Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Google econometrics: Nowcasting euro area car sales and big data quality requirements[1]

Per Nymand-Andersen,

European Central Bank

The paper has been published as an ECB statistics working paper

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

BANK INDONESIA
BANK SENTRAL REPUBLIK INDONESIA

EUROPEAN CENTRAL BANK
EUROSYSTEM

Irving Fisher Committee on
Central Bank Statistics
BANK FOR INTERNATIONAL SETTLEMENTS

Per Nymand-Andersen
Adviser, European Central Bank

Google econometrics: Nowcasting euro area car sales and quality requirements

Picture: INDOLINK Consulting's blog

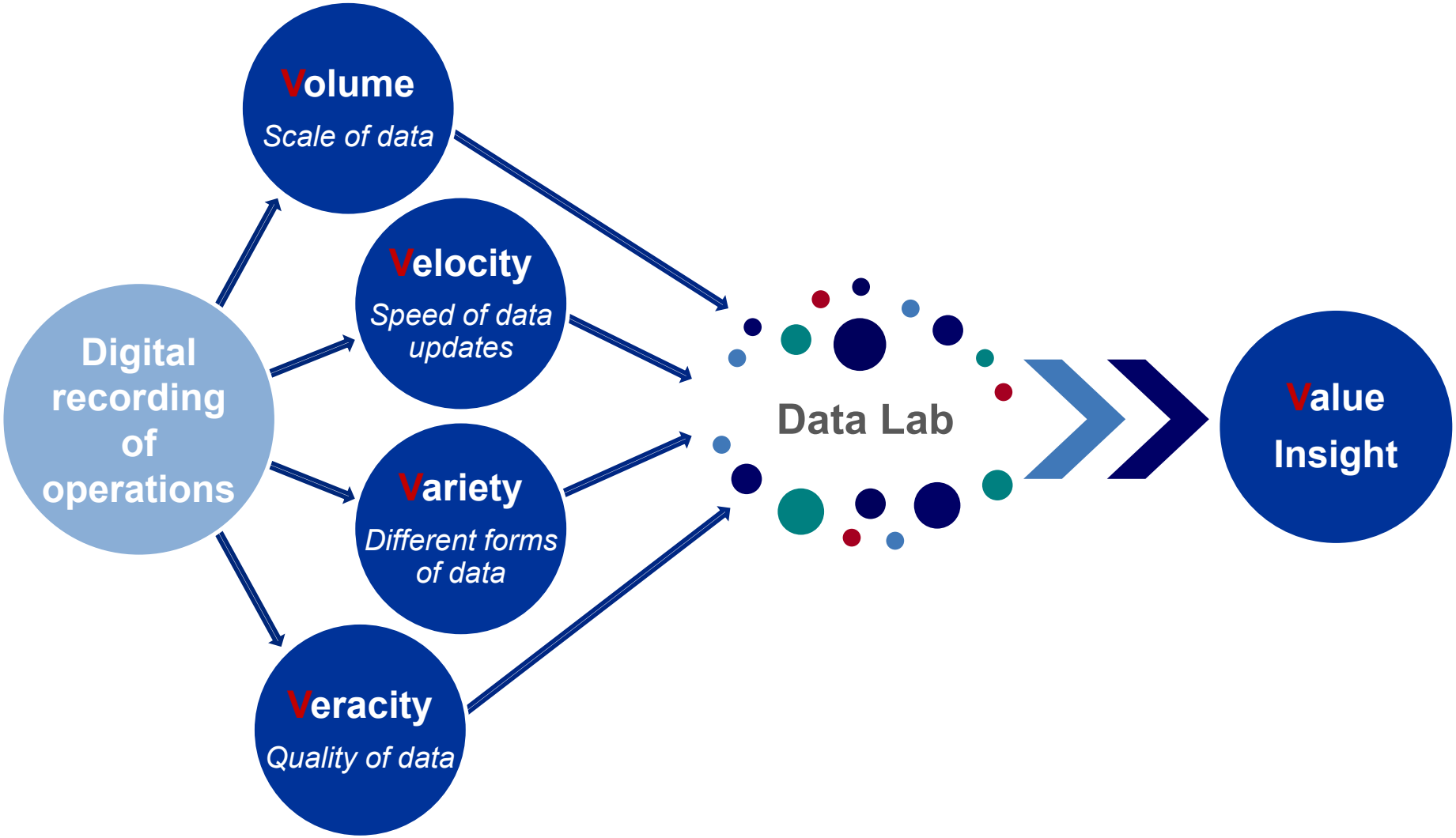
Bank Indonesia - IFC/BIS Workshop on big data for central bank policies, 23-25 July 2018

Overview

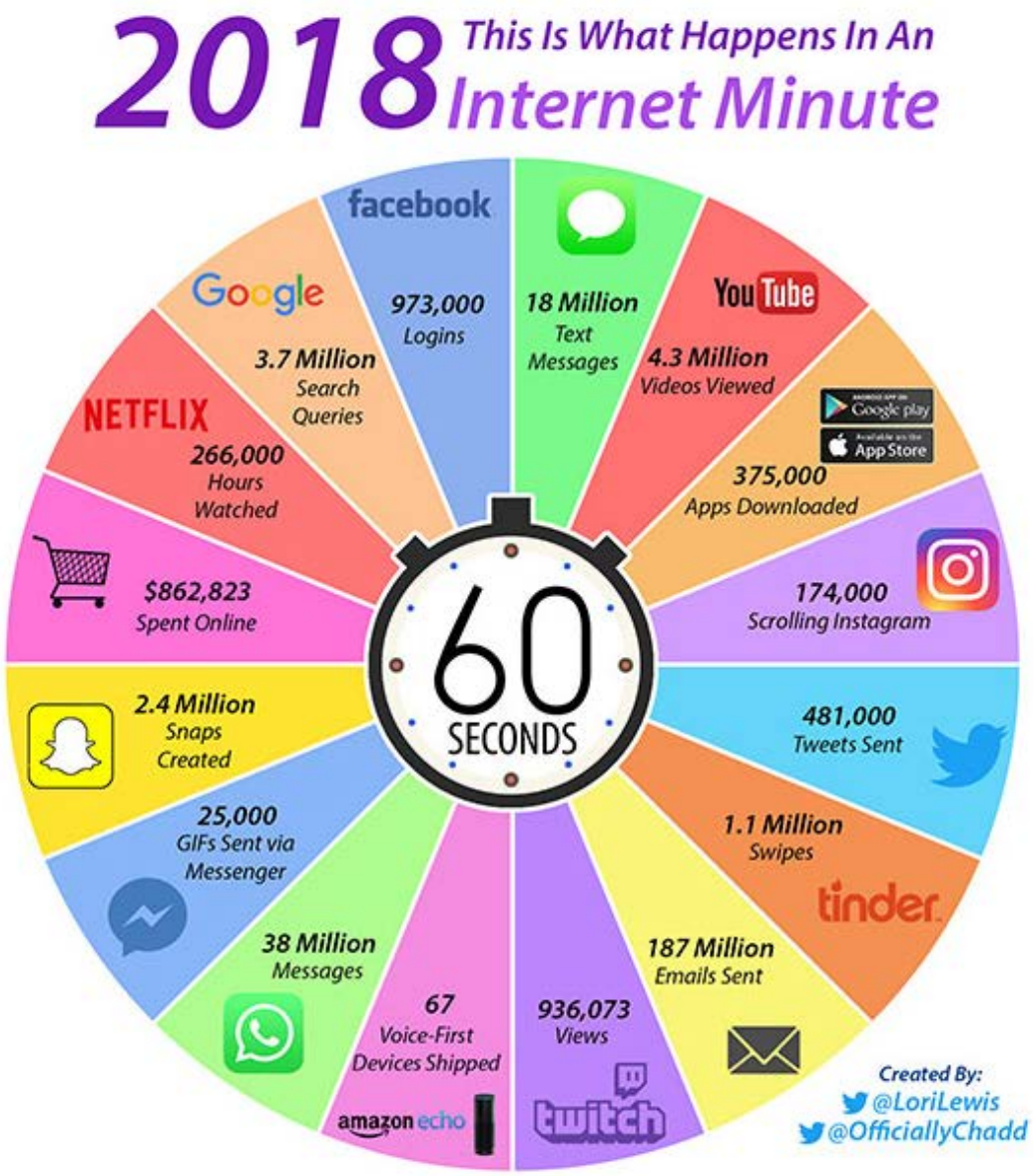
1	Pretty Big Data	- Data never sleeps	
2	Google econometrics	- Data mania versus phobia	
3	Discovery	- Big Data analytics	

Disclaimer: The opinions expressed in this presentation are not necessarily those of the European Central Bank (ECB) or the European System of Central Banks (ESCB)

Pretty Big Data – 5 Vs – Adding Value to data



Data never sleeps



Digital exploration

- Accessing
- Structuring datasets
- Linking data sets
- Slice & dice across time and datasets
- Standards (concepts, semantics)
- Transparency in methods
- Representatively
- Robustness of findings

Which preparations are needed today to have the capacity and functionality required in 3 years time?

- From experimenting to central banking tool kits?
- Linking current and past data
- Querying variety of formats
- Analytical techniques & tools
- Technical independent
- Skill sets

Source: 2018 Update: What Happens In An Internet Minute, Lori Lewis

Data mania versus phobia – a paradigm of borderless records

Digital transformation in finance and economics

E- trade

Settlement systems

Credit cards

Mobile transaction

Lending & financing

Big data

Today's talk

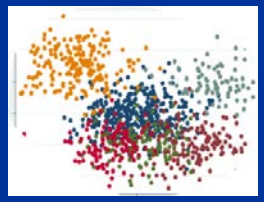
S-media

Price scans

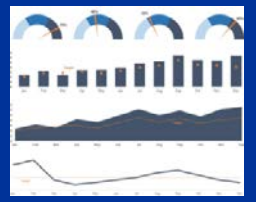
- Fintech
- Crypto assets
- DLT
- S-contracts

Data lab

Systematic acquire, Structure, Process,



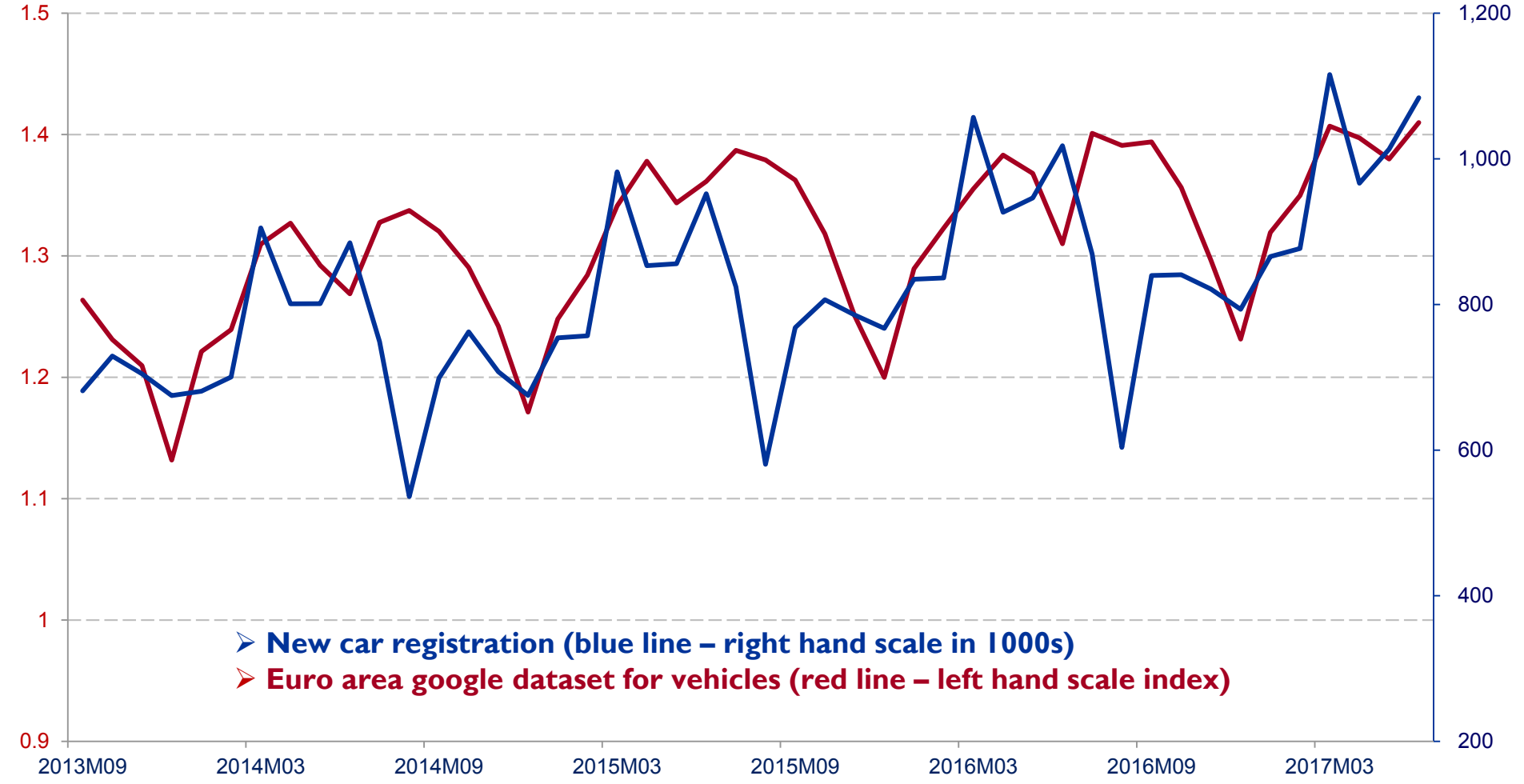
Statistical algorithm and data explorations



Packaging data for Insights & business

Google econometrics – Nowcasting car sales

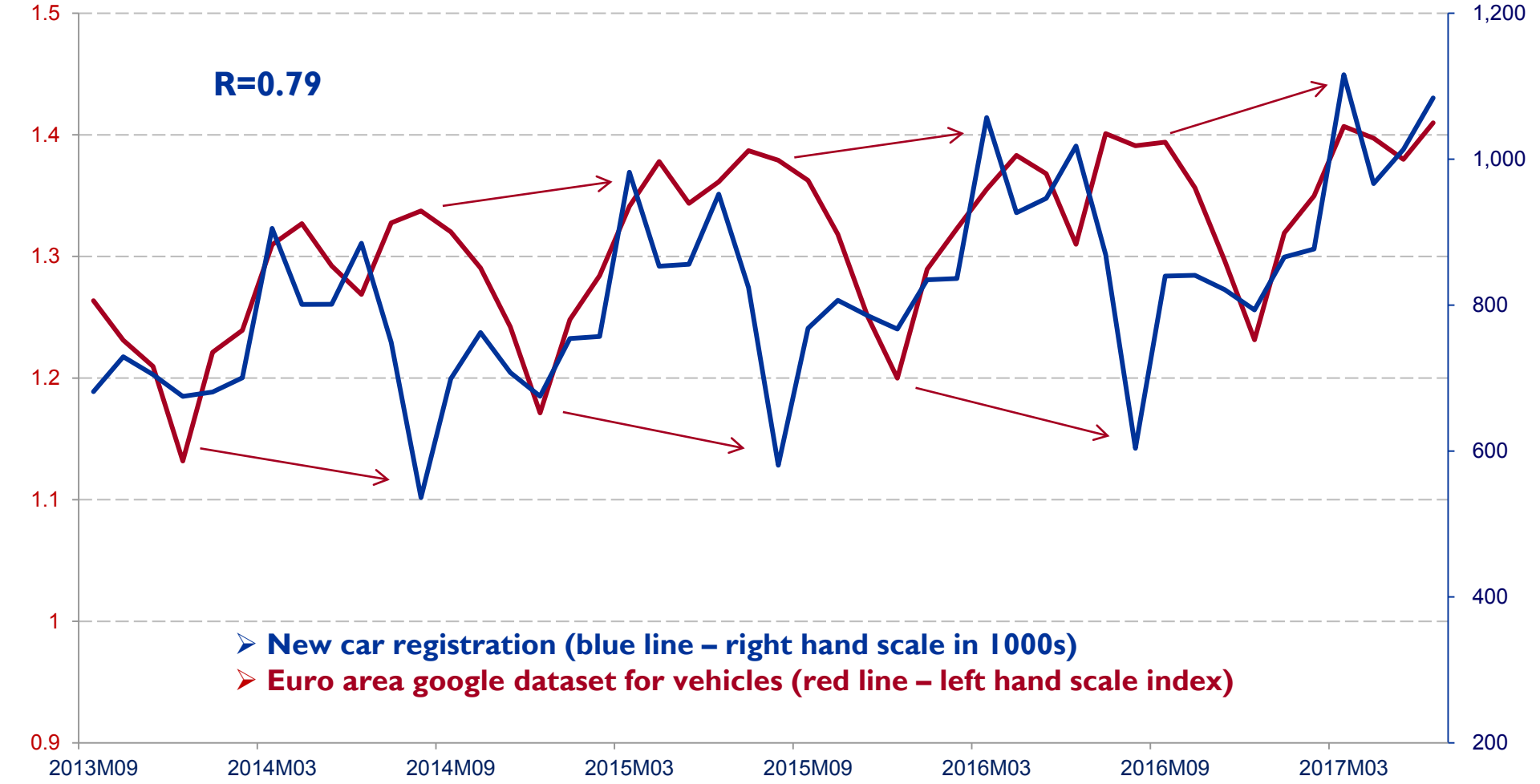
- Google data: "Autos & Vehicles",
- 10 euro area countries, weighted by national car registrations, monthly series, indexed, Represents 96% of euro area car registrations



Google econometrics: nowcasting euro area car sales and big data quality requirements, Nymand-Andersen, P., Pantelidis, E.,ECB SPS no 30

Google econometrics – Nowcasting car sales

1. There is a short term and long term dynamics between the two series
2. Two-way causal relationship and five lag lengths
3. Google data accounts for ≈ 22% of the variance in the car sales (2nd months & onwards)



Google econometrics – Nowcasting car sales

- ***7 seasonal autoregressive models*** *are used for nowcasting car sales with and without Google data and compared against a* ***seasonal autoregressive model*** (baseline)

Baseline model: $\mathbf{B_t} = \alpha + \beta_1 \log(y_{t-1}) + \beta_2 \log(y_{t-12}) + \varepsilon_t$

with Google data: $Y_t = \alpha + \beta_1 \log(y_{t-1}) + \beta_2 \log(y_{t-12}) + G_t + \varepsilon_t$, **Where**

$$\mathbf{G_t} = \beta_3 \Delta(G_t) + \beta_4 \Delta(G_{t-1}) + \beta_5 \Delta(G_{t-2}) + \beta_6 \Delta(G_{t-3}) + \beta_7 \Delta(G_{t-4}) + \beta_8 \Delta(G_{t-5})$$

Other indicators: $Y_t = B_t + [(I_i); \log(I_i)]$ V $Y_t = B_t + [(I_i); \log(I_i)] + G_t$

$I_{i=1,2,3}$ **Three indicators without and with Google data**

- **"Harmonised Index of Consumer Prices" for "Motor cars" with *one lag*.**
- **"Industrial Confidence Indicator"**
- **"Disposable Income of Households" log with *five lags*.**

The quarterly income data is converted into monthly (linear interpolation method)

Google econometrics – Nowcasting car sales

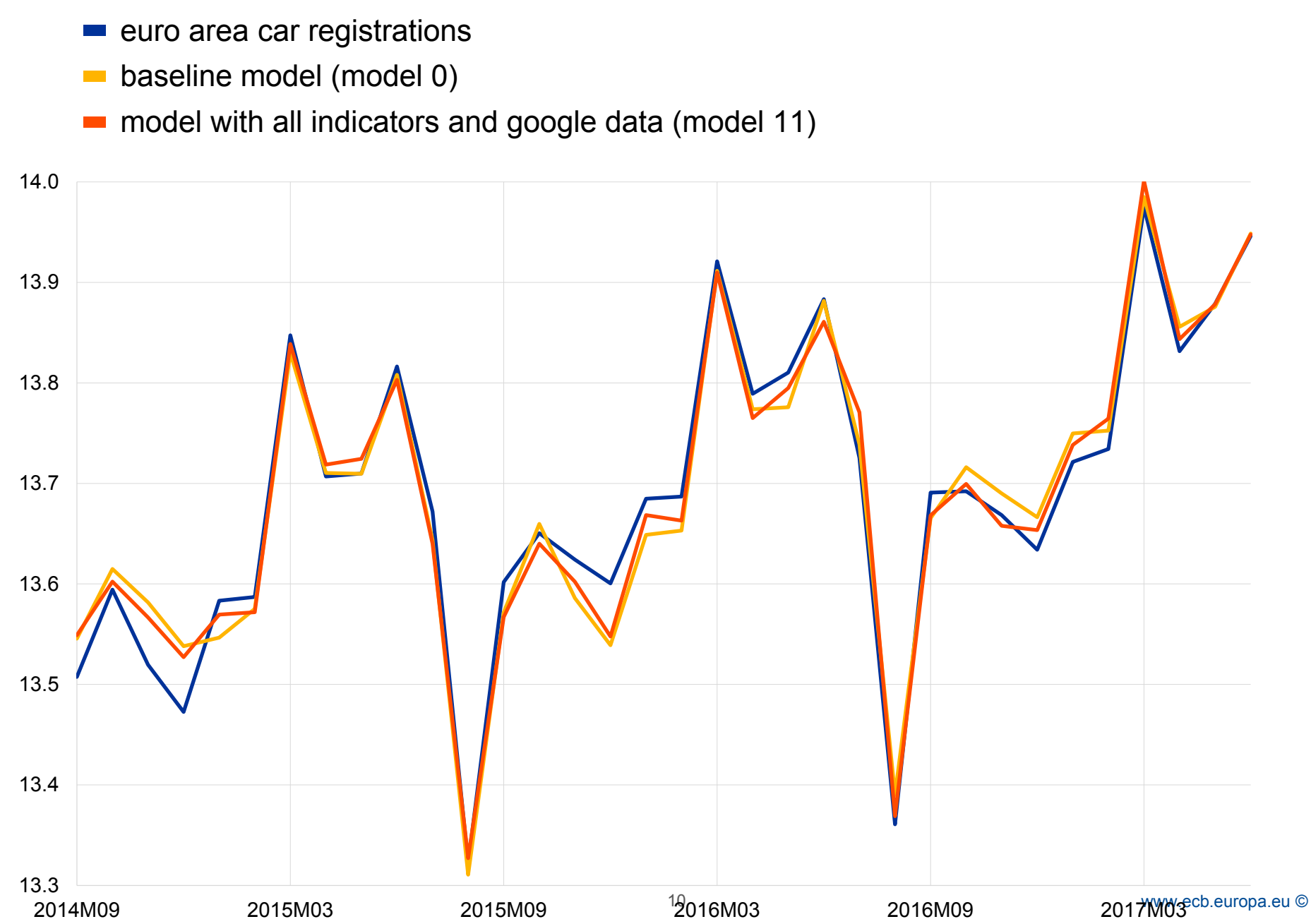
Overview of model performance and nowcasting accuracy

7 take aways

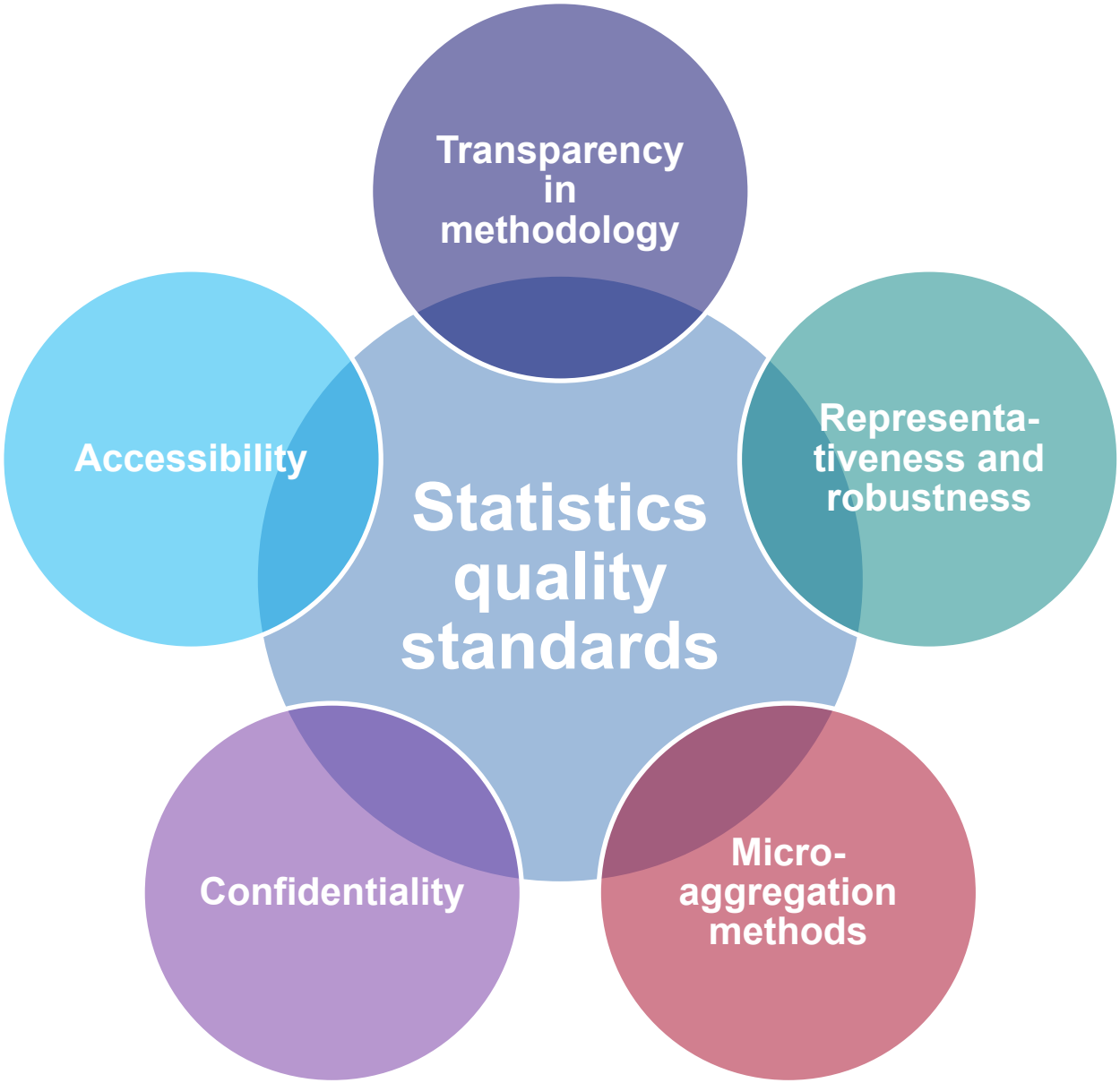
	Model/criteria	RMSE	MAE	MAPE	Improvement	DM# Base	DM# Pair (6)
0	Baseline	0.029019	0.023711	0.174652	(1)	(4)	0.0331
1	Baseline & Google (2)	0.026812	0.020303	0.149357	15.7%	0.0331	
2	With inflation rate	0.027585	0.023092	0.169944	3.4%	0.3357	0.0204
3	With inflation rate & Google	0.024885	0.019445	0.14288	21.5%	0.0383	
4	With confidence indicator	0.028942	0.023738	0.174715	4.6%	(5) 0.5289	0.0307
5	With confidence indicator & Google	0.026747	0.020334	0.149567	15.6%	0.0347	
6	With income	0.027437	0.02231	0.164112	5.8%	0.1791	0.0453
7	With income & Google	0.02296	0.018139	0.133189	(3) 30.5%	0.0195	
8	With household savings	0.027669	0.02122	0.15623	10.9%	0.0383	(7) 0.1965
9	With household savings & Google	0.025713	0.019338	0.142174	21.4%	0.0218	
							Including all explanatory variables
10	All indicators	0.022811	0.017993	0.13253	(3) 31.2%	0.0086	0.0019
11	All indicators & Google	0.012257	0.010332	0.075806	131%	0.00004	

DM = Diebold and Mariano's test - Exploring the usefulness of internet search data on future car sales

Google econometrics – Nowcasting car sales



Big Data analytics – Quality assessment



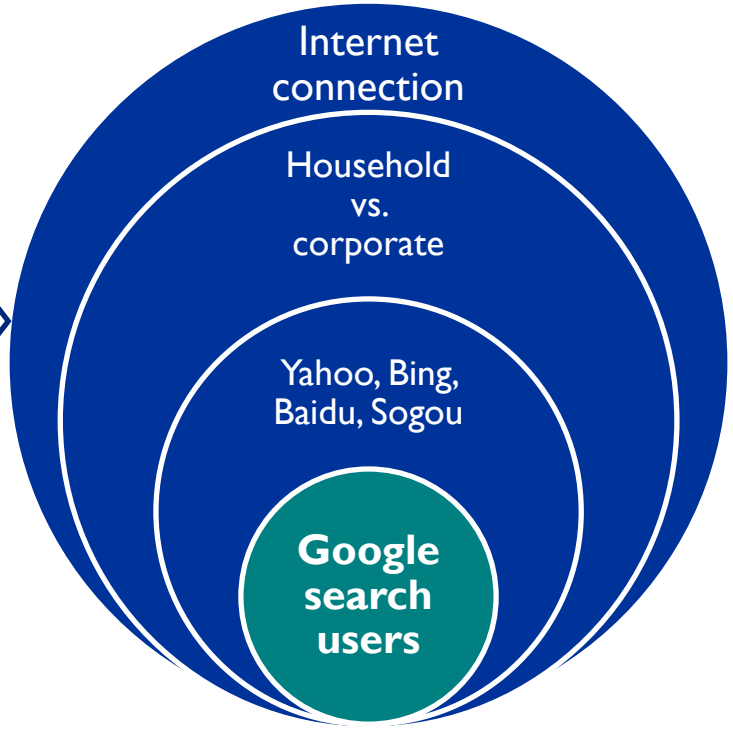
Big Data analytics – Quality assessment

One **misperception** of big data is that we **do not need** to worry about **sample bias and representativeness**, as large volumes of information supersede standard sampling theory, since big data provide census-type information

1

Digital recording of operations

Census?



2 **Unit Measurement** → **Not people - but searches**

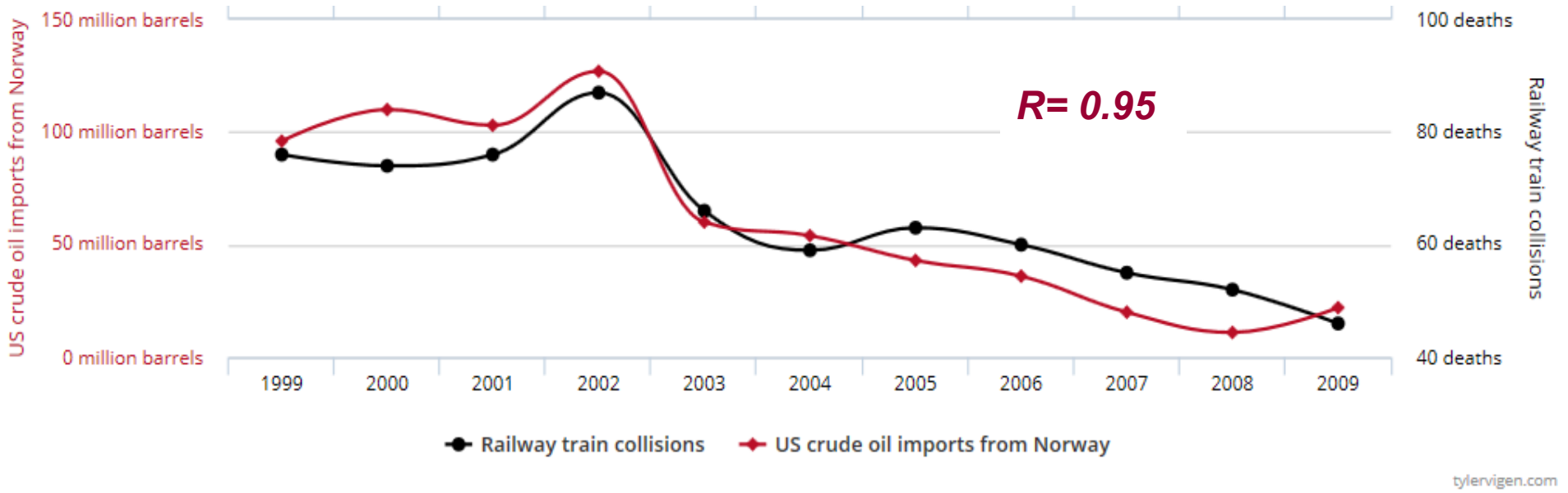
3 **Event driven** → **Volume changes may not refer to changes in demand (e.g. emission checks)**

Google logo is downloaded from *Wikipedia*.

Big Data analytics – Quality assessment

Correlation is *not (necessary)* causation

US crude oil imports from Norway correlates with drivers killed in collision with railway train



No conclusion can be drawn simply on the basis of correlations between two variables. *The similarity is a coincidence*. We *should* say that there is no causation

"The invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning"
Stephen Jay Gould, American evolutionary biologist and author, 1981

Three takeaways

Progress lies in experimenting

Amara's law

Potential valuable source for nowcasting economic activities

Foster transparency in data quality standards

Moving from experimenting to central banking tool kits

Collaborations - build partnerships for excellence

Google econometrics

Any question?