Multivariate Hidden Markov model for cryptocurrency log-returns

Francesco Bartolucci

Department of Economics, Italy, Università di Perugia (IT)

francesco.bartolucci@unipg.it

Ferdinando Ametrano

Gianfranco Forte

ferdinando.ametrano@unimib.it

Fulvia Pennoni

Department of Statistics and Quantitative Methods, Università di Milano-Bicocca (IT)

fulvia.pennoni@unimib.it

Abstract

We provide an analysis of the market data of the major crypto currencies by assuming a multivariate hidden Markov process also known as latent Markov process. We model jointly the daily log-returns of BTC, ETH, XRP, LTC and BCH. The observed log-returns are assumed to be correlated according to a variance-covariance matrix conditionally on a latent Markov process of first-order having a discrete number of latent states. In order to compare states according to their volatility we estimate the specific variance-covariance matrix of each state. Maximum likelihood estimation of the model parameters is carried out by the Expectation-Maximization algorithm. The latent states can be ordered according to expected average values of the log-returns and their estimated volatility. We consider different model specifications in terms of number of latent states, which are identified in terms of expected log-returns and level of volatility. Under each considered scenario we also predict the latent state by the maximum a posteriori rule.

Keywords: Expectation-Maximization algorithm, latent variables,time series

1 Introduction

We propose to monitor financial asset price series constructed over a multivariate Hidden Markov model (HMM) (Bartolucci et al., 2014, 2017; **?**). We outline the Expectation-Maximization (EM) algorithm (Dempster et al., 1977; Welch, 2003) for the maximum likelihood estimation. We provide an example based on the analysis of market data referred to five cryptocurrencies.

The remainder of the paper is organized as follows. In Section 2 we define the notation of the quantities of interest for the proposed model and some theoretical results. In Section 3 we describe the data and the markets. In Section 4 we show the results of the proposal. We conclude the paper with Section **??** where we provide some conclusions.

2 Proposed model

We consider a real valued sequence x_t representing prices of the crypto at time $t = 1, 2, \ldots$. For some strictly increasing and differentiable real function $g(\cdot)$, we define $y_t = g(x_t)$ as the $g-$transformed prices, and let $r_t = y_t - y_{t-1}$, with $t = 2, 3, \ldots$, denote their first order difference. As notable examples of this framework, setting $g(\cdot) = 1$ imply working with raw prices and their differences, and setting $g(\cdot) = \log(\cdot)$ imply working with log-prices and log-returns.

These assumptions concern only a limited segment of data, those relative to times $t - s + 1, \ldots, t$, therefore the whole process can also violate those restrictions globally (in particular the whole process can be a locally-stationary sequence). In fact, under the above assumptions (1) and (2), the sequence r_t is unconditionally distributed as a locally stationary process with piecewise constant mean, which becomes stationary (constant mean) in the time intervals $t - s + 1, \ldots, t$, while the (possibly transformed) prices y_t form a sub/super martingale. On the other hand, these assumptions give rise to

a general (possibly non-Gaussian) sub/super martingale (SSM) for the transformed prices y_t. We propose a HMM is based on a kernel transition function that follows a Gaussian distribution with certain mean and standard deviation.

3 Data

In what follows we show the results of the HMM model to the time series of log-returns market data concerning the daily closing price of BTC, ETH, XPR, LTC and BCH.

The data provided by the Crypto Currency Lab are related to the following daily quotes (940) over a three year period from 2^{nd} August 2017 - 27^{th} February 2020

- BTC: Bitcoin;
- ETH: Ethereum;
- XRP;
- LTC: Litecoin;
- BCH: Bitcoin cash;

For each of them we model the daily log-returns

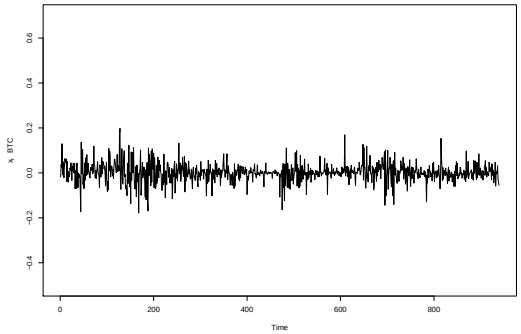
$$x_t = \log\left(\frac{y_t}{y_{t-1}}\right) \quad t = 1, \ldots, 940,$$

with y_t denoting the closing price on day t.

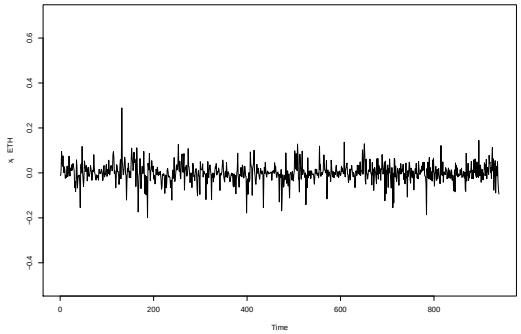
From Figure 2 we observe the daily log-returns of the crypto currencies taken into account and we notice the high level of volatility of each one.

From Figure **??** we observe the daily log-returns of the BTC crypto currency and the corresponding prices for the whole period.

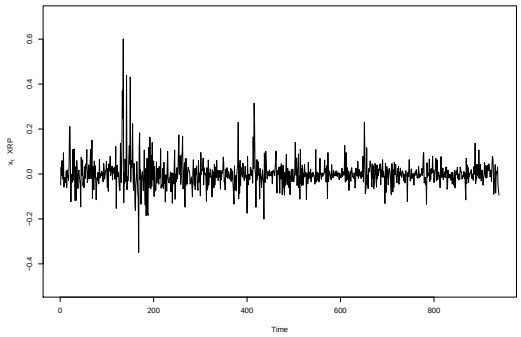
Table 1 reports the observed variance-covariance matrix, Table 2 reports the observed correlations and Table 3 reports the partial correlations.



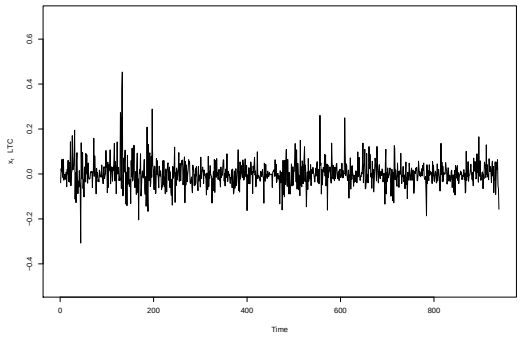
(a) BTC



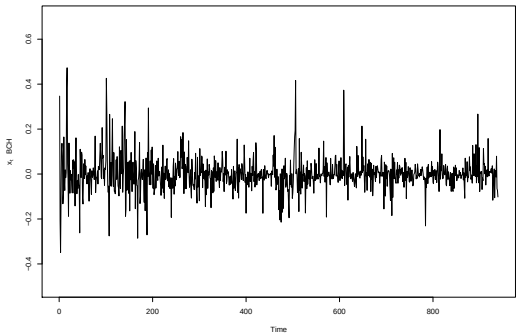
(b) ETH



(c) XRP

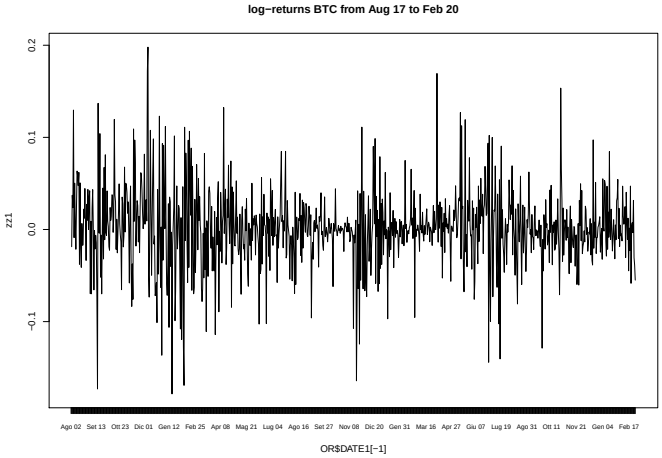


(d) LTC



(e) BCH

Figure 1: *Time series of log-returns on the BTC, ETH, XRP, LTC, BCH crypto closing prices at each day (complete observations are referred to the period from 2^{nd} August 2017 to 27^{th} February 2020).*



(a) BTC

(b) ETH

Figure 2: *Time series of log-returns on the BTC crypto closing prices at each day and prices (complete observations are referred to the period from 2nd August 2017 to 27th February 2020).*

Table 1: *Observed variance-covariance matrix.*

BTC	0.15				
ETH	0.13	0.38			
XRP	0.09	0.23	0.28		
LTC	0.16	0.29	0.21	0.29	
BCH	0.19	0.45	0.27	0.35	0.61

Table 2: *Observed correlation matrix.*

	BTC	ETH	XRP	LTC	BCH
BTC	1.00				
ETH	0.55	1.00			
XRP	0.44	0.71	1.00		
LTC	0.74	0.86	0.73	1.00	
BCH	0.62	0.94	0.66	0.82	1.00

4 Results of the HMM models specific variance-covariance matrix for each latent state

We estimated the model where a specific variance-covariance matrix is considered for each latent state. The HMM is estimated by adapting suitable functions of the `R` package `LMest` (Bartolucci et al., 2017).

The order of the latent distribution is selected according of the Bayesian Information Criterion (Schwarz, 1978). The proposed multivariate HMM is estimated for a number of latent states ranging from 1 to 7 and the results are displayed in Table 4. We performed a model selection strategy and we account for the multimodality of the likelihood function allowing for different sets of starting values of the EM algorithm. We also report the values of the Akaike Information Criterion (**?**).

The best model according to the results showed in 4 is that with four latent states.

Table 3: *Observed partial correlation matrix.*

	BTC	ETH	XRP	LTC	BCH
BTC	1.00				
ETH	-0.38	1.00			
XRP	-0.16	0.14	1.00		
LTC	0.63	0.46	0.37	1.00	
BCH	0.34	0.82	-0.04	-0.12	1.00

Table 4: *Results from the fitting of the multivariate HMMs with specific variance-covariance matrix for each latent state of the daily log-returns for BTC, ETH, XRP, LTC, BCH for increasing number of latent states.*

k	log-likelihood	parameters	BIC	AIC
1	7785.46	15	-15468.25	-15540.93
2	9044.86	43	-17795.40	-18003.73
3	9334.79	68	-18204.13	-18533.58
4	9455.19	95	-18260.12	-18720.37
5	9553.41	124	-18258.05	-18858.81
6	9635.06	155	-18209.18	-18960.13
7	9718.65	188	-18150.47	-19061.29

4.1 HMM with 3 hidden states

When the order of the latent chain is 4 the corresponding averages of each state ordered with respect to BTC are showed in Table 5. Tables from 6 to 8 show the estimated marginal correlations and partial correlations matrices for each latent state. Table 9 shows the estimated matrix of transition probabilities and Figure 3 reports the corresponding heat map. Figure 4 depicts the decoded states across days.

Table 5: *Estimated averages for the HMM with $k = 3$ latent states.*

	1	2	3
BTC	-0.0013	-0.0005	0.0167
ETH	-0.0024	-0.0022	0.0164
XRP	-0.0064	-0.0041	0.0400
LTC	-0.0037	-0.0040	0.0297
BCH	-0.0072	-0.0045	0.0415

Table 6: *Estimated marginal correlations in the lower triangle and partial correlations given all remaining variables in the upper triangle of the HMM for state 1.*

	BTC	ETH	XPR	LTC	BCH
BTC	1.0000	0.1042	0.0422	0.4080	-0.0027
ETH	0.2250	1.0000	-0.0409	0.2070	0.0358
XRP	0.1892	0.0644	1.0000	0.2646	0.0636
LTC	0.4778	0.2901	0.3295	1.0000	0.1585
BCH	0.1082	0.0952	0.1300	0.2203	1.0000

Table 7: *Estimated marginal correlations in the lower triangle and partial correlations given all remaining variables in the upper triangle of the HMM for state 2.*

	BTC	ETH	XPR	LTC	BCH
BTC	1.0000	0.2966	0.1538	0.1559	0.3009
ETH	0.6550	1.0000	0.2131	0.1265	0.0485
XRP	0.6784	0.6336	1.0000	0.2927	0.2697
LTC	0.6829	0.6102	0.7278	1.0000	0.3661
BCH	0.7200	0.6002	0.7279	0.7553	1.0000

Table 8: *Estimated marginal correlations in the lower triangle and partial correlations given all remaining variables in the upper triangle of the HMM for state 3.*

	BTC	ETH	XPR	LTC	BCH
BTC	1.0000	0.0422	0.1293	0.4035	0.3083
ETH	0.4384	1.0000	0.1129	0.1120	0.0310
XRP	0.7510	0.4555	1.0000	0.3344	0.2418
LTC	0.8389	0.4743	0.8073	1.0000	0.3666
BCH	0.8112	0.4404	0.7772	0.8417	1.0000

Table 9: *Estimated transition probabilities for the HMM with $k = 3$ latent states.*

	1	2	3
1	0.6933	0.2156	0.0911
2	0.2973	0.7027	0.0000
3	0.3745	0.0021	0.6234

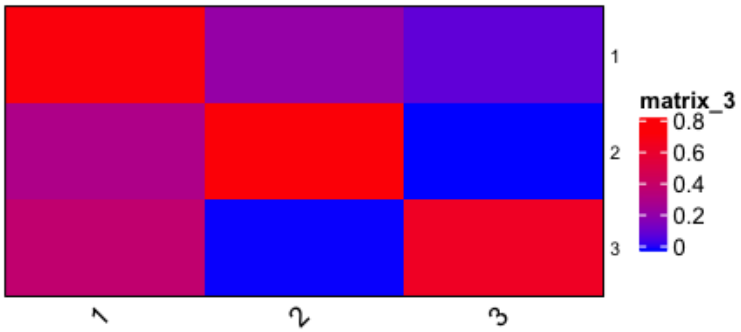


Figure 3: *Heat transition matrix with 3 hidden states.*

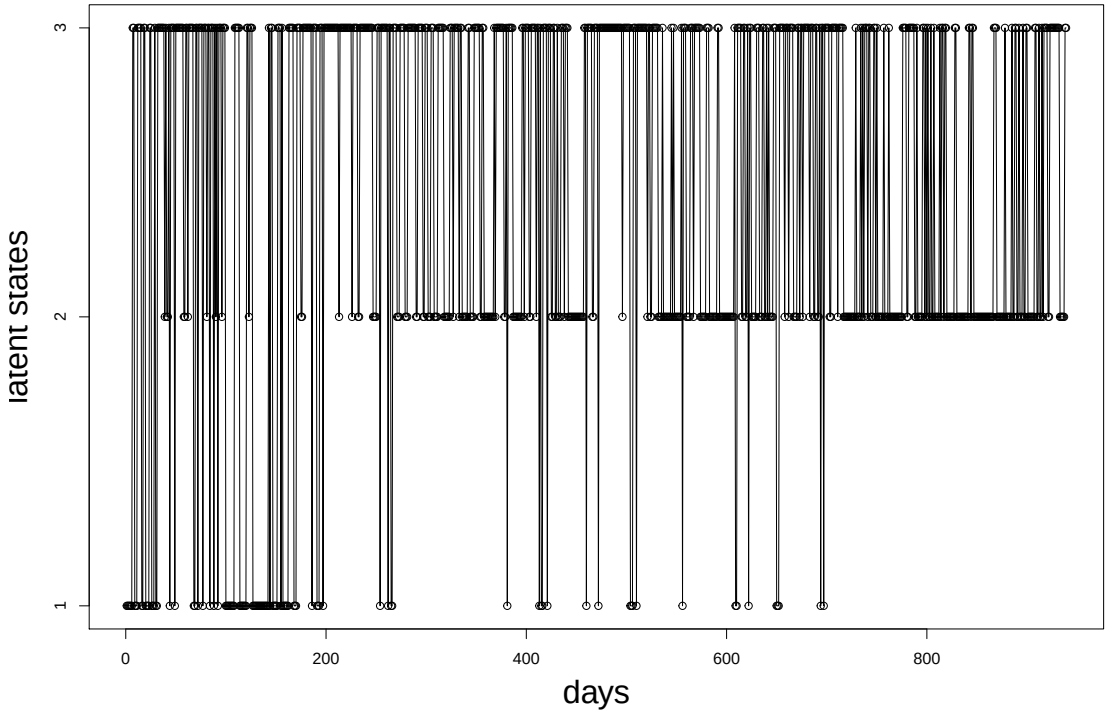


Figure 4: *Decoded states for the HMM with 3 states.*

4.2 Results with 4 hidden states

In the following, we show the results of the HMM with 4 states.

When the order of the latent chain is 4 the corresponding averages of each state ordered with respect to BTC are showed in Table 10. Tables from 11 to 14 show the estimated marginal correlations and partial correlations matrices for each latent state. Table 15 shows the estimated matrix of transition probabilities and Figure 5 reports the corresponding heat map. Figure 6 depicts the decoded states across days.

Table 10: *Estimated averages for the HMM with specific variance-covariance matrix with $k = 4$ latent states.*

	1	2	3	4
BTC	-0.0033	-0.0014	0.0042	0.0185
ETH	-0.0023	-0.0018	-0.0025	0.0159
XRP	-0.0060	-0.0042	-0.0053	0.0408
LTC	-0.0065	-0.0034	0.0013	0.0309
BCH	-0.0069	-0.0037	-0.0070	0.0420

Table 11: *Estimated marginal correlations in the lower triangle and partial correlations given all remaining variables in the upper triangle of the HMM for state 1.*

	BTC	ETH	XPR	LTC	BCH
BTC	1.0000	0.3023	0.1547	0.1073	0.3038
ETH	0.6570	1.0000	0.2454	0.1214	0.0589
XRP	0.6703	0.6554	1.0000	0.3079	0.2733
LTC	0.6439	0.6060	0.7229	1.0000	0.3475
BCH	0.7039	0.6090	0.7268	0.7342	1.0000

Table 12: *Estimated marginal correlations in the lower triangle and partial correlations given all remaining variables in the upper triangle of the HMM for state 2.*

	BTC	ETH	XPR	LTC	BCH
BTC	1.0000	0.1201	-0.1036	0.2234	0.3492
ETH	0.3244	1.0000	-0.0377	0.1627	0.1071
XRP	0.2414	0.1463	1.0000	-0.0121	0.4908
LTC	0.5219	0.3607	0.3279	1.0000	0.4042
BCH	0.5700	0.3412	0.5633	0.6358	1.0000

Table 13: *Estimated marginal correlations in the lower triangle and partial correlations given all remaining variables in the upper triangle of the HMM for state 3.*

	BTC	ETH	XPR	LTC	BCH
BTC	1.0000	0.0422	0.1293	0.4035	0.3083
ETH	0.4384	1.0000	0.1129	0.1120	0.0310
XRP	0.7510	0.4555	1.0000	0.3344	0.2418
LTC	0.8389	0.4743	0.8073	1.0000	0.3666
BCH	0.8112	0.4404	0.7772	0.8417	1.0000

Table 14: *Estimated marginal correlations in the lower triangle and partial correlations given all remaining variables in the upper triangle of the HMM for state 4.*

	BTC	ETH	XPR	LTC	BCH
BTC	1.0000	-0.0438	0.4256	0.1400	-0.2087
ETH	0.2539	1.0000	0.3108	-0.0433	-0.0292
XRP	0.6621	0.4297	1.0000	0.6080	0.1601
LTC	0.5788	0.3035	0.7728	1.0000	0.0183
BCH	-0.0811	0.0294	0.1103	0.0788	1.0000

Table 15: *Estimated transition probabilities for the HMM with specific variance-covariance matrix with $k = 4$ latent states.*

	1	2	3	4
1	0.6891	0.1604	0.0499	0.1005
2	0.1960	0.7149	0.0891	0.0000
3	0.1437	0.1327	0.7175	0.0061
4	0.3007	0.0003	0.0469	0.6520

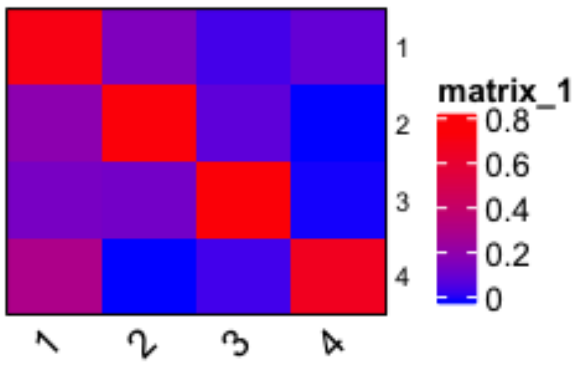


Figure 5: *Heat transition matrix with 4 hidden states.*

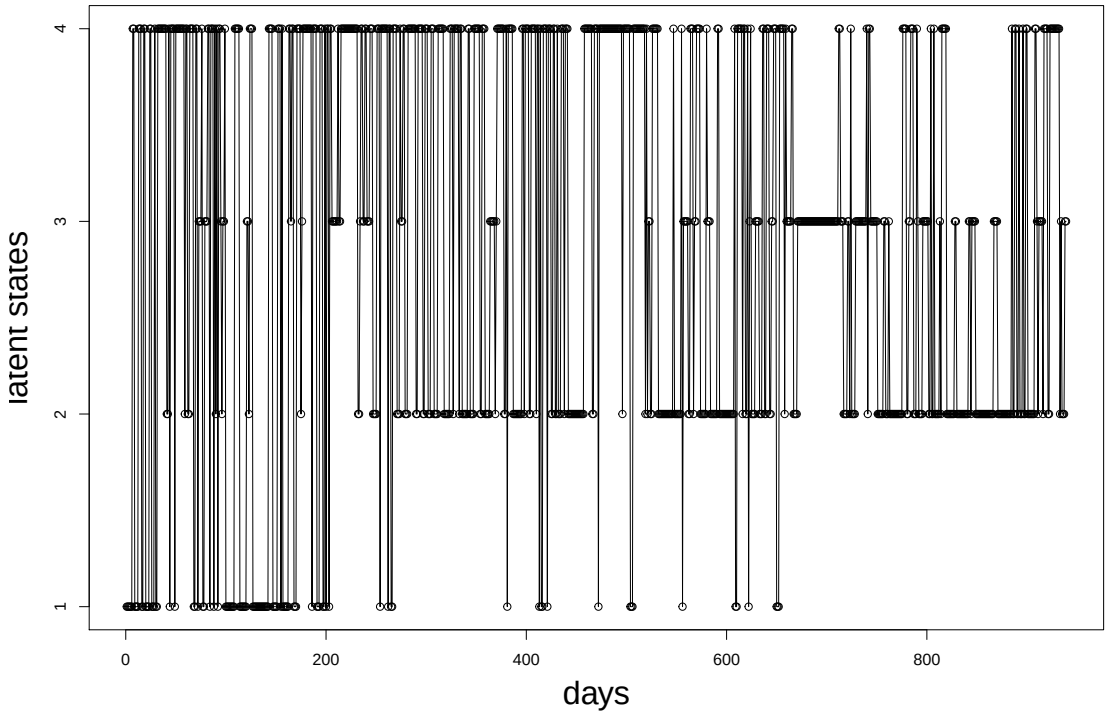


Figure 6: *Decoded states for the HMM with 4 states with specific variance-covariance matrix.*

4.3 HMM with 5 hidden states

When the order of the latent chain is 5 the corresponding averages of each state ordered with respect to BTC are showed in Table 16. Tables from 17 to 21 show the estimated marginal and partial correlations matrices for each latent state. Table 22 shows the estimated matrix of transition probabilities and Figure 7 reports the corresponding heat map. Figure 8 depicts the decoded states across days.

Table 16: *Estimated averages for the HMM with $k = 5$ latent states.*

	1	2	3	4	5
BTC	-0.0031	-0.0011	-0.0008	0.0022	0.0153
ETH	-0.0034	-0.0020	-0.0018	-0.0034	0.0176
XRP	-0.0054	-0.0036	-0.0072	-0.0052	0.0375
LTC	-0.0094	-0.0028	-0.0032	-0.0007	0.0305
BCH	-0.0146	-0.0035	0.0012	-0.0074	0.0395

Table 17: *Estimated marginal correlations in the lower triangle and partial correlations given all remaining variables in the upper triangle of the HMM for state 1.*

	BTC	ETH	XPR	LTC	BCH
BTC	1.0000	-0.0428	0.0773	0.5244	0.2160
ETH	0.3583	1.0000	0.1011	0.0862	0.0592
XRP	0.7706	0.3892	1.0000	0.3905	0.0435
LTC	0.9029	0.4063	0.8306	1.0000	0.4864
BCH	0.8498	0.3870	0.7599	0.8947	1.0000

Table 18: *Estimated marginal correlations in the lower triangle and partial correlations given all remaining variables in the upper triangle of the HMM for state 2.*

	BTC	ETH	XPR	LTC	BCH
BTC	1.0000	-0.1498	0.3587	0.3002	-0.2612
ETH	0.1058	1.0000	0.2631	-0.0053	-0.0954
XRP	0.6813	0.2988	1.0000	0.5969	0.1338
LTC	0.6742	0.2088	0.8023	1.0000	0.0908
BCH	-0.1197	-0.0342	0.0760	0.0709	1.0000

Table 19: *Estimated marginal correlations in the lower triangle and partial correlations given all remaining variables in the upper triangle of the HMM for state 3.*

	BTC	ETH	XPR	LTC	BCH
BTC	1.0000	0.3604	-0.1934	-0.1147	0.5220
ETH	0.7838	1.0000	0.3143	0.2663	0.1223
XRP	0.6803	0.8043	1.0000	0.0768	0.5330
LTC	0.6085	0.7333	0.7072	1.0000	0.2868
BCH	0.8252	0.8388	0.8581	0.7542	1.0000

Table 20: *Estimated marginal correlations in the lower triangle and partial correlations given all remaining variables in the upper triangle of the HMM for state 4.*

	BTC	ETH	XPR	LTC	BCH
BTC	1.0000	0.1006	-0.1251	0.2123	0.3688
ETH	0.3067	1.0000	-0.0920	0.1482	0.1434
XRP	0.2121	0.0934	1.0000	-0.0096	0.4985
LTC	0.5064	0.3406	0.3054	1.0000	0.3938
BCH	0.5680	0.3340	0.5490	0.6240	1.0000

Table 21: *Estimated marginal correlations in the lower triangle and partial correlations given all remaining variables in the upper triangle of the HMM for state 5.*

	BTC	ETH	XPR	LTC	BCH
BTC	1.0000	0.2728	0.2246	0.2674	0.2776
ETH	0.6307	1.0000	0.1670	0.0812	0.0233
XRP	0.7326	0.5894	1.0000	0.3172	0.2140
LTC	0.7560	0.5743	0.7561	1.0000	0.3486
BCH	0.7392	0.5406	0.7181	0.7634	1.0000

Table 22: *Estimated transition probabilities for the HMM with $k = 5$ latent states.*

	1	2	3	4	5
1	0.3413	0.1104	0.4413	0.0385	0.0684
2	0.0000	0.7030	0.1938	0.1032	0.0000
3	0.4128	0.1307	0.2748	0.0591	0.1227
4	0.0616	0.1312	0.0503	0.7333	0.0237
5	0.3402	0.0000	0.0000	0.0198	0.6400

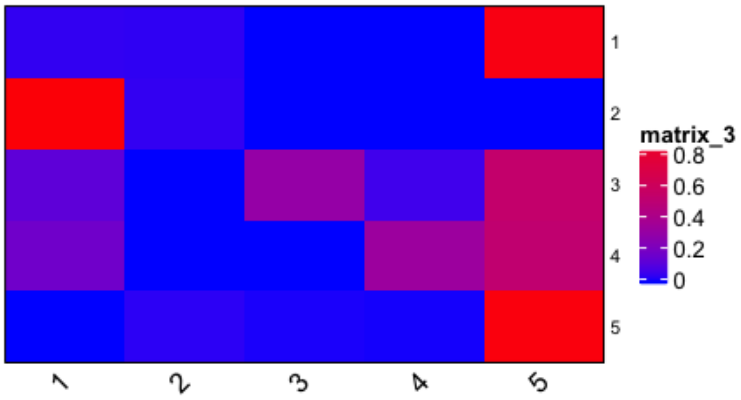


Figure 7: *Heat transition matrix with 5 hidden states.*

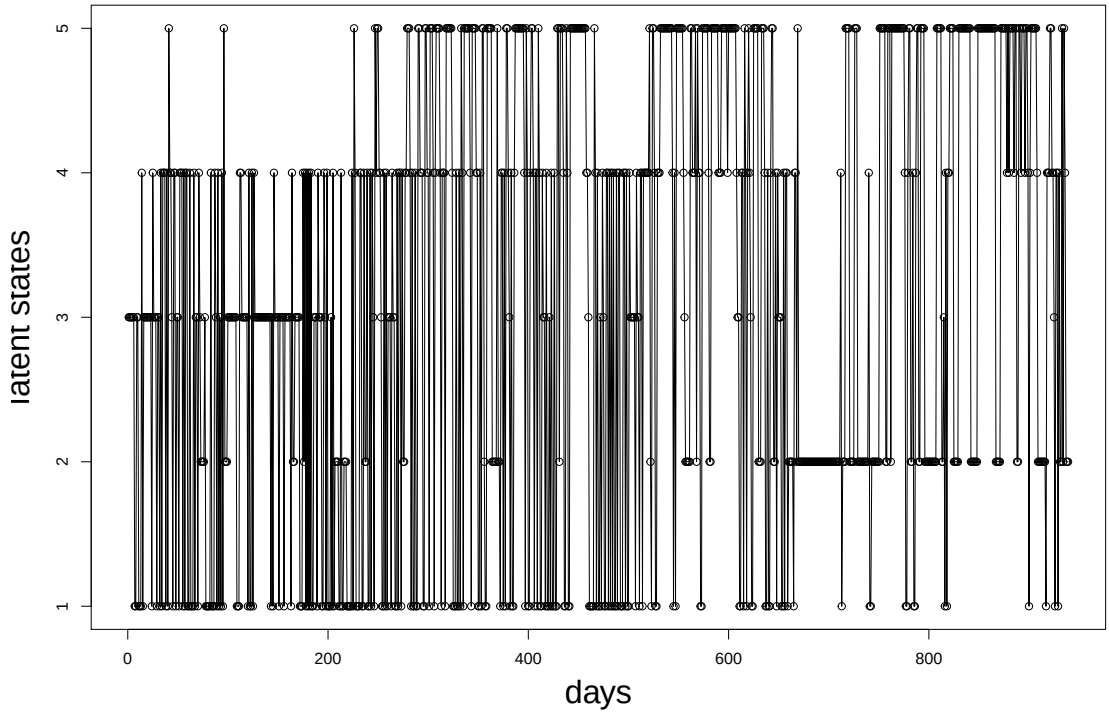


Figure 8: *Decoded states for the HMM with 5 states.*

References

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2014). Latent Markov models: A review of a general framework for the analysis of longitudinal data with covariates. *TEST*, 23:433–465.

Bartolucci, F., Pandolfi, S., and Pennoni, F. (2017). `LMest`: An r package for latent Markov models for longitudinal categorical data. *Journal of Statistical Software*, 81:1–38.

Dempster, A. P., Laird, N. M., and 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.

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

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