''Geco'' and its Potential for Real Estate Research: Evidence from the U.S. Housing Market

Authors

Ralf Hohenstatt, Manuel Käsbauer, and Wolfgang Schäfers

Abstract

Google econometrics (Geco) is a powerful tool for research based on individuals rational. Following the seminal work of Ginsberg et al. (2009), this is another paper based on search query data from Google Insights for Search (I4S). Information on the home buying process is embedded in existing literature on the price-volume relationship in the housing market. The main findings are: I4S subcategories yield inferences about prices and transactions in the near future. While the ''Real Estate Agency'' subcategory serves as a very robust indicator of transaction volume, ''Home Financing'' provides interesting insights into the corresponding financing decisions. Therefore, this study seeks to improve the informational efficiency of a relatively imperfect market and is addressed to policymakers as well as real estate professionals.

Having recently experienced the creation and collapse of two bubbles within a ten-year period, this field of research has become a particularly significant topic for economists. This is also due to the fact that traditional theories cannot explain economic development solely by means of fundamental real-world data. Stiglitz (1990, p. 13) states: ''If the reason that the price is high today is *only* because investors believe that the selling price will be high tomorrow—when 'fundamental' factors do not seem to justify such a price—then a bubble exists.''

Applying this view to the latest housing boom in the United States, the belief of housing market participants in increasing prices, ignored the relationship (or indeed the lack thereof) to fundamental values. In the case of the American housing bubble, it was the ''widespread perception that houses are a great investment, and the boom psychology that helped spread such thinking,'' (Shiller, 2007, p. 7). Alternatively expressed, both psychological and social factors are extremely important for the creation of a bubble. Hard real-world data cannot capture such socio-demographic changes, at least not without a substantial time lag. The decision of individuals for or against buying a house is an aggregate of many subordinate decision-making processes. These may include the birth of a child, marriage, migration for various reasons (e.g., employment opportunities), age (e.g., retirement), investment in a second home, and so forth.

However, it may not be possible to measure individual rationales accurately at the time that they co-determine future economic developments. According to Fishbein and Ajzen (1975), there is no alternative to asking people for their intended behavior through traditional surveys. However, such interviewing is inevitably associated with certain inadequacies, such as costly information, a limited population and low return rate, interview bias, and time lags.

A possible remedy may be found in global Internet connections and the substantially increasing number of users over the past decade. There is now barely an aspect of modern life in which the Internet is not involved. These influences relate mainly to social networking, economic activity of all kinds, and information gathering. The latter is usually achieved through the use of Internet search engines like Yahoo!, Bing or Google, and the information can easily be tracked and thus observed almost immediately, therefore eliminating the mentioned time lag. Although such information has been analyzed for over a decade, this issue became especially important with the supply of publically available search query data by Internet firms. The rising popularity of Google's search engine, as well as the easy access to and application of its search query reporting tools ''Google Trends'' and ''Google Insights for Search'' (Google I4S), offer new perspectives for information gathering about people's intentions. Accordingly, Google data may have the potential to account contemporaneously for changes in consumer sentiment, and therefore enable the making of inferences about the near future in the real world.

Hence, this study aims to use this new data source in order to overcome the problems of interviewing and to account simultaneously for changes in sentiment. In the light of the latest housing bubble in the U.S., existing information on the Home Buying Search Process (HBSP[1]), in combination with house prices and the volume of transactions, serve as a framework for exploring the explanatory power of this new potential means of sentiment measurement.

The paper is organized as follows: Section 2 summarizes the existing research to date, based on an analysis of Google search query data. After that, the characteristics and behavior of Americans during the process of buying a home are taken into consideration. In this context, the hypothesis is developed that individuals reveal their intentions to Google, which are in turn influenced by and drive the real world. Section 3 describes the generation of Google data, as well as the real-world data used in the analysis. Section 4 provides a detailed analysis of interactions between the real world and Google search query data, as well the implications of Google data for house prices and the volume of transactions. Section 5 checks the robustness of our findings and Section 6 concludes.

Literature Review

Research with Google Data

Theoretical analysis, based on data generated from Internet activity, has been conducted for over a decade and mainly in the field of computer science. However, research using online search query data, has became an emerging discipline since Google first launched its public search query analysis tool Google Trends in 2006, thus providing a tool for the broader public. In summer 2008, Google came out with the even more user-friendly Google I4S, which led again to a growing number of scientific investigations. This present research is the first conclusive review on Google data research (excluding computer science topics), categorized by Google data usage on non-economic considerations, socio-economic topics as well as Google data principles.

Ginsberg et al. (2009) conducted the first academic work based on analyzing Google's search query data. They identify regional illness hot spots in the U.S. through observing search queries related to the most common ''influenza-like'' symptoms. This approach led to the development of a tool (Google Flue Trends[2]) for tracking epidemics, that can be utilized worldwide, assuming that there is a sufficient population of search-engine users in that particular area. Hence, this paper can be seen as revolutionary in its dedication to this new field of academic research.

Constant and Zimmermann (2008) use Google I4S to analyze political and socioeconomic questions, specifically to observe the popularity of U.S. presidential candidates in 2008. Moreover, they are able to identify the financial crisis and recession as fundamental to American election campaigns, as recession and unemployment issues become more and more important. Remaining in the field of socioeconomic research, Della Penna and Huang (2009) demonstrate the predictive power of Google search queries for indentifying consumer sentiment in the U.S. They construct an index with theoretically identified search queries related to consumer behavior. The resulting Google search-based index (SBI) is more up-to-date and predicts the two major American sentiment indices both individually and combined, namely the University of Michigan's Index of Consumer Sentiment (ICS) and the Conference Board's Consumer Confidence Index (CCI). However, the reverse is not true. One month later, Schmidt and Vosen (2009) confirm the predictive power while relating the ICS and CCI to an alternatively derived Google indicator, based on 56 consumption categories, according to the BEA's[3] national income and product accounts. The same month, Kholodilin, Podstawski, Siliverstovs, and Bürgi (2009) also attempt to forecast private consumption with a Google indicator in the U.S. In contrast to earlier Google sentiment articles, they use the Organisation for Economic Co-operation and Development (OECD) consumer confidence indicator, rather than the ICS, and split their sample into a normal and a crisis period. They find that the forecast precision for all indicators is almost the same in a normal economic environment. By contrast, in phases of abnormal economic activity, the Google indicator outperforms the traditional one.

Considering unemployment matters, Askitas and Zimmermann (2009a) investigate the predictive power of Google I4S for the German labor market. Their idea is theoretically to derive keywords that individuals who had just become unemployed would use for their job search. Their results suggest that Google data are a good predictor of unemployment. Applying their fundamental article, Askitas and Zimmermann (2009b) analyze the current unemployment situation for May and June 2009, showing turning points as they appear through the expansion of short-time work. They concur that Internet data are indeed useful, because of its rapid availability and adjustment, as well as its predictive power, even under volatile economic circumstances. Later in 2009, the authors provide a more thorough analysis and compare the predictive power of Google with the German stock market index (DAX). Askitas and Zimmermann (2009c) conclude that Google data alone outperform the indications of the DAX in terms of forecasting unemployment, but they obtain the best explanatory power by using a combination of the two types of data. Choi and Varian (2009b) apply the standard approach from their recent paper (Choi and Varian, 2009a) to the U.S. unemployment time series, establishing that data from Google Trends are a useful tool for predicting initial claims. D'Amuri (2009) demonstrates the implementation of a Google unemployment indicator, using the example of the Italian labor market. With its quarterly reporting interval, Italy differs from the German and the American monthly frequency, in that its sample is consistently smaller. Nevertheless, using a Google indicator in predicting unemployment in this particular case, also leads to estimation and forecasting gains in accuracy, compared to traditional indicators. Furthermore, Google data inclusive models for small samples, perform better than identical conventional models that use a sample collected over an even longer period of time (with or without leading indicators). D'Amuri and Marcucci (2009) build on the above research by assessing the unemployment forecasting accuracy of online search queries using the traditional ''Initial Jobless Claims'' similar to Choi and Varian (2009b) (as a leading indicator), as well as a Google index (GI) (both in varying combinations), finding that models augmented with the GI significantly exceed traditional models in terms of forecast accuracy as well as predictive ability, and that the GI is ''the best leading indicator to predict the U.S. unemployment rate.''[4]

After the potential offered by Google data became more established, Choi and Varian (2009a) wrote a fundamental article on the use of search query data for research matters. Their main objective is to clarify questions for researchers interested in conducting academic work with Google data, which are beyond the scope of the information provided on their website (i.e., forecasting methods or architecture of the data). They find that adding relevant Google data outperforms forecasting models that abandon Google predictors. Nevertheless, they point to the existence of sampling errors, caused by possible day-to-day variations in the database. Likewise, Suhoy (2009) tests the predictive power of Israeli query indices, starting from the published nowcasting model of Choi and Varian (2009a). Beside some valuable prediction results (especially for the ''Human Resources'' category), she also points out some potential problems associated with using search query data, such as non-stationarity, and varying predictive power in the long run. In addition, she refers to the state of competition between social-search (e.g., Facebook, Twitter), knowledge-search (Q&A sites[5]), and Google searches. Focusing on the fundamental predictability of search trends with Google query data, the results of Shimshoni, Efron, and Matias (2009) suggest predictability[6] for more than half of the most popular Google search queries, because aggregated search queries are more predictable than individual ones and approximately 90% of the categories include predictable queries. Their article is also part of the Google project implementing an automatic forecasting option for predictable trends into Google's I4S tool. In conclusion, Tierney and Pan (2009) consider both the benefits and problems emerging from the use of Google search query data, especially with regard to the way the data are rehashed, prepared, and provided. Extending on previous articles, they state that Google data, as directly measured consumer data, are helpful for the avoidance of time lags in predicting economic activity. On the other hand, they point out some privacy issues with respect to the raw data and indicate that, through the varying normalization and scaling process at each aggregation level, the interpretation of estimation results might be ambiguous.

Home Buying Search Process

When it comes to an investment in their own property, the behavior of Americans is not really comparable to that of other countries. Americans go where the jobs are and where they can most likely afford home ownership, that is, they are relatively mobile and flexible. According to the annual American Census Geographical Mobility/Migration report, 37.1 million people changed residence in 2009 in the U.S. (equal to 12.5%).[7] These statistics indicate that the average American moves 11.7 times during his lifetime.[8] The Profile of Home Buyers and Sellers report for the U.S. is published annually by the National Association of Realtors (NAR) and is used to verify the assumptions in this paper. The currently available report for 2009 gathered information from 9,138 home buyers (out of a sample of 120,038 interviewees) who purchased a home between July 2008 and June 2009, and is therefore one of the largest national questionnaire-based surveys with an in-depth analysis of the HBP across the country.[9]

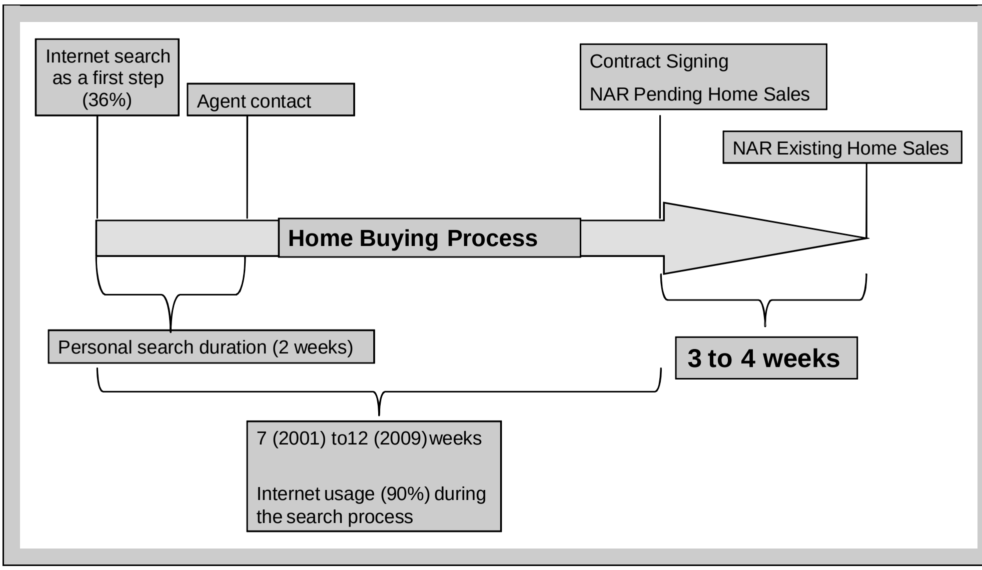
Initially, the decision process with respect to buying a first owner-occupied home is driven by many individual decisions, such as marital status, job situation, economic environment, net wealth, and so forth. Decisions to buy as an investment or other repeat-buying decisions differ slightly from those of a first-time buyer. The four most influential drivers for first-time buyers are the desire to own a home, affordability, a change in the family situation, and making use of a first-home-buyer tax credit. This also relates to repeat-buyers for whom a job-related relocation or move, desire for a larger home, to own a home, and a change in the family situation are of prime importance. This pre-decision process is associated with the gathering of information about the HBP (i.e., from friends, Internet search engines or Q&A sites, newspapers, specialized institutions, real estate agents, home buying seminars, etc.). Once the evaluation has been completed and the decision to buy a home made, various channels are available for finding the desired home. According to the NAR, the first steps taken during the HBP by all buyers (first-time and repeat) are, among others, to look online for properties for sale (36%), contact a real estate agent (18%), look online for information about the home-buying process (11%), or contact a bank or mortgage lender (8%). Considering the information sources utilized during the entire HBP, the use of Internet sources (90%) is the leading one, followed by the real estate agent (87%), and yard signs (59%), taking all buyers into account. The actions that follow the Internet search are to inspect the home of interest (77%), gather more information by viewing the home online (61%) or to contact the agent responsible for the property (28%). The time a buyer spends searching the Internet before contacting an agent is, on average, two weeks. Traditional media such as newspapers, magazines or home buying guides (3% as a first step) and print newspaper advertisement (40% over the entire HBP) are declining in usage and significance. The proportion of buyers who find the home they purchase through the Internet (36%) is the same as for real estate agents (36%). However, it has to be emphasized that the relevance of a real estate agent for the HBP remained stable in recent years, while Internet use was rising. In terms of the total length of searching, Internet searchers need twice as long (12 weeks) before the purchasing decision is made, with no differences between first-time and repeat buyers, and also independent of the involvement of an agent. The overall search duration increased from seven weeks (2001) to eight weeks (2003 to 2007) and 10 weeks in 2008. Moreover, Internet searchers actually visit three times as many homes (15), compared to those who do not use the Internet.[10] However, apart from the rising use of the Internet in the HBP, buying a house is still associated with consulting a real estate agent. To date, 79% of buyers who use the Internet to search for a home still ultimately purchase their home through a real estate agent (77% independently from the Internet search).[11] Concentrating on Internet searchers in the HBP, among others, mostly Multiple Listing Service (MLS) websites (60%) and real estate company websites (46%) are used during the search process.

Exhibit 1 illustrates that HBP lasts between 10 and 16 weeks from the beginning of the HBSP, contract signing, and the existing home sale. Moreover, after the actual sale takes place, it needs a further four weeks until NAR reports its own gathered transaction data publically.

Determinants and Characteristics of the Housing Market

Because the Google data are fitted to real world data, two variables and the existing theory on their relationship are addressed in more detail, namely house

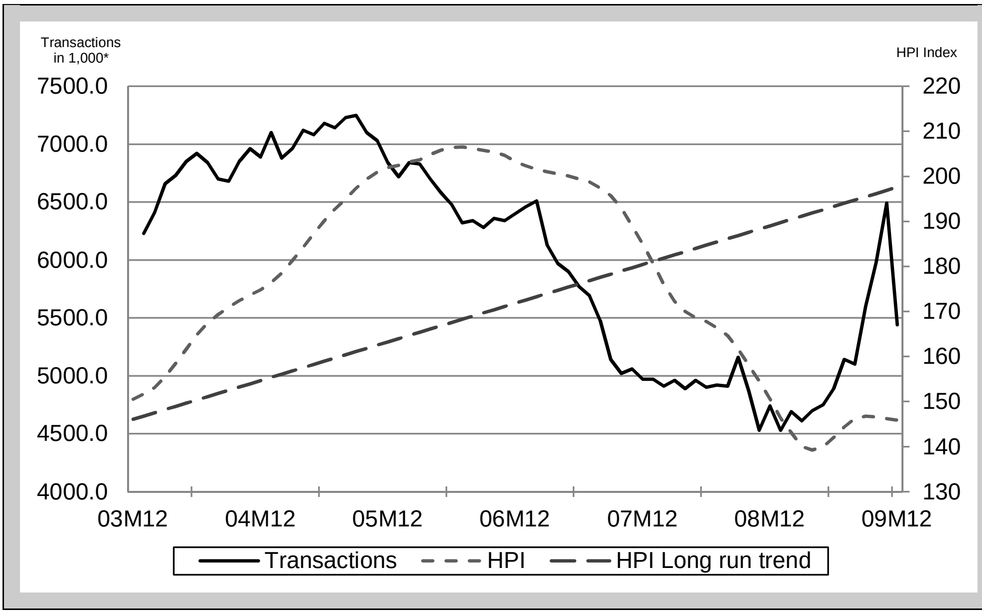
Exhibit 1 | Home Buying Process



prices and the volume of transactions. Beside Stein's (1995) model, which is based on the so-called down payment effect, as well as the approach of Genesove and Mayer (2001), who argue in terms of the loss aversion of homeowners, the work of Berkovec and Goodman (1996), promoting the so-called search model seems to fit best to this study. The main reason is that all three models agree on a positive price-volume relationship, but only the search model predicts that the volume of transactions leads prices.[12] Exhibit 2 confirms this perspective for the underlying sample period, where transactions decline from August 2005 onwards while prices still increase until July 2006.

A number of recent studies on this topic are of concern for this study. Adopting the search model, Novy-Marx (2009) argues that a first-order impact of a demand shock might be smoothed by price adjustments, but changes in the relative shares of market participants (buyer-to-seller ratio) and their expected value from market participation (second order impact) tightens overall market conditions. In the present paper, actual market participants are a subset of Google searchers, because Google accounts for those who are going to buy, as well as those interested in buying, but who decide to delay the purchase for any reason related to market conditions or private and unobservable circumstances. According to Arbel, Ben-Shahar, and Sulganik (2009), the correlation between prices and volume depends on the state of the market with respect to the interaction of house-price returns and the long-run mean of house prices.[13] Recently, Shi, Young, and Hargreaves (2010) also find support for the search model based on disaggregated data, but emphasize the applicability of the two alternative theories for some local markets.[14]

Exhibit 2 | House Prices and Volume of Transactions

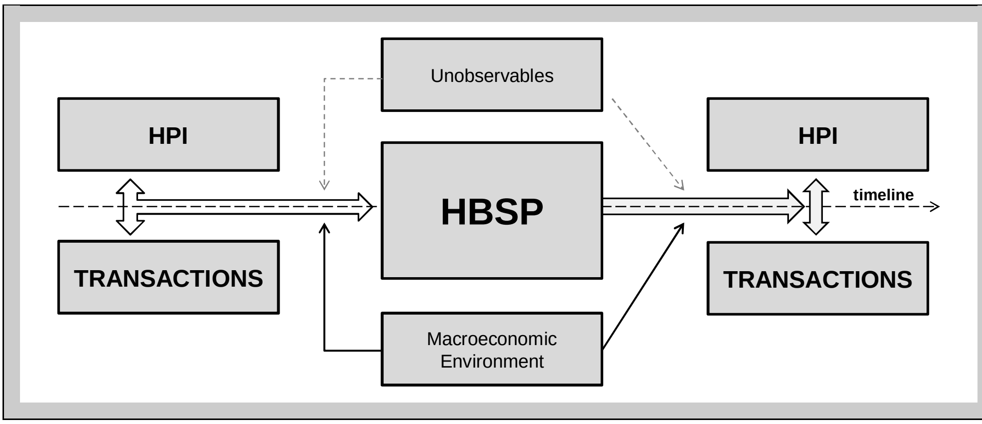


* Seasonally adjusted and annualized by the NAR.

In the light of the arguments in the introduction, with respect to euphoric behavior in the housing market, the issues relating to the specific sample period need to be addressed further. Starting in the late 1990s on the West Coast, the boom spread throughout the U.S. until Phoenix experienced a price rise of 43% in 2005. The appearance of a bubble was clear, because the economic fundamentals did not justify the price increases. Real rents and construction costs matched one another, but neither with prices.[15] But what drives people to continue investing in houses, specifically as a second home for investment, even when such behavior no longer seems rational? Shiller (2007) argues that it is the fear of decreasing affordability in a rising house-price environment, which is only mitigated partly by policy programs. Stevenson (2008) argues that individuals perceive the implicit costs of not reaping the benefits of house price appreciation. Both arguments highlight the relevance of a belief in further appreciation, which render an individual's decision rational and logically comprehensible (even if imprudent) with adaptive expectations. Clayton, MacKinnon, and Peng (2008) emphasize again the role of sentiment among market participants for market liquidity. Recent research highlights the need for reliable indicators, capturing market fads and therefore the great potential of real-time information as provided by Google I4S.

Taking these arguments together and those from the previous section, the time setting can be formulated. Households might, among other things, be driven by prices, as well as by transactions, and behave in such a way as to drive the real-world data (Exhibit 3).

Exhibit 3 | Model Time Setting



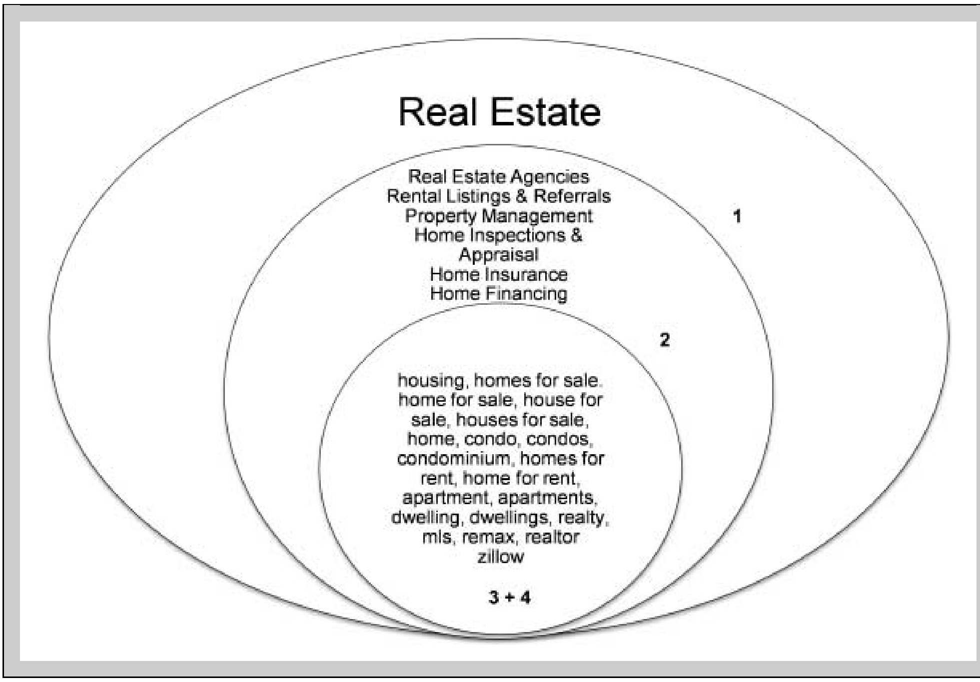
Data

Data Extraction from Google Insights for Search

In order to gather appropriate search query data for the empirical analysis, the following approach is employed.[16] The data are extracted at the unspecified search query level (without a category relation), with regard to logically derived keywords, separately for each time period from January 2004 to December 2009 (on a weekly basis).[17] At that level, the results are relatively undirected, due to the huge variety of meanings and relationships with regard to certain search queries, and are therefore specified by using the categories suggested by Google.[18] In addition, the derived categorized single queries are investigated further with respect to their singular/plural notation, by comparing them to one another. Appendix 2 shows their relative relevance, as well as their statistical characteristics and relation to each other. It follows that households generally search in plural terms (i.e., searches for ''homes for sale'' exceed the query ''home for sale'' six times). The same approach is used to identify typical search patterns in the HBP (e.g., houses for sale or homes for sale) resulting in the variables considered in the empirical analysis. In both cases, the ''impact factor''[19] serves as the decision criterion, which indicates the average volume of a search query over the selected sample period.

Beyond the use of single queries, the top-level category ''Real Estate,'' as well as its six related additional subcategories are generated.[20] Exhibit 4 illustrates the initially derived search queries and categories, as well as the selection process and final data set.

Exhibit 4 | Hierarchy of the Search Query Data



Note: The data is structured top-down as follows: 1. top level category, 2. second level category, 3. single search query data, 4. single search query data (categorized with Real Estate).

Real World and Macroeconomic Data

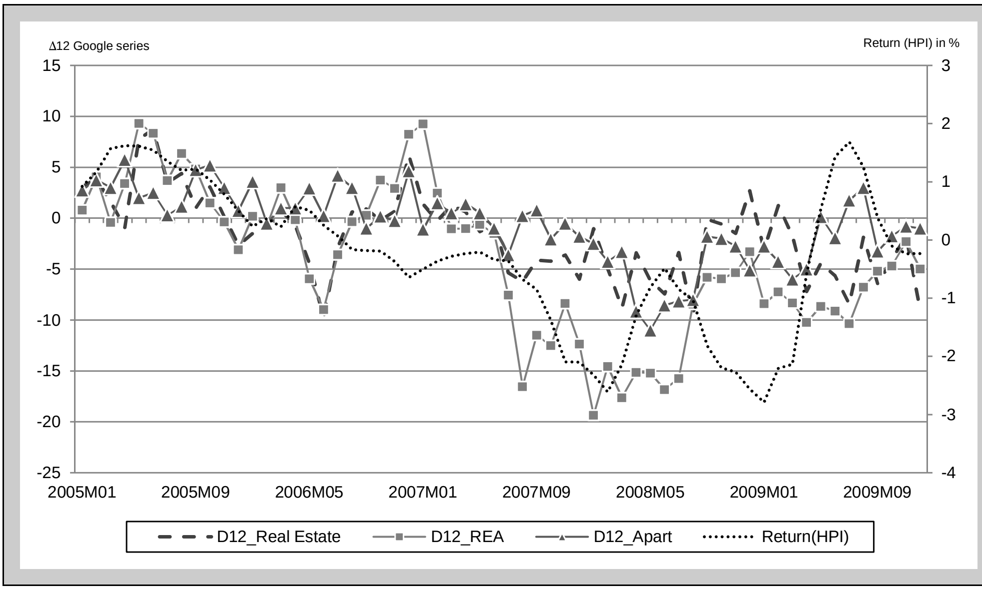
The S&P's Case-Shiller Index Composite for 20 MSAs (DS code USCSHP20F) serves for the analysis of Google data on prices. To consider houses and apartments, the unadjusted series of existing home sales for single family and condominium (DS code USEXHOMEP) are used. At a later point in the analysis section, the market index S&P's 500 composite (DS code S&PCOMP), disposable income (DS code USPERDISB), total employment (DS code USEMPTOTO), and the FHA effective mortgage rate (DS code USMEGFH) are introduced to account for overall market conditions.

Analysis

Preliminary Steps

All Google data, as well as transactions, exhibit an obvious seasonal pattern.[21] In order to account for seasonality, two approaches are considered: The first is the

Exhibit 5 | Selected Time Series Adjusted by the 12th Difference

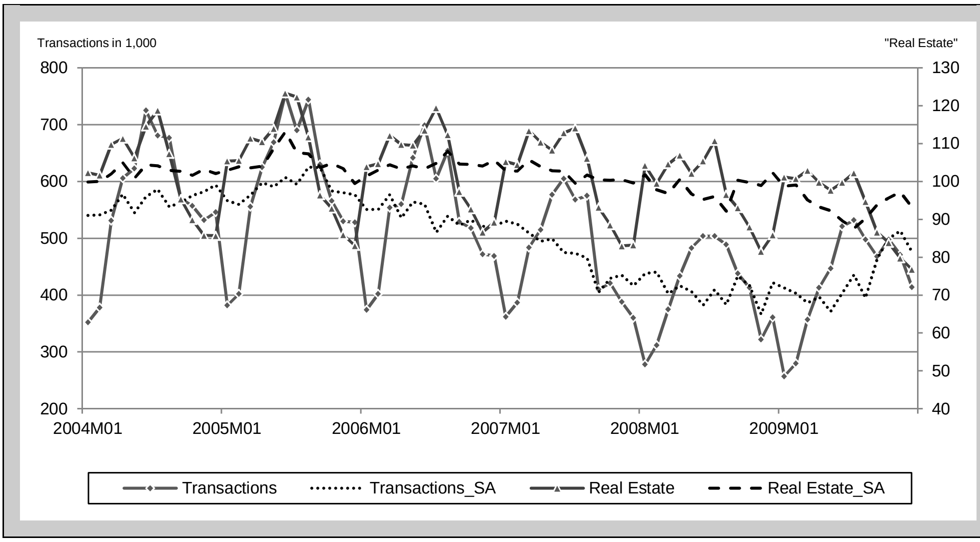


12th difference of each variable [i.e., Askitas and Zimmermann (2009a)], the other is the transformation by the Census X11 ARIMA model. While the former transformation is associated with a further reduction in the already small sample by twelve observations, the advantage is the intuition behind this approach, as well as the fact that this transformation leaves the time series with more information in terms of variance.

Exhibit 5 illustrates the relationship between the 12th difference transformations (scaled on the left ordinate) of Real Estate (top level category), Real Estate Agency (subcategory), and Apartments (single search query) to house price returns (scaled on the right ordinate). A further advantage of the D12 series would be that they are already stationary, while the adjustment by Census X11 additionally requires first differencing, in order to obtain stationary time series. At first glance, all series are slightly above the time axis during the boom and fall along with returns during the bust (Exhibit 2). However, the D12 series seem to be technically stationary at a 5% or 10% level, but econometric reasoning raises doubt that reliable inferences can be drawn from these data. Hence, this approach is rejected, accepting the greater loss of information from the Census X11 and first differencing, in order to obtain a stationary series, and moreover, allowing for a change in seasonal factors over time, as well as to considering a trend in the series.

Exhibit 6 compares the level data with the seasonally adjusted data of the Real Estate category (right ordinate) and transactions (left ordinate). First comparing the unadjusted series, one might suppose that Real Estate tends to lead transactions

Exhibit 6 | Real Estate Category and Volume of Transactions in Level and Adjusted by the Census X11



when comparing the peaks and lows. Therefore, the next step is to perform Granger Causality tests, to get a feeling for the lead-lag-structure and interactions of the Google data, house prices, and transactions.

Because Granger Causality tests are sensitive to the lag specification, the lead-lag-relationships are analyzed separately by lags, as suggested by the Schwarz and Akaike criterion.

The basic model for testing Granger Causality is:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_{p1} y_{t-p} + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \varepsilon_t, \tag{1}$$

with $\varepsilon_t = error_t$ and p is the selected lag order with respect to an information criterion, e.g., AIC Granger Causality null-hypothesis: H_0: $\alpha_1 = \cdots = \alpha_p = 0$.

However, Granger Causality tests do not address attributes with respect to the exact lag structure of two variables (assuming that there is no constant and predefined gap), and the direction of the correlation as well as the significance of contemporaneous relationships (Granger Causality tests only for joint significance of lagged variables). Therefore, the next step in the analysis is to simply regress all possible combinatorial relationships on each other using the following approach:[22]

Exhibit 7 | Lead-lag Structure of House Prices, Transactions, and the Google Data

Dependent Variable	Independent Variable	Lag (G, AIC)	Lag (max. t-stat)	Relationship
House Price	Transactions	4	3	+
Transactions	**Real Estate Agency**	2	1	+
Real Estate	House Price	4	0	–
	Transactions	2	0	+
	Real Estate Agency	2	0	+
	Apartments	5	4	+
Real Estate Agency	House Price	4	4	+
	Transactions	2	2	+
Homes for Sale	Transactions	3	3	–
Apartments	House Price	4	3	+
	Transactions	2	3	–
	Real Estate Agency	1	0	+
	Home Financing	1	–	0

Note: The first column variable is Granger caused by second column variables. The second column variables in bold are insensitive to the lag specification criterion. The third and fourth column show the lags suggested by AIC in the Granger Causality tests and the maximum t-stat of the simple OLS regressions. The last column shows the direction of the relationship.

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \beta_4 x_{t-i} + \varepsilon_t, \qquad (2)$$

for $i = \{0, 1, 2, ..., 11\}$.

Exhibit 7 summarizes the combined results of these two steps.[23]

As expected, transactions Granger cause house prices, but the reverse is not true (Exhibit 2).[24] The remainder of the discussion concentrates on the reasoning behind dropping the Real Estate and Homes for Sale categories, while retaining the Real Estate Agency and Apartments series. The case of Home Financing is somewhat puzzling, but leads an interesting discussion at the end of this section.

The Real Estate results are the converse: First, it is caused by house prices, transactions, Real Estate Agency and Apartments, but it only Granger causes the latter. Second, the negative contemporaneous relationship of house prices to the Real Estate category turns out to be positive in a higher lag structure.[25] These results can be explained by a very high noise factor in the top-level category (e.g., a rise of the search query ''subprime'' would lead to an increase in the category as a whole, although prices decrease). Finally, the Homes for Sale search query appears to have no implications for any of the time series.

The Apartments search query is, among other variables, Granger caused by house prices and transactions, but differently signed. A rise in the volume of transactions seems to lead to a saturation of the market for apartments. A rise in prices leads to lower home ownership affordability and drives the demand for apartments. This implicitly assumes that the search query Apartments is largely influenced by searches for rental apartments.[26]

By contrast, the results for Real Estate Agency seem to be more consistent. This only affects transactions and is caused by prices as well as transactions. Additionally, the (un)reported OLS-regressions never contradict to this consistency.

Although Home Financing is not caused by any variable considered here and only causes Apartments, it has a contemporaneous impact on transactions and on house prices, again, differently signed with regard to the direction of relationship in the OLS regressions. That is, a positive impact of Home Financing on transactions can be explained by the fact that the specific financing decision is the last element of the HBP before the transaction takes place and therefore, a monthly frequency cannot detect a lag structure. By contrast, the case of the Home Financing subcategory on prices is ambiguous. Initially, the correlation between Home Financing and the mortgage rate is negative. Therefore, the refinancing decisions of households, due to a mortgage rate decrease, seem to be associated with a higher interest (preference) for Home Financing. Assuming that individuals extend their outstanding loans, rather than reducing their interest payments, such refinancing does not affect prices.[27] However, considering the sample period, one of the top search queries in this subcategory is ''foreclosure'' or the rising query ''reo.'' Thus, taking this variable as an approximation of the amount of stressed mortgages justifies the negative impact of Home Financing on house prices in the sample.[28]

These preliminary steps indicate the direction of the various relationships, as well as the lag structures. Most importantly, Google data are not only able to explain house prices and transactions, but are also driven by them. This can be explained simply by the fact that people are certainly driven by market occurrences and therefore their interest in properties is, among other things, both dependent on and a result of price-volume dynamics. This introduces an endogeneity problem, especially in the light of a time set-up according to Exhibit 3.

Modeling the Home Buying Process

A VAR model is employed to address this endogeneity issue. According to the results presented in the previous section, the variables Real Estate and Homes for Sale are dropped. Thus, basically, apart from house prices and transactions, the relevant variables are Real Estate Agency, which is assumed to be a very robust indicator for transactions, Home Financing, because of its contemporary impact on house prices, as well as its technical exogeneity in the Granger Causality tests,

Exhibit 8 | Relative Explanatory Power of Google and Real-world Data for Transactions

R-squared	0.245	0.263	0.306	0.442	0.330	0.501	0.515
Adj. R-squared	0.209	0.177	0.225	0.344	0.199	0.370	0.337
Akaike AIC	−3.028	−2.935	−2.994	−3.125	−2.913	−3.120	−3.030
Schwarz SC	−2.897	−2.674	−2.733	−2.766	−2.521	−2.630	−2.410
House Price			x		x	x	x
Google-Data				x		x	x
Macro-Data		x			x		x

Exhibit 9 | Relative Explanatory Power of Google and Real-world Data for House Prices

R-squared	0.407	0.434	0.470	0.554	0.503	0.577	0.600
Adj. R-squared	0.380	0.369	0.409	0.475	0.405	0.466	0.454
Akaike AIC	−8.595	−8.526	−8.587	−8.670	−8.532	−8.607	−8.545
Schwarz SC	−8.466	−8.266	−8.326	−8.311	−8.140	−8.117	−7.925
Transactions			x		x	x	x
Google-Data				x		x	x
Macro-Data		x			x		x

and Apartments, which covers the searches for rental properties.[29] Furthermore, the S&P 500, disposable income, employment, and the mortgage rate are used to account for overall macroeconomic conditions.[30]

The following section compares the explanatory power of three sets of variables for house prices and transactions, apart from their own lags: the just introduced macroeconomic variables, the Google data, and transactions for house prices and vice versa. Exhibits 8 and 9 show the changes in information criterions for different included variables.[31] In this step, all specifications are included in a VAR3 model with the macroeconomic data and Home Financing treated as exogenous. Furthermore, house prices are included in Exhibit 8 (transactions in Exhibit 9), so as to control for a contemporaneous price-volume relationship.[32]

Focusing on the adjusted R-squared, one observes a decrease in the goodness-of-fit by including the macro data only. House prices marginally increase the explanatory power by 1.6 percentage points, but Google data have a huge impact, pushing up the adjusted R-squared by 13.5 percentage points.[33] However, as concluded in previous research with Google data, the best result is obtained by the combination of Google data and house prices. Based on this, the lowest Akaike values are obtained by using Google data only, because of its higher "penalty" for additional regressors, compared to the adjusted R-squared. This finding is confirmed even more strongly in Exhibit 9 for house prices.

Again, the macro data do not improve the model and according to the previous results, transactions add some explanatory power, but the increase in the adjusted R-squared caused by including only the Google data is again noticeably higher. Contrary to the transaction analysis, no other combination of the sets leads to a higher adjusted R-squared (lower Akaike respectively) and Google data alone provide the best goodness-of-fit. However, the statement that Google data outperform real-world data must be interpreted with caution. The fact that macroeconomic data cannot explain extreme movements of house prices is not surprising. It might well be the case that the analysis of normal economic conditions leads to the conclusion that a combination of both, real world and Google data, performs best.[34]

Combining the results of the analysis so far yields the determination of the final model. For a start, a secondary result of the last step is that the contemporaneous price-volume relationship can be neglected. This is also the case for the macroeconomic variables. However, they are included in the system as they have explanatory power for the Google data. Finally, the mortgage rate, which is (highly) negatively correlated with Home Financing is redundant.[35] In order to capture the entire HBSP a VAR-4-model (= 16 weeks) is employed. A lag exclusion test suggests a need to drop the third lag. Hence, the architecture of the model looks as follows:

$$\begin{pmatrix} P_t \\ Q_t \\ G1_t \\ G2_t \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ \beta_{04} \end{pmatrix} + \mathbf{A} \begin{pmatrix} P_{t-i} \\ Q_{t-i} \\ G1_{t-i} \\ G2_{t-i} \end{pmatrix} + \mathbf{B} \begin{pmatrix} G3_t \\ Emp_t \\ Inc_t \\ SP500_t \end{pmatrix} + \begin{pmatrix} u_{Pt} \\ u_{Qt} \\ u_{G1t} \\ u_{G2t} \end{pmatrix}. \tag{3}$$

for $i = \{1, 2, 4\}$; P_t refers to house prices (Equation 1), Q_t to transactions (Equation 2), $G1_t$ to Real Estate Agency (Equation 3), $G2_t$ to Apartments (Equation 4), $G3_t$ to Home Financing, Emp_t to employment, Inc_t to income, and $SP500_t$ to S&P's 500 in t. **A** is a 12×4 coefficient matrix of the endogenous variables, **B** is a 4×4 coefficient matrix of the exogenous variables, u_{xt} are the error terms of the respective equation. See Appendix 8 for the detailed estimates.

In Equation 1 for house prices, the fourth transactions lag has a significant impact. Apartments and the Real Estate Agency show insignificant (negative or positive) and significant negative estimates, which is puzzling. The issue is addressed at the end of this section. As expected from the previous step in the analysis, Home Financing is negatively related to house prices. In Equation 2 (transactions), the Real Estate Agency has a significant positive impact. This estimate is not only very robust in all of the specifications so far, but is also in line with the duration of the HBP suggested by the NAR survey. Additionally, the positive relationship between Home Financing and transactions can be explained by the fact that the financing decision is (temporally) very close to the actual transaction where a monthly frequency in the data cannot detect any lag relationship. Finally, Real Estate Agency (Equation 3) is positively driven by transactions and prices, while in Equation 4 for Apartments, house prices are positively signed, but only transactions are significant. At this point, the initial time setting of Exhibit 3 turns out to be valid: house prices and transactions, among other things, drive the rational of households, recorded by Google I4S, which in turn drives the housing market dynamics. With respect to the impact of the macroeconomic variables on Google data, employment is positively related to Real Estate Agency and Apartments, while income drives Apartments only. The negative direction of the latter relationship can be explained by a higher preference to rent an apartment, after disposable income decreases. This argument is in line with the negative impact of Apartments on house prices, but not aligned to transactions. Hence, these results again suggest that Apartments is driven mainly by a preference to rent an apartment, while Real Estate Agency seems to be a quite robust indicator of transactions.

Taking all findings into consideration, five conclusions can be drawn: First, Google data improve the quality of explaining house prices, but the impact of the lagged variables is not clearly directed—probably a result of the extreme market environment investigated. Second, the relationships between Home Financing and different variables with respect to lending in the housing market yield useful insights into the dynamics of the HBP, both, quantitative and qualitative. Third, Real Estate Agency serves as a very good predictor of transactions. Also, assuming a lasting effect of transactions on house prices, this time series has implications for the overall housing market. Fourth, the disadvantage of informational time lags described in Exhibit 1 can be solved by using real-time search query data. Finally, the time setting introduced in Exhibit 3 turns out to be empirically valid: Housing market dynamics influence search query data, which in turn influence the real world.

Robustness Checks

This section consists of two steps, which both challenge the findings and solve a potential shortcoming.[36]

Although theory based on the NAR survey allows the assumption that interactions between Google data and transactions do not exceed a three-month time span, the interactions between the two sets of variables and prices is not clear. This might especially be the case for prices and transactions. Accordingly, the analyses use a maximum lag structure of 12 months.[37] The AIC criterion suggests a lag structure of eight. While it does not make sense to estimate a model with 32 variables, the sequential elimination of regressors strategy is used to reduce the number of estimates. For this reason, the information criterion of interest (here:

AIC) is used to decide whether a regressor should be included in the model (see Lütkepohl and Krätzig, 2004, p. 122). The top-down procedure of this sequential elimination starts with the highest lag and determines whether deleting a regressor improves the information criterion. As suspected, prices and transactions interact with each other in a higher (six-to-eight) lag order. Moreover, although all estimators show the expected sign and significance, a negative impact of the sixth lag of transactions on Real Estate Agency is an exception. However, the inconsistency can be regarded as minor, since this direction of causality is not the focus of the research. Otherwise, the impact of house prices on Apartments is positive up to lag eight. While lacking in significance in the VAR model, this presumption stems from the simple OLS regressions and now confirms the affordability argument statistically.

In comparison to previous research, one might claim that the data-generating process is too restricted with respect to the potential set of variables. Previous studies of Google I4S constructed indices based on a variety of different search queries, but the findings in this study are basically reduced to the Real Estate Agency series. However, the exact construction of these indices is not generally transparent in terms of reproducibility. Google constructs its own Real Estate Index: ''[It] tracks queries related to 'real estate, mortgage, rent, apartments' and so forth.''[38] This index is transformed equally to the other series (seasonal adjustment and first log differences) and replaces Real Estate Agency. It turns out to perform similarly to Real Estate Agency: the significance of the first two lags for transactions, but no explanatory power for house prices. In the direct comparison, Google's Real Estate Index performs worse than single Real Estate Agency variable.

A final remark concerns a potential selections bias, because the analysis is restricted to Google searches only. The authors are not aware of any competitive alternatives to Google I4S (i.e., Yahoo! Buzz-Index or Bing! xRank). Furthermore, Google has a worldwide market share of 85.15% (Yahoo! 6.33%).[39] Hence, a sufficient representativeness of the investigation may be assumed. A question of greater concern is the growing importance and competition between Google and social networks (e.g., Facebook, Twitter) or knowledge search (Q&A sites) and their implications of potential biases (Suhoy, 2009). Yet, to date, there is no adequate information on how individuals use these sources in comparison to search engines.

Conclusion

To the best of the authors' knowledge, this is the first real estate-specific study based on Google I4S search queries. Similar to previous research, a combination of Google and real-world data yield the best specification of empirical models. However, Google data in fact add significantly more explanatory power to the models than the real-world data. This, together with the absence of a reporting lag, renders them valuable in making inferences about the near future, hence improving the efficiency of a relatively fragile market. These findings complement and can improve the decision-making processes of real estate professionals and policymakers.

The initial time setting derived from the existing literature and real-world interviews from the NAR are verified empirically. Individual interests and preferences are driven by—and drive—the real world in a reciprocal interaction. During the entire analysis, Real Estate Agency endures as a very robust indicator of transactions. That can be explained by the fact that the aggregation level is not as undirected as Real Estate, but is still abstract enough, compared to special search queries. This is also true for Home Financing, which yields interesting implications for the different financing aspects of the housing market. Furthermore, the search duration suggested by the NAR surveys fit very well with the investigation, even when a higher lag structure is considered.

Finally, the results must be seen in the light of the extreme market conditions present in the sample. Nonetheless, the findings might also to be valid under normal economic circumstances. Further research could usefully be conducted at a disaggregated level for the U.S., because existing surveys offer cross-sectional information as well. Especially when larger samples of Google data are available, research on the identification of special search queries indicating a decreasing soundness of the housing sector (e.g., ''foreclosure'' or ''reo,'') could help detect bubbles in the future. In general, the use of search query data can be extended to any research question involving consumers.

Concerning developments of Google I4S, additional tools further simplifying the analysis and extraction of data seem very promising, especially for final consumers (e.g., a tool for seasonal adjustment of the data). With respect to the real estate market, individual indices for housing and retail adapted to real-world relations would allow a simultaneous evaluation of changes in individual decisions and rationales for buying. However, at the same time, the future of this new field of research should also be seen in the light of some rather delicate privacy issues associated with this source of information.

Appendix 1

Google Insights for Search Description

Google is the most frequently used search engine *worldwide*. It provides aggregated historical logs of online search queries with a time series of weekly counts. Since 2006, the data has been publically accessible through the Google Trends module and since summer 2008, through Google Insights for Search (beta version). The main advantage of Google I4S, compared to Google Trends, is the user friendliness, without any differences in the underlying data set. Both provide the option of downloading the data as a .csv data file for further editing. The search query data, depending on availability, have been retrievable on a weekly basis at worldwide, national, state, sub-regional or MSA level since January 2004. Beside the normal single-query keyword search, the data can alternatively be searched by physical location [using the Internet protocol (IP)] for each query, or for different time frames. Beyond the option of comparing up to five different search queries, keywords can be compared within a search placeholder, by combining the words in common ways (by space, +, ‘‘). Subsequent to the analysis of a single search query, the search process can be specified by a suggested, corresponding category.[40] In addition to searching for a specific search query, the data can also be searched directly, isolated for 27 top level categories and 241 categories at the second level (subcategory), or a certain search query can be specified by using a related category.

Once the analysis for the desired search queries is complete, Google I4S provides several items of information. The chosen search query trend is shown in a chart, where the original Google search data is normalized and scaled from 0 to 100, compared to the total volume of searches, for the selected region and time period.[41] Hence, it might be the case that requesting the data for the same keywords are not necessarily identical over time, once a new relative maximum is reached. Moreover, some additional results, such as the regional distribution (numbers/ visualized in a geographical heat map), changes in time (numbers/visualized), top search queries and queries with the highest growth rates can be observed (assuming a sufficient search volume). The total data generation is subject to the Google data privacy protection. As opposed to the single query results, the chart for categories is depicted in percentage changes (starting with 0), with the entire series relating to the starting point. Furthermore, applying a category or a category filter provides the option of switching between an ‘‘interest-level’’ view and a ‘‘Growth relative to the selected category’’ view.

Appendix 2

Search Query Selection: Single versus Plural

	Impact Factor	Mean	Std. Dev.	Correlation
Home for Sale	13	76.87	10.97	0.80
Home for Sale	77	69.23	10.12	
House for Sale	25	68.43	10.69	0.95
House for Sale	77	65.13	10.61	
Home for Rent	12	64.76	14.74	0.94
Homes for Rent	72	66.95	16.77	
Apartment	31	68.27	14.62	0.76
Apartments	79	78.58	10.51	
Condo	72	71.57	14.15	0.81
Condominium	13	59.58	19.40	

Notes: The search queries are in the original level. The impact factor shows the relationship between the two corresponding search queries.

Appendix 3

Correlation Matrix

	Emp	SP500	MRATE	INC	HPI	Trans	RE	REA	HF	HfS	Apart	GREX
Emp	1.00											
	—											
SP500	0.16	1.00										
	1.21	—										
MRATE	0.02	−0.10	1.00									
	0.18	−0.85	—									
INC	0.18	0.27	−0.04	1.00								
	1.41	2.34	−0.35	—								
HPI	−0.02	0.23	0.16	0.10	1.00							
	−0.13	1.98	1.38	0.82	—							
Trans	0.02	0.04	−0.10	−0.03	−0.16	1.00						
	0.18	0.33	−0.81	−0.25	−1.30	—						
RE	0.20	−0.13	−0.14	−0.12	−0.18	0.42	1.00					
	1.56	−1.05	−1.17	−0.98	−1.50	3.82	—					
REA	0.21	0.12	−0.10	0.02	−0.16	0.13	0.46	1.00				
	1.62	0.99	−0.86	0.13	−1.33	1.08	4.32	—				
HF	0.06	−0.14	−0.35	−0.05	−0.24	0.43	0.66	0.20	1.00			
	0.44	−1.18	−3.11	−0.44	−2.05	3.95	7.33	1.73	—			

Appendix 3 (continued)

Correlation Matrix

	Emp	SP500	MRATE	INC	HPI	Trans	RE	REA	HF	HfS	Apart	GREX
HfS	0.00	−0.01	−0.11	0.04	0.05	0.02	0.28	0.46	0.12	1.00		
	0.02	−0.06	−0.88	0.33	0.43	0.19	2.44	4.33	1.03	—		
Apart	−0.02	−0.03	−0.12	−0.15	−0.11	−0.09	0.22	0.13	0.13	0.39	1.00	
	−0.12	−0.25	−1.02	−1.30	−0.92	−0.75	1.85	1.12	1.10	3.50	—	
GREX	0.20	0.06	0.00	0.08	−0.11	0.11	0.45	0.71	0.20	0.39	0.08	1.00
	1.51	0.43	0.01	0.62	−0.82	0.80	3.75	7.60	1.51	3.18	0.63	—

Notes: HPI: S&P's Case-Shiller Index for 20 MSAs, Trans: Transactions reported by the NAR including single family and condominiums, RE: Google I4S Real Estate category, REA: Google I4S Real Estate Agency subcategory, HF: Google I4S Home Financing subcategory, HfS: Google I4S Homes for Sale specific search query, Apart: Google I4S Apartments specific search query, GREX: Google Real Estate Index. S&P's 500 composite (Datastream code S&PCOMP), disposable income (USPERDISB), unemployment rate (Datastream code USUN%TOTQ) and the mortgage rate (FHA MORTGAGE RATE, Datastream Code USMEGFH). Values in table are *t*-stats.

Appendix 4

Granger Causality Tests

Independent Variables	Dependent Variables HPI	Trans	RE	REA	HF	HfS	Apart
Panel A: *F*-stats of joint significance							
HPI	—	1.627	2.737**	4.323***	1.505	1.228	4.388***
Trans	2.180*	—	3.510**	4.162**	4.541**	3.273**	2.835*
RE	0.352	0.009	—	0.055	0.048	1.709	0.741
REA	0.834	7.949***	8.135***	—	2.719	—	—
HF	1.392	0.626	0.069	2.257	—	1.231	2.931*
HfS	0.013	1.811	1.277	—	0.538	—	—
Apart	1.029	1.066	4.283***	—	0.936	—	—
Panel B: AIC value and the suggested lag for the respective pair							
HPI	—	−2.979, 4	−4.688, 4	−4.462, 4	−2.667, 4	−4.150, 1	−4.760, 4
Trans	−8.680, 4	—	−4.656, 2	−4.458, 2	−2.668, 1	−4.182, 3	−4.691, 2
RE	−8.564, 4	−2.943, 2	—	−4.337, 2	−2.603, 1	−4.156, 1	−4.595, 5
REA	−8.596, 4	−3.164, 2	−4.778, 2	—	−2.642, 1	—	—
HF	−8.631, 4	−2.943, 1	−4.591, 1	−4.421, 1	—	−4.150, 1	−4.718, 1
HfS	−8.483, 1	−3.025, 3	−4.609, 1	—	−2.611, 1	—	—
Apart	−8.608, 4	−2.975, 2	−4.799, 5	—	−2.616, 1	—	—

Notes: The independent variables in the first column are regressed on the dependent variables in the respective rows, as well as lagged dependent variables. The lag structure is suggested by the AIC Info criteria. In the upper part are the *F*-stats of joint significance. The lower part shows the AIC value and the suggested lag for the respective pair. HPI: S&P's Case-Shiller Index for 20 MSAs, Trans: Transactions reported by the NAR including single family and condominiums, RE: Google I4S Real Estate category, REA: Google I4S Real Estate Agency subcategory, HF: Google I4S Home Financing subcategory, HfS: Google I4S Homes for Sale specific search query, Apart: Google I4S Apartments specific search query.

*p-value < 0.1

**p-value < 0.05

***p-value < 0.01

Appendix 5

Simple OLS Regressions according to Granger-Causalities

Independent Variables	Dependent Variable: House Prices						Dependent Variable: Transactions					
		Trans	RE	REA	HF	Apart		HPI	RE	REA	HF	Apart
Constant	0.000						−0.003					
	(−0.421)						(−0.472]					
Dep. Var.$_{t-1}$	0.573***						−0.412***					
	(4.993)						(−3.406)					
Dep. Var.$_{t-2}$	0.166						−0.179					
	(1.241)						(−1.377)					
Dep. Var.$_{t-3}$	−0.326***						0.229*					
	(−2.812)						(1.854)					
Indicator$_t$		−0.012*	−0.016	−0.005	**−0.016****	−0.027		−2.484	**0.914*****	0.350	**0.298*****	0.259
		(−1.863)	(−0.978)	(−0.302)	**(−2.831)**	(−1.494)		(−1.588)	**(3.647)**	(1.384)	**(3.077)**	(0.894)
Indicator$_{t-1}$		0.001	−0.025	−0.023	−0.005	0.002		−0.675	−0.191	**0.632****	−0.032	0.228
		(0.108)	(−1.502)	(−1.541)	(−0.876)	(0.133)		(−0.423)	(−0.626)	**(2.549)**	(−0.289)	(0.813)
Indicator$_{t-2}$		0.012*	0.016	0.014	0.009	**−0.030***		0.879	0.095	0.611**	−0.138	0.028
		(1.775)	(0.983)	(0.928)	(1.552)	**(−1.740)**		(0.563)	(0.332)	(2.384)	(−1.338)	(0.100)
Indicator$_{t-3}$		**0.017****	−0.016	−0.012	−0.006	0.003		2.678*	0.182	−0.283	−0.023	0.126
		(2.471)	(−0.979)	(−0.751)	(−1.079)	(0.166)		(1.760)	(0.650)	(−1.036)	(−0.224)	(0.461)
Indicator$_{t-4}$		0.017**	0.014	−0.014	0.012**	−0.014		3.995**	−0.316	−0.069	−0.066	0.225
		(2.255)	(0.833)	(−0.894)	(2.206)	(−0.798)		(2.562)	(−1.179)	(−0.260)	(−0.681)	(0.819)

Appendix 5 (continued)

Simple OLS Regressions according to Granger-Causalities

Independent Variables	Dependent Variable: House Prices						Dependent Variable: Transactions					
		Trans	RE	REA	HF	Apart		HPI	RE	REA	HF	Apart
Indicator$_{t-5}$		0.007 (0.788)	0.014 (0.826)	0.002 (0.120)	0.008 (1.387)	0.014 (0.780)		**5.326*** (3.338)**	0.000 (0.001)	−0.087 (−0.320)	−0.131 (−1.376)	−0.168 (−0.609)
AIC	−8.595	−8.642	—	—	−8.684	−8.613	−3.028	−3.153	−3.190	−3.097	−3.138	—
SC	−8.466	−8.479	—	—	−8.522	−8.451	−2.897	−2.988	−3.027	−2.933	−2.975	—
R^2 Adj.	0.380	0.418	—	—	0.440	0.399	0.209	0.313	0.337	0.272	0.301	—
S.E. of regression	0.003	0.003	—	—	0.003	0.003	0.052	0.048	0.047	0.050	0.049	—
Sum squared resid.	0.001	0.001	—	—	0.001	0.001	0.171	0.144	0.142	0.155	0.149	—
LM-Stat (Ser. Corr.)	1.420	0.908	—	—	1.407	1.758	0.939	0.683	0.515	1.599	1.505	—

Notes: The variables are regressed on the three lagged dependent variables, as well as the indicators suggested by the Granger causality tests for each lag separately. The information criterions and statistics in the lower part of the table relate to the regression where the lagged independent variable (indicator) had the highest *t*-stat and is marked in bold. HPI: S&P's Case-Shiller Index for 20 MSAs, Trans: Transactions reported by the NAR including single family and condominiums, RE: Google I4S Real Estate category, REA: Google I4S Real Estate Agency subcategory, HF: Google I4S Home Financing subcategory, Apart: Google I4S Apartments specific search query. The Homes for Sale search query as an indicator, as well as all other Google data a not reported but available upon request. *t*-stats are in parentheses.

* p-value < 0.1

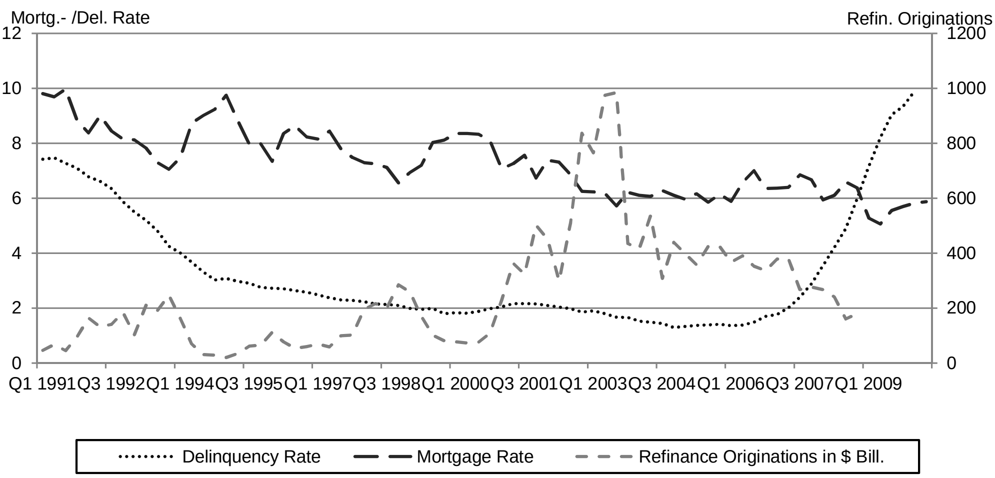
** p-value < 0.05

*** p-value < 0.01

Appendix 6

Illustration of Mortgage-market Variables with respect to ''Home Financing''

The graph shows the U.S. delinquency rate on real estate loans (DS Code USBDLRE.Q), the U.S. mortgages refinance originations (DS Code USMRREFOA), and the FHA effective mortgage rate (DS Code USMEGFH), as well as the correlation matrix of these variables.



Correlation Matrix: 1991:Q1–2008:Q4

	Delinquency Rate	Refinance Originations	Mortgage Rate
Delinquency Rate	1.00		
Refinance Originations	−0.42	1.00	
Mortgage Rate	0.57	−0.74	1.00

Appendix 7

Relative Explanatory Power of Google Data for House Prices and Transactions

	Sample: 2004M02 2009M04 (72 observations)							Sample: 2004M02 2009M04 (72 observations)						
Independent Variables	Dependent Variable: House Price							Dependent Variable: Transactions						
Dep. Var.$_{t-1}$	0.573***	0.529***	0.601***	0.703***	0.566***	0.609***	0.668***	−0.412***	−0.418***	−0.444***	−0.415***	−0.450***	−0.484***	−0.505***
	(4.993)	(4.334)	(5.009)	(5.502)	(4.514)	(4.344)	(4.734)	(−3.406)	(−3.321)	(−3.587)	(−2.982)	(−3.506)	(−3.447)	(−3.323)
Dep. Var.$_{t-2}$	0.166	0.151	0.174	−0.121	0.138	0.093	−0.084	−0.179	−0.146	−0.195	−0.230	−0.133	−0.244*	−0.224
	(1.241)	(1.112)	(1.280)	(−0.823)	(0.990)	(0.595)	(−0.523)	(−1.377)	(−1.053)	(−1.450)	(−1.591)	(−0.921)	(−1.667)	(−1.393)
Dep. Var.$_{t-3}$	−0.326***	−0.357***	−0.354***	−0.176***	−0.397***	−0.278**	−0.263*	0.229*	0.259*	0.226*	0.128	0.290**	0.184	0.150
	(−2.812)	(−2.812)	(−2.973)	(−1.519)	(−3.243)	(−2.057)	(−1.958)	(1.854)	(1.952)	(1.778)	(0.947)	(2.062)	(1.363)	(0.979)
House Price / Transactions$_{t-1}$			0.000		0.002	0.001	−0.004			−0.567		−0.568	−1.452	−1.511
			(−0.058)		(0.205)	(0.069)	(−0.362)			(−0.242)		(−0.234)	(−0.574)	(−0.565)
House Price / Transactions$_{t-2}$			0.007		0.003	0.009	0.003			0.347		0.786	1.889	2.317
			(0.791)		(0.312)	(0.898)	(0.319)			(0.154)		(0.337)	(0.793)	(0.926)
House Price / Transactions$_{t-3}$			−0.011		0.017**	−0.004	−0.009			2.526		3.174	3.013	2.772
			(−1.462)		(2.002)	(−0.377)	(−0.919)			(1.226)		(1.459)	(1.505)	(1.282)
Real Estate Agency$_{t-1}$				−0.015		−0.008	−0.008				0.489**		0.550**	0.589**
				(−1.020)		(−0.462)	(−0.447)				(1.974)		(2.222)	(2.253)
Real Estate Agency$_{t-2}$				0.019		0.014	0.026				0.490*		0.648**	0.724**
				(1.276)		(0.764)	(1.343)				(1.925)		(2.480)	(2.506)
Real Estate Agency$_{t-3}$				−0.017		−0.022	−0.024				−0.145		0.000	−0.005
				(−1.160)		(−1.195)	(−1.350)				(−0.557)		(−0.001)	(−0.018)
Apartments$_{t-1}$				0.005		−0.001	0.010				0.168		−0.108	−0.005
				(0.303)		(−0.060)	(0.466)				(0.611)		(−0.356)	(−0.014)
Apartments$_{t-2}$				−0.038**		−0.029	−0.033				−0.017		−0.254	−0.194
				(−2.079)		(−1.361)	(−1.583)				(−0.062)		(−0.811)	(−0.582)
Apartments$_{t-3}$				0.015		−0.006	0.010				−0.180		−0.255	−0.373
				(0.797)		(−0.324)	(0.517)				(−0.659)		(−0.877)	(−1.182)
Constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	−0.003	−0.003	−0.004	0.000	−0.004	0.001	0.000
	(−0.421)	(−0.236)	(−0.501)	(−0.811)	(−0.289)	(−0.606)	(−0.577)	(−0.472)	(−0.373)	(−0.579)	(0.033)	(−0.595)	(0.174)	(0.037)
House Price$_t$										−2.087		−1.323	−0.510	−0.618
										(−0.995)		(−0.598)	(−0.239)	(−0.275)

Appendix 7 (continued)

Relative Explanatory Power of Google Data for House Prices and Transactions

	Sample: 2004M02 2009M04 (72 observations)							Sample: 2004M02 2009M04 (72 observations)						
Independent Variables	Dependent Variable: House Price							Dependent Variable: Transactions						
Transactions,			−0.008		−0.005	−0.010	−0.002							
			(−0.995)		(−0.598)	(−1.140)	(−0.275)							
Home Financing,				−0.021***		−0.021***	−0.017**				0.254**		0.243**	0.228*
				(−3.379)		(−2.945)	(−2.144)				(2.479)		(2.095)	(1.789)
Employment,		−0.077			−0.144		−0.200		1.622			2.070		−0.386
		(−0.506)			(−0.921)		(−1.201)		(0.632)			(0.794)		(−0.145)
Income,		−0.023			−0.030		−0.023		−0.085			0.296		0.233
		(−0.540)			(−0.708)		(−0.558)		(−0.125)			(0.421)		(0.354)
Mortgage Rate,		0.011			0.008		0.004		−0.128			−0.166		−0.147
		(1.226)			(0.886)		(0.390)		(−0.896)			(−1.103)		(−0.872)
S&P500,		0.013			0.018*		0.016		−0.113			−0.167		0.070
		(1.331)			(1.763)		(1.486)		(−0.728)			(−0.973)		(0.398)
R^2	0.407	0.434	0.470	0.554	0.503	0.577	0.600	0.245	0.263	0.306	0.442	0.330	0.501	0.515
Adj. R^2	0.380	0.369	0.409	0.475	0.405	0.466	0.454	0.209	0.177	0.225	0.344	0.199	0.370	0.337
Akaike AIC	−8.595	−8.526	−8.587	−8.670	−8.532	−8.607	−8.545	−3.028	−2.935	−2.994	−3.125	−2.913	−3.120	−3.030
Schwarz SC	−8.466	−8.266	−8.326	−8.311	−8.140	−8.117	−7.925	−2.897	−2.674	−2.733	−2.766	−2.521	−2.630	−2.410
House Price / Transactions			x		x	x	x			x		x	x	x
Google-Data				x		x	x				x		x	x
Macro-Data		x			x		x		x			x		x

Notes: Disposable income, employment, S&P500, mortgage rate, Real Estate Agency, Apartments, and Home Financing are regressed on house prices and transactions separately and in combination. The second (ninth) column with three lagged dependent variables is for benchmarking purposes only. House Price: S&P's Case-Shiller Index for 20 MSAs, Transactions: Transactions reported by the NAR including single family and condominiums, Real Estate Agency: Google I4S Real Estate Agency subcategory, Home Financing: Google I4S Home Financing subcategory, Apartments: Google I4S Apartments specific search query, S&P500: S&P's 500, Income: disposable income, Employment: total employment, Mortgage Rate: FHA Mortgage Rate. *t*-stats are in parentheses.

* *p*-value < 0.1

** *p*-value < 0.05

*** *p*-value < 0.01

Appendix 8

Final VAR Model with Search Query and Macroeconomic Data

	House Price	Transactions	Real Estate Agency	Apartments
House Price$_{t-1}$	0.595***	−0.382	−0.735	1.090
	(4.642)	(−0.177)	(−0.717)	(1.164)
House Price$_{t-2}$	−0.059	2.048	−0.375	1.285
	(−0.423)	(0.873)	(−0.336)	(1.263)
House Price$_{t-4}$	−0.339***	3.061*	2.841***	−0.182
	(−2.874)	(1.784)	(3.013)	(−0.211)
Transactions$_{t-1}$	−0.006	−0.499***	−0.003	0.188***
	(−0.731)	(−3.534)	(−0.045)	(3.071)
Transactions$_{t-2}$	0.005	−0.323***	0.168***	0.159***
	(0.713)	(−2.623)	(2.869)	(2.985)
Transactions$_{t-4}$	0.015*	−0.050	0.026	0.056
	(1.748)	(−0.337)	(0.377)	(0.867)
Real Estate Agency$_{t-1}$	0.005	0.481*	0.041	0.078
	(0.334)	(1.835)	(0.333)	(0.684)
Real Estate Agency$_{t-2}$	0.024	0.672**	−0.262*	−0.264**
	(1.391)	(2.335)	(−1.918)	(−2.119)
Real Estate Agency$_{t-4}$	−0.005	0.132	−0.067	−0.085
	(−0.321)	(0.547)	(−0.586)	(−0.809)
Apartments$_{t-1}$	0.022	−0.050	−0.120	−0.345
	(1.064)	(−0.145)	(−0.731)	(−2.304)
Apartments$_{t-2}$	−0.037**	−0.042	0.031	0.040
	(−2.198)	(−0.145)	(0.226)	(0.319)
Apartments$_{t-4}$	−0.021	0.207	−0.139	0.088
	(−1.316)	(0.757)	(−1.072)	(0.747)
Constant	0.000	0.001	−0.004	0.002
	(−0.025)	(0.138)	(−1.189)	(0.838)
Home Financing$_t$	−0.014**	0.227**	0.036	0.118**
	(−2.141)	(2.005)	(0.671)	(2.406)
Employment$_t$	−0.146	−0.980	3.163**	2.086*
	(−0.990)	(−0.393)	(2.674)	(1.930)
Income$_t$	−0.017	0.077	0.177	−0.564**
	(−0.446)	(0.120)	(0.578)	(−2.024)
SP500$_t$	0.012	0.101	−0.004	−0.046
	(1.306)	(0.649)	(−0.055)	(−0.683)

Appendix 8 (continued)

Final VAR Model with Search Query and Macroeconomic Data

	House Price	Transactions	Real Estate Agency	Apartments
R^2	0.640	0.494	0.444	0.407
Adj. R^2	0.525	0.332	0.266	0.218
Sum sq. resids.	0.000	0.113	0.026	0.021
S.E. equation	0.003	0.048	0.023	0.021
F-Statistic	5.560	3.050	2.493	2.149
Log likelihood	308.021	118.759	168.702	174.793
Akaike AIC	−8.687	−3.038	−4.528	−4.710
Schwarz SC	−8.128	−2.478	−3.969	−4.151
Mean dependent	0.000	−0.002	−0.002	−0.001
S.D. dependent	0.004	0.058	0.026	0.023
Determinant resid covariance (dof adj.)			0.000	
Determinant resid covariance			0.000	
Log likelihood			775.737	
Akaike information criterion			−21.126	
Schwarz criterion			−18.889	

VAR Residual Serial Correlation LM Tests	Lags	LM Stat	Prob.
	1	11.348	0.788
	2	14.018	0.597
	3	11.840	0.755
	4	15.035	0.522
	5	12.902	0.680
	6	10.743	0.825

VAR Residual Heteroscedasticity Tests: No Cross Terms (only levels and squares)

	Chi. Sq.	df	Prob.
	343.587	320	0.175

Appendix 8 (continued)

Final VAR Model with Search Query and Macroeconomic Data

Notes: House Price: S&P's Case-Shiller Index for 20 MSAs, Transactions: Transactions reported by the NAR including single family and condominiums, Real Estate Agency: Google I4S Real Estate Agency subcategory, Home Financing: Google I4S Home Financing subcategory, Apartments: Google I4S Apartments specific search query, S&P500: S&P's 500, Income: disposable income, Employment: total employment. *t*-stats are in parentheses. Sample: 2004M02 2009M04 (72 observations).

* *p*-value < 0.1

** *p*-value < 0.05

*** *p*-value < 0.01

Endnotes

1 HBP is used as an abbreviation for the home buying process and HBSP for home buying search process in this article.

2 http://www.google.de/search?q=google+flue+trends.

3 http://www.bea.gov/national/pdf/NIPAhandbookch5.pdf.

4 D'Amuri and Marcucci (2009, p. 19).

5 Question and answering sites (e.g., Yahoo! Answers, WikiAnswers, Askville etc.).

6 Among other preconditions, a mean absolute prediction error (MAPE) smaller than 25% for a special time series.

7 http://www.census.gov/population/www/socdemo/migrate/cal-mig-exp.html.

8 http://www.census.gov/newsroom/releases/archives/mobility_of_the_population/cb10-67.html.

9 The information presented in this section is based on NAR (2009, 20–108).

10 Zumpano, Johnson, and Anderson (2003, 147) and NAR (2009, 52).

11 The relevance of real estate agents in the housing market is question of research for about two decades. See, among others, Zietz and Newsome (2001) for adverse selection due to lower commissions in lower-priced properties; see Benjamin, Jud, and Sirmans (2000) for an extensive review on real estate agent literature, subsumed under six different questions of research; see Gwin (2004) for the trade-off between supply of information on brokers' websites, dependent on the search costs of prospective buyers, and the risk of disintermediation; see Zietz and Newsome (2002) for the varying influences of agency representation on the sale price of a residential property for different property sizes; see Benjamin, Chinloy, Jud, and Winkler (2005) for the relationship between the financial performance of residential real estate brokerage firms and their Internet utilization.

[12] See Leung, Lau, and Leong (2002) or Shi, Young, and Hargreaves (2010).

[13] However, their analysis is limited to interactions between buyers and developers on the market for new housing only.

[14] Apart from different samples, a reason for the confirmation or rejection of different theories is often the frequency of the underlying data. For example, Clayton, MacKinnon, and Peng (2008) cannot support the search-model on a quarterly, but on a yearly basis. Shi, Young, and Hargreaves (2010) also confirm causality in the long run, only.

[15] Shiller (2007, p. 3ff.) See also Clayton (1998), who already claimed that the difference of observed and fundamental prices (i.e., justified by rents) serves as a good predictor for a potential correction in the near future.

[16] Appendix 1 offers a brief summary about the Google I4S application and its capabilities.

[17] Beyond that, key words suggested by the NAR (2009) report, as well as all search queries suggested by Google with respect to their relevance for the category and the top rising search queries are considered.

[18] For any of the search queries, the real estate category is suggested in the first place.

[19] The impact factor ''[...] indicates a total of the search term and presents the average of all points on the graph for that search term.'' http://www.google.com/support/insights/bin/answer.py?hl=en&answer=90657.

[20] The Google data are generated in the second week of February 2010. To adjust the frequency of the Google data to transactions and house prices, the weekly search query results are converted to a monthly level using SAS (accounting for a varying monthly intersection).

[21] Shimshoni, Efron, and Matias (2009) suggest a positive (negative) correlation of 0.80 (−0.94) between seasonal pattern (high level of outliers) and high (low) predictability while analyzing 10 root categories and their corresponding 1,000 most popular queries.

[22] Additionally, the model $y_t = \beta_0 + \beta_1 x_{t-i}$ is estimated with HAC standard errors but results are misleading. Lags of the independent variables are significant up to the tenth lag, which is contrary to the NAR survey (NAR, 2009). The approach of a smoothed effect of a three-month period (about the duration of the HBP) is considered in the model $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \beta_4 x_{t-i} + \beta_5 x_{t-i-1} + \beta_6 x_{t-i-2}$. A joint significance test (Wald test) of $\beta_4 = \beta_5 = \beta_6 = 0$ suggests quite the same relationships as the clear defined lag model. The results of these regressions are available on request.

[23] See Appendix 4 for the detailed results of the Granger Causality Tests based on AIC. The corresponding results based on the Schwarz criterion are available on request. In order to clarify the concept presented here, only the OLS regressions (Equation 2) for house prices and transactions are reported (see Appendix 5); for all Google data as dependent variables, the estimates are available on request.

[24] However, there is a feedback of prices on transactions within a five-month lag. Even though this seems to be contrary to the implications of Exhibit 2 (transactions lead prices), the visibility of causality does not necessarily contradict with a statistical interaction of prices and volume, with the causality from volume to price being stronger. This is simply the reverse of the finding by Zhou (1997), who relied on the minimal forecasting error as a decision rule for the number of lags.

[25] Also, *t*-statistics of the contemporaneous relationships are much higher, compared to the lagged effects in all (unreported) OLS-regressions including Real Estate as the dependent variable.

[26] Moreover, the second lag of Apartments on house prices is significant and negative. This has to be seen in the light of the sample period, when there was a huge demand for apartments in the period of massive house-price depreciation.

[27] Even if they were to invest the money in their homes, i.e., renovation, and increase its value, any house price index would not account for this simultaneously. Moreover, assuming a negative relationship between the mortgage rate and house prices the Home Financing coefficient is even incorrectly signed (notice that there is no clear relationship between mortgage rates and house prices in literature). The argument of lower mortgage rates inducing a higher demand for housing is based on dynamics, which can hardly be regarded as simultaneous in a monthly frequency. Moreover, the (unreported) OLS regressions imply a significant positive impact in Home Financing on prices with a four-month lag. Therefore, an increase of Home Financing in *t minus four,* accompanied by a mortgage rate decrease, results in a higher demand for housing, hence higher prices four months later.

[28] Delinquency rates rose from the beginning of 2006 onwards dramatically. See Appendix 6 for an illustration of delinquency rates, mortgage rates, and refinancing volume.

[29] Beside this heuristic reasoning, one should also notice three further aspects. First, Real Estate Agency has a huge impact on Real Estate and therefore they are somewhat arbitrary. Second, the correlations between the Google data are very high for Real Estate, which might cause multicollinearity in the VAR model and the impact of small changes are huge due to that (see Appendix 3). Third, the greatest impact of Real Estate on transactions is contemporaneous and considering the results so far, one cannot treat it as exogenous and include in the model.

[30] See, among others, Wheaton and Nechayev (2008), Clayton, Miller, and Peng (2008), and Arbel, Ben-Shahar, and Sulganik (2009). The latter two contributions also considered the trend of explanatory variables. However, because of the short time period and the aggregation level, none of the trend dummies are significant. All variables are equally transformed by taking the first log-differences.

[31] The detailed estimates for each specification can be found in Appendix 7.

[32] Upon that, different lag specifications without controlling for a contemporaneous relationship of transactions and house prices have been tested. Of course, this leads to a change in the information criterions, but the ranking of the corresponding combinations stays the same.

[33] Notice by comparing these two specifications: Regressions including Google data have seven additional variables—three lags of Real Estate Agency and Apartments, as well as the contemporary effect of Home Financing; specifications with transaction include only four additional variables—the contemporary effect and three lags.

[34] Kholodilin, Podstawski, Siliverstovs, and Bürgi (2009) among others confirm the outperformance of Google data to real world data in the abnormal economic environment.

[35] Basically, one might wish to include only lagged regressors in a VAR model. But the monthly frequency of the data justifies that the relationships are contemporaneous. Also, coefficients of matrix **A** in Equation 3 are robust with respect to this issue.

[36] Traditionally, impulse-response analysis and regressions for the boom and bust period separately are performed as well. According to Exhibit 2, the upturn is defined from January 2004 to July 2006 (30 observations) and the downturn from August 2006 to April 2009 (33 observations). Findings, for example Google data, perform better in

economic downturns (Askitas and Zimmermann, 2009a, 2009b) are of lower relevance and available on request. The same is true for Granger Causality tests in the final VAR model, which adds no additional information to the Causality test above. However, Baryla, Zumpano, and Elder (2000) state that the search duration of broker-assisted searchers increases in environments of low economic activity and high interest rates. Hence, these robustness issues rise in importance, once longer time series are available.

[37] All model specifications so far allow for a maximum lag structure of six months only.

[38] http://www.google.com/finance?q=GOOGLEINDEX_US:RLEST.

[39] http://marketshare.hitslink.com/search-engine-market-share.aspx?qprid=4.

[40] For example, if someone is searching for the query "homes for sale" in the U.S., Google evaluates the query distribution and suggests the following categories (top level): Real Estate (25%–50%), Local (25%–50%), Travel (0–10%), and Industries (0–10%). If the query is, for example, related to the Real Estate category, another specification can be made (second level): Real Estate Agencies (25%–50%), Home Financing (25%–50%), and Rental Listings and Referrals (10–25%).

[41] Google I4S analyzes a share out of the total Google search queries. Based on these results, the entered search queries are compared with the total volume of Google search queries for a certain time frame. For example, if someone is searching for homes for sale in a time period from January 2004 to January 2010 at the U.S. national aggregation level, the search query results for homes for sale are compared to the total volume of search queries in the U.S. for the same time period and depicted on a scale of 0 to 100. The results can be then specified by using a recommended category.

References

Arbel, Y., D. Ben-Shahar, and E. Sulganik. Mean Reversion and Momentum: Another Look at the Price-Volume Correlation in the Real Estate Market. *The Journal of Real Estate Finance and Economics*, 2009, 39:3, 316–35.

Askitas, N. and K.F. Zimmermann. Google Econometrics and Unemployment Forecasting. Institute for the Study of Labor (IZA), Discussion Papers, 2009a, June, No. 4201.

——. Prognosen aus dem Internet: Weitere Erholung am Arbeitsmarkt erwartet. Institute for the Study of Labor (IZA). 2009b, June, Standpunkte, No. 13.

——. Googlemetrie und Arbeitsmarkt in der Wirtschaftskrise. Institute for the Study of Labor (IZA), 2009c, August, No. 17.

Baryla, E.A. Jr., L.V. Zumpano, and H.W. Elder. An Investigation of Buyer Search in the Residential Real Estate Market under Different Market Conditions. *Journal of Real Estate Research*, 2000, 20:1, 75–91.

Benjamin, J.D., P. Chinloy, D. Jud, and D.T. Winkler. Technology and Real Estate Brokerage Firm Financial Performance. *Journal of Real Estate Research*, 2005, 27:4, 409–26.

Benjamin, J.D., D. Jud, and G.S. Sirmans. What Do We Know About Real Estate Brokerage? *Journal of Real Estate Research*, 2000, 20:1, 5–30.

Berkovec, J. and J. Goodman. Turnover as a Measure of Demand for Existing Homes. *Real Estate Economics*, 1996, 24:4, 421–40.

Choi, H. and H. Varian. Predicting the Present with Google Trends. Google Research. Google Inc., 2009a, Draft date April 2009.

——. Predicting Initial Claims for Unemployment Benefits. Google Research. Google Inc. 2009b, July).

Clayton, J. Further Evidence on Real Estate Market Efficiency. *Journal of Real Estate Research*, 1998, 15:1/2, 41–58.

Clayton, J., G. MacKinnon, and L. Peng. Time Variation of Liquidity in the Private Real Estate Market: An Empirical Investigation. *Journal of Real Estate Research*, 2008, 30:2, 125–60.

Clayton, J., N. Miller, and L. Peng. Price-Volume Correlation in the Housing Market: Causality and Co-movements. *Journal of Real Estate Finance and Economics*, 2008, 40: 1, 14–40.

Constant, A. and K.F. Zimmermann. Im Angesicht der Krise: US-Präsidentschaftswahlen in transnationaler Sicht. In: Wochenbericht, 2008, 75:44, 688–701.

D'Amuri, F. Predicting Unemployment in Short Samples with Internet Job Search Query Data. *Munich Personal RePEc Archive (MPRA)*, October 2009, No. 18403.

D'Amuri, F. and J. Marcucci. ''Google it!'' Forecasting the US Unemployment Rate with a Google Job Search index. Fondazione Eni Enrico Mattei, Working Papers, November 2009, No. 2010.31.

Della Penna, N. and H. Huang. Constructing Consumer Sentiment Index for U.S. Using Google Searches. University of Alberta, Working Papers, October 2009, No. 2009-26.

Fishbein, M. and I. Ajzen. *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Addison-Wesley, Reading, MA, 1975, 369.

Genesove, D. and C.J. Mayer. Loss Aversion and Seller Behavior: Evidence from the Housing Market. National Bureau of Economic Research (NBER), Working Papers, March 2001, No. 8143.

Ginsberg, J., M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski, and L. Brilliant. Detecting Influenza Epidemics Using Search Engine Query Data. *Nature*, February 2009, 457:7232, 1012–14.

Gwin, C.R. International Comparisons of Real Estate E-nformation on the Internet. *Journal of Real Estate Research*, 2004, 26:1, 1–24.

Kholodilin, K.A., M. Podstawski, B. Siliverstovs, and C. Bürgi. Google Searches as a Means of Improving the Nowcasts of Key Macroeconomic Variables. German Institute for Economic Research (DIW), Discussion Papers, November 2009, No. 946.

Leung, C.K.Y., G.C.K. Lau, and Y.C.F. Leong. Testing Alternative Theories of the Property Price-Trading Volume Correlation. *Journal of Real Estate Research*, 2002, 23:3, 253–64.

Lütkepohl, H. and M. Krätzig. *Applied Time Series Econometrics*. New York: Cambridge University Press, 2004.

National Association of Realtors (NAR). Profile of Home Buyers and Sellers 2009. November 2009.

Novy-Marx, R. Hot and Cold Markets. *Real Estate Economics*, 2009, 37:1, 1–22.

Schmidt, T. and S. Vosen. Forecasting Private Consumption: Survey-based Indicators vs. Google Trends. *Ruhr Economic Papers*, November 2009, No. 0155.

Shi, S., M. Young, and B. Hargreaves. House Price-Volume Dynamics: Evidence from 12 Cities in New Zealand. *Journal of Real Estate Research*, 2010, 32:1, 75–96.

Shiller, R.J. Understanding Recent Trends in House Prices and Home Ownership. National Bureau of Economic Research (NBER), Working Paper Series, October 2007, No. 13553.

Shimshoni, Y., N. Efron, and Y. Matias. On the Predictability of Search Trend. Google Inc.-Israel Labs. Draft Date August 2009.

Stein, J.C. Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects. *The Quarterly Journal of Economics*, 1995, 110:2, 379–406.

Stevenson, S. Modeling Housing Market Fundamentals: Empirical Evidence of Extreme Market Conditions. *Real Estate Economics*, 2008, 36:1, 1–29.

Stiglitz, J.E. Symposium on Bubbles. *The Journal of Economic Perspectives*, 1990, 4:2, 13–18.

Suhoy, T. Query Indices and a 2008 Downturn: Israeli Data. Research Department Bank of Israel. July 2009, Discussion Paper, No. 2009.06.

Tierney, H.L.R. and B. Pan. A Poisson Regression Examination of the Relationship between Website Traffic and Search Engine Queries. Munich Personal RePEc Archive (MPRA), 2009.

Wheaton, W.C. and G. Nechayev. The 1998–2005 Housing ''Bubble'' and the Current ''Correction'': What's Different This Time? *Journal of Real Estate Research*, 2008, 30:1, 1–26.

Zhou, Z. Forecasting Sales and Price for Existing Single-Family Homes: A VAR Model with Error Correction. *Journal of Real Estate Research*, 1997, 14:1/2, 155–67.

Zietz, J. and B. Newsome. A Note on Buyer's Agent Commission and Sales Price. *Journal of Real Estate Research*, 2001, 21:3, 245–54.

——. Agency Representation and the Sale Price of Houses. *Journal of Real Estate Research*, 2002, 24:2, 165–92.

Zumpano, L.V., K.H. Johnson, and R.I. Anderson. Internet Use and Real Estate Brokerage Market Intermediation. *Journal of Housing Economics*, 2003, 12, 134–50.

The authors are grateful to Prof. Richard Reed from the University of Deakin (Melbourne) and Prof. Richard Peiser from Harvard for their useful comments. Last but not least, we thank Dr. Brian Bloch for his comprehensive editing of the manuscript.

Ralf Hohenstatt, University of Regensburg, D-93040 Regensburg, Germany or ralf.hohenstatt@irebs.de.

Manuel Käsbauer, University of Regensburg, D-93040 Regensburg, Germany or manuel.kaesbauer@irebs.de.

Wolfgang Schäfers, University of Regensburg, D-93040 Regensburg, Germany or wolfgang.schaefers@irebs.de.