Received: 20 February 2021 | Accepted: 6 October 2021

DOI: 10.1111/ecno.12193

ORIGINAL ARTICLE

WILEY

Exploring the dependencies among main cryptocurrency log-returns: A hidden Markov model

Fulvia Pennoni[1] | **Francesco Bartolucci**[2] | **Gianfranco Forte**[3] | **Ferdinando Ametrano**[4]

[1]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milano, Italy

[2]Department of Economics, University of Perugia, Perugia, Italy

[3]Department of Business and Law and Crypto Asset Lab, University of Milano-Bicocca, Milano, Italy

[4]Crypto Asset Lab and Department of Business and Law, University of Milano-Bicocca, Milano, Italy

Correspondence

Fulvia Pennoni, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8, Milano 20126, Italy.
Email: fulvia.pennoni@unimib.it

Abstract

A hidden Markov model is proposed for the analysis of time-series of daily log-returns of the last 4 years of Bitcoin, Ethereum, Ripple, Litecoin, and Bitcoin Cash. These log-returns are assumed to have a multivariate Gaussian distribution conditionally on a latent Markov process having a finite number of regimes or states. The hidden regimes represent different market phases identified through distinct vectors of expected values and variance–covariance matrices of the log-returns, so that they also differ in terms of volatility. Maximum-likelihood estimation of the model parameters is carried out by the expectation–maximisation algorithm, and regimes are singularly predicted for every time occasion according to the maximum-a-posteriori rule. Results show three positive and three negative phases of the market. In the most recent period, an increasing tendency towards positive regimes is also predicted. A rather heterogeneous correlation structure is estimated, and evidence of structural medium term trend in the correlation of Bitcoin with the other cryptocurrencies is detected.

KEYWORDS

Bitcoin, Bitcoin cash, decoding, Ethereum, expectation-maximisation algorithm, Litecoin, Ripple, time-series

1 | INTRODUCTION

Following the seminal paper of Satoshi Nakamoto (2008) and the creation of the Bitcoin network in 2009, an increasing number of crypto-assets have appeared. Almost all are of little interest, being just clones of the first without any real functional innovation and/or trading liquidity. A few exceptions exist that are relevant enough to be considered as investable assets. Therefore, crypto-assets time-series nowadays consist of multidimensional and complex data, and these assets represent the most volatile and challenging financial market (Borri, 2019).

Cryptocurrencies are peculiar assets in terms of both their origin and operating mechanism. Phenomena such as digital scarcity, periodic halving, emergence of many cryptocurrencies trying to replicate the success of Bitcoin, which remains the centre of this universe, and the absence of regulation, contribute to make crypto-assets unique. Many papers analysed the behaviour of the prices of a single crypto; among others, Chan et al. (2017), Osterrieder (2017), and Szczygielski et al. (2020) investigated the distributional properties, showing a considerable and expected distance from the assumption of a Gaussian distribution. Bariviera et al. (2017) found long-term memory in the Bitcoin time-series only up to 2014 highlighting a convergence process towards efficiency, while Zhang et al. (2018) extended their analysis to the time-series variance, identifying long memory.

Many other empirical profiles of cryptocurrencies have been investigated and deserve attention in their high risk/reward profile. In particular, it is worth mentioning the works focusing on the efficiency of the cryptocurrency market and its speculative traits resulting in periodic bubbles. Efficiency side is of particular interest, given the youthfulness of this market. Some notable works are Nadarajah and Chu (2017), Almudhaf (2018), Alvarez-Ramirez et al. (2018), and Tiwari et al. (2018). Applying different techniques, they shed light on the inefficiency of Bitcoin in weak form, while pointing out a marked trend towards greater efficiency.

The speculative nature of cryptocurrencies is a key feature of this market; without this and the resulting hype and stellar returns, the experiment in decentralised digital currencies would have been short-lived. About this issue Cheung et al. (2015), Cheah and Fry (2015), Corbet et al. (2018), and Agosto and Cafferata (2020) found strong evidence of periodic bubbles through an analysis of the data based on the approaches adopted in Phillips et al. (2011) and Phillips et al. (2015), which rely on the variants of the augmented Dickey–Fuller test, and on modelling the cost of mining as in Xiong et al. (2020). The latter also showed that it is possible to predict these bubbles with a remarkable level of accuracy.

In this study we consider five main cryptocurrencies and monitor their log-returns jointly. We avoid going into the debate on the representativeness of the different cryptocurrencies and we focus on the market data referred to Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Litecoin (LTC), and Bitcoin Cash (BCH). This choice is mainly related to their trading volume, as will be further explained in the following. We propose a popular unsupervised statistical method based on a multivariate Hidden Markov (HM) model; see Cappé et al. (1989), Mamon and Elliott (2007), and Zucchini et al. (2017) for details on the model in the context of time-series data and Bartolucci et al. (2013) in the context of longitudinal data. This model may be seen as an extension of the finite mixture model (McLachlan & Peel, 2000) with a particular dependence structure across variables referred to different time points. The HM approach provides a flexible framework for many financial applications and it allows incorporating stochastic volatility in a rather simple form. From the pioneering work of Akaike (1998), showing that an ARMA process can be represented by a Markovian structure, many other extensions have been proposed in the literature. Hamilton (1989) was the first to formulate a model where the latent regime follows a Markov process, and many other works following this proposal appeared in the literature; see, among others, Genon-Catalot et al. (2000), Bartolucci and De Luca (2003), Rossi and Gallo (2006), Mamon and Elliott (2007), Langrock et al. (2012), Giudici and Abu Hashish (2020), and Lin et al. (2020).

Unlike the prevailing literature, in which applications of switching models are focused exclusively on the estimation and prediction of volatility and the expected log-returns are considered as unpredictable (see, among others Ang & Bekaert, 2002), we proceed as proposed in De Angelis and Paas (2013) by also modelling the conditional means of the time-series. In particular, stable periods, crises, and financial bubbles differ significantly for mean returns and structural levels of covariance. Furthermore, in line with the most recent literature, we model the log-returns of crypto-assets considering their correlation structure. In fact the study of interconnectedness, comovements, and volatility spillovers between cryptocurrencies has received considerable attention in recent research; see Corbet et al. (2018), Giudici and Polinesi (2019), and Chen et al. (2020) for a network analysis; Yi et al. (2018) and Giudici and Pagnottoni (2019, 2020) for a Value-at-Risk analysis; Katsiampa et al. (2019) for a multivariate GARCH analysis; and Sifat et al. (2019) for the ARMA analysis. Considering multiple time-series jointly instead of a single series as proposed in Huang et al. (2019) allow us to account for the sideway movements in the long-term trends and to identify actual trend change signals in the market.

The multivariate HM assumes that the daily log-returns of the five mentioned cryptos comes from a specific probabilistic distribution associated with the hidden regimes. The expectation-maximisation (EM) algorithm (Baum et al., 1970; Dempster et al., 1977; Welch, 2003) is employed for maximum-likelihood estimation of the model parameters. The conditional distributions of the observed log-returns for the states of the hidden variables are taken from the Gaussian family with different means, variances, and covariances. This study is a first attempt to estimate a multivariate model of the main cryptocurrencies using the market prices collected from 3 August 2017, to 30 June 2021. We shed light on the structure of their connection, simultaneously seizing the dynamics of the regimes. The proposal is consistent with some of the stylised facts typical of financial time-series (Rydén & Teräsvirta, 1998) and also accounts for the structural links between assets (Dias et al., 2015).

The remainder of the paper is organised as follows. In Section 2 we define the notation and the main assumptions of the HM model, as well as its maximum-likelihood estimation via the EM algorithm. In Section 3 we describe the data and the reference market. In Section 4 we show the main results. In Section 5 we provide some conclusions.

2 | MULTIVARIATE HIDDEN MARKOV MODEL

We consider multiple time-series with data collected in vectors $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$, where each element y_{tj}, with $t = 1, 2, \ldots$, $j = 1, \ldots, r$, corresponds to the log-return of asset j among those considered. We will use $\boldsymbol{y}_t$ to denote either the random vector at time t or one of its realisations in a way that will be clear from the context; the same convention will be applied to scalar random variables. In the following, we first describe the HM model assumptions for our specific formulation and then outline the steps of the EM algorithm for its estimation.

The main assumption of the HM model is that the random vectors $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$ are conditionally independent given a hidden process $u_1, u_2, \ldots$ that follows a Markov chain with k hidden states, labelled from 1 to k. The model includes two different submodels, named measurement and structural models, which are described in more detail in the following. The measurement model corresponds to the conditional distribution of every vector $\boldsymbol{y}_t$ given the underlying variable u_t, $t = 1, 2, \ldots$. In this regard, we assume a multivariate Gaussian distribution for the overall log-returns of the criptocurrencies, that is,

$$\boldsymbol{y}_t | u_t = u \sim N_r(\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u), \quad u = 1, \ldots, k,$$

where $\boldsymbol{\mu}_u$ and $\boldsymbol{\Sigma}_u$ are, for hidden state u, the specific mean vector and variance–covariance matrix, respectively. Obviously, the conditional means in $\boldsymbol{\mu}_u$ define the expected log-returns when the underlying chain is in state u, while the elements of $\boldsymbol{\Sigma}_u$ provide measures of volatility of each asset and correlation between pairs of assets. Different constraints may be conceived on these matrices that can also be useful to test some hypotheses, the main of which is that of homoskedasticity: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_u$, $u = 1, \ldots, k$. Other formulations could be adopted (Banfield & Raftery, 1993;

Celeux & Govaert, 1995). Denoting by T the last observation time, the previous assumptions imply that the conditional distribution of the time-series given the sequence of hidden states may be expressed as follows:

$$f(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T, |u_1, \ldots, u_T) = \prod_{t=1}^{T} \phi(\boldsymbol{y}_t; \boldsymbol{\mu}_{u_t}, \boldsymbol{\Sigma}_{u_t}),$$

where, in general, $\phi(\cdot;\cdot,\cdot)$ denotes the density of the multivariate Gaussian distribution, in our case of dimension r, with a certain mean vector and variance–covariance matrix.

The structural model for the distribution of the hidden Markov process is based on two sets of parameters, which are the initial and transition probabilities. They are denoted by $\lambda_u = p(u_1 = u)$, $u = 1, \ldots, k$, and $\pi_{(v|u)} = p(u_t = v|u_{t-1} = u)$, $u, v = 1, \ldots, k$, respectively, and are collected in the initial probability vector $\boldsymbol{\lambda}$ and in the transition matrix $\boldsymbol{\Pi}$. Joint together, the measurement and structural models imply that the manifest distribution of the time-series has a density that is computed by a forward recursion (Baum et al., 1970; Welch, 2003). This recursion requires a number of operations that linearly increases with the number of observation times needed to exploit the factorisation of the manifest distribution.

Maximum-likelihood estimation of the model based on the previous assumptions is considered. With reference to the observed time-series of log-returns, the log-likelihood function of this model is defined as follows:

$$\ell(\boldsymbol{\theta}) = \log f(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T),$$

where $\boldsymbol{\theta}$ is the vector of all model parameters, that is, $\boldsymbol{\mu}_u$, $\boldsymbol{\Sigma}_u$, for $u = 1, \ldots, k$, and $\boldsymbol{\lambda}$, and $\boldsymbol{\Pi}$. By maximising $\ell(\boldsymbol{\theta})$ we estimate these parameters and, for this aim, we employ the EM algorithm. The latter is based on the so-called complete-data log-likelihood denoted by $\ell^*(\boldsymbol{\theta})$, which may be expressed as the sum of three components that are maximised separately. The EM algorithm alternates two steps until convergence in $\ell(\boldsymbol{\theta})$: *E-step*: compute the conditional expected value of $\ell^*(\boldsymbol{\theta})$, given the observed data and the value of the parameters at the previous step; *M-step*: maximise the expected value of $\ell^*(\boldsymbol{\theta})$ obtained from the E-step and then update the model parameters. In particular, the parameters of the measurement model are updated in a simple way as follows:

$$\boldsymbol{\mu}_u = \frac{1}{\sum_{t=1}^{T} \hat{w}_{tu}} \sum_{t=1}^{T} \hat{w}_{tu}\, \boldsymbol{y}_t,$$
$$\boldsymbol{\Sigma}_u = \frac{1}{\sum_{t=1}^{T} \hat{w}_{tu}} \sum_{t=1}^{T} \hat{w}_{tu} (\boldsymbol{y}_t - \boldsymbol{\mu}_u)(\boldsymbol{y}_t - \boldsymbol{\mu}_u)',$$

for $u = 1, \ldots, k$, where $\hat{w}_{tu} = p(u_t = u|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$ is the posterior probability that the time-series is in state u at occasion t. Regarding the parameters of the structural model, we simply have

$$\lambda_u = \hat{w}_{1u}, \qquad u = 1, \ldots, k,$$
$$\pi_{v|u} = \frac{1}{\sum_{t=2}^{T} \hat{w}_{t-1,u}} \sum_{t=2}^{T} \hat{z}_{tuv}, \qquad u, v = 1, \ldots, k,$$

where $\hat{z}_{tuv} = p(u_{t-1} = u, u_t = v|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$. The overall vector of estimates obtained at convergence is denoted by $\hat{\boldsymbol{\theta}}$.

Since the EM algorithm may converge to a local maximum not corresponding to the global maximum, common initialisation strategies involve a multistart rule from appropriate deterministic and random starting values. Deterministic starting values of the parameters of the measurement model, $\boldsymbol{\mu}_u$ and $\boldsymbol{\Sigma}_u$, $u = 1, \ldots, k$, are defined on the basis of descriptive statistics (mean vector and variance–covariance matrix) of the observed log-returns. The starting values for the initial probabilities λ_u are chosen as $1/k$, for $u = 1, \ldots, k$, whereas for the transition probabilities we adopt the following rule: $\pi_{v|u} = (h+1)/(h+k)$ when $v = u$ and $\pi_{v|u} = 1/(h+k)$ when $v \neq u$, where h is a suitable positive constant (e.g., an integer from 5 to 10). The random starting rule is instead based on values drawn from a multivariate Gaussian distribution for $\boldsymbol{\mu}_u$, $u = 1, \ldots, k$, and on normalised random numbers drawn from a uniform distribution between 0 and 1 for initial and transition probabilities. The starting values for the variance–covariance matrices are again based on their sample counterpart. Overall, for a given k, inference is based on the solution corresponding to the largest value of the log-likelihood at convergence, which typically is the global maximum.

A further important aspect concerns model selection in terms of choice of the appropriate number of regimes, denoted by k. When there are not substantial reasons to assume a predefined value for k, we rely on the Bayesian information criterion (BIC; Schwarz, 1978), which is based on the following index:

$$\mathrm{BIC}_k = -2\hat{\ell}_k + \log(T)\#\mathrm{par}_k, \tag{1}$$

where $\hat{\ell}_k$ denotes the maximum of the log-likelihood of the model with k states and $\#\mathrm{par}_k$ denotes the number of free parameters equal to $k[r + r(r+1)/2] + k^2 - 1$ for the heteroskedastic model. Another common criterion is the Akaike information criterion (AIC, Akaike, 1973), which is based on the following index

$$\mathrm{AIC}_k = -2\hat{\ell}_k + 2\#\mathrm{par}_k.$$

It is worth mentioning that other indices could be considered for selecting the model that represents the best compromise between goodness-of-fit and complexity. Among them we mention the corrected AIC (AIC_c), proposed in Hurvich and Tsai (1993), and defined with a different penalisation term:

$$\mathrm{AIC}_{c,k} = -2\hat{\ell}_k + 2\frac{\#\mathrm{par}_k(\#\mathrm{par}_k - 1)}{n - \#\mathrm{par}_k - 1}.$$

In practice we estimate a series of HM models for increasing value k and we select the number of hidden states corresponding to the minimum value of the chosen index.

Among the previous ones, it is worth noting that, once the maximum-likelihood estimates are obtained, prediction of the most likely sequence of hidden states can be performed through the so-called *local decoding*. This is a maximum-a-posteriori prediction based on probabilities obtained from the EM algorithm. *Global decoding* may also be performed through the Viterbi algorithm (Juang & Rabiner, 1991; Viterbi, 1967) that is able to predict the entire sequence of latent states. Computational tools required for the estimation are implemented by adapting suitable functions of package LMest (Bartolucci et al., 2017) in R (R Core Team, 2021); they are available upon request from the authors.

3 | DATA AND REFERENCE MARKET

The crypto-asset market lacks regulation and established best practices, and it is quite common to have manipulated asset prices and trading volumes. For these reasons, suitable criteria must be used to select crypto-assets for quantitative analyses. Approaches based on the so-called market capitalisation are unreliable because it is not easy to define the real free-floating capital for every asset. Descriptive statistics in the market indices are available, but they are generally not relevant from an economic point of view. In fact, due to the anonymity/pseudonymity of the blockchain protocol, it is difficult to determine the amount of capital lost forever, making the calculations of the actual capital available for trading almost always impossible.

The adopted criteria to select the cryptocurrencies for the illustrative example are the same underlying the Crypto Asset Lab Index defined by the Crypto Asset Lab (CAL).[1] Each crypto-assets must satisfy the following requirements: (i) being a scarce digital bearer asset that; (ii) having a value that is not pegged to any other asset or currency; (iii) being traded on at least two reliable exchanges;[2] (iv) having no more than 80% of its combined 90-day trading volume on a single reliable exchange; and (v) being actively traded on reliable exchanges against traditional fiat currencies, stablecoins (i.e., crypto-assets pegged to fiat currencies), BRC, or ETH.

In light of the above remarks, we consider BTC, ETH, XRP, LTC, and BCH focusing on a recent time span of almost 4 years for the seek of comparability on the liquidity side. In this way, we account for the recent introduction and development of XRP and LTC. The data provided by the CAL are referred to daily quotes from 3 August 2017, to 30 June 2021; these data are available upon request from the CAL. Prices are collected daily every midnight, and they come from eight exchanges among the 81 currently active ones, namely: BitFlyer, BitStamp, Bittrex, Coinbase, Gemini, itBit, Kraken, and Poloniex. The selection of these exchanges is justified by their relevance, reliability, and ethical standards. It is known that the cryptocurrency market is affected by distinct phenomena of volume manipulation aimed at attracting customers displaying false liquidity. Therefore, we have followed exclusion criteria based on manipulation, taken also into account by the CAL for the design of its index, and similar to those of Bitwise submitted to the Securities and Exchange Commission (SEC).[3]

Overall, we dispose of a series of $T = 1{,}429$ log-returns for $r = 5$ assets. Figure 1 shows the daily prices of the five cryptocurrencies over the entire period of observation. It is immediate to recognise the high volatility that characterises this market and the waves of exponential price increases. The first wave started in mid-2017, the first big "bubble" that brought BTC above 20k USD, another one in mid-2019, and the most recent one started in late 2020. Figure 2 depicts the daily log-returns of the five cryptocurrencies, confirming their high volatility, a strong degree of comovement, and showing the typical volatility clustering common to financial assets.

Table 1 reports the observed correlations and partial correlations. The correlations are above 0.5, and very high for the pairs BTC-ETH, ETH-LTC, and LTC-BCH. However, the partial correlation structure is not so obvious to interpret, suggesting that the BTC dominance does not necessarily result in a unique comoving driver.

4 | RESULTS

Results for the HM model estimated through the procedure outlined in Section 2 show that the minimum value of the Bayesian information criterion (BIC) is reached considering a six-state heteroskedastic structure. The estimation is performed for a number of states ranging from 1 to 7, and Table 2 shows for each of these models, the maximum-log-likelihood ($\hat{\ell}_k$), the number of free parameters ($\#\text{par}_k$), the BIC_k, AIC_k, and $\text{AIC}_{c,k}$ values. The sequence of BIC values leads to selecting an HM model with $k = 6$ latent states.

[1]The Crypto Asset Lab (CAL) is an independent academic lab established at the University of Milano-Bicocca; for more details, see the web page at https://cryptoassetlab.diseade.unimib.it/. CAL developed an index (CAL Index) of the cryptocurrency market in 2019 which is reliable in terms of data inclusion criteria and independence. The price of each cryptocurrency originates from a double weighting scheme for trading volumes. The first step is to convert the prices of each cryptocurrency expressed in EUR, CAD, GBP, JPY into USD (using the European Central Bank [ECB] exchange rates) to have all cryptocurrency pairs expressed in one currency. Then the price of every coin for each of the eight exchanges stems from a volume-based weighted average price in each fiat currency (EUR, CAD, GBP, JPY, and USD). The final price is then calculated by taking the volume-weighted average of the single prices on the different exchanges.

[2]An exchange is considered reliable when it: (i) has not been charged for publishing false or inflated trading volumes; (ii) is registered and has obtained the license to operate in its jurisdiction; (iii) provides open and reliable functioning application programming interface; and (iv) applies trading fees.

[3]The presentation is available at the following link https://www.sec.gov/comments/sr-nysearca-2019-01/srnysearca201901-5164833-183434.pdf. We also note that the selected exchanges coincide with those used by the Chicago Mercantile Exchange (CME) to calculate BTC futures prices.

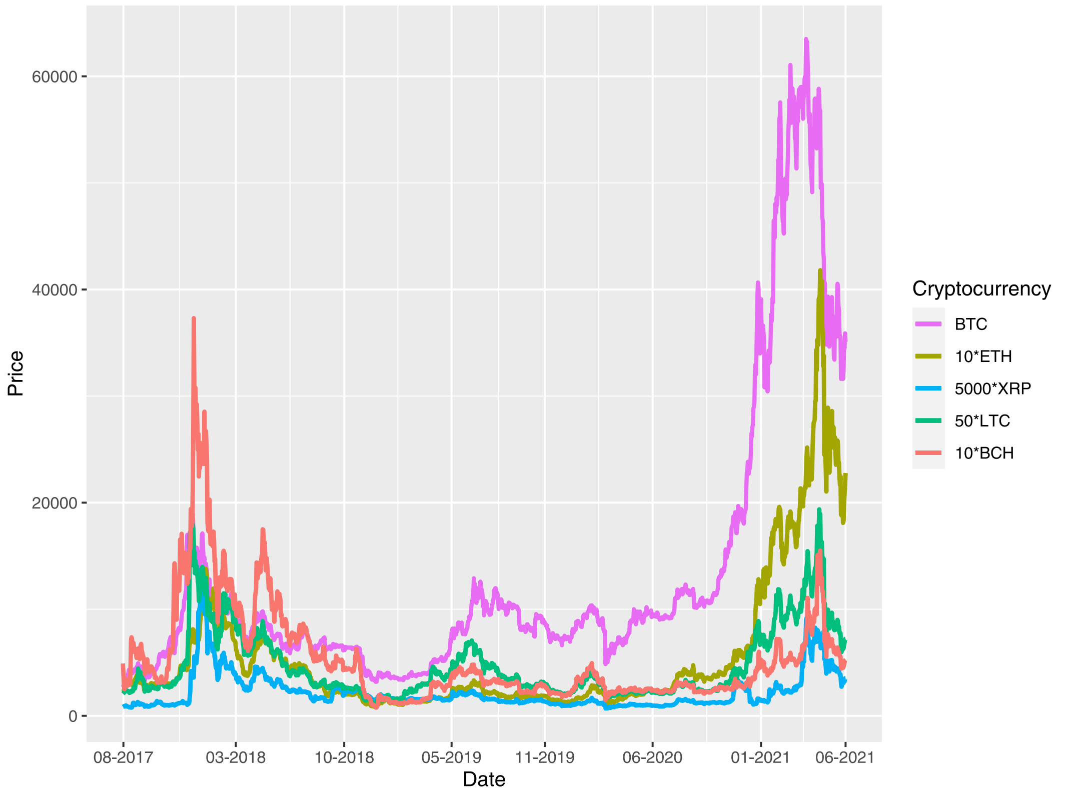

FIGURE 1 Daily time-series of prices of BTC, ETH, XPR, LTC, and BCH. Complete observations are referred to the period from 3 August 2017, to 30 June 2021 (ETH, XPR, LTC, and BCH prices have been appropriately scaled for comparison). BCH, Bitcoin cash; BTC, Bitcoin; ETH, Ethereum; LTC, Litecoin; XRP, Ripple

4.1 | Hidden Markov model with six regimes

The vectors of the expected conditional log-returns estimated under the selected model with k = 6 regimes are shown in Table 3. According to the elements of such vectors, three negative regimes (1, 2, 3) and three positive regimes (4, 5, 6) are estimated, which represent a variety of situations in the market. The first relevant evidence that emerges is that, when looking at cryptocurrencies as a whole, their average return from State 1 to 6 is increasing in a monotonic and exponential trend. This is in agreement with the evidence of speculative bubbles as discussed in Section 1. Considering cryptocurrencies individually, however, this regularity is not strictly observed. Regime 3 shows a negative overall average return; however, three of the five cryptocurrencies have positive returns, meaning that this regime reflects a more mixed pattern rather than a common negative expectation.

The results in Table 4, characterise each regime in terms of estimated conditional variances, correlations, and partial correlations. In particular, Regimes 3 and 5 are the most volatile among those with negative and positive average log-returns, respectively. Considering the correlations, these are very high for all regimes, with the only exception of Regime 6, which is also the one characterised by the highest average returns. Looking at the partial correlations, ETH and XRP, often show a divergent behaviour with respect to the other cryptocurrencies.

Table 5 shows the estimated matrix of transition probabilities among regimes. We remark that the highest persistence is observed for Regimes 2, 4, and 6, whereas Regime 1 and especially 5 are less

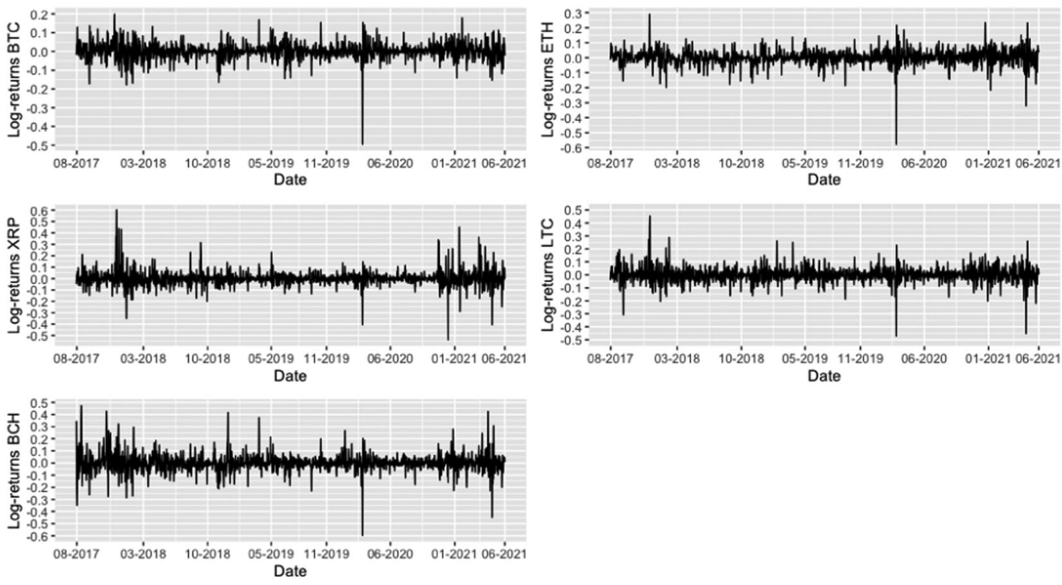

FIGURE 2 Daily time-series of log-returns of the BTC, ETH, XRP, LTC, and BCH cryptocurrencies based on closing prices (complete observations are referred to the period from 3 August 2017, to 30 June 2021). BCH, Bitcoin cash; BTC, Bitcoin; ETH, Ethereum; LTC, Litecoin; XRP, Ripple

TABLE 1 Observed correlation (left panel) and partial correlation (right panel) matrices of the five cryptocurrencies

	Correlations					Partial correlations				
	BTC	ETH	XRP	LTC	BCH	BTC	ETH	XRP	LTC	BCH
BTC	1.000					1.000				
ETH	0.840	1.000				0.732	1.000			
XRP	0.463	0.693	1.000			−0.319	0.270	1.000		
LTC	0.721	0.857	0.787	1.000		0.463	0.040	0.474	1.000	
BCH	0.378	0.684	0.689	0.789	1.000	−0.575	0.429	−0.044	0.571	1.000

Abbreviations: BCH, Bitcoin cash; BTC, Bitcoin; ETH, Ethereum; LTC, Litecoin; XRP, Ripple.

persistent. There is a quite high probability of transition towards State 3 that has average log-returns close to zero and which may be considered as a "centre of gravity." Concerning the highest estimated transition probability from the less persistent Regime 5 to Regime 3, we notice that this switch is not surprising, as Regime 5 represents positive log-returns. This transition can be read like the typical pullback following a substantial price increase.

Figure 3 illustrates the estimated posterior probabilities of being in latent state u, at time t, and $u = 1, \ldots, k$, with $t = 1, \ldots, T$, conditional on the observed time-series and with over-imposed trend according to a smoothed local regression. Through these probabilities we are able to characterise the assets along time at the different market phases. In the most recent period, considering the trend line we notice the increasing tendency for Regimes 4, 5, and 6, and a decreasing tendency especially for Regime 2. Moreover, apart from few exceptions, there are no stable periods.

TABLE 2 Summary information for selecting the number of regimes for the multivariate Hidden Markov model

k	$\hat{\ell}_k$	$\#\text{par}_k$	BIC_k	$\text{AIC}_{c,k}$	AIC_k
1	11,945.160	15	−23,781.360	−23,860.320	−23,890.023
2	13,901.684	43	−27,491.014	−27,717.367	−27,800.759
3	14,361.176	68	−28,228.397	−28,586.351	−28,715.651
4	14,507.426	95	−28,324.769	−28,824.852	−29,001.453
5	14,653.901	124	−28,407.061	−29,059.801	−29,284.408
6	**14,766.747**	**155**	**−28,407.570**	**−29,223.495**	**−29,495.993**
7	14,868.592	188	−28,371.531	−29,361.180	−29,693.150

Note: Figures in bold correspond to the selected model.

TABLE 3 Estimated expected conditional log-returns under the Hidden Markov model with $k = 6$ hidden states (states are ordered according to increasing average values of the expected log-returns)

	1	2	3	4	5	6
BTC	−0.0044	0.0000	0.0002	0.0067	−0.0002	0.0178
ETH	−0.0024	−0.0007	0.0067	0.0010	−0.0077	0.0142
XRP	−0.0089	−0.0030	0.0056	−0.0004	0.0489	−0.0005
LTC	−0.0069	−0.0029	−0.0053	0.0058	0.0287	0.0219
BCH	0.0005	−0.0040	−0.0120	−0.0029	−0.0162	0.0752
Average	−0.0044	−0.0021	−0.0010	0.0020	0.0107	0.0257

Abbreviations: BCH, Bitcoin cash; BTC, Bitcoin; ETH, Ethereum; LTC, Litecoin; XRP, Ripple.

Figure 4 depicts the decoded states obtained through *local decoding* by considering the maximum of the posterior probabilities showed in Figure 3. The predicted trajectories indicate that States 1, 2, and 3, representing negative phases of the market, are visited the 19%, 27%, and 22% of the overall period, respectively. States 4, 5, and 6, related to phases of a market with rise in prices, are visited the 20%, 4%, and 7% of the $T = 1{,}429$ days, respectively. Therefore, States 5 and 6 occur in a small fraction of the time occasions.

On the basis of the previous results some conclusions can be drawn. The distribution of states closely mirrors the dynamics of the cryptocurrency market. In particular, the speculative waves are caught by States 5 and 6, which have been less persistent and show clusters during the “explosive” trends, at the end of 2017, the first big bubble, in mid-2019, and the most recent surge in the second half of 2020 that brought the price of BTC to $60K. Such periods of extreme price movements are characterised by rapid price accelerations with an exponential or even explosive behaviour and are one of the primary concerns for investors due to the high risk represented by the subsequent bubble burst with extremes losses. Bouri et al. (2019) and Agosto and Cafferata (2020) estimated a relevant interconnection between couples of cryptos during both the price surge and the subsequent bubble burst. As a result, protection opportunities in the cryptocurrency market during these periods are limited. The estimated posterior probabilities in Figure 3 suggest that the proposed model adequately succeeds in identifying a trend of a sharp decrease of these episodes. This is important

TABLE 4 Estimated conditional correlations (lower triangle), variances (diagonal, figures in bold), and partial correlations given all remaining variables (upper triangle, figures in italics) under the Hidden Markov model with k = 6 hidden states

	BTC	ETH	XPR	LTC	BCH
State 1					
BTC	**0.0331**	*0.0276*	*0.2916*	*0.4796*	*0.0039*
ETH	0.5098	**0.0565**	*0.1951*	*0.0639*	*−0.0775*
XRP	0.8949	0.5479	**0.0417**	*0.3814*	*0.1401*
LTC	0.9221	0.5214	0.9204	**0.0499**	*0.5179*
BCH	0.8442	0.4564	0.8577	0.9080	**0.0639**
State 2					
BTC	**0.0166**	*0.3406*	*0.1437*	*0.0924*	*0.2909*
ETH	0.7782	**0.0216**	*0.2039*	*0.1679*	*0.1218*
XRP	0.7548	0.7602	**0.0172**	*0.3541*	*0.2168*
LTC	0.7591	0.7623	0.8213	**0.0225**	*0.3989*
BCH	0.7922	0.7608	0.8017	0.8382	**0.0231**
State 3					
BTC	**0.0520**	*−0.1693*	*−0.1352*	*0.5073*	*0.3364*
ETH	0.5241	**0.0487**	*−0.1230*	*0.3795*	*0.1279*
XRP	0.6083	0.4397	**0.0821**	*0.4131*	*0.1388*
LTC	0.8632	0.6624	0.7369	**0.0627**	*0.3944*
BCH	0.8368	0.6091	0.6809	0.8842	**0.0727**
State 4					
BTC	**0.0407**	*0.3350*	*−0.0947*	*−0.0045*	*0.4603*
ETH	0.7961	**0.0395**	*0.3541*	*0.1405*	*0.2357*
XRP	0.6782	0.7950	**0.0392**	*0.1375*	*0.3360*
LTC	0.6505	0.7138	0.6902	**0.0508**	*0.3146*
BCH	0.8293	0.8406	0.8029	0.7616	**0.0399**
State 5					
BTC	**0.1002**	*0.2956*	*−0.1371*	*0.1731*	*0.3267*
ETH	0.6987	**0.1253**	*−0.0047*	*0.1620*	*0.2869*
XRP	0.3787	0.4327	**0.2151**	*−0.0514*	*0.4864*
LTC	0.6806	0.6786	0.4533	**0.1617**	*0.4480*
BCH	0.7349	0.7409	0.6313	0.7865	**0.1495**
State 6					
BTC	**0.0497**	*−0.0361*	*0.3656*	*0.2929*	*−0.2580*
ETH	0.1965	**0.0372**	*0.3270*	*−0.0826*	*−0.0110*
XRP	0.6271	0.3821	**0.0494**	*0.5177*	*0.1521*
LTC	0.6013	0.2124	0.7200	**0.0633**	*0.1202*
BCH	−0.0884	0.0500	0.1465	0.1389	**0.1509**

Abbreviations: BCH, Bitcoin cash; BTC, Bitcoin; ETH, Ethereum; LTC, Litecoin; XRP, Ripple.

TABLE 5 Estimated transition probabilities under the Hidden Markov model with k = 6 hidden states, figures in italics correspond to the probabilities in the main diagonal and in bold to the largest probability in each row out of the main diagonal

	1	2	3	4	5	6
1	*0.4762*	**0.2056**	0.1822	0.1035	0.0136	0.0188
2	**0.1933**	*0.6988*	0.0197	0.0882	0.0000	0.0000
3	**0.1544**	0.0282	*0.5813*	0.0913	0.0853	0.0594
4	0.1013	**0.1491**	0.0253	*0.6816*	0.0276	0.0151
5	0.0000	0.0000	**0.5300**	0.0295	*0.3847*	0.0557
6	0.0000	0.0000	**0.2745**	0.0114	0.0476	*0.6664*

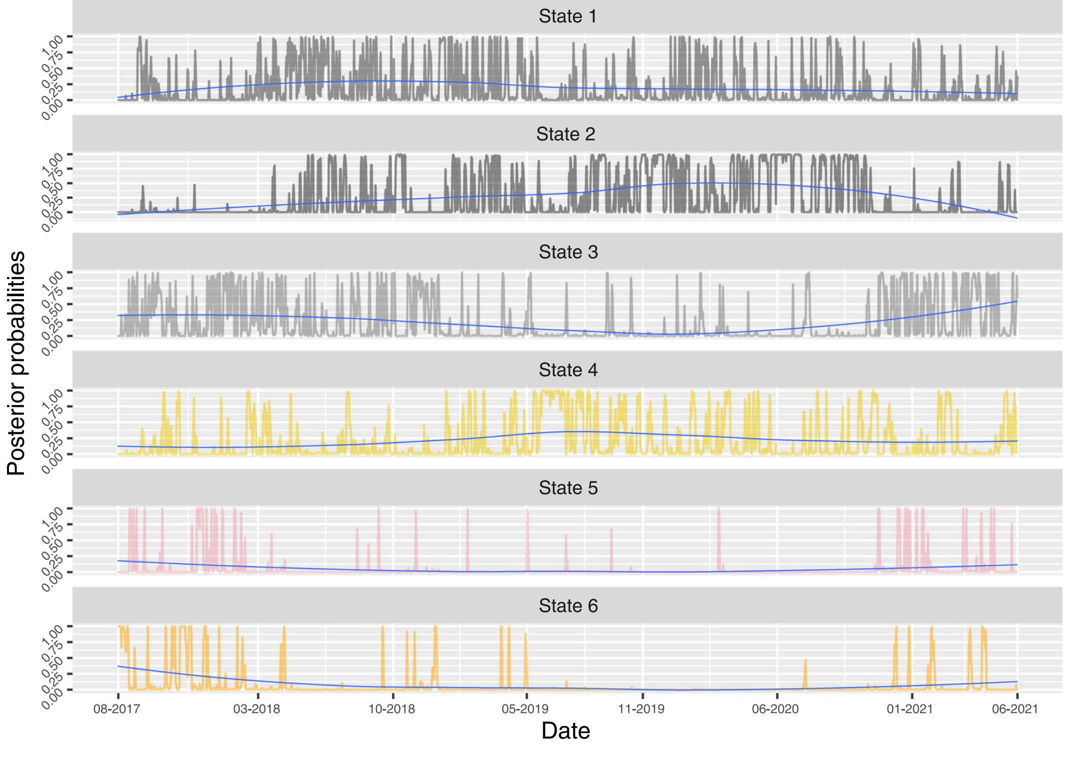

FIGURE 3 Predicted posterior probabilities under the Hidden Markov model with k = 6 states with over-imposed trend according to smoothed local regression (in blue)

since these bubbles poses a serious limit for retail and institutional investors in considering cryptocurrencies as investable assets.

Figures from 1 to 5 reported in the supplementary information depict the observed log-returns, the predicted averages, and *standard deviations* over the entire time period for each cryptocurrency. In our intention, the HM model with six regimes is not meant to provide precise univariate predictions of log-returns or volatility, but the results are truly comforting; the model seems able to detect regimes of high or low returns and volatilities timely.

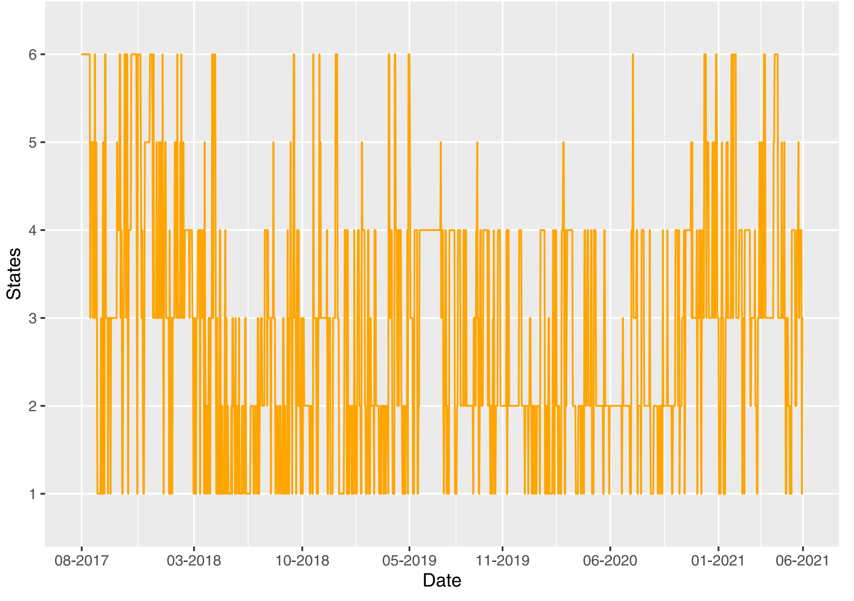

FIGURE 4 Predicted sequence of hidden states over time under the Hidden Markov model with k = 6 states

Finally, Figure 5 shows the estimates of daily correlations between BTC and the other cryptocurrencies, namely, ETH, XPR, LTC, BCH, along with a trend line inferred according to a smooth local regression. A clear contribution of the multivariate HM model is to detect structural trends in the interconnection between assets, and we show that in absence of speculative bubbles, correlations remain very high, with marked decreases only during those periods. A possible interpretation of this result is related to the maturity stage of the cryptocurrency market. Indeed, our analysis focuses on the most recent period of market development from 2017 to 2021. The strengthening of the systematic risk can be explained by factors such as the highest number of institutional players involved in transactions, the spread of derivatives on BTC, and the increased diffusion of cryptocurrency-based indices. All factors were pushing for investments in a portfolio logic. Another explanation could be the rise of stablecoins pair dominance during the year 2018, concurring to the overall decline in the contribution of BTC pairs to total industry trade volume, with a pressure to a stronger link with USD of all cryptocurrencies and a subsequent increase in correlation.

5 | CONCLUSION

We propose a multivariate Hidden Markov model to analyse log-returns of the main five cryptocurrencies: Bitcoin, Ethereum, Ripple, Litecoin, and Bitcoin cash in the period 2017–2021. The narrow universe of cryptocurrencies and exchanges we selected fulfills the intention to concentrate on the more reliable, liquid, and less manipulated crypto-assets in the market. The choice of recent observations (nearly 4 years of data) followed similar criteria of homogeneity between time-series, especially with reference to the liquidity profile. The proposed approach is coherent with many stylised facts in finance. The advantage of employing an HM model that includes expected log-returns and state-specific variance–covariance matrices lies in the use of the surplus of information provided by this model compared to traditional regime-switching models that focus exclusively on volatility.

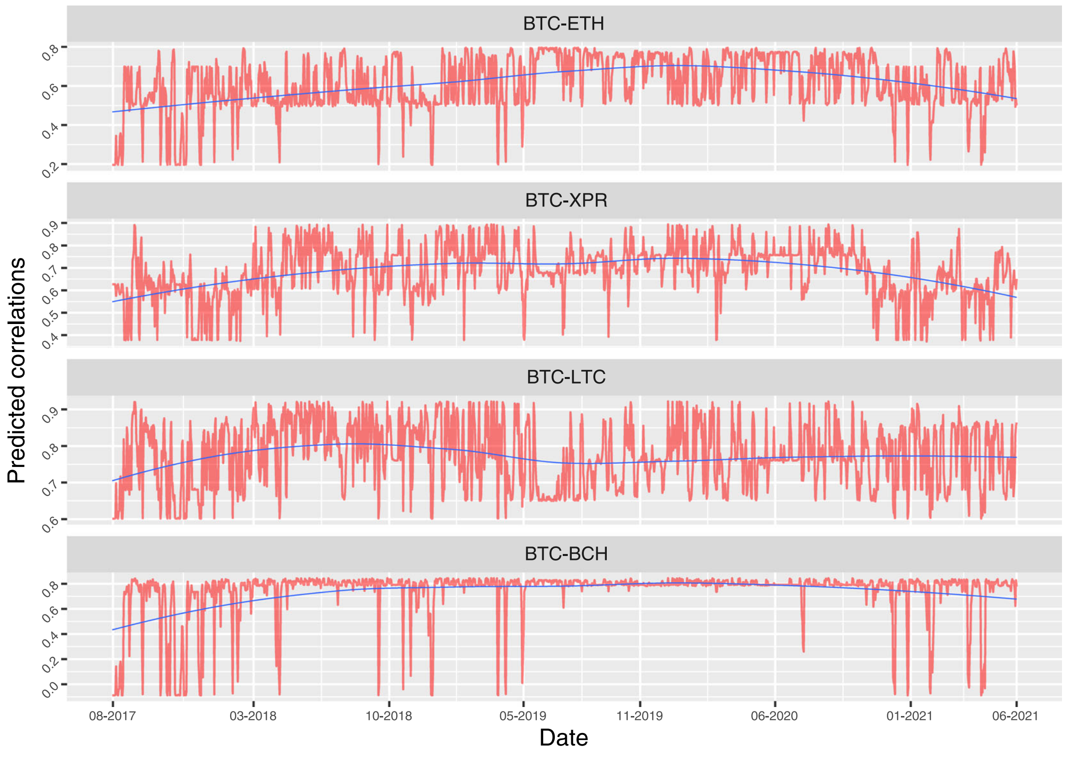

FIGURE 5 Predicted correlations between BTC and the other cryptocurrencies under the Hidden Markov model with k = 6 hidden states with over-imposed smooth trend according to a local regression (blue line). BCH, Bitcoin cash; BTC, Bitcoin; ETH, Ethereum; LTC, Litecoin; XRP, Ripple

To choose the appropriate number of regimes, we rely on the Bayesian information criterion and select a model with six regimes. We showed that the model performs well in capturing the different regime dynamics of this market. The model proves capable of segmenting into six regimes with consistent monotonic dynamics of average log-returns, thus providing rather remarkable univariate predictions of log-returns and volatility. Three of these regimes represent a negative phase of the market, and Regime 3 shows high volatility. Regimes 4, 5, and 6 are related to phases of a marked rise in price. Finally, we spot a trend of increase of the market correlation from the predicted correlations of the cryptocurrencies coupled to Bitcoin. This evidence is coherent with the hypothesis of an increasing association observed in more mature markets, and the induced stronger link with the USD starting from 2018 due to the rise of stablecoins. The robustness of the proposal concerning variation in the data was empirically assessed by estimating the model with prices considered for a shorter time interval (from 3 August 2017, to 27 February 2020); see Pennoni et al. (2021).

Further improvements of the present approach may consider a generalisation of the model using different conditional distributions of the log-returns, among which is the multivariate Student's t distribution. We may also consider comparing of the proposed approach with regime-switching copula models where the states are characterised on the basis of the correlation structures. This approach may be more suitable to model the distribution tails and thus the probabilities of simultaneous occurrence of rare events (Demarta & McNeil, 2005; Nasri & Rémillard, 2019). Another promising approach, especially when portfolio allocation is of interest, may be based on the estimation procedure proposed in Nystrup et al. (2020) to reduce the estimated switching patterns.

ACKNOWLEDGMENTS

Open access funding provided by Universita degli Studi di Milano-Bicocca within the CRUI-CARE Agreement.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the Crypto Asset Lab, see the web page at: https://cryptoassetlab.diseade.unimib.it/.

ETHICS STATEMENT

The authors have followed the required ethical standard.

ORCID

Fulvia Pennoni https://orcid.org/0000-0002-6331-7211
Francesco Bartolucci https://orcid.org/0000-0001-7057-1421
Gianfranco Forte https://orcid.org/0000-0002-3412-4162
Ferdinando Ametrano https://orcid.org/0000-0001-8521-0063

REFERENCES

Agosto, A., & Cafferata, A. (2020). Financial bubbles: A study of co-explosivity in the cryptocurrency market. *Risks*, *8*, 1–14.
Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov, & C. F. (Eds.), *Second international symposium on information theory* (pp. 267–281). Akademiai Kiado.
Akaike, H. (1998). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. In *Selected papers of Hirotugu Akaike* (pp. 223–247). Springer.
Almudhaf, F. (2018). Pricing efficiency of Bitcoin trusts. *Applied Economics Letters*, *25*, 504–508.
Alvarez-Ramirez, J., Rodriguez, E., & Ibarra-Valdez, C. (2018). Long-range correlations and asymmetry in the Bitcoin market. *Physica A: Statistical Mechanics and its Applications*, *492*, 948–955.
Ang, A., & Bekaert, G. (2002). International asset allocation with regime shifts. *The Review of Financial Studies*, *15*, 1137–1187.
Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*, *49*, 803–821.
Bariviera, A. F., Basgall, M. J., Hasperué, W., & Naiouf, M. (2017). Some stylized facts of the Bitcoin market. *Physica A: Statistical Mechanics and its Applications*, *484*, 82–90.
Bartolucci, F., & De Luca, G. (2003). Likelihood-based inference for asymmetric stochastic volatility models. *Computational Statistics & Data Analysis*, *42*, 445–449.
Bartolucci, F., Farcomeni, A., & Pennoni, F. (2013). *Latent Markov models for longitudinal data*. Chapman & Hall/CRC Press.
Bartolucci, F., Pandolfi, S., & Pennoni, F. (2017). LMest: An R package for latent Markov models for longitudinal categorical data. *Journal of Statistical Software*, *81*, 1–38.
Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, *41*, 164–171.
Borri, N. (2019). Conditional tail-risk in cryptocurrency markets. *Journal of Empirical Finance*, *50*, 1–19.
Bouri, E., Shahzad, S. J. H., & Roubaud, D. (2019). Co-explosivity in the cryptocurrency market. *Finance Research Letters*, *29*, 178–183.
Cappé, O., Moulines, E., & Rydén, T. (1989). *Inference in hidden Markov models*. New York: Springer-Verlag.
Celeux, G., & Govaert, G. (1995). Gaussian parsimonious clustering models. *Pattern Recognition*, *28*, 781–793.
Chan, S., Chu, J., Nadarajah, S., & Osterrieder, J. (2017). A statistical analysis of cryptocurrencies. *Journal of Risk and Financial Management*, *10*, 1–23.
Cheah, E.-T., & Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters*, *130*, 32–36.
Chen, Y., Giudici, P., HadjiMisheva, B., & Trimborn, S. (2020). Lead behaviour in Bitcoin markets. *Risks*, *8*, 1–14.
Cheung, A., Roca, E., & Su, J.-J. (2015). Crypto-currency bubbles: An application of the Phillips-Shi-Yu (2013) methodology on Mt. Gox Bitcoin prices. *Applied Economics*, *47*, 2348–2358.
Corbet, S., Meegan, A., Larkin, C., Lucey, B., & Yarovaya, L. (2018). Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*, *165*, 28–34.
De Angelis, L., & Paas, L. J. (2013). A dynamic analysis of stock markets using a hidden Markov model. *Journal of Applied Statistics*, *40*, 1682–1700.
Demarta, S., & McNeil, A. J. (2005). The t-copula and related copulas. *International Statistical Review*, *73*, 111–129.
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Dias, J. G., Vermunt, J. K., & Ramos, S. (2015). Clustering financial time series: New insights from an extended hidden Markov model. *European Journal of Operational Research*, *243*, 852–864.

Genon-Catalot, V., Jeantheau, T., Larédo, C., et al. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli*, *6*, 1051–1079.

Giudici, P., & Abu Hashish, I. (2020). A hidden Markov model to detect regime changes in cryptoasset markets. *Quality and Reliability Engineering International*, *36*, 2057–2065.

Giudici, P., & Pagnottoni, P. (2019). High frequency price change spillovers in Bitcoin markets. *Risks*, *7*, 1–18.

Giudici, P., & Pagnottoni, P. (2020). Vector error correction models to measure connectedness of Bitcoin exchange markets. *Applied Stochastic Models in Business and Industry*, *36*, 95–109.

Giudici, P., & Polinesi, G. (2019). Crypto price discovery through correlation networks. *Annals of Operations Research*, *299*, 443–457.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*, 357–384.

Huang, J.-Z., Huang, W., & Ni, J. (2019). Predicting Bitcoin returns using high-dimensional technical indicators. *The Journal of Finance and Data Science*, *5*, 140–155.

Hurvich, C. M., & Tsai, C.-L. (1993). A corrected Akaike information criterion for vector autoregressive model selection. *Journal of Time Series Analysis*, *14*, 271–279.

Juang, B. H., & Rabiner, L. R. (1991). Hidden Markov models for speech recognition. *Technometrics*, *33*, 251–272.

Katsiampa, P., Corbet, S., & Lucey, B. (2019). High frequency volatility co-movements in cryptocurrency markets. *Journal of International Financial Markets, Institutions and Money*, *62*, 35–52.

Langrock, R., MacDonald, I. L., & Zucchini, W. (2012). Some nonstandard stochastic volatility models and their estimation using structured hidden Markov models. *Journal of Empirical Finance*, *19*, 147–161.

Lin, Y., Xiao, Y., & Li, F. (2020). Forecasting crude oil price volatility via a HM-EGARCH model. *Energy Economics*, *87*, 104–693.

Mamon, R. S., & Elliott, R. J. (2007). *Hidden Markov models in finance*. Springer.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Wiley.

Nadarajah, S., & Chu, J. (2017). On the inefficiency of Bitcoin. *Economics Letters*, *150*, 6–9.

Nasri, B. R., & Rémillard, B. N. (2019). Copula-based dynamic models for multivariate time series. *Journal of Multivariate Analysis*, *172*, 107–121.

Nystrup, P., Lindström, E., & Madsen, H. (2020). Learning hidden Markov models with persistent states by penalizing jumps. *Expert Systems with Applications*, *150*, 1–14.

Osterrieder, J. (2017). The statistics of Bitcoin and cryptocurrencies. In *2017 International Conference on Economics, Finance and Statistics (ICEFS 2017)*, Hong Kong, 14–15 January, 2017, Atlantis Press.

Pennoni, F., Bartolucci, F., Forte, G., & Ametrano, F. (2021). Exploring the dependencies among main cryptocurrency log-returns: A hidden Markov model. Technical report. https://econpapers.repec.org/paper/pramprapa/106150.htm

Phillips, P. C., Shi, S., & Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, *56*, 1043–1078.

Phillips, P. C., Wu, Y., & Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review*, *52*, 201–226.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rossi, A., & Gallo, G. M. (2006). Volatility estimation via hidden Markov models. *Journal of Empirical Finance*, *13*, 203–230.

Rydén, T., Teräsvirta, T., & Åsbrink, S. (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics*, *13*, 217–244.

Satoshi, N. (2008). Bitcoin: A peer-to-peer electronic cash system. *Consulted*, *1*, 28.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Sifat, I. M., Mohamad, A., & Shariff, M. S. B. M. (2019). Lead-lag relationship between Bitcoin and Ethereum: Evidence from hourly and daily data. *Research in International Business and Finance*, *50*, 306–321.

Szczygielski, J. J., Karathanasopoulos, A., & Zaremba, A. (2020). One shape fits all? A comprehensive examination of cryptocurrency return distributions. *Applied Economics Letters*, *27*, 1567–1573.

Tiwari, A. K., Jana, R. K., Das, D., & Roubaud, D. (2018). Informational efficiency of Bitcoin—An extension. *Economics Letters*, *163*, 106–109.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory*, *13*, 260–269.

Welch, L. R. (2003). Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter*, *53*, 1–13.

Xiong, J., Liu, Q., & Zhao, L. (2020). A new method to verify Bitcoin bubbles: Based on the production cost. *The North American Journal of Economics and Finance*, *51*, 1–11.
Yi, S., Xu, Z., & Wang, G.-J. (2018). Volatility connectedness in the cryptocurrency market: Is Bitcoin a dominant cryptocurrency? *International Review of Financial Analysis*, *60*, 98–114.
Zhang, W., Wang, P., Li, X., & Shen, D. (2018). Some stylized facts of the cryptocurrency market. *Applied Economics*, *50*, 5950–5965.
Zucchini, W., MacDonald, I. L., & Langrock, R. (2017). *Hidden Markov models for time series: An introduction using R*. New York: Springer-Verlag.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Pennoni, F., Bartolucci, F., Forte, G., & Ametrano, F. (2022). Exploring the dependencies among main cryptocurrency log-returns: A hidden Markov model. *Economic Notes*, 51, e12193. https://doi.org/10.1111/ecno.12193