Universitetet i Stavanger

UIS BUSINESS SCHOOL

MASTER'S THESIS

STUDY PROGRAM:	THESIS IS WRITTEN IN THE FOLLOWING SPECIALIZATION/SUBJECT:
Master of Science in Business Administration	Economics and Applied Finance
	IS THE ASSIGNMENT CONFIDENTIAL? (**NB**! Use the red form for confidential theses)

TITLE:

Can Google Trends predict gold returns and its implied volatility?

AUTHOR(S)		SUPERVISOR:
Candidate number:	Name:	Peter Molnár
4030 …………………	Jenny Ho …………………………………….	
4056 …………………	Lars Henrik Kristiansen …………………………………….	

Acknowledgements

We would like to thank our thesis supervisor, Peter Molnár. His guidance and knowledge on Google Trends and the gold market has been of tremendous help. We cannot emphasize enough how much we appreciate his availability, encouragement and enthusiasm. Additionally, we would like to thank our friends and family for their love and support.

Abstract

We investigate the impact of investor attention and economic uncertainty on gold price changes and their volatility. We use Google searches for "gold" as a measure of investor attention, considering searches originated in US, India, and globally. We find that Google searches originated in India are more relevant for the gold market than global searches or searches originated in the US. Increased Google searches are associated with gold price declines and increased volatility. Moreover, this relationship is not just a correlation. It also holds in predictive manner in both directions: 1) Increased Google searches are followed by gold price declines and high volatility and 2) gold price declines and high volatility are followed by increased Google searches. We consider three measures of economic uncertainty, the VIX index, the Economic Policy Uncertainty index and the TED spread, and find that the VIX index has the strongest impact on gold. High levels of the VIX index are both contemporaneously related and also predict gold price increases and high volatility.

Table of contents

List of tables and figures

1. Introduction

Google's search engine is widely used all over the world, covering approximately 90% of the global search engine market (Harford, 2017). The search engine provides search volume indexes (henceforth: SVI) through their platform Google Trends (henceforth: GT). GT provides useful information, as keyword-searches can be utilized in decision making for various subjects, including producers, consumers, politicians and investors. The usage of GT data has grown tremendously since its launch in 2006, and has been proven to have a vast variety of applications.

Throughout history, gold has been used as a store of value, a currency and a medium of exchange. In countries facing relatively high inflation, such as India, gold is considered as an important saving and hedge tool. Moreover, gold symbolizes social status and prosperity, and is often passed down from generation to generation. Almost one-third of the global demand for gold derives from India (Rani & Vijayalakshmi, 2014), which emphasizes the importance of gold in the Indian society.

Although the significance of gold and the vast usage of search engine data is well established, the literature on the dynamics of internet search trends and gold is limited. This thesis aims to contribute to the studies regarding gold and GT, by using SVI data from three different geographic regions and adding three market uncertainty indexes. We investigate the impact of Google's SVI from India, the US and globally on gold returns and its volatility.

We find that the relationship between Google searches and the gold market is strongest for Google searches originated in India. Further, increased Google searches for "gold" are associated with negative current and future gold returns, and high current and future implied volatility of gold. As we include the VIX index, EPU index and the TED spread as control variables, we have a few additional findings. The VIX index is positively related with current and future gold returns, as well as current and future implied volatility of gold. Moreover, positive gold returns predicts an increase in the implied volatility of gold.

The remainder of this thesis is structured as follows. Previous literature is reviewed in section 2. Section 3 presents the data. The methodology is discussed in section 4. Section 5 discuss our findings. We conclude in section 6.

2. Literature review

Previous studies on Google search trends has investigated its predictive power in various fields, such as house prices in India (Venkataraman, Panchapagesan, & Jalan, 2018) and the US (Kulkarni, Haynes, Stough, & Paelinck, 2009; Wu & Brynjolfsson, 2015), influenza epidemics (Ginsberg et al., 2008; Polgreen, Chen, Pennock, & Nelson, 2008), unemployment rates (D'amuri & Marcucci, 2017) and gasoline prices (Molnár & Bašta, 2017). Choi and Varian (2012) were one of the first to study if Google search queries could predict economic activity. They claimed that data obtained from GT could predict the present values of different macro-economic indicators, which can be useful in short-term economic prediction. The indicators investigated were home sales, travel destination planning, automotive sales and unemployment claims in the US.

Google search trends has also been widely used to study financial markets. Researchers has utilized search query data from GT to predict the dynamics of stock markets (Bijl, Kringhaug, Molnár, & Sandvik, 2016; Kim, Lučivjanská, Molnár, & Villa, 2019) investor attention (Da, Engelberg, & Gao, 2011), investor sentiment (Joseph, Babajide Wintoki, & Zhang, 2011) and stock market volatility (Dimpfl & Jank, 2016; Kim et al., 2019). Bijl et al. (2016) studied if Google Trends data could forecast stock returns in the US market and whether a trading strategy based on Google search queries would be profitable. They found the trading strategy to be profitable when the transaction costs was not included, but not profitable when the transaction costs were taken into account. The study also revealed that high Google search volumes predicts negative returns. The recent study of the impact of Google search activity on the Norwegian stock market, by Kim et al. (2019), yielded no relationship between Google search activity and stock returns. This differs from the previous findings of Bijl et al. (2016) and Da et al. (2011). However, they found that Google search queries can predict trading volume and volatility in the Norwegian stock market. Another recent study by Alborg, Molnár, and de Vries (2018), investigated the predictive ability of Google searches on Bitcoin return, its volatility and trading volume. They found that Google searches for "Bitcoin" is not able to predict return and volatility. However, Google Trends proved to be a predictor of Bitcoin trading volume.

Da et al. (2011) were one of the first to use Google's SVI as a direct measure for investor attention by using data from Russel 3000 stocks for a four-year period (2004-2008). They found that the SVI predict a higher stock price in the weeks following an increase in search volume, and also proved that SVI is better in capturing investor attention than other measures of investor attention. Later, Joseph et al. (2011) used the same framework to investigate whether data from Google trends could capture investor sentiment. They proved that Google's search volume index can predict stock prices in the following weeks and that SVI can explain investor sentiment, in line with the previous findings of Da et al. (2011).

There exists a lot of studies related to explaining the gold market and the price of gold. Former studies has focused on investigating volatility (Batten & Lucey, 2010; Baur, 2012; Bentes, 2015; Fang, Yu, & Xiao, 2018), key determinants of the gold price (Elfakhani, Baalbaki, & Rizk, 2009; Levin & Wright, 2006) and the relationship between the price of gold and crucial indicators such as the US dollar, oil price and inflation rate (Kannan & Dhal, 2008; Mo, Nie, & Jiang, 2018; Shafiee & Topal, 2010). Levin and Wright (2006) investigated the key short- and long-term determinants of the gold price. They found a statistically significant long-term connection between the general price level in the US and the gold price. This is a one-to-one relationship, meaning that a one percent increase in the US price level leads to the same increase in the price of gold. Additionally, they found short-term relationships between movements in the gold price and changes in inflation, its volatility and credit risk in the US. Recently, macroeconomic variables has been utilized to forecast volatility in the gold futures market in the US by using mixed data sampling (Fang et al., 2018). The study proved that macroeconomic variables, such as inflation and employment rate impacts volatility significantly, both during and after periods of economic turmoil. In relation to forecasting the gold price, the predictive ability of several models has been investigated (Aye, Gupta, Hammoudeh, & Kim, 2015; Hassani, Silva, Gupta, & Segnon, 2015; Shafiee & Topal, 2010). However, none of these studies included internet search trends in their prediction models.

Literature on the dynamics between Google search trends and precious metals, such as gold, is limited. It is, however growing. Baur and Dimpfl (2016) found a positive relationship between gold price volatility and Google search trends. Their findings also suggested a strong connection between gold returns and investors demand for information. Additionally, they found that investors are more likely to Google bad news rather than good news in relation to gold. When comparing gold to silver, palladium and platinum, they concluded that the results for gold were distinctive. Recently, a study from India (Jain & Biswal, 2018) examined the dynamics between Google searches on gold and the gold spot price, the stock index price and the US dollar – Indian Rupee exchange rate. Google search queries was used as a measure for investor interest in gold. They observed negative correlation between the search trends for gold and the gold spot prices in India, which suggests that investor interest increases when gold prices are decreasing.

3. Data

We employ weekly time series from December 28, 2008 to December 30, 2018. The data for the ten-year sample period was retrieved from Google Trends and The Federal Reserve Bank of St. Louis. We also collected data for the years prior to 2008, in order to standardize the SVI. In the sections following, we describe the variables used in our regression model, and how the data on these variables was retrieved, transformed and standardized.

3.1 Gold Return

The gold price data was obtained from the Federal Reserve Bank of St. Louis, which retrieved their data from the London Bullion Market Association (LBMA). The LBMA was the source of choice because it is the biggest gold exchange in the world, with estimated 70% of the global notional trading volume (World Gold Council, 2016). The price of gold is quoted in US dollars per troy ounce, and was selected due to the US dollar's nature as a comparable currency. The gold price was transformed to the rate of return on a weekly basis with the following method:

$$R_t = 100 * ln(\frac{P_t}{P_{t-1}}) \tag{1}$$

where R_t is the natural log return on gold at week t, P_t is the price of gold at week t and P_{t-1} is the price of gold at the previous week.

3.2 Google Trends

Google Trends is a service provided by Google, allowing anyone to access data about the volume of searches made on any topic. The search volume index uses a standardized scale from 0 to 100, where 0 is the lowest search volume for a given period and 100 is the highest search volume. The data for a chosen period shows SVI on a daily, weekly or monthly basis. The SVI is relative to the search volume for the chosen period. This means that overlapping data sets will most likely yield different values of SVI for a particular observation. GT can also provide search volumes on a country, regional and global level. We opted to investigate the predictive power of Google search trends on the gold price, using GT data from the US, India and globally, hence our data was retrieved on a country and global level.

Additionally, GT can filter search information based on the following categories: (1) Arts & Entertainment, (2) Autos & Vehicles, (3) Beauty & Fitness, (4) Books & Literature, (5) Business & Industrial, (6) Computer & Electronics, (7) Finance, (8) Food & Drink, (9) Games, (10) Health, (11) Hobbies & Leisure, (12) Home & Garden, (13) Internet & Telecom, (14) Jobs & Education, (15) Law & Government, (16) News, (17) Online Communities, (18) People & Society, (19) Pets & Animals, (20) Real Estate, (21) Reference, (22) Science, (23) Shopping, (24) Sports and (25) Travel. In order to exclude searches that are not related to the gold price, such as gold medals, gold jewelry, the color gold and gold memberships, we opted to apply the finance filter. Previous studies has argued that the results obtained using a filtered SVI does not improve the unfiltered searches (Bijl et al., 2016). However, we found that filtered searches provided a better predictive model for gold returns.

We collected weekly SVI, as in Bijl et al. (2016) and Da et al. (2011). Weekly data for searches made in the US, India and globally, from December 28, 2008 to December 30, 2018, was retrieved from Google trends. It was also necessary to retrieve data from 2007 due to the standardization methods. In order to obtain weekly data for the sample period, we collected data using three-year periods, starting in 2006. Further, the SVI was transformed to abnormal SVI (ASVI) using two different methods. The first method follows the formula by Bijl et al. (2016), where the average of the previous 52 weeks is subtracted from the raw SVI, and their difference is divided by the standard deviation of the previous year:

$$ASVI_t^B = \frac{\mathrm{SVI_t} - \frac{1}{52}\sum_{i=1}^{52} SVI_{t-1}}{\sigma_{SVI,t}} \quad (2)$$

where $\mathrm{SVI_t}$ is the search volume at week t and $\sigma_{SVI,t}$ is the standard deviation of the search volume index for the previous 52 weeks.

The second method used to standardize the SVI is based on Da et al. (2011), where the logarithm of the SVI is standardized relatively to its median over past 52 weeks:

$$ASVI_t^D = ln\ (\mathrm{SVI_t}) - \ln[Median(\mathrm{SVI_{t-1}}, \dots, \mathrm{SVI_{t-52}})] \quad (3)$$

3.3 Volatility

By including implied gold volatility, we can study the pure effect of Google trends on gold prices. The implied volatility of gold price is in US dollars and was obtained from the Federal

Reserve Bank of St. Louis, which retrieved their data from the Chicago Board Options Exchange. The implied volatility was transformed, using the following equation:

$$Volatility_t^D = \ln(Implied\ gold\ volatility_t) \tag{4}$$

3.4 EPU Index

The Economic Policy Uncertainty Index (EPU) is an index provided by Baker, Bloom, and Davis (2016), and was retrieved from policyuncertainty.com. The index is based on articles published in the 10 largest newspapers in the US, and use words such as: “economy” or “economic”, “uncertain” or “uncertainty”, and one or more of “congress”, “deficit”, “Federal Reserve”, “legislation”, “regulation” or “White House”. For an article to be included in the EPU index it must contain all three terms of economic, policy and uncertainty, in addition to other category-relevant terms such as “the Fed”, “central bank” and “inflation”.

Figure 1. EPU index

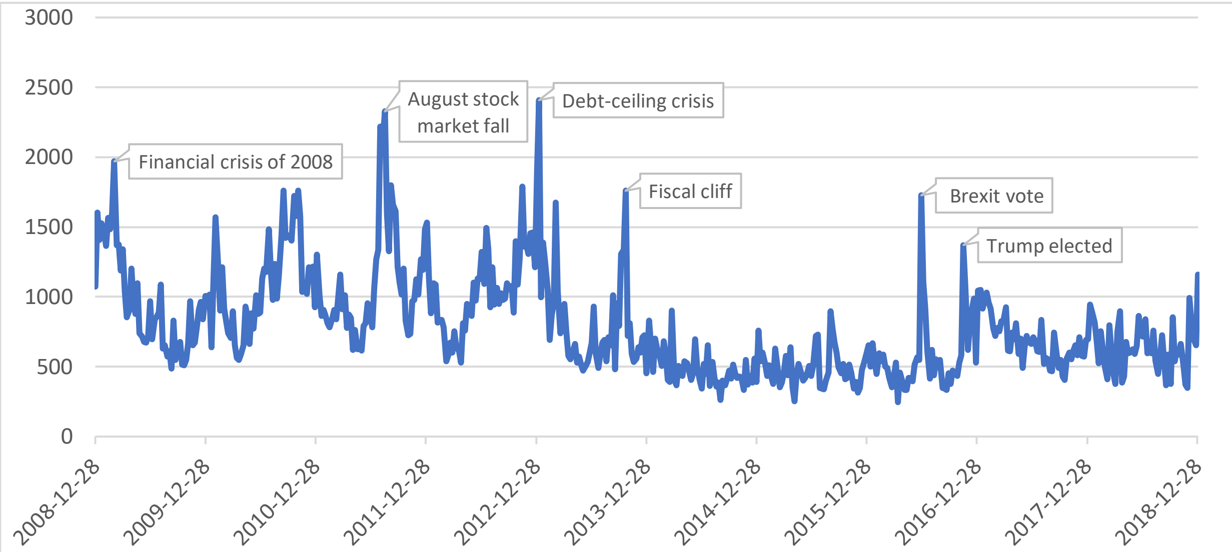

Figure 1 illustrates how major economic and political events affect the EPU index. The daily EPU data was converted to the sum of each week. Further, the EPU was transformed, using the equation below:

$$EPU_t^D = ln\ (EPU_t) \tag{5}$$

Where $\ln(EPU_t)$ is the logarithm of the sum of values of EPU during week t.

3.5 TED spread

The TED spread is the spread between two financial instruments: the Eurodollar futures contract and the 3-month Treasury Bill (T-Bill). When the spread is high, the banks believe other banks' borrowing from them are at a higher risk of default so they charge a higher interest rate, and vice versa. In this thesis we utilize the data provided, as well as the natural log of the weekly values. The TED spread was transformed to natural log as follows:

$$TED_t^D = \ln(TED_{\mathrm{t}}) \qquad (6)$$

where $ln(TED_{\mathrm{t}})$ is the weekly natural log value of the TED spread during week t.

3.6 VIX Index

The Volatility Index (VIX) was constructed by the Chicago Board Options Exchange on January 19, 1993. It attempts to predict the volatility in the financial market 30-days in the future, by using the S&P 500 index option. The index is widely used by investors to measure market risk and the sentiment of other investors. We use the logarithmic transformation of the VIX index:

$$VIX_t^D = \ln(VIX_{\mathrm{t}}) \qquad (7)$$

3.7 Summary statistics

Table 1. Summary statistics for all variables

	N	Mean	St.Dev	min	max	skewness	kurtosis
Gold return	523	0.034	0.796	-4.993	2.760	-0.793	7.044
ln(Volatility)	523	-1.728	0.292	-2.309	-0.840	0.402	3.018
ln(EPU)	523	6.575	0.428	5.505	7.787	0.236	2.526
ln(VIX)	523	2.853	0.351	2.234	3.915	0.793	3.231
ln(TED)	523	-1.248	0.448	-2.303	0.378	0.785	3.785
$ASVI_d$ Global	523	0.024	0.115	-0.181	0.796	1.691	8.759
$ASVI_d$ US	523	0.006	0.104	-0.237	0.585	1.304	6.880
$ASVI_d$ India	523	0.030	0.178	-0.352	1.155	1.328	7.211
$ASVI_b$ Global	523	0.192	2.264	-2.002	43.515	13.894	259.086
$ASVI_b$ US	523	-0.010	1.565	-1.764	25.098	8.778	130.779
$ASVI_b$ India	523	0.150	2.660	-1.650	55.003	17.059	347.73

In table 1, the summary statistics for all variables are presented. When comparing the two standardized SVIs, $ASVI_d$ and $ASVI_b$, we observed that the skewness and kurtosis of $ASVI_b$ was significantly higher. This can be explained by examining the SVI data and the method of standardization. The standardization method by Bijl et al. (2016), does not take large spikes in the SVI into account, resulting in inflated ASVI values. Thus, we concluded that the standardization method by Da et al. (2011) was more suitable for our regression analysis, and we use $ASVI_d$ in our analysis.

Further, we tested the variables for correlation before conducting regressions:

Table 2. Correlation matrix for all variables

Variables	Gold return	ln volatility	ln EPU	ln VIX	ln TED	$ASVI_d$ Global	$ASVI_d$ US	$ASVI_d$ India
Gold return	1.000							
ln(volatility)	0.024	1.000						
ln(EPU)	0.091	0.361	1.000					
ln(VIX)	0.102	0.731	0.473	1.000				
ln(TED)	-0.005	0.172	0.108	0.329	1.000			
$ASVI_d$ Global	-0.032	0.291	0.167	0.175	-0.034	1.000		
$ASVI_d$ US	0.065	0.276	0.226	0.201	-0.035	0.916	1.000	
$ASVI_d$ India	-0.152	0.218	0.036	0.093	-0.023	0.831	0.651	1.000

The correlations for all variables are presented in table 2. High correlation between the ASVI variables was observed, therefore we always utilize only one of these three variables in our models. VIX and gold volatility are correlated, which can be explained by the fact that volatility increases in general in uncertain times. Thus, there is a significant co-movement in implied volatility of stock market and implied volatility of gold.

3.8 Figures monthly rate

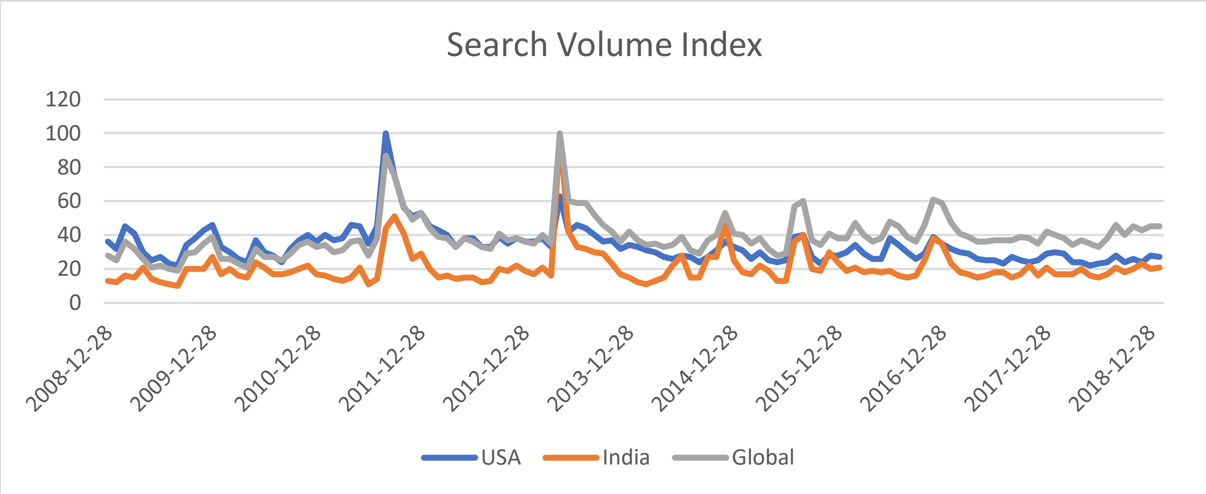

Figure 2. The Monthly search volume index for the U.S., India and Global from December 2008 to December 2018.

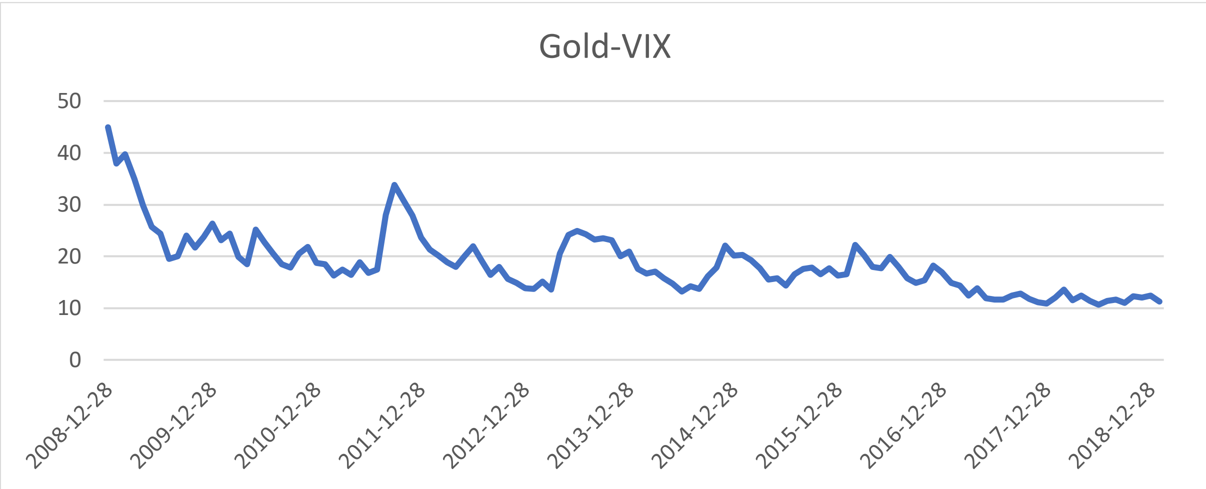

Figure 3. Gold ETF Volatility Index retrieved from the Chicago Board Options Exchange from December 2008 to December 2018.

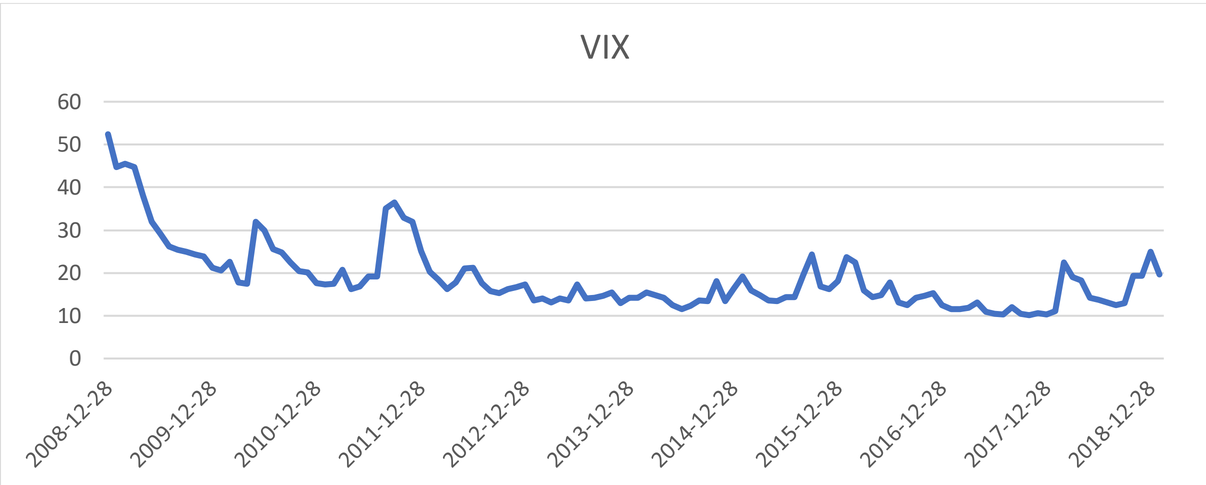

Figure 4. The monthly Volatility Index in the U.S. stock market from December 2008 to December 2018.

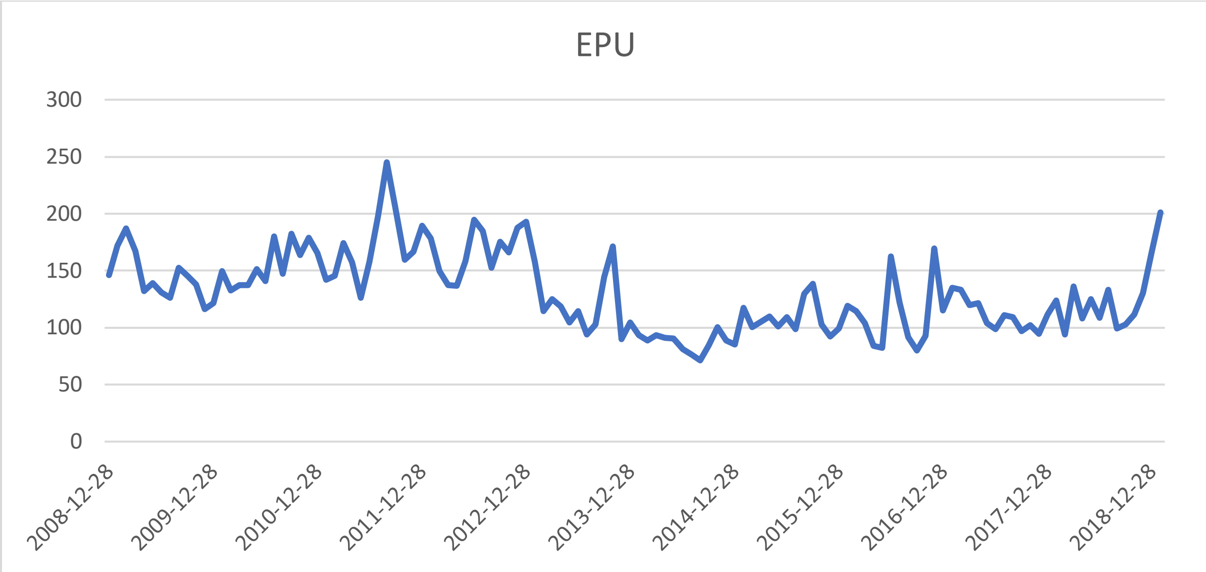

Figure 5. The monthly Economic Policy Uncertainty index in the U.S. from December 2008 to December 2018.

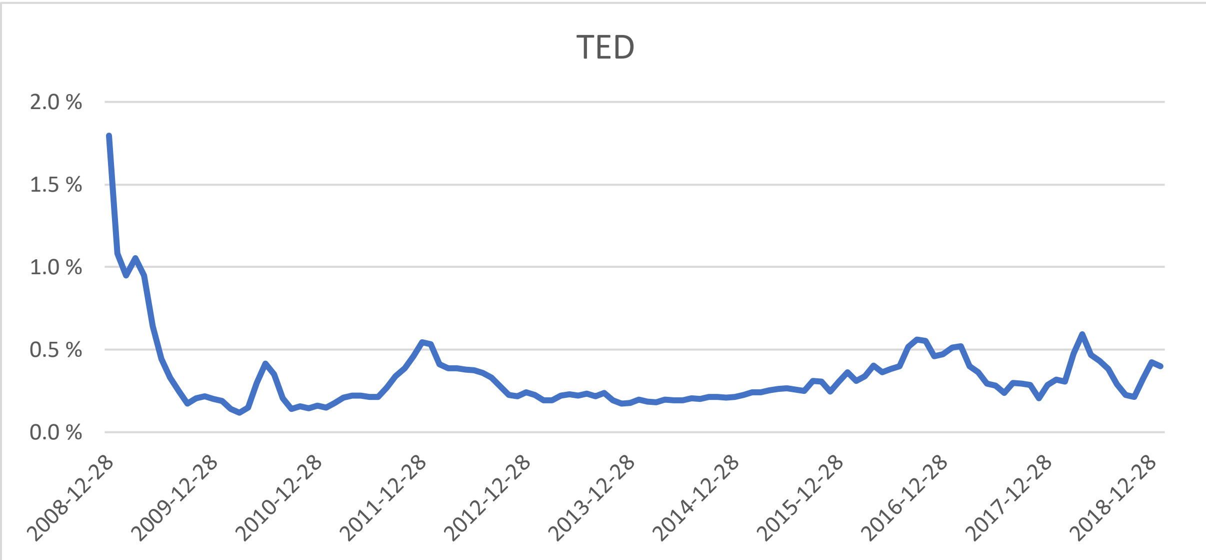

Figure 6. The monthly TED spread from December 2008 to December 2018.

In figures 2 – 6, we noticed several noteworthy observations. In the end of 2008 huge spikes appears on the uncertainty indexes due to the global financial crisis. Further, we observed that all indexes spike in August 2011, which could be explained by the stock market fall in Europe, the US, Asia and the Middle East. In addition, the SVI, Gold ETF Volatility Index and EPU index spike in August of 2013.

4. Methodology

The results were obtained using the statistical software Stata. We tested for heteroscedasticity and autocorrelation, by conducting the Breusch-Pagan and Breusch-Godfrey tests. Both autocorrelation and heteroscedasticity were detected in the dataset, thus we present the results with robust standard errors. Our models were carried out using datasets consisting of SVI from three geographic regions: India, the US and globally. We included the CBOE volatility index (VIX), TED spread (TED) and Economic Policy Uncertainty index (EPU) in our models.

4.1 Descriptive models of gold returns

In the descriptive regression model, we conducted regressions using gold returns as the dependent variable. We regressed gold returns against the ASVI, implied gold volatility, VIX index, TED spread and EPU index during the same week, which resulted in the following model:

$$R_t = \alpha + \beta_1 ASVId_t + \beta_2 \ln(Volatility_t) + \beta_3 \ln(EPU_t) + \beta_4 ln\ (VIX_t) \qquad (8)$$

$$+\beta_5 ln\ (TED_t) + \epsilon_t$$

where R_t is the return at time *t*, and βs are the coefficients for abnormal SVI, volatility, EPU index, VIX index and TED spread.

4.2 Predictive model of gold returns

Further, we wanted to investigate the predictive ability of the ASVI. In order to predict future values of gold returns, we use lagged variables. The predictive regression model of gold is stated below.

$$R_t = \alpha + \beta_1 ASVId_{t-1} + \beta_2 ln\ (Volatility_{t-1}) + \beta_3 ln\ (EPU_{t-1}) \qquad (9)$$

$$+\beta_4 ln\ (VIX_{t-1}) + \beta_5 \ln(TED_{t-1}) + \epsilon_t$$

4.3 Descriptive models of implied gold volatility

Additionally, we wanted to study the relationship between implied volatility of gold and ASVI. Due to the significant correlation past volatility has on current volatility, we decided to create two separate descriptive models. This allows us to investigate how the other independent variables explain the implied volatility of gold. In these descriptive models we used the logarithmic transformation of implied gold volatility as the dependent variable, which resulted in the following models:

$$\ln(Volatility_t) = \alpha + \beta_1 ASVId_t + \beta_2 Gold\ Return_t + \beta_3 \ln(EPU_t) + \beta_4 ln\ (VIX_t) + \beta_5 ln\ (TED_t) + \epsilon_t \tag{10}$$

$$\ln(Volatility_t) = \alpha + \beta_1 ln\ (Volatility_{t-1}) + \beta_2 ASVId_t + \beta_3 Gold\ Return_t + \beta_4 ln\ (EPU_t) + \beta_5 ln\ (VIX_t) + \beta_6 ln\ (TED_t) + \epsilon_t \tag{11}$$

4.4 Predictive model of implied gold volatility

We also investigated the predictive ability of ASVI on implied gold volatility. In order to predict future implied gold volatility, we use lagged variables. The predictive model of implied gold volatility and is presented below.

$$\ln(Volatility_t) = \alpha + \beta_1 \ln(Volatility_{t-1}) + \beta_2 ASVId_{t-1} + \beta_3 Gold\ Returns_{t-1} + \beta_4 ln(EPU_{t-1}) + \beta_5 ln\ (VIX_{t-1}) + \beta_6 ln(TED_{t-1}) + \epsilon_t \tag{12}$$

4.5 Predictive model of ASVI India

Lastly, we included a model with ASVI India as the dependent variable, due to the fact that one third of the worlds demand for gold derives from India. We wanted to investigate whether gold return and its implied volatility can predict the search volume for gold in India, which resulted in the following model:

$$ASVId_t\ India = \alpha + \beta_1 Gold\ Return_{t-1} + \beta_2 ln(Volatility_{t-1}) \tag{13}$$

5. Results

5.1 Descriptive model of gold returns

The analysis of gold returns, presented in table 3, shows interesting results, with three significant independent variables. The VIX index is statistically significant at a 5% level of confidence in all three models, with ASVI from India, the US and globally. This indicates that an increase in volatility and uncertainty in the stock markets, is associated with positive gold returns. This could be explained by the movement of risk averse investors from the stock market to the safer gold market, creating an upward shift in demand for gold.

Interestingly, VIX appears to be a better explanatory variable in comparison to the volatility of gold. The results that the VIX index (the implied volatility of the stock market) is more relevant for gold returns than implied volatility of the gold market makes sense. High implied volatility of gold does not have an obvious impact on gold returns, as there are no good safer alternatives to gold. On the other hand, high implied volatility of the stock market means that investors might choose a safer alternative (gold) to stocks, and therefore it is expected that high levels of the VIX index should be associated with positive gold returns.

Further, ASVI for India is the only significant ASVI variable at a 5% confidence level. In table 3 we observe that both the VIX index and GT India are statistically significant. However, univariate regression with GT India has R-squared of 0.023, whereas univariate regression with the VIX index has R-squared of 0.010, indicating that GT India is economically the most significant factor. The same conclusion is confirmed later in Table 4 with predictive regressions. It is not surprising that GT in India are highly important for what is happening with the gold price. In India, inflation is rather high, and gold is a rather important saving tool there.

Table 3. Descriptive model of gold volatility

	Dependent variable: Gold Return$_t$									
Independent variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
ln(Volatility)$_t$	0.0657							-0.282	-0.361*	-0.188
	(0.160)							(0.191)	(0.209)	(0.188)
ln(EPU)$_t$		0.169**						0.109	0.0854	0.0956
		(0.0857)						(0.0885)	(0.0878)	(0.0853)
ln(VIX)$_t$			0.233*					0.398**	0.412**	0.366**
			(0.123)					(0.164)	(0.167)	(0.162)
ln(TED)$_t$				-0.00827				-0.0938	-0.0799	-0.0985
				(0.0876)				(0.0850)	(0.0846)	(0.0846)
ASVId$_t$ Global					-0.220			-0.303		
					(0.509)			(0.465)		
ASVId$_t$ USA						0.504			0.412	
						(0.459)			(0.416)	
ASVId$_t$ India							-0.682**			-0.695**
							(0.317)			(0.294)
Constant	0.148	-1.080*	-0.631*	0.0238	0.0396	0.0309	0.0543*	-2.417**	-2.428**	-2.066**
	(0.297)	(0.560)	(0.342)	(0.118)	(0.0308)	(0.0336)	(0.0316)	(0.952)	(0.996)	(0.904)
Observations	522	522	522	522	522	522	522	522	522	522
R-squared	0.001	0.008	0.010	0.000	0.001	0.004	0.023	0.022	0.023	0.043

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

5.2 Predictive model of gold returns

In the predictive models of gold returns, presented in table 4, we observed that the coefficient of lagged ASVI India is significant at a 5% confidence level. This implies that investors in India investigate the gold price, when questioning the safe haven ability of gold. High Google searches reflects the situation when attention (and therefore buying pressure) for gold is already very high. Afterwards, when this attention (and therefore also buying pressure) declines, price partly reverts, which is demonstrated by negative returns. Again, we observe the significance of the search volume index in India, due to its R-squared being relatively higher than the other independent variables, emphasizing the importance of the Indian market.

Further, the lagged VIX index coefficient is significant and positive when investors are less confident in the stock market, because the implied volatility is high. This results in an upward shift in demand for gold and an increase in gold returns. The VIX index is widely used by investors to observe the current sentiment of other investors, which could explain the preference to shift their investments from the stock market to the gold market, even 30 days before the predicted increase in stock market volatility.

Table 4. Predictive model of gold returns

	Dependent variable: Gold Return$_t$									
Independent variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
ln(volatility)$_{t-1}$	0.0946							-0.215	-0.293*	-0.155
	(0.136)							(0.176)	(0.168)	(0.171)
ln(EPU)$_{t-1}$		0.145*						0.0750	0.0662	0.0473
		(0.0870)						(0.0891)	(0.0893)	(0.0881)
ln(VIX)$_{t-1}$			0.258**					0.453***	0.476***	0.427***
			(0.120)					(0.150)	(0.149)	(0.150)
ln(TED)$_{t-1}$				-0.0326				-0.141*	-0.133	-0.139
				(0.0877)				(0.0843)	(0.0840)	(0.0844)
ASVId$_{t-1}$ Global					-0.648			-0.791		
					(0.698)			(0.718)		
ASVId$_{t-1}$ USA						-0.142			-0.315	
						(0.732)			(0.752)	
ASVId$_{t-1}$ India							-0.821**			-0.856**
							(0.395)			(0.399)
Constant	0.198	-0.918	-0.701**	-0.00660	0.0499*	0.0350	0.0585*	-2.281***	-2.429***	-1.912**
	(0.252)	(0.566)	(0.334)	(0.117)	(0.0303)	(0.0329)	(0.0311)	(0.843)	(0.828)	(0.856)
Observations	522	522	522	522	522	522	522	522	522	522
R-squared	0.001	0.006	0.013	0.000	0.009	0.000	0.034	0.035	0.025	0.058

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

5.3 Descriptive models of implied gold volatility

In the descriptive models of implied gold volatility, presented in tables 5 and 6, we observed that implied volatility is best explained by the VIX index, even without including past volatility in model. The coefficient is positive, meaning that the increase in the VIX index results in an increase in implied gold volatility. An increase in google searches in period t is associated with an increase in volatility in the same period. Note that the VIX index as an independent variable has the highest R-squared of all the independent variables in the single regression models.

Table 5. Descriptive model of implied gold volatility

	Dependent variable: ln(volatility)$_t$									
Independent variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Gold Return$_t$	0.00871							-0.0166	-0.0217*	-0.0114
	(0.0216)							(0.0106)	(0.0114)	(0.0111)
ln(EPU)$_t$		0.245***						0.000576	-0.00204	0.0127
		(0.0285)						(0.0220)	(0.0221)	(0.0219)
ln(VIX)$_t$			0.607***					0.605***	0.609***	0.610***
			(0.0225)					(0.0267)	(0.0265)	(0.0269)
ln(TED)$_t$				0.113***				-0.0407**	-0.0423**	-0.0448**
				(0.0374)				(0.0196)	(0.0195)	(0.0195)
ASVId$_t$ Global					0.732***			0.403***		
					(0.122)			(0.0736)		
ASVId$_t$ USA						0.772***			0.366***	
						(0.129)			(0.0748)	
ASVId$_t$ India							0.356***			0.233***
							(0.0712)			(0.0447)
Constant	-1.730***	-3.339***	-3.462***	-1.588***	-1.748***	-1.735***	-1.740***	-3.518***	-3.509***	-3.616***
	(0.0128)	(0.184)	(0.0644)	(0.0525)	(0.0122)	(0.0122)	(0.0123)	(0.132)	(0.132)	(0.128)
Observations	522	522	522	522	522	522	522	522	522	522
R-squared	0.001	0.130	0.534	0.030	0.085	0.076	0.048	0.567	0.558	0.562

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6. Descriptive model of implied gold volatility

	Dependent variable: ln(volatility)$_t$										
Independent variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**
ln(volatility)$_{t-1}$	0.955***								0.857***	0.863***	0.860***
	(0.0111)								(0.0159)	(0.0156)	(0.0161)
Gold Return$_t$		0.00871							-0.00543	-0.00840	-0.00257
		(0.0216)							(0.00795)	(0.00876)	(0.00815)
ln(EPU)$_t$			0.245***						-0.0103	-0.0131	-0.00343
			(0.0285)						(0.00811)	(0.00819)	(0.00808)
ln(VIX)$_t$				0.607***					0.106***	0.105***	0.108***
				(0.0225)					(0.0150)	(0.0150)	(0.0153)
ln(TED)$_t$					0.113***				-0.0165**	-0.0165**	-0.0189***
					(0.0374)				(0.00667)	(0.00676)	(0.00685)
ASVId$_t$ Global						0.732***			0.232***		
						(0.122)			(0.0466)		
ASVId$_t$ USA							0.772***			0.240***	
							(0.129)			(0.0453)	
ASVId$_t$ India								0.356***			0.130***
								(0.0712)			(0.0275)
Constant	-0.0805***	-1.730***	-3.339***	-3.462***	-1.588***	-1.748***	-1.735***	-1.740***	-0.510***	-0.473***	-0.558***
	(0.0198)	(0.0128)	(0.184)	(0.0644)	(0.0525)	(0.0122)	(0.0122)	(0.0123)	(0.0729)	(0.0720)	(0.0767)
Observations	522	522	522	522	522	522	522	522	522	522	522
R-squared	0.925	0.001	0.130	0.534	0.030	0.085	0.076	0.048	0.941	0.940	0.939

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

5.4 Predictive model of implied gold volatility

In the predictive models of implied gold volatility (table 7), the lagged volatility coefficient is positive and significant at the 1% level of confidence. This is not surprising, as we expected that past volatility would impact current volatility. Moreover, the gold return coefficient is positive and significant at a 1% confidence level. These results implies that an increase in gold price in the previous week results in a higher volatility the following week. This differs from the leverage effect commonly found in the stock market, where an increase in price causes a decrease in stock market volatility.

Further, an increase in economic policy uncertainty is associated with decreasing gold market volatility. The decrease in volatility could be an incentive for investors to invest in the gold market due to it being a more predictable investment. The lagged VIX coefficient is positive and statistically significant at 5% confidence level in all models. These results suggest that the gold market volatility increases when there is an increase in stock market volatility. This could be explained by the increasing number of investors interested in pursuing a safer investment, such as gold, and reflects our results for the predictive model of gold return.

Additionally, all lagged ASVI coefficients are positive and significant at the 1% level of confidence, implying that an increase in search volume for gold price predicts an increase in gold price volatility the following week. This confirms our findings from the predictive model of gold return (table 4), which yielded negative returns when SVI is increasing, resulting in a larger price deviation from the mean.

Table 7. Predictive model of implied gold volatility

	Dependent variable: ln(volatility)$_t$										
Independent variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**
ln(volatility)$_{t-1}$	0.955***								0.896***	0.902***	0.909***
	(0.0111)								(0.0169)	(0.0155)	(0.0172)
Gold return$_{t-1}$		0.0210							0.0136***	0.00989**	0.0163***
		(0.0202)							(0.00512)	(0.00494)	(0.00493)
ln(EPU)$_{t-1}$			0.240***						-0.0120	-0.0167**	-0.00390
			(0.0286)						(0.00843)	(0.00838)	(0.00880)
ln(VIX)$_{t-1}$				0.600***					0.0371**	0.0333**	0.0346**
				(0.0225)					(0.0155)	(0.0149)	(0.0161)
ln(TED)$_{t-1}$					0.114***				-0.00287	-0.00186	-0.00603
					(0.0367)				(0.00694)	(0.00689)	(0.00713)
ASVId$_{t-1}$ Global						0.944***			0.277***		
						(0.0938)			(0.0596)		
ASVId$_{t-1}$ USA							1.036***			0.324***	
							(0.107)			(0.0615)	
ASVId$_{t-1}$ India								0.452***			0.131***
								(0.0609)			(0.0399)
Constant	-0.0805***	-1.730***	-3.310***	-3.441***	-1.587***	-1.753***	-1.736***	-1.743***	-0.219***	-0.161**	-0.244***
	(0.0198)	(0.0127)	(0.185)	(0.0647)	(0.0514)	(0.0120)	(0.0118)	(0.0122)	(0.0789)	(0.0769)	(0.0842)
Observations	522	522	522	522	522	522	522	522	522	522	522
R-squared	0.925	0.003	0.126	0.525	0.031	0.141	0.137	0.077	0.937	0.938	0.932

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

5.5 Predictive model of ASVI India

ASVI India proved to be the most significant independent variable in our descriptive and predictive models. This is not surprising due to the relatively higher market demand for gold in India compared other countries. The high demand for gold does not only reflect gold as a consumption good and financial commodity, but also its social-cultural dimensions. Thus, we opted to investigate the predictive ability of gold return and implied gold volatility on ASVI India.

We find that an increase in gold price is associated with decreased Google searches for gold, presented in table 8. However, increased implied volatility of gold predicts an increase in search volume. The positive relationship between volatility and ASVI India implies that investors search for gold price information when the uncertainty in the gold market is high.

Table 8. Predictive model of ASVI India

Dependent variable: $ASVId_t$ India			
Independent variables	(1)	(2)	(3)
Gold returns$_{t-1}$	-0.0236*		-0.0244**
	(0.0124)		(0.0116)
ln(volatility)$_{t-1}$		0.0876***	0.0891***
		(0.0258)	(0.0255)
Constant	0.0304***	0.181***	0.184***
	(0.00780)	(0.0474)	(0.0469)
Observations	522	522	522
R-squared	0.011	0.021	0.033

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6. Conclusion

We investigate the relationship between investor attention to gold, and gold returns and its implied volatility. As a measure of investor attention to gold, we make use of Google's search volume index for the word "gold". We consider Google searches in three areas: the US, India, and globally. Additionally, we study the impact of three market uncertainty indexes (EPU index, VIX index and TED spread) on returns and volatility of gold.

We find that increased Google searches for "gold" are associated with negative present and future gold returns, and high present and future implied volatility of gold. Considering the Google searches originated in three areas (the US, India, global), we find that Google searches from India are most strongly related to what is occurring in the gold market. These findings emphasizes the impact of the Indian gold demand and its relevance as a global indicator for investor attention to gold. The negative relationship of the SVI for India with gold returns is in line with the previous findings of Jain and Biswal (2018). In addition, we find that an increase in search volume of gold are associated with high present and future implied volatility of gold.

Furthermore, our results implies that high levels of the VIX index (high implied stock market volatility) has a significant impact on present and future gold return, in contrary to high levels of implied gold volatility, and is associated with positive gold returns. In the presence of high stock market uncertainty, risk averse investors seek out safe haven commodities, such as gold.

Regarding further research we suggest including countries that are large importers of gold by using internet search trends from their preferred search engines, such as China and the search engine Baidu. It is also noteworthy that applying gold related news to search volume indexes might produce more robust estimates and models. In relation to the market uncertainty indexes used in our regressions, it is important to consider that the EPU index only reflects the economic uncertainty in the US. This suggests that the EPU index might not capture the various regional uncertainties, and that applying it with search volume indexes from India and globally could be sub optimal. In addition, the VIX index only reflects the US stock market, and it could be useful to include volatility indexes of stock markets in other countries in the models.

References

Aalborg, H. A., Molnár, P., & de Vries, J. E. (2018). What can explain the price, volatility and trading volume of Bitcoin? *Finance Research Letters*. doi:10.1016/j.frl.2018.08.010

Aye, G., Gupta, R., Hammoudeh, S., & Kim, W. J. (2015). Forecasting the price of gold using dynamic model averaging. *International Review of Financial Analysis, 41*(C), 257-266. doi:10.1016/j.irfa.2015.03.010

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The quarterly journal of economics., 131*(4), 1593-1636. doi:10.1093/qje/qjw024

Batten, J. A., & Lucey, B. M. (2010). Volatility in the gold futures market. *Applied Economics Letters, 17*(2), 187-190. doi:10.1080/13504850701719991

Baur, D. G. (2012). Asymmetric volatility in the gold market. *Journal of Alternative Investments, 14*(4), 26-38. doi:10.3905/jai.2012.14.4.026

Baur, D. G., & Dimpfl, T. (2016). Googling gold and mining bad news. *Resources Policy, 50*, 306-311. doi:10.1016/j.resourpol.2016.10.013

Bentes, S. R. (2015). Forecasting volatility in gold returns under the GARCH, IGARCH and FIGARCH frameworks: New evidence. *Physica A, 438*, 355-364. doi:10.1016/j.physa.2015.07.011

Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis, 45*, 150-156. doi:10.1016/j.irfa.2016.03.015

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record, 88*(Supp.1), 2-9. doi:10.1111/j.1475-4932.2012.00809.x

D'amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting, 33*(4), 801-816. doi:10.1016/j.ijforecast.2017.03.004

Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *Journal of Finance, 66*(5), 1461-1499. doi:10.1111/j.1540-6261.2011.01679.x

Dimpfl, T., & Jank, S. (2016). Can Internet Search Queries Help to Predict Stock Market Volatility? *European Financial Management, 22*(2), 171-192. doi:10.1111/eufm.12058

Elfakhani, S., Baalbaki, I., & Rizk, H. (2009). *Gold price determinants: empirical analysis and implications* (Vol. 4).

Fang, L., Yu, H., & Xiao, W. (2018). Forecasting gold futures market volatility using macroeconomic variables in the United States. *Economic Modelling, 72*, 249-259. doi:10.1016/j.econmod.2018.02.003

Ginsberg, J., H. Mohebbi, M., S. Patel, R., Brammer, L., S. Smolinski, M., & Brilliant, L. (2008). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012. doi:10.1038/nature07634

Harford, T. (2017). Just google it: The student project that changed the world. Retrieved from https://www.bbc.com/news/business-39129619

Hassani, H., Silva, E. S., Gupta, R., & Segnon, M. K. (2015). Forecasting the price of gold. *Applied Economics, 47*(39), 1-12. doi:10.1080/00036846.2015.1026580

Jain, A., & Biswal, P. C. (2018). Does internet search interest for gold move the gold spot, stock and exchange rate markets? A study from India. *Resources Policy*. doi:10.1016/j.resourpol.2018.04.016

Joseph, K., Babajide Wintoki, M., & Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting, 27*(4), 1116-1127. doi:10.1016/j.ijforecast.2010.11.001

Kannan, R., & Dhal, S. (2008). *India's demand for gold: some issues for economic development and macroeconomic policy* (Vol. 7).

Kim, N., Lučivjanská, K., Molnár, P., & Villa, R. (2019). Google searches and stock market activity: Evidence from Norway. *Finance Research Letters, 28*, 208-220. doi:10.1016/j.frl.2018.05.003

Kulkarni, R., Haynes, K., Stough, R., & Paelinck, J. (2009). *Forecasting Housing Prices with Google Econometrics*.

Levin, E. J., & Wright, R. (2006). *Short-run and Long-run Determinants of the Price of Gold*.

Mo, B., Nie, H., & Jiang, Y. (2018). Dynamic linkages among the gold market, US dollar and crude oil market. *Physica A: Statistical Mechanics and its Applications, 491*, 984-994. doi:10.1016/j.physa.2017.09.091

Molnár, P., & Bašta, M. (2017, 6-9 June 2017). *Google searches and Gasoline prices.* Paper presented at the 2017 14th International Conference on the European Energy Market (EEM).

Polgreen, P. M., Chen, Y., Pennock, D. M., & Nelson, F. D. (2008). Using internet searches for influenza surveillance. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 47*(11), 1443. doi:10.1086/593098

Rani, A. S., & Vijayalakshmi, K. (2014). Importance of Gold in Indian Economy. *Facts For You*.

Shafiee, S., & Topal, E. (2010). An overview of global gold market and gold price forecasting. *Resources Policy, 35*(3), 178-189. doi:10.1016/j.resourpol.2010.05.004

Venkataraman, M., Panchapagesan, V., & Jalan, E. (2018). Does internet search intensity predict house prices in emerging markets? A case of India. *Property Management, 36*(1), 103-118. doi:10.1108/PM-01-2017-0003

World Gold Council. (2016). Major global trading hubs. Retrieved from https://www.gold.org/what-we-do/gold-market-structure/global-gold-market?fbclid=IwAR33N0lxORbY9xfQpF8fFMkZa2FrwR6gFwRwqKznohp2jXBRbbcvUsQfX44

Wu, L., & Brynjolfsson, E. (2015). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. In *Economic Analysis of the Digital Economy* (pp. 89-118): National Bureau of Economic Research, Inc.