

ÉTUDE DE LA RELATION ENTRE LA SOMME DES SIGNAUX ET L'ÉVOLUTION FUTURE DES PRIX

La structure de la somme des signaux fournis par nos systèmes de prise de position sur les marchés financiers évolue en fonction de la configuration des marchés. L'étude des sommes des signaux confirme le bon comportement de nos systèmes lors des périodes de tendance. Sa mise en relation avec l'évolution future des prix sur les marchés financiers ne permet pas d'améliorer notre outil de détermination de la position optimale à investir sur les marchés.



Sommaire

Table des illustrations	3
1. Fonctionnement de nos systèmes de prise de position	4
1.1. Contexte	4
1.2. Calcul des signaux	4
1.3. Calcul de la position sur le marché	5
1.4. Objet de l'étude	5
2. Spécificités liées à notre approche métier	5
2.1. Back-testing	5
2.2. Ajustement des prix des contrats	5
2.3. Standardisation des prix	6
3. Analyse des signaux	7
3.1. Des marchés majoritairement haussiers	7
3.2. Intérêt à analyser les signaux par indicateur de trading	8
4. Analyse des sommes de signaux	9
4.1. Confirmation du biais haussier sur la période étudiée	9
4.2. Différenciation par catégorie	10
4.3. L'indicateur FOT comme facteur de référence	12
4.4. Caractérisation des marchés	13
5. Mise en relation de la somme des signaux et de l'évolution future des prix	13
5.1. Partitionnement espéré	14
5.2. Performance de nos systèmes de suivi de tendance	14
5.3. Modélisations envisageables	17
6. Modélisations	17
6.1. Deux échantillons de données aux caractéristiques différentes	17
6.2. Conserver notre formule actuelle de calcul des positions	19
Glossaire des termes financiers	20
Bibliographie	21

Table des illustrations

Figure 1 : Nombre de signaux calculés	4
Figure 2 : Calcul de la position sur le marché	5
Figure 3 : Illustration des gaps de roll	6
Figure 4 : Différences de volatilité annuelle par classe d'actifs	6
Figure 5 : Répartition des signaux	7
Figure 6 : Distribution des signaux par catégorie	8
Figure 7 : Distribution des signaux par indicateur de trading et par horizon temporel	9
Figure 8 : Histogramme de distribution de la somme des signaux	10
Figure 9 : Dispersion de la moyenne des sommes des signaux par catégorie	10
Figure 10 : Distribution de la somme des signaux par contrat	11
Figure 11 : Distribution de la somme des signaux par indicateur	12
Figure 12 : Illustration de la réaction des systèmes à l'évolution future des prix	13
Figure 13 : Partitionnement espéré avec 9 clusters	14
Figure 14a : Partitionnement obtenu avec l'évolution des prix à 1 mois	15
Figure 14b : Partitionnement obtenu avec l'évolution des prix à 2 mois	15
Figure 14c : Partitionnement obtenu avec l'évolution des prix à 3 mois	16
Figure 15 : Dispersion de la moyenne des sommes de signaux par catégorie, dans l'échantillon basé sur les dates	18
Figure 16 : Distribution des sommes de signaux par contrat, dans l'échantillon basé sur les dates	18

1. Fonctionnement de nos systèmes de prise de position

1.1. Contexte

Au sein de notre société, nous calculons une fois par semaine des signaux d'ordres sur des marchés financiers. Ces produits financiers sont des contrats à terme. Nous les nommerons contrats par la suite. Nos systèmes en traitent 24 mais nous disposons des données quotidiennes pour plus de 50 contrats.

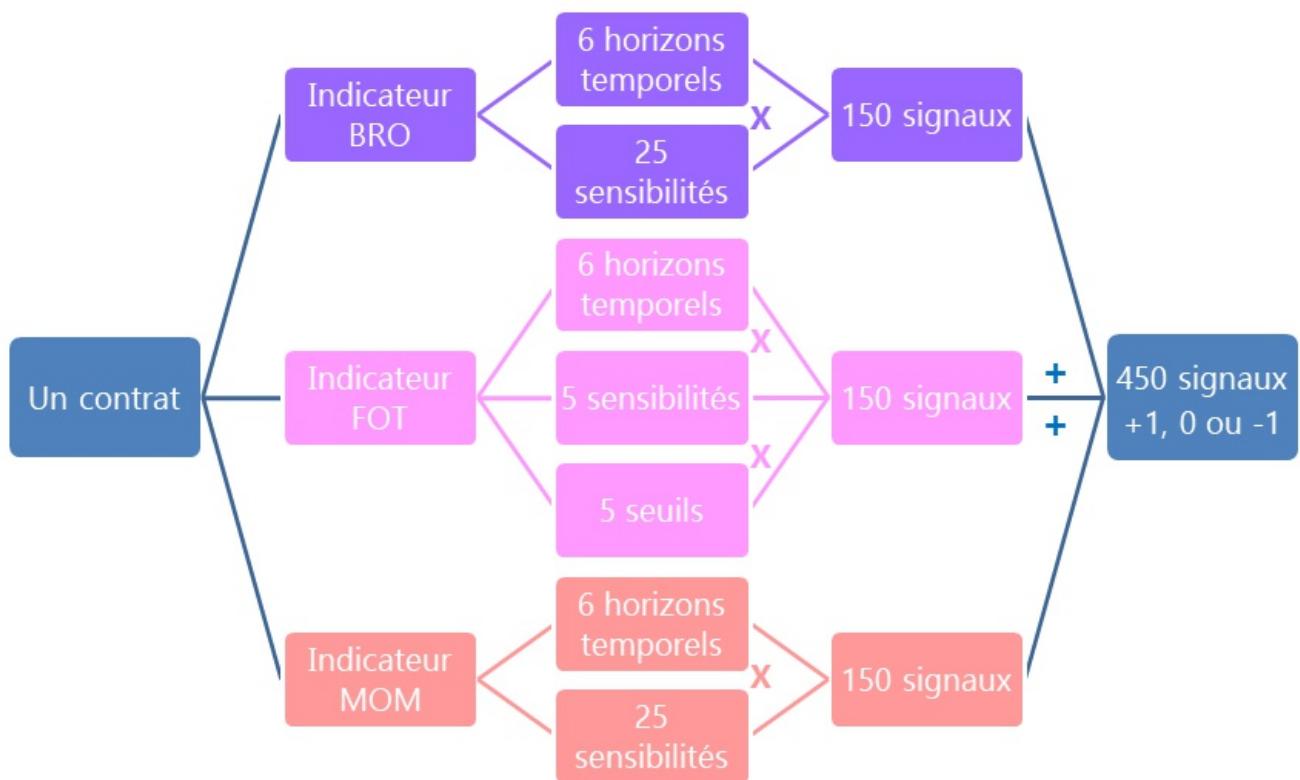
Les signaux d'ordre sont hebdomadaires mais ils pourraient être quotidiens. Ils cherchent à capter les tendances à moyen-long terme, à travers des positions qui sont maintenues entre 1 mois et 6 mois.

1.2. Calcul des signaux

Pour chaque contrat traité, nous calculons des signaux d'ordre à partir de 3 indicateurs techniques que nous appelons BRO (Break-Out), MOM (Momentum) et FOT (Force Of Trend). Chaque indicateur dispose de 150 réglages différents et produit donc 150 signaux. Chaque signal peut prendre la valeur -1, 0 ou 1.

Ces 150 réglages différents proviennent du croisement de 2 ou 3 paramètres : l'horizon temporel, la sensibilité (réactivité à l'évolution du marché), et éventuellement un seuil.

Figure 1 : Nombre de signaux calculés



Les trois indicateurs utilisent les 6 mêmes valeurs pour l'horizon temporel :

- 3 horizons à moyen-terme (MT1, MT2 et MT3) ;
- 3 horizons à long-terme (LT1, LT2 et LT3).

Deux indicateurs (BRO et MOM) utilisent les 25 mêmes valeurs de sensibilité. Le croisement des 6 horizons temporels et des 25 sensibilités donne 150 signaux.

Le dernier indicateur (FOT) utilise 5 valeurs de sensibilité et 5 valeurs de seuil. La combinaison des 6 horizons temporels, des 5 sensibilités et des 5 valeurs de seuil aboutit également à 150 signaux.

Au total, le système calcule donc 450 signaux, égaux chacun à -1, 0 ou +1, pour un contrat donné à une date donnée.

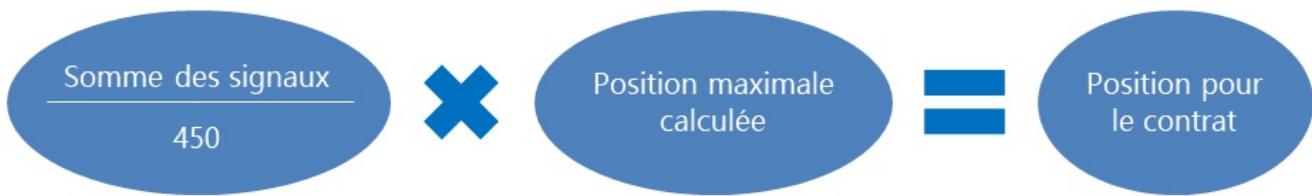
1.3. Calcul de la position sur le marché

Le système génère la taille de la position sur le marché à partir de la somme de ces 450 signaux et des conditions de marché, en croisant deux facteurs :

- une formule de money management détermine la position maximale autorisée sur le contrat ;
- la moyenne des signaux pondère cette position maximale pour obtenir la position sur le marché.

Si tous les signaux sont nuls (0) ou si j'ai autant de signaux acheteurs (+1) et vendeurs (-1), nous obtenons une moyenne nulle et donc une position nulle. A l'inverse extrême, si tous les signaux sont dans le même sens, la position sur le marché correspondra à la position maximale autorisée sur le marché. Il existe une multitude de situations intermédiaires entre ces extrêmes.

Figure 2 : Calcul de la position sur le marché



1.4. Objet de l'étude

Ce projet vise à étudier le remplacement de la moyenne des signaux par une autre pondération. Cette nouvelle pondération serait basée sur un modèle prenant en compte un ou plusieurs facteurs tels que la somme des signaux, les sommes par indicateur, ou les sommes par horizon temporel. Pour chacun de ces facteurs, le modèle peut également intégrer l'évolution récente de la somme considérée.

Nous devons définir une variable endogène afin mener nos modélisations. Etant donné que nous cherchons à identifier les tendances haussières ou baissières, nous nous basons sur l'évolution du prix dans le temps futur par rapport au jour de l'analyse.

Pour refléter notre horizon d'investissement à moyen-long terme, nous menons plusieurs modélisations avec l'évolution du prix à 1 mois, à 2 mois et à 3 mois comme variables endogènes.

2. Spécificités liées à notre approche métier

2.1. Back-testing

Les données sur les signaux sont générées par les programmes de back-testing de nos modèles de prise de décision. Le code est adapté afin d'obtenir les variables suivantes : l'ID du contrat, la catégorie d'actif, la date de référence des données, l'indicateur de trading, l'horizon temporel, et le signal SNL obtenu.

Chaque ligne de donnée correspond donc au signal obtenu pour un contrat donné par un indicateur avec un réglage précis.

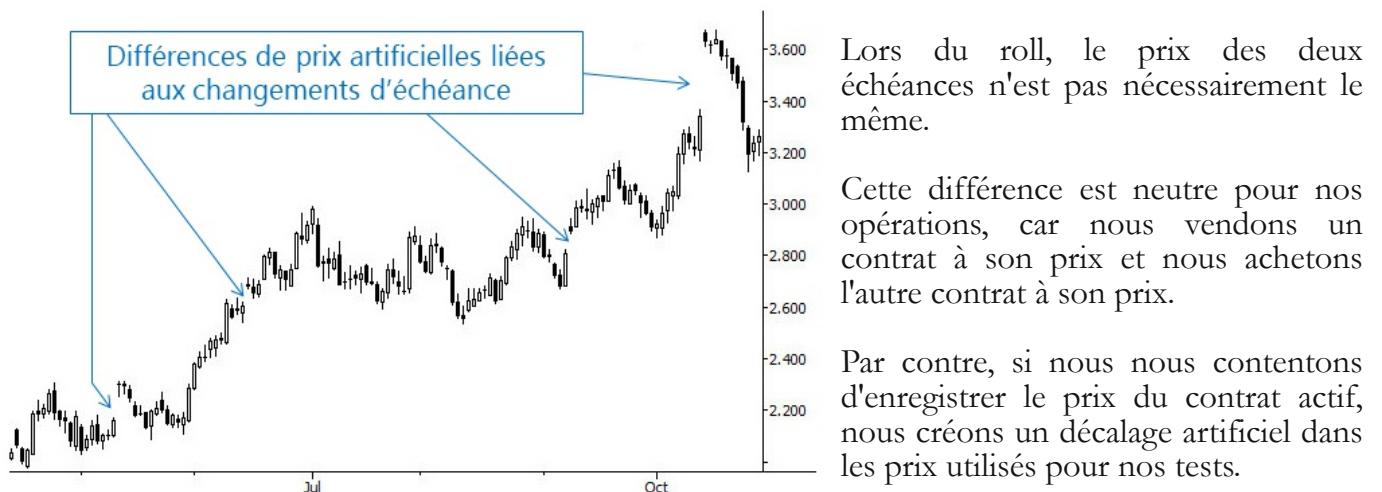
2.2. Ajustement des prix des contrats

Nous disposons des cotations depuis le début des années 1990 pour l'ensemble des contrats. Les données les plus anciennes ont été achetées et nettoyées il y a plusieurs années. Les données les plus récentes ont été collectées quotidiennement à l'aide d'outils professionnels payants, et vérifiées tous les jours. Nous détenons donc des données sur les prix propres et de bonne qualité.

Les contrats à terme possèdent la spécificité d'avoir une échéance. Plusieurs échéances différentes sont cotées en même temps pour un même sous-jacent et n'ont pas le même prix. Lorsque le contrat

approche de son expiration, les positions sont clôturées et reportées sur l'échéance suivante grâce à un contrat de spread. Ce changement d'échéance s'appelle le roll.

Figure 3 : Illustration des gaps de roll



Pour résoudre ce problème, deux séries de données sont stockées :

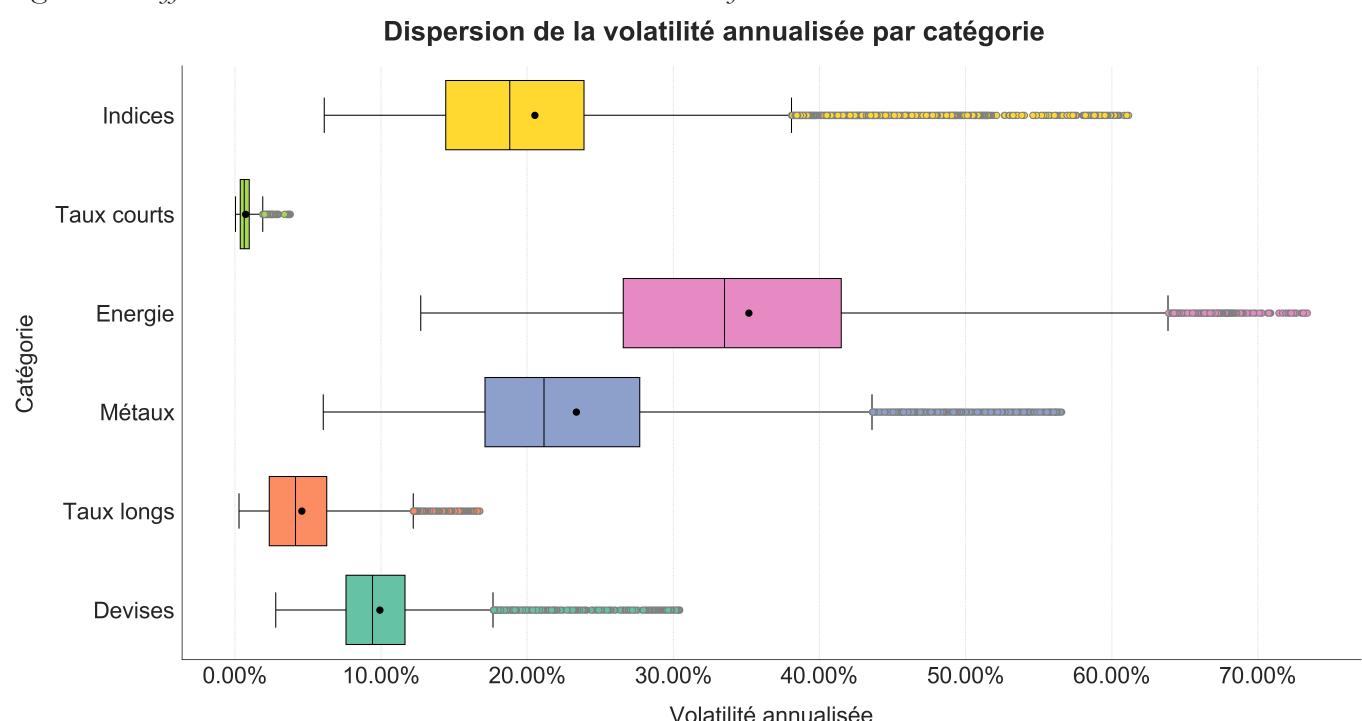
- le prix réel, qui garde en mémoire le prix pour chaque échéance ;
- le prix continu, pour lequel nous décalons les valeurs passées à chaque fois que nous changeons d'échéance. Ce prix reflète la variation réelle subie par le marché et nos positions, sans le décalage artificiel lié au changement d'échéance. Pour ce faire, les valeurs passées sont décalées de la différence entre les valeurs de clôture des deux échéances la veille du jour de roll.

Pour notre étude, nous utilisons le prix continu et le prix réel. Chaque ligne est donc constituée des variables suivantes : l'ID du contrat, la date de référence du jour de cotation, le prix de clôture continu (close_price), le prix moyen continu (mid_price), le prix de clôture réel (real_price).

2.3. Standardisation des prix

Les différentes classes d'actifs présentent des profils de risque très différents, qui se traduisent par des volatilités plus ou moins importantes. La figure 4 permet de classer les classes d'actifs par volatilité croissante : taux courts, taux longs, devises, indices boursiers, métaux et énergie.

Figure 4 : Différences de volatilité annuelle selon la classe d'actifs



La volatilité se définit par l'écart-type de la variation du prix en pourcentage sur une période donnée. La variation du prix en pourcentage est obtenue par la différence des prix continus divisée par le prix réel non rollé.

$$\text{Volatilité} = \sigma \left(\frac{\text{Prix continu du jour} - \text{Prix continu de la veille}}{\text{Prix réel de la veille}} \right)$$

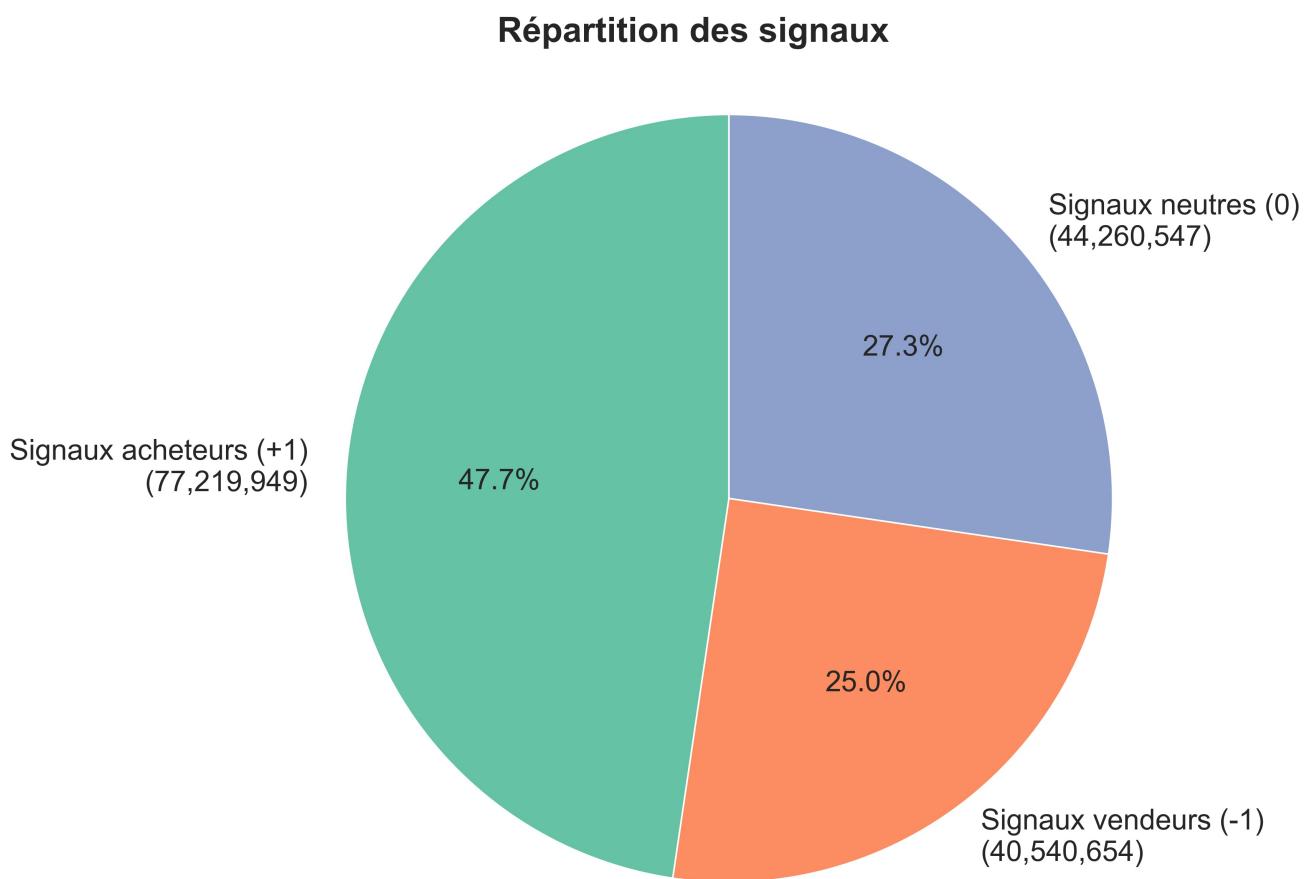
Afin de comparer l'évolution des prix d'un marché à l'autre, et d'étudier tous les contrats avec le même modèle, nous standardisons l'évolution des prix en la divisant par la volatilité.

3. Analyse des signaux

Notre base de données contient 162 021 150 signaux provenant du calcul du back-testing de nos systèmes sur 51 contrats entre 1989 et 2019.

3.1. Des marchés majoritairement haussiers

Figure 5 : Répartition des signaux



Il y a à peu près autant de signaux neutres et vendeurs. Le nombre de signaux acheteurs est largement supérieur et représente presque la moitié du total.

Les signaux neutres correspondent à des phases où les marchés n'ont pas de tendance marquée ou sont dans des périodes de consolidation ou d'hésitation.

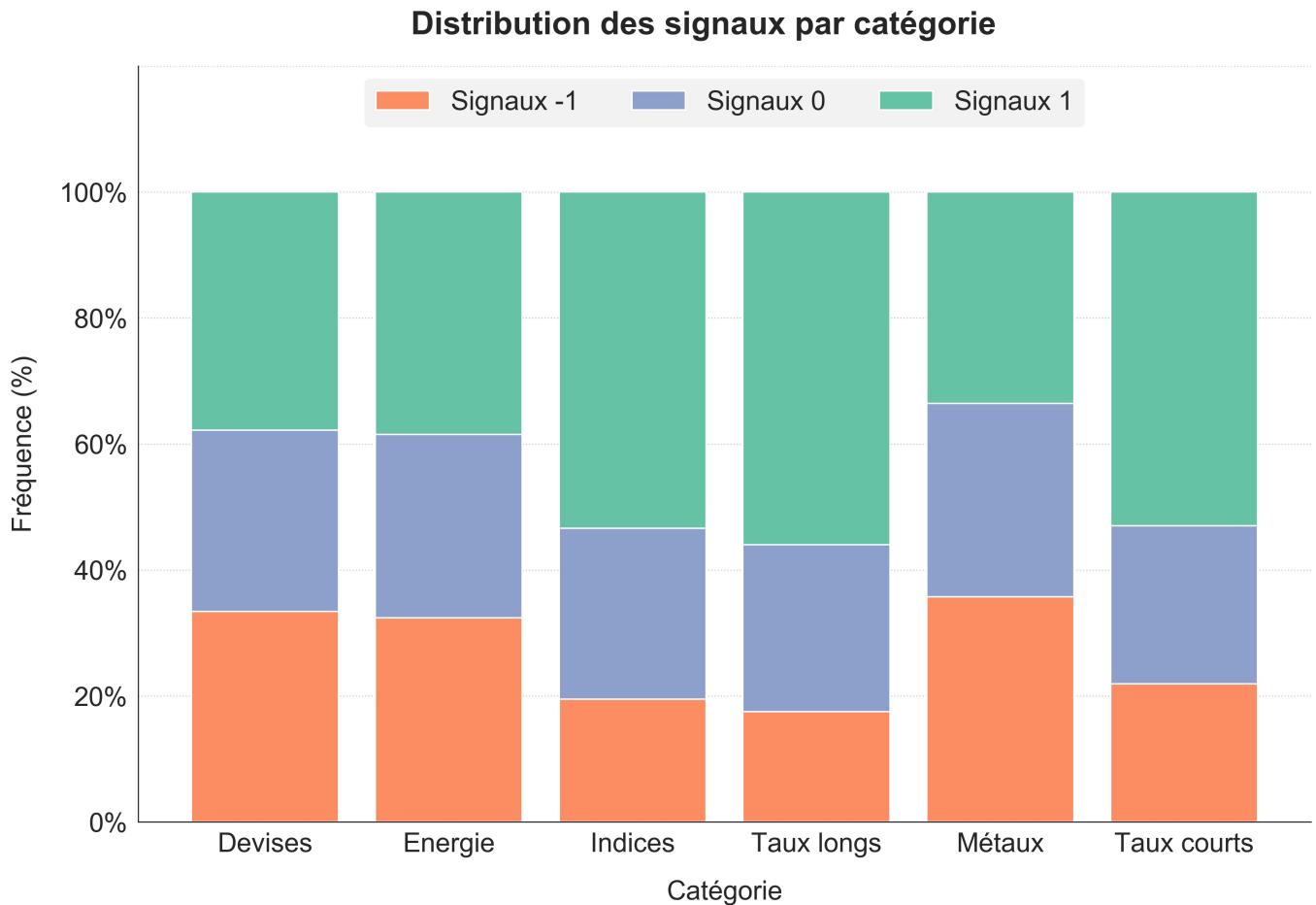
Le grand nombre de signaux acheteurs s'explique facilement par la structure générale des marchés. La plupart des marchés ont des phases haussières assez progressives et longues alors que les périodes baissières sont souvent plus rapides et fortes.

De plus, sur la période étudiée, tous les sous-jacents liés aux taux d'intérêt (à court terme comme à long terme) ont connu plusieurs grandes tendances haussières quasi-continues. Ces contrats

représentent 45% (23 sur 51) des contrats analysés et biaisen donc le nombre de signaux acheteurs vers le haut.

De la même manière, les indices boursiers ont connu plus d'années haussières que baissières. Même si certaines phases baissières ont été brusques en termes de variations, elles ont souvent été limitées dans le temps.

Figure 6 : Distribution des signaux par catégorie



Cette analyse métier est confortée par l'analyse de la distribution des signaux par catégorie. Le nombre de signaux acheteurs et vendeurs est à peu près égal pour les catégories des devises, de l'énergie et des métaux. Par contre, il y a une majorité de signaux positifs pour les catégories des indices boursiers, des taux courts et des taux longs (plus de 50% du total des signaux de ces catégories). La proportion de signaux neutres est à peu près constante d'une catégorie à l'autre.

3.2. Intérêt à analyser les signaux par indicateur de trading

Les indicateurs de trading BRO et MOM se comportent différemment de l'indicateur FOT.

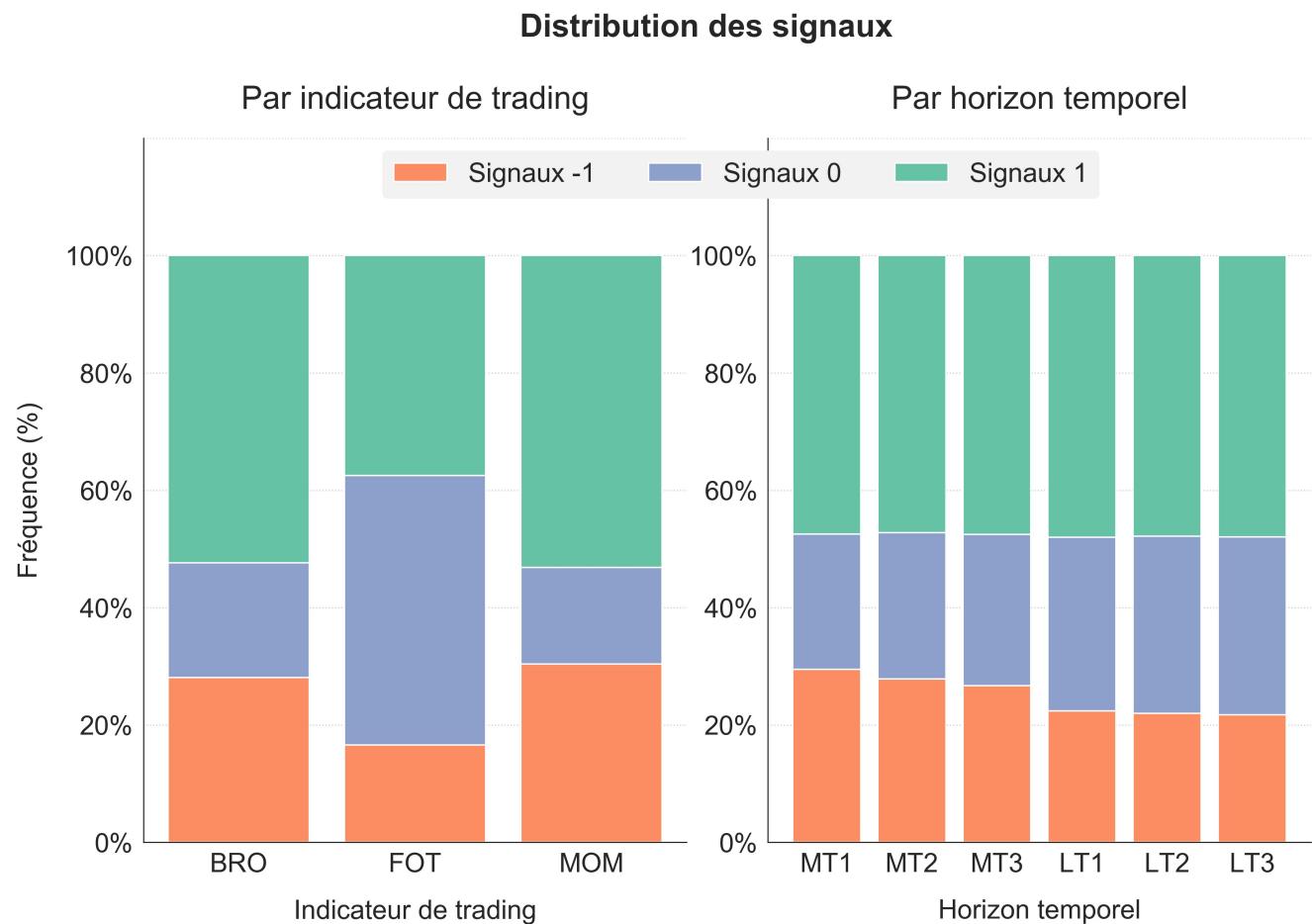
Les deux premiers nommés ont très peu de signaux neutres (moins de 20%). Ils présentent une majorité de signaux acheteurs et environ 30% de signaux vendeurs. Ils restent donc peu souvent hors du marché, même lors des phases de retournement ou de tendance peu marquée.

A l'inverse, l'indicateur FOT est plus sélectif pour générer des signaux acheteurs ou vendeurs. Il déclenche des signaux acheteurs ou vendeurs seulement lorsque la tendance est bien marquée. Environ 45% des signaux qu'il génère sont des signaux neutres. Parmi les 55% restants, les signaux acheteurs sont environ deux fois plus nombreux que les signaux vendeurs.

L'analyse des signaux par horizon de trading est beaucoup moins intéressante. Ce facteur ne semble pas très discriminant. Les signaux acheteurs représentent entre 45% et 50% du total, quel que soit l'horizon de trading. La seule petite nuance est la proportion des signaux vendeurs qui augmente au

détriment des signaux neutres à mesure que l'horizon temporel se réduit. Cela s'explique par la structure des corrections baissières, qui sont souvent brusques et fortes. Elles déclenchent donc plus facilement des signaux sur des horizons plus courts. Les signaux sur les horizons plus longs se retournent rarement sur ces corrections, mais plutôt sur de vraies tendances baissières durables.

Figure 7 : Distribution des signaux par indicateur de trading et par horizon temporel



4. Analyse des sommes de signaux

Notre modèle agrège les signaux par contrat et par jour afin de calculer les positions sur le marché. L'étude de la somme des signaux est donc plus importante que l'étude des signaux individuels dans l'optique qui anime ce projet. Nous nous attendons à retrouver dans les sommes des signaux certaines caractéristiques observées sur les signaux purs.

Pour rappel, la somme des signaux est nécessairement comprise entre -450 et +450.

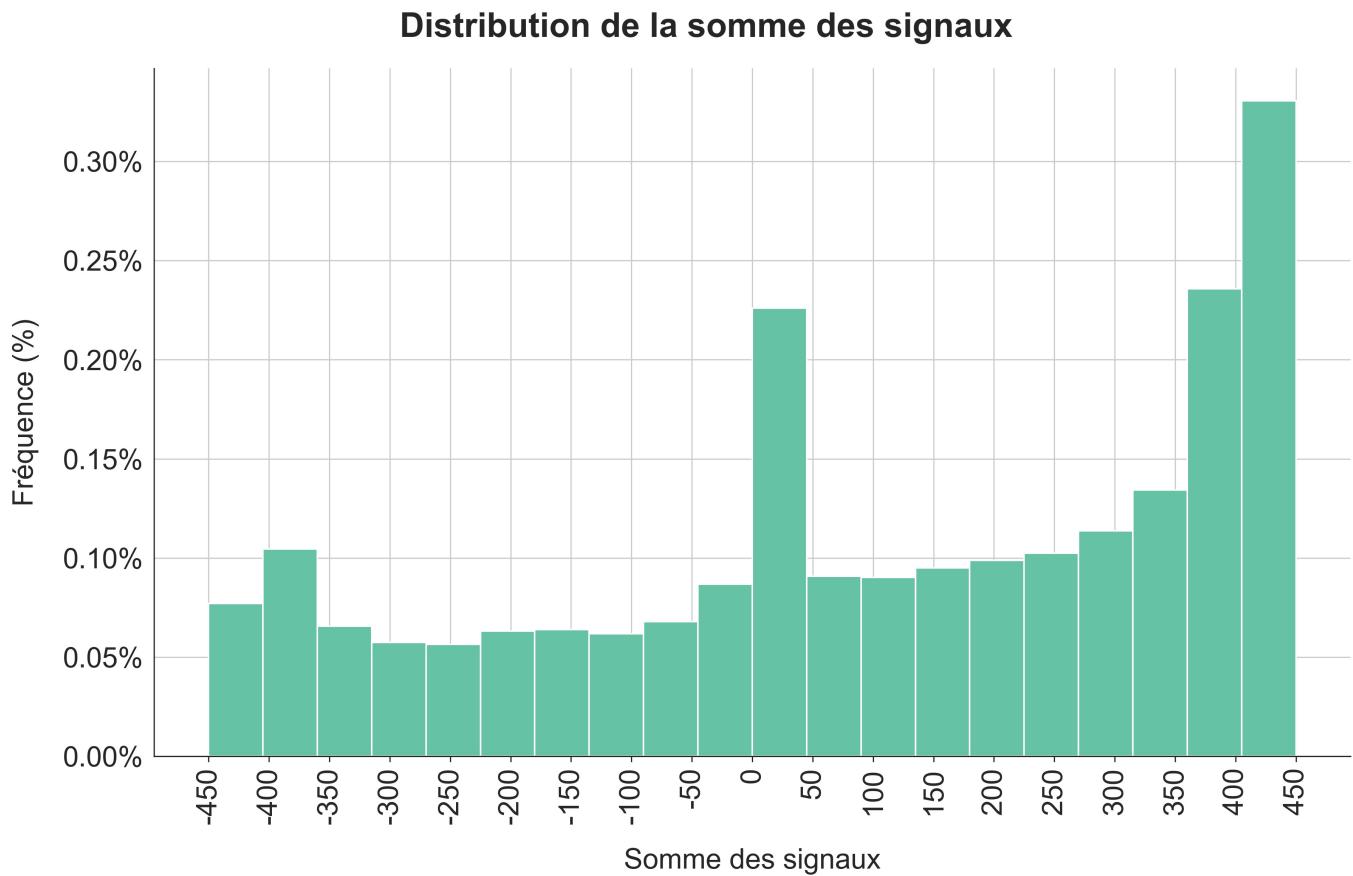
4.1. Confirmation du biais haussier sur la période étudiée

La somme des signaux s'étend entre -448 et +450. La moyenne des sommes est égale à 101,87 tandis que la médiane s'établit à 134,0.

Ces statistiques confirment l'observation précédente que les marchés ont globalement connu plus de tendances haussières que de tendances baissières sur la période étudiée.

L'histogramme de la somme des signaux montre une distribution qui ne suit pas du tout une loi normale. La distribution est multimodale et décalée vers la droite. Les sommes positives sont surreprésentées, en particulier au-delà de 350. Nous notons également un mode important entre 0 et 50.

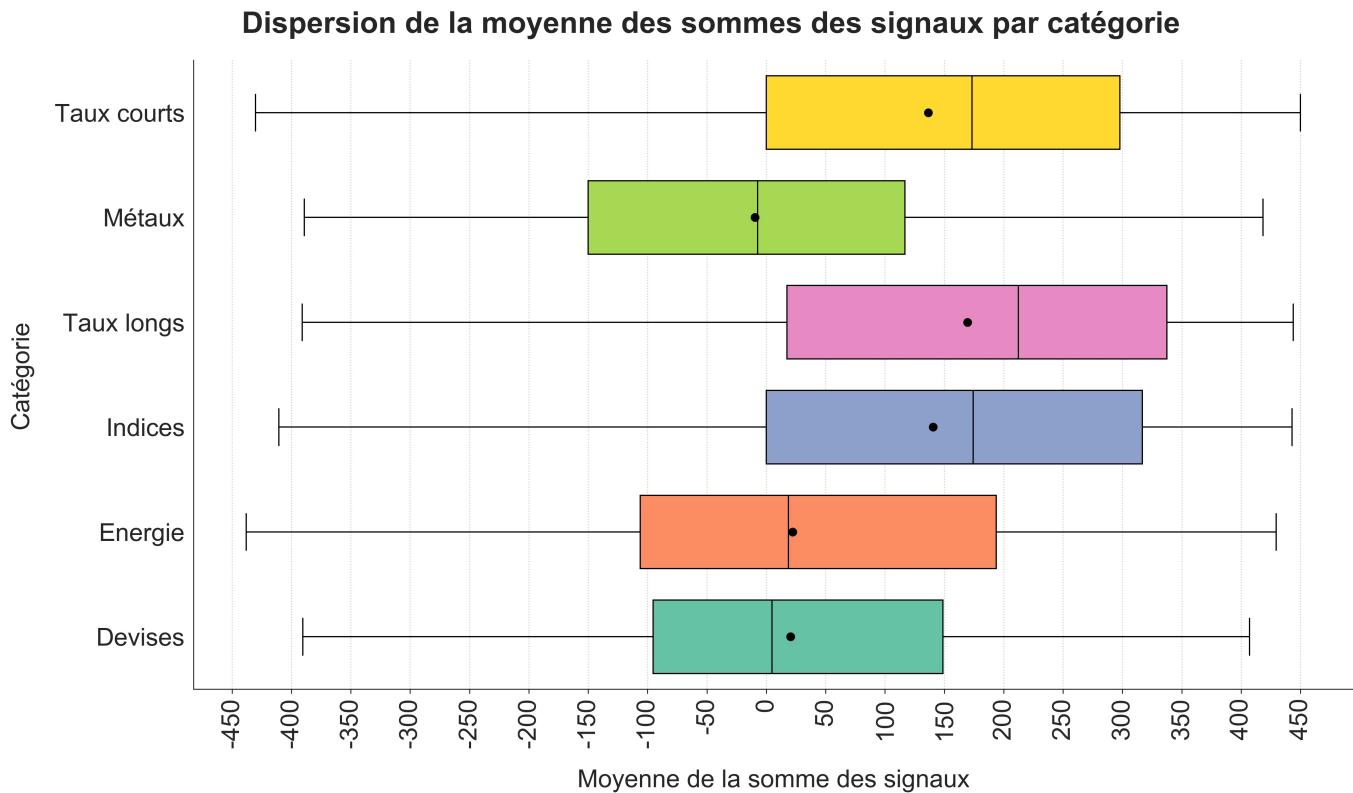
Figure 8 : Histogramme de distribution de la somme des signaux



4.2. Différenciation par catégorie

Afin d'observer les différences entre catégories, nous calculons chaque jour la moyenne des sommes de signaux de tous les contrats qui composent chaque catégorie.

Figure 9 : Dispersion de la moyenne des sommes des signaux par catégorie



La moyenne et la médiane sont proches de zéro pour trois classes d'actifs : les métaux, les devises et l'énergie. Seule la catégorie des métaux présente un corps de boîte à moustache plus orienté vers les valeurs négatives.

Les taux courts, les indices boursiers et les taux longs ont des moyennes très positives. Leurs médianes sont nettement supérieures à leurs moyennes, signe d'un skewness négatif. Ces mesures confirment la prédominance des tendances haussières pour ces trois catégories sur la période étudiée.

L'analyse par contrat montre une certaine similitude entre les contrats appartenant à une même classe d'actifs.

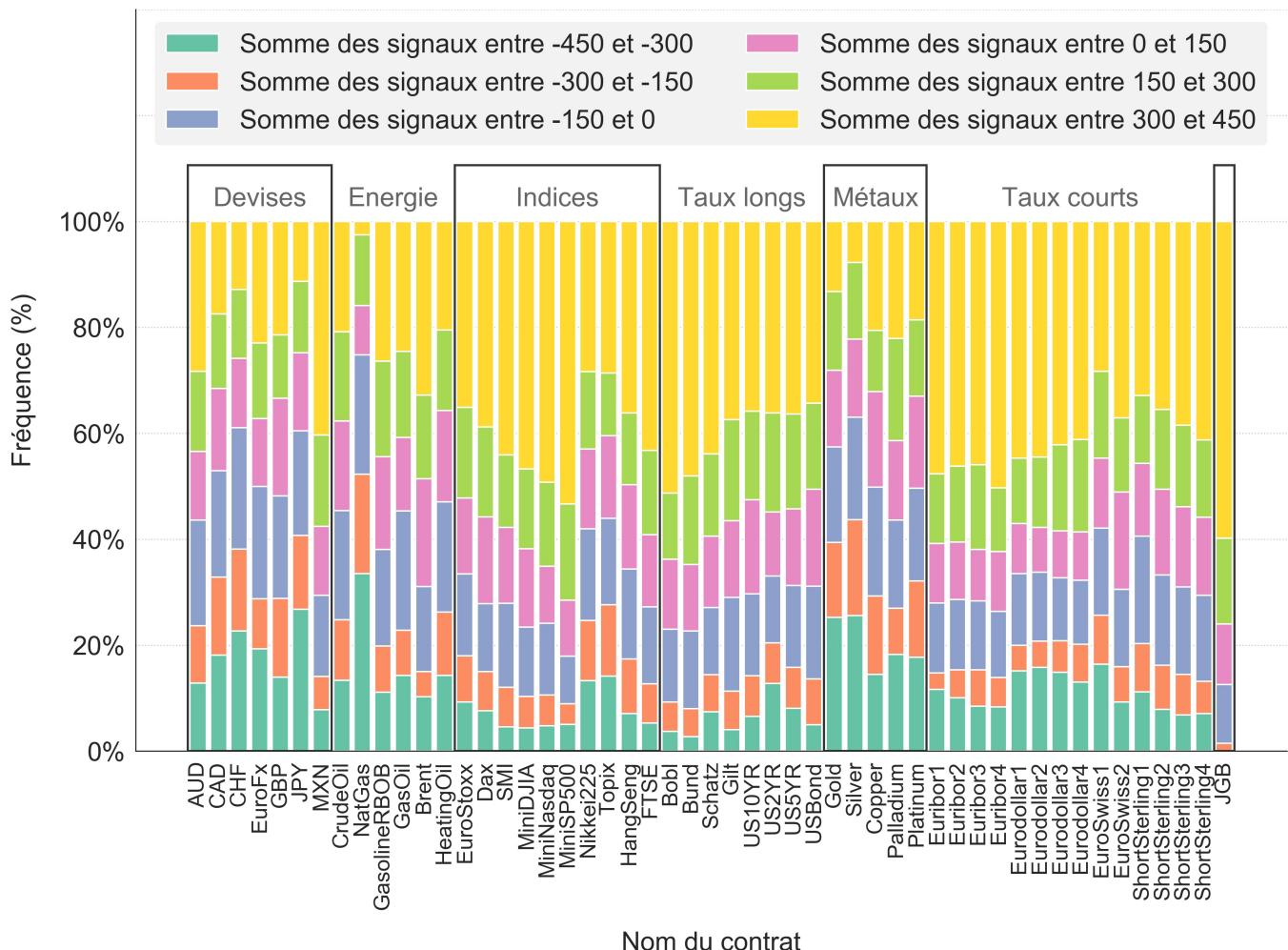
Les contrats se distinguent peu sur la fréquence des sommes comprises entre -150 et +150. Ils se différencient essentiellement sur la fréquence des sommes les plus extrêmes, à savoir des sommes supérieures à +300 et inférieures à -300.

Comme attendu, les plus fortes fréquences de sommes supérieures à +300 se retrouvent parmi les contrats des classes d'actifs des indices boursiers, des taux courts et des taux longs.

Les quelques contrats qui possèdent une fréquence des sommes négatives supérieure à la fréquence des sommes positives appartiennent principalement aux catégories des métaux et des devises.

Figure 10 : Distribution de la somme des signaux par contrat

Distribution de la somme des signaux par contrat



Deux contrats se démarquent des autres.

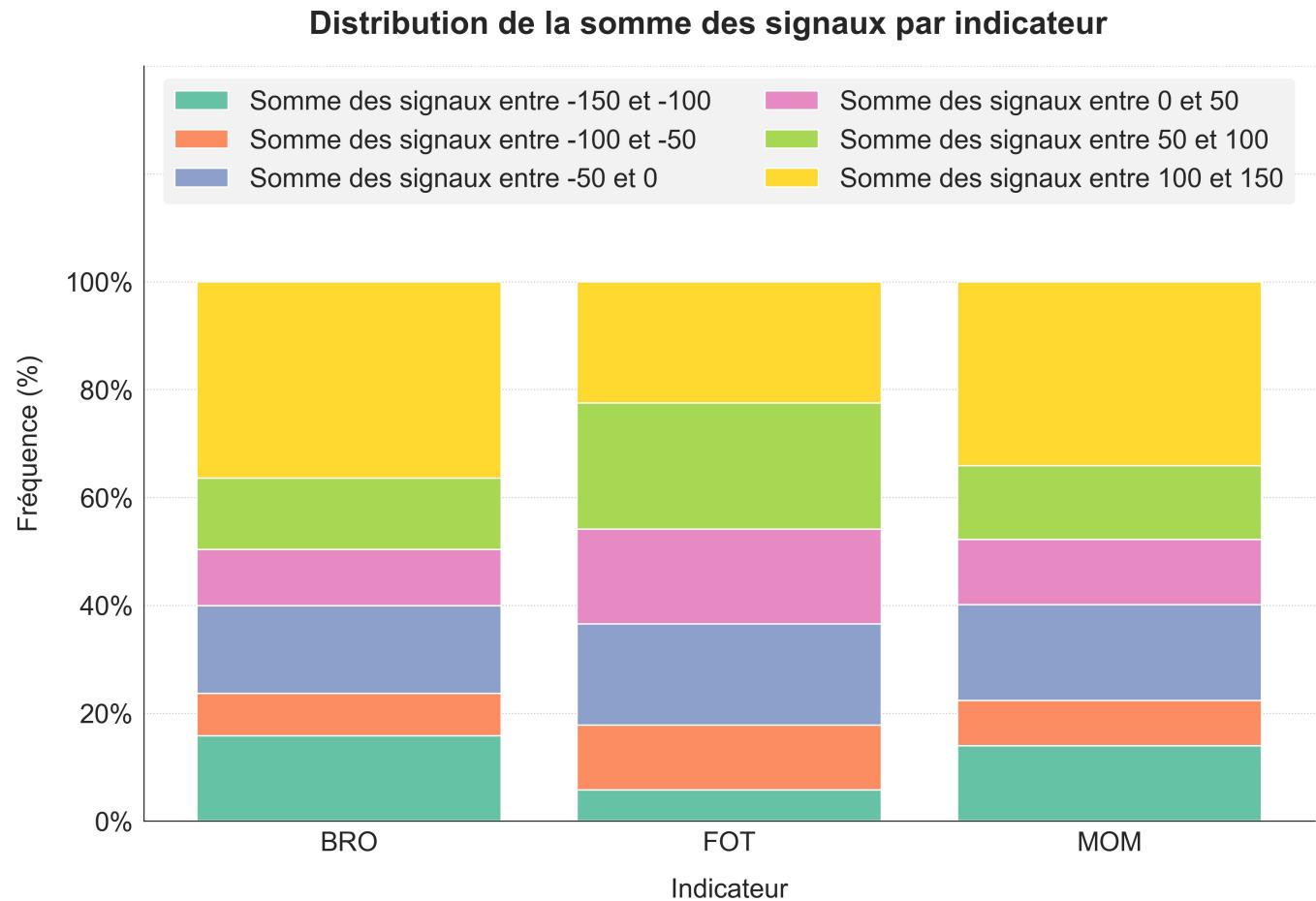
D'une part, le JGB (taux longs japonais à 10 ans) n'a aucune somme des signaux inférieure à -300, et très peu de sommes inférieures à -150. La conséquence logique est que les sommes supérieures à +300 sont largement sur-représentées, avec une fréquence d'environ 60%.

D'autre part, le gaz naturel est le seul contrat où la fréquence des sommes inférieures à -300 dépasse le seuil des 30%. Ce contrat est également le seul dont les sommes très positives présentent une fréquence inférieure à 5%.

4.3. L'indicateur FOT comme facteur de référence

L'analyse par indicateur confirme la sur-représentation des sommes de signaux positives, dont la fréquence dépasse 60% quel que soit l'indicateur.

Figure 11 : Distribution de la somme des signaux par indicateur



Les trois indicateurs se différencient sur les sommes les plus extrêmes.

L'indicateur FOT présente les fréquences les plus basses dans les sommes extrêmes. Cela signifie qu'une partie de ses signaux est moins réactive à se déclencher lors de l'apparition de tendances ou plus prompte à s'annuler lors des retournements de configuration. C'est l'indicateur le plus sélectif du point de vue de l'analyse des tendances.

A l'inverse, l'indicateur BRO est le plus rapide à se déclencher ou le plus long à se retourner. L'indicateur MOM se comporte de façon assez similaire à l'indicateur BRO, même s'il est très légèrement moins réactif.

Ces observations mettent en évidence que la somme des signaux de l'indicateur FOT peut constituer un facteur de référence pour détecter les phases de tendance avérée et prononcée des marchés. Deux options s'offrent à nous : considérer cette somme comme telle ou la considérer comme un booléen indicateur de tendance si elle dépasse un certain seuil.

L'analyse des sommes par horizon temporel n'apporte aucune information de valeur supplémentaire. Elle confirme le caractère peu discriminant de cette mesure, entrevu lors de l'analyse des signaux individuels.

4.4. Caractérisation des marchés

La somme des signaux permet de caractériser les conditions de marché. Sa distribution varie d'une catégorie à l'autre et d'un indicateur à l'autre, alors qu'elle est peu sensible à l'horizon temporel.

Par rapport à notre approche initiale, cela conforte l'utilisation de la somme des signaux et des sommes par indicateur dans nos essais de modélisation. Par contre, les sommes par horizon temporel n'apportent pas d'information supplémentaire et nous abandonnons la piste de leur utilisation.

5. Mise en relation de la somme des signaux et de l'évolution future des prix

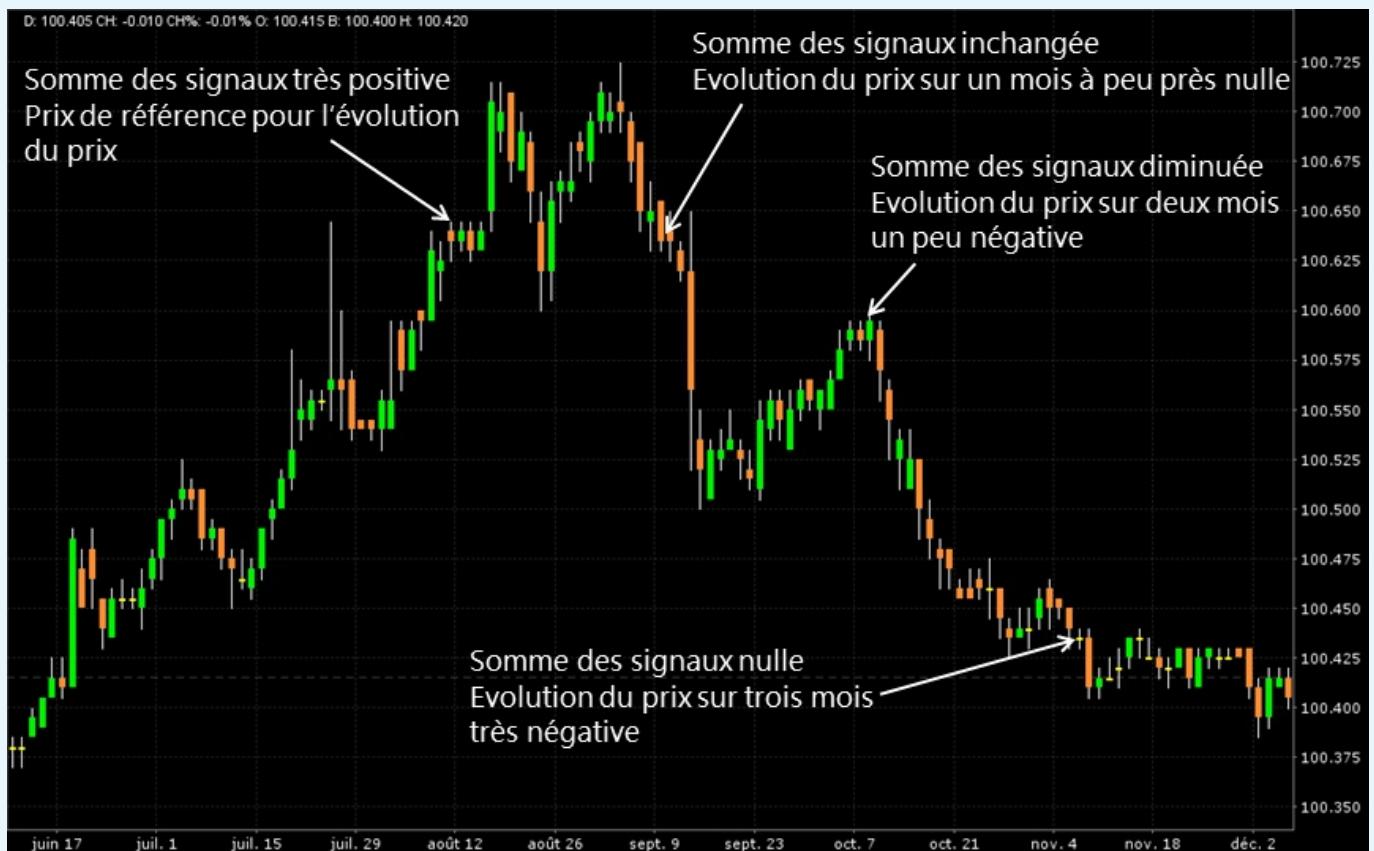
Nous partitionnons les données en nous basant sur deux paramètres : la somme des signaux à un moment donné et l'évolution des prix dans le futur. Nous considérons l'évolution des prix standardisée, telle qu'elle est définie précédemment.

L'évolution des prix standardisée est calculée sur des horizons d'un mois (21 jours de marché), de deux mois (42 jours de marché) et de trois mois (64 jours de marché), afin d'étudier différentes configurations de marché.

Il est d'usage de considérer qu'environ un tiers des positions d'un système suiveur de tendance sont profitables. Les profits globaux viennent du fait que les positions gagnantes sont conservées longtemps pendant les tendances à moyen-long terme, alors que les positions perdantes sont coupées plus vites. Cette analyse du partitionnement des données vise à vérifier que nos systèmes sont aptes à capter les tendances futures du marché et se comportent comme attendu.

Focus : L'adaptation de nos systèmes à l'évolution du marché n'est pas prise en compte

Figure 12 : Illustration de la réaction des systèmes à l'évolution future des prix



Il faut garder à l'esprit que les statistiques étudiées sont une somme de photos instantanées prises jour après jour. Lorsque nous comparons la somme des signaux à un moment donné à l'évolution

future des prix, nous faisons abstraction de la réaction de nos systèmes aux évolutions futures des prix.

Dans la réalité, si notre système a une somme de signaux élevée, et donc une position très acheteuse, et si les prix se mettent à baisser pendant les mois qui suivent, nos systèmes vont progressivement s'adapter à la nouvelle configuration. La somme des signaux va progressivement diminuer, voire devenir négative.

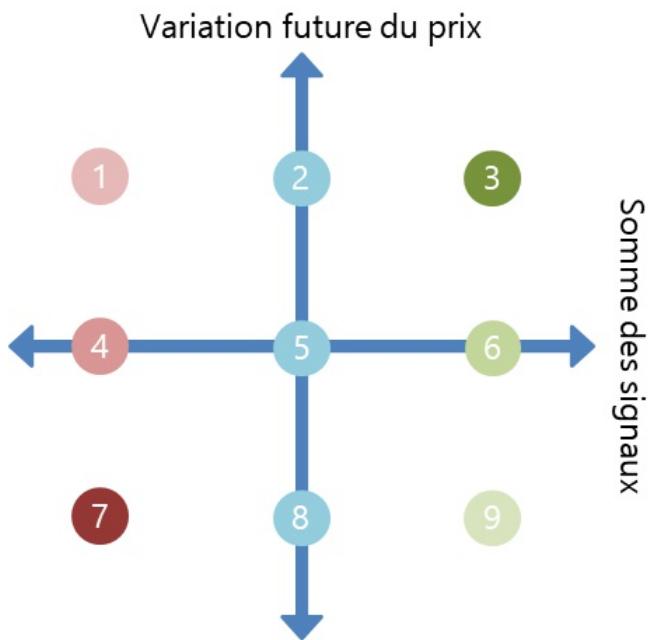
La figure 12 illustre l'adaptation des systèmes aux conditions de marché. Si nous regardons les statistiques au 12 août dans cet exemple, la somme des signaux est très positive, l'évolution du prix à un mois nulle, l'évolution du prix à 2 mois légèrement négative et l'évolution du prix à 3 mois très négative.

Dans les faits, la somme des signaux est très positive depuis longtemps car ce marché a connu une hausse assez régulière depuis plusieurs mois, permettant de générer des profits. Un mois plus tard, la somme des signaux n'est pas modifiée car les modèles n'ont pas encore réagi. Deux mois plus tard, certains systèmes ont commencé à se désactiver ou à s'inverser, et la somme des signaux a diminué. Enfin, trois mois plus tard, la somme des signaux est désormais nulle. Les systèmes avec les horizons temporels les plus longs sont toujours acheteurs alors que les systèmes avec les horizons temporels les plus courts sont devenus vendeurs.

5.1. Partitionnement espéré

Les outils permettant de déterminer le nombre optimal de classes de regroupement amènent à choisir entre 8 et 12 clusters quel que soit l'horizon de variation de prix considéré.

Figure 13 : Partitionnement espéré avec 9 clusters



Idéalement, nous aimerais que les données soient regroupées en 9 clusters de manière assez symétrique sur chacun des axes, afin d'obtenir une matrice 3x3 avec :

- les sommes des signaux regroupées selon qu'elles sont significativement négatives, significativement positives ou entre les deux (donc à peu près nulles) ;
- les évolutions standardisées du prix regroupées selon qu'elles sont significativement négatives, significativement positives ou entre les deux (peu significatives).

Les clusters 1 et 9 constituent une approximation des positions perdantes, mais ils ne prennent pas en compte la réactivité de nos systèmes aux évolutions du marché.

Les clusters 3 et 7 forment une estimation des positions gagnantes.

Les clusters 4 et 6 représentent les positions sans impact significatif sur les gains du système. Enfin les clusters 2, 5 et 8 regroupent les situations de marché sans prise de position.

L'analyse du partitionnement avec 9 clusters n'aboutit malheureusement pas à une telle matrice 3x3. Nous n'obtenons que 2 clusters dans les sommes de signaux significativement négatives, quel que soit l'horizon temporel de l'évolution future du prix.

5.2. Performance de nos systèmes de suivi de tendance

L'analyse avec 10 clusters est plus probante. Nous obtenons en effet la matrice espérée, que ce soit avec l'évolution future des prix à 1 mois, à 2 mois ou à 3 mois. La dixième classe regroupe des

Figure 14a : Partitionnement obtenu avec l'évolution des prix à 1 mois

Résumé du k-means avec 10 clusters - Evolution sur 1 mois

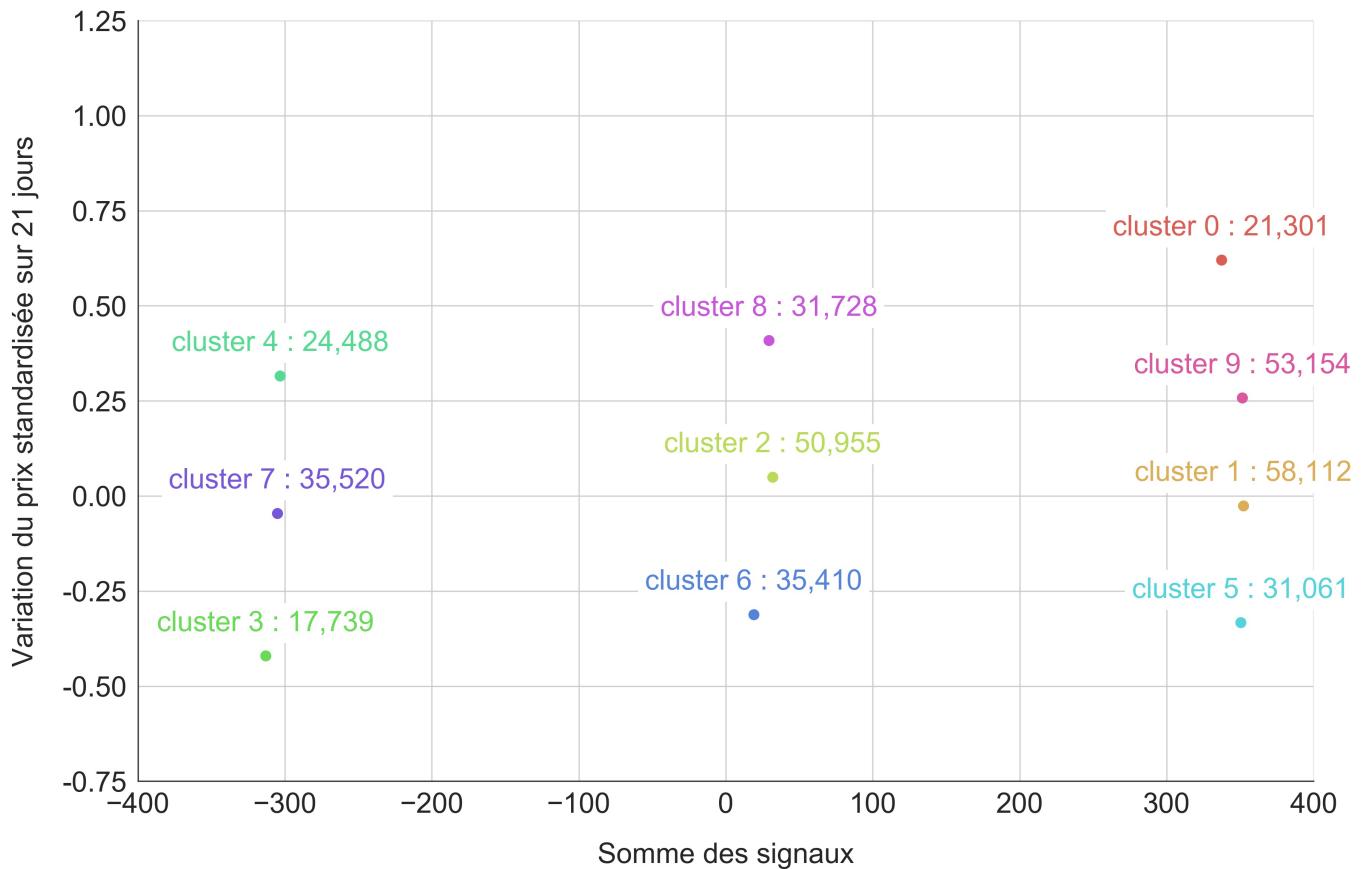
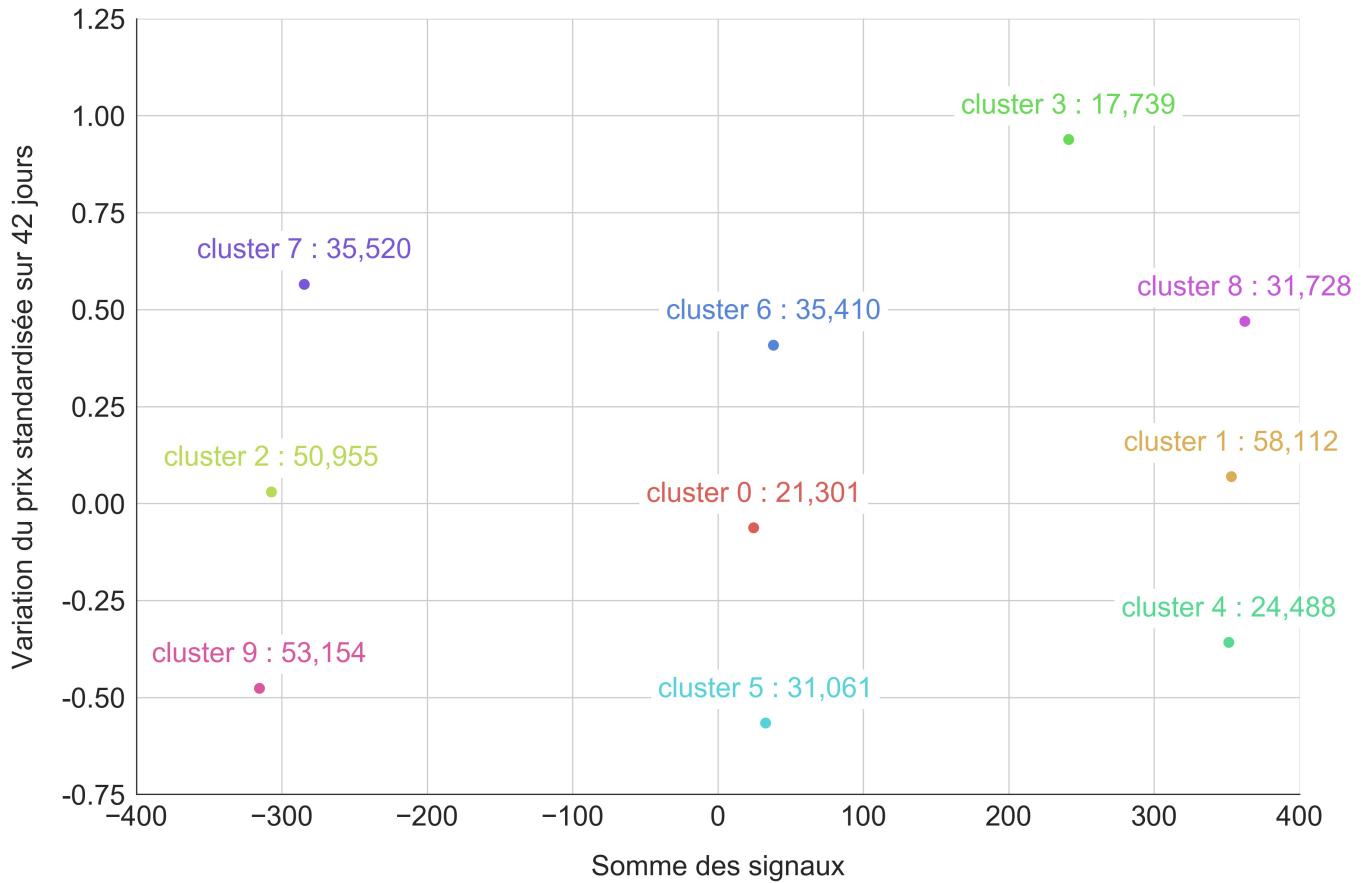


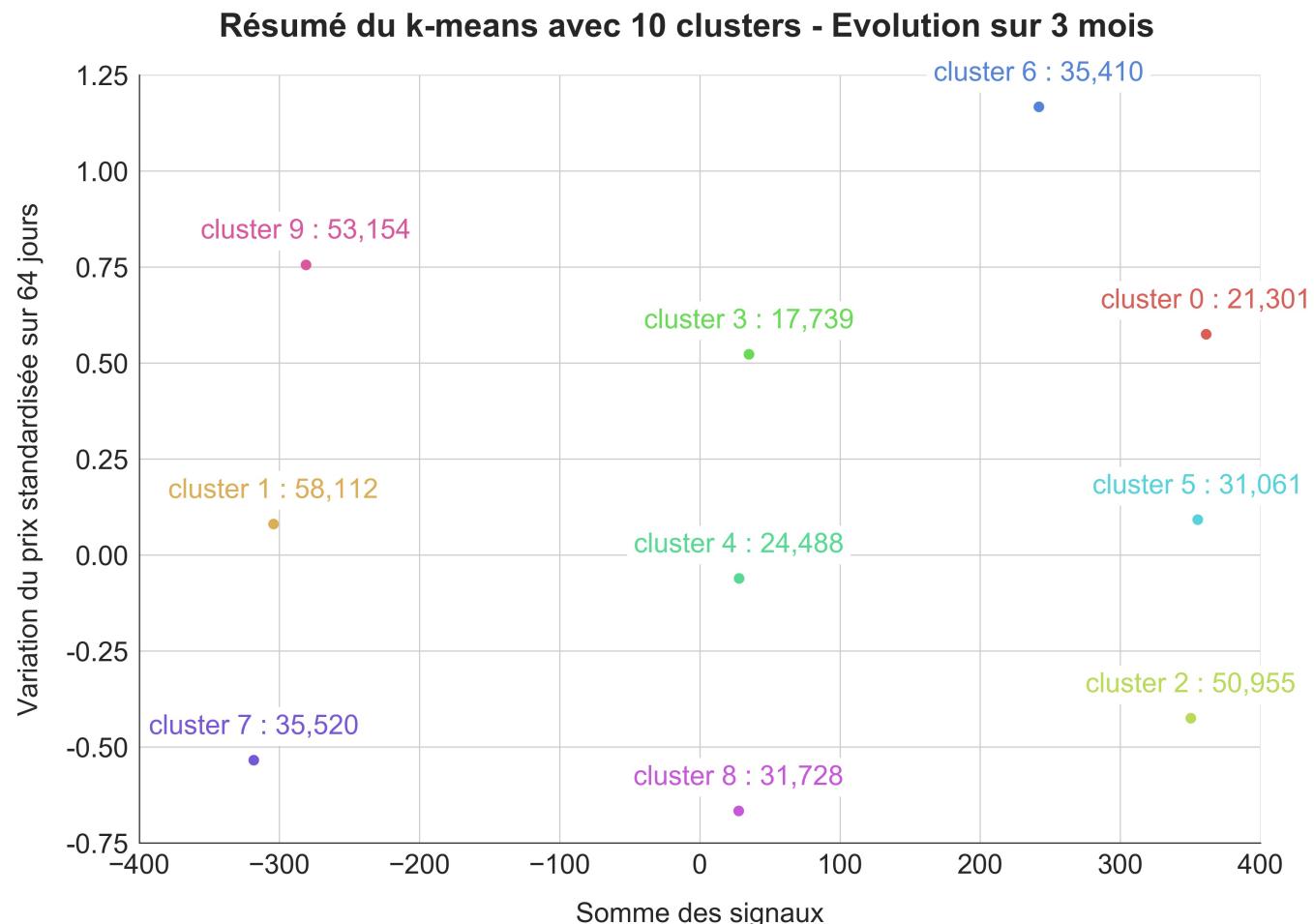
Figure 14b : Partitionnement obtenu avec l'évolution des prix à 2 mois

Résumé du k-means avec 10 clusters - Evolution sur 2 mois



sommes de signaux significativement positives et des variations futures positives très élevées. Ce dixième cluster regroupe donc les positions les plus profitables. Il présente un intérêt certain si nous réussissons à l'identifier et à l'isoler, par exemple à l'aide d'une régression logistique.

Figure 14c : Partitionnement obtenu avec l'évolution des prix à 3 mois



Nous calculons à partir de ces clusters une estimation du pourcentage des positions gagnantes, qui ne prend pas en compte l'adaptation des positions par les systèmes lorsque les conditions de marché changent.

Dans chacun des cas, 3 clusters représentent les positions gagnantes : les deux clusters avec des sommes de signaux significativement positives et des variations futures de prix significativement positives, et le cluster avec une somme des signaux significativement négative et une variation future du prix significativement négative. Les clusters avec une somme des signaux proche de zéro ne sont pas pris en compte dans le total des positions.

Avec un horizon de variation des prix d'un mois, nous obtenons une approximation de 38,2% (92 194 / 241 375) de positions gagnantes.

Avec un horizon de deux mois, l'estimation est similaire avec 37,8% (102 621 / 271 696) de positions gagnantes.

Enfin, à un horizon de trois mois, nous trouvons un résultat un peu dégradé avec 32,3% (92 231 / 285 513) de positions gagnantes. Le plus long horizon temporel se traduit par un plus grand nombre de marchés qui se retournent dans l'intervalle de temps. Mais rappelons-nous que cette estimation ignore l'adaptation des signaux aux nouvelles configurations de marché qui interviennent pendant dans ces trois mois.

Dans les trois cas, les résultats obtenus sont conformes, voire meilleurs, au ratio d'un tiers de positions gagnantes propre aux systèmes suiveurs de tendance.

5.3. Modélisations envisageables

L'ensemble des analyses précédentes nous amènent à considérer plusieurs modélisations possibles :

- une régression linéaire afin d'estimer la variation standardisée future du prix à partir de la somme globale des signaux, de la somme des signaux par indicateur et/ou de la variation de ces sommes. Nous ne nous attendons pas à de bons résultats à la vue de la dispersion observée lors de la projection des clusters sur le repère constitué de la somme des signaux et de la variation future des prix ;

- une régression logistique afin de déterminer si un marché est en tendance à partir de la somme globale des signaux, de la somme des signaux par indicateur et/ou de la variation de ces sommes. Cette tâche est assez similaire à la précédente, et nous ne nous attendons pas à de bons résultats, pour la même raison.

- une régression logistique, basée sur les mêmes variables explicatives, afin de déterminer si une configuration de marché appartient au dixième cluster, qui regroupe les situations générant le plus de profit. Cette régression ne sera modélisée que sur les variations de prix à 2 mois et à 3 mois, où ce cluster se distingue plus significativement des autres, notamment sur la somme des signaux. Nous cherchons donc à l'isoler. Ici aussi, la tâche s'annonce ardue sur la base d'une régression logistique.

6. Modélisations

6.1. Deux échantillons de données aux caractéristiques différentes

Etant donné le grand nombre de données dont nous disposons, nous générerons deux types d'échantillons pour étudier nos modélisations.

Echantillon aléatoire

Nous constituons le premier échantillon par tirage au sort aléatoire de 20% des données, en respectant la proportion des différents contrats (basé sur l'ID du contrat) et des différents clusters obtenus lors du partitionnement. Nous générerons un échantillon par horizon temporel des variations futures de prix.

Nous utilisons les données de ces échantillons pour effectuer nos essais de modélisation. Leur analyse montre une distribution des signaux et des sommes de signaux fidèle à ce que nous avons observé sur l'ensemble des données.

Echantillon basé sur la connaissance métier

Le deuxième échantillon est le même pour tous les horizons temporels étudiés. Il est basé sur la connaissance métier. Nous sélectionnons la période continue de 4 années qui va de 2014 à 2017, qui possède des particularités intéressantes pour notre analyse :

- les années 2014 et 2015 sont très bonnes : 2014 est la troisième meilleure année historique depuis 1995 (et la meilleure sur les quinze dernières années) et le rendement de 2015 se situe 20% au-dessus du rendement annuel moyen ;

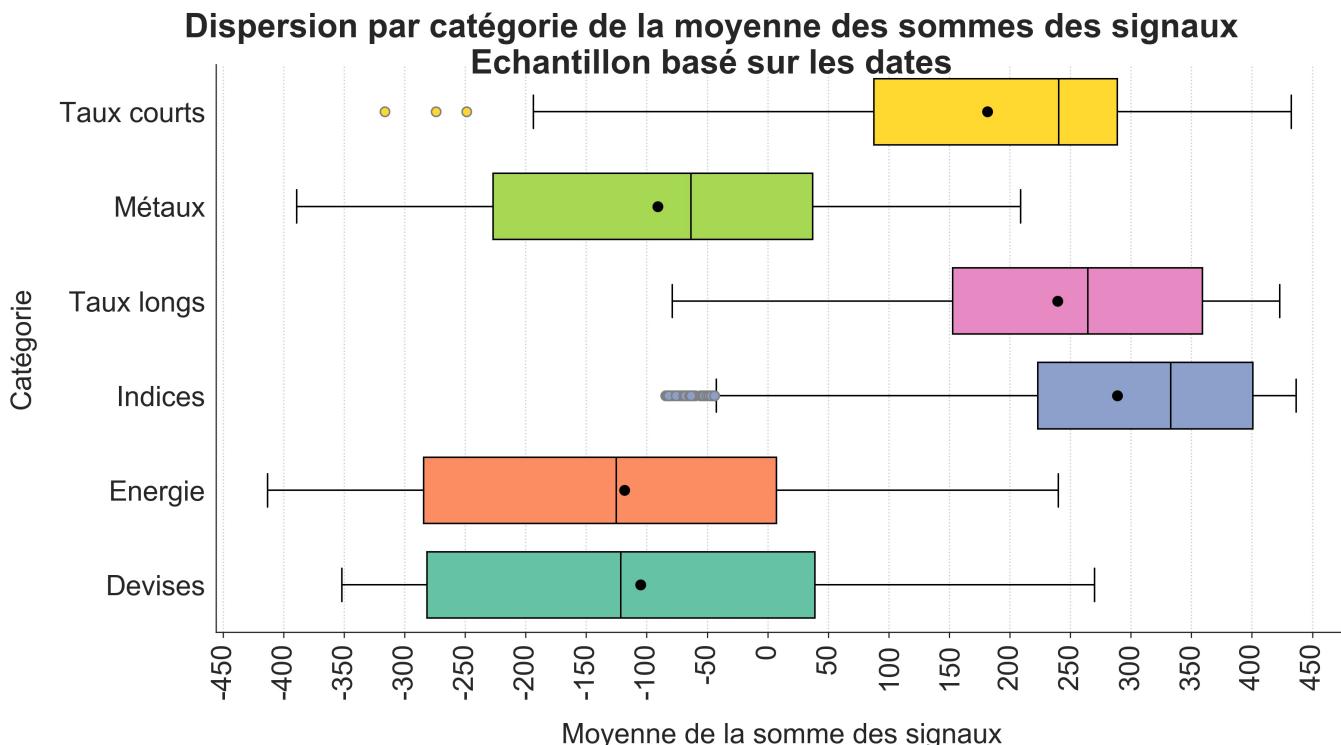
- les années 2016 et 2017 sont au contraire mauvaises, le rendement de chacune de ces années étant inférieur à la moitié du rendement annuel moyen.

Cet échantillon présente ainsi deux sous-périodes distinctes : la première favorable à nos systèmes, la seconde défavorable.

Nous utilisons les données de ce deuxième échantillon pour vérifier les modélisations qui paraissent intéressantes sur la base du premier échantillon. Elles constituent une sorte de stress-test.

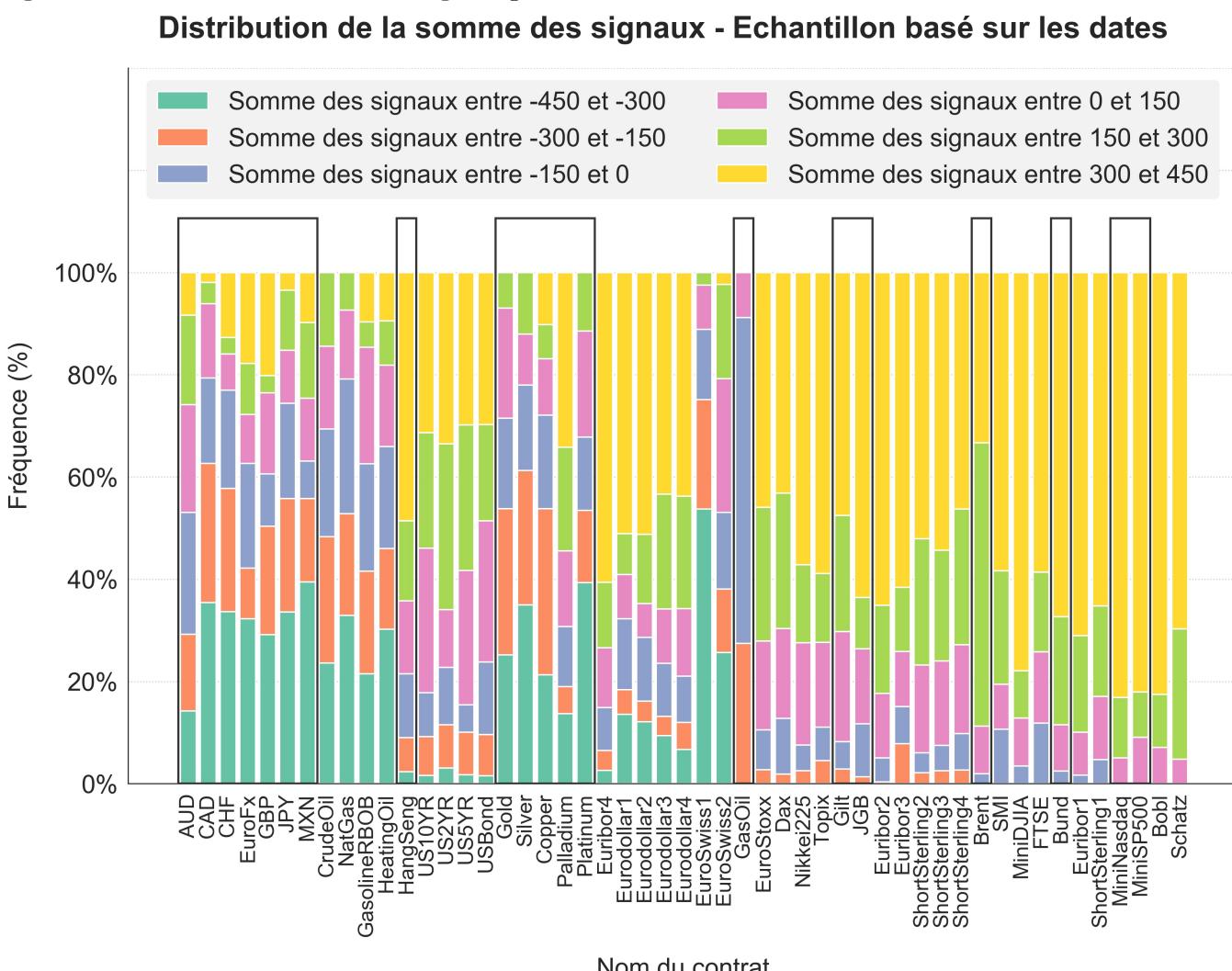
L'analyse des sommes de signaux par catégorie et par contrat confirme que les configurations de marché de cet échantillon diffèrent de la moyenne historique. Beaucoup de contrats n'ont pas de sommes de signaux très négatives (inférieures à -300) ou très positives (supérieures à +300).

Figure 15 : Dispersion de la moyenne des sommes de signaux par catégorie, dans l'échantillon basé sur les dates



La dispersion des sommes de signaux est fortement orientée vers les valeurs négatives pour les catégories des devises, métaux et énergies, alors qu'elle est fortement biaisée vers les valeurs positives pour les catégories des taux courts, des taux longs et des indices boursiers.

Figure 16 : Distribution des sommes de signaux par contrat, dans l'échantillon basé sur les dates



6.2. Conserver notre formule actuelle de calcul des positions

Les modélisations sont peu concluantes. Aucune des modélisations envisagées ne donne de résultat intéressant. Le R² ne dépasse pas 10% dans la plupart des cas, et ce pour les trois modélisations envisagées.

Les R² les plus élevés approchent 30%, pour la régression logistique modélisant l'existence d'une tendance sur le mois à venir, en fonction de la somme des signaux par indicateur. Les variables exogènes de ce modèle sont la somme des signaux sur l'indicateur FOT et l'interaction entre les sommes de signaux sur les indicateurs BRO et MOM.

Le même modèle pour détecter l'existence d'une tendance à 2 mois ou à 3 mois donne de très mauvais résultats, avec des R² à nouveau inférieurs à 10%.

A la vue des conclusions de nos analyses sur les sommes de signaux, et surtout de leur mise en perspective avec l'évolution future des prix, nous nous attendions à ces mauvais résultats des tentatives de modélisation.

Notre système de détermination de la taille des positions sur les marchés ne peut tirer aucune amélioration robuste des résultats obtenus lors de ces analyses. De nouveaux essais de modélisation pourraient être menés ultérieurement, en se basant sur des systèmes plus complexes de machine-learning.

Glossaire des termes financiers

Analyse technique : étude de l'historique des prix d'un actif financier pour identifier des configurations de marché et pour tenter de prédire leurs évolutions futures. Il existe deux types d'analyse technique : l'analyse technique graphique et l'analyse technique mathématique et statistique.

Back-testing (ou back-test) : ensemble de procédures informatiques consistant à tester les résultats d'une stratégie à partir de données historiques.

Break-out : le break-out se produit lorsque le prix passe au-dessus d'un niveau de résistance, ce qui déclenche un ordre d'achat, ou en-dessous d'un niveau de support, ce qui déclenche un ordre de vente. La détermination des niveaux de résistance et de support est subjective et peut s'appuyer sur différentes approches.

Contrat à terme : produit financier qui permet de prendre un engagement de vendre ou d'acheter un sous-jacent à un prix et à une date déterminés à l'avance. Les contrats à terme sont des produits standardisés et cotés : ils sont négociables sur un marché réglementé.

Force of trend : indicateur propriétaire qui mesure la force de la tendance. La variation des prix entre deux dates est pondérée par la volatilité entre ces deux mêmes dates. Un signal d'achat est déclenché lorsque l'indicateur passe au-dessus d'un seuil positif, un signal de vente lorsqu'il passe sous un seuil négatif.

Indicateur technique : série d'opérations mathématiques appliquées aux cours ou aux volumes d'une valeur et pouvant être représentée graphiquement.

Momentum : le momentum mesure la rapidité d'évolution des prix sur une période donnée. Dans sa version la plus basique, il s'agit de la différence des prix entre deux dates. En suivi de tendance, un momentum positif est un indicateur d'achat, un momentum négatif un indicateur de vente.

Money management : gestion du capital investi sur les marchés financiers en vue d'accroître son espérance de rentabilité et de minimiser le risque de perte.

Roll : action de fermer une position sur l'échéance courante d'un contrat à terme pour ouvrir la même position sur l'échéance suivante

Sous-jacent : actif sur lequel un produit dérivé porte.

Tendance : mouvement clair et soutenu, à la hausse ou à la baisse, d'un marché. Les actifs évoluent en général en formant des minima (creux) et maxima (sommets) locaux. Une tendance haussière se met en place lorsque les creux et sommets sont de plus en plus élevés. À l'inverse, une tendance baissière est observée si les creux et sommets sont de plus en plus faibles.

Volatilité : écart-type de la variation des prix en pourcentage. Elle mesure l'amplitude des variations d'un produit financier.

Bibliographie

- BARON, François. *Le Chartisme*. 2ème éd. Editions d'Organisation, 2008, 496 p.
- CLENOW, Andreas. *Following the Trend : Diversified Managed Futures Trading*. Editions Wiley Trading, 2013, 300 p.
- COVEL, Michael. *Trend Following : How to Make a Fortune in Bull, Bear, and Black Swan Markets*. 5ème éd. Editions Wiley Trading, 2017, 688 p.
- FAITH, Curtis. *Way of the Turtle : The Secret Methods that Turned Ordinary People into Legendary Traders*. Editions McGraw-Hill Professional, 2007, 320 p.
- GRIMES, Adam. *The Art and Science of Technical Analysis : Market Structure, Price Action, and Trading Strategies*. Editions Wiley Trading, 2012, 480 p.
- MURPHY, John. *L'Analyse Technique des Marchés Financiers*. Editions Valor, 2004, 565 p.