Article

Do Cryptocurrency Prices Camouflage Latent Economic Effects? A Bayesian Hidden Markov Approach [†]

Constandina Koki [1],[*] , Stefanos Leonardos [2] and Georgios Piliouras [2]

[1] Department of Statistics, School of Information Sciences and Technology, Athens University of Economics and Business, 10434 Athens, Greece

[2] Engineering Systems and Design, Singapore University of Technology and Design, Singapore 487372, Singapore; stefanos_leonardos@sutd.edu.sg (S.L.); georgios@sutd.edu.sg (G.P.)

[*] Correspondence: kokiconst@aueb.gr

[†] A preliminary version of this paper appeared Decentralized 2019 and awarded with the Best Paper Award.

Abstract: We study the Bitcoin and Ether price series under a financial perspective. Specifically, we use two econometric models to perform a two-layer analysis to study the correlation and prediction of Bitcoin and Ether price series with traditional assets. In the first part of this study, we model the probability of positive returns via a Bayesian logistic model. Even though the fitting performance of the logistic model is poor, we find that traditional assets can explain some of the variability of the price returns. Along with the fact that standard models fail to capture the statistic and econometric attributes—such as extreme variability and heteroskedasticity—of cryptocurrencies, this motivates us to apply a novel Non-Homogeneous Hidden Markov model to these series. In particular, we model Bitcoin and Ether prices via the non-homogeneous Pólya-Gamma Hidden Markov (NHPG) model, since it has been shown that it outperforms its counterparts in conventional financial data. The transition probabilities of the underlying hidden process are modeled via a logistic link whereas the observed series follow a mixture of normal regressions conditionally on the hidden process. Our results show that the NHPG algorithm has good in-sample performance and captures the heteroskedasticity of both series. It identifies frequent changes between the two states of the underlying Markov process. In what constitutes the most important implication of our study, we show that there exist linear correlations between the covariates and the ETH and BTC series. However, only the ETH series are affected non-linearly by a subset of the accounted covariates. Finally, we conclude that the large number of significant predictors along with the weak degree of predictability performance of the algorithm back up earlier findings that cryptocurrencies are unlike any other financial assets and predicting the cryptocurrency price series is still a challenging task. These findings can be useful to investors, policy makers, traders for portfolio allocation, risk management and trading strategies.

Keywords: cryptocurrencies; bitcoin; ethereum; bayesian modeling; logistic regression; non-homogeneous hidden markov models; variables selection; forecasting

JEL Classification: C11; C52; C53; E42; O39

1. Introduction

What are cryptocurrencies? How do they compare to traditional financial instruments? Are they like traditional money, like commodities, a hybrid of the former or an utterly new type of asset that merit their own definition and understanding? Early research, mainly focusing on Bitcoin

(henceforth BTC), provides mixed insights. While the creation of new BTCs resembles the mining process of gold—or precious metals in general—its attributes clearly differentiate it from conventional commodities [1]. The claim that BTC is fundamentally different from valuable metals like gold is also backed by Klein et al. [2] due to its shortage in stable hedging capabilities. Along with [3], Cheah and Fry [1] also argue that standard economic theories cannot explain BTC price formation and using data up to 2015, they provide evidence that BTC lacks the qualities necessary to be qualified as money. However, using GARCH models, Dyhrberg [4] demonstrates that BTC has similarities to both gold and the US dollar (USD) and somewhat surprisingly that it may be ideal for risk-averse investors. Also, while the BTC is useful to diversify financial portfolios—due to the negative correlation to the US implied volatility index (VIX)—it otherwise has limited *safe-haven* properties [5–7]. Using data from a longer period (between 2010 and 2017), Demir et al. [8] conclude the opposite, namely that BTC may indeed serve as a hedging tool, due to its relationship to the Economic Policy Uncertainty Index (EUI).

The fact that cryptocurrencies are different from any other asset in the financial market is further supported by [9–11]. High volatility, speculative forces and large dependence on social sentiment at least during its earlier stages—as measured by social media and Internet data (Google trends, Wikipedia searches and Twitter posts)—are qualified by many as some of the main determinants of BTC prices [12,13]. Yet, a large amount of price variability remains unaccounted for. Moreover, the proliferation of cryptocurrencies other than BTC that are supported by different technologies, i.e., variations of the standard Proof-of-Work distributed consensus of the BTC blockchain, e.g., [14], calls for a more comprehensive research approach. Despite the high documented correlation in the price of the various cryptocurrencies, it is highly debated whether this trend will also continue into future or not [15,16].

In the present paper, we make an effort towards understanding the correlation between a set of traditional assets and cryptocurrencies. We adopt an economic/financial perspective and use a set of 14 financial and economic predictors comprising main exchange rates (4 variables), equity indices (4 variables), commodity future prices (oil and gold) and economic uncertainty indicators (2 variables) along with 2 quasi-economic and 2 cryptocurrency specific variables: the hash rate which captures the amount of investment on mining equipment and hence accounts for the economic size of the network and the average block size which implicitly measures the amount of transactions and hence the activity in the respective cryptocurrency. All the variables and the applied transformations are summarized in Table 1. Also, we report the correlations between the explanatory variables in Table 2.

Earlier studies highlight the scarcity of results on cryptocurrencies other than BTC and underline the need for a better understanding of the entire cryptocurrency ecosystem and its properties (statistical and economic), see e.g., [11,17]. Studies that go beyond the BTC prices and confirm via various financial models the importance of using diverse cryptocurrencies—rather than a single one—in portfolio optimization include but are not limited to [18] and [15]. In view of the above, in the present study, apart from Bitcoin (BTC), we also focus on Ether (ETH), the native coin of the Ethereum blockchain [14,19], and currently the second largest cryptocurrency in terms of market capitalization [20]. Unlike the BTC blockchain, the Ethereum blockchain has been launched eponymously and is governed, or more aptly researched and developed, by the Ethereum Foundation [21], a non-profit organization-based in Switzerland. The architecture of the Ethereum ecosystem has far-reaching implications on its long-term development and sustainability that clearly differentiate it from BTC. Supporting smart contract execution—execution of code snippets that go beyond the simple monetary transactions of BTC—Ethereum has scheduled a transition from the currently computationally heavy Proof of Work (introduced by BTC and followed by most cryptocurrencies) to the computationally efficient alternative of Proof of Stake, which saves on energy resources and provides a scalable infrastructure while retaining the same security guarantees as Proof of Work. Without going further into the technical details, the main motivation to study ETH that stems from these considerations is the following. Given the different technological advancements that are promised by Ethereum, will ETH become independent from BTC and follow its own path as a cryptocurrency or are after all the values of all

cryptocurrencies inevitably tied, as they are up to now [16]? Keeping in mind that ETH—i.e., the native coin—is only one of the main applications of the ETH blockchain—and the blockchain technology in general—it should also be noted that price movements of ETH may not necessarily align in the future with technological advancements in the Ethereum blockchain.

From a methodological perspective, we perform a two-layer Bayesian analysis. First, we transform the cryptocurrency series into a binary series and apply a logistic regression model on the transformed series. Specifically, if the current price return, i.e., $Y_t - Y_{t-1}$, exceeds a predefined threshold then we assign the value 1 and 0 otherwise. Then, we investigate whether the logistic regression model—which is widely used by applied statisticians and econometricians for analyzing binary data, see [22] and references therein—with a specific covariate set is an appropriate model for estimating the probability of observing the value 1 in these binary series. We use the methodology of Polson et al. [23] to make inference on the model's parameters with an additional reversible jump step to allow for model uncertainty, cf. Section 2.1. Secondly, we model the log-price series data using a novel Hidden Markov (regime switching) model, namely the non-homogeneous Pólya Gamma Hidden Markov model (NHPG) of [24], cf. Section 2.2. Hidden Markov models introduce time-variation in the parameters through an underlying unobserved discrete process. In brief, at any given time *t*, the observed log-price data point depends on a latent (hidden) state. Hence, conditionally on the hidden states, the parameters of the data generating process vary and thus allowing for a flexible data representation. In our setting, the underlying process follows a binomial process with exogenous variables. It has been shown that the NHPG model outperforms similar models in forecasting conventional financial data, cf. [25]. Also, it uses Bayesian Model Averaging (BMA) approach for inference which has been shown to possess desirable properties for forecasting applications [26–29].

With all these in mind, the questions we aim to address are the following:

Q1. Does the underlying information from fiat currencies, commodities, stock indices and blockchain specific variables explain/predict the probability of positive returns?
Q2. Do the same variables have explanatory/predictive power on both the BTC and ETH cryptocurrencies?
Q3. Do the same explanatory variables affect the BTC price series both on the long and short run?

We use daily data (for both the response and the explanatory variables) between 2017 and 2019. For question Q3, we compare the BTC data of the whole 2014–2019 period to the 2017–2019 period (also used in Q1–Q2). As in most of the recent studies, we exclude the period up to 2014 which exhibits markedly different characteristics.

The findings of our experiments can be summarized as follows. The logistic model is not suitable to model the probabilities of positive daily returns of BTC and ETH. However, changing the magnitude of returns, we observe that (a) the logistic model has improved performance (b) the statistical significant covariates in the logistic regression model change and (c) the in-sample (fitting) results are different for the BTC and ETH series.

Considering the second experiment, we find that the NHPG model identifies periods of different volatility and accounts well for the heteroskedacity of all three price series (BTC short and long periods and ETH). Graphically, this is illustrated in later figures. The hidden states—which may be described as periods of high and low volatility—are not persistent, i.e., the transitions between the two states are frequent. Based on the same figures, the in-sample performance of the NHPG algorithm is good. However, the set of included predictors—predictors with posterior probability of inclusion above 0.5—is large, which implies that each predictor explains only a small fraction of the volatility of the series. Concerning specific predictors, the exclusion of some of the fiat currency exchange rates for the ETH series suggests a (still) more geographically restricted interest for the currency in comparison to BTC. It is also worth mentioning that the cryptocurrency specific variables, hash rate and average block size are not significant for modeling the BTC and ETH price series. This may indicate a more mature and stabilizing mining network that is less responsive to price expectations, sentiment or extreme speculation.

Finally, as shown in the last figure, the mean posterior out-of-sample predictions, although better for ETH than for BTC, are in general not good as they frequently even miss the direction of movement of the series. However, this is a common outcome in exchange rates [30,31]. In sum, our results confirm that the Hidden Markov approach is promising in the understanding of cryptocurrencies price formation and back earlier findings that cryptocurrencies are unlike any existing financial asset and hence that their understanding requires novel tools and ideas.

In the related literature, the first layer of our methodology, i.e., the logistic regression model, falls into the binary regression models literature. They have only been applied in the cryptocurrency context, by [32] to forecast the daily price direction of BTC, by [33] to study the price co-explosivity in leading cryptocurrencies and by [34] to study the herding behavior of BTC. As far as the second layer of our methodology, the present NHPG model falls into the Markov-switching literature that is the benchmark for predicting exchange rates, see [35] and [36,37] and explaining financial time series, see [38] and references therein. This class of models, account for the non-stationarities and non-linearities of the time series. Although standard in financial applications ([38]), Hidden Markov models have been applied in the cryptocurrency context by [39] as Markov-switching GARCH models to model the volatility dynamics of BTC, by [40], as a state-space model for representing the BTC price series, by [18] as multivariate state state-space models in forecasting cryptocurrencies, as homogeneous Hidden Markov, i.e., hidden Markov models with constant transition probabilities, by [41,42], and in the understanding of price bubbles by [43]. Also, [44], study the BTC and ETH prices under structural break setting while [45] study the cryptocurrency returns and volatility under stochastic volatility model with discontinuous jumps. The use of the NHPG model in explaining and predicting the BTC and ETH price series is also supported by the findings of various articles. For example the authors of [46,47] and [8], demonstrate the non-stationarity of the BTC index and volume and underline the importance of modeling non-linearity in Bitcoin prediction models. This is further elaborated by Beckmann and Schüssler [26] who suggest that model selection and the use of averaging criteria are necessary to avoid poor forecasting results. Following the similar reasoning, Phillip et al. [11] posit that standard models are inadequate to capture the extreme variability of cryptocurrencies and argue in favor of more composite approaches. In an important finding, Ciaian et al. [3] show that the Bitcoin price series exhibits structural breaks and identifies periods of data (prior to 2013 and between 2013 and 2015) of markedly different variance and other econometric characteristics. Their findings further suggest that significant price predictors may vary over time. Pichl and Kaizoji [48] use data from various time periods to demonstrate, among other results, that the BTC price series exhibits heteroskedasticity.

Finally, the present paper falls into the strand of literature that studies the explanatory and predictive power of traditional financial and economic indices on the cryptocurrency price series. To name a few, Refs. [4,49] analyze the relationship between BTC, gold and USD, Ref. [50] study the predictive power of a large set of exogenous variables, such as commodities, volatility indices, stock indices. Subsets of the studied indices are studied under various settings, see e.g., [7,18,40,47,48,51–53].

All in all, our aim is to contribute to the literature that studies the modeling and prediction of cryptocurrencies, using a novel Bayesian elaborate econometric model and to try to gain understanding in the statistical, econometric and financial properties of existing cryptocurrencies.

The rest of the paper is structured as follows. In Section 2, we describe the two econometric models of this study: the logistic regression model is described analytically in Section 2.1 and the NHPG model and simulation scheme is described in Section 2.2. The empirical study is presented in Section 3. In detail, the data set that we used is described in Section 3.1, the results regarding the logistic model are presented in Section 3.2 and lastly, the results regarding the NHPG model are presented in Section 3.3. We conclude the paper with a discussion of the limitations of the present model and directions for future work in Section 4.

This paper considerably extends its earlier conference version. Concerning the applied methodology, we provide a rigorous description of the logistic regression model for studying the

probabilities of positive returns (Section 2.1) and of the NHPG model (Section 2.2). In addition, we have updated the data set—BTC and ETH series—and the covariate set. Specifically, in the covariate set, (Table 1), we have included the Russel 2000 index, excluded the autoregressive terms and applied different transformations on the variables. More importantly, concerning the results, this paper includes the novel analysis of the logistic model (Section 3.2) and based on the new covariate set, it offers more enriched outcomes and more comprehensive insight from the analysis of the NHPG model (Section 3.3).

2. Methodology

2.1. The Logistic Regression Model

Let Y_t be the ETH or the BTC price series with realization y_t. Also, consider a set of $r-1$ available predictors $\{X_t\}$ with realization $x_t = (1, x_{1t}, \ldots, x_{r-1t})$ at time t. The explanatory variables (predictors) $\{X_t\}$ that are used in the present analysis are described in Table 1. We transform the cryptocurrency price series as a binary series, i.e., a series that takes the values 1 or 0, as follows. Let

$$U_{t,\alpha} = I\{Y_t - Y_{t-1} \geq \alpha\}, \tag{1}$$

where I denotes the indicator function that takes the value 1 if $Y_t - Y_{t-1} \geq \alpha$ and 0 otherwise and α is a predefined threshold. Intuitively, we study the connection of the predictors $\{X\}$ with the probability of having positive daily returns $\{Y_t - Y_{t-1} \leq \alpha\}$. We perform our analysis for various positive thresholds, $\alpha \in \{0, 1, \ldots, 5\%\}$.

We treat the binary series $U_{t,\alpha}$ as a Bernoulli(p_t) variable. From the class of the generalized linear models, we use a logit link to model the probabilities (p_t). The standard logistic regression model is defined as

$$U_{t,\alpha} \sim Bernoulli(p_t), \tag{2}$$

$$p_t = g^{-1}(\eta_t).$$

with $\eta_t = x_t\beta$, β the logistic regression coefficients and $g(z) = \log\frac{z}{1-z}$ the logit link function. Then, the probabilities are modeled as,

$$\log\frac{p_t}{1-p_t} = x_t\beta \Leftrightarrow p_t = \frac{\exp(x_t\beta)}{1+\exp(x_t\beta)}. \tag{3}$$

We use the recently proposed latent variable scheme, namely the Pólya-Gamma data augmentation method of [23] which has significantly improved results.

The authors introduce of [23] proved that binomial likelihoods—or Bernoulli likelihoods—parametrized by log odds can be represented as mixtures of Gaussian distributions with respect to the Pólya-Gamma distribution. Their main result is that letting $p(\omega)$ be the density of a Pólya-Gamma latent variable ω, with $\omega \sim \mathcal{PG}(b,0)$, for $b > 0$, the following identity holds for all $a \in \mathbb{R}$,

$$\frac{\exp(\psi)^a}{(1+\exp(\psi))^b} = 2^{-b}\exp(k\psi)\int_0^\infty \exp\left(-\omega\psi^2/2\right)p(\omega)\,d\omega, \tag{4}$$

with $k = a - b/2$. Furthermore, the conditional distribution of $\omega \mid \psi$ is also Pólya-Gamma, $\mathcal{PG}(b,\psi)$.

When $\psi = x\beta$, the previous identity gives rise to a conditionally conjugate augmentation scheme for Bernoulli likelihoods of logistic parameters. The likelihood is given by

$$\mathcal{L}(\beta) = \prod_{t=1}^N \left\{\frac{\exp(x_t\beta)}{1+\exp(x_t\beta)}\right\}^{u_t}\left\{\frac{1}{1+\exp(x_t\beta)}\right\}^{1-u_t} = \prod_{t=1}^N \frac{\exp(x_t\beta)^{u_t}}{1+\exp(x_t\beta)}. \tag{5}$$

Using the result of [23] with $k_t = u_t - 1/2$ and setting $\Omega = diag\{\omega_1, \ldots, \omega_T\}$, the augmented likelihood is proportional to

$$\mathcal{L}(\beta, \omega) \propto \prod_{t=1}^{T} \frac{1}{2} \exp(k_t x_t \beta) \int_0^{\infty} \exp\left\{-\omega_{t,s}(x_t\beta)^2/2\right\} p(\omega_t) d\omega_t. \tag{6}$$

Assuming as prior distributions $\omega \sim \mathcal{PG}(1,0)$ and $\beta \sim \mathcal{N}(m_{\beta_0}, V_{\beta_0})$, simulation from the posterior distributions can be done iteratively in two steps:

$$\omega_t \mid u_t \sim \mathcal{PG}(1, x_t\beta), \ t = 1, \ldots, T \tag{7}$$

$$\beta \mid U, \Omega \sim \mathcal{N}(m_{\beta}, V_{\beta}),$$

where $V_{\beta} = \left(X'\Omega X + V_{\beta_0}^{-1}\right)^{-1}$ and $m_{\beta} = V_{\beta}\left(X'k + V_{\beta_0}^{-1} m_{\beta_0}\right)$, and $k = (u_1 - 1/2, \ldots, u_T - 1/2)$.

To account for model uncertainty, on the predictor set, we perform a reversible jump step withing the augmented Pólya-Gamma data augmentation scheme, as proposed in [24]

2.1.1. Evaluation Metrics

For each iteration, we kept an in-sample replication of the binary series and compared it with the actual binary series U_t. Using a 1-0 loss function, we measure the accuracy of the logistic regression model—by means of the average number of misestimated data values per iteration—in representing the studied series. We report the results of this study in Section 3.2.

2.2. The Non-Homogeneous Polya-Gamma Hidden Markov Model

Given a time horizon $T \geq 0$ and discrete observation times $t = 1, 2, \ldots, T$, we consider an observed random process $\{Y_t\}_{t \leq T}$ and a hidden underlying process $\{Z_t\}_{t \leq T}$. The hidden process $\{Z_t\}$ is assumed to be a two-state non-homogeneous discrete-time Markov chain, $s = 1, 2$, that determines the states of the observed process. In our setting, the observed process is either the BTC or the ETH logarithmic prices series. The description of the hidden states is not pre-determined and is subject to the interpretation of the results.

Let y_t and z_t be the realizations of the random processes $\{Y_t\}$ and $\{Z_t\}$, respectively. We assume that at time t, $t = 1, \ldots, T$, y_t depends on the current state z_t and not on the previous states. Consider also the set of predictors $\{X_t\}$ of Section 2.1. A subset of the predictors $X_t^{(1)} \subseteq \{X_t\}$ of length $r_1 - 1$ affects the cryptocurrency linearly. In addition, a subset $X_t^{(2)} \subseteq \{X_t\}$ of length $r_2 - 1$ is used to describe the dynamics of the time-varying transition probabilities, i.e., the probabilities of moving from hidden state $s = 1$ to the hidden state $s = 2$ and vice versa. Thus, we allow the predictors to affect the series $\{Y_t\}$ linearly and non-linearly. Given the above, the cryptocurrency price series $\{Y_t\}$ can be modeled as

$$Y_t \mid Z_t = s \sim \mathcal{N}(x_{t-1}^{(1)} B_s, \sigma_s^2), \ s = 1, 2, \tag{8}$$

where $B_s = (b_{0s}, b_{1s}, \ldots, b_{r_1-1s})'$ are the regression coefficients and $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. The dependence of the observed process on the unobserved states, allows the model to capture the non-stationarities, non-linearities and the changes in the volatility, i.e., heteroskedasticity of the cryptocurrency series.

The dynamics of the unobserved process $\{Z_t\}$ can be described by the time-varying (non-homogeneous) transition probabilities, which depend on the predictors $X_t^{(2)}$ and are given by the following relationship

$$P(Z_{t+1} = j \mid Z_t = i) = p_{ij}^{(t)} = \frac{\exp(x_t^{(2)}\beta_{ij})}{\sum_{j=1}^{2} \exp(x_t^{(2)}\beta_{ij})}, \ i, j = 1, 2, \tag{9}$$

where $\beta_{ij} = (\beta_{0,ij}, \beta_{1,ij}, \ldots, \beta_{r_2-1,ij})'$ is the vector of the logistic regression coefficients to be estimated. Please note that for identifiability reasons, we adopt the convention of setting, for each row of the transition matrix, one of the β_{ij} to be a vector of zeros. Without loss of generality, we set $\beta_{ij} = \beta_{ji} = \mathbf{0}$ for $i, j = 1, 2, i \neq j$. Hence, for $\beta_i := \beta_{ii}$, $i = 1, 2$, the probabilities can be written in a simpler form

$$p_{ii}^{(t)} = \frac{\exp(x_t^{(2)} \beta_i)}{1 + \exp(x_t^{(2)} \beta_i)} \text{ and } p_{ij}^{(t)} = 1 - p_{ii}^{(t)}, \; i, j = 1, 2, \; i \neq j. \tag{10}$$

Summing up, the unknown quantities of the NHPG are $\{\theta_s = (B_s, \sigma_s^2), \beta_s, s = 1, 2\}$, i.e., the parameters in the mean predictive regression equation and the parameters in the logistic regression equation for the transition probabilities of the unobserved process $\{Z_t\}$, $t = 1, \ldots, T$. We follow the Bayesian methodology of [24], for joint inference on model specification, model parameters and predictions. Specifically, the authors in [24] use conditional conjugate analysis for the parameters in the mean predictive regression equation, i.e.

$$\sigma_s^2 \sim \mathcal{IG}(p, q), \; B_s \mid, \sigma_s^2 \sim \mathcal{N}(L_0, \sigma_s^2 V_0), \; s = 1, 2, \tag{11}$$

where $\mathcal{IG}$ denotes the Inverted-Gamma distribution. The conditional and the marginal posterior distributions for the state specific parameters σ_s and B_s,

$$\sigma_s^2 \sim \mathcal{IG}\left(p + \frac{n_s}{2}, q + \frac{1}{2}\left(L_{0s}' V_{0s}^{-1} L_{0s} + Y_s' Y_s - L_s' V_s^{-1} L_s\right)\right), \tag{12}$$

$$B_s \mid \sigma_s^2, z^T, y^T \sim \mathcal{N}\left(L_s, \sigma_s^2 V_s\right), \tag{13}$$

with $V_s = \left(V_{0s}^{-1} + X_s^{(1)'} X_s^{(1)}\right)^{-1}$, $L_s = V_s\left(V_{0s}^{-1} L_{0s} + X_s^{(1)'} Y_s\right)$. To make inference about the logistic regression coefficients, the authors model the probabilities of staying at the same state for two consecutive time periods, i.e., p_{ss}^t. They define, for $t = 1 \ldots, T - 1$, the quantity $\tilde{Z}_{t+1}^s = I[Z_{t+1} = Z_t = s]$ with the sum $\sum_t \tilde{Z}_{t+1}^s$, be the number of times that the chain was at state s for two consecutive time periods. Then,

$$p\left(\tilde{Z}_{t+1}^s = 1 \mid x_t^{(2)}\right) = p_{ss}^t = \frac{\exp\left(x_t^{(2)} \beta_s\right)}{1 + \exp\left(x_t^{(2)} \beta_s\right)} \Leftrightarrow logit\left(p_{ss}^{(t)}\right) = x_t^{(2)} \beta_s, \; s = 1, 2, \tag{14}$$

and inference for the probabilities falls to the case of the logistic model of Section 2.1

Finally, for every dataset we kept L out-of-sample observations and compare the estimated forecasts using the NHPG model. Given model M, the predictive distribution of y_{T+L}, $L \leq 1$ is

$$f_p\left(y_{T+L} \mid y^{T+L-1}\right) = \int f\left(y_{T+L} \mid y^{T+L-1}, z^{T+L-1}, M, \beta_M, \theta_M\right) \pi\left(\beta_M, \theta_M \mid y^{T+L-1}\right) d\beta_M d\theta_M, \tag{15}$$

where $f\left(y_{T+1} \mid y^T, z^T, \beta_M, \theta_M\right) = \sum_{s=1}^2 P\left(Z_{T+1} = s \mid Z_T = z_T\right) f_s\left(y_{T+1}\right)$.

All in all, the MCMC sampling scheme is constructed with recursive updates of (i) the latent variables z^T given the current value of the model parameters by using the scaled Forward–Backward algorithm (Scott [54]) (ii) the logistic regression coefficients by adopting the auxiliary variables method of Polson et al. [23] given the sequence of states z^T, (iii) the mean regression coefficients conditional on z^T by using the Gibbs sampling algorithm (iv) the covariate set using a couple of reversible jump steps and (v) the predictive distributions given the parameters and hidden states. The MCMC steps are detailed in Algorithm 1.

Algorithm 1 MCMC Sampling Scheme for Inference on Model Specification and Parameters

---

1: *% After each procedure the parameters and model space are updated conditionally on the previous quantities*

2: **procedure** SCALED FORWARD–BACKWARD$((\theta, y^t))$

3: 　　*%Simulation of a realization of the hidden states z_t*

4: 　　**for** $t = 1, \ldots, T$ and $i = 1, 2$ **do**

5: 　　　　$\pi_t(i \mid \theta) \leftarrow \frac{\alpha_t(s)}{\sum_{j=1}^2 \alpha_t(j)} = P(z_t = i \mid \theta, y^t)$ 　(▷) Simulation of the scaled forward probabilities

6: 　　**for** $t = T, T - 1, \ldots, 1$ **do**

7: 　　　　$z_t \leftarrow P(z_t \mid z_{t+1}) = \frac{p_{iz_{t+1}} \pi_t(i|\theta)}{\sum_{j=1}^m p_{jz_{t+1}} \pi_t(j|\theta)}$ 　　　　　　　　　　(▷) Backwards simulation of z_t

8: **procedure** MEAN_REGRES_PARAM$(\beta_s, \sigma_s, s = 1, 2)$

9: 　　*%Simulation of the mean regression parameters*

10: 　　**for** $s = 1, 2$ **do** 　　　　　　　　　　　　　　(▷) Conjugate analysis with Gibbs sampler

11: 　　　　$\beta \mid \sigma^2 \sim f_B, \sigma^2 \sim IG$ 　　　　　　　　(▷) $f_B \equiv$ Normal and $f_\sigma \equiv$ Inverse-Gamma

12: **procedure** LOG_REGRES_COEF$((\beta_s, \omega_s))$

13: 　　*%Simulation of the logistic regression coefficients*

14: 　　**for** $s = 1, 2$ **do** 　　　　　　　　　　　　　(▷) Pólya-Gamma data augmentation scheme

15: 　　　　augment the model space with ω_s 　　(▷) Conjugate analysis on the augmented space

16: 　　　　sample from $\beta_s \sim f_{\beta_s|\omega}$

17: 　　　　　　and $\omega_s \mid \beta_s \sim \mathcal{PG}$ 　　(▷) Posteriors $f_{\beta_s|\omega} \equiv$ Normal and $\mathcal{PG} \equiv$ Pólya-Gamma

18: **procedure** DOUBLE_REV_JUMP$(X^{(1)}, X^{(2)})$

19: 　　*%Variable selection with double reversible jump step*

20: 　　**for** $i = 1, 2$ **do** *%Propose to add/remove a covariate*

21: 　　　　add: choose $X^{candidate}$ from $X \cap X^{(i)^c}$ 　　　　(▷) Calculate acceptance probability α

22: 　　　　**if** $\alpha < rand(0, 1)$ **then** $X^{(i)} \leftarrow X^{(i)} \cup X^{can}$

23: 　　　　remove: choose $X^{candidate}$ from $X^{(i)}$ 　　　　(▷) Calculate acceptance probability α

24: 　　　　**if** $\alpha < rand(0, 1)$ **then** $X^{(i)} \leftarrow X^{(i)} \cap X^{can^c}$

25: **procedure** PREDICT

26: 　　*% Make L-steps-ahead predictions*

27: 　　**for** $t = T + 1, \ldots, T + L$ **do**

28: 　　　　$\hat{y}_t \sim f$ with $f(y_{T+1} \mid y^T, z^T, \beta_M, \theta_M) = \sum_{s=1}^2 P(Z_{T+1} = s \mid Z_T = z_T) f_s(y_{T+1})$.

---

3. The Empirical Application

3.1. The Data

We assess the explanatory power of 12 financial/economic and 2 cryptocurrency specific variables, outlined in Table 1, on the BTC and ETH series, through two different experimental exercises: (a) the logistic regression model and (b) NHPG model. We analyze the daily BTC and ETH price series for the period ranging from 1/1/2017 to 16/11/2019. Missing data due to the non-business days are filled with the last available value. For the first exercise, we transform the price series into binary series using the transformation $\{Y_t - Y_{t-1} \geq \alpha\}$.

Table 1. List of explanatory variables along with the applied transformations and the online resources. The Hash Rate (HR) and Average Block Size (AVS) have been retrieved [55] for Bitcoin and from [56] for Ether.

Explanatory Variables			
Description	Symbol	Transformation	Retrieved from
US dollars to Euros exchange rate	USD/EUR	Normalized	investing.com
US dollars to GBP exchange rate	USD/GBP	Normalized	investing.com
US dollars to Japanese Yen exchange rate	USD/JPY	Normalized	investing.com
US dollars to Chinese Yuan exchange rate	USD/CNY	Normalized	investing.com
Russel 2000 index	R2000	Normalized	finance.yahoo.com
Standard & Poor's 500 index	SP500	Normalized	finance.yahoo.com
NASDAQ Composite index	NASDAQ	Normalized	finance.yahoo.com
Dow Jones Industrial Average	DOW	Normalized	finance.yahoo.com
Crude Oil Futures price	OIL	Normalized	finance.yahoo.com
Price of Gold	GOLD	Normalized	finance.yahoo.com
CBOE Volatility index	VIX	Normalized	finance.yahoo.com
Equity market Economic Uncertainty index	EUI	None	fred.stlouisfed.org
Hash Rate	HR	Percentage of change	quandl.com/etherscan.io
Average Block Size	AVS	Percentage of change	quandl.com/etherscan.io

The aim of this experiment is to model/explain the probability of a positive return. We repeat the experiment for different magnitudes (α) of returns, i.e., gains more than $0\%, 1\%, \ldots, 5\%$. For the second exercise, we use the logarithm of the price series of the two aforementioned cryptocurrencies. Additionally to the previous data sets, we apply the NHPG methodology in a larger BTC log-price series (for the period ranging from 1/1/2014 until 16/11/2019). The closing BTC prices were downloaded from [20] and the ETH prices were downloaded from [56].

Table 2. Correlation matrix of the explanatory variables.

Variables	USD/EUR	USD/GBP	USD/JPY	USD/CNY	R2000	SP500	NASDAQ	DOW	OIL	GOLD	VIX	EUI	HR	AVS
USD/EUR	1.0	0.63	0.6	0.54	0.16	0.25	0.26	0.2	−0.82	−0.33	0.08	0.13	0.01	−0.02
USD/GBP	0.63	1.0	0.92	−0.07	0.66	0.73	0.73	0.71	−0.56	−0.18	−0.09	0.18	−0.01	−0.02
USD/JPY	0.6	0.92	1.0	−0.09	0.55	0.66	0.65	0.64	−0.54	−0.31	−0.03	0.23	0.00	−0.01
USD/CNY	0.54	−0.07	−0.08	1.0	−0.06	−0.08	−0.06	−0.1	−0.42	−0.21	0.12	−0.05	0.00	−0.02
R2000	0.16	0.66	0.55	−0.06	1.0	0.95	0.95	0.95	−0.08	−0.07	−0.23	0.03	−0.02	−0.01
SP500	0.25	0.73	0.66	−0.08	0.95	1.0	1.0	0.95	−0.25	−0.19	−0.11	0.12	−0.02	−0.02
NASDAQ	0.26	0.73	0.65	−0.06	0.95	0.95	1.0	0.99	−0.25	−0.18	−0.07	0.13	−0.02	−0.01
DOW	0.2	0.71	0.64	−0.1	0.95	1.0	0.99	1.0	−0.2	−0.16	−0.09	0.12	−0.02	−0.02
OIL	−0.82	−0.56	−0.54	−0.42	−0.08	−0.25	−0.25	−0.2	1.0	0.6	−0.21	−0.18	−0.00	0.00
GOLD	−0.33	−0.18	−0.31	−0.21	−0.07	−0.19	−0.18	−0.16	0.6	1.0	−0.12	−0.09	−0.00	0.00
VIX	0.08	−0.09	−0.03	0.12	−0.23	−0.11	−0.07	−0.09	−0.21	−0.12	1.0	0.39	0.02	0.00
EUI	0.13	0.18	0.23	−0.05	0.03	0.12	0.13	0.12	−0.18	−0.09	0.39	1.0	−0.02	−0.01
HR	0.01	−0.01	0.00	0.00	−0.02	−0.02	−0.02	−0.02	0.00	−0.01	0.02	−0.02	1.0	−0.18
AVS	−0.02	−0.02	−0.01	−0.02	−0.01	−0.02	−0.01	−0.02	0.02	0.01	0.00	−0.01	−0.18	1.0

3.2. Results: The Logistic Regression Model

We report the in-sample performance for the logistic regression model in Table 3 for every threshold and for both cryptocurrencies. Even though the in-sample performance—based on the large number of incorrect point estimates of the series—is poor, we see that when the threshold α increases, the in-sample performance is improved.

Table 3. Mean incorrect estimations out of the total $T = 1017$ observations, per iteration of the BTC and ETH binary series. Increasing the magnitude of minimum returns the average number of misestimations decreases. This shows that the covariate set has explanatory power on defining the probability of larger returns. In parenthesis we report the error rate.

	Mean Incorrect Estimations Per Iteration					
	Thresholds					
Coin	$\alpha = 0\%$	$\alpha = 1\%$	$\alpha = 2\%$	$\alpha = 3\%$	$\alpha = 4\%$	$\alpha = 5\%$
BTC	501 (0.49%)	485 (48%)	399 (39%)	314 (31%)	234 (23%)	174 (17%)
ETH	504 (0.49%)	428 (42%)	472 (46%)	413 (40.7%)	331 (32.7%)	259 (25.5%)

To be more precise, we find that for $\alpha = 0\%$ the covariate set has no explanatory power on the $U_{t,\alpha}$ since the probability of incorrect estimations is almost 50% for both coins. However, for $\alpha = 5\%$ the probability of incorrect estimations drops to 17% for the binary BTC series and to 25.5% for the binary ETH series. This result is an indication that the accounted covariate set, which includes fiat currencies, stock indices and commodities can be used to explain/predict the possibility of larger positive returns. A visualization of the best in-sample performance for the $U_{t,5}$ series for BTC and ETH is in Figures 1 and 2.
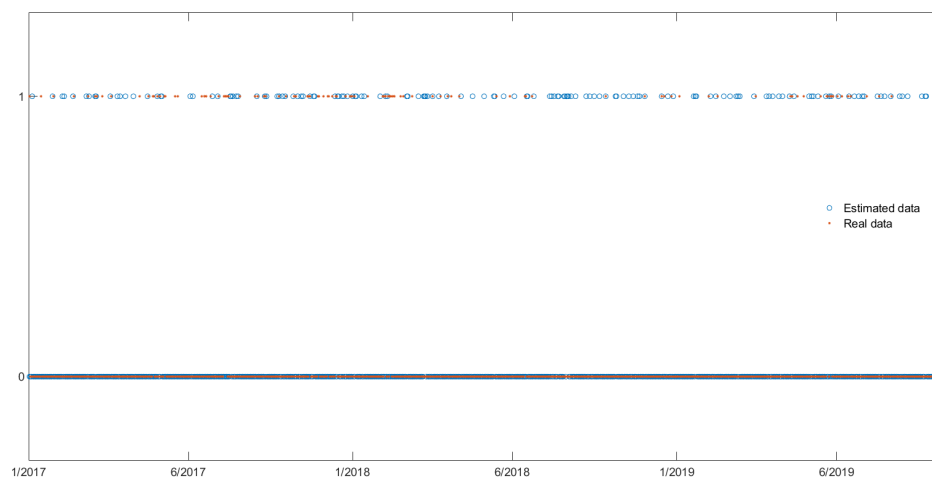

Figure 1. Realization of the binary BTC series with $\alpha = 5\%$ using the logistic regression model. The blue circles represent the realized series whereas the red dots are the actual data points.
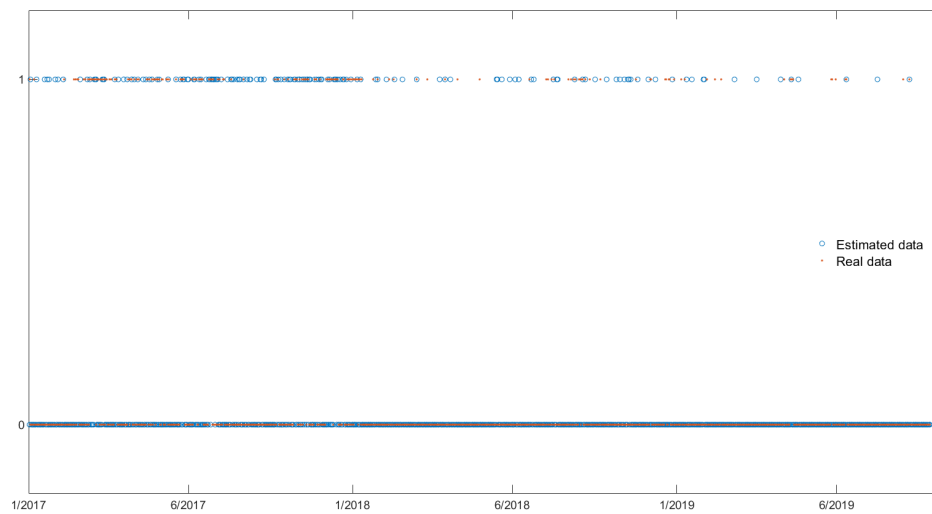
Figure 2. Realization of the binary ETH series with $\alpha = 5\%$ using the logistic regression model. The blue circles represent the realized series whereas the red dots are the actual data points.

In Tables 4 and 5, we report the posterior probabilities of inclusion π_k, $k = 1, \ldots, K$ of the K explanatory variables for the studied logistic models for the BTC and ETH binary series, respectively. For both coins, we observe that the studied covariated set does not affect nor predict the probability of positive return, i.e., $Y_t - Y_{t-1} \geq 0$.

Table 4. Posterior probabilities of inclusion of the explanatory variables for the binary BTC series for the period ranging from 1/2016 to 11/2019 The probabilities of the variables that are included in the median probability model, i.e., the variables with probability of inclusion above 0.5, are highlighted with bold.

Predictors	Posterior Probabilities of Inclusion					
	Return's magnitude					
	$\alpha = 0\%$	$\alpha = 1\%$	$\alpha = 2\%$	$\alpha = 3\%$	$\alpha = 4\%$	$\alpha = 5\%$
USD/EUR	0	0	0	0	**1.00**	**0.65**
USD/GBP	0	0	0	0	0	0
USD/JPY	0	0	0	0	0	0
USD/CNY	0	0.01	0	0	0	0.02
R2000	0	0.01	0.02	**1.00**	**1.00**	**0.72**
SP500	0	0. 3	0.02	0	0.01	0.02
NASDAQ	0	0.03	0.07	0	0	0
DOW	0	0.08	0.03	0	0	0
OIL	0	0.07	**0.77**	0	0.01	0.01
GOLD	0	0.11	**0.80**	0.01	0.02	0.07
VIX	0	0	0	0.3	0	0
EUI	0	0	0	0	0	0
HR	0	0	0	0	0	0
AVS	0	0	0	0	0.01	0.15

Table 5. Posterior probabilities of inclusion of the explanatory variables for the binary ETH series for the period ranging from 1/2016 to 11/2019 The probabilities of the variables that are included in the median probability model, i.e., the variables with probability of inclusion above 0.5, are highlighted with bold.

	Posterior Probabilities of Inclusion					
Predictors	$\alpha = 0\%$	$\alpha = 1\%$	$\alpha = 2\%$	$\alpha = 3\%$	$\alpha = 4\%$	$\alpha = 5\%$
USD/EUR	0.01	0	0.11	0.01	0.18	**0.55**
USD/GBP	0.04	0.02	0	0.01	0.01	0.14
USD/JPY	0.01	0	0.01	0	0.03	0.03
USD/CNY	0.03	0.05	0.02	0.01	0.16	0.47
R2000	0.03	0.05	0.01	0.01	0.03	0.07
SP500	0.22	**0.78**	0.37	0.28	0.36	**0.60**
NASDAQ	0.30	**0.53**	**0.71**	**0.92**	**0.89**	**0.97**
DOW	0	0.02	0.45	0.09	0.24	0.37
OIL	0	0.08	0.02	0	0.03	0.11
GOLD	0.07	0.13	0	0	0.01	0.32
VIX	0.01	0.01	**0.58**	**0.80**	**0.55**	**0.83**
EUI	0	0	0	0	0	0
HR	0	0	0	0	0	0.07
AVS	0	0.03	0.07	0.01	0.19	**0.94**

However, if we change the magnitude of return (α), we find that there is a correlation between the covariate set and the probabilities of observing $Y_t - Y_{t-1} \geq \alpha$, $\alpha = \{1, 2, 3, 4, 5\%\}$. We find that the sets including the covariates that affect the binary BTC and ETH series—covariates with posterior probability of inclusion above 0.5—are different. Also, we observe that even though the binary ETH series is correlated with more covariates than the covariates of the binary BTC series, the logistic model has worse performance in explaining the studied series (as seen in Table 3).

The results of this experiment imply that the accounted covariate set has some explanatory power on the series but a more elaborated and more complicated model, such as the NHPG model, needs to be considered.

3.3. Results: The NHPG Model

In this section, we present the results of the NHPG model using the logarithmic ETH and BTC price series. Figures 3–5 plot the log ETH, log BTC and extended log BTC datasets (blue line) along with the estimated in-sample time series (gray line). This shows graphically the good in-sample performance of the NHPG model to replicate the log BTC and log ETH series. Shaded bars indicate the time periods that the underlying hidden process is in state 1. The states alternate between 1 and 2 rather frequently, confirming previous studies on the heteroskedasticity of the series, see e.g., [11], and on the existence of structural breaks and regime switches, see e.g., [39,44].
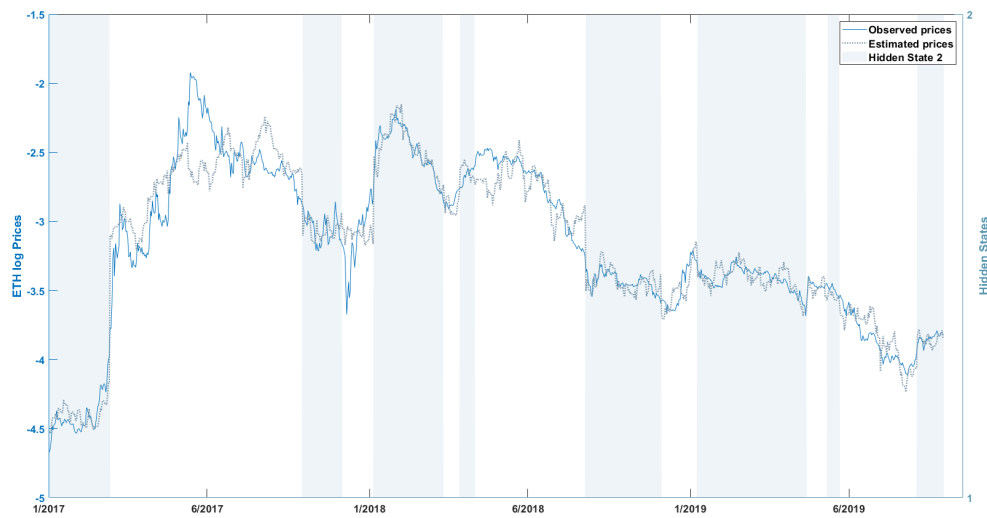
Figure 3. Logarithmic ETH price series (blue line) and in-sample estimated logarithmic ETH price series for the period 6/2016–5/2019 (gray dotted line). Shaded bars mark times with hidden state 1 (smoothed probability above 0.5). The model accounts for the heteroskedasticity of the series.
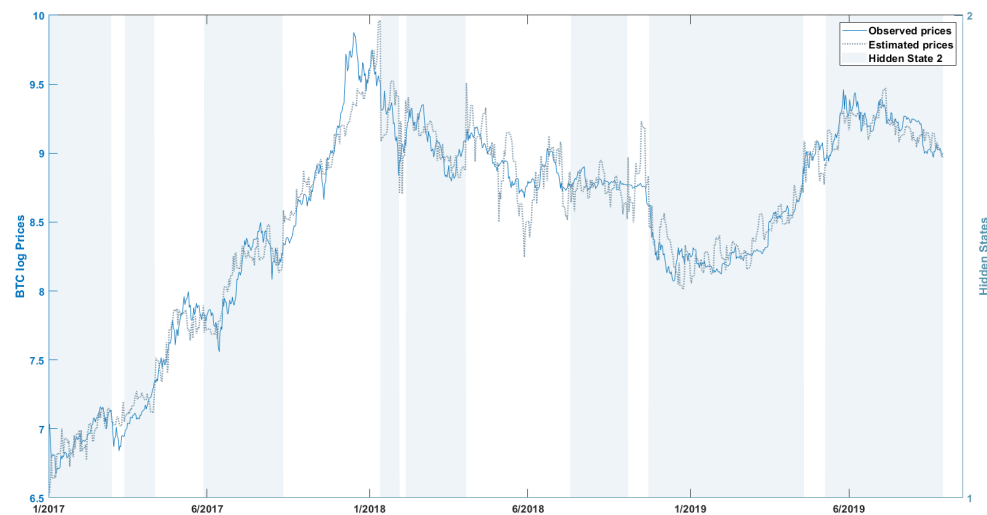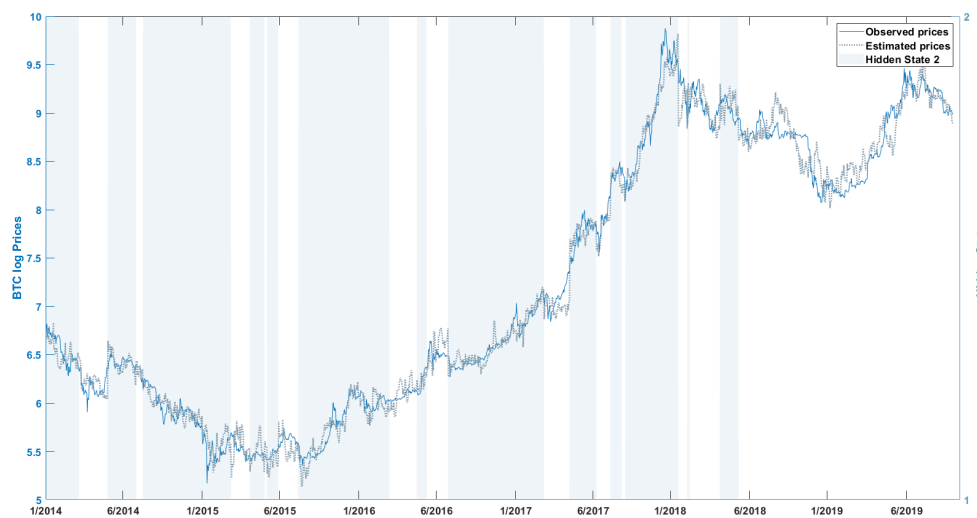

Figure 4. Logarithmic BTC prices series (blue line) and in-sample estimated logarithmic BTC price series for the period 6/2016–5/2019 (gray dotted line). Shaded bars mark times with hidden state 1 (smoothed probability above 0.5).

Figure 5. Logarithmic BTC prices series (blue line) and in-sample estimated logarithmic BTC price series for the period 5/2013–5/2019 (gray dotted line). Shaded bars mark times with hidden state 1 (smoothed probability above 0.5). The change of the sample sizes has a significant impact on the distribution of the unobserved process.

However, the out-of-sample performance of the NHPG is poor. Figure 6, shows the posterior mean (gray lines), of the 30 empirical predictive distributions, along with the actual out-of sample log prices of ETH, BTC (blue line). The mean posterior out-of-sample predictions are in general not good, as they frequently miss the direction of movement of the series.

(a) ETH   **(b)** BTC
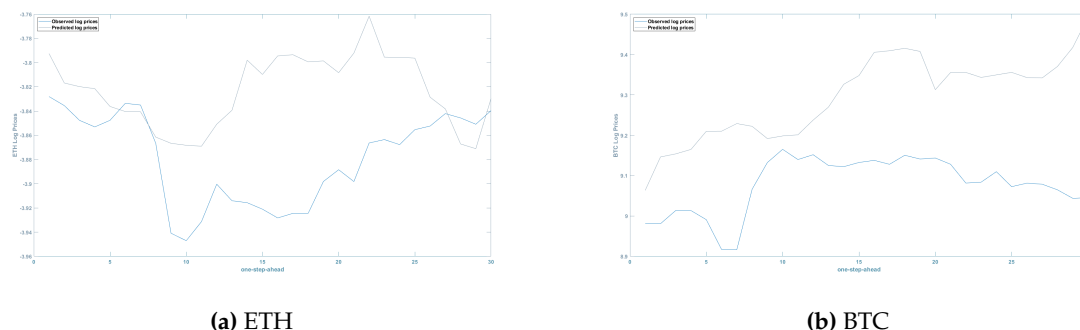
Figure 6. Mean posterior out-of-sample predictions (gray line) for $L = 30$ days both for the (**a**) ETH and (**b**) BTC log-transformed price series (blue line). While the predictions for ETH are better than those for BTC, both are not satisfactory as they frequently miss the direction of price movement. The BTC predictions are essentially the same for both the 2016–2019 and 2013–2019 data sets (the second not shown here).

Even worse, when we examine the posterior prediction intervals with boundaries the 2.5% and 97.5% quantiles of the empirical predictive densities—instead of the mean point forecasts—we find, for some cases, that they do not include the actual out-of-sample values. Hence, we confirm the claim of previous studies that financial and economic variables do not accurately predict the price fluctuation of cryptocurrencies. The good in-sample and poor out-of-sample performance of the two-state Non-Homogeneous Hidden Markov models is also observed in the exchange rate literature, see for example [30,31].

In Table 6, we report the posterior probabilities of inclusion of the explanatory variables for the mean equation—first number—and the logistic regression—second number in every cell. Variables with posterior probability of inclusion above 0.5 in either the mean equation or the transition

probabilities are marked with bold. These variables make up the median probability model (MPM). Although the BTC and ETH are correlated [15], the variables that affect the series (by means of the MPM) are not the same. The MPM of ETH consists of 7 covariates: the USD/EUR, USD/GBP and USD/CNY exchange rates, Russel 2000, S&P 500, Dow Jones, NASDAQ indices. The MPM of BTC additionally contains the USD/JPY exchange rate and VIX. This difference on the exchange rates indicates that ETH is more geographically restricted. Moreover, the hash rate (HR) and Average Block size is insignificant for both cryptocurrencies. Regarding the extended BTC dataset, we find that Gold and Crude Oil future prices are also significant. Furthermore, it is worth mentioning that all the effects are linear in BTC series and hence the transition probabilities of hidden states are homogeneous (constant through time). This is in difference with ETH price series where S&P 500 and USD/GBP affects the series also non-linearly. The inclusion of these two variables in the transition probabilities equation results to the non-constant transition probabilities and consequently indicates that the Non-Homogeneous Hidden Markov model is promising in the understanding of cryptocurrencies price formation. The non-constant transition probabilities along with the fact that there exist other variables with non-negligible posterior probabilities of inclusion (above 0.3), imply that the are other drives that drive changes in the underlying process that go beyond the financial aspects that have been considered in the present setting. Finally, we observe that even though the number of statistical important variables is large, for both coins, the forecasting performance of this model is very modest. This is an indication that non-traditional financial and economic variables need to be considered.

Table 6. Posterior probabilities of inclusion of the explanatory variables. The first value in each cell is the posterior probability of inclusion in the mean equation and the second the probability of inclusion in the transition probabilities of the underlying Markov process. The probabilities of the variables that are included in the median probability model, i.e., variables with probability above 0.5 in either the mean equation or the logistic regression equation are highlighted with bold.

	Posterior Probabilities of Inclusion		
	Data Sets		
Predictors	BTC	BTC	ETH
Sample period	1/2014-11/2019	1/2017-11/2019	1/2017-11/2019
USD/EUR	**1.00 0**	**1.00 0**	**1.00 0.12**
USD/GBP	**1.00 0**	**1.00 0.01**	**0.97 0.54**
USD/JPY	**1.00 0**	**1.00 0**	0 0.06
USD/CNY	**1.00 0**	**1.00 0**	**1.00 0.11**
R2000	**1.00 0.06**	**1.00 0.01**	**1.00 0.26**
SP500	**1.00 0.08**	**1.00 0**	**1.00 0.63**
VIX	**1.00 0.07**	**0.70 0.08**	0 0.12
DOW	**1.00 0.06**	**0.90 0.01**	**1.00 0.32**
NASDAQ	**1.00 0.06**	**1.00 0.01**	**1.00 0.34**
GOLD	**1.00 0.02**	0.32 0.05	0 0.45
CO	**1.00 0.01**	0.24 0.01	0.00 0.08
EUI	0 0.01	0 0.01	0 0
HR	0 0.02	0 0	0 0.01
AVS	0 0.02	0 0	0 0

To sum up, even though we observe a large number of statistical significant explanatory variables, the insufficient forecasting performance of the NHPG model confirms that the cryptocurrencies are still decoupled from the mainstream financial and economic assets [57].

4. Concluding Remarks

We applied a logistic regression model with a predefined covariate set to model the probabilities of observing daily returns exceeding a predefined threshold for the Bitcoin (BTC) and Ether (ETH) series. We show empirically that the logistic model has weak fitting performance. However, we find that changing the magnitude of positive returns, improves the fitting performance of the logistic regression model. This result motivated us to incorporate the logistic regression model into a more complex model. Therefore, we applied a specific instance of the Non-Homogeneous Hidden Markov models to the logarithm of BTC and ETH price series.

We used the non-homogeneous Pólya-Gamma Hidden Markov Model (NHPG) of [24]. Focusing on a data set of financial/economic predictors, we studied general properties of the cryptocurrency price series. While the NHPG algorithm exhibited good in-sample performance, it revealed that changes in the underlying two-state Markov process are frequent, thus indicating that the states are not persistent, contributing to the already high heteroskedasticity of both the Bitcoin and the Ether data series. Notably, both cryptocurrency specific variables were not found significant for BTC and ETH. Significance of exchange rates revealed a more geographically restricted interest for ETH than for BTC.

From a modeling point of view, the median probability model included too many covariates, thus, indicating data with high variability and confirming that financial and economic variables—even if cryptocurrency specific—are not enough to explain the formation of cryptocurrency prices. Along with the poor out-of-sample predictions, these findings show that even algorithms with good performance on conventional financial data do not capture all aspects of cryptocurrencies. In the main takeaway of this study, these results back earlier findings that cryptocurrencies are unlike any other financial asset and that the understanding of their properties requires not only the combination of more sophisticated models but also the inception of novel ideas and tools.

While the current study offers a novel perspective on the hidden states—and hence on the underlying forces—that drive cryptocurrency markets, it also suggests that the analysis of their price formation requires more elaborate tools. Recent advances in deep neural networks provide methods to identify hidden layers that approximate complex non-linear relationships. Specifically, by exploring electronic high-frequency data of supply, demand and prices in financial markets, Deep Learning models can uncover universal price formation mechanisms, [58]. This approach seems particularly promising for cryptocurrency markets. Along these lines, the current model may prompt a more extensive application of the rich Hidden Markov theory and analytical toolbox on cryptocurrency markets.

All in all, the investigation of the exogenous variables that affect or drive the cryptocurrency market can be useful to investors, policy makers, traders for portfolio allocation, risk management and trading strategies.

References

1. Cheah, E.T.; Fry, J. Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Econ. Lett.* **2015**, *130*, 32–36. doi:10.1016/j.econlet.2015.02.029. [CrossRef]

2. Klein, T.; Thu, H.; Walther, T. Bitcoin is not the New Gold – A comparison of volatility, correlation, and portfolio performance. *Int. Rev. Financ. Anal.* **2018**, *59*, 105–116. doi:10.1016/j.irfa.2018.07.010. [CrossRef]

3. Ciaian, P.; Rajcaniova, M.; Kancs, A. The economics of BitCoin price formation. *Appl. Econ.* **2016**, *48*, 1799–1815. doi:10.1080/00036846.2015.1109038. [CrossRef]

4. Dyhrberg, A.H. Bitcoin, gold and the dollar – A GARCH volatility analysis. *Financ. Res. Lett.* **2016**, *16*, 85–92. doi:10.1016/j.frl.2015.10.008. [CrossRef]

5. Bouri, E.; Gupta, R.; Tiwari, A.K.; Roubaud, D. Does Bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. *Financ. Res. Lett.* **2017**, *23*, 87–95. doi:10.1016/j.frl.2017.02.009. [CrossRef]

6. Bouri, E.; Azzi, G.; Dyhrberg, A.H. On the return-volatility relationship in the Bitcoin market around the price crash of 2013. *Econ. Open Access Open Assess. J.* **2017**, *11*, 1–16. doi:10.5018/economics-ejournal.ja.2017-2. [CrossRef]

7. Bouri, E.; Molnár, P.; Azzi, G.; Roubaud, D.; Hagfors, L.I. On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? *Financ. Res. Lett.* **2017**, *20*, 192–198. doi:10.1016/j.frl.2016.09.025. [CrossRef]

8. Demir, E.; Gozgor, G.; Lau, C.M.; Vigne, S.A. Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation. *Financ. Res. Lett.* **2018**, *26*, 145–149. doi:10.1016/j.frl.2018.01.005. [CrossRef]

9. Katsiampa, P. Volatility estimation for Bitcoin: A comparison of GARCH models. *Econ. Lett.* **2017**, *158*, 3–6. doi:10.1016/j.econlet.2017.06.023. [CrossRef]

10. Hayes, A.S. Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. *Telemat. Inform.* **2017**, *34*, 1308–1321. doi:10.1016/j.tele.2016.05.005. [CrossRef]

11. Phillip, A.; Chan, J.; Peiris, S. On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin. *Econom. Stat.* **2018**. doi:10.1016/j.ecosta.2018.10.003. [CrossRef]

12. Georgoula, I.; Pournarakis, D.; Bilanakos, C.; Sotiropoulos, D.; Giaglis, G.M. Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices. In Proceedings of the 9th Mediterranean Conference on Information Systems , Samos, Greece, 3–5 October 2015. doi:10.2139/ssrn.2607167. [CrossRef]

13. Kraaijeveld,O. and Smedt, J.D. The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *J. Int. Financ. Mark. Inst. Money* **2020**, doi:0.1016/j.intfin.2020.101188. [CrossRef]

14. Buterin, V.; Reijsbergen, D.; Leonardos, S.; Piliouras, G. Incentives in Ethereum's Hybrid Casper Protocol. In Proceedings of the 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Seoul, Korea, 15–17 May 2019, pp. 236–244. doi:10.1109/BLOC.2019.8751241. [CrossRef]

15. Borri, N. Conditional tail-risk in cryptocurrency markets. *J. Empir. Financ.* **2019**, *50*, 1–19. doi:10.1016/j.jempfin.2018.11.002. [CrossRef]

16. Vora, R. Ethereum Price Analysis: Ethereum (ETH) Needs To Discover The Magic Spell To Surge On Its Own Available Online: https://www.cryptonewsz.com/ethereum-price-analysis-ethereum-eth-needs-to-discover-the-magic-spell-to-surge-on-its-own/29402/ (accessed on 27 July 2019).

17. Gandal, N.; Halaburda, H. Can We Predict the Winner in a Market with Network Effects? Competition in Cryptocurrency Market. *Games* **2016**, *7*, 16. doi:10.3390/g7030016. [CrossRef]

18. Hotz-Behofsits, C.; Huber, F.; Zörner, T.O. Predicting crypto-currencies using sparse non-Gaussian state space models. *J. Forecast.* **2018**, *37*, 627–640. doi:10.1002/for.2524. [CrossRef]

19. Buterin, V. A Next-Generation Smart Contract and Decentralized Application Platform, 2014. Available online: https://github.com/ethereum/wiki/wiki/White-Paper (accesed on 1 September 2019).

20. Crypto.com. Available Online: https://coinmarketcap.com/all/views/all/ (accessed on 30 July 2019).

21. Ethereum Foundation. Available Online: https://www.ethereum.org/ (accessed on 30 July 2019).

22. Frühwirth-Schnatter, S.; Frühwirth, R. Auxiliary mixture sampling with applications to logistic models. *Comput. Stat. Data Anal.* **2007**, *51*, 3509 – 3528. doi:10.1016/j.csda.2006.10.006. [CrossRef]

23. Polson, N.G.; Scott, J.G.; Windle, J. Bayesian Inference for Logistic Models Using Polya-Gamma Latent Variables. *J. Am. Stat. Assoc.* **2013**, *108*, 1339–1349. doi:10.1080/01621459.2013.829001. [CrossRef]

24. Koki, C.; Meligkotsidou, L.; Vrontos, I. Forecasting under model uncertainty: Non-homogeneous hidden Markov models with Pòlya-Gamma data augmentation. *J. Forecast.* **2018**.10.1002/for.2645. [CrossRef]

25. Meligkotsidou, L.; Dellaportas, P. Forecasting with non-homogeneous hidden Markov models. *Stat. Comput.* **2011**, *21*, 439–449. doi:10.1007/s11222-010-9180-5. [CrossRef]

26. Beckmann, J.; Schüssler, R. Forecasting exchange rates under parameter and model uncertainty. *J. Int. Money Financ.* **2016**, *60*, 267–288. doi:10.1016/j.jimonfin.2015.07.001. [CrossRef]

27. Groen, J.J.J.; Paap, R.; Ravazzolo, F. Real-Time Inflation Forecasting in a Changing World. *J. Bus. Econ. Stat.* **2013**, *31*, 29–44. doi:10.1080/07350015.2012.727718. [CrossRef]

28. Wright, J.H. Forecasting US inflation by Bayesian model averaging. *J. Forecast.* **2009**, *28*, 131–144. doi:10.1002/for.1088. [CrossRef]

29. Wright, J.H. Bayesian Model Averaging and exchange rate forecasts. *J. Econom.* **2008**, *146*, 329 – 341. doi:10.1016/j.jeconom.2008.08.012. [CrossRef]

30. Yuan, C. Forecasting exchange rates: The multi-state Markov-switching model with smoothing. *Int. Rev. Econo. Financ.* **2011**, *20*, 342–362. doi:10.1016/j.iref.2010.09.002. [CrossRef]

31. Marsh, I.W. High-frequency Markov switching models in the foreign exchange market. *J. Forecast.* **2000**, *19*, 123–134. doi:10.1002/(SICI)1099-131X(200003)19:2<123::AID-FOR750>3.0.CO;2-C. [CrossRef]

32. Atsalakis, G.S.; Atsalaki, I.G.; Pasiouras, F.; Zopounidis, C. Bitcoin price forecasting with neuro-fuzzy techniques. *Eur. J. Oper. Res.* **2019**, *276*, 770–780. doi:10.1016/j.ejor.2019.01.040. [CrossRef]

33. Bouri, E.; Shahzad, S.J.H.; Roubaud, D. Co-explosivity in the cryptocurrency market. *Financ. Res. Lett.* **2019**, *29*, 178–183. doi:0.1016/j.frl.2018.07.005. [CrossRef]

34. Bouri, E.; Gupta, R.; Roubaud, D. Herding behaviour in cryptocurrencies. *Financ. Res. Lett.* **2019**, *29*, 216–221. doi:10.1016/j.frl.2018.07.008. [CrossRef]

35. Engel, C. Can the Markov switching model forecast exchange rates? *J. Int. Econ.* **1994**, *36*, 151–165. doi:10.1016/0022-1996(94)90062-0. [CrossRef]

36. Lee, H.Y.; Chen, S.L. Why use Markov-switching models in exchange rate prediction? *Econ. Model.* **2006**, *23*, 662–668. doi:10.1016/j.econmod.2006.03.007. [CrossRef]

37. Frömmel, M.; MacDonald, R.; Menkhoff, L. Markov switching regimes in a monetary exchange rate model. *Econ. Model.* **2005**, *22*, 485 – 502. doi:10.1016/j.econmod.2004.07.001. [CrossRef]

38. Mamon, R.; Elliott, R. (Eds.) *Hidden Markov Models in Finance: Further Developments and Applications*; Springer: Berlin/Heidelberg, Germany, 2014; Volume II.

39. Ardia, D.; Bluteau, K.; Rüede, M. Regime changes in Bitcoin GARCH volatility dynamics. *Financ. Res. Lett.* **2019**, *29*, 266–271. doi:10.1016/j.frl.2018.08.009. [CrossRef]

40. Poyser, O. Exploring the dynamics of Bitcoin's price: a Bayesian structural time series approach. *Eurasia. Econ. Rev.* **2019**, *9*, 29–60. doi:10.1007/s40822-018-0108-2. [CrossRef]

41. Koutmos, D. Market risk and Bitcoin returns. *Ann. Oper. Res.* **2019**. doi:10.1007/s10479-019-03255-6. [CrossRef]

42. Koutmos, D. Liquidity uncertainty and Bitcoin's market microstructure. *Econ. Lett.* **2018**, *172*, 97–101. doi:10.1016/j.econlet.2018.08.041. [CrossRef]

43. Phillips, R.C.; Gorse, D. Predicting cryptocurrency price bubbles using social media data and epidemic modelling. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 November–1 December 2017. doi:10.1109/SSCI.2017.8280809. [CrossRef]

44. Mensi, W.; Al-Yahyaee, K.H.; Kang, S.H. Structural breaks and double long memory of cryptocurrency prices: A comparative analysis from Bitcoin and Ethereum. *Financ. Res. Lett.* **2019**, *29*, 222–230. doi:10.1016/j.frl.2018.07.011. [CrossRef]

45. Chaim, P.; Laurini, M.P. Volatility and return jumps in Bitcoin. *Econ. Lett.* **2018**, *173*, 158–163. doi:10.1016/j.econlet.2018.10.011. [CrossRef]

46. Balcilar, M.; Bouri, E.; Gupta, R.; Roubaud, D. Can volume predict Bitcoin returns and volatility? A quantiles-based approach. *Econ. Model.* **2017**, *64*, 74–81. doi:10.1016/j.econmod.2017.03.019. [CrossRef]

47. Jang, H.; Lee, J. An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information. *IEEE Access* **2018**, *6*, 5427–5437. doi:10.1109/ACCESS.2017.2779181. [CrossRef]

48. Pichl, L.; Kaizoji, T. Volatility Analysis of Bitcoin Price Time Series. *Quant. Financ. Econ.* **2017**, *1*, 474–485. doi:10.3934/QFE.2017.4.474. [CrossRef]

49. Baur, D.G.; Dimpfl, T.; Kuck, K. Bitcoin, gold and the US dollar—A replication and extension. *Financ. Res. Lett.* **2018**, *25*, 103 – 110. doi:10.1016/j.frl.2017.10.012. [CrossRef]

50. Walther, T.; Klein, T.; Bouri, E. Exogenous drivers of Bitcoin and Cryptocurrency volatility—A mixed data sampling approach to forecasting. *J. Int. Financ. Mark. Inst. Money* **2019**, *63*, 101133. doi:10.1016/j.intfin.2019.101133. [CrossRef]

51. Van Wijk, D. What Can Be Expected from the BitCoin. Ph.D. Thesis, Erasmus Universiteit Rotterdam, Rotterdam, The Netherlands, 2013.

52. Yermack, D. Chapter 2–Is Bitcoin a Real Currency? An Economic Appraisal. In *Handbook of Digital Currency*; Chuen, D.L., Ed.; Academic Press: San Diego, CA, USA, 2015; pp. 31–43. doi:10.1016/B978-0-12-802117-0.00002-3. [CrossRef]

53. Estrada, J.C.S. Analyzing Bitcoin Price Volatility. Ph.D. Thesis, University of California: Berkeley, CA, USA, 2017.

54. Scott, S.L. Bayesian Methods for Hidden Markov Models. *J. Am. Stat. Assoc.* **2002**, *97*, 337–351. doi:10.1198/016214502753479464. [CrossRef]

55. Quantl.com. The World'S Most Powerful Data Lives on Quandl. Available Online: https://www.quandl.com/ (accessed on 30 July 2019).

56. Etherscan.io. Available Online: https://etherscan.io (accessed on 30 July 2019).

57. Gil-Alana, L.A.; Abakah, E.J.A.; Rojo, M.F.R. Cryptocurrencies and stock market indices. Are they related? *Res. Int. Bus. Financ.* **2020**, *51*, 101063. doi:10.1016/j.ribaf.2019.101063. [CrossRef]

58. Sirignano, J.; Cont, R. Universal features of price formation in financial markets: perspectives from deep learning. *Quant. Financ.* **2019**, *19*, 1449–1459. doi:10.1080/14697688.2019.1622295. [CrossRef]