Timely Indices for Residential Construction Sector

Attilio Gardini and Enrico Foscolo[1]

Abstract The paper proposes a method for constructing Internet search activities–based indices of economic production in the Italian construction sector. The aim consists in representing current unobserved housing supply dynamics. Online data management tools and filters are described and an econometric analysis is carried out in order to establish statistical significant relationships between Internet indicators and actual data by official statistical agencies (the target series). The proposed methodology can therefore fill the temporal gap in national statistical institutes released publications and provides real time indices of trend–cycle dynamics in the Italian construction sector.

Key words: Index numbers, Residential Construction, Internet, Cointegration

1 Introduction

The empirical evidence on residential construction production in Italy is far less convincing. We have little direct and lagged evidence and no wide standard data sets to observe construction activity dynamics. Nevertheless, since real estate represents about 15% of Italian Gross Domestic Product, updated information are important to assess and predict overall economic activity.

Internet indices provide an answer to this thirst for knowledge. However, the amount of data may be misleading. One requires management tools and filters to capture the signal and separate it from noise.

A simple and easy solution to represent trend–cycle comes from the most popular and used search engine, Google. We already recognize some contributors on the potential of Google data as exogenous variable. The relevance of a Google job–search

[1]Enrico Foscolo

Free University of Bozen–Bolzano, School of Economics and Management

University of Bologna, Department of Statistical Sciences, Rimini Campus Branch

email: enrico.foscolo2@unibo.it

index as an indicator for unemployment dynamics in the United States was tested in [1,6] and [10,11] introduced monthly consumer indicators based on Google search activity data, providing significant benefits to forecasts comparing with common survey–based counterparts. Health–seeking behavior in the form of on–line Google search queries was monitored by [7].

The possibility of adding the search indices to a simple autoregressive model in order to improve the forecasting of new home sales was first explored in [5]. In this paper we provide new indices for residential construction production based on weekly Internet search activities. We test the relations with reference series and the exogeneity of search data is not assumed a priori. We propose a dynamic specification for the Italian residential construction production based on a cointegrated Vector Error Correction Model (hereafter, VECM; see [9]), where the cointegration rank is subject to inference. The main result of this analysis is to establish a statistical significant relationship between Google and real data. Thus, the Google leading indicators provide a double benefit; i.e., improving future predictions and allowing better understanding for current unobserved dynamics in the construction sector.

2 Managing Query Data on Construction Sector

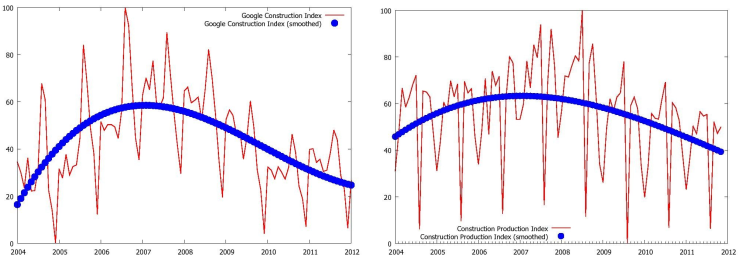
Google Insights for Search (http://www.google.com/insights/search/?hl=en-US) is the system provided by Google in order to analyze portions of worldwide Google web searches from all Google domains starting from January 2004. This mechanism computes how many searches have been done for the entered keywords, relative to the total number of searches done on Google over time. Insights for Search provides normalized not seasonally adjusted weekly indexed series. In order to avoid uncertainty about the criteria by which the Insights for Search determines the context of the terms, we specify some categories provided by an automated engine (see http://support.google.com/insights/?hl=en). When filters are applied, Google system only evaluates queries that are related to that category. For our aims, the system provides several options, such as Real Estate Agencies or Real Estate Listings.

In order to exploit as much information from Google time series as possible, we extract common unobserved variables from search engine data. To identify residential construction production factors we first select 5 housing relevant categories (cf. Table 1) and we employ asymptotically distribution–free estimation methods (e.g., [2,3]). Keywords such as home–like words and sale are included in our queries. Finally, we select the number of factors by means of the parallel analysis (e.g., [8]). For our case, parallel analysis suggests that 2 factors for residential construction production might be most appropriate. Nevertheless, we only choose the latent variable with higher factor loadings in correspondence with Real Estate Agencies and Real Estate Listings subcategories, in our opinion the best matches with the construction sector.

Table 1: The selected Google categories for residential construction production.

	Google Categories	***Google Subcategories***
Residential Construction	Real Estate	Real Estate Agencies Real Estate Listings Timeshares & Vacation Properties Property management

Figure 1: The Google Construction index (left panel) and the Construction Production index (right panel).



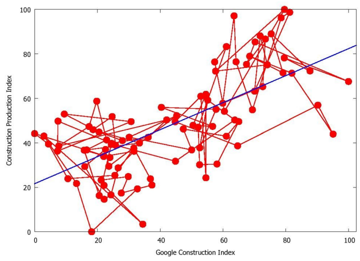
3 From Query Data to Econometric Framework

Checking the connection between Google factors and official housing related data is the necessary step in order to consider these indices as proxies of housing market dynamics. In order to do that, we exploit the vector autoregressive specification (hereafter, VAR) and we test the presence of stochastic common trends. When cointegration is detected, variables will show a tendency to co–move over time. Such cointegrated relations can often be interpreted as long–run relations and are therefore of considerable economic interest for our purpose. In this connection we use the VECM which gives a convenient reformulation of VARs.

The proposed application refers to monthly Google Italian activity over the period 2004:04 – 2011:11 (T = 92 observations), and monthly construction production index (not seasonally adjusted; shortly, CPI) provided by Eurostat, the Statistical Office of the European Union (see Fig. 1). We test the unrestricted VAR with k = 3 lags by maximum likelihood. We also include seasonal dummies and unrestricted drift term. The adequacy of the model is checked by residual analysis where it is seen that there is no autocorrelation, no heteroskedasticity, and no seriously large normalized residuals.

Testing for the cointegration rank r (see [9]) we find that $r = 0$ should be rejected at 5% level for all tests, but $r = 1$ is not rejected for a significant level of 5%. Thus, the analysis indicates that our variables are nonstationary but cointegrate. Indeed, Figure 2 suggests that the two indices move together around the identify line for the whole period. To conclude, we check parameter constancy throughout the sample period. The Chow forecast (shortly, CF) test against the alternative that all coefficients including the residual covariance matrix may vary is performed. In order to take into account small sample deviations we calculate bootstrap p–values obtained by means of 1000 replications as in [4]. Anyway the null hypothesis of constant parameters is not rejected. Thus, the analysis confirms the use of the Google index as proxy of CPI trend–cycle.

Figure 2: The scatter plot between Google index and CPI.



References

1. Askitas, N., Zimmermann, K.F.: Google Econometrics and unemployment forecasting. Appl. Econ. Q. **55**, 107 – 120 (2009)
2. Browne, M.W.: Generalized least squares estimators in the analysis of covariance structures. S. Afr. Stat. J. **8**, 1 – 24 (1974)
3. Browne, M.W.: Asymptotically distribution–free methods for the analysis of covariance structures. Brit. J. Math. Stat. Psy. **37**, 62 – 83 (1984)
4. Candelon, B., Lütkepohl, H.: On the reliability of Chow type tests for parameter constancy in multivariate dynamics models. Econ. Lett. **73**, 155 – 160 (2001)
5. Choi, H., Varian, H.: Predicting the present with Google Trends. Google Technical Report (2009)
6. D'Amauri, F., Marcucci, J.: "Google it!" Forecasting the US unemployment rate with a Google job search index. MPRA Paper 18248, University Library of Munich, Germany (2009)
7. Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brillant, L.: Detecting influenza epidemics using search engine query data. Nat. **457**, 1012 – 1014 (2009)
8. Horn, J.L.: A rational test for the number of factors in factor analysis. Psychom. **30**, 179 – 185 (1965)
9. Johansen, S.: Likelihood–based inference in cointegrated vector autoregressive models, 2. edn. Advanced Text in Econometrics, Oxford University Press, Oxford (1996)
10. Vosen, S., Schmidt, T.: Forecasting private consumption: survey–based indicators vs. Google trends. J. Forecast. **30**, 565 – 578 (2011)
11. Vosen, S., Schmidt, T.: A monthly consumption indicator for Germany based on Internet search query data. Appl. Econ. Lett. **19**, 683 – 687 (2012)