

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, 487372 Singapore, Singapore; stefanos_leonardos@sutd.edu.sg (S.L.); georgios@sutd.edu.sg (G.P.)

* Correspondence: kokiconst@aueb.gr

[†] Presented at the 3rd annual Decentralized Conference, Athens, Greece, 30 October–1 November 2019.

Published: 21 October 2019



Abstract: With Bitcoin, Ether and more than 2000 cryptocurrencies already forming a multi-billion dollar market, a proper understanding of their statistical and financial properties still remains elusive. Traditional economic theories do not explain their characteristics and standard financial models fail to capture their statistic and econometric attributes such as their extreme variability and heteroskedasticity. Motivated by these findings, we study Bitcoin and Ether prices via a Non-Homogeneous Pólya Gamma Hidden Markov (NHPG) model that has been shown to outperform its counterparts in conventional financial data. The NHPG algorithm has good in-sample performance and identifies both linear and non-linear effects of the predictors. Our results indicate that all price series are heteroskedastic with frequent changes between the two states of the underlying Markov process. In a somewhat unexpected result, the Bitcoin and Ether prices, although correlated, are significantly affected by different variables. We compare long term to short term Bitcoin data and find that significant covariates may change over time. Limitations of the current approach—as expressed by the large number of significant predictors and the poor out-of-sample predictions—back earlier findings that cryptocurrencies are unlike any other financial asset and hence, that their understanding requires novel tools and ideas.

Keywords: cryptocurrencies; blockchain; bitcoin; ethereum; non-homogeneous hidden markov models; model selection; forecasting

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

1. Introduction

A decade after the pseudonymous Satoshi Nakamoto [1] launched Bitcoin, the first digital coin to attract widespread public attention, and cryptocurrencies have become a multi-billion dollar market. Blockchain, the technology underlying Satoshi's decentralized monetary system, has been considered by many as the most revolutionising concept since the internet and is currently extensively researched by scientists, academics [2] and not the least by financial institutions and governments. With more than 2000 cryptocurrencies currently in circulation [3], and many more about to emerge, [4], the questions are still more than the answers.

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 assets 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 a conventional commodity, cf. [5]. The claim that BTC is fundamentally different from valuable metals like gold is also backed by [6] due to its shortage in stable hedging capabilities. Along with [5,7] 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 [8], 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, [9–11]. Using data from a longer period (between 2010 and 2017), Demir et al. [12] 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 [13–15]. 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, cf. [16,17]. 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—calls for a more comprehensive research approach. Despite the high documented correlation in the price of the various cryptocurrencies, [18], it is highly debated whether this trend will also continue into future or not, [19].

Summary & Results

In the present paper, we make an effort towards understanding price formation of the two largest cryptocurrencies, Bitcoin (BTC) and Ether (ETH), in terms of market capitalization. Our analysis uses a specific instance of the Non-Homogeneous Hidden Markov models, namely the Non-Homogeneous Pólya Gamma Hidden Markov model (NHPG) of [20], which has been shown to outperform similar models in conventional financial data, cf. [21]. The present model, cf. Section 2, falls into the Markov-Switching literature with two possible states that is the benchmark for predicting exchange rates, see [22–24]. It uses Bayesian Model Averaging (BMA) approach for inference which has been shown to possess desirable properties for forecasting applications, cf. [25–28].

Hidden Markov models have been shown to be a powerful tool in explaining financial data, [29]. Although standard in financial applications, this approach has only been applied in the cryptocurrency context by [30] as state space model and [31] in the context of price bubbles. Yet, the benefits of using Hidden Markov models to understand price formation of cryptocurrencies go beyond these applications. Their application is further supported by the specific characteristics of cryptocurrency data that have been identified by earlier researchers. For example, the non-stationarity of the BTC index and volume indicates the importance of modeling the non-linearities of the data, [12,32,33]. This is further elaborated by [15,25,34] who suggest that model selection and the use of averaging criteria are necessary to avoid poor forecasting results in view the cryptocurrencies' extreme and non-constant volatility. More importantly, Ciaian et al. [7] show that the Bitcoin price series exhibits structural breaks and suggest that significant price predictors may vary over time.

In the current setting, we adopt an economic/financial perspective and use a set of 11 financial and economic predictors comprising main exchange rates (4 variables), equity indices (3 variables), commodity prices (oil and gold) and economic uncertainty indicators (2 variables) along with 2 quasi-economic and 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 are summarized in Table 1. In the related cryptocurrency literature, subsets of these indices are studied under various settings, see e.g., [11,30,33–38].

This mixture of financial and macroeconomic indicators, economic uncertainty and volatility indices and finally, blockchain specific variables aims to achieve a comprehensive coverage of the forces that affect cryptocurrency prices. More specifically, while social sentiment and global financial advancements affect cryptocurrency prices to a certain degree, they do not capture the dependencies of price fluctuations to the hidden game theoretic foundations of the underlying protocols. The sustainability of a blockchain and hence, of the supported cryptocurrency depends not only on common financial considerations but more importantly, on the actions of the various involved actors, e.g., users, developers and miners, who interact in the presence of sometimes not perfectly aligned or conflicting profit maximization incentives. Understanding the influence of these incentives—which are not necessarily restricted to economic profits—on cryptocurrency markets is precisely the motivation for the blend of financial and blockchain specific variables that is employed in the current study. With this in mind, the questions that we aim to address are the following:

- Q1. Do the same explanatory variables affect both the BTC and ETH cryptocurrencies?
- Q2. What is the predictive power of the NHPG model on the BTC and ETH price series?
- Q3. Do the same explanatory variables affect the BTC price series both on the long and short run?

For questions Q1–Q2, we use daily data (for both the response and the explanatory variables) between 2016 and 2019, i.e., after the initial coin offer of ETH and a reasonable market-price adjustment period. For question Q3, we compare the BTC data of the whole 2013–2019 period to the 2016–2019 period (also used in Q1–Q2). As in most of the recent studies, we exclude the period up to 2013 which exhibits markedly different characteristics. To account for the non-stationarity of the price series, we use log-transformation of the data. Our aim is to contribute to the literature that studies the modeling and prediction of cryptocurrencies, [13,15,16,30,33,38,39], with an elaborate econometric model and to try to gain understanding in the statistical, econometric and financial properties of existing cryptocurrencies.

The findings of our experiments can be summarized as follows. 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). 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 frequent, thus account for the heteroskedacity of the series, cf. Figures 1–3. 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. On the other hand, the inclusion of NASDAQ in all three cases is satisfactory due to the relation of NASDAQ to technological advancements. It is also worth mentioning, that the hash rate is no more significant for modeling the BTC price when we restrict to the 2016–2019 period. 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 Figure 4, 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 rate, [40,41]. 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.

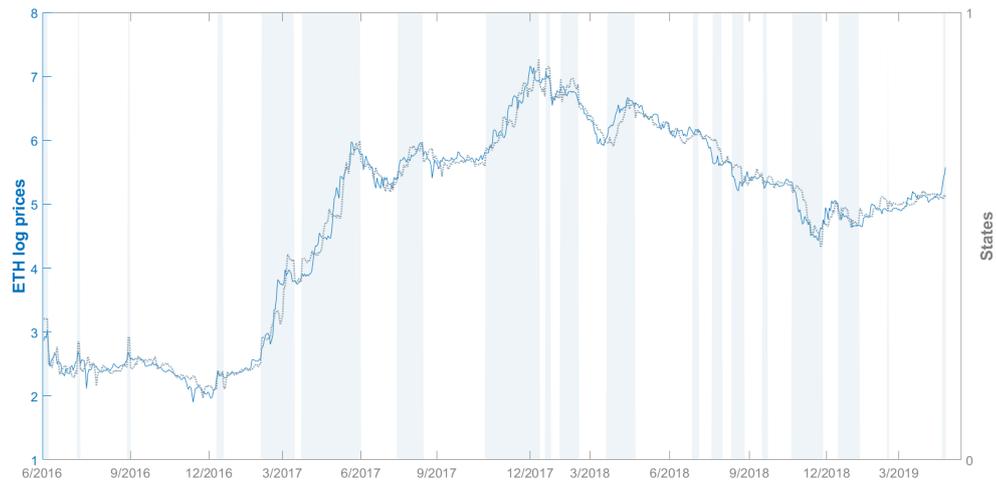


Figure 1. Logarithmic ETH price series (blue line) and in-sample estimated logarithmic ETH price series for the period 6/2016–5/2019 (gray dotted line). Shaded bars mark times with hidden state 1 (smoothed probability above 0.5). The model accounts for the heteroscedasticity of the series.

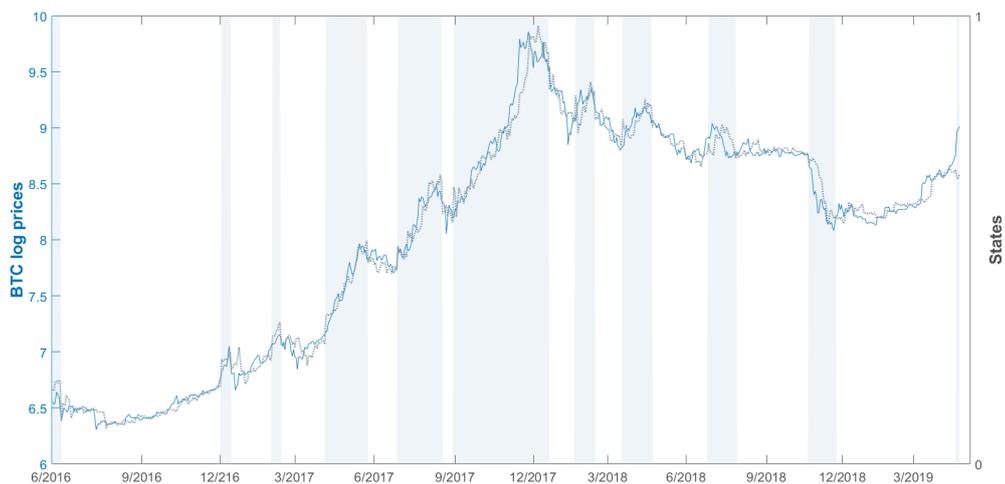


Figure 2. Logarithmic BTC prices series (blue line) and in-sample estimated logarithmic BTC price series for the period 6/2016–5/2019 (gray dotted line). Shaded bars mark times with hidden state 1 (smoothed probability above 0.5).

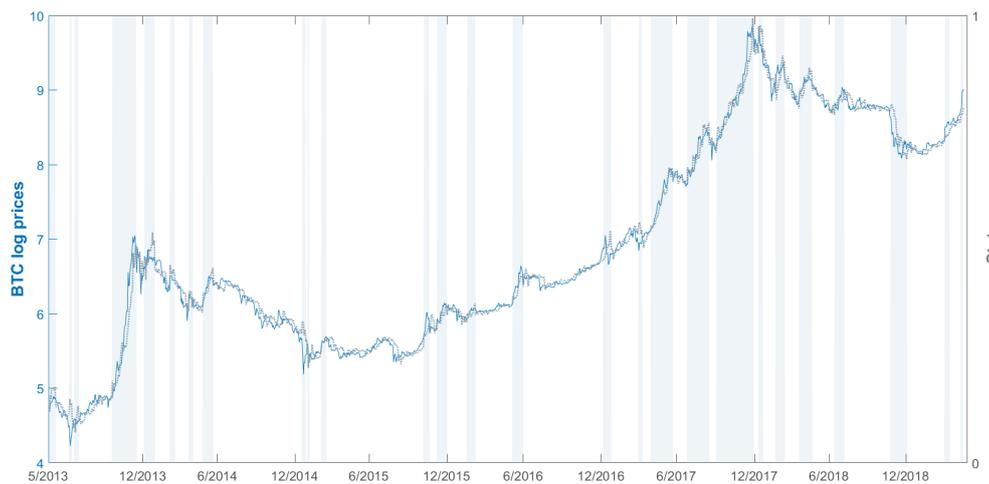


Figure 3. Logarithmic BTC prices series (blue line) and in-sample estimated logarithmic BTC price series for the period 5/2013–5/2019 (gray dotted line). Shaded bars mark times with hidden state 1 (smoothed probability above 0.5). The change of the sample sizes has a significant impact on the distribution of the unobserved process.

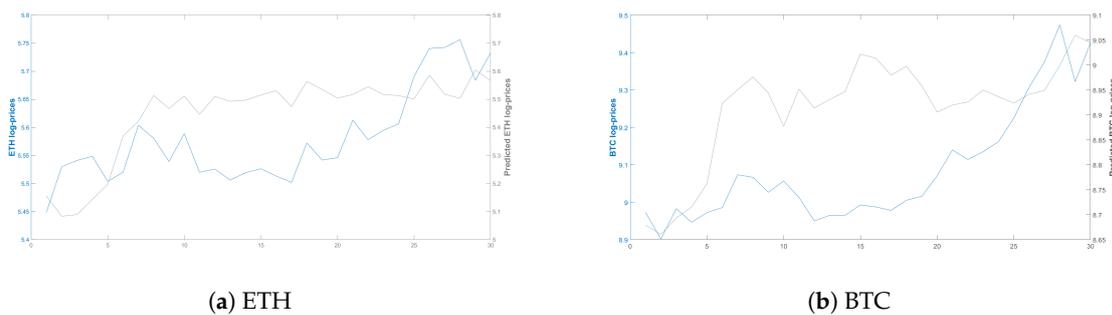


Figure 4. Mean posterior out-of-sample predictions (gray line) for $L = 30$ days both for the (a) ETH and (b) BTC log-transformed price series (blue line). While the predictions for ETH are better than those for BTC, both are not satisfactory as they frequently miss the direction of price movement. The BTC predictions are essentially the same for both the 2016–2019 and 2013–2019 data sets (the second not shown here).

2. Modeling Cryptocurrency Price Series

Given a time horizon $T \geq 0$ and discrete observation times $t = 1, 2, \dots, T$, 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 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 , $t = 1, \dots, T$, y_t depends on the current state z_t and not on the previous states. Consider also a set of $r - 1$ available predictors $\{X_t\}$ with realization $x_t = (1, x_{1t}, \dots, x_{r-1t})$ at time t . The explanatory variables (predictors) $\{X_t\}$ that are used in the present analysis are described in Table 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

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

where $B_s = (b_{0s}, b_{1s}, \dots, b_{r_1-1s})'$ are the regression coefficients and $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2 . The 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

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

where $\beta_{ij} = (\beta_{0,ij}, \beta_{1,ij}, \dots, \beta_{r_2-1,ij})'$ is the vector of the logistic regression coefficients to be estimated. 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

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

The Non-Homogeneous Pólya-Gamma Hidden Markov Model

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. We follow the methodology of [20]. In brief, the authors propose the following MCMC sampling scheme for joint inference on model specification and model parameters.

1. Given the model's parameters, the hidden states are simulated using the Scaled Forward Backward of algorithm of [42].
2. The posterior mean regression parameters are simulated using the standard conjugate analysis, via a Gibbs sampler method.
3. The logistic regression coefficients are simulated using the Pólya-Gamma data augmentation scheme [43], as a better and more accurate sampling methodology compared to the existing schemes.
4. The set of covariates that affect the model linearly and non-linearly (via the transition probabilities) are updated using a double reversible jump algorithm.
5. Predictions are made conditional on the simulated unknown quantities.

The steps 1–5 of the MCMC algorithm 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  $t = 1, \dots, T$  and  $i = 1, 2$  do
5:      $\pi_t(i | \theta) \leftarrow \frac{\alpha_t(s)}{\sum_{j=1}^2 \alpha_t(j)} = P(z_t = i | \theta, y^t)$  ( $\triangleright$ ) Simulation of the scaled forward probabilities
6:   for  $t = T, T - 1, \dots, 1$  do
7:      $z_t \leftarrow P(z_t | z_{t+1}) = \frac{p_{iz_{t+1}} \pi_t(i|\theta)}{\sum_{j=1}^m p_{jz_{t+1}} \pi_t(j|\theta)}$  ( $\triangleright$ ) 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 ( $\triangleright$ ) Conjugate analysis with Gibbs sampler
11:     $\beta | \sigma^2 \sim f_B, \sigma^2 \sim IG$  ( $\triangleright$ )  $f_B \equiv$  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 ( $\triangleright$ ) Pólya-Gamma data augmentation scheme
15:     augment the model space with  $\omega_s$  ( $\triangleright$ ) Conjugate analysis on the augmented space
16:     sample from  $\beta_s \sim f_{\beta_s|\omega}$ 
17:     and  $\omega_s | \beta_s \sim \mathcal{PG}$  ( $\triangleright$ ) 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  $X \cap X^{(i)c}$  ( $\triangleright$ ) Calculate acceptance probability  $\alpha$ 
22:     if  $\alpha < rand(0,1)$  then  $X^{(i)} \leftarrow X^{(i)} \cup X^{can}$ 
23:     remove: choose  $X^{candidate}$  from  $X^{(i)}$  ( $\triangleright$ ) Calculate acceptance probability  $\alpha$ 
24:     if  $\alpha < rand(0,1)$  then  $X^{(i)} \leftarrow X^{(i)} \cap X^{canc}$ 
25: procedure PREDICT
26:   % Make L-steps-ahead predictions
27:   for  $t = T + 1, \dots, T + L$  do
28:      $\hat{y}_t \sim f$  with  $f(y_{T+1} | y^T, z^T, \beta_M, \theta_M) = \sum_{s=1}^2 P(Z_{T+1} = s | Z_T = z_T) f_s(y_{T+1})$ .

```

3. The Data-Experiment

We assess the predictive ability of 11 financial/economic and 2 cryptocurrency specific variables, outlined in Table 1, in explaining and forecasting the prices of BTC and ETH via the NHPG model. We run the model using the prices of BTC and ETH and standardized explanatory variables.

However, as shown in Table 2 the estimated posterior variance of the two states is considerably high.

To stabilize the variance of the series, we perform a logarithmic transformation on the complete dataset. Due to space limitation, we only report the results of the log-transformation. We use values of the predictors lagged 7, along with an autoregressive term of lag 7. Since we have daily data, we use information of seven days prior to the studied day. We perform two experiments. In the first experiment, we use BTC and ETH prices for the same period—since the inception of Ether and after an initial market adjustment period. Specifically, we use daily data from 6/2016 until 6/2019, in which we excluded the data from the non-business days to synchronize the time series. From the data set, we keep 30 observations to perform our out-of-sample analysis and assess the forecasting ability of the model. In the second experiment, we compare the prices of BTC starting from 5/2013 until 6/2019 to the smaller dataset of the previous experiment, i.e., the dataset from 6/2016 until 6/2019.

The closing BTC prices were downloaded from coinmarket.com and the ETH prices were downloaded from etherscan.io.

Table 1. List of variables and online resources. The Hash Rate (HR) and Average Block Size (AVS) have been retrieved from quandl.com for Bitcoin and from etherscan.io for Ether.

Explanatory Variables		
Description	Symbol	Retrieved from
US dollars to Euros exchange rate	USD/EUR	investing.com
US dollars to GBP exchange rate	USD/GBP	investing.com
US dollars to Japanese Yen exchange rate	USD/JPY	investing.com
US dollars to Chinese Yuan exchange rate	USD/CNY	investing.com
Standard & Poor’s 500 index	SP500	finance.yahoo.com
Dow Jones Industrial Average	DOW	finance.yahoo.com
NASDAQ Composite index	NASDAQ	finance.yahoo.com
Crude Oil Futures price	CO	finance.yahoo.com
Price of Gold	GOLD	finance.yahoo.com
CBOE Volatility index	VIX	finance.yahoo.com
Equity market related Economic Uncertainty index	EUI	fred.stlouisfed.org
Hash Rate	HR	quandl.com/etherscan.io
Average Block Size	AVS	quandl.com/etherscan.io

Table 2. Measures of in-sample and out-of-sample estimation: mean posterior variance of the two states and mean square forecast error for the three datasets.

Sample period	Mean Posterior Variance		
	BTC 4/2013–6/2019	BTC 6/2016–6/2019	ETH 6/2016–6/2019
σ_1^2	0.001×10^6	2.63×10^6	7.05×10^3
σ_2^2	1.49×10^6	0.03×10^6	0.002×10^3
Mean Square Forecast Error			
MSFE	4.528×10^6	7.986×10^6	2.756×10^4

4. Results

Figures 1–3 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 the heteroscedasticity of the series.

However, the out-of-sample performance of the NHPG is poor. Figure 4, 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. 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 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 [40,41].

In Table 3, 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, [18], the variables that affect the series (by means of the MPM) are not the same. The MPM of ETH consists of 8 covariates: the USD/EUR and USD/CNY exchange rates, the price of Crude Oil and Gold, the VIX and NASDAQ indices and both quasi-economic variables HR and AVS. The MPM of BTC additionally contains the USD/GBP and USD/JPY exchange rates. This difference on the exchange rates indicates that ETH is more geographically restricted. Moreover, while the hash rate (HR) is significant for the relatively newer Ethereum blockchain, it is not anymore significant for the BTC blockchain when we restrict to the more recent 2016–2019 period.

Table 3. 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						
Predictors	BTC		BTC		ETH	
Sample period	4/2013 - 6/2019		6/2016-6/2019		6/2016-6/2019	
USD/EUR	0.65	0.39	0.72	0.50	0.58	0.50
USD/GBP	1.00	0.36	0.62	0.48	0.47	0.48
USD/JPY	1.00	0.17	0.59	0.27	0.36	0.22
USD/CNY	1.00	0.39	0.81	0.44	0.67	0.36
CO	1.00	0.07	0.97	0.24	1.00	0.15
VIX	1.00	0.07	0.70	0.16	1.00	0.12
SP500	0.46	0.13	0.42	0.20	0.47	0.18
DOW	0.95	0.08	0.48	0.16	0.45	0.11
NASDAQ	1.00	0.13	0.78	0.23	0.77	0.11
GOLD	1.00	0.12	0.97	0.32	1.00	0.13
EUI	0.05	0.01	0.07	0.01	0.00	0.00
HR	0.55	0.02	0.32	0.22	1.00	0.01
AVS	1.00	0.02	1.00	0.01	0.57	0.06

Concerning the extended dataset of BTC, only the EUI and SP500 variables are excluded from the MPM. Hence, the explanatory variables that affect the log BTC series are not the same on the long and on the short run. From the market indices, only NASDAQ is found to be significant in all 3 time series. The high inclusion probabilities of this index are supported by the fact that NASDAQ is an index that consists mostly of technology and AI firms. The inclusion of the VIX in all three MPM’s confirms the relationship between cryptocurrency price formation and volatility in conventional financial markets. Finally, it is worth mentioning that all the effects are linear with only the USD/EUR marginally making it to the transition probabilities equation. The inclusion of at least one variable 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 there are other drives that drive changes in the underlying process that go beyond the financial aspects that have been considered in the present setting.

5. Conclusions

Using the Non-Homogeneous Pólya-Gamma Hidden Markov Model (NHPG) of [20] on the Bitcoin (BTC) and Ether (ETH) price series and 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, the hash rate was not found significant for BTC in the period 2016–2019, while NASDAQ was the only equity index to significantly affect both currencies. 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, [44]. 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. It can be extended to comprise more hidden states, high frequency data, a larger set of covariates and the aforementioned price formation Deep Learning methods. Via the inclusion of selected blockchain specific metrics—quantified technological advancements and transaction graph analysis—future work may further disentangle the underlying game theoretic interactions and dynamics that influence cryptocurrencies.

Funding: Stefanos Leonardos and Georgios Piliouras were supported by MOE AcRF Tier 2 Grant 2016-T2-1-170 and in part by the National Research Foundation (NRF), Prime Minister’s Office, Singapore, under its National Cybersecurity R&D Programme (Award No. NRF2016NCR-NCR002-028) and administered by the National Cybersecurity R&D Directorate. Georgios Piliouras acknowledges SUTD grant SRG ESD 2015 097 and NRF 2018 Fellowship NRF-NRFF2018-07.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System. 2008. Available online: <https://bitcoin.org/bitcoin.pdf> (accessed on 30 July 2019).
2. Garay, J.; Kiayias, A.; Leonardos, N. The Bitcoin Backbone Protocol: Analysis and Applications. In *Advances in Cryptology—EUROCRYPT 2015*; Oswald, E., Fischlin, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2015; pp. 281–310.
3. Crypto.com. 2019. Available online: <https://coinmarketcap.com/all/views/all/> (accessed on 30 July 2019).
4. Paper, L.W. 2019. Available online: <https://libra.org/en-US/white-paper/?noredirect=en-US> (accessed on 30 July 2019).

5. 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.
6. 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.
7. Ciaian, P.; Rajcaniova, M.; Kancs, A. The economics of BitCoin price formation. *Appl. Econ.* **2016**, *48*, 1799–1815. doi:10.1080/00036846.2015.1109038.
8. 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.
9. 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.
10. Bouri, E.; Azzi, G.; Dyhrberg, A.H. On the return-volatility relationship in the Bitcoin market around the price crash of 2013. *Economics* **2017**, *11*, 1–16. doi:10.5018/economics-ejournal.ja.2017-2.
11. 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.
12. 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.
13. 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.
14. Hayes, A.S. Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. *Telemat. Inf.* **2017**, *34*, 1308–1321. doi:10.1016/j.tele.2016.05.005.
15. 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.
16. Georgoula, I.; Pournarakis, D.; Bilanakos, C.; Sotiropoulos, D.; Giaglis, G.M. Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices. *MCIS* **2015**. doi:10.2139/ssrn.2607167.
17. Colianni, S.G.; Rosales, S.M.; Signorotti, M. *Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis*; CS229 Project; Stanford University: Stanford, CA, USA, 2015.
18. Borri, N. Conditional tail-risk in cryptocurrency markets. *J. Empir. Financ.* **2019**, *50*, 1–19. doi:10.1016/j.jempfin.2018.11.002.
19. Vora, R. Ethereum Price Analysis: Ethereum (ETH) Needs To Discover The Magic Spell To Surge On Its Own. 2019. Available online: <https://www.cryptonews.com/ethereum-price-analysis-ethereum-eth-needs-to-discover-the-magic-spell-to-surge-on-its-own/29402/> (accessed on 27 July 2019).
20. Koki, C.; Meligkotsidou, L.; Vrontos, I. Forecasting under model uncertainty: Non-homogeneous hidden Markov models with Polya-Gamma data augmentation. *arXiv* **2019**, arXiv:1802.02825.
21. Meligkotsidou, L.; Dellaportas, P. Forecasting with non-homogeneous hidden Markov models. *Stat. Comput.* **2011**, *21*, 439–449. doi:10.1007/s11222-010-9180-5.
22. 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.
23. 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.
24. 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.
25. 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.
26. 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.
27. Wright, J.H. Forecasting US inflation by Bayesian model averaging. *J. Forecast.* **2009**, *28*, 131–144. doi:10.1002/for.1088.
28. Wright, J.H. Bayesian Model Averaging and exchange rate forecasts. *J. Econ.* **2008**, *146*, 329–341. doi:10.1016/j.jeconom.2008.08.012.

29. Mamon, R.; Elliott, R. (Eds.) *Hidden Markov Models in Finance: Further Developments and Applications, Volume II*; International Series in Operations Research & Management Science; Springer: Berlin/Heidelberg, Germany, 2014.
30. Poyser, O. Exploring the dynamics of Bitcoin's price: a Bayesian structural time series approach. *Eurasian Econ. Rev.* **2019**, *9*, 29–60. doi:10.1007/s40822-018-0108-2.
31. 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; pp. 1–7. doi:10.1109/SSCI.2017.8280809.
32. Balciar, 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.
33. 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.
34. 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.
35. Van Wijk, D. What can be expected from the BitCoin. Ph.D. Thesis, Erasmus Universiteit Rotterdam, Rotterdam, The Netherlands, 2013.
36. 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.
37. Estrada, J.C.S. Analyzing Bitcoin Price Volatility. Ph.D. Thesis, University of California, Berkeley, CA, USA, 2017.
38. 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.
39. McNally, S.; Roche, J.; Caton, S. Predicting the Price of Bitcoin Using Machine Learning. In Proceedings of the 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), Cambridge, UK, 21–23 March 2018; pp. 339–343. doi:10.1109/PDP2018.2018.00060.
40. Yuan, C. Forecasting exchange rates: The multi-state Markov-switching model with smoothing. *Int. Rev. Econ. Financ.* **2011**, *20*, 342–362. doi:10.1016/j.iref.2010.09.002.
41. 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.
42. Scott, S.L. Bayesian Methods for Hidden Markov Models. *J. Am. Stat. Assoc.* **2002**, *97*, 337–351. doi:10.1198/016214502753479464.
43. 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.
44. 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.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).