V. Solé. 2003. "Least Effort and the Origins of Scaling in Human Language." *Proceedings of the National Academy of Sciences* 100 (3): 788–91.

Gromes, Thorsten. 2019. "Does Peacekeeping Only Work in Easy Environments? An Analysis of Conflict Characteristics, Mission Profiles, and Civil War Recurrence." *Contemporary Security Policy* 40 (4): 459–80.

Gutenberg, Beno, and Charles F. Richter. 1944. "Frequency of Earthquakes in California." *Bulletin of the Seismological Society of America* 34 (4): 185–88.

Hadi, Mohammed A., Jacob Aruldhas, Lee-Fang Chow, and Joseph A. Wattleworth. 1995. "Estimating Safety Effects of Cross-Section Design for Various Highway Types Using Negative Binomial Regression." *Transportation Research Record* 1500: 169.

Hegre, Håvard, Marie Allansson, Matthias Basedau, Michael Colaresi, Mihai Croicu, Hanne Fjelde, and Frederick Hoyles et al. 2019. "Views: A Political Violence Early-Warning System." *Journal of Peace Research* 56 (2): 155–74.

Hilbe, Joseph M. 2011. *Negative Binomial Regression*. Cambridge: Cambridge University Press.

Hultman, Lisa, Jacob Kathman, and Megan Shannon. 2013. "United Nations Peacekeeping and Civilian Protection in Civil War." *American Journal of Political Science* 57 (4): 875–91.

Lacina, Bethany, and Nils Petter Gleditsch. 2005. "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths." *European Journal of Population/Revue européenne de Démographie* 21 (2): 145–66.

Lange, Kenneth. 1995. "A Gradient Algorithm Locally Equivalent to the EM Algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (2): 425–37.

Litvinenko, Yuri E. 1996. "A New Model for the Distribution of Flare Energies." *Solar Physics* 167 (1–2): 321–31.

Liu, Chuanhai, and B. Rubin Donald. 1994. "The ECME Algorithm: A Simple Extension of EM and ECM with Faster Monotone Convergence." *Biometrika* 81 (4): 633–48.

Liu, Haibin, Rachel A. Davidson, David V. Rosowsky, and Jery R. Stedinger. 2005. "Negative Binomial Regression of Electric Power Outages in Hurricanes." *Journal of Infrastructure Systems* 11 (4): 258–67.

Mandelbrot, Benoit B. 1983. *The Fractal Geometry of Nature*, vol. 173. New York, NY: W. H. Freeman.

Myhre, Gunnar, Kari Alterskjær, Camilla Weum Stjern, Øivind Hodnebrog, Louis Marelle, Bjørn Hallvard Samset, and Jana Sillmann et al. 2019. "Frequency of Extreme Precipitation Increases Extensively with Event Rareness under Global Warming." *Scientific Reports* 9 (1): 16063.

Pettersson, Therese, and Magnus Öberg. 2020. "Organized Violence, 1989–2019." *Journal of Peace Research* 57 (4): 597–613.

Sundberg, Ralph, Kristine Eck, and Joakim Kreutz. 2012. "Introducing the UCDP Non-State Conflict Dataset." *Journal of Peace Research* 49 (2): 351–62.

Vestby, Jonas, Sebastian Schutte, Andreas Foro Tollefsen, and Halvard Buhaug. (2024). "Societal Determinants of Flood-Induced Displacement." *Proceedings of the National Academy of Sciences* 121 (3): e2206188120.

Weidmann, Nils B., and Espen Geelmuyden Rød. 2019. *The Internet and Political Protest in Autocracies*. Oxford: Oxford University Press.

Wu, C.F. Jeff. 1983. "On the Convergence Properties of the EM Algorithm." *The Annals of Statistics* 11 (1): 95–103.

Randahl, David, and Johan Vegelius. (2024) Inference with Extremes: Accounting for Extreme Values in Count Regression Models. *International Studies Quarterly*, https://doi.org/10.1093/isq/sqae137