



universität  
wien

# Search Interest for Gold Prices and Realized Gold Prices

Macroeconometrics

**Supervisor:**

Univ.-Prof. Dipl.-Ing. Dr. Robert Kunst

**Authors:**

Hashem Zehi, Samuel (Student ID 12012285)

Hochholzer, Matthias (Student ID 11724853)

Master of Science Economics (M.Sc.)

Vienna, June 26, 2021

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data</b>	<b>2</b>
<b>3</b>	<b>Gold Price Modeling</b>	<b>3</b>
3.1	Best-Prediction Modeling . . . . .	4
3.2	Volatility Analysis . . . . .	4
<b>4</b>	<b>Joint Modeling for Gold and Search Interest Index</b>	<b>5</b>
4.1	Exploratory Analysis of the Gold Search Interest Index . . . . .	5
4.2	VAR Model for Gold Price and Search Interest Index . . . . .	5
<b>5</b>	<b>Conclusion</b>	<b>6</b>
	<b>References</b>	<b>7</b>

# 1 Introduction

This project investigates the relationship between the realized price of gold and a standardized index measuring the relative online search interest for gold prices. More specifically, can search interest help improve prediction of one-period ahead gold prices. Due to the technological advances in the financial industry nowadays more and more algorithms work based on internet sentiment in order to buy and sell financial products. One possible measure for such automated algorithms is using data provided by companies like Google or Microsoft. Large spikes in search interest may trigger such algorithms. As financial news outlets like Bloomberg or the Financial Times write articles on price changes more and more people will look up prices of goods and commodities, again triggering the algorithms, resulting in a feedback loop.

This paper is structured as follows. First we explain the data collection process then do some exploratory data analysis. Then we attempt to find the best prediction model for gold prices. Afterwards we model the gold price and the relative search interest jointly. Lastly we give some closing remarks on our results and issues with modeling and data.

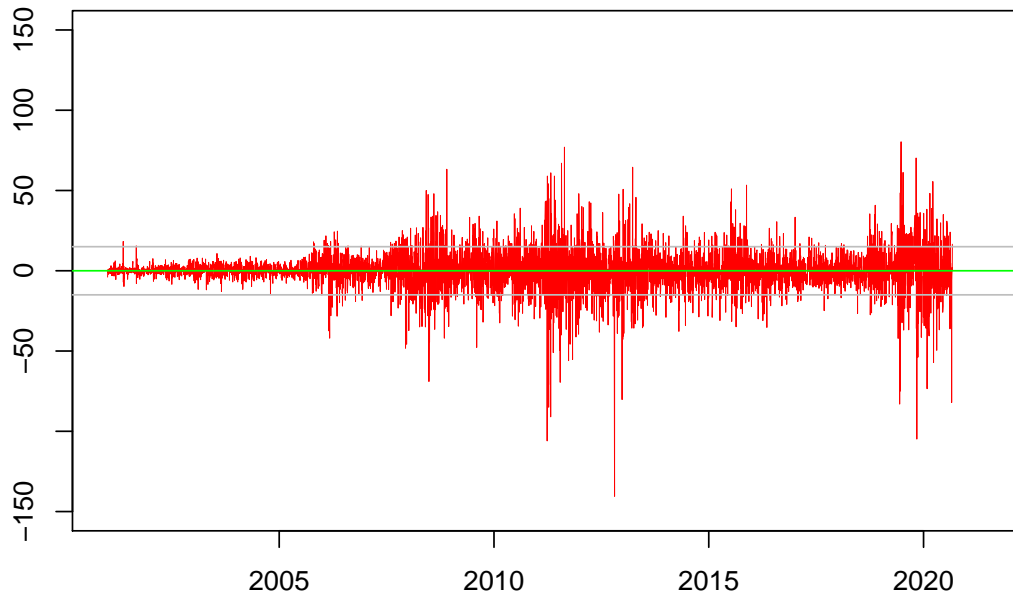
## 2 Data

The data used in this project comes from two sources. Firstly the gold price search interest is provided to the public by Google Trends (2021). This data is standardize to a range from zero to 100 over the entire time range. 100 represents the point in time with the highest search interest and zero either no interest at all or simply not enough data for this specific point in time. 50 then simply means that the popularity is at 50% of the peak value. Due to the ability to collect a wide range of data on the users of the website Google removes datapoints where the same user looked the term up multiple times within a short period of time. Furthermore, regional differences are accounted for due to the data only being relative search interest, thus regions with larger populations and particular search interests cannot skew the data. It is noteworthy that if the data is split up by regions India, one of the largest markets for gold, appears to have the highest relative search interest, with only countries like Oman, Pakistan, and Myanmar being close. For this analysis we focus on the worldwide interest in the search term “gold price” over the period 2004 until June 2021.

Secondly, the data on gold prices is provided by ICE Benchmark Administration Limited (2021) and published for use by the Economic Research Division of the Federal Reserve Bank St. Louis and collected by the London Bullion Market Association. The price is set twice daily, in the morning and the afternoon, and measured in US dollars per troy ounce. For our analysis we use the afternoon price given two separate frequencies. When not modeling the data jointly we use a daily frequency as it allows us to make a more detailed analysis and gives a more accurate reflection of the volatility. It should be noted that there around 200 of the 5,300 observations contain “NA” values. As there appears to be not specific pattern such as for example no prices being reported on weekends, we have elected to replace these values with the price from the previous day.

In the joint modeling we are constrained by the frequency with which Google makes search interest data available: monthly. Thus we use a monthly frequency for the realized prices as well, to be more precise we use the prices at the end of each month.

Figure 3.1: Daily Gold Prices (First-Differences)

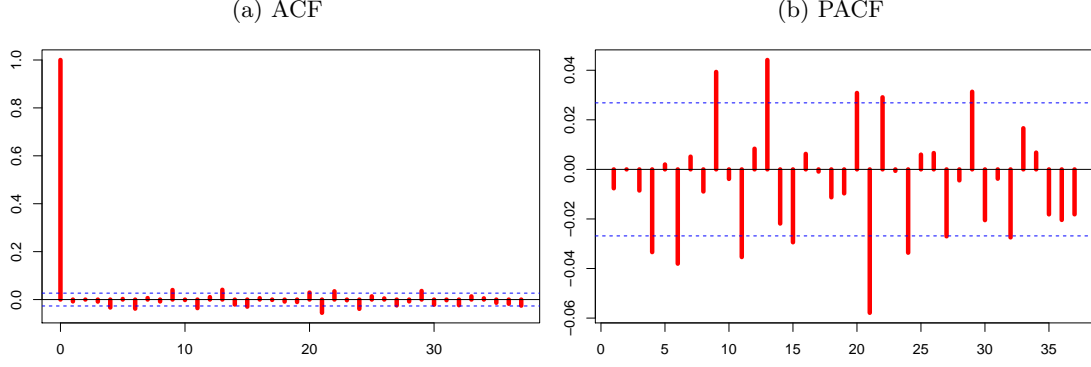


### 3 Gold Price Modeling

In this section look at the daily price of gold.

Now we consider the first-differenced gold prices. At first glance they appear to be stationary around zero with phases of higher and lower volatility. The overall level of volatility appears to have increased starting from 2006 onward, as indicated by the gray lines. Before 2006 there were very few observations outside of the  $\pm 15$  band, compared to this occurring very frequently afterwards. In order to illustrate the high frequency of data we chose very thin lines for Figure 3.X. It appears that there is some volatility clustering in the daily prices of gold. Four clusters appear very prominent, namely one at around 2007-2008, 2011, 2013, and 2019-2020. Every now and then there are very large outliers, one notable in 2013 with a day-on-day decrease by over USD 100.

Figure 3.2: Empirical (Partial) Autocorrelation Functions of the Gold Price FDs



### 3.1 Best-Prediction Modeling

In this section we use the AIC in order to determine the optimal lag order. As our data are of somewhat high frequency over a longer period of time, any concerns over the performance of the information criterion in small samples can be disregarded. When modeling prices we are often concerned with trying to predict prices, which justifies the choice of the AIC as it selects the best-forecasting model in large samples rather than leading to a consistent choice of lag orders. We estimate all  $ARIMA(p, q)$  models within a range of  $p$  and  $q$  values and then choose the one with the lowest AIC value, where the criterion is defined as follows:

$$AIC = \log \hat{\sigma}^2 + 2 \frac{p + q + 1}{2}.$$

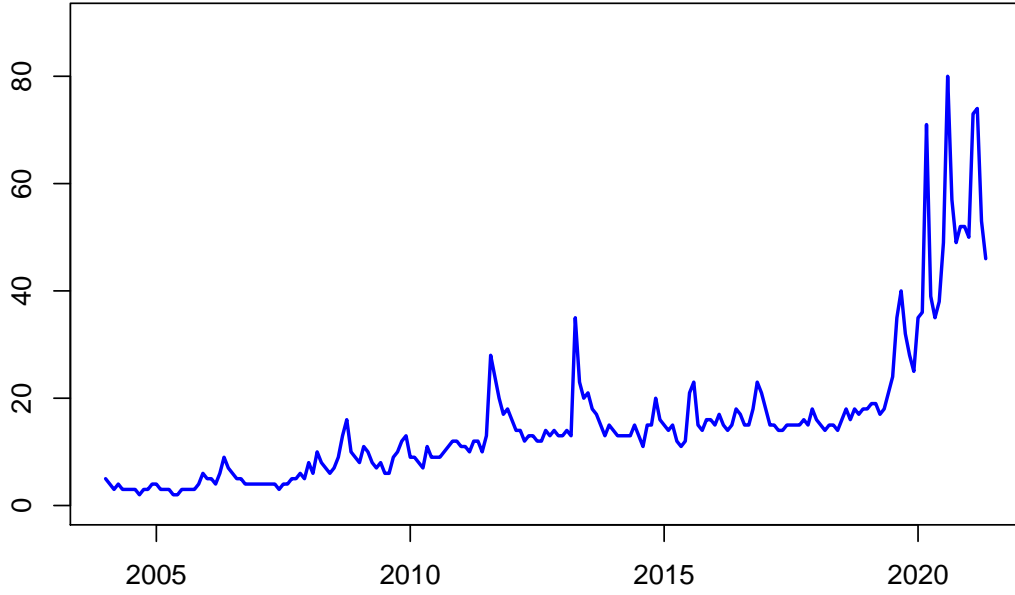
### 3.2 Volatility Analysis

As mentioned in the exploratory analysis of the data it does not appear as if the autocovariance is constant over time. Thus it is a natural consequence to further look at ways of modelling the volatility. We do so by considering (G)ARCH models.

- ARCH:  $E(\epsilon_t^2 | \mathcal{I}_{t-1}) = \sigma_t^2 = 119.54 + 0.11\epsilon_{t-1}^2$
- GARCH:  $E(\epsilon_t^2 | \mathcal{I}_{t-1}) = \sigma_t^2 = 0.14 + 0.07\epsilon_{t-1}^2 + 0.93\sigma_{t-1}^2$ , according to t-series (slight difference with the "rugarch" package)
- iGARCH:  $E(\epsilon_t^2 | \mathcal{I}_{t-1}) = \sigma_t^2 = 0.16 + 0.07\epsilon_{t-1}^2 + 0.93\sigma_{t-1}^2$
- eGARCH:  $\log \sigma_t^2 = 0.02 + 0.04 \frac{|\epsilon_{t-1}|}{\sigma_{t-1}} + 0.10 \log \sigma_{t-1}^2 + 0.13 \frac{\epsilon_{t-1}}{\sigma_{t-1}}$

According to the customary tool AIC, GARCH(1,1), eGARCH(1,1) and iGARCH(1,1) outperform ARCH(1). No further conclusions can be drawn. We would opt for the simplest, the GARCH(1,1) model. The eGARCH model, however, could give quite reasonable results for prizes.

Figure 4.1: Relative Search Interest in the Term “gold price” from 2004-2021



## 4 Joint Modeling for Gold and Search Interest Index

### 4.1 Exploratory Analysis of the Gold Search Interest Index

We start by looking a bit closer at the search interest index. This is plotted over the entire observational period available in Figure 4.1. It is easy to see that there was a slight upwards trend in the data over the period from 2004 to 2019. In this period there are two spikes in the data, one around late 2012 and one around late 2013 - early 2014. According to The Indian Express (2016) the first spike appears around the time right before the wedding and festival season in India coinciding with a weak local currency. A similar line of reasoning appears to be the reason for the spike in late 2013 - early 2014. Around late 2019 there appears to be a very strong search interest increase, joint with an increase in volatility. This however is not particularly surprising as gold is often hailed as a good, safe investment in times of increased uncertainty. It is likely that political unrest in India, followed by the start of the Corona crisis are the main drivers for this development. Skepticism about the effects of loosened monetary policy on the worth of local currencies in large parts of the world may lead to a flight from local currencies to gold due to its historically stable value.

### 4.2 VAR Model for Gold Price and Search Interest Index

## 5 Conclusion

## References

## References

- Google Trends (2021). *Google Trends: 'gold price', 2004-2021*. URL: <https://trends.google.com/trends/explore?date=all&geo=US&q=gold%20price>.
- ICE Benchmark Administration Limited (2021). *Gold Fixing Price 3:00 P.M. (London time) in London Bullion Market, based in U.S. Dollars*. Ed. by Federal Reserve Bank of St. Louis. URL: <https://fred.stlouisfed.org/series/GOLDPMGBD228NLBM\#0>.
- The Indian Express (24/06/2016). "Past instances when gold prices skyrocketed in India". In: *The Indian Express*. URL: <https://indianexpress.com/article/business/commodities/past-instances-when-gold-prices-skyrocketed-in-india-2872995/>.