HMM

tilian bourachot

April 2025

1 Introduction

Dans cette partie, nous avons choisi d'utiliser un modèle de Markov caché (Hidden Markov Model, HMM) pour détecter des changements de régime dans les prix des stablecoins. Ce type de modèle probabiliste est bien adapté à la modélisation de systèmes dynamiques où l'on suppose que les observations (ici, les variations de prix) dépendent d'un état caché du marché, comme un état stable ou un état instable.

L'objectif de cette partie est donc de mettre en place un détecteur de *depeg* fondé sur un HMM, capable d'identifier automatiquement des phases anormales dans l'évolution des prix de stablecoins, et en particulier celles correspondant à une perte temporaire ou durable de la parité.

2 Modèles de Markov cachés et algorithme EM

Un *modèle de Markov caché* (HMM) est un modèle statistique permettant de représenter un processus stochastique dont l'évolution dépend d'un *état latent* (non observable), qui change au cours du temps selon une chaîne de Markov. À chaque instant, cet état caché génère une observation visible selon une certaine loi de probabilité. L'objectif est donc d'inférer les états cachés à partir des observations.

Dans le cadre de ce projet, on suppose que le marché du stablecoin peut se trouver dans deux *régimes* :

- un **état stable**, où le prix reste proche de la parité ;
- un **état de depeg**, correspondant à une perte temporaire ou durable de cette parité.

Ces régimes ne sont pas observés directement, mais les *rendements logarithmiques* observés, notés r_t, dépendent de l'état caché du marché s_t.

Un HMM est défini par :

- Un nombre d'états cachés $N = 2$;

- Une matrice de transition $A = [a_{ij}] \in \mathbb{R}^{2\times 2}$ telle que
$$a_{ij} = \mathbb{P}(s_{t+1} = j \mid s_t = i);$$
- Une loi de probabilité pour les observations dans chaque état, ici une loi normale :
$$r_t \mid s_t = j \sim \mathcal{N}(\mu_j, \sigma_j^2);$$
- Un vecteur de probabilités initiales $\pi = [\pi_0, \pi_1]$ tel que $\pi_i = \mathbb{P}(s_1 = i)$.

Algorithme EM (Baum-Welch)

L'entraînement du HMM repose sur l'algorithme EM (*Expectation-Maximization*), plus précisément sa version appelée *algorithme de Baum-Welch*. Il s'agit d'un algorithme itératif en deux étapes :

- **E-step (Expectation)** : calcul de l'espérance de la log-vraisemblance complète, conditionnellement aux observations et aux paramètres estimés $\theta^{(k)}$:
$$Q(\theta \mid \theta^{(k)}) = \mathbb{E}_{s|r,\theta^{(k)}} \left[\log \mathbb{P}(s, r \mid \theta)\right];$$
- **M-step (Maximization)** : mise à jour des paramètres θ en maximisant Q.

Ces étapes sont répétées jusqu'à convergence.

Une fois le modèle entraîné, on peut prédire à chaque instant t l'état le plus probable :
$$\hat{s}_t = \arg\max_{i\in\{0,1\}} \mathbb{P}(s_t = i \mid r_{1:T}).$$

Dans notre cas, les observations sont les rendements logarithmiques :
$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right)$$
calculés à partir des séries de prix P_t. Le modèle permet alors de détecter automatiquement les périodes correspondant à des phases de stabilité ou de *depeg*.

3 Application du modèle HMM aux rendements horaires de l'USDT sur Coinbase

Dans cette partie, nous nous intéressons aux prix horaires de l'USDT sur la plateforme **Coinbase** au cours de l'année **2022**. Bien que l'USDT soit généralement perçu comme un stablecoin fiable, il a connu une période de tension importante au mois de **mai 2022**, dans un contexte de panique liée à l'effondrement du stablecoin algorithmique TerraUSD (UST). Ce stress de marché a brièvement

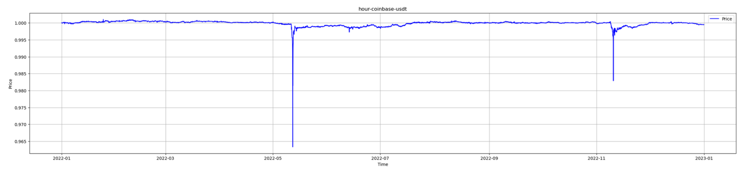

Figure 1: Prix en heure de l'USDT sur Coinbase en 2022

affecté la stabilité de l'USDT, provoquant un désancrage momentané (ou *depeg*) par rapport au dollar. Cette période constitue donc un bon cas d'étude pour tester notre modèle de détection de changement de régime.

À partir des prix agrégés à une fréquence horaire, nous calculons les **rendements logarithmiques** entre chaque paire de points successifs. Ce passage au logarithme permet de transformer les variations relatives de prix en une série additive, ce qui est statistiquement plus stable et mieux adapté aux modèles probabilistes comme les HMM. Contrairement aux variations absolues, les rendements logarithmiques présentent également une symétrie entre hausses et baisses, et une meilleure gestion des changements d'échelle. Ils sont définis par :

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right)$$

où P_t représente le prix de l'USDT à l'instant t.

Nous entraînons ensuite un **modèle de Markov caché à deux états** (stable / depeg) à l'aide de la bibliothèque `hmmlearn`. Chaque état est modélisé par une loi normale sur les rendements, avec estimation de la moyenne et de la variance propres à chaque régime. Une fois le modèle ajusté, nous prédisons, pour chaque instant, l'*état caché le plus probable*.

Les résultats sont visualisés à l'aide d'un graphique interactif `Plotly`. Chaque point représente un prix horaire, coloré selon l'état détecté par le modèle. On peut ainsi repérer visuellement les phases de stabilité et les phases suspectes où l'USDT aurait pu perdre sa parité, sans avoir besoin de fixer arbitrairement un seuil sur le prix. Cette approche permet une détection dynamique et robuste des épisodes de *depeg*.

4 Filtrage de Kalman avec paramètres dépendants de l'état

Dans cette étape, nous affinons l'analyse en appliquant un **filtre de Kalman** dont les paramètres *évoluent dynamiquement en fonction de l'état caché* détecté par le HMM. L'idée est d'adapter la sensibilité du filtre selon que le marché est stable ou en situation de *depeg*.

Pour cela, deux jeux de paramètres sont définis :

- Un jeu pour l'**état stable**, avec une faible variance du bruit de processus

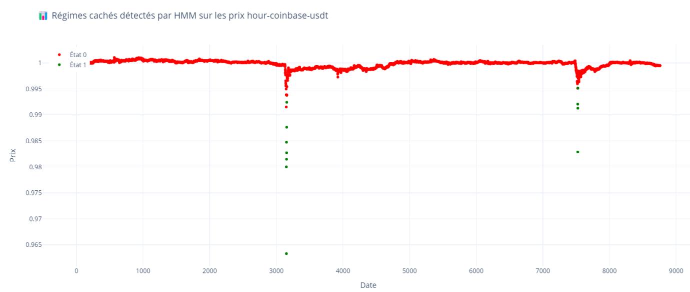

Figure 2: Prix en heure de l'USDT sur Coinbase en 2022 avec HMM

(Q) et du bruit d'observation (R), traduisant une confiance élevée dans la stabilité du système ;

- Un autre jeu pour l'**état instable**, avec des valeurs de Q et R plus élevées, traduisant une incertitude accrue dans les observations et dans l'évolution du prix.

Le filtre de Kalman est ensuite appliqué sur les prix de l'USDT, en recalibrant ses paramètres à chaque pas de temps selon l'état le plus probable donné par le HMM. Cela permet d'**ajuster dynamiquement la réactivité du filtre**, le rendant plus souple dans les phases agitées, et plus rigide dans les phases calmes.

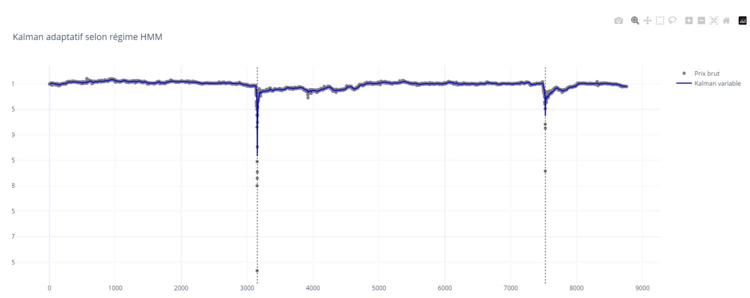

Figure 3: Kalman adaptatif en fonction du régime HMM

Cette méthode ne constitue pas le cœur de notre analyse, mais elle illustre comment le HMM peut être utilisé comme **module de pilotage adaptatif** pour des modèles dynamiques tels que le filtre de Kalman. Elle permet une estimation plus lissée du prix sous-jacent, mais ne sera pas approfondie dans la suite du projet.

5 Étude du depeg de l'USDC en 2023

Après l'analyse de l'USDT en 2022, nous appliquons notre modèle de Markov caché à l'**USDC** sur l'année **2023**, une période marquée par un événement majeur de *depeg* au mois de mars. Cet épisode a été déclenché par la faillite de la **Silicon Valley Bank (SVB)**, dans laquelle l'émetteur de l'USDC, **Circle**, détenait une part importante de ses réserves. Cette situation a provoqué une perte de confiance temporaire, menant à une chute rapide du prix de l'USDC jusqu'à environ 0,88 $.

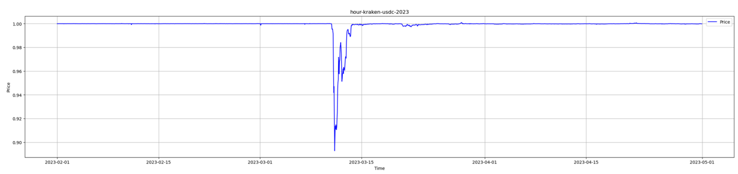

Figure 4: Prix de l'USDC en 2023 sur Kraken

Dans cette étape, nous conservons le **modèle HMM à deux états** préalablement entraîné sur les rendements de l'USDT, sans effectuer de nouvel entraînement sur l'USDC. Cette approche permet de tester la capacité du modèle à se *généraliser* à un autre actif, en transférant sa compréhension des régimes de marché à une série de prix différente.

Les rendements horaires de l'USDC pour l'année 2023 sont donc calculés, puis injectés directement dans le modèle entraîné. Les états cachés sont prédits pour chaque instant, et la période de mars 2023 est effectivement identifiée comme un changement de régime, ce qui confirme l'utilité du modèle en détection automatique de *depeg*, même sans supervision explicite.

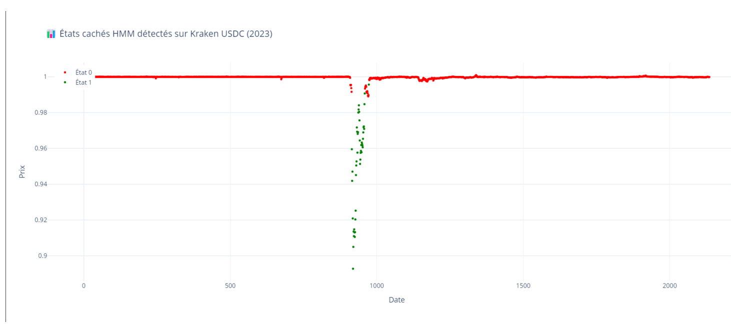

Figure 5: HMM sur le prix de l'USDC en 2023 sur Kraken

Enfin, dans la dernière partie de notre notebook, nous soulevons une question importante liée à l'usage du modèle en conditions réelles : *le modèle reste-t-il pertinent lorsque la série de prix est observée au fur et à mesure, comme c'est le cas dans la vraie vie ?* Dans ce contexte *en ligne*, les décisions doivent être prises sans connaître les données futures. Bien que notre modèle soit actuellement appliqué de manière *hors ligne* (c'est-à-dire sur des séries complètes), une vérification simple, réalisée en appliquant le modèle progressivement sur les données, montre qu'il **reste capable de détecter efficacement le changement de régime**, même sans accès à toute la série. Cela suggère que le HMM pourrait être utilisé en pratique comme outil de surveillance en quasi temps réel.

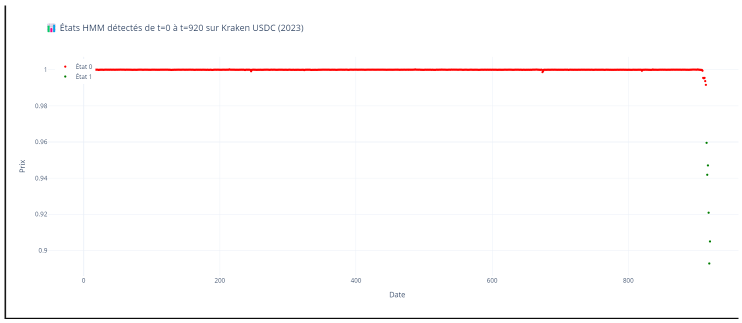

Figure 6: HMM progressif sur le Prix de l'USDC en 2023 sur Kraken