HETEROSEDASTICITY IN EXCESS BITCOIN RETURN DATA: GOOGLE TREND VS. GARCH EFFECTS

Chamil W. Senarathne
School of Economics, Wuhan University of Technology, Wuhan, China
e-mail: chamil@whut.edu.cn
ORCID: 0000-0002-4813-0213

Tijana Šoja
Central Bank of Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina
e-mail: tijana.soja@cbbh.ba
ORCID: 0000-0002-5289-2173

© 2019 Chamil W. Senarathne, Tijana Šoja
This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivs license (http://creativecommons.org/licenses/by-nc-nd/3.0/)
DOI: 10.15611/fins.2019.3.04
JEL Classification: C58, D53, E51, G12, G14, G17

Abstract: This paper examines the mixture of distribution properties associated with heteroskedastic excess Bitcoin return data, using the volume of Google search queries as a proxy for the information arrival time, from a monthly data sampling period of June 2010 to May 2019. The results show that the volatility coefficients become highly statistically insignificant when the lagged volume of search queries is included in the conditional variance equation of the GJR-GARCH-M model. This clearly suggests that the volume of search queries is shown to provide significant explanatory power regarding the variance of heteroskedastic excess Bitcoin return, which can be traced from the ARCH process defined in the GJR-GARCH-M specification. A significant negative relationship between the conditional volatility and the volume of search queries indicates that Internet (online) information arrival reduces the risk premium in the Bitcoin market, which may improve market stability.

Keywords: Bitcoin, information flow, GARCH-in-Mean, GARCH effects, Google trend.

1. Introduction

Nowadays, the price of Bitcoin is changing at an unprecedented pace. As of January 1, 2017, the price of Bitcoin was $995.44 and it increased significantly in 2017 to reach its highest value of $19,345.49 on 16 December 2017. An investor who had held Bitcoin during this period would have earned a return of 1843.41% for just about a one-year holding period. In such a speculative market, there is little that absolute return volatility models could do in asset pricing. Along these lines, Roll [1988] was
the first to show how firm-specific stock return variation could be explained by goodness-of-fit from an asset pricing regression in the presence of a mixture of distributions of stock returns. Although Bitcoin does not have an underlying asset, as stocks have underlying firms, security specific variation could be tested by Generalized Autoregressive Conditional Heteroskedasticity (GARCH) methodology because price changes tend to cluster in time [Urquhart 2017]. A number of studies show that the Bitcoin return is very weakly associated with financial assets (e.g. stocks), economic variables (e.g. economic growth) or return on various commodities (e.g. gold) (see [Trautman, Dorman 2018; Baur et al. 2018; Chan et al. 2019]) and, as such, no market model could accurately forecast the magnitude of Bitcoin price changes based on economic data. In contrast to these research findings, Kjærland et al. [2018] find a significant relationship between variables such as the CBOE volatility index (VIX), oil, gold, and Bitcoin transaction volume, whereas Bouoiyour et al. [2016] conclude that Bitcoin, as a speculative asset, is mainly driven by long-term fundamentals.

Pavel et al. [2016] find that the market forces and attractiveness of Bitcoin for investors have a significant impact on Bitcoin’s price. They also find that the impact is not stable and varies over time. This suggests that macro and microeconomic dynamics have some kind of impact on Bitcoin price formation. Georgoula et al. [2015] explore the relationship between Bitcoin price and fundamental economic variables. Their results show that the value of Bitcoin is negatively affected by the exchange rate between the USD and the euro, and the results also show that the Bitcoin price is positively associated with the number of Bitcoins in circulation and negatively associated with the S&P 500.

The notion of randomness in stock price changes was first identified by Jules Regnault, a French stockbroker, in 1863 before his immediate successor Louis Jean-Baptiste Alphonse Bachelier advanced the idea further in his PhD thesis in 1900. Clark [1973] then showed why the probability distribution of the daily price change is leptokurtic, giving particular reference to the randomness in the number of within-day transactions (i.e. number of transactions). His theory suggests that the daily price changes are generated from a stochastic process, in which the price changes are driven by a single latent mixing variable (i.e. volume of trading). Hence the mean of the variance of daily number of transactions is proportional to the mean number of daily transactions. It is likely that this phenomenon could be observed in other imminent proxies such as the volume of online search queries.

Figure 1 shows the relationship between Bitcoin price and Google trend data. At every point in time the average volume of Google search queries is closely associated with the price of Bitcoin. More importantly, there is no significant difference between the distributions of the average volume of 17 search terms and ‘Bitcoin’ search term data as they follow the same pattern. This observation suggests that Google search query data may contain much valuable information about the volatility dynamics of Bitcoin price changes. Kristoufek [2013] shows a strong
positive correlation between Bitcoin search query data and Bitcoin’s price. He utilizes a wavelet coherency analysis and finds correlations between Bitcoin’s exchange price and various economic factors [Kristoufek 2015]. Internet of things contributes enormously to the performance and efficiency of the cryptocurrency market. In particular, Garcia et al. [2014] examine the impact of various IT-based social interactions on the exchange rate of Bitcoin and find that Bitcoin’s exchange rates are influenced by online information searches and information sharing over online social networks. More recently, Urquhart [2018] examined the predictive power for realized volatility and returns using Google trend data and found no significant contemporaneous relationship between the variables. However, they have found that attention as proxied by Google trend data of Bitcoin is significantly influenced by the previous day’s high realized volatility and volume. This provides a clear motivation to study the heteroskedasticity in excess Bitcoin return data using the framework of Lamoureux and Lastrapes [1990].

The tendency that large changes in market prices are followed by large changes, of either sign, and small changes in prices are followed by small changes, as noted by Mandelbrot [1963], could be effectively observed in the Bitcoin market [Urquhart 2017]. Scholars such as Kurihara and Fukushima [2018] show that these clusters persist over a long-term in the Bitcoin market and, as such, it encourages
the application of stochastic volatility models, such as the GARCH model, for the forecast of return and variance. GARCH, as a mean-reverting stochastic volatility model, could accurately capture the phenomenon of volatility clustering and can explain the effects of kurtosis (e.g. non-normality) on the forecast of past errors.

The objective of this paper is to examine whether the heteroskedasticity in excess Bitcoin return data could be explained by the time dependence in the Google trend data, using the framework of Lamoureux and Lastrapes [1990], from a monthly data sampling period of June 2010 to May 2019. The results show that ARCH and GARCH effects vanish and the total volatility persistence becomes negligible when the average Google search volume is included in the conditional variance equation. Furthermore, the average Google search volume is significantly and negatively related to conditional volatility. The paper is organized as follows. Section 2 provides the methodological framework. Section 3 describes the data set including its empirical properties. Section 4 discusses the findings, and Section 5 provides the concluding remarks along with the practical implications.

2. Theoretical specification

Following Lamoureux and Lastrapes [1990] and many others (see e.g. [Sharma et al. 1996; Choi et al. 2012; Zhang et al. 2014; Senarathne, Jayasinghe 2017; Senarathne, Wei 2018; Senarathne 2019a; Senarathne 2019b; Senarathne 2019c]) define $\delta_{jt}$ and denote the $j^{th}$ intraday equilibrium excess market price increment in day $t$ summed up over a monthly data horizon as follows. These increments determine the relation between mean and variance of excess Bitcoin returns.

$$
\varepsilon_t = \sum_{j=1}^{n_t} \left( \delta_{jt} - \sqrt{\sigma_{jt}^2} \right).
$$

(1)

The mixing variable $n_t$ reflects the aggregate amount of online information arrival at the Bitcoin market, subject to the assumption that the information arrives sequentially rather than simultaneously. Define $\Omega$, the persistence of conditional variance estimated by GJR-GARCH-in-Mean (GJR-GARCH-M) model, such that, $\Omega = E(\varepsilon_t^2 | n_t)$. The mixture model requires, $\Omega = \sigma^2 n_t$ and $\varepsilon_t | n_t \sim N(0, \sigma^2 n_t)$.

Consider the Engle et al. [1987] version of variance estimate with an asymmetric term for the excess Bitcoin return in the sense of Glosten et al. [1993]. The model outlined below is helpful in describing the risk-return tradeoff under time-varying risk premium.

$$
R_t = \mu_{t-1} + \delta \sqrt{h_t} + \varepsilon_t,
$$

(2)

$$
\varepsilon_t \backslash (\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) \sim N(0, h_t),
$$

(3)
Heteroskedasticity in excess Bitcoin return data: Google trend vs. GARCH effects

\[ h_t = \varphi + \pi_1 \varepsilon_{t-1}^2 + \pi_2 I_{(t-1)} \varepsilon_{(t-1)}^2 + \pi_3 \sigma_{(t-1)}^2. \] (4)

Usually, \( R_t \equiv (R_t - R_f) \), where the excess return is \( \varepsilon_t = R_t - \sqrt{h_t} \) on holding risky asset/s when \( \mu_{t-1} \) is constrained to zero. The GJR-GARCH-M model described above offers a more meaningful way to analyze the relationship between time-varying risk premium and its conditional volatility in the Bitcoin market. That is, the standard deviation of each observation is allowed to affect the heteroscedastic disturbance drawn from a mixture model. The model allows the conditional variance to affect the mean of return and parameterizes the evolution of mean and the variance of excess Bitcoin return simultaneously subject to,

\[ R_t | \Phi_{t-1} \sim N(\mu_{t-1} + \delta \sqrt{h_t}, h_t), \] (5)

where \( \Phi \) is the information set available to investors at time \( t - 1 \) and \( \varphi \) is the intercept term of the conditional variance. \( h_t \) is the conditional variance at time \( t \) and the non-negative terms, \( \pi_1 \) and \( \pi_3 \) are the short-term (ARCH) and long-term volatility (GARCH) coefficients of the conditional variance equation. Any asymmetric volatility effect is accounted for by coefficient \( \pi_2 \) which is expected to be negative and statistically significant if a negative shock has a greater impact on volatility than the positive shocks of the same magnitude. \( I \) is a dummy variable employed under condition \( \begin{cases} 1 \text{ if } \varepsilon_{t-1} < 0 \\ 0 \text{ otherwise} \end{cases} \). In the process of estimating variance, \( \mu_{t-1} \) is constrained to zero in the sense of Lamoureux and Lastrapes [1990] and Senarathe and Wei [2018, p. 140].

Urquhart [2018] finds that the attention of Bitcoin as proxied by Google search volume is significantly influenced by the previous day’s high realized volatility. It is therefore plausible to surmise that the heteroskedasticity based volatility (i.e. GJR-GARCH) of excess Bitcoin return could be traced by the serial correlation associated with Google trend data. In order to understand such time dependence in the rate of information arrival at the Bitcoin market, the average volume of Google search queries is introduced into the conditional variance equation (4) in the sense of Lamoureux and Lastrapes [1990]. Then, the conditional variance equation (4) becomes,

\[ h_t = \varphi + \pi_1 \varepsilon_{(t-1)}^2 + \pi_2 I_{(t-1)} \varepsilon_{(t-1)}^2 + \pi_3 \sigma_{(t-1)}^2 + \lambda GoogleTrend_{\text{Average}}^\text{Average}. \]

Under null hypothesis of excess Bitcoin return is characterized by the type of asymmetric GARCH-M model described above and reflects the time dependence in the rate of information arrival at the Bitcoin market on Google search queries, the total volatility persistence as measured by the sum of coefficients \( \pi_1 + \pi_2 + \pi_3 \) should be negligible and statistically insignificant.
3. Data and empirical results

Daily USD Bitcoin quotes (BTC-USD) are obtained from the Yahoo Finance webpage covering a monthly data sampling period of June 2010 to May 2019. The volume of Google search query data (the resulting numbers are scaled on a range of 0 to 100 based on topics’ proportion to all searches by Google) are obtained from Google webpage (https://trends.google.com) for the 25 top search terms of Bitcoin. When the word ‘bitcoin’ is typed on the Google trend search bar, it suggests the most searched 25 related queries but the suggestions vary instantly depending on the latest trends and updates. The average volume of all search queries is taken as the proxy for the mixing variable. The search terms that are not directly relevant for Bitcoin returns were removed, and only directly relevant search terms were considered (i.e. 17 search terms). The access to dataset and the Eviews workfile is available at http://dx.doi.org/10.17632/t6z3jkh494.1. Some descriptive statistics of the sample data are given in Table 2.

Table 1. Top 25 search terms suggested by Google trend

<table>
<thead>
<tr>
<th>No</th>
<th>Search term</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>price bitcoin</td>
<td>√</td>
</tr>
<tr>
<td>2</td>
<td>bitcoin usd</td>
<td>√</td>
</tr>
<tr>
<td>3</td>
<td>mining bitcoin</td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>bitcoin news</td>
<td>√</td>
</tr>
<tr>
<td>5</td>
<td>buy bitcoin</td>
<td>√</td>
</tr>
<tr>
<td>6</td>
<td>bitcoin to usd</td>
<td>√</td>
</tr>
<tr>
<td>7</td>
<td>bitcoin kurs</td>
<td>√</td>
</tr>
<tr>
<td>8</td>
<td>bitcoin value</td>
<td>√</td>
</tr>
<tr>
<td>9</td>
<td>what is bitcoin</td>
<td>√</td>
</tr>
<tr>
<td>10</td>
<td>bitcoin cash</td>
<td>X</td>
</tr>
<tr>
<td>11</td>
<td>bitcoin wallet</td>
<td>√</td>
</tr>
<tr>
<td>12</td>
<td>btc</td>
<td>X</td>
</tr>
<tr>
<td>13</td>
<td>ethereum</td>
<td>X</td>
</tr>
<tr>
<td>14</td>
<td>free bitcoin</td>
<td>X</td>
</tr>
<tr>
<td>15</td>
<td>bitcoins</td>
<td>√</td>
</tr>
<tr>
<td>16</td>
<td>bitcoin usd price</td>
<td>√</td>
</tr>
<tr>
<td>17</td>
<td>bitcoin stock</td>
<td>√</td>
</tr>
<tr>
<td>18</td>
<td>bitcoin chart</td>
<td>√</td>
</tr>
<tr>
<td>19</td>
<td>litecoin</td>
<td>X</td>
</tr>
<tr>
<td>20</td>
<td>bitcoin today</td>
<td>√</td>
</tr>
<tr>
<td>21</td>
<td>bitcoin exchange</td>
<td>√</td>
</tr>
<tr>
<td>22</td>
<td>price of bitcoin</td>
<td>√</td>
</tr>
<tr>
<td>23</td>
<td>bitcoin euro</td>
<td>√</td>
</tr>
<tr>
<td>24</td>
<td>bitcoin miner</td>
<td>X</td>
</tr>
<tr>
<td>25</td>
<td>coinbase</td>
<td>X</td>
</tr>
</tbody>
</table>

Source: own work.
Table 2. Empirical description of the sample data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Max.</th>
<th>Min.</th>
<th>JB</th>
<th>ADF</th>
<th>Q (36)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0.113</td>
<td>0.072</td>
<td>1.742</td>
<td>−0.492</td>
<td>152.08</td>
<td>−7.95</td>
<td>51.26</td>
</tr>
<tr>
<td>GT</td>
<td>6.48</td>
<td>2.65</td>
<td>100.00</td>
<td>0.00</td>
<td>4635.19</td>
<td>−3.47</td>
<td>194.71</td>
</tr>
</tbody>
</table>

Note: JB – Jarque-Bera test statistic for normality. Under null hypothesis for normality, critical value of χ² (2) distribution at 5% significance level is 5.99. ADF – augmented Dickey-Fuller test statistic for stationarity of data for maximum 18 lags. Under null hypothesis for data having unit root, the critical value at 5% significance level is −2.87. LM is the ARCH LM test statistic for number of observations multiplied by the R-squared value for 3 lags. Under null hypothesis, critical value of χ² (3) distribution at 5% significance level is 7.815 (OLS equation). Q (36) is the Ljung-Box Q statistic for serial correlation up to 36 lags. Under null hypothesis for no serial correlation, the critical value of χ² (36) distribution at 5% significance level is 50.99.

Source: own work.

Table 2 outlines the statistical properties of the sample data. According to the descriptive statistics, Bitcoin return and Google trend data are highly non-normal as the test statistics exceed the critical value of 5.99 under the Jarque-Bera test. The null hypothesis of data having unit root is nearly rejected at 5% significance level for Google trend data and Bitcoin returns are clearly stationary as the test statistic substantially exceeds its critical value of −2.87. The test results of the Ljung-Box Q test are reported for detecting serial correlation in the time series data up to 36 lags. Bitcoin return is approximately serially correlated as the test statistic marginally exceeds its critical value of 50.99 at 5 percent significance level. However, Google trend data are highly serially correlated as hypothesized under the methodological framework. The test statistic of Ljung-Box Q exceeds its critical value substantially.

Table 3. Estimation results

<table>
<thead>
<tr>
<th>Equation</th>
<th>Maximum likelihood estimation of GJR-GARCH-M model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>π₁</td>
</tr>
<tr>
<td>hₜ without GT</td>
<td>0.020**</td>
</tr>
<tr>
<td>hₜ with GT</td>
<td>0.551</td>
</tr>
</tbody>
</table>

Note: The coefficient of equity risk premium in the mean as in equation (4) and (6) are 0.291 ** (t-statistic, 3.023) and 0.216* (t-statistic, 2.562) respectively. **Statistically significant at 1% assuming returns are conditionally normally distributed. *Statistically significant at 5%. The coefficients are estimated using the methods described by Bollerslev and Wooldridge (1992) for obtaining quasi-maximum likelihood (QML) covariances and robust standard errors. The residual diagnostic test statistics of JB (2), ARCH-LM (Obs.*(3)), Q (20), Log likelihood, Durbin-Watson statistic respectively for each equation are; conditional variance equation without GT (39.80, 0.88, 26.58, −21.17, 1.58); conditional variance equation with GT (88.99, 0.24, 24.05, −36.47, 1.65). The critical value of χ² (20) distribution at 5% significance level is 31.41.

Source: own work.
Table 3 enumerates the estimation outcome from the GJR-GARCH-M model as specified above. The volatility coefficients $\pi_1$ and $\pi_3$ that measure ARCH and GARCH effects in the conditional variance equation are highly statistically significant at 1 percent significance level, when the lagged volume of Google search queries is excluded from the conditional variance equation. Coefficient $\pi_2$ which measures the leverage effect in excess Bitcoin return is negative and statistically significant at 1 percent significance level. This finding suggests that negative shocks have a greater impact on volatility than do positive shocks of the same magnitude. The persistence of negative shocks in the excess Bitcoin return indicates that investors are more reactive to negative news than to positive news. Bollerslev et al. [2011] show that, the leverage effect is more pervasive when investors have more diversifiable stocks in the market. In respect of the cryptocurrency market, Liu [2018] and Brauneis and Mestel [2019] specifically identify diversification potentials among cryptocurrencies. The total volatility persistence as measured by the sum of ARCH, GARCH and leverage coefficient is 0.761 – which suggests strong volatility persistence in the excess Bitcoin return.

After the lagged volume of Google search queries is included in the conditional variance equation as in (6), all volatility coefficients (including leverage coefficient) become highly statistically insignificant. More importantly, the total volatility persistence reduces substantially (from 0.761 to 0.163), confirming that the volume of Google search queries reflects an uneven but persistent flow information arrival at the Bitcoin market. Therefore it can be said that the ARCH properly captures this time series property associated with the volume of Google search queries, as it shows significant explanatory power regarding the variance of excess Bitcoin return. Moreover, the volume of Google search queries is negatively and significantly related to conditional volatility, which indicates that the online information arrival reduces the excessive risk (i.e. risk premium) in the Bitcoin market. Therefore the online information arrival is vital for the stabilization of the Bitcoin market because it may provide diversification possibilities among cryptocurrencies or other speculative assets due to a reduction of excessive volatility.

Although the residuals of the two GJR-GARCH-M models (with and without proxy) are non-normally distributed, they are serially uncorrelated as the null hypothesis of the ARCH-LM test and the Ljung-Box Q test is clearly accepted. The log-likelihood ratios indicate that the goodness-of-fit of the two models is very high. Residuals are, however, subject to a small amount of positive autocorrelation as the test statistic falls between the range of 0 to 2.

4. Conclusion

The Internet has become an integral part of everyday life and it is impossible to imagine a day without access to it. People need information for various economic activities (e.g. consumption, investment etc.) and Internet provides this information
much faster than other modes of access. The volume of queries being processed by
search engines may therefore contain much valuable information about the
disequilibrium dynamics of various speculative asset prices (e.g. stocks [Zhang et al.
2014; Shen et al. 2016; Shen et al. 2018]). This paper provides the first evidence for
heteroskedasticity in excess Bitcoin return data using Google search volume as the
mixing variable for the information arrival time.

The results show that, when the lagged volume of Google search queries
is included in the conditional variance equation, ARCH and GARCH effects vanish
and the total volatility persistence becomes negligible as hypothesized. These
findings suggest that the volume of Google search queries reflects an uneven but
persistent flow information arrival at the Bitcoin market and is shown to have
significant explanatory power regarding the variance of excess Bitcoin return.

The cryptocurrency operation has failed to achieve the mainstream adoption as
the medium of exchange due to high volatility and lack of scalability. A significant
negative relationship between conditional volatility and volume of Google search
queries indicates that Internet (i.e. online) information arrival reduces the excessive
risk (i.e. risk premium or volatility) in the Bitcoin market, which may improve
market stability and provide diversification possibilities (among cryptocurrencies
or other speculative assets). As such, the policymakers must promote the availability
of timely and transparent online information pertaining to cryptocurrency exchange.

Acknowledgement

The authors would like to thank the Editor-in-Chief, Professor Jacek Uchman and
the Managing Editor, Professor Ilona Falat-Kilińska. The authors would also like
to thank the members of the editorial team for their friendly editorial assistance.
This paper has enormously benefited from the helpful and constructive comments
and suggestions of two anonymous reviewers. Any remaining errors are the authors’
responsibility.

References

Baur D.G., Hong K., Lee A.D., 2018, Bitcoin: Medium of exchange or speculative assets?, Journal
Bollerslev T., Sizova N., Tauchen G., 2011, Volatility in equilibrium: Asymmetries and dynamic depen-
dencies, Review of Finance, 16(1), pp. 31-80.
letin, 36(2), pp. 843-850.
Brauneis A., Mestel R., 2019, Cryptocurrency-portfolios in a mean-variance framework, Finance
Chan W.H., Le M., Wu Y.W., 2019, Holding Bitcoin longer: The dynamic hedging abilities of Bitcoin,


HETEROSKEDASTYCZNOŚĆ PUNKTÓW ZWROTNYCH BITCOINÓW: TREND GOOGLE VS. EFEKTY GARCH

Streszczenie: W artykule przeanalizowano właściwości dystrybucji związanych z heteroskedastycznymi punktami zwrotnymi nadwyżki bitcoinów, wykorzystując zapytania w wyszukiwarce Google jako przybliżenie czasu dostarczenia informacji w okresach miesięcznych od czerwca 2010 r. do maja 2019 r. W analizowanym okresie badanie pokazuje, że współczynniki zmienności stają się wysoce nieistotne statystycznie, gdy opóźniona liczebność zapytań w Google jest uwzględniona w równaniu wariancji warunkowej modelu GJR-GARCH-M. To sugeruje, iż liczba zapytań zapewnia znaczną moc wyjaśniającą w odniesieniu do wariantu heteroskedastycznego zwrotu nadmiaru bitcoinów, który można prześledzić na podstawie procesu ARCH zdefiniowanego w specyfikacji GJR-GARCH-M. Znaczny negatywny związek między zmiennością warunkową a liczbą zapytań wyszukiwania wskazuje, że pojawienie się informacji w Internecie (online) zmniejsza premię za ryzyko na rynku bitcoinów, co może poprawić stabilność rynku.

Słowa kluczowe: bitcoin, przepływ informacji, GARCH-in-Mean, efekty GARCH, trend Google.