IJRM Special Session

Open Science (OS): A meta-analysis of modelling practices

30 May 2025

Hannes Datta Harald van Heerde Oliver Koll Koen Pauwels Karin Teichmann





Agenda

- Part 1: OS Brand Price Elasticity Project
 - Recap: Motivation, research question, and data
 - Updates to delivering your work
 - Q&A
- Part 2: OS in Marketing Research: Why (and Why Not) to Do It?
- Part 3: Discussion with Session Participants on OS Theses



Part 1

Meta-Scientific Research in Marketing: Understanding the Impact of Researcher Choices







Motivation (I)

- Progress in science stems from researcher decisions
 - theoretical framework or a set of hypotheses (if any),
 - operationalizations of constructs, sampling,
 - modeling and interpretation...
- Recent attention to key researcher decisions: choosing a pathway amongst the many analytical options (e.g., Huntington-Klein et al. 2021, Sarstedt et al. 2024).
- Potential heterogeneity in learnings derived from the same set of data (Menkveld et al. 2022) is particularly important since...
 - its size is unknown, and
 - limited reasoning of a (group of) researcher(s) for picking a specific analytical path (among many available)



Motivation (II)

- Analytical pathways often informed by the data
 - e.g., whether a variable has many zeros, is skewed, or is multicollinear
 - e.g., prevalence of missing data, prompting aggregation decisions
- Researcher choices are relevant to both the Empirics First (EF) approach (<u>Golder et al. 2023</u>) and the Theory First (TF) approach
- Necessity to motivate and document research choices precisely and comprehensively



This project

- Each of the 200+ participating research team (with max. 2 researchers) analyzes
 - the same household panel data set,
 - to estimate price elasticities for the same set of brands,
 - providing estimates, standard errors and *p*-values (or Bayesian equivalents),
 - along with the **code and survey answers** as to their research choices.
- Collaboration with
 - AiMark: non-profit institution promoting the use of consumer/household/scanner data in enhancing our understanding of marketing and its effectiveness since 1999
 - International Journal of Research in Marketing: publishes this multi-authored paper if it passes the review process

Participating teams

- 207 teams registered initially
- 196 teams completed the NDA with AiMark
 - 118 two-person teams
 - 78 one-person teams





Research Question (I)

What is the brand price elasticity (BPE*) for a preselected set of 68 brands in the product category of meat substitutes?

In this project, we are interested in estimating the EFFECT of a 1% brand price increase on brand volume sales, while keeping everything else constant.

* BPE is defined as the <u>percentage change in volume brand sales</u> (in weight) due to a <u>1% increase in</u> the brand's price.



Research Question (II)

- Ethical, health and environmental concerns regarding meat production and consumption
 - rapid increase of meat substitutes (Godfray et al., 2018).
 - expected growth: 8.7% annually within the next 5 years worldwide (<u>Statista, 2025</u>).
- Meat substitutes
 - meat-like products, sharing characteristics with 'real' meat (e.g., texture, taste, appearance, chemical compositions)
 - sold at comparably higher prices than meat → lowering prices as a strategic lever to encourage consumption aligned with environmental preservation and public health (<u>Jahn, Guhl, and Erhard, 2024</u>).
- Research has investigated price elasticities for meat substitutes; e.g., category price elasticities for meat substitutes (<u>Zhao et al., 2023</u>; Jahn et al. 2024).
- In our project, we are interested in **<u>brand</u>** price elasticities for meat substitutes.



Data

- Household panel data, provided by YouGov (formerly GfK Panel Services) and Kantar
 - contain a set of households representative of a country's population (Austria, Belgium, France, Germany, Great Britain, the Netherlands, Spain, and the United States)
 - households equipped with home scanners to scan all their grocery purchases at home (2014 – 2023*)
- Selected data contains purchases of households in the focal category ("meat substitutes")
 - Examples are: (all vegetarian!) burgers, sausages, gyros, etc.
- To keep the research approach parsimonious and the number of possible pathways limited, we refrained from adding more information about the panelists and their purchases in other categories.

* varies by country: Spain only 2017-2023, USA 2018-2020.

Submission process



Project Timeline and Submission Procedure





Upload your material by June 15, 2025

You should have **received an email around May 25**, containing a link to a survey and submission details.

- 1. Your elasticity estimates
- 2. Description of your model specification and variable definitions
- 3. Your replication package

Upload your material (I): Estimates

Template_Estimates.xlsx

	A	В	С	D	E	F	G	
1	Brand Number	Country	Brand Name	Brand Price Elasticity Estimate	Standard Error	Two-sided p-value	Comments (if any)	
2				defined as the percentage change in volume brand sales (in weight) due to a 1% increase in the brand' price.	of the brand price elasticity estimate s	testing the null hypothesis that the brand price elasticty equals 0		
3	1	Austria	JUST VEG!					
4	2	Austria	SPAR VEGGIE					
5	3	Austria	VEGAVITA					
6	4	Austria	HERMANN					
7	5	Austria	VEGINI					ip: If you generate your
8	6	Austria	GARDEN GOURMET				r	esults with a "nineline"
9	7	Austria	VEGANZ					
10	8	Austria	LANDHOF				t	hat automaticallywrites
11	9	Belgium	COLRUYT					our estimates to a CSV
12	10	Belgium	GARDEN GOURMET				ý	
13	11	Belgium	QUORN				fi	le, please STILL
14	12	Belgium	ALDI					manually) ontor your
15	13	Belgium					()	nanualiy) enter your
17	14						е	lasticity estimates here
10	15							
19	10						S	o_we nave a common
20	18	Belgium	IGLO				S	ubmission format to
21	19	Germany	RUEGENWALDER MU	EHLE			<i>p</i>	rocess in batch.

Upload your material (II): Model spec & var's

Template_Model.xlsx

,	ALL EQUATIONS (i.e., your model) of your model (best to use a screenshot from Word/Overleaf/PDE)	brand b and time t and potentially other subscripts. Specifical equations that are needed to represent the model. Please use the in-built equation editor or copy a screenshot with the model specification from
Please specifiy all variables f	rom your equation above (including subscripts for brai	nd b and time t and potentially other subscripts)
Variable Symbol (from equation)	Short Description	Precise Operationalization

Tip: If you want to include more extensive notes on your model (say, a "modeling note" you have written in Latex and saved as a PDF), you can include that note in your replication package (see next slides).

Enter your variables ("variable operationalization") in the table.

Upload your material (III): Replication package

- a) all **code** used to process the raw data and generate the final estimates with detailed comments, and
- b) A **README** with instructions on how to run your code, the order in which the scripts should be executed, and any required software packages.

We ask you to follow <u>https://datacodestandard.org/</u>. → See our explanations on the next slides.

Data and Code Availability Standard

https://datacodestandard.org/ (I)

	Data		
1	Data Availability Statement	A Data Availability Statement is provided with detailed enough information such that an independent researcher can replicate the steps needed to access the original data, including any limitations and the expected monetary and time cost of data access.	Only rec
2	Raw data	Raw data used in the research (primary data collected by the author and secondary data not otherwise available) is made publicly accessible. Exceptions are explained under Rule 1.	Please r this will
3	Analysis data	Analysis data is provided as part of the replication package unless they can be fully reproduced from accessible data within a reasonable time frame. Exceptions are explained under Rule 1.	Try to ke file size fully rep
4	Format	The data files are provided in any format compatible with commonly used statistical package or software. Some journals require data files in open, non-proprietary formats.	We pref what yo
5	Metadata	Description of variables and their allowed values are publicly accessible.	Please ii variable
6	Citation	All data used in the paper are cited.	Only rec

Only required for **auxiliary data**, <u>not</u> for the AiMark data.

Please **re-include the AiMark data** as well – this will allow us to easily run your workflow.

Try to keep any temporary files generated "on the way", unless there are file size issues. You can exclude temporary files only if **they can be fully reproduced with your code and the data** contained in your package.

We prefer **CSV**. But the principal rule is: supply what you used in your code; do not "convert" it for us at the end.

Please include **variables** mentioned earlier, and any additional variables (e.g., from auxilary data) you have used.

Only required for auxilary data added by you.

https://datacodestandard.org/ (II)

	Code		<i>Tip:</i> You can also submit code that is not central to
7	Data transformation	Programs used to create any final and analysis data sets from raw data are included.	generating your final results, but that you have used to explore the data
8	Analysis	Programs producing the computational results (estimation, simulation, model solution, visualization) are included.	(e.g., to create plots). You can describe these files in the README (see next
9	Format	Code is provided in source format that can be directly interpreted or compiled by appropriate software.	snaes).

e.g., R, .py, SPSS syntax, etc.

https://datacodestandard.org/ (III)

	Supporting materials		
10	Instruments	If collecting original data through surveys or experiments, survey instruments or experiment instructions as well as details on subject selection are included.	Optional
11	Ethics	If applicable, details are shared about ethics approval.	Optional
12	Pre-registration	If applicable, pre-registration of the research is identified and cited.	Optional
13	Documentation	A README document is included, containing a Data Availability Statement, listing all software and hardware dependencies and requirements (including the expected run time), and explaining how to reproduce the research results. The README follows the schema provided by the Social Science Data Editors' template README.	See next slide

Your readme

DOI 10.5281/zenodo.7293838



A template README for social science replication packages.

The template README provided on this website is in a form that follows best practices as defined by a number of data editors at social science journals.

Authors: Lars Vilhuber, Miklós Koren, Joan Llull, Marie Connolly, Peter Morrow

This project is maintained at socialscience-data-editors/template_README This page displays the latest release-candidate. For official releases, see the Releases page or the above DOI.

A template README for social science replication packages

The template README provided on this website is in a form that follows best practices as defined by a number of data editors at social science journals. A full list of endorsers is listed in Endorsers.

Versions

The most recent version is available at https://social-science-dataeditors.github.io/template_README/. Specific releases can be found at https://github.com/social-science-data-editors/template_README/releases.

Formats

The template README is available in a variety of formats:

Release candidates (which may be the latest version)

- HTML (best for reading)
- LaTeX
- Word
- PDF
- Markdown

Template README and Guidance

INSTRUCTIONS: This README suggests structure and content that have been approved by various journals, see Endorsers. It is available as Markdown/txt, Word, LaTeX, and PDF. In practice, there are many variations and complications, and authors should feel free to adapt to their needs. All instructions can (should) be removed from the final README (in Markdown, remove lines starting with > INSTRUCTIONS). Please ensure that a PDF is submitted in addition to the chosen native format. Please ensure that the README is called "README" plus the appropriate suffix, not some non-standard name. This helps replicators immediately locate the necessary document.

Overview

INSTRUCTIONS: The typical README in social science journals <u>serves the</u> <u>purpose of guiding</u> a reader through the available material and a route to replicating the results in the research paper. Start by providing a brief overview of the available material and a brief guide as to how to proceed from beginning to end.

Please <u>modify</u> the template as you see fit. We will now go through the README together.

→ Download at <u>https://social-science-data-</u>editors.github.io/template_README.

Submitting your package - UPDATE

- Via survey link, sent around May 25, 2025 to all participating teams
 - 50 MB upload limit on Qualtrics
 - including your elasticities (template_estimates.xlsx), the modeling notes (templates_model.xlsx), and your zipped replication package.
 - Does your submission exceed the file size limit?
 - Always upload to Qualtrics...
 - Your elasticities (templates_estimates.xlsx) and modeling notes (templates_model.xlsx)
 - Your replication package *without* any data (but *with* readme & code)
 - Then...
 - Contact us at <u>info@elasticity-open-science.com</u>, and you will receive a special submission link for large replication packages (anything between 50 MB and a few GBs).

Q&A on the big team science project

Part 2

Why (and why not) to do open science?!

Goal of OS practices

Increase transparency, accountability, equity and collaboration, and knowledge production by increasing access to research results, methods, and tools.

Ross-Hellauer, 2022, p. 363

Significant investment in academic research

2% of GDP

in the European Union

Accessibility to research







Source: https://retractionwatch.com/2021/12/30/2021-a-review-of-the-years-3200-retractions/

But... perceived risk of (data) sharing



Source: Vilhuber, Lars (2020). "Reproducibility and replicability in economics." Harvard Data Science Review.

Learning costs

coding documentation managing teams time





Doing OS isn't that hard



Deer, L., Adler, S. J., Datta, H., Mizik, N., & Sarstedt, M. (2025). Toward open science in marketing research. *International Journal of Research in Marketing*, 42(1), 212-233.

Preregister hypotheses



Make data accessible/searchable

	Missed our webin	ar with tips &	tricks fo	or researd	chers? <u>Wa</u>	itch it n	ow on You	ITube!				×
web-scraping.org Da	tabases 🗸 🛛 Tutorial	s∨ Comn	nunity ∽	Othe	r Resou	irces	Contac	t				
 DATABASES Database of published paper Web datasets API directory TUTORIALS RESOURCES 	s Discov Journal: Type: Year: 313 r	rabas er publishe JM JMR Web scraping 2001 2002 2012 2013 results	C O d pape JCR API 2003 2014	f pu rs in ma JCP M Both web 2004 2015	arketin ktSci 2005 2016	g using and API 2006 2017	ed g web / Manu 2007 2018	pap data. al extraction 2008 2019	ers 2009 2020	2010	2011	
	Schlosser Can incl increase persuas reviews ² of rating Published Data colle	, AE (2011) uding pros ar the helpfulne iveness of onl The interact and argume l in JCP ection: Web scra	nd cons ess and ine ive effect ents	Huang Polma Effec psych const evalu onlin Publis Data c	g, N; Buri n, E (201 nologica trual and trual	ich, G; Ho 6) Iltiple Il distanc d consum field sta vs P : Both web	ces on ner udy of	Henk Hoeg (2011 Disc Con Less Publi 9 Data	xel, AP; B gg, J; Aqu sounting sumers scious, Humar ished in J collection	ioegersha uino, K; Le Human are Price Employe CP n: Web sc	usen, J; emmink, J ity: When e ees Appea	r

and API

Source: https://web-scraping.org

Update paper, after publication

A Primer on Bayesian Inference for Accounting Research

Home Paper Code 🗸 Additional Material 🗸 🖓

A Primer on Bayesian Inference for Accounting Research

Welcome to the companion website for the paper *What Can Bayesian Inference Do for Accounting Research?* I hope this site and the accompanying <u>GitHub repo</u> will help those interested in Bayesian methods to find a starting point. Hopefully you find it useful.

AUTHOR Harm Schuett PUBLISHED Oct. 4, 2021

AFFILIATION

Tilburg University

Source: https://hschuett.github.io/BayesForAccountingResearch/

Make code reusable

Ξ readme.md

marketingtools

The goal of marketingtools is to collect and make accessible code for conducting empirical (marketing) research.

Features

- Computation of Gaussian Copula correction terms, using their ECDF (Park & Gupta 2012) (make_copula)
- Estimation of Feasible Generalized Least Squares (FGLS) with unbalanced panels (itersur)
- Estimation of market share attraction models (Fok 2001) o Transformation of wide panels into their basebrand representation o Support for homogenous and heterogenous coefficients o Restricted competition; fully extended model not implemented o Estimation via itersur() (FGLS)

Installation

You can install the released version of musiclabels from GitHub with:

install.packages("devtools")
devtools::install_github("hannesdatta/marketingtools")

hannesdatta Hannes Datta



Languages

0

• R 100.0%

Make your work interactive and transparent

Comparison between Specifications

Please pick a specification:	Choose to observe Read	sults Summary or List		
lorder-Strategy	Summary	•		
Adjust the default specification:	Section 1: Co	efficient Distr	ibution	
Data Level Time Controls	Minimum		Maximum	
Store Border-Month FE	-0.7	•	0.7	•
Norkat Eirod Effect		Dis	stribution (0 Outliers Excluded)	
Yes Yes	0.6			
eature-Display Control Price Controls	0.5		0.0136	

ROI Analyses Distribution of Best Delta Residual Variation Semi-Parametric Estimation

Source: https://advertising-effects.chicagobooth.edu/

Overview

All Regression Results

Teach open science by practicing open science

.

	Team Project	Search Q Upload Communities
Online Data Collections (oDCM)	Overview and proj	June 14, 2022 Dataset Open Acces
Search	Collecting data via web scrapi members, you plan and execu	Web scraping and API projects from "Online Data Collection and Management" (Spring
Course +	closely following the recomm	2022)
Schedule and Course Material	documentation.	 Datta, Hannes As part of "Online Data Collection and Management" (taught at Tilburg University, Spring 2022), students collected publicl
Tutorials •	The focus lies on <i>completir</i> of your project <i>manageable</i>	available datasets for use in academic research projects. With this repository, I am sharing (a) the documentation of thes data sets, and (b) the associated source code that led to the collection of the data. The repository also contains the collected datasets.
Team Project	up as a proper data docum	The data consists of the following projects:
Workplan and coaching	template download. Head	Autoscout (electric cars vs gasoline cars in the Dutch market) Mediamarkt (e-commarce)
Grading		Steam API
More resources	Getting started	 Twitch (chat capture) Zalando (e-commerce)
Exam 🕨	Workplan, deliverables and coa	Course website: https://odcm.hannesdatta.com. Archived at https://doi.org/10.5281/zenodo.6641811 (check for more recent versions if available).
About	Tips and examples	Sources: https://zenodo.org/record/6641811 and https://odcm.hannesdatta.cor

Getting started requires retraining...

- Way of working
 - Larger, more diverse teams
 - Project management
 - Built-in transparency
- Platforms
 - For coding: e.g., GitHub
 - For data: e.g., Zenodo, Dataverse
 - For preregistration: e.g., aspredicted.org
- Coding
 - Coding quality
 - Documentation
 - Automation
 - Portability

Discussion

"Journals should publish more studies with imperfect findings."

"IJRM should require authors to submit replication packages for all empirical papers."

"Open Science Increases Heterogeneity in Findings, which undermines perceived trust in science among the public." "Pre-registration does not mitigate, but amplifies the file-drawer problem."

Discussion

- Evaluation Open Science: Cost vs. Benefits
 - Can you identify any significant costs or challenges?
 - Do the benefits of open science outweigh its costs, in your experience?
- Overcoming Drawbacks of Open Science
 - Have you encountered drawbacks or challenges yourself?
 - What strategies could be employed to mitigate these drawbacks?
 - Share an example where overcoming a challenge led to a positive outcome.



Thanks!

For more information, consult our website (<u>https://www.elasticity-open-science.com</u>) or email us (<u>info@elasticity-open-science.com</u>).



Koen Pauwels

Northeastern University

Boston, USA

Editor-in-Chief, IJRM



Karin Teichmann

University of Innsbruck Austria

Big team & open science



Oliver Koll

University of Innsbruck Austria

> AiMark Scientific Board, panel data



Hannes Datta

Tilburg University Netherlands

Reproducible & open science, marketing analytics



Harald van Heerde

University of New South Wales Sydney, Australia

Vice-chair, AiMark

References

- Golder, Peter N., Marnik G. Dekimpe, Jake T. An, Harald J. van Heerde, Darren Kim, and Joseph W. Alba (2023), "Learning from Data: An Empirics-First Approach to Knowledge Generation in Marketing," Journal of Marketing, 87 (3), 319–336
- Godfray, H. C. J., Aveyard, P., Garnett, T., Hall, J. W., Key, T. J., Lorimer, J., Pierrehumbert, R. T., Scarborough, P., Springmann, M., and Jebb, S. A. (2018). Meat consumption, health, and the environment. Science, 361(6399):eaam5324
- Huntington-Klein, N., Arenas, A., Beam, E., Bertoni, M., Bloem, J. R., Burli, P., ... & Stopnitzky, Y. (2021). The influence of hidden researcher decisions in applied microeconomics. Economic Inquiry, 59(3), 944-960.
- Jahn, S., Guhl, D., and Erhard, A. (2024). Substitution patterns and price response for plant-based meat alternatives. Proc. Natl. Acad. Sci. U.S.A., 121 (50) e2319016121, https://doi.org/10.1073/pnas.2319016121.
- Menkveld, Albert J. et al. Non-Standard Errors (May 31, 2023). Journal of Finance, Forthcoming, https://dx.doi.org/10.2139/ssrn.3961574
- Sarstedt, M., Adler, S. J., Ringle, C. M., Cho, G., Diamantopoulos, A., Hwang, H., & Liengaard, B. D. (2024). Same model, same data, but different outcomes: Evaluating the impact of method choices in structural equation modeling. Journal of Product Innovation Management, forthcoming.
- Zhao, S., Wang, L., Hu, W., & Zheng, Y. (2023). Meet the Meatless: Demand for New Generation Plant-Based Meat Alternatives. Applied economic perspectives and policy, 45(1), 4–21. https://doi.org/10.1002/aepp.13232