



'How can I possibly follow the Cohort Component Model?'

Ridhi Kashyap, LCDS Away Day, 2023.

Current Projects Nearing/Bearing Fruition

- I'd first begin by talking about what's currently coming through:



- The SHAPE of Research Impact.

Current Projects Nearing/Bearing Fruition

- I'd first begin by talking about what's currently coming through:

nature climate change

Brief Communication

Shoring emissions through used vehicle exports

<https://doi.org/10.1038/s41558-024-0043-1>

Received: 30 July 2023

Accepted: 28 January 2024

Published online: 20 February 2024

Check for updates

Riad Jedin Newman^{1,2}, Kayla Schulte^{1,2}, Muel Marilou Morelles^{1,2,3}, Charles Reibel^{1,2} & Douglas R. Lesonsky^{1,2}

Policies to reduce transport emissions often overlook the international flow of used vehicles. We quantify the rate at which used vehicles generated CO₂ and pollution for all used vehicles exported from Great Britain – a globally leading used vehicle exporter – across 2005–2021. Destined for low- to middle-income countries, exported vehicles fall much below US standards and, even under extremely optimistic functioning as new assumptions, generate at least 13–53% more emissions than scrapped or on-road vehicles.

Transport is the largest emitting sector of greenhouse gases, accounting for a quarter to a third of all emissions at developed countries¹, with serious consequences for both climate and health^{2–4}. Air pollutants such as nitrogen oxides (NO_x) and particulate matter (PM_{2.5}), which are often byproducts of combustion, are emitted and cause millions of deaths each year⁵. These impacts fall most heavily on lower- and middle-income countries (LMICs), which suffer more overall and per capita premature deaths from air pollution^{6,7} and stand to suffer the greatest impacts from climate change⁸. The source of vehicles in LMICs is dominated by imported used vehicles^{9–11}. As of 2020, LMIC countries receiving used vehicles had no vehicle emission standards¹² and only 8 had “very good” emissions regulations. However, the United States, European Union, Japan and United Kingdom collectively supply 90% of used vehicles exported to these LMICs¹³. The potential for rapid regulation to reduce emissions is just these problems, all of which already maintain high vehicle emissions standards.

Using comprehensive government databases, we quantify per kilometer rates at which vehicles generate carbon and pollution for every vehicle in a LMIC country (Great Britain) from January 2005 to December 2020. We compare these vehicle emissions to those of new vehicles sold in the same period and those that would have been driven if they had not been scrapped.

These data reveal substantially higher rates of CO₂ and pollution generation in exported vehicles, even under optimistic assumptions. Our new emissions intensity estimates that account for vehicle modifications or vehicle degradation with age (Fig. 1 and Extended Data Fig. 1) export can generate at least 21 g CO₂ per kilometer (km) for vehicles exported for the same period (Fig. 1a) and Supplementary Table 1 and at least 29 g CO₂ per kilometer

than the corresponding on-road used car fleet (Fig. 1a, mean 1870, 214 and 164 g per km² for exported, scrapped and on-road fleets, respectively, interquartile range (IQR) 1510–2213, 1473–1991 and 1242–1801 g km²).

Emission figures were even more striking for other pollutants. Exported cars used smaller amounts of high-toxicity particulates (Fig. 1b) but 46% more “toxic NO_x CO₂ higher Fig. 1b) than scrapped cars. Likewise, observed engine capacities were larger (Fig. 1c) and fuel efficiency at least 9% worse (by 0.3 miles per gallon (MPG), mean 18.5, 16.3 and 14.4 mpg for exported, scrapped and on-road used fleets respectively; IQR 10–47 mpg exported, 10–47 mpg scrapped and 17–49 mpg on-road fleet).

A substantial fraction (42%) of exported diesel vehicles were predicted to fail the current 1000 g a emission standard¹⁴ that limits the high-massivity maximum for all vehicles registered after 2004. A surprising 33% were predicted to fail the 1000 g a “toxic NO_x CO₂ higher Fig. 1b) standard, and 9% failed the 1000 g a carbon intensity and NO_x standards. These differences were for the worst of our age-adjusted, vehicle age, engine capacity and fuel efficiency. Shifting to the average, similar or even larger gaps in pollution rates were observed for the median pollution rate of exported, scrapped and on-road used cars or on-road used vehicles (Fig. 1).

Early results that for ex-ante export can reveal that the gap between exported, on-road and scrapped fleets continues over time (Fig. 1a, 1b). Upon further parsing and thus could expansion of this gap over 2005–2021 alongside decreases of road prices, used car prices and vehicle ending regulations during the COVID-19 pandemic (Fig. 1b). That is, Great Britain previously issues lower emissions vehicles while receiving higher emissions vehicles (Fig. 1b). Whereas geographical disparities due to sources

- The SHAPE of Research Impact.

- Nature Climate Change.

¹Environmental Center for Sustainable Science, Huxford Department of Population Health, University of Oxford, UK, ²University College, University of Oxford, Oxford, UK, ³MRC Centre for Environment and Health, Imperial College London, London, UK, ⁴Health Data University of Oxford, Oxford, UK, ⁵Department of Biostatistics, University of Oxford, Oxford, UK, ⁶in-mail: douglas.lesonsky@ox.ac.uk

Current Projects Nearing/Bearing Fruition

- I'd first begin by talking about what's currently coming through:

- The SHAPE of Research Impact.

- Nature Climate Change.vspace

- InterModel Vigorish (x2).

The InterModel Vigorish (IMV) as a flexible and portable approach for quantifying predictive accuracy with binary outcomes

Benjamin W. Domingue^{1,2}, Charles Rahul^{3,4}, Jessica Fust^{1,2}, Jeremy Frame^{4,2}, Khat Kanopke^{4,2}, Alexandre Rigau^{4,2}, Ben Stenhouse^{4,2}, and Ajay Shastri Tripathi^{4,2}

¹Stanford University

²University of Oxford

³University of Michigan

⁴Institute for Future Studies, Stockholm

⁵Umeå University

⁶benaminge@stanford.edu & charles.rahul@biology.ox.ac.uk

⁷Alphabetinc

October 23, 2023

Acknowledgements

This work was supported by the Jacobs Foundation (B.D.), the Leverhulme Center for Demographic Science (The Leverhulme Trust) (C.R.) and Wellcome College (C.R.), and from Bundesministerium für Bildung und Forschung (BMBF, A.R.). The authors would like to acknowledge: Lisa Bell, Richard Brown, Kjetil Christen, David Clancy, Tim Eapen, George Jernan, David Kellard, Mike Siler, Nilsen Trank, Mark Verhaeg, Rikman Moun, and Yuhua Wu. We would like to thank the authors of the manuscript, and Yuhua Wu and Yuhua Wu for assistance with the InterModel. The IMV (InterModel and InterModel Study) is sponsored by the National Institute on Aging (grant number NIA U01AG059740) and is conducted by the University of Michigan.

Author Contribution Statement

BD conceived the original idea and study design, wrote simulations, conducted empirical analyses, and InterModel and revised the manuscript. CR undertook code reviews, empirical analyses, visualizations, and contributed to drafting and revising the manuscript. NK worked on empirical applications, and developed a Shiny application. AST and AR worked on the key motivating features of the IMV, including but not limited to statistical aspects and analyses to the InterModel. JF, JP, NK, AR, BS and AST all commented on the manuscript.

Current Projects Nearing/Bearing Fruition

- I'd first begin by talking about what's currently coming through:

- The SHAPE of Research Impact.

- Nature Climate Change.vspace

- InterModel Vigorish (x2).

- Procurements Data (JRSSA)

1 The Role of the Third Sector in Public Health Service Provision:
2 Evidence from 25,338 heterogeneous procurement datasets

3 Charles Rahal^{1,2} and John Mahan³

4 ¹Leverhulme Centre for Demographic Science, University of Oxford

5 ²Nuffield College, University of Oxford

6 ³Third Sector Research Centre, University of Birmingham

7 February 19, 2024

8 Abstract

9 We examine the role of non-profits within publicly funded healthcare which runs parallel to
10 private provision, in a "two-tier" system through a unique Big Data pipeline. Scrapping tens of
11 thousands of heterogeneous transaction datasets across a consolidating hierarchy, the processed
12 dataset contains over 2.1Bb worth of transaction across 1.8m+ rows of client data, spanning
13 2013-2020. Information includes but is not limited to date of procurement transaction, supplier
14 name, and transaction value. We utilize this dataset to test a range of hypotheses related to the
15 introduction of the Health and Social Care Act of 2012. The proportion of contracts placed with
16 nonprofit organisations is relatively low, with limited evidence as to whether this has increased
17 the involvement of third sector organisations in line with NHS "marketisation". We analyse the
18 patterns of procurement by supplier category, skill of service delivery (ICD10) and ICD10 code,
19 and regions used to show the range types of services which voluntary organisations supply. We
20 also analyse the patterns of commissioning across entity class and rate distributions of regional
21 clusters. We conclude with a consideration of high-value Community Interest Companies, and
22 discuss potential further areas of research within a healthcare context, which such government
23 transaction data makes possible.

24 **Keywords:** Healthcare Supply, National Health Service, Data Visualisation, Big Data

25 **JEL Codes:** H10, H41, H51, I11, I10

For correspondence: Charles Rahal, Leverhulme Centre for Demographic Science, Demographic Science Unit,
University of Oxford, OX1 1LN, United Kingdom. Tel: +44(0) 1865 280775. Email: charles.rah@demographic.ox.ac.uk.
The research code library which accompanies this work can be found at github.com/charal/SHAPES. The authors
are grateful to co-authors on earlier versions of the work from Mark Everell, David Brinkley, Martin Miles,
Lucy Reynolds and James Ross. Technical research assistance provided by Iain Kewster. The origin of this work
arises from a meeting and participating sessions funded by the ESRC grant number ES/S000000/1 and research
ES/S000000/1, with ongoing funding largely and gratefully acknowledged from the British Academy and the
Leverhulme Trust Grant R2-2019-001. The Leverhulme Centre for Demographic Science (LCDS), and Nuffield College,
Oxford, are grateful to the ESRC for their support. The authors are grateful to participants at the International Conference for Administrative
Data Research, the Economic Society, the ESRC for National Institute, the Spatial Unit at the Department
for Levelling Up, and the Government Data Science Community Support.

Current Projects Nearing/Bearing Fruition

- I'd first begin by talking about what's currently coming through:

- The SHAPE of Research Impact.

- Nature Climate Change.vspace

- InterModel Vigorish (x2).

- Procurements Data (JRSSA)

- Life Course Prediction

Short Essays Predict Later Life Cognition and Educational
Attainment Close to or Better than Genomics or Expert
Assessment

Tobias Wolfgram¹, Felix C. Tropf^{2,3}, and Charles Rahni⁴

¹Institute of Sociology, Heidelberg University, Germany

²Centre for Longitudinal Studies, Social Research Institute, University College London, UK

³Department of Sociology, Purdue University, USA

⁴Levinsholm Centre for Demographic Science and Nuffield College, University of Oxford, UK

January 10, 2024

Abstract

Previous research using standard social survey data has emphasized a relative lack of power when predicting educational and psychological outcomes. Leveraging a unique longitudinal dataset, we explore the potential predictive collaps for educational attainment, cognitive abilities, and non-cognitive traits using recent advances in deep learning-based natural language processing, the *Frontiers of genetic techniques*, and transfer-based assessment. We show that short navigational concepts of around 200 words written at age 11 accurately predict cognitive and non-cognitive traits at the same and later age, similar to teacher assessment. The same is true for predicting final educational attainment. Genetic prediction is substantially less accurate. When these three non-standard measures are combined into an ensemble model, they can nearly perfectly predict cognitive ability (F1), close to test-retest reliability, and explain 38% of individual differences in educational attainment. These findings show that recent advances in large language modelling *explore* behavioural and social outcomes with tools for accurate prediction, strongly motivating responsible research into a deeper understanding of the information which such models are accurately able to contextualise.

Keywords: Psychological Methods, Longitudinal Surveys, Deep Learning

Data Availability: All non-genetic data associated with the project can be downloaded from the website of the UK Data Service, organized in sets of checklists, or just one for *harry* 0.3 and additional ones for each subsequent wave. Genes are available separately. Access to genetic data requires a separate application to the data access committee of the centre for longitudinal studies (UCL). **Code Availability:** All statistical software used for generating the results in this paper are currently listed in a generic GitHub repository. The code will be made available following the article's acceptance for publication. **Acknowledgements:** F.C.T. was supported by the Laboratory of Excellence in Economics and Decision Sciences (Ladex Excellence) funded under the French National Research Agency (ANR) Investments of the Future (Grant ANR-11-LABX-0907) and the UCL project FINDER (EP/T01006/1). C.R. is grateful for support from the Levinsholm Trust (Grant RC-S018-000) for the Levinsholm Centre for Demographic Science, and Nuffield College. The authors declare no competing interests. **Correspondence:** Correspondence concerning this article should be addressed to Tobias Wolfgram, Faculty of Sociology, Heidelberg University, Universitätsstrasse 25, 69115 Heidelberg, Germany. Email: tobias.wolfgram@uni-heidelberg.de.

Work-In-Progress

- I am *obsessed* with P/QRNGs. **Very** exciting preprint coming soon!
- New ML-IMV paper with \$SHAP, model ablation and more.
- We're developing libraries to use scientometric data *at scale*.
- New ML-based advancements to the (causal) SCM literature.
- RobustiPy! Our new library for model uncertainty. Great hackathon!
- We've cleaned and linked 18m probate records; LCDS seminar in HT 2024!
- GWAS Diversity Monitor v2.0 is hopefully coming soon...
- Post-SHAPE-REF 2021 analysis papers are also a high priority.

Work-In-Progress

- I am *obsessed* with P/QRNGs. **Very** exciting preprint coming soon!
- New ML-IMV paper with \$SHAP, model ablation and more.
- We're developing libraries to use scientometric data *at scale*.
- New ML-based advancements to the (causal) SCM literature.
- RobustiPy! Our new library for model uncertainty. Great hackathon!
- We've cleaned and linked 18m probate records; LCDS seminar in HT 2024!
- GWAS Diversity Monitor v2.0 is hopefully coming soon...
- Post-SHAPE-REF 2021 analysis papers are also a high priority.

Work-In-Progress

- I am *obsessed* with P/QRNGs. **Very** exciting preprint coming soon!
- New ML-IMV paper with \$SHAP, model ablation and more.
- We're developing libraries to use scientometric data *at scale*.
- New ML-based advancements to the (causal) SCM literature.
- RobustiPy! Our new library for model uncertainty. Great hackathon!
- We've cleaned and linked 18m probate records; LCDS seminar in HT 2024!
- GWAS Diversity Monitor v2.0 is hopefully coming soon...
- Post-SHAPE-REF 2021 analysis papers are also a high priority.

Work-In-Progress

- I am *obsessed* with P/QRNGs. **Very** exciting preprint coming soon!
- New ML-IMV paper with \$SHAP, model ablation and more.
- We're developing libraries to use scientometric data *at scale*.
- New ML-based advancements to the (causal) SCM literature.
- RobustiPy! Our new library for model uncertainty. Great hackathon!
- We've cleaned and linked 18m probate records; LCDS seminar in HT 2024!
- GWAS Diversity Monitor v2.0 is hopefully coming soon...
- Post-SHAPE-REF 2021 analysis papers are also a high priority.

Work-In-Progress

- I am *obsessed* with P/QRNGs. **Very** exciting preprint coming soon!
- New ML-IMV paper with \$SHAP, model ablation and more.
- We're developing libraries to use scientometric data *at scale*.
- New ML-based advancements to the (causal) SCM literature.
- RobustiPy! Our new library for model uncertainty. Great hackathon!
- We've cleaned and linked 18m probate records; LCDS seminar in HT 2024!
- GWAS Diversity Monitor v2.0 is hopefully coming soon...
- Post-SHAPE-REF 2021 analysis papers are also a high priority.

Work-In-Progress

- I am *obsessed* with P/QRNGs. **Very** exciting preprint coming soon!
- New ML-IMV paper with \$SHAP, model ablation and more.
- We're developing libraries to use scientometric data *at scale*.
- New ML-based advancements to the (causal) SCM literature.
- RobustiPy! Our new library for model uncertainty. Great hackathon!
- We've cleaned and linked 18m probate records; LCDS seminar in HT 2024!
- GWAS Diversity Monitor v2.0 is hopefully coming soon...
- Post-SHAPE-REF 2021 analysis papers are also a high priority.

Work-In-Progress

- I am *obsessed* with P/QRNGs. **Very** exciting preprint coming soon!
- New ML-IMV paper with \$SHAP, model ablation and more.
- We're developing libraries to use scientometric data *at scale*.
- New ML-based advancements to the (causal) SCM literature.
- RobustiPy! Our new library for model uncertainty. Great hackathon!
- We've cleaned and linked 18m probate records; LCDS seminar in HT 2024!
- GWAS Diversity Monitor v2.0 is hopefully coming soon. . .
- Post-SHAPE-REF 2021 analysis papers are also a high priority.

Work-In-Progress

- I am *obsessed* with P/QRNGs. **Very** exciting preprint coming soon!
- New ML-IMV paper with \$SHAP, model ablation and more.
- We're developing libraries to use scientometric data *at scale*.
- New ML-based advancements to the (causal) SCM literature.
- RobustiPy! Our new library for model uncertainty. Great hackathon!
- We've cleaned and linked 18m probate records; LCDS seminar in HT 2024!
- GWAS Diversity Monitor v2.0 is hopefully coming soon. . .
- Post-SHAPE-REF 2021 analysis papers are also a high priority.

On the Horizon

- On the 'Limits to Prediction in Social Systems':
 - Involves innovating theoretical paradigms and tests to ML-based CSS.
 - Based on the idea of 'learning curves', but more work required.
 - **Major Question:** How predictable are *all* facets of people's lives?
- On the 'Evolution of Science':
 - Europe lacks a major lab which analyses the scientific record at scale.
 - **Major Question:** How has scientific collaboration evolved?
 - Has it evolved in a 'good' way? Linked data to evaluate.
- On 'The Supply of Health and Social Care':
 - There are some great epidemiological ABM models.
 - So much more work is needed to model the supply side.
 - **Major Question:** How can health and social care systems function efficiently?

On the Horizon

- On the 'Limits to Prediction in Social Systems':
 - Involves innovating theoretical paradigms and tests to ML-based CSS.
 - Based on the idea of 'learning curves', but more work required.
 - **Major Question:** How predictable are *all* facets of people's lives?
- On the 'Evolution of Science':
 - Europe lacks a major lab which analyses the scientific record at scale.
 - **Major Question:** How has scientific collaboration evolved?
 - Has it evolved in a 'good' way? Linked data to evaluate.
- On 'The Supply of Health and Social Care':
 - There are some great epidemiological ABM models.
 - So much more work is needed to model the supply side.
 - **Major Question:** How can health and social care systems function efficiently?

On the Horizon

- On the 'Limits to Prediction in Social Systems':
 - Involves innovating theoretical paradigms and tests to ML-based CSS.
 - Based on the idea of 'learning curves', but more work required.
 - **Major Question:** How predictable are *all* facets of people's lives?
- On the 'Evolution of Science':
 - Europe lacks a major lab which analyses the scientific record at scale.
 - **Major Question:** How has scientific collaboration evolved?
 - Has it evolved in a 'good' way? Linked data to evaluate.
- On 'The Supply of Health and Social Care':
 - There are some great epidemiological ABM models.
 - So much more work is needed to model the supply side.
 - **Major Question:** How can health and social care systems function efficiently?

On the Horizon

- On the 'Limits to Prediction in Social Systems':
 - Involves innovating theoretical paradigms and tests to ML-based CSS.
 - Based on the idea of 'learning curves', but more work required.
 - **Major Question:** How predictable are *all* facets of people's lives?
- On the 'Evolution of Science':
 - Europe lacks a major lab which analyses the scientific record at scale.
 - **Major Question:** How has scientific collaboration evolved?
 - Has it evolved in a 'good' way? Linked data to evaluate.
- On 'The Supply of Health and Social Care':
 - There are some great epidemiological ABM models.
 - So much more work is needed to model the supply side.
 - **Major Question:** How can health and social care systems function efficiently?