Surveys or digital trace data, which one should we use?

Using MultiTrait-MultiMethod models to simultaneously estimate the measurement quality of surveys and digital trace data.

Oriol J. Bosch| Department of Methodology, LSE & RECSM

Melanie Revilla |Institut Estudis Barcelona Internacional (IBEI)

Patrick Sturgis| Department of Methodology, LSE

Jouni Kuha| Department of Methodology, LSE

o.bosch-jover@lse.ac.uk

orioljbosch

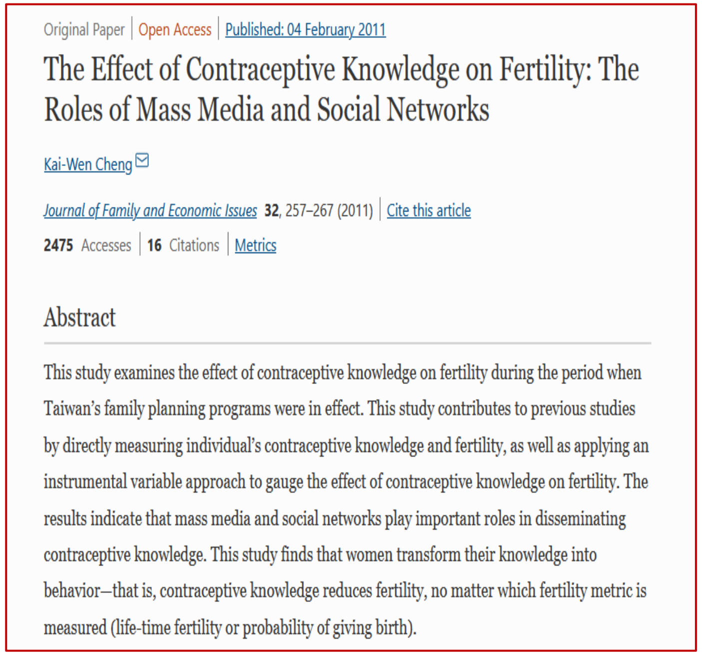
https://orioljbosch.com/

Acknowledgements: I would like to thank Melanie Revilla, Mariano Torcal, Patrick Sturgis and Jouni Kuha

Funding: This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No 849165; PI: Melanie Revilla); the Spanish Ministry of Science and Innovation under the "R+D+i projects" programme (grant number PID2019-106867RB-I00 /AEI/10.13039/501100011033 (2020-2024), PI: Mariano Torcal); and the BBVA foundation under their grant scheme to scientific research teams in economy and digital society, 2019 (PI: Mariano Torcal).

The importance of measuring what people do online

- Increased importance of understanding **the extent** and **the type of media/content people are exposed** to
- As well as its **effect** on how people **think, feel,** and **behave**


Original Paper | Open Access | Published: 04 February 2011

The Effect of Contraceptive Knowledge on Fertility: The Roles of Mass Media and Social Networks

Kai-Wen Cheng

Journal of Family and Economic Issues 32, 257–267 (2011) | Cite this article

2475 Accesses | 16 Citations | Metrics

Abstract

This study examines the effect of contraceptive knowledge on fertility during the period when Taiwan's family planning programs were in effect. This study contributes to previous studies by directly measuring individual's contraceptive knowledge and fertility, as well as applying an instrumental variable approach to gauge the effect of contraceptive knowledge on fertility. The results indicate that mass media and social networks play important roles in disseminating contraceptive knowledge. This study finds that women transform their knowledge into behavior—that is, contraceptive knowledge reduces fertility, no matter which fertility metric is measured (life-time fertility or probability of giving birth).

Digital trace data to understand online behaviours

- Survey self-reports are still the **most common approach**

The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure

Get access ›

Markus Prior

Public Opinion Quarterly, Volume 73, Issue 1, Spring 2009, Pages 130–143, https://doi.org/10.1093/poq/nfp002

Published: 18 March 2009

Cite Permissions Share ▾

Abstract

Many studies of media effects use self-reported news exposure as their key independent variable without establishing its validity. Motivated by anecdotal evidence that people's reports of their own media use can differ considerably from independent assessments, this study examines systematically the accuracy of survey-based self-reports of news exposure. I compare survey estimates to Nielsen estimates, which do not rely on self-reports. Results show severe overreporting of news exposure. Survey estimates of network news exposure follow trends in Nielsen ratings relatively well, but exaggerate

Digital trace data to understand online behaviours

- Survey self-reports are still the **most common approach**

- More and more availability of **digital traces to directly observe media exposure**

Individual-level approach: web trackers

Direct observations of online behaviours using tracking solutions, or *meters.*

Group of tracking technologies (plug-ins, apps, proxies, etc)

Installed on participants devices

Collect traces left by participants when **interacting with their devices online: URLs, apps visited, cookies…**

Great, we will get unbiased measures!

Is web tracking data actually unbiased?

Little but growing evidence that **web tracking data is affected by errors**

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistics in Society
Series A

ORIGINAL ARTICLE | Open Access

When survey science met web tracking: Presenting an error framework for metered data

Oriol J. Bosch, Melanie Revilla

First published: 06 November 2022 | https://doi.org/10.1111/rssa.12956

Funding information: Fundación BBVA, H2020 European Research Council, , Grant/Award Number: 849165; Ministerio de Ciencia e Innovación, , Grant/Award Number: PID2019-106867RB- I00 /AEI/10.13039 /501100011033

SECTIONS PDF TOOLS SHARE

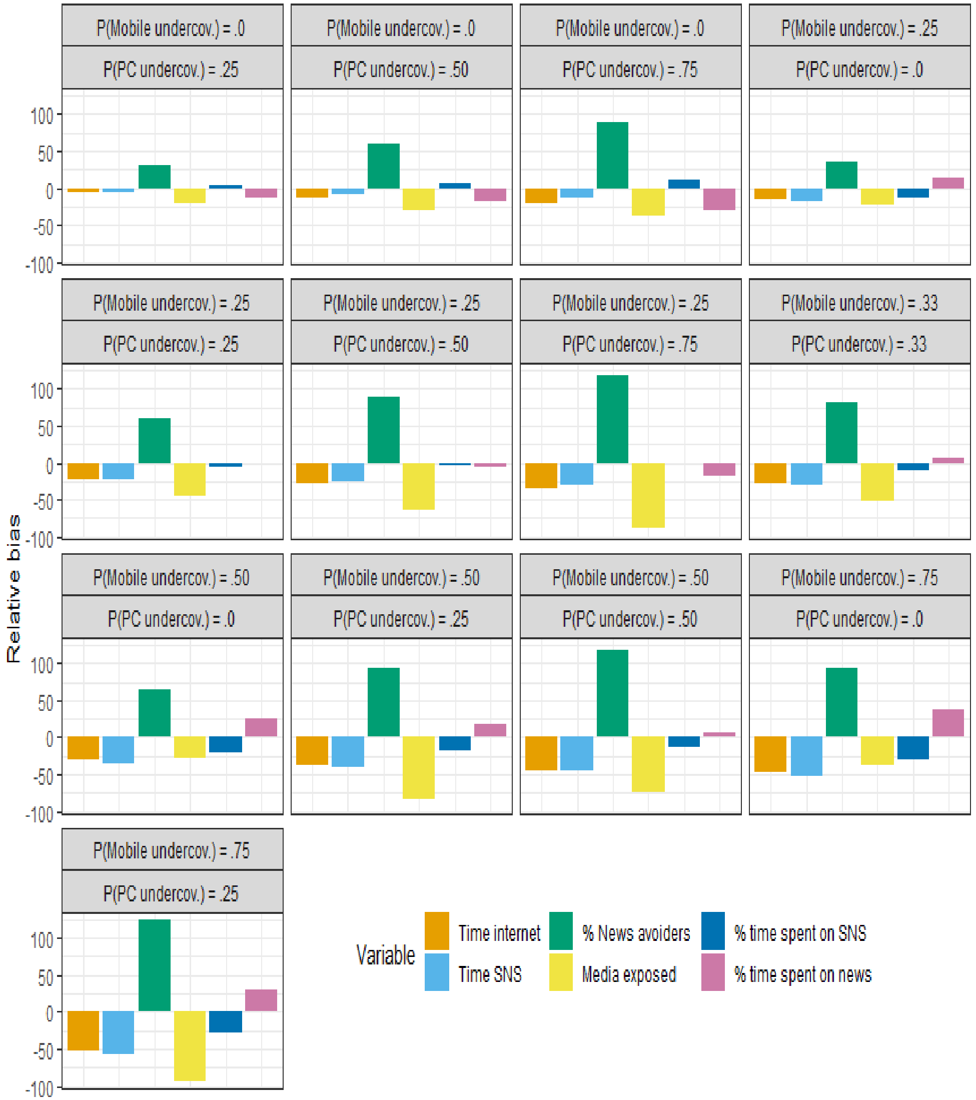
Abstract

Metered data, also called web-tracking data, are generally collected from a sample of participants who willingly install or configure, onto their devices, technologies that track digital traces left when people go online (e.g., URLs visited). Since metered data allow for the observation of online behaviours unobtrusively, it has been proposed as a useful tool to understand what people do online and what impacts this might have on online and offline phenomena. It is crucial, nevertheless, to understand its limitations. Although some research have explored the potential errors of metered data, a systematic categorisation and conceptualisation of these errors are missing. Inspired by the Total Survey Error, we present a Total Error framework for digital traces collected with Meters (TEM). The TEM framework (1) describes the data generation and the analysis process for metered data and (2) documents the sources of bias and variance that may arise in each step of this process. Using a case study we also show how the TEM can be applied in real life to identify, quantify and reduce metered data errors. Results suggest that metered data might indeed be affected by the error sources identified in our framework and, to some extent, biased. This framework can help improve the quality of both stand-alone metered data research projects, as well as foster the understanding of how and when survey and metered data can be combined.

Is web tracking data actually unbiased?

Little but growing evidence that **web tracking data is affected by errors**

We know that these errors can introduce measurement errors of a considerable size

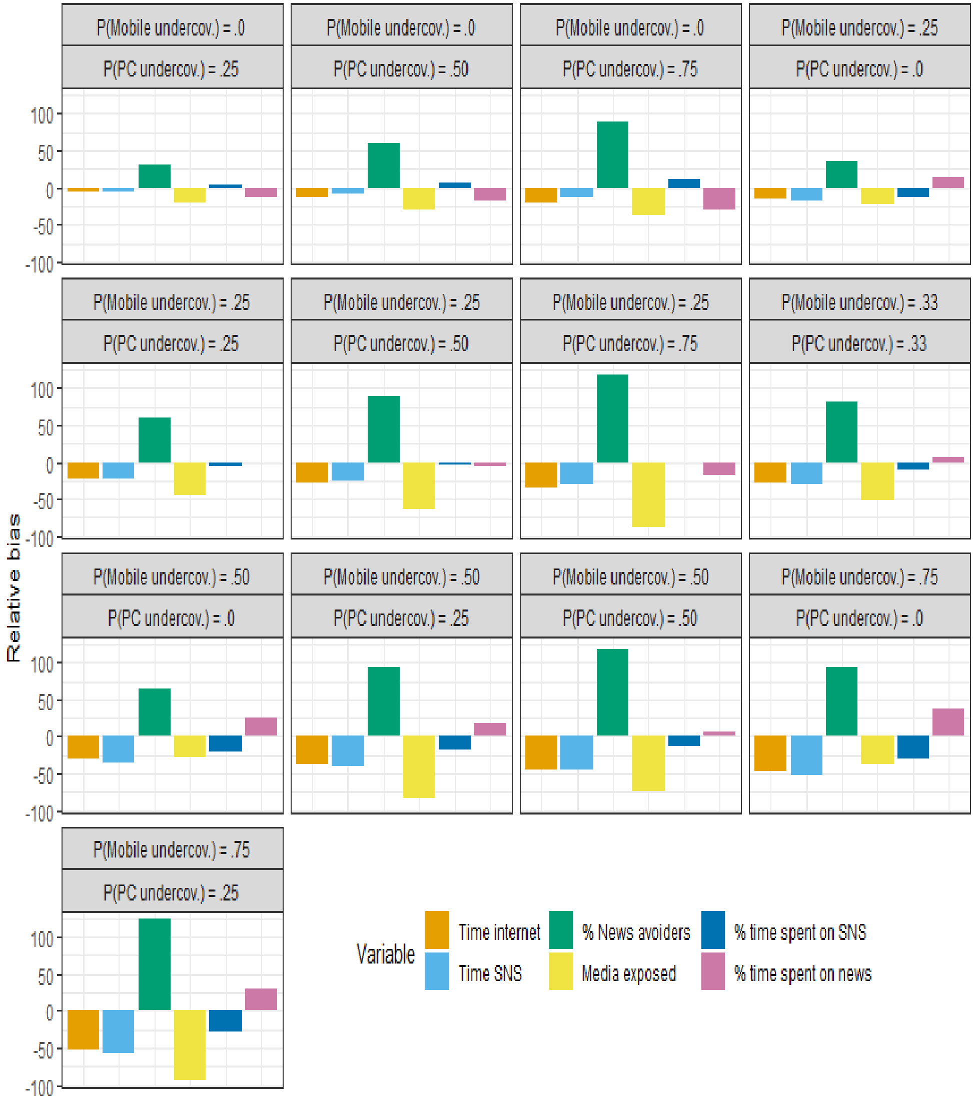


Is web tracking data actually unbiased?

Little but growing evidence that **web tracking data is affected by errors**

We know that these errors can introduce measurement errors of a considerable size

But still not near what we know about surveys!



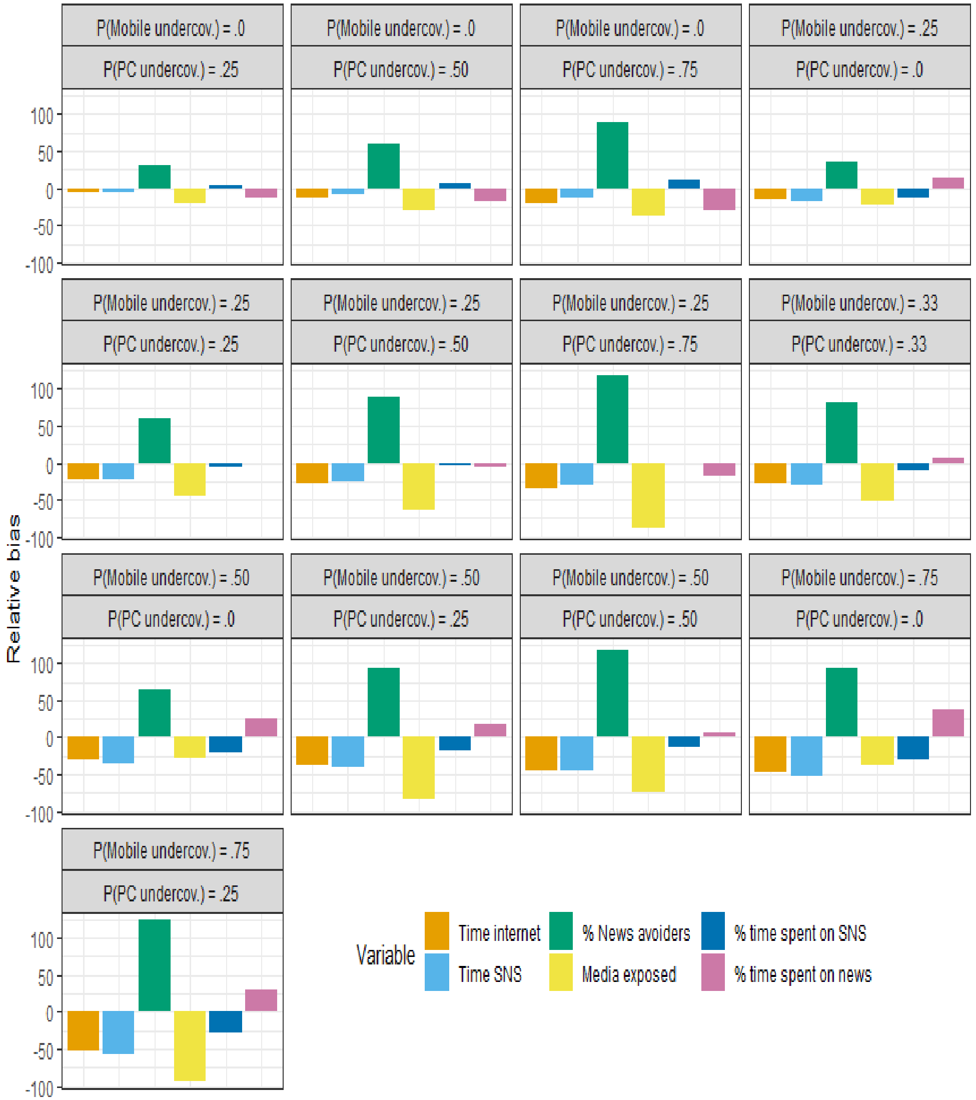
Is web tracking data actually unbiased?

Little but growing evidence that **web tracking data is affected by errors**

We know that these errors can introduce measurement errors of a considerable size

But still not near what we know about surveys!

My pitch: adapt decades of knowledge in psychometrics and survey methodology to improve how we use digital trace data



Simultaneously estimating the measurement quality of digital trace data and surveys using MultiTrait-MultiMethod (MTMM) models

Measurement quality

Quality = part of variance explained by the latent concept of interest
→ complement of measurement errors

t = true score for 11 point scale

Reliability coefficient

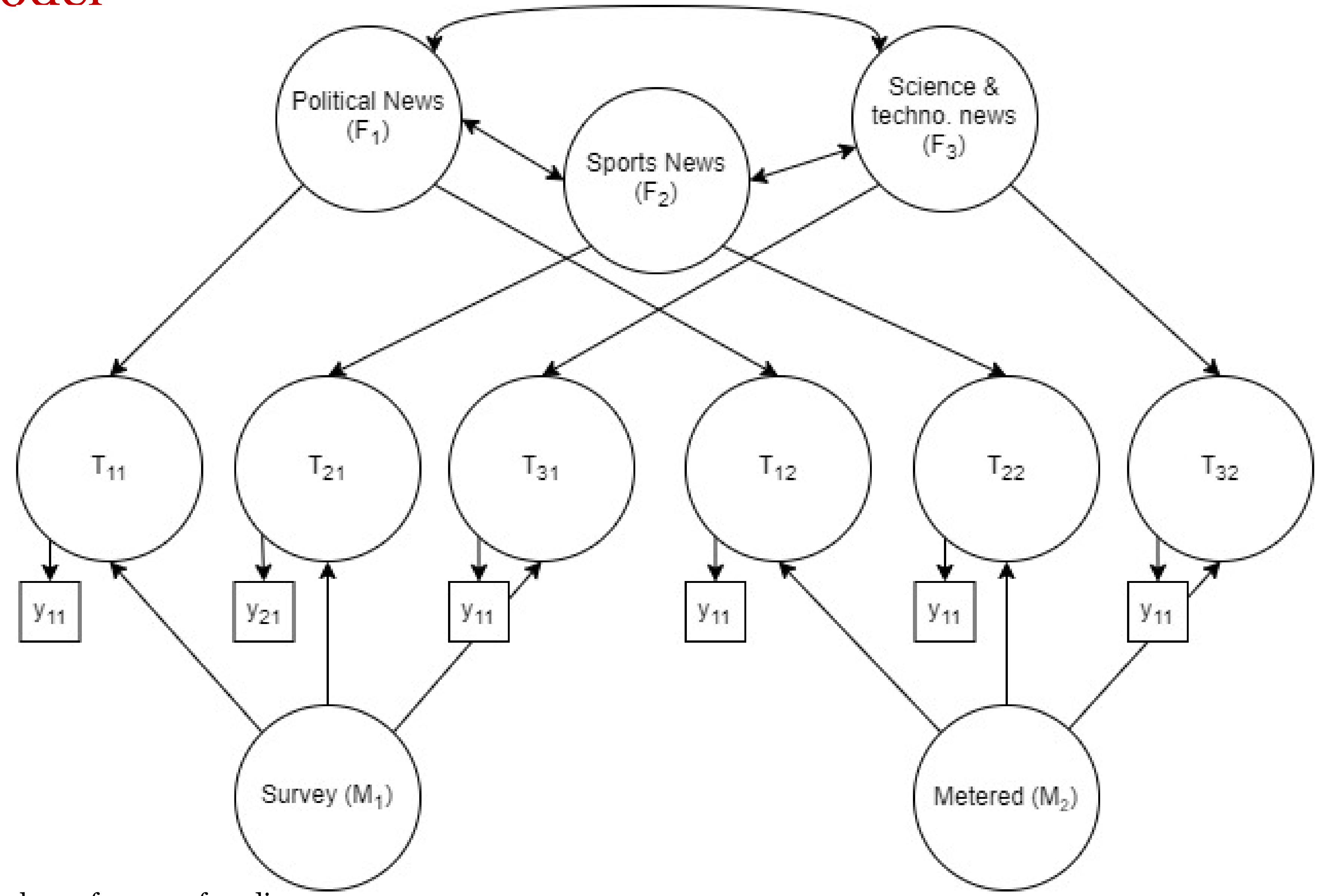
y = observed response 11 point scale

e = random error 11 point scale

Quality = reliability *x* measurement validity

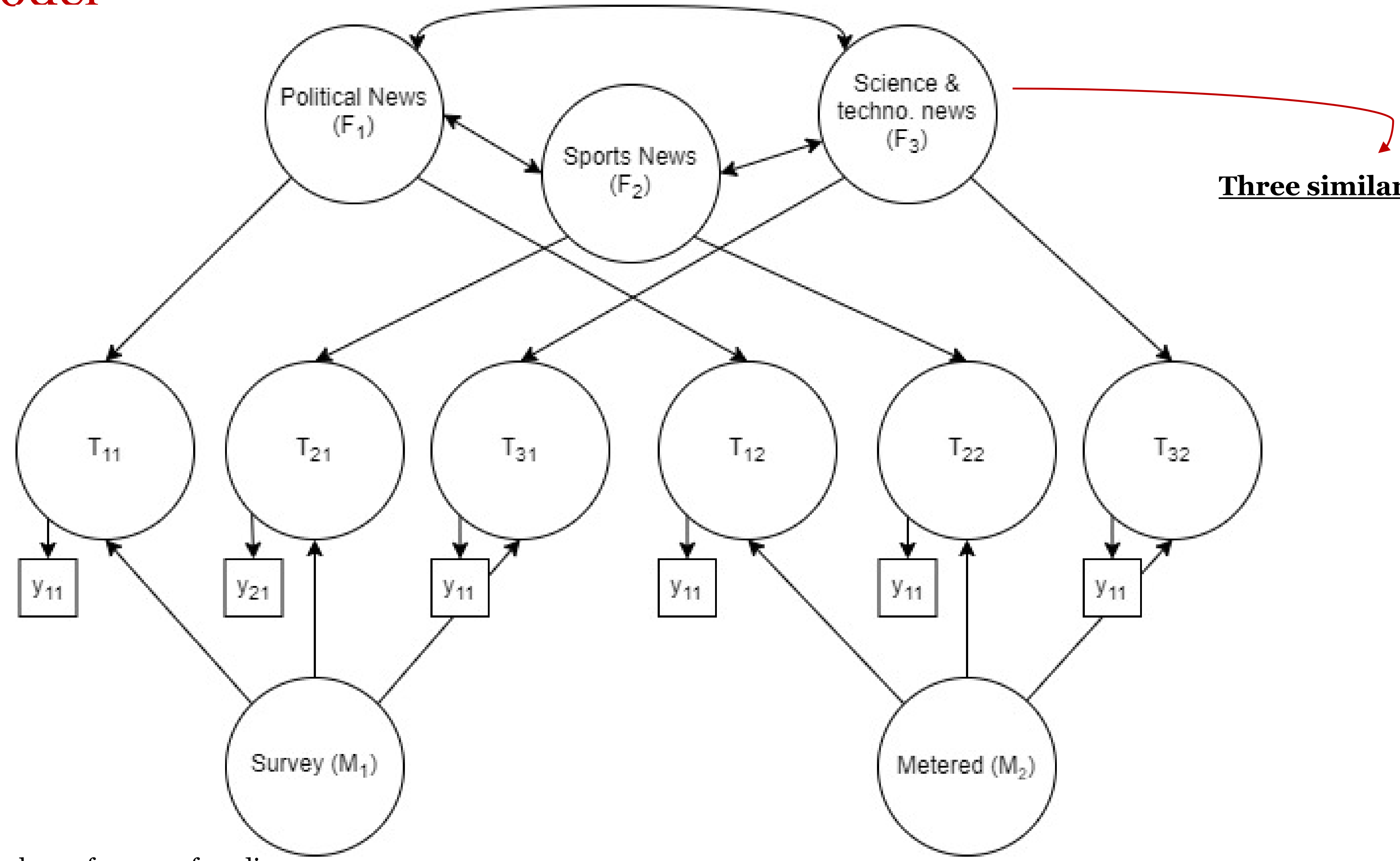
Quality = strength of the relationship between the latent concept of interest and the observed answers

The model



Residuals are not shown for ease of reading

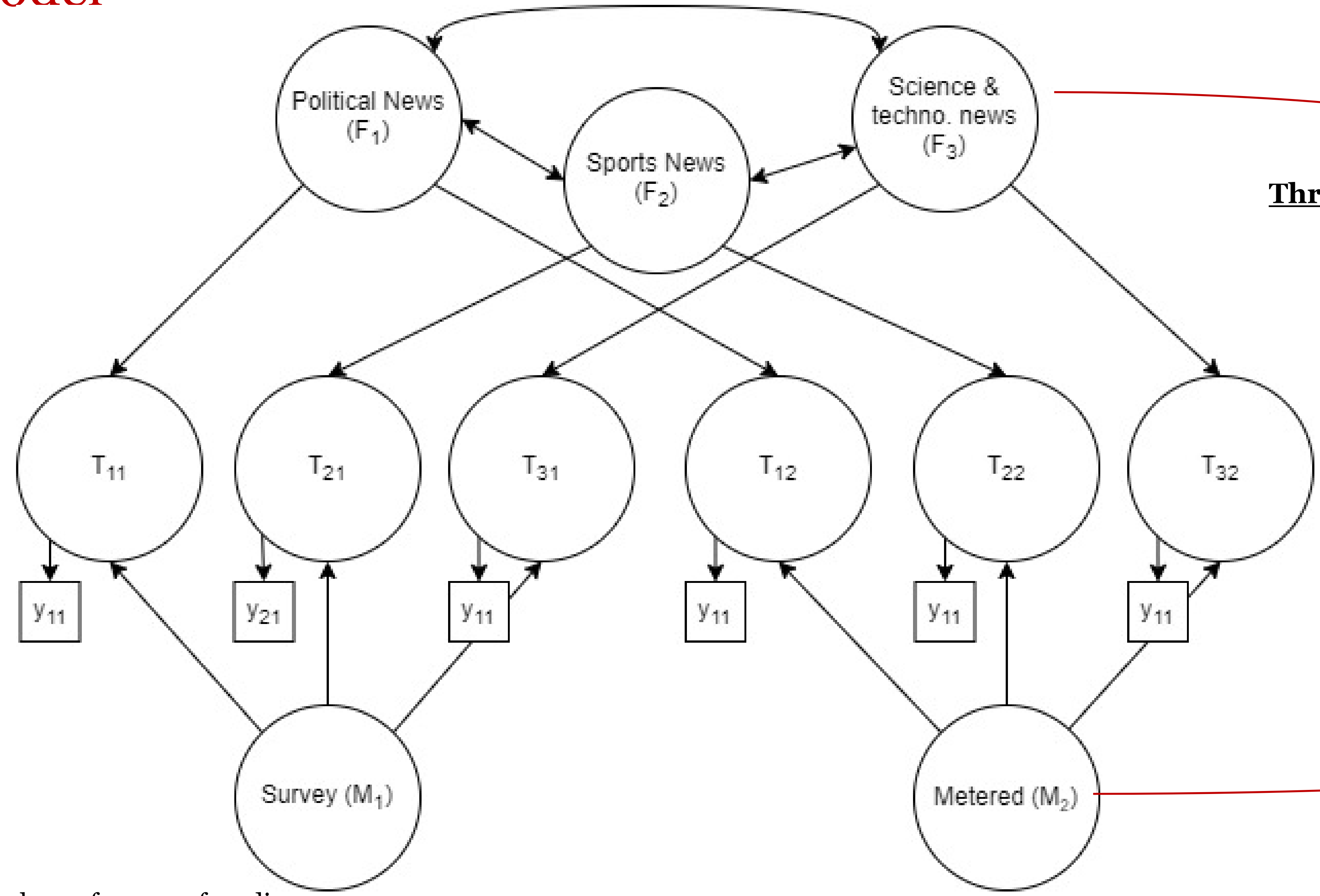
The model



Three similar traits

Residuals are not shown for ease of reading

The model

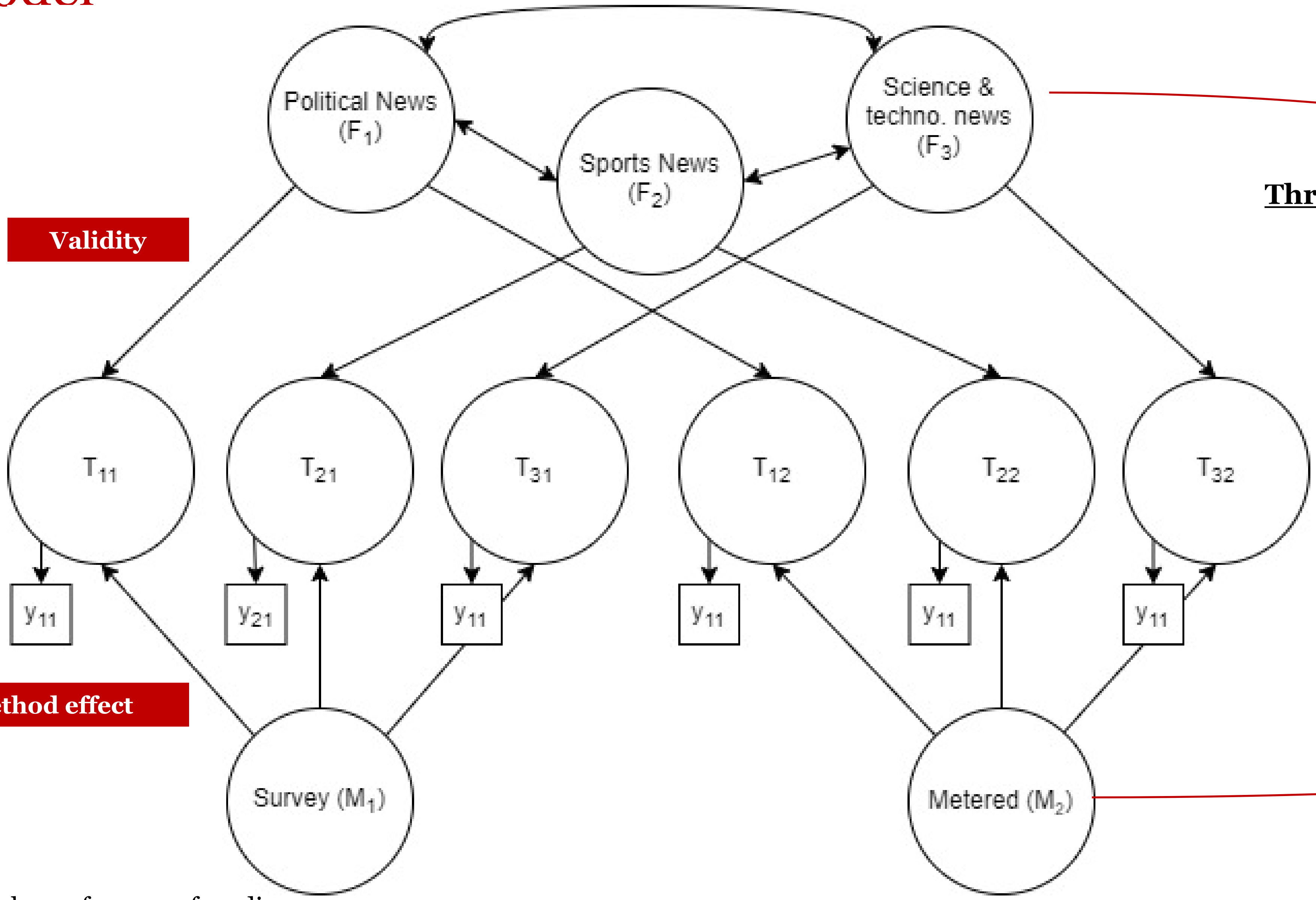


Three similar traits

Two methods to measure these traits

Residuals are not shown for ease of reading

The model



Validity

Reliability

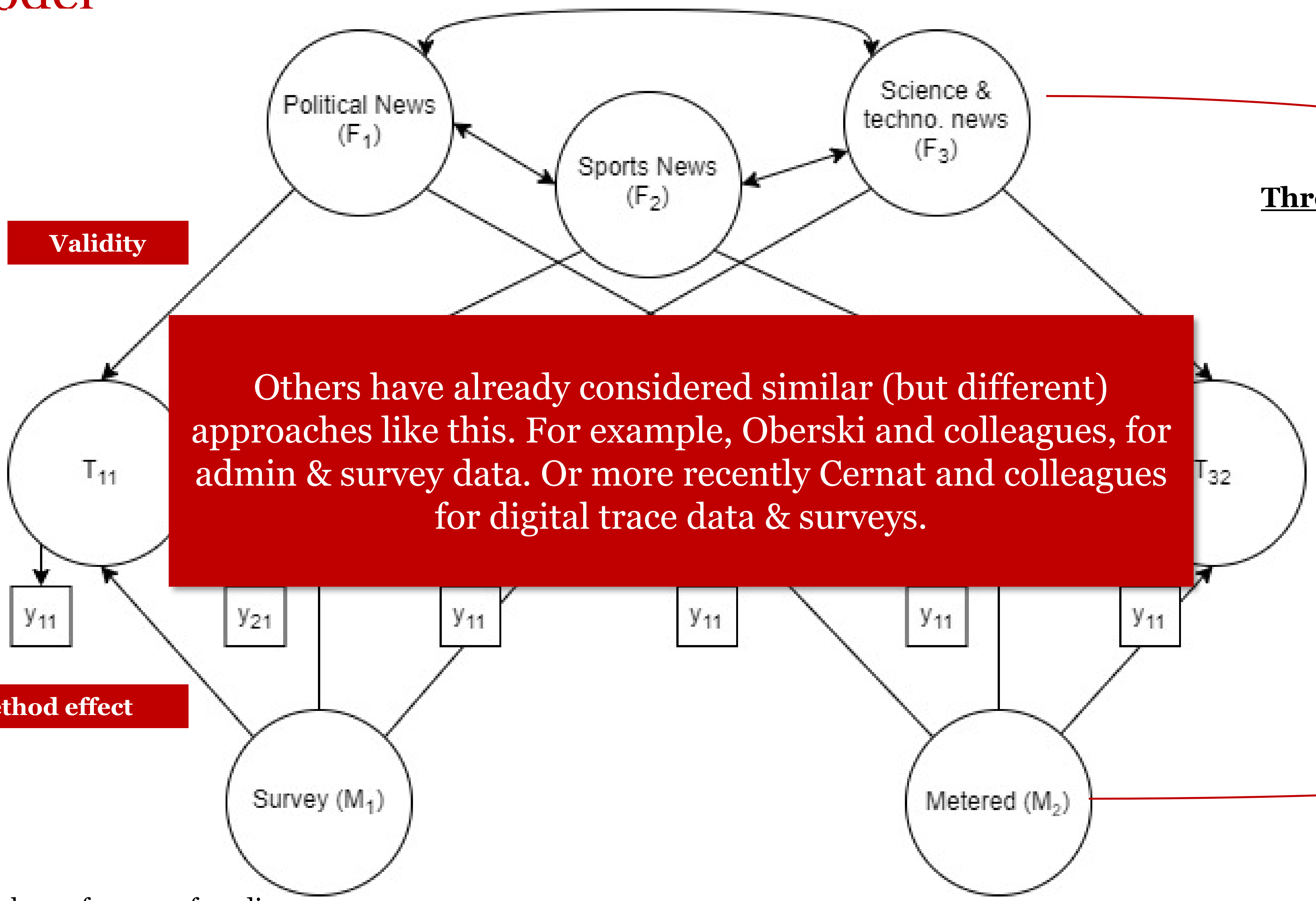
Method effect

Three similar traits

Two methods to measure these traits

Residuals are not shown for ease of reading

The model



Three similar traits

Reliability

Method effect

Two methods to measure these traits

Others have already considered similar (but different) approaches like this. For example, Oberski and colleagues, for admin & survey data. Or more recently Cernat and colleagues for digital trace data & surveys.

Residuals are not shown for ease of reading

This study

Research questions

Research questions

What is the overall validity, reliability, method effect and measurement quality of several measurements computed with digital trace data? (**RQ.1**)

And how do these compare with the quality estimates from equivalent survey questions? (**RQ.2**)

Data

web data opp

- **Survey** combined with **web tracking data** at the individual level

- Netquest metered panel in Spain
  - **Cross-quotas:** gender, age, and education
  - **Sample size:** 1,200
  - **Fieldwork:** Late May – Early June 2023

- Tracking technologies installed in both **mobile and desktop devices**

- Part of the ERC project **WEB DATA OPP**

Three differ groups of traits of interest

1. News exposure traits

- Exposure to news about politics
- Exposure to news about sports
- Exposure to news about science and technology

2. Communication traits:

- Use of social media
- Use of instant messaging
- Use of e-mails

3. Entertainment traits:

- Use of video platforms (YouTube, Vimeo, Twitch)
- Use of audio streaming (Spotify, Audible, Apple podcast)
- Use of TV/Movie streaming (Netflix, BBC online)

The measurements

1. Survey questions

More specifically, on average, how much time per day do you spend on the Internet reading news and articles...

- **MC4_1.** ... about politics and current affairs?
- **MC4_1_HH.** Hours: *[SMALL NUMERIC OPEN BOX]* **MC4_1_MM.** Minutes: *[SMALL NUMERIC OPEN]*

The measurements

1. **Survey questions**

2. **Web tracking data**

Characteristics	My choices
Metric	Minutes
List of traces	
List of media	Tranco
Top media	All
Information	Those identified as specific concept
Exposure	
Time threshold	1 second
Devices	All devices (with or without app)
Tracking period	31 days

The measurements

1. **Survey questions**

2. **Web tracking data**

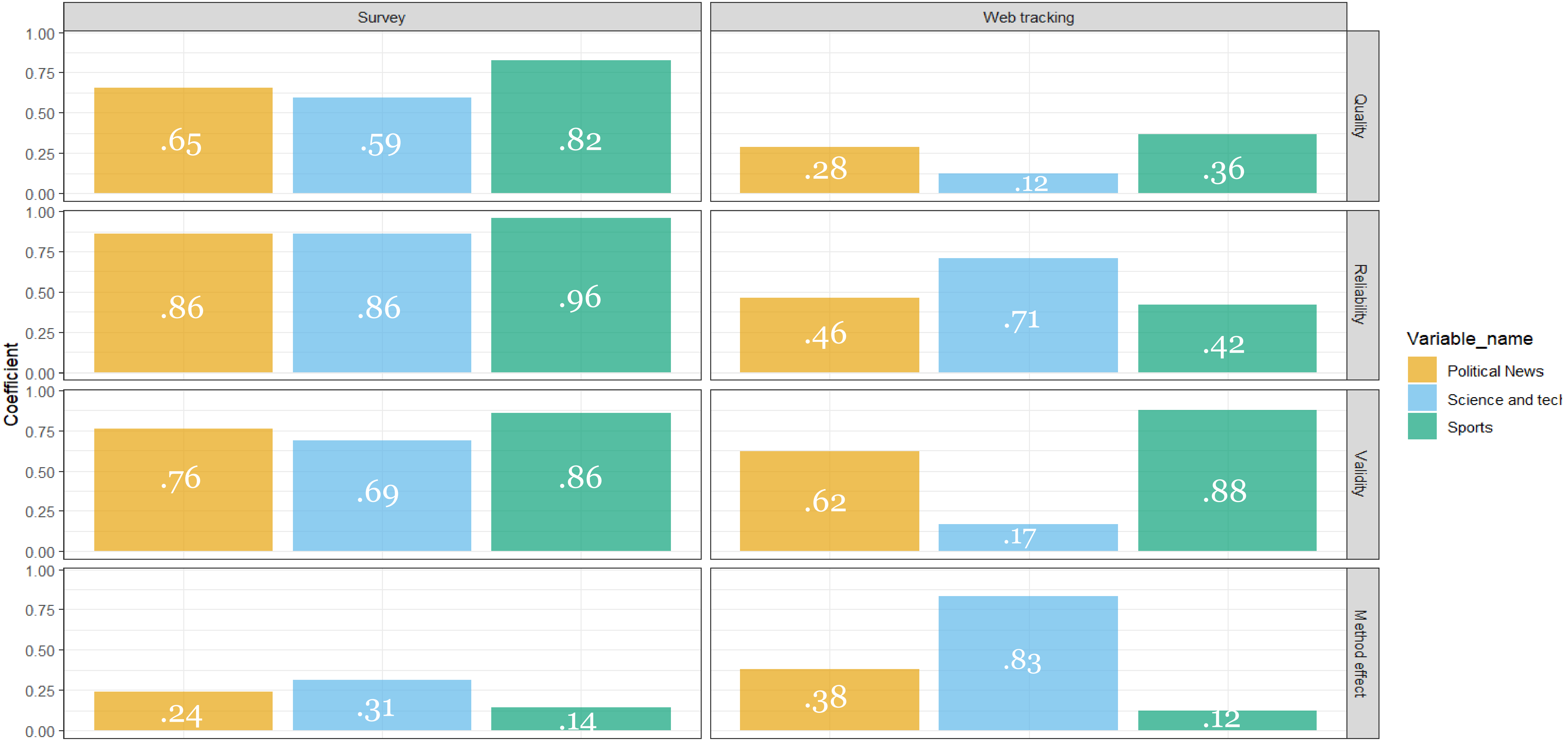
I use the log of these measures

Results

Correlation matrix

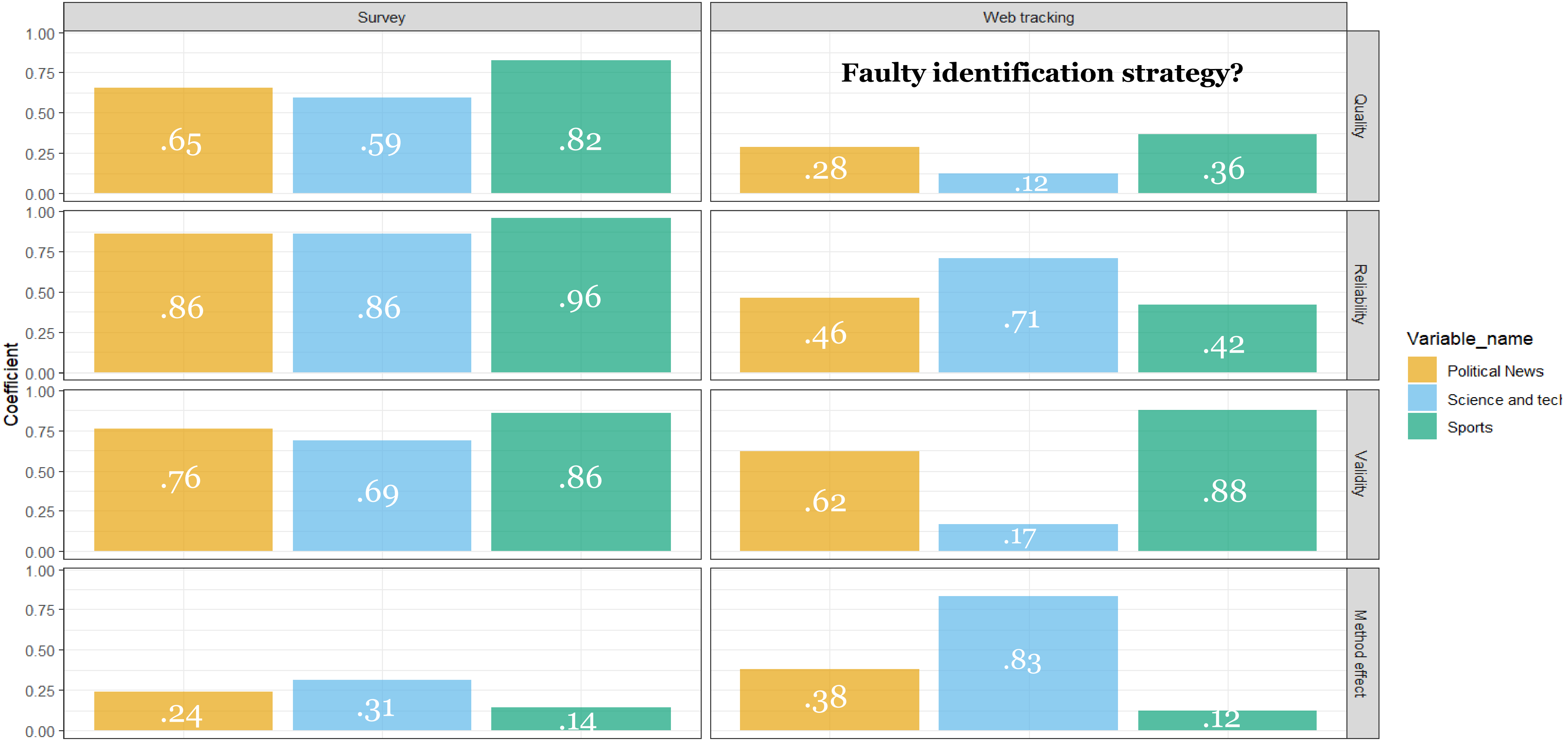
Web tracking:	Politics	Sports	Science	Social Media	Messages	Emails	Videos	Audio	TV
Survey:									
Politics	.17								
Sports		.33							
Science			.07						
Social Media				.41					
Messages					.14				
Emails						.11			
Video							.17		
Audio								.29	
TV									.34

#1 News: quality estimates



RESULTS

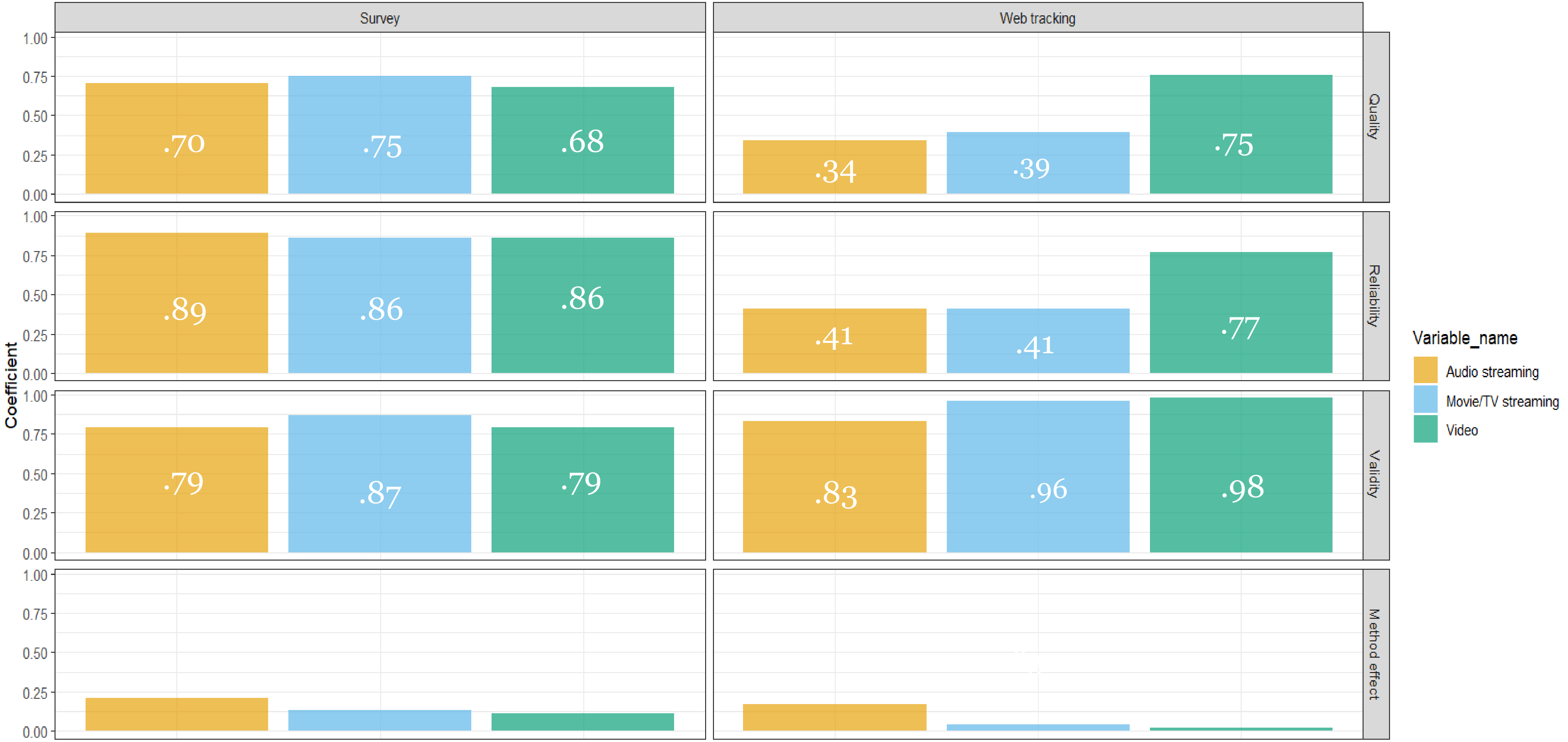
#1 News: quality estimates



RESULTS

#2 Communication: quality estimates

#3 Entertainment: quality estimates



RESULTS

#3 Entertainment: quality estimates

CONCLUSIONS

Take-home messages

webdataopp

- Results **put into question** the measurement quality of web tracking measurements
  - Some concepts are measures very accurately: **communication and video streaming**

    → **Variance explained by trait: +/- 80%**

  - While others are extremely off: **news media exposure and some entertainment**

    → **Variance explained by trait: 12-39% !!!**

Take-home messages

- Results **put into question** the measurement quality of web tracking measurements
    - Some concepts are measures very accurately: **communication and video streaming**

      → **Variance explained by trait: +/- 80%**

    - While others are extremely off: **news media exposure and some entertainment**

      → **Variance explained by trait: 12-39% !!!**

Surveys, on the other hand, perform acceptably well. They also struggle more with news, but their quality is never below .50 and generally around .70 (agrees w/ Alwin)

Take-home messages

- Results **put into question** the measurement quality of web tracking measurements
  - Some concepts are measures very accurately: **communication and video streaming**
  - While othe **entertainment**

Even if surprising, some of these results make logical sense when we think about the theory of the potential error causes of web tracking data!

Surveys, on the other hand, perform acceptably well. They also struggle more with news, but their quality is never below .50 and generally around .70 (agrees w/ Alwin)

The limits of this approach

- **VERY preliminary results…take with a pinch of salt**

- We need to think much more about the MTMM models used, how to fine tune them, and their limitations
    1. Is it biased towards surveys?
    2. Is it of any value if surveys and web tracking do not measure the same to start with?
    3. True score model??
    4. Differential measurement errors!

Thanks!

Questions?

Oriol J. Bosch| PhD Candidate, The London School of Economics

o.bosch-jover@lse.ac.uk

orioljbosch

https://orioljbosch.com/