

The Application of Neural Networks to the Pricing of Credit Derivatives

Alessandro Ludovici*

Università "G. d'Annunzio", Chieti-Pescara

The present paper deals with a new approach to the pricing of credit derivatives, which are innovative financial instruments able to immunize a securities portfolio from the default risk of the issuers, using neural networks. After an essential analysis of the most important topics inherent to these nonlinear statistical instruments, particular emphasis, due to their diffusion, has been put on the characters of Credit Default Swaps and on the particularities of the structural and reduced form approaches proposed for their analysis. In the final part of the paper the effectiveness of neural networks in approximating the evaluation of credit derivatives and in improving the timing in the default prevision is illustrated. [JEL Classification: C45, G12, G32, G33]

1. - Introduction

The recent history of financial markets shows how the impetuous development of the financial innovation process, which concerns all of their structural components, has been associated with the constant engagement of operators in finding more efficient computational methodologies to support the analysis. In this paper a possible way to study this phenomenon goes deeper, focusing attention on the market of derivative instruments, which has been the origin of almost all the most important innovative phenomena for decades, and on neural networks, a computational

* <alessandro.ludovici@prometeia.it>. The Author is grateful to Prof. Eliana Angelini for her constant support during the writing of this paper. Many thanks also go to Prof. Andrea Roli for his precious suggestions and to Dott. Maila Bozzetto for her help in analyzing the data. Finally, he thanks the referees of the Costa Prize for their comments, and in memory of Prof. Antonio Cirillo, whose teachings have made this paper possible. The sole responsibility for the contents of this paper lies with the Author.

method that has increasingly attracted the interest of operators. More in particular, we try to analyze the capability of these instruments in close approximation to the results of complex and nonlinear financial calculation methodologies. In the following section, we analyze the architecture of neural networks, mainly focusing the attention on the concepts that are most useful for their application to financial instruments.

The first part of the third section deals with credit derivatives, because of the fast development that has characterized their recent history as effective ways of credit risk management: particular emphasis has been put on the analysis of Credit Default Swaps, because they are the most diffused; in the second part we describe the CreditGradesTM model, formulated by important financial institutions as the first attempt to introduce greater transparency in the evaluation of these instruments. The paper ends with the conclusion that, concerning this field of the financial markets, neural networks constitute a valid instrument of calculation: in fact there still does not exist in literature a formula of evaluation for the CDSs, able to tie the quoted spreads to the specific underlying variables of each examined firm, and the neural network can face the problem of the functional form from a statistical point of view. A further elaboration made in this paper concerns one of the biggest international aerial carriers, the American company UAL Corporation, between 1997 and 2004: the data needed have been obtained through the database of the Federal Reserve System and BloombergTM. The elaboration carried out shows how, even if all the proposed models signal the deterioration of the creditworthiness with a certain anticipation, CreditGradesTM turns out particularly efficient under this profile. The efficiency is improved by the use of the neural network, which remarkably increases in advance the number of trimesters with which the deterioration of the creditworthiness is indicated.

2. - Neural Networks: Architecture and Applications

Neural networks have been used in different fields of study, such as engineering, medicine, physics and others. As far as the

reasons that have led in this study to use this type of instruments among the several types of nonlinear approximation methods known in literature, it is necessary to remember at first that neural networks do not preliminarily request to specify the existing relation between the input and the output of the studied phenomenon. Neural networks are a highly adaptable and flexible instrument, as they are able to determine a functional approximation of nonparametric type. Moreover, neural networks are well known in the economic and financial field¹ and constitute the main representative of the family of nonlinear learning systems. In any case, it is necessary to emphasize that there are no studies in the available literature that apply such instruments to credit derivatives, and this reason has naturally led to the exploration of this innovative field of analysis. Finally, it is important to remember that, beyond the effectiveness that will be better illustrated in the next parts, neural networks are connoted also for relatively small training, tuning and execution times, because they use optimization techniques which have been known for a long time in literature. Although the relative structures differ remarkably from one another, it is possible to point out some fundamental principles regarding the functioning of these operative instruments. The following part of this section is therefore dedicated to the presentation of the principal results, inherent to this topic, available in literature. It is important to start by emphasizing that, in order to analyze financial dynamics, relatively little complex networks are effective, at least compared to those of other fields².

A. - Architecture of Neural Networks

A neural network relates a set of input variables $\{x_i\}$, $i = 1, 2, \dots k$ to a set of one or more output variables $\{y_j\}$, $j = 1, 2, \dots h$. An

¹ Among the first studies in the economic and financial field *cfr.* MALLIARIS M.E. - SALCHENBERGER L. (1993).

² DOLCINO F. - GIANNINI C. - ROSSI E. (1998, page 1). For a useful description of the phenomenon in general terms, *cfr.* FLOREANO D. - NOLFI S. (1993, pages 15-50) and GORI M. (2003).

essential characteristic of a neural network, differently from other approximation methods, 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 processes³.

In the feedforward neural network, parallel elaboration is associated with 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.

Therefore, the neurons process the input variables in two ways: firstly by forming linear combinations and then by transforming these combinations with a particular function, typically the logsigmoid function, illustrated in Graph 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 small changes in a high (or low) value of the independent variables, show little changes in the dependent variables. At the analytical level, the neural network can be described by the following equations⁴:

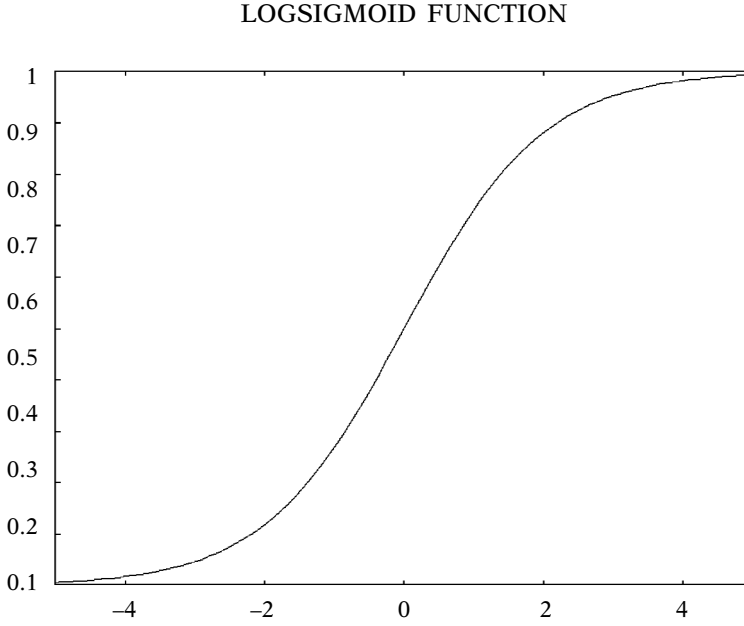
$$(1) \quad n_{k,t} = \omega_{k,0} + \sum_{i=1}^m \omega_{k,i} x_{i,t}$$

$$(2) \quad N_{k,t} = L(n_{k,t}) = \frac{1}{1 + e^{-n_{k,t}}}$$

³ McNELIS P.D. (2005, page 21).

⁴ McNELIS P.D. (2005, page 25).

GRAPH 1



$$(3) \quad y_t = \gamma_0 + \sum_{k=1}^q \gamma_k N_{k,t}$$

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 γ_k and added to the constant term γ_0 , in order to obtain the output y_t , which represents the prediction of the neural network concerning time or observation t . 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 using two (respectively N and P) or more hidden layers⁵:

$$(4) \quad n_{k,t} = \omega_{k,0} + \sum_{i=1}^m \omega_{k,i} x_{i,t}$$

$$(5) \quad N_{k,t} = L(n_{k,t}) = \frac{1}{1 + e^{-n_{k,t}}}$$

$$(6) \quad p_{l,t} = \rho_{l,0} + \sum_{k=1}^s \rho_{l,k} N_{k,t}$$

$$(7) \quad P_{l,t} = \frac{1}{1 + e^{-p_{l,t}}}$$

$$(8) \quad y_t = \gamma_0 + \sum_{l=1}^q \gamma_l P_{l,t}$$

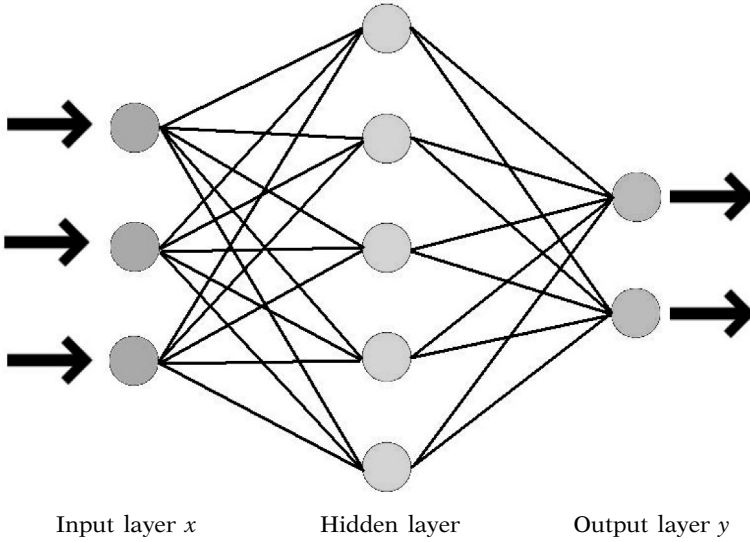
Adding another hidden layer increases the number of parameters (weights) to be estimated by the factor $(s + 1)(q - 1) + (q + 1)$; in fact the net with a single hidden layer, with m input variables and s neurons, has $(m + 1)s + (s + 1)$ parameters, while the same net, with two hidden layers and q neurons 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, but of the greater probability that the net converges to a local rather than global optimum. Moreover, increasing the number of parameters uses up the degrees of freedom if the sample size is limited, and requires a longer training time. In any case it has been demonstrated that a neural network with two layers is able to replicate any nonlinear function⁶: in fact it does not just approximate a phenomenon by adapting a fixed functional form, but it determines which functional form is able to best describe the studied phenomenon.

⁵ McNELIS P.D. (2005, page 33).

⁶ DOLCINO F. - GIANNINI C. - ROSSI E. (1998, page 20); WHITE H. (1990, pages 540, ss.).

GRAPH 2

NEURAL NETWORK WITH ONE HIDDEN LAYER
AND TWO OUTPUT NEURONS



In Graph 2 a net with a multiple number of output variables is illustrated⁷. A neural network with a hidden layer and two output variables is described by the following equations⁸:

$$(9) \quad n_{k,t} = \omega_{k,0} + \sum_{i=1}^m \omega_{k,i} x_{i,t}$$

$$(10) \quad N_{k,t} = L(n_{k,t}) = \frac{1}{1 + e^{-n_{k,t}}}$$

$$(11) \quad y_{1,t} = \gamma_{1,0} + \sum_{k=1}^q \gamma_{1,k} N_{k,t}$$

⁷ DOLCINO F. - GIANNINI C. - ROSSI E. (1998, page 2); MARCO G. - VARETTO F. (1994, page 24).

⁸ McNELIS P.D. (2005, page 36).

$$(12) \quad y_{2,t} = \gamma_{2,0} + \sum_{k=1}^q \gamma_{2,k} N_{k,t}$$

It is possible to observe that adding an output variable leads to the evaluation of $(q + 1)$ more parameters, which is equal to the number of neurons of the hidden layer increased of one unit. Therefore, another output variable causes an increasing number of parameters that have to be estimated, equal to the number of the neurons of the hidden layer, not equal to the input variables. Using a neural network with multiple outputs makes sense only if they are closely correlated to the same set of input variables: for example, the temporal structure of inflation or interest rates. One of the most common criticisms made is that neural networks are substantially black boxes: it is not possible to find the answer to questions pertaining to the nature of the parameters, the reasons of the choice of their number, the number of the neurons, the number of the hidden layers and the reasons that relate the architecture of the net to the structure of the underlying problem⁹.

The risk, when models are based on a high number of parameters, is that their flexible nature seems to explain nearly everything, but in reality explains nothing. In any case, 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 difficult to identify a precise cause and effect relationship. Hence, the black box criticism constitutes, paradoxicallly, also one of the greatest qualities of neural networks. However, even though it is easy to increase the number of the parameters of the net, the importance of the clarity of the assumptions must never be forgotten in any model¹⁰.

⁹ McNELIS P.D. (2005, page 55).

¹⁰ McNELIS P.D. (2005, page 57).

B. - Data Scaling

A neural network is not able to analyze data or to give solutions in absolute value: if there are data of an unusually elevated or reduced value, problems of overflow or underflow could happen. Instead, when sigmoid functions are used, it becomes indispensable to preprocess data: this family of functions in fact has a codomain of type $[0, 1]$ (or $[-1, 1]$ in the case of the logsigmoid function), so the values must be scaled to these intervals, otherwise the output of the net would become useless. In fact, it would be equal to the superior or inferior threshold for all of the values higher or lower than a determined limit. In other words, for a great amount of data not standardized 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 methods¹¹, the linear reduction transforms the series of values x_k in the serie \hat{x}_k , using the following formulas:

$$(13) \quad \hat{x}_{k,t} = \frac{x_{k,t} - \min(x_k)}{\max(x_k) - \min(x_k)}$$

if the range is between 0 and 1, and

$$(14) \quad \hat{x}_{k,t} = 2 \frac{x_{k,t} - \min(x_k)}{\max(x_k) - \min(x_k)} - 1$$

if the desired range is between -1 and 1 .

C. - 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 network¹².

¹¹ MCNELIS P.D. (2005, page 64).

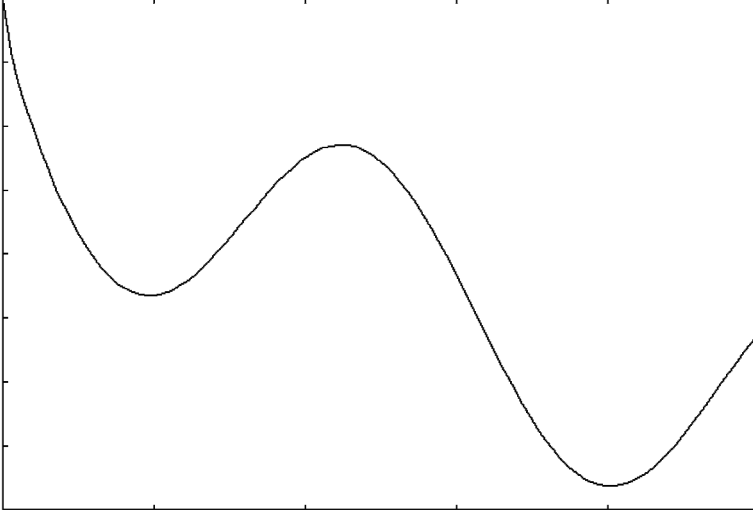
¹² DOLCINO F. - GIANNINI C. - ROSSI E. (1998, pages 33, ss.); MCNELIS, P.D. (2005, page 65); MARCO G. - VARETTO F. (1994, page 25).

Preliminarily, it is necessary to notice that the ways with which such process can be carried out are two: the first one, called “not supervised”, consists in feeding the net with only the input data of the sample, so that it can carry out a decomposition in cluster with them picking the existing affinities; the second one, called “supervised”, is focused on the analysis of both the input and the output data: because of the importance that it has in this paper, the considerations which follow are mainly referred to it. The training of the net is a problem much more complex than the evaluation of the parameters of a linear model, since neural networks have a high nonlinear complexity nature. 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. In short, in any nonlinear model it is necessary to begin the evaluation of the parameters on the basis of random values of them. However, as it will be shown, the capability of the evaluation process to converge to a global optimum depends on the goodness of this initial hypothesis: in fact, if it is situated near a local optimum rather than the global one¹³, it is likely that the first one will be reached. This is illustrated in Graph 3: the initial choice of the parameters (or weights of the neurons) could be situated anywhere on the x-axis: if it is near a local minimum, the training process of the net would stop there. Later on, it will be observed that the process ends in a point where 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. On the other hand, 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, where it is therefore necessary to determine the set of parameters $\Omega = \{\omega_{k,i}, \rho_{l,k}, \gamma_l\}$.

¹³ DOLCINO F. - GIANNINI C. - ROSSI E. (1998, pages 29, ss.).

GRAPH 3

EXAMPLE OF A SUCCESSION OF LOCAL AND GLOBAL MINIMUMS



The problem consists of minimizing the loss function¹⁴, defined by the sum of the squares of the differences between the observed data y and the prediction of the net \hat{y} :

$$(15) \quad \min_{(\Omega)} \Psi(\Omega) = \sum_{t=1}^T (y_t - \hat{y}_t)^2$$

$$(16) \quad \hat{y}_t = f(x_t; \Omega)$$

in which T is the number of the observations of the output vector y , and $f(x_t; \Omega)$ represents the neural network. Ψ is a nonlinear function of Ω . All nonlinear optimizations begin with an initial random choice of the parameters and then try to reach the solution by finding the best possible value within a reasonable number of iterations. Different methodologies have been proposed in order to

¹⁴ McNELIS P.D. (2005).

lead this search¹⁵: some make reference to metaheuristics, e.g. genetic algorithms, in alternative to the classic gradient descent method, or Newton-Raphson. However, 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 goes below the predefined tolerance value¹⁶. In order to avoid local optimums, a first convergence can be determined, and then the process is repeated 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, the most important problems are when the number of the parameters increases or when the architecture of the network becomes particularly complex. Paul John Werbos proposed, at the beginning of 1970's, an alternative to the gradient descent called backpropagation¹⁷. It is a very flexible method that avoids the problems caused by the evaluation of the Hessian matrix in the gradient descent, and it is the most widely used method. In fact, in the training process, the inverse Hessian matrix is replaced by an identity matrix having a dimension equal to the number k of the parameters, multiplied by the learning coefficient ρ :

$$(17) \quad (\Omega_1 - \Omega_0) = -H_0^{-1} Z_0 = -\rho Z_0$$

In order to avoid oscillations this coefficient is chosen in the range $[0.05, 0.5]$ and it can also be endogenous, in fact it can assume various values when the gradient drops and the process seems to converge; or finally different parameters can be adopted. So, the problem of this choice along with the existence of local minimums must be solved. Moreover, low values of the learning coefficient, although capable of avoiding oscillations, can prolong the minimizing process. This can however be accelerated by adding a 'momentum', so that at iteration n we will have¹⁸:

¹⁵ McNELIS P.D. (2005, page 67).

¹⁶ DOLCINO F. - GIANNINI C. - ROSSI E. (1998, page 44).

¹⁷ McNELIS P.D. (2005, page 69); MARCO G. - VARETTO F. (1994, page 27).

¹⁸ McNELIS P.D. (2005, page 70).

$$(18) \quad (\Omega_n - \Omega_{n-1}) = -\rho Z_{n-1} + \mu(\Omega_{n-1} - \Omega_{n-2})$$

Therefore, with μ generally equal to 0.9, the calculation of the parameters moves more quickly outside a plateau in the error surface.

Now we will briefly discuss the methods used to estimate the effectiveness of the net output. Regarding the evaluation of the goodness of the predictions, the most common index is the *R*-squared (goodness of fit) as far as the capability of the net to predict the training data. As regards the evaluation of the predictions outside the training sample, a common index is the root mean squared error. In other words, after dividing the sample into two parts, the first (in sample) will be used in order to train the net. The second part (out of sample) will not be used to train the net, but to estimate its capability of predicting data coming from the same population, but not included in the training set. Usually, about 25% of total data is used for out of sample testing.

However, concerning the amount of necessary data¹⁹, a neural network undoubtedly requires the evaluation of many more coefficients than e.g. a linear model, and this leads to the necessity of a wider sample. 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 is not always a positive aspect: in the financial field, using very old data can distort the models, because they tend to change rapidly. As a consequence, remote data could not be related to the present ones anymore.

3. - Credit Derivatives

“Credit derivatives”²⁰ are contracts whose final value depends on the creditworthiness of one or more trade or sovereign entities”²¹. In this section, we will analyze the principal

¹⁹ DOLCINO F. - GIANNINI C. - ROSSI E. (1998, page 24), where the concepts of “evaluation error” and “approximation error” are analyzed.

²⁰ MORGAN J.P. (2000, pages 7, ss.).

²¹ HULL J.C. (2003, page 708).

contributions available in literature concerning their pricing, also to introduce the CreditGrades™ model, a first class methodology in the evaluation of default probability. We, therefore, begin with credit default swaps (CDSs), which in recent years have conquered the superiority²² in terms of volumes exchanged in the market of credit derivatives. However, this market remains essentially over the counter because the realization of a centralized system of exchange has still not been achieved: in a limited way, in Europe the iTraxx index is an example of such a system.

In general, CDSs²³ are contracts which offer protection against the default risk (credit event)²⁴ of a specific firm, called reference entity. The buyer of the protection obtains the right to sell at par, on verification of the credit event²⁵, a specific obligation (reference obligation)²⁶, issued by the firm. The nominal value of this bond is the notional principal of the contract, and the right is obtained through a series of periodic payments. They are calculated by applying the quoted spread to the notional principal: the series ends when the contract expires or if the credit event takes place, which usually involves a final accrual payment. As in every derivative contract, the settlement²⁷ can take place with the physical delivery of the obligations and the payment of their nominal value, or in cash. In the last case, a complex procedure is carried out, in which the calculation agent determines the average price Z of the obligation²⁸ at a prefixed date, successive to the verification of the credit event, and pays a sum equal to $(100 - Z)\%$ of the notional principal to the protection buyer. As mentioned, the market of credit derivatives is essentially an OTC market, and therefore financial intermediaries carry out a determining role²⁹ as market makers: at

²² FRANCIS C. - KAKODKAR A. - MARTIN B. (2003, pages 4, ss.).

²³ CAPUTO NASSETTI F. - FABBRI A. (2000, page 89); O'KANE D. - TURNBULL S. (2003, page 1); MORGAN J.P. (2000, pages 12, ss.7; FRANCIS C. - KAKODKAR A. - MARTIN B. (2003, pages 10, ss.).

²⁴ CAPUTO NASSETTI F. - FABBRI A. (2000, pages 31, ss. and page 59).

²⁵ HULL J.C., *ibidem*.

²⁶ CAPUTO NASSETTI F. - FABBRI A. (2000, page 31).

²⁷ CAPUTO NASSETTI F. - FABBRI A. (2000, page 34), where the various types of settlements are detailed.

²⁸ HULL J.C. (2003, page 709).

²⁹ HULL J.C. (2003); FRANCIS C. - KAKODKAR A. - MARTIN B. (2003, pages 8, ss.).

any moment they quote a bid price to buy the protection and an ask price to sell the protection, for a given expiry date and reference entity. Credit derivatives are effective instruments in order to manage and to modify the risk profile of a financial portfolio, which is reduced or increased by buying or selling protection. Moreover, a fast portfolio diversification can be obtained by trading protections concerning different reference entities.

The following sections are dedicated to two of the most important models proposed in literature for the evaluation of these instruments: the first is relative to the reduced form approach and the last to the structural approach. However, it is necessary to underline that these two models can only be applied to the analysis of quoted societies.

A. - The Hull and White Model

Reduced form approaches assume the existence of a correlation between the default of a firm and some particular indicators of the economic conjuncture, called background factors. More specifically, the Hull and White model is based on a fundamental consideration about the value of the bonds issued by different entities. In fact, a bond issued by the Government is by definition risk free, so the difference in value of a similar bond, issued by a firm, constitutes the market evaluation of the default costs. Therefore, analyzing the course of this difference, it is possible to extrapolate at anytime the evaluation of the default probability expressed by the market, on the base of the quotations of the issued bonds. Having thus stated, in the following part of this section the model will be illustrated, with three simplified hypotheses: risk neutrality, mutual independence of the variables and possible verification of the credit event only at the payment dates, $i = 1, \dots, n$, so as to avoid the complications related to the calculation of the accrued interest. Moreover, it is assumed that the default probability of the protection seller is null, but on this subject we will return later. We will indicate with:

T : residual life of the contract;

- p_i : default probability at time t_i ;
 \hat{R} : expected recovery rate, that is the sums expected from the reference entity if the credit event takes place;
 $u(t_i)$: present value, at the rate of 1 € per year, of payments between time 0 and time t_i ;
 $e(t_i)$: present value of an accrual payment at time t_i equal to $t_i - t^*$ euro, where t^* represents the date immediately preceding time t_i ;
 $v(t_i)$: the present value of 1 € received at time t_i ;
 w : annual payment made by the buyer per euro of notional principal;
 s : value of w that causes the credit default swap to have a value of zero;
 π : the probability that within the expiry date the credit event does not take place;
 $A(t_i)$: accrued interest on the reference obligation at time t_i as a percentage of face value.

As already stated, the approach obtains the default probabilities from the quotations of the bonds issued by the reference entity or from derivative contracts³⁰, such as asset swaps. Therefore:

$$(19) \quad \pi = 1 - \sum_{i=1}^n p_i$$

The present value of the payments is:

$$(20) \quad w \sum_{i=1}^n (u(t_i) + e(t_i)) p_i + wu(T) \pi$$

while the expected payoff of the CDS is:

$$(21) \quad 1 - \hat{R}(1 + A(t_i)) = 1 - \hat{R} - A(t_i) \hat{R}$$

³⁰ HULL J.C. (2003, pages 676, ss.).

Then, its present value, given the hypothesis of risk neutrality, is equal to³¹:

$$(22) \quad \sum_{i=1}^n (1 - \hat{R} - A(t_i) \hat{R}) v(t_i) p_i$$

Consequently, the current value of a CDS, for the protection buyer, is equal to the difference between the present value of the expected payoff and the present value of the payments:

$$(23) \quad \sum_{i=1}^n (1 - \hat{R} - A(t_i) \hat{R}) v(t_i) p_i - w \sum_{i=1}^n (u(t_i) + e(t_i)) p_i - w u(T) \pi$$

The buyer will have to pay on an annual basis s (CDS spread)³², which is the value of w that makes the current value of the contract equal to 0:

$$(24) \quad s = \frac{\sum_{i=1}^n (1 - \hat{R} - A(t_i) \hat{R}) v(t_i) p_i}{\sum_{i=1}^n (u(t_i) + e(t_i)) p_i + u(T) \pi}$$

In Table 1 a numerical example of this formula is shown³³; it concerns a five-year CDS, with payments at times t_i , and a reference obligation with a coupon of 12% per year. Assuming that the default probabilities obtained from market data, as stated above, and the expected recovery rate are the ones shown in the table, and that the default can only happen at the end of each year, the spread is 542 basis points per year.

If the default can take place at anytime, p_i will be replaced by q_i , that is by the default density probability at time t_i , and therefore:

³¹ HULL J.C. (2003, page 712).

³² HULL J.C. (2003).

³³ HULL J.C. (2003); BENKERT C. (2004, page 81); O'KANE D. - TURNBULL S. (2003, pages 5, ss.); FRANCIS C. - KAKODKAR A. - MARTIN B. (2003, pages 16, ss.) for alternative approaches.

TABLE 1

EXAMPLE OF A CDS CONTRACT
(our elaboration)

	T_i	R	p_i	$A(t_i)$	$v(t_i)$	$(1-R-A(t_i)R)p_i v(t_i)$	$u(t_i)$	$u(t_i)p_i$
	1	0.4	0.05000	0.12000	0.95238	0.02629	0.95238	0.04762
	2	0.4	0.07000	0.12000	0.90703	0.03505	1.85941	0.13016
	3	0.4	0.09000	0.12000	0.86384	0.04292	2.72325	0.24509
	4	0.4	0.10500	0.12000	0.82270	0.04768	3.54595	0.37232
	5	0.4	0.12000	0.12000	0.78353	0.05190	4.32948	0.51954
			0.43500			0.20383		1.31473
π			0.56500					
$\sum u(t_i)p_i + u(T)\pi$			3.76089					
S			0.05420					

$$(25) \quad s = \frac{\int_0^T (1 - \hat{R} - A(t_i)\hat{R})v(t)q(t)dt}{\int_0^T (u(t) + e(t))q(t)dt + u(T)\pi}$$

In order to better understand the nature of the spread s , it is possible to use a quasi-arbitrage argumentation³⁴. A contemporary purchase of the reference obligation and a CDS of equal expiry date, written on it, eliminates the default risk. If y indicates the yield to maturity of the obligation, it is evident that the cancellation of the credit risk implies a reduction s of y . This reduction can eventually last until the credit event, instead of the maturity date of the obligation. In short, $y - s$ turns out to be a risk free rate of return, and therefore it must be aligned to the risk free rate r of the same expiry date, in order to avoid arbitrages. However, the protection buyer would not gain any interest between the date of payment of the last coupon and the date of the credit event. In addition, it is unsure that the rate of return that the buyer gains from that moment until the original expiry date of the contract is equal to r ³⁵.

The expected recovery rate turns out to be the only variable

³⁴ HULL J.C. (2003, page 713).

³⁵ HULL J.C. (2003, page 714).

that is estimated in order to obtain the spread s^{36} , as all the other variables are defined in the CDS contract ($T, w, A(t)$) or available on the market ($p_i, u(t), e(t), v(t)$). However, it has a limited incidence on the CDS spread, because its influences tend to offset each other³⁷. In fact, as the expected recovery rate increases, estimates of the default probability increase and the payoffs provided by the CDS decrease. The incidence can be elevated for nonstandard CDSs, e.g. binary CDSs³⁸, which guarantee, when the credit event takes place, an independent payment from the expected recovery rate, which only influences the default probability. Other nonstandard CDSs are basket CDSs³⁹, in which there are multiple reference entities and the payment is made when the default of whichever of them takes place. The add-up basket CDS is a portfolio of ordinary CDSs; the first-to-default basket CDS⁴⁰ ceases to exist when the first default takes place. These last instruments have a very complex nature, and to evaluate them it is necessary to proceed through Monte Carlo simulations⁴¹: nevertheless, when the correlation of the defaults grows⁴², the value of the first-to-default basket CDS diminishes, since it turns out to be less “diversified”. Instead, if the reference entities are diversified, a default will be more likely to take place; consequently, the buyer will also be more likely to receive a payment, so the value of the first-to-default basket CDS increases⁴³.

Finally, it is necessary to also consider the default probability of the protection seller⁴⁴. In fact, in that case, the buyer will have to stipulate a new CDS with another seller, and,

³⁶ HULL J.C. (2003, page 715); CAPUTO NASSETTI F. - FABBRI A. (2000, pages 142, ss.) where the relative evaluation technique is detailed.

³⁷ HULL J.C. - WHITE A. (2000, page 19).

³⁸ HULL J.C. (2003, page 716).

³⁹ HULL J.C. (2003).

⁴⁰ MORGAN J.P. (2000, page 28); FRANCIS C. - KAKODKAR A. - MARTIN B. (2003, pages 84, ss.).

⁴¹ HULL J.C. (2003); CAPUTO NASSETTI F. - FABBRI A. (2000, pages 190, ss.), where the evaluation techniques based on the conditional probability of events are detailed.

⁴² FRANCIS C. - KAKODKAR A. - MARTIN B. (2003, pages 89, ss.).

⁴³ HULL J.C. (2003).

⁴⁴ HULL J.C. (2003, pages 717, ss.).

if the creditworthiness of the reference entity has diminished, the spread will be higher. Therefore, the effects of the default of the seller on the CDS depend both on the probability of reduction of the creditworthiness of the reference entity and on its correlation with this event. This case can only be analyzed through Monte Carlo simulations, and it has been demonstrated⁴⁵ that the incidence of the default risk of the seller on the CDS spread is directly proportional to the mentioned correlation. However, it is obvious that the protection must be bought from a seller whose default is as little correlated as possible with that of the reference entity.

B. - The CreditGrades™ Model

The CreditGrades™ model for credit risk assessment derives from the operating experience of these four main financial institutions: Deutsche Bank, Goldman Sachs, Riskmetrics Group and J.P. Morgan. It is an attempt to create a “standard of transparency” on the credit markets, after the release of the CreditMetrics™ model in 1997, which is now considered a *de facto* standard for the risk assessment of a credit portfolio. According to these institutions⁴⁶, in 2002 there was a strong need for an instrument that could express the credit risk assessment of a single exposure. This was a result of the wider controls by the authorities in this important field of financial intermediation, and the increasing role that risk plays on the markets. On this topic, it is possible to mention the impetuous development of credit derivatives, as seen in the previous part. In this context, an essential factor has been the increasing number of defaults of important corporations, which in 2001 became truly worrisome. Apart from the famous “Enron case”, “Defaults in 2001 were notable for their individual size as well as their frequency. Twenty-

⁴⁵ HULL J.C. (2003, page 718); about these concepts cfr. ANGELINI E. (2002, pages 276, ss.).

⁴⁶ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 1).

nine issuers had defaults totaling over one billion dollars in debt apiece”.⁴⁷

The CreditGrades™ model belongs to the structural approach, which uses the option pricing theory to estimate the default probability, starting from the Black, Scholes and Merton studies. In this way, it tries to create a link between the credit market and the securities market. For this approach, the shareholders will reimburse the debt only if the asset value of the company exceeds it (and therefore the option is in the money), while, in the other case, they will abandon the assets to the creditors. Therefore, their position is similar to the holders of a call option sold by the creditors. Hence, the expiry date and the strike price of the option are equal to the expiry date and the amount of the debt. Consequently, the probability that the option is in the money is opposite to the default probability of the firm. The intention stressed by the authors of this model was to amplify the characteristics of this approach, in order to use only market data for the evaluation. In this way data from proprietary databases are not necessary, as they could diminish the transparency of the model⁴⁸. This section is dedicated to its essential analysis, as was presented in the technical document by its authors.

One of the peculiar features of the model is its definition of the default point, which is supposed to be stochastic. As stated in the technical document: “We cannot expect to know the exact level of leverage of a firm except at the time the firm actually defaults. The uncertainty in the barrier admits the possibility that the firm’s asset value may be closer to the default point than we might otherwise believe. This leads to higher short-term spreads than are produced without the barrier uncertainty”⁴⁹, in this way avoiding a point of weakness in the structural approach. Briefly, we summarize its essential characteristics⁵⁰:

⁴⁷ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002).

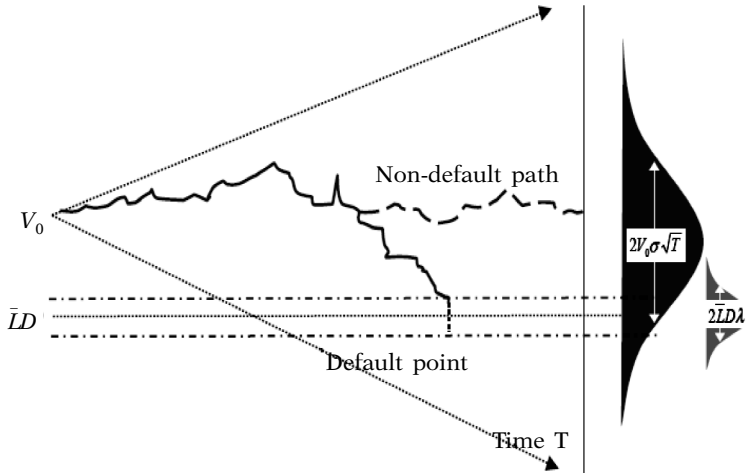
⁴⁸ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, pages 2, ss.).

⁴⁹ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 5).

⁵⁰ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 6); for a brief description of the model *cfr.* STAMICAR R. - FINGER C.C. (2005, pages 2, ss.).

GRAPH 4

STOCHASTIC PROCESS OF THE ASSETS OF THE FIRM
AND DEFAULT POINT*



* FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 7).

- the assets of the firm follow a stochastic process V ;
- the default happens when this process crosses the threshold equal to the default point;
- this default point must necessarily correspond to a certain part of the debts; in the model it depends on the average recovery rate L , which is supposed to be stochastic.

As regards the last point, the stochastic nature of L is also based on the empirical analysis carried out, which are described in the technical document. It is shown that this value is extremely variable, because it is influenced e.g. by the operating or financial nature of the default, by the possibilities of reorganization or liquidation of the firm and so on. Therefore, L is distributed in a lognormal way with average \bar{L} and standard deviation λ ; denoting with D the debt per share of the reference entity, we have

$$(26) \quad LD = \bar{L}De^{\left(\lambda Z - \frac{\lambda^2}{2}\right)}$$

where Z is a standard normal random variable.

Then, the survival probability at time t is equal to the probability that V does not cross the default point LD before time t , and so we will have⁵¹

$$(27) \quad P(t) = N\left(-\frac{A_t}{2} + \frac{\log(d)}{A_t}\right) - dN\left(-\frac{A_t}{2} - \frac{\log(d)}{A_t}\right)$$

$$(28) \quad d = \frac{V_0 e^{\lambda^2}}{LD}$$

$$(29) \quad A_t^2 = \sigma^2 t + \lambda^2$$

In these equations, $N(x)$ is the standard normal cumulative distribution function of the variable x , while σ represents the volatility of the assets of the reference entity. Once obtained the default probability $(1 - P(t))$, the corresponding credit spread must be calculated⁵². Hence, it is necessary to introduce the recovery rate R : while L is the medium recovery rate of all the various debts, R only regards the underlying credit. We can define the default density probability function⁵³

$$(30) \quad f(t) = -\frac{dP(t)}{dt}$$

so the cumulative default probability within time t will be

$$(31) \quad 1 - P(0) + \int_0^t ds f(s)$$

Focusing the analysis on a CDS with expiry date t and spread c , the present value of the expected loss payments will be

⁵¹ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002).
⁵² FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 45).

⁵³ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002).

$$(32) \quad (1-R)(1-P(0)) + \int_0^t ds f(s) e^{-rs}$$

where, as always, r represents the risk free interest rate, which is supposed to be constant. The present value of the expected spread payments will be

$$(33) \quad c \int_0^t ds P(s) e^{-rs}$$

Therefore, we obtain the price of the CDS from the difference between these two payments⁵⁴; then, the equilibrium spread makes the expected spread payments equal to the expected loss payments, so that the value of the CDS is null⁵⁵:

$$(34) \quad \hat{c} = r(1-R) \frac{1-P(0)+H(t)}{P(0)-P(t)e^{-rt}-H(t)}$$

where

$$(35) \quad H(t) = e^{r\xi} (G(t+\xi) - G(\xi))$$

$$(36) \quad G(t) = d^{z+\frac{1}{2}} N \left(-\frac{\log(d)}{\sigma\sqrt{(t)}} - z\sigma\sqrt{(t)} \right) + d^{-z+\frac{1}{2}} N \left(-\frac{\log(d)}{\sigma\sqrt{(t)}} + z\sigma\sqrt{(t)} \right)$$

$$(37) \quad z = \sqrt{\left(\frac{1}{4} + \frac{2r}{\sigma^2} \right)}$$

$$(38) \quad \xi = \frac{\lambda^2}{\sigma^2}$$

The determination of the asset value and the asset volatility is faced by the model in a peculiar way, in the attempt to associate the methodological rigor with an easier evaluation of the

⁵⁴ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 46).

⁵⁵ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002).

parameters⁵⁶. First of all, the attention is focused on the long term, that is $t > \lambda^2 / \sigma^2$: otherwise the default probability would be mainly influenced by λ ; moreover, the distance from default is defined as the number of standard deviations separating the present value of the assets of the firm from the default point; as a consequence of Itô's lemma, we have⁵⁷:

$$(39) \quad \eta = \frac{1}{\sigma} \log \left(\frac{V}{LD} \right) = \frac{V}{\sigma_s S} \frac{\partial S}{\partial V} \log \left(\frac{V}{LD} \right)$$

where S represents the equity price and σ_s represents the equity volatility of the reference entity. At this point, it is possible to examine the following boundary conditions⁵⁸:

– near the default point, S/LD is much smaller than 1, hence, we have $V = S + LD \approx LD$ and it is possible to write the approximation

$$(40) \quad V \approx LD + \frac{\partial V}{\partial S} S$$

Replacing in (39) we will have:

$$(41) \quad \eta \approx \frac{1}{\sigma_s}$$

– instead, for high values of S/LD , it is possible to assume that $S/V \rightarrow 1$, that is the rate of increase of V is equal to the one of S , so

$$(42) \quad \eta \approx \frac{1}{\sigma_s} \log \left(\frac{S}{LD} \right)$$

It follows that all these conditions can be satisfied at the same time by these expressions:

⁵⁶ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 10).

⁵⁷ STAMICAR R. - FINGER C.C. (2005, page 10).

⁵⁸ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002).

$$(43) \quad V = S + LD$$

$$(44) \quad \eta = \frac{(S+LD)}{\sigma_s S} \log \left(\frac{S+LD}{LD} \right)$$

and for V_0 we will have

$$(45) \quad V_0 = S_0 + \bar{LD}$$

$$(46) \quad \sigma = \sigma_s \frac{S}{S + \bar{LD}}$$

The last equation explains that, if the asset volatility is constant, the equity price and the equity volatility are inversely related. Near the default, equity volatility is very high, and this turns out to be coherent⁵⁹ with the so called volatility smile. Since erratic values can also be observed on the market, the model suggests the non-transitory values \hat{S} and $\hat{\sigma}_s$, in order to determine a stable level of the asset volatility⁶⁰. It is important to observe that equity volatility can be either historical or implied. Therefore, we will have

$$(47) \quad \sigma = \hat{\sigma}_s \frac{\hat{S}}{\hat{S} + \bar{LD}}$$

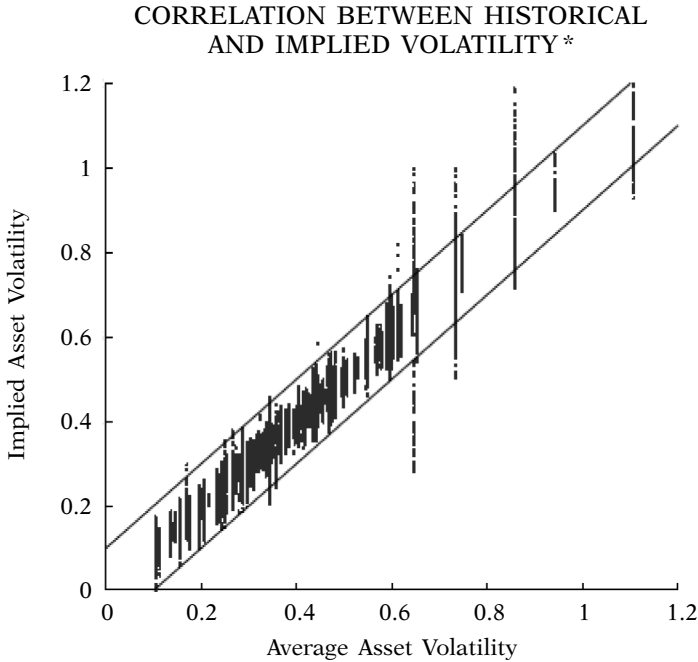
For the empirical testing of the model, which is described with details in the technical document, it is preliminarily necessary to identify an asset volatility estimator⁶¹, because it is not directly observable, and it is supposed to be constant in the long term. However, this estimator must depend on equity volatility in some way, because the model belongs to the structural approach. Thus, the CDS spread observed on the market has been chosen as the verification parameter. The authors of the model have used a sample made up of 122 firms, operating both in the

⁵⁹ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 11).

⁶⁰ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002).

⁶¹ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 17).

GRAPH 5



* FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 19).

industrial and in the financial field, and they have examined the five-year CDS spread between May 2000 and August 2001. The data concerning the specific recovery rates of the reference obligations have been taken from the J.P. Morgan database, and a total of 6,194 quotations have been used. With the standard values $\bar{L} = 0.5$ and $\lambda = 0.3$ and the boundary conditions (40)-(42), the implied asset volatility is the value of σ which makes the CDS spread equal to the one observed on the market, and this value seems to be rather stable in the long term⁶².

In Graph 5 it is evident that only for 4 firms the difference between the long term value and the one observed exceeds 10%. Therefore, historical volatility has been calculated for each firm,

⁶² FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002).

and, according to the authors of the model, the best results are obtained⁶³ on the basis of an observation period of between 750 and 1,000 days. The speculative grade firms achieve the best predictions because they show a strong relationship between equity and debt, while there is a high⁶⁴ *volatility skew* for the firms characterized by an elevated creditworthiness. In fact, their default would be similar to a case of deep out of the money. Therefore, with a temporal window of 1,000 days, 88% of the estimates fall within 10% of the value assumed to be real, that is the implied volatility of the five-year CDS spread observed on the market. Further tests have shown that this result is strong both for the industrial and the financial field, being invariant with respect to the historical epoch of observation⁶⁵.

It has been observed that the deterioration of creditworthiness is often accompanied by increments in the volatility and in the spread applied on debt: given the difficulties using historical volatility, the model has been extended⁶⁶ in order to use the one-year option implied volatility. This leads to greater accuracy of the predictions and to overcome the problems concerning data on debt, which are only available on a quarterly basis⁶⁷. Therefore, implied volatility⁶⁸ allows to capture the signals of market quotations with greater rapidity, even if the predictions become more variable. In Graph 6, it is evident that implied volatility (dash line) supplies a better evaluation of the CDS spread (black line) than the one obtained by using historical volatility (grey line), even not considering the correction for volatility skew (dots line).

Finally, the technical document describes the association measures between the predictions of the model and the data observed on the market⁶⁹: as far as the correlation index, we must

⁶³ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 18).

⁶⁴ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 19).

⁶⁵ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002).

⁶⁶ STAMICAR R. - FINGER C.C. (2005, page 1).

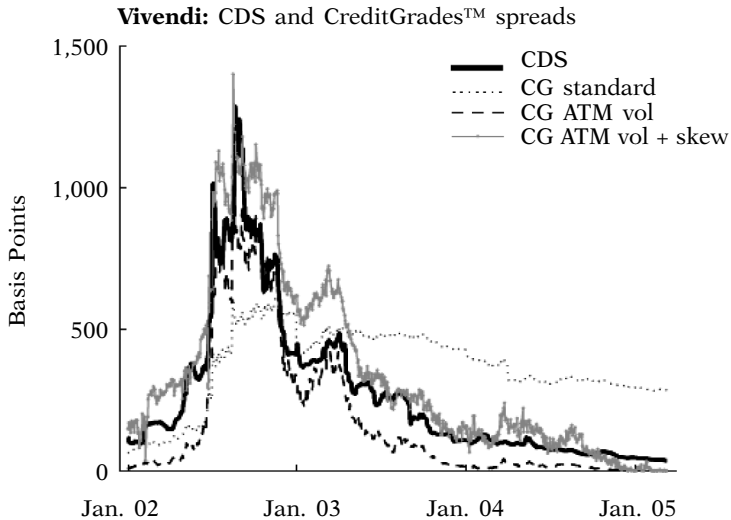
⁶⁷ HULL J.C. - NELKEN I. - WHITE A. (2004, pages 6, ss.).

⁶⁸ STAMICAR R. - FINGER C.C. (2005, page 26).

⁶⁹ FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J. (2002, page 28).

GRAPH 6

**EVALUATED SPREADS CONCERNING VIVENDI
ON THE BASIS OF THE DIFFERENT TYPES OF VOLATILITY***



* STAMICAR R. - FINGER C.C. (2005, page 18).

remember that the high default probabilities of the speculative grade firms have a greater relative weight. In any case, even considering these distortions, the value of the index is always over 60%, with a minimum probability of correct classification of the creditworthiness of 75%.

C. - Application of Neural Networks to the Credit Grades™ Model

The final part of this paper illustrates the results of the application of the neural networks carried out in the study. Topics including the architecture of the net, the number of layers and neurons, the learning process, the measures of efficiency and the evaluation of the predictive capability of the net have a very

complex nature⁷⁰, and only the verification of the various existing methodologies can help in finding their solutions. Therefore, the proposed approaches are to be seen essentially as starting points, or obligatory steps toward more detailed analysis, which are difficult to describe in general terms. In the rest of this section, the potentialities of neural networks in approximating the pricing of credit derivatives will be shown, using artificial data⁷¹, which are generated on the basis of the CreditGradesTM model. The primary reason of this choice is that in this way there are no limits for the dimensions of the available samples, neither for the in sample nor for the out of sample; moreover, we must remember the adhesion characteristics to the reality of this model.

It is therefore necessary to generate a random sample concerning the independent variables of the model (S , D , σ_S , t , r , R): because we are in the presence of homogenous equations in D ⁷², we can put $D = 1$ and replace S with S/D . As far as the range is concerned, it has been chosen $25\% < S/D < 625\%$ and $10\% < \sigma_S < 90\%$: they are values⁷³ widely diffused on the market, and in the same way it has been chosen $3\% < r < 7\%$ and $3 < t < 7$ years. The output of the net is made up of two values: the first one represents the CDS spread (which will not exceed the more than adequate value of 1,500 basis points, because of the restrictions mentioned above), and the second indicates the default probability. These two variables are closely correlated, and therefore it is possible to use a single neural network for their evaluation. The training sample is made up of 500 observations.

Table 2 shows R-squared and root mean squared error: the values are highly coherent, as it is easy to observe. The multi-layer perceptron neural network has been trained in supervised mode for 1,000 epochs of learning by using the backpropagation algorithm, with a single hidden layer of 19 neurons. We have determined both of these values with a detailed preliminary study,

⁷⁰ McNELIS P.D. (2005, page 109).

⁷¹ McNELIS P.D. (2005, page 115); HANKE M. (1997, page 21).

⁷² STAMICAR R. - FINGER C.C. (2005, page 14).

⁷³ Alternatively, it would be possible to employ volatility data from related markets to narrow down the scope of the problem for the neural network.

TABLE 2

EVALUATION OF THE APPROXIMATION
OF THE NEURAL NETWORK
(*our elaboration*)

Error tipology	Value
R-squared spread	0.99
R-squared default pr.	0.99
Root mean squared error spread	0.00147
Root mean squared error prob. default	0.00744

according to the cross-validation method⁷⁴, that divides the sample into two parts, as described in section 2. The in sample part has been used to train different nets: the differences were in the number of neurons, layers and epochs of learning; while the out of sample part has been used in order to determine which net minimizes the forecast error. According to the results mentioned in section 2, that state it is possible to approximate any nonlinear phenomenon by using neural networks, the new application to the category of credit derivatives effectively captures the market dynamics. An interesting development of the analysis is to replace the artificial sample with real market data, and the future studies are oriented in this way. In the present paper, it has instead been chosen to use artificial data, and one of the reasons is that the application to credit derivatives is new. In fact, apart from considerations about the availability of data, any attempt of evaluation of economic phenomena using neural networks must be preceded by⁷⁵ the determination of a basis model. This model is needed to relate the variables which are supposed to be relevant. In this context, the description of the CreditGrades™ model has been considered particularly useful, because of its characteristics of clarity and strength and also for its wide diffusion among

⁷⁴ McNELIS P.D. (2005, pages 115, ss.); about these concepts cfr. BISHOP C.M. (1995) and HECHT-NIELSEN R. (1991).

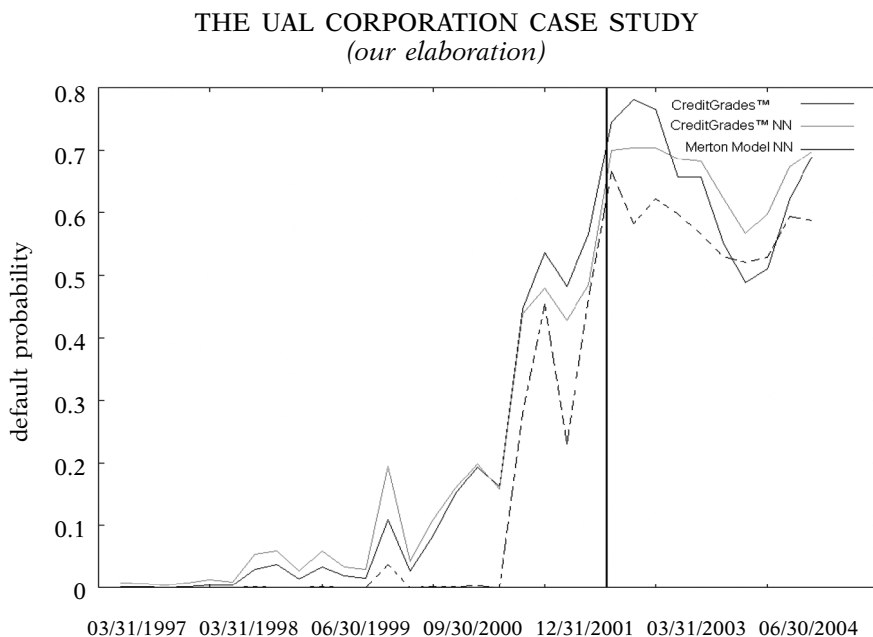
⁷⁵ For various examples cfr. McNELIS P.D. (2005, pages 145, ss.), in particular pages 148, ss.

financial operators. However, the use of market data should not lead to any problems, on the basis of the mentioned coherence of this model. Because in literature there is no unanimity on the determination of the CDS pricing function, neural networks can be seen as effective instruments, which can satisfy this lack from a statistical point of view.

D. - Prevision of the Default: the UAL Corporation Case

In conclusion of the paper, we show in Graph 7 an elaboration regarding UAL Corporation, one of the biggest international air carriers. On September, 12th 2002, this American firm applied for Chapter 11 of US Bankruptcy Code, and announced the financial reorganization on February, 1st 2006. The annualized default probabilities of this firm have been calculated on a quarterly basis. Graph 7 shows the three different models which have been used:

GRAPH 7



they are the CreditGrades™ model (in black), the neural net trained using it (in grey), and another neural net trained using the classic model of Merton (dash line). Table 3 shows the data which have been used for the elaboration: they cover the period March 1997 - December 2004. The black vertical line in Graph 7

TAB. 3

DATA CONCERNING THE UAL CORPORATION CASE STUDY
(*our elaboration*)

Date	Leverage D/E	Imp. Vol. %	Risk free rate
03/31/97	0.9008	35.5400	0.0602
06/30/97	0.8120	36.4100	0.0567
09/30/97	0.7095	35.9500	0.0547
12/31/97	0.7878	38.3700	0.0551
03/31/98	0.9293	40.2400	0.0541
06/30/98	1.1247	33.0900	0.0538
09/30/98	1.4788	46.5000	0.0441
12/31/98	1.7558	43.5600	0.0453
03/31/99	1.3761	39.3000	0.0472
06/30/99	1.5614	46.5200	0.0507
09/30/99	1.5238	38.9200	0.0522
12/31/99	1.3534	40.3100	0.0598
03/31/00	1.7447	68.9900	0.0628
06/30/00	1.6977	38.7800	0.0608
09/30/00	2.5322	44.4500	0.0607
12/31/00	3.6114	43.7700	0.0532
03/31/01	4.2346	46.1000	0.0409
06/30/01	3.9570	39.0200	0.0372
09/30/01	9.3055	78.9300	0.0249
12/31/01	13.6767	92.4500	0.0217
03/31/02	11.3719	68.7800	0.0270
06/30/02	16.0735	90.6400	0.0206
09/30/02	72.5431	287.1200	0.0153
12/31/02	5.9430	325.9800	0.0132
03/31/03	9.4754	345.5400	0.0119
06/30/03	9.3012	233.1100	0.0109
09/30/03	6.1978	226.0700	0.0115
12/31/03	4.7573	170.0500	0.0126
03/31/04	4.8408	143.7500	0.0120
06/30/04	4.9365	152.0600	0.0209
09/30/04	9.9011	203.4300	0.0221
12/31/04	7.9689	256.7400	0.0275

represents the date of default, and it is easy to observe that all the three models indicate in advance the deterioration of the creditworthiness of the firm, by increasing the default probability. However, the identifying capability of the CreditGrades™ model emerges with respect to the classic Merton model. In a certain measure, the particular neural network trained with its results boosts the good predictive characteristics of the CreditGrades™ model, by the timing of increasing the default probability. The data used for the analysis have been obtained from Bloomberg™ as regards the leverage and the implied equity volatility, and from the database of the Federal Reserve System as regards the interest rate of Treasury Bills with a constant maturity of one year, which is assumed to be the risk free rate. The specific recovery rate has been put equal to 50%.

BIBLIOGRAPHY

- ANGELINI E., «Il profilo di rischio di un portafoglio con derivati creditizi: l'elaborazione di un modello di simulazione Montecarlo», *Banca Impresa Società*, no. 2, XXI, 2002, pages 259-282.
- BISHOP C.M., *Neural Networks for Pattern Recognition*, Oxford, Oxford University Press, 1995.
- CAPUTO NASSETTI F. - FABBRI A., *Trattato sui contratti derivati di credito*, Milano, EGEA, 2000.
- DOLCINO F. - GIANNINI C. - ROSSI E., *Reti neurali artificiali per l'analisi e la previsione di serie finanziarie*, Milano, Collana studi del Credito Italiano, 1998.
- FINGER C. - FINKELSTEIN V. - LARDY J.P. - PAN G. - TA T. - TIERNEY J., *CreditGrades Technical Document*, New York, RiskMetrics Group, 2002.
- FLOREANO D. - NOLFI S., «Reti neurali: algoritmi di apprendimento, ambiente di apprendimento, architettura», *Giornale Italiano di Psicologia*, a. XX, February 1993, pages 15-50.
- FRANCIS C. - KAKODKAR A. - MARTIN B., *Credit Derivative Handbook 2003*, London, Merrill Lynch Global Securities Research and Economics Group, 2003.
- GORI M., «Introduzione alle reti neurali artificiali», *Mondo Digitale*, AICA, no. 4, settembre 2003.
- HANKE M., «Neural Network Approximation of Option-Pricing Formulas for Analytically Intractable Option-Pricing Models», *Journal of Computational Intelligence in Finance*, September/October, 1997.
- HECHT-NIELSEN R., *Neurocomputing*, Reading (MA), Addison-Wesley Publishing Company, 1991.
- HULL J.C., *Opzioni, futures e altri derivati*, Milano, Il Sole 24Ore S.p.A., 2003.
- HULL J.C. - NELKEN I. - WHITE A., «Merton's Model, Credit Risk, and Volatility Skews», *Journal of Credit Risk*, no. 1, 2004, pages 1-27.
- HULL J.C. - WHITE A., «Valuing Credit Default Swaps I: No Counterparty Default Risk», *Journal of Derivatives*, Vol. 8, no. 1, Autumn 2000, pages 29-40.
- J.P. MORGAN, *The J.P. Morgan Guide to Credit Derivatives*, Londra, Risk Publications, 2000.
- MALLIARIS M.E. - SALCHENBERGER L., «A Neural Network Model for Estimating Option Prices», *Journal of Applied Intelligence*, no. 3, 1993, pages 193-206.
- MARCO G. - VARETTO F., «Diagnosi delle insolvenze e reti neurali», Roma, Banca Editrice, 1994.
- MCNELIS P.D., *Neural Networks in Finance*, San Diego, Elsevier Academic Press, 2005.
- O'KANE D. - TURNBULL S., «Valuation of Credit Default Swaps», *Quantitative Credit Research Quarterly*, Lehman Brothers, vol. 2003-Q1/Q2, April, 2003.
- STAMICAR R. - FINGER C.C., *Incorporating Equity Derivatives into the CreditGrades Model*, New York, RiskMetrics Group, 2005.
- WHITE H., «Connectionist Nonparametric Regression: Multilayer Feed-Forward Networks Can Learn Arbitrary Mappings», *Neural Networks*, no. 3, 1990, pages 535-549.

