

## **Dimension reduction of technical indicators for the prediction of financial time series - Application to the BEL20 Market Index**

AMAURY LENDASSE<sup>1</sup>, JOHN LEE<sup>2</sup>, ÉRIC DE BODT<sup>3</sup>, *Member IEEE*,  
VINCENT WERTZ, *Member IEEE*, AND MICHEL VERLEYSEN, *Member IEEE*,

**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. 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. The methodology developed in the paper is validated on data from the BEL20 market index, using systematic cross-validation results.

**Classification Codes:** G00, G14.

### **1. Introduction**

Since the beginning of this century, the question of the predictability of financial series (at least of stock market prices) has been the subject of a highly controversial debate in

---

<sup>1</sup> Centre for Systems Engineering and Applied Mechanics, Engineering Faculty, Université catholique de Louvain, Louvain-la-Neuve, Belgium. E-mail: lendasse@auto.ucl.ac.be; wertz@auto.ucl.ac.be

<sup>2</sup> Microelectronics Laboratory, Engineering Faculty, Université catholique de Louvain, Louvain-la-Neuve, Belgium. E-mail: lee@dice.ucl.ac.be; verlaysen@dice.ucl.ac.be

<sup>3</sup> Institut d'Administration et de Gestion (IAG), Faculty of Economic, Social and Political Sciences, Université catholique de Louvain, Louvain-la-Neuve, Belgium and ESA - Université de Lille 2, France. E-mail: debodt@fin.ucl.ac.be

*Keywords:* Time series, prediction, exogenous indicators, nonlinear projection, principal component analysis, market index, cross-validation.

finance. Fama (1965), in its seminal paper, recalls the meaning of the random walk hypothesis<sup>4</sup> (first proposed by Bachelier, 1900) and presents different empirical tests of it. He concludes in those terms: “The main conclusion will be that the data seem to present consistent and strong support for the model. This implies, of course, that chart reading, though perhaps an interesting pastime, is of no real value to the stock market investor.” Many empirical works, mainly based on linear statistical tests, have conducted to the same conclusion in the years sixties and seventies, despite the heavy use of charts and technical indicators<sup>5</sup> by the professional community. However, as underlined by Campbell *et al.* (1997), “Recent econometric advances and empirical evidence seem to suggest that financial asset returns are predictable to some degree”. Among those works, three of them have constituted main advances in this field. Brock *et al.* (1992) test two popular technical trading rules on the Dow Jones market index on the period going from 1897 to 1986. They use a bootstrap methodology to validate their results and conclude that “their results provide strong support for the technical strategies”. Sullivan *et al.* (1999) propose new results on the same data set (extended with 10 new years of data). Their methodology, still relying on heavy use of bootstrap, allows avoiding (at least to some extent) the data-snooping bias<sup>6</sup> and is applied to a universe of 26 trading rules. They confirm that the results of Brock *et al.* (1992) stand up to inspection against data-snooping effects. The recent contribution of Lo *et al.* (2000), using a new approach based on nonparametric kernel regression, confirms that “several technical indicators do provide incremental information and may have some practical value”. On the basis of all of those empirical evidences, we will consider that there is some interest in trying to predict the evolution of financial asset prices, as do Refenes *et al.* (1997) in their introduction to the methods used in financial engineering.

When time series prediction is viewed as a regression problem, the inputs being past values of the series and exogenous variables, one may expect useful information (for the prediction of the series) to be contained in these inputs. Nevertheless, it is difficult to know if the information content of specific inputs is relevant, redundant or useless. Furthermore, it is well known that any regression method (in particular non-linear ones, like multi-layer perceptrons and radial-basis function networks), are difficult to use when the number of inputs is large. There is thus a strong interest in reducing the number of inputs; the question is how to reduce the number of inputs without losing relevant information.

---

<sup>4</sup> Following the random walk hypothesis, successive price changes are statistically independent from each other: the prices of a financial asset, if the market is informationally efficient, walk at random. The main justification of the hypothesis is the following: If the market is efficient, at any time, all the available information is reflected in the current price level. The price movements are therefore caused by the arrival of new information, which is by definition due to hazard.

<sup>5</sup> Technical indicators are numerical values calculated on the basis of past prices, volumes, and other market statistics and used to forecast future prices movements.

<sup>6</sup> The concept of data-snooping bias has been proposed by Lo and MacKinlay in 1990 (at least in the financial literature). This bias in the evaluation of forecasting method appears as soon as a data set is used more than once for purposes of inference of model selection. It can be illustrated by the “Give me the data, I will give you the model” sentence.

In this paper, we will focus on this question and we will compare the classical linear data compression approach with a new non-linear one. The objective will be to keep as much as possible the information contained in the initial inputs, while reducing as much as possible the number of new “constructed” variables (or the dimension of the projection space). The new “constructed” variables are then used as input to the prediction algorithm, as illustrated in Figure 1.

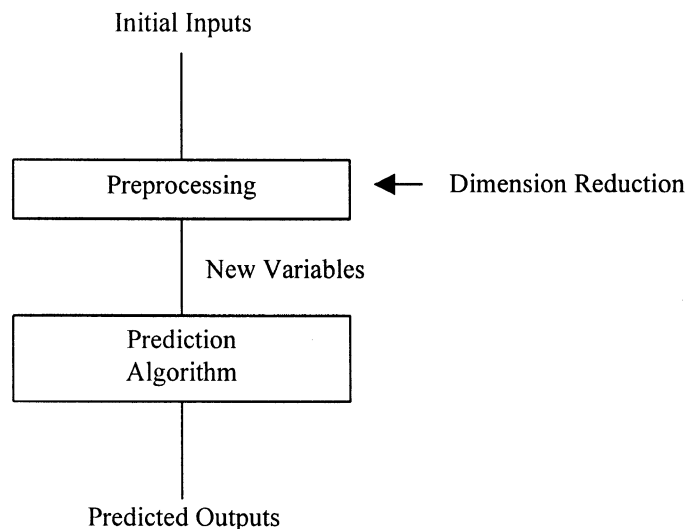


Fig. 1. The two steps of the methodology.

To perform the transformation between the initial inputs and the new variables, we may choose to use a linear method or a non-linear one. In this paper, we will use Principal Component Analysis as a linear transformation and Curvilinear Component Analysis (Lo and MackKinley, 1990) as a non-linear one.

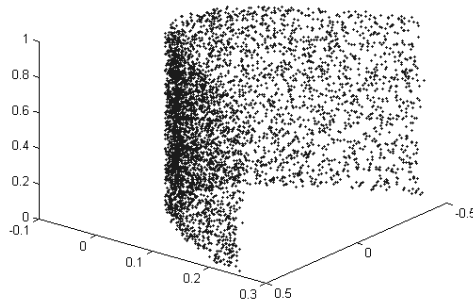
In the following, we will successively present possible projection methods, evaluation criteria that may be used to assess the quality of the projection, and the methodology that we used to perform the projection on real data sets. The methodology is of particular importance, as most of the methods first require the adjustment of parameters, and secondly could be evaluated through different criteria. Establishing a sound methodology, based on objective criteria and applicable in various fields, is the key point of this paper.

## 2. Dimension reduction

### 2.1. Intrinsic Dimension

First, it is important to evaluate the projection dimension, *i.e.* the dimension of the space of new variables. If the estimation of this dimension is too small, information will be lost

in the projection. If it is too large, the usefulness of the method is lost. To evaluate this dimension, the concept of intrinsic dimension is used. The intrinsic dimension is the effective number of degrees of freedom of a set; a definition of the intrinsic dimension can be found in Takens (1985). This concept is presented here with the well-known horseshoe distribution (Fig. 2): for this data set, the intrinsic dimension is equal to two as two degrees of freedom are sufficient to uniquely determine any data in the set, although the data live in  $\mathbb{R}^3$ . The computation of the intrinsic dimension is explained in Grassberger and Procaccia (1983), but its determination remains very difficult to apply, not to say approximate, for high dimensional data sets. Therefore, the intrinsic dimension will be only considered here as a rough approximation of the dimension that should be used for the projection.



*Fig. 2. Horseshoe distribution.*

## **2.2. Projection Methods**

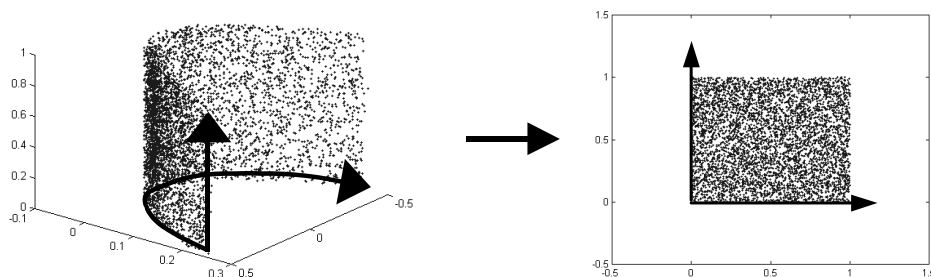
### *2.2.1. Principal Component Analysis*

Principal Component Analysis (PCA) of a set of data is based on the computation of the covariance matrix of this set. The singular value decomposition of this matrix provides the desired linear transformation and it remains to the analyst to choose the number of principal components to keep (that is to say, the dimension of the projection space). Hence, in this case, the intrinsic dimension does not need to be evaluated. As PCA is a linear transformation, any dimension reasonably estimated by PCA (for example by keeping 95 or 99% of the variance of the initial set) is usually larger than the intrinsic dimension of the set.

### *2.2.2. Curvilinear Component Analysis*

This nonlinear extension of the Principal Component Analysis (also named Vector Quantization and Projection, see Demartines and Héroult (1997 and 1993)) spreads out the manifold that contains the data and projects it from a high dimensional space to a smaller dimensional one. The projection of the horseshoe distribution carried out by Curvilinear Component Analysis (CCA) is illustrated in Figure 3. The details of the algorithm will not be explained here but it must be mentioned that the first step consists in a

quantization of the input space using centroids. The quantization provides a partition of this space in some clusters in which the data manifold can be considered as linear; Curvilinear Component Analysis may thus be considered as a piecewise Principal Component Analysis. The number of centroids is the important parameter that one must select to perform an adequate projection. A precise way to choose this parameter will be explained in the described below an applied to the BEL20 Market Index prediction example.



*Fig. 3. Projection carried out by CCA from  $R^3$  to  $R^2$ .*

### 2.3. Evaluation Criteria

The goal of the projection is the reduction of the number of inputs (for the prediction method) and the preservation of the initial information. Several ways can be used to evaluate how the projection is successful. Looking to Figure 1 again, one can try to measure if the new variables contain the same information as the initial inputs, without taking care of the subsequent prediction algorithm. The quality of the projection can also be evaluated by looking at the end of the chain, *i.e.* the predicted outputs. The advantage of this last approach is that the ultimate goal itself is measured, instead of an intermediate criterion that is not guaranteed to be adequate; the drawback is that any shortcoming in the prediction method itself (for example difficulties in the convergence of non-linear methods) will be measured together with the quality of the projection. The first criterion detailed below belongs to the first approach, while the three last ones belong to the second approach.

#### 2.3.1. Deprojection Measure

The projection by PCA or CCA is a reversible operation; nevertheless the result of a deprojection (inverse of the projection) will not correspond to the initial inputs. The mean distance between the initial inputs and the results of the deprojection is called the deprojection measure; it should be as small as possible and should correspond in ideal situations to the noise surrounding the information. If the dimension of the projection space is decreased below the intrinsic dimension, then the deprojection measure will increase; this will validate or invalidate the rough intrinsic dimension estimation. The deprojection Measure is also a way to rank several projection methods, as Principal Component Analysis and Curvilinear Component Analysis, used with the same projection dimension. The main advantage of this criterion is that it is totally independent of

the prediction method used afterwards. Unfortunately, this advantage has a counterpart: there is no guarantee that a projection dimension chosen according to this measure will be optimal when the best prediction outputs are looked for.

### 2.3.2. Determination Coefficient

This criterion is often used especially in financial applications, in particular for the volatility forecasting (see for example Brailsford and Faff, 1996). Contrary to the previous criterion, it requires the prediction of the target variable and measures the difference between the target and the prediction. Let  $y_t$  be the target (output to be predicted),  $\hat{y}_t$  its prediction,  $\bar{y}$  the average of  $y_t$ . In order to evaluate how the predictions correspond to the targets, we compute a linear regression according to equation (1) on the whole data set:

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

A and B are the coefficients of this regression. We then compute the determination coefficient  $r^2$  according to

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

The  $r^2$  coefficient measures how the points in the  $y_t$  vs.  $\hat{y}_t$  plane are close to the straight line defined by the A and B coefficients. The more  $r^2$  is close to one, the more the points are close to the line and the approximation is considered as correct. This criterion, while used a lot in empirical works, seems not really appropriate in financial applications. Indeed, financial data are very noisy; the deterministic part (*i.e.* the part that can be predicted) is generally far less important than the noisy part. Hence, the standard deviation of  $\hat{y}_t$  will be lower than the standard deviation of  $y_t$ , resulting in a badly conditioned regression (1). The use of the determination coefficient is not robust in this kind of application, as it will be illustrated on the BEL20 market index application.

### 2.3.3. Mean Square Error

The Mean Square Error (*MSE*) is defined by

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

It must be made as small as possible for a successful prediction. Unfortunately, in financial applications the Mean Square Error is usually very high since the series are very noisy.

#### 2.3.4. Percentage of Correct Approximations of Sign

As the amplitude of financial time series is always difficult to predict, it is interesting to calculate the Percentage of Correct Approximations of Sign (*PCAS*) defined by

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

The *PCAS* has to be as large as possible. It has also to verify

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

otherwise, this would mean that the approximation would be worst than a constant approximation (either 100% of positive or of negative signs).

### 2.4. Methodology

The procedure used to project the initial inputs on new relevant variables is described here, together with the tests that determine if the projection improves the results of prediction, and if a nonlinear projection must be used instead of classical linear one.

In the following, we denote by:

- $u_t^i$ ,  $1 \leq i \leq n$ , the technical indicators or exogenous variables used in the prediction algorithm; note that selected past values of the series may belong to the technical indicators;
- $z_t$  the input vector containing the values of the series and of the exogenous variables at time  $t$ :

$$z_t = [u_t^1 u_t^2 \dots u_t^n y_t];$$

- $T$  the total number of samples available in the series ( $1 \leq t \leq T$ );
- $D$  the intrinsic dimension of the learning set.

#### 2.4.1. Mixture

The input vectors  $z_t$  are randomly mixed according to a draw without replacement  $f(t)$  law:

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

### 2.4.2. Learning and Validation sets

The data must be divided in learning and validation sets. The validation set is used to check the results obtained on the learning set and justify that no overfitting has occurred. The validation results can then be used to tune parameters (like the number of centroids or the projection dimension) of the learning method. The division is made between the learning set  $z^L$  and the validation set  $z^V$  according to

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

and

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

The  $\alpha$  parameter ( $\alpha T$  needs to be an integer) must be carefully chosen: it must be large enough to ensure an efficient number of elements in the learning set while paying attention to keep a sufficient number of elements in the validation set. Some advices will be given further on how to choose this parameter.

### 2.4.3. Normalization of Technical Indicators

In the learning set, each of the  $n$  indicators  $u_t^{iL}$  ( $1 \leq i \leq n$ ) must be normalized according to

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

where  $\bar{u}^{iL} = \frac{\sum_t u_t^{iL}}{\sum_t 1}$  is the mean value of indicator  $u_t^{iL}$  evaluated on the learning set,

$\sigma^{iL} = \sqrt{\frac{\sum_t (u_t^{iL} - \bar{u}^{iL})^2}{\sum_t 1}}$  is its standard deviation and  $\tilde{u}_t^{iL}$  is the normalized value.



As the normalization is part of the learning, points from the validation set are not included in the computation of the mean and standard deviation. The validation set is thus normalized by using the same parameters as those computed on the learning set:

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

#### 2.4.4. Intrinsic Dimension

The intrinsic dimension  $d$  of the technical indicators of the learning set is computed and will be used as a first rough approximation of the projection dimension.

#### 2.4.5. First Principal Component Analysis

CCA may be directly used as projection method. Nevertheless, it must be mentioned that, in practical situations, a preliminary PCA greatly helps the CCA learning. A Principal Component Analysis of the technical indicators in the learning set is hence performed. The dimension  $n_1$  is chosen as the minimal dimension for which the loss of information after PCA is negligible. The technical indicators of the learning and validation sets are projected from a  $n$ -dimensional space to a  $n_1$ -dimensional space according to the parameters defined by the Principal Component Analysis on the learning set.

#### 2.4.6. First Curvilinear Component Analysis

A first Curvilinear Component Analysis is then performed on the  $n_1$ -dimensional space, reducing it to a  $d$ -dimensional space. The Curvilinear Component Analysis algorithm is used, taking only the learning set into account. As a first approximation, the number of centroids is chosen so that, in average, the local intrinsic dimension in each cluster of the quantization is equal to  $d$ .

#### 2.4.7. Optimization of $n_1$

The projection from a  $n$ -dimensional space to a  $d$ -dimensional space is repeated using firstly Principal Component Analysis and secondly Curvilinear Component Analysis (PCA + CCA), but the intermediate dimension  $n_1$  is varied around its initial value. Each of the projections is also applied on the validation set. The deprojection measure obtained on the validation set is studied. Usually, this measure has a minimum (or a bend) and this minimum (or bend) is chosen as the optimal value for  $n_1$ .

#### 2.4.8. Error Criteria on the validation set

New projections are computed from the  $n$ -dimensional to the  $d$ -dimensional space by varying the projection method (PCA-only or PCA + CCA), the number of centroids (in the nonlinear case) and the projection dimension  $d$ . In each case, the deprojection measure on the validation set is computed. Then, the determination coefficient, the mean

square error and the percentage of correct approximations of sign are evaluated on the validation set. For the purpose of computing these three last criteria, we need the predicted outputs  $\hat{y}_i$ ; these values can be computed by any prediction algorithm, including linear and non-linear (MLP, RBF, etc.) ones. In the example developed in Section 3, we will limit ourselves to a basic linear prediction model, the aim of this paper being to compare linear and non-linear data compression techniques in the field of financial applications.

#### 2.4.9. Cross-validation

Operations 2.4.1 to 2.4.8 are repeated  $k$  times; the four measures and error criteria computed in 2.4.8 are averaged. Several tuning parameters have still to be optimized. Since a global optimization is hardly feasible, we propose to determine first the optimal number of centroids (used in the CCA) by minimizing the deprojection measure, which is independent of the prediction model. The other errors are used to choose the final projection dimension ( $d_L$  in the PCA-only case and  $d_{NL}$  in the PCA + CCA one). In fact, the projection dimension cannot be calculated without the knowledge of the number of centroids. The choice between the determination coefficient, the mean square error or the percentage of correct approximations of sign is done according to the goal of the time series prediction application. The analysis of the influence of these criteria on  $d_L$  or  $d_{NL}$  will be detailed on the BEL20 Market Index application. The number  $k$  of repetitions must be large enough to obtain smooth graphs (criteria vs. dimension); moreover, the larger  $\alpha$  (splitting coefficient between learning and validation sets), the larger  $k$  must be large too (if the learning set is large, the quality of the learning phase is improved but the validation set is not large enough to obtain reliable results). If the computing time is not too long, it is thus appropriate to increase  $\alpha$  and  $k$  to obtain simultaneously better analysis and better performances.

### 3. Application to the BEL20 Market Index

The methodology is tested on the BEL20 Market Index<sup>7</sup> from December 1st, 1987 to February 14th, 1998 (2663 daily observations). This index is represented in Figure 4. Its choice has been motivated by two main reasons: first, the Belgian Stock Market has not been the subject of numerous studies in the field of technical analysis. As such, the risk of data-snooping bias is reduced. Second, being a small market, the level of efficiency is probably not so high as with the major international stock markets, offering therefore more opportunities to the use of technical analysis approaches. 42 technical indicators have been used (they are described in Tab. 1). These indicators are strongly correlated (the condition number of the covariance matrix is almost infinite, the smallest eigenvalue being almost equal to zero).

---

<sup>7</sup> The twenty most representative shares of the Belgian Stock Market compose this market index.

**TABLE I.** Short description of the 42 technical indicators.

Indicators	Number
Lagged value of the returns	4
Lagged value of the variance of the returns	3
Moving average of the returns	3
Slope of the lagged value of the prices	3
Lagged value of the short and long interest rate level	3
Lagged value of the variance of the short and long interest rate	3
Slope of the lagged value of the prices	3
Technical indicators based on the SP 500 market index	12
Technical indicators based on the currency exchange rate	8

Predicting this series is not easy: the phenomenon is not stationary<sup>8</sup>. Furthermore, the correlation coefficient correlation between the current price and the next day price is 0.999, while the one between the current return and the next day return is 0.099; two results that seem at first sight to confirm the random walk hypothesis. The daily return  $y_t$  of this market is calculated according to

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

and is represented in Figure 5. The mean of  $y_t$  is close to zero.

The methodology described in the previous section is then applied.

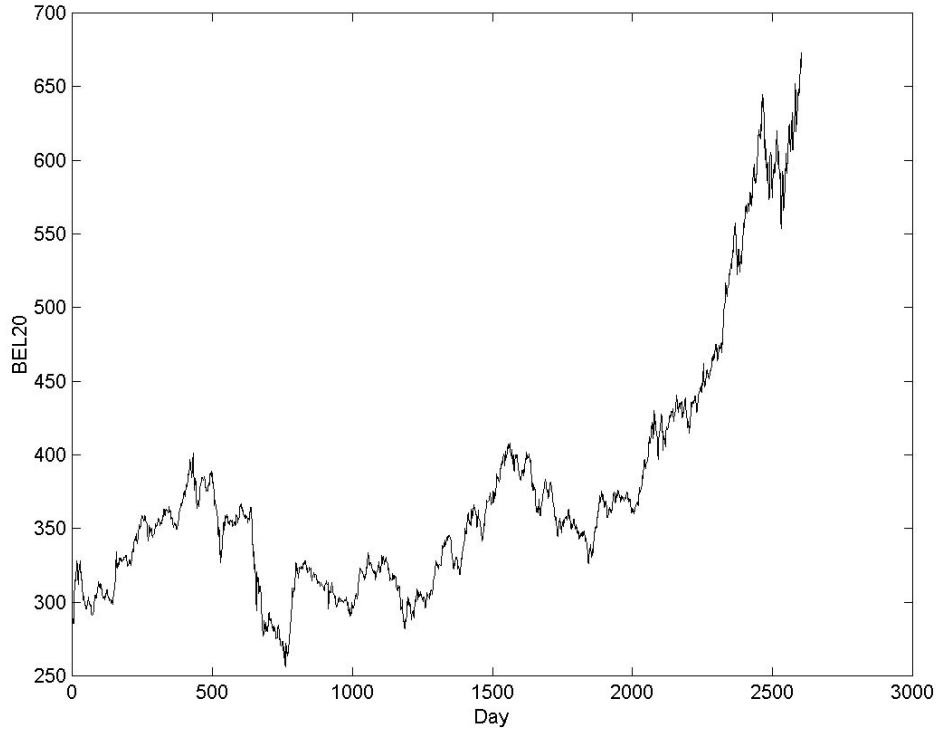
- Stages 2.4.1 to 2.4.3 are performed.
- Stage 2.4.4: The intrinsic dimension of the technical indicators is 7.
- Stage 2.4.5: The result of the Principal Component Analysis is shown in Figure 6. This result is the percentage of information kept by a Principal Component Analysis versus the projection dimension.

The initial projection dimension  $n_1$  is chosen equal to 30; a linear projection from a 42-dimensional space to a 30-dimensional space is performed (keeping 98% of the variance of the initial set).

- Stage 2.4.6: A nonlinear projection (CCA) is performed from a 30-dimensional space to a 7-dimensional space. An initial number of centroids is computed: 125.
- Stage 2.4.7: The projection from a 42-dimensional space to a 7-dimensional space is repeated using Principal Component Analysis and Curvilinear Component Analysis; the intermediate dimension  $n_1$  is swept around the initial value (30). The deprojection

---

<sup>8</sup> An augmented Dickey-Fuller test confirms the presence of a unit root at a level of confidence of 1%.



*Fig. 4. BEL20 Market Index from December 1st, 1987 to February 14th, 1998.*

measure obtained on the validation set is represented in Figure 7. The optimal value of  $n1$  corresponds to the bend on the curve: 27.

- Stages 2.4.8 and 2.4.9: First, the number of centroids is optimized, using the deprojection measure. It turns out that the initial value (125) is confirmed. Then, a new set of projections is performed from the 42-dimensional space to the  $d$ -dimensional space when varying  $d$  around 7. Note that, for the sake of validation of our methodology, the computations have been done for  $d = 2$  to 24. Each projection has been repeated 20 times ( $k = 20$ ), while varying the learning and validation sets. For each trial, a linear prediction model of  $y_t$  based on the projected data is estimated on the learning set. Then, the determination coefficient, the mean square error and the percentage of correct approximations of sign are calculated on the validation set; these three criteria are shown in Figures 8, 9 and 10 respectively.

#### 4. Discussion

It can be noticed that the PCA-only method gives slightly better results than the PCA + CCA method. We might have expected the opposite, or at least equivalent results. In fact, this phenomenon may be explained by a too small number of data and by the

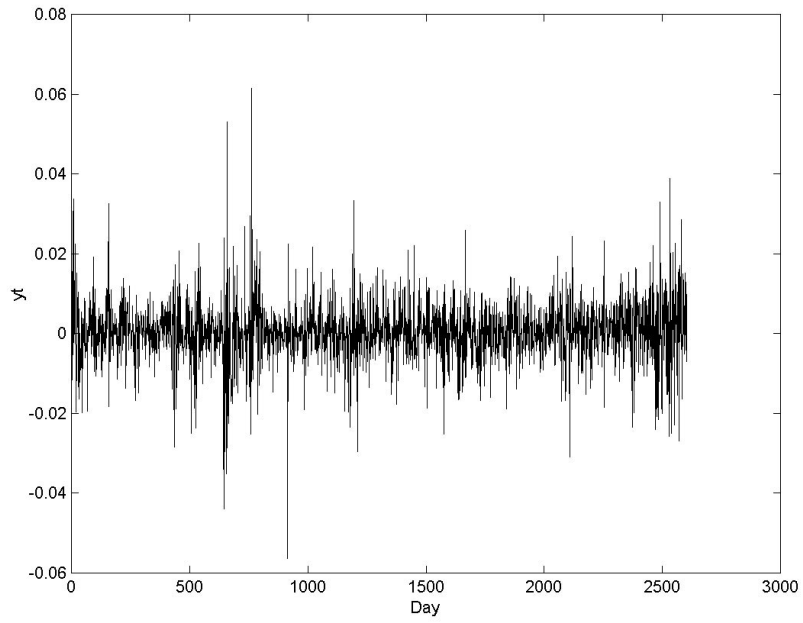


Fig. 5. Daily return of BEL20 Market Index from December 1st, 1987 to February 14th, 1998.

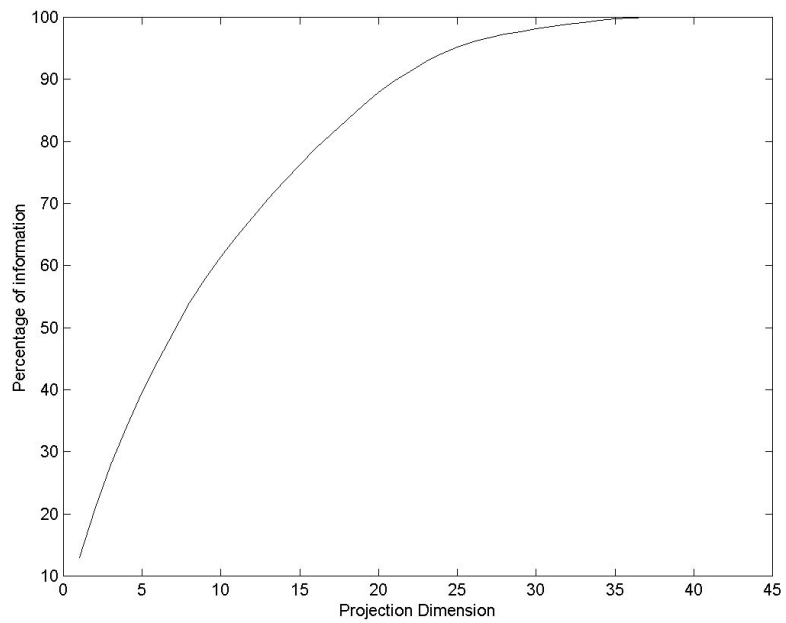
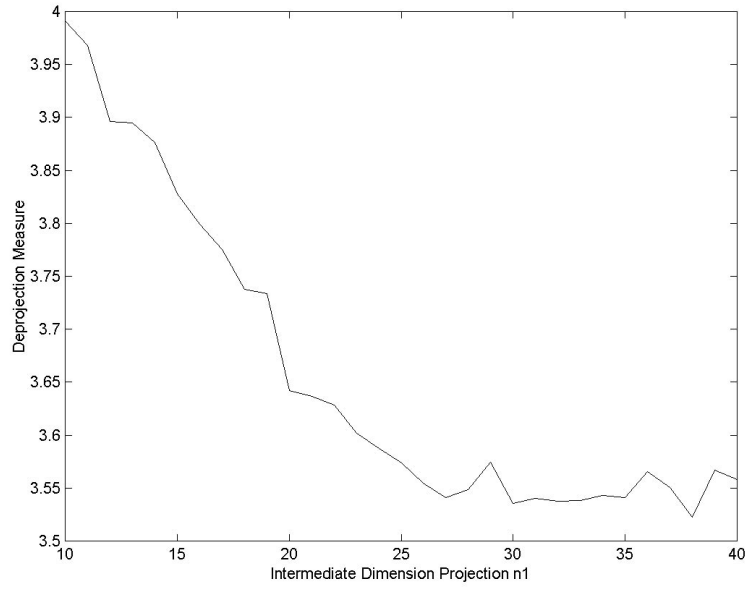
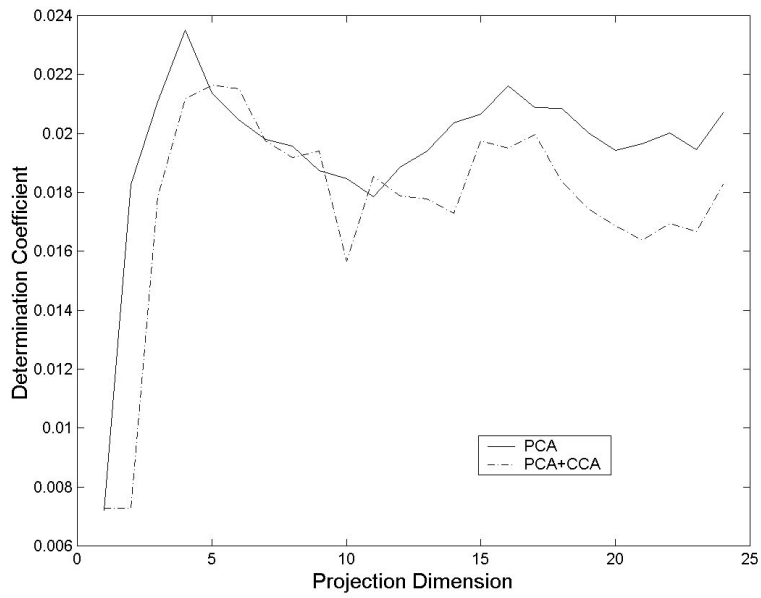


Fig. 6. Percentage of information kept by a PCA versus the projection dimension.



**Fig. 7.** The deprojection measure obtained on the validation set versus the intermediate projection dimension  $n_1$ .



**Fig. 8.** Determination coefficient on the validation set versus the projection dimension.

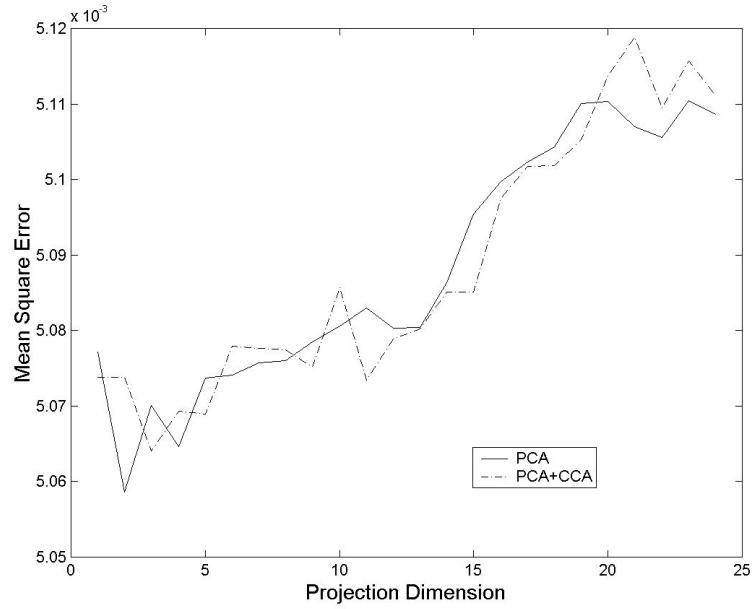


Fig. 9. Mean square error on the validation set versus the projection dimension.

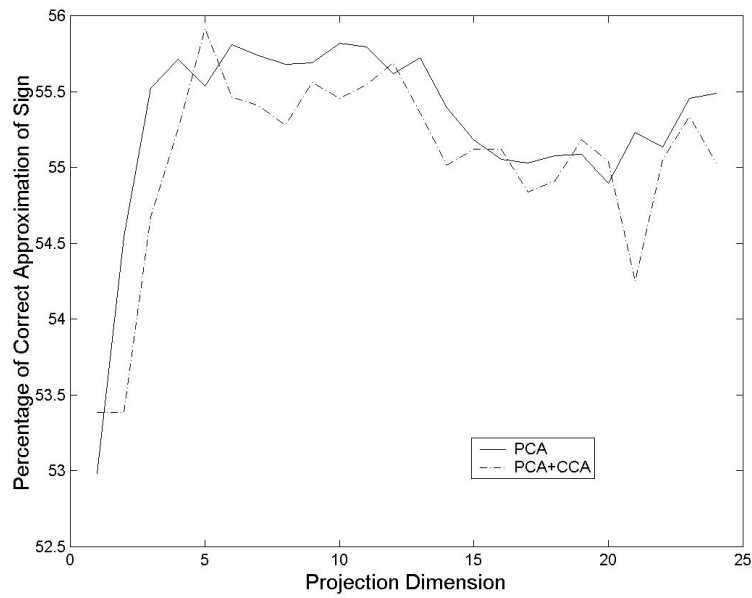
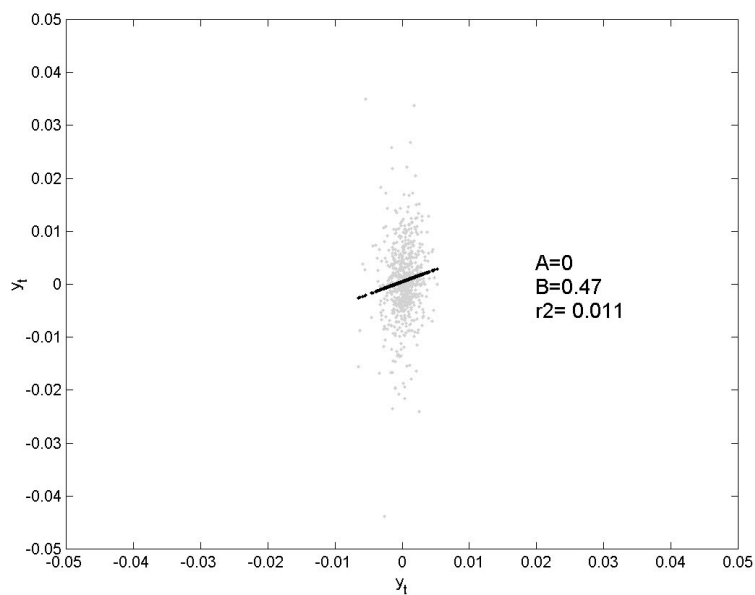
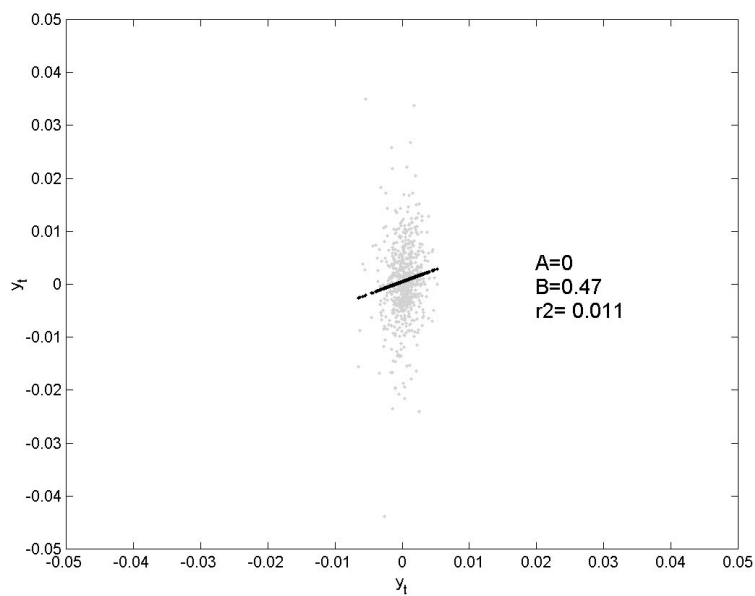


Fig. 10. Percentage of correct approximations of sign on the validation set versus the projection dimension.



*Fig. 11.  $y_t$  versus  $\hat{y}_t$  in gray and the regression result in black.*



*Fig. 12. Modified  $y_t$  versus  $\hat{y}_t$  in gray and the new regression result in black.*



important noise contained in the series. Another explanation is the use of a linear model for prediction. One may expect that a nonlinear prediction method would give improved results, especially when the initial inputs are preprocessed in a nonlinear way; some results obtained with a nonlinear model may be found in Lendasse *et al.* (2000). Nevertheless, it must be mentioned that problems related to a difficult convergence of nonlinear models add to the difficulty of choosing the parameters in our methodology, making the results more difficult to illustrate. Furthermore, the number of trials that must be averaged in the cross-validation stage has to be increased in order to reduce the effects of local minima with nonlinear predictors.

The best projection dimension depends on the criterion: 4 with the determination coefficient, 2 with the mean square error and 4 to 6 with the percentage of correct approximations of sign; this last one is the only criterion for which a nonlinear projection gives improved results. Referring to (5), it must be mentioned that this series increases about 53% of the time. It must be stressed that obtaining results on the forecasting of the sign of daily returns time series is particularly important. The anticipation of the orientation of the market is at the basis of any market timing strategies, which justify the technical analysis approach. Furthermore, as mentioned above, a non-linear prediction algorithm (Verleysen *et al.*, 1999) would still improve the results (the level of improvement remains limited here because a simple linear prediction model is used).

To obtain correct results with the determination coefficient, the number of trials must be large because this coefficient is very sensitive to small perturbations on the data. To calculate it, a regression between  $y_t$  and  $\hat{y}_t$  is done according to (1). An example of this regression taken from our application is represented in Figure 11. The regression coefficients are  $A = 0$  and  $B = 0.47$ ; the determination coefficient is  $r^2 = 0.011$ . If only one data is changed (see the circle in Fig. 12), the regression coefficients become  $A = 0$  and  $B = 0.34$ , and the determination coefficient becomes  $r^2 = 0.006$ . The reason of this ill-conditioned result is the small amplitude of  $\hat{y}_t$  compared to the amplitude of  $y_t$ . Unfortunately, a lot of financial applications present such a peculiarity. So, our point of view is that the use of the determination coefficient criterion should be avoided in financial applications as a quality of forecasting criterion.

## 5. Conclusions

Even if it is considered that there is some interest in trying to predict the evolution of financial asset prices, the use of a large number of technical indicators remains difficult with any prediction tool. This paper shows how to use – possibly nonlinear – data compression techniques to reduce the number of technical indicators used for the prediction.

As the use of non-linear 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. It is also shown that the use of the determination coefficient is dangerous in the specific case of financial predictions.

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.

*Acknowledgement.* Michel Verleysen is a Research Associate of the Belgian National Fund for Scientific Research (FNRS). This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with its authors.

## References

- Bachelier L., *Théorie de la spéculation*, Gauthier-Villars, Paris.
- Brailsford T.J., Faff R.W. (1996) An evaluation of volatility forecasting techniques, *Journal of Banking and Finance* **20**, pp. 419-438.
- Brock W., Lakonishok J., Le Baron B. (1992) Simple technical trading rules and the stochastic properties of stock returns, *The Journal of Finance* **XLVII**, pp. 1731-1764.
- Campbell J., Lo A., MacKinlay A.C. (1997) *The Econometrics of Financial Markets*. Princeton University Press, Princeton.
- Demartines P., Héroult J. (1993) Vector quantization and projection, *Lecture Notes in Computer Sciences* **686**, Springer-Verlag, Berlin, pp. 328-333.
- Demartines P., Héroult J. (1997) Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets, *IEEE Trans. on Neural Networks* **8**, pp. 148-154.
- Fama E. (1900) The Behavior of stock market prices, *Journal of Business* **38**, pp. 34-105, (1965).
- Grassberger P., Procaccia I. (1983) Measuring the strangeness of strange attractors, *Physica D* **56**, pp. 189-208.
- Lendasse A., de Bodt E., Wertz V., Verleysen M. (2000) Nonlinear financial time series forecasting – Application to Bel20 stock market index, *European Journal of Economic and Social Systems* **14**, pp 81-91.
- Lo A., MacKinlay A.C. (1990) Data-snooping biases in tests of financial asset pricing models, *The Review of Financial Studies* **3**, pp. 431-467.
- Lo A., Mamaysky H., Wang J. (2000) Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation, *The Journal of Finance* **55**, pp. 1705-1770.
- Refenes A.P., Burgess A.N., Bentz Y. (1997) Neural networks in financial engineering: A study in methodology, *IEEE Transactions on Neural Networks* **8**, pp. 1222-1267.
- Sullivan R., Timmerman A., White H. (1999) Data-snooping, technical trading rule performance, and the bootstrap, *The Journal of Finance* **54**, pp. 1647-1691.
- Takens F. (1985) On the numerical determination of the dimension of an attractor, *Lecture Notes in Mathematics* **1125**, Springer-Verlag, Berlin, pp. 99-106.
- Verleysen M., de Bodt E., Lendasse A. (1999) Forecasting financial time series through intrinsic dimension estimation and nonlinear data projection, *Engineering Applications of Bio-inspired Artificial Neural Networks II*, pp. 596-605.