

Détection Probabiliste de Manipulation de Marchés

Soutenance Intermédiaire - Projet de Recherche

Adonis Jamal Jean Martini
Encadrant : Damien Challet

CentraleSupélec – Laboratoire MICS

4 décembre 2025

Sommaire

① Introduction

Contexte

Hypothèses

Objectifs

Avancement et Plan

② État de l'art

③ Méthodologie

Finance Quantitative

Machine Learning

Les marchés et le Carnet d'Ordres

- **Limit Order Book (LOB)** : Répertoire central des intentions d'achat (Bid) et de vente (Ask).
- **Dynamique** : Le prix médian (*mid-price*) réagit à l'état instantané et au flux d'ordres passé (Order Flow).
- **Données** :
 - Données Haute Fréquence (Level 3 / Market-by-order).
 - Nécessité de traiter des millions d'événements par jour.

Problématique

La microstructure du marché peut être exploitée pour induire les autres participants en erreur.

Le spoofing : une manipulation structurelle

- **Définition** : Placer des ordres non-bona fide (sans intention d'exécution) pour créer une fausse impression de liquidité ou de pression directionnelle.
- **Tactiques courantes** :
 - *Layering* : Empilement d'ordres à différents niveaux de prix.
 - *Vacuuming* : Création de vides de liquidité pour provoquer des mouvements brusques.
- **Mécanisme** :
 - 1 Placer un gros ordre d'achat (Fake) loin du prix.
 - 2 Attendre que le prix monte (réaction des algos).
 - 3 Vendre réellement (Bona fide) à un prix plus élevé.
 - 4 Annuler l'ordre d'achat initial.

Hypothèses de travail

Hypothèses Microstructurelles

- La distance de placement des ordres limites joue un rôle critique dans la formation des prix (contrairement aux modèles basés uniquement sur le volume au best bid/ask) [Tao+20].
- Un agent manipulateur agit rationnellement pour minimiser son coût d'exécution espéré.

Hypothèse Comparative

En l'absence de régulation stricte, les carnets d'ordres de crypto-monnaies (CEX) présentent :

- Une fréquence de spoofing plus élevée.
- Une sensibilité (impact prix) plus forte aux ordres de spoofing que les marchés actions traditionnels.

Objectifs du projet

Le projet vise à développer une meilleure caractérisation de la **"Spoofability"** des marchés.

Objectifs principaux :

- 1 **Implémentation de modèles et features** : Réseaux de neurones [FC25; PCM24; Tao+20], exploration de l'état de l'art.
- 2 **Comparaison inter-marchés** :
 - Marchés Actions (Régulés, surveillés).
 - Marchés Crypto (Non régulés, suspicion de forte manipulation).
- 3 **Validation de l'hypothèse** : Les marchés actions sont-ils vraiment moins sensibles (spoofables) que les marchés crypto ? Arrivons-nous à détecter ces manipulations ?

Avancement actuel et plan de travail

Réalisé (Mois 1) :

- Étude bibliographique approfondie (Fabre, Tao, Poutré).
- Prise en main des données (L3 fournies par l'encadrant).
- Architecture du code Python structurée (pré-traitement, feature engineering, modèles).

Planification (Mois 2-5) :

- **Décembre - Janvier** : Implémentation complète et calibration des modèles sur les données. Exploration des features.
- **Février** : Analyse de sensibilité. Quantification de la spoofability comparée.
- **Mars** : Exploration avancée (anomalies, clustering, autres modèles).
- **Avril** : Rédaction du rapport final et préparation soutenance.

Synthèse Bibliographique

Notre approche s'appuie sur trois axes de recherche récents :

- **Approche Probabiliste & Features Hawkes** [FC25] : Utilisation de réseaux de neurones pour prédire non pas un prix, mais les paramètres d'une distribution de prix (μ, σ, α) , en utilisant des variables de flux d'ordres multi-échelles (temps, volume, distance).
- **Détection Non-Supervisée (Deep Learning)** [PCM24] : Utilisation d'Auto-encodeurs (Transformers) pour apprendre la "normalité" du marché et détecter les anomalies via la distance de reconstruction ou SVM One-Class.
- **Modélisation de l'Imbalance** [Tao+20] : Étude théorique de l'impact de l'imbalance multi-niveaux sur le mouvement des prix et stratégies optimales de spoofing.

Critère de Détection Financier

Nous adoptons le point de vue du manipulateur rationnel.

Fonction de Coût Espéré $\mathbb{E}[C]$

Le spoofer compare deux scénarios pour exécuter un ordre cible q :

- 1 **Exécution normale** : Coût d'exécution sans intervention.
- 2 **Avec Spoofing** : Coût d'exécution après insertion d'un faux ordre Q à distance δ .

Règle de Détection

Un ordre (Q, δ) est suspect si son insertion améliore significativement le coût d'exécution espéré sur le côté opposé :

$$\Delta C(Q, \delta) = \mathbb{E}[C_{\text{normal}}] - \mathbb{E}[C_{\text{spoofer}}] > 0$$

Réseau de Neurones Probabiliste

L'objectif n'est pas de classer (Spoof/Pas Spoof) mais d'apprendre la dynamique des prix $P(\Delta p | \mathcal{F}_t)$.

- **Entrée** x : Vecteur de features (Spread, Flux d'ordres Limites multi-échelles, Flux d'ordres Marché).
- **Modèle** : Feed-forward Neural Network (MLP) léger pour le temps réel.
- **Sortie** $\Theta(x)$: Paramètres d'une distribution (ex: Gaussienne Asymétrique).

$$\Delta p \sim \mathcal{SN}(\mu(x), \sigma(x), \alpha(x))$$

- **Entraînement** : Maximisation de la vraisemblance (Maximum Likelihood Estimation) sur données historiques saines.

Le modèle PNN sert à calculer les espérances en prédisant le mouvement de prix Δp induit.

Merci pour votre attention

Bibliographie I

- [Tao+20] Xuan Tao et al. *On Detecting Spoofing Strategies in High Frequency Trading*. 2020. arXiv: 2009.14818 [q-fin.TR]. URL: <https://arxiv.org/abs/2009.14818>.
- [PCM24] Cédric Poutré, Didier Chételat, and Manuel Morales. “Deep Unsupervised Anomaly Detection in High-Frequency Markets”. In: *The Journal of Finance and Data Science* 10 (Apr. 2024), p. 100129. DOI: 10.1016/j.jfds.2024.100129.
- [FC25] Timothée Fabre and Damien Challet. *Learning the Spoofability of Limit Order Books With Interpretable Probabilistic Neural Networks*. 2025. arXiv: 2504.15908 [q-fin.TR]. URL: <https://arxiv.org/abs/2504.15908>.

Organisation du travail d'équipe

Organisation et dispositif

- Réunions hebdomadaires avec l'encadrant (D. Challet).
- Gestion de projet agile : ré-évaluation périodique des objectifs.
- Outils : Teams (Comms), Git (Versionning Code), Overleaf (Rapport).

Répartition du travail

- **Adonis** : Code, travail sur les méthodes, orienté finance quantitative.
- **Jean** : Code, travail sur les méthodes, orienté machine learning.

Construction des variables explicatives

Pour capturer la dynamique temporelle et spatiale du carnet, nous utilisons des variables inspirées des processus de Hawkes marqués.

Soit N_t le processus de comptage des ordres. On définit les variables de flux L_t (Limit) et M_t (Market) :

$$L_t(\phi) = \int_{(-\infty, t]} \phi(t - s, v_s, \delta_s) dN_s \quad (1)$$

- **Noyau ϕ** : Intègre trois dimensions cruciales :
 - **Temps** : Décroissance exponentielle $e^{-\beta t}$ (mémoire du marché).
 - **Volume (v_s)** : Impact de la taille de l'ordre.
 - **Distance (δ_s)** : Distance au mid-price (innovation majeure par rapport à l'imbalance classique).
- **Multi-échelles** : Combinaison de plusieurs horizons temporels (β) et échelles de distance (η).