Expectile Hidden Markov Regression Models for Analysing Cryptocurrency Returns

XIII Giornata della Ricerca MEMOTEF

Beatrice Foroni[1], Luca Merlo[2], Lea Petrella[1]

June 27-28, 2023

[1]MEMOTEF Department, Sapienza, University of Rome

[2]Department of Human Sciences, European University of Rome

Introduction

Introduction

- In this work we develop a linear expectile hidden Markov model (EHMM) for the analysis of cryptocurrency time series with a risk management perspective.
- To the best of our knowledge, a HMM for estimating conditional expectiles has not yet been proposed in the literature.
- The methodology allows to focus on the tails of returns distribution and describe their temporal evolution by introducing the model time-dependent coefficients evolving according to a latent discrete homogeneous Markov chain.

Introduction (cont.)

- Estimation of the model parameters is based on the asymmetric normal distribution.
- Maximum likelihood estimates are obtained via an Expectation-Maximization algorithm using efficient M-step update formulas for all parameters.
- We evaluate the introduced method with both artificial data and real data investigating the relationship between daily Bitcoin returns and major world market indices.

Financial Applications

- HMMs have become very popular in the financial time series literature since the seminal works of Hamilton (1989).
- Univariate applications of HMMs to asset allocation, stock returns, and financial data are discussed, for example, in Langrock et al. (2012); Cavaliere et al. (2014); Nystrup et al. (2017); Maruotti et al. (2019).
- Multivariate extensions have been proposed in Bernardi et al. (2017).

Goals

- Capture stylized facts of financial time series, as for example serial correlation and heterogeneity, by letting the parameter selection process be driven by an unobserved (i.e. hidden) Markov chain.
- Build dynamic models in a risk management framework, which may be of greater importance for investors and policymakers, especially after the financial crisis of the last 15 years.

Hidden Markov Model (HMM)

Hidden Markov Model (HMM)

- A HMM is a particular kind of dependent mixture model consisting of two parts: an underlying unobserved process and a state-dependent process (Zucchini et al., 2016)
- Let $\{S_t\}_{t=1}^{T}$ be a first-order Markov chain defined on the discrete state space $\{1, ..., K\}$.
- The process $\{S_t\}$, which represents the underlying unobserved process of the HMM, fulfills the Markov property

$$\mathbb{P}(S_t = s_t | S_{t-1} = s_{t-1}, ..., S_1 = s_1) = \mathbb{P}(S_t = s_t | S_{t-1} = s_{t-1}) \tag{1}$$

Hidden Markov Model (HMM)

- Let $\{X_t\}_{t=1}^{T}$ denote a sequence of observations
- The process $\{X_t\}$ represents the state-dependent process of the HMM and fulfills the conditional (on the hidden states) independence property

$$\begin{aligned} &\mathbb{P}(X_t = x_t | X_1 = x_1, ..., X_{t-1} = x_{t-1}, S_1 = s_1, ..., S_t = s_t) \\ &= \mathbb{P}(X_t = x_t | S_t = s_t) \end{aligned} \quad (2)$$

- Then, the pair of stochastic processes $\{(S_t, X_t)\}$ is called a K-state Hidden Markov Model.

Expectile Hidden Markov Model

Expectile Hidden Markov Model

- Proposed by Newey and Powell (1987), expectile regression is a "quantile-like" generalization of standard mean regression based on asymmetric least-squares estimation.
- Is an alternative to quantile regression approach for characterizing the entire conditional distribution of a response variable.
- Formally, the expectile of order $\tau \in (0, 1)$ of a continuous response Y given the P-dimensional vector of covariates $\mathbf{X} = \mathbf{x}$, is defined as

$$\mu_{\mathbf{x}}(\tau) = \arg\min_{\mu \in \mathbb{R}} \mathbb{E}[\omega_\tau(Y - \mu_{\mathbf{x}}(\tau))] \quad (3)$$

Expectile Hidden Markov Model (cont.)

- $\omega_\tau(u) = u^2|\tau - \mathbb{I}(u < 0)|$ is the asymmetric square loss function.
- In a regression framework, for a given τ, a linear expectile model is defined as $\mu_{\mathbf{x}}(\tau) = \mathbf{x}'\beta(\tau)$.
- If $\tau = 1/2$ expectile regression reduces to the standard mean regression.
- Advantages: uniqueness of the ML solutions, differentiability of the squared loss function.

Expectile Hidden Markov Model (cont.)

- Let $\{S_t\}_{t=1}^{T}$ be a latent, homogeneous, first-order Markov chain defined on the discrete state space $\{1, ..., K\}$.
- We collect the initial and transition probabilities in the K-dimensional vector π and in the $K \times K$ Π matrix, respectively.
- Let Y_t denote a continuous observable response variable and $\mathbf{X}_t = (1, X_{t2}, \ldots, X_{tP})'$, be a vector of P exogenous covariates, at time $t = 1, ..., T$.
- The proposed EHMM is defined as $Y_t = \mathbf{X}_t' \beta_k(\tau) + \epsilon_{tk}(\tau)$.

Expectile Hidden Markov Model (cont.)

- Extending the approach of Waldmann et al. (2017) to the HMM setting, we use the AN distribution to describe the conditional distribution of the response given covariates and the state occupied by the latent process at time t.

AN density function

$$f_Y(y_t|\mathbf{X}_t = \mathbf{x}_t, S_t = k) = \frac{2\sqrt{\tau(1-\tau)}}{\sqrt{\pi\sigma_k^2}(\sqrt{\tau}+\sqrt{1-\tau})} \exp\left[-\omega_\tau\left(\frac{y_t-\mu_{tk}}{\sigma_k}\right)\right], \tag{4}$$

- μ_{tk} is defined by the linear model $\mu_{tk} = \mathbf{x}_t'\beta_k(\tau)$.
- We use the AN distribution as a working likelihood and we develop an EM algorithm (Baum et al., 1970) to estimate the parameters.
- The density kernel of AN corresponds to the expectile loss function.

Expectile Hidden Markov Model (cont.)

- For a given number of hidden states K, the EM algorithm runs on the complete log-likelihood function of the model introduced, which is defined as

$$\ell_c(\theta_\tau) = \sum_{k=1}^{K} \gamma_1(k) \log \pi_k + \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{j=1}^{K} \xi_t(j,k) \log \pi_{k|j} + \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_t(k) \log f_Y(y_t | \mathbf{x}_t, S_t = k), \tag{5}$$

- $\theta_\tau = (\beta_1, \ldots, \beta_K, \sigma_1, \ldots, \sigma_K, \pi, \Pi)$ represents the vector of all model parameters.
- $\gamma_t(k)$ denotes a dummy variable equal to 1 if the latent process is in state k at occasion t and 0 otherwise.
- $\xi_t(j,k)$ is a dummy variable equal to 1 if the process is in state j in $t-1$ and in state k at time t and 0 otherwise.

Expectile Hidden Markov Model (cont.)

- To estimate θ_τ, the algorithm iterates between, the E- and M-steps, until convergence.

E-step:

- At iteration (h+1) the unobservable indicator variables $\gamma_t(k)$ and $\xi_t(j,k)$ in (5) are replaced by their conditional expectations given the observed data and the current parameter estimates $\theta_\tau^{(h)}$.
- To obtain the conditional expectations, we obtain that

$$\gamma_t^{(h)}(k) = P_{\theta_\tau^{(h)}}(S_t = k | y_1, \ldots, y_T), \tag{6}$$

$$\xi_t^{(h)}(j,k) = P_{\theta_\tau^{(h)}}(S_{t-1} = j, S_t = k | y_1, \ldots, y_T) \tag{7}$$

using the Forward-Backward algorithm of Welch (2003).

Expectile Hidden Markov Model (cont.)

- The conditional expectation of the complete log-likelihood function in (5) given the observed data and the current estimates is

$$Q(\theta_\tau|\theta_\tau^{(h)}) = \sum_{k=1}^{K} \gamma_1^{(h)}(k) \log \pi_k + \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{j=1}^{K} \xi_t^{(h)}(j,k) \log \pi_{k|j} + \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_t^{(h)}(k) \log f_Y(y_t|\mathbf{x}_t, S_t = k).$$

Expectile Hidden Markov Model (cont.)

M-step:

- In the M-step we maximize $Q(\theta_\tau|\theta_\tau^{(h)})$ with respect to θ_τ to obtain the update parameter estimates $\theta_\tau^{(h+1)}$.
- The maximization of $Q(\theta_\tau|\theta_\tau^{(h)})$ can be partitioned into orthogonal subproblems.
- Initial probabilities π_k and transition probabilities $\pi_{k|j}$ are updated using:

$$\pi_k^{(h+1)} = \gamma_1^{(h)}(k), \quad k = 1, \dots, K$$

and

$$\pi_{k|j}^{(h+1)} = \frac{\sum_{t=1}^{T} \xi_t^{(h)}(j,k)}{\sum_{t=1}^{T} \sum_{k=1}^{K} \xi_t^{(h)}(j,k)}, \quad j,k = 1, \dots, K.$$

Expectile Hidden Markov Model (cont.)

- the M-step update expression for β_k and σ_k^2 given the first-order condition yields, for $k = 1, ..., K$ are

$$\beta_k^{(h+1)} = \Big(\sum_{t=1}^{T} \gamma_t^{(h)}(k)|\tau - \mathbb{I}(y_t < \mathbf{x}_t'\beta_k)|\mathbf{x}_t\mathbf{x}_t'\Big)^{-1}\Big(\sum_{t=1}^{T} \gamma_t^{(h)}(k)|\tau - \mathbb{I}(y_t < \mathbf{x}_t'\beta_k)|\mathbf{x}_t y_t\Big), \quad (8)$$

$$\sigma_k^{2(h+1)} = \frac{2}{\sum_{t=1}^{T}\gamma_t^{(h)}(k)} \sum_{t=1}^{T} \gamma_t^{(h)}(k)|\tau - \mathbb{I}(y_t < \mathbf{x}_t'\beta_k^{(h+1)})|(y_t - \mathbf{x}_t'\beta_k^{(h+1)})^2, \quad (9)$$

- which can be computed using Iterative Reweighted Least Squares data with appropriate weights.
- For fixed τ and K we initialize the EM algorithm by providing the initial states partition, $\{S_t^{(0)}\}_{t=1}^{T}$, according to a Multinomial distribution with probabilities $1/K$ (Maruotti et al., 2021; Merlo et al., 2022).

Expectile Hidden Markov Model (cont.)

- From the generated partition, the elements of $\Pi^{(0)}$ are computed as proportions of transition.
- We obtain $\beta_k^{(0)}$ and $\sigma_k^{(0)}$ by fitting mean regressions on the observations within state k.
- To deal with the possibility of multiple roots we fit the proposed EHMM using a multiple random starts strategy with different starting partitions and retain the solution corresponding to the maximum likelihood value.
- To estimate the standard errors we employ the parametric bootstrap scheme of Visser et al. (2000).

Empirical Application

- We apply the methodology proposed to analyze the Bitcoin daily returns as a function of global leading financial indices.
- As predictors we employ Crude Oil, Standard & Poor's 500 (S&P500), Gold COMEX daily closing prices and Volatility Index (VIX) from September 2014 to October 2022.
- We fit the proposed EHMM for different values of K varying from 2 to 5 at three expectile levels $\tau = 0.10, 0.50, 0.90$.
- To compare models with differing number of states we compare three widely employed penalized likelihood selection criteria for K (AIC, BIC, ICL).
- ICL, which favours a more parsimonious choice, select K=2 states for $\tau = \{0.1, 0.5, 0.9\}$.

Results

	Intercept	Crude Oil	S&P500	Gold	VIX	σ_k
State 1						
$\tau = 0.10$	**-1.036 (0.280)**	0.024 (0.021)	**0.595 (0.096)**	**0.189 (0.072)**	**0.029 (0.012)**	1.433 (0.040)
$\tau = 0.50$	0.122 (0.158)	0.031 (0.072)	0.409 (0.383)	0.263 (0.249)	0.009 (0.036)	1.695 (0.062)
$\tau = 0.90$	**1.297 (0.061)**	-0.009 (0.020)	**0.589 (0.088)**	**0.134 (0.065)**	0.014 (0.011)	1.335 (0.041)
State 2						
$\tau = 0.10$	**-6.52 (0.06)**	**-0.256 (0.096)**	**2.072 (0.476)**	**1.032 (0.320)**	-0.055 (0.058)	4.964 (0.157)
$\tau = 0.50$	0.242 (0.092)	-0.056 (0.055)	**1.087 (0.357)**	**0.613 (0.214)**	-0.025 (0.026)	6.164 (0.169)
$\tau = 0.90$	**6.244 (0.229)**	0.017 (0.079)	**0.948 (0.291)**	**0.835 (0.249)**	-0.002 (0.041)	4.692 (0.128)

Table 1: State-specific parameter estimates for three expectile levels, with bootstrapped standard errors (in brackets) obtained over 1000 replications. Point estimates are displayed in boldface when significant at the standard 5% level.

Conclusions

- We develop a linear expectile hidden Markov model for the analysis of time series where temporal behaviors of the data are captured via time-dependent coefficients that follow an unobservable discrete homogeneous Markov chain.
- We analyze the association between Bitcoin and a collection of global market indices, not only at the average, but also during times of market distress.
- Empirically, we find evidence of strong and positive interrelations among Bitcoin returns and S&P500 and Gold at the tails of the distribution, while no connection emerges during tranquil periods ($\tau = 0.5$).

References

Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171, 1970.

Mauro Bernardi, Antonello Maruotti, and Lea Petrella. Multiple risk measures for multivariate dynamic heavy–tailed models. *Journal of Empirical Finance*, 43:1–32, 2017.

Giuseppe Cavaliere, Michele Costa, and Luca De Angelis. Investigating stock market behavior using a multivariate markov-switching approach. In *Advances in Latent Variables*, pages 185–196. Springer, 2014.

James D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the econometric society*, pages 357–384, 1989.

Roland Langrock, Iain L MacDonald, and Walter Zucchini. Some nonstandard stochastic volatility models and their estimation using structured hidden markov models. *Journal of Empirical Finance*, 19(1):147–161, 2012.

Antonello Maruotti, Antonio Punzo, and Luca Bagnato. Hidden markov and semi-markov models with multivariate leptokurtic-normal components for robust modeling of daily returns series. *Journal of Financial Econometrics*, 17(1):91–117, 2019.

Antonello Maruotti, Lea Petrella, and Luca Sposito. Hidden semi-Markov-switching quantile regression for time series. *Computational Statistics & Data Analysis*, 159: 107208, 2021.

Luca Merlo, Antonello Maruotti, Lea Petrella, and Antonio Punzo. Quantile hidden semi-Markov models for multivariate time series. *Statistics and Computing*, 32(4): 1–22, 2022.

Whitney K Newey and James L Powell. Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, pages 819–847, 1987.

Peter Nystrup, Henrik Madsen, and Erik Lindström. Long memory of financial time series and hidden markov models with time-varying parameters. *Journal of Forecasting*, 36(8):989–1002, 2017.

Ingmar Visser, Maartje EJ Raijmakers, and Peter CM Molenaar. Confidence intervals for hidden Markov model parameters. *British Journal of Mathematical and Statistical Psychology*, 53(2):317–327, 2000.

Elisabeth Waldmann, Fabian Sobotka, and Thomas Kneib. Bayesian regularisation in geoadditive expectile regression. *Statistics and Computing*, 27(6):1539–1553, 2017.

Lloyd R Welch. Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter*, 53(4):10–13, 2003.

Walter Zucchini, Iain L MacDonald, and Roland Langrock. *Hidden Markov models for time series: an introduction using R*. Chapman and Hall/CRC, 2016.

Appendix

Expectile hidden Markov regression models for analyzing cryptocurrency returns

Simulation Study

- We conduct a simulation study to validate the performance of our method under different scenarios in terms of
  1. recovering the true values of the parameters;
  2. assessing the classification behavior of the proposed model;
  3. evaluating the capability of penalized likelihood criteria in selecting the optimal number of hidden states K.
- We analyze two different sample sizes ($T = 500$, $T = 1000$) and two different distributions for the error term for 500 Monte Carlo simulations, similar to Maruotti et al. (2021).

Simulation Study

- We draw observations from a two state HMM ($K = 2$) using

$$Y_t = \begin{cases} -1 + 2X_t + \epsilon_{t1}, & S_t = 1 \\ 1 - 2X_t + \epsilon_{t2}, & S_t = 2, \end{cases} \tag{10}$$

  with $X_t \sim \mathcal{N}(0, 1)$.
- In the first scenario, ϵ_{tk} is generated from a normal distribution with standard deviation 1. In the second one, ϵ_{tk} is generated from a skew-t distribution with 5 degrees of freedom and asymmetry parameter 2, for k = 1,2.
- Matrix of transition probabilities is set equal to $= \left(\begin{smallmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{smallmatrix}\right)$.
- We fit the proposed EHMM at five expectile levels, i.e., $\tau = \{0.10, 0.25, 0.50, 0.75, 0.90\}$ and we calculate bias and standard deviation.

Simulation results

τ	0.10		0.25		0.50		0.75		0.90	
	Bias	Std.Err	Bias	Std.Err	Bias	Std.Err	Bias	Std.Err	Bias	Std.Err
Panel A: T=500										
State 1										
$\beta_{1,1} =$ -1	0.020	0.093	0.010	0.076	-0.002	0.073	-0.018	0.080	-0.048	0.100
$\beta_{2,1} =$ 2	0.001*	0.109	0.001	0.093	0.004	0.087	0.010	0.089	0.023	0.101
State 2										
$\beta_{1,2} =$ 1	0.040	0.055	0.013	0.041	-0.002	0.037	-0.013	0.039	-0.027	0.047
$\beta_{2,2} =$ -2	-0.008	0.067	0.001*	0.058	0.001	0.055	-0.003	0.058	-0.012	0.068
Panel B: T = 1000										
State 1										
$\beta_{1,1} =$ -1	0.021	0.068	0.010	0.055	-0.001	0.051	-0.016	0.056	-0.042	0.070
$\beta_{2,1} =$ 2	0.006	0.071	0.003	0.060	0.004	0.057	0.008	0.060	0.017	0.069
State 2										
$\beta_{1,2} =$ 1	0.039	0.038	0.014	0.029	0.001*	0.026	-0.010	0.028	-0.023	0.034
$\beta_{2,2} =$ -2	-0.012	0.053	-0.004	0.045	-0.002	0.043	-0.005	0.044	-0.014	0.050

Table 2: Bias and standard error values of the state-regression parameter estimates with Gaussian distributed errors for $T = 500$ (Panel A) and $T = 1000$ (Panel B). * represents values smaller (in absolute value) than 0.001.

Simulation results (cont.)

τ	0.10		0.25		0.50		0.75		0.90	
	Bias	Std.Err	Bias	Std.Err	Bias	Std.Err	Bias	Std.Err	Bias	Std.Err
Panel A: T=500										
State 1										
$\beta_{1,1} =$ -1	-0.064	0.228	-0.009	0.119	-0.004	0.099	-0.018	0.166	0.083	0.489
$\beta_{2,1} =$ 2	-0.153	0.319	-0.049	0.183	0.001	0.125	0.027	0.161	-0.045	0.471
State 2										
$\beta_{1,2} =$ 1	0.181	0.144	0.066	0.073	0.013	0.051	-0.018	0.060	-0.052	0.131
$\beta_{2,2} =$ -2	-0.060	0.100	-0.029	0.080	-0.015	0.074	-0.015	0.079	-0.038	0.103
Panel B: T = 1000										
State 1										
$\beta_{1,1} =$ -1	-0.053	0.162	0.001	0.077	-0.004	0.069	-0.022	0.123	0.025	0.369
$\beta_{2,1} =$ 2	-0.133	0.236	-0.024	0.111	0.013	0.082	0.039	0.116	0.027	0.319
State 2										
$\beta_{1,2} =$ 1	0.168	0.101	0.057	0.048	0.010	0.035	-0.019	0.041	-0.059	0.083
$\beta_{2,2} =$ -2	-0.067	0.067	-0.033	0.052	-0.016	0.048	-0.015	0.054	-0.031	0.072

Table 3: Bias and standard error values of the state-regression parameter estimates with skew-t distributed errors for $T = 500$ (Panel A) and $T = 1000$ (Panel B).

Simulation results (cont.)

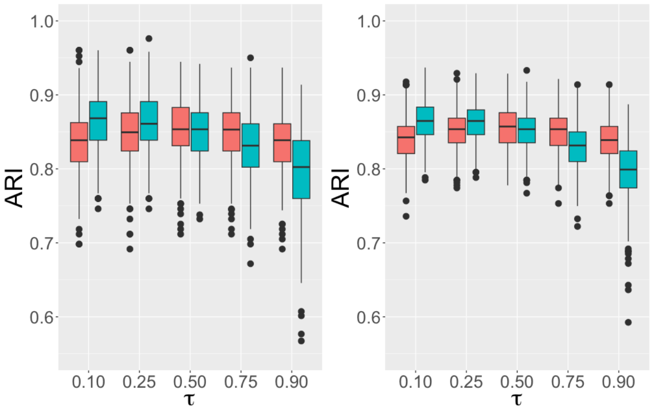

Figure 1: From left to right, box-plots of ARI for the posterior probabilities for Gaussian (red) and skew-t (blue) distributed errors with $T = 500$ and $T = 1000$.

Simulation results (cont.)

τ	0.10			0.50			0.90		
	AIC	BIC	ICL	AIC	BIC	ICL	AIC	BIC	ICL
Panel A: Gaussian errors									
K = 1	0	0	0	0	0	0	0	0	0
K = 2	0	84	94	42	84	84	0	81	94
K = 3	10	13	3	35	6	6	9	14	1
K = 4	90	4	4	23	10	10	91	5	5
Panel B: skew-t errors									
K = 1	0	0	0	0	0	0	0	0	0
K = 2	0	0	52	0	3	86	0	0	55
K = 3	0	79	44	0	70	12	1	77	39
K = 4	100	21	4	100	28	2	99	23	7

Table 4: Percentage frequency distribution of the selected number of hidden states K under Gaussian and skew-t errors over 300 replications. We draw observations from a two state HMM ($K = 2$), and we fit the EHMM with $K = 1, 2, 3, 4$ in order to select the best K associated to the lowest penalized likelihood criteria.