Regime-switching factor models for cryptocurrencies

Brian Zhu

22 December 2021

Abstract

In this paper, we consider a Markov regime-switching factor model with the goal of explaining asset returns while considering cyclical patterns present in the market. To estimate the model's parameters, we formulate an expectation-maximization algorithm that only requires running forward-backward algorithms and linear regressions. The model and algorithm are empirically tested on cryptocurrency data over the past year with a three-factor model. The results suggest that a moderate number of regimes are optimal based on the Bayesian information criterion and that our algorithm has desirable properties, while justifying the choice of factors and providing another interpretation of explained asset returns via regime-dependent parameters.

1 Introduction

A *regime-switching model* associates observed data to latent variables representing the regime (e.g. a bullish or bearish market) that the observed data comes from, where the latent regime variable may determine some or all parameters of a model for the data. The latent variables are typically modelled with a Markov chain, because of which regime-switching models are also known as *Markov-switching models* or *Markov regime-switching models*. In finance, regime-switching models have been prized for their ability to the capture cyclical patterns of financial markets. The first application of a regime-switching model in finance was in Hamilton (1989), where such a model was proposed to detect the breakpoints of business cycles in autoregressive time series. Another notable application is the family of Markov-switching conditional heteroskedasticity models first explored by Hamilton & Sumsel (1994), where the regime-path dependence of the conditional variance poses a tractability problem and thus has been the subject of further research (see Augustyniak (2014) for expectation-maximization algorithms or Billio et al. (2016) for Gibbs sampling methods, for example).

A *factor model* relates N variables, typically macroeconomic variables or asset returns, to M factors by fitting a *loading matrix* and a diagonal *idiosyncratic risk matrix*. When $N \gg M$, factor models can be seen as a form of dimension reduction. Some common factor models for stock returns are the *capital asset pricing model (CAPM)*, first published in Sharpe (1964), with market returns as the sole factor, and the Fama-French three-factor model of Fama & French (1993), which considers market returns, size (market capitalization), and value (book-to-market ratio) as factors.

The first appearance of a regime-switching factor model was in Kim & Yoo (1995) in a macroeconomic context, proposing a nested model $y_t = \Lambda f_t + u_t$ where the factors f_t and idiosyncratic risk components u_t were further modeled with regime-dependent vector autoregression models. Since then, regime-switching models have appeared in financial contexts, with between two (Nystrup et al. 2018) and four (Bae et al. 2014) regimes identified as suitable for best explaining stock returns. Costa & Kwon (2019) and Costa & Kwon (2020) proposed models $r_t = \alpha + \sum_s B_s f_t \mathbb{I}_s + \varepsilon_t$ and $r_t = \sum_s (\alpha_s + B_s f_t + \varepsilon_{s,t})\mathbb{I}_s$, respectively (where s denotes a state-dependent parameter or indicator function), for the purpose of improving estimates of the mean vector and covariance matrix of asset returns for mean-variance portfolio optimization.

In this paper, we consider the model of Costa & Kwon (2020). To estimate the parameters of this model, we describe an expectation-maximization algorithm to maximize the observed likelihood that only requires running *forward-backward algorithms (FBAs)* and linear regressions, as opposed to numerical optimization algorithms. To our knowledge, this is the first time that the maximization step has been explicitly described; it is notable that linear regressions can still be used to fit factor models when there is regime-switching. We then fit models using cryptocurrency data.

Simply speaking, cryptocurrencies are digital currencies, i.e. binary data designed to serve as a medium of exchange and storage of value. The "crypto" part refers to the cryptographic *blockchain* technology that is used to record transactions. Notably, cryptocurrencies use a decentralized form of control via networks of computers "mining" cryptocurrencies, i.e. verifying transactions, as opposed to traditional currencies or central bank digital currencies controlled by a central bank. While Liu & Tsyvinski (2021) suggests that cryptocurrency returns behave differently compared to traditional currency returns and casts doubt that cryptocurrencies serve the two aforementioned purposes of a currency, the cryptocurrency market has grown substantially in value and notoriety, especially over the past two years. The total market capitalization has grown from circa \$200 billion in November 2019 to over \$2500 billion as of November 2021, and the total number of cryptocurrencies has nearly tripled from 2817 to 7557 over the same time period.[1,2] Social media and celebrities have induced sudden periods of rapid appreciation or depreciation in value. These peculiarities of cryptocurrencies have made the cryptocurrency market appealing to traders and academics alike.

Urquhart (2016) finds that Bitcoin returns are predictable, suggesting that the *efficient market hypothesis* does not hold for the cryptocurrency market, with interesting examples of good predictors including Google Trends data (Urquhart (2018)) and Twitter activity (Shen et al. (2019)). This has inspired the development of factor models to find risk factors that may explain cryptocurrency returns. Hubrich (2017) proposes a four-factor model of market returns, value (transaction volume to market capitalization ratio), momentum (prior returns), and carry (discrepancies between spot and futures markets), and Shen et al. (2020) examines a three-factor model of market returns, size, and reversals (prior returns). Given the recent growth and volatility of the cryptocurrency market, a factor model analysis with more recent data and an introduction of regime-switching framework with the goal of learning said volatility is relevant and is the motivation of this paper.

In the empirical part of this paper, we consider a three-factor model of market returns, size, and momentum. We fit regime-switching factor models with varying numbers of regimes, before analyze the one- and two-regime models in detail. The rest of this paper is organized as follows: Section 2 describes the regime-switching factor model details and inference for hidden Markov models, Section 3 describes the algorithm, Section 4 covers the empirical results, and Section 5 concludes.

[1]https://www.statista.com/statistics/730876/cryptocurrency-maket-value/
[2]https://www.statista.com/statistics/863917/number-crypto-coins-tokens/

2 Model

2.1 The regime-switching factor model

Suppose there are N assets and M factors. Let $r_{i,t}$ be the observed return on asset i and $f_{t,j}$ be the observed factor j at time t for $t \in \{1, \ldots, T\}$. Denote $r_t = (r_{1,t}, \ldots, r_{N,t})$ and $f_t = (f_{t,1}, \ldots, f_{t,M})$. Let $z_t \in \{1, \ldots, S\}$ be a latent state variable at time t governed by an S-state discrete-time Markov chain with initial distribution $p \in \mathbb{R}^S$ and transition matrix $P \in \mathbb{R}^{S \times S}$. The *regime-switching factor model (RSFM)* with S states and state-dependent idiosyncratic risk is

$$r_t = \sum_{s=1}^{S} (\alpha_s + B_s f_t + \varepsilon_{s,t}) \cdot \mathbb{I}[z_t = s]$$

where $\alpha_s = (\alpha_{1,s}, \ldots, \alpha_{N,s})^\top \in \mathbb{R}^N$ is the excess return vector, $B_s \in \mathbb{R}^{N \times M}$ is the loading matrix, and $\varepsilon_{s,t} \sim \mathcal{N}(0_N, \Omega_s)$ where $\Omega_s = \mathrm{diag}(\sigma^2_{1,s}, \ldots, \sigma^2_{N,s})$ is the idiosyncratic risk matrix in state s.

2.2 Inference for hidden Markov models

Per Kuan (2002), we define important quantities for inference in *hidden Markov models (HMMs)*. Let $\mathcal{R}_t = (r_1, \ldots, r_t)$ and $r = (r_1, \ldots, r_T)$. The *prediction probabilities* $\mathbb{P}[z_t = s | \mathcal{R}_{t-1}, \theta]$ predict the latent variable distributions at time t given past observations and satisfy

$$\mathbb{P}[z_t = s | \mathcal{R}_{t-1}, \theta] = \sum_{s'=1}^{S} \mathbb{P}[z_{t-1} = s' | \mathcal{R}_{t-1}, \theta] \cdot P(s', s)$$

where $P(s', s)$ is the (s', s)-th entry of P. The *filtering probabilities* $\mathbb{P}[z_t = s | \mathcal{R}_t, \theta]$ predict the latent variable distributions at time t given past and current observations, and satisfy

$$\mathbb{P}[z_t = s | \mathcal{R}_t, \theta] = \frac{\mathbb{P}[z_t = s | \mathcal{R}_{t-1}, \theta] \cdot \mathbb{P}[r_t | z_t = s, \mathcal{R}_{t-1}, \theta]}{\sum_{s'=1}^{S} \mathbb{P}[z_t = s' | \mathcal{R}_{t-1}, \theta] \cdot \mathbb{P}[r_t | z_t = s', \mathcal{R}_{t-1}, \theta]}$$

where $\mathbb{P}[r_t | z_t = s, \mathcal{R}_{t-1}, \theta]$ is a conditional density of r_t. Given p and P, the *forward-backward algorithm for hidden Markov models (HMM-FBA)* first computes prediction and filtering probabilities iteratively starting from $t = 1$ and going forward:

$$\begin{aligned}
\mathbb{P}[z_1 = s | \mathcal{R}_0, \theta] &= p(s) \\
\mathbb{P}[z_1 = s | \mathcal{R}_1, \theta] &= \frac{\mathbb{P}[z_1 = s | \mathcal{R}_0, \theta] \cdot \mathbb{P}[r_1 | z_1 = s, \mathcal{R}_0, \theta]}{\sum_{s'=1}^{S} \mathbb{P}[z_1 = s' | \mathcal{R}_0, \theta] \cdot \mathbb{P}[r_1 | z_1 = s', \mathcal{R}_0, \theta]} \\
\mathbb{P}[z_2 = s | \mathcal{R}_1, \theta] &= \sum_{s'=1}^{S} \mathbb{P}[z_1 = s' | \mathcal{R}_1, \theta] \cdot P(s', s) \\
&\vdots \\
\mathbb{P}[z_{T+1} = s | \mathcal{R}_T, \theta] &= \sum_{s'=1}^{S} \mathbb{P}[z_T = s | \mathcal{R}_T, \theta] \cdot P(s', s)
\end{aligned}$$

where $p(s)$ is the s-th entry of p. The *smoothing probabilities* $\mathbb{P}[z_t = s|r,\theta]$ assign distributions for the latent variable at time t given all observed data and, per Kim (1994), is approximately

$$\mathbb{P}[z_t = s|r,\theta] \approx \mathbb{P}[z_t = s|\mathcal{R}_t,\theta] \sum_{s'=1}^{S} \frac{P(s,s') \cdot \mathbb{P}[z_{t+1} = s'|r,\theta]}{\mathbb{P}[z_{t+1} = s'|\mathcal{R}_t,\theta]}$$

Given the prediction and filtering probabilities, the smoothing probabilities can be iteratively computed starting from $t = T$ and going backward, completing the HMM-FBA:

$$\begin{aligned}
\mathbb{P}[z_T = s|r,\theta] &= \mathbb{P}[z_T = s|\mathcal{R}_T,\theta] \sum_{s'=1}^{S} \frac{P(s,s') \cdot \mathbb{P}[z_{T+1} = s'|r,\theta]}{\mathbb{P}[z_{T+1} = s'|\mathcal{R}_T,\theta]} \\
\mathbb{P}[z_{T-1} = s|r,\theta] &= \mathbb{P}[z_{T-1} = s|\mathcal{R}_{T-1},\theta] \sum_{s'=1}^{S} \frac{P(s,s') \cdot \mathbb{P}[z_T = s'|r,\theta]}{\mathbb{P}[z_T = s'|\mathcal{R}_{T-1},\theta]} \\
&\vdots \\
\mathbb{P}[z_1 = s|r,\theta] &= \mathbb{P}[z_1 = s|\mathcal{R}_1,\theta] \sum_{s'=1}^{S} \frac{P(s,s') \cdot \mathbb{P}[z_2 = s'|r,\theta]}{\mathbb{P}[z_2 = s'|\mathcal{R}_1,\theta]}
\end{aligned}$$

Note that $\mathbb{P}[z_{T+1} = s'|r,\theta] = \mathbb{P}[z_{T+1} = s'|\mathcal{R}_T,\theta]$ in the first equation above.

2.3 Practical considerations

One problem that arises in practice when computing the prediction and filtering probabilties is that $\mathbb{P}[r_t|z_t = s,\mathcal{R}_{t-1},\theta]$ may be very small. To resolve this issue, we work with the log-probability:

$$\begin{aligned}
&\log \frac{\mathbb{P}[z_t = s|\mathcal{R}_{t-1},\theta] \cdot \mathbb{P}[r_t|z_t = s,\mathcal{R}_{t-1},\theta]}{\sum_{s'=1}^{S} \mathbb{P}[z_t = s'|\mathcal{R}_{t-1},\theta] \cdot \mathbb{P}[r_t|z_t = s',\mathcal{R}_{t-1},\theta]} \\
&= \log \mathbb{P}[z_t = s|\mathcal{R}_{t-1},\theta] + \log \mathbb{P}[r_t|z_t = s,\mathcal{R}_{t-1},\theta] - \log \sum_{s'=1}^{S} \mathbb{P}[z_t = s'|\mathcal{R}_{t-1},\theta] \cdot \mathbb{P}[r_t|z_t = s',\mathcal{R}_{t-1},\theta]
\end{aligned}$$

Now the first two terms above are easy to compute, but the third term may still be intractable. For $x_1,\ldots,x_n \le 0$, define $\mathrm{LSE}(x_1,\ldots,x_n) = \log(\exp x_1 + \cdots + \exp x_n)$ and observe that

$$\begin{aligned}
\log(x_1 + \cdots + x_n) &= \mathrm{LSE}(\log x_1,\ldots,\log x_n) \\
\mathrm{LSE}(x_1,\ldots,x_n) &= x_{\max} + \log(\exp(x_1 - x_{\max}) + \cdots + \exp(x_n - x_{\max}))
\end{aligned}$$

where $x_{\max} = \max\{x_1,\ldots,x_n\}$. Let $\ell_t(s) = \log \mathbb{P}[z_t = s|\mathcal{R}_{t-1},\theta] + \log \mathbb{P}[r_t|z_t = s,\mathcal{R}_{t-1},\theta]$. Then

$$\log \sum_{s'=1}^{S} \mathbb{P}[z_t = s'|\mathcal{R}_{t-1},\theta] \cdot \mathbb{P}[r_t|z_t = s',\mathcal{R}_{t-1},\theta] = \mathrm{LSE}(\ell_t(1),\ldots,\ell_t(S))$$

Note that all $\ell_t(s)$ are tractable and the $\ell_t(s) - \ell_t(\max)$ where $\ell_t(\max) = \max\{\ell_t(1),\ldots,\ell_t(S)\}$ are much closer to zero than the $\ell_t(s)$, so $\mathrm{LSE}(\ell_t(1),\ldots,\ell_t(S))$ is tractable, unless in extreme cases.

3 Method

3.1 Motivation

We would like to estimate a RSFM's *factor model parameters* α_s, B_s, and Ω_s for $s \in \{1, \ldots, S\}$ and *Markov chain parameters* p and P by maximizing the observed likelihood $\mathbb{P}[r|\theta]$ via an *expectation-maximization (EM)* algorithm. Let $\theta = (\alpha_1, B_1, \Omega_1, \ldots, \alpha_S, B_S, \Omega_S, p, P)$ be the parameter vector and let $z = (z_1, \ldots, z_T)$. Given an initial or previous θ_0, the E-step computes the posterior distribution $\mathbb{P}[z|r, \theta_0]$. The M-step then maximizes $\mathcal{Q}(\theta|\theta_0)$, the *evidence lower bound* given θ_0:

$$\begin{aligned}
\mathcal{Q}(\theta|\theta_0) = \mathbb{E}_{z|r,\theta_0}\left[\log \mathbb{P}[r, z|\theta]\right] &= \mathbb{E}_{z|r,\theta_0}\left[\sum_{t=1}^{T} \log \mathbb{P}[r_t, z_t|\theta]\right] \\
&= \sum_{t=1}^{T} \mathbb{E}_{z|r,\theta_0}[\log(\mathbb{P}[z_t|\theta] \cdot \mathbb{P}[r_t|z_t, \theta])] \\
&= \sum_{t=1}^{T} \sum_{s=1}^{S} \mathbb{P}[z_t = s|r, \theta_0](\log \mathbb{P}[z_t = s|\theta] + \log \mathbb{P}[r_t|z_t = s, \theta])
\end{aligned}$$

As the $\mathbb{P}[z_t = s|r, \theta_0]$ term above is the smoothing probability of state s at time t under θ_0, it thus suffices to compute the smoothing probabilities for the E-step. Let $q_t(s) = \mathbb{P}[z_t = s|r, \theta_0]$. Then

$$\begin{aligned}
\mathcal{Q}(\theta|\theta_0) &= \sum_{t=1}^{T} \sum_{s=1}^{S} q_t(s) \log \mathbb{P}[z_t = s|\theta] + \sum_{t=1}^{T} \sum_{s=1}^{S} q_t(s) \log \mathbb{P}[r_t|z_t = s, \theta] \\
&= \sum_{t=1}^{T} \sum_{s=1}^{S} q_t(s) \log\big(P^t p\big)(s) + \sum_{t=1}^{T} \sum_{s=1}^{S} q_t(s) \log \mathbb{P}[r_t|z_t = s, \theta]
\end{aligned}$$

We first focus on the factor model parameters, so we drop the first double sum as none of its terms depend on them. Let $B_{i,s}$ be the i-th row of B_s. As $r_t - \alpha_s - B_s f_t \sim \mathcal{N}(0_N, \Omega_s)$ for $s \in \{1, \ldots, S\}$,

$$\begin{aligned}
\mathcal{Q}_F(\theta|\theta_0) &= \sum_{t=1}^{T} \sum_{s=1}^{S} q_t(s) \log \mathbb{P}[r_t|z_t = s, \theta] \\
&= \sum_{t=1}^{T} \sum_{s=1}^{S} q_t(s) \left(-\frac{N \log 2\pi}{2} - \frac{\log \det \Omega_s}{2} - \frac{(r_t - \alpha_s - B_s f_t)^\top \Omega_s^{-1} (r_t - \alpha_s - B_s f_t)}{2}\right) \\
&= \sum_{t=1}^{T} \sum_{s=1}^{S} q_t(s) \left(-\frac{\sum_{i=1}^{N} \log 2\pi}{2} - \frac{\sum_{i=1}^{N} \log \sigma_{i,s}^2}{2} - \sum_{i=1}^{N} \frac{(r_t - \alpha_{i,s} - B_{i,s}^\top f_t)^2}{2\sigma_{i,s}^2}\right) \\
&= \sum_{t=1}^{T} \sum_{s=1}^{S} q_t(s) \sum_{i=1}^{N} \left(-\frac{\log 2\pi}{2} - \frac{\log \sigma_{i,s}^2}{2} - \frac{(r_{i,t} - \alpha_{i,s} - B_{i,s}^\top f_t)^2}{2\sigma_{i,s}^2}\right) \\
&= \sum_{s=1}^{T} \sum_{t=1}^{S} \sum_{i=1}^{N} q_t(s) \left(-\frac{\log 2\pi}{2} - \frac{\log \sigma_{i,s}^2}{2} - \frac{(r_{i,t} - \alpha_{i,s} - B_{i,s}^\top f_t)^2}{2\sigma_{i,s}^2}\right)
\end{aligned}$$

Define $\beta_{i,s} = (\alpha_{i,s}, B_{i,s})^\top \in \mathbb{R}^{M+1}$, $x_t = (1, f_{t,1}, \dots, f_{t,M}) \in \mathbb{R}^{M+1}$, and $v_{i,s} = \log \sigma^2_{i,s}$. Using this parameterization and changing the order of summation, we have

$$\mathcal{Q}_F(\theta|\theta_0) = \sum_{i=1}^{N} \sum_{t=1}^{S} \sum_{s=1}^{T} q_t(s) \left(-\frac{\log 2\pi}{2} - \frac{v_{i,s}}{2} - \frac{(r_{i,t} - \beta_{i,s}^\top x_t)^2}{2 \exp v_{i,s}} \right)$$

Note that $\mathcal{Q}_F(\theta|\theta_0)$ is concave in the $\beta_{i,s}$ and $v_{i,s}$, so the first-order conditions (FOCs) for the global maximizer of $\mathcal{Q}_F(\theta|\theta_0)$ are

$$\nabla_{\beta_{i,s}} \mathcal{Q}_F(\theta|\theta_0) = \sum_{t=1}^{T} \frac{q_t(s)(r_{i,t} - \beta_{i,s}^\top x_t) x_t}{\exp v_{i,s}} = 0 \iff \sum_{t=1}^{T} (\sqrt{q_t(s)} r_{i,t} - \beta_{i,s}^\top \sqrt{q_t(s)} f_t) \sqrt{q_t(s)} f_t = 0$$

$$\nabla_{v_{i,s}} \mathcal{Q}_F(\theta|\theta_0) = \sum_{t=1}^{T} \left(-\frac{q_t(s)}{2} + \frac{(r_{i,t} - \beta_{i,s}^\top x_t)^2}{2 \exp v_{i,s}} \right) = 0 \iff \frac{\sum_{t=1}^{T} (r_{i,t} - \beta_{i,s}^\top x_t)^2}{\exp v_{i,s}} = \sum_{t=1}^{T} q_t(s)$$

A quick way of solving the first FOC is to observe that in the ordinary least squares (OLS) setting of $y_t = \beta^\top x_t + \varepsilon_t$, the FOC for β is $\sum_{t=1}^{T} (y_t - \beta^\top x_t) x_t = 0$. By matching terms between the OLS and RSFM settings, it follows that the M-step update for the $\beta_{i,s}$ is

$$\beta^\star_{i,s} = (\tilde{X}_s^\top \tilde{X}_s)^{-1} \tilde{X}_s^\top \tilde{r}_{i,s}$$

where $\tilde{X}_s$ and $\tilde{r}_{i,s}$ are defined as

$$\tilde{X}_s = \begin{bmatrix} \sqrt{q_1(s)} & \sqrt{q_1(s)} f_{1,1} & \cdots & \sqrt{q_1(s)} f_{1,M} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{q_T(s)} & \sqrt{q_T(s)} f_{T,1} & \cdots & \sqrt{q_T(s)} f_{T,M} \end{bmatrix} \quad \tilde{r}_{i,s} = \begin{bmatrix} \sqrt{q_1(s)} r_{i,1} \\ \vdots \\ \sqrt{q_T(s)} r_{i,T} \end{bmatrix}$$

Since $\exp v_{i,s} = \sigma^2_{i,s}$, solving the second FOC gives the M-step update for the $\sigma_{i,s}$:

$$\sigma^\star_{i,s} = \frac{\|\tilde{r}_{i,s} - \tilde{X}_s \beta^\star_{i,s}\|}{(\sum_{t=1}^{T} q_t(s))^{1/2}}$$

We then focus on the Markov chain parameters, updated using the *Baum-Welch forward-backward algorithm (BW-FBA)* of Baum et al. (1970), which first computes $\phi_{s,t} = \mathbb{P}[r_1, \dots, r_t, z_t = s|\theta_0]$ and $\psi_{s,t} = \mathbb{P}[r_{t+1}, \dots, r_T | z_t = s, \theta_0]$, the *forward* and *backward densities*, respectively.

$$\phi_{s,1} = p_0(s) \cdot \mathbb{P}[r_t | z_t = s, \theta_0]$$

$$\phi_{s,t} = \sum_{s'=1}^{S} \phi_{s',t-1} \cdot \mathbb{P}[r_t | z_t = s', \theta_0] \cdot P_0(s', s) \text{ for } t = 2, \dots, T$$

$$\psi_{s,T} = 1$$

$$\psi_{s,t} = \sum_{s'=1}^{S} \psi_{s,t+1} \cdot P_0(s, s') \cdot \mathbb{P}[r_{t+1} | z_{t+1} = s', \theta_0] \text{ for } t = T-1, \dots, 1$$

The smoothing probabilities here are given by

$$q_t(s) = \frac{\phi_{s,t}\psi_{s,t}}{\sum_{s'=1}^{S} \phi_{s',t}\psi_{s',t}}$$

We can also compute *smoothing transition probabilities* $q_t(s, s') = \mathbb{P}[z_t = s, z_{t+1} = s'|r, \theta_0]$ with

$$q_t(s, s') = \frac{\phi_{s,t} \cdot P_0(s, s') \cdot \psi_{s',t} \cdot \mathbb{P}[r_{t+1}|z_{t+1} = s', \theta_0]}{\sum_{\varsigma=1}^{S} \sum_{\varsigma'=1}^{S} \phi_{\varsigma,t} \cdot P_0(\varsigma, \varsigma') \cdot \psi_{\varsigma',t} \cdot \mathbb{P}[r_{t+1}|z_{t+1} = \varsigma', \theta_0]}$$

The Markov chain parameter updates are then given by

$$p^\star(s) = q_1(s)$$

$$P^\star(s, s') = \frac{\sum_{t=1}^{T-1} q_t(s, s')}{\sum_{t=1}^{T-1} q_t(s)}$$

completing the BW-FBA. In practice, the $\phi_{s,t}$ and $\psi_{s,t}$ may be very small, so we can use the methods described in Section 2.3 here as well.

3.2 Description

We now provide a concise description of our algorithm. Note that the E-step can be computed with either the HMM-FBA or BW-FBA. If the HMM-FBA is used in the E-step, note that a BW-FBA still must be used in the M-step for the Markov chain parameters. However, a benefit of the HMM-FBA is that inference probabilities can be collected by iteration.

Algorithm RSFM-FIT.

- Initialize θ_0.
- E-step: compute the $q_t(s)$ and $q_t(s, s')$ under θ_0 according to the HMM-FBA (see Section 2.2) and/or BW-FBA (see Section 3.1).
- M-step: update the parameters according to

$$\beta^\star_{i,s} = (\tilde{X}_s^\top \tilde{X}_s)^{-1} \tilde{X}_s^\top \tilde{r}_{i,s}$$

$$\sigma^\star_{i,s} = \frac{\|\tilde{r}_{i,s} - \tilde{X}_s \beta^\star_{i,s}\|}{(\sum_{t=1}^{T} q_t(s))^{1/2}}$$

$$p^\star(s) = q_1(s)$$

$$P^\star(s, s') = \frac{\sum_{t=1}^{T-1} q_t(s, s')}{\sum_{t=1}^{T-1} q_t(s)}$$

- Set $\theta_0 = \theta^\star$. Repeat the E-step and M-step until a convergence criterion is attained.

We can run the algorithm for multiple random initializations of θ_0 to mitigate the problem of local maxima in the observed likelihood landscape.

4 Results

4.1 Data and factor construction

Cryptocurrency data is obtained with the CoinGecko API and one-month Treasury bill rates, to be used as the risk-free asset, are obtained from the U.S. Department of the Treasury.[3,4] The sample time period, spanning one year and two months, is from 14 October 2020 to 13 December 2021 at a daily frequency. There are 425 total observation periods, the first 15 of which are used to compute returns and factors, so $T = 410$. We select roughly the largest $N = 100$ cryptocurrencies by market capitalization that have available data over the sample period.

As mentioned in Section 1, we consider a three-factor model of the market, size, and momentum, inspired by Hubrich (2017) and Shen et al. (2020), that we will henceforth call the *cryptocurrency three-factor model (C3FM)*. The market factor refers to the return on the *market portfolio*, a portfolio with all sample cryptocurrencies weighted by individual market capitalization. Letting $c_{i,t}$ denote the market capitalization of coin i at time t and $c_t = \sum_{i=1}^{N} c_{i,t}$ denote the total sample market capitalization at time t, the market factor MKT_t is

$$\mathrm{MKT}_t = \sum_{i=1}^{N} r_{i,t} \frac{c_{i,t}}{c_t}$$

The size and momentum factors are constructed in a manner similar to the construction of the factors of Fama & French (1993). We use a holding period duration of $H = 14$ days, i.e. two weeks. Before the start of the sample period and at the end of every holding period, we sort the coins by market capitalization at that time and by mean prior returns over the past H days. Let Portfolio B ("big") contain the top 20 coins by market capitalization and portfolio S ("small") contain the remaining 80. Let Portfolio W ("winner") contain the top 30 coins by prior mean returns, portfolio L ("loser") contain the bottom 30, and portfolio M ("middle") contain the middle 40. We then take intersections of these portfolios to get portfolios BW, BM, BL, SW, SM, and SL. Using the same notation to denote the returns of these portfolios, the size factor SMB_t ("small-minus-big") is

$$\mathrm{SMB}_t = \frac{1}{3}(\mathrm{SW}_t + \mathrm{SM}_t + \mathrm{SL}_t) - \frac{1}{3}(\mathrm{BW}_t + \mathrm{BM}_t + \mathrm{BL}_t)$$

and the momentum factor WML_t ("winner-minus-loser") is

$$\mathrm{WML}_t = \frac{1}{2}(\mathrm{BW}_t + \mathrm{SW}_t) - \frac{1}{2}(\mathrm{BL}_t + \mathrm{SL}_t)$$

The intersection portfolios are updated every H days. The C3FM equation is thus

$$r_t - \mathrm{RF}_t \cdot 1_N = \alpha + B \begin{pmatrix} \mathrm{MKT}_t - \mathrm{RF}_t \\ \mathrm{SMB}_t \\ \mathrm{WML}_t \end{pmatrix} + \varepsilon_t$$

[3] https://www.coingecko.com/en/api

[4] https://home.treasury.gov/

where $\varepsilon_t \sim \mathcal{N}(0, \Omega)$ where $\Omega = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2)$. We then add a regime-switching compoment to the C3FM for the *S-regime regime-switching cryptocurrency three-factor model (RS-C3FM(S))*:

$$r_t - \mathrm{RF}_t \cdot 1_N = \sum_{s=1}^{S} \left(\alpha_s + B_s \begin{pmatrix} \mathrm{MKT}_t - \mathrm{RF}_t \\ \mathrm{SMB}_t \\ \mathrm{WML}_t \end{pmatrix} + \varepsilon_{s,t} \right) \cdot \mathbb{I}[z_t = s]$$

We now present empirical evidence to help justify the choice of factors. Table 1 shows the mean returns of size and momentum quintile portfolios and the t-test results for the highest quintile less the lowest quintile. The quintile portfolios are formed prior to the start of the sample period and at the end of every holding period by sorting the coins by market capitalization or mean prior returns then partitioning the sorted coins into their respective quintiles. There is a statistically significant trend where smaller coins have higher mean returns than bigger coins, and where winning coins have higher mean returns than losing coins. Table 2 shows the mean returns of the intersected size and momentum portfolios that are used to construct the factors. While not as many high-low difference combinations are statistically significant due to the coarser partition, the return anomalies for small and winning coins are still present in moderation.

	Q1	Q2	Q3	Q4	Q5	t-test
Size	0.4789	0.7523	0.6866	0.9303	1.2381	<0.001 (Q1<Q5)
Momentum	0.8121	0.9226	0.5320	0.4227	0.4081	0.031 (Q1>Q5)

Table 1. Mean returns of size and momentum quintile portfolios with paired t-test p-values.

	W	M	L	t-test
B	0.7989	0.4936	0.3561	0.033 (BW>BL)
S	0.8274	0.7567	0.7655	0.342 (SW>SL)
t-test	0.440 (BW<SW)	0.032 (BM<SM)	0.011 (BL<SL)	

Table 2. Mean returns of intersected size and momentum portfolios with paired t-test p-values.

The presence of the size return anomalies is line with previous literature in cryptocurrencies and in asset pricing in general. The presence of the momentum return anomalies is in line with Hubrich (2017), but its relation to Shen et al. (2020) is more complex since it uses a reversal factor, which is akin to a reserved momentum factor LMW ("loser-minus-winner"). While the direction may be reversed, our empirical analysis, Hubrich (2017), and Shen et al. (2020) share the common ground that prior returns are a good predictor of future returns.

4.2 Optimal number of regimes

We fit a RS-C3FM on our data for $S \in \{1, \ldots, 5\}$ regimes using $K = 10$ random initializations and 50 iterations per regime. We record the observed log-likelihood (OLL) of the output θ_k given by

$$\mathrm{OLL}(\theta_k) = \log \mathbb{P}[r|\theta_k] = \sum_{t=1}^{T} \sum_{s=1}^{S} (\log \mathbb{P}[z_t = s|\theta_k] + \log \mathbb{P}[r_t|z_t = s, \theta_k])$$

and the posterior log-likelihood (PLL) given by

$$\text{PLL}(\theta_k) = \sum_{t=1}^{T}\sum_{s=1}^{S}(\log \mathbb{P}[z_t = s|r, \theta_k] + \log \mathbb{P}[r_t|z_t = s, \theta_k])$$

where the latent variables follow the smoothing probabilities $\mathbb{P}[z_t = s|r, \theta_k]$ instead. We compute the Akaike information criterion (AIC) and Bayesian information criterion (BIC) of the best output $\theta^\star = \max\{\theta_1, \dots, \theta_K\}$ given by

$$\begin{aligned}\text{AIC} &= -2 \cdot \text{OLL}(\theta^\star) + 2(2NS + MNS + S^2 - 1)\\ \text{BIC} &= -2 \cdot \text{OLL}(\theta^\star) + (2NS + MNS + S^2 - 1)\log T\end{aligned}$$

and evaluate the model with the BIC due to its asymptotic properties described in Schwarz (1978).[5]

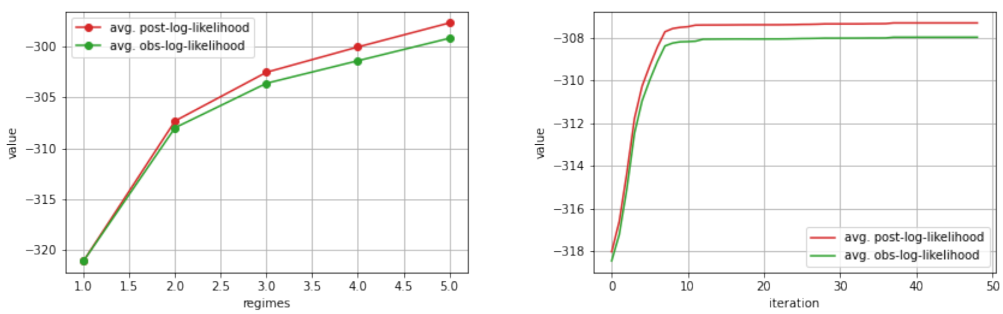

Figure 3. OLL and PLL against number of regimes (left) and EM iterations (right).

The left plot of Figure 3 plots the OLL and PLL over a varying number of regimes. As expected, the OLL and PLL increase in the regime number due to increased model complexity. The right plot of Figure 3 plots the OLL and PLL against iterations in an instance of the EM algorithm described in Section 3. The values increase rapidly at first before converging, supporting the ability of our algorithm to find a local maxima. The OLL and PLL move in tandem in both plots, which suggests that they are good proxies for each other.

The AIC and BIC values for varying numbers of regimes are shown in Table 4. The BIC values suggest that between two and four regimes appear suitable for the RS-C3FM, with three regimes being the most BIC-optimal. The range of regimes that are the most suitable for the cryptocurrency market is similar to that for the stock market per Costa & Kwon (2020), Nystrup et. al (2018), and Bae et. al (2014). Figure 5 plots the AIC and BIC values against the number of regimes.

	1 Regime	2 Regimes	3 Regimes	4 Regimes	5 Regimes
AIC	264294.63	254553.86	251974.75	251167.45	250373.81
BIC	266302.71	258582.07	258031.11	259260.01	260510.59

Table 4. AIC and BIC values for varying numbers of regimes.

[5] Note that the α_s contribute NS degrees of freedom (df), the B_s contribute MNS degrees of freedom, the $\sigma_{i,s}$ contribute NS df, p contributes $(S-1)$ df, and P contributes $S(S-1)$ df towards an S-state M-factor RSFM.

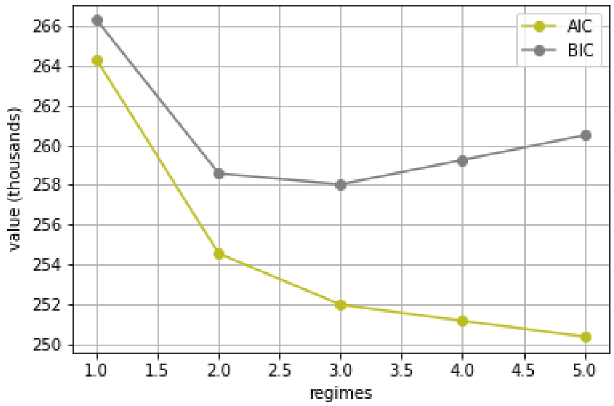

Figure 5. AIC and BIC against number of regimes.

Per Raftery (1995), a difference in BIC of over 10 provides "very strong evidence" for the model with the smallest BIC, so we can consider the BIC-optimal choice of three regimes as significant.

4.3 One-regime factor model analysis

We take a closer look at one-regime factor models to evaluate the explanatory power of the C3FM and establish a baseline of comparison for the two-regime factor models that will be examined later. As a benchmark to assess the C3FM, we first fit the CAPM on our data, given by

$$r_{i,t} - \mathrm{RF}_t = \alpha_i + \beta_{\mathrm{MKT}} \cdot (\mathrm{MKT}_t - \mathrm{RF}_t) + \varepsilon_{i,t}$$

where $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_i^2)$. The C3FM in individual asset form is

$$r_{i,t} - \mathrm{RF}_t = \alpha_i + \beta_{\mathrm{MKT}} \cdot (\mathrm{MKT}_t - \mathrm{RF}_t) + \beta_{\mathrm{SMB}} \cdot \mathrm{SMB}_t + \beta_{\mathrm{WML}} \cdot \mathrm{WML}_t + \varepsilon_{i,t}$$

where $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_i^2)$. Recall that a factor model explains asset returns well for lower values of $|\alpha_i|$, representing abnormal returns not explained by the factors, and higher values of the adjusted R-squared or similar explanatory measures. Lower levels of σ_i, the idiosyncratic risk, are also preferable for most investors. We thus compare the mean absolute α_i, adjusted R-squared, and mean σ_i fit by these two models in Table 6.

	CAPM	C3FM	t-test
mean $\lvert\alpha_i\rvert$	0.3333	0.2960	0.003 (CAPM>C3FM)
mean R^2_{adj}	0.3783	0.4097	<0.001 (CAPM<C3FM)
mean σ_i	6.6653	6.4777	<0.001 (CAPM$\neq$C3FM)

Table 6. Selected parameter means of the CAPM and C3FM with paired t-test p-values.

The mean absolute α_i values are lower with significance and mean adjusted R-squared values are higher with significance for the C3FM compared to the CAPM, suggesting that the C3FM is able to explain cryptocurrency returns better than the CAPM. This constitutes more evidence justifying the choice of factors for the C3FM. The mean σ_i values are also lower with significance.

4.4 Two-regime factor model analysis

We first fit a two-state CAPM on our data. Figure 7 shows the filtering and smoothing probabilities for regime 1, and Figure 8 overlays the most probable regime at each time (with regime 1 blank and regime 2 shaded), based on the smoothing probabilities, on the raw and squared factor time series. In Table 9, using the most probable regime assignments, we fit regular CAPMs to each regime with the observations assigned to that regime. We then compute the means for selected parameters and test them against the one-state model.

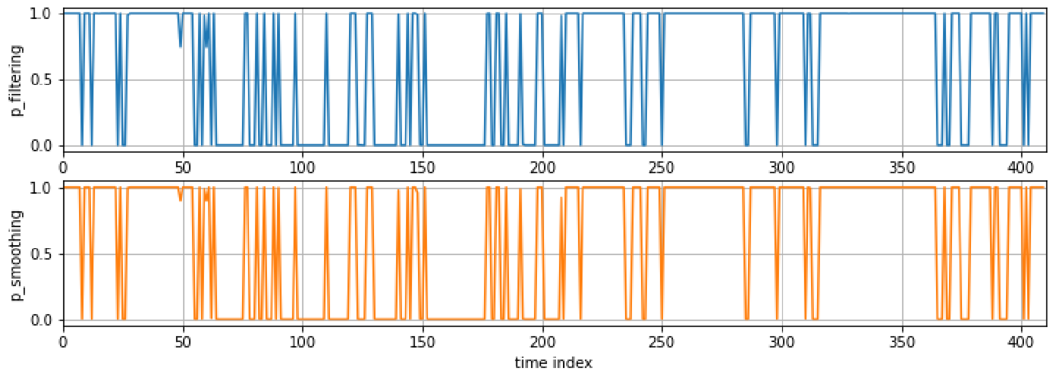

Figure 7. Filtering and smoothing probabilities for regime 1 over time.

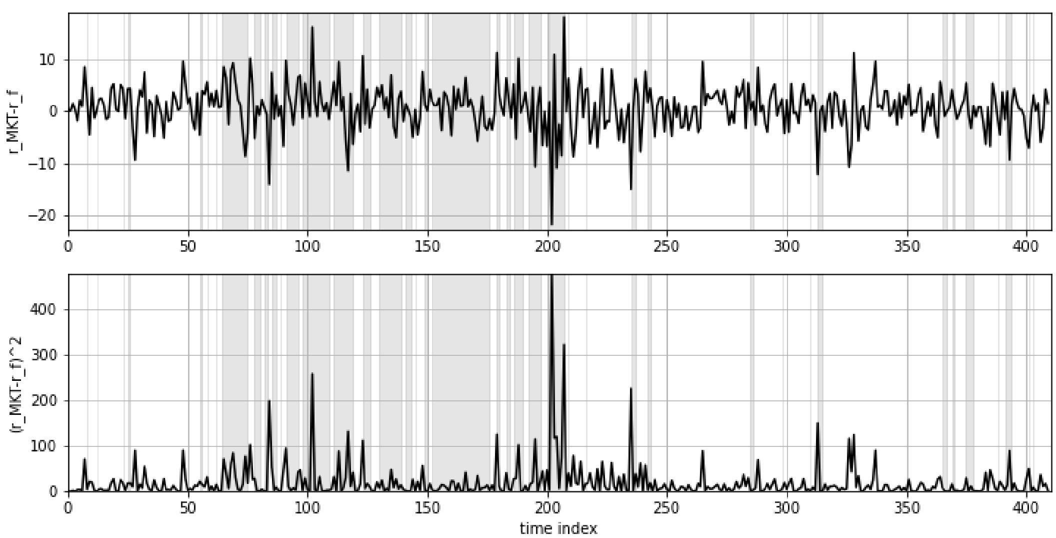

Figure 8. Raw and squared factor time series with most probable regimes overlaid.

The filtering and smoothing probabilities are very similar and decisive (i.e. close to zero or one), implying that the current information plays a decisive role in determining the regime. Figure 7 and Table 8 suggest that the market consists of a regime with low volatility (regime 1) and a regime with high volatility (regime 2). For regime 1, the mean absolute alphas are higher with weak significance while the mean adjusted R-squared values are lower with significance. This means that while the abnormal returns are of greater magnitude when using the factor loadings for regime 1 than when using the factor loadings for a one-regime model, the variation in the returns is still explained better by the regime-specific loadings *given* the abnormal returns. This yields a key practical insight gained from the RSFM: if the market is in a period of low volatility, a trader or investor can take advantage of the higher regime-specific absolute alphas and make trades or investments accordingly. For regime 2, the absolute alphas are lower and the adjusted R-squared values are higher with significance, which is expected since there is greater market uncertainty during a period of increased volatility; thus a practical application of the RSFM here is be the ability to isolate such periods.

	CAPM	RS-CAPM(2) Regime 1	RS-CAPM(2) Regime 2
mean $\lvert\alpha_i\rvert$	0.3333	0.3969 (0.077)	0.8893 (<0.001)
mean R^2_{adj}	0.3783	0.4360 (<0.001)	0.3633 (<0.001)
mean σ_i	6.6653	4.6952 (<0.001)	8.6998 (<0.001)

Table 9. Selected parameter means of the RS-CAPM(2) with paired two-tailed p-values.

We then fit a two-state C3FM on our data, and repeat the same analysis with the CAPM. The results are displayed in Figure 10-13 and Table 14.

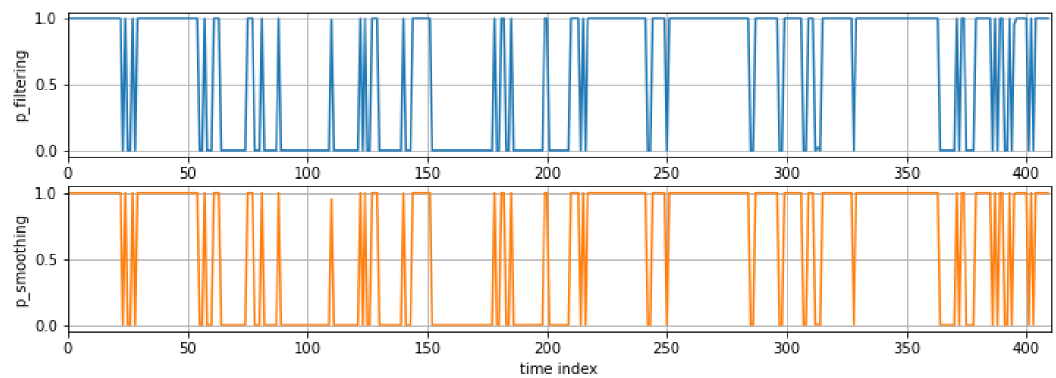

Figure 10. Filtering and smoothing probabilities for regime 1 over time.

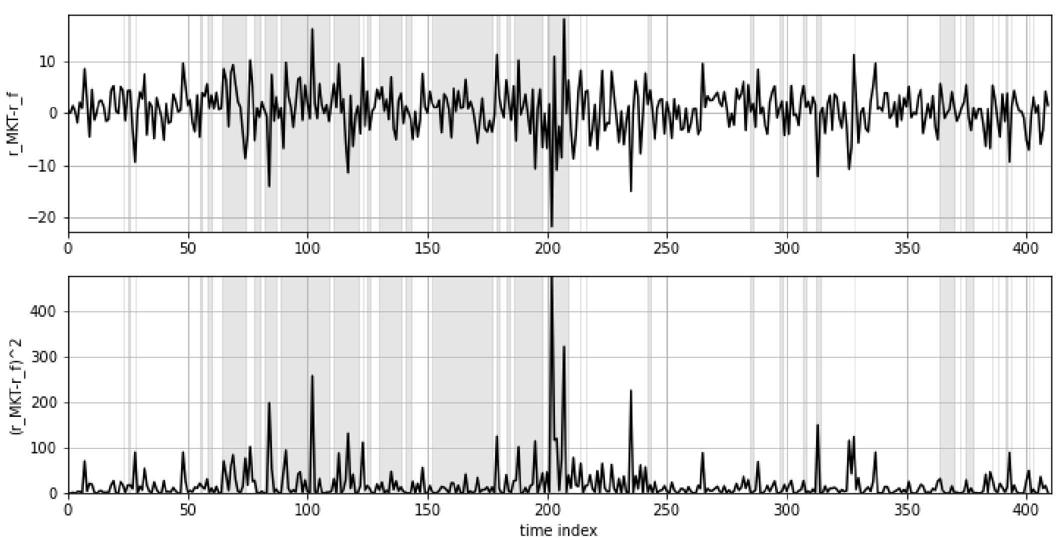

Figure 11. Raw and squared time series for $\text{MKT}_t - \text{RF}_t$ with most probable regimes overlaid.

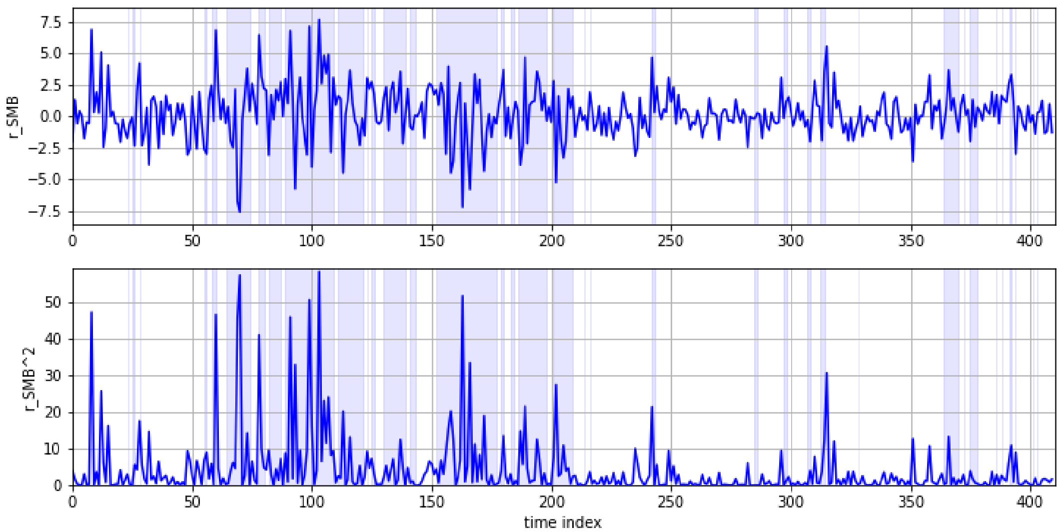

Figure 12. Raw and squared time series for SMB_t with most probable regimes overlaid.

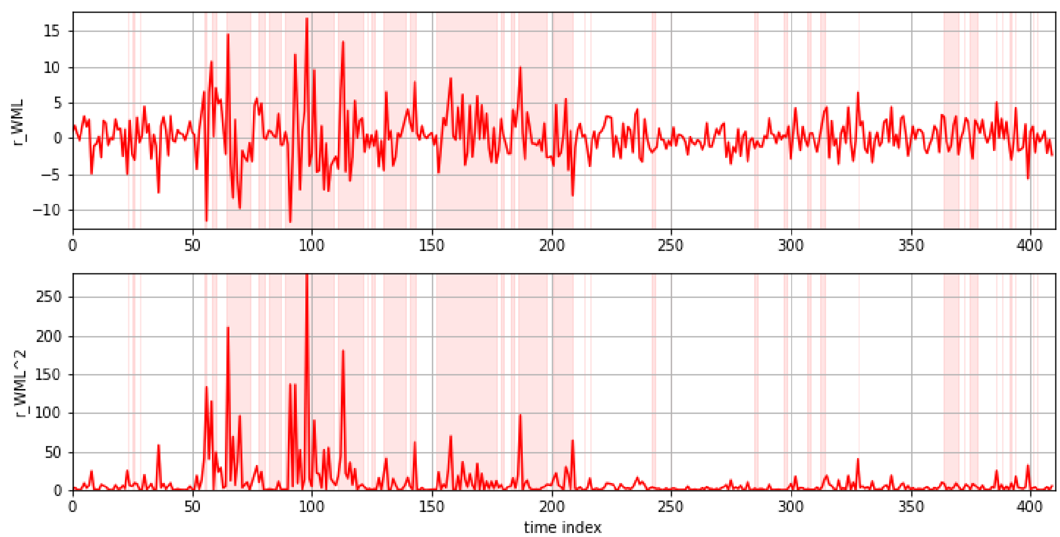

Figure 13. Raw and squared time series for WML_t with most probable regimes overlaid.

	C3FM	RS-C3FM(2) Regime 1	RS-C3FM(2) Regime 2
mean $\|\alpha_i\|$	0.2960	0.3827 (0.005)	0.7656 (<0.001)
mean R^2_{adj}	0.4097	0.4799 (<0.001)	0.3895 (<0.001)
mean σ_i	6.4777	4.4277 (<0.001)	8.3790 (<0.001)

Table 14. Selected parameter means of the RS-C3FM(2) with paired two-tailed p-values.

The findings for the two-regime CAPM hold for the two-regime C3FM as well, giving a sort of robustness check on the findings. The filtering and smoothing probabilities are similar and decisive; the regimes are classified on the basis of volatility, with the model also picking up volatile periods of the size and momentum factors; the trends in the tables for the selected parameter means are the same as well.

5 Conclusion

The cryptocurrency market, with its recent growth and susceptibility to sudden volatile periods, is a a great case study for regime-switching models. In this paper, we have considered a regime-switching factor model and devised an expectation-maximization algorithm to estimate it consisting of Baum-Welch updates for the E-step and M-step on the Markov chain, and weighted linear regressions for the M-step on the factor model (that is similar to how regular factor models are estimated). With the most recent data on cryptocurrencies, we consider a three-factor model and justify the choice of factors empirically. We then demonstrate the versatility of our algorithm by fitting the three-factor model for various numbers of regimes, from which we conclude that three regimes is BIC-optimal. One-regime and two-regime factor models are analyzed further. Compared to one-regime models, the mean adjusted R-squared values are lower (higher) and mean idiosyncratic risk levels are higher (lower) for the high (low) volatility regime, based on the regime-specific parameters. While this is to be expected, the higher absolute alphas in both regimes for both the CAPM and C3FM presents a phenomenon that warrants further exploration. But for now, this phenomenon can potentially be taken advantage of in trading, investing, or portfolio optimization strategies.

References

Augustyniak, M. (2014). Maximum likelihood estimation of the Markov-switching GARCH model. *Computational Statistics & Data Analysis*, 76, 61–75.

Bae, G. I., Kim, W. C., and Mulvey, J. M. (2014). Dynamic asset allocation for varied financial markets under regime switching framework. *European Journal of Operational Research*, 234(2), 450-458.

Baum, L. E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1), 164-171.

Billio, M., Casarin R., Osuntuyi, A. (2016). Efficient Gibbs sampling for Markov switching GARCH models. *Computational Statistics & Data Analysis*, 100, 37-57.

Costa, G. and Kwon, R. H. (2019). Risk parity portfolio optimization under a markov regime-switching framework. *Quantitative Finance*, 19(3), 453-471.

Costa, G. and Kwon, R. H. (2020). A regime-switching factor model for mean-variance optimization. *Journal of Risk*, 22(4).

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3-56.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357-384.

Hamilton, J. D. and Sumsel R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1-2), 307-333.

Hubrich, S. (2017). 'Know When to Hodl 'Em, Know When to Fodl 'Em': An Investigation of Factor Based Investing in the Cryptocurrency Space. *SSRN Electronic Journal*, 10.2139/ssrn.3055498.

Kim, C. J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60, 1–22.

Kim, M.-J. and Yoo, J.-S. (1995). New index of coincident indicators: A multivariate Markov switching factor model approach. *Journal of Monetary Economics*, 36(3), 607-630.

Kuan, C.-M. (2002). *Lecture on the Markov-switching model.* Unpublished manuscript, Institute of Economics, Academia Sinica. Retrived from https://homepage.ntu.edu.tw/~ckuan/pdf/Lec-Markov_note.pdf.

Liu, Y. and Tsyvinski A. (2021). Risks and returns of cryptocurrencies. *The Review of Financial Studies*, 34(6), 2689-2727.

Nystrup, P., Madsen, H., and Lindström, E. (2018). Dynamic portfolio optimization across hidden market regimes. *Quantitative Finance*, 18(1), 83-95.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111-163.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3), 425-442.

Shen, D., Urquhart, A., and Wang, P. (2019). Does twitter predict bitcoin? *Economics Letters*, 174, 118-122.

Shen, D., Urquhart A., and Wang P. (2020). A three-factor pricing model for cryptocurrencies. *Finance Research Letters*, 34, 101248.

Urquhart, A. (2016). The inefficiency of bitcoin. *Economics Letters*, 148, 80-82.

Urquhart, A. (2018). What causes the attention of bitcoin? *Economics Letters*, 166, 40–44.

Appendix

Code and data for Section 4 can be found at https://github.com/zhubrian/regime-switching.