University of Pretoria
Department of Economics Working Paper Series

Predicting Bitcoin Returns: Comparing the Roles of Newspaper- and Internet Search-Based Measures of Uncertainty

Elie Bouri
Holy Spirit University of Kaslik
Rangan Gupta
University of Pretoria
Working Paper: 2019-55
July 2019

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413

Predicting Bitcoin Returns:
Comparing the Roles of Newspaper- and Internet Search-Based Measures of Uncertainty

Elie Bouri [*] and Rangan Gupta [**]

Abstract

We compare the ability of two measures of uncertainty, a newspaper-based measure and an internet search-based measure, to predict Bitcoin returns. Using monthly data from July 2010 to May 2019 and a predictive regression model characterized by a heteroskedastic error structure and, we show that Bitcoin is a hedge against both measures. However, the predictive content of the internet-derived uncertainty related queries measure is statistically stronger than the measure of uncertainty based on newspapers for predicting Bitcoin returns, which is possibly due to the fact that the measure of uncertainty is now directly obtained from individual investors via internet searches.

Keywords: Bitcoin; Hedging; Predictability; Economic Uncertainty

JEL Codes: C32, G12

[*] Corresponding author. USEK Business School, Holy Spirit University of Kaslik, Jounieh, Lebanon. Email: eliebouri@usek.edu.lb.
[**] Department of Economics, University of Pretoria, Pretoria, South Africa. Email: rangan.gupta@up.ac.za.

1. Introduction

Over the past three years, there has been a tremendous growth in research into the role of Bitcoin as a hedge against macroeconomic and financial uncertainties (see Bouri et al. (2017, 2018), Aysan et al. (2019) and Fang et al. (2019) for detailed reviews of this literature). Demir et al. (2018) show that increases in the newspaper-based measure of economic policy uncertainty (EPU) of the United States (US) developed by Baker et al. (2016) tend to predict higher Bitcoin returns.[1] Notably, the EPU index is based on search results from 10 large newspapers (USA Today, the Miami Herald, the Chicago Tribune, the Washington Post, the Los Angeles Times, the Boston Globe, the San Francisco Chronicle, the Dallas Morning News, the Houston Chronicle and the Wall Street Journal) for terms related to economic and policy uncertainty. Baker et al. (2016) search for articles containing the terms 'uncertainty' or 'uncertain', 'economic' or 'economy' and one or more of 'congress', 'legislation', 'White House', 'regulation', 'Federal Reserve' or 'deficit'.

In this paper, we add to the related literature by hypothesizing that when we replace the frequency of newspaper articles that contain specific terms related to economic and policy uncertainty (as in the EPU index of Baker et al. (2016)) with the intensity of individual searches on the internet for similar words as a measure of uncertainty (as in the newly developed uncertainty measure of Bontempi et al. (2019) described below), the latter approach is likely to have relatively stronger predictive content (hedging impact) for Bitcoin returns. This is because, an index that measures the volume of internet searches for uncertainty-related topics, involves a shift in focus from the channel through which the message is conveyed (i.e. newspapers) to the receivers of the message (i.e. individual investors). If indeed our hypothesis is validated, then relying on the information from the EPU is likely to lead to future underprediction of Bitcoin returns, and hence, inaccurate hedging strategies.

To aid us in our objective, we use the Economic Uncertainty Related Queries (EURQ) index developed by Bontempi et al. (2019), which measures volumes of "economic uncertainty related queries" and thus reflects the intensity of individual searches of the internet for specific terms related to economic and financial uncertainty, and compare its predictive impact on Bitcoin returns with that of EPU. Methodologically, we use a predictive regression model characterized by an exponential generalized autoregressive conditional heteroskedasticity (EGARCH)-based error structure. The heteroskedastic model not only controls for the well-known volatility in Bitcoin

[1] Wang et al., (2018) analyse risk spillovers from EPU to Bitcoin, and find negligible impact to suggest that Bitcoin can act as a safe-haven or a diversifier under EPU shocks.

returns, but also controls for possible biases due to omitted variables, which, in turn, are strictly related to heteroskedasticity effects (Caporin, et al., 2018). To the best of our knowledge, this is the first paper to compare the predictive impact of EPU and EURQ in the US on Bitcoin returns. The remainder of the paper is organized as follows: Section 2 discusses the data and methodology; Section 3 presents the results; and Section 4 concludes.

2. Data and Methodology

Our main variable of interest is Bitcoin return, defined as logarithmic-returns ($r_t = \ln(p_t)-\ln(p_{t-1})$, where p_t denotes Bitcoin price in US dollars. The corresponding data is obtained from CryptoCompare at: https://www.cryptocompare.com. Figure A1 in the Appendix plots the Bitcoin return, while Table A1 provides summary statistics for the same. Bitcoin return is found to have positive skewness and excess kurtosis, resulting in a non-normal distribution as indicated by the overwhelming rejection of the null of normality under the Jarque-Bera test. Data for EPU and EURQ are available for download from the following links: http://policyuncertainty.com/us_monthly.html and http://policyuncertainty.com/EURQ_monthly.html. Our sample of analysis covers the monthly period from July 2010 to May 2019 (i.e. 107 observations), with the start date determined by the availability of Bitcoin price data, and the end date by the two measures of uncertainty. The natural logarithms of EPU (LEPU) and EURQ (LEURQ) are plotted in Figure A1 and summarized in Table A1. EPU has a lower mean but higher volatility than EURQ. Neither of the uncertainty measures are non-normally distributed based on the Jarque-Bera test. Since we want to compare the relative strengths of EPU and EURQ, we standardize them to have a unit variance when estimating the EGARCH model.[2]

To relate Bitcoin returns to the EPU and EURQ of the US, we use an exponential GARCH (EGARCH) model (Nelson, 1991). Notably, the choice of the EGARCH model over the family of other GARCH models is based on the ability of the former to better fit the data, in terms of standard goodness-of-fit measures.[3] This, in turn, is possibly a reflection of the impact of negative price movements on future volatility being different from that of positive price movements.

[2] We do not make any further transformations to the natural logarithms of EPU and EURQ, as both uncertainty measures are stationary based on standard unit root tests. The results of these tests are available from the authors upon request.

[3] Complete details of the estimations of various symmetric and asymmetric GARCH models are available upon request from the authors.

Formally, the EGARCH model used in this paper is described by assuming that the return process of Bitcoin (r_t) is given by:

$$r_t = \mu + \theta_1 LEPU_{t-1}^{std} + \theta_2 LEURQ_{t-1}^{std} + \sigma_t \varepsilon_t \tag{1}$$

where, ε_t is a sequence of $N(0, 1)$ *i.i.d.* random variables, and

$$\ln(\sigma_t^2) = \alpha_0 + \frac{\alpha_1 a_{t-1} + \gamma \mid a_{t-1} \mid}{\sigma_{t-1}} + \beta \ln(\sigma_{t-1}^2) \tag{2}$$

where $a_t = \sigma_t \varepsilon_t$. Notice that equation (2) allows us to capture an asymmetric effect between positive and negative returns. Also, to avoid the possibility of a negative variance, the model is an $AR(1)$ on $\ln(\sigma_t^2)$ rather than σ_t^2. If Bitcoin indeed does serve as a hedge to the two measures of uncertainty, we would expect both θ_1 and θ_2 to be positive in a statistically significant manner in equation (1). Our hypothesis that standardized LEURQ ($LEURQ^{std}$) provides a stronger predictive impact than standardized LEPU ($LEPU^{std}$), which would require us to have $\theta_2 > \theta_1$ in the statistical sense.

3. Empirical Results

To motivate the use of a model with heteroskedastic error structure, we first use ordinary least squares to estimate the predictive regression model: $r_t = \mu + \theta_1 LEPU_{t-1}^{std} + \theta_2 LEURQ_{t-1}^{std} + u_t$, with $u_t \sim N(0, \sigma^2)$, and perform diagnostic tests of serial correlation and heteroskedasticity on the residual, i.e., u.[4] As shown in Table A2 in the Appendix of the paper, while there is no evidence of serial correlation, the null of no-heteroskedasticity cannot be rejected (at least at the 10% level of significance). These results provide strong motivation for the usage of a GARCH-based predictive regression model.[5]

Hence, we now turn next to the results from the estimation of the EGARCH model, which in turn are reported in Table 1. As can be seen from the volatility equation, γ is negative and significant, which highlights the fact that negative innovations are more destabilizing than positive innovations. Furthermore, the impact of both lagged EPU and EURQ are positive and strongly significant in the mean equation, suggesting that Bitcoin does serve as a hedge against uncertainty.

[4] Interestingly, neither of θ_1 and θ_2 were found to be significant even at the 10% level, though they were both positive (0.0174 and 0.0343, respectively). Understandably, the existence of strong heteroskedasticity, as shown in Table A2, resulted in the non-significance.

[5] The ARCH test on the residual of the EGARCH model however, showed no evidence of any further heteroskedasticity, given the F-statistic of 0.1004, with a p-value of 0.7520.

More importantly, we find that the predictability of EURQ is stronger (0.0350) than that of EPU (0.0179), with the coefficient of the former being larger than the latter by 0.0171 (i.e. almost double). In fact, the null of $\theta_2=\theta_1$ is rejected at the highest level of significance, based on Wald-type test of coefficient restriction, which has a $F(1,99)$-statistic of 130.1980, with a corresponding p-value of 0.000.[6]

Insert Table 1 around here

As an additional analysis, we conduct a forecasting exercise over the out-of-sample of January 2015 to May 2019, with an in-sample of July 2010 to December 2014 (a 50% split of the whole sample as suggested by Rapach et al. (2005)). Basically, we estimate the model given by equations (1) and (2), by first setting $\theta_2 = 0$, and then setting $\theta_1 = 0$, and produce recursive one-step-ahead forecasts over the out-of-sample period. Unreported results show that the root mean square errors (RMSEs) for Bitcoin returns produced under the first case, i.e. with information based only on EPU, are slightly higher (at 0.2119) than under the second case (at 0.2116), i.e. when the model uses information based only on EURQ.

In summary, EURQ is found to be a relatively more important predictor of Bitcoin returns than EPU (both in- and out-of-sample)[7], which adds to prior findings that limit their analyses to a news-based measure of uncertainty (e.g., Demir et al., 2018).

[6] We estimate equation (1) with contemporaneous values of EPU and EURQ, and find $\theta_1 = 0.0470$ and $\theta_2 = 0.0734$, with both being statistically significant at the 1% level, and also with $\theta_1 < \theta_2$ in a statistical fashion, given the $F(1,100)$-statistic being 525.3594 and a p-value of 0.0000. Note the contemporaneous responses of Bitcoin returns to EPU and EURQ are stronger than the lagged responses. In addition, following Bouri et al. (2017) and Aysan et al. (2019), we include the lagged Chicago Board Options Exchange (CBOE) Volatility Index (VIX) and the geopolitical risks (GPR) index of Caldara and Iacoviello (2018) respectively in equation (1), along with the lagged EPU and EURQ. The VIX data comes from the FRED database, while the GPR data is downloaded from: https://www2.bc.edu/matteo-iacoviello/gpr.htm. Interestingly, our basic result of the stronger hedging ability of EURQ relative to EPU continues to hold, with $\theta_1 = 0.0201$ and $\theta_2 = 0.0308$, with both being statistically significant at the 1% level, and also with $\theta_1 < \theta_2$ in a statistical fashion, given the $F(1,97)$-statistic being 3.6940 and a p-value of 0.0575. Complete details of these results are available upon request from the authors.

[7] We also estimate EGARCH models for gold returns (with gold prices derived from the FRED database of the Federal Reserve Bank of St. Louis) with EPU and EURQ as predictors over the monthly period from January 2004 (which corresponds to the starting date of the EURQ index) to May 2019. Interestingly, the impact of the two measures of lagged uncertainties is positive but not significant, even at the 10% level, but the contemporaneous effect is positive and significant at the 5% level. The coefficient of EPU is found to be 0.0060 (p-value = 0.0350) and that of EURQ 0.0061 (p-value = 0.0250). However, these two coefficients are not statistically significant even at the 10% level. In other words, unlike Bitcoin, the impact of the news-based measure of uncertainty and internet-based search queries on uncertainty are similar for gold's hedging ability. Complete details of these results are available upon request from the authors.

4. Conclusion

In this paper, we analyse the predictive ability of two alternative measures of uncertainties for predicting Bitcoin returns. While the first is a news-based measure, the second is obtained from internet searches of uncertainty related queries. We postulate that the latter index is likely to have a stronger positive impact on Bitcoin returns, as it involves a shift in focus from newspapers, i.e., the channel through which the message is conveyed to individual investors who receive the message. When we test this hypothesis using a predictive regression model accounting for heteroskedasticity, we find that our hypothesis is indeed validated by both in-sample and out-of-sample analyses. This finding can be explained by the fact most of investors in the Bitcoin market are individual and inexperienced investors (Bouri et al., 2019), who often make investment decisions based on the information-content of search engines (Kristoufek, 2013). Our results imply that using the intensity of individual searches on the internet of words aiming to measure uncertainty, investors should be able to design better hedging strategies associated with Bitcoin in their portfolios, compared to a metric of uncertainty based on newspaper articles. As part of future research, it would be interesting to extend our analysis to other cryptocurrencies, and check if our results continue to hold.

References

Aysan, A.F., Demir, B., Gozgor, G., Lau, C.K.M. (2019). Effects of the geopolitical risks on Bitcoin returns and volatility. Research in International Business and Finance, 47, 511-518.

Baker, S.R., Bloom, N., Davis, S.J. (2016). Measuring economic policy uncertainty. The Quarterly Journal of Economics, 131(4), 1593-1636.

Bontempi, M.E., Frigeri. M., Golinelli., R., Squadrani, M. (2019). Uncertainty, Perception and Internet. University of Bologna, Mimeo. Available for download from: https://www.dropbox.com/s/w4qhwk303hufueu/BGS_Interest.pdf?dl=0.

Bouri, E., Gupta, R., Lau, C.K.M., Roubaud, D., Wang, S. (2018). Bitcoin and global financial stress: A copula-based approach to dependence and causality in the quantiles. The Quarterly Review of Economics and Finance, 69, 297-307

Bouri, E., Gupta, R., Roubaud, D. (2019). Herding behaviour in cryptocurrencies. Finance Research Letters, 29, 216-221.

Bouri, E., Gupta, R., Tiwari, A.K., Roubaud, D. (2017). Does Bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. Finance Research Letters, 23, 87-95.

Caldara, D., Iacoviello, M. (2018). Measuring Geopolitical Risk. Board of Governors of the Federal Reserve System, International Finance Discussion Paper No. 1222.

Caporin, M., Pelizzon, L., Ravazzolo, F., Rigobon, R. (2018). Measuring sovereign contagion in Europe. Journal of Financial Stability, 34, 150–181.

Demir, E., Gozgor, G., Lau, C.K.M., Vigne, S.A. (2018). Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation. Finance Research Letters, 26, 145-149.

Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of U.K. inflation. Econometrica, 50, 987-1008.

Fang, L., Bouri, E., Gupta, R., Roubaud, D. (2018). Does global economic uncertainty matter for the volatility and hedging effectiveness of Bitcoin? International Review of Financial Analysis, 61, 39-36.

Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. Scientific reports, 3, 3415.

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach. Econometrica, 59, 347-370.

Rapach D.E., Wohar, M.E., Rangvid, J. (2005). Macro variables and international stock return predictability. International Journal of Forecasting, 21(1), 137–166.

Wang, G-J., Xie, C., Wen, D., and Zhao, L. (2018). When Bitcoin meets economic policy uncertainty (EPU): Measuring risk spillover effect from EPU to Bitcoin. Finance Research Letters. DOI: https://doi.org/10.1016/j.frl.2018.12.028.

Table 1. Estimation Results

Parameter	Coefficient	Standard Error	z-Statistic	p-value
Mean Equation				
μ	-1.9889	0.0002	-8905.3980	0.0000
θ_1	0.0179	0.0013	13.6384	0.0000
θ_2	0.0350	1.70E-05	2058.7590	0.0000
Volatility Equation				
α_0	0.0568	0.0914	0.6215	0.5343
γ	-0.3164	0.1015	-3.1187	0.0018
α_1	0.2151	0.0759	2.8343	0.0046
β	0.9297	0.0005	1868.9570	0.0000
Log-Likelihood	-9.4590			
AIC	0.3105			

Note: The mean and volatility equations of the model are respectively: $r_t = \mu + \theta_1 LEPU_{t-1}^{std} + \theta_2 LEURQ_{t-1}^{std} + \sigma_t \varepsilon_t$, and $\ln(\sigma_t^2) = \alpha_0 + \frac{\alpha_1 a_{t-1} + \gamma \mid a_{t-1} \mid}{\sigma_{t-1}} + \beta \ln(\sigma_{t-1}^2)$. AIC: Akaike information criterion.

APPENDIX:

Table A1. Summary Statistics

Statistic	Bitcoin Log-Returns (*r*)	LEPU	LEURQ
Mean	0.1122	4.8995	5.1834
Median	0.0721	4.9008	5.1762
Maximum	1.7421	5.6495	5.5373
Minimum	-0.4921	4.1570	4.9836
Std. Dev.	0.3616	0.3027	0.1014
Skewness	1.6295	0.2344	0.3985
Kurtosis	7.8901	2.5481	3.2731
Jarque-Bera	153.9635 (0.0000)	1.8904 (0.3886)	3.1640 (0.2056)
Observations	107		

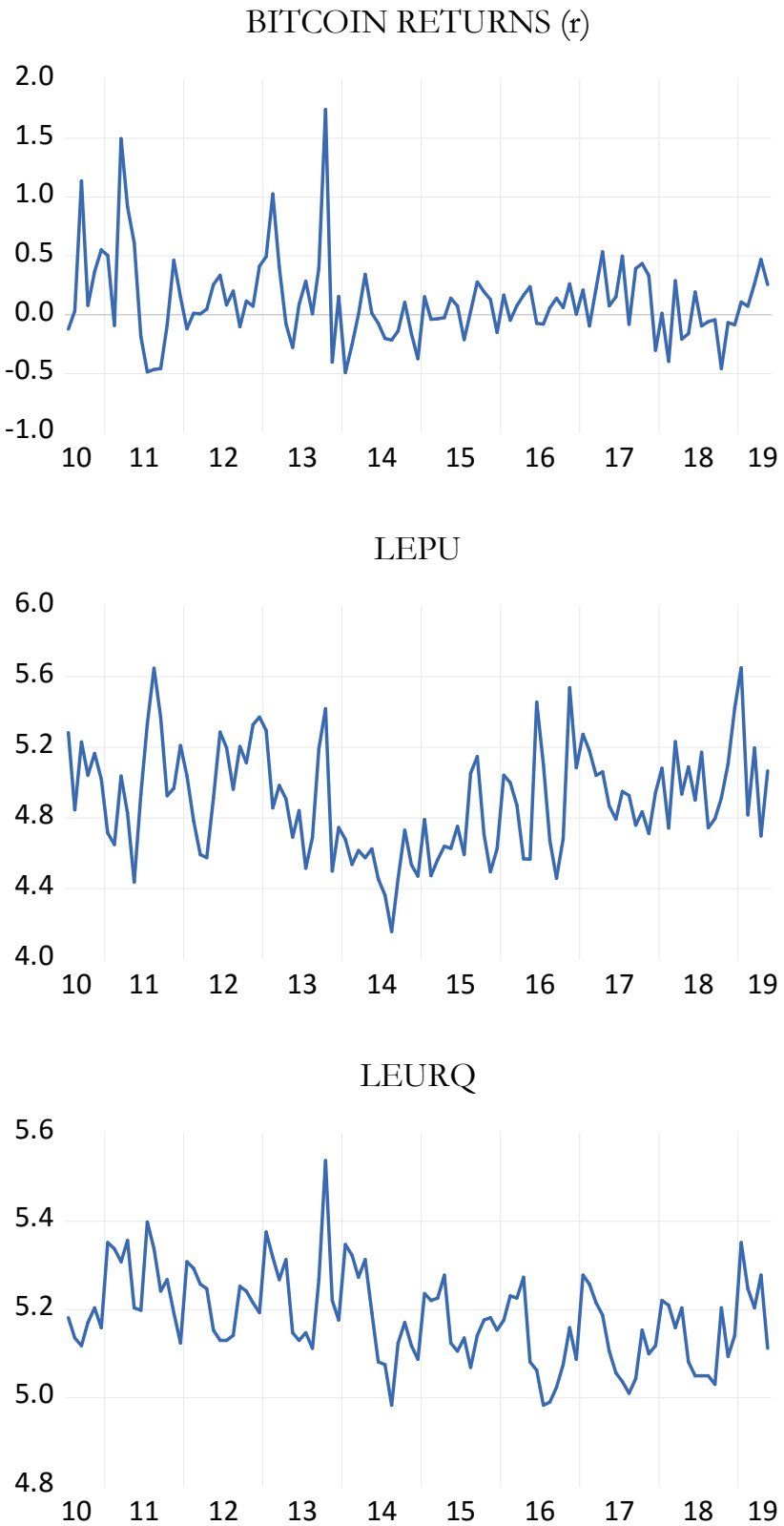
Note: LEPU and LEURQ are the natural logarithms of the uncertainty indices of Baker et al. (2016) and Bontempi et al. (2019) derived from newspapers and internet search queries, respectively; Jarque-Bera test statistic corresponds to a test of the null hypothesis of normality.

Table A2. Residual Diagnostic Tests of the Ordinary Least Squares Estimation of the Predictive Regression Model

Serial Correlation Test	*F*-statistic
Breusch-Godfrey	0.6121 (0.5442)
Heteroskedasticity Tests	*F*-statistic
Breusch-Pagan-Godfrey	4.1728 (0.0181)
Harvey	5.0155 (0.0083)
Glejser	6.6215 (0.0020)
ARCH	3.5178 (0.0635)
White	2.1446 (0.0662)

Note: Tests performed on the residual of: $r_t = \mu + \theta_1 LEPU_{t-1}^{std} + \theta_2 LEURQ_{t-1}^{std} + u_t, u_t \sim N(0, \sigma^2)$; Null hypothesis of the tests are no-serial correlation and no-heteroskedasticity; Entries in parentheses correspond to the *p*-values of the various test statistics.

Figure A1. Data Plots



Note: See Notes to Table A1.