Improving Maritime Shipping Emissions Estimates Using Machine Learning

Allen Peters*, Hiroyuki Kasahara, Oliver Xu, Jasper Zhao

DRAFT

February 27, 2024

Abstract

We explore the potential of machine learning algorithms to improve upon engineering estimates of CO_2 emissions from maritime shipping. Traditional estimates rely on engineering approximations that may not entirely capture actual fuel use. We match reported annual ship-level emissions from a European Union emissions reporting program with tracking data and technical characteristics for the global fleet of dry bulk ships. As a baseline, we follow industry standard procedures to calculate engineering estimates of annual ship-level emissions. We then train various machine learning algorithms on the residual—the discrepancy between reported and calculated emissions—and are able to improve out-of-sample prediction.

*Corresponding author.
Vancouver School of Economics, University of British Columbia, 6000 Iona Drive, Vancouver, BC, Canada.
Email addresses: apeters@protonmail.com (A. Peters), hkasahar@mail.ubc.ca (H. Kasahara), xjy099@student.ubc.ca (O. Xu), zhaoxinbo1999@outlook.com (J. Zhao).
We thank ... for excellent research assistance. Data acquisition was funded by a generous grant from the Centre for Innovative Data in Economics Research.

1 Introduction

Motivation:

- Shipping emissions are a significant contributor to global CO_2 emissions.
- Traditional estimates rely on engineering approximations that may not entirely capture actual fuel use.
- Higher frequency estimates (typically published with lag)
- Existing estimates may suffer from poor data quality

What we do:

- Provide a machine learning approach to more accurately estimate CO_2 emissions
- Higher frequency estimates (typically published with lag)
- Take emissions reports as truthful and accurate
- Predict residual between reported and calculated emissions - thereby guarantee improvement upon standard bottom-up approach
- Dealing with missing data for extrapolation
- We are able to explain x% of variation in residual
- Better untracked (no AIS data) predictions???

Contribution and results preview:

- Machine learning provides a flexible functional form that can:
  - Capture divergence from engineering relationships
  - Help to mitigate error from imperfect/incomplete tracking data
- Bottom-up vs. top-down
  - How do reported values compare?
  - How do our estimates compare?

Literature:

- IMO Fourth GHG Study
- precursors for estimation method:
  -
- Paper on detecting misreporting in EU MRV
- Other literatures???

2 Data

- AIS data
  - Descriptive statistics on: missing observations, ...?
- Vessel characteristics from Clarkson Research WFR
- EU MRV
  - Context
  - Potential for misreporting
  - descriptive statistics on number of reporting ships, comparison of reporting to non-reporting over various dimensions (which ones?)

3 Methodology

- Matching AIS data to EU trips
  - Compare to page 140 of IMO Fourth GHG Study (Faber et al., 2020)
- Replicating IMO bottom-up approach (Highlight differences)
  - Interpolating missing data
- Machine Learning
  - Variable selection
  - Algorithms
    - Linear: lasso, ridge, (elastic net?)
    - discuss why not neural net
    -
  - Hyper parameter tuning (cross-validation)
- Extrapolation
  - Comparison of MRV reporting fleet to non-reporting

4 Results

- Linear vs. non-linear models
- discuss log-additive discrepancy
- Why do some models work better than others?

5 Conclusion