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

Per Nymand-Andersen,
European Central Bank

*The paper has been published as an [ECB statistics working paper](http://ecb.europa.eu)*

---

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.
Per Nymand-Andersen
Adviser, European Central Bank

Google econometrics:
Nowcasting euro area car sales and quality requirements

Picture: INDOLINK Consulting’s blog
<table>
<thead>
<tr>
<th></th>
<th>Pretty Big Data</th>
<th>- Data never sleeps</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Google econometrics</td>
<td>- Data mania versus phobia</td>
</tr>
<tr>
<td>3</td>
<td>Discovery</td>
<td>- Big Data analytics</td>
</tr>
</tbody>
</table>

**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)
Digital recording of operations

Volume
Scale of data

Velocity
Speed of data updates

Variety
Different forms of data

Veracity
Quality of data

Data Lab

Value
Insight
Data never sleeps

2018: This Is What Happens In An Internet Minute

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

Today’s talk

Big data

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

- New car registration (blue line – right hand scale in 1000s)
- Euro area google dataset for vehicles (red line – left hand scale index)

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 (2\textsuperscript{nd} months & onwards)

R=0.79

- New car registration (blue line – right hand scale in 1000s)
- Euro area google dataset for vehicles (red line – left hand scale index)
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: \[ B_t = \alpha + \beta_1 \log(y_{t-1}) + \beta_2 \log(y_{t-12}) + \varepsilon_t \]

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

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

Other indicators: \[ Y_t = B_t + [\{I_i\}; \log(I_i)] \]

\[ Y_t = B_t + [\{I_i\}; \log(I_i)] + G_t \]

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

<table>
<thead>
<tr>
<th>Model/criteria</th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>Improvement</th>
<th>DM# Base</th>
<th>DM# Pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Baseline</td>
<td>0.0290</td>
<td>0.0237</td>
<td>0.1746</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Baseline &amp; Google</td>
<td>0.0268</td>
<td>0.0203</td>
<td>0.1493</td>
<td>15.7%</td>
<td>0.0331</td>
<td>0.0331</td>
</tr>
<tr>
<td>2 With inflation rate</td>
<td>0.0276</td>
<td>0.0230</td>
<td>0.1699</td>
<td>3.4%</td>
<td>0.3357</td>
<td>0.3357</td>
</tr>
<tr>
<td>3 With inflation rate &amp; Google</td>
<td>0.0249</td>
<td>0.0194</td>
<td>0.1429</td>
<td>21.5%</td>
<td>0.0383</td>
<td>0.0204</td>
</tr>
<tr>
<td>4 With confidence indicator</td>
<td>0.0289</td>
<td>0.0237</td>
<td>0.1747</td>
<td>4.6%</td>
<td>0.5289</td>
<td>0.0307</td>
</tr>
<tr>
<td>5 With confidence indicator &amp; Google</td>
<td>0.0267</td>
<td>0.0203</td>
<td>0.1496</td>
<td>15.6%</td>
<td>0.0347</td>
<td>0.0347</td>
</tr>
<tr>
<td>6 With income</td>
<td>0.0274</td>
<td>0.0223</td>
<td>0.1641</td>
<td>5.8%</td>
<td>0.1791</td>
<td>0.0453</td>
</tr>
<tr>
<td>7 With income &amp; Google</td>
<td>0.0229</td>
<td>0.0181</td>
<td>0.1332</td>
<td>30.5%</td>
<td>0.0195</td>
<td>0.0195</td>
</tr>
<tr>
<td>8 With household savings</td>
<td>0.0277</td>
<td>0.0212</td>
<td>0.1562</td>
<td>10.9%</td>
<td>0.0383</td>
<td>0.1965</td>
</tr>
<tr>
<td>9 With household savings &amp; Google</td>
<td>0.0257</td>
<td>0.0193</td>
<td>0.1422</td>
<td>21.4%</td>
<td>0.0218</td>
<td>0.0218</td>
</tr>
<tr>
<td>10 All indicators</td>
<td>0.0228</td>
<td>0.0179</td>
<td>0.1325</td>
<td>31.2%</td>
<td>0.0086</td>
<td>0.0086</td>
</tr>
<tr>
<td>11 All indicators &amp; Google</td>
<td>0.0122</td>
<td>0.0103</td>
<td>0.0758</td>
<td>131%</td>
<td>0.0004</td>
<td>0.0019</td>
</tr>
</tbody>
</table>

DM = Diebold and Mariano’s test - Exploring the usefulness of internet search data on future car sales
Google econometrics – Nowcasting car sales

- euro area car registrations
- baseline model (model 0)
- model with all indicators and google data (model 11)
Big Data analytics – Quality assessment

Statistics quality standards

- Transparency in methodology
- Representativeness and robustness
- Micro-aggregation methods
- Confidentiality
- Accessibility
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.
**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
Google econometrics: Nowcasting euro area car sales and quality requirements

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
FACTS COUNT

The **ECB Statistics Paper Series (SPS)** is a channel for statisticians, economists, researchers and other professionals to publish innovative work undertaken in the area of statistics and related methodologies of interest to central banks.

Fact-check your talk before you walk

WHAT ABOUT YOU WRITING?