



**Laboratoire de Mathématiques et Informatique pour la Complexité et les  
Systèmes  
MICS**

**Présente**

## **L'AVIS DE SOUTENANCE**

**De Monsieur Timothée Fabre**

Laboratoire MICS, CentraleSupélec, Université Paris Saclay, soutiendra  
publiquement ses travaux de thèse de doctorat intitulés :

***“High-frequency manipulation detection in cryptocurrency limit order books.”***

Sous la Direction de Monsieur Damien Challet et Monsieur Ioane Muni Toke.

**Le lundi 2 décembre 2024 à 14h00**

À l'école CentraleSupélec, dans l'**amphi V** - Bâtiment Eiffel.

### **Membres du jury :**

Charles-Albert LEHALLE, Rapporteur & Examineur, Ecole Polytechnique

Thomas WEBER, Rapporteur & Examineur, Ecole Polytechnique Fédérale de Lausanne

Laurence CARASSUS, Examinatrice, CentraleSupélec

Mathieu ROSENBAUM, Examineur, Ecole Polytechnique

### **Résumé :**

La première partie de cette thèse s'intéresse à l'étude de la microstructure des carnets d'ordre de cryptomonnaie ainsi que de leurs spécificités. Par l'analyse d'un problème d'exécution en haute fréquence, nous démontrons que la petite taille du tick et les frais de transaction importants des marchés de cryptomonnaies poussent les agents à coter leurs ordres de manière agressive. Puis, nous nous intéressons à l'impact du volume des ordres marché sur le taux d'arrivée des transactions, ainsi qu'aux relations de meneur-suiveur au sein des cryptomonnaies. Pour cela, nous développons une méthode d'estimation non-paramétrique

des processus de Hawkes linéaires en haute dimension. Nous démontrons que l'impact global d'une exécution d'ordre marché sur le taux d'arrivée des transactions est une fonction concave du volume, puis nous étudions la structure de causalité d'un univers de 15 paires au moyen de nouvelles mesures de quantification des relations de meneur-suiveur. La deuxième partie est consacrée au développement de nouvelles méthodes non-supervisées de détection de manipulation en environnement haute fréquence. Dans un premier temps, nous nous intéressons au repérage de transactions potentiellement fausses, au sens où les deux contreparties de l'exécution sont une seule et même personne. Pour ce faire, nous introduisons une nouvelle classe de processus ponctuels auto-excitants à changement de régime et examinons sa capacité à inférer le régime suspect. Nous caractérisons ce régime en étudiant la réponse du marché ainsi que l'état de la liquidité. Dans un second temps, nous nous concentrons sur la détection d'ordres limites de grande taille qui sont insérés dans l'unique but de tromper le marché. Inspirés par les processus de Hawkes, nous construisons de nouvelles variables explicatives construites à partir du flux d'ordres, puis nous entraînons un réseau de neurones probabiliste à prédire la distribution de probabilité de la variation du prix. Ce modèle nous permet de calculer la fonction de coût d'un agent manipulateur, aboutissant à la construction d'un critère de détection d'ordre posté de mauvaise foi. Nous démontrons par des illustrations de cas détectés que notre méthodologie est capable de repérer des comportements anormaux qui peuvent être classifiés comme de la manipulation. Nous clôturons cette thèse par une discussion des résultats, ainsi que par une ouverture sur des questions qui peuvent faire l'objet de recherches futures.

### **Abstract :**

The first part of this thesis focuses on the study of the microstructure of cryptocurrency order books and their particularities. Through the analysis of an execution problem in high-frequency trading, we demonstrate that the small tick size and the significant transaction costs in cryptocurrency markets drive agents to quote their orders aggressively. We then investigate the impact of the volume of market orders on the transaction arrival rate, as well as the lead-lag relationships within cryptocurrencies. To do this, we develop a non-parametric estimation method for linear Hawkes processes in high dimensions. We demonstrate that the overall impact of a market order execution on the transaction arrival rate is a concave function of volume. Finally, we introduce new measures of causality, calculated from the Hawkes kernel estimates, and illustrate their relevance in a practical case. The second part is dedicated to the development of new unsupervised methods for detecting market manipulation in a high-frequency environment. First, we focus on identifying potentially fake transactions where both sides of the execution are the same individual. To this effect, we introduce a new class of self-exciting point processes with regime shifts and examine its ability to infer the suspicious regime. We show that the model can identify an extreme-burst regime, and characterize this regime by

studying market response and liquidity state. Secondly, we focus on detecting spoofing, *i.e.* large limit orders that are placed solely to deceive the market. Inspired by Hawkes processes, we construct new explanatory variables derived from the order flow. Then, a probabilistic neural network is trained to predict the probability distribution of price variation. This model allows us to compute the cost function of a manipulative agent, leading to the construction of a detection criterion for non-*bona fide* orders. We demonstrate through case illustrations that our methodology is capable of identifying abnormal behaviours that can be classified as manipulation. We conclude this thesis with a discussion of the main contributions as well as an exploration of questions that may be the subject of future research.