Predicting bank loan recovery rates with neural networks

João A. Bastos*

CEMAPRE, ISEG, Technical University of Lisbon
Rua do Quelhas 6, 1200-781 Lisbon, Portugal

Abstract

This study evaluates the performance of feed-forward neural networks to model and forecast recovery rates of defaulted bank loans. In order to guarantee that the predictions are mapped into the unit interval, the neural networks are implemented with a logistic activation function in the output neuron. The statistical relevance of explanatory variables is assessed using the bootstrap technique. The results indicate that the variables which the neural network models use to derive their output coincide to a great extent with those that are significant in parametric regression models. Out-of-sample estimates of prediction errors suggest that neural networks may have better predictive ability than parametric regression models, provided the number of observations is sufficiently large.

JEL Classification: G17; G21; G33; C45

Keywords: Loss given default; Recovery rate; Forecasting; Bank loan; Fractional regression; Neural network

This version: September 2010

1 Introduction

With the advent of the new Basel Capital Accord (Basel Committee on Banking Supervision, 2006), banking organizations may choose between two approaches for determining credit risk capital requirements: a *standardized* approach relying on ratings attributed by external agencies for risk-weighting assets, and an *internal ratings based* (IRB) approach in which institutions may implement their own internal models to calculate credit risk capital charges. Banks that adopt the advanced variant of the IRB approach are expected to provides estimates of the *loss given default*: the credit that is lost when a borrower defaults, expressed as a fraction of the exposure at default. Modeling loss given default

*E-mail address: `jbastos@iseg.utl.pt`

(or its complement, the *recovery rate*) on defaulted obligations is a challenging task. First, excluding the costs of workout processes or potential gains in asset sales, recovery rates are only observed on the interval $[0, 1]$. This imposes the use of econometric techniques that take into account the bounded nature of recovery rates. Second, recovery rate distributions are frequently bimodal, containing many observations with very low recoveries and many with complete or near complete recoveries (see, e.g., Asarnow and Edwards, 1995; Felsovalyi and Hurt, 1998; Davydenko and Franks, 2008; Araten et al., 2004; Dermine and Neto de Carvalho, 2006; Caselli et al., 2008). Third, empirical studies show that it is not easy to find explanatory variables that strongly influence recovery rates.

The most straightforward technique for modeling recoveries is the linear regression model estimated by ordinary least squares methods. For example, this approach is employed in Caselli et al. (2008), Davydenko and Franks (2008) and Grunert and Weber (2009). However, modeling and predicting recoveries with a linear model has serious limitations. First, because the support of the linear model is the real line it does not ensure that predicted values lie in the unit interval. Also, given the bounded nature of the dependent variable, the partial effect of any explanatory variable cannot be constant throughout its entire range. These limitations can be overcome by employing an econometric methodology specifically developed for modeling proportions, such as the (nonlinear) fractional regression estimated using quasi-maximum likelihood methods (Papke and Wooldridge, 1996). In the context of credit losses, this approach was adopted in Dermine and Neto de Carvalho (2006) and Chalupka and Kopecsni (2009). An alternative procedure is to perform the regression on appropriately transformed recoveries. The most eminent example of this technique is Moody's LossCalc$^{\text{TM}}$ V2 (Gupton and Stein, 2005), in which recoveries are normalized via a beta distribution and a linear regression is carried out on the transformed data set. A distinct approach is offered by nonparametric models, in which the functional form for the conditional mean of the response variable is not predetermined by the researcher but is derived from information provided by the data. For example, Bastos (2010) suggested the use of nonparametric regression trees for modeling recoveries on bank loans. The advantage of this technique is its interpretability, since tree models resemble 'look-up' tables containing historical recovery averages. Furthermore, because the predictions are given by recovery averages, they are inevitably bounded to the unit interval.

The purpose of this study is to investigate the performance of artificial neural networks to forecast bank loan recoveries. An artificial neural network is a nonparametric mathematical model that attempts to emulate the functioning of biological neural networks. It consists of a group of interconnected processing units denoted by *neurons*. Due to their good capability of approximating arbitrary complex functions, neural networks have been applied in a wide range of scientific domains. In particular, neural networks have been successfully employed in modeling the *probability of default* (see, e.g., Altman et al., 1994), which, together with the recovery rate, determines the expected credit loss of a financial asset. In this study, neural networks are trained to identify and learn patterns in a set of recovery rates of defaulted bank loans. In regression problems, off-the-shelf neural network implementations typically employ linear activation functions in the output neuron. In this analysis, on the contrary, the neural networks are implemented with a logistic activation function in the output neuron, since this choice guarantees that the predicted values are constrained to the unit interval.

The data set employed in this study contains the monthly history of cash flows recovered by the bank during the workout process. This allows for the estimation of out-of-sample predictive accuracies at several recovery horizons after default, and to understand the properties of neural networks under different recovery rate distributions and number of observations. The parametric model against which the performance of neural networks is benchmarked is the fractional regression of Papke and Wooldridge (1996). The performance of these techniques is also benchmarked against simple predictions given by average recoveries. It is shown that neural networks have better predictive ability than parametric regressions, provided the number of observations is sufficiently large.

The statistical relevance of explanatory variables and the direction of partial effects given by the neural network and fractional regression models are compared. Because the opaqueness of neural networks precludes the derivation of valid statistical tests to assess the importance of input variables, the sampling distributions and appropriate critical values are obtained using a resampling technique known as the *bootstrap* (Efron, 1979). It is shown that the significance of the explanatory variables and the direction of the partial effects given by the neural network models are, in general, compatible with those given by the fractional regressions.

The remainder of this paper is organized as follows. The next section describes the data set of bank loans and the explanatory variables. Section 3 reviews the parametric fractional regression and neural network techniques. The determinants of recovery rates given by the parametric and neural network models are discussed in Section 4. In Section 5, a comparison of in-sample and out-of-sample predictive accuracies given by these models is presented. Finally, Section 6 provides some concluding remarks.

2 Data sample and variables

This study is based on the bank loan data set of Dermine and Neto de Carvalho (2006). It consists of 374 loans granted to small and medium size entreprises (SMEs) by a bank in Portugal. The defaults occurred between June 1995 and December 2000, and the mean loan amount is 140,874 euros.[1]

Borrowing firms are classified into four groups according to their business sector: (i) the *real* sector (activities with real assets, such as land, equipment or real estate), (ii) the *manufacturing* sector, (iii) the *trade* sector, and (iv) the *services* sector. To each individual loan is attributed a rating by the bank's internal rating system. The rating reflects not only the probability of default of the loan but also the guarantees and collateral that support the operation. There are seven classes of rating: A (the best), B, C1, C2, C3, D and E (the worst). These alpha-numeric rating notches were transformed into numeric values by an ordinal encoding that assigns the value 1 to rating A, the value 2 to rating B, and so on. Nearly half of the loans had no rating attributed. In order to avoid the exclusion of loans with missing rating class, which would reduce significantly the number of available observations, surrogate ratings were given to unrated loans by

[1]Because this study is essentially methodological, only a brief review of the data is given here. For a comprehensive description of the data, see Dermine and Neto de Carvalho (2006).

multiple imputation.[2] Fifty eight per cent of the loans are covered by personal guarantees. These are written promises that grant to the bank the right to collect the debt against personal assets pledged by the obligor. Fifteen per cent of the loans are covered by several varieties of collateral. These include real estate, inventories, bank deposits, bonds and stocks. Thirty six per cent of the loans are not covered by personal guarantees or any form of collateral. The loans are also characterized by the contractual lending rate, the age of the borrowing firm and the number of years of relationship with the bank. The mean age of the firms is 17 years while the mean age of relationship with the bank is 6 years. The values of the rating, collateral and personal guarantees are those recorded at default.

	1995	1996	1997	1998	1999	2000	Total
All data	65	89	59	57	47	57	374
12 month horizon	65	89	59	57	47		317
24 month horizon	65	89	59	57			270
36 month horizon	65	89	59				213
48 month horizon	65	89					154

Table 1: Number of loans organized by year of default, and the number of loans that are used for each recovery horizon.

Recovery rates are estimated using the discounted value of cash-flows recovered by the bank after default.[3] The database contains the monthly history of cash-flows received by the bank after the loans became non-performing. These cash-flows include incoming payments due to realizations of collateral. For some loans, those defaulted in June 1995, a long recovery history of 66 months is available. As the default date approaches the end of year 2000, the recovery history is shortened. The dates when the defaulted loans were officially written-off by the bank are not available and, therefore, ultimate recoveries cannot be calculated, as required by Basel II. Instead, cumulative recovery rates are calculated for horizons of 12, 24, 36 and 48 months after default. It should be noted that, since most cash-flows are received shortly after default (see Dermine and Neto de Carvalho, 2006), the distribution of recoveries for the longest recovery horizons are good approximations of the distribution of ultimate recoveries. Furthermore, calculating recovery rates at several horizons after default is useful for understanding the performance of neural networks under different recovery rate distributions and number of observations. The first row in Table 1 shows how the 374 loans included in the data are distributed across the years. The second row shows that, when cumulative recoveries are calculated for an horizon of 12 months after default, defaults that occurred in year 2000 are not considered since they do not include 12 months of recovery history. In this case, the data sample is reduced to 317 observations. For longer recovery horizons, the number of available observations is further reduced, as indicated in Table 1.

The discount rate that is used to compute the present value of the post-default cash-

[2]Five imputations were generated using the program Amelia II (`http://gking.harvard.edu/amelia`). As discussed in King et al. (2001), 5 imputations are usually sufficient, unless the number of missing values in the complete data set is exceptionally high. The prediction errors in Section 5

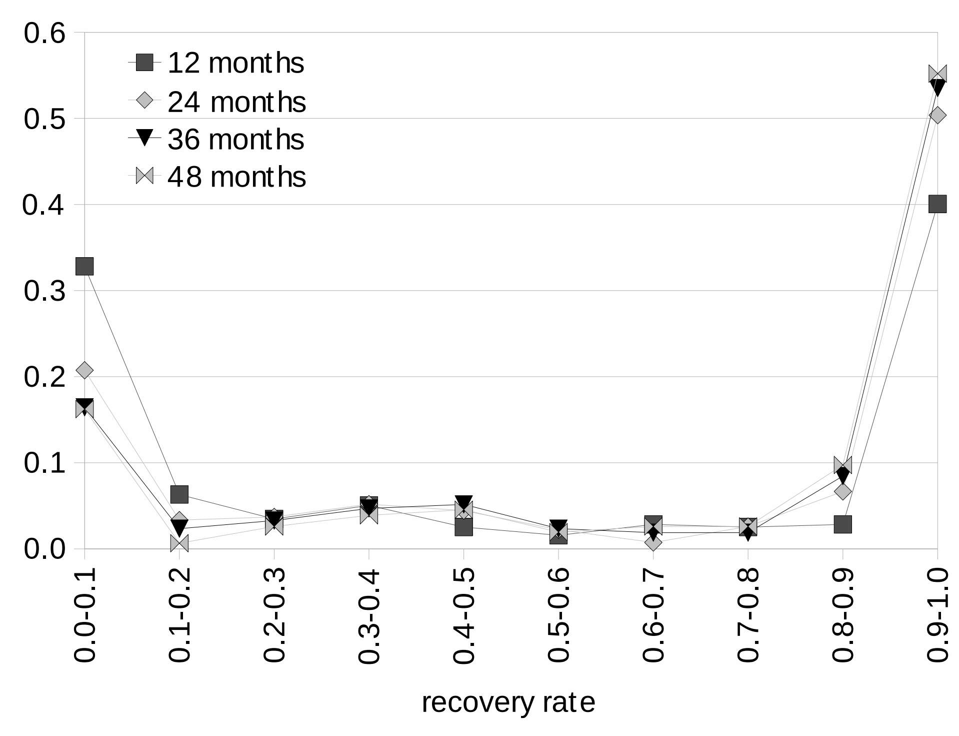

Figure 1: Distribution of the cumulative recovery rate for recovery horizons of 12, 24, 36 and 48 months after the default event.

flows is the loan-specific contractual lending rate. While this rate may not capture the total risk of the firm after default, a substantial part of the total recovery is collected in the first months of the work-out process and, therefore, calculated recoveries should not change dramatically with the discount rate. The costs of the resolution process are not considered in the calculation of the recovery rates.[4] Figure 1 shows the distribution of cumulative recovery rates for horizons of 12, 24, 36 and 48 months after the default event. It can be seen that the distributions are bimodal with many observations with low recoveries and many with complete or near complete recoveries. Also, there are no substantial differences between recovery rate distributions for horizons of 24, 36 and 48 months. Naturally, this is a consequence of marginal recovery rates that decrease rapidly with time after default.

correspond to average values over the five imputed data sets.

[3]Secondary market prices at the time of emergence from default were not available.

[4]Data on workout costs incurred in the recovery process are not available for individual loans and, therefore, all results refer to gross recoveries. Dermine and Neto de Carvalho (2006) estimate that the average recovery cost incurred by the bank is 2.6% of the amount recovered.

3 Models

3.1 Parametric fractional regression

As mentioned in the introductory section, linear models estimated with ordinary least squares methods are not appropriate for modeling recovery rates, since the predicted values are not guaranteed to be bounded to the unit interval. An appropriate parametric model for recovery rates is the fractional regression of Papke and Wooldridge (1996) which was specifically developed for modeling fractional response variables. Let y be the variable of interest (i.e., recovery rates) and $\mathbf{x}$ the vector of explanatory variables (i.e., firm and contract characteristics). The fractional regression model is

$$E(y|\mathbf{x}) = G(\mathbf{x}\beta), \tag{1}$$

where $G(\cdot)$ is some nonlinear function satisfying $0 < G(z) < 1$ for all $z \in \mathbb{R}$. As suggested by Papke and Wooldridge (1996), a consistent and asymptotically normal estimator of β may be obtained by maximization of the Bernoulli quasi-likelihood function,

$$L(\beta) \equiv y \log[G(\mathbf{x}\beta)] + (1 - y) \log[1 - G(\mathbf{x}\beta)]. \tag{2}$$

Common choices for $G(\cdot)$ are the cumulative normal distribution, the logistic function, and the log-log function.

3.2 Artificial neural network

A feed-forward neural network consists of a group of elementary processing units (denoted by *neurons*) interconnected in such way that the information always moves in one direction. The most prominent type of feed-forward network is the multilayer perceptron, in which the neurons are organized in layers and each neuron in one layer is directly connected to the neurons in the subsequent layer. In practice, multilayer perceptrons with three layers are mostly used, as illustrated in Figure 2. The input layer consists of $m + 1$ inputs corresponding to m explanatory variables and an additional constant input called the *bias*. The output layer contains a number of neurons equal to the number of dependent variables (one in this case). The layer between the input and output layers is called the *hidden layer*. The number of neurons in the hidden layer h is determined by optimization of the network performance. The hidden layer also includes a constant bias.

The output of each neuron is a weighted sum of its inputs that is put through some *activation function*. Denoting by $w^{(1)}$ the weights of the connections between the inputs and the hidden neurons, and by $w^{(2)}$ the weights of the connections between the hidden neurons and the output neuron, the network's output is given by

$$\hat{f}(\mathbf{x}; \mathbf{w}) = \Phi_2 \left(\sum_{k=1}^{h} w_k^{(2)} \Phi_1 \left(\sum_{j=1}^{m} w_{jk}^{(1)} x_j + w_{m+1,k}^{(1)} \right) + w_{h+1}^{(2)} \right), \tag{3}$$

where the function $\Phi_1(\cdot)$ and $\Phi_2(\cdot)$ are the activation functions of the neurons in the hidden and output layers, respectively. Commonly chosen activation functions for the hidden neurons are the logistic function and the hyperbolic tangent. Also, typical neural

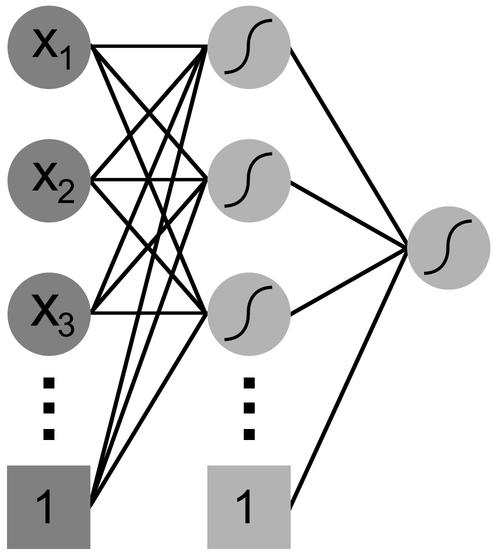

Figure 2: Scheme of a multilayer perceptron with three layers. The dark circles represent the network's inputs. The light circles represent the neurons. The input layer contains $m + 1$ inputs corresponding to m explanatory variables and an additional constant input called the *bias*, which is represented by a square. The output layer contains a number of neurons equal to the number of dependent variables (one in this case). The layer between the input and output layers is called the *hidden layer*. The number of neurons in the hidden layer is determined by optimization of the network performance. The hidden layer also includes a constant bias.

network implementations employ a linear activation function in the output neuron. In this study, a logistic activation function is used in the hidden layer neurons. In order to ensure that the network predictions are mapped into the unit interval, a logistic activation function is also employed in the output neuron.

Given n observations, learning occurs by comparing the network output $\hat{f}$ with the desired output y, and adjusting iteratively the connection weights in order to minimize the loss function

$$L(\mathbf{w}) = \frac{1}{2n} \sum_{i=1}^{n} \left[y_i - \hat{f}(\mathbf{x}_i; \mathbf{w}) \right]^2 . \tag{4}$$

After a sufficiently large number of iterations, the network weights converge to a configuration where the value of the loss function is small. The weights are adjusted by a non-linear optimization algorithm, called *gradient descent*, that follows the contours of the error surface along the direction with steepest slope.

4 Significance of explanatory variables

In Papke and Wooldridge (1996) fractional regression model, the partial effects of explanatory variables on the response variable are not constant, given that the function $G(\cdot)$ in Equation 1 is nonlinear. However, the chain rule gives that the partial effect of variable x_j is

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} = \frac{dG(\mathbf{x}\beta)}{d(\mathbf{x}\beta)}\beta_j. \tag{5}$$

Since $G(\cdot)$ is strictly monotonic, the sign of the coefficient gives the direction of the partial effect. The quasi-maximum likelihood estimator of β is consistent and asymptotically normal regardless of the distribution of the response variable conditional on $\mathbf{x}$ (Gourieroux et al., 1984).

With respect to the neural network models, it is not trivial to derive the direction of the partial effects and understand if explanatory variables have significant effects on the network's output. In order to circumvent this problem, Baxt and White (1995) suggested the bootstrap technique. Denoting by $\hat{f}$ the trained network output function, the partial effect on recovery rates of perturbing variable x_j is approximated by the *network sample mean delta*

$$\Delta_j \hat{f} \equiv \frac{1}{n}\sum_{i=1}^{n} \Delta_j \hat{f}(\mathbf{x}_i), \tag{6}$$

where $\Delta_j \hat{f}(\mathbf{x}_i)$ denotes the change in the network's output by perturbing the jth component of the ith observation. Because $\hat{f}$ is an estimate of the true relation between recovery rates and explanatory variables, it is subject to sampling variation. Therefore, partial effects that are truly zero may seem to be nonzero and vice versa. The sampling variation may be derived by drawing a large number of pseudosamples of size n with replacement from the original sample. For each of the pseudosamples the network sample mean deltas are calculated and the bootstrap distribution and corresponding critical values of $\Delta_j \hat{f}$ are obtained.[5]

Table 2 shows the statistical significance and direction of the partial effects of the explanatory variables for the fractional regression (FR) models and the neural network (NN) models, and for recovery horizons of 12, 24, 36 and 48 months after default.[6] The symbols -, - - and - - - indicate that a variable has a *negative* effect on recoveries with a statistical significance of 10%, 5% and 1%, respectively; the symbols +, ++ and +++ indicate that a variable has a *positive* effect on recoveries with a statistical significance of 10%, 5% and 1%, respectively; and a bullet (•) means that a variable is *not* statistically significant. The results for the fractional regression models were obtained with the logistic function $G(\mathbf{x}\beta) = 1/(1 + \exp(-\mathbf{x}\beta))$. With respect to the neural network models, three parameters were optimized: (i) the number of neurons in the hidden layer, (ii) the *learning rate*, which determines the size of the changes in the network weights during the learning

[5]Baxt and White (1995) also suggest an alternative approach based on residual sampling. This involves creating pseudosamples in which the input patterns from the original sample are maintained, but the recovery rates are perturbed using n residuals obtained by sampling with replacement from the sample of residuals given by the neural network trained on the original data. However, this approach is not feasible here since the recovery rates of these pseudosamples would not be necessarily constrained to the interval [0,1].

[6]The coefficients of the fractional regression models are average values over the five imputed data sets. Also, the standard errors of the coefficients are corrected in order to account for the variance of the coefficients across the five imputed data sets (For details, see, King et al., 2001). With respect to the neural network models, the critical values of the bootstrap distributions are average values over the five imputed data sets.

	12 months		24 months		36 months		48 months	
	FR	NN	FR	NN	FR	NN	FR	NN
Loan size	•	- - -	- -	- - -	- - -	- -	- - -	- - -
Collateral	•	•	+	++	+	++	+++	+++
Personal guarantee	•	•	-	-	•	•	-	-
Manufacturing sector	•	•	•	-	•	- -	- - -	- - -
Trade sector	•	•	•	-	- -	- - -	- - -	- - -
Services sector	•	•	•	•	•	•	•	•
Lending rate	•	•	•	•	•	•	•	•
Age of firm	+	•	+	•	+	+	++	+
Rating	- -	- - -	- -	- - -	-	- -	•	•
Years of relationship	•	++	+	+++	•	+	•	+

Table 2: Statistical significance and direction of partial effects of explanatory variables given by the fractional regression (FR) models and the neural network (NN) models, and for recovery horizons of 12, 24, 36 and 48 months. The symbols -, - - and - - - indicate that a variable has a negative effect on recoveries with a statistical significance of 10%, 5% and 1%, respectively; the symbols +, ++ and +++ indicate that a variable has a positive effect on recoveries with a statistical significance of 10%, 5% and 1%, respectively; and the symbol • means that a variable is not statistically significant.

process, and (iii) the *momentum term*, which determines how past network weight changes affect current network weight changes. Neural network training is stopped when the out-of-sample error is no longer improved. The amount of training cycles is defined by a parameter called the *number of epochs*. The critical values for the neural network models were obtained from 1000 bootstrap pseudosamples.

The results in Table 2 show that, to a large extent, the neural network models are in agreement with the parametric models in terms of which variables determine recovery rates. For instance, excluding the fractional regression for 12 months horizon, all models indicate that the size of the loan has a statistically significant negative effect on recovery rates at all recovery horizons.[7] Also, both techniques suggest that collateral has a statistically significant positive impact on recovery rates for the longest recovery horizons of 24, 36 and 48 months and personal guarantees have a statistically significant negative impact on recoveries for 24 and 48 months horizon.

The models also suggest that the manufacturing and trade sectors present lower recoveries with respect to the base case (the real sector) for longer horizons. However, for a 24 months horizon, the neural networks indicate that these sectors may also have a negative impact on recoveries while the fractional regressions suggest that they are not significant. On the other hand, both techniques reveal that the services sector is not statistically significant at all recovery horizons. Similarly, the contractual lending rate does not appear to have any effect on recovery rates across all horizons.

The fractional regressions indicate that the age of the firm has a positive effect on recoveries regardless of the recovery horizon. That is, according to these models, older firms should exhibit better recoveries. On the other hand, the neural networks indicate

[7]See Dermine and Neto de Carvalho (2006) for the empirical interpretation of these effects.

that this variable is only important for longer recovery horizons. Both techniques indicate that the rating is significant for 12, 24 and 36 months recovery horizons. The sign of the coefficients is in agreement with the expected direction of the partial effect: poor creditworthiness results in lower recoveries. According to the neural network models, the age of the obligor's relationship with the bank is relevant for all recovery horizon, and longer relationships result in better recoveries. On the other hand, for the fractional regression this variable is only significant for a recovery horizon of 24 months.

5 Forecasting performance

The predictive accuracy of the models is assessed using two widespread measures: the root-mean-squared error (RMSE) and the mean absolute error (MAE). These are defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \quad \mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|, \tag{7}$$

where y_i and $\hat{y}_i$ are the actual and predicted values of observation i, respectively, and n is the number of observations in the sample. Models with lower RMSE and MAE have smaller differences between actual and predicted values and predict actual values more accurately. However, RMSE gives higher weights to large errors and, therefore, this measure may be more appropriate when these are particularly undesirable.

Because the developed models may overfit the data, resulting in over-optimistic estimates of the predictive accuracy, the RMSE and MAE must also be assessed on a sample which is independent from that used for estimating the models. In order to develop models with a large fraction of the available data and evaluate the predictive accuracy with the complete data set, a 10-fold cross-validation is implemented. In this approach, the original sample is partitioned into 10 subsamples of approximately equal size. Of the 10 subsamples, a single subsample is retained for measuring the predictive accuracy (the *test set*) and the remaining 9 subsamples are used for estimating the model. This is repeated 10 times, with each of the 10 subsamples used exactly once as test data. Then, the errors from the 10 folds can be averaged or combined to produce a single estimate of the prediction error.

Table 3 shows in-sample and out-of-sample RMSEs and MAEs of the recovery rate predictions given by the fractional regressions and the neural networks. The out-of-sample errors correspond to average values over 100 test sets obtained from 10 random 10-fold cross validations. Also shown are the errors given by a simple model in which the predicted recovery is given by the average of actual recoveries (*Historical* model). The last row in Table 3 shows the corrected resampled T-test (Nadeau and Bengio, 2003) for the null hypothesis that the out-of-sample prediction errors of the fractional regressions and the neural networks are equal.[8]

As anticipated, in-sample errors are typically smaller than out-of-sample errors since the models overfit the data, giving over-optimistic estimates of the predictive accuracy.

[8] Denote by $\varepsilon_i^{(1)}$ and $\varepsilon_i^{(2)}$ the prediction errors in test set i given by models 1 and 2, respectively, and let N denote the total number of test sets. The corrected resampled test for the equality of mean errors

	12 months		24 months		36 months		48 months	
Model	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
In-sample								
FR	0.4139	0.3804	0.3841	0.3381	0.3577	0.3053	0.3371	0.2732
NN	0.3750	0.3235	0.3541	0.2970	0.3312	0.2688	0.3166	0.2510
Out-of-sample								
Historical	0.4365	0.4163	0.4099	0.3803	0.3840	0.3444	0.3736	0.3243
FR	0.4297	0.3951	0.3983	0.3526	0.3749	0.3224	0.3580	0.2947
NN	0.4145	0.3586	0.3870	0.3272	0.3671	0.2965	0.3631	0.2946
$T_{\text{NN}-\text{FR}}$	-2.24*	-5.55**	-1.51	-3.75**	-0.97	-3.82**	1.04	-0.02

Table 3: In-sample and out-of-sample root-mean-squared errors (RMSE) and mean absolute errors (MAE) of the recovery rate estimates, for recovery horizons of 12, 24, 36 and 48 months, and for the fractional regression (FR) models, the neural networks (NN) models and the model in which the predicted recovery is equal to the historical average. The numbers for out-of-sample evaluation refer to average values over 100 test sets obtained from 10 random 10-fold cross-validations. Also shown is the corrected resampled T-test for the null hypothesis that the errors of the fractional regressions and the neural networks are equal. One (*) and two (**) asterisks mean that the null is rejected with 5% and 1% significance level, respectively.

Therefore, the models not only fit the "true" relationship between recovery rates and the explanatory variables but also capture the idiosyncrasies ("noise") of the data employed in their estimation. Both fractional regressions and neural networks give better forecasts than simple predictions based on historical averages. The neural network models have a statistically significant better predictive accuracy than the fractional regressions for a recovery horizon of 12 months, both in term of RMSE and MAE. For horizons of 24 and 36 months, the neural networks also outperform the fractional regression, but only in terms of MAE. Finally, both models exhibit comparable prediction errors for a horizon of 48 months.

These results suggest that the neural network's accuracy may be penalized by the decreasing number of observations as the recovery horizon increases (see last column in Table 1). In order to test this hypothesis, the analysis was repeated on three random subsets of the 12 months horizon set, containing 270, 213 and 154 observations (that is, the number of observation in 24, 36 and 48 months horizon sets). The results shows that on these "reduced" 12 months horizon data sets the predictive advantage of the neural

$m^{(1)} = \frac{1}{N}\sum_i \varepsilon_i^{(1)}$ and $m^{(2)} = \frac{1}{N}\sum_i \varepsilon_i^{(2)}$ is given by

$$T = \frac{m^{(1)} - m^{(2)}}{\sqrt{\left(\frac{1}{N} + q\right) S}},$$

where $S = \mathrm{Var}(\varepsilon^{(1)} - \varepsilon^{(2)})$ and q is the ratio between the number of observations in the test set and the number of observations in the training set. Here, because 10 random 10-fold cross-validations are generated, $q = 0.1/0.9$ and $N = 100$. The corrected resampled T-test follows a Student's t-distribution with $N - 1$ degrees of freedom.

networks with respect to the fractional regressions is generally lost. In fact, the neural networks only have a statistically significant advantage over the fractional regression in the subset with 270 observations and solely in terms of MAE.

6 Conclusions

This study evaluates the performance of neural networks to forecast bank loan credit losses. The properties of the neural network models are compared with those of parametric models obtained from fractional regressions. The neural networks are implemented with a logistic activation function in the output neuron, in order to guarantee that the predictions are mapped into the unit interval. Recovery rate models for several recovery horizons are implemented and analyzed. In the neural network models, the statistical relevance of explanatory variables is assessed using the bootstrap technique. It is shown that there are few divergences with respect to which variables the network models use to derive their output and those that are statistically significant in fractional regression models. Furthermore, when a variable is significant according to both techniques, the direction of the partial effect is usually the same. Nevertheless, some discrepancies can be found. For instance, while neural networks suggest that the age of relationship of the bank with the client has a positive effect on recoveries across all horizons, the fractional regressions indicate that this variable only has a positive impact on recoveries for the 24 months horizon. On the other hand, the fractional regressions suggest that the age of the firm is relevant across the four horizons, while according to the neural networks this variable is only important for the longest horizons.

Out-of-sample estimates of the prediction errors are evaluated. The results indicate that neural networks models have a statistical significant predictive advantage over regression models for a recovery horizon of 12 months in terms of RMSE and MAE, and for recovery horizons of 24 and 36 months in terms of MAE. For a recovery horizon of 48 months, the predictive ability of the two techniques is comparable. However, the decline of the neural networks performance for longer horizons may be related to the reduced number of observations when the recovery horizon is increased.

Acknowledgments

I am grateful to Cristina Neto de Carvalho for kindly supplying the data for this study. This research was supported by a grant from the Fundação para a Ciência e a Tecnologia.

References

Altman, E., Marco, G., and Varetto, F., 1994. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks. Journal of Banking & Finance 18, 505-529.

Araten, M., Jacobs Jr., M., Varshney, P., 2004. Measuring LGD on commercial loans: An 18-year internal study. The RMA Journal 4, 96-103.

Asarnow, E., Edwards, D., 1995. Measuring loss on defaulted bank loans: A 24-year study. Journal of Commercial Lending 77, 11-23.

Basel Committee on Banking Supervision, 2006. International convergence of capital measurement and capital standards. Bank for International Settlements.

Bastos, J.A., 2010. Forecasting bank loans loss-given-default. Journal of Banking & Finance 34, 2510-2517.

Baxt, W.G., and White, H., 1995. Bootstrapping confidence intervals for clinical input variable effects in a network trained to identify the presence of acute myocardial infarction. Neural Computation 7, 624-638.

Caselli, S., Gatti, S., Querci, F., 2008. The sensitivity of the loss given default rate to systematic risk: New empirical evidence on bank loans. Journal of Financial Services Research 34, 1-34.

Chalupka, R., Kopecsni, J., 2009. Modelling bank loan LGD of corporate and SME segments: A case study. Czech Journal of Economics and Finance 59, 360-382.

Davydenko, S.A., Franks, J.R., 2008. Do bankruptcy codes matter? A study of defaults in France, Germany and the U.K. The Journal of Finance 63, 565-608.

Dermine, J., Neto de Carvalho, C., 2006. Bank loan losses-given-default: A case study. Journal of Banking & Finance 30, 1291-1243.

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. The Annals of Statistics 7(1), 1-26.

Felsovalyi, A., Hurt, L., 1998. Measuring loss on Latin American defaulted bank loans: A 27-year study of 27 countries. Journal of Lending and Credit Risk Management 80, 41-46.

Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo-maximum likelihood methods: Theory. Econometrica 52, 681-700.

Grunert, J., Weber, M., 2009. Recovery rates of commercial lending: Empirical evidence for German companies. Journal of Banking and Finance 33, 505-513.

Gupton, G.M., Stein, R.M., 2005. LossCalc V2: Dynamic prediction of LGD. Moody's Investors Service.

King, G., Honaker, J., Joseph, A., Scheve, K., 2001. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. American Political Science Review 95, 49-69.

Nadeau, C., Bengio, Y., 2003. Inference for the generalization error. Machine Learning 52, 239-281.

Papke, L.E., Wooldridge, J.M., 1996. Econometric methods for fractional response variables with an application to 401(K) plan participation rates. Journal of Applied Econometrics 11, 619-632.