Climate Change, Biodiversity Loss*

A replication analysis

Julia Kim

April 2, 2024

Abstract

Abstract here.

Table of contents

*Code and data are available at: https://github.com/julia-ve-kim/US_Climate_Change_Biodiversity; Minimal and Full Replication on Social Science Reproduction platform available at https://doi.org/10.48152/ssrp-9vc2-gp37 and https://doi.org/10.48152/ssrp-qvfh-0r62, respectively.

1 Introduction

2 Data

2.1 Source

The paper and data used for replication are obtained from “Noah’s Arc in a Warming World: Climate Change, Biodiversity Loss, and Public Adaptation Costs in the United States” (Moore et al. 2022), published in the *Journal of the Association of Environmental and Resource Economists*. Their analysis sought, in part, to explore how the probability of a species being listed under the ESA was influenced by its conservation status, utility value to humans, and evolutionary distinctiveness. Our replication seeks to address the validity of these three findings and to discuss the projected increases in species listing due to climate change.

2.2 Methodology

The original dataset is made publicly available by Moore et al. (2022). As part of our reproduction, we removed variables not of interest to our analysis and clarified the names of variables to make them easier to work with. A further discussion of the variables of interest that were retained in the final dataset, the source of their data, and the methodology employed to clean them, follow.

2.2.1 Listing Status of Species

2.2.2 Endangerment of Listed and Nonlisted Species

2.2.3 Google Ngrams for Listed and Nonlisted Species

An n-gram is defined as a sequence of n words in some particular order: a 2-gram, for instance, is a two-word sequence like “*Balaena mysticetus*”, whilst a 3-gram is three-word sequence like “*Lithobates areolatus circulosus*.” An on-line search engine, the Google Books Ngram Viewer uses text processing to provide the frequency of such n-grams in a corpus of over 25 million digitised books published over the course of more than 500 years (Google 2020). These books are published in eight languages; in particular, the English corpus of 2019 consists over over 16 million of such books published between 1470 and 2019 (Michel et al. 2011).

As argued by Moore et al. (2022), the rich variety of ways in which a species may provide utility – through cultural significance, commercial value, scientific interest, and so forth – should likely directly influence the frequency with which the species is written about over time. As a result, they argue that it is reasonable to use common and scientific name n-grams as an additional estimate of the utility value of species to humans. We adopt their reasoning, and provide potential limitations of this argument in Section 5.5.

To determine the n-gram frequencies for species names, Moore et al. (2022) performed case-insensitive searches in Google Books’ English corpus of 2019 for all common and scientific names present in the NatureServe and ESA archives between 1800 and 2016. They assigned a frequency of 0 across all years whenever a name failed to return valid data. Here, the purpose of treating common and scientific name n-grams separately owes to the fact that common names pose a few challenges. In particular, a species may be designated colloquially by multiple common names or may lack a common name altogether (Cheese 2021). Conversely, a single common name may be used for multiple different species (e.g., a “millipede” referring to any of the 10,000 species in the anthropod class Diplopoda) or may have additional, unrelated meanings or uses (e.g., “British soldier” referring to an army serviceman of the United Kingdom or to the lichen *Cladonia cristatella*) (Cheese 2021). To account for such complicating factors, common and scientific names were thus treated separately.

As described by Moore et al. (2022), n-gram frequencies were then aggregated to the species level. More specifically, for unlisted species, the two n-gram frequencies were taken to be their respective averages from 1950 to 2016. For listed species, they were calculated as their respective average frequencies from 1950 to 10 years *prior* to the date of the listing decision. This was done to minimise the bias to the n-gram values following any publicity generated by the listing decision (Moore et al. 2022). Owing to how small the raw n-gram frequencies were, all frequencies were z-score normalised before analysis: this was accomplished by calculating the difference between each raw frequency and the corresponding n-gram mean and dividing the whole by the corresponding n-gram standard deviation. At last, to further account for the challenges posed by common names, as discussed previously, any species whose standardised common name n-gram had a frequency > 10 times its corresponding standardised scientific n-gram frequency were discarded in the dataset (Moore et al. 2022).

A plot of the mean standardised common and scientific n-gram frequencies per taxon is shown in Figure 1.

2.2.4 Evolutionary Distinctiveness

3 Model

3.1 Model set-up

3.1.1 Model justification

4 Results

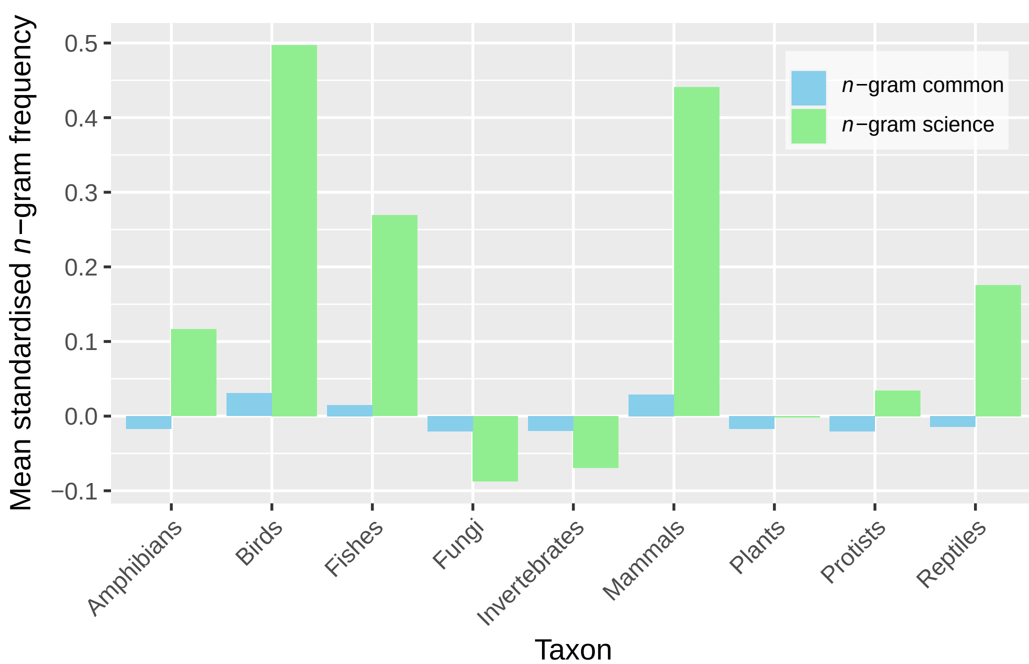

Figure 1: Distribution of the mean standardised common and scientific n-gram frequencies per taxon.

5 Discussion

5.1 Findings

Our paper has successfully replicated three of their major findings:

(1) The probability of listing changes with conservation status, decreasing monotonically from the most to the least endangered.

(2) There is some evidence that the probability of listing of a species is associated with its utility value to humans. In particular, imperiled vertebrates – including reptiles, fish, birds, amphibians, and mammals – are much more likely to be listed than plants and invertebrates. However, listing probability appears to be significantly affected by neither scientific nor common name n-grams.

(3) There is conflicting evidence as to whether the probability of listing is related to the distinctiveness of a species. Whilst species in smaller genera are more likely to be listed, there is no evidence that the species' phylogenetic diversity plays a role in the listing decision.

5.2 Case study

5.3 Ethical Implications

5.4 Accounting for Bias

5.5 Limitations

Although the ESA mandates the listing all species at risk of extinction, capacity and budgetary constraints often slow down the the listing process, leading to a backlog of species awaiting assessment (Alexander 2010). In some cases, USFWS will designate a species as "warranted but precluded" to acknowledge that listing the species is necessary but of lesser priority than other species in need of greater protection (Alexander 2010).

A limitation of the n-gram measure is that it captures only English-language media and includes only material published in books, excluding popular media such as amagazines, websites or newspapers.

Limitations to using ngram as a proxy for species value

5.6 Future Research

6 Appendix

7 Additional data details

References

Alexander, Kristina. 2010. "Warranted but Precluded: What That Means Under the Endangered Species Act (ESA)." Washington, DC: Congressional Research Service.

Cheese, Tyler. 2021. "What's in a Name? Problematic Names in the World of Wildlife." *Canadian Geographic*. Canadian Geographic. https://canadiangeographic.ca/articles/whats-in-a-name-problematic-names-in-the-world-of-wildlife/.

Google. 2020. "Google Books Ngram Viewer." https://storage.googleapis.com/books/ngrams/books/datasetsv3.html.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K Gray, Joseph Porter Pickett, Dale Hoiberg, et al. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331 (6014): 176–82. https://doi.org/10.1126/science.1199644.

Moore, Frances C, Arianna Stokes, Marc N Conte, and Xiaoli Dong. 2022. "Noah's Ark in a Warming World: Climate Change, Biodiversity Loss, and Public Adaptation Costs in the United States."

Journal of the Association of Environmental and Resource Economists 9 (5): 981–1015.