# Bayesian Insights into Aerial Bombing Strategies: An Ordered Logistic Regression Analysis of WWII Target Prioritization Against Germany*

Yunzhao Li

April 17, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

## 1 Introduction

Target selection in WWII aerial campaigns was a crucial aspect that had far-reaching implications for the war's outcome. This paper aims to dissect the multi-faceted considerations behind target prioritization using Bayesian statistical modeling, providing insights into historical strategies as well as methodological implications for contemporary analysis. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2….

## 2 Data

The data for this study was meticulously curated from declassified military records, comprising numerous raids, their targets, and associated strategic variables.

*Code and data are available at: https://github.com/yunzhaol/aerial_bomb_priority.git.

Table 1: 2022 CES (Cooperative Election Study) Data (Cultural)

## 2.1 Variable

Our analysis focuses on the following variables: target priority, target industry, country of the mission's origin, total tons of bombs dropped, and the number of aircraft partaking in the mission.

## 2.2 Justification

The selected variables are integral for understanding the strategic matrix that guided target selection, enabling a thorough investigation into the strategic landscape of WWII aerial campaigns.

# 3 Model

The goal of our modelling strategy is twofold. Firstly,…

We constructed a Bayesian ordered logistic regression model to analyze the nuanced hierarchy in target prioritization, utilizing the rstanarm package to accommodate the ordinal nature of target priority levels.

Our model statistically infers the relationship between target priority and various strategic factors, providing a probabilistic assessment of their impacts.

Background details and diagnostics are included in Appendix B.

## 3.1 Model set-up

Define $y_i$ as the number of seconds that the plane remained aloft. Then $\beta_i$ is the wing width and $\gamma_i$ is the wing length, both measured in millimeters.

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \alpha + \beta_i + \gamma_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta &\sim \text{Normal}(0, 2.5) && (4)\\
\gamma &\sim \text{Normal}(0, 2.5) && (5)\\
\sigma &\sim \text{Exponential}(1) && (6)
\end{aligned}
$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

### 3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance $\theta$.

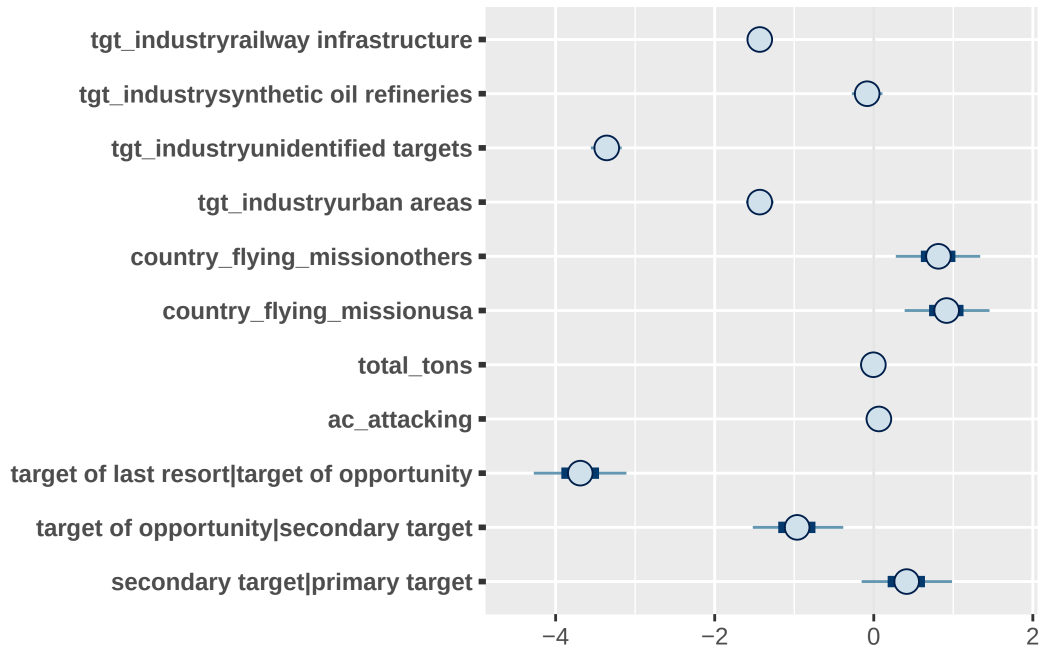
# 4 Results

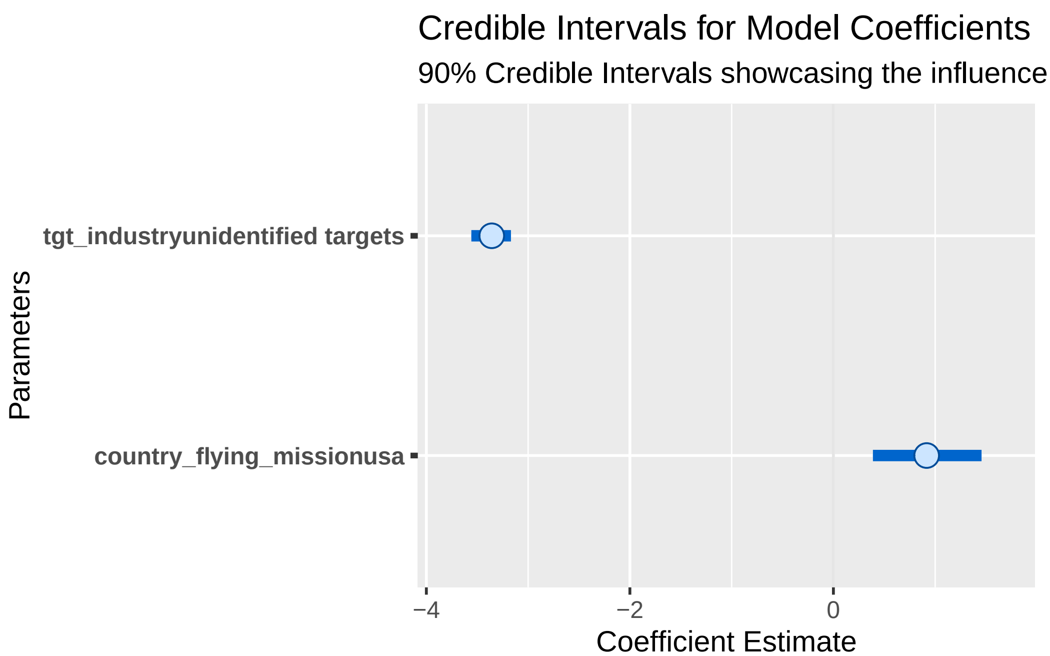
The analysis revealed distinct patterns in target prioritization, highlighting the role of industry type and the mission's originating country in determining target selection, while the impact of tonnage and aircraft numbers was relatively subdued. Table 2.

```
Model Info:
 function:     stan_polr
 family:       ordered [logistic]
 formula:      tgt_priority_explanation ~ tgt_industry + country_flying_mission +
       total_tons + ac_attacking
 algorithm:    sampling
 sample:       4000 (posterior sample size)
 priors:       see help('prior_summary')
 observations: 5000

Estimates:
                                             mean   sd   10%   50%   90%
tgt_industryrailway infrastructure         -1.4    0.1 -1.6  -1.4  -1.3
tgt_industrysynthetic oil refineries       -0.1    0.1 -0.2  -0.1   0.1
tgt_industryunidentified targets           -3.4    0.1 -3.5  -3.4  -3.2
tgt_industryurban areas                    -1.4    0.1 -1.6  -1.4  -1.3
country_flying_missionothers                0.8    0.3  0.4   0.8   1.2
country_flying_missionusa                   0.9    0.3  0.5   0.9   1.3
total_tons                                  0.0    0.0  0.0   0.0   0.0
ac_attacking                                0.1    0.0  0.1   0.1   0.1
target of last resort|target of opportunity -3.7    0.4 -4.1  -3.7  -3.2
target of opportunity|secondary target     -1.0    0.3 -1.4  -1.0  -0.5
secondary target|primary target             0.4    0.3  0.0   0.4   0.9

Fit Diagnostics:
                                mean   sd   10%   50%   90%
mean_PPD:target of last resort 0.0    0.0  0.0   0.0   0.1
```

```
mean_PPD:target of opportunity 0.2    0.0  0.2   0.2   0.3
mean_PPD:secondary target      0.2    0.0  0.2   0.2   0.2
mean_PPD:primary target        0.5    0.0  0.5   0.5   0.5

The mean_ppd is the sample average posterior predictive distribution of the outcome variable

MCMC diagnostics
                                            mcse Rhat n_eff
tgt_industryrailway infrastructure          0.0  1.0  3583
tgt_industrysynthetic oil refineries        0.0  1.0  4392
tgt_industryunidentified targets            0.0  1.0  3608
tgt_industryurban areas                     0.0  1.0  3491
country_flying_missionothers                0.0  1.0  5105
country_flying_missionusa                   0.0  1.0  4972
total_tons                                  0.0  1.0  5818
ac_attacking                                0.0  1.0  3987
target of last resort|target of opportunity 0.0  1.0  5262
target of opportunity|secondary target      0.0  1.0  5068
secondary target|primary target             0.0  1.0  4950
mean_PPD:target of last resort              0.0  1.0  3779
mean_PPD:target of opportunity              0.0  1.0  3433
mean_PPD:secondary target                   0.0  1.0  3451
mean_PPD:primary target                     0.0  1.0  4147
log-posterior                               0.1  1.0   721

For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective

Priors for model 'aerial_priority_model'
------

Coefficients
 ~ flat

Counts
 ~ dirichlet(concentration = [1,1,1,...])
------
See help('prior_summary.stanreg') for more details
```

Table 2: Explanatory models of flight time based on wing width and wing length



# 5 Discussion

## 5.1 First discussion point

The implications of our study stretch beyond historical curiosity, offering a lens through which modern strategic decisions can be evaluated, and the methodology refined.

## 5.2 Second discussion point

## 5.3 Third discussion point

## 5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

# Appendix

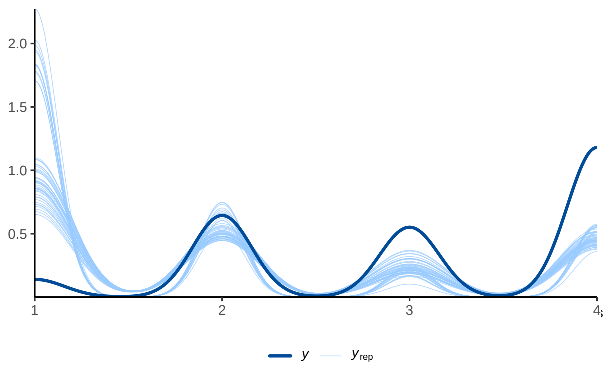
## A Additional data details

## B Model details

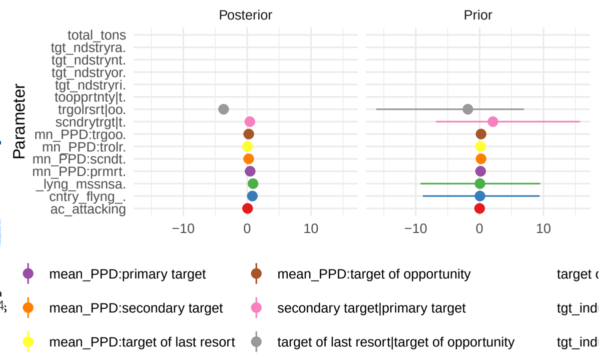
### B.1 Posterior predictive check

In Figure 1a we implement a posterior predictive check. This shows...

In Figure 1b we compare the posterior with the prior. This shows...




(a) Posterior prediction check



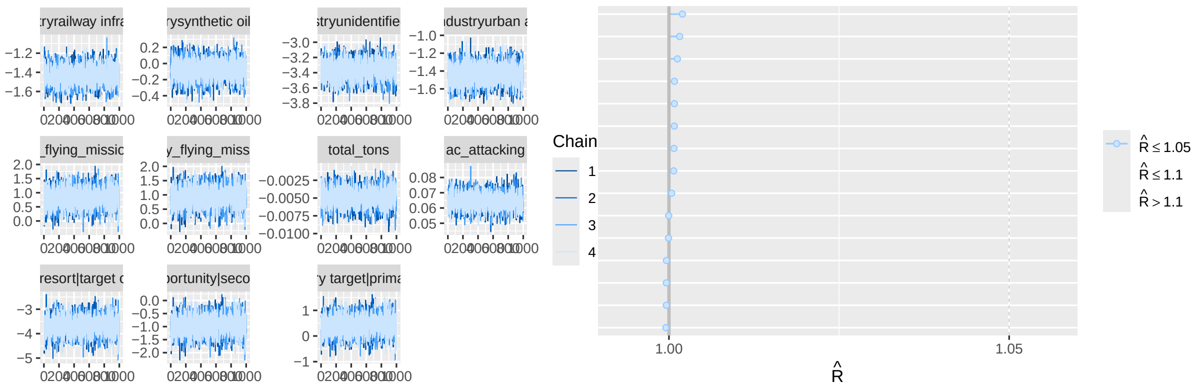

(b) Comparing the posterior with the prior

Figure 1: Examining how the model fits, and is affected by, the data

### B.2 Diagnostics

Figure 2a is a trace plot. It shows... This suggests...

Figure 2b is a Rhat plot. It shows... This suggests...




(a) Trace plot

(b) Rhat plot

Figure 2: Checking the convergence of the MCMC algorithm

## References

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.