

# Spécification détaillée – Stratégie de Market-Making en trois versions

---

## 1. Contexte et objectifs

Nous souhaitons déployer un moteur de market-making systématique qui évolue en trois étapes :

- **V1** : capturer le spread sans parti pris directionnel en appliquant le modèle d'Avellaneda & Stoikov (2008).
- **V1.5** : accroître la rentabilité via un signal micro-alpha (Order-Flow Imbalance) et un pilotage dynamique de l'inventaire.
- **V2** : confier la décision de cotation à un agent d'apprentissage par renforcement profond (Soft-Actor-Critic) muni de garde-fous « risk-first ».

Toutes les versions reposent sur un carnet **Binance** reconstruit localement à partir des flux WebSocket *depth* et *aggTrade*.

---

## 2. Fondements théoriques (exposé complet)

### 2.1 Modèle d'Avellaneda & Stoikov (2008)

#### Hypothèses structurelles

- Le mid-price suit une diffusion géométrique sans drift à très court terme :  $dS_t = \sigma dW_t$ .
- Les ordres « prenants » (market orders) frappant nos quotes suivent deux processus de Poisson indépendants :  $N_t^{\text{bid}}, N_t^{\text{ask}}$ , dont les intensités décroissent exponentiellement avec l'écart au mid :  $\lambda_{\text{bid}}(\delta) = A e^{-k\delta}$ ,  $\lambda_{\text{ask}}(\delta) = A e^{-k\delta}$ .  $A > 0$  traduit l'activité générale,  $k > 0$  la sensibilité des preneurs au prix.
- L'inventaire  $q_t$  évolue par pas discrets  $\pm 1$  à chaque exécution. Le teneur de marché supporte un **coût quadratique d'inventaire** paramétré par  $\gamma$ .

**Formulation de contrôle** On maximise l'espérance de la valeur terminale des avoirs moins une pénalité d'inventaire :  $\max_{\delta_a, \delta_b} \mathbb{E} \left[ X_T + q_T S_T - \frac{\gamma}{2} q_T^2 \right]$ , avec  $X_t$  le cash et  $T$  l'horizon de liquidation.

**HJB et solution fermée** La résolution de l'équation de Hamilton-Jacobi-Bellman, après linéarisation (hypothèse de séparation des variables), conduit aux quotes optimales :

$$\begin{aligned} \text{Reservation price } r_t &= S_t - \gamma q_t \sigma^2 (T-t) - \frac{1}{1+\gamma/k} \frac{P_{\text{bid}}}{P_{\text{ask}}} \\ P_{\text{bid}} &= r_t - \delta_{\text{bid}} - \gamma q_t \sigma^2 (T-t) \\ P_{\text{ask}} &= r_t + \delta_{\text{ask}} - \gamma q_t \sigma^2 (T-t) \end{aligned}$$

Le **spread théorique**  $2\delta_*$  ne dépend que de  $\gamma$  et  $k$ , tandis que le **déplacement** du centre dépend du produit  $q_t \sigma^2$  : plus l'inventaire est éloigné de 0 ou la volatilité élevée, plus on se fait « payer » pour accepter le risque.

**Interprétation économique** : le teneur de marché exige un surplus (le spread) pour compenser le risque d'exécution adverse, et il décale ses prix pour attirer les ordres qui réduisent sa position nette.

## 2.2 Micro-alpha haute fréquence (Order-Flow Imbalance)

**Concept** : l'OFI mesure, sur une fenêtre courte  $\Delta$ , la différence normalisée entre volumes agressifs côté achat et côté vente : 
$$\text{OFI}_t(\Delta) = \frac{\sum_{i \in \text{BUY}} v_i - \sum_{j \in \text{SELL}} v_j}{\sum_{i \in \text{BUY}} v_i + \sum_{j \in \text{SELL}} v_j}.$$

### Résultats empiriques

- Sur crypto, futures FX et actions à haute liquidité, la corrélation OFI  $\rightarrow$  prochain tick est statistiquement significative jusqu'à 2-3 s.
- La distribution est leptokurtique; on applique donc un **clamp** (p. ex.  $\pm 3$  écarts-types) pour éviter les extrêmes.
- Dans 55-60 % des cas, le signe de l'OFI prédit la direction immédiate du mid-price.

## 2.3 Modèle d'impact linéaire de Kyle (1985) et extension multivariée

**Kyle univarié** : un trader informé, un bruit de trading et un teneur de marché fixent les prix. L'équilibre Nash conduit à un impact linéaire  $\Delta S = \Lambda v$  où  $\Lambda = \sigma/\beta$ .

**Extension multi-actifs** : l'impact devient matrice SPD  $\mathbf{\Lambda}$ . Estimation empirique :  $\mathbf{\Lambda} = \arg \min_{\mathbf{\Lambda}} \|\mathbf{r} - \mathbf{\Lambda} \mathbf{v}\|^2$ . La contrainte SPD garantit l'absence d'arbitrage.

## 2.4 Apprentissage par renforcement profond appliqué au MM

**Cadre MDP** : états (prix relatifs, inventaire, volatilité, signaux, profondeur L5, temps), actions (offsets  $\pm N$  ticks), récompense ( $\text{PnL} - \lambda |q|^2$ ).

**SAC** : politique gaussienne, replay buffer, entropie  $\alpha$  auto-tuned, risques de **distribution shift** et nécessité de garde-fous.

## 2.5 Ingestion temps-réel du carnet Binance

**Flux** : `depth@100ms`, `aggTrade`, snapshot REST.

**Synchronisation** : snapshot  $\rightarrow$  buffer diff  $\rightarrow$  appliquer U/u  $\rightarrow$  resnapshot 30 min.

**Architecture** : listener WS, ring buffer, moteur carnet, worker snapshot, publisher interne, métriques Prometheus.

**Tolérance pannes** : heartbeat, CRC, resnapshot sur mismatch, horloge PTP.

---

### 3. Version V1-α — Avellaneda-Stoikov avec décalage OFI

#### 3.1 Contexte et principe général

La version **V1-α** sert de **socle conceptuel** : elle applique le modèle d'Avellaneda & Stoikov (2008) et **ajoute un décalage minimal du centre** fondé sur le signal Order-Flow Imbalance (OFI). L'objectif reste de capturer le spread sans complexifier la gestion du risque inventaire. & Stoikov (2008) pour fixer deux quotes symétriques autour d'un prix de réserve. L'objectif est de **capturer le spread affiché** tout en gardant l'inventaire proche de zéro, sans chercher d'alpha directionnel.

#### 3.2 Flux d'entrée indispensables

Catégorie	Variables	Description	Cadence recommandée
Prix	Mid-price $S_t$	Moyenne best bid / best ask du carnet local	$\leq 50$ ms
Volatilité	$\sigma_t$	EWMA 100 observations (ajustée every tick)	continu
Inventaire	$q_t$	Position nette en unités de base	mise à jour sur fill
Horloge	$t$	Timestamp UNIX $\mu$ s	continu
Paramètres	$\gamma, k, T$	Aversion, liquidité, horizon	reload à chaud possible

#### 3.3 Cadre décisionnel

1. **Coût d'inventaire** : le teneur paie une pénalité quadratique  $\gamma q^2$ .
2. **Flux preneur** : intensité décroissante  $\lambda(\delta) = A e^{-k\delta}$  ;  $k$  mesure l'élasticité de la demande.
3. **Optimisation analytique** : maximisation de l'espérance d'utilité  $\Rightarrow$  Prix de réserve :  
 $r_t = S_t - \gamma q_t \sigma_t^2 (T - t)$  \ Écart optimal unique :  $\delta^* = (1/\gamma) \ln(1+\gamma/k)$  \ Les quotes sont donc **équidistantes** du prix de réserve et leur écart ne dépend pas de l'inventaire.

#### 3.3bis Décalage OFI minimal

Afin d'ajouter un premier edge directionnel, on décale simplement le centre des quotes :

$$\text{centre}_t = r_t + \beta_{\text{ofi}} \cdot \text{OFI}_t$$

- **Plage de  $\beta_{\text{ofi}}$**  : bornée à  $\pm 1$  tick pour ne pas transformer la stratégie en preneuse.
- **Aucun changement du spread** : on conserve  $\text{spread} = 2 \delta^*$  issu d'Avellaneda-Stoikov.
- **Pas de skew d'inventaire** : le prix de réserve reste la seule composante qui gère le risque.

### 3.4 Paramétrage et calibration

Paramètre	Sens économique	Méthode d'estimation	Ordre de grandeur (BTC-USDT)
$k$	Courbe de profondeur effective	Régression log-linéaire : $P(\text{fill})$ vs écart	0.4 → 1.2 ticks
$\gamma$	Averse au risque d'inventaire	Grid-search : maximiser Sharpe sous RMS q	1e-4 → 5e-4
$T$	Délai de liquidation cible	60 → 300 s selon la taille	typ. 120 s

**Astuce pratique** : déduire d'abord  $k$  sur données historiques (probabilité d'exécution), puis balayer  $\gamma$  pour tracer PnL vs RMS q. Choisir le couple qui offre le meilleur ratio retour/risque avant frais.

### 3.5 Boucle opérationnelle en temps réel

Observation → Décision → Cotation → Mise à jour :

1. Mesurer  $S_t$ ,  $\sigma_t$ ,  $q_t$ .
2. Calculer  $r_t$  et  $\delta^*$ .
3. Publier  $\text{bid} = r_t - \delta^*$ ,  $\text{ask} = r_t + \delta^*$ .
4. À chaque exécution : mettre à jour  $q_t$ , cash et latence ; retourner en 1.

### 3.6 Contrôles de risque

- **Limite inventaire** :  $|q_t| \leq q_{\text{max}}$  (déterminé par capital et volatilité).
- **Spread adaptatif** : si  $\sigma_t \uparrow > \text{factor } 2$ , multiplier  $\delta^*$  par 1.5.
- **Kill-switch latence** : si ACK > 300 ms (P99) → retirer quotes et réévaluer.

### 3.7 Métriques de performance

KPI	Interprétation	Seuils cibles
Spread capturé (%)	Part du spread affiché réellement perçue	$\geq 70 \%$
RMS inventaire	Volatilité de la position nette	$\leq 0.4 q_{\text{max}}$
Fill ratio	Exec / quotes	$\geq 5 \%$
Cancel ratio	Quotes annulées / totales	$\leq 70 \%$
Latence send→ACK (P99)	Robustesse infrastructure	$\leq 300 \text{ ms}$

### 3.8 Plan de validation

1. **Back-test hors-ligne** : relecture tick-by-tick 7 jours, mêmes paramètres.
2. **Stress test** : volatilité  $\times 2$ , liquidité  $\div 2$ , latence réseau 200 ms.
3. **Paper-trade testnet** : comparer PnL net fees, vérifier respect des limites.
4. Critères : PnL > 0, KPI conformes, aucun dépassement inventaire.

## 4. Version 1.5 — AS enrichi : OFI + Depth Imbalance + vol-scaled spread

### 4.1 Contexte et objectif

Cette version intermédiaire vise à **booster la rentabilité** sans complexité excessive avant le passage au RL. Elle s'appuie sur :

1. **OFI** (Order-Flow Imbalance) déjà présent en V1-α.
2. **Depth Imbalance (DI)** : asymétrie de profondeur entre le bid et l'ask.
3. **Spread dynamique proportionnel à la volatilité instantanée.**
4. **Quote ageing** : cadence de rafraîchissement calée sur l'évolution des signaux.

### 4.2 Flux d'entrée supplémentaires

Catégorie	Variables	Description	Fenêtre / cadence
Profondeur top-N	depth_bid_L1-L5, depth_ask_L1-L5	Somme des tailles visibles sur 5 niveaux	$\leq 100$ ms
Depth Imbalance	DI_raw	$(\text{depth\_bid} - \text{depth\_ask}) / (\text{depth\_bid} + \text{depth\_ask})$	Recalcul à chaque update depth
DI filtré	DI_t	Z-score, clamp $\pm 3\sigma$ , EMA 3 obs	identique OFI
Volatilité courte	$\sigma_t$	EWMA 100 ticks	continu

### 4.3 Construction du signal DI

1. **Aggregation** : volumes L1-L5 bid et ask.
2. **Normalisation** : ratio volumique pour obtenir DI\_raw.
3. **Clamp & lissage** :  $\pm 3\sigma$  puis EMA 3 observations.
4. **Test de pertinence** : corrélation DI  $\rightarrow \Delta\text{tick}$ , p-value  $< 0.05$ .

### 4.4 Mécanique de cotation améliorée

```
centre_t = r_t + \kappa_{inv} \cdot q_t + \beta_{ofi} \cdot OFI_t + \beta_{di} \cdot DI_t
spread_t = base_spread + \kappa_{vol} \cdot \sigma_t + \kappa_{inv} \cdot |q_t|
```

- **centre\_t** : prix de réserve AS déplacé par l'inventaire et deux signaux de flux (OFI, DI).
- **spread\_t** : part fixe ( $2\delta^*$ ) + élargissement linéaire à la fois de la volatilité ( $\kappa_{vol}$ ) et du risque d'inventaire ( $\kappa_{inv}$ ).

### 4.5 Quote ageing et refresh

- *Cancel/replace* toute quote qui :
- reste vivante **> 750 ms** ;
- ou si le **signe** de OFI ou DI bascule.

## 4.6 Calibration des paramètres

Paramètre	Sens	Méthode	Valeur initiale (BTC)
$\beta_{ofi}$	OFI $\rightarrow$ ticks	Régression Ridge	0.3 tick/u
$\beta_{di}$	DI $\rightarrow$ ticks	Grid-search sur Sharpe	0.2 tick/u
$\kappa_{inv}$	Pénalité inventaire centre & spread	Optimiser RMS q	0.1 tick/lot
$\kappa_{vol}$	Sensibilité volatilité	Régression spread vs $\sigma$	1.2

## 4.7 Contrôles de risque spécifiques

- **Clamp offset global** :  $|\kappa_{inv} q_t + \beta_{ofi} OFI_t + \beta_{di} DI_t| \leq 3 \text{ ticks}$ .
- **Spread plancher** :  $spread_t \geq 2 \text{ ticks}$  pour éviter exécutions gratuites.
- **Latency guard** : quote ageing 750 ms et ACK P99  $\leq 300$  ms.

## 4.8 Indicateurs de performance cibles

KPI	Objectif V1.5+	Gain vs V1- $\alpha$
PnL / trade	+30 %	$\uparrow$
Spread capturé	$\geq 78 \%$	+8 pts
RMS inventaire	$\leq 0.35 q_{max}$	-
Hit ratio OFI + DI	$\geq 60 \%$	n/a
Cancel ratio	$\leq 70 \%$	idem

## 4.9 Procédure de validation

1. **Back-test A/B** 14j : V1- $\alpha$  vs V1.5+.
2. **Out-sample** 7j : vérifier robustesse paramètres.
3. **Paper-trade** testnet : latence, fill, ageing.
4. **Canary** 1 % : PnL  $\geq$  baseline + latence ok.
5. KPI review hebdo et ajustement  $\beta_{ofi}$ ,  $\beta_{di}$ ,  $\kappa_{vol}$ .

# 5. Version 2 — Agent RL Soft-Actor-Critic (SAC) amélioré

## 5.1 Contexte et objectif

Cette version délègue la décision de cotation à un agent SAC **enrichi** : état plus riche, fonction de récompense multi-objectif et double critic afin de mieux gérer le compromis rentabilité / risque / latence.

## 5.2 Espace d'observation

Catégorie	Variables principales (normalisées)	Pourquoi ?
Prix & dérivés	$\Delta_{mid} 50 \text{ ms}$ , $\Delta_{mid} 250 \text{ ms}$ , log-retours cumulés	Momentum immédiat

Catégorie	Variables principales (normalisées)	Pourquoi ?
Flux d'ordres	<b>OFI 1 s &amp; 3 s</b> , compteurs aggTrade	Prévoir direction tick
Carnet (depth)	Profondeurs L1-L5 bid & ask, pente, <b>Depth Imbalance (DI)</b>	Pression liquidité et asymétrie
Volatilité	$\sigma$ EWMA 1 s & 10 s (vol-scaled)	Adapter spread
Position	Inventaire q et <b>position queue</b> (rang dans la profondeur)	Gestion risque & priorité
Temps	Time-to-close, minute intra-heure	Patterns saisonniers

**Dimensions typiques** : 50–70 features après concaténation.

### 5.3 Espace d'action (toutes les 50-100 ms)

- Centre shift  $\Delta c$  :  $-5 \rightarrow +5$  ticks.
- Log-spread  $\Delta s$  (exponentiel) :  $-1 \rightarrow +1$ .
- Refresh booléen  $p$  : 0 = laisse vivre, 1 = cancel/replace.

### 5.4 Récompense multi-objectif & double critic

$$\text{reward} = \Delta \text{PnL} - \lambda_q \cdot q^2 - \lambda_c \cdot \text{cancel} - \lambda_l \cdot \text{latency}$$

- **$\Delta \text{PnL}$**  : mark-to-mid variation sur le pas.
- **$\lambda_q$**  : pénalité inventaire (contrôle RMS  $q$ ).
- **$\lambda_c$**  : coût d'annulation (discipline d'ordres).
- **$\lambda_l$**  : coût latence (ms excédant budget).

Le réseau **actor** est entraîné à maximiser ce reward, tandis que nous utilisons **deux critics** :

1. Critic-mean : estime l'espérance du retour (standard SAC).
2. Critic-var : estime la **variance** cumulée du PnL afin de stabiliser le Sharpe (objectif auxiliaire : minimiser la variance).

### 5.5 Pipeline d'entraînement hors-ligne

1. 60 j de tick-data (side-bandit)  $\rightarrow$  buffer 1–2 M transitions.
2. Réseaux  $2 \times 256$  ReLU ; entropie  $\alpha$  auto-tuned.
3. Loss totale = Loss\_mean +  $\eta \cdot$  Loss\_var ( $\eta \simeq 0.1$ ).
4. Curriculum : actions bornées  $\pm 2$  ticks puis élargies à  $\pm 5$ .
5. Early-stop : reward moyen saturant  $\leq 0.3 \sigma$  sur 50 k steps.

### 5.6 Safety layer (identique, renforcé)

- Clamp  $|\Delta c| \leq 3$  ticks & spread final  $\geq$  min.
- $\Delta \text{offset}$  step-to-step  $\leq 2$  ticks.
- Quote ageing  $\geq 500$  ms.
- $|q| > q_{\text{max}} \rightarrow$  forcer réduction.
- ACK P99  $> 300$  ms  $\rightarrow$  fallback V1.5.

## 5.7 Déploiement progressif

Phase	Expo max	Durée	Promotion	Recul
Paper-trade	0 %	5 j	Latence & reward OK	-
Canary	1 %	3 j	PnL > 0 & KPI conformes	DD > 2 $\sigma$
Ramp-up	5 % → 25 %	+1 j/palier	PnL $\geq$ V1.5 & RMS q stable	KPI hors bornes
Prod	100 %	-	Revue mensuelle	guard-rail

## 5.8 Indicateurs de succès

KPI	Cible vs V1.5
PnL net / trade	$\geq +15$ %
RMS inventaire	$\leq 60$ % V1
Sharpe ajusté var	+20 %
Latence quote→ACK (P99)	$\leq 300$ ms
Cancel ratio	$\leq 70$ %

## 5.9 Plan de validation

1. Back-test V1.5 vs V2 sur même historique.
2. Shadow-mode 48 h en prod (V2 virtuel).
3. Canary réel selon § 5.7.
4. Stress tests :  $\sigma \times 2$ , depth  $\div 2$ , latence +200 ms.
5. KPI daily first week, then weekly.

## 6. Tests & validation

Un même **processus multi-étapes** s'applique aux trois versions, avec des exigences de plus en plus strictes.

### 6.1 Back-test hors-ligne

Étape	Objectif	Détails	Seuils de succès
Relecture tick-by-tick ( $\geq 14$ j)	Vérifier logique de cotation et calculs PnL	Side-bandit : on ne pousse pas nos quotes dans l'historique	PnL $\geq 0$ , RMS q $\leq 0.4$ q_max
Grid-search params	Trouver ( $\gamma$ , $k$ , $\beta$ , $\kappa$ ) optimaux	Maximiser Sharpe sous contrainte inventaire	Sharpe 1 décile > baseline
Stress scénarios $\sigma \times 2$ , depth $\div 2$	Robustesse vol/liquidité	Rejouer les pires journées	PnL reste positif, drawdown < 2 $\sigma$



## 6.2 Paper-trade (testnet ou « shadow-mode »)

Focus	Métriques suivies	Durée min
Latence infra	send→ACK P99, TTL snapshot	≥ 48 h
Discipline ordres	Cancel ratio, ageing effectif	
Cohérence signaux	Drift OFI/DI vs prod live	

## 6.3 Canary en prod (1 % du capital)

- Durée : 3 jours minimum.
- Promotion si : PnL net  $\geq 0$  et KPI latence / inventaire conformes.
- Recul (fallback version précédente) si : drawdown  $> 2\sigma$  baseline ou ACK P99  $> 300$  ms.

## 6.4 Ramp-up progressif

Palier	Exposition max	Condition passage
P1	5 %	PnL $\geq 0$ , RMS q conforme 24 h
P2	10 %	idem + cancel ratio $\leq 70$ %
P3	25 %	Sharpe 7j $\geq$ baseline
P4	100 %	Revue mensuelle risk & ops

## 6.5 Tests de stress en ligne

1. **Spike volatilité** :  $\sigma$  1 min  $> 3 \times \sigma$  10 min  $\rightarrow$  spread multiplié  $\times 1.5$ .
2. **Dégradation réseau** : RTT  $> 200$  ms  $\rightarrow$  monitoring fallback.
3. **Anomalie carnet** : depth L1  $<$  seuil  $\rightarrow$  suspendre cotation.

## 6.6 Calendrier de revue KPI

- **Quotidien** première semaine live : PnL, latence, inventaire RMS, cancels.
- **Hebdomadaire** ensuite : Sharpe, drawdown, hit ratio  $\alpha$ -signals.
- **Mensuel** : revue complète risk/compliance et recalibration éventuelle.

## 6.7 Critères de rollback

**Tout dépassement persistant** ( $> 15$  min) d'un garde-fou inventaire, latence ou cancel ratio déclenche le fallback automatique vers la version antérieure, suivi d'une analyse post-mortem.

## 7. Glossaire

Terme	Définition opérationnelle	Importance dans l'algo
<b>Mid-price</b>	Moyenne du meilleur bid et ask visibles ( $best\_bid + best\_ask / 2$ )	Point neutre pour calculer prix de réserve et PnL mark-to-mid

Terme	Définition opérationnelle	Importance dans l'algo
<b>Spread capturé</b>	Différence entre le prix d'exécution et le mid-price au moment de la cotation	Mesure l'efficacité du market-maker à « prendre » le spread
<b>OFI</b> ( <i>Order-Flow Imbalance</i> )	Ratio $(vol\_buy - vol\_sell) / (vol\_buy + vol\_sell)$ sur une fenêtre courte (0,5-3 s)	Micro-alpha directionnel dans V1-α, V1.5 et V2
<b>Depth Imbalance (DI)</b>	Asymétrie de profondeur L1-L5 : $(depth\_bid - depth\_ask) / (depth\_bid + depth\_ask)$	Signal secondaire de pression de carnet (V1.5, V2)
<b>RMS inventaire</b>	Racine carrée de la moyenne des carrés de la position nette sur une période	Indicateur clé de risque ; contraint par $\lambda\_q$ ou $\kappa\_inv$
<b>Canary</b>	Exposition très limitée ( $\approx 1\%$ ) utilisée pour tester une version en prod réelle	Permet rollback rapide si KPI dégradés
<b>CRC</b> ( <i>Cyclic Redundancy Check</i> )	Contrôle d'intégrité du snapshot Binance pour vérifier la cohérence du carnet local	Déclenche resynchronisation si mismatch
<b>Quote ageing</b>	Durée max pendant laquelle une quote reste active avant cancel / replace	Limite l'exposition latence et protège contre stale quotes
<b>Safety layer</b>	Ensemble de règles déterministes filtrant les actions du RL	Garantit la conformité latence, inventaire, spread
<b>Position queue</b>	Rang de la quote dans la profondeur au niveau de prix où elle est postée	Influence la probabilité d'exécution et la latence de remplissage