Diskussionspapierreihe
Working Paper Series

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

# DETERMINANTS OF HOUSE PRICE DYNAMICS. WHAT CAN WE LEARN FROM SEARCH ENGINE DATA?

LARS BENNÖHR
MARCO OESTMANN

No. 153
OCTOBER 2014

Department of Economics
Fächergruppe Volkswirtschaftslehre

Autoren / Authors

**Lars Bennöhr**
Helmut Schmidt Universität Hamburg
Department of Economics
Holstenhofweg 85, 22043 Hamburg
Germany

**Marco Oestmann (corresponding author)**
Helmut Schmidt Universität Hamburg
Department of Economics
Holstenhofweg 85, 22043 Hamburg
Germany
oestmann@hsu-hh.de

Redaktion / Editors

Helmut Schmidt Universität Hamburg / Helmut Schmidt University Hamburg
Fächergruppe Volkswirtschaftslehre / Department of Economics

Eine elektronische Version des Diskussionspapiers ist auf folgender Internetseite zu finden/
An electronic version of the paper may be downloaded from the homepage:
http://fgvwl.hsu-hh.de/wp-vwl

Koordinator / Coordinator

Klaus B. Beckmann
wp-vwl@hsu-hh.de

Helmut Schmidt Universität Hamburg / Helmut Schmidt University Hamburg
Fächergruppe Volkswirtschaftslehre / Department of Economics

Diskussionspapier Nr. 153
Working Paper No. 153

# Determinants of house price dynamics. What can we learn from search engine data?

LARS BENNÖHR

MARCO OESTMANN

**Zusammenfassung/ Abstract**

There is a broad literature about fundamental determinants of house prices, which received increasing attention in the aftermath of the subprime crisis. However, there might be several other partly unobservable socio-demographic, psychological or individual factors influencing real estate price dynamics. Using quarterly data, we try to capture such effects by including relevant Google search engine query information into a set of standard fundamental variables. We perform fixed-effects regressions for a panel of 14 EU-countries comprising the years 2005-2013. As dependent variable the house price index (HPI) from Eurostat is employed. We find that Google data as a single aggregate measure of unobserved variables plays a substantial role in explaining house price developments.

**Danksagungen**

Financial support from the Fritz Thyssen Foundation is gratefully acknowledged. We would like to thank Christian Pierdzioch, Michael Berlemann, Max Steinhardt and Claudia Buch for helpful comments. Egle Wahl and Karol Sewielski provided excellent research assistance.

**JEL-Klassifikation / JEL-Classification**: C23, C82, R21, R31

**Schlagworte /Keywords:** Google Trends, House Price Index, Real Estate, Google Econometrics

# 1 Introduction

In recent years, a vast number of studies have been conducted to explain the dynamics on real estate markets. Although some relevant fundamental factors driving house prices have been identified in the past decades, the recent worldwide financial crisis, triggered by a collapse of the US-House price bubble, showed in an impressive way that Economists seem to be quite far away from a complete understanding of the price determination process on real estate markets. The standard models incorporating the core set of these consensus fundamental variables, both supply and demand side, which are assumed to influence aggregate house price dynamics, (e.g. interest rates, economic activity, population size and distribution or inflation) especially fail regularly to explain the evolvement of house price inflation and other observed price anomalies on real estate markets (Hohenstatt et al. (2011)). Thus, because of the outstanding importance of the real estate sector for the whole economy, models better explaining recent and hopefully future price dynamics are urgently needed. With this paper we want to take a step into this direction.

Beside the mentioned traditional (macro-)indicators psychological ("investor sentiment"), socio-demographic or individual factors (birth of a child, marriage, migration) exist, which obviously play a role in the house/property buying process and thus could have an impact on prices. These variables are mostly very hard to measure or even unobservable. Exploitable information often can only be retrieved at high cost (e.g. using surveys) with many problems such as limited reliability and considerable time lags. As an easy way out, we propose to use search engine data, i.e. a single search frequency index obtained from Google Trends (www.google.com/trends/) to incorporate the aggregate influence of these variables. In our paper we mainly address the question if the integration of such kind of search data contributes to a better understanding of house price dynamics by increasing the explanatory power of conventional empirical models. Using the house price index (HPI) from Eurostat we are able to show for a sample of 14 EU-countries for the years 2005-2013 that the additional integration of Google data from a predefined search category ("Real Estate Agencies") can improve the explanatory power in nearly all specifications significantly. Thus, Google search intensities might help mitigating the problem of insufficient data availability modeling the house buying and price determination process.

Our results suggest also that Google may be used as a leading indicator for house prices. In a second step this finding may be used to make better forecasts, but this analysis is beyond the scope of our paper. Our aim is to analyze and elaborate conditions under which Google indicators, i.e. the related search intensities may improve existing models and better explain observed house price developments.

The underlying idea is simple, convincing and well established in the literature. We use revealed internet search activity as reliable signal for future real world economic transactions (Choi and Varian (2009)). Applied to our real estate context, we imply that potential buyers of real estate property inform themselves via the Internet before they show their demand on the market and finally acquire a dwelling if matching has been successful. The decision of buying a house or to be more general property, is the result of many different individual subordinate decisions of households and strongly influenced by sentiment and social (peer group) norms, informational cascades and herd behavior. Google search indicators should be able to capture all these rational and irrational factors quite well in real time without any delay. Of course, some search queries may be mainly induced by hard objective rational facts (such as low mortgage rates for example) and thus reflect classical rational decision making. But since observed search activities seem to be always motivated by a mixture of all variables and factors playing a role in the context of decision making, they surely are also influenced by (and reflect) other quite irrational motives such as "sentiment". Recording and categorizing billions of these search queries over time and geographic regions, Google Trends aggregates signals of decision-makers' intentions and thus measures at least partly an overall level of "investor sentiment" (Wu and Brynjolfsson (2013)).

All in all, our idea is not to model these individual decision processes and find variables via surveys on an individual level or proxies on macro level. We only look on real observed behavior, the clicks for real estate related search terms recorded and categorized by Google, the leading search engine worldwide[1] and assume that expressed preferences via clicks in fact reveal underlying economic intentions and activities.

Thus, for us, processing a search query is a "honest signal" for actual interest (Wu and Brynjolfsson (2013), Pentland (2010)). Since search activities of individuals are time consuming and associated with considerable opportunity cost, they should at least in most cases reveal serious intentions and preferences. Usually, they are also conducted in an anonymous way (in the sense that nobody can see what searches are processed without criminal effort) and should therefore be largely unbiased by bargaining, gaming, strategic signaling or other distorting factors. This allows us to make viable inferences from Google

[1] According to publicly available assessments, for most European countries included in the paper the market share of Google lies well above 90%, e.g. Germany, UK, France, Spain, Italy (Haucap and Kehder (2013); Greenlight Digital (2010); Edelman (2011)).

search intensities concerning subsequent real world economic transactions which show up presumably with a certain delay in prices.

This is especially true in a real estate context, because the decision to acquire property for the majority of households is the most important financial decision in lifetime. Consequently, intense search activities nearly always precede the purchase decision and also accompany the acquiring process. Recent studies and surveys (see for example NAR (2012, 2013); NAR and Google (2012); OFT (2012) or Hess (2011, 2012))[2] show that nowadays such activities are mostly carried out using multiple channels, with the Internet taking a prominent and dominating position. According to the National Association of Realtors (NAR) 90.0% of home buyers in the US used the Internet to search for a home in 2012. In Germany 97.5% of people who search for real estate use online property marketing portals like “Immonet” (www.immonet.de) or “Immobilienscout24” (www.immobilienscout24.de), where private sellers and professional agents can present their real estate offers (Hess (2011, 2012)). For the UK, the Office of Fair Trading (OFT) estimates, that about 40.0% of estate agents' sales are originating from such portals in 2010. It finds that additional to most buyers, who use these portals as a starting point when searching also nearly all agents (92.0%) use property portals 'very often' (OFT (2010)). Obviously, the Internet provides easier access to information for both buyers and sellers and reduces transaction costs, but in many cases this information is not that easy to retrieve and cannot be found directly. In such cases, usually Google is used as a standard search vehicle. We assume that these search engine related activities primarily mirror changes in market demand. Changes in market demand in turn should have a large impact on prices, especially when supply is rather sluggish as it is normally the case on real estate markets. In our paper we try to find empirical evidence for this conjecture.

Our paper is organized as follows. Section 2 provides a brief overview on recent research using search engine data. At this, we put special emphasis on related papers covering the use of Google indicators in a real estate context. In section 3 we discuss some standard variables in the literature of house price determinants and inspect the possible connection between house price dynamics and search activities in more detail to motivate our research strategy. In the following section 4 we introduce the dataset and present some basic descriptive statistics with special emphasis on the HPI and the relevant Google indicator. In

2 For Germany, Hess (2011, 2012) analyzes trends in digital media use. The OFT (2010) provides a comprehensive market study for the UK. The NAR (2012, 2013) delivers recent information about the buyer and seller characteristics on the US market, whereas NAR and Google (2012) highlight for the US the digital media use in the consumer home search process.

section 5 the econometric methodology is characterized and our baseline results for the whole country sample are presented. Section 6 looks at the relation between HPI and the Google indicator in more detail and introduces some variations of our baseline regressions in order to check the robustness of our results. At first, we split the sample into two groups of countries according to the level of internet use and access. Additionally, we investigate two time horizons, a “pre-crisis period” until the collapse of Lehmann Brothers Inc. ranging from Q1_2005 until Q3_2008 and a following “crisis-period” ranging to Q1_2013. Section 7 concludes and discusses some implications for further research.

## 2 A brief review of the literature

Since the introduction of the first public accessible analysis tool Google Trends in 2006 and the presentation of the even more user friendly extension Google Insights for search (which was 2012 merged with Google Trends) in 2008, research with search query data has well established in the academic literature. Many studies in various different fields of interest have been published in the meantime, a considerable number on economic topics.

In the following we present a small overview on the literature using search engine data. Since Google has by far the largest market share and there is to our knowledge no competitive alternative to Google tools, all of the following studies were carried out using Google’s database. We mainly concentrate on economic papers and lay special emphasis on studies using Google information in real estate research. More references can be found in Choi and Varian (2012), a more detailed description of most of the mentioned papers is provided in Hohenstatt et al. (2011).

One of the first and most popular academic paper using Google search data was published by Ginsberg et al. (2009). In this work search queries related to most common flu symptoms are analyzed and used to identify and track illness hotspots and epidemics in the US.

Of Course, Google data have also been applied to various other problems and topics. Concerning economic issues, most research has been carried out in the field of (forecasting) private consumption and unemployment. Beside other topics, for example Guzman (2011), who examines Google data as a predictor of inflation, housing markets received also some attention.

Concerning consumption, the strand of the literature mainly deals with the construction of consumer sentiment indicators using Google search data to better forecast private consumption. Schmidt and Vosen (2011) derive a Google indicator based on 56 consumption categories according to the national income and product accounts coming from the U.S. Bureau of Economic Analysis (BEA). They compare the forecasting performance of the new indicator with traditional survey based indices like the University of Michigan's Index of Consumer Sentiment and the Conference Board's Consumer Confidence Index. Their finding is that the Google indicator outperforms traditional survey based indicators. Kholodilin et al. (2010) forecast private consumption using a Google indicator and comparing it with the properties of the OECD consumer confidence indicator. They find that the Google indicator performs better in episodes of unusual or extreme economic activity.

There is also a broad literature employing Google indicator in labor market research. To simplify, the main idea here is that people fearing unemployment or have just become unemployed, will search the Internet for information on benefit systems or new jobs. So monitoring the relevant search terms and queries may be a useful indicator and predictor of unemployment. Choi and Varian (2009) try to explain initial jobless benefit claims by using unemployment and welfare related internet searches. They find that using Google data can improve predictions and forecast accuracy. Other studies finding similar results have been conducted for Germany, Italy and Israel (Askitas and Zimmermann (2009); D'Amuri (2009); Suhoy (2009)).

Because it tackles some interesting questions and received certain attention from the media, we would finally like to mention the recently published dissertation from Seth-Davidowitz (2013) covering topics from Political Economy and Public Health. He uses, for example Google data to measure racial animus against African-Americans. In a second paper he tries to predict election turnouts with Help of Google. These predictions prove stronger than other available indicators.

Concerning housing markets, we are only aware of five papers that at least partly analyze real estate market issues with Google data. Probably, the initial work in this field of research comes from Choi and Varian (2012). They conclude that real estate related searches can improve nowcasts for house sales in the US.

Wu and Brynjolfsson (2013) explain and predict US home sales and house prices on state level with help of a seasonal AR-models incorporating both variables and a Google indicator. As Google measure they use the predefined categories "Real Estate Agencies" and "real estate listings". To account for any time-invariant influences population, as well as regional and time fixed effects are employed as controls. Prediction power of different models with and without Google indicators is investigated using the mean absolute error (MAE). The resulting Google augmented model outperforms forecasts of quarterly housing sales of the National association of realtors (NAR). For sales the volume of sales from existing single family housing units from NAR on US-State level from Q1_2006 until Q3_2011 is used. The state level house price index on a quarterly basis comes from the Federal Housing Finance Agency (www.fhfa.gov).

Wu and Brynjolfsson (2013) find that online search frequencies can improve the accuracy of prediction for present and future sales, though including Google data is more effective for predicting the future. They conclude that todays' search activities can be useful for predicting future housing indicators. Search frequency data are more effective for predicting sales volume than for predicting state level house price indices. According to them, this is partly because supply and demand shifts which influence home price shifts cannot be exactly identified by Google indicators.

Hohenstatt et al. (2011) provide a comprehensive analysis of the US housing market for 20 Metropolitan Statistical Areas (MSA) with monthly data ranging from M2_2004 until M4_2009. They use a VAR-Analysis to address occurring endogeneity issues, modelling house prices and transactions as well as two Google indicators "Rreal Estate Agencies" and the single search query "Apartments" as endogenous variables. Search frequencies from the Google category "home financing", employment, income and the S&P500 index are used to augment the model with exogenous (macro) data, introduced to account for overall market conditions. In order to motivate the structure of their VAR, they undertake a thorough Granger causality analysis, experimenting with different Google categories and single search terms.

As data input for their analysis the S&P's Case-Shiller Index Composite for 20 MSAs is used to model real estate prices. Sales are measured with the unadjusted series of existing home sales for single family and condominium from the NAR. The market index S&P's 500 composite comes from Datastream (S&PCOMP), as well as disposable income (USPERDISB) and total employment (USEMPTOTO).

Hohenstatt et al. (2011) find that Google data improve the quality of explaining house prices, but the impact of the lagged variables is not clearly directed which is attributed to the "extreme market environment" by the authors. The Google category "Real Estate Agencies" serves as a very good predictor of transactions and assuming an effect of transactions on house prices has also implications for the overall housing market. The disadvantage of informational time lags can be at least partly mitigated by using real-time search query data. Finally, they find evidence that housing market dynamics influence search query data, which in turn influence the real world.

McLaren and Shanbhogue (2011) try to explain the house price development in the UK on country level with monthly data for the period M5_2004 to M3_2011. They use a simple AR-model regressing monthly house price change coming from Halifax and Nationwide on lagged endogenous variables. They extend this baseline model with the Google search indicator and two other house price indicators from the Home Builders Federation (HBF) and the Royal Institution of Chartered Surveyors (RICS). As Google indicator the single search term “estate agents” was chosen.

McLaren and Shanbhogue (2011) find that incorporating the Google indicator into the baseline model significantly boosts the information content and explanatory power of the model. These results are supported by the out-of-sample one month ahead nowcast tests using the RMSE criterion. Here the search term variable outperforms the existing House price indicators from HBF and RICS. They conclude that Google search data can improve the understanding of the current state of the housing market.

Finally Webb (2009) finds that searches for foreclosure are highly correlated with actual US home foreclosures. He suggests to employ search trends as an early warning indicator for potential problems on the US Housing market. Since this topic lies not exactly in our focus we do not want to present more details here.

Summing up, at first we can conclude that all related studies concerning real estate markets have been carried out on national level using US or UK data. Furthermore, all studies obviously focus (at least partly) on forecasting or nowcasting issues. Our paper in contrast is to our knowledge the first one, which examines in detail the connection between Google search intensities and real estate price dynamics in a multinational and European context. Involving 14 developed, but still quite different European countries, we try to shed some light on the Google-House price-nexus in a completely new regional and methodological setting. This allows us finally to better assess the information content of real estate internet search activities for the corresponding development of prices and augment the picture drawn for the US.

# 3 Google indicators and classical determinants of house prices

Following classical theoretical studies (for a selection of first generation models see for example Muth (1960), Huang (1966), and Smith (1969). Famous second stage models comprise Kearl (1979), Buckley and Ermisch (1982), Dougherty and Van Order (1982), or Poterba (1984)) and taking a look on recent empirical studies, mostly based on these considerations, (see among the broad and fast growing literature beside many others for example Kajuth et al. (2013) for Germany or Gattini and Hiebert (2010) for the euro area)[3], surveys or overview articles on house price determination (see e.g. Hilbers et al. (2008) or Girouard et al. (2006))[4] we are able to identify a set of variables that seem to be "fundamental", i.e. theoretically well motivated and regularly found significant in (macro-) econometric models explaining real estate price dynamics. Usually, models including these variables provide a "fundamental" or basic benchmark evaluating and classifying recent price developments. On the demand side, most prominent indicators employed in such models are (real) interest rates, (real per capita) disposable income and demographic changes (like population growth or number and size of households). On the supply side we often find variables like housing stock, housing investment, or building permits et cetera. Additionally to this core set of variables many other indicators like vacancy rates, construction costs, taxes, unemployment measures, or variables capturing the conditions on the rental market respectively the cost of external finance are used occasionally to assess and explain price movements on the housing markets. A detailed discussion of relevant indicators and variables can be found e.g. in Hilbers et al. (2008) and Girouard et al. (2006).

In order to check whether Google search frequencies possess information concerning house prices, we proceed as follows. In a first step, we build a baseline explanation model including

---

3 Gattini and Hiebert (2010) emphasize the role of fundamentals. They build a quarterly vector error correction model which is estimated over 1970-2009 using supply and demand forces, i.e. housing investment, real disposable income per capita and a mixed maturity measure of the real interest rate.

4 Focusing on OECD countries, Girouard et al. (2006) provide a comprehensive overview on theoretical foundations and the role of fundamentals in empirical studies. Hilbers et al. (2008) analyze house price developments from a European perspective. They identify and discuss a broad range of indicators and factors describing and measuring (a) general housing market conditions and trends, (b) demand, (c) supply, (d) the rental market, (e) taxation and finally (f) the financial sector as a whole and thus (could) influence prices on the housing market. Additionally, they estimate an own empirical model incorporating user costs, demographic pressures, and per capita income.

some of these conventional standard indicators, which should have an impact on house price development. In a second step we augment this baseline model with our Google indicator, i.e. search intensities from Google Trends and check whether we gain explanatory power. In a third step, we consider some alternative model specifications to test for robustness of our results.

Before we take a closer look on our set of potential indicators modeling real estate prices in chapter 4 we want to point out why and how Google search data is well suited to extend this set of explanatory factors, capturing rather hard to observe measureable factors. We examine the relationship between Google search queries and their involvement in the property acquiring process in more detail to assess their possible impact on house prices and deepening our preliminary ideas presented in the motivation in a structured way.

The starting point of our considerations is the home buying process (HBP). Although the process differs between countries in detail, there are some similarities which are characterized in the following. In general, the process starts with the decision to search or to offer (existing or newly built) property. The next or sometimes concurrent step usually is to gather information (often via Internet) to get an idea about prices of relevant property for sale. Subsequently, some sellers decide to sell on their own without any assistance, the majority in contrast contacts immediately or sometimes later a real estate agent who markets the property and provides assistance during the selling process. On the demand side, some potential buyers contact a real estate agent, too. The rest organizes their search activities privately. During the following matching process, all parties, real estate agents but also private sellers and buyers nowadays mainly use real estate internet marketplaces which have become a very important and popular platform to facilitate the transaction process in most countries. If potential buyers find an interesting offer, they normally inspect the home of interest. If expectations are met, price negotiations follow. In most cases these activities have to be repeated several times until seller and buyer finally come together. Then usually some legal and official transactions take place: the contract has to be signed, the solicitor comes into play, taxes have to be paid, the new owner has to register the land and property transfer.

Of course, this process is accompanied by several other activities related to the home buying process. Just to mention a few: gathering information about financing conditions, contacting a bank or a mortgage lender, searching for a solicitor or lawyer, obtaining information about general or special aspects of the HBP, making appointments, conducting marketing activities, contacting sellers, buyers, lawyers, and especially real estate agents and so on. In general, the Internet can be involved in nearly every activity related to the matching process.

And every time the Internet is used at least some people use Google. These search queries then are categorized by Google.

But how can we relate these recorded and observable search activities with price movements on the real estate market? At first, as on every market prices and transaction volumes are the results from demand meeting supply. Every time when demand for property meets adequate supply, i.e. the matching process is successful, we observe prices and the transaction is officially recorded. We assume that the demand and supply side are driven by several macro-variables like interest rates, economic activity, inflation, and so on, but also by unobservable individual or other factors, covering e.g. personal or psychological issues, social developments or just rather irrational sentiment, as mentioned above. These indicators also exert influence on all activities involved with the home buying process which can be considered, at least partly, as kind of matching process after formation of demand and supply. Generally speaking, when demand or supply are increasing, for example due to macroeconomic shocks, we expect that activities in the matching process also increase.

We argued already above that all of these activities can be separated in internet related and non-internet activities. The Internet related activities in turn can be divided in search activities using Google and other internet activities. Although presumably most of these internet activities do not involve Google, we assume that a sufficiently large number utilize Google to find the relevant information. Considering the activities in the matching process, it seems appropriate to assume that most of the recorded searches can be attributed to the demand side (McLaren and Shanbhogue (2011)). As a consequence increasing search frequencies can be primarily interpreted as an increase in demand relative to supply and thus serve as an indicator for increasing prices in the (near) future.

Every search query is categorized by Google within main categories consisting of several subcategories. It should not be concealed, that there exist at least three problems with this classification. Firstly, and mostly severe, the complete categorization process is a black box to us. Unfortunately, Google only provides very little information about this procedure making it impossible to directly assess the quality. Secondly, classification errors might occur: it is possible that search queries may be wrongly categorized as real estate related or the other way round, queries might not be properly identified as real estate related and thus are falsely assigned. Thirdly, and probably a minor problem, “fun queries” without serious interest may exist, which also can be source of distortion and thus misleading.

Backed with various sophisticated studies using successfully Google search frequency data, we assume that overall the categorization seems to work quite properly and represents no fundamental problem. So, all in all, it seems reasonable to use search frequencies (described in detail in chapter 4) as integrated proxy variable for unobserved indicators and include them as additional determinant explaining house price developments. They, as kind of a melting pot, for sure include information about macroeconomic conditions, but as well, and this is our main point and novelty, information about all the other factors influencing real estate markets. In the following we want to exploit this information. Figure 1 illustrates and summarizes our considerations.

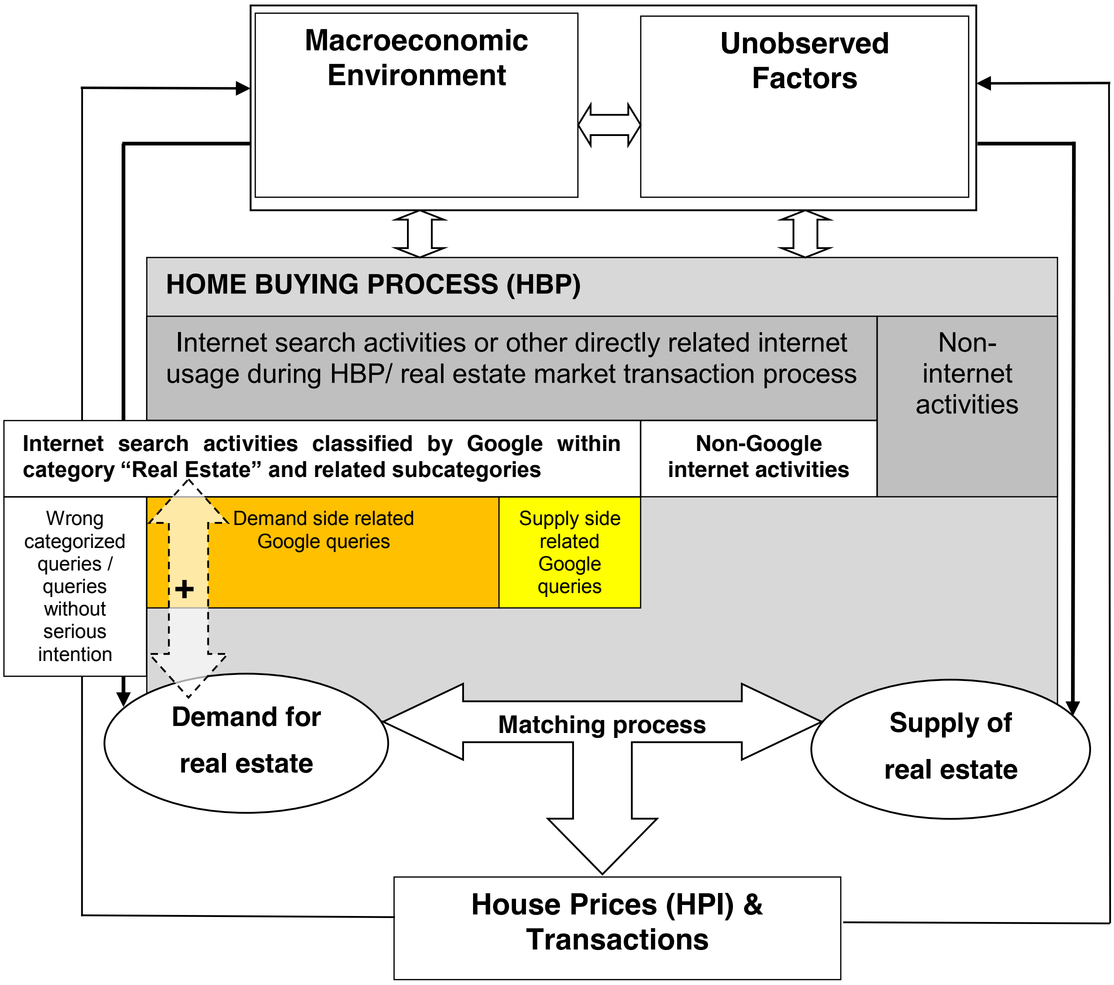


*Figure 1: Determinants of house prices and the home buying process (HBP)*

# 4 Dataset description

Building our empirical benchmark model, we primarily need some of the standard explanatory factors, which were introduced in the previous chapter. Since we want to focus on the additional effect including Google data, we keep our variable set rather small allowing us to incorporate more countries and thus adding more information to our analysis. In the end we decided to employ inflation, interest and unemployment rates as basic explanatory factors, which are all presented in detail in section 4.1. Secondly, serving as dependent variable in the following, we added the HPI from Eurostat which describes the real estate price dynamics (section 4.2). Finally, as additional explanatory factor of special interest, we collected and tested several search intensities from Google Inc. as well. As described in detail in chapter 4.3, we use the search interest for a predefined category ("Real Estate Agencies") as Google indicator in our analysis. The result is an unbalanced panel including 14 European countries (Austria, Belgium, Denmark, France, Germany, Hungary, Ireland, Italy, Netherlands, Portugal, Slovenia, Spain, Sweden, United Kingdom) ranging from the first quarter 2005 until the first quarter 2013. Short summary statistics for the whole dataset are provided in Table 1.

| Variable | Source | Obs. | Mean | Min | Max | Standard Deviation |
|---|---|---|---|---|---|---|
| HPI Change Quarterly | Eurostat | 368 | 0.3397 | -10.00 | 15.10 | 2.5740 |
| Real Estate Agencies | Google | 462 | 91.4583 | 0.00 | 171.08 | 32.1355 |
| HCPI Inflation | Eurostat | 462 | 2.3502 | -2.50 | 7.90 | 1.3206 |
| Short Term Interest Rate | Eurostat | 456 | 2.6069 | 0.16 | 10.49 | 2.0217 |
| Long Term Interest Rate | Eurostat | 462 | 4.2474 | 1.16 | 13.22 | 1.7037 |
| Harmonised Unemployment | Eurostat | 462 | 8.0978 | 3.03 | 26.37 | 3.7345 |

*Table 1: Descriptive Statistics*

## 4.1 Standard explanatory factors

Composing our database, we face some restrictions. Google search frequencies are only available back to the beginning of 2004, the HPI from Eurostat starts in 2005. Due to this time limitation we are forced to use data with quarterly frequency and a quite large country set. This implies that some important drivers of house price development (such as demographic factors or housing stocks) are missing in our analysis because they are not available on a quarterly basis or not existing for a sufficiently large number of countries. We experimented with some additional factors (e.g. GDP or disposable income) but balancing results, data availability, time and frequency restrictions we finally end up with a set of four explanatory variables: inflation, short and long term interest rates, as well as unemployment.

Concerning inflation and house prices usually a positive relationship is assumed. Firstly, inflation might affect house prices because of substitution effects. Secondly, it might also reflect the stance of the economy, i.e. high inflation caused by a small output gap. And finally, another linkage might be the investment in property as an inflation hedge, for example suggested by Wurtzebach et al. (1991).

The effect of short term interest rates on real estate prices in contrast is quite ambiguous. Since short term interest rates are highly influenced by monetary policy, small rates are usually set, when actual production is far away from its potential and/or the risk of high inflation is small. Price changes for houses are a part of overall inflation, so the HPI should be influenced by monetary policy (see e.g. Taylor (2007), Aoki et al. (2004), Goodhart and Hofmann (2008)). Additionally, on the one hand low short term rates could indicate low household incomes and therefore lower upward pressure on real estate prices, measured for example by the HPI. On the other hand low short term rates might reduce the costs of mortgage credits and boost the demand for houses. The latter one holds also true for long term interest rates (Hirata et al. (2013)).

Finally, considering our last explanatory variable, unemployment should reduce the ability of households to finance house purchases as pointed out for example by Hlaváček and Komarek (2009). Typically, unemployment is also negatively correlated with overall level of economic activity. So usually it is assumed, that higher unemployment should lead to a price decrease in the real estate sector.

### 4.2 House price index

In order to measure house price changes we use the house price index (HPI) for residential properties from Eurostat, which is market price based and reflects the price developments of all residential properties purchased by households (flats, detached houses, terraced houses, etc.), both new and existing, independently of their final use and independently of their previous owners. Self-built dwellings are excluded to rule out non-market transactions. The index is quality adjusted.[5] Figure 2 shows the development of real estate prices in our 14 countries measured by the Eurostat HPI as well as the quarterly change of the index.

### 4.3 Google search data

Google provides comprehensive and publicly available search frequencies via its tool Google Trends (http://www.google.com/trends/). Beginning in 2004, it is possible to track single keywords for different geographical locations on a weekly basis. Additionally, every search term entered is recorded and classified by Google Trends into a set of predetermined categories and subcategories based on the potential search results of the search term. If a search term is ambiguous, a proportional attribution to each involved category according to the proportion of search results that relate to that category will be applied by Google.

The main output variable of Google Trends is a normalized search intensity called web search interest over time. This intensity is calculated by dividing the number of Google searches for a certain search term (or in our case the number of searches attributed to a special category) in a certain region and time, divided by the total volume of searches originating from that particular region in the corresponding period. Before made public, the data is normalized by Google. If a single search term is analyzed, the highest intensity in the period is set to 100. If a whole category or subcategory is analyzed, Google reports the weekly percentage change of search intensity with respect to the first week of the analyzed period.

Considering intensities rather than absolute number of searches has several advantages. It allows for example to account for increasing computer usage or increased popularity of Google services. Sometimes Google reports search interest to be zero. Since Google Trends only analyzes popular search terms, this happens normally when the search volume is too low to calculate interest and make it accessible. Furthermore, Google Trends eliminates repeated queries from a certain user over a short period of time.

Because we have a multilateral framework with quite heterogeneous countries and different languages, the use of a single search term or several keywords as Google indicator seems not to be adequate. Thus, for us, categories, which are standardized and permanently maintained by Google, are of major interest since they allow to compare countries abstracting from semantic differences and other distorting country specific factors concerning

[5] For methodological details see the technical report of Eurostat available under the following URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/methodology/hps/house_price_index

the search behavior. In our following analysis we employ the main category “Real Estate” and the corresponding subcategories (Apartments & Residential Rentals, Commercial & Investment Real Estate, Property Development, Property Inspections & Appraisals, Property Management, Real Estate Agencies, Real Estate Listings, Timeshares & Vacation Properties) which are available for all countries ensuring the compatibility of search frequencies over countries and time.[6] Starting in 2004, we compute for all categories a quarterly index from the weekly raw data described above.

Although all subcategories as well as the main category definitely comprise demand and supply motivated searches, there are obviously some groups which are supposed to be more demand oriented and some rubrics which reflect probably more supply side issues. Consequently, we assume that the information content for prices differs substantially between categories, which is impressively confirmed by our findings presented in the next chapter. Drawing on the results of Hohenstatt et al. (2011), we use the index for subcategory "Real Estate Agencies" as Google indicator in the following. It turns out that this subcategory is best suited to explain house prices changes. Figure 3 shows the development of our Google measure over time compared to the main category “Real Estate” for the 14 countries in our sample.

Except for Slovenia, we observe for all countries seasonal fluctuations, which have to be addressed in the econometric analysis. Although seasonal patterns slightly differ between countries, generally search interest is very low in winter (Q4) reaching its downward peak in December. A final remark concerning Slovenia is in order. The disputable search frequencies for the years 2005 and 2006 represent no problem for our analysis because Slovenia enters our regression not before 2007 due to missing HPI data.

[6] Because of the differences between American and British English (BE), the category names slightly differ when the language of Google Trends is set to „English (UK)“. The main category for example is labeled „Property" in BE.

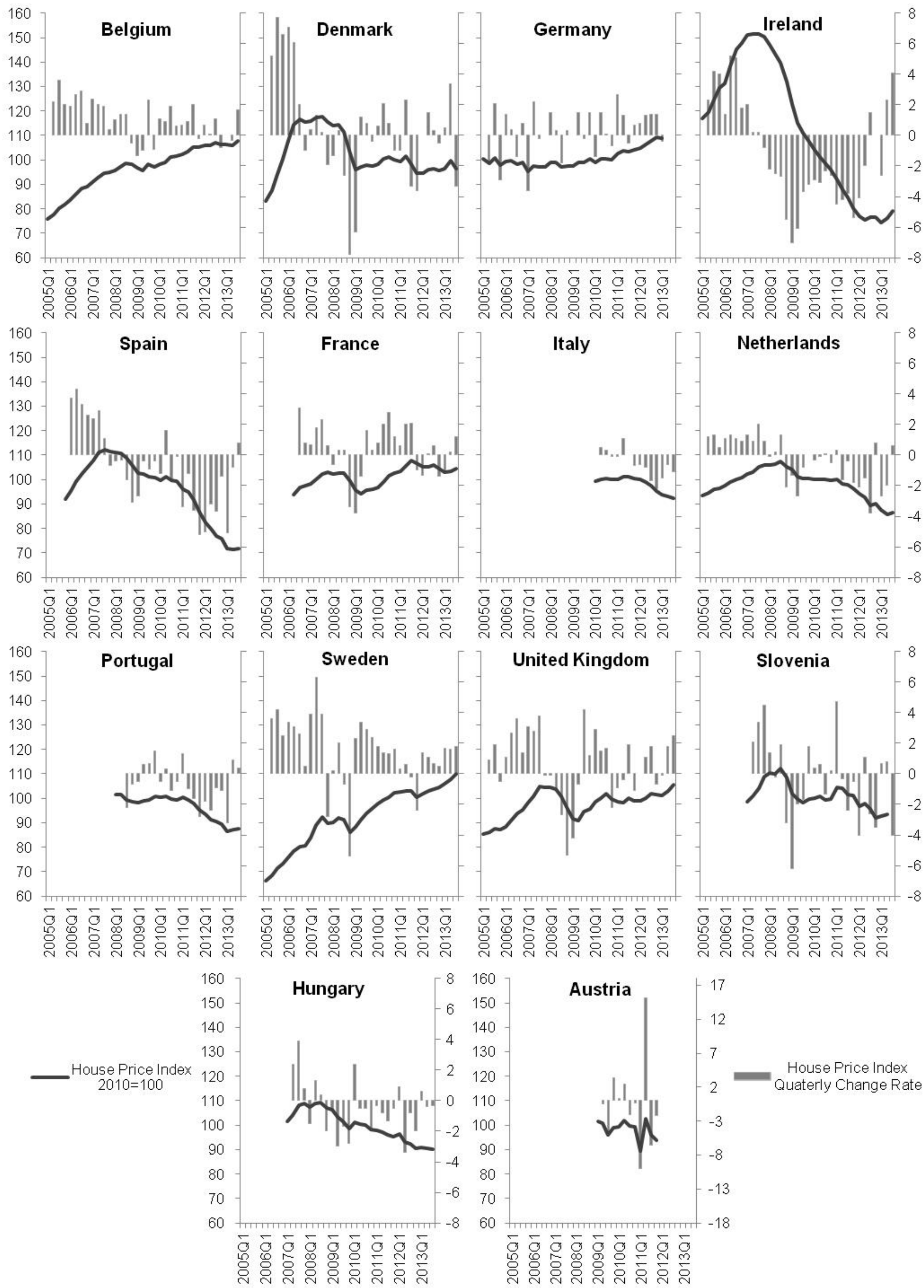


*Figure 2: Development of House Price Index (2010=100) and quarterly change rate.*

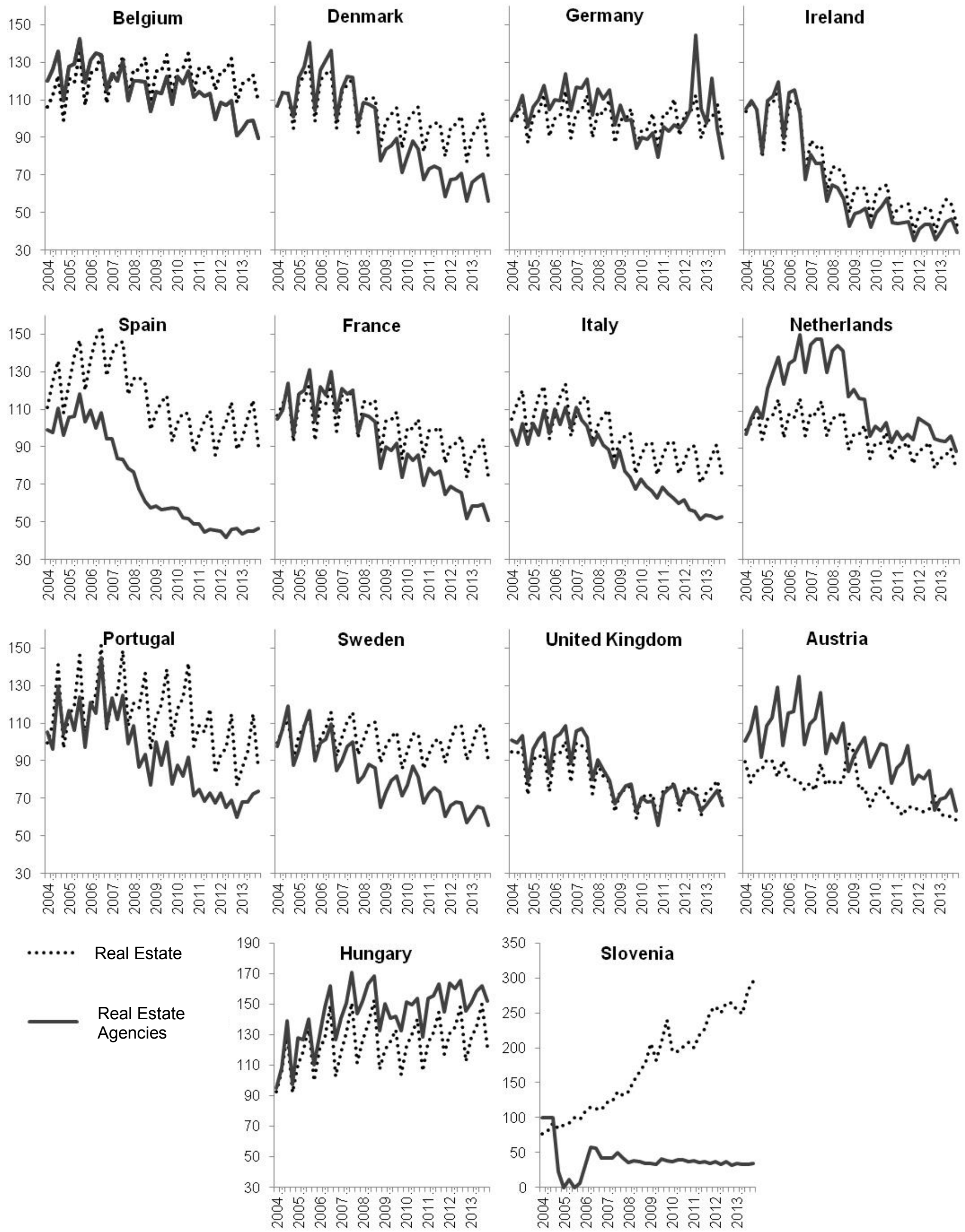


*Figure 3: Quarterly Google search intensities for the main category "Real Estate" and the subcategory "Real Estate Agencies" (Index, 01.01.2004 =100; own calculations).*

## 5 Econometric methodology and baseline results

To illuminate the role of Google data in explaining house prices we use our dataset presented in the previous chapter to perform a standard panel analysis. Since the classical Hausman (1978) test indicates the inconsistency of a random effects model, we estimate in the following different specifications of a fixed effects model using equation (1).

$$(1) \qquad Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + u_{it}$$

As dependent variable *Y* we employ the quarterly HPI change to rule out non-stationarity problems for the index-level indicated e.g. by a Maddala and Wu (1999) test.

As independent variables *X* inflation, interest rates and unemployment are utilized. All exogenous variables enter the regression lagged by one period (3 months) to model a certain delay of the price reaction. Additionally, this design allows us to deal with possible endogeneity problems. In our benchmark specification (I) without search engine data we also include seasonal dummies to account for a possible seasonal structure and a crisis dummy to capture effects caused by the global financial crisis, which certainly had a large impact on real estate markets in most countries of our sample. Following for example Hirata et al. (2013) there is some evidence for higher uncertainty affecting house price developments. Also Shiller (2007) argues that institutional changes followed by a crisis might have an influence on house price developments. The crisis dummy switches from 0 to 1 after the collapse of Lehman Brothers Inc. (Q4_2008) until the end of the sample period, accounting for possible structural breaks triggered by the crisis.

In specification (II) the benchmark model is augmented with Google data, i.e. the search intensities for the predefined category “Real Estate Agencies”. In model (III) we excluded the crisis dummy assessing its impact on the results. Finally, in specification (IV) we estimate a regression with a full set of time dummies abandoning crisis and seasonal dummy variables. Table 2 summarizes the results for our four baseline regressions. To account for occurring heteroscedasticity we compute robust standard errors using the Huber and White (1980) sandwich estimator. We also checked for serial correlation, which is absent in all specifications according to a Woolridge (2002) test.

| Variable | Benchmark Model (I) | Benchmark Model with Google Data (II) | Without Crisis Dummy (III) | With Time Dummies (IV) |
|---|---|---|---|---|
| **Real Estate Agencies (t-1)** | — | **0.0426***<br>(0.0065)** | **0.0785***<br>(0.0000)** | **0.0360***<br>(0.0045)** |
| Inflation (t-1) | -0.2314*<br>(0.0619) | -0.1515<br>(0.2005) | -0.1501<br>(0.2003) | 0.0558<br>(0.7225) |
| Short Term Interest Rate (t-1) | -0.8244***<br>(0.0013) | -0.8092***<br>(0.0004) | -0.5721***<br>(0.0089) | -0.5177**<br>(0.0147) |
| Long Term Interest Rate (t-1) | -0.1382<br>(0.2300) | -0.1515<br>(0.2378) | -0.1684<br>(0.2978) | -0.1901*<br>(0.0768) |
| Harmonised Unemployment (t-1) | -0.2754***<br>(0.0014) | -0.1928**<br>(0.0155) | -0.1636**<br>(0.0482) | -0.1629***<br>(0.0092) |
| Seasonal Dummy Q1 | 0.1949<br>(0.5527) | 0.8304*<br>(0.0554) | 1.3907***<br>(0.0070) | — |
| Seasonal Dummy Q2 | 0.9588**<br>(0.0265) | 1.2285***<br>(0.0075) | 1.6048***<br>(0.0021) | — |
| Seasonal Dummy Q3 | 0.5809<br>(0.1464) | 0.8417*<br>(0.0511) | 1.2223***<br>(0.0076) | — |
| Financial Crisis Dummy | -3.050***<br>(0.0001) | -2.0803***<br>(0.0028) | — | — |
| Constant | 7.0872***<br>(0.0000) | 1.4903<br>(0.4273) | -4.1793***<br>(0.0083) | -0.8410<br>(0.5854) |
| Observations<br>$R^2$<br>Adj. $R^2$ | 366<br>0.3837<br>0.3699 | 366<br>0.4113<br>0.3964 | 366<br>0.3615<br>0.3472 | 366<br>0.4579<br>0.3738 |

p-values in brackets (***p<0.01; **p<0.05; *p<0.10)

*Table 2: Baseline estimation results (Google indicator: "Real Estate Agencies")*

The integration of search data substantially increases the explanatory power of our benchmark model (I). As expected, we observe a highly significant positive relationship between lagged search intensities in the subcategory "Real Estate Agencies" and subsequent HPI changes, i.e. higher search interest is followed by a delayed but accelerated price increase. This result, which can be similarly found in the related literature (Wu and Brynnjolfson (2013)), turns out to be quite robust as our specifications (III) and (IV) show. Obviously adding our Google variable bears some valuable information explaining the price dynamics in our 14 countries. Furthermore, this result can be interpreted as evidence for the dominance of demand oriented search queries, indicating rather high demand than supply, leading to subsequent price jumps.

Taking a look on our four exogenous control variables, most noticeably we find a robust negative connection between lagged short term interest rates and HPI changes. This expected and quite strong relationship (significance at least at 5%-level in all specifications I –IV) is also in line with the literature (see e.g. Goodhart and Hofmann (2008)). The influence of the unemployment rate on house price changes is also negative as expected. This result is also robust to specification changes. Long term interest rates in contrast seem to have no strong effect on HPI change. Surprisingly, the same holds true for the inflation rate, which only turns out to be significant at 10%-level in our benchmark model (I) without Google data.

Seasonal differences play a role in our baseline regressions. In the benchmark model (I) the second quarter turns out to be different, in (II) an (III), when the Google index enters the model all seasonal dummies become significant, capturing the seasonal patterns in the data.

The financial crisis dummy turns out to be highly significant as well, explaining some of the price dynamics when included in (I) and (II). Excluding the "Lehman dummy" in specification (III) removes a large amount of explanatory power from our model. All in all, this supports the thesis of a structural break on real estate markets in the aftermath of the crisis.

Finally, incorporating a complete set of time dummy variables leads to the highest absolute explanatory power of nearly 46%. At the same time, we observe that the adjusted $R^2$ is even lower than in our simple Google augmented model (II).

In order to test alternative measures of search interest, we augmented our benchmark model (I) with the search indices for the main category "Real Estate" (Model IIa) and the seven other subcategories (Specification IIb – IIh) introduced in the previous chapter. As Table 3 shows, employing other indicators instead of our reference measure "Real Estate Agencies" seems not to be a promising approach. Altering the Google measure leaves the influence of the other variables widely unaffected, but no other search measure contains more information about HPI development. With exception of "Commercial & Investment Real Estate" (Model IIb) and "Property Development" (Model IIc) all alternative Google measures turn out to be insignificant, adding no or only marginal explanatory power to our baseline model (I). Thus, we are prone to assume that most categories including the main category "Real Estate" are mixed information groups, i.e. not clearly demand or supply dominated, which eliminates the information content for prices.

| | Real Estate Agencies | Real Estate (IIa) | Commercial & Investment Real Estate (IIb) | Property Development (IIc) | Property Management (IId) | Property Inspections & Appraisals (IIe) | Real Estate Listings (IIf) | Timeshares & Vacation Properties (IIg) | Apartments & Residential Rentals (IIh) |
|---|---|---|---|---|---|---|---|---|---|
| Google Indicator (t-1) | 0.043*** | 0.001 | 0.027* | 0.020* | 0.008 | -0.003 | 0.015 | 0.013 | 0.011 |
| | (0.007) | (0.915) | (0.085) | (0.084) | (0.172) | (0.620) | (0.395) | (0.172) | (0.230) |
| Inflation (t-1) | -0.151 | -0.229* | -0.246** | -0.203 | -0.227* | -0.242** | -0.221* | -0.216* | -0.233* |
| | (0.201) | (0.085) | (0.040) | (0.101) | (0.066) | (0.040) | (0.066) | (0.071) | (0.055) |
| Short Term Interest Rate (t-1) | -0.809*** | -0.821*** | -0.791*** | -0.820*** | -0.822*** | -0.827*** | -0.793*** | -0.819*** | -0.820*** |
| | (0.000) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) |
| Long Term Interest Rate (t-1) | -0.151 | -0.139 | -0.146 | -0.117 | -0.136 | -0.136 | -0.126 | -0.139 | -0.149 |
| | (0.238) | (0.216) | (0.177) | (0.321) | (0.237) | (0.232) | (0.328) | (0.224) | (0.200) |
| Harmonised Un-employment (t-1) | -0.193** | -0.272*** | -0.225*** | -0.237*** | -0.262*** | -0.284*** | -0.238*** | -0.263*** | -0.265*** |
| | (0.016) | (0.000) | (0.008) | (0.008) | (0.003) | (0.002) | (0.002) | (0.003) | (0.002) |
| Seasonal Dummy Q1 | 0.830* | 0.224 | 0.450 | 0.385 | 0.252 | 0.196 | 0.466 | 0.784 | 0.509 |
| | (0.055) | (0.609) | (0.197) | (0.268) | (0.446) | (0.546) | (0.350) | (0.207) | (0.140) |
| Seasonal Dummy Q2 | 1.228*** | 0.973** | 1.105** | 1.029** | 0.969** | 0.980** | 1.073** | 1.255** | 1.122** |
| | (0.008) | (0.033) | (0.014) | (0.022) | (0.025) | (0.018) | (0.015) | (0.019) | (0.014) |
| Seasonal Dummy Q3 | 0.842* | 0.591 | 0.688* | 0.646 | 0.600 | 0.607 | 0.679 | 0.732* | 0.712* |
| | (0.051) | (0.183) | (0.090) | (0.121) | (0.139) | (0.117) | (0.133) | (0.090) | (0.076) |
| Financial Crisis Dummy | -2.080*** | -3.039*** | -2.609*** | -2.811*** | -3.065*** | -3.048*** | -2.979*** | -2.890*** | -3.024*** |
| | (0.003) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Constant | 1.490 | 6.882*** | 4.088*** | 4.562*** | 6.180*** | 7.495*** | 4.621 | 5.413*** | 5.743*** |
| | (0.427) | (0.004) | (0.009) | (0.004) | (0.000) | (0.000) | (0.154) | (0.001) | (0.000) |
| Observations | 366 | 366 | 366 | 366 | 366 | 366 | 366 | 366 | 366 |
| $R^2$ | 0.411 | 0.384 | 0.395 | 0.390 | 0.386 | 0.384 | 0.388 | 0.389 | 0.385 |
| Adj. $R^2$ | 0.396 | 0.368 | 0.380 | 0.374 | 0.371 | 0.369 | 0.372 | 0.373 | 0.370 |

p-values in brackets (***p<0.01; **p<0.05; *p<0.10)

*Table 3: Estimation results for alternative Google indicators based on Google Trends "Real Estate" categories*

# 6 Some variations and extensions

## 6.1 The role of internet access and use

It is reasonable to assume that the additional explanatory power of our Google indicator is influenced by the overall importance of search engine use within a country. This importance in turn mainly depends on the access and the intensity of internet use. In countries where a large part of the population has no access or does not frequent the Internet regularly our measure is likely to lose its information content and ability to indicate demand and following price switches. On the opposite, in “high importance” countries where the prevalence and use of the Internet is strongly pronounced, our Google measure should be able to map a sufficient part of demand driven activities, adding valuable information to existing models. In such countries, we expect that more activities within the HBP are carried out online. Thus, our Google measure here is assumed to better reflect mood swings or other changes leading to shifts in demand (and prices) than in countries with lower internet importance.

In order to check whether this is the case for our country sample, we consider three indicators taken from the Information Society Statistics (Eurostat), assessing the relevance of the Internet in our countries. Firstly, we take into account “Households with internet access” (isoc_ci_in_h), indicating disparities in access. Secondly, we look at the “Daily use of Internet” (isoc_ci_ifp_fu), telling us additionally something about the frequency or intensity of use and thirdly, perhaps the most informative indicator for our purpose, we incorporate “Individuals who have used a search engine to find information” measuring cross county differences in search engine use (isoc_sk_iskl_i).

Although the level of access has risen dramatically between 2005 and 2013, especially in low level countries, the share of households with access to the Internet still varies substantially, as shown by Figure 4.

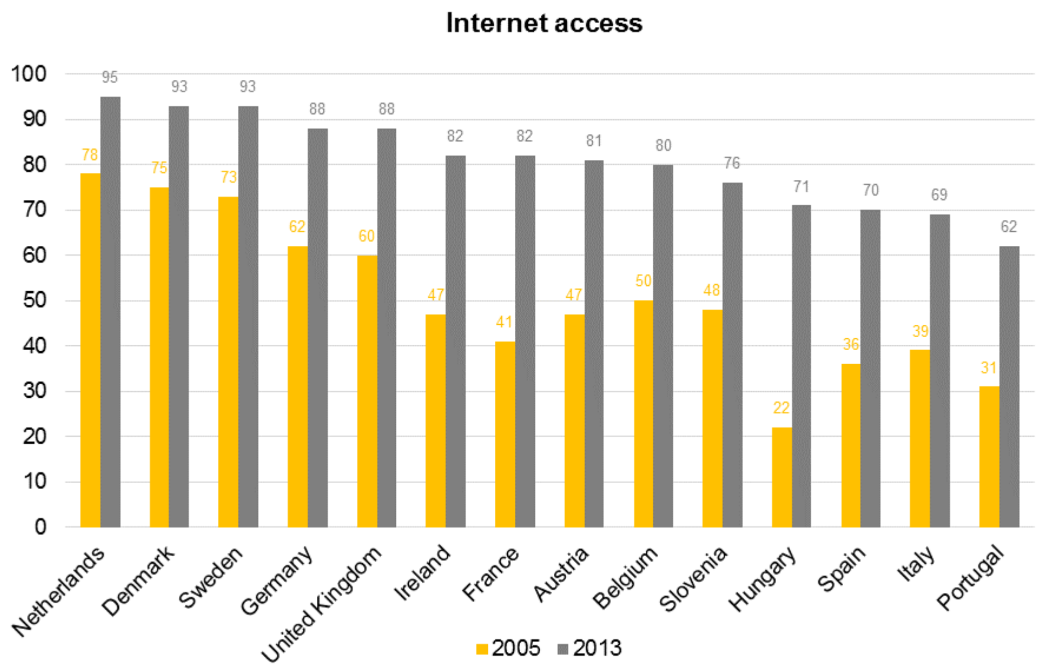


*Figure 4: Households with Internet access 2005 and 2013 (% of all households; France: 2006 and 2013); Source: Eurostat.*

In 2013, there are top-ranking countries achieving nearly complete internet access, like the Netherlands (95 %) or Denmark and Sweden (93 %). At the end of the list, we find three southern European countries, Spain (70 %), Italy (69 %) and Portugal (62 %) with a comparably low share.

Not surprisingly, taking a look on the intensity of internet use in Figure 5, we almost find the same order of countries. In Denmark (84 %), the Netherlands (83 %) and Sweden (81 %) more than 4 out of 5 people are online daily. In contrast to these results, we find again the lowest proportions in Spain (54 %), Italy (54 %) and Portugal (48 %), where approximately only one out of two persons can be considered as "heavy user", accessing the Internet daily, which can be at least partly attributed to the low access quotas in these countries.

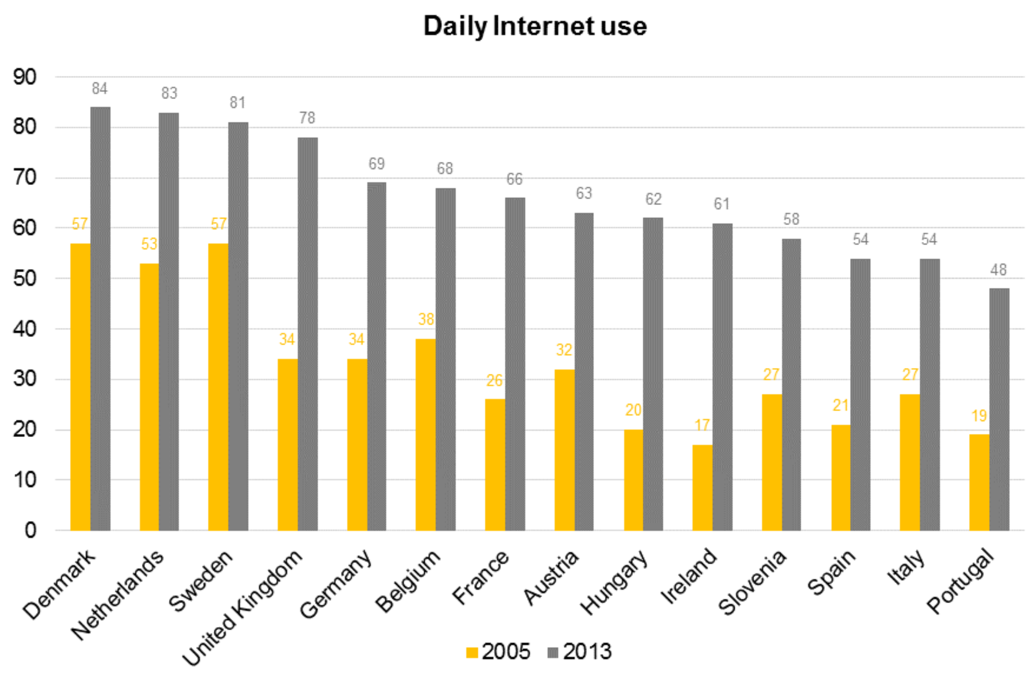


*Figure 5: Daily Internet use 2005 and 2013 (% of all individuals; France: 2006 and 2013); Source: Eurostat*

Finally, inspecting our third indicator "Search engine use" in detail in Figure 6, we again find for 2013 the usual suspects at the top (Denmark, the Netherlands and Sweden, all 92 %) and at the bottom of the country ranking (Portugal 65 % and Italy 62 %). Comparing the 2013 figures with 2005, it shows that the overall use of search engines has increased significantly over the past eight years which should improve the reliability of our Google indicator. Nevertheless, even in 2013 we still observe large differences within our country sample which are again driven mainly by cross-country disparities in internet access. The differences almost completely level out, when we relate the people who have used a search engine to individuals who ever used the Internet, since nearly everybody who uses the Internet uses search engines.

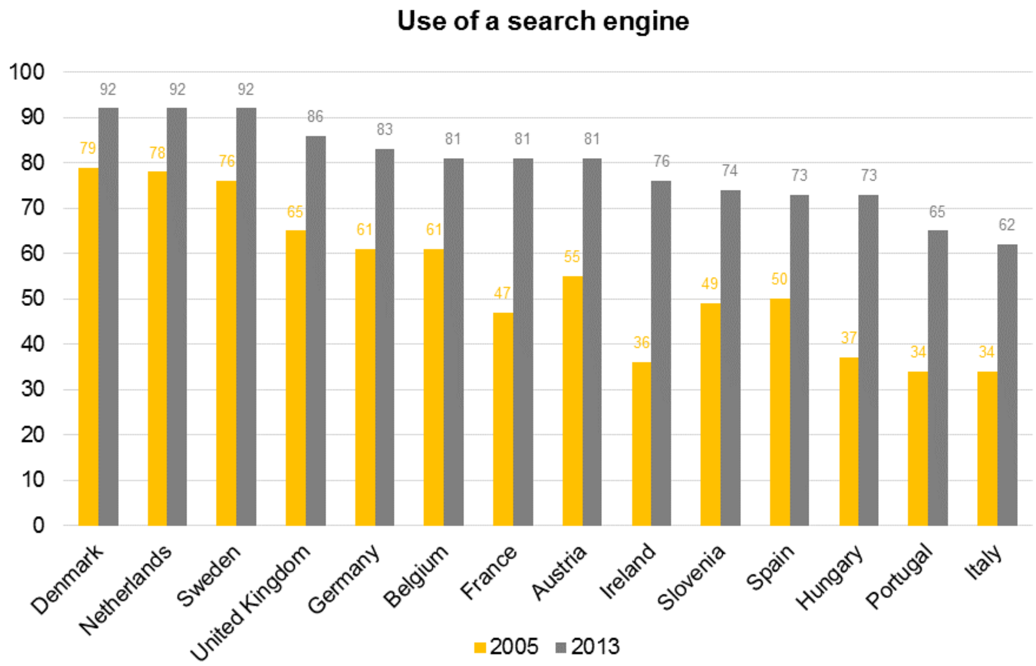


*Figure 6: Individuals who have used a search engine to find information 2005 and 2013 (in % of all individuals; Belgium, France and Spain: 2006 and 2013); Source: Eurostat.*

All in all, concerning internet access and (search engine) use there is quite a great deal of heterogeneity within our sample, which can influence the quality of our Google indicator. To account for these differences, we use a standard k-means cluster analysis based on average internet access between 2005 and 2013 to split the sample into a country group "H" with high internet relevance (Denmark, Netherlands, Sweden, United Kingdom, Germany) and a group "L" with comparatively low importance (Austria, Belgium, France, Ireland, Hungary, Slovenia, Spain, Italy, Portugal)[7]. In order to inspect how our Google indicator performs in these two different settings, we conducted two separate estimations, one for countries with high

[7] We receive the same groups, when using search engine use or a combination of both or all three indicators. We also experimented with a distinction into three (high, medium and low affinity) or even four groups accepting a quite low number of observations within the subsamples. The results, which are available upon request, point qualitatively into the same direction.

internet importance (V) and one for countries with low relevance (VI). Table 4 summarizes the results.

| Variable | Benchmark Model (all countries) (I) | Country sample H "High relevance" (V) | Country sample L "Low relevance" (VI) |
|---|---|---|---|
| **Real Estate Agencies (t-1)** | **0.0426***** **(0.0065)** | **0.0883**** **(0.0268)** | **0.0339** **(0.1299)** |
| Inflation (t-1) | -0.1515 (0.2005) | -0.5484 (0.1581) | -0.1850 (0.2988) |
| Short Term Interest Rate (t-1) | -0.8092*** (0.0004) | -0.7508** (0.0292) | -0.7227** (0.0100) |
| Long Term Interest Rate (t-1) | -0.1515 (0.2378) | -0.3226 (0.3588) | -0.1162 (0.3916) |
| Harmonised Unemployment (t-1) | -0.1928** (0.0155) | -0.0197 (0.8831) | -0.1927 (0.1375) |
| Seasonal Dummy Q1 | 0.8304* (0.0554) | 2.3924** (0.0140) | 0.1158 (0.7959) |
| Seasonal Dummy Q2 | 1.2285*** (0.0075) | 2.1866* (0.0511) | 0.7538* (0.0700) |
| Seasonal Dummy Q3 | 0.8417* (0.0511) | 1.6450** (0.0314) | 0.4033 (0.4924) |
| Crisis Dummy | -2.0803*** (0.0028) | -0.5363 (0.6428) | -2.4598*** (0.0041) |
| Constant | 1.4903 (0.4273) | -5.0986 (0.1837) | 3.0397 (0.2836) |
| Observations | 366 | 157 | 209 |
| $R^2$ | 0.4113 | 0.4742 | 0.4183 |
| Adj. $R^2$ | 0.3964 | 0.4420 | 0.3920 |

p-values in brackets (***p<0.01; **p<0.05; *p<0.10)

*Table 4: Estimation results for countries with high and low Internet relevance*

In contrast to the short term interest rate, which seems to be the most important driver of house prices in all specifications, the Google indicator only proves significant in subsample "H" (at 5%-level). So, our measure of search interest seems to bear sufficient information explaining some parts of the price dynamics only in a high use and access context.

As expected, compared to the whole country sample, we gain overall explanatory power of the model when restricting to high level countries. For the low relevance countries in contrast, our model seems to work not that properly. The adjusted R² shrinks slightly, indicating a worse fit.

The crisis dummy is only relevant (and significant at 1%-level) in countries with low internet importance. This might point into the direction that structural breaks stemming from the financial crisis are much more pronounced in the low importance sample which contains countries like Portugal, Spain, Ireland and Italy, whose economies and real estate sectors were hit very hard by the financial crisis over the past years. Interestingly, in those countries the importance of seasonal effects nearly completely vanishes, i.e. all seasonal dummies except Q2 become insignificant. Furthermore, “Harmonised Unemployment“, our measure of economic activity becomes irrelevant in both country samples, which is a little bit puzzling on first sight.

## 6.2 The Google indicator in troubled times

As seen in chapter 2, there is some empirical evidence that the explanatory power of Google indicators might be influenced also by the recent financial crisis. Since the real estate sector was strongly and heterogeneously affected in many of our countries, this is also an important question for our analysis. While it could be that adding search query information works even better in times of economic uncertainty, the opposite might also be true. In times of trouble, which are usually associated with huge uncertainty, people generally are looking for any kind of advice and possibly evolve additional online search activities expanding our data basis. If these activities are mostly informative, i.e. prudent, systematic and with connection to subsequent real world decisions this should increase the quality of our indicator. But if they are mostly panic driven with no systematic consequences for real world activities this simply inflates our data pool diluting the information content of our Google measure and making it harder to extract valuable information. Thus, ex ante, the impact of the crisis is not quite clear.

In Table 5 we investigate the consequences of the financial crisis for our Google measure in more detail, splitting the sample into a pre-crisis period ranging from Q1_2005 until Q3_2008 and a crisis period (Q4_2008-Q1_2013) after the collapse of Lehman Brothers Inc. in September 2008. Restricting our attention to the early non-crisis era obviously boosts the explanatory power of our Google augmented standard model. The Google measure is still highly significant, as well as all seasonal dummies. Interestingly, in the pre-crisis era short term interest rates and unemployment become irrelevant, whereas the lagged inflation rate now seems to exert some influence on real estate prices.

In contrast to these findings, our model seems not well suited to explain house price dynamics in troubled times. Only short term interest rates and a single seasonal dummy are observed to be significant, whereas all other factors including the search query variable lose their relevance. Furthermore, the explanatory power of our model decreases dramatically. Although our standard model seems to be misspecified for this exceptional situation, this result can be considered as first evidence for the existence of a dilution effect, making our Google indicator uninformative in rather uncertain crisis times.

| **Variable** | **Benchmark Model** (Q1_2005-Q1_2013) (I) | **Pre-Crisis Period** (Q1_2005-Q3_2008) (VII) | **Crisis Period** (Q4_2008-Q1_2013) (VIII) |
|---|---|---|---|
| **Real Estate Agencies (t-1)** | **0.0426***<br>(0.0065)** | **0.0841***<br>(0.0069)** | **0.0149<br>(0.5294)** |
| Inflation (t-1) | -0.1515<br>(0.2005) | -0.7837**<br>(0.0196) | -0.2075<br>(0.2030) |
| Short Term Interest rate (t-1) | -0.8092***<br>(0.0004) | -0.5592<br>(0.1698) | -0.6426**<br>(0.0104) |
| Long Term Interest Rate (t-1) | -0.1515<br>(0.2378) | -0.6784<br>(0.2964) | -0.1638<br>(0.1631) |
| Harmonised Unemployment (t-1) | -0.1929**<br>(0.0155) | -0.2879<br>(0.4514) | -0.0902<br>(0.5568) |
| Seasonal Dummy Q1 | 0.8304*<br>(0.0554) | 2.1375**<br>(0.0118) | 0.2729<br>(0.5546) |
| Seasonal Dummy Q2 | 1.2285***<br>(0.0075) | 1.8425***<br>(0.0058) | 1.1335**<br>(0.0202) |
| Seasonal Dummy Q3 | 0.8417*<br>(0.0511) | 1.9737***<br>(0.0001) | 0.4412<br>(0.4146) |
| Crisis Dummy | -2.0803***<br>(0.0028) | — | — |
| Constant | 1.4903<br>(0.4273) | -0.5694<br>(0.9463) | 0.8774<br>(0.7584) |
| Observations<br>$R^2$<br>Adj. $R^2$ | 366<br>0.4113<br>0.3964 | 132<br>0.5165<br>0.4850 | 234<br>0.2204<br>0.1927 |

p-values in brackets (***p<0.01; **p<0.05; *p<0.10)

*Table 5: Estimation results before and after the collapse of Lehman Brothers Inc.*

## 6.3 A supply side augmented model

Finally, we want to investigate the influence of supply side factors on our Google-HPI nexus. Because of the unavailability of quarterly data on housing stock, we are forced to draw on data on building permits. Eurostat provides two quarterly indices (2010=100) on permissions, which are available for our whole country sample (sts_cobp_q). The first one relates to the number of dwellings, the second time series refers to m² of useful floor area. Both comprise residential buildings, except residences for communities. To check if the integration of a supply side measure has an impact on the information content of our Google indicator for house prices, we proceed as follows. In a first step we enlarge our standard model (I) with the lagged quarterly growth rate of building permits, both in terms of numbers (Model IX) and square meters (Model XI).[8] In a second step we estimate two further models (X) and (XII) inserting our well known Google indicator into these two extended baseline models. The results for all specifications are summarized in Table 6.

| **HPI change** | **Supply Side Benchmark Model #1 (IX)** | **Model #1 with Google Data (X)** | **Supply Side Benchmark Model #2 (XI)** | **Model #2 with Google Data (XII)** |
|---|---|---|---|---|
| **Real Estate Agencies (t-1)** | — | **0.0417**** **(0.0101)** | — | **0.0413**** **(0.0106)** |
| Inflation (t-1) | -0.2216* (0.0723) | -0.1419 (0.2472) | -0.2202* (0.0725) | -0.1413 (0.2484) |
| Short Term Interest Rate (t-1) | -0.8025*** (0.0015) | -0.7944*** (0.0005) | -0.7925*** (0.0013) | -0.7854*** (0.0004) |
| Long Term Interest Rate (t-1) | -0.1497 (0.2011) | -0.1628 (0.2169) | -0.1510 (0.2000) | -0.1639 (0.2164) |
| Harmonised Unemployment (t-1) | -0.2569*** (0.0021) | -0.1799** (0.0222) | -0.2531*** (0.0021) | -0.1772** (0.0236) |
| Seasonal Dummy Q1 | 0.1038 (0.7228) | 0.7319* (0.0739) | 0.0770 (0.7837) | 0.7019* (0.0758) |
| Seasonal Dummy Q2 | 1.0420** (0.0181) | 1.3070*** (0.0046) | 1.0242** (0.0173) | 1.2881*** (0.0041) |
| Seasonal Dummy Q3 | 0.4662 (0.2307) | 0.7288* (0.0876) | 0.3853 (0.3045) | 0.6525 (0.1168) |
| Financial Crisis Dummy | -3.0807*** (0.0001) | -2.1396*** (0.0015) | -3.0616*** (0.0001) | -2.1308*** (0.0014) |
| Building Permits (t-1) [QC, Number] | 0.8494** (0.0354) | 0.7835** (0.0479) | — | — |
| Building Permits (t-1) [QC, Square meter] | — | — | 1.1826** (0.0147) | 1.0866** (0.0183) |
| Constant | 6.9920*** (0.0000) | 1.5664 (0.4142) | 6.9597*** (0.0000) | 1.5869 (0.4071) |
| Observations | 359 | 359 | 359 | 359 |
| $R^2$ | 0.3838 | 0.4101 | 0.3867 | 0.4125 |
| Adj. $R^2$ | 0.3679 | 0.3931 | 0.3709 | 0.3956 |

p-values in brackets (***p<0.01; **p<0.05; *p<0.10)

*Table 6: Regression results for supply side extended baseline models*

[8] In order to model a delayed price reaction resulting from a gap between the grant of a permission and the time the dwelling is available at the market, we tested specifications up to 4 lags (12 months) with levels and growth rates. We find that only one period lagged quarterly change rates of building permits exert a significant influence on house prices.

Although significant at 5%-level, we find that introducing data on building permits does not add much explanatory information to our regressions. The Google indicator is observed to be significant in both extensions (X) and (XII) increasing clearly the coefficient of determination of the supply side augmented baseline specifications. Thus, we conclude that the robust connection between search interest and house prices seems to be nearly unaffected by including our supply side measure.

## 7 Conclusion and implications for further research

One of the major shortcomings of traditional models explaining house prices is the lack of information. Most relevant data on macro level such as population or housing stocks are only available at quarterly or even lower frequency. Additionally, and even more severe, other useful data which may describe important drivers of house price dynamics, for example investor sentiment or socio-demographic or even individual characteristics are not observable or only to retrieve at very high cost. As a simple way out to improve empirical models explaining house prices we propose to include Google search frequencies as an aggregate measure to better account for such factors. Our study, which is to our knowledge the first one incorporating search engine information into a multi-country framework on house prices, suggests that Google data at least partly maps these missing factors and thus helps to close this informational gap.

In contrast to earlier studies which focus on regional or country wide data of the US and UK we show for a sample of 14 EU-countries comprising the years 2005-2013 that the additional integration of search interest data into a standard model of fundamentals helps to explain changes of the Eurostat HPI.

Analyzing and comparing information on real estate related searches provided from Google via its public web facility Google Trends we find that the predefined search category "Real Estate Agencies" possesses the highest information content for our purpose. Using an index based on this category significantly boost the explanatory power of our standard model in nearly all specifications. In a robustness analysis, we are able to show, that this is effect is especially pronounced for countries with high internet relevance concerning use and access. We also checked the information content of our Google indicator in troubled times. Splitting the sample into a crisis and pre crisis period before the crash of Lehman Brothers Inc. in autumn 2008, reveals first evidence for a dilution effect, making our Google indicator uninformative in rather uncertain crisis times. Finally, we also tested a model extension using building permits in order to model the supply side. This procedure leaves our Google-HPI nexus nearly unaffected.

All in all, our Google indicator seems to contain a great deal of information concerning European house prices, especially when the importance and diffusion of the Internet is rather high and in times of economic stability. Since Google data is easy to retrieve, free of charge and available without any delay, we are convinced that real estate professionals, researchers and policy makers should not ignore these valuable source of information. Our approach is a first promising step to show how search interest data can attenuate informational problems of empirical house price models.

In contrast to other studies we focus on explaining house price dynamics until now. Naturally, the next step on our research agenda is to conduct a fore- and nowcasting analysis to check in a multinational context if models using search interest data outperform traditional fore- or nowcasting models. In this context we intend to address also the question if the use of Google data is more effective in nowcasting house prices, which is shown by Wu and Brynjolfsson (2013) for the US. Additionally, to shed more light on causality and timing issues and better exploit the weekly raw data structure of our Google index we consider to apply more sophisticated estimation techniques (e.g. mixed frequencies methods, panel VAR analysis). Finally, a more formal structural break analysis could be introduced to better capture the regime switch triggered by the financial crisis.

## References


Aoki, K., Proudman, J. and Vlieghe, G. (2004), “House Prices, Consumption, and Monetary Policy: A Financial Accelerator Approach”, *Journal of Financial Intermediation* 13(4), 414–35.

Askitas, N. and Zimmermann, K.F. (2009), “Google Econometrics and Unemployment Forecasting”, *Applied Economics Quarterly* 55(2), 107–20.

Buckley, R and Ermisch, R. (1982), “Government Policy and house Prices in the United Kingdom: An Econometric Analysis”, *Oxford Bulletin of Economics and Statistics* 44(4), 273-304.

Choi, H. and Varian, H. (2009), “Predicting initial claims for unemployment benefits”, Google Inc., Technical report.

Choi H. and Varian, H. (2012), “Predicting the Present with Google Trends”, *Economic Record* 88 (Supplement 1), 2-9.

D’Amuri, F. (2009), “Predicting unemployment in short samples with Internet job search query data”, *MPRA Paper,* No. 18403.

Dougherty, A., and van Order, R. (1982), “Inflation, Housing Costs and Consumer Price Index”, *American Economic Review* 72(1), 154-165.

Drukker, D. M. (2003), “Testing for serial correlation in linear panel-data models”, *The Stata Journal* (3)2, 1-10.

Edelman, B. (2011), “Bias in Search Results? Diagnosis and Response”, *The Indian Journal of Law and Technology* 7, 16-32.

Gattini, L. and Hiebert, P. (2010), “Forecasting and assessing Euro area house prices through the lens of key fundamentals”, *ECB Working Paper Series,* No. 1249.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M. S. and Brilliant, L. (2009), “Detecting influenza epidemics using search engine query data”, *Nature* 457, 1012–14.

Girouard, N., Kennedy, M., van den Noord, P. and André, C. (2006), “Recent House Price Developments: The Role of Fundamentals”, *OECD Economics Department Working Papers*, No. 475.

Goodhart, C. and Hofmann, B. (2008), “House Prices: Money, Credit and the Macroeconomy", *ECB Working Paper Series,* No. 888.

Greenlight Digital (2010), “Search engine market shares around the world - Q4 2010”, available online:
http://www.greenlightdigital.com/assets/images/market-share-large.png
(last accessed: 15.10.2014)

Guzman, G. (2011), “Internet search behavior as an economic forecasting tool: the case of inflation expectations”, *The Journal of Economic and Social Measurement* 36(3), 119-167.

Haucap, J. and Kehder, C. (2013), “Suchmaschinen zwischen Wettbewerb und Monopol: Der Fall Google”, *DICE Ordnungspolitische Perspektiven*, No. 44.

Hausman, J.A. (1978), “Specification Tests in Econometrics”, *Econometrica* 46(6), 1251-1271.

Hess, T. and Lehrer, C. (2011), “Digitale Medien in der Immobilienvermittlung: von ortsbezogenen Diensten bis hybrid TV”, *Ludwigs-Maximilians-Universität München, Immostudie 2011.*

Hess, T. and Wagner, T. (2012), “Digitale Medien in der Immobilienvermittlung: von ortsbezogenen Diensten bis Hybrid-TV”, *Ludwigs-Maximilians-Universität München, Immostudie 2012.*

Hilbers P., Hoffmaister, A., Angana Banerji, A. and Shi, H. (2008), “House Price Developments in Europe: A Comparison“, *IMF Working Paper*, No. 08/211

Hirata, H., Kose, M.A., Otrok, C. and Terrones, M.E. (2013), "Global House Price Fluctuations: Synchronization and Determinants," *NBER International Seminar on Macroeconomics*, University of Chicago Press 9(1), 119 - 166.

Hlaváček, M. and Komarek, L. (2009), “Housing Price Bubbles and their Determinants in the Czech Republic and its Regions”, Czech National Bank, *Working Paper Series,* No.12.

Hohenstatt, R., Käsbauer, M. and Schäfers, W. (2011), “’Geco’ and its Potential for Real Estate Research: Evidence from U.S. Housing Market”, *Journal of Real Estate Market Research* 33(4), 471-506.

Huang, D. (1966), “The Short-Run Flows of Nonfarm Residential Mortgage Credit”, *Econometrica* 34(2), 433-459.

Kajuth, F., Knetsch, T. and Pinkwart, N. (2013), “Assessing house prices in Germany: evidence from an estimated stock-flow model using regional data”, *Bundesbank Discussion Paper,* No. 46/2013.

Kearl, J. (1979), “Inflation, Mortgages, and Housing”, *Journal of Political Economy* 87(5), 1115-1138.

Kholodilin, K. ,Podstawski, M., Siverstovs, B. and Bürgl, C. (2009), “Google Searches as a Means of Improving the Nowcasts of Key Macroeconomic Variables”, *DIW Discussion Papers,* No. 946.

Maddala, G.S. and Wu, S. (1999), “A Comparative Study of Unit Root Tests With Panel Data and A New Simple Test”, *Oxford Bulletin of Economics and Statistics* 61(1), 631-652.

McLaren, N. and Shanbhogue, R. (2011), “Using Internet search data as economic indicators”, *Bank of England Quarterly Bulletin* 51(2), 134-140.

Muth, R. (1960), “The Demand for Non-Farm-Housing” in A. C. Harberger (ed.): *The Demand for Durable Goods*, Chicago: University of Chicago Press, 29-96.

National Association of Realtors (2012), “Profile of Home Buyers and Sellers 2012”, available online:
http://www.realtor.org/sites/default/files/Highlights-NAR-2012-HBS-2012-11-12.pdf
(last accessed 15.10.2014).

National Association of Realtors and Google Inc. (2013), “The Digital House Hunt: Consumer and Market Trends in Real Estate“, available online:
http://www.realtor.org/sites/default/files/Study-Digital-House-Hunt-2013-01_1.pdf
(last accessed: 15.10.2014).

Office of Fair Trading (2010), “Home Buying and Selling: A Market Study“, OFT 1186.

Pentland, A.S. (2010), “Honest signals”, Cambridge, Massachusetts: The MIT Press.

Poterba, J. (1984), “Tax Subsidies to Owner-Occupied Housing: An Asset Market Approach”, *Quarterly Journal of Economics* 99(4), 729-752.

Shiller, R.J. (2007), "Understanding recent trends in house prices and homeownership", Federal Reserve Bank of Kansas City, *Proceedings*, 89-123.

Smith, L. (1969), “A Model of the Canadian Housing and Mortgage Markets”, *Journal of Political Economy* 77(5), 795-816.

Stephens-Davidowitz, S. (2013), “Essays using Google Data”, *Dissertation*, Harvard University, Cambridge, Massachusetts.

Suhoy, T. (2009), “Query indices and a 2008 downturn”, *Bank of Israel Discussion Paper,* No. 2009.06.

Taylor, J.B. (2007), "Housing and monetary policy", Federal Reserve Bank of Kansas City, *Proceedings*, 463-476.

Vosen, S. and Schmidt, T. (2011), “Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends“, *Journal of Forecasting* 30(1), 565–578.

Webb, G.K. (2009), “Internet search statistics as a source of business intelligence: searches on foreclosure as an estimate of actual home foreclosures”, *Issues in Information Systems* 10(2), 82-87.

White, H. (1980), “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity”, *Econometrica* 48(4), 817–38.

Wooldridge, J.M. (2002), “Econometric Analysis of Cross Section and Panel Data”, Cambridge, Massachusetts: The MIT Press.

Wu, L. and Brynjolfsson, E. (2013), “The future of prediction: how Google searches foreshadow housing prices and sales”, *Economics of Digitization*, NBER.

Wurtzebach H.C., Mueller, R.G. and Machi, D. (1991), "The Impact of Inflation and Vacancy of Real Estate Returns", *Journal of Real Estate Research* 6(2), 153-168.

DISKUSSIONSPAPIERE DER FÄCHERGRUPPE VOLKSWIRTSCHAFTSLEHRE

DISCUSSION PAPERS IN ECONOMICS

Die komplette Liste der Diskussionspapiere ist auf der Internetseite veröffentlicht / for full list of papers see: http://fgvwl.hsu-hh.de/wp-vwl

**2014**

152 Dewenter, Ralf; Giessing, Leonie: The Effects of Elite Sports on Later Job Success, October 2014

151 Dewenter, Ralf; Rösch, Jürgen; Terschüren, Anna: Abgrenzung zweiseitiger Märkte am Beispiel von Internetsuchmaschinen, October 2014

150 Berlemann, Michael; Jahn, Vera: Governance, firm size and innovative capacity: regional empirical evidence for Germany, August 2014

149 Dewenter, Ralf; Rösch, Jörg: Net neutrality and the incentives (not) to exclude competitors, July 2014

148 Kundt, Thorben: Applying "Benford's law" to the Crosswise Model: Findings from an online survey on tax evasion, July 2014

147 Beckmann, Klaus; Reimer, Lennart: Dynamiken in asymmetrischen Konflikten: eine Simulationsstudie, July 2014

146 Herzer, Dierk: Unions and income inequality: a heterogeneous panel cointegration and causality analysis, July 2014

145 Beckmann, Klaus; Franz, Nele; Schneider, Andrea: Intensive Labour Supply: a Menu Choice Revealed Preference Approach for German Females and Males, June 2014

144 Beckmann, Klaus; Franz, Nele; Schneider, Andrea: On optimal tax differences between heterogenous groups, May 2014

143 Berlemann, Michael; Enkelmann, Sören: Institutions, experiences and inflation aversion, May 2014

142 Beckmann, Klaus; Gattke, Susan: Tax evasion and cognitive dissonance, April 2014

141 Herzer, Dierk; Nunnenkamp, Peter: Income inequality and health – evidence from developed and developing countries, April 2014

140 Dewenter, Ralf; Heimeshoff, Ulrich: Do Expert Reviews Really Drive Demand? Evidence from a German Car Magazine, March 2014

139 Dewenter, Ralf; Heimeshoff, Ulrich: Media Bias and Advertising: Evidence from a German Car Magazine, March 2014.

138 Beckmann, Klaus; Reimer, Lennart: Dynamics of military conflict from an economics perspective, February 2014.

**2013**

137 Christmann, Robin: Tipping the Scales - Conciliation, Appeal and the Relevance of Judicial Ambition.

136 Hessler, Markus; Loebert, Ina: Zu Risiken und Nebenwirkungen des Erneuerbare-Energien-Gesetzes, June 2013.

135 Wesselhöft, Jan-Erik: The Effect of Public Capital on Aggregate Output- Empirical Evidence for 22 OECD Countries -, June 2013.

134 Emrich, Eike; Pierdzioch, Christian; Rullang, Christian: Zwischen Ermessensfreiheit und diskretionären Spielräumen: Die Finanzierung des bundesdeutschen Spitzensports – eine Wiederholungsstudie, April 2013.

133 Christmann, Robin: Vertragliche Anreize und die Fehlbarkeit des Richters – Der ungewisse Gang vor Gericht und sein Einfluss auf eine Verhaltenssteuerung im BGB-Vertragsrecht, March 2013.

132 Gerrits, Carsten: Internetnutzer und Korruptionswahrnehmung - Eine ökonometrische Untersuchung, February 2013.

131 Freese, Julia: The regional pattern of the U.S. house price bubble - An application of SPC to city level data, January 2013.

**2012**

130 Kruse, Jörn: Unabhängige staatliche Institutionen: Funktionalität und demokratische Legitimation, November 2012.

129 Andrae, Jannis: Unabhängigkeit von Institutionen - Gründe bzw. Ursachen und Kriterien zur Beurteilung der Unabhängigkeit von öffentlichen Institutionen im demokratischen Rechtsstaat, November 2012.

128 Andrae, Jannis: Ideengeschichtliche Aspekte unabhängiger Institutionen, November 2012.

127 Pfeiffer, Christoph P.: Causalities and casualties: Media attention and terrorism, 1970–2010, November 2012.

126 Pierdzioch, Christian; Emrich, Eike: A Note on the International Coordination of Anti-Doping Policies, November 2012.

125 Berlemann, Michael; Wesselhöft, Jan-Erik: Estimating Aggregate Capital Stocks Using the Perpetual Inventory Method – New Empirical Evidence for 103 Countries –, October 2012.

124 Berlemann, Michael; Freese, Julia; Knoth, Sven: Eyes Wide Shut? The U.S. House Market Bubble through the Lense of Statistical Process Control, October 2012.

123 Pierdzioch, Christian; Emrich, Eike; Klein, Markus: Die optimierende Diktatur – Politische Stabilisierung durch staatlich verordnetes Doping am Beispiel der DDR, August 2012.

122 Flatau, Jens; Emrich, Eike; Pierdzioch, Christian: Zum zeitlichen Umfang ehrenamtlichen Engagements in Sportvereinen – sozioökonomische Modellbildung und empirische Prüfung, August 2012.