Sentiment-Based Predictions of Housing Market Turning Points with Google Trends

Marian Alexander Dietzel

ERES Istanbul 2015

Google Trends

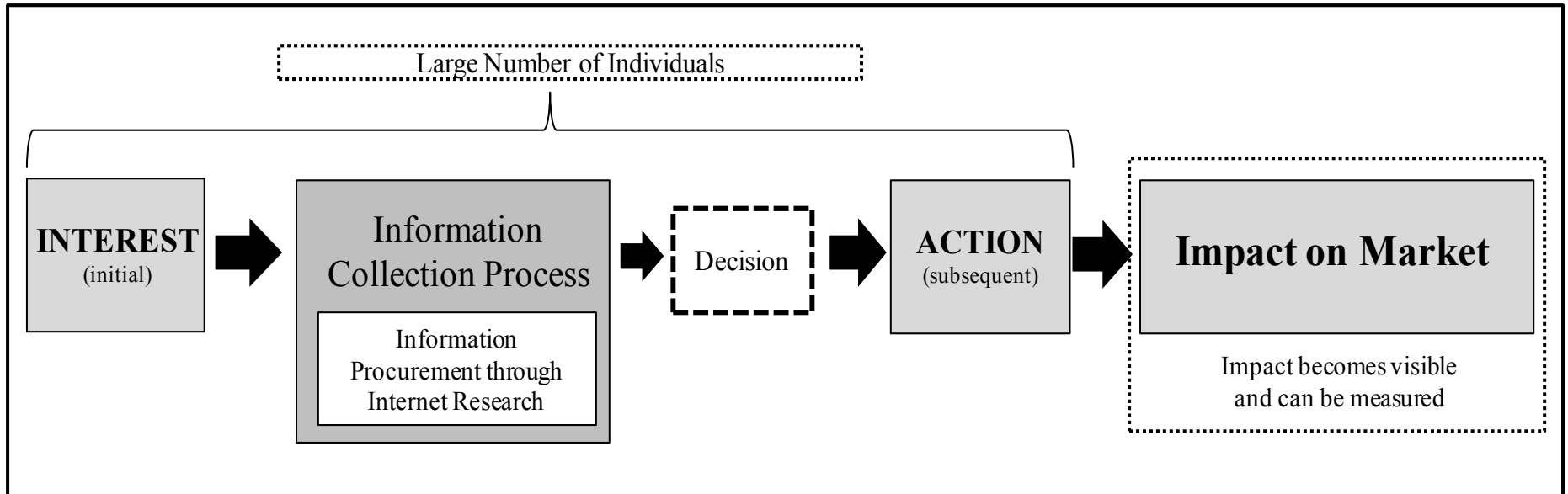## Google searches as a new dataset

- Search Volume Indices (SVI) derived from Google Trends (http://www.google.com/trends/)

- Normalized values, scaled measured between 0 and 100

- The weekly data covers search queries conducted from Sunday to Saturday.

- Google Trends makes the newest weekly data available with an approximate two day delay.
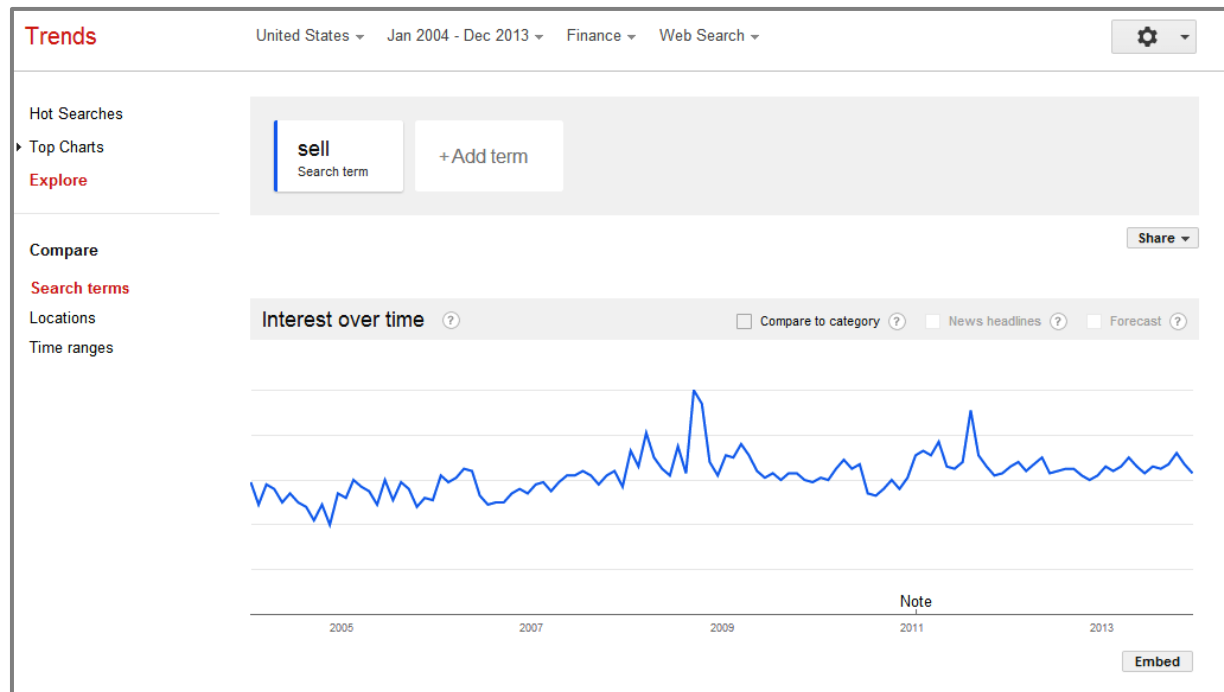
Motivation and Theoretical Background

Google search interest as a new dataset

- Every **(free) Market** is really only **influenced** by its **Participants**

- Every **Action** starts with **Interest** and is (usually) followed by an **Information Collection Process**

- The subsequent **Action** is what actually impacts on the **Market** and **can be measured**

Large Number of Individuals

| INTEREST (initial) | → | Information Collection Process

Information Procurement through Internet Research | → | Decision | → | ACTION (subsequent) | → | Impact on Market

Impact becomes visible and can be measured |

Motivation and Theoretical Background

Motivation and Research Question

Tsolacos, S. (2012), "**The role of sentiment indicators for real estate market forecasting**", Journal of European Real Estate Research, Vol. 5, No. 2, pp. 109-20.

▶ **Shows that the behavior of real estate markets can be predicted with the help of sentiment indicators.**

▶ **Uses probability models i.e. Probit models**

Can **probit-models** based on search volume data predict whether the housing **market is going to rise or fall**?

Preliminary Analysis

Graphical Inspection in annual differences

Legend: Case-Shiller House Price Index — Google Real Estate Category
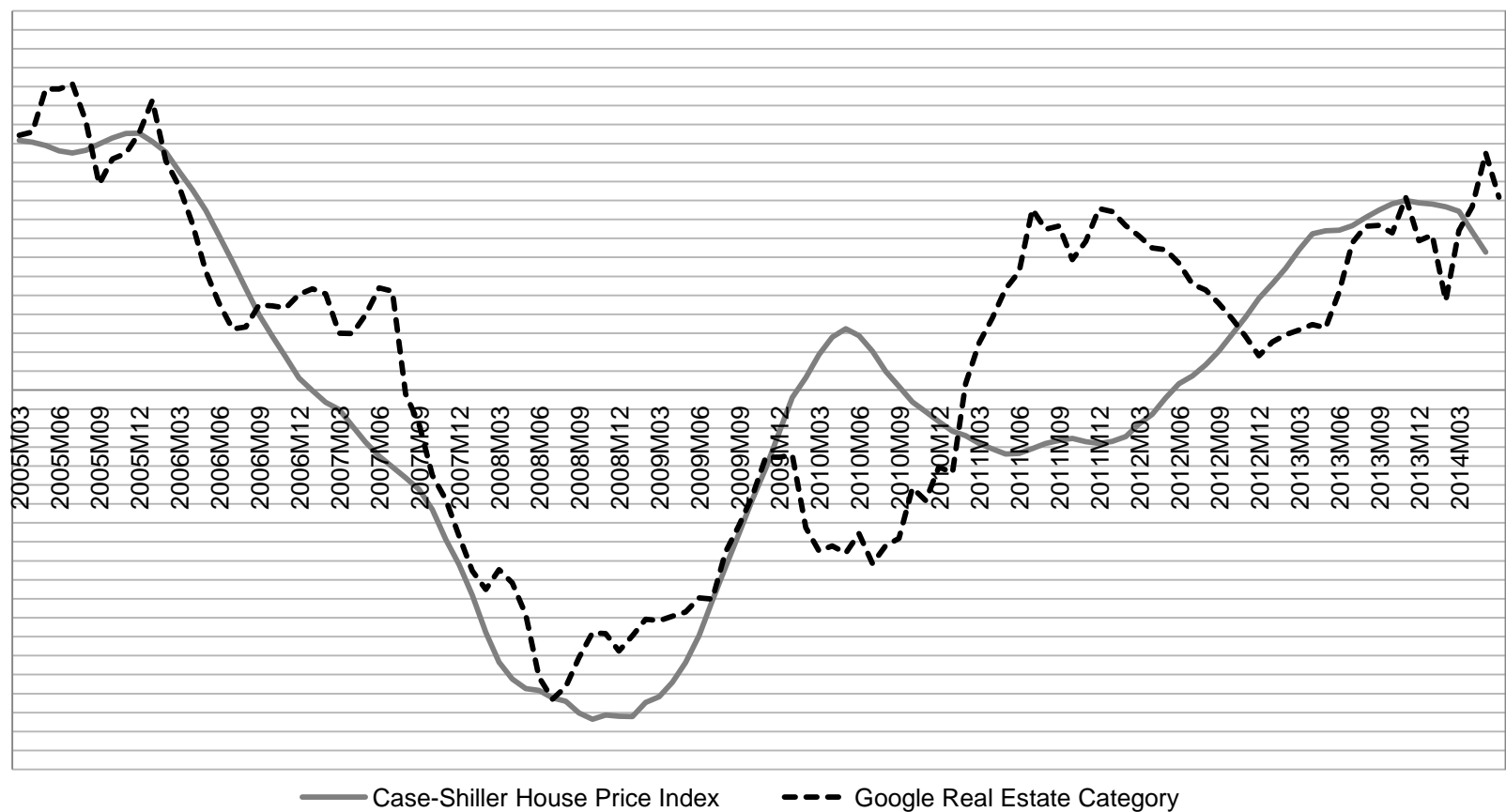
Research Design and Methodology

Data

Housing Data:

Case-Shiller 20-City House Price Index

Google Data:

Search Volume Indices (SVI) derived from Google Trends (http://www.google.com/trends/):

Google search indices

Search Volume index	Regression Label	Category	Subcategory	Topic	individual search terms
Real Estate Category	G_RE	Real Estate		-	-
Property Inspections & Appraisals Subcategory	G_APR	Real Estate	Property Inspections & Appraisals	-	-
Real Estate Agencies Subcategory	G_AG	Real Estate	Real Estate Agencies	-	-
Real Estate Listings Subcategory	G_LIST	Real Estate	Real Estate Listings	-	-
Keywords Housing Market	G_K_HM	*no category filter*			housing market+real estate market+real estate trends
Construction	G_CONS	Business & Industrial	Construction & Maintenance	-	-
Home (Building function)	G_HOM	Real Estate	-	Topic: Building Function	-

Research Design and Methodology

Optimum lag lengths

Search Volume Index	Regression Label	Lag	z-stat	p-value	AIC	Relationship
Real Estate Category	G_RE	k=10	2.690	0.009	1.356	-
Property Inspections & Appraisals Subcategory	G_APR	k=12	4.681	0.000	1.149	-
Real Estate Agencies Subcategory	G_AG	k=11	5.533	0.000	1.087	-
Real Estate Listings Subcategory	G_LIST	k=4	-3.126	0.002	1.326	+
Keywords Housing Market	G_K_HM	k=1	3.962	0.000	1.253	-
Construction	G_CONS	k=9	5.251	0.000	1.123	-
Home (Building function)	G_HOM	k=4	-3.536	0.001	1.301	+

Research Design and Methodology

Selection Process: 1) **Start with given univariate model with a determined lag order**

2) **Every additional variable has to**

      a) **decrease the AIC (Akaike Information Criterion) and**

      b) **all variables have to remain statistically significant**

3) **The best performing model is chosen by the lowest AIC**

Variable	Coefficient	z-Statistic	Prob.
Constant	0,788	2,999	0,003
G_APR(-12)	52,950	3,741	0,000
G_LIST(-4)	-105,016	-4,371	0,000
G_CONS(-9)	147,380	5,253	0,000

McFadden R-squared	0,659
Akaike info criterion	0,544
Prob(LR statistic)	0,000
Observations	112

Sample period: 2005M03 2014M06

Empirical Results

Expectation-prediction evaluation

Cut off point (C) = 0.5	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
Probability (Dep=1)<=C	51	6	57	57	55	112
Probability (Dep=1)>C	6	49	55	0	0	0
Total	57	55	112	57	55	112
Correct	51	49	100	57	0	57
% Correct	**89.47**	**89.09**	**89.29**	100	0	**50.89**
% Incorrect	10.53	10.91	10.71	0	100	49.11
Total Gain*	**-10.53**	**89.09**	**38.39**			
Percent Gain**	NA	89.09	78.18			

*Change in "% Correct" from default (constant probability) specification

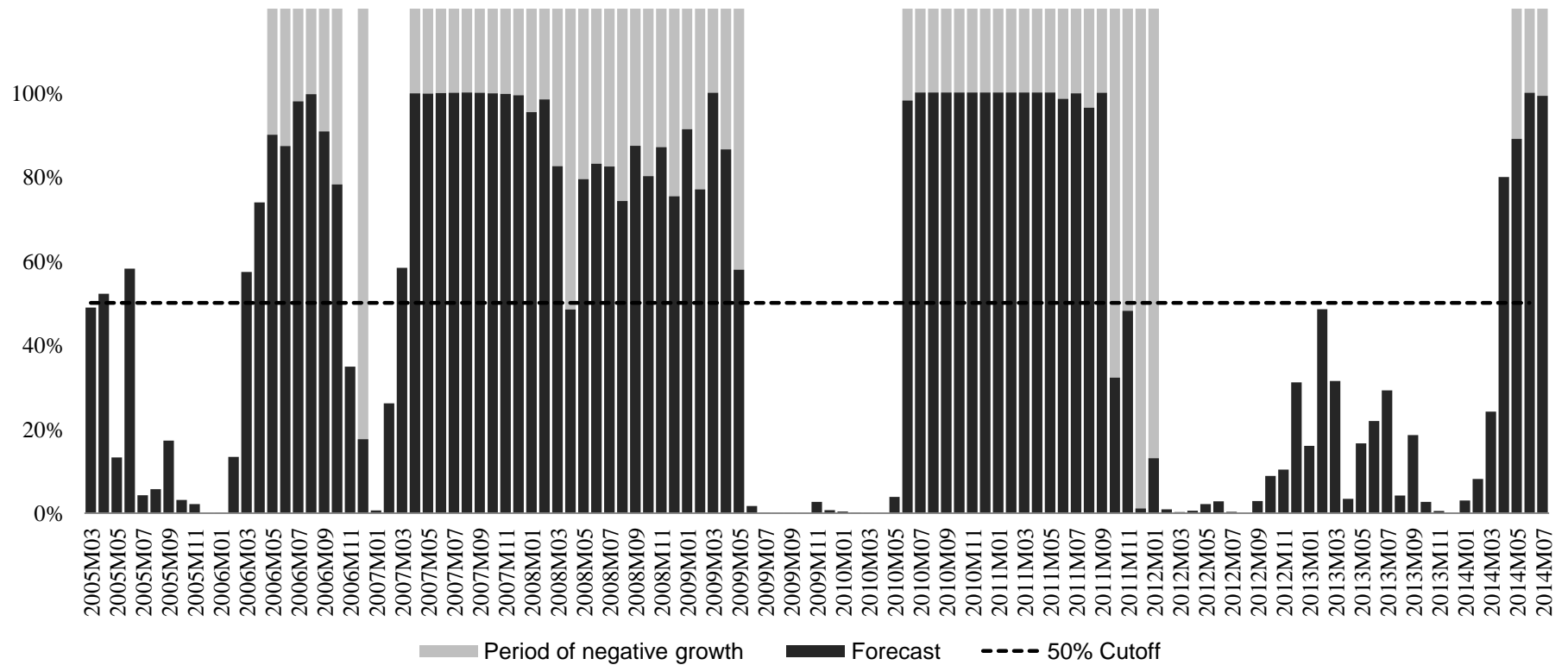**Percent of incorrect (default) prediction corrected by equation

Empirical Results

In-sample Forecast

In-sample Forecast
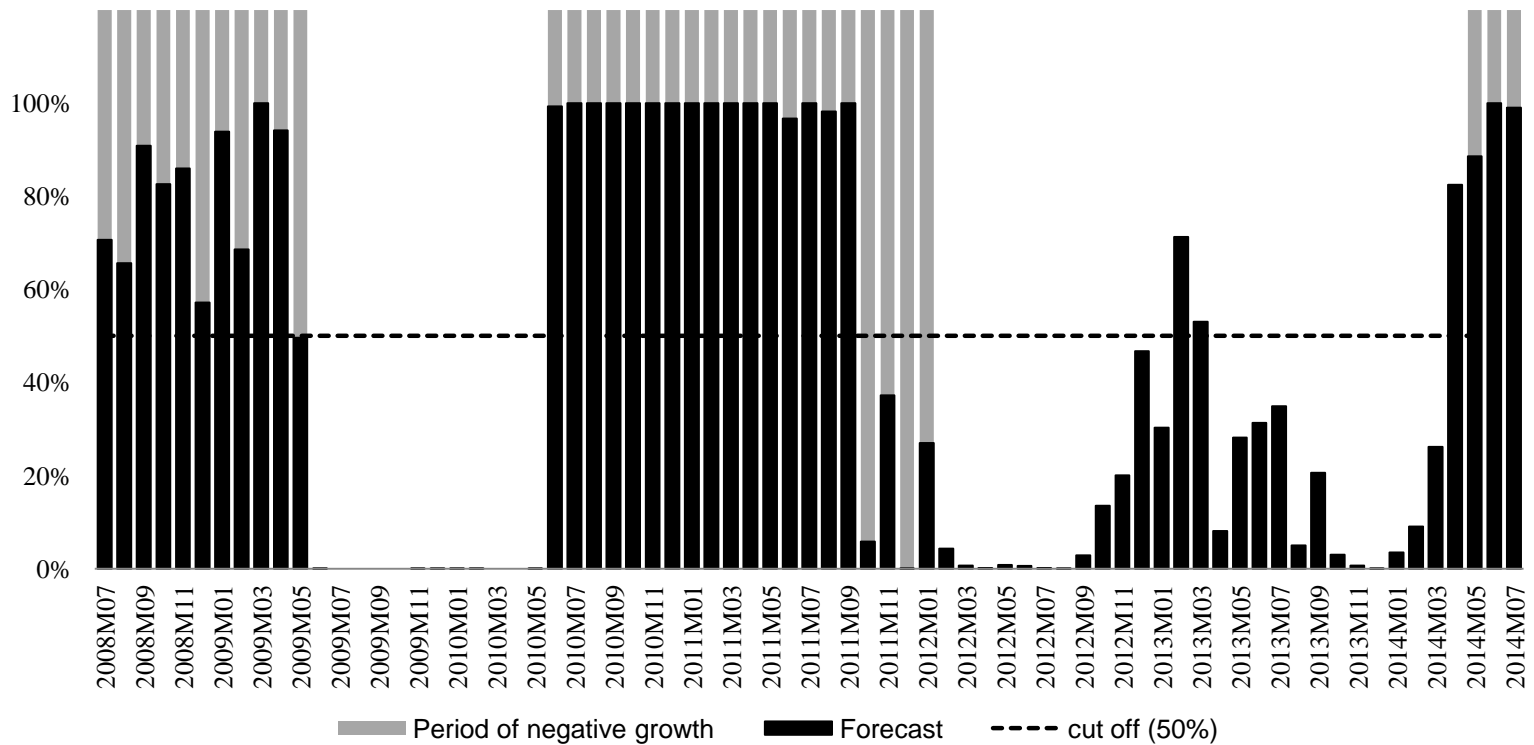
Prediction accuracy: 89.3 %
Mean Squared Error (MSE): 7.0 %

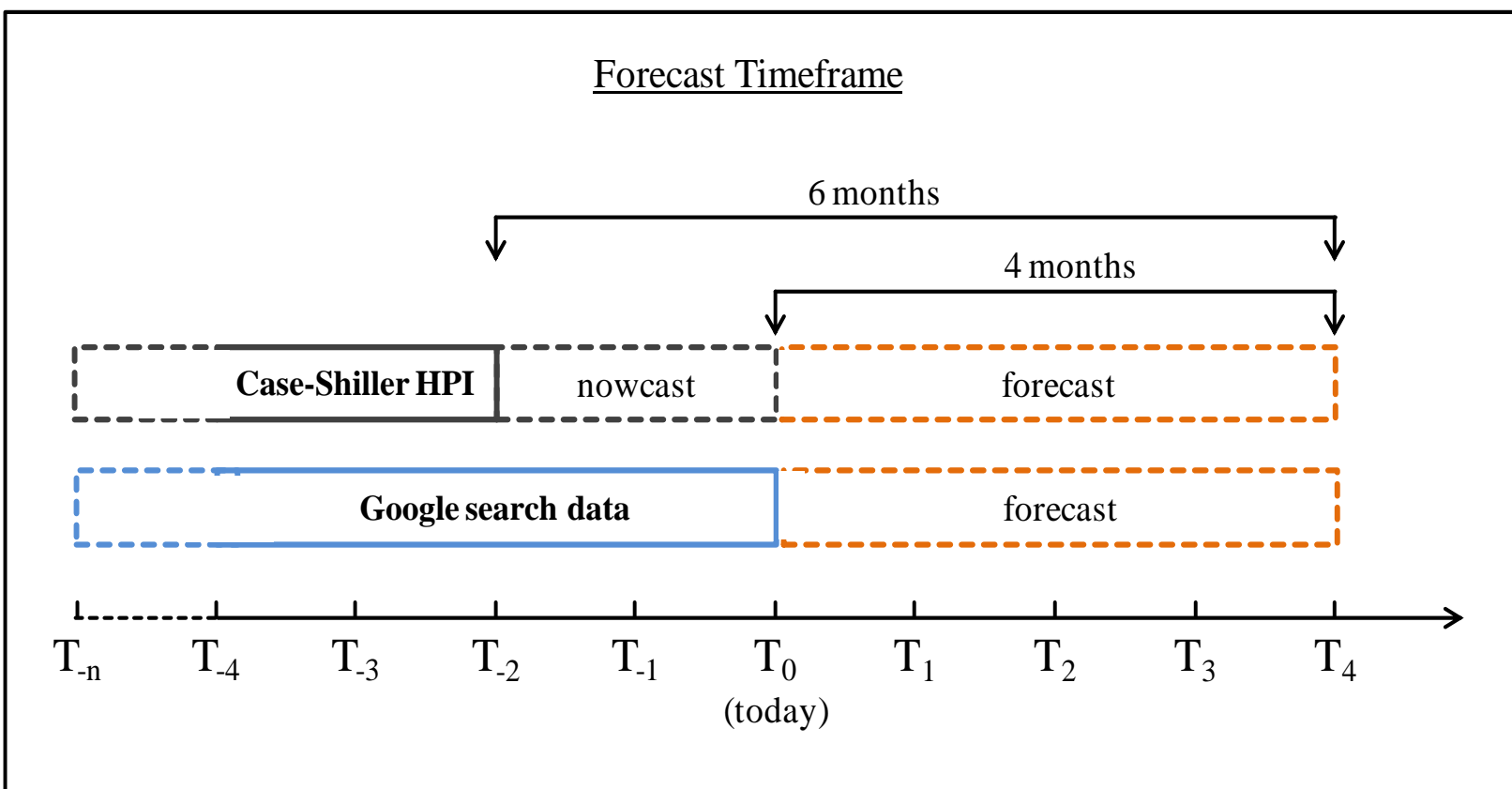Legend: Period of negative growth — Forecast — — — 50% Cutoff

12

Empirical Results

Out-of-sample Forecast (one month ahead)

Prediction accuracy: 88.9 %
Mean Squared Error (MSE): 8.2 %

Empirical Results

Conclusion

Main Findings

- **Google** data would have reliably **predicted turning points** in the **housing market** (this includes the bust of the housing bubble in 2006)

- The **probit prediction** models predict correctly in **about 90 %** of the cases

- The presented model could be **applied in practice** as **Google data** are available with a **time delay** of **only two days**

Questions

Sentiment-Based Commercial Real Estate Forecasting with Google Search Volume Data

Thank you for your attention! Remarks?