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:

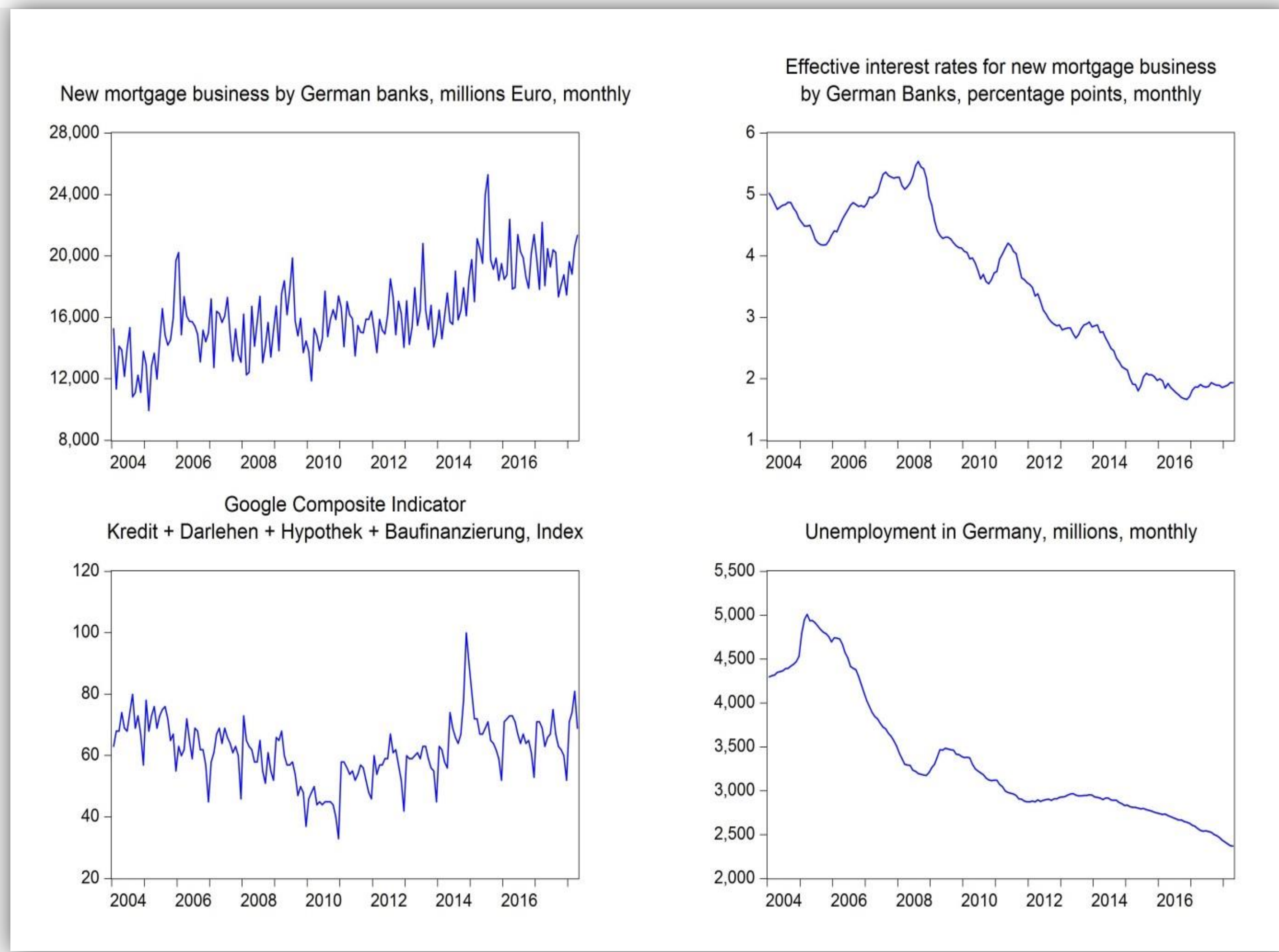
In how far can Google search data explain the variation in volumes of mortgage transactions at the federal level in Germany?

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

- $\boldsymbol{I(Kredit_t) = \frac{R(Kredit_t)}{\max\{R(Kredit_t)\}} \times 100}$ with $R(Kredit_t) = \frac{Kredit_t}{Google_t}$

2. Google Data



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)**

- $\Delta mortgages_t = \beta_m L^m \Delta mortgages_t$
- $\Delta mortgages_t = \beta_m L^m \Delta mortgages_t + \gamma_m L^m \Delta interest_t + \theta_m L^m \Delta unempl_t$
- $\Delta mortgages_t = \beta_m L^m \Delta mortgages_t + \gamma_m L^m \Delta interest_t + \theta_m L^m \Delta unempl_t + \delta_m L^m \Delta Google_t$

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

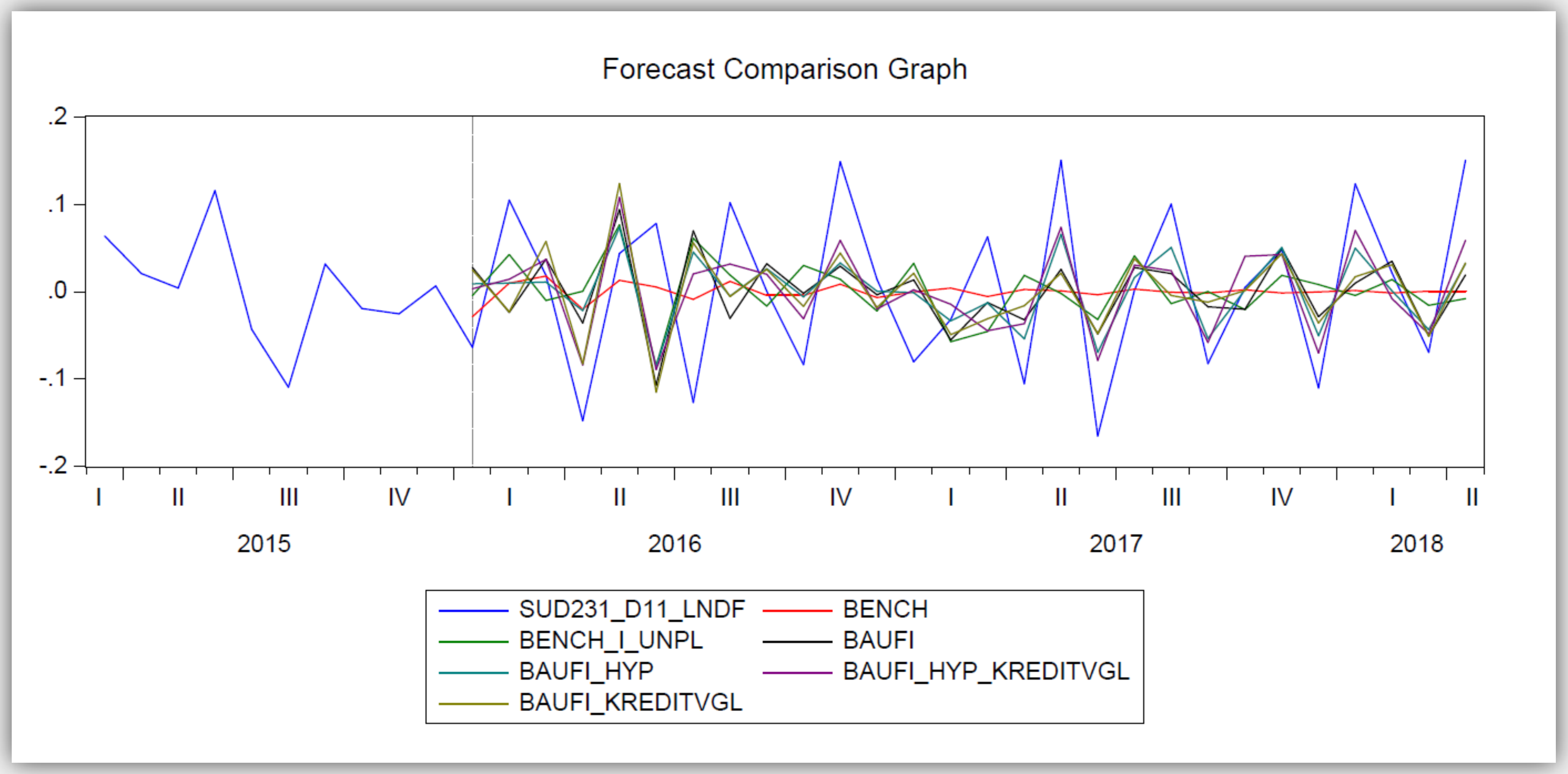
Combination tests
Null hypothesis: Forecast i includes all information contained in others

Equation	F-stat	F-prob
BENCH	9.892147	0.0000
BENCH I UNPL	7.354970	0.0003
BAUFI	13.18475	0.0000
BAUFI HYP	8.076732	0.0002
BAUFI_HYP_KREDIT	5.965349	0.0012
BAUFI_KREDITVGL	12.60686	0.0000

Evaluation statistics

Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
BENCH	0.090353	0.076160	112.5150	167.2447	0.874054	0.957484
BENCH_I_UNPL	0.102884	0.088681	282.8204	166.0138	0.803323	1.298619
BAUFI	0.095041	0.079783	306.7584	141.2235	0.698204	0.927898
BAUFI_HYP	0.079904	0.064154	222.1042	120.8835	0.590645	0.839572
BAUFI_HYP_KREDIT	0.072731	0.061808	252.6928	116.4031	0.498878	0.658610
BAUFI_KREDITVGL	0.093064	0.078871	290.7800	136.7531	0.656967	0.857488

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.

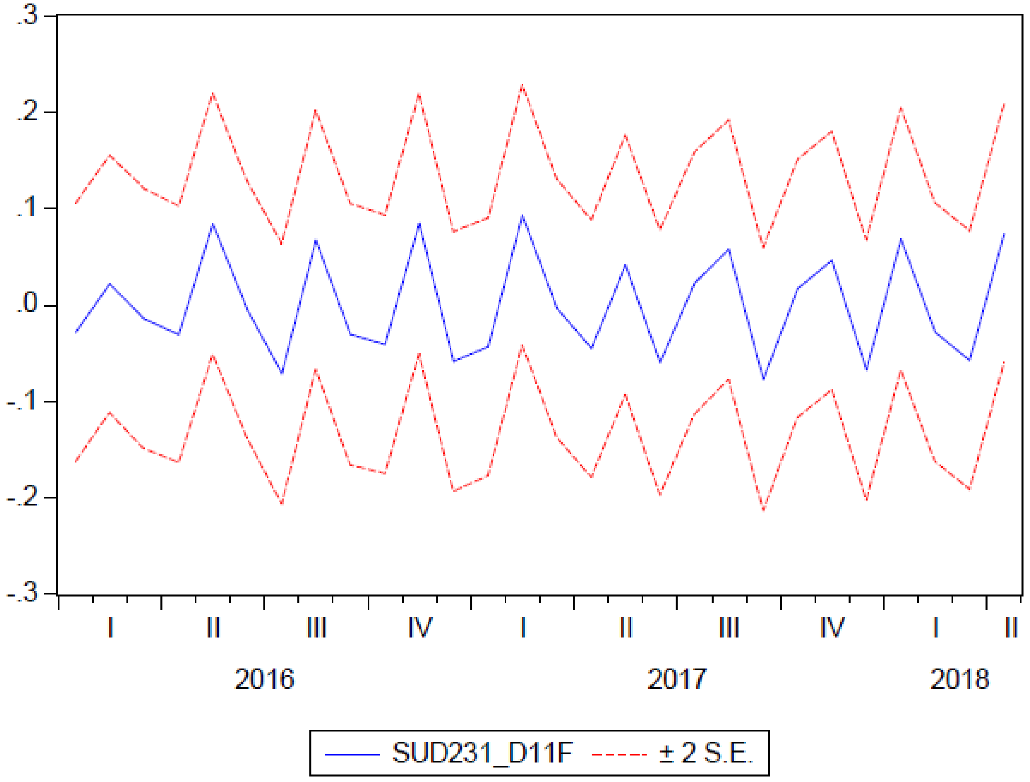
- Further robustness checks are needed. In particular:

  → Evaluate GoogleTrends relative to survey indicators
  → Further variable selection procedures to be applied

Thank you for your attention!

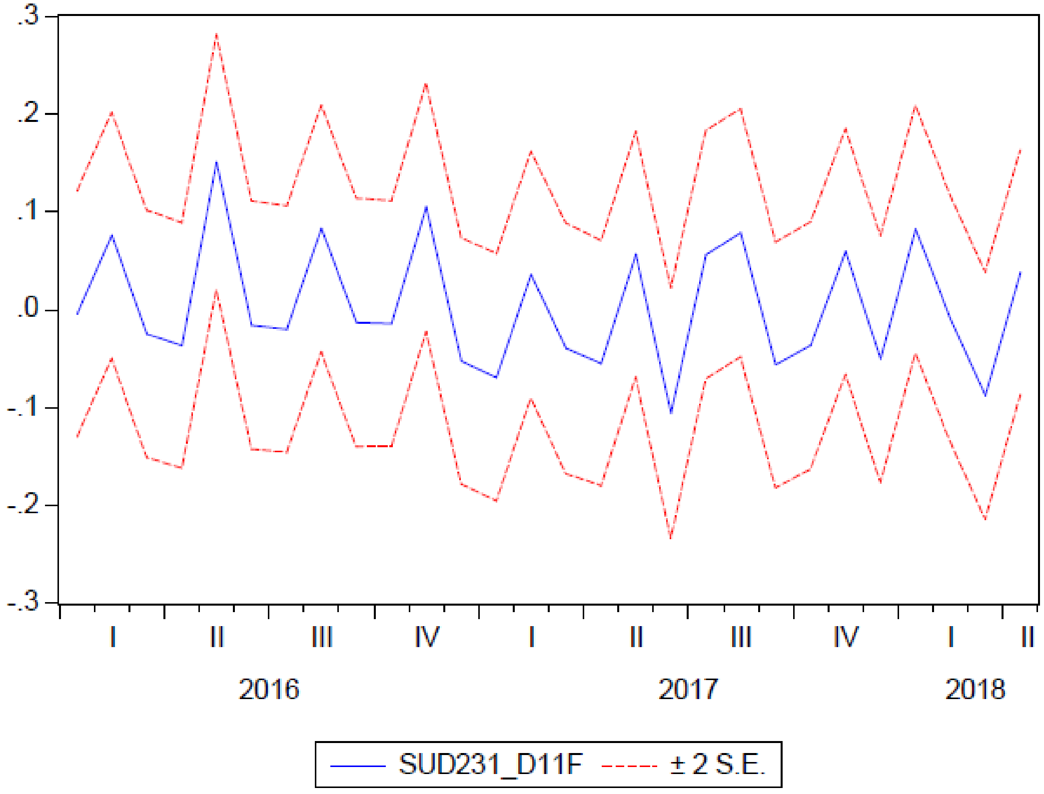
E-mail: simon.oehler@bundesbank.de

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

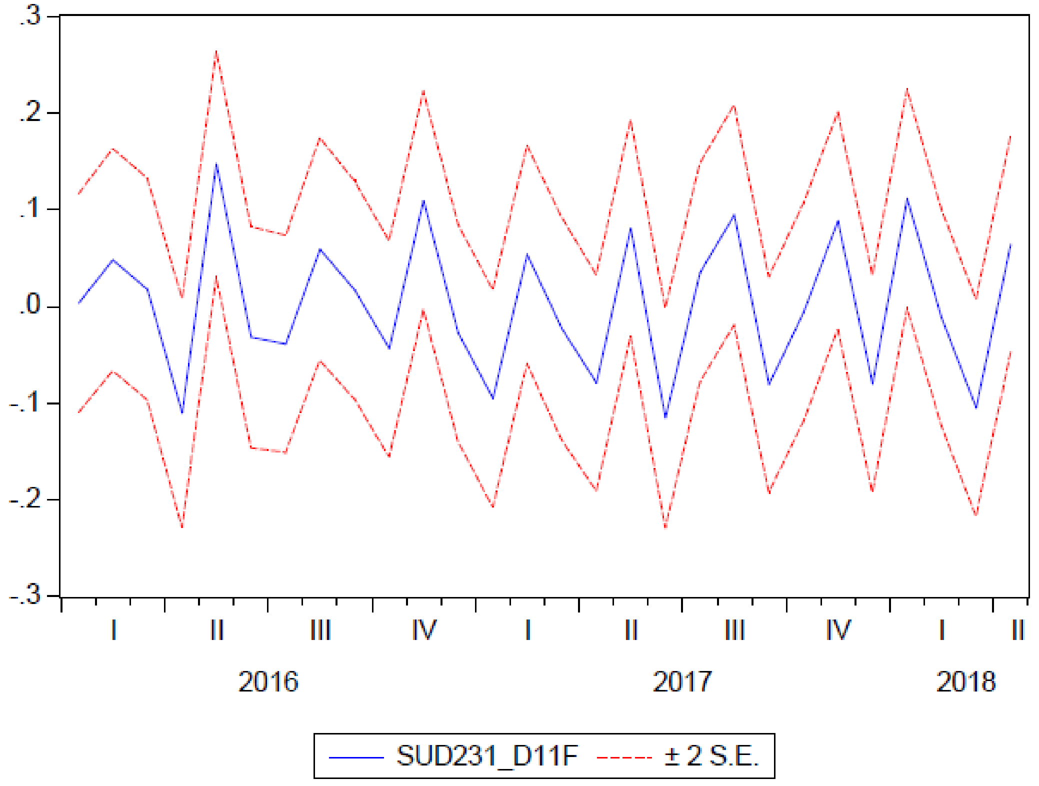
Backup II
Static Forecasts Benchmark with interest rate



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

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

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

R-squared	0.558962	Mean dependent var	0.004244
Adjusted R-squared	0.531180	S.D. dependent var	0.080372
S.E. of regression	0.055031	Akaike info criterion	-2.897948
Sum squared resid	0.384611	Schwarz criterion	-2.705199
Log likelihood	206.0605	Hannan-Quinn criter.	-2.819619
Durbin-Watson stat	1.996244		