

outlier

tilian bourachot

April 2025

1 Introduction

Dans cette section, nous nous intéressons à la qualité des données collectées sur les stablecoins. Les données que nous utilisons sont issues de plusieurs plateformes d'échange (Binance, Coinbase, Kraken, FTX) et couvrent différentes périodes. Chaque enregistrement contient des informations telles que le prix d'un stablecoin, la quantité échangée, le moment de la transaction ainsi que le type d'opération (achat ou vente).

Ces données sont essentielles pour nos futures analyses de dynamique de prix, de détection de *depeg* ou encore de prévision. Cependant, comme tout jeu de données transactionnel à grande échelle, elles contiennent des **valeurs aberrantes (outliers)**. Ces outliers peuvent être dus à des erreurs de saisie, des anomalies techniques ou encore des événements économiques exceptionnels. Leur présence peut fausser les résultats d'analyses statistiques classiques, notamment les tests de stationnarité ou les modèles de prévision.

Avant toute modélisation, il est donc fondamental de procéder à un **prétraitement rigoureux des données**. Cette étape de *preprocessing* vise à :

- identifier les outliers,
- les visualiser clairement dans les séries de prix,
- et décider d'une stratégie de traitement adaptée (suppression, marquage ou remplacement).

Ce travail préparatoire permet d'assurer la robustesse des analyses ultérieures et de mieux comprendre les phénomènes atypiques liés aux mouvements de marché.

2 Méthode de traitement des outliers

Dans un premier temps, nous appliquons une méthode simple et robuste pour identifier les valeurs aberrantes dans les séries de prix : la **méthode des trois écarts-types**. Cette technique repose sur l'hypothèse que, dans une distribution normale, environ 99,7 % des valeurs se situent à moins de trois écarts-types

de la moyenne. Toute valeur en dehors de cet intervalle est donc considérée comme un outlier potentiel.

Nous avons mis en œuvre deux fonctions complémentaires :

- **remove_outliers** : cette fonction supprime les lignes contenant des valeurs considérées comme aberrantes dans les colonnes spécifiées. Elle permet un nettoyage rapide des données avant analyse.
- **mark_outliers** : cette variante marque simplement les outliers à l'aide d'une nouvelle colonne binaire (`{nom_colonne}_outlier`). Elle permet une visualisation des anomalies sans supprimer d'observations, ce qui est utile pour l'étude descriptive.

Ce premier filtrage permet d'éliminer les erreurs manifestes (erreurs techniques, entrées incohérentes), mais il ne capture pas tous les cas d'outliers. En particulier, il reste des **événements extrêmes mais légitimes**, tels que :

- des épisodes de panique de marché provoquant des mouvements de prix brutaux mais réels ;
- des sauts de volatilité autour d'annonces majeures ;
- des comportements spécifiques sur certaines plateformes ou à certaines heures.

Ces situations ne peuvent pas être traitées par un simple critère statistique global. Elles nécessitent une **analyse contextuelle plus fine**, que nous présenterons dans la suite.

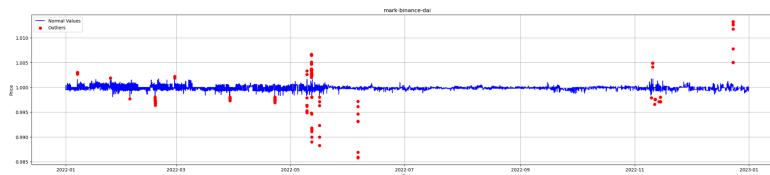


Figure 1: DAI avec outliers marqués

3 Études de cas : vrais outliers à conserver

Le depeg de l'USDD en décembre 2022

Parmi les valeurs extrêmes qui persistent après l'application du filtre statistique, certaines correspondent à des phénomènes réels et documentés sur les marchés. Un exemple significatif est celui de l'**USDD (Decentralized USD)**, une stablecoin algorithmique lancée en 2022 par Justin Sun, fondateur de la DAO Tron. Indexé sur le dollar américain dans un rapport de 1:1, l'USDD repose sur un

mécanisme de surcollatéralisation via des actifs tels que TRX, BTC ou USDC, et circule sur plusieurs blockchains (Ethereum, Tron, BNB Chain).

À la **mi-décembre 2022**, l'USDD s'est écarté de sa parité avec le dollar. Selon les données de Coingecko, la valeur de l'USDD a chuté à **0,971 \$ le 13 décembre 2022**, suite à plusieurs jours de fluctuations. En réponse, Justin Sun a déclaré que des fonds supplémentaires avaient été mobilisés pour restaurer la stabilité. Malgré cette intervention, de nouveaux décrochages ont eu lieu début janvier 2023, notamment le 6 janvier (0,972 \$) et le 10 janvier (0,977 \$), confirmant un épisode de tension prolongée.

Contrairement à des erreurs techniques ou à des valeurs isolées, ces écarts de prix sont liés à une situation économique réelle : perte temporaire de confiance dans le mécanisme algorithmique, arbitrages défensifs, et réactions du marché aux annonces de l'émetteur. Il s'agit donc d'un **vrai outlier**, qu'il ne faut surtout pas supprimer du jeu de données. Au contraire, ce type d'événement mérite une attention particulière car il reflète un risque de *depeg* partiel, caractéristique des stablecoins algorithmiques.

Réaction de marché sur l'USDC

L'évolution du prix de l'**USDC sur Binance** en décembre 2022 révèle un **pic soudain autour du 13 décembre**, où le prix dépasse momentanément 1,0025 \$, avant de retomber rapidement à son niveau habituel autour de 1 \$.

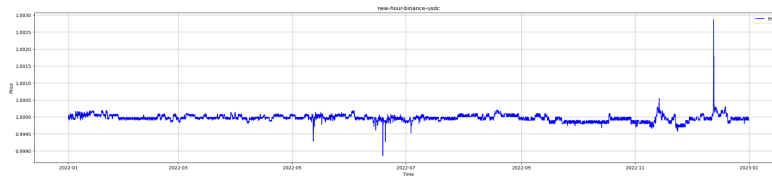


Figure 2: USDC Binance

Cet épisode peut être interprété comme une **conséquence directe du depeg de l'USDD** observé à la même période. Dans un contexte d'incertitude, les investisseurs ont pu chercher refuge dans des stablecoins perçus comme plus sûrs, notamment l'USDC, entraînant ainsi une pression à l'achat temporaire sur certaines plateformes.

Ce pic ne reflète donc pas une erreur ou un dysfonctionnement technique, mais plutôt une **réaction rationnelle du marché face à un stress sur un autre actif**. Il s'agit là encore d'un **vrai outlier**, qu'il convient de conserver dans les données, car il illustre un comportement systémique pertinent à modéliser, en lien avec la contagion entre stablecoins.

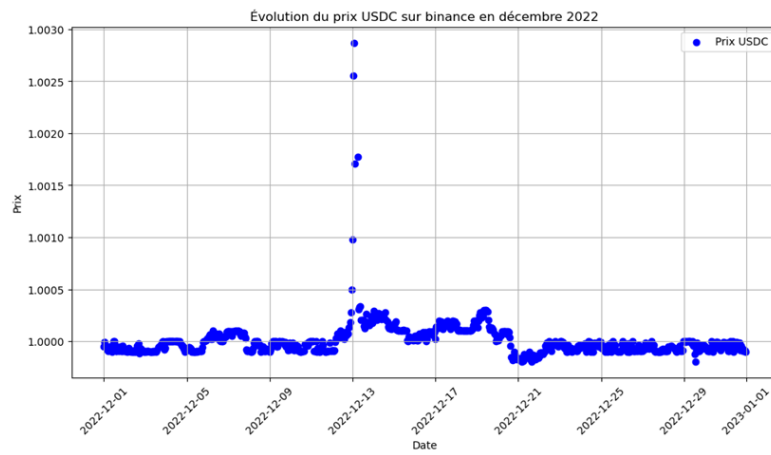


Figure 3: USDC Binance en decembre