Developments in the residential mortgage market in Germany – What can Google data tell us?

9th IFC Conference, „Are post-crisis statistical initiatives completed?“, Session 5 – Big Data

Simon Oehler, Deutsche Bundesbank
Agenda

1. Motivation & Literature Review
2. Google Data
3. Econometric Approach
4. Results
5. Conclusion
1. Motivation & Literature Review

- In recent years interest in internet search data has increased & research has started to investigate the potential of this new data source.
- Examples comprise:
  - Choi, Varian (2011); Predicting the present with Google Trends
  - Schmidt, Vosen (2009); Forecasting Private Consumption, Survey-based Indicators vs. Google Trends
  - McLaren, Shanbhogue (2011); Using internet search data as economic indicators, BoE Quarterly Bulletin, Q2
  - Askitas, Zimmermann (2014); Detecting Mortgage Delinquencies with Google Trends
  - Chauvet, Gabriel, Lutz (2016); Mortgage default risk: New evidence from internet search queries
  - Saxa (2014); Forecasting Mortgages, CNB Working Paper
1. Motivation & Literature Review

Why Google search data?

“An individual's interest in certain documents (and not in others) is a function of the individual's state and so are search queries which are used to locate them. These queries are therefore utterances worth being investigated […]” - Askitas, Zimmermann (2014)

“We have found that [search] queries can be useful leading indicators for subsequent consumer purchases in situations where consumers start planning purchases significantly in advance of their actual purchase decision.” - Choi, Varian (2011), Predicting the Present with Google Trends

– Real estate & the financing thereof should meet this condition

Research question:

In how far can Google search data explain the variation in volumes of mortgage transactions at the federal level in Germany?
2. Google Data

- **37 Google series** are downloaded from https://trends.google.de/trends
- Selection is not solely “data driven”. A priori “economic/human reasoning” involved as selection of time series is restricted to search terms relating to “mortgage” or “housing”.
- Geography: Germany
- Language: German
- Frequency: Monthly
- Period: 2004 – April 2018
- Sampling: random sample of total searches is drawn by Google
- **Index**: no information about actual volumes or query shares

\[ I(Kredit_t) = \frac{R(Kredit_t)}{\max\{R(Kredit_t)\}} \times 100 \quad \text{with} \quad R(Kredit_t) = \frac{Kredit_t}{Google_t} \]
2. Google Data

New mortgage business by German banks, millions Euro, monthly

Effective interest rates for new mortgage business by German Banks, percentage points, monthly

Google Composite Indicator
Kredit + Darlehen + Hypothek + Baufinanzierung, Index

Unemployment in Germany, millions, monthly

Simon Oehler, Deutsche Bundesbank
Basel, 30-31 August 2018
3. Econometric approach

- All time series are log-transformed and first differenced.

- Seasonal adjustment:
  - **Response:** New mortgage business with seasonal patterns, particularly in July
  - **Controls:**
    - Effective Interest rate: no seasonality
    - Unemployment: seasonally adjusted
  - **Google:**
    - Almost all Google series with (strong) seasonal pattern around the end of the year: large drop in December and sharp rise in January of the subsequent year.

- Modeling approach: Benchmark augmented by controls and Google data (stepwise forward selection procedure)

\[
\Delta \text{mortgages}_t = \beta_m L^m \Delta \text{mortgages}_t \\
\Delta \text{mortgages}_t = \beta_m L^m \Delta \text{mortgages}_t + \gamma_m L^m \Delta \text{interest}_t + \theta_m L^m \Delta \text{unempl}_t \\
\Delta \text{mortgages}_t = \beta_m L^m \Delta \text{mortgages}_t + \gamma_m L^m \Delta \text{interest}_t + \theta_m L^m \Delta \text{unempl}_t + \delta_m \Delta \text{Google}_t
\]
4. Results
Out-of-sample forecasts

Forecast Evaluation
Date: 08/14/18  Time: 16:41
Sample: 2016M01 2018M04
Included observations: 28
Evaluation sample: 2016M01 2018M04
Number of forecasts: 6

<table>
<thead>
<tr>
<th>Equation</th>
<th>F-stat</th>
<th>F-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>BENCH</td>
<td>9.892147</td>
<td>0.0000</td>
</tr>
<tr>
<td>BENCH</td>
<td>UNPL</td>
<td>7.354970</td>
</tr>
<tr>
<td>BAUFI</td>
<td>13.18475</td>
<td>0.0000</td>
</tr>
<tr>
<td>BAUFI</td>
<td>HYP</td>
<td>8.076732</td>
</tr>
<tr>
<td>BAUFI</td>
<td>HYP</td>
<td>KREDIT</td>
</tr>
<tr>
<td>BAUFI</td>
<td>KREDIT</td>
<td>VGL</td>
</tr>
</tbody>
</table>

Combination tests
Null hypothesis: Forecast i includes all information contained in others

<table>
<thead>
<tr>
<th>Equation</th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>SMAPE</th>
<th>Theil U1</th>
<th>Theil U2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BENCH</td>
<td>0.090353</td>
<td>0.076160</td>
<td>112.5150</td>
<td>167.2447</td>
<td>0.874054</td>
<td>0.957484</td>
</tr>
<tr>
<td>BENCH</td>
<td>UNPL</td>
<td>0.102884</td>
<td>0.088681</td>
<td>282.8204</td>
<td>166.0138</td>
<td>0.080323</td>
</tr>
<tr>
<td>BAUFI</td>
<td>0.095041</td>
<td>0.079783</td>
<td>306.7584</td>
<td>141.2235</td>
<td>0.698204</td>
<td>0.927898</td>
</tr>
<tr>
<td>BAUFI</td>
<td>HYP</td>
<td>0.079904</td>
<td>0.064154</td>
<td>222.1042</td>
<td>120.8835</td>
<td>0.590645</td>
</tr>
<tr>
<td>BAUFI</td>
<td>HYP</td>
<td>KREDIT</td>
<td>0.072731</td>
<td>0.061808</td>
<td>252.6928</td>
<td>116.4031</td>
</tr>
<tr>
<td>BAUFI</td>
<td>KREDIT</td>
<td>VGL</td>
<td>0.093064</td>
<td>0.078871</td>
<td>290.7800</td>
<td>136.7531</td>
</tr>
</tbody>
</table>
4. Results
Out-of-sample forecasts
5. Conclusion

- Results suggest that Google data contain (short term) cyclicality which can be exploited for forecasting/nowcasting.

- In particular the search terms „Baufinanzierung“, „Hypothek“, „Kreditvergleich“, „Kreditrechner“ proved to be significant and relevant indicators for the change in growth rates of mortgage business in Germany under the tested model specifications.

- Thus far, the models presented here control for mortgage market interest rates and unemployment as a macroeconomic indicator.

- Further robustness checks are needed. In particular:
  - Evaluate GoogleTrends relative to survey indicators
  - Further variable selection procedures to be applied
Thank you for your attention!

E-mail: simon.oehler@bundesbank.de
Backup I
Static Forecasts Benchmark

Forecast: SUD231_D11F
Actual: SUD231_D11_LNDF
Forecast sample: 2016M01 2018M04
Included observations: 28
Root Mean Squared Error 0.063301
Mean Absolute Error 0.054191
Mean Abs. Percent Error 287.2026
Theil Inequality Coefficient 0.427125
  Bias Proportion 0.002628
  Variance Proportion 0.404280
  Covariance Proportion 0.593092
Theil U2 Coefficient 0.490265
Symmetric MAPE 103.7871
Backup II
Static Forecasts Benchmark with interest rate

Forecast: SUD231_D11F
Actual: SUD231_D11_LNDF
Forecast sample: 2016M01 2018M04
Included observations: 28
Root Mean Squared Error 0.064830
Mean Absolute Error 0.056045
Mean Abs. Percent Error 298.0209
Theil Inequality Coefficient 0.410584
  Bias Proportion 0.000123
  Variance Proportion 0.224556
  Covariance Proportion 0.775321
Theil U2 Coefficient 0.723895
Symmetric MAPE 111.4840

---

Simon Oehler, Deutsche Bundesbank
Basel, 30-31 August 2018
Page 13
Backup III
Static Forecast Google augmented II

Forecast: SUD231_D11F
Actual: SUD231_D11_LNDF
Forecast sample: 2016M01 2018M04
Included observations: 28
Root Mean Squared Error 0.054638
Mean Absolute Error 0.045272
Mean Abs. Percent Error 225.9676
Theil Inequality Coefficient 0.323992
  Bias Proportion 0.000322
  Variance Proportion 0.132201
  Covariance Proportion 0.867477
Theil U2 Coefficient 0.416872
Symmetric MAPE 93.20970
### Backup IV

#### Regression Output BAUFI_HYP_KREDITVGL

**Dependent Variable:** SUD231_D11_LNDF  
**Method:** Least Squares  
**Date:** 08/10/18  
**Time:** 17:43  
**Sample (adjusted):** 2004M09 2015M12  
**Included observations:** 136 after adjustments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUD231_D11_LNDF(-2)</td>
<td>-0.387154</td>
<td>0.067556</td>
<td>-5.730879</td>
<td>0.0000</td>
</tr>
<tr>
<td>SUD231_D11_LNDF(-1)</td>
<td>-0.503458</td>
<td>0.068513</td>
<td>-7.348352</td>
<td>0.0000</td>
</tr>
<tr>
<td>SUD231_D11_LNDF(-7)</td>
<td>-0.270081</td>
<td>0.063962</td>
<td>-4.222515</td>
<td>0.0000</td>
</tr>
<tr>
<td>SUD131_LNDF(-2)</td>
<td>-0.991662</td>
<td>0.204519</td>
<td>-4.848749</td>
<td>0.0000</td>
</tr>
<tr>
<td>GOOGLE_BAUFI_D11_LNDF(-3)</td>
<td>0.119307</td>
<td>0.044107</td>
<td>2.704970</td>
<td>0.0078</td>
</tr>
<tr>
<td>GOOGLE_HYP_D11_LNDF(-3)</td>
<td>-0.083612</td>
<td>0.020631</td>
<td>-4.052762</td>
<td>0.0001</td>
</tr>
<tr>
<td>GOOGLE_BAUFI_D11_LNDF(-1)</td>
<td>0.155329</td>
<td>0.043875</td>
<td>3.540303</td>
<td>0.0006</td>
</tr>
<tr>
<td>GOOGLE_KREDITVGL_D11_LNDF(-1)</td>
<td>0.086669</td>
<td>0.027745</td>
<td>3.123732</td>
<td>0.0022</td>
</tr>
<tr>
<td>GOOGLE_HYP_D11_LNDF(-1)</td>
<td>-0.042178</td>
<td>0.019847</td>
<td>-2.125220</td>
<td>0.0355</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>Description</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-squared</td>
<td>0.558962</td>
<td>Mean dependent var</td>
<td>0.004244</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.531180</td>
<td>S.D. dependent var</td>
<td>0.080372</td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>0.055031</td>
<td>Akaike info criterion</td>
<td>-2.897948</td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>0.384611</td>
<td>Schwarz criterion</td>
<td>-2.705199</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>206.0605</td>
<td>Hannan-Quinn criter.</td>
<td>-2.819619</td>
</tr>
<tr>
<td>Durbin-Watson stat</td>
<td>1.996244</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>