European Journal of Operational Research 207 (2010) 1702–1716

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Pairs trading and outranking: The multi-step-ahead forecasting case

Nicolas Huck [*,1]

ESSCA School of Management, 1, Rue Lakanal, 49003 Angers Cedex 01, France

ARTICLE INFO

Article history:
Received 16 September 2009
Accepted 30 June 2010
Available online 6 July 2010

Keywords:
Finance
Pairs trading
Multi-criteria decision-aiding
Electre III
Multi-step-ahead forecasting

ABSTRACT

Pairs trading is a popular speculation strategy. Several implementation methods are proposed in the literature: they can be based on a distance criterion or on co-integration. This article extends previous research in another direction: the combination of forecasting techniques (Neural Networks) and multi-criteria decision making methods (Electre III). The key contribution of this paper is the introduction of multi-step-ahead forecasts. It leads to major changes in the trading system and raises new empirical and methodological questions. The results of an application based on S&P 100 Index stocks are promising: this methodology could be a powerful tool for pairs selection in a highly non-linear environment.

1. Introduction

Pairs trading is one of Wall Street's quantitative methods of speculation which dates back to the mid-1980s. At this time, the first automated trading systems were developed by Nunzio Tartaglia, a Wall Street quantitative analyst at Morgan Stanley, and his group of mathematicians, physicists and computer scientists. Pairs trading was one of the key features of these systems. They generated hundreds of millions of dollars in profit until 1989 when the group disbanded. As underlined by Engelberg et al. (2008), financial economists have long been interested in understanding the profitability underlying various forms of statistical arbitrage because it is a way to question market efficiency and the behaviour of stock prices.

In its most common form, pairs trading involves forming a portfolio of two related stocks whose relative pricing departs from its "equilibrium" state. It is linked to co-integration (Bossaerts, 1988) and correlation in stock prices, mean reversion (DeBondt and Thaler, 1985), contrarian strategies (Jegadeesh and Titman, 1993) and also to the law of the one price. Pairs trading is one way to select and build stocks for a long/short dollar neutral portfolio. In reality, even such strategies require some outlay, if only to meet margin calls and brokerage fees.

By going long on the relatively undervalued stock and short on the relatively overvalued stock, a profit may be made by unwinding the position upon "convergence" of the spread. The success of pairs trading, especially statistical arbitrages, depends heavily on the modelling and forecasting of the spread time series. The ability to anticipate the "direction" of this spread is a key point. As observed in Leitch and Tanner (1995), the ability to forecast direction is essential for the success of trading strategies.

Whilst the strategy appears simple and has, in fact, been widely implemented by traders and hedge funds, owing to the proprietary nature of the area there has been a limited amount of published research until a recent burst of interest in the last few years. This includes several books on the subject (Erhman, 2006; Whistler, 2004). The key points of the empirical studies dealing with pairs trading are as follows: these strategies exhibit, at least according to the first studies, large and significant risk-adjusted returns which are not due exclusively to a short-term reversion phenomenon. The literature can be divided into three main categories according to the methodology they discussed to select and trade pairs:

- The distance approach.
- The modelling of mean reversion.
- Combined forecasts and Multi-Criteria Decision Methods (MCDM).

The first category of papers includes Andrade et al. (2005), Gatev et al. (1999, 2006), Engelberg et al. (2008), Papadakis and Wisocky (2008) and Do and Faff (2008).

The Gatev et al. (1999, 2006) papers are the most cited papers on pairs trading. Like many traders, they envision a simple algorithm for choosing pairs. The rule follows the general outline of first "find stocks that move together" then "take a long short position when

* Tel.: +33 2 41734747.
E-mail address: Nicolas.Huck@essca.fr

[1] The author has benefited from helpful discussions with J.L. Bertrand, X. Brusset, S. Freulon and C. Montagné. He also wishes to thank three anonymous reviewers for valuable comments and suggestions.

0377-2217/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2010.06.043

they diverge". These two papers include a trading system and the management of a portfolio and consider a very large number of stocks using the CRSP database (about 2300 securities).

Gatev et al. (1999, 2006) also show that pairs trading after costs can be profitable. The first version has been known for about 10 years and Do and Faff (2008), replicating the Gatev et al. (2006) methodology with more recent data, report that the results of this strategy are declining. Engelberg et al. (2008) indicate that the profitability of this strategy decreases exponentially over time. Furthermore, a large part of the profits can be made in the first 10 days.

This approach requires the selection and the trading steps to be parameterized in some way. It uses a simple standard deviation strategy to select and trade stocks. With daily data, they form pairs over a 12 month period and trade them over the next six months. Among the candidates chosen during the first stage, if prices diverge by more than two standard deviations, a long/short position is open. As underlined by the authors, both the 12 month and six month periods are chosen arbitrarily.

The distance approach merely exploits the statistical relationship of a pair at a price level. As the approach is normative and economic free, it has the advantage of not being exposed to model mis-specification and mis-estimation. On the other hand, this strategy lacks forecasting ability: if a "divergence" is observed, the assumption is that prices should converge in the future because of the law of the one price. When equilibrium is reached or at the end of the six months trading period, the positions are closed out.

Two methodologies, coming from an econometric standpoint, attempt to model the expected mean reversion phenomenon.

The co-integration approach can be an attempt to parametrize pairs trading, by exploring the possibility of co-integration (Engle and Granger, 1987; Johansen, 1988). Works using this methodology include Vidyamurthy (2004), Lin et al. (2006) and Galenko et al. (2007).

Generally speaking, the framework is as follows: first, choose two co-integrated stock price series, then open a long/short position when stocks deviate from their long term equilibrium and finally, close the position after convergence or at the end of the trading period.

Consider two shares whose prices are integrated of order 1. P_i^t refers to the price[2] of the ith asset called A_i at time t.

If the share prices P_1^t and P_2^t are co-integrated, co-integration coefficients 1 and β exist so that a co-integration relationship can be constructed as follows:

$$P_1^t - \beta, \quad P_2^t = \epsilon_t, \tag{1}$$

where ϵ_t is a stationary process. When a divergence (based on the standard deviation of ϵ_t) from the equilibrium state is observed, the trading involves buying one share 1 and selling β shares 2 as in Lin et al. (2006). In that case, the strategy is not perfectly dollar neutral, which is a major difference compared to other frameworks.

A stochastic approach is used by Elliott et al. (2005) and Do et al. (2006). In a continuous setting, the first of these articles models the difference between the two stock prices using a mean reverting Gaussian Markov chain model. The second studies the behaviour of the series at the return level.

Most of the time, this group of works suffers, from the financial and practical point of view, from the fact that the real data application, if there is one, considers only a very limited number of stocks. The articles based on distance criteria or on combined forecasts and MCDM propose fairly developed trading systems and the management of a portfolio over a period of many years. On the other hand, these mathematical approaches can provide analytical results about the supposed speed of convergence of a given series, the first time passage or the optimal threshold for opening and closing positions.

The present article belongs to the third category of papers according to the segmentation given above and follows Huck (2009). This framework is general and flexible and can be seen as a sort of combined forecast which is specially designed for pairs trading. The combination of forecasts can be done using Multi-Criteria Decision Methods (MCDM). The approach used in Huck (2009), from a non-technical point of view, can be described as follows:

- Consider n stocks, $\frac{n*(n-1)}{2}$ different pairs can thus be formed.
- For each pair, forecast the difference of return the stocks should have at the end of the trading period.
- Define an anti-symmetric matrix of size n. Each element of the matrix will be the anticipated spread computed during the forecasting step.
- This matrix will be the input of an MCDM in order to rank stocks in terms of anticipated returns.
- The strategic behaviour is thus quite simple:
  - Firstly, buy the first stocks of the ranking and sell the last ones (for one stock that is bought, an equal dollar value of another stock is sold at the same time). All pairs have the same weight in the portfolio. The strategy is, by construction, dollar neutral.
  - Secondly, at the end of the trading period, close all positions.
  - If the ranking was relevant, a profit is made.

In brief, the method is based on three phases: forecasting, ranking and trading. This framework differs from the others on one key point: it has been developed without reference to any equilibrium model. This methodology is positive whereas most of the literature is clearly normative with reference to an equilibrium state. This approach is nevertheless complementary with other existing techniques. The framework discussed in this paper provides much more trading possibilities:

- Even if we know that prices "diverge" according to an equilibrium based strategy, which would indicate that a mean reversion phenomenon is expected in the short or in the long term, a forecasting based approach could indicate an expansion of the divergence in the short term.
- Furthermore, this approach could detect the "birth" of the divergence, as long as an equilibrium state exists, which is a trading opportunity the normative frameworks cannot consider.

The approach developed in this article is more short term oriented than the rest of the literature (the pairs may stay open for 50 or 100 days). It is not necessarily a problem because, as underlined in Engelberg et al. (2008), much of the profits come in the first 10 days.

Data snooping is a major concern in this type of study and will have an impact on the design of the method and on the trading system: for example, during the forecasting step, a unique specification of the forecasting method is used for the different information sets.

The main empirical conclusions of Huck (2009) were the following:

- This work, based on weekly data and limited to one step-ahead forecasts and to a one week holding period for each pair, shows that significant and positive spreads can be captured.
- The lower the number of pairs selected in the portfolio, the better the ability to anticipate direction and the excess return. The method could provide a tool for selecting some pairs among a large number of securities.

[2] In the following, the return of stock i at time t will be written as r_i^t. $r_i^{t+h|t}$ will be the cumulative return between dates t and $t+h$.

The contributions of the present article are multiple, being both methodological and empirical. Multi-step forecasts (up to four step-ahead) are performed. In this part of the methodology, we still consider weekly data which means the holding period may last for four weeks. The use of multi-step forecasts leads to important changes in the trading system: the management of the portfolio will be more complex and detailed. Closing thresholds are also introduced and positions can be unwound each day if a threshold is reached.

One conscious weakness of this paper is the absence of a direct comparison between our framework and, for example, the distance based approach. This is because the method discussed in this paper is still under development and we wish to focus on it: the introduction of multi-step forecasts is a serious improvement but more work needs to be done before a fully satisfactory framework can be produced. We are aware that the possible extensions briefly discussed in this paper (length of the trading period, forecast combination, etc.) will increase the complexity of the system, but we believe they are necessary before a relevant comparison/competition can be performed. The introduction of new elements will always be clearly justified from a trading, financial or methodological point of view in order to keep the system from being a black box.

The remainder of the paper is as follows: Section 2 describes the methodology and its different steps. The article focuses on the way some well-known techniques (Electre III and Neural Networks) can be part of a global pairs trading system. The application (data, trading system and performance indicators) is presented in Section 3. Results and comments are provided in Section 4. Section 5 concludes.

2. Methodology

Multi-criteria decision methods (Figueira et al., 2005a; Wallenius et al., 2008) are now used in various fields and are becoming increasingly popular in finance as underlined in different surveys (Zopounidis, 1999; Steuer and Na, 2003; Spronk et al., 2005; Xidonas and Psarras, 2009; Duygun Fethi and Pasiouras, 2010).

A multi-criteria decision making problem is concerned with the task of analysing a finite number of alternatives (A_1,A_2,...), each of which is explicitly described in terms of different characteristics or criteria (C_1,C_2,...). The techniques of multiple criteria decision making are especially useful for ranking actions in order of preference. The methodology developed in this article intends to perform a ranking of n stocks (the alternatives), based on anticipated returns/spreads (the criteria), in order to select and trade pairs.

Section 2.1 presents, with some modifications, the framework proposed in Huck (2009), introduces briefly Neural Networks and the Electre III method and how to use it for pairs trading. Some important issues concerning multi-step forecasting are discussed in Section 2.2.

Although, in practice, the ranking step follows the forecasting step, the presentation starts with Electre III. We have chosen this order because the link between MCDM and the pairs trading system that will be designed is essential in this work whereas Neural Networks are just a forecasting method that is used for reasons of convenience. The mathematical aspects (optimization) of Electre III and Neural Networks will not be presented.

2.1. A general framework

The multi-criteria analysis method used in this paper is Electre III. It was introduced by Roy (1978). For a complete description of the method, the reader may refer to Figueira and Greco (2004) and Figueira et al. (2005b). It is an outranking method based on the concept of fuzzy logic. This technique takes into account the uncertainty and vagueness which are usually inherent in data produced by predictions and estimations. It has a long history of successful real-life applications. Other ranking techniques could be used without changing the key points discussed in this paper. Nevertheless, Electre III seems quite appropriate for pairs trading because it aggregates preference relations on pair of alternatives whereas, for example, the Multi-Attribute Utility Theory (Keeney and Raiffa, 1976) aggregates unidimensional utility functions into a single global utility function combining all the criteria.

Table 1
General structure of the decision matrix used in the article.

Alternatives (assets/stocks)	Criteria (Performances – spreads between stock return forecasts)			
	C_1	C_2	...	C_n
A_1	0	$\hat{r}^{t+h\|t}_{1\|1,2} - \hat{r}^{t+h\|t}_{2\|1,2}$	...	$\hat{r}^{t+h\|t}_{1\|1,n} - \hat{r}^{t+h\|t}_{n\|1,n}$
A_2	$\hat{r}^{t+h\|t}_{2\|1,2} - \hat{r}^{t+h\|t}_{1\|1,2}$	0	...	$\hat{r}^{t+h\|t}_{2\|2,n} - \hat{r}^{t+h\|t}_{n\|2,n}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
A_n	$\hat{r}^{t+h\|t}_{n\|1,n} - \hat{r}^{t+h\|t}_{1\|1,n}$	$\hat{r}^{t+h\|t}_{n\|2,n} - \hat{r}^{t+h\|t}_{2\|2,n}$	...	0

The way the method has been implemented and the choice of some technical parameters will be described in detail in Section 4.

As pointed out in Section 1, modelling and forecasting of spread time series are crucial in pairs trading. If we consider n as being the number of stocks in the whole information set, there are $\frac{n*(n-1)}{2}$ anticipated spreads and pairs. Following Huck (2009), we use a decision matrix, see Table 1, based on multiple forecasts and spreads. There is no restriction on forecasting method: in the decision matrix, we merely assume the use of a quantitative forecasting model.

The performance of the ith asset on the jth criterion is the spread between the return forecasts of assets A_i and A_j conditionally to the past returns of assets i and j. $\hat{r}^{t+h|t}_{i|i,j}$ is one estimator of $r^{t+h|t}_{i}$. These spreads can be defined for various values of h, the forecast horizon. A spread is noted $s^{t+h|t}_{i,j}$ with:

$$s^{t+h|t}_{i,j} = \hat{r}^{t+h|t}_{i|i,j} - \hat{r}^{t+h|t}_{j|i,j}. \quad (2)$$

Let us call this decision matrix $S^{t+h|t}$. It gathers all spreads and forecasts computed at time t with an h period forecast horizon. Each value in the decision matrix is a relative performance. In this case, the number of alternatives equals the number of criteria; the matrix is thus anti-symmetric. Several extensions and modifications to the decision matrix are proposed in Huck (2009). It is possible to add new criteria to the decision matrix: an important case is discussed in Section 2.2.

The ranking obtained after using Electre III helps to detect potentially "under- and overvalued" stocks and to select pairs. Our approach is "integrated" (Jacobs et al., 1999), in the sense that "losing" and "winning" securities are treated and chosen in a single analysis. The methodology presented in this paper can be seen as a sort of bridge between forecast combinations (Timmermann, 2006) and multi-criteria decision methods (Electre III) but in our framework, the output is not a new point or density forecast; it is a ranking. This methodology is a stratagem to manage a large data base: although we only use bivariate information sets for forecasting, the whole information set is indirectly considered.

Because a large number of bivariate information sets[3] has to be analysed, an efficient solution is to use Artificial Neural Networks for forecasting instead of parametric models that need to be designed precisely, one by one. The drawback is that Neural Networks could be seen as a black box.

[3] The application of this article examines 90 stocks, which means a total of $4005\left(\frac{90*(90-1)}{2}\right)$ bivariate information sets/pairs must be considered.

Elman Networks (Elman, 1990) are considered in this work. These are somewhat more complex than the feedforward networks which are the most popular networks. These networks allow the neurons to depend not only on the input variables but also on their own lagged values. Elman networks, compared to feedforward networks, are specific to data that have a time dimension. It is a way of capturing memory in financial markets.

For a general introduction to ANNs, the reader may refer to Zhang et al. (1998). ANNs are considered to be a universal approximator in a wide variety of non-linear patterns which, is especially important because non-linearities are common features in financial data. Furthermore, Neural Networks are good predictors (Swanson and White, 1995, 1997; Sharda, 1994).

The precise specification of the networks (number of input nodes, layers, nodes in the hidden layer, etc.) is given in the next section.

2.2. Some remarks on multi-step forecasting

As underlined in Section 1, the use of multi-step forecasts is the main methodological contribution of this paper compared to Huck (2009): for example, four step-ahead forecasts with weekly data will provide the trading system with much more opportunity and flexibility. In this case, the holding period of a pair, based on the forecast horizon, could last for 20 trading days. It clearly remains a short-term strategy but it covers what should be the most profitable days of pairs trading strategies (Engelberg et al., 2008).

An important choice has to be made because, in the literature, several approaches are reported for this type of problem with Neural Networks. Discussions on multi-step forecasting, briefly presented here, can, for example, be found in Zhang et al. (1998), Kline (2004) and Chevillon (2007).

The three approaches, using Kline's (2004) terminology, are:

- The iterative method which uses a single step-ahead model to iteratively generate forecasts.
- The independent method which uses a dedicated network to forecast each forecast horizon.
- The joint method which uses a single network to forecast all forecast horizons.

In this very general discussion of multi-step forecasting, we simplify notations as much as possible:

- Only one stock is considered.
- The past returns of a given asset are the only information used to perform forecasts.

P^t is the price of a stock at time t, r^t is the return at time t, h the forecast horizon ($h \geqslant 1$), $r^{t+h|t}$ is the return between dates t and $t+h$ $\left(r^{t+h|t} = \frac{P^{t+h}-P^t}{P^t}\right)$. If h equals one, we can note $r^{t+1|t} = r^{t+1}$. We thus have $r^{t+h|t} = \prod_{i=1}^{h}(1 + r^{t+i|t+i-1}) - 1 = \prod_{i=1}^{h}(1 + r^{t+i}) - 1$.

In our trading context, the return of a series at time $t+h$ is much less important than the cumulative return of a series between t and $t+h$.

The first approach is called the iterative forecasting as used in the Box–Jenkins model in which the forecast values are iteratively used as inputs for the next forecasts. If we assume a simple forecasting model with two lags and one output node[4] where f is the function estimated by ANN:

$$\hat{r}^{t+1} = f(r^t, r^{t-1}).$$

$\hat{r}^{t+1}$ is the estimate of r^{t+1}. To estimate two periods ahead, the model becomes:

$$\hat{r}^{t+2} = f(\hat{r}^{t+1}, r^t),$$
$$\hat{r}^{t+2|t} = (1 + \hat{r}^{t+1}) * (1 + \hat{r}^{t+2}) - 1.$$

To estimate three periods forward, the model becomes:

$$\hat{r}^{t+3} = f(\hat{r}^{t+2}, \hat{r}^{t+1}),$$
$$\hat{r}^{t+3|t} = (1 + \hat{r}^{t+1}) * (1 + \hat{r}^{t+2}) * (1 + \hat{r}^{t+3}) - 1.$$

As the forecasts move forward, past observations are discarded. Instead, forecasts rather than observations are used to forecast further future points. It involves a problem of propagating early errors into future forecasts: a poor forecast in an early period has an adverse effect on later forecasts and on the cumulative anticipated returns.

This approach is commonly used for short-term forecast horizons where the one-lag correlation is dominant over other lag correlations. Typically, the longer the forecasting horizon, the less accurate the iterative method.

Approaches 2 and 3 can be used in two ways in our trading context:

- The first is to forecast returns for each sub-period. A product is thus needed.
- In the second case, the global/cumulative return is forecast directly.

The second approach in multi-step forecasting is to create independent models for each forecast horizon. In this case, if l is the number of lags, the forecasts, up to three steps, could be written in the first case, as:

$$\hat{r}^{t+1} = f_1(r^t, \ldots, r^{t-l}),$$
$$\hat{r}^{t+2} = f_2(r^t, \ldots, r^{t-l}),$$
$$\hat{r}^{t+3} = f_3(r^t, \ldots, r^{t-l}),$$
$$\hat{r}^{t+3|t} = (1 + \hat{r}^{t+1}) * (1 + \hat{r}^{t+2}) * (1 + \hat{r}^{t+3}) - 1$$

or, in the second case, as:

$$\hat{r}^{t+1|t} = f_1(r^t, \ldots, r^{t-l}),$$
$$\hat{r}^{t+2|t} = f_2(r^t, \ldots, r^{t-l}),$$
$$\hat{r}^{t+3|t} = f_3(r^t, \ldots, r^{t-l}).$$

The third option is to create a single model that simultaneously computes forecasts for all possible forecast horizons. The model for generating three step-ahead forecasts would be, in the first case, written as:

$$(\hat{r}^{t+3}, \hat{r}^{t+2}, \hat{r}^{t+1}) = f(r^t, \ldots, r^{t-l}),$$
$$\hat{r}^{t+3|t} = (1 + \hat{r}^{t+1}) * (1 + \hat{r}^{t+2}) * (1 + \hat{r}^{t+3}) - 1$$

or, in the second case, as:

$$(\hat{r}^{t+3|t}, \hat{r}^{t+2|t}, \hat{r}^{t+1|t}) = f(r^t, \ldots, r^{t-l}). \tag{3}$$

These forecasts could thus be used for an independent trading system with different maximal holding periods. The parameter estimation is accomplished by pooling the different forecast errors. In this approach, parameters are shared across forecast horizons. A significant appeal of this method is that it can strongly reduce the number of networks to be estimated, and the computation time, because only one function f is needed. This forecasting strategy could be especially useful in the extension briefly presented below.

Zhang et al. (1998), Kline (2004) and the references therein show none of the methods dominates the others in all circum-

[4] For reasons of brevity, the following equations of this subsection consider models with one output node and lag data coming from only one time series whereas models with two output nodes and a bivariate information set will be used in the application. In that case, the first equation of this subsection should be written as: $(\hat{r}_1^{t+1}, \hat{r}_2^{t+1}) = f(r_1^t, r_1^{t-1}, r_2^t, r_2^{t-1})$.

stances. Linked to the trading system that is detailed in the next section and to the nature of the data (the presence of non-linearities) to be considered, we believe the independent method – the second approach without product – to be the most relevant. Arguments in favour of this proposition are:

- Each model is dedicated to forecasting its own forecast horizon. It avoids products between forecasts.
- With this method, the models developed and ultimately chosen may differ substantially across the horizon to be forecast (functional form, lags, estimated parameters). As has already been mentioned in Section 1, data snooping is an important issue. In the application, a first solution to circumvent this problem will be to use the same network design and number of lags for the different information sets and forecast horizons. Too much optimization would reduce the appeal of this work.

The way multi-step forecasts are integrated in this work is quite simple: for each forecasting and trading horizon (up to four steps/weeks in this article), an independent application will be proposed and discussed in the following.

A source of improvement in the selection process that would be relevant in our trading context would be the combination, in a single system, of the different forecasts/horizons in the same decision matrix. In the application of this article an $n * n$ decision matrix is used (see Table 1) with n being the number of stocks in the whole information set.

An alternative decision matrix, say $D^{t+2,t}$, considering for example, at the same time, one and two step-ahead forecasts, could be written as follows using the notations of the beginning of this section:

- $D^{t+2,t} = [S^{t+1|t}, S^{t+2|t}]$.
- This matrix has n lines (alternatives) and $2 * n$ columns (criteria) instead of n. The elements $d_{i,j}^{t+2|t}$ of this matrix are computed as:
  - For $j \leqslant n, d_{i,j}^{t+2|t} = \hat{r}_{i|i,j}^{t+1|t} - \hat{r}_{j|i,j}^{t+1|t}$.
  - For $j > n, d_{i,j}^{t+2|t} = \hat{r}_{i|i,j-n}^{t+2|t} - \hat{r}_{j|i,j-n}^{t+2|t}$.

With this approach, several forecast horizons can easily be combined. In these cases, the debate between independent and joint models stays open. This alternative presents some advantages and drawbacks:

- The probability of a stock being among the "most over – or undervalued" stocks will increase if the spread forecasts, with respect to the forecast horizons, are large and coherent (always in the same direction). In behavioural terms and in a univariate setting, an investor might interpret this solution as follows: if the return forecasts of one stock over different horizons are all "positive", it is more likely that the agent will buy this equity compared to a stock with "mixed" forecasts. The choice of weightings between the different $n * n$ matrices included in the whole matrix will be an important issue but some intuitive heuristics could be interesting: giving the same weight to each horizon or considering a weight proportional to the forecast horizon.
- The optimization process will be more complicated and the computation time will probably increase by more than the number of sub-matrices. It could also have some impact on the choice of the different thresholds in the Electre III method.

3. Design of the application

This section presents in detail the trading system developed in this article, based on the methodology introduced in the previous section. The aim is to establish:

- whether the strategy could be profitable even if some aspects have not yet been integrated;
- what might be a relevant design for the trading system and the problems during the implementation of the method;
- whether there is a relationship between performance and forecast horizon.

The data used are the returns of S&P 100 stocks[5] between January 1993 and December 2006. These stocks are among the most liquid in the world: the transaction costs would be relatively low. Only 90 stocks are considered in the analysis: if the quotation of a stock started in 1992 or later, it has been removed from the sample (Google, for example). The whole list of stocks used is provided in Table 2. Data are considered on a weekly basis (Friday closing price) for forecasting purposes and on a daily basis in the trading system. This different use of the data between the two steps (forecasting and trading) is a consequence of several points:

- An arbitrage is necessary between the need for data for the estimation of Neural Networks and the need to forecast several trading days ahead. The use of weekly data divides by 5 (compared to daily data) the possible number of observations for the estimation and forecasting stages but increases by the same coefficient the forecast horizon (in days).
- Christoffersen and Diebold (2006) explain that sign dependence is not likely to be found in very high-frequency (e.g. daily) or very low-frequency (e.g. annual) returns; instead, it is more likely to be found at intermediate return horizons. Weekly data seem to be an interesting trade-off during the forecasting step.
- If, for example, we want a 10 day-ahead indication for pairs selection and trading, two options might be:
  - Using weekly data and performing two step-ahead forecasts.
  - Using daily data and performing 10 step-ahead forecasts.
- The idea that the higher the forecast horizon the lower the quality of the forecast is a serious drawback for the first example above. Forecasting up to four step-ahead is the limit chosen in this article.
- If we have 5, 10, 15 or 20 day-ahead trading indications based on 1, 2, 3 and 4 week-ahead forecasts, daily data can nevertheless be used in the trading system: if a closing threshold is reached, a position could be closed out on a day which is not a multiple of 5. It is one of the reasons why the trading strategy, in this article, is much more realistic and flexible than Huck (2009).

The trading system will be active for 8 years: between 1999 (beginning) and 2006 (end). The computational period is, in fact, slightly greater. A few weeks must be added before and after in order to always have h overlapping open portfolios. If $h = 2$, there is a one week lag before the two parts of the portfolio are active.

3.1. Trading system and performance measures

Each week, after the first two stages (forecasting and ranking[6]), "p" stocks are bought and sold for a similar amount: the first ("undervalued stocks") and last ("overvalued stocks") of the ranking. The possible values of "p" are the following: 5, 10 or 15. The definition of pairs, and the way we count them, derives directly from "p": "p" long and short positions define "p^2" pairs.[7] They are independent after opening and, if needed, can be closed individually. The follow-

[5] The data were extracted from Datastream in January 2007.

[6] The computation of the Electre III method is performed using the Matlab code written by Francesco DiPierro.

[7] The way we count and define pairs in this article differs, for example, from Gatev et al. (2006) and Huck (2009).

Table 2
List of stocks used in the application.

3M	Bristol Myers Squibb	EMC	Intl. Paper	Rockwell Auto.
AES	Burl. Nthn. Santa Fe C	Eastman Kodak	JPMorgan Chase & Co.	Sara Lee
AT&T	CBS	Entergy	Jonhson & Jonhson	Schlumberger
Abbott Labs.	Cigna	Exelon	Limited Brands	Southern
Alcoa	Campbell Soup	Exxon Mobil	McDonalds	Sprint Nextel
Allegheny Techs.	Caterpillar	Fedex	Medimmune	Target
Altria Group Inco.	Chevron	Ford Motor	Medtronic	Texas Intsts
Amer. Elec. Pwr.	Cisco Systems	General Dynamics	Merck & Co.	Tyco Intel.
American Express	Citigroup	General Electric	Merrill Lynch & Co.	US Bancorp
American Intl. Gp.	Clear Chl.Comms.	General Motors	Microsoft	United Techno.
Amgen	Coca Cola	Halliburton	National Semicon.	Verizon Comms.
Anheuser-Busch	Colgate-Palm.	Harrahs Entm.	Norfolk Southern	Wachovia
Avon Products	Comcast	Heinz	Oracle	Wal Mart Stores
Baker Hughes	Computer Scis.	Hewlett–Packard	PepsiCo	Walt Dysney
Bank of America	ConocoPhillips	Home Depot	Pfizer	Wells Fargo & Co
Baxter Intl.	Dell	Honeywell Intl.	Procter & Gamble	Weyerhaeuser
Black & Decker	Dow Chemicals	Intel	Raytheon	William Cos.
Boing	Du Pont	IBM	Regions Finl. New	Xerox

ing particular values (25,100,225) can be found, by construction, in the first table of results (Table 3 considers one step-ahead forecasts, $h = 1$) when there is no-closing threshold (a pair can only be closed out after five trading days). "p", given the full database (90 stocks), could be in the range $[1:45]$. The examination of the whole spectrum is not necessary for two main reasons:

- The relationship between the number of stocks/pairs and performance will be adequately documented with three relevant values of "p". The negative link observed in Huck (2009) will appear partly in the result. The forecast horizon, as we will see, plays a major role.
- From a strategic point of view, the trading system must select a limited number of stocks, i.e. only "top under and overvalued assets". The idea is that these "top stocks" could provide interesting performances but the chosen values of "p" are such that the diversification of the portfolio remains acceptable.

Forecasts and rankings are computed, each week, with an up to 4 week horizon. At a given time, the complete portfolio is the sum of overlapping portfolios open each week. If $h = 2$, the complete portfolio is defined by the positions opened during the current week and the week before. If we look at Tables 4–6 and still the cases without closing threshold, the number of open pairs are simply h times (50, 200 and 450 in Table 4) the values found in Table 3.

These numbers are one of the indicators of diversification of the portfolio. A second measure, also provided in the tables of results, is the number of different stocks in the portfolio. This value, computed daily, includes all long and short equities and takes into account the fact that a stock can be bought in two consecutive weeks. It explains why in Table 3, when $h = 1$ in the no-closing threshold case, the number of stocks in the portfolio equals 10, 20 and 30 if $p = 5$, 10 or 15 respectively. The relationship between p and the number of different stocks in the portfolio is thus simply $2 * p$. If $h > 1$ (overlapping portfolios) or if closing thresholds are introduced, the corresponding values found are necessarily below $2 * p * h$. There are two reasons for this:

- In the trading system, a stock (eventually the two legs of a pair) can be chosen and open several weeks consecutively: this is a consequence of the forecasting model used. As in Section 2.2, we simplify the notations as much as possible here and consider only one stock, the use of 4 lags as inputs in the Neural Networks and h, the forecast horizon, equals 3.
  At date t, the inputs of the forecasting model are $(r^t, r^{t-1}, r^{t-2}, r^{t-3})$. At time $t+1$, they are $(r^{t+1}, r^t, r^{t-1}, r^{t-2})$. Let us assume now that r^{t-1} is a value with an important impact on the cumulative return of the series several steps ahead. At times t and $t+1$, this particular value r^{t-1} can lead to high values when computing $(\hat{r}^{t+3|t})$ and $(\hat{r}^{t+4|t+1})$ and thus to the selection of the same asset (pair) over two consecutive weeks if the variation, anticipated the first time, has not yet occurred.
  Furthermore, when $h = 1$, a pair could open during a full week, close after 5 days and re-open instantaneously because the new forecasting and ranking sequences indicate that this stock/pair is still among the best candidates. In that case, transaction costs should only be considered once. All pairs have the same weight, are open for the same amount but, due to overlapping portfolios, a stock may be overweighted in the portfolio.
- If the closing threshold is reached, the pair must be closed. Pairs are quite often closed by groups. If a particular asset experiences a rapid, large variation, it may lead to the closing of all pairs concerned by this particular asset.

As already mentioned, closing thresholds have been introduced: they are symmetric for profits and losses. The chosen values, 10% and 20%, are large enough to capture the anticipated spread returns and to avoid too huge losses. A third case, the absence of a closing threshold is also developed. Three important remarks must be made:

- Tables 3–6 report maximum profits and losses, per pair, greater, in absolute value, than the predefined threshold. This is explained by the nature of the data used in the article: these are only daily closing prices and not intraday prices. From a strategic point of view, we consider all transactions occur at the end of the day at the closing price. It would have been easy to strictly limit, from a computational point of view, profits and losses to the threshold. This would have introduced two new assumptions: the possibility of transaction during the trading day and an approximation. Even if it would be very marginal, during a very volatile day a spread return could reach the profit threshold, say 10%, in the morning and the loss threshold when the market closes. In this example, a loss would be recorded whereas a profit would be more in line with the use of the closing threshold.
- The introduction of fixed thresholds presents advantages and drawbacks. The same threshold is used for pairs with low and high levels of volatility. This can be inadequate for some pairs. A 5% variation on a low volatile pair could be "more significant", in a trading context, than a 10% movement on a highly volatile pair. A fixed threshold ensures that the variation needed before closing the position is much greater than the transaction costs.

Table 3
Results with one step-ahead forecasts ($h = 1$).

Design									
(p)	5	5	5	10	10	10	15	15	15
$(h * p^2)$	25	25	25	100	100	100	225	225	225
$(2 * p * h)$	10	10	10	20	20	20	30	30	30
Closing threshold (%)	10	20	None	10	20	None	10	20	None
Return per pair									
Mean return	0.62	0.65	0.73	0.45	0.47	0.50	0.30	0.32	0.33
Pairs with positive return (%)	52.78	52.54	52.49	52.26	52.12	52.09	51.61	51.52	51.50
Max	55.32	56.18	220.30	55.32	59.55	228.61	57.42	59.55	228.61
Min	−36.34	−36.34	−45.67	−36.34	−55.40	−63.83	−41.04	−56.62	−63.83
Standard deviation	7.55	7.99	9.48	6.96	7.40	8.28	6.63	6.98	7.59
VAR 1%	−15.92	−20.33	−19.36	−15.73	−19.90	−18.91	−15.31	−18.56	−17.87
VAR 5%	−11.67	−11.44	−11.45	−11.23	−10.89	−10.89	−10.99	−10.46	−10.46
Skewness	0.48	0.68	6.53	0.33	0.47	5.04	0.27	0.41	4.34
Kurtosis	5.85	7.20	139.02	5.38	7.25	126.91	5.48	7.40	119.20
Opening time per pair (days)									
Mean	4.64	4.95	5	4.69	4.96	5	4.73	4.97	5
Max	5	5	5	5	5	5	5	5	5
Min	1	1	5	1	1	5	1	1	5
Standard deviation	0.89	0.36	0	0.83	0.31	0	0.78	0.28	0
Skewness	−2.52	−7.23	NaN	−2.80	−8.53	NaN	−3.04	−9.68	NaN
Kurtosis	8.40	57.58	NaN	10.03	80.31	NaN	11.57	103.67	NaN
Threshold reached (+)(%)	12.09	2.46	0	10.28	1.83	0	8.88	1.42	0
Threshold reached (−)(%)	9.25	1.09	0	8.10	0.98	0	7.42	0.82	0
Daily returns of the portfolio									
Mean	0.122	0.131	0.146	0.092	0.092	0.099	0.062	0.062	0.066
Days with positive return (%)	49.52	50.00	49.90	50.72	50.63	50.58	50.34	50.29	50.43
Max	14.94	13.49	35.39	8.46	8.77	17.19	6.39	6.94	12.06
Min	−9.42	−7.79	−7.79	−6.46	−6.41	−6.41	−5.35	−5.49	−5.49
Standard deviation	1.93	1.96	2.13	1.39	1.43	1.50	1.14	1.17	1.21
VAR 1%	−4.91	−4.97	−4.97	−3.37	−3.57	−3.75	−2.77	−2.85	−2.87
VAR 5%	−2.82	−2.97	−3.01	−2.11	−2.16	−2.16	−1.76	−1.80	−1.81
Skewness	0.64	0.63	2.64	0.43	0.41	1.03	0.41	0.43	0.81
Kurtosis	7.94	7.13	41.45	5.46	5.60	13.06	5.35	5.51	9.71
Number of pairs in the portfolio per day									
Mean	23.21	24.73	25	93.89	99.20	100	212.88	223.54	225
Max	25	25	25	100	100	100	225	225	225
Min	3	12	25	31	68	100	89	175	225
Standard deviation	3.42	1.23	0	10.78	2.96	0	21.02	5.13	0
Skewness	−2.48	−5.89	NaN	−2.44	−5.06	NaN	−2.50	−4.84	NaN
Kurtosis	9.77	42.91	NaN	9.61	33.67	NaN	10.04	30.66	NaN
Number of stocks in the portfolio per day									
Mean	9.90	9.98	10	19.88	19.98	20	29.87	29.98	30
Max	10	10	10	20	20	20	30	30	30
Min	4	8	10	14	18	20	24	28	30
Standard deviation	0.50	0.17	0	0.54	0.17	0	0.55	0.20	0

A threshold based on the standard deviation of each pair, often used in the distance approach (Gatev et al., 2006), could be considered. An alternative could be a function like $Max(T, \delta\sigma)$ where T is a fixed threshold, σ the standard deviation of the spread time series and δ a coefficient indicating the size of the deviation needed before closing the position.

- As a comparison, Gatev et al. (2006), in a distance based method, open positions for the top 20 pairs when prices have diverged, on average, by 5.3%. This is also the threshold used to consider whether stocks have converged. The authors underline that this is a relatively narrow gap and they have not experimented to find out the optimal trigger in terms of profitability, which may be much higher than two standard deviations. In the present article, the thresholds are only considered as the reference to close the positions. The reason for opening positions is that we are anticipating a large spread (and the direction of the spread) between the returns of two stocks in the near future.

One part of the tables is concerned with the opening time of the pairs. The indication "Threshold reached (+)(%)" indicates the proportion of pairs reaching the closing threshold in a favourable way ("Threshold reached (+)(%)" refers to losses): the higher the closing threshold, the smaller this indicator will be. Thresholds reached on the last day of the allowed trading period are part of this indicator: in that case, the pair closes for two reasons.

The last clarifications required concerning the tables of results deal with the performance indicators. In order to have a trading system as simple as possible, additional hypotheses about the margin requirements and transaction costs have not been introduced in our long-short context. This has some impact on the definition of what we call "return" in the tables of results.

- The "return per pair" indicates the average performance of a pair at the end of the allowed trading period. In the tables, this value can be seen as the "profit", in dollars, if the pair is initiated with 100 dollars of long and short positions. Potentially, the maximal loss on a pair is greater than 100%.
- The "daily return of a pair" is, in fact, computed as the difference between the spreads captured by a pair between two consecutive dates.
- The "daily return of the portfolio", at a given date, is the mean of the "daily returns" among active pairs. It takes into account that pairs open and close at different dates and the fact that some

Table 4
Results with two step-ahead forecasts ($h = 2$).

Design									
(p)	5	5	5	10	10	10	15	15	15
$(h * p^2)$	50	50	50	200	200	200	450	450	450
$(2 * p * h)$	20	20	20	40	40	40	60	60	60
Closing threshold (%)	10	20	None	10	20	None	10	20	None
Return per pair									
Mean return	0.59	0.51	0.46	0.45	0.34	0.27	0.29	0.21	0.16
Pairs with positive return (%)	52.54	51.90	51.73	52.04	51.26	51.13	51.19	50.62	50.51
Max	54.99	59.95	96.41	54.99	61.43	100.39	55.34	61.43	106.84
Min	−51.94	−55.39	−58.68	−55.10	−67.48	−89.61	−57.47	−67.48	−94.35
Standard deviation	9.55	11.13	11.69	8.95	10.24	10.59	8.54	9.66	9.90
VAR 1%	−18.79	−25.39	−30.02	−18.03	−24.53	−27.92	−17.42	−23.83	−26.08
VAR 5%	−13.80	−20.25	−17.87	−13.26	−17.56	−16.26	−12.91	−16.03	−15.22
Skewness	0.01	0.09	0.42	−0.03	0.06	0.21	0.01	0.08	0.19
Kurtosis	3.00	4.12	7.58	3.06	4.59	8.55	3.13	4.83	8.66
Opening time per pair (days)									
Mean	7.95	9.58	10	8.25	9.67	10	8.46	9.73	10
Max	10	10	10	10	10	10	10	10	10
Min	1	1	10	1	1	10	1	1	10
Standard deviation	2.89	1.43	0	2.74	1.28	0	2.61	1.16	0
Skewness	−1.06	−3.76	NaN	−1.29	−4.34	NaN	−1.47	−4.82	NaN
Kurtosis	2.59	17.06	NaN	3.16	22.22	NaN	3.70	27.07	NaN
Threshold reached (+)(%)	23.94	6.56	0	20.99	5.16	0	18.57	4.29	0
Threshold reached (−)(%)	20.34	5.43	0	17.80	4.25	0	16.49	3.53	0
Daily return of the portfolio									
Mean	0.079	0.050	0.044	0.060	0.032	0.025	0.040	0.020	0.014
Days with positive return (%)	52.63	50.84	50.70	51.86	51.08	51.13	52.19	50.65	50.60
Max	9.60	8.22	11.79	5.62	6.32	7.14	4.52	4.88	5.20
Min	−8.26	−8.76	−9.13	−5.22	−5.52	−6.97	−3.79	−4.14	−3.75
Standard deviation	1.47	1.44	1.52	1.11	1.11	1.15	0.89	0.90	0.92
VAR 1%	−3.78	−3.82	−4.02	−2.85	−2.83	−2.95	−2.40	−2.43	−2.43
VAR 5%	−2.24	−2.17	−2.24	−1.71	−1.73	−1.75	−1.35	−1.42	−1.50
Skewness	0.08	0.01	0.16	0.16	0.17	0.18	0.19	0.22	0.25
Kurtosis	6.53	6.49	8.63	5.64	6.02	7.11	5.46	5.75	5.86
Number of pairs in the portfolio per day									
Mean	39.74	47.88	50	165.02	193.36	200	380.49	437.73	450
Max	50	50	50	200	200	200	450	450	450
Min	6	25	50	44	124	200	133	301	450
Standard deviation	8.74	3.45	0	30.97	10.53	0	62.51	19.44	0
Skewness	−0.85	−2.22	NaN	−1.05	−2.27	NaN	−1.13	−2.36	NaN
Kurtosis	3.21	9.35	NaN	3.57	9.33	NaN	3.73	10.00	NaN
Number of stocks in the portfolio per day									
Mean	14.97	15.42	15.48	28.99	29.33	29.40	42.92	43.17	43.23
Max	20	20	20	38	38	38	56	56	56
Min	7	11	12	20	24	24	34	36	36
Standard deviation	1.77	1.59	1.57	2.50	2.46	2.45	3.32	3.34	3.33

positions have been closed out during the h possible weeks of holding because the closing threshold has been reached.

3.2. Specification details

We will now present the way the first two parts, forecasting with Neural Networks and ranking with Electre III, has been implemented. The parameters will be discussed and defended.

A general and unique specification strategy is proposed for the neural networks: we recall that 4005 information sets have to be considered. In this application, the aim is to forecast the cumulative returns of two stocks. As a consequence, there are two output nodes. Neural Networks have one hidden layer. This is the most common choice in the literature. It is influenced by theoretical works showing that a single hidden layer is sufficient for ANN's to approximate any complex non-linear function (Cybenko, 1989; Hornik et al., 1989).

Five lagged returns for each asset are introduced as input which means that the models have 10 input nodes. Using the notations of Section 2, the inputs to compute $s_{1,2}^{t+h|t}$ are $r_1^t,\ r_1^{t-1},\ r_1^{t-2},\ r_1^{t-3},\ r_1^{t-4}$ and $r_2^t,\ r_2^{t-1},\ r_2^{t-2},\ r_2^{t-3},\ r_2^{t-4}$. We recall that the number of autoregressive terms in the Box–Jenkins model for a univariate time series is not appropriate for the non-linear relationships modelled by neural networks. The choice of 5 lags seems reasonable to forecast efficiently from one to four step-ahead. We do not try to find out the optimal values of the number of lags which is probably not the same for the different values of the forecast horizon. In order to avoid overfitting, researchers have provided empirical rules to restrict the number of hidden nodes (Lippmann, 1987; Hecht-Nielsen, 1990; Wong, 1991; Kang, 1991; Tang and Fishwick, 1993). Of course, none of these heuristics works well for each and every problem. One of the most popular rules for the one hidden layer feedforward network is to use a number of hidden nodes equal to the number of inputs. Elman networks are known as needing more neurons than feedforward networks because of the recurrent part. As a consequence, the number of hidden nodes considered here is the double of the number of inputs. Networks, in our application, have thus 20 neurons. A tan-sigmoid transfer function is used.

The training of Elman Networks is performed using a 6 year mobile window ($6 * 52 = 312$ points/weeks). The debate on the optimal size of the estimation window is left open in order to avoid

Table 5
Results with three step-ahead forecasts ($h = 3$).

Design									
(p)	5	5	5	10	10	10	15	15	15
($h * p^2$)	75	75	75	300	300	300	675	675	675
($2 * p * h$)	30	30	30	60	60	60	90	90	90
Closing threshold (%)	10	20	None	10	20	None	10	20	None
Return per pair									
Mean return	0.37	0.40	0.35	0.51	0.48	0.38	0.41	0.37	0.25
Pairs with positive return (%)	51.50	50.86	50.63	52.25	51.51	51.30	51.74	51.07	50.84
Max	41.71	58.95	84.91	51.64	58.95	96.74	52.89	61.67	96.74
Min	−46.90	−59.77	−67.67	−54.52	−59.77	−76.81	−58.68	−59.77	−76.82
Standard deviation	10.21	12.70	13.57	9.79	11.84	12.37	9.48	11.27	11.72
VAR 1%	−18.98	−27.25	−36.09	−18.48	−26.40	−33.17	−18.20	−25.77	−31.83
VAR 5%	−14.35	−21.79	−21.23	−13.87	−21.12	−19.38	−13.52	−20.61	−18.19
Skewness	−0.04	0.05	0.22	−0.07	0.01	0.16	−0.07	0.01	0.11
Kurtosis	2.14	3.13	5.88	2.34	3.37	6.45	2.46	3.59	6.72
Opening time per pair (days)									
Mean	10.69	13.88	15	11.21	14.11	15	11.52	14.24	15
Max	15	15	15	15	15	15	15	15	15
Min	1	1	15	1	1	15	1	1	15
Standard deviation	4.86	2.82	0	4.71	2.56	0	4.58	2.37	0
Skewness	−0.59	−2.67	NaN	−0.78	−3.08	NaN	−0.91	−3.41	NaN
Kurtosis	1.78	9.26	NaN	2.05	11.81	NaN	2.27	14.15	NaN
Threshold reached (+)(%)	29.48	10.36	0	27.22	8.52	0	25.43	7.47	0
Threshold reached (−)(%)	26.22	8.79	0	23.53	7.21	0	22.12	6.10	0
Daily return of the portfolio									
Mean	0.035	0.027	0.021	0.052	0.034	0.024	0.042	0.026	0.016
Days with positive return (%)	50.48	50.82	50.39	50.82	50.63	50.72	50.92	51.45	51.30
Max	6.60	5.94	5.09	4.90	4.31	4.74	3.91	3.38	3.60
Min	−7.16	−5.64	−5.70	−4.46	−4.31	−4.37	−3.30	−3.53	−3.71
Standard deviation	1.26	1.16	1.18	0.91	0.88	0.90	0.72	0.71	0.73
VAR 1%	−3.53	−3.12	−2.98	−2.44	−2.35	−2.44	−1.83	−1.88	−1.95
VAR 5%	−1.94	−1.84	−1.94	−1.41	−1.37	−1.37	−1.09	−1.11	−1.14
Skewness	0.08	0.08	0.13	0.29	0.10	0.06	0.33	0.08	0.03
Kurtosis	5.96	5.02	4.95	5.67	5.25	5.64	5.95	5.43	5.61
Number of pairs in the portfolio per day									
Mean	53.41	69.39	75	223.99	282.10	300	518.27	640.68	675
Max	75	75	75	300	300	300	675	675	675
Min	9	42	75	60	198	300	171	481	675
Standard deviation	14.38	6.67	0	52.45	20.88	0	111.14	40.48	0
Skewness	−0.47	−1.32	NaN	−0.63	−1.33	NaN	−0.68	−1.40	NaN
Kurtosis	2.39	4.15	NaN	2.47	4.13	NaN	2.55	4.30	NaN
Number of stocks in the portfolio per day									
Mean	18.67	19.71	19.86	36.74	37.48	37.59	53.16	53.71	53.80
Max	26	26	26	46	46	46	65	67	67
Min	9	13	13	28	28	29	44	44	44
Standard deviation	2.59	2.24	2.21	3.08	2.94	2.91	3.73	3.73	3.75

data snooping problems. The chosen value is a result of an arbitrage between the need for relatively recent data, the need to have enough data for the estimation of the weights in the Neural Networks, and the length of the period during which the pairs can be opened. Networks are re-estimated each year. The whole application involves the estimation of 32,040 ($4005 * 8$) networks because eight years of trading system are examined.

If we consider that the estimation of one network takes a few tenths of a second with a standard computer, this indicates that our method is highly time consuming from a computational point of view. An historical work, as in this paper, using, for example, the data used in Gatev et al. (2006) (2300 securities, more than two million pairs) and a large number of re-estimations is currently particularly difficult. On the other hand, the methodology, with a more reduced database, remains fully operational even in a real-time application if re-estimations are not performed too often. When using Neural Networks, about 98% of the computational time is dedicated to the estimation of the weights and the rest to the forecasting part.

The different thresholds and weights in Electre III are chosen according to some simple rules. We call q_j and p_j the indifference and preference thresholds.

- $q_j = \sigma\left(\left[s_{1,j}^{t+h|t}, \ldots, s_{n,j}^{t+h|t}\right]\right)$, where σ denotes the standard deviation.
- $p_j = 2 * \sigma\left(\left[s_{1,j}^{t+h|t}, \ldots, s_{n,j}^{t+h|t}\right]\right) = 2 * q_j$.
- Very high values for veto thresholds are chosen so that veto phenomena are totally ineffective.
- Rogers and Bruen (1998) underline that the assignment of importance weightings to each criterion is a crucial step within Electre III, which is a non-compensatory decision-aid model. They propose a method for choosing weights based on Hinkle's resistance to change (1965). In our work, an intuitive and relevant solution is simply to give the same weight to the different criteria/stocks.

During the ranking step, the discrimination threshold of the credibility index and the computation time are linked. As a first attempt, the discrimination threshold $s(\lambda)$ of the credibility index is as follows:

$$s(\lambda) = \alpha - \beta\lambda \quad \text{with} \quad \alpha = 0.28, \quad \beta = 0.15, \quad \lambda = \rho(b, a). \tag{4}$$

Table 6
Results with four step-ahead forecasts ($h = 4$).

Design									
(p)	5	5	5	10	10	10	15	15	15
$(h * p^2)$	100	100	100	400	400	400	900	900	900
$(2 * p * h)$	40	40	40	80	80	80	120	120	120
Closing threshold (%)	10	20	None	10	20	None	10	20	None
Return per pair									
Mean return	0.34	0.16	−0.08	0.43	0.32	0.05	0.36	0.31	0.11
Pairs with positive return (%)	51.58	50.40	50.17	51.88	50.94	50.63	51.79	51.08	50.85
Max	39.94	52.11	85.01	56.53	68.57	85.32	56.53	68.57	88.93
Min	−54.17	−59.70	−84.87	−54.17	−59.70	−86.85	−56.94	−59.73	−87.09
Standard deviation	11.01	14.29	15.73	10.56	13.42	14.33	10.22	12.84	13.55
VAR 1%	−20.21	−28.45	−43.23	−19.27	−27.69	−40.04	−18.65	−26.91	−37.76
VAR 5%	−14.91	−22.86	−26.11	−14.37	−22.30	−23.35	−13.98	−21.82	−21.65
Skewness	−0.15	−0.06	−0.08	−0.09	−0.04	−0.10	−0.08	−0.02	−0.05
Kurtosis	2.17	2.70	5.32	2.13	2.90	5.58	2.19	3.08	5.64
Opening time per pair (days)									
Mean	12.60	17.77	20	13.29	18.14	20	13.81	18.41	20
Max	20	20	20	20	20	20	20	20	20
Min	1	1	20	1	1	20	1	1	20
Standard deviation	6.76	4.53	0	6.71	4.21	0	6.61	3.90	0
Skewness	−0.21	−1.98	NaN	−0.36	−2.27	NaN	−0.49	−2.54	NaN
Kurtosis	1.49	5.73	NaN	1.56	7.01	NaN	1.67	8.47	NaN
Threshold reached (+)(%)	35.13	13.57	0	32.77	11.77	0	30.66	10.61	0
Threshold reached (−)(%)	31.96	13.13	0	29.43	10.86	0	27.99	9.46	0
Daily return of the portfolio									
Mean	0.030	0.009	−0.003	0.034	0.016	0.002	0.029	0.016	0.005
Days with positive return (%)	50.56	51.53	50.90	51.91	51.96	51.38	52.49	51.48	52.11
Max	5.93	4.40	6.12	6.18	3.89	3.55	3.97	3.13	3.07
Min	−7.01	−5.44	−5.82	−3.50	−3.26	−3.51	−3.04	−2.78	−3.15
Standard deviation	1.21	1.05	1.04	0.85	0.78	0.79	0.68	0.64	0.65
VAR 1%	−3.59	−3.06	−2.78	−2.23	−2.09	−2.09	−1.85	−1.66	−1.70
VAR 5%	−1.88	−1.62	−1.76	−1.39	−1.29	−1.34	−1.08	−1.06	−1.07
Skewness	−0.14	−0.32	−0.16	0.39	−0.01	−0.06	0.30	−0.04	−0.09
Kurtosis	6.49	5.56	5.43	6.38	4.80	4.74	6.09	4.91	4.88
Number of pairs in the portfolio per day									
Mean	62.88	88.82	100	265.41	362.61	400	620.47	827.98	900
Max	100	100	100	398	400	400	884	900	900
Min	12	50	100	71	221	400	198	571	900
Standard deviation	19.01	10.80	0	72.52	36.99	0	156.35	72.47	0
Skewness	−0.29	−1.03	NaN	−0.41	−1.08	NaN	−0.46	−1.07	NaN
Kurtosis	2.22	3.34	NaN	2.14	3.33	NaN	2.12	3.15	NaN
Number of stocks in the portfolio per day									
Mean	19.88	21.46	21.68	38.51	39.71	39.89	55.23	56.05	56.20
Max	29	32	32	53	55	57	77	78	78
Min	11	15	15	28	29	29	44	44	44
Standard deviation	2.90	2.87	2.93	4.43	4.70	4.81	5.59	5.86	5.95

For a limited number of rankings among the more than 1600 we performed for the four possible values of h, the forecast horizon, and 8 years of weekly data, the values of the discrimination parameters were too restricting and a ranking could not be established in a reasonable time (about one or two minutes) or the computation could cease because the maximal allowed number of recursions was reached.

In fact, several values for α have been introduced and tested (0.28; 0.26; 0.24; 0.22; 0.20; 0.19; 0.18) and the highest possible α was kept. Ninety percent of the rankings used $\alpha = 0.28$, 96% an α greater than 0.24. The most problematic cases (say $\alpha \leqslant 0.20$) concern less than 2% of the rankings and mainly occur when $h = 1$ or $h = 2$. Roy and Bouyssou (1993) advocate $\alpha = 0.30$ and $\beta = 0.15$ as parameters of the discrimination threshold of the credibility index. This part of our methodology tries to follow this recommendation as far as possible.

The number of stocks (90 in this application) plays a major role in these problems. The computation time would be much reduced if the number of stocks was more limited. With only 50 stocks, the computation time of the rankings would be divided by more than 10 and much fewer problems would appear with the discrimination parameters of the credibility index which are the only degrees of freedom we could/would choose in the application:

- By construction, in this financial application, we consider that all securities must have the same weight.
- For q_j and p_j, the indifference and preference thresholds, we believe that one and two standard deviations are relevant with data produced by forecast.

If a direct application of the method proposed in this article, with the information set considered in Gatev et al. (2006) (about 2300 securities), is unrealistic, a solution to circumvent this problem is to divide the full database by industry groups. According to these sector rankings, a global sector-weighted portfolio could then be managed. In that context, an important question would be to establish whether and why some sectors should be more or less suitable for pairs trading.

4. Results

The results are reported in Tables 3–6 and in two graphs: the first one provides historical indications concerning performance whereas the second gives details, day by day, of the cumulative return per pair captured by the different strategies. Each table fo-

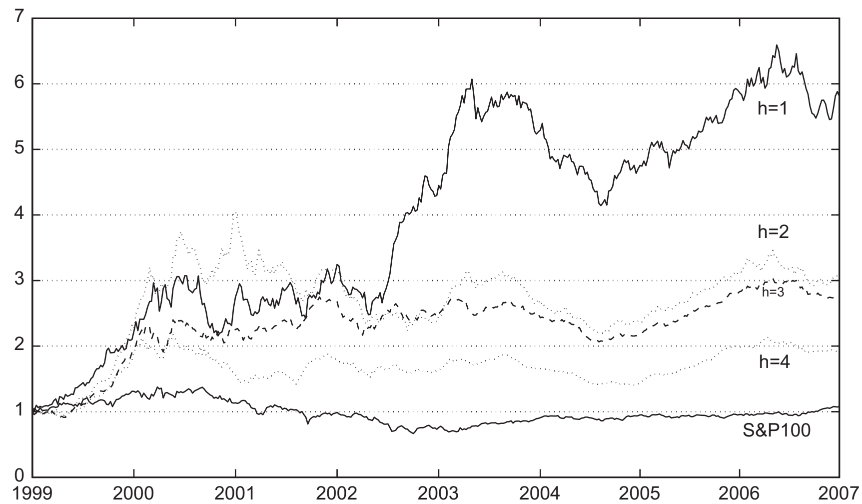

Fig. 1. Cumulative performances between 1999 (beginning) and 2006 (end) of the trading strategies with $p = 10$ and a 10% closing threshold and comparison with the S&P 100 Index.

cuses on a particular forecast horizon, h, from 1 to 4, which means the maximal holding periods of a pair go from 5 to 20 days.

Before detailing the results, the main comments are as follows:

- Twenty-three of the 24 trading strategies generate positive returns per pair. The only "failure" occurs when the portfolio is weakly diversified ($p = 5$), with the highest forecast horizon ($h = 4$) and a without closing threshold.
- The best performances are realized when $h = 1$. The one period ahead forecasts would produce the best trading indications.
- The returns per pair, at the end of the allowed trading period, and the daily return of the portfolio decline strongly with the forecast horizon, especially between the one and the two week-ahead forecasts. Engelberg et al. (2008) indicate that profit should decrease exponentially over time. There is no clear link between their conclusion and our paper because our methodology is very different (the absence of any reference to an equilibrium state). The extensions mentioned in Section 2, i.e. improvements in the selection process, would be useful to clarify the relationship. Furthermore, the information provided in Fig. 2 lead us to think that the maximal trading period allowed for each pair in this article is perhaps too short.
- The negative relationship between returns and number of stocks observed in Huck (2009) is confirmed up to two step-ahead forecasts.
- For $h = 3$ or 4, the most diversified portfolios perform slightly better and strategies with closing threshold (10% or 20%) dominate the ones without closing threshold.

Tables 3–6 provide minimal and maximal values for several return indicators: they are within a very wide range, especially when $h = 1$. The introduction of closing thresholds during the holding period, of course, reduces the absolute values of the extreme returns of the return per pair in the different situations. An increase of p, the number of long and short positions, leads to the reduction of the Value at Risk of the daily returns of the portfolio.

For example, considering all possible forecast horizons, a profit per pair up to 220% or a loss of 94% can be made. When $h = 1$ and $p = 5$, the best daily return of the portfolio is greater than 35%. These extreme values can be largely explained by one stock: the behaviour of Williams Companies[8] in July 2002. If one point, the 35%, was deleted from the series of daily returns of the portfolio when $h = 1$, $p = 5$ and without closing threshold, the kurtosis would fall from 41 to less than 7 which is much more consistent with the other results.

Whatever the design, most of the systems generate positive returns per pair. With a short forecasting horizon, $h = 1$ and $p = 5$, the mean return per pair is above 0.62%. In this case, the closing threshold has only a small influence on the results because the holding period may last 5 days at most. Among the different designs, one of them is especially interesting in terms of performance, independently of the forecast horizon. When $p = 10$ and with a closing threshold set to 10%, the return per pair always exceeds 0.43%. According to our empirical results, this case would be a relevant trade-off between the number of long and short positions and the management of the risk via the closing thresholds. If we consider the current transaction costs proposed by brokerage firms, some of the designs presented in the paper could be profitable for orders of a few thousands of shares.

For each design of the trading system, at least 50% of the pairs generate positive returns: the proportion is greater than 52% in the best situations. These proportions are significantly greater than 50% at a 5% rate if they exceed 50.96%, 50.48% and 50.32% when $p = 5$, 10, 15 respectively. It is always true when the horizon forecast equals 1 or 2.

Whatever the forecast horizon, h, and with closing thresholds up to 20%, some pairs close after one day. From a more general point of view, when the closing threshold is set to 10% and 20% and if we consider an average between the results according to the different values of p; the pairs reaching the threshold are about 18% and 3%, 40% and 10%, 51% and 16%, and 63% and 23% if h, the forecast horizon, equals respectively to 1, 2, 3 and 4.

- These values may seem quite important: a significant part of the pairs closes before the end of the allowed trading period (5, 10, 15 or 20 days). This of course occurs more often when the market exhibits a period of great volatility because the closing thresholds, in the application, are constant. A second point to note is that the way stocks/pairs are ranked plays a major role when we look at the results for the different values of p, and the number of long and short positions. As mentioned earlier, the performance values in the decision matrix are spreads between forecasts. This approach increases the probability that

[8] The share price was 6.32$ on 18 July, 0.88 on the 25th and 4.15 on 1 August.

stocks/pairs with a high level of volatility will first. As a consequence, when p is weak, the proportion of pairs reaching the threshold is higher.

- As underlined in the first comments, when the forecast horizon, h, equals 3 or 4, designs with only the top five long and short positions do not overcome the situations with $p = 10$ or 15. This could indicate that the predictability of top pairs/stocks, with the performance values used in the decision matrix, is very short term, 2 weeks at most.
- Among the pairs closing because the threshold is reached, whatever the case, more than 50% lead to a positive profit.

When the forecast horizon equals 1, even with closing thresholds (10% or more), the mean opening time is very close to the maximal allowed time. On the other hand, when h increases the trend is not proportionate, especially if the closing threshold is weak. For example, with $p = 10$ and a 10% closing threshold, when h changes from 3 to 4, the mean opening time only increases by about 2 days, from 11, 21 to 13, 29.

As the returns per pair are far from increasing proportionally with the forecast horizon, the mean daily return of the portfolio decreases with h. When h equals 1, the results appear very promising with a mean daily return of the portfolio greater than 0.09% with 10 long and short positions or less. If we now consider the case mentioned earlier, $p = 10$ and a 10% closing threshold, but $h = 4$, the mean daily return of the portfolio falls to 0.034%.

Surprisingly, the only designs where the proportion of days with positive excess returns is (slightly) lower than 50% occurs when $h = 1$ and $p = 5$: these are nevertheless the most profitable in terms of return. A positive skewness of the series of daily returns explains this point. On the other hand, among the different designs, the proportion of days with positive returns never exceeds 52.63%. The proportions of days with positive returns are significantly greater than 50% at a 5% rate if they exceed 52.1%. Only three of the 36 designs exceed this threshold with about 2000 trading days.

In the tables of results, the diversification is illustrated by the number of pairs and the number of different stocks in the portfolio. We recall that in the application, if, for example, we open five long and short positions, it corresponds to 25 pairs (5^2). More specifically, if we look at the minimal values when $h = 1$, $p = 5$ and with a 10% closing threshold, we get 3 for the number of pairs and 4 for the number of stocks. This means that the long or the short part of the portfolio is, in fact, the same asset. A similar situation occurs when $h = 2$. The pairs are equally-weighted in the portfolio but the weight of each stock, in absolute value, is not necessarily the same.

Fig. 1 provides some indications concerning the way the performance progresses over time. We recall that the returns (per pair, for the portfolio) computed in this application are far from sufficient to obtain the real return of the portfolio, owing to the fact that new assumptions need to be made. The aim of this graph is to extract possible trends. The figure presents only one design ($p = 10$ and a 10% closing threshold) but for the four possible forecast horizons. By way of comparison, the progress of the S&P 100 during the trading period (1999–2006) is also reported.

All series are normalized to one at the beginning of 1999. The four curves are based on the daily returns of the portfolios. If R^d is the daily return of a portfolio at time d, as defined before, R^1 is the first daily return we consider, I_t is the value of the index at time t:

$$I_t = \prod_d^t (1 + R^d).$$

The key informations provided in Fig. 1 are as follows:

- The "long-term" forecasts, in our application, the design with $h = 4$, are clearly dominated from the beginning.
- During the first three years, the strategies with $h \leqslant 3$ perform well. The $h = 2$ design even outperforms the case with the shortest forecast horizon during a large part of this subperiod.
- The split takes place around 2003 when the index doubles for $h = 1$ and stays relatively stable for $h = 2$ or 3. While it is true that the portfolio with $h = 1$ is the least diversified, the difference cannot be explained by a unique day/outlier. There are at least 14 stocks and 31 pairs open in the portfolio as indicated in Table 3. The reasons for this over-performance are unknown.
- In the last three years (2004, 2005, 2006), the performances, regardless of h ($h = 4$ included), are very similar and weak. The indexes show a rise of only 10/15%. This strong fall in the results is perhaps consistent with the conclusions of Do and Faff (2008).
- As we might suppose, the correlations between the returns of the four strategies are positive and highly significant. Using weekly returns, they are between 0.15 and 0.46. The correlations are highest when there is a difference of only one between the forecast horizon. Furthermore, even if the trading systems have not been developed to be market-neutral, we found that the correlation between the returns of the strategies and the return of the S& P 100 index is not significantly different from 0.

Fig. 2 is composed of 12 graphs which represent the 36 situations (three values for the number of stocks, four forecast horizons and three types of closing thresholds). It details the return captured per pair, in mean, day by day, from opening to closing. If a pair reaches the closing threshold and is closed, this last return is repeated until the end of the maximal allowed trading period. If closing thresholds are introduced, there is consequently a certain inertia in the results; this explains why the solid lines corresponding to the strategies with a 10% closing threshold are less volatile. The last value of each curve is, by construction, already given in Tables 3–6 as the mean return per pair. Additional conclusions provided by this figure compared to the four tables of results are as follows:

- In nearly all the designs, the trend of excess returns captured per pair through time is positive. Furthermore if we consider that the accuracy/relevance of the forecasts decreases, based on the return per pair, with the forecast horizon, as we could conclude from our empirical results, we need to question one of the assumptions made in this work. We have assumed that the maximal allowed trading period for one pair was equivalent to the forecast horizon: this is perhaps not optimal. It might be worth investigating whether the results would not be better if the allowed trading period was longer than the forecast horizon. Using a one week-ahead forecast (five trading days), the pair could stay open for 10 or 20 days.
- Gatev et al. (2006) proposed a variant to their main pairs trading system based on a distance criterion. They postpone the trades to the day following the crossing/divergence. The monthly excess return falls, in that case, from 1.44% to 0.90% for what they call the "top 20 portfolio". They consider this fall in the excess returns as an estimate of the average bid-ask spread and hence the transaction costs of trading in their sample. Although actual transaction costs may be different, it is informative to know whether the trading profits are large enough to survive this estimate of transaction costs. For our article, the details provided in Fig. 2 are especially informative. If we delayed our transactions by one day, it would not change our results to any large degree.

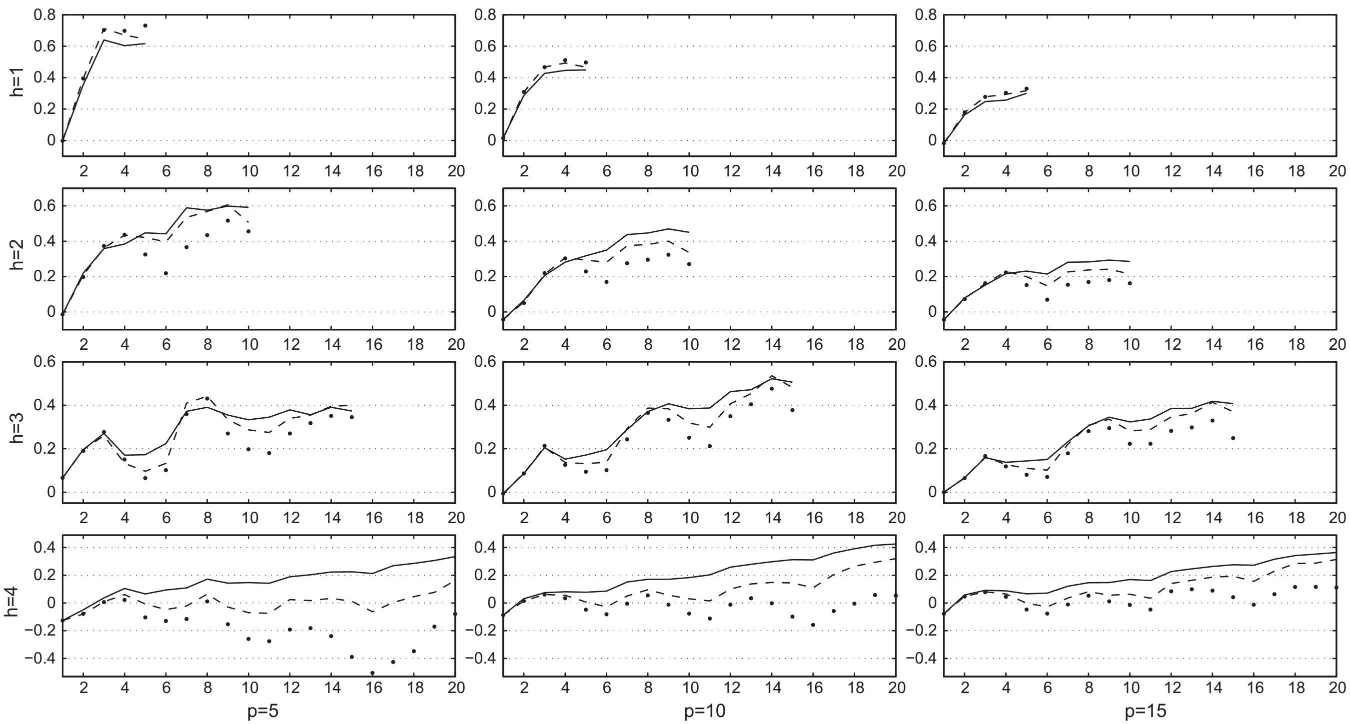

Fig. 2. Daily cumulative returns (%) of pairs for various pairs trading strategies. The solid lines refer to strategies with a 10% closing threshold, the dashed lines to a 20% closing threshold and the dotted line to the absence of threshold.

5. Conclusion

This article deals with an equity long-short trading strategy called pairs trading. We extend the literature on a particular approach of pairs trading based on forecasting and multi-criteria decision methods. This framework remains very general, positive and flexible. It is defined without reference to any equilibrium model. The process can be divided into three steps: the forecasting of spread returns between all pairs; the ranking/selection of stocks/pairs; and the trading of pairs in real time. Pairs are made up of the first and last assets of the ranking.

The most important contribution of this paper is the introduction of multi-step-ahead forecasts. The methodological and empirical implications are discussed. In an application considering S&P 100 stocks and up to four week-ahead forecasts, the main performance indicators (return per pair, daily return of the portfolio) are promising especially in the very short term. A negative relationship between the number of pairs and performance is observed, especially for a forecast horizon of 1 or 2 weeks. It indicates that this ranking-based method could detect potential pockets of predictability.

The introduction of multi-step forecasts makes the trading system much more realistic: pairs can be opened until 20 days (4 weeks) against only 5 days in previous research. Furthermore, closing thresholds have also been added to the trading system.

The detailed examination of the results raises a number of new questions:

- The trading system is active for 8 years, from 1999 to 2006. The first two/three years are especially good. After 2002, the performances are weaker. Does this mean that the markets are more efficient?
- Understanding of the sources (liquidity, news, co-integration, non-linearities, etc.) of profit is crucial. This point needs to be explored in detail and compared with the two other mains approaches.
- A joint approach involving our own and the other two is of course necessary and will be performed when the design has been improved in order to make a fair comparison possible. At least two points need to be modified in further research: the approach to ranking and selecting stocks, in this article favours highly volatile assets; furthermore the maximal holding period is probably not optimal. We have seen that when the maximal holding period equals the forecast horizon, the trend of the return per pair during the holding period is, in general, positive. This means that even with a one week ahead trading indication, it might be better to leave the pairs open for a longer period.

There are several possible extensions and improvements to the selection process. One is to use risk adjusted values in the decision matrix. Another would be to define a larger decision matrix in order to integrate forecasts defined over several horizons.

In the same spirit, a co-integration-based approach could be combined with our method. Additional criteria in the decision matrix could indicate whether two stocks have diverged: the higher this criterion in absolute value, the higher the probability that the stocks of this pair should be selected. The practical problems (computation time, etc.) discussed in the paper would increase.

References

Andrade, S., Di Pietro, V., Seasholes, M., 2005. Understanding the profitability of pairs trading. Working Paper, UC Berkeley Haas School.

Bossaerts, P., 1988. Common nonstationary components of asset prices. Journal of Economic Dynamics and Control 12, 347–364.

Chevillon, G., 2007. Direct multi-step estimation and forecasting. Journal of Economic Surveys 21 (4), 746–785.

Christoffersen, P., Diebold, X., 2006. Financial asset returns, direction-of-change forecasting, and volatility dynamics. Management Science 52 (8), 1273–1287.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Mathematical Control Signals System 2, 303–314.

DeBondt, W., Thaler, R., 1985. Does the market overreact? Journal of Finance 40, 793–805.

Do, B., Faff, R., 2008. Does nave pairs trading still work? Working Paper, Monash University.

Do, B., Faff, R., Hamza, K., 2006. A new approach to modeling and estimation for pairs trading. Working Paper, Monash University.

Duygun Fethi, M., Pasiouras, F., 2010. Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. European Journal of Operational Research 204 (2), 189–198.

Elliott, R., van der Hoek, J., Malcolm, W., 2005. Pairs trading. Quantitative Finance 5 (3), 271–276.

Elman, J.L., 1990. Finding structure in time. Cognitive Science 14, 179–211.

Engelberg, J., Gao, P., Jagannathan, R., 2008. An Anatomy of Pairs Trading: The Role of Idiosyncratic News, Common Information and Liquidity (November 17, 2008). Available at SSRN: <http://ssrn.com/abstract=1330689>.

Engle, R., Granger, C., 1987. Co-integration and error correction: Representation, estimation, and testing. Econometrica 55 (2), 251–276.

Erhman, D.S., 2006. The Handbook of Pairs Trading: Strategies Using Equities, Options and Futures. John Wiley & Sons.

Figueira, J., Greco, S., 2004. A bi-polar ELECTRE III method, manuscript, 24p.

Figueira, J., Greco, S., Ehrgott, M., 2005a. Multiple Criteria Decision Analysis: State of the Art Surveys. Springer Science Business Media, Inc., New York.

Figueira, J., Mousseau, V., Roy, B., 2005b. ELECTRE methods. In: Figueira, J., Greco, S., Ehrgott, M. (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys. Springer-Verlag, Boston, Dordrecht, London, pp. 133–162.

Galenko, A., Popova, E., Popova, I., 2007. Trading in the presence of cointegration. Available at SSRN: <http://ssrn.com/abstract=1023791>.

Gatev, E., Goetzmann, W., Rouwenhorst, G., 1999. Pairs trading: Performance of a relative value trading arbitrage rule. Working Paper, Yale School of Management.

Gatev, E., Goetzmann, W., Rouwenhorst, G., 2006. Pairs trading: Performance of a relative value trading arbitrage rule. The Review of Financial Studies 19 (3), 797–827.

Hecht-Nielsen, R., 1990. Neurocomputing. Addison-Wesley, Menlo Park, CA.

Hinkle, D., 1965. The change of personal constructs from the viewpoint of a theory of construct implications. Ph.D. Dissertation, Ohio State University.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.

Huck, N., 2009. Pairs selection and outranking: An application to the *S&P* 100 Index. European Journal of Operational Research 196 (2), 819–825.

Jacobs, B., Levy, K., Starer, D., 1999. Long-short portfolio management: An integrated approach. Journal of Portfolio Management 25 (2), 23–32.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. Journal of Finance 48, 65–91.

Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231–254.

Kang, S., 1991. An investigation of the use of feedforward neural networks for forecasting. Ph.D. Thesis, Kent State University.

Keeney, R.L., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. John Wiley & Sons, New York.

Kline, D., 2004. Multi-step time series forecasting with Neural Networks. In: Zhang, G.P. (Ed.), Neural Networks in Business Forecasting. Idea Group Inc., Hershey, PA, USA, pp. 226–250.

Leitch, G., Tanner, J., 1995. Professional economic forecasts: Are they worth their costs? Journal of Forecasting 14, 143–157.

Lin, Y.-X., McRae, M., Gulati, C., 2006. Loss protection in pairs trading through minimum profits bounds: A cointegration approach. Journal of Applied Mathematics and Decision Sciences, 1–14.

Lippmann, R.P., 1987. An introduction to computing with neural nets. IEEE ASSP Magazine (April), 4–22.

Papadakis, G., Wisocky, P., 2008. Pairs trading and Accounting information. Working Paper, MIT Sloan School of Management.

Rogers, M., Bruen, M., 1998. A new system for weighting environmental ELECTRE III. European Journal of Operational Research 107, 552–563.

Roy, B., 1978. Electre III: Un algorithme de classements fond sur une reprsentation floue des prfrences en prsence de critres multiples. Cahiers du Centre d'Etudes de Recherche Oprationnelle 20 (1).

Roy, B., and D. Bouyssou., 1993, Aide multicritre la dcision: Mthodes et cas, Economica, Paris.

Sharda, R., 1994. Neural Networks for the MS/OR analyst: An application bibliography. Interfaces 24, 116–130.

Spronk, J., Steuer, R.E., Zopounidis, C., 2005. Multicriteria decision aid/analysis in finance. In: Figueira, J., Greco, S., Ehrgott, M. (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys. Springer Science Business Media, Inc., New York.

Steuer, R.E., Na, P., 2003. Multiple criteria decision making combined with finance: A categorized bibliographic study. European Journal of Operational Research 150, 496–515.

Swanson, N., White, H., 1995. A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. Journal of Business and Economic Statistics 13 (3), 265–275.

Swanson, N., White, H., 1997. Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. International Journal of Forecasting 13, 439–461.

Tang, Z., Fishwick, P.A., 1993. Feedforward neural nets as models for time series forecasting. ORSA Journal on Computing 5 (4), 374–385.
Timmermann, A., 2006. Forecast Combinations. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting. North Holland, Amsterdam, pp. 135–196.
Vidyamurthy, G., 2004. Pairs Trading, Quantitative Methods and Analysis. John Wiley & Sons, Canada.
Wallenius, J., Dyer, J., Firshburn, P., Steuer, R., Zionts, S., Deb, K., 2008. Multiple criteria decision making and multiattribute utility theory: Recent accomplishments and what lies ahead. Management Science 54, 1336–1349.
Whistler, M., 2004. Trading Pairs – Capturing Profits and Hedging Risk with Statistical Arbitrage Strategies. John Wiley & Sons.
Wong, F.S., 1991. Time series forecasting using backpropagation neural networks. Neurocomputing 2, 147–159.
Xidonas, P., Psarras, P., 2009. Equity portfolio management within the MCDM frame: A literature review. International Journal of Banking, Accounting and Finance 1 (3), 285–309.
Zhang, G., Patuwo, E., Hu, M., 1998. Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting 14, 35–62.
Zopounidis, C., 1999. Multicriteria decision aid in financial management. European Journal of Operational Research 119 (2), 404–415.