Google searches for portfolio management: a risk and return analysis

Mario Maggi[1] and Pierpaolo Uberti[2]

[1] Department of Economics and Management, University of Pavia, Pavia 27100, Italy, mario.maggi@unipv.it, ORCID: 0000-0002-7233-4443
WWW home page: https://sites.google.com/a/unipv.it/mariomaggi/
[2] Department ofEconomics (DIEC), University of Genoa, Genoa 16126, Italy, uberti@economia.unige.it, , ORCID: 0000-0001-8426-0368

Abstract. Google search data has proven to be useful in portfolio management. The basic idea is that high search volumes are related to bad news and risk increase. This paper shows additional evidence about the use of Google search volumes in risk management, for the Standard & Poor Industrial index component,s from 2004 to 2017. To overcome the (time-series and cross-section) limitations Google imposes on the data download, a re-normalization procedure is presented, to obtain a multivariate sample of volumes which preserve their relative magnitude. The results indicate that the volumes' normalization and the starting portfolio are decisive for the portfolio performances. Correctly normalized Google search volumes yield poor results. This may lead to revise the interpretation of the search volume: it can be considered a risk indicator, but when used in a *equally risk contribution* portfolio, no evidence of the improvement of the risk-return performances is found.

Keywords: online searches; Google Trends; portfolio management

1 Introduction

In the recent years the increasing availability of web data fueled a wave of studies which analyzed the relations between web searches and many aspects of the social sciences. Since [2, 8] who first used the Google search volumes to forecast the influenza diffusion, the information contents of the Google queries has been analyzed also for other phenomena. In particular, some studies focused on the relations between web searches and financial data (for example, see [3, 5, 6]). In fact, as [1] showed, there is evidence of the information flow from media to the financial market. The information content of the Google search volumes has been documented, among others, by [3, 5, 6]. Recently, different works explored the possibility of exploiting the forecasting power of web data to set up asset allocation and trading strategies. For instance, in [4, 9] the Google data are used with the aim of improving the return or the risk-return combination of a financial portfolio.

This paper deepens the analysis by [4] of the contribution of Google data to the asset allocation performances. We focus on the need to obtain a multivariate series of search volumes whose sizes are proportional to the (undisclosed) real volumes. The application of re-normalized series yields different results with respect to [4]. The use of the Google search volumes as risk indicator can reduce the standard deviation of the portfolio return, rather than improve the risk-return performances.

2 Google Trends

Google collects data about every query users type on its web search engine and decided to disclose a part of these data through its service Google Trends[3]. Data about the Google search volumes are available from January 2004. The region and time window may be customized and multiple series can be downloaded in `csv` format. The data Google allows to download are not the raw volumes, but a normalized index (the Google Index, GI in the following) taking integer values between 0 and 100. The maximum volume attained on the selected window is set to 100; all the other data are normalized accordingly and rounded to the nearest integer. This way the dynamic properties are retained, but the absolute size of the volume is lost.

We remark that this rounding may lead to a strong information loss when the series' volumes have a large difference in the overall size: the resolution of the smallest series is reduced. As an example, we downloaded the series of the queries "Italy" only and the couple "Italy" and "United States" (monthly, from 2004 to present). "Italy" has a GI ranging from 32 to 100, instead when downloaded together with "United States", "Italy" ranges from 3 to 9 (only 7 different values), with a clear loss of information about the dynamics of the smallest series.

Google imposes other limitations on the disclosed data, for instance: (i) the possibility to download up to 5 multiple series; (ii) the longest the time window, the lowest the frequency of the data (monthly over 5 years, weekly from 3 months to 5 years, daily from 7 to 270 days, and so on).

In order to obtain a multivariate sample of weekly data from 2004 to 2017, we follow a re-normalization procedure. First we downloaded all the series as unique queries for three periods, each period not longer than 5 years (obtaining weekly data). The three series overlap at the extreme dates. Then, we concatenate the series matching the values for the overlapping weeks. We do not round the results. Once the univariate series have been composed, to obtain a multivariate sample which preserves the relative size of the search volumes, we downloaded multiple series. We include in each search up to 5 series with comparable size, to limit the information loss due to the rounding. Each query must belong to (at least) 2 different multiple series, to allow a cross normalization, similar to the one operated along time. Again, we do not round the results.

[3] See `https://trends.google.com/trends/`

3 Asset allocation based on Google search volumes

Following [4], we use the GIs data to find the weights of a (long only) portfolio composed by the Standard and Poor Industrial index (SPI) stocks. The basic idea is that the search volume is related to bad news, so it is a risk indicator: when the interest on a given stock increases, many people look for information to trade; there is evidence of the relation between web searches and trading volumes [3, 5]; an increase in trading, produces a possible increase in the price volatility.

For this reason, it is possible to extend to search volumes the *equal risk contribution* (ERC) rule, proposed by [7] to manage the risk. Let $V_{i,t}$ be the GI and $w_{i,t}$ be the portfolio weight for stock i, at time t. The ERC rule, yields the portfolio weights

$$w_{i,t} = \frac{V_{i,t}^{-\alpha}}{\sum_{j=1}^{N} V_{j,t}^{-\alpha}}, \tag{1}$$

where α controls for the relevance given to $V_{i,t}$: for $\alpha = 0$, the portfolio is uniform, $w_i = \frac{1}{N},\ i = 1, \ldots, N$; for $\alpha > 0$, w_i decreases with $V_{i,t}$, underweighting stocks with large GIs; for $\alpha < 0$, w_i increases with $V_{i,t}$, overweighting stocks with large GIs.

Starting from an initial portfolio $w_{i,0}$, a recursive version of (1) can be obtained applying the updating rule

$$w_{i,t} = \frac{w_{i,t-1} e^{-\alpha g_{i,t}}}{\sum_{j=1}^{N} w_{j,t-1} e^{-\alpha g_{j,t}}}, \quad t \geq 1, \tag{2}$$

where $g_{i,t} = \ln V_{i,t} - \ln V_{i,t-1}$ is the log-rate of variation of the Google search volumes (remark that $g_{i,t}$ is scale free). This way we can explicitly control for the starting portfolio: (1) and (2) are equivalent for the same $w_{i,0}$.

We set up portfolios applying (1) and (2) on the stocks listed on the SPI, from July 2004 to July 2017 (GI query = "Company Name"). The value of α ranges from -2 to 2 to vary the strength and the sign of the GIs contribition.

Figure 1 shows the averages, standard deviations and Sharpe ratios for portfolio returns based on the GIs. Remark that without the re-normalization (solid lines) the average return increases, the standard deviation decreases and the Sharpe ratio increases with α; that is the stonger GIs are used as risk indicator, the better the portfolio performances. This result is in line with [4], who downloaded the GIs 5 by 5, without (we suppose) an overall cross normalization. Moreover, the normalized GIs and and the case of uniform starting portfolio produce opposite results, suggesting that a negative α could work better.

Remark that the normalized GIs produce a very unbalanced portfolio (the weigts' Gini coefficient is 0.98). Moreover, we conjecture that the results obtained in the non normalized case may depend on chance. Therefore, we run a Monte Carlo experiment, using the GIs to update a random initial portfolio with the rule (2). We find that in $\alpha = 0$: the average return is decreasing 80.2% of times, the standard deviation is decreasing 75.4% of times, the Sharpe ratio is

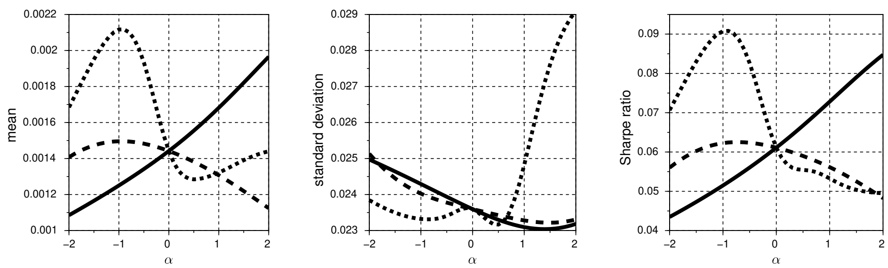

Fig. 1. Returns's average, standard deviation and Sharpe ratio of the portfolios composed through (1) and (2), for $\alpha \in [-2, 2]$: univariate GIs (solid); normalized GIs (short dashed); rule (2) starting from the uniform portfolio (long dashed).

decreasing 77.2% of times. We conclude that if the GI is used as a risk measure, there is no evidence supporting the improvement of risk-return performances, unless GIs are used with the opposite meaning, i.e. $\alpha < 0$, so the risk is decreasing with GI.

References

1. Alanyali, M., Moat, H. S., Preis, T.: Quantifying the Relationship Between Financial News and the Stock Market. Sci. Rep. **3**, 3578 (2013). doi:10.1038/srep03578
2. Ginsberg, J., Mohebbi, M. H., Patel, R. S, Brammer, L., Smolinski, M. S., Brilliant, L.: Detecting influenza epidemics using search engine query data. Nature **457**, 1012–1014. (2009). doi:10.1038/nature07634
3. Heiberger, R. H.: Collective Attention and Stock Prices: Evidence from Google Trends Data on Standard and Poor's 100. PLoS One **10(8)**, (2015). doi:10.1371/journal.pone.0135311
4. Kristoufek, L.: Can Google Trends search queries contribute to risk diversification?. Sci. Rep. **3**, 2713 (2013). doi:10.1038/srep02713
5. Kristoufek, L.: Power-law correlations in finance-related Google searches, and their cross-correlations with volatility and traded volume: Evidence from the Dow Jones Industrial components. Phisica A **428)**, 194–205 (2015), doi:10.1016/j.physa.2015.02.057
6. Li, X., Ma, J., Wang, S., Zhang, X.: How does Google search affect trader positions and crude oil prices? Econ. Model. **49**, 162–171 (2015). doi:10.1016/j.econmod.2015.04.005
7. Maillard, S., Roncalli, T., Teïletche, J.: The properties of equally weighted risk contribution portfolios. J. Port. Manag. **36(1)**, 60–70 (2010)
8. Polgreen, P. M, Chen, Y., Pennock, D. M.: Using internet searches for influenza surveillance. Clin. Infect. Dis. **47**, 1443–1448 (2008). doi:10.1086/593098
9. Preis, T., Moat, H. S., Stanley, E.: Quantifying Trading Behavior in Financial Markets Using *Google Trends*. Sci. Rep. **3**, 1684 (2014). doi:10.1038/srep01684