# The application of neural networks to the pricing of credit derivatives

Eliana Angelini and Alessandro Ludovici*[1]
University "G. d'Annunzio", Chieti-Pescara ISTAO, Ancona

**Introduction**

The recent history of financial markets shows how, to the impetuous development of the financial innovation process, which has invested all the structural components of the same, has been associated the constant engagement of the operators in finding more efficient computational methodologies, able to be an effective dynamic support of the analysis. Growing concerns about credit risk have created the need for sophisticated credit risk analysis and management tools. Credit risk measurement models and credit risk management tools are both of significant importance to the further development of credit markets.

Credit derivatives are flexible and efficient instruments that enable users to isolate and trade credit risk. Credit derivatives allow users to isolate credit risk from other quantitative and qualitative factors associated with owing an exposure. Hence, they can be used to transfer and hedge credit risk in an efficient and flexible manner, customized to a client's requirements. This transfer of credit risk may be complete or partial, and may be for the life of the asset or for a shorter period. Credit risk includes not just default or insolvency risk but also changes in credit spreads and thereby market values, changes in credit ratings and generic changes in credit quality. Credit derivatives can be used when a sale in the cash market is either not efficient or not possible. Even when cash market alternatives exist, credit derivatives may be preferred because they do not require funding. Furthermore, since derivatives are over-the-counter contracts, transactions are confidential. Finally, speed of settlement and liquidity are reasons why credit derivatives are a better alternative to the reinsurance market.

Credit derivatives are swaps, forward and option contracts, particularly credit default swaps (CDS); they can be used to hedge against all these types of credit risk. For a simple credit default swap, over some time period, one counterparty (the protection seller) receives a predetermined fee payment from another counterparty (the protection buyer); in return, the protection seller agrees that in the case of a credit event of a reference entity, it will pay the seller the loss on a bond of the reference entity, that is the bond's par value less its recovery.

A common question when considering the use of credit derivatives, as an investment or a risk management tool, is how they should correctly be priced. In general, the pricing of a credit default swap will depend on the

[1]* The authors would like to thank Andrea Roli for the precious suggestions; Elisa Coletti and Maila Bozzetto for the collaboration to the collection of data. Sole responsibility for the contents of this analysis rests with the authors.

credit quality of the reference entity and the length of the swap contract.

This article proposes a non parametric model for estimating pricing of credit derivatives, using learning networks. The recent application of nonlinear methods, such as neural networks to credit risk analysis, shows promise of improving on traditional credit models. Neural networks differ from classical credit systems mainly in their black box nature and because of they assume a non-linear relation among variables. The two main issues to be defined in a neural network application are the network typology and structure and the learning algorithm. The connections (links) among neurons have associated a weight which determines type and intensity of the information exchanged. As regards the independent variables of the model we start from the typical assumption of the structural approach of Merton, for which the relevant information in order to evaluate credit risk can be obtained from the market data of the analyzed firm. Therefore, option pricing theory is used in order to create a link between the credit market and the securities market.

The model that we are introducing turns out peculiar not only for the use of the neural network, but also for the use of the implied volatility of one-year options written on the shares of the analyzed firm, instead of historical volatility: this leads to a higher capability of getting the signals launched by the market about the creditworthiness of the firm; in fact historic volatility, being a medium value, brings in temporal lags in the evaluation.

In any case, our model differs from the structural approach for the fact that it consider the 30-month historical series for CDS spreads, and this turns out to be the main advantage of our study, as we show in the final section.

The paper is organized as follows. The paper begins (*section 1*) by stating the implications of credit derivatives in portfolio credit risk management. In *section 2* we first briefly overview the main principles and characteristics of neural networks, focusing the attention above all on the concepts that are most useful for the application to financial instruments; then we describe the pricing model we developed and tested for credit derivatives. The architecture of the neural network is feed-forward, trained for 17000 learning epochs using the back-propagation algorithm, with two hidden layers of 9 and 10 neurons each: by the study carried out it turns out obvious that neural networks are able to totally capture the variability relative to the market dynamics of credit derivatives. In *section 3*, the effectiveness of neural network in approximating the evaluation of credit derivatives is illustrated. As regards the sample, it includes 18 American firms, relative to various fields, including financial institutions which, operating typically with a high leverage due both to the activity carried out and to the laws concerning the capital of banks, usually introduce remarkable factors of distortion in parametric models. We shall show that neural networks are not affected by this problem. The temporal range embraces the period September 2002 - March 2006: we have considered the five-year CDS spread relative to each firm, for a total of 180 observations on a quarterly basis obtained through the Fitch™ database: as already pointed out, among the variables implied volatility has a determining role, in fact we have obtained a positive correlation with CDS spreads equal to 0.6338. Leverage is another key variable, obtained dividing the face value of the debt of the firm by the total of its liabilities (including the market capitalization), getting the data from the Bloomberg™ database. We have considered the risk free rate equal to the one-year constant maturity Treasury Bills yield, taken from the Federal Reserve System database. We then discuss in detail the experimental settings and the results we obtained, leading to considerable accuracy in prediction. The paper ends evidencing that, as far as this field of the financial markets is concerned, neural networks for sure constitute a valid instrument of calculation: in fact there still does not exist in literature a formula of evaluation for the CDS, able to tie the quoted spreads to the specific underlying variables of each examined firm, and the neural network can, as will be shown, satisfy this lack with high effectiveness, facing the problem of determination of the functional form from a statistical point of view. As we will show, it is easy to calculate the sensitivity of the CDS spread to each independent variable, in order to determine a statistical pricing formula for CDS.

The paper concludes with a discussion of advantages and limitations of the solution achieved.

## 1. Credit derivatives: new financial instruments

Credit derivatives are financial instruments used to transfer credit risk of loans and other assets. They are bilateral financial contracts with payoffs linked to a credit related event such as a default, credit downgrade or bankruptcy. There are various types, but the basic structures of all credit derivatives are swaps, options and forwards. Due to their high flexibility credit derivatives can be structured according to the end-users' needs. For instance, the transfer of credit risk can be effected to the whole life of the underlying asset or for a shorter time, and the transfer can be a complete or a partial one. Delivery can take place in the form of over the counter contracts or embedded in notes. Moreover, the underlying can consist of a single credit-sensitive asset or a pool of credit-sensitive assets[2].

### 1.1Credit derivatives: products and structures

The most important and widely used credit derivative is a credit default swap[3]. It is an agreement in which the one counterparty (the protection buyer) pays a periodic fee, typically expressed in fixed basis points on the notional amount, in return for a contingent payment to the other counterparty (the protection seller) in the event that a third-party reference credit defaults. A default is strictly defined in the contract to include, for example, bankruptcy, insolvency, and/or payment default. The definition of a credit event, the relevant obligations and the settlement mechanism used to determine the contingent payment are flexible and determined by negotiation between the counterparties at the inception of the transaction. Since 1991, the International swap and Derivatives association (ISDA) has

[2] See HENKE, S., BURGHOF, H.P., RUDOLPH, B., (1998).

[3] The credit default swap is also known as credit default put, credit swap, default swap, credit put or default put.

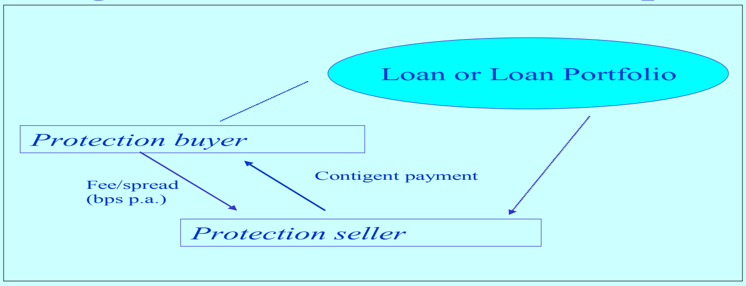
made available a standardized letter confirmation allowing dealers to transact credit swaps under the umbrella of an ISDA Master Agreement. The evolution of increasingly standardized terms in the credit derivatives market has been a major growth because it has reduced legal uncertainty that hampered the market's growth.

The contingent payment in the event of default can be identify as either:

- a payment of par by the protection seller in exchange for physical delivery of the defaulted underlying;
- a payment of par less the recovery value of the underlying as obtained from dealers;
- a payment of a binary, i.e. fixed, amount.

Credit default swaps can be viewed as an insurance against the default of the underlying or a put option on the underlying. Figure 1 exhibits the basic structure of a credit default swap.

Figure 1: Credit default swap




Also there is the total return swap, in which one counterparty (total return payer) pays the other counterparty (total return receiver) the total return of an asset (the reference obligation) for receiving a regular floating rate payment, such as Libor plus a spread. "Total return" comprises the sum of interest, fees and any change-in value payments (any appreciation or depreciation) with respect to the reference obligation.

In contrast to the credit default swap, the total return swap does not only transfer the credit risk but also the market risk of the underlying; it effectively creates a synthetic credit-sensitive instrument. A total return swap allows an investor to enjoy all of the cash flow benefits of a security without actually owing the security.

Credit spread option is an option on a reference credit's spread in the loan or bond market. In a spread put option one party pays a premium for the right to sell a bond to a counterparty at a certain spread at a definite time in the future. A credit spread option gives the buyer protection in the event of any unfavourable credit migration. In a default option, the asset can be put only on default. The credit spread is the yield differential between the reference credit and a pre-determined benchmark rate. Thus, in credit spread derivatives, payment is based on the movement of the value of one reference credit against another. A ratings or downgrade option is a credit spread option that pays out if a specified company's rating is downgraded. This type of an option is sometimes embedded in bond structures.

Finally, credit linked notes are created by embedding credit derivatives in notes. Credit derivatives have the advantage that funding is not necessary, whereas credit linked notes have the benefit of avoiding counterparty risk. Credit linked notes are frequently issued by special purpose vehicles (corporations or trusts) that hold some form of collateral securities financed through the issuance of notes or certificates to the investor. The investor receives a coupon and par redemption, provided there has been no credit event of the reference entity. The vehicle enters into a credit swap with a third party in which it sells default protection in return for a premium that subsidizes the coupon to compensate the investor for the reference entity default risk.

## 1.2. **Fundamental attractions of using credit derivatives**

In theory, credit derivatives are tools that enable financial operators to manage their portfolio of credit risks more efficiently; they enable market participants to devise flexible personal approaches to the management of credit risk associated with a variety of underlying financial assets. The promise of these important instruments has not escaped regulators and policymakers. "Credit derivatives and other complex financial instruments have contributed to the development of a far more flexible, efficient, and hence resilient financial system than existed just a quarter-century ago" (Greenspan, 2004).

The credit derivatives market offers its users a range of tools which enable the transfer of credit risk. A brief review of the available products reveals that in most cases one party to a transaction receives a fee and commits to provide the other party with a payment should the credit quality of a third party deteriorate. Whilst the mechanism contained in these products are easy to understand, the broad range of applications is not immediately obvious.

The users of the risk-management benefits of credit derivatives tend to be quite diverse. An increasingly important user group includes financial institutions, corporates and fund managers. Financial institutions have embraced the full range of benefits; the use of credit derivatives by banks has been motivated by the desire to improve portfolio diversification and to improve the management of credit portfolios. Corporates are also looking to reduce the credit exposure to key trading partners and specifically they are interested in using credit derivatives to isolate credit risks in project financing. For fund managers, although the asset benefits of credit derivatives still suffer from lack of liquidity, the use of structures that hedge out spread risk has some appeal.

This paragraph focus on a range of uses for credit derivatives and divides them between credit risk management and asset opportunities[4].

### 1.2.1 Using credit derivatives for managing credit risk

The principal feature of these instruments is that they separate and isolate

[4] For more detailed information on the characteristics of credit derivatives see DAS, S., (1998); TAVAKOLI, J.M., (1998).

credit risk facilitating the trading of credit risk with the purpose of:

- replicating credit risk;
- transferring credit risk;
- hedging credit risk.

In practice, the rationale behind a transaction may relate to the management of credit lines, to regulatory capital offsets, to balance sheet optimization, portfolio hedging and diversification or pure risk reduction itself. Credit derivatives can be used as a risk management tool by portfolio managers to:

- Achieve portfolio diversification: credit derivatives can be used to achieve portfolio diversification by allowing access to previously unavailable credits. They can also be used to diversify across a range of borrowers and to gain exposure to an asset without owing it.
- Reduce concentration risk: investors can reduce portfolio credit risk concentrations using derivatives structures; they can thus manage country and industry risks. Reducing credit concentration in loan portfolios is commonly viewed as the main use of credit derivatives. However, to date credit derivatives are generally referenced to assets which are widely traded, i.e. for which market prices are readily available, or for which a rating by an international agency is at hand.
- Manage exposures while maintaining client relationships. Changes to credit risk management in the banking sector are an additional factor contributing to greater use of credit derivatives. Investors can use credit derivatives to reduce exposures without selling them. This effectively frees up credit lines, allowing more business to be done with a customer. Furthermore, a bank that is concerned about credit loss on a particular loan can protect itself by transferring the risk to someone else while keeping the loan on its books. As part of their credit risk management, banks are viewing credit derivatives more and more often as tradable products, which can be transferred to third parties before the maturity date[5].
- Manage regulatory capital: the new supervisory rules provided for by Basel II are also increasing the incentives for banks to use credit derivatives. Where guarantees or credit derivatives are direct, explicit, irrevocable and unconditional, and supervisors are satisfied that banks fulfil certain minimum operational conditions relating to risk management processes they may allow banks to take account of such credit protection in calculating capital requirements. A guarantee or credit derivative must represent a direct claim on the protection provider and must be explicitly referenced to specific exposures or a pool of exposures, so that the extent of the cover is clearly defined and incontrovertible. Other than non-payment by a protection purchaser of money due in respect of the credit protection contract it must be irrevocable; there must be no clause in the contract that would allow the protection provider unilaterally to cancel the credit cover or that would increase the effective cost of cover as a result of deteriorating

[5] See DUFFEE, G.R., ZHOU, C., (2001), STULTZ, R., (2003), MINTON, B.A., STULTZ, R., WILLIAMSON, R., (2005).

credit quality in the hedged exposure. It must also be unconditional; there should be no clause in the protection contract outside the direct control of the bank that could prevent the protection provider from being obliged to pay out in a timely manner in the event that the original counterparty fails to make the payment due. There are cases where a bank obtains credit protection for a basket of reference names and where the first default among the reference names triggers the credit protection and the credit event also terminates the contract. In this case, the bank may recognise regulatory capital relief for the asset within the basket with the lowest risk-weighted amount, but only if the notional amount is less than or equal to the notional amount of the credit derivative. In the case where the second default among the assets within the basket triggers the credit protection, the bank obtaining credit protection through such a product will only be able to recognise any capital relief if first-default-protection has also be obtained or when one of the assets within the basket has already defaulted[6].

### 1.2.2 Asset opportunities

Credit derivatives have evolved to become an important financial asset class. As already argued, credit derivatives enable credit risk to be separated from the funding component of its underlying instrument; as it is often the form of the underlying instrument that creates obstacles for the investor, this separation of the credit risk creates important opportunities. The decision to use the asset opportunities of credit derivatives tends to be based on one of the following needs:

- Access to new markets: investors can create new assets with a specific maturity not currently available in the market;
- Obtain tailored investments: credit derivatives can be used to create instruments with exact risk- return profile sought. Maintaining diversity in credit portfolios can be challenging. This is particularly true when the portfolio manager has to submit with constraints such as currency denominations, listing considerations or maximum or minimum portfolio duration. Credit derivatives are being used to address this problem by providing tailored exposure to credits that are not otherwise available in the wished form or not available at all in the cash market.
- Improve the risk-return profile of portfolios: credit derivatives offer new possibilities of turning a given market opinion into an investment strategy. This particularly entails assumption of specific types of credit risk without the acquisition of the asset itself. Instead of purchasing a specific bond, a market participant who considers some credit risks to be overvalued can earn an attractive premium as a protection seller in the credit default swap market. Premiums are generated without having to tie up any capital for the purchase of a bond issue (at least as long as no credit event occurs). On the other hand, market participants who consider risks to be underestimated can purchase protection by paying a premium. Owing to the limited possibilities for short sales in the

[6] See BANK FOR INTERNATIONAL SETTLEMENT, (2005).

bond market, hedge funds are increasingly entering into positions in credit derivative market to implement their financial strategies. In particular:

- to hedge dynamic risks: exposures that change with market movements can be hedged using credit derivatives;
- to manage illiquid credits: credit derivatives can be utilized to actively manage risk in large illiquid loans portfolios;
- to execute short credit positions: credit derivatives can be employed to execute short credit positions without the risk of a short squeeze or high financing costs. Hence, investors can use them to hedge or take advantage of deteriorating credit qualities;
- to hedge declining credit quality: default and spread options and swaps can be used to hedge failing credit qualities. Credit spread options and swaps can be used to hedge fluctuations in credit spreads and not have to wait for default to get a payout.

## 2 Neural networks: architecture and applications

Neural networks have been used in different fields of study, such as engineering, medicine, physics and others. Although the relative structures differ remarkably with one another, it is possible to point out some fundamental principles regarding essentially the functioning of such operative instruments. Moreover, it is important to start the treatment emphasizing that, in order to analyze the financial dynamics, relatively little complex networks are effective, at least compared to those of other fields[7].

### 2.1 Architecture of neural networks

A neural network relates a set of input variables $\{x_i\}$, $i$=1,2,..$k$ to a set of one or more output variables $\{y_j\}$, $j$=1,2,..$h$. An essential characteristic of a neural network, differently from other methods of approximation, is that it uses one or more hidden layers, in which the input variables are transformed by a logistic or logsigmoid function: this characteristic, as shown later, gives to these instruments a particular efficiency in modeling nonlinear statistical processes.

In the feedforward neural network parallel elaboration is associated to the typical sequential elaboration of the linear methods of approximation. In fact while in the sequential elaboration particular weights are given to the input variables through the neurons of the input layer, in the parallel one the neurons of the hidden layer operate further transformations in order to improve the predictions. The connectors (between the input neurons and the neurons in the hidden layers, and between these and the output neurons) are called synapses. The feedforward neural network with a single hidden layer is the simplest and at the same time the most used network in the economic and financial field.

[7]DOLCINO, F., GIANNINI, C., ROSSI, E., (1998). For a useful description of the phenomenon in general terms, see FLOREANO, D., NOLFI, S., (1993) and GORI, M., (2003).

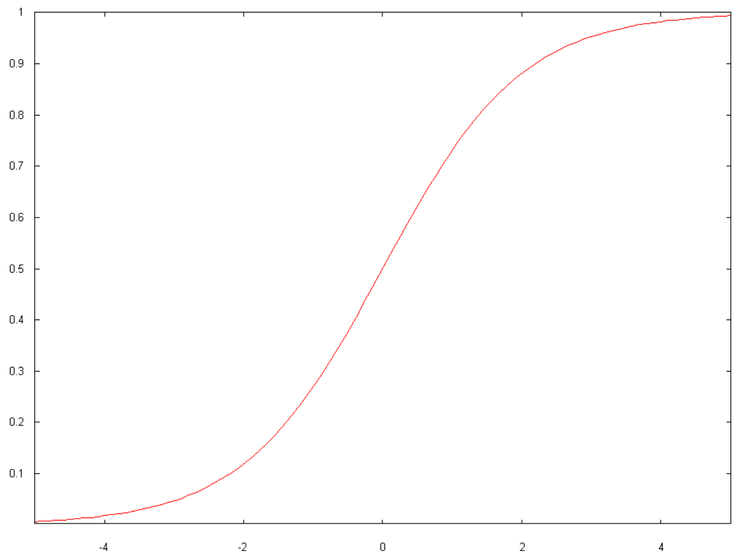



Fig. 2.1 Logsigmoid function.

Therefore the neurons process the input variables in two ways: firstly forming linear combinations and lastly transforming these combinations through a particular function, typically the logsigmoid function, illustrated in fig. 2.1. An essential characteristic of this function is the threshold behavior near values 0 and 1, which turns out to be particularly suitable to economic problems, which usually, for very high (or very low) values of the independent variables, show little changes in response to small changes of the variables. At the analytical level, the neural network can be described by the following equations[8]:

$$n_{k,t}=\omega_{k,0}+\sum_{i=1}^{m}\omega_{k,i}x_{i,t}$$

$$N_{k,t}=L(n_{k,t})=\frac{1}{1+e^{-n_{k,t}}}$$

$$y_t=\Upsilon_0+\sum_{k=1}^{q}\Upsilon_k N_{k,t}$$

where $L(n_{k,t})$ represents the logsigmoid activation function. It is a system with $m$ input variables $x_i$ and $q$ neurons. A linear combination of these input variables, observed at time $t$, with the weights of the input neurons $\omega_{k,i}$ and the constant term (*bias*) $\omega_{k,0}$ forms the variable $n_{k,t}$. Then this variable is transformed by the logistic function and becomes the neuron $N_{k,t}$ at time or observation $t$. The set of $q$ neurons at time or observation $t$ is therefore linearly combined with the coefficient vector $k$ and added to the constant term $\omega_{k,0}$ in order to obtain the output $y_t$ concerning time or observation $t$, representing the prediction of the neural network for the analyzed variable. The feedforward neural network used with the logsigmoid activation function is often called multi-layer perceptron or MLP network. A highly complex problem could be treated widening this structure, and therefore

[8]MCNELIS, P.D. , (2005).

using two (respectively N and P) or more hidden layers[9]:

$$n_{k,t}=\omega_{k,0}+\sum_{i=1}^{m}\omega_{k,i}x_{i,t}$$

$$N_{k,t}=L(n_{k,t})=\frac{1}{1+e^{-n_{k,t}}}$$

$$p_{l,t}=\rho_{l,0}+\sum_{k=1}^{s}\rho_{l,k}N_{k,t}$$

$$P_{l,t}=\frac{1}{1+e^{-p_{l,t}}}$$

$$y_t=\Upsilon_0+\sum_{l=1}^{q}\Upsilon_l P_{l,t}$$

Adding another hidden layer increases the number of parameters (weights) to be estimated by the factor ($s$+1) ($q$-1) + ($q$+1), since the net with a single hidden layer, with $m$ input variables and $s$ neurons has ($m$+1)$s$ + ($s$+1) parameters, while the same net with two hidden layers and $q$ neurons in the second hidden layer has ($m$+1)$s$ + ($s$+1)$q$ + ($q$+1) parameters. However the disadvantage of these models for complexity does not consist of the number of parameters, which in any case use up degrees of freedom if the sample size is limited and requires a longer training time, but of the greater probability that the net converges to a local rather than global optimum. Anyway it has been demonstrated that a neural network with two layers is able to approximate any nonlinear function[10]. A further quality of this instrument consists exactly of the fact that it does not just approximate a phenomenon on the basis of a presumed functional form to be adapted, but at the same time it determines the functional form and proceeds to the evaluation of the weights.

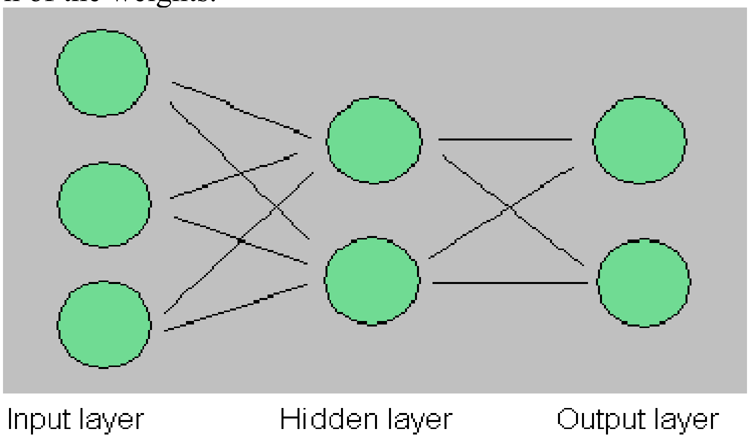



Fig. 2.2 Neural network with one hidden layer and two output neurons. (our elaboration)

In fig. 2.2 a net with a multiple number of output variables is illustrated. A neural network with a hidden layer and two output variables is described by the following equations:

[9]MCNELIS, P.D. , (2005).
[10]BELTRATTI, A., SERIO, M., TERNA, P., (1996).

$$n_{k,t}=\omega_{k,0}+\sum_{i=1}^{m}\omega_{k,i}x_{i,t}$$

$$N_{k,t}=L(n_{k,t})=\frac{1}{1+e^{-n_{k,t}}}$$

$$y_{1,t}=Y_{1,0}+\sum_{k=1}^{q}Y_{1,k}N_{k,t}$$

$$y_{2,t}=Y_{2,0}+\sum_{k=1}^{q}Y_{2,k}N_{k,t}$$

It is possible to observe that adding an output variable implies the evaluation of ($q$+1) parameters more, equal to the number of neurons of the hidden layer increased of one unit. Therefore adding an output variable implies an increasing number of parameters to be estimated, equal to the number of the neurons of the hidden layer, not to the input variables. Using a neural network with multiple outputs makes sense only if these are closely correlated to the same set of input variables: as an example we could mention the temporal structure of the rates of inflation or of the rates of interest. One of the most common criticisms made to these instruments is that they are substantially black boxes: questions regarding the nature of the parameters, the reasons of the choice of their number, of the number of the neurons, of the number of the hidden layers, the reasons that relate the architecture of the net to the structure of the underlying problem to be explained do not find an answer.

The risk, when models are based on a high number of parameters, is that their extreme flexibility[11], being able to explain anything and its opposite, ends up in not carrying any knowledge contribution. However, we must underline that the same criticism can be made to any statistical approximation method: therefore not only to neural networks, but also to linear models, univariate and multivariate regression and so on. Neural networks, in particular, are able to explain very irregular processes, on which it is therefore difficult to identify a precise relation of cause-effect. Therefore the black box criticism constitutes, paradoxicalally, also one of the greatest qualities of neural networks. In any case, the simplicity with which it is possible to increase the number of the parameters of the net must never make forget the importance, in any model, of the clarity of the assumptions.

## 2.2 Data scaling

A neural network is not able to analyze data or to give solutions in absolute value: especially if there are data of an unusually elevated or reduced value, problems of overflow or underflow could happen. When instead sigmoid functions are used, it becomes indispensable to preprocess data: this family of functions in fact has a codominy of type [0,1] (or [-1,1] in the case of the logsigmoid function), for which the values must be scaled to these intervals otherwise the output of the net would become useless, being equal to the

[11]HYKIN, S., (1999).

superior or inferior threshold in correspondence of all the different values higher or lower than a determined limit. In other words, for a great amount of data not standardize to the interval the neurons would simply transmit the threshold value, so a wide part of the information would be lost. As far as the methods, the linear reduction transforms the series of values $x_k$ in the serie $\hat{x}_k$ , using the following formulas:

$$\hat{x_{k,t}}=\frac{x_{k,t}-min(x_k)}{max(x_k)-min(x_k)}$$

if the range is between 0 and 1, and

$$\hat{x_{k,t}}=2\frac{x_{k,t}-min(x_k)}{max(x_k)-min(x_k)}-1$$

if the desired range is between -1 and 1, while the logaritmic reduction uses the formula

$$\hat{x_{k,t}}=\frac{\log(1+x_{k,t})}{\log max(x_k)}$$

**2.3 Learning process**

After the data have been scaled, we have to deal with the problem of the evaluation of the parameters (weights) through the process known as learning (training) of the neural network. Certainly it is a problem much more complex than the evaluation of the parameters of a linear model, as for the nature of high nonlinear complexity of neural networks. For these reasons numerous optimal solutions can exist, but they do not minimize the difference between the predictions of the net and the effective values to be evaluated. In short, in any non linear model it is necessary to begin the evaluation of the parameters on the basis of conditions which represent a guess of the value of the same. However, as it will be shown, the capability of the process of evaluation of the parameters to converge to a global optimum depends on the goodness of these initial hypothesis: in fact if it is situated near a local optimum instead of the global one[12], it is likely that the first one will be reached.

[12]DOLCINO, F., GIANNINI, C., ROSSI, E., (1998).

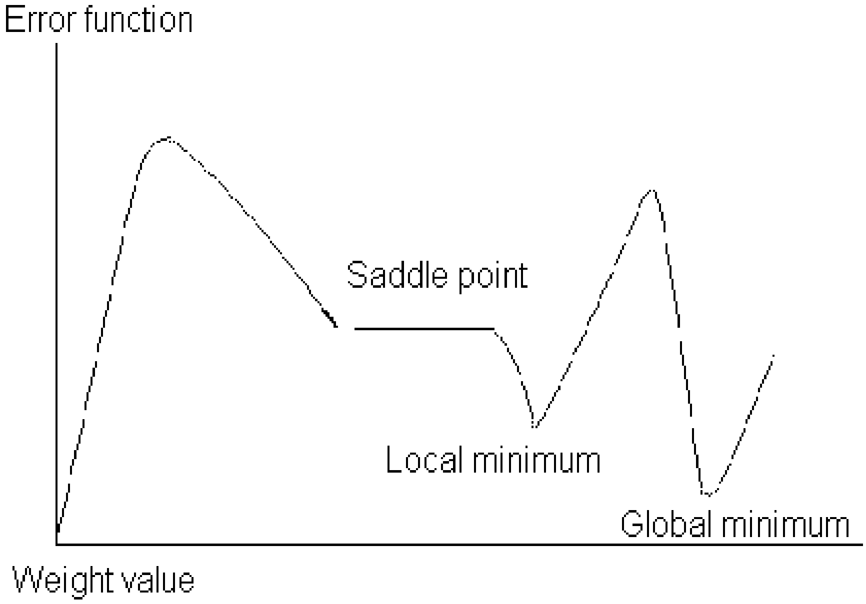



Fig. 2.3 Example of succession of local and global minimums

This is illustrated in fig. 2.3: the initial guess of the parameters (or weights of the neurons) could accidentally be situated wherever on the x-axis: if it is near a local minimum, the training process of the net would lead towards this. Later on, it will be observed that the training process of the network is completed when a point is reached in which the derivative of the loss function is null: we must remember that this condition, beyond the global optimum, identifies also the local ones and the saddle points. So it can be anticipated that if the learning coefficient, which indicates the sensibility of the net to the training process, is too low, this would lead to the impossibility of the network to escape from local optimums; while if it is too high, it could carry the training process to oscillate continuously far away from the optimum point, and therefore the network would diverge. In analytical terms, it is possible to illustrate the learning process of a net with two hidden layers, for which it is therefore necessary to determine the set of parameters $\Omega=\{\omega_{k,i}, \rho_{l,k}, \gamma_l\}$.

The problem consists of[13] the minimizing of the loss function, defined as the sum of the squares of the differences between the observed data sample $y$ and the prediction of the net $\hat{y}$:

$$min_{(\Omega)} \psi(\Omega)=\sum_{t=1}^{T}(y_t-\hat{y}_t)^2$$
$$\hat{y}_t=f(x_t;\Omega)$$

in which T is the number of the observations of the output vector $y$, and $f(x_t;\Omega)$ represents the neural network. $\Psi$ is a nonlinear function of $\Omega$ . All nonlinear optimizations begin with an initial guess about the solution and try further, better solutions until finding the best possible within a reasonable

[13] WERBOS, P., (1988).

number of iterations. Different methodologies have been proposed in order to lead this search: some make reference to complex results of logical-numerical analysis, e.g. genetic algorithms, in alternative to the classic method of the reduction of the gradient or Newton-Raphson method. In any case the chosen algorithm continues until the last iteration *n*, or in alternative a tolerance criterion can be set up, stopping the iterations when the reduction of the error function comes down a predefined tolerance value. In order to avoid local optimums, a solution could be to determine a first convergence of the process, and then to repeat it with a set of different initial parameters in order to verify whether the solution changes. Alternatively, numerous processes could be carried out to determine the best solution. However, there are the most important problems when the number of the parameters increases or the architecture of the network becomes particularly complex. Paul John Werbos proposed in the beginning of 1970's an alternative to the gradient method called backpropagation method. It is a very flexible method, able to avoid the problems caused by the evaluation of the Hessian matrix in the reduction of the gradient, and surely it is the most used method. In the passage from an iteration to the successive one in the process of evaluation of the parameters, the inverse Hessian matrix is in fact replaced by an identity matrix having dimension equal to the number *k* of the parameters, multiplied by the learning coefficient $\rho$:

$$(\Omega_1 - \Omega_0) = -H_0^{-1} Z_0 = -\rho Z_0$$

In order to avoid oscillations this coefficient is chosen in the range [0,05, 0,5] and it can also be endogenous, that is it can assume various values when the gradient comes down and the process seems to converge; or finally different coefficients for the various parameters can be adopted. However the problem of the choice of this coefficient remains, together with the existence of local minimums. Moreover, low values of the learning coefficient, although as anticipated are able to avoid oscillations, can extend uselessly the convergence of the minimizing process. This can however be accelerated adding a 'momentum' for which at iteration *n* we will have:

$$(\Omega_n - \Omega_{n-1}) = -\rho Z_{n-1} + \mu (\Omega_{n-1} - \Omega_{n-2})$$

Therefore, with $\mu$ generally equal to 0,9, the calculation of the parameters moves more fastly outside a plateau in the error surface. Now we will briefly discuss the methods used to estimate the effectiveness of the output of the net. Relatively to the evaluation of the goodness of the predictions of the net, the most common index is R-squared (goodness of fit) especially as far as the capability of the net to predict the data with which it has been trained is concerned, and the root mean squared error (Rmse) as for the capability to generalize the predictions outside the data samples used for the training; in other words, divided the sample into two parts, the first (in sample) will be used in order to train the net, and the other (out of sample), in general equal to about 25% of total data, will be used to estimate the capability of the net to predict data coming from the same population but not used for the training.

However as to the total amount of necessary data[14], undoubtedly a neural network requires the evaluation of many more coefficients than, for example, a linear model, and this leads to the necessity of a wide sample. Surely the availability of wide samples improves the predictive abilities of the net, but it also implies longer training times. Moreover, the availability of a wide sample not always is a positive aspect, especially in the financial field where using very old data brings distortions in the models, because they tend to vary with extreme rapidity and therefore very remote data are no more in any relations with the present ones.

## 3 Application of neural networks to Credit Derivatives

In this section the potentialities of neural networks in the approximation of the pricing of credit derivatives will be shown using real market data, collected from Fitch™ and Bloomberg™ data bases.

Starting from September 2002, we have collected on a quarterly basis data regarding 5-year maturity CDS spreads of 18 companies from various economic sectors, together with data concerning the leverage of the firms, the implied volatility of 1-year maturity call options written on the equities of the firms, and the risk free rate assumed to be equal to the 1-year constant maturity Treasury Bill yield. As regards the recovery rate, we have used the most commonly values adopted by the operators to price CDS, depending on the economic sector to which the reference entity belongs to. In the following diagrams we show the sample collected until March 2006, therefore covering 14 quarters.

**Sample description**

| N | Ticker | Name | Market Cap. (bln $) |
|---|---|---|---|
| 1 | AA | ALCOA Inc. | 30,18 |
| 2 | BA | Boeing Company (The) | 71,91 |
| 3 | CCL | Carnival Corporation | 30,13 |
| 4 | COX | Cox Communications Inc. * | 5,9 |
| 5 | CTX | Centex Corporation | 6,15 |
| 6 | CVS | CVS Corporation | 26,96 |
| 7 | CZN | Citizens Communications Corporation | 4,81 |
| 8 | FD | Federated Department Stores Inc. | 23,16 |
| 9 | GPS | Gap, Inc. (The) | 16,23 |
| 10 | IBM | International Business Machines Corporation | 149,11 |
| 11 | JPM | JPMorgan Chase & Co. | 177,41 |
| 12 | JWN | Nordstrom Incorporated | 15,03 |
| 13 | LEH | Lehman Brothers Holdings Inc. | 43,46 |
| 14 | LEN | Lennar Corporation | 6,74 |
| 15 | MAR | Marriott International, Inc. | 19,51 |
| 16 | MCD | McDonald's Corporation | 56,05 |
| 17 | SBC | AT&T Inc. | 233,83 |
| 18 | TXT | Textron Financial Corporation | 12,21 |

* Company was delisted on December, 9th 2004. This fact does not affect in any way our results.

Tab. 3.1 Details of the companies included in the sample

[14] DOLCINO, F., GIANNINI, C., ROSSI, E., (1998), where the concepts of "evaluation error" and "approximation error" are analyzed.

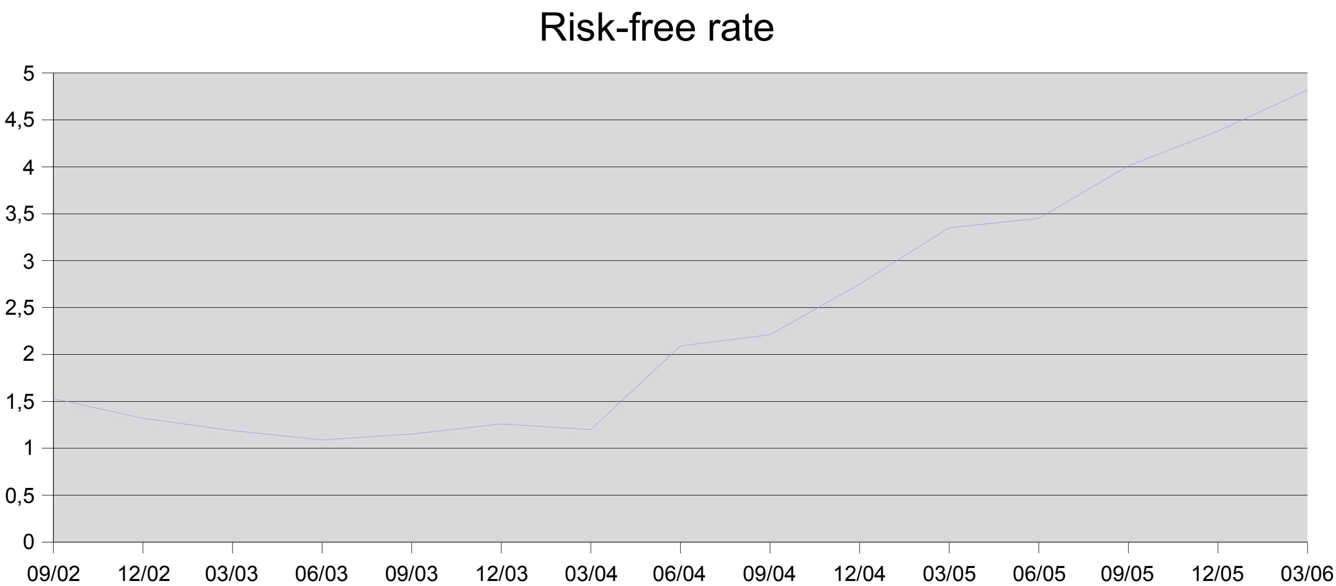



Fig. 3.1 Risk free rate during the period covered by our study  Source: Federal Reserve System

| Economic sector | Recovery rate |
|---|---|
| Hotel chains | 0,26 |
| Department stores | 0,33 |
| Finance | 0,36 |
| Telecommunications | 0,37 |
| Constructions | 0,39 |
| Metal and mechanic | 0,42 |
| Food | 0,45 |

Tab. 3.2 Recovery rates  Source: Altman and Kishore (1996)

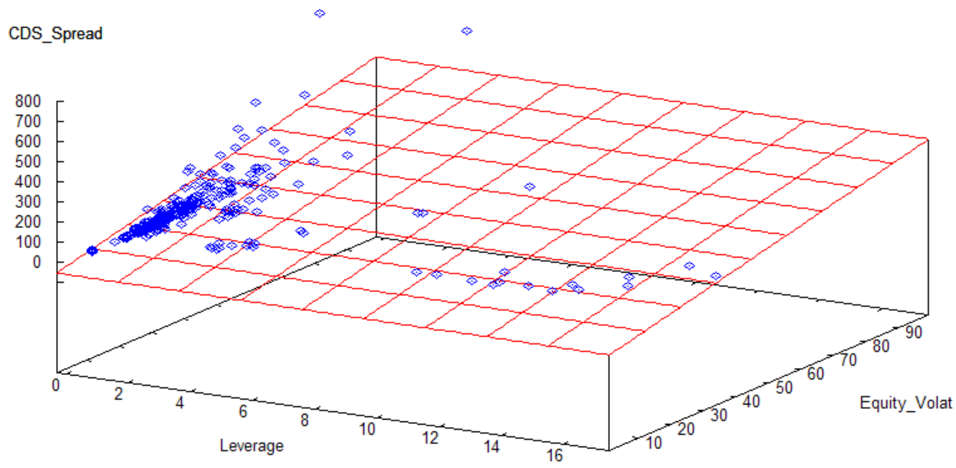



Fig. 3.2 Relationship between CDS Spread, Leverage and Equity volatility in our sample

As regards the risk free rate, we must consider that a portfolio made up of a risky bond with yield equal to *i* and a CDS written on it with a spread equal to *sp* is virtually free of any credit risk, so its yield must be equal to the risk free rate; therefore we have the following approximation:

$$r_f = i - sp$$

showing an inverse relationship between *sp* and *rf*, confirmed by market data. We have the following correlation values:

| ***Variable*** | ***Corr. with CDS Spread*** |
|---|---|
| rf | -0,2187 |
| R | -0,1475 |
| Lev | -0,0485 |
| Vol | 0,6338 |

Of course we can notice a negative correlation with R (the recovery rate) and a strong positive correlation with Vol (the implied volatility which in our study proves to be very effective in predicting creditworthiness deterioration). The absence of a correlation with the leverage should not seem strange: our sample in fact includes financial companies too, which typically have a very high gearing ratio and a low CDS spread due to prudential regulation: in any case the neural network can solve this problem very well because of its nonparametric capabilities. Without considering the financial firms, the correlation of Lev and Sp would rise to 0,317.

The sample is made up of companies coming from different economic sectors, as it is easy to catch reading the recovery rates applied: of course we consider only big (or at least medium)-caps, the only ones for which a liquid market for CDS exists. In fig. 3.2 we show the relationship between CDS spread, Leverage and Equity volatility. It is evident that there is no linear relation between them. Moreover, only a few data are characterized by a leverage of more than 2: of course these can only be banks, which for prudential regulation can have a high gearing ratio. In the following part we will show how neural networks are able to price both industrial and financial firms at the same time, even if they show a strongly different leverage.

We have used a feedforward neural network, with the backpropagation algorithm; it is a 4-layer network, with two hidden layers and therefore an output layer of only one node (the CDS spread).

The input layer consists of 18 nodes: in the first four nodes we have the risk free rate, the recovery rate, the leverage and the implied volatility of the firm; in the remaining 14 nodes we have the serie of quarterly CDS spreads of the firm. If there is a lack in the data, we just use the value of the preceeding quarter. This approach merges data coming from the firm with data (the CDS spreads) coming from the market, giving great effectiveness to the predictions of the network. Moreover the power of this approach can be appreciated observing that in this way the network is able to price CDS with reference entities coming both from the industrial field (which usually have low leverages and high CDS spreads) and from the financial field (which have an extremely high gearing ratio but are characterized by a history of low CDS spreads because of the prudential regulation, using this detail to discriminate between them). Fig. 3.3 shows the structure of the network.  The sample has of course been shuffled; the learning parameter has been settled to 0.5 and the initial parameters of the neurons have been chosen in the range [-2,2]. Our study shows that a logaritmic reduction is more efficient, because our sample consists of extremely variable data, so a simple linear reduction would enhance the distortions brought by the so-called outliers, that is data very different from the rest of the sample.

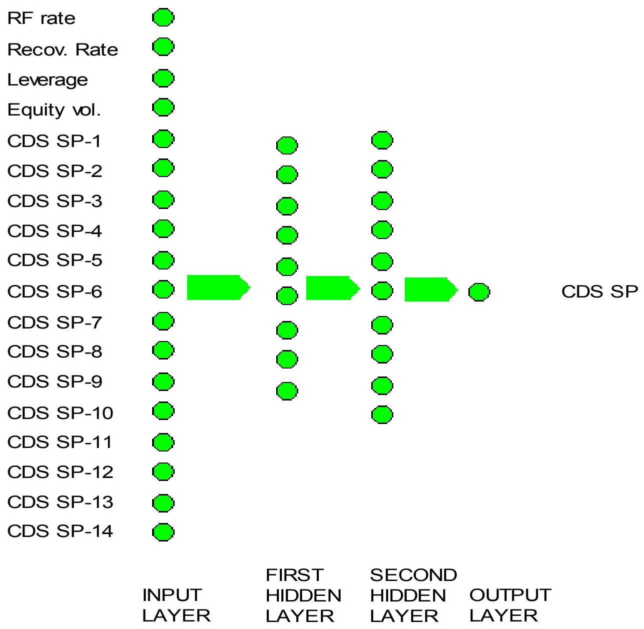



Fig. 3.3 Structure of the neural network

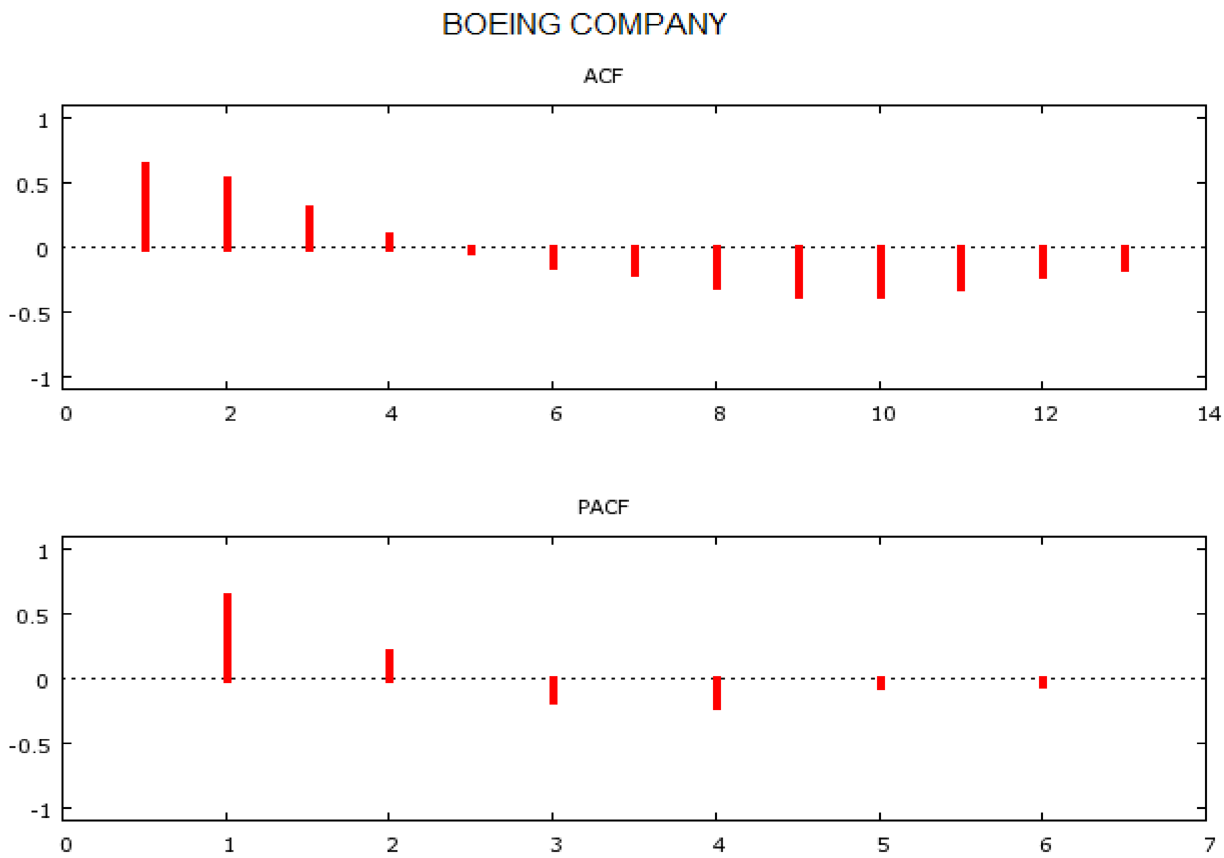



Fig. 3.4 Typical correlogram of a CDS spread time serie

In fig. 3.4 we show as an example the correlogram for the CDS spread time serie of The Boeing Company only, for the sake of simplicity, but we obtained the same structure for all the companies included in our sample: in the first part we can see the correlation between each value and a delayed value (the delay being expressed on the x-axis); the second part shows the correlation between each value and *p* preceeding values, with *p* on the x-axis. It is therefore evident that the correlation between values, even if decreasing, is strong, so the serie is autoregressive; we can then express each value in terms of the preceeding ones. In this sense a CDS spread is more similar to an interest rate than to an equity price, so that it shows a mean reversion process which tends to pull spreads higher (lower) than some long-run average level back to this value over time. Obviously we shall have a negative (positive) drift. The sinusoidal cycle observable in the correlogram explains this phenomenon: moreover, it is a consequence of the strict relationship between CDS spreads and risk-free interest rates already discussed[15].

[15]HULL, J. C., (2003).

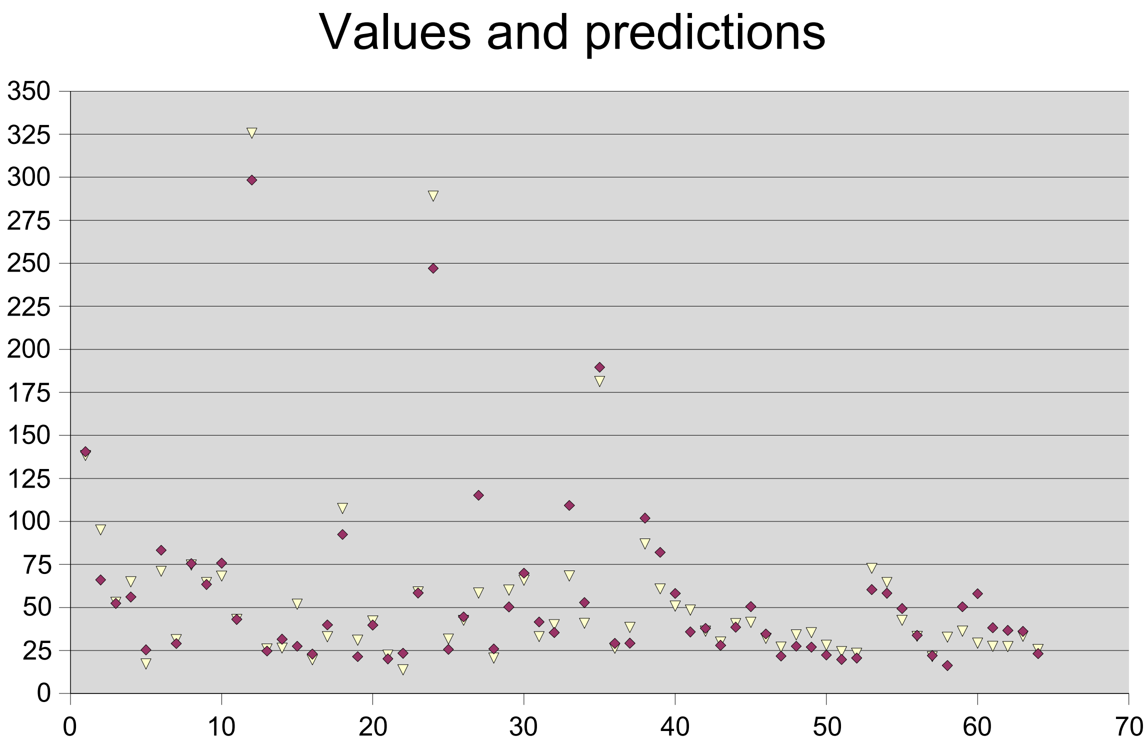



Fig. 3.5 Market data (in yellow) and predictions of the neural network (in red)

Fig 3.5, showing in red the neural network predictions and in yellow the real market data, confirms the effectiveness of the neural network in predicting CDS spreads.

| ***Error typology*** | ***Value*** |
|---|---|
| R-squared | 0,9082 |
| Root mean squared error | 14,3988 |

Tab. 3.3 Evaluation of the approximation of the neural network. (our elaboration)

| | ***NN*** | ***CreditGrades*** | ***Linear regression*** |
|---|---|---|---|
| Correlation | 0,9636 | -0,02 | 0,9309 |
| Rmse | 14,3988 | >100 | 30,86 |
| R-square | 0,9086 | >1 | 0,8566 |

Tab. 3.4 Comparison between neural networks and other models: effectiveness of fit. (our elaboration)

In tab. 3.3 and 3.4 the values of R-squared and rmse are shown: as it is easy to observe, the results are highly coherent. We must stress the point that using traditional models such as Creditgrades™ and others we would obtain predictions almost useless, even excluding banks from the sample; neural networks surely are a great pricing instrument in order to evaluate credit spreads. The architecture of the neural network is feedforward, trained for 17000 learning epochs using the backpropagation algorithm. Therefore it turns out obvious that neural networks are able to totally capture the variability relative to the market dynamics of credit derivatives: because of the fact that in literature there is no unanimity on the determination of the form of the CDS spread evaluation function, neural networks can therefore be seen as effective instruments of elaboration able to satisfy this lack from a statistical point of view.

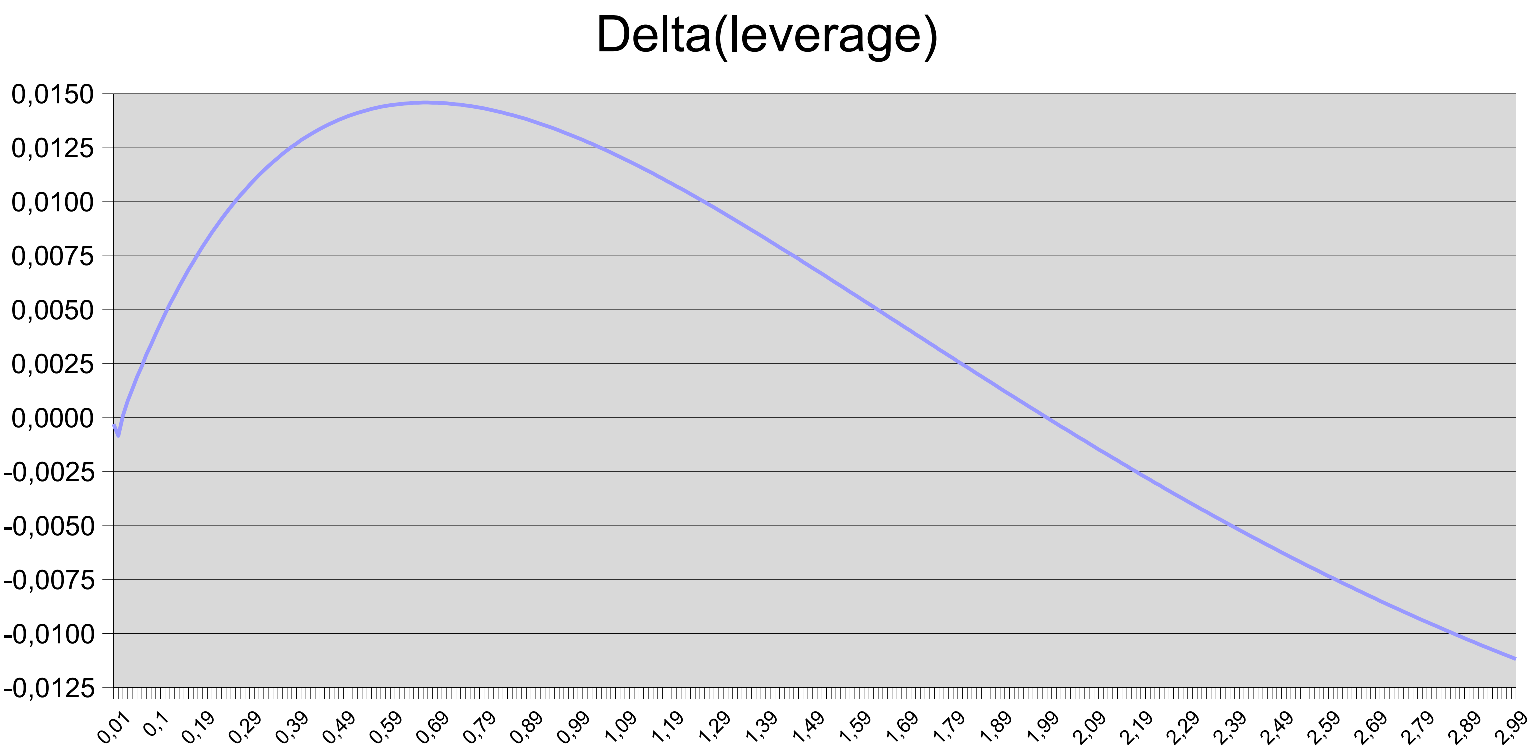


Fig. 3.6 Relationship between delta and leverage

Fig. 3.6 shows a “delta” for a CDS contract: in fact we find on the x-axis the leverage, and on the y-axis the value calculated with the finite differences method, that is

$$\Delta = \frac{SP(Lev+h) - SP(Lev)}{h} ; h \to 0$$

In a similar manner we can calculate for a CDS all the “greek” letters typical of derivative contracts using the outputs of the neural nework with $h \sim 10^{-6}$. It is evident in the diagram that for high leverages “delta” becomes negative: in fact we must remember that highly leveraged companies belong usually to the financial sector, so that they are less risky because of the prudential regulation. This effect is explained very well by the network, in fact for low leverages (typical of the industrial field) we see a direct relationship between leverage and CDS spreads. In general “delta” tends to be low for CDS because the autoregressive part of the model explains a large part of their variability. In other words, the neural network is able to recognize the risk of the activity carried out by the company using the time serie of its CDS spread: in the part of our study covering the correlation, we obtained a value for each observation and the preceeding one of 0,90 , as it is evident from the correlogram shown above. This correlation explains the major part of the variability of CDS spreads, while the remaining part is explained by the independent variables relative to the specific structure of each company.

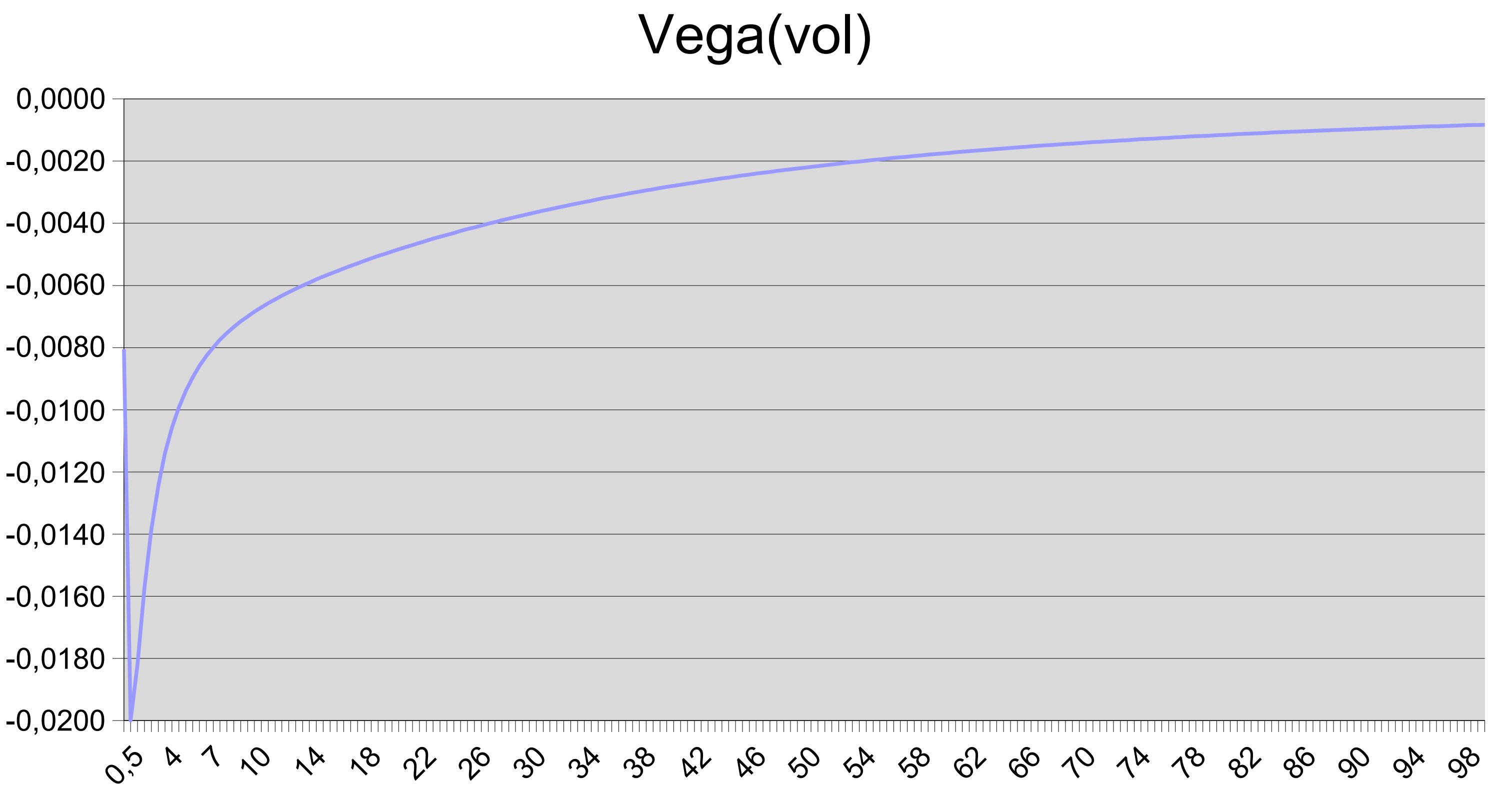



Fig. 3.7 Relationship between vega and equity volatility

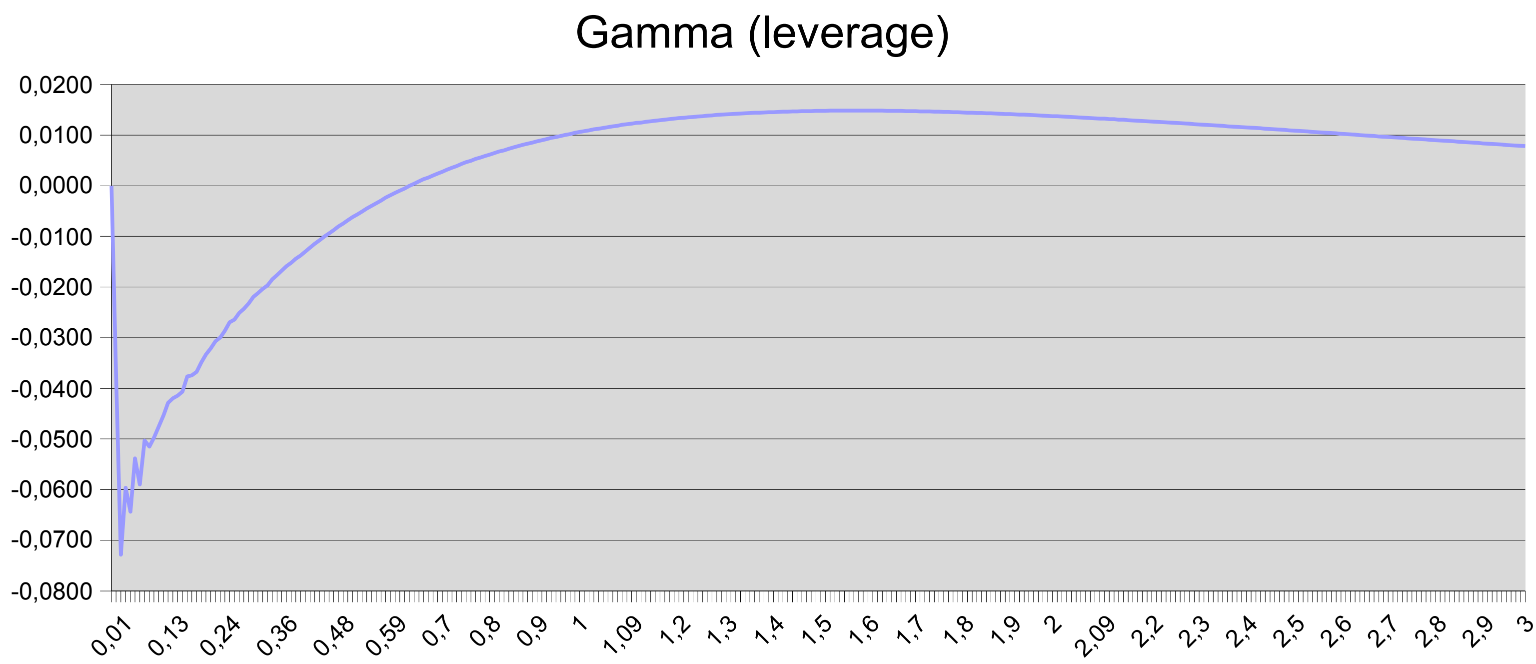



Fig. 3.8 Relationship between gamma and leverage

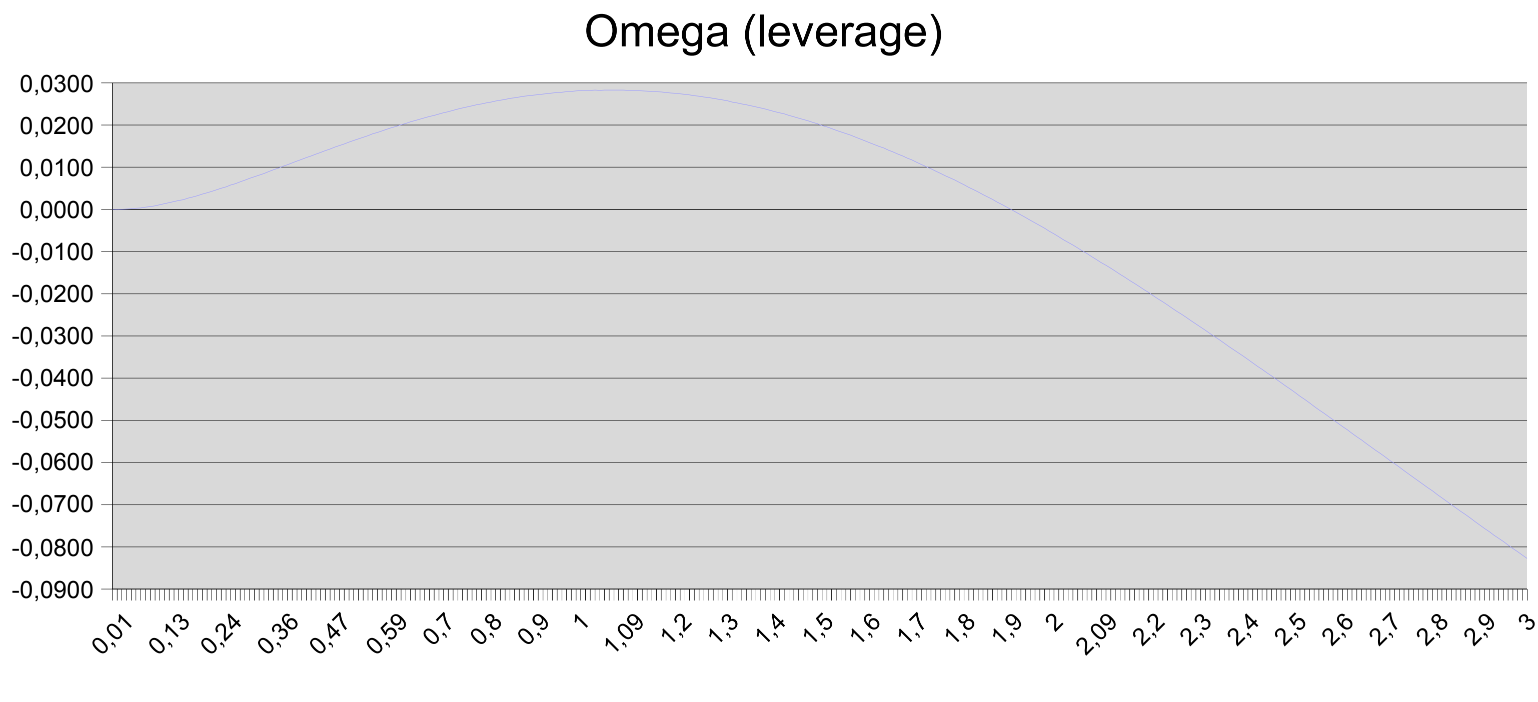



Fig. 3.9 Relationship between omega and leverage

## 5 Conclusion and future work

In this paper we have discussed an innovative approach to the study of CDS spreads using neural networks, non-parametric statistical instruments: in the first part we have explained the structure of the market and the general topics relating to neural networks, while in the last part, starting from the high correlation observed between each CDS spread value and the preceeding one in the time serie of each company, we have trained a neural network based both on these time series and on the structural details of the firms, that is leverage, option-implied equity volatility and recovery rates. Our results in terms of r-squared and rmse are highly coherent and are confirmed by the empirical data; in the future our work will be focused on the application of this network to other samples, as soon as the market for CDS becomes deeper and more liquid, in order to show higher volumes.

# BIBLIOGRAPHY


ALTMAN, E. I., KISHORE, V. M. "Almost everything you wanted to know about recoveries on defaulted bonds", *Financial Analysts Journal*, November-December 1996, 57-64.

ANGELINI, E. "Il profilo di rischio di un portafoglio con derivati creditizi: l'elaborazione di un modello di simulazione Montecarlo", in *Banca Impresa Società*, a. XXI, 2002, n° 2, pg. 259-82.

BANK FOR INTERNATIONAL SETTLEMENT, "International Convergence of Capital Measurement and Capital Standards", Basel Committee on Banking Supervision, A Revised Framework, Update November 2005.

BELTRATTI, A., SERIO, M., TERNA, P. "Neural networks for economic and financial modelling", International Thomson Computer Press, 1996.

CAPUTO NASSETTI, F., FABBRI, A. "Trattato sui contratti derivati di credito", EGEA, 2000.

DAS, S. "Credit derivatives. Trading & Management of Credit & Default Risk", John Wiley & Sons, Chicago, 1998.

DUFFEE, G.R., ZHOU, C. "Credit Derivatives in Banking: Useful Tools for Managing Risk?", *Journal of Monetary Economics*, n. 48, 2001.

DOLCINO, F., GIANNINI, C., ROSSI, E. "Reti neurali artificiali per l'analisi e la previsione di serie finanziarie", Collana studi del Credito Italiano, 1998.

FLOREANO, D., NOLFI, S. "Reti neurali: algoritmi di apprendimento, ambiente di apprendimento, architettura", in *Giornale Italiano di Psicologia*, a. XX, febbraio 1993, 15-50.

GORI, M. "Introduzione alle reti neurali artificiali", in *Mondo Digitale* n° 4, AICA, settembre 2003.

GREENSPAN, A. "Economic flexibility", Speech to HM Treasury Enterprise Conference, London, UK, 2004.

MARCO, G., VARETTO, F. "Diagnosi delle insolvenze e reti neurali", BANCARIA EDITRICE, 1994.

HANKE, M. "Neural network approximation of option-pricing formulas for analytically intractable option-pricing models", in *Journal of Computational Intelligence in Finance*, September/October 1997.

HENKE, S., BURGHOF, H.P, RUDOLPH, B. "Credit Securitization and Credit Derivatives: Financial Instruments and the Credit Risk management of Middle market Commercial Loan Portfolios", CFS Working paper Nr. 98/07.

HYKIN, S. "Neural networks: a comprehensive foundation", Prentice Hall International, 1999

HULL, J.C. "Opzioni, futures e altri derivati, Il Sole 24Ore S.p.A., 2003.

HULL, J.C. "Valuing correlation-dependent derivatives using a structural approach", *INI Conference*, February 25, 2005.

J.P. MORGAN, *"*The J.P. Morgan guide to credit derivatives*",* 2000.

MCNELIS, P.D. "Neural networks in finance", Elsevier Academic Press, 2005.

MERRYL LYNCH*,* "Credit derivative handbook 2003", 16 april 2003.

MINTON B.A., STULTZ R., WILLIAMSON R. "How much do bank use credit derivatives to reduce risk?", Working Papers, 2005.

MORRISON, " Credit derivatives: New Financial Instruments for Controlling Credit Risk", Economic Review of the Federal Reserve. 2005.

O'KANE, D., TURNBULL, S. "Valuation of credit default swaps", in *Quantitative credit research quarterly*, Lehman Brothers, vol. 2003-Q1/Q2, april 2003.

RAI, S., HOLOPPA, H. "Credit derivatives: the new wave in Risk Management"*,* in The Journal of Lending & Credit Risk Management, march 1997.

RANSON, B.J. "Credit derivatives. A user's observations", in The Journal of Lending & Credit Risk Management, may 1997.

REOCH, R., MASTERS, B. "Credit derivatives: structures and application", in Financial Derivatives and Risk Management, march 1996.

RISKMETRICS GROUP*, "*CreditGrades Technical Document*",* May 2002.

STAMICAR, R., FINGER, C.C. *"*Incorporating equity derivatives into the CreditGrades model*",* RiskMetrics Group*,* 2005.

STULTZ, R. "Risk management and Derivatives", South-Western Publishing, 2003.

TAVAKOLI, J.M. "Credit derivatives. A guide to instruments and applications", John Wiley & Sons, Chicago, 1998.

WERBOS, P. "Backpropagation, past and future", in Proceedings of the IEEE International conference on neural networks, IEEE press, 1988.