# Replicating Nel and Righarts

Kevin Wagner*

December 15, 2024

**Abstract**

This replication paper examines the relationship between natural disasters and the risk of violent civil conflict, replicating and extending the analysis conducted by Philip Nel and Marjolein Righarts in their 2008 study, "Natural Disasters and the Risk of Violent Civil Conflict." Using cross-sectional time-series data from the EM-DAT International Disaster Database and the Uppsala/PRIO Armed Conflict Dataset, Nel and Righarts argue that natural disasters significantly increase the likelihood of civil conflict in low- and middle-income countries, particularly those with high inequality, mixed regimes, and weak economic growth. This replication reproduces their key findings using rare event logistic regression and assesses the robustness of their statistical models.

*PhD Student. Northwestern University. Email: kwagner@u.northwestern.edu

# 1 Introduction

Natural disasters—from earthquakes and hurricanes to floods and droughts—have long been associated with profound socio-political consequences. Researchers in environmental security, political ecology, and conflict studies have debated the degree to which environmental shocks translate into instability and violence. Such shocks may exacerbate underlying grievances, weaken already fragile state institutions, and create new opportunities for political actors to mobilize around resource scarcities or grievances. With the frequency and intensity of certain natural disasters projected to rise due to climate change, understanding the link between environmental shocks and civil conflict risk has become increasingly urgent.

In their 2008 study, *Natural Disasters and the Risk of Violent Civil Conflict*, Philip Nel and Marjolein Righarts examine the onset of violent civil conflict across 187 political units between 1950 and 2000. Using cross-sectional time-series data, they argue that natural disasters elevate the likelihood of civil conflict onset, especially in low- and middle-income countries characterized by high levels of inequality, mixed political regimes, and weak economic growth. Their work represents a significant contribution to the literature that investigates how environmental shocks interact with structural and political conditions to influence conflict dynamics.

This replication paper seeks to reproduce Nel and Righarts's key findings using their original data and methods, while also critically assessing their modeling choices and exploring alternative estimation strategies. Specifically, I aim to (1) replicate their main results using rare events logistic regression (relogit) as they did, ensuring that their approach and conclusions are transparent and reproducible; (2) verify their modeling choices and data preparation steps; and (3) examine the robustness of their results by introducing an alternative method—Firth's penalized logistic regression—that addresses small-sample bias without relying on the now less-supported relogit approach. By comparing results across estimation methods, this replication clarifies how methodological decisions influence inferences about the relationship between natural disasters and conflict.

## 2 Background

Nel and Righarts's methodological framework integrates environmental shock data (from EM-DAT) with political and economic indicators, leveraging cross-sectional time-series analysis to estimate the effects of disasters on civil conflict onset. The ambition of this design is noteworthy, but a careful review reveals several methodological challenges. These include the specification of the statistical models, the handling of observational data, the need for appropriate inference techniques, and the influence of data biases.

A core concern in studies like this is the risk of "kitchen-sink" modeling—introducing numerous controls without robust theoretical justification. Overfitting can obscure the underlying relationships that the theory attempts to capture. Achen's (2005) critique emphasizes that every variable added to a model without clear theoretical grounding increases the risk of misinterpretation. Similarly, Berk (2008) reminds us that if the model does not match the data-generating process, results may be biased or spurious. In the context of environmental shocks and conflict, where pathways are complex and multi-causal, parsimony and clear theoretical linkage between predictors and outcomes are crucial.

Data handling also raises substantial concerns. Nel and Righarts rely on observational data from EM-DAT and Uppsala/PRIO. Pepinsky (2018) warns that listwise deletion of missing data can bias results by disproportionately excluding data-poor contexts, often in the low-income regions most relevant to the theory. While multiple imputation can be helpful, Arel-Bundock and Pelc (2018) point out that such techniques rest on strong assumptions. Without careful validation, imputation can introduce new distortions.

Statistical inference complexities further challenge the analysis. Aronow and Miller (2019) advocate for agnostic approaches that limit strong parametric assumptions—particularly important in studies of conflict, where rare events and nonlinear relationships are common. Powell and Wooldridge highlight potential issues in calculating standard errors in dynamic models, and Mooney's endorsement of bootstrapping suggests that more robust approaches are often needed to confirm the stability of estimates.

Taken together, these methodological critiques underscore the importance of scrutinizing Nel and Righarts's choices, verifying their results, and considering alternative methods to handle the data's complexity and rarity of events.

### 2.0.1 Critique of Data Usage in Nel and Righarts' Analysis of Natural Disasters and Conflict

Nel and Righarts rely on the EM-DAT database for disaster information and the Uppsala/PRIO dataset for conflict data. While these datasets are standard sources in the literature, they contain biases and limitations that can shape conclusions about the disaster-conflict nexus.

**Biases in Disaster Data (EM-DAT):**

Harrington and Otto (2020) highlight that EM-DAT often underreports disasters in regions with weak observational networks, such as remote or low-income areas. This underreporting affects slow-onset disasters—like droughts or heatwaves—which are frequently less visible than large, rapid-onset events. Gall et al. (2009) argue that media and institutional biases lead to overrepresentation of certain disaster types (e.g., floods, hurricanes) while undercapturing others. Over time, changes in reporting standards and technological advancements in detection may introduce temporal inconsistencies.

These biases can skew inferences. If some disasters remain underreported, the observed correlation between disasters and conflict may reflect data artifacts. Moreover, the conceptual framework linking disasters to resource scarcity and social stress might be undercut if the most subtle but impactful events are missing or systematically misrepresented.

**Biases in Conflict Data (Uppsala/PRIO):**

Miller et al. (2022) note similar issues with conflict data. Highly visible conflicts—urban, heavily mediated, or internationally monitored—are more likely to appear in datasets, while rural, low-intensity, or sporadic conflicts are often overlooked. Such reporting patterns create a biased picture of where and when conflicts occur. In contexts relevant to disaster-triggered instability, this omission may be critical.

Together, these biases undermine the causal claims that Nel and Righarts advance. If disasters and conflicts are both mismeasured, patterns might emerge from systematic reporting distortions rather than real-world processes. Given these constraints, more robust data-handling techniques, sensitivity analyses, or triangulation with qualitative evidence could strengthen the study's claims.

# 3 Replication

To assess the accuracy and validity of the original results, I began by replicating Nel and Righarts's analysis as closely as possible. Since their original estimation employed the relogit function in Stata—no longer readily available in standard R packages—the first step was to reproduce their results in Stata. Achieving an identical replication in Stata ensures that the authors' use of relogit and their temporal data structuring are properly understood.

After confirming the original results in Stata, I migrated the analysis to R, using alternative rare events modeling techniques. This two-stage process allows for a rigorous test of the reliability and transferability of their findings.

**Data Preparation in Stata:**

The replication included using the provided dataset (from Nel and Righarts) and replicating the transformations required for time-series cross-sectional analysis. The code snippet below (adapted and condensed for illustration) shows how key variables were lagged and leaded:

```
* Stata variable setup

clear all

use "\\replication_paper\Nel and Righarts ND and VCC final 20 June 2007"

* Ensure the data is sorted by state and year for lagged variables
sort state year

* Generate lagged variables
gen imr_lag1 = imr[_n-1] if state == state[_n-1]
gen imr_sq_lag1 = imr_sq[_n-1] if state == state[_n-1]

* Generate t+1 dependent variable
gen vccall_tp1 = vccall[_n+1] if state == state[_n+1]
```

This setup ensures that each observation aligns with the temporal structure described by Nel and Righarts.

Rare Events Logistic Regression (relogit) in Stata:

The following Stata code runs rare events logistic regression models (relogit) to analyze the relationship between natural disasters and the likelihood of violent civil conflict. The first model (Model 1) includes predictors such as natural disasters per capita (allND_pc), lagged infant mortality rate (imr_lag1), its squared term (imr_sq_lag1), political regime type (mixedx), GDP growth (gdpgrox), and a measure of the brevity of prior conflicts (brevity). The second model (Model 2) features a similar structure but uses a slightly different specification, substituting alternative predictors for political regime type (mixed) and GDP growth (gdpgro). The results from both models are appended to a document titled regression_results.doc using the outreg2 command, allowing for organized presentation of regression outputs with titles identifying each model. This code is integral to testing and comparing hypotheses within the context of rare events logistic regression.

```
* Stata relogit models


*Model 1
relogit vccall allND_pc imr_lag1 imr_sq_lag1 mixedx gdpgrox brevity outreg2
using "regression_results.doc", append ctitle(Model 1)


* Model 2
relogit vccall allND_pc imr_lag1 imr_sq_lag1 mixed gdpgro brevity outreg2
using "regression_results.doc", append ctitle(Model 2)


* Same logic is used for remaining three models - code available in .do files
```

By comparing the replication outputs with the original tables in Nel and Righarts's article, I confirmed that most coefficients and standard errors matched closely. One discrepancy appeared in Model 4, where the originally reported coefficient of 0.089 seems to be a typographical error and should likely be 0.89. Furthermore, the authors did not specify that GDP per capita was log-transformed; this was discovered through trial and error. Finally, attempts to replicate Model 5's coefficients did not yield exact matches, though significance levels were consistent. This discrepancy was investigated by altering lead/lag specifications and transformations without success. An inquiry to the authors remains unanswered.

Overall, these minor inconsistencies do not undermine the central findings. The successful replication of Models 1-4 confirms that Nel and Righarts correctly implemented rare event logistic regression and that their principal conclusions about the relationship between disasters and conflict risk are robust to replication efforts.

TABLE 1. Natural Disasters and the Risk of Violent Civil Conflict, 1950–2000

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| *Dependent* | *All violent civil conflict (Smaller sample)†* | *All violent civil conflict* | *All violent civil conflict* | *All violent civil conflict* | *All violent civil conflict t + 1* |
| No. natural disasters per capita | 1.075** (0.280) | 1.186** (0.268) | 0.913** (0.287) | 0.089** (0.279) | 0.950** (0.306) |
| Infant mortality rate $t-1$ | 0.023** (0.007) | 0.023** (0.006) | | | 0.024** (0.006) |
| Infant mortality rate squared $t-1$ | −0.000* (0.000) | −0.000** (0.000) | | | −0.000** (0.000) |
| GDP per capita | | | −0.428** (0.083) | | |
| Mixed regime | 0.422** (0.164) | 0.524** (0.155) | 0.583** (0.149) | 0.568** (0.155) | 0.375* (0.160) |
| GDP growth | −0.060** (0.020) | −0.068** (0.020) | −0.058** (0.019) | −0.070** (0.019) | −0.048* (0.020) |
| "Youth bulge" | | | | 0.052** (0.015) | |
| Brevity of peace | 0.863** (0.181) | 0.872** (0.164) | 0.908** (0.168) | 0.978** (0.161) | 0.877** (0.165) |
| Constant | −4.992** (0.273) | −5.187** (0.255) | −0.505 (0.682) | −5.507** (0.467) | −5.172** (0.257) |
| *N* | 5893 | 7840 | 8203 | 7928 | 7663 |

Relogit regression coefficients with robust standard errors in parentheses.
$*p < .05$, $**p < .01$.
†Sample with missing values on regime and GDP growth variables.
*Note*: All violent civil conflict = onset of violent civil conflict with minimum of 25 deaths.

Table 2. Stata Replication Table — Natural Disasters and the Risk of Violent Civil Conflict, 1950–2000

| | 1 | | 2 | | 3 | | 4 | | 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Dependent* | *All violent civil conflict (Smaller sample)* | | *All violent civil conflict* | | *All violent civil conflict* | | *All violent civil conflict* | | *All violent civil conflict t+1* | |
| No. natural disasters per capita | 1.075*** | (0.280) | 1.186*** | (0.268) | 0.913*** | (0.287) | 0.892*** | (0.279) | 1.107*** | (0.286) |
| Infant mortality rate t-1 | 0.023*** | (0.007) | 0.023*** | (0.006) | | | | | 0.0257*** | (0.006) |
| Infant mortality rate squared t-1 | 0.000** | (0.000) | 0.000*** | (0.000) | | | | | 0.000*** | (0.000) |
| GDP per capita | | | | | -0.428*** | (0.083) | | | | |
| Mixed regime | 0.442*** | (0.164) | 0.524*** | (0.155) | 0.583*** | (0.149) | 0.568*** | (0.155) | 0.387** | (0.160) |
| GDP growth | -0.060*** | (0.021) | -0.068*** | (0.020) | -0.058*** | (0.019) | -0.070*** | (0.019) | -0.0515** | (0.020) |
| "Youth bulge" | | | | | | | 0.0515*** | (0.015) | | |
| Brevity of peace | 0.863*** | (0.181) | 0.872*** | (0.164) | 0.908*** | (0.168) | 0.978*** | (0.161) | 0.557*** | (0.166) |
| Constant | -4.991*** | (0.273) | -5.186*** | (0.255) | -0.505 | (0.682) | -5.507*** | (0.467) | -5.182*** | (0.258) |
| *N* | 5,892 | | 7,839 | | 8,203 | | 7,928 | | 7,662 | |

*Notes:* Robust standard errors in parentheses. *p <.1, **p <.05, ***p <.01.
Sample with missing values on regime and GDP growth variables.
All violent civil conflict = onset of violent civil conflict with minimum of 25 deaths.

# 4 An Alternative Method

While King and Zeng's relogit technique was an innovation for dealing with rare events, it has since been superseded in many respects by alternative approaches. Firth's penalized likelihood method is one such alternative, often proving more stable and widely implemented in contemporary statistical software. Firth's method reduces small-sample bias by introducing a penalty to the likelihood function, ensuring that parameter estimates remain finite and less biased, even in cases of extremely rare outcomes.

To evaluate the robustness of Nel and Righarts's findings, I applied the Firth method using R's `logistf` package. After loading the data and replicating the required transformations (lags, leads), I estimated a model analogous to Nel and Righarts's specification:

```
#R Code Sample using Firth method


rep_df <- load("nel_righarts_data.RData")
x <- x %>%
  group_by(state) %>%
  mutate(imr_t_1 = lag(imr),
         imr_sq_t_1 = lag(imr_sq),
         vccall_tp1 = lead(vccall),
         vccmajor_tp1 = lead(vccmajor))


#Model 1
t1_1 <- logistf::logistf(vccall ~ allND_pc + imr_t_1 + imr_sq_t_1 + mixedx
  + gdpgrox + brevity, data = x, firth = TRUE)


#Model 2
```

```r
t1_2 <- logistf::logistf(vccall ~ allND_pc + imr_t_1 + imr_sq_t_1 + mixed
  + gdpgro + brevity, data = rep_df, firth = TRUE)
```

Comparing these estimates to the Stata relogit results revealed similar patterns and levels of statistical significance. This convergence strongly suggests that Nel and Righarts's findings are not an artifact of the particular rare events methodology they employed. Instead, the fundamental relationship they identify—between natural disasters and the increased likelihood of civil conflict onset—holds across different analytical frameworks.

**Conclusion**

This replication confirms and extends the analysis offered by Nel and Righarts. The primary result—that natural disasters significantly increase the risk of violent civil conflict in certain socio-political and economic contexts—proves robust, surviving both the original relogit approach and the alternative Firth penalized regression method.

Nevertheless, the replication also highlights issues in the data and methods. Minor discrepancies in coefficient values, unnoted transformations of key variables, and incomplete matches in one of the models underscore the importance of thorough documentation. Data biases in EM-DAT and Uppsala/PRIO challenge causal claims and indicate the need for additional sensitivity analyses, bootstrapping, or the integration of complementary data sources.

Ultimately, this replication demonstrates that while Nel and Righarts's core conclusions are sound, future research should pay careful attention to data quality, model specification, and methodological choices. Incorporating methods designed to handle rare events data, validating assumptions through sensitivity analyses, and acknowledging data limitations will lead to more nuanced and reliable insights into how natural disasters shape the risk of violent civil conflict.