

Amélioration des indicateurs techniques pour l'analyse du marché financier

Hunor Albert-Lorincz***, Jean-François Boulicaut**

* CIC Recherche Quantitative Structurés de Taux
6, avenue de Provence, 75009 Paris, France
alberthu@sdm.cic.fr
<http://liris.cnrs.fr/~halbert/>

** INSA Lyon, LIRIS CNRS UMR 5205
Batiment Blaise Pascal
F-69621 Villeurbanne cedex, France
Jean-Francois.Boulicaut@insa-lyon.fr
<http://liris.cnrs.fr/~jboulica/>

Résumé. La technique des motifs fréquents a été utilisée pour améliorer le pouvoir prédictif des stratégies quantitatives. Innovant dans le contexte des marchés financiers, notre méthode associe une signature aux configurations de marché fréquentes. Un système de « trading » automatique sélectionne les meilleures signatures par une procédure de « back testing » itérative et les utilise en combinaison avec l'indicateur technique pour améliorer sa performance. L'application des motifs fréquents à cette problématique des indicateurs techniques est une contribution originale. Au sens du test t de Student, notre méthode améliore nettement les approches sans signatures. La technique a été testé sur des données journalières type taux d'intérêt et actions. Notre analyse des indicateurs (Williams%R, BN et croisement des moments) a montré que qu'une approche par signatures est particulièrement bien adaptée aux stratégies à mémoire courte.

1 Introduction

Les indicateurs techniques sont des fonctions des données de marché, historiques et actuelles, qui produisent un signal d'achat ou de vente. Ce sont les « briques » qui permettent de construire des stratégies de « trading » en réaction aux indicateurs et en fonction de la composition du portefeuille de l'investisseur.

Certaines techniques de fouille de données permettent d'attribuer une signature aux configurations de marché précédant le déclenchement d'un indicateur technique. La comparaison des performances de l'indicateur seul et du même indicateur précédé d'une signature permet de choisir les signatures qui améliorent les performances de l'indicateur. La stratégie de « trading » peut donc exploiter un filtrage pour se restreindre aux transactions qui se

déclenchent après une signature et qui ont produit des gains dans le passé. Les signatures sont testées sur une période d'apprentissage afin d'exclure celles pour lesquelles l'indicateur technique a sous performé. Les meilleures signatures sont gardées pour la période suivante, dite de validation ou de « trading » pendant laquelle l'indicateur technique est pris en compte uniquement s'il est précédé par une « bonne » signature. L'analyse des signatures montre une remarquable stabilité temporelle, et l'indicateur filtré par les « bonnes » signatures de la période d'apprentissage sur performe régulièrement l'indicateur seul pendant la période de validation.

Nous utilisons une technique de fouille de données classique (i.e., la recherche de motifs fréquents) pour identifier les signatures qui caractérisent les quelques jours précédant le déclenchement de l'indicateur technique. Nous obtenons ainsi des règles d'analyse technique spécifiques au sous-jacent, basées sur plusieurs jours et adaptées à la période considérée.

Une première contribution de ce travail consiste à caractériser les configurations de marché afin que l'extraction de motifs fréquents puisse s'appliquer sur ce type de données. A notre connaissance, l'idée d'améliorer un indicateur technique par la méthode des signatures est également originale. Cet article ne propose pas un nième indicateur technique mais plutôt une méthodologie générale pour améliorer des stratégies quantitatives existantes.

L'amélioration des indicateurs techniques est un objectif bien différent de la prédiction des cours boursiers et a reçu moins d'attention. F. Allen et R. Karjalainen ont travaillé sur la combinaison et le calibrage des seuils de déclenchement des indicateurs par la programmation génétique (Allen et Karjalainen 1999). Les travaux plus récents de Becker rapportent que les règles de « trading » générées par la programmation génétique peuvent battre la stratégie « acheter et garder » en tenant compte des frais de transaction (Becker 2003). Ces travaux concernent le paramétrage d'un ensemble de règles de « trading » déjà connues et ne cherchent pas à caractériser une configuration de marché en addition d'un indicateur technique. Notre approche est donc complémentaire à ce type de travaux.

La Section 2 présente la méthodologie appliquée et le système de « trading » réalisé est rapidement présenté en Section 3. La Section 4 présente les résultats expérimentaux obtenus sur des données réelles. La Section 5 est une brève conclusion.

2 Méthodologie

Pour mettre en place la stratégie de « trading » basée sur des signatures il faut d'abord définir le contenu d'une configuration de marché (scénario) et le transformer dans le format attendu par l'algorithme d'extraction des motifs utilisé (discretisation et agrégation)..

2.1 La description des scénarii

Un scénario regroupe l'ensemble des données de marché historiques sur « N » jours qui constitue le domaine de l'indicateur. Les indicateurs techniques sont donc des fonctions des scénarii f qui associent à un scénario une valeur dans l'ensemble {Acheter, Vendre, Rien}. Par exemple, la fonction « f » pourrait être de type Acheter, si la moyenne mobile sur 2 semaines est supérieure à la moyenne mobile sur 30 jours, sinon Rien. Il s'agit évidemment d'un exemple hypothétique.

Comme dans la plupart des travaux dans ce domaine, chaque journée composant un scénario sera caractérisée par la valeur de l'indicateur technique, le cours d'ouverture (O) et de

clôture (C) ainsi que les maxima (H) et minima (L) journaliers. Le nombre de jours « N » composant un scénario est déterminé empiriquement, et s'étend généralement sur le domaine de l'indicateur technique.

De nombreuses techniques de fouille de données travaillent sur des données discrètes et supposent un pré-traitement. Ainsi, nous voulons transformer les données de marché composant les scénarii en une représentation non numérique. L'association de chaque valeur à sa q-quantile¹ apparaît comme un choix de discrétisation simple facilement mis en œuvre. Les deux scénarii quantilisés associés aux journées J1 et J2 seront graphiquement représentés comme suit :

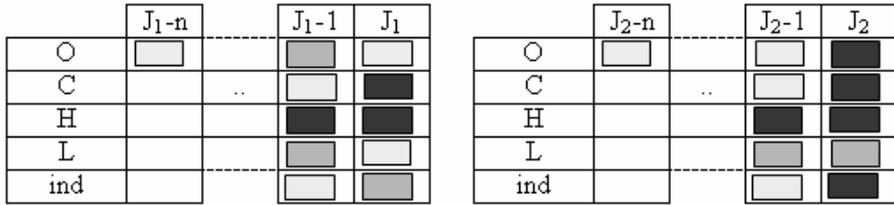


FIG. 1 – Scénarii 3-quantilisés. Les carrés foncés indiquent des valeurs du tiers supérieur, le semi foncé représente une valeur moyenne, et le claire une donnée du tiers inférieur.

La représentation de la Figure 1 caractérise les niveaux absolus des quantiles. Il est quelquefois intéressant de « normaliser » cette représentation, par exemple en soustrayant le plus petit quantile des autres quantiles. Cette alternative généralise la représentation précédente et permet l'expression en valeurs relatives. L'intuition derrière cette nouvelle représentation est de regrouper les configurations similaires à deux niveaux de marché différents (e.g., clôture supérieure d'un quantile à l'ouverture dans un marché bas et un marché haut) et d'en extraire un signal plus fort.

La méthodologie exposée dans cet article associera une signature aux scénarii fréquents en passant par l'extraction de motifs fréquents. L'extraction de motifs fréquents est l'une des tâches de fouille de données qui a été la plus étudiée depuis une dizaine d'années et nous supposons que le lecteur est familiarisée avec cette technique. Une annexe rappelle cependant le principe algorithmique qui a été utilisé.

Un motif f-fréquent (noté simplement fréquent dans la suite) est un sous-ensemble commun à une collection d'ensembles (scénarii dans le cas présent, collection de transactions dans le cadre classique des données de ventes) qui apparaît en au moins f ensembles. L'extraction des motifs fréquents intervient sur des ensembles de valeurs et il faut donc représenter chaque scénario sous une forme transactionnelle. On préfère souvent une représentation suffisamment riche pour pouvoir reconstituer le scénario de départ à partir d'un ensemble non ordonné, i.e., une bijection. La bijection a l'avantage de pouvoir retrouver la configuration de marché à partir de sa signature, et permet à un expert humain de comprendre a posteriori les règles générées par la technique de fouille de données.

En introduisant un système de coordonnées (x, y) comme sur la Figure 2, il est possible de définir une bijection pour chaque valeur composant un scénario.

¹ La q-quantilisation d'une suite de valeurs ordonnée signifie le partitionnement des valeurs par l'addition de q-1 points médians. Chaque valeur est ensuite remplacée par le numéro de sa partition.

Amélioration des indicateurs techniques

Un exemple de transformation est donné par la formule :

$$\text{Code_ens} = 5 * x * q + y * q + \text{value}(x, y)$$

où " q " est le nombre de quantiles et 5 le nombre de données par jour (indicateur, ouverture, clôture, maxima, minima).

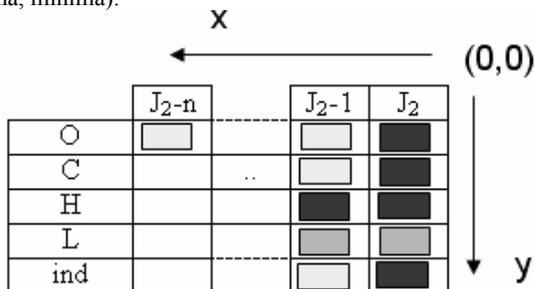


FIG. 2 – Transformation de la configuration de marché en une représentation ensembliste.

Pour les deux dernières colonnes des scénarii de la Figure 1, on obtient les ensembles suivants : ($q=3$). Le calcul de la seconde valeur de la colonne J_{2-1} a été explicité.

$$J1 = \{2 ; 6 ; 9 ; 11 ; 13 ; 16 ; 5 * 3 * 1 + 3 * 1 + 2 = 20 ; 24 ; 25 ; 29\}$$

$$J2 = \{3 ; 6 ; 9 ; 10 ; 15 ; 17 ; 20 ; 24 ; 25 ; 29\}$$

Après l'étape de discrétisation et de codification, chaque jour est représenté par un ensemble de $5 \times N$ symboles obtenus par la formule précédente. C'est bien un ensemble et non pas une séquence, l'ordre des symboles n'a plus d'importance.

2.2 Les signatures

On souhaite caractériser les scénarii fréquents qui peuvent être à la base d'une stratégie quantitative. Dans le contexte d'amélioration des indicateurs techniques il n'est pas intéressant de signer les événements rares : il est souhaitable que la fréquence du motif soit comparable à la fréquence de l'indicateur et les motifs fréquents apparaissent donc opportuns.

La signature d'un scénario est un sous-ensemble (sous-motif) fréquent (dans les ensembles des scénarii) des motifs qui le décrivent. Les scénarii peu fréquents n'ont pas forcément de signature et un scénario peut avoir plusieurs signatures.

Nous donnons un exemple de signature (i.e., un sous-motif commun) des deux scénarii de la Figure 1. Le sous-ensemble commun de $J1$ et de $J2$ est $\{6 ; 9 ; 20 ; 24 ; 25 ; 29\}$ ce qui, grâce à la bijection, correspond au motif visuel donné dans la Figure 3.

Notons que les sous-ensembles des signatures sont également des signatures. Dans le contexte de l'amélioration des indicateurs, la période étudiée comprend quelques centaines d'ensembles et les seuils de fréquences utilisés sont de l'ordre de quelques dizaines.

J-1	J

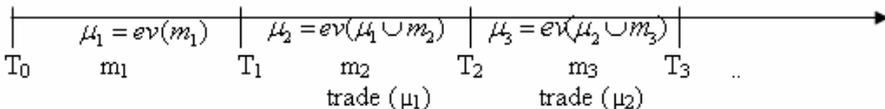
FIG. 3 – Signature des scénarii de la Figure 1.

2.3 Algorithme d'extraction

L'extraction des sous-ensembles d'une collection de scénarii est une tâche difficile car les approches naïves se heurtent à une explosion combinatoire. De très nombreux algorithmes ont été proposés pour réaliser des extractions complètes de tous les ensembles fréquents dans des contextes réalistes. Pour notre système, nous avons utilisé notre propre implémentation des « FP-Tree » (Han et al. 2004) (voir Annexe) étendue à la gestion des motifs fermés (Pasquier et al. 1999). La structure « FP-tree » a fait ses preuves lors des extractions de très grande échelle. Elle est adaptée aux tâches courantes en finance quantitative qui ne concernent généralement que quelques milliers d'ensembles. L'extraction se fait quasiment en temps réel, ce qui rend la méthode proposée éligible pour le « trading » haute fréquence.

3 Le système de « trading »

Le comportement du sous-jacent est assujéti à des phénomènes de mode et à des régimes de « trading » qui varient dans le temps. Aussi, est-il nécessaire de suivre la validité des signatures dans le temps, d'enrichir la base des signatures par les nouveaux motifs qui apparaissent dans les cours et d'écarter (invalider) les motifs qui commencent à sous performer. Nous décrivons maintenant ce mécanisme d'apprentissage progressif. Les dates T_0 à T_n déterminent $n-1$ périodes. La première est utilisée pour l'initialisation du système (apprentissage des « bonnes » signatures) qui seront utilisées pour le « trading » pendant la période suivante. A partir de l'apprentissage initial, chaque période sert à traiter avec les bonnes signatures identifiées jusqu'au début de la période et alimente les signatures de la période suivante.



La première période fournit les motifs m_1 . Cet ensemble est testé contre la performance de l'indicateur seul pendant cette même période pour n'en garder que les motifs μ_1 qui améliorent les performances de l'indicateur. Le contenu de μ_1 générera des transactions de la deuxième période, et sera ajouté aux motifs m_2 de cette période pour calculer l'ensemble μ_2 , contenant les signatures de bonne qualité identifiées jusqu'à T_2 (la fin de la 2^e période).

Amélioration des indicateurs techniques

Le système de « trading » bâti autour de ce re-balancement est illustré sur la Figure 4. Suivons d'abord les flèches qui émergent de la période p-1 jusqu'au rectangle « filtrage ». Cette partie correspond à l'apprentissage du système. L'étape « extraction » fournit les motifs fréquents de la période p-1. La mesure de performance évalue les gains de chaque motif fréquent identifié. Pour ce faire, la stratégie de « trading » à optimiser est effectuée sur la période p-1 et l'on calcule la performance de l'indicateur seul. Ensuite, pour chaque motif fréquent extrait, la stratégie est à nouveau appliquée. Cette fois, l'indicateur n'est retourné vers la stratégie que si son scénario d'occurrence contient le motif en cours d'étude. Si la performance de la stratégie précédée par un motif est supérieure à la performance de l'indicateur seul, le motif est ajouté à l'ensemble μ_{p-1} , i.e., la collection des bonnes signatures, et utilisé par le « trading » de la période suivante. Les indicateurs de la période p sont pris en compte uniquement si leur scénario d'occurrence contient une « bonne » signature de la période précédente. Dans ce cas, le filtrage génère un signal qui est envoyé à la stratégie. Après le trading, la période p devient une période d'apprentissage pour la période p+1.

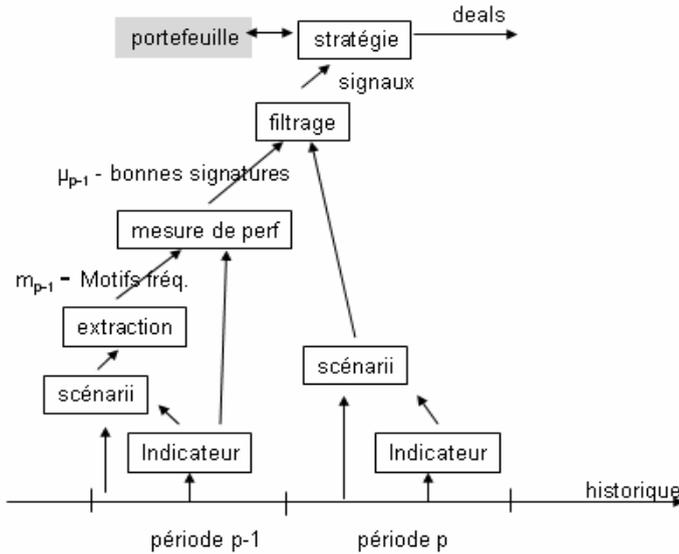


FIG. 4 – *Système de trading.*

Un système de « trading » sur trois périodes (apprentissage, sélection, trading) est difficile de mettre en œuvre sans une diminution drastique de la fréquence des transactions. En fait, les signatures sont attachées à un régime de marché et leur apport disparaît si elles ne sont pas appliquées immédiatement en « trading » (impossibilité d'avoir une période de sélection). Le raccourcissement de la période d'apprentissage est délicat, car la diminution de points pour l'étape de fouille de données conduit à un sur-apprentissage, i.e., un phénomène de sur-spécialisation qui produit des bonnes performances pendant la période d'apprentissage mais qui se comporte de manière médiocre ultérieurement.

4 Résultats expérimentaux

Pour démontrer le caractère générique de la méthode, elle a été appliquée à des données financières de type « action » et « fixed income », qui ont des comportements (volatilité, retour à la moyenne, etc.) très différents. Les indicateurs utilisés sont simples et bien connus dans le monde financier : Williams %R (Achelis 2000), croisement de moyennes mobiles, et un indicateur confidentiel, appelé boîte noire (BN). La période étudiée s'étend sur cinq ans. La performance a été mesurée comme le gain réalisé en « N » jours.

Le Williams%R permet de déterminer quand un actif est sur vendu ou sur acheté sur une période (d'habitude 2 semaines).

$$W\%R = -100 * \frac{\text{plus haut des plus hauts} - \text{cloture}}{\text{plus haut des plus hauts} - \text{plus bas des plus bas}}$$

Le portefeuille est composé d'un dépôt « cash » et de quelques titres. Dans la gestion, la valeur du portefeuille (« wealth process ») ne tient pas compte de la rémunération du dépôt « cash » (ni de l'effet des dividendes pour les données de type « indice » ou « action »), car l'étude vise à montrer l'efficacité du filtrage toutes choses égales par ailleurs. Les tests ont été faits sur des stratégies de type « long », donc sans vente à découvert.

Données de type action. Le tableau suivant compare les performances de W%R seul au W%R signé pour une action française. Les signatures réduisent le nombre de mauvaises transactions et sur performant de manière répétitive le W%R seul.

Année	W%R signé		W%R seul	
	Perf	nbDeals	Perf	nbDeals
2	101.74	88	99.81	144
3	107.06	142	107.14	174
4	100	0	102.06	190
5	98.3	80	97.25	96
6	104.57	134	104.05	150
7	98.55	108	99.44	118
8	112.58	166	113.26	184
9	146.95	180	146.92	196
10	88.62	70	88.68	130
11	96.84	104	92.41	112
12	80.45	72	78.68	80
13	89.37	90	92.97	100
14	99.16	130	96.67	144
15	95.38	38	95.58	48

TAB. 1 – Le W%R signé sur performe le W%R seul sur 15 ans. L'ordre de grandeur des transactions restant est comparable au nombre de transactions au départ « sans signatures ».

Amélioration des indicateurs techniques

On voit bien que l'indicateur signé ne surpasse pas l'indicateur seul sur toutes les périodes. Elle a néanmoins un meilleur comportement en moyenne. La somme des retours journaliers (dS/S) pour le Williams %R signé termine à 12.59 en retour cumulé alors que l'indicateur seul ne produisait que 7.29 sur 10 ans, soit un gain de 5.3% avec la méthode des signatures.

L'analyse des actions composant l'indice CAC40 depuis le 1^{er} janvier 2000 a permis de valider l'apport de la technique des signatures sur un grand nombre de périodes et d'actions. La méthode proposée a apporté 6% de gains par rapport au seul W%R sur 5 ans. Compte tenu du nombre de périodes comparées, le test t de Student a permis de conclure que les performances de la technique signée sont supérieures à l'indicateur seul de manière significative. Pour une stratégie quantitative dans la finance, ce gain est énorme ! La moyenne des performances de l'indicateur W%R signé est 96.93, comparé au W%R qui fait 90.63. La $P(T \leq t)$ unilatéral est 10,6%. Les tests peuvent être obtenus auprès des auteurs.

Nous avons également testé l'apport des signatures sur un indicateur propre à l'entreprise CIC et qui est appelé boîte noire (BN). L'amélioration est encore plus significative (seuil de 1%). La moyenne de la BN signé est 104.22, alors que la BN seul n'a fait que 98.13%. Cette fois la $P(T \leq t)$ unilatéral valait 0.011, ce qui est un très beau résultat.

Données de type taux d'intérêt. Compte tenu de l'efficacité accrue des marchés de capitaux, le W%R ne dégage pas de performance sur la période étudiée. Cependant, les signatures permettent de baisser les pertes. Comme les graphiques suivants le montreront, par périodes le W%R signé produit des gains alors que le W%R s'inscrit dans une tendance baissière. La Figure 5 illustre le « wealth process » de la stratégie, i.e., la valeur du portefeuille produit par la stratégie tout au cours de sa vie. Le W%R signé se distingue pendant les périodes où le W%R perd, voir il réalise des gains par période (de 92.93 à 96.13, soit 3.2% de mieux).

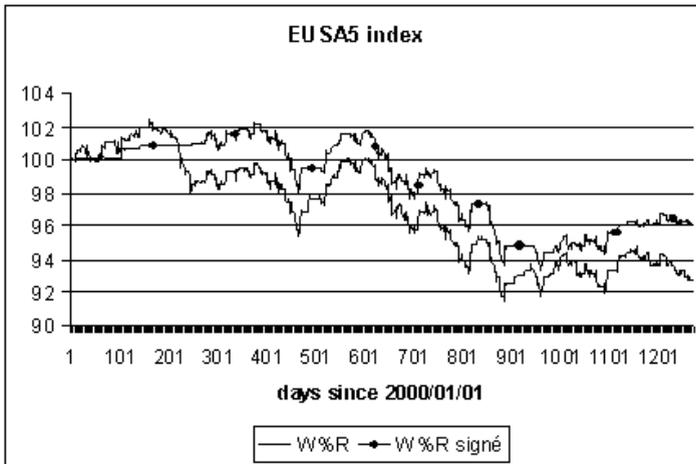


FIG. 5 – Wealth process du taux swap 5 ans.

Le Tableau 2 illustre les arguments développés pour la Figure 5. Les chiffres ont été obtenus avec remise des capitaux, i.e., le portefeuille est vidé en début de chaque période et le dépôt « cash » est réinitialisé à 100.

W%R signé		W%R	
perf	nDeals	Perf	nbDeals
100.966	24	98.6248	356
98.3502	306	98.7312	334
98.8325	346	98.4007	382
96.8069	266	98.1424	310
101.185	268	98.8317	356

TAB. 2 – Performances de W%R signé et du W%R appliqué sur les taux swap 5 ans entre début 2000 et fin 2004. (L'année 1999 a été utilisée pour initialiser l'apprentissage).

4.1 Discussion

La stratégie de « trading » en réponse de l'indicateur a été paramétrée pour mesurer les performances à N jours après la transaction. (5 jours pour les statistiques précédentes). Les signatures caractérisent une configuration de marché ponctuelle et ont un pouvoir de prédiction d'environ une semaine. Il est normal que leur validité soit limitée dans le temps : l'ensemble qui décrit la configuration de marché n'inclut aucune information sur le comportement historique du cours. La "mémoire longue" et l'auto corrélation ne sont pas utilisées. Nous pensons qu'en enrichissant la description des scénarii par des éléments du passé (minimum et maximum sur un passé plus long, moyenne mobile), le pouvoir de prédiction peut être étendu de quelques jours, mais l'extraction des motifs fréquents devient plus difficile à cause d'un alphabet plus large. La durée de vie d'une signature est assez courte, elle dépasse rarement trois périodes (y compris celle d'apprentissage).

Dans le cas du W%R, les signatures surperforment la stratégie non signée en perdant un peu moins d'argent. On doit alors se demander si l'amélioration des performances n'est pas essentiellement liée au filtrage ? En effet, appliquer moins souvent une stratégie perdante améliore forcément les performances. Cependant, l'étude visuelle intuitive du comportement des signatures sur la Figure 5 nous laisse à penser que non ! En effet, dans le cas d'un filtrage aléatoire, l'écart entre les deux courbes devrait se constituer progressivement, alors que l'indicateur filtré se distingue aux moments où l'indicateur seul commence à perdre. De plus, le filtrage produit des résultats significativement meilleurs pour l'indicateur BN, et présente des gains alors que la stratégie de base est légèrement déficitaire.

Stratégies croisement des moyennes mobiles. Ce type d'indicateur est basé sur la mémoire longue non corrélée avec les signatures qui caractérisent la configuration présente. Aussi, ces indicateurs ne sont améliorés par la méthode des signatures.

Frais de transactions. L'effet du filtrage par les signatures est de limiter le nombre des transactions en gardant seulement celles qui ont le plus de chances d'aboutir. En conséquence, il réduit également les frais de transaction. L'étude présente sous-estime donc l'apport des signatures d'une part en négligeant les économies sur les frais de transactions et d'autre part en ignorant la rémunération du compte de dépôt.

5 Conclusion

Les tests effectués promettent d'aboutir à un système de « trading » automatique. Le « back testing » des indicateurs avec re-balancement s'est montré capable d'identifier les signatures caractéristiques à chaque période. Elles sont suffisamment stables dans le temps pour pouvoir améliorer les performances de l'indicateur technique durant la période suivante. Les techniques décrites apportent un moyen pour améliorer les stratégies quantitatives existantes et ne demandent pas la mise en œuvre de nouvelles stratégies. Dans les exemples traités, la technique des signatures a produit 3% et 6% de gains en plus par rapport à l'indicateur seul pour trois indicateurs différents. D'autres mesures sont en cours pour tester l'impact de la méthode sur davantage d'indicateurs techniques. L'application à des données « haute fréquence » est également à l'étude.

Remerciements. Les auteurs souhaitent remercier Swann Chmil (BNP Paribas Asset Management) pour sa contribution à cette recherche et les relecteurs anonymes pour leurs suggestions.

Références

- Achelis, S. B. (2000). *Technical Analysis from A to Z*. McGraw-Hill.
- Allen, F., R. Karjalainen (1999). Using genetic algorithms to find technical trading rules. *Journal of Financial Economics*, 51;245-271.
- Backer, L. A., M. Seshadri (2003). GP-evolved Technical Trading rules Can OutPerform Buy and Hold. Proceedings of the 6th International Conference on Computational Intelligence and Natural Computing, Cary, North Carolina, USA.
- Pasquier, N., Y. Bastide, R. Taouil, L. Lakhal (1999). Efficient Mining of Association Rules Using Closed Itemset Lattices. *Information Systems* 24(1):25-46.
- Han J., J. Pei, Y. Yin, R. Mao (2004). Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach, *Data Mining and Knowledge Discovery* 8(1):53-87.

Annexe

L'exemple suivant introduit l'extraction des motifs fréquents avec la structure FP-tree.

No.	Extrait de scénario	<i>ensembles réordonnés</i>
1	f a d ..	<i>f d a</i>
2	d b f c ..	<i>f d c b</i>
3	d a c b ..	<i>d c a b</i>
4	e c f ..	<i>f c</i>

TAB. A1 - Extrait de scénario pour illustrer l'extraction des motifs fréquents.

Pour simplifier la présentation, chaque scénario sera représenté par un extrait de l'ensemble qui le décrit. Les lettres tiennent pour des valeurs discrétisées précédemment. On cherchera à identifier tous les sous-ensembles 2-fréquents (présents dans au moins deux ensembles) des extraits du Tableau 1 avec l'algorithme FP-growth (Han et al. 2004).

A) La première étape consiste à parcourir tous les ensembles pour trouver la fréquence de chaque symbole. Les ensembles sont ensuite réorganisés selon un ordre donné par la fréquence des symboles : $(f:3) > (d:3) > (c:3) > (a:2) > (b:2)$, en laissant de côté les symboles qui ont une fréquence d'apparition inférieure au seuil requis (3^e colonne du Tableau A1). Le symbole "e" disparaît. On note que l'ordre et le FP-tree résultant ne sont pas uniques.

B) Les ensembles réordonnés sont ensuite insérés dans une structure arborescente. Chaque ensemble est représenté par un chemin dans cet arbre, de manière à ce que les sous-ensembles communs soient proches de la racine. Chaque nœud de l'arbre contient un symbole et un compteur des ensembles qui le contiennent sur cette branche de l'arbre. Concrètement, le premier scénario va créer la branche $\langle f:1, d:1, a:1 \rangle$. Le deuxième scénario partage un préfixe commun avec le premier. Il fait augmenter les compteurs sur la portion commune et créer un fils du nœud "d" pour "c". L'insertion des ensembles restant produit l'arbre à gauche de la Figure A1.

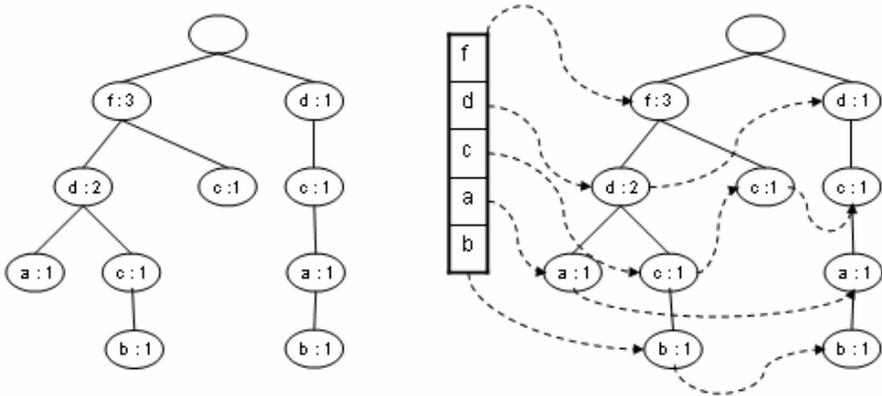


FIG. A1 - Structure FP-tree muni de pointeurs transversaux.

Pour des raisons exposées plus loin, cet arbre est enrichi de pointeurs transversaux (Figure A1 à droite) qui réunissent toutes les occurrences d'un même symbole. Le vecteur des pointeurs respecte l'ordre établi au point (A). Toute l'information présente dans les scénarii du Tableau A1 est contenue dans la structure FP-tree. Les chemins de l'arbre regroupent les symboles qui paraissent souvent ensemble, mais il faut utiliser les pointeurs transversaux pour trouver tous les motifs fréquents. En effet, la collection des branches (entre la racine et le symbole) le long des pointeurs d'un symbole "X" quelconque contient tous les ensembles où X est fréquent. Un traitement récursif permet de les extraire.

C) Si l'arbre contient plusieurs branches, l'extraction commence par chercher les motifs contenant le symbole le moins fréquent, "b" dans le cas présent. La sommation des compteurs le long de la liste chaînée fournira la fréquence de "b". En suivant cette liste, tous les chemins entre la racine de l'arbre et le nœud contenant le symbole courant "b" $\{fdc:1, dca:1\}$

sont regroupés. La fréquence de chaque branche est égale à la fréquence du symbole en cours sur la branche. Après l'extraction des branches, la procédure recommence à l'étape A avec les ensembles correspondant aux branches pour trouver tous les motifs fréquents contenant le symbole "b". Comme ni "f", ni "a" n'apparaissent suffisamment souvent avec "b", l'arbre contiendra un seul chemin (dc:2) qui donnera les motifs fréquents bcd:2, bc:2 et bd:2. Quand l'arbre contient une seule branche, e.g., <d:2, c:2>, les motifs fréquents sont donnés par tous les sous-ensembles des symboles le long de la branche, e.g., d:2, c:2, et dc:2.

Après le traitement de "b", l'extraction continue avec le symbole immédiatement plus fréquent que "b", "a" dans le cas présent. Cette étape livrera les motifs contenant "a" mais ne contenant pas "b". Les chemins fd:1 et dc:1 permettent de retrouver les motifs fréquents ad:2 et a:2. Les trois chemins de "c" {fd:1, f:1 et d:1} contribuent à la construction d'un arbre à deux branches. Celui-ci est parcouru avec le même algorithme et contient les motifs fréquents d:2 et f:2. S'agissant de l'arbre construit à partir des chemins de "c", on obtient les motifs c:3, cd:2 et cf:2.

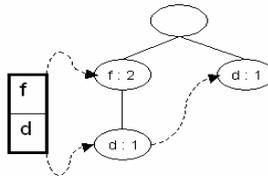


FIG. A2 - Sous FP-Tree correspondant aux branches contenant "c".

Le symbole "d" apparaît trois fois dans les chemins fd:2 et d:1. L'arbre mono branche en résulte contient le motif fréquent f:2. Le symbole "d" génère donc deux motifs: d:3 (le long des pointeurs) et df:2. Notons que bd:2 a été trouvé à l'étape précédente. Enfin, le dernier motif fréquent est f:3. L'extraction s'achève et tous les motifs fréquents ont été trouvés.

Summary

Frequent itemsets were used to improve the performances of the technical trading indicators. Our method assigns a signature to the frequent market configurations, and overrides the usual trading rules by the information deduced from the signatures. An iterative back testing selects the best signatures, and combines them with the technical indicator to improve its performances. Combining frequent itemsets with technical trading indicators is an original contribution. The method statistically improves the performances of the trading strategies, and the authors recommend it for any automatic trading system based on technical indicators. This study was based on daily interest rate and stock market data, and three different indicators: Williams % R, moment crossing, and a confidential black box strategy. Signatures are particularly well suited for short memory processes.