

Bitcoin Price Predictive Modeling Using Expert Correction

Bohdan M. Pavlyshenko

Ivan Franko National University of Lviv, Ukraine
b.pavlyshenko@gmail.com, www.linkedin.com/in/bpavlyshenko/

January 11, 2022

Abstract

The paper studies the linear model for Bitcoin price which includes regression features based on Bitcoin currency statistics, mining processes, Google search trends, Wikipedia pages visits. The pattern of deviation of regression model prediction from real prices is simpler comparing to price time series. It is assumed that this pattern can be predicted by an experienced expert. In such a way, using the combination of the regression model and expert correction, one can receive better results than with either regression model or expert opinion only. It is shown that Bayesian approach makes it possible to utilize the probabilistic approach using distributions with fat tails and take into account the outliers in Bitcoin price time series.

1 Introduction

One of the main goals in the Bitcoin analytics is price forecasting. There are many factors which influence the price dynamics. The most important factors are: interaction between supply and demand, attractiveness for investors, financial and macroeconomic indicators, technical indicators such as difficulty, the number of blocks created recently, etc. A very important impact on the cryptocurrency price has trends in social networks and search engines. Using these factors, one can create a regression model with good fitting of bitcoin price on the historical data. Paper [1] shows that views of Wikipedia Bitcoin related pages correlate with Bitcoin price movements. It reflects the potential investors' interests in cryptocurrency. Google trends of search Bitcoin related keywords show different effects, including investors' interests, speculators' activities, etc. In [2], Bitcoin price was analyzed. The paper [3] studies different drivers of Bitcoin price. In [4], a significant correlation between Bitcoin price and social and web search media trends was shown. In [5], the analysis shows that bitcoin has many similarities to both gold and the dollar. In [6], the use of Bayesian regression for Bitcoin price analytics was studied. The impact of news on investors' behavior is studied in [7]. Bitcoin economy, behavior and mining are considered in [8, 9]. Bitcoin market behaviour, especially price dynamics is the subject of different studies ([10, 1]). Different factors affecting bitcoin price are analyzed in the [10]. The specific feature of Bitcoin is that this cryptocurrency is neither issued nor controlled by financial or political institutions such as Central Bank, government, etc. Bitcoin is being mined without economic underlying factors. In [10], the economics of BitCoin Price Formation has been analyzed.

Price dynamics is mostly affected by speculative behaviour of investors. One of the main bitcoin drivers is the news in the Internet. According to efficient market theory, the stock and financial markets are not predictable, since all available information is already reflected in stock prices. But nowadays the dominance of efficient market theory is not so obvious. Some influential scientists argue that market can be partially predictable ([11]). Behind modern market prediction, there are behavioral and psychological theories. Some economists believe that historical prices, news, social network activities contain patterns that make it possible to partially predict financial market. Such theories and approaches are considered in the survey [11].

In this paper, we consider an approach for building regression predictive model for bitcoin price using expert correction by adding a correction term. It is assumed that an experienced expert can make model correction relying on his or her experience.

2 Bitcoin Price Modeling

Let us consider a regression model for Bitcoin price. As the regressors in our model, we used historical data which describe Bitcoin currency statistics, mining processes, Google search trends, Wikipedia pages visits. As the currency statistics, we took *total_bitcoins* - the total number of bitcoins that have already been mined, *price* - average USD market price across major bitcoin exchanges, *volume* -the total USD value of trading volume on major bitcoin exchanges. As mining information, we took *difficulty*, which is a relative measure of how difficult it is to find a new block. As network activity, we took *n_unique_addresses* which is the total number of unique addresses used on the Bitcoin blockchain. Time series for mentioned above variables was taken from Bitcoin.info site. As the factors of price formation, we also considered Google trend for 'bitcoin' keyword and number of visits of Wikipedia page about 'cryptocurrency'. Time series for chosen features are shown on the Fig. 1 . A target variable *price* and all the regressors are considered in the logarithmic scale, so $price' = \ln(price + 1)$, $x'_i = \ln(x_i + 1)$ After logarithmic transformation all the regressors were normalized by extracting mean values and then dividing by standard deviation. We can write the regression model as:

$$\begin{aligned}
 price' = & \alpha + \beta_{gtrend} \cdot gtrend' + \\
 & \beta_{wiki_cryptocurrency} \cdot wiki_cryptocurrency' + \\
 & \beta_{difficulty} \cdot difficulty' + \\
 & \beta_{n_unique_addresses} \cdot n_unique_addresses' + \\
 & \beta_{total_bitcoins} \cdot total_bitcoins' + \\
 & \beta_{volume} \cdot volume'
 \end{aligned} \tag{1}$$

For the numerical modeling, we used Python with the packages *pandas*, *numpy*, *scipy*, *matplotlib*, *seaborn*, *sklearn*, *quandl*, *pystan*. To get time series of Wikipedia pages visits, we used Python package *mwviews*. To find coefficients α, β_i , we used linear regression with Lasso regularization from python package *scikit-learn*. Fig. 2 shows real dynamics of Bitcoin price and the price predicted by regression model (1). Fig. 3 shows obtained linear regression coefficients for features. The results show the importance of Google trends for Bitcoin keyword searches, and the importance of views of Wikipedia pages about the cryptocurrency. For this analyzed model, we received the error value RMSE=1277.8.

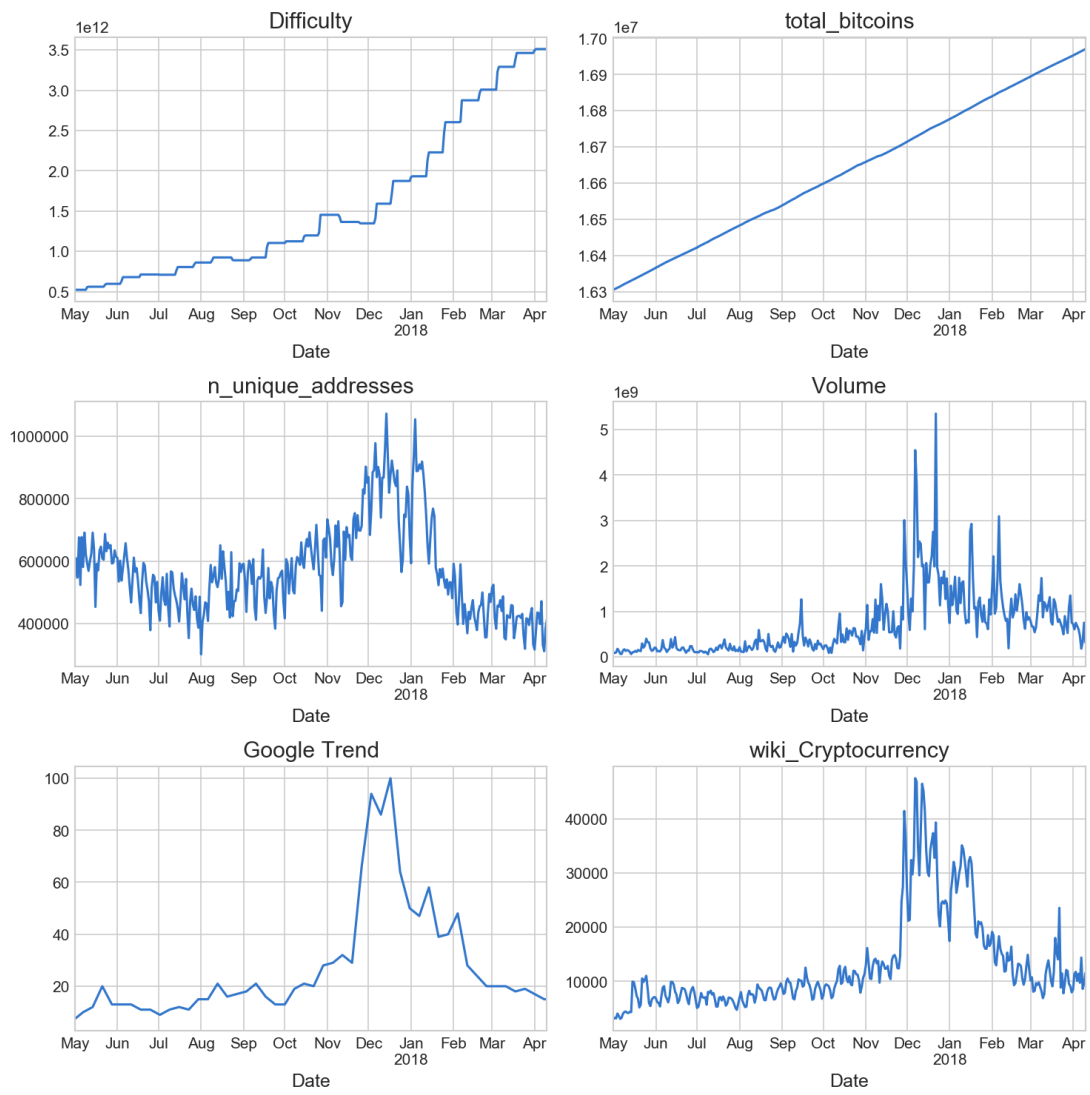


Figure 1: Time series of features



Figure 2: Bitcoin price dynamics and prediction

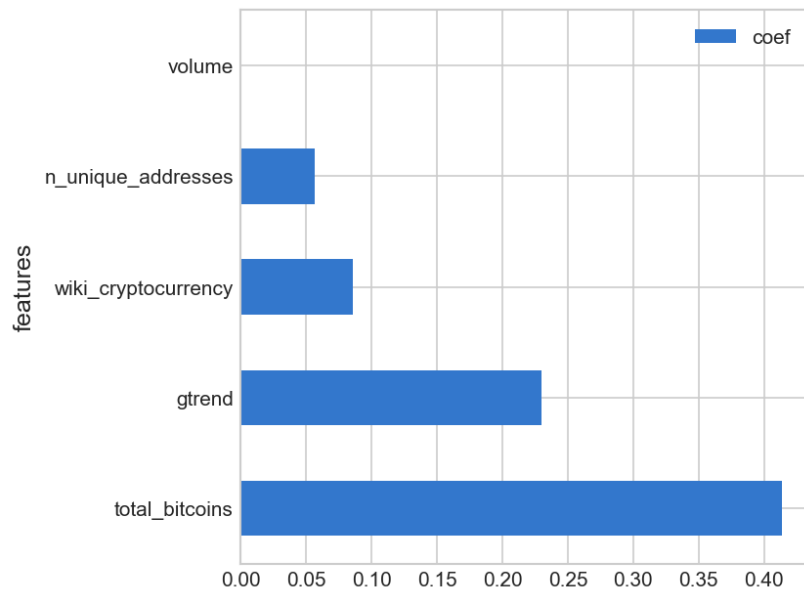


Figure 3: Features regression coefficients

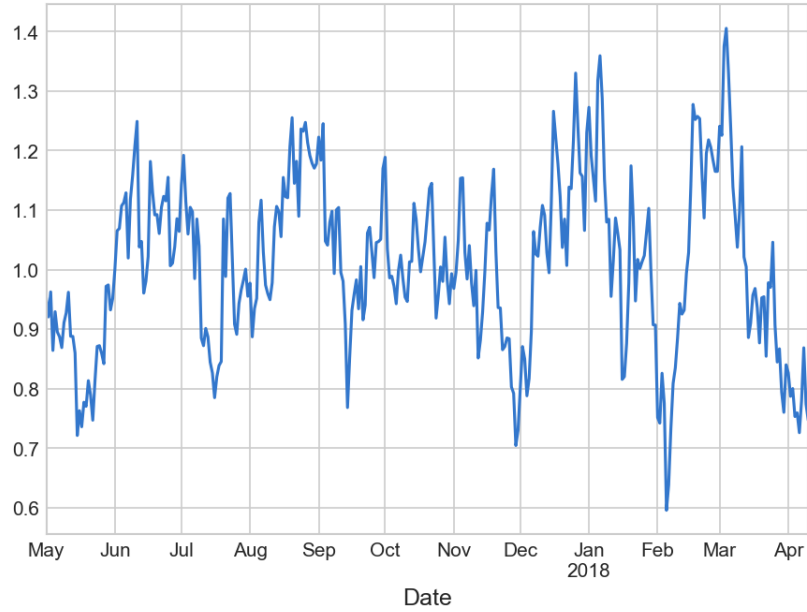


Figure 4: The ratio of real Bitcoin price to predicted price

3 Modeling of expert correction

Fig. 2 shows that in some time periods, predicted price is higher comparing to real price, in some time periods it is lower. Fig. 4 shows the ratio of real and predicted price. We can see that this ratio has periodic oscillations which describe the deviation of predictions from real values. It can be explained by existence of some factors which impact price but they are not included into the model (1). It can be the factors of complex behavior of investors. Suppose that some experienced expert understands such type of behavior. As a result, he or she can explain the deviation dynamics of regression model (1) forecasting from real values. Expert correction can be involved into the model by an additional term in the regression model (1). This term describes the dynamics of model deviation. To describe this term, the expert needs to define the local extremum on deviation time series which are pivot points for deviation trends. We suppose that the expert can define such points correctly relying on his or her personal experience. The time series for possible expert correction is shown on Fig. 5. The results of the calculation of regression model (1) with added such expert correction term are shown on the Fig. 6. For this case, we received the error value $RMSE=856.4$. Fig. 7 shows the regression coefficient in case of expert correction term included into regression model. Obtained results show that adding the expert term improves regression model performance.

4 Bayesian regression model

For probabilistic approach, which makes it possible to get risk assessments, one can use Bayesian inference approach. Bayesian regression is a method which gives advantages when we need to take into account non Gaussian statistics for the variables under study ([12]). For Bayesian modeling, we used Stan software with *pystan* python package. To do the model robust to outliers, we can consider Bitcoin price distributed with

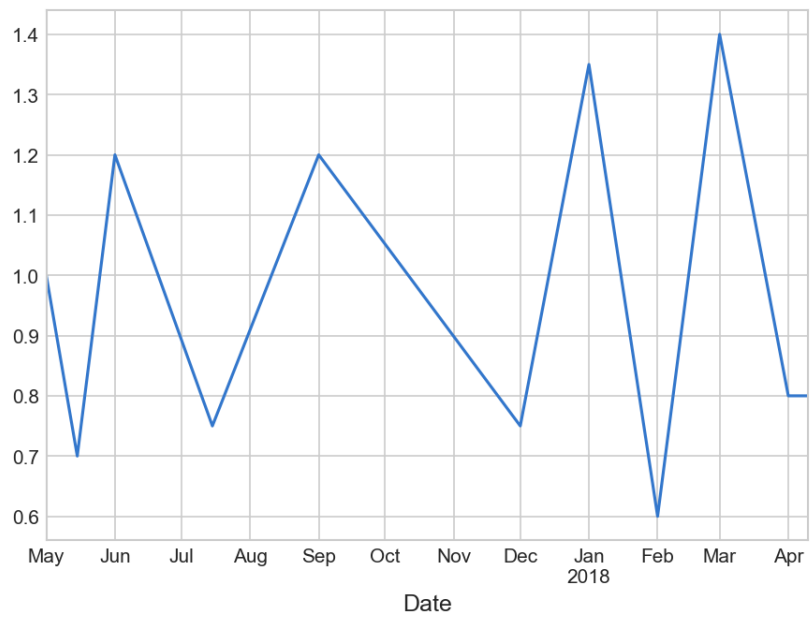


Figure 5: Expert correction term time series



Figure 6: Bitcoin price dynamics & prediction with expert correction

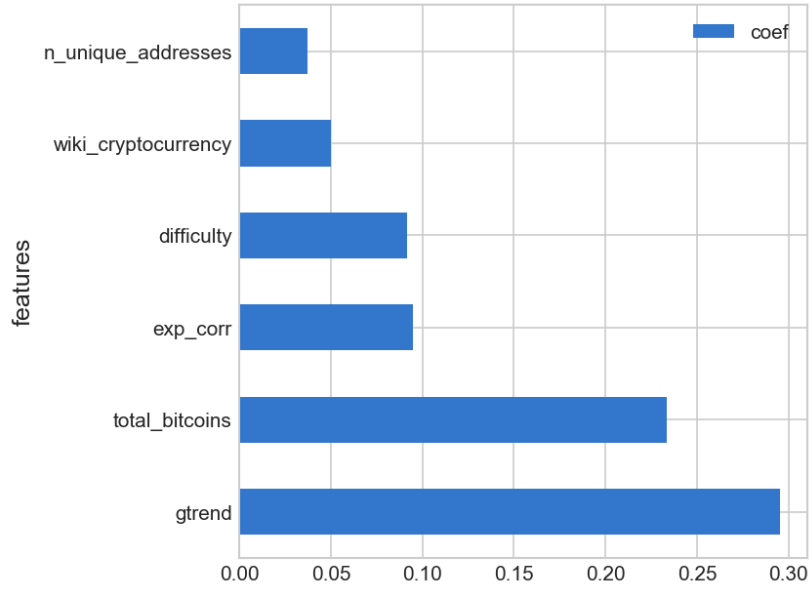


Figure 7: Feature coefficients for regression model with expert correction

Student's t-distribution. Student's t-distribution is similar to normal distribution but it has heavier tails. Let us consider that the Bitcoin price is a stochastic variable which is distributed by Student's t-distribution

$$price'_s \sim Student.t(\mu, \sigma, \nu),$$

where ν denote the degree of freedom, μ is a location parameter, σ is a scale parameter. In Bayesian approach, we consider that $\mu = price'$, where $price'$ is defined by the regression model (1). Fig. 8 shows the box plots for received parameters of regression model. Fig. 9 shows the box plots for the parameters in case of added expert correction term. For scale parameter σ , the value 0.14 was received in case of the absence of expert correction term and the value 0.1 was received in case of added expert correction term. It shows that adding expert correction term into Bayesian regression model decreases the width of the probability distribution function for the target variable.

5 Conclusion

In our study, we considered the linear model for Bitcoin price which includes regression features based on Bitcoin currency statistics, mining processes, Google search trends, Wikipedia pages visits. The pattern of deviation of regression model prediction from real prices is simpler comparing to price time series. So, we assume that this pattern can be predicted by an experienced expert. In such a way, combining the regression model and expert correction, one can receive better results than with either regression model or expert opinion only. We assume that the expert can catch the complex time series patterns that relies on financial behavioral theories, economics and politics. These patterns cannot be caught from time series historical data since they exist for a short time period. After these patterns are published or considered between investors, they soon become self-destructed. The received results show that the correct expert definition of time pivot points for the regression model deviation can essentially improve the prediction of Bitcoin price. In the proposed approach, the expert has to define time

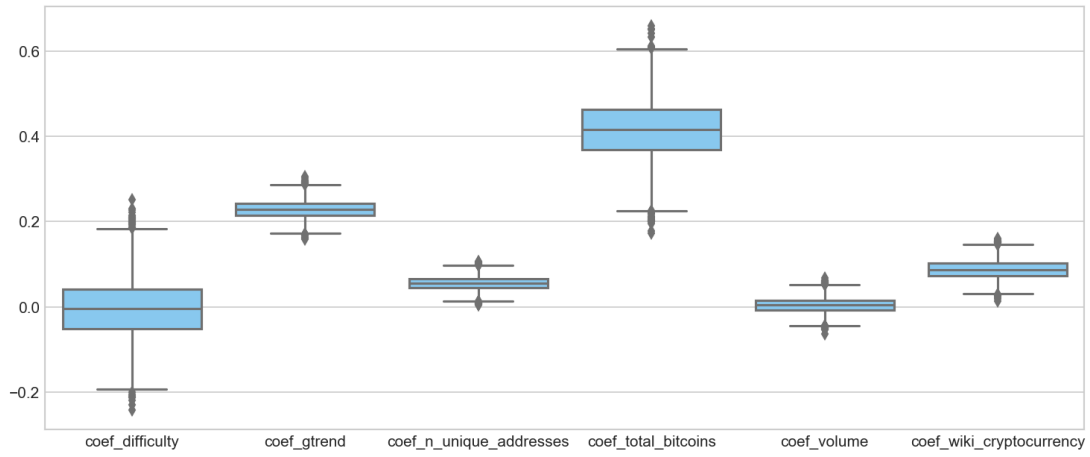


Figure 8: Boxplot for regression model coefficients

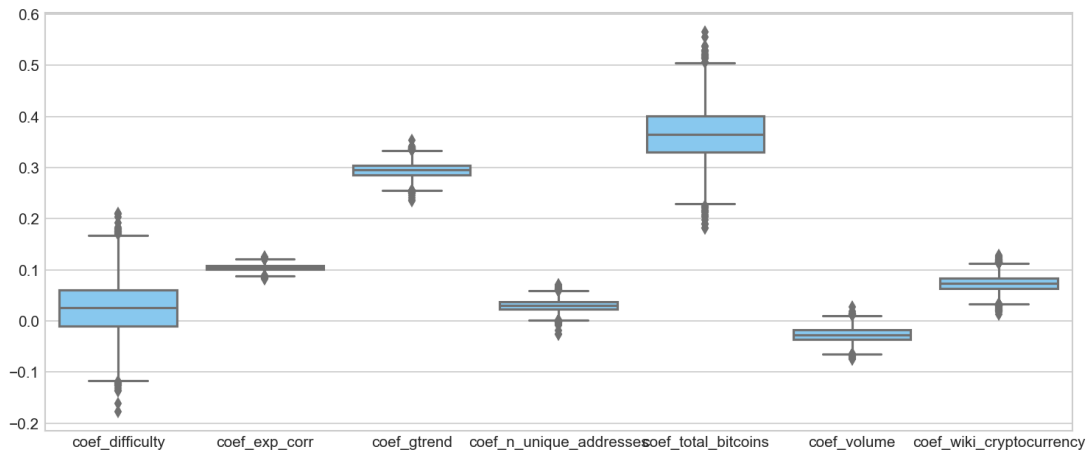


Figure 9: Boxplot for coefficients of regression model with expert correction term

pivot points which describe the deviation of regression model based on the historical data comparing to real price time series. With Bayesian inference, one can utilize the probabilistic approach using distributions with fat tails and take into account outliers in Bitcoin price time series. Having probability density distributions for price and model parameters, one can make risk assessment via calculating value at risk characteristics.

References

- [1] L. Kristoufek, “Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era,” *Scientific reports*, vol. 3, p. 3415, 2013.
- [2] J. Bouoiyour, R. Selmi, and A. K. Tiwari, “Is bitcoin business income or speculative foolery? new ideas through an improved frequency domain analysis,” *Annals of Financial Economics*, vol. 10, no. 01, p. 1550002, 2015.
- [3] J. Bouoiyour, R. Selmi, A. K. Tiwari, O. R. Olayeni, *et al.*, “What drives bitcoin price,” *Economics Bulletin*, vol. 36, no. 2, pp. 843–850, 2016.
- [4] M. Matta, I. Lunesu, and M. Marchesi, “Bitcoin spread prediction using social and web search media.,” in *UMAP Workshops*, 2015.
- [5] A. H. Dyrberg, “Bitcoin, gold and the dollar—a garch volatility analysis,” *Finance Research Letters*, vol. 16, pp. 85–92, 2016.
- [6] D. Shah and K. Zhang, “Bayesian regression and bitcoin,” in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, pp. 409–414, IEEE, 2014.
- [7] B. M. Barber and T. Odean, “All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors,” *The review of financial studies*, vol. 21, no. 2, pp. 785–818, 2007.
- [8] R. Grinberg, “Bitcoin: An innovative alternative digital currency,” *Hastings Sci. & Tech. LJ*, vol. 4, p. 159, 2012.
- [9] J. A. Kroll, I. C. Davey, and E. W. Felten, “The economics of bitcoin mining, or bitcoin in the presence of adversaries,” in *Proceedings of WEIS*, vol. 2013, p. 11, 2013.
- [10] P. Ciaian, M. Rajcaniova, and d. Kancs, “The economics of bitcoin price formation,” *Applied Economics*, vol. 48, no. 19, pp. 1799–1815, 2016.
- [11] B. G. Malkiel, “The efficient market hypothesis and its critics,” *Journal of economic perspectives*, vol. 17, no. 1, pp. 59–82, 2003.
- [12] J. Kruschke, *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press, 2014.