

RÉDUCTION OF THE DIMENSION FOR FINANCIAL TIME SERIES
FORECASTING - APPLICATION TO THE BEL 20 MARKET INDEX

Lendasse Amaury, Lee John, de Bodt Eric, Wertz Vincent, Verleysen Michel

Université Catholique de Louvain-la-Neuve, Belgique

lendasse@auto.ucl.ac.be

lee@dice.ucl.ac.be

debodt@fin.ucl.ac.be

wertz@auto.ucl.ac.be

verleysen@dice.ucl.ac.be

Abstract- Prediction of financial time series using artificial neural networks has been the subject of many publications, even if the predictability of financial series remains a subject of scientific debate in the financial literature. Facing this difficulty, analysts often consider a large number of exogenous indicators, which makes the fitting of neural networks extremely difficult. In this paper, we analyze how to aggregate a large number of indicators in a smaller number using -possibly nonlinear- projection methods. As the use of nonlinear projection tools involves the tuning of several coefficients, the core of this work is the development of a sound methodology to adjust these parameters, based on objective criteria. Several of these criteria are presented, and used in appropriated circumstances. Nonlinear projection methods are shown to be equivalent to the linear Principal Component Analysis when the prediction tool used on the new variables is linear. Furthermore, the computation of the nonlinear projection gives an objective way to evaluate the number of resulting indicators needed for the prediction. Finally, the advantages of nonlinear projection could be further exploited by using a subsequent nonlinear prediction model. It is also shown that the use of the determination coefficient is dangerous in the specific case of financial predictions. The methodology developed in the paper is validated on data from the BEL20 market index, using systematic cross-validation results.

The application of this methodology on the BEL20 Market Index shows that comparable results are obtained when using a linear projection method or a non-linear one, when a subsequent linear prediction tool is used. The advantage of the methodology presented here is that it automatically evaluates the number of new variables that must be kept after projection, in order to keep the necessary and relevant information needed for the prediction. Finally, the advantages of a nonlinear projection could be further exploited by using a subsequent nonlinear prediction tool, even if the results would be more difficult to illustrate because of the inherent difficulties of nonlinear predictors.

Key words – Time series, prediction, Dimension Reduction and Technical Indicators.

References

1. E. Fama (1965) The Behavior of Stock Market Prices, *Journal of Business*, 38, pp. 34-105.
2. L. Bachelier (1900), *Théorie de la spéculation*, Gauthier-Villars, Paris.
3. J. Campbell, A. Lo and A. C. MacKinlay (1997) *The Econometrics of Financial Markets* Princeton University Press.
4. W. Brock, J. Lakonishok and B. Le Baron (1992) Simple Technical Trading Rules and the Stochastic Properties of Stock Returns, *The Journal of Finance*, vol. XLVII, n°5, pp. 1731-1764.
5. R. Sullivan, A. Timmerman and H. White (1999) Data-Snooping, Technical Trading Rule Performance, and the Bootstrap, *The Journal of Finance*, vol. 54, issue 5, pp. 1647-1691.
6. A. Lo, H. Mamaysky and J. Wang (2000) Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation, *The Journal of Finance*, vol. 55, n° 4, août 2000, sous presse.
7. A.P. Refenes, A.N. Burgess and Y. Bentz (1997) Neural Networks in Financial Engineering: A Study in Methodology, *IEEE Transactions on Neural Networks*, vol. 8, n°6, pp. 1222-1267.
8. P. Demartines, J. Héroult (1997) Curvilinear Component Analysis: A self-organizing neural network for nonlinear mapping of data sets, *IEEE Trans. on Neural Networks*, 8(1), pp. 148-154.
9. A.Lo., A.C. MackKinlay (1990) Data-snooping Biases in Tests of Financial Asset Pricing Models, *The Review of Financial Studies*, vol. 3, p. 431-467.
10. F. Takens (1985) On the numerical Determination of the dimension of an attractor, *Lecture Notes in Mathematics*, Vol. 1125, Springer-Verlag, pp. 99-106.
11. P. Grassberger and I. Procaccia (1983) Measuring the Strangeness of Strange Attractors, *Physica D*, 56, pp. 189-208.
12. P. Demartines and J. Héroult (1993) Vector Quantization and Projection, *Lecture Notes in Computer Sciences*, Vol. 686, Springer-Verlag, pp. 328-333.
13. T.J. Brailsford and R.W. Faff (1996) An evaluation of volatility forecasting techniques, *Journal of Banking and Finance*, vol. 20, pp. 419-438.
14. A. Lendasse, E. de Bodt, V. Wertz and M. Verleysen (2000) Nonlinear Financial Time Series Forecasting – Application to Bel20 Stock Market Index, *European Journal of Economic and Social Systems*, 14, n°1, pp 81-91.
15. M. Verleysen, E. de Bodt and A. Lendasse (1999) Forecasting Financial Time Series through Intrinsic Dimension Estimation and Nonlinear Data Projection, *Engineering Applications of Bio-inspired Artificial Neural Networks*, vol. II, 1607: pp. 596-605.

REDUCTION DE LA DIMENSION D'UN ENSEMBLE D'INDICATEURS
TECHNIQUES EN VUE DE LA PREDICTION DE SERIES TEMPORELLES
FINANCIERES - APPLICATION A L'INDICE DE MARCHE BEL 20

Lendasse Amaury, Lee John, de Bodt Eric, Wertz Vincent, Verleysen Michel

Université Catholique de Louvain-la-Neuve, Belgique

lendasse@auto.ucl.ac.be

lee@dice.ucl.ac.be

debodt@fin.ucl.ac.be

wertz@auto.ucl.ac.be

verleysen@dice.ucl.ac.be

Résumé – *La prévision des séries financières par l'utilisation de réseaux de neurones artificiels a fait l'objet de nombreuses publications, et ce malgré le fait que la prévisibilité même de ces séries reste l'objet d'un débat scientifique important dans la littérature financière. Dans ce domaine, le chercheur prend souvent en considération un grand nombre d'indicateurs exogènes, ce qui rend l'estimation des paramètres d'un réseau de neurones très complexe. Dans cet article, nous analysons comment agréger un grand nombre d'indicateurs en utilisation des méthodes de projection, notamment non-linéaires, afin de réduire le nombre de variables d'entrée dans le réseau de neurones. Nous montrons que les méthodes de projection non-linéaires produisent des résultats au moins équivalents à ceux obtenus à l'aide de la très classique analyse en composantes principales, lorsque l'on utilise un outil de prévision linéaire pour effectuer les comparaisons. L'approche non-linéaire fournit en outre un moyen objectif d'évaluer le nombre de facteurs à conserver pour effectuer la prévision. Enfin, les avantages de l'approche non-linéaire pourraient être exploités plus avant en utilisant un outil de prévision non-linéaire. La méthodologie proposée dans ce travail est appliquée sur des données provenant de l'indice de marché BEL20, en utilisant de manière systématique une technique de cross-validation pour s'assurer de la stabilité des résultats.*

Key words – **Time series, prediction, Dimension Reduction and Technical Indicators.**

1 Introduction

Depuis le début de ce siècle, la question de la prévisibilité des séries financières (au moins de l'évolution des prix observés sur les marchés d'actions) est l'objet d'un vif débat en finance. Fama [1], dans son article fondateur, rappelle la signification précise de l'hypothèse de marche aléatoire des cours boursiers¹ (proposée initialement par Bachelier [2]) et en présente différents tests empiriques. Il conclut dans ces termes : "La principale conclusion est que les données semblent présenter un support consistant et fort de cette hypothèse. Cela implique, évidemment, que la lecture des graphiques, malgré qu'il puisse s'agir d'un passe-temps agréable, n'est d'aucune valeur réelle pour l'investisseur dans un marché d'action". De nombreux travaux empiriques, principalement basés sur des tests statistiques linéaires, ont conduit à la même conclusion dans les années soixante et septante et ce malgré l'utilisation abondantes des approches chartistes et des indicateurs techniques² par la communauté professionnelle. Toutefois, comme le soulignent Campbell, Lo et MacKinlay [3], "De récents travaux économétriques semblent suggérer que les prix des actifs financiers sont prévisibles, au moins jusqu'à un certain degré". Parmi ces travaux, trois ont constitué des avancées importantes. Brock, Lakonishok et Le Baron [4] testent deux règles techniques populaires sur l'indice de marché Dow Jones sur la période 1897 à 1986. Ils utilisent une approche de type bootstrap pour valider leurs résultats et concluent que "les résultats confortent l'intérêt de l'utilisation des règles d'analyse technique". Sullivan, Timmerman et White [5] proposent de nouveaux résultats sur les mêmes données (étendues avec dix nouvelles années). Leur méthodologie, qui repose toujours sur l'utilisation du bootstrap, permet d'éviter (au moins dans une mesure significative) le biais de *data-snooping*³ et est appliquée cette fois à 26 règles techniques. Ils confirment que les résultats de Brock, Lakonishok et Le Baron résistent à l'inspection du biais de *data-snooping*. La récente contribution de Lo, Mamaysky et Wang [6], fondée sur une utilisation d'une nouvelle approche de régression non-paramétrique, confirme que "différents indicateurs techniques fournissent un supplément d'information et peuvent avoir une valeur réelle". Sur la base de ces résultats empiriques nouveaux, nous considérerons dans ce travail, tout comme le font Refenes, Burgess et Bentz [7] dans leur introduction à l'utilisation des réseaux de neurones en finance, qu'il peut y avoir un sens à tenter de prévoir l'évolution du prix d'actifs financiers.

Lorsque le problème de la prévision des séries temporelles est abordé comme un problème de régression, les variables à considérer sont les valeurs passées de la série ainsi qu'un certain nombre d'autres variables exogènes. On pose évidemment l'hypothèse que ces variables contiennent une information qui permettra la prévision de la série en question. En l'absence

¹ Selon l'hypothèse de marche aléatoire des cours boursiers, les changements de prix successifs sont statistiquement indépendants : les prix d'un actif financier, si le marché est efficient, marchent au hasard. La principale justification de cette hypothèse est la suivante : si le marché est efficient, à chaque moment du temps, toute l'information disponible est prise en compte par les opérateurs pour déterminer le prix effectivement observé. Les mouvements ultérieurs du prix sont dès lors causés par l'arrivée de nouvelles informations, processus qui par définition est strictement aléatoire.

² Les indicateurs techniques sont des valeurs numériques calculées sur la base des prix passés, des volumes et d'autres statistiques de marché et utilisés pour prévoir les mouvements futurs des prix.

³ Le concept de biais de *data-snooping* a été proposé par Lo et MacKinlay [8] (dans la littérature financière). Ce biais dans l'évaluation de la performance des méthodes de prévision apparaît dès que des données sont utilisées plus d'une fois afin d'effectuer la sélection de modèles. Il peut être illustré par l'assertion : "Donnez-moi les données, je vous donnerai le modèle".

d'un modèle théorique fort, il est néanmoins souvent en pratique difficile de déterminer a priori quelles sont les variables qui sont pertinentes, celles qui sont redondantes et celles qui sont simplement inutiles. De plus, l'utilisation de modèles de régression (et c'est d'autant plus vrai qu'il s'agit de modèles de régression non-linéaire, telles des perceptions ou des réseaux à fonctions radiales) est d'autant plus difficile que le nombre de variables est important. La réduction du nombre de ces variables (qui conditionne directement la complexité du modèle de régression) est d'un intérêt tout particulier. La question est dès lors de savoir comment réduire le nombre de variables sans perdre (trop) d'information utile pour effectuer des prévisions.

Dans cet article, nous centrons notre attention sur ce point et nous proposons une comparaison de l'approche linéaire classique de compression de l'information (l'analyse en composantes principales) avec une nouvelle approche non-linéaire. L'objectif est de conserver autant que possible l'information contenue dans les variables d'origine, tout en réduisant le plus possible le nombre de facteurs (ou variables construites) utilisés pour effectuer la prévision (à savoir, le nombre de dimensions de l'espace de projection). Les nouvelles variables "construites" sont celles utilisées comme entrées pour le modèle de prévision (cfr fig. 1).

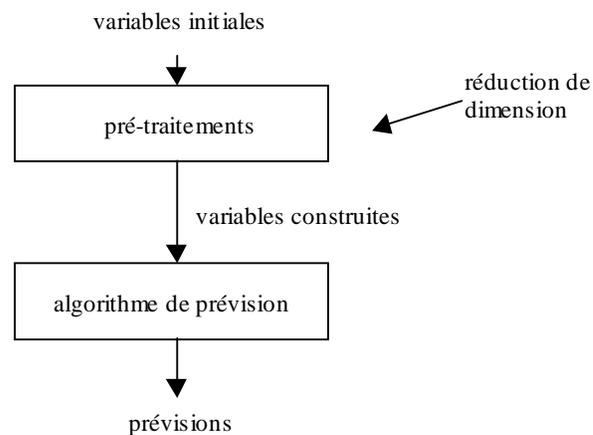


Fig. 1. Les deux étapes de la méthodologie.

Pour réaliser la transformation entre les variables indépendantes initiales et les variables construites⁴, on peut utiliser soit une approche linéaire, soit une approche non-linéaire. Dans cet article, nous utiliserons l'analyse en composantes principales (ACP ci-après) comme approche linéaire de référence et l'analyse en composantes curvi-linéaires [9] (CCA ci-après) comme approche non-linéaire.

Dans un premier temps, nous rappellerons quelles sont les différentes approches de projection possibles et quelles sont les critères d'évaluation que l'on peut utiliser pour évaluer la qualité de la projection obtenue. Ensuite, nous présenterons la méthodologie que nous avons utilisée, ce point constituant le point central de notre travail. Nous présenterons enfin les résultats obtenus sur des données réelles, issues du marché boursier belge.

⁴ Nous utiliserons par la suite le terme "facteurs", repris du domaine l'analyse en composantes principales mais étendu au contexte non-linéaire.

2 Techniques de réduction du nombre de dimensions

2.1 La dimension intrinsèque

Il est dans un premier temps important de chercher à évaluer la dimension de projection (la dimension de l'espace des facteurs). Si l'estimation de cette dimension est trop petite, beaucoup d'information sera perdue dans le cadre de la projection et cela pénalisera inévitablement les performances de l'algorithme de prévision. Si son estimation est trop importante, la complexité du modèle de prévision croîtra très rapidement (d'autant plus vite que l'algorithme de prévision utilisé est non-linéaire) et son estimation deviendra très instable. Pour évaluer la dimension de l'espace de projection, le concept de dimension intrinsèque est utilisé. La dimension intrinsèque est le nombre effectif de degrés de liberté d'un ensemble; une définition plus précise de concept est proposée dans [10]. Nous l'illustrons ici par l'exemple bien connu du "fer à cheval" (cfr fig. 2). Pour ce jeu de données, la dimension intrinsèque est deux (deux degrés de libertés sont suffisants pour déterminer de manière unique chaque point dans l'espace des données et cela malgré que les données soient situées dans un cube), alors que la dimension apparente est trois (le cube). Le calcul de la dimension intrinsèque est expliqué dans [11], mais sa détermination reste très délicate en pratique pour des jeux de données de grande dimension. Nous considérerons dès lors ici qu'il ne s'agit que d'une approximation assez grossière du nombre de dimensions qui doit être utilisé pour réaliser la projection.

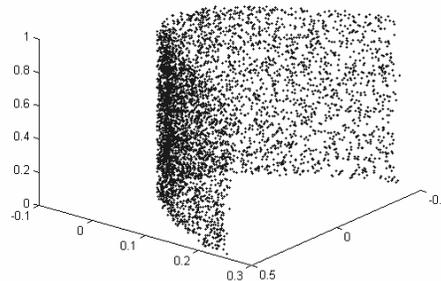


Fig. 2. La distribution en "fer à cheval".

2.1 Les méthodes de projection

2.2.1 L'analyse en composantes principales

L'ACP est basée sur le calcul de la matrice de variance / covariance entre les variables initiales du jeu de données. La décomposition en valeurs singulières de la matrice fournit la transformation linéaire désirée et il reste au chercheur à choisir le nombre de composantes principales (le nombre de facteurs) à conserver (il s'agit du choix de la dimension de l'espace de projection). Dans ce contexte, l'estimation de la dimension intrinsèque n'est pas une étape préalable indispensable. L'ACP étant une transformation linéaire, le nombre de dimensions nécessaires pour capter une part significative de l'information présente dans le

jeu de données initial (en termes plus techniques, le nombre de composantes principales nécessaire pour capter 90 ou 95% de variance totale) est en général supérieur à la dimension intrinsèque du jeu de données.

2.2.2 L'analyse en composante curvi-linéaire (CCA)

Cette extension de l'ACP (appelée également *Vector Quantization and Projection* [8,12]) déploie l'espace qui contient les données d'origine et réalise une projection (au sens large du terme) vers un espace de dimension plus petite. La projection de la distribution en "fer à cheval" obtenue à l'aide de l'algorithme CCA est illustrée à la fig. 3. Les détails de l'algorithme ne seront pas expliqués ici mais il doit être mentionné que la première étape consiste en une quantification de l'espace d'entrée (ou espace des variables indépendantes initiales) à l'aide de centroïdes. Cette quantification produit une partition de l'espace d'entrée en sous-groupes au sein desquels on pose l'hypothèse qu'une représentation linéaire de la distribution des données représente une approximation acceptable. La CCA doit être considérée comme une forme d'ACP par parties, chaque sous-groupe de données obtenu après quantification étant l'objet d'une compression par ACP. Le nombre de sous-groupes est un paramètre important qui doit être déterminé a priori pour effectuer la projection. Une approche précise pour déterminer ce paramètre sera décrite ci-dessous et appliquée aux données issues du marché boursier belge.

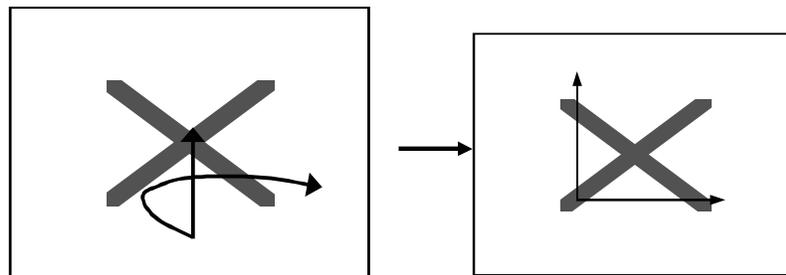


Fig. 3. Projection de la distribution en "fer à cheval" réalisée par l'algorithme CCA.

2.3 Les critères d'évaluation

L'objectif de la projection est la réduction du nombre de variables explicatives nécessaire pour effectuer la prévision tout en préservant (autant que possible) l'information contenue dans le jeu de données initial. Il y a plusieurs approches possibles pour évaluer la qualité de la projection obtenue. En se référant à la fig. 1, on peut chercher à mesurer si le résultat de la projection contient la même information que le jeu des variables indépendantes de départ, sans se préoccuper de l'étape ultérieure de prévision. L'alternative est d'évaluer la qualité des prévisions obtenues, en utilisant les facteurs issus de la projection comme variables explicatives pour l'algorithme de prévision. L'avantage de cette seconde approche est de concentrer l'attention sur l'objectif même du problème, à savoir la prévision. Son inconvénient est bien entendu que toute lacune au niveau de l'étape de prévision (par exemple, des problèmes de convergence numérique dans le contexte des algorithmes non-linéaires) se répercutera sur la mesure de la qualité de la projection. Le premier critère abordé ci-dessous appartient à la première des deux approches tandis que les trois suivants relèvent de la seconde approche.

2.3.1 La mesure de dé-projection

La projection par ACP ou CCA est une opération réversible; toutefois, les résultat d'une "dé-projection" ne permettent pas de reconstruire strictement la situation de départ (sauf évidemment si l'espace de projection est de même dimension que l'espace initial). La distance moyenne entre les variables indépendantes initiales et le résultat d'une dé-projection est appelée mesure de dé-projection. L'objectif est que cette distance soit aussi petite que possible et qu'elle corresponde (ce qui serait idéal) au bruit qui entoure l'information. La mesure de dé-projection peut être utilisée pour comparer les performances de différents algorithmes de projection (tels l'ACP, la CCA, ...), pour autant que l'on utilise le même nombre de dimensions pour l'espace de projection. Le principal avantage de ce critère est qu'il est totalement indépendant de la mesure de prévision utilisée par après. Cet avantage a malheureusement un coût : il n'y a pas de garantie que la dimension de projection choisie, voire l'algorithme de projection retenu, soit le meilleur pour obtenir les prévisions les meilleures possibles.

2.3.2 Le coefficient de détermination

Ce critère est souvent utilisé dans le cadre des applications financières, en particulier dans le domaine de la prévision de la volatilité (cfr, par exemple, Brailsford and Faff [13]). Contrairement au critère précédant, il requiert une prévision de la variable cible (ou variable dépendante) et est fondé sur une comparaison entre les valeurs prévues et les valeurs réelles de cette dernière. Désignons y_t est la cible (ou variable dépendante), \hat{y}_t sa prévision et \bar{y} la valeur moyenne des y_t . Afin d'évaluer si les prévisions correspondent aux valeurs réelles, on évalue la régression linéaire suivante (1) sur l'ensemble du jeu de données :

$$y_t \cong A + B\hat{y}_t . \quad (1)$$

A et B sont les coefficients de cette régression. On calcule alors le coefficient de détermination selon l'expression :

$$r^2 = \frac{\sum_t (y_t - \bar{y})^2 - \sum_t (y_t - A - B\hat{y}_t)^2}{\sum_t (y_t - \bar{y})^2} . \quad (2)$$

Le coefficient r^2 mesure le degré avec lequel les points définis par (y_t, \hat{y}_t) sont proches de la ligne droite définie par les coefficients A et B. Le plus r^2 est proche de 1, le plus les points sont proches de la droite, le plus la prévision est considérée comme correcte. Ce critère, bien que souvent utilisé dans les travaux empiriques, est toutefois peu approprié, en particulier dans le contexte des applications financières. En effet, les données financières soit très bruitées : la part déterministe (la partie du processus qui peut faire l'objet d'une prévision) est généralement beaucoup moins importante que la part qui correspond au bruit (en tout cas pour les données de marché, ce qui n'est pas sans nous renvoyer à la problématique de l'efficacité évoquée en introduction). Pas conséquent, l'écart-type de \hat{y}_t est souvent très inférieur à celui de y_t , conduisant à une régression mal conditionnée (1). Nous illustrerons ce point lors de l'application de notre méthodologie aux données issues de l'indice boursier BEL20.

2.3.3 L'erreur moyenne quadratique (ou *Mean Square Error* - MSE)

Le critère MSE se définit comme :

$$\text{MSE} = \frac{\sqrt{\sum_t (y_t - \hat{y}_t)^2}}{\sum_t 1}. \quad (3)$$

L'objectif est de la minimiser afin d'obtenir des prévisions aussi précises que possible. Malheureusement, en pratique, comme les données financières sont très bruitées, le MSE est souvent très élevé.

2.3.4 Le pourcentage du nombre de signes corrects (ou *Percentage of Correct Approximations of Sign* - PCAS)

L'amplitude des mouvements des séries financières étant très difficile à prévoir, il peut être intéressant de calculer le PCAS (c'est-à-dire de se concentrer sur la prévision de la direction des mouvements). Le PCAS s'évalue comme :

$$\text{PCAS} = \frac{\sum_t (\text{sign}(y_t) \text{sign}(\hat{y}_t) + 1)}{2 \sum_t 1} 100. \quad (4)$$

L'objectif est de la maximiser. La prévision doit aussi vérifier la condition suivante :

$$\text{PCAS} \geq \max \left(\frac{\sum_t (\text{sign}(y_t) + 1)}{2 \sum_t 1} 100, \frac{\sum_t (1 - \text{sign}(y_t))}{2 \sum_t 1} 100 \right); \quad (5)$$

Si tel n'est pas le cas, cela signifie qu'une prévision naïve et constante (toujours la hausse ou toujours la baisse, selon le cas) serait plus efficace que la prévision obtenue !

2.4 La méthodologie

La procédure utilisée pour projeter les variables indépendantes initiales (compresser le signal) et le test utilisé pour déterminer si la projection améliore la prévision sont les suivants.

Par la suite, nous noterons :

- u_t^i , $1 \leq i \leq n$, l'indicateur technique ou la variable exogène utilisée dans l'algorithme de prévision (les valeurs passées de la série à prévoir peuvent faire partie des indicateurs techniques);
- z_t , le vecteur d'entrée contenant les valeurs de la série et les variables exogènes à la période t : $z_t = [u_t^1 \ u_t^2 \ \dots \ u_t^n \ y_t]$
- T , le nombre total d'observations disponibles dans la série ($1 \leq t \leq T$);

- D, la dimension intrinsèque de l'ensemble d'apprentissage.

2.4.1 Mélange

Les vecteurs d'entrées z_t sont tirés au hasard sans remplacement, ce qui conduit à la loi $f(t)$:

$$z_t = [u_t^1 \quad u_t^2 \quad \dots \quad u_t^n \quad y_t] = [u_{f(t)}^1 \quad u_{f(t)}^2 \quad \dots \quad u_{f(t)}^n \quad y_{f(t)}]. \quad (6)$$

2.4.2 Ensemble d'apprentissage et de validation

Les données doivent être divisées en deux sous-ensembles: un ensemble d'apprentissage et un ensemble de validation. L'ensemble de validation est utilisé pour vérifier la stabilité des résultats obtenus sur l'ensemble d'apprentissage et l'absence de phénomène de sur-apprentissage (*overfitting*). Les résultats de validation peuvent alors être utilisés pour déterminer les paramètres de la méthode d'apprentissage (tels le nombre de centroides utilisés pour effectuer la projection). La division faite entre l'ensemble d'apprentissage Z^L et l'ensemble de validation Z^V est réalisée selon le principe suivant :

$$Z^L = \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_{\alpha T} \end{bmatrix} = \begin{bmatrix} u_1^1 & u_1^2 & \dots & u_1^n & y_1 \\ u_2^1 & u_2^2 & \dots & u_2^n & y_2 \\ \dots & \dots & \dots & \dots & \dots \\ u_{\alpha T}^1 & u_{\alpha T}^2 & \dots & u_{\alpha T}^n & y_{\alpha T} \end{bmatrix} \quad (7)$$

et

$$Z^V = \begin{bmatrix} z_{\alpha T+1} \\ z_{\alpha T+2} \\ \dots \\ z_T \end{bmatrix} = \begin{bmatrix} u_{\alpha T+1}^1 & u_{\alpha T+1}^2 & \dots & u_{\alpha T+1}^n & y_{\alpha T+1} \\ u_{\alpha T+2}^1 & u_{\alpha T+2}^2 & \dots & u_{\alpha T+2}^n & y_{\alpha T+2} \\ \dots & \dots & \dots & \dots & \dots \\ u_T^1 & u_T^2 & \dots & u_T^n & y_T \end{bmatrix}. \quad (8)$$

Le paramètre α (αT doit être une valeur entière) doit être choisi avec précaution. Il doit être suffisamment important pour s'assurer d'un nombre suffisant d'observations dans l'ensemble d'apprentissage (chaque observation placée dans l'ensemble de validation limitant d'autant l'information disponible pour évaluer les paramètres du modèle) tout en laissant assez d'observation dans l'ensemble de validation pour évaluer sérieusement la stabilité des résultats. Certains conseils seront donnés ultérieurement concernant le choix de ce paramètre.

2.4.3 Normalisation des indicateurs techniques

Dans l'ensemble d'apprentissage, chacun des n indicateurs u_i^{iL} ($1 \leq i \leq n$) doit être normalisé selon :

$$\tilde{u}_t^{iL} = \frac{u_t^{iL} - \bar{u}^{iL}}{\sigma^{iL}} \quad (9)$$

où $\bar{u}^{iL} = \frac{\sum_t u_t^{iL}}{\sum_t 1}$ est la valeur moyenne de l'indicateur u_t^{iL} évalué sur l'ensemble

d'apprentissage, $\sigma^{iL} = \sqrt{\frac{\sum_t (u_t^{iL} - \bar{u}^{iL})^2}{\sum_t 1}}$ est son écart-type et \tilde{u}_t^{iL} est sa valeur normalisée.

Comme la normalisation fait partie de l'apprentissage, les observations de l'ensemble de validation ne sont pas utilisées pour le calcul de la moyenne et de l'écart-type. L'ensemble de validation est donc normalisé en utilisant les mêmes paramètres que ceux calculés pour l'ensemble d'apprentissage (ce qui suppose la stationnarité de la série) :

$$\tilde{u}_t^{iV} = \frac{u_t^{iV} - \bar{u}^{iL}}{\sigma^{iL}}. \quad (10)$$

2.4.4 La dimension intrinsèque

La dimension intrinsèque D des indicateurs techniques de l'ensemble d'apprentissage est calculée et utilisée comme approximation "grossière" de la dimension de projection adéquate.

2.4.5 L'ACP initiale

La CCA pourrait être utilisée directement comme méthode de projection. Il doit toutefois être souligné qu'en pratique, réaliser une ACP préalable facilite largement la mise en œuvre de la CCA. Nous réalisons donc une ACP initiale sur les indicateurs techniques de l'ensemble d'apprentissage. La dimension n_1 est choisie comme la dimension minimale qui limite la perte d'information à un niveau négligeable. Les indicateurs techniques de l'ensemble d'apprentissage et de l'ensemble de validation sont alors projetés de l'espace initial (de dimension n) vers un sous-espace de dimension n_1 , par application des coefficients qui caractérisent les facteurs obtenus par l'ACP sur l'ensemble d'apprentissage.

2.4.6 La CCA initiale

Une première CCA est alors réalisée sur l'espace de dimension n_1 afin de le ramener à un espace de dimension D . L'algorithme CCA est utilisé, en prenant en compte exclusivement les données de l'ensemble d'apprentissage. En première approximation, le nombre de centroides est choisi de manière telle que, en moyenne, la dimension intrinsèque locale de chaque sous-groupe issu de la quantification soit égal à D .

2.4.7 Optimisation de n_1

La projection de l'espace initial de dimension n vers un espace de dimension D en utilisant la procédure décrite aux points 6 (réalisation d'une ACP) et 7 (réalisation d'une CCA) est répétée, mais cette fois en faisant varier la dimension n_1 autour de sa valeur initiale. Chaque projection est également appliquée à l'ensemble de validation. La mesure de dé-projection obtenue sur l'ensemble de validation est étudiée. Habituellement, cette mesure a un minimum (ou un coude) et ce minimum (ou coude) est choisi comme valeur optimale pour n_1 .

2.4.8 Critère d'erreur sur l'ensemble de validation

De nouvelles projections sont réalisées de l'espace initial de dimension n à l'espace de dimension D en faisant varier la méthode de projection (ACP uniquement ou ACP + CCA), le nombre de centroides (dans le cas non-linéaire) et la dimension de projection D . Dans chaque cas, la mesure de dé-projection sur l'ensemble de validation est calculée. Ensuite, le coefficient de détermination, le MSE et le PCAS sont évaluées, également sur l'ensemble de validation. Pour la mesure de ces trois derniers critères, nous devons avoir à notre disposition les prévisions \hat{y}_i ; ces valeurs peuvent être obtenues par tout algorithme de prévision, linéaire ou non-linéaire (perceptrons, réseaux à fonction radiale, ...). Dans l'exemple développé à la section III, nous nous limitons à un modèle linéaire, l'objectif de notre travail étant de comparer les approches de projection linéaire et non-linéaire dans le cadre des applications financières.

2.4.9 Validation croisée

Les opérations 1 à 8 sont répétées k fois; on calcule ensuite la moyenne des quatre mesures d'erreur calculées au point 8. Différents paramètres doivent toutefois encore être optimisés. Comme une optimisation globale est difficilement envisageable en pratique, nous proposons de d'abord déterminer le nombre optimal de centroides (utilisés dans la CCA) en minimisant la mesure de dé-projection, qui est indépendante du modèle de prévision. Les autres erreurs sont utilisées pour choisir la dimension de projection finale (D_L dans le cas de l'ACP et D_{NL} dans le cas de la combinaison ACP et CCA). On notera que la mesure de dé-projection ne peut être calculée sans une connaissance préalable du nombre de centroides. Le choix entre le coefficient de détermination, le MSE et le PCAS doit être réalisé selon l'objectif de l'application en question. L'analyse de l'influence de ces critères sur le choix de D_L et D_{NL} dans le cadre de l'application aux données issues du marché boursier belge sera détaillée à la section III. Le nombre k de répétitions doit être suffisamment important pour obtenir des résultats représentatifs. De plus, plus α est grand (coefficient de répartition des observations entre les ensembles d'apprentissage et de validation), plus k doit également être élevé (plus α est grand, moins il y a d'observations dans l'ensemble de validation et plus les résultats obtenus à ce niveau sont instables). Dans la mesure où le temps de calcul le permet, la stratégie suivie consiste donc à augmenter simultanément α et k , pour obtenir à la fois une meilleure estimation des paramètres et une robustesse plus forte des résultats.

3 Application à l'indice de marché BEL20

La méthodologie proposée à la section précédente est testée sur l'indice de marché BEL20⁵. La période couverte va du 1^o décembre 1987 au 14 février 1998, soit 2663 observations journalières. Les données utilisées sont présentées à la fig. 4. Le choix de l'indice BEL20 est motivé essentiellement par deux considérations : d'une part, l'indice belge n'a pas fait jusqu'à présent l'objet de nombreuses études et comme tel, le risque de *data-snooping* semble réduit; d'autre part, le marché boursier belge étant de taille réduite, son niveau d'efficience n'est probablement pas aussi élevé que celui des marchés boursiers les plus importants de notre planète et il devrait offrir plus d'opportunités pour l'utilisation des approches qui relèvent de l'analyse technique. 42 indicateurs techniques ont été utilisés (ils sont décrits au tableau 1). Ces indicateurs sont fortement corrélés (l'indice de conditionnement de la matrice variance / covariance est quasi infini – la plus faible valeur propre étant quasi-nulle).

Indicateurs	Nombre
Valeurs retardées des rendements	4
Valeurs retardées de la variance des rendements	3
Moyennes mobiles des rendements	3
Pentes des valeurs retardées de rendements	3
Valeurs retardées des taux d'intérêts courts et longs	3
Valeurs retardées de la variances des taux d'intérêts courts et longs	3
Pente des valeurs retardées des prix	3
Indicateurs techniques basés sur l'indice de marché SP500	12
Indicateurs techniques basés sur le taux de change de la devise belge	8

Tab. 1. Courte description des 42 indicateurs techniques utilisés.

⁵ Cet indice de marché est calculé à partir de l'évolution du prix des 20 actions les plus représentatives de l'activité de la bourse de Bruxelles.

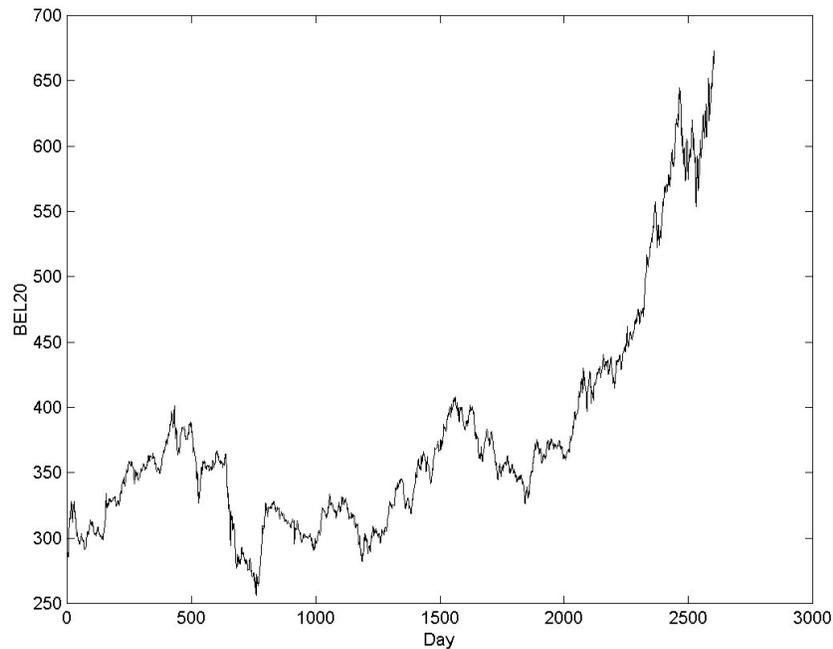


Fig. 4. Evolution de l'indice de marché BEL20 du 1^o décembre 1987 au 14 février 1998.

Prévoir l'évolution de cette série est une tâche complexe. Le premier constat est que cette série n'est pas stationnaire⁶. De plus, le coefficient de corrélation entre le prix à l'instant t et le prix à l'instant $t+1$ est de 0.999, tandis que celui observé entre le taux de rentabilité à l'instant t (la variation du prix entre $t-1$ et t) et le taux de rentabilité à l'instant $t+1$ est de 0.099. Ces deux résultats semblent, en première analyse, confirmer l'hypothèse de marche aléatoire. Le taux de rentabilité journalier y_t de cet indice est calculé selon la formule :

$$y_t = \log\left(\frac{\text{BEL20}_t}{\text{BEL20}_{t-1}}\right) \quad (11)$$

et est présenté à la fig. 5. Sa moyenne est proche de 0. La méthodologie présentée à la section précédente y est appliquée :

⁶ Un test augmenté de Dickey-Fuller confirme la présence d'une racine unitaire à un niveau de confiance de 1%.

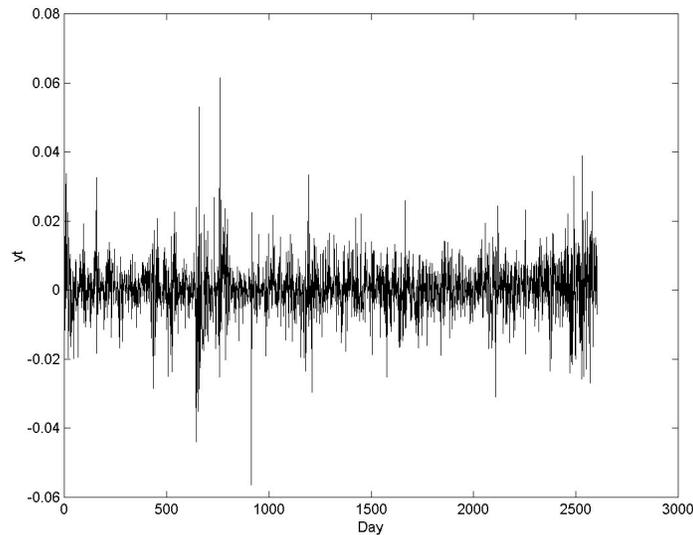


Fig. 5. Taux de rentabilité journaliers de l'indice de marché BEL20 du 1^o décembre 1987 au 14 février 1998.

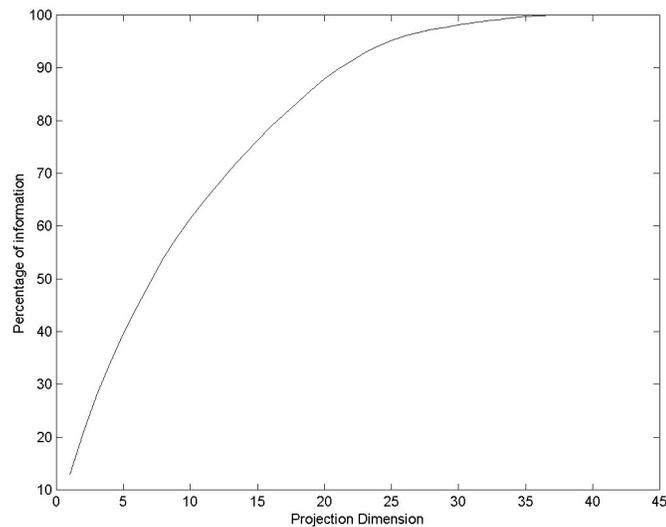


Fig. 6. Pourcentage d'information (mesuré par le rapport entre la variance expliquée et la variance totale) conservé à l'aide de l'ACP en fonction du nombre retenu de dimensions de projection.

- les étapes 1 à 3 sont réalisées;
- étape 4 : la dimension intrinsèque des indicateurs techniques est de 7;
- étape 5 : le résultat de l'ACP sont présentés à la fig. 6. La figure reprend le pourcentage d'information conservé (mesuré par le rapport entre la variance expliquée et la variance totale) à l'aide de l'ACP en fonction du nombre de dimensions.

La dimension initiale de projection n_1 retenue est de 30; une projection linéaire de l'espace d'entrée (42 dimensions) dans l'espace en 30 dimensions est réalisée (conservant 98% de la variance totale du jeu initial de données).

- étape 6 : une projection non-linéaire (CCA) est réalisée de l'espace en 30 dimensions vers l'espace en 7 dimensions. Le nombre initial de centroides est de 125.
- étape 7 : la projection de l'espace en 42 dimensions vers l'espace en 7 dimensions est répétées en utilisant l'ACP et la CCA; on fait varier la dimension intermédiaire n_1 autour de sa valeur initiale (30). La mesure de dé-projection obtenue sur l'espace de validation est présentée à la fig. 7. La valeur optimale de n_1 correspond au coude sur la courbe, soit la dimension 27.

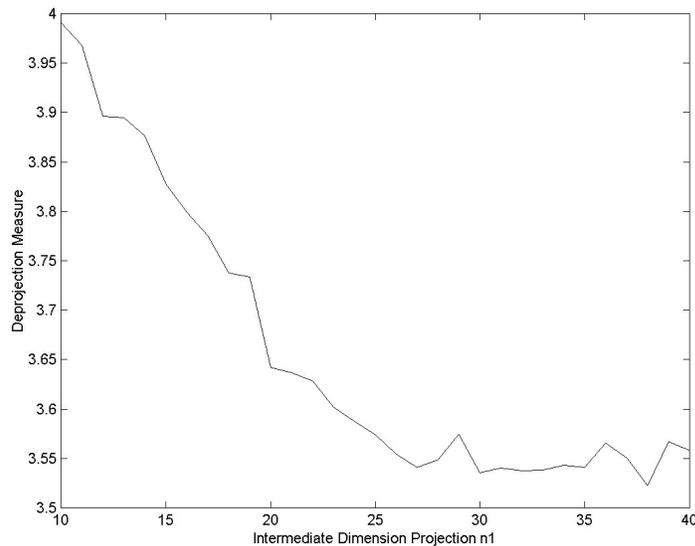


Fig. 7. Mesures de dé-projection obtenues sur l'ensemble de validation en faisant varier n_1 autour de sa valeur initiale.

- étapes 8 et 9 : dans un premier temps, le nombre de centroides est optimisé, en utilisant la mesure de dé-projection. Il apparaît que la valeur 125 est confirmée. Ensuite, de nouvelles projections sont réalisées de l'espace d'entrée (42 dimensions) vers un espace de sortie de dimension D , en faisant varier D autour de 7. Pour valider l'approche proposée, nous avons fait varier D entre 2 et 24. Chaque projection a été répétée 20 fois ($k=20$), tout en faisant varier les ensembles d'apprentissage et de validation. Pour chaque essai, un modèle de prévision linéaire de y_t construit à partir des facteurs obtenus à l'aide de la projection, est estimé sur l'ensemble d'apprentissage. Ensuite, le coefficient de détermination, le MSE et le PCAS sont calculés sur l'ensemble de validation. Les résultats obtenus pour ces trois critères sont présentés aux figs. 8, 9 et 10 respectivement.

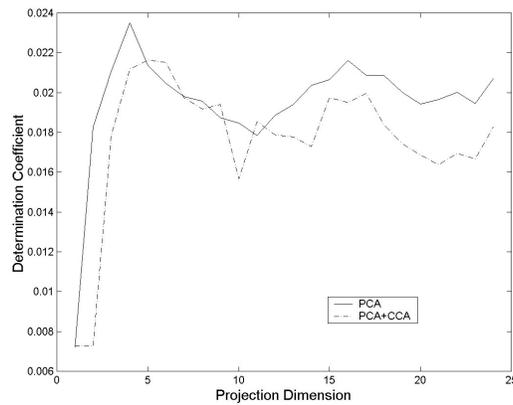


Fig. 8. Coefficients de détermination obtenus sur l'ensemble de validation en fonction de la dimension de projection.

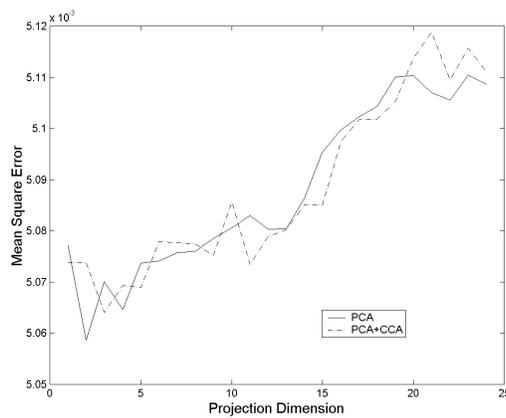


Fig. 9. MSE obtenus sur l'ensemble de validation en fonction de la dimension de projection.

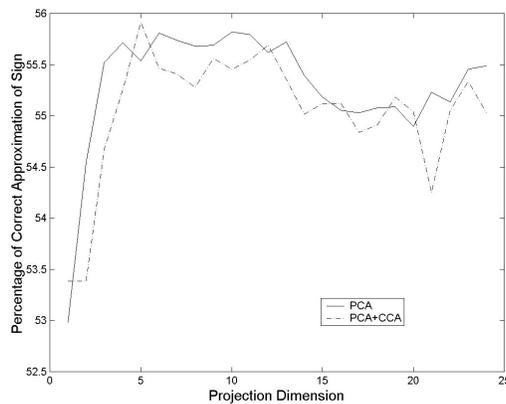


Fig. 10. PCAS obtenus sur l'ensemble de validation en fonction de la dimension de projection

4 Discussion des résultats

On constate tout d'abord que l'approche fondée uniquement l'ACP produit des résultats légèrement meilleurs que ceux obtenus en combinant l'ACP et la CCA. On aurait pu s'attendre au résultat inverse ou à tout le moins à des performances comparables. En fait, ce phénomène peut être expliqué par le faible nombre de données utilisées et par le niveau important du bruit dans la série. De manière générale, on peut s'attendre à ce que la technique de projection non linéaire permette une amélioration des résultats, en particulier lorsque les variables initiales font l'objet de prétraitements non linéaires; certains résultats dans ce sens sont présentés dans [14]. Il faut toutefois souligner que l'utilisation d'une approche de prévision non linéaire aurait, dans ce travail, rendu difficile la comparaison des résultats obtenus à l'aide de l'ACP et de la CCA. Plusieurs problèmes propres à ce type de modèles (problèmes de convergence, choix des valeurs initiales des paramètres, choix de la complexité du modèle, ...) auraient pu influencer fortement les résultats obtenus, rendant très délicate leur interprétation.

On constate que la meilleure dimension de projection dépend du critère d'évaluation retenu : elle est de 4 si l'on se fonde sur la coefficient de détermination, de 2 selon le MSE et de 4 à 6 selon le PCAS. Ce dernier critère est le seul pour lequel l'approche non linéaire (ACP et CCA) produit une amélioration des résultats. En se reportant à l'équation 5, il doit être mentionné que les taux de rentabilité sont positifs 53% du temps. Il faut souligner également que la prévision du signe des taux de rentabilité est un objectif particulièrement important d'un point de vue concret. L'anticipation de l'orientation du marché est à la base des stratégies dites de *market timing*, qui justifient très largement l'utilisation des indicateurs techniques. De plus, comme nous l'avons déjà souligné, l'utilisation d'un algorithme de prévision non linéaire [15] aurait plus que probablement permis d'obtenir des résultats encore plus convaincants.

On soulignera également que pour obtenir des résultats stables avec le coefficient de détermination, le nombre d'itérations doit être augmenté. Cette mesure est en effet très sensible à de petites perturbations dans les données. Pour le calculer, une régression de y_t sur \hat{y}_t est réalisée selon (1). Un exemple d'une telle régression dans le cadre de notre application est présenté à la fig. 11. Les coefficients obtenus sont de 0 pour A et de 0.47 pour B. Le coefficient de détermination est de 0.011. Si une donnée seulement est modifiée (le point entouré dans la fig. 12), les coefficients de la régression deviennent 0 pour A et 0.34 pour B. Le coefficient de détermination devient quant à lui 0.006. La raison de cette instabilité des résultats est la faible amplitude des variations de \hat{y}_t par rapport à l'amplitude des variations de y_t . Malheureusement, de nombreuses applications sur des séries financières présentent cette particularité. Le coefficient de détermination est, dans ce contexte, une mauvaise mesure de la qualité des prévisions obtenues.

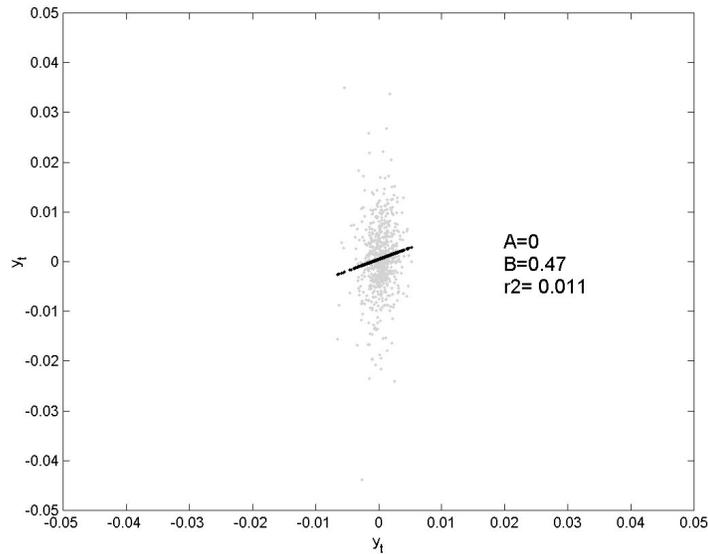


Fig. 11. y_t en fonction de \hat{y}_t en gris et les résultats de la régression en noir.

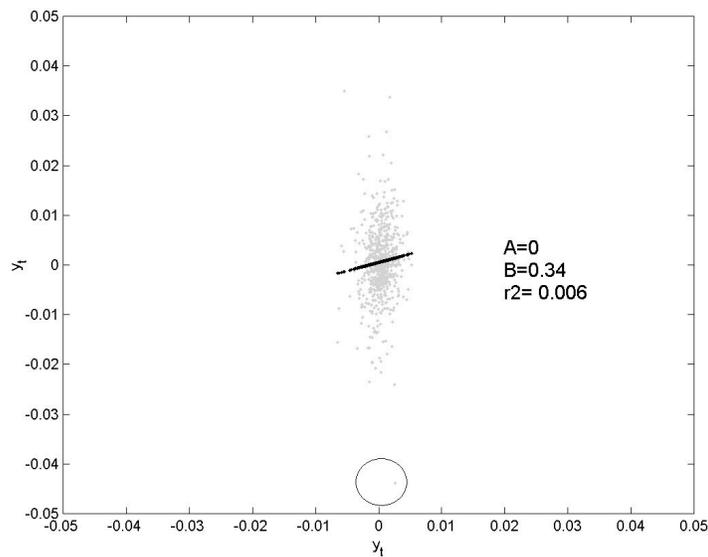


Fig. 12. y_t en fonction de \hat{y}_t en gris et les nouveaux résultats de la régression en noir.

5 Conclusions

Si l'on admet que tenter de prévoir l'évolution future des séries de prix sur des marchés financiers a un sens, l'utilisation des indicateurs techniques est une approche séduisante. Leur très grand nombre rend toutefois difficile l'estimation des modèles de prévision, qu'ils soient

linéaires ou non linéaires. Cet article montre comment l'utilisation de techniques de compression de données, linéaires ou non linéaires, peut aider à réduire leur nombre.

Comme l'utilisation d'une approche de projection non linéaire suppose la détermination d'un certain nombre de paramètres, nous proposons une méthodologie robuste d'ajustement de ces paramètres, basée sur des critères objectifs. Nous soulignons aussi que l'utilisation du coefficient de détermination pour évaluer la qualité des prévisions obtenues est une mauvaise approche dans le cadre des séries de financières (au moins pour les données de marché).

L'application de notre méthodologie sur les données de l'indice BEL20 montre que des résultats comparables sont obtenus en utilisant une approche de projection linéaire (l'analyse en composantes principales) et une approche de projection non linéaire (l'analyse en composantes curvi-linéaires), si la prévision est réalisée à l'aide d'un modèle linéaire. L'avantage de l'approche non linéaire présentée ici est qu'elle permet d'évaluer automatiquement le nombre de facteurs (le nombre de dimensions de l'espace de projection) qui doivent être conservés afin de garder l'information nécessaire à la réalisation d'une prévision. Enfin, il doit être souligné que les avantages de l'approche de projection non linéaire en termes de prévision auraient pu être exploités davantage par l'utilisation d'un modèle de prévision non linéaire, mais qu'un tel choix aurait rendu l'analyse des résultats présentés dans ce travail plus délicat.

Remerciements

Michel Verleysen est chercheur qualifié du Fonds National de la Recherche Scientifique belge.

Références

1. E. Fama (1965) The Behavior of Stock Market Prices, *Journal of Business*, 38, pp. 34-105.
2. L. Bachelier (1900), *Théorie de la spéculation*, Gauthier-Villars, Paris.
3. J. Campbell, A. Lo and A. C. MacKinlay (1997) *The Econometrics of Financial Markets* Princeton University Press.
4. W. Brock, J. Lakonishok and B. Le Baron (1992) Simple Technical Trading Rules and the Stochastic Properties of Stock Returns, *The Journal of Finance*, vol. XLVII, n°5, pp. 1731-1764.
5. R. Sullivan, A. Timmerman and H. White (1999) Data-Snooping, Technical Trading Rule Performance, and the Bootstrap, *The Journal of Finance*, vol. 54, issue 5, pp. 1647-1691.
6. A. Lo, H. Mamaysky and J. Wang (2000) Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation, *The Journal of Finance*, vol. 55, n° 4, août 2000, sous presse.
7. A.P. Refenes, A.N. Burgess and Y. Bentz (1997) Neural Networks in Financial Engineering: A Study in Methodology, *IEEE Transactions on Neural Networks*, vol. 8, n°6, pp. 1222-1267.

8. P. Demartines, J. Héroult (1997) Curvilinear Component Analysis: A self-organizing neural network for nonlinear mapping of data sets, *IEEE Trans. on Neural Networks*, 8(1), pp. 148-154.
9. A.Lo., A.C. MackKinlay (1990) Data-snooping Biases in Tests of Financial Asset Pricing Models, *The Review of Financial Studies*, vol. 3, p. 431-467.
10. F. Takens (1985) On the numerical Determination of the dimension of an attractor, *Lecture Notes in Mathematics*, Vol. 1125, Springer-Verlag, pp. 99-106.
11. P. Grassberger and I. Procaccia (1983) Measuring the Strangeness of Strange Attractors, *Physica D*, 56, pp. 189-208.
12. P. Demartines and J. Héroult (1993) Vector Quantization and Projection, *Lecture Notes in Computer Sciences*, Vol. 686, Springer-Verlag, pp. 328-333.
13. T.J. Brailsford and R.W. Faff (1996) An evaluation of volatility forecasting techniques, *Journal of Banking and Finance*, vol. 20, pp. 419-438.
14. A. Lendasse, E. de Bodt, V. Wertz and M. Verleysen (2000) Nonlinear Financial Time Series Forecasting – Application to Bel20 Stock Market Index, *European Journal of Economic and Social Systems*, 14, n°1, pp 81-91.
15. M. Verleysen, E. de Bodt and A. Lendasse (1999) Forecasting Financial Time Series through Intrinsic Dimension Estimation and Nonlinear Data Projection, *Engineering Applications of Bio-inspired Artificial Neural Networks*, vol. II, 1607: pp. 596-605.