Kim Ristolainen

Were the Scandinavian Banking Crises Predictable? A Neural Network Approach

Aboa Centre for Economics
Discussion paper No. 99
Turku 2015

The Aboa Centre for Economics is a joint initiative of the economics departments of the University of Turku and Åbo Akademi University.

ISSN 1796-3133

Printed in Uniprint
Turku 2015

Kim Ristolainen

Were the Scandinavian Banking Crises Predictable? A Neural Network Approach

Aboa Centre for Economics
Discussion paper No. 99
January 2015

ABSTRACT

The early warning system literature on banking crises has often relied on linear classifiers such as the logit model, which are usually estimated with large datasets of multiple regions of countries. We construct an EWS based on an artificial neural network model with monthly data from the Scandinavian countries to tackle the poor generalization ability of the usual models that might be due to regional heterogeneity of the countries and a nonlinear decision boundary of the classification problem. We show that the Finnish and Swedish banking crises in 1991 were quite predictable with an artificial neural network model when information from earlier crises in Denmark and Norway was used. We also use cross validation in the model selection process to get the optimal amount of complexity to the models. Finally the area under the ROC-curve is used as the model assessment criteria and in this framework we show that the artificial neural network outperforms the logit regression in banking crises prediction.

JEL Classification: G21, C45, C52

Keywords: Early Warning System, Banking Crises, Scandinavia, Neural Networks, Validation

Contact information

Kim Ristolainen
Department of Economics
University of Turku
FI-20014, Finland
Email: kkrist (at) utu.fi

1 Inroduction

The financial crises of emerging economies in the 1980s spurred a literature of early warning systems for currency crises. These predictive models are supposed to give a warning signal of an upcoming financial crises in some given time window so that the policy makers can take precautionary measures in a good time. The models are constructed using various econometric and statistical methods with information from crisis indicators, which are usually taken from economic theory.

The two seminal papers in the field use different approaches in constructing these early warning models. Kaminsky, Lizondo & Reinhart (1998) use a nonparametic univariate signals approach, where the idea is to compare the behaviour of economic fundamentals during normal periods of the economy and in pre-crisis periods. In practice an optimal threshold for the crisis signal is calculated for each indicator by minimizing the noise-to-signal ratio[1] e.g. inflation over 10 procent would give a signal of an upcoming crisis in some given time window. Berg & Pattillo (1999) use a probit model for the same purpose. The probit/logit model has the advantage of taking into account the correlation between the indicators and it can aggregate the information into a composite index more sufficiently. They find that the probit model has superior performance in predicting the currency crises of the 1990s, when compared to the nonparametric KLR model. After these two papers the literature has spread to different types of financial crises and more complex/advanced models[2]. Manasse, Roubini & Schimmelpfennig (2003) use classification and regression tree analysis to construct an EWS for debt crises. Nag & Mitra (1999) use dynamic artificial neural network (ANN) to predict currency crises. Frank & Schmied (2003) do similar analysis with a different dataset. Both papers find ANN to be a superior predictor compared with the logit model. Fioramanti (2008) uses ANN to predict sovereign debt crises with a dataset of 46 emerging countries. He concludes that ANN outperforms the traditional probit EWS in crises prediction.

Although the costs of systemic banking crises are compelling[3], the EWS literature on banking crises is rather succinct compared with the one for currency crises. Demirguc-Kunt & Detragiache (1998) have the seminal paper in the banking crises EWS literature. They use a multivariate logit model with a dataset of 77 countries. Kaminsky & Reinhart (1999) use the signals approach to study the occurance of twin crises[4] with a dataset of 22 emerging countries. Davis & Karim (2008) compare these methods with an updated dataset. They conclude that the signals approach might be better for country specific banking crises EWS and the multivariate logit for a global early warning system.

For our knowledge, artificial neural networks have not been used in constructing an EWS for banking crises. This paper tries to fill this gap in the literature and builds an

[1] Ratio of correctly called crises and incorrectly called crises periods.

[2] For a comprehensive literature review of artificial neural network based EWS see Sarlin (2012).

[3] Average bailout costs 10% of GDP (Caprio & Klingebiel, 1996) and average estimated cumulative output losses 5.6% of GDP (Hoggart & Saporta, 2001).

[4] Occurrence of banking and currency crises at the same time.

early warning system with two different approaches to predict the Scandinavian banking crises of the early 1990s. Three of these crises were so severe that they were labelled by Reinhart & Rogoff (2008b) to be part of "the big five" banking crises of the post-Second World War era. Recent EWS studies (Davis, Karim & Liandze, 2011) have also argued that EWSs should be build for each region by themselves, because of regional heterogeneity of the indicator variables signalling the crises. That is why in this paper data is from four Scandinavian countries that experienced systemic banking crises in the late 1980s and early 1990s. The objective of this paper is to examine if the out-of-sample prediction of the Finnish and Swedish banking crises in 1991 could have been possible given the information from the previous Scandinavian crises in Denmark and Norway. In the process we will compare the classification ability of the artificial neural network model and the usual logit model with the area under the ROC curve as the criteria.

So why use artificial neural network for early warning systems? Many early warning systems have good in-sample results, but poor out-of-sample predictions. This might be due to the fact that the usual methods in EWS literature - such as the multivariate logit - model the banking crises probability with linear combinations of the crises indicators, which are transformed with a link function to be between zero and one. This means that the model can only solve linearly separable classification problems. However these problems might not in many cases be linearly separable and a non-linear function of the indicators should be used before the transformation to a probability with a link function. The relationship between the banking crises probability and the indicators might be nonlinear[5] e.g. the marginal increase of the crises probability might be larger for more severe declines of real GDP growth. Artificial neural networks are nonparametric statistical models originally inspired by neuroscience, that can approximate any continious function - linear or nonlinear - with any degree of accuracy[6]. We also tackle the poor generalization ability of the EWS models by using cross validation when choosing the amount of complexity for both models to get the best possible out-of-sample predictions. Cross validation is often used in machine learning, but seldom in economics. The area under the receiver operating characteristic (AUROC) curve is used as the criteria to the validation process and the assessment of prediction results.

The paper is organized as follows: In section 2 we formulate the logit regression and the artificial neural network. The model selection process is also described in this section. Section 3 introduces the Scandinavian crises and the dataset. The results of the analysis are presented in section 4 and section 5 concludes.

[5]See Lo Duca & Peltonen (2012).

[6]See Bishop (1995).

2 Econometric specifications

2.1 Logit model for classification

We use a simple logit regression as the benchmark model for the comparison with the artificial neural network model. This multivariate parametric statistical method seems to be the most sensible option for the comparison, because of its popularity in the Early warning system literature[7] and its often found superiority in out-of-sample prediction compared with the nonparametric signals approach[8]. Fixed effects logit model was not used, because of the unbalanced nature of the panel dataset and the fact that our analysis of the question were the Scandinavian crises predictable relies heavily on the assumption that the four countries and the banking crises they experienced were as homogeneous as possible.

The logit regression can be used to model the conditional probability of the occurrence of a banking crisis. This is done by modelling the crisis probability p with linear combinations of the explanatory variables X, which are transformed with a link function to be in the interval $[0, 1]$ and to sum up to 1. In logit regression the link function is a logistic function. The outcome y is a binary variable, which is equal to 1 for banking crises and 0 otherwise.

$$y = \begin{cases} 1 & with\, probability \quad p \\ 0 & with\, probability \quad 1-p \end{cases} \tag{1}$$

One important thing when moving from simple classification of crises and tranquil periods to using these models as early warning systems is the choice of the dependent variable. We would like to have the warning signals rather early than in the same month as the banking crises is predicted to happen. Usually this time window is chosen[9] to be between 12 and 24 months. We follow here Berg & Pattillo (1999) and Kaminsky et al. (1998), who use 24 months as the time horizon for the EWS to signal a crises. In this case the dependent variable y is a pre-crisis dummy variable that gets the value one for the 24 months preceding the crisis beginning and zero otherwise. We also calculated the results for 12 and 18 month horizons to check the robustness of the results.

In the current classification problem the conditional probability of banking crisis for observation i is given by

$$p_i = Pr[y_i = 1|X_i] = F(X_i'\beta) \tag{2}$$

$$F(v) = \frac{e^v}{1+e^v} \tag{3}$$

[7]See Berg & Pattillo (1999) and Demirguc-Kunt & Detragiache (1998).
[8]See Davis & Karim (2008).
[9]Usually an Ad Hoc procedure.

The parameters of the logit model are estimated with maximum likelihood. The binary response is Bernoulli distributed with a probability mass function

$$f(y_i|x_i) = p_i^{y_i}(1-p_i)^{1-y_i} \tag{4}$$

From (4) we can derive the log-likelihood function for N independent observations

$$L_N(\beta) = \sum_{i=1}^{N}[y_i lnF(X_i'\beta) + (1-y_i)ln(1-F(X_i'\beta)] \tag{5}$$

The maximum likelihood estimator can be derived from the likelihood function by differentiating w.r.t β and setting the equation equal to zero. From this equation we can solve numerically a consistent estimator $\widehat{\beta}_{ml}$, when the conditional probability p is correctly specified.

Finally the classification is done by setting an arbitrary threshold c for the conditional banking crisis probability to classify an observation to a crisis or a normal period e.g. the model signals a banking crises, if the conditional probability given by the logit regression is over 0.5. Usually the threshold is set to 0.5 or the frequency of crises in the dataset (Fioramanti, 2008). Some authors (Sarlin, 2012) have used a policy makers loss function to estimate the optimal threshold given his or her preferences concerning type 1 and type 2 errors. This isn't a good way to assess the classification performance of different models, because the arbitrary choice of the threshold can be seen as an ad hoc solution to get a better fit for the model.

We take a more general approach in assessing the prediction ability of different models. The Receiver operating characteristic (ROC) -curve[10] plots the correctly called crises periods (true positive rate,TP) and incorrectly called normal periods (false negative rate,FN) of the model for all possible values of the probability threshold c. The ROC-curve displays the trade off between type 1 error and type 2 error when the threshold c is changed. The area under the ROC-curve (AUROC) can be used to compare the classification ability of different models without the need to set an arbitrary threshold for p. Under certain conditions[11] AUROC can be be interpreted as the probability, that a randomly chosen banking crisis period is ranked higher by the model than a randomly chosen normal period of the economy. More formally

$$TP(c) = Pr[X_i'\beta \geq c|y_i = 1] \tag{6}$$

$$FP(c) = Pr[X_i'\beta \geq c|y_i = 0] \tag{7}$$

[10]For more information on the ROC-curve see Berge & Jorda (2011).

[11]See Hsieh & Turnbull (1996).

$$AUROC = \int_0^1 [TP(c) * FN(c)]d[FN(c)] \tag{8}$$

A perfect classifier would have an $AUROC$ of 1. A model that has no predictive information at all and is equal to a random guess has an $AUROC$ equal to 0.5. If the models AUROC is under 0.5, then we would just use the opposite prediction that the model is giving to get an AUROC over 0.5. If model A has a higher ROC-curve than model B for every threshold c

$$ROC_A(c) > ROC_B(c) \quad \forall c \in [0, 1], \tag{9}$$

then model A stochastically dominates model B. This gives a more sensible and sufficient way to compare the prediction ability of two different models.

2.2 Artificial neural network

The artificial neural network is a nonparametric nonlinear statistical method, which can be used both in regression and classification problems. It has been used in numerous fields and application including economics and finance. The power of this method comes from its ability to learn different relations between variables, if the model is given a sufficient amount of complexity. The motivation of the model came from desire to model the functions of human brain[12].

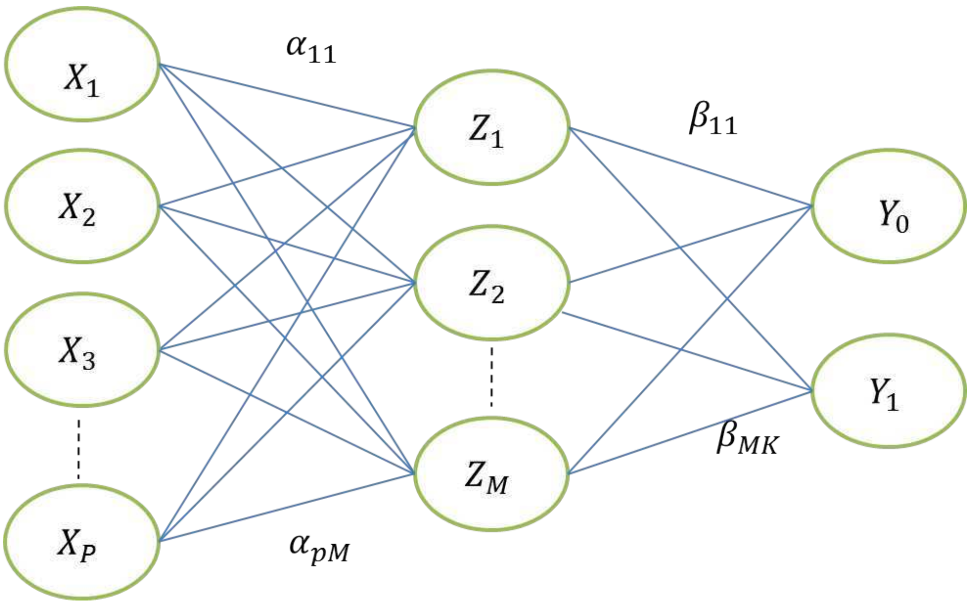


Figure 1: a Feed-forward Multilayer Perceptron with one hidden layer

In this paper we use the most widely used neural network model called feed-forward multilayer perceptron (Figure 1). The model consists of multiple layers of computing

[12] For a comprehensive introduction of neural networks see Rojas (1996).

units called neurons/nodes, which are connected with weights $w = (\alpha, \beta)$. First is the input layer, which is where the information from each input variable X_j is sent to the next layer first multiplied with weights. Next is the hidden layer which holds the hidden neurons Z_m that act as computing units with the weights. The hidden layer also makes a nonlinear transformation to the data with an activation function, which is usually a sigmoid function. The sigmoid function fits the information to be in the interval $[0, 1]$. Finally the output layer gives the output p of the model after the information is again multiplied with weights and put through an output function. In a classification problem there are as many output neurons as there are classes in the dependent output variable. The output function is in this case usually a softmax function which produces positive estimates that sum up to one. The formal definition of a multilayer perceptron with J input variables, M hidden neurons, a binary output variable, a sigmoid activation and output function is

$$Z_{mi} = \sigma(\alpha_{0m} + \sum_{j=1}^{J} \alpha_{jm} X_{ji}) \tag{10}$$

$$p_i = Pr[y_i = 1|X_i] = \sigma(\beta_0 + \sum_{m=1}^{M} \beta_m Z_{mi}) \tag{11}$$

$$\sigma(v) = \frac{1}{1 + \exp^{-v}} \tag{12}$$

The optimal weights in the MLP are solved with backpropagation algorithm. This algorithm uses the derivatives of the output error w.r.t to the weights to adjust the weights at each iteration/training epoch. The algorithm is stopped after certain amount of iterations, when specific criteria is met or if the process has converged. The error function for the classification problem is a cross-entropy.

The neural network architecture choice of one hidden layer can be motivated by Bishop (1995), who said that a classification neural network with one hidden layer and a sufficient amount of hidden neurons can approximate any continuous function with any degree of accuracy. The choice of the backprobagation algorithm was motivated by the fact that it shrinks the weights of irrelevant inputs into zero. To insure that all explanatory variables are treated equally in the network, the data is pre-processed with rescaling so that each variable has a mean of zero and variance of one. The network was also trained with 10 different starting values for the backpropagation algortihm to avoid infeasible local minima. The overfitting was managed with weight decay that adds a penalty term to the error function, which will penalize larger weights. The choice of the weight decay parameter and the number of hidden neurons is discussed in the next section.

2.3 Model selection process

The model selection process is a crucial part in finding a classifier that generalizes well, in other words gives also good out-of-sample predictions. In this paper we are using cross validation that is often used in machine learning literature, but not so much in the early warning system literature. Usually the dataset is split into three different sets. The training set is used to fit the model e.g. estimation of parameters in the logit model or weights in the ANN model. After this a validation set is used to get an prediction error from a different dataset that was used to fit the model. In this step the level of complexity of the model is determined so that it minimizes the error given from the validation set. The complexity would mean the number of hidden neurons and hidden layers in the case of neural networks and the subset of predictors in the logit model. Finally the test set is used to get an estimate of the true error of the model. The test set is not used in the model selection in anyway. The error given from the validation set can not be used to estimate the true error of the final model, because it is used in the model selection process and is therefore biased downwards compared with the true error rate of the model.

Rather than using a single validation set in the second step, we use k-fold cross validation to estimate the expected prediction error of the model. In this method the remaining dataset after removing the test set is sliced into k equally sized folds. After this the model is fit with $k-1$ folds and tested with the remaining fold. This is repeated k times so that each fold is used as the test set once. Now the average of these k prediction errors is an estimate of the expected prediction error of the model[13]. Instead of minimizing an error measure such as cross-entropy or SSE in the cross-validation phase, we use the $AUROC$ as a measure of the classification/prediction ability of different model architectures and choose the one that maximizes this criteria. We use 10-fold cross validation, because of its popularity in the literature. In the logit model a stepwise selection process is implemented with $AUROC$ as the criteria to get the best subset of predictors. Two schemes of explanatory variables is used for the ANN models. One where all the original indicator candidates are used to fit the neural network and one with the same indicators as in the best logit model. In the neural network model the number of hidden neurons and the weight decay parameter are chosen in the cross validation phase[14]. After the best model specification is found, all 10 folds are used to fit the final model and to get the training/in-sample AUROC. When the best models for both logit and ANN have been chosen, the comparison of these methods is done with the test set. The performance on the validation phase is also compared, because it also gives some kind of measure of how well the model has learned the data.

[13]For more information on model selection, model assessment and cross validation see Hastie, Tibshirani & Friedman (2009).

[14]See figure 2. The number of possible hidden neurons ranges from 1 to 30. The weight decay parameter can have values of 0, 0.1 and 0.3. Many other values was also implemented for the weight decay parameter, but they were found to have consistently poorer results.

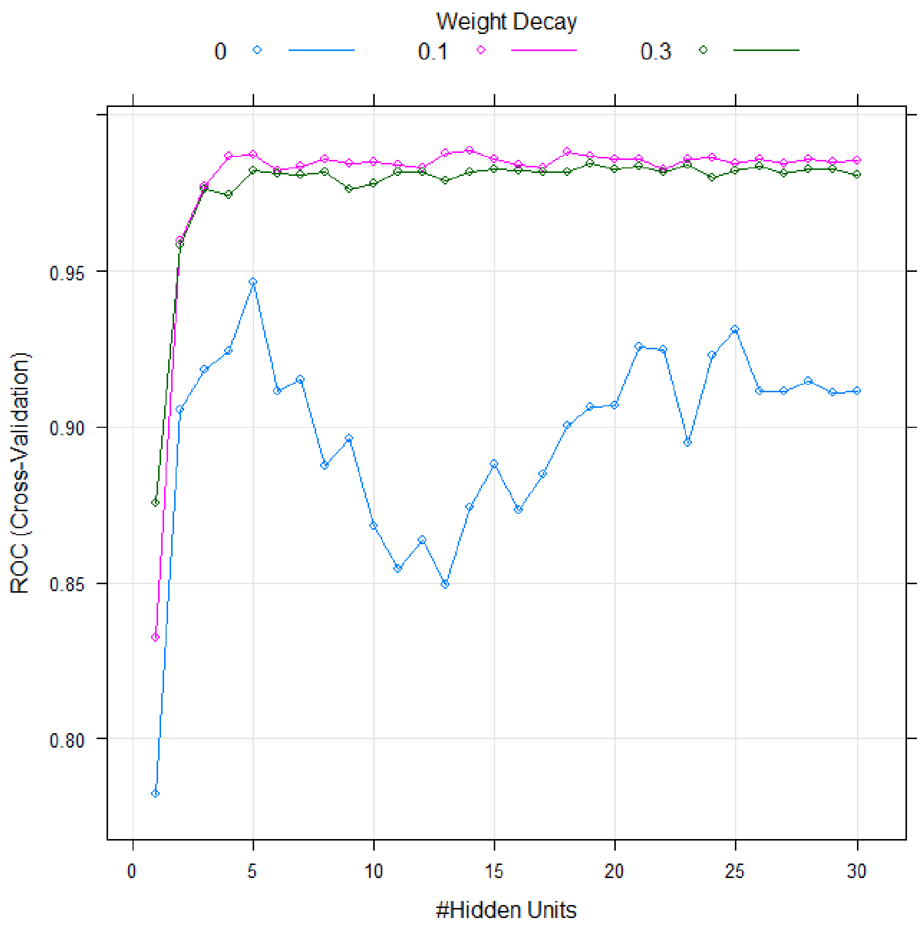


Figure 2: Cross validation AUROC of the ANN2 model for the classification of Scandinavian banking crises when the number of neurons and the weight decay parameter is varied.

3 Data

In this paper we use macroeconomic data from Kaminsky (2006) database, that was originally constructed from various data sources including IFS, WEO, WDI, BIS and different central banks. The data is at monthly frequency from a timespan of 1970M1-2003M6. We use data from four Scandinavian countries, which all experienced a systematic banking crises in the late 1980s or early 1990s. The use of only these four countries was motivated by Davis et al. (2011), who state that when constructing an early warning system for banking crises the model should be estimated for each region separately. They found that different variables were signalling the crises in different regions and that a model with combined regions performed poorly.

The binary dependent banking crisis variable was constructed by taking the starting months of the crises from Kaminsky (2006) crisis chronology and the ending years from Demirguc-Kunt & Detragiache (1998). The pre-crisis dependent variable is constructed given the same information e.g. binary dependent variable for 24 months pre-crisis gets

Table 1: Description of the banking crisis

Country	Begin date	Description[15]	D&D end year	Bailout cost[16] % GDP
Denmark	March, 1987	Two small banks collapsed, which caused a decline in bank lending; two thirds of the problem banks were merged.	1992[17]	-
Norway	November, 1988	Two regional saving banks fail and they are eventually merged.	1993	3.3
Finland	September, 1991	Skopbank collapses on September 19. Government intervenes and takes control of 3 banks accounting for 31% of total savings bank system deposits.	1994	8.4
Sweden	November, 1991	The government rescues Nordbanken, which is the second largest bank in Sweden. Total 5 out of 6 largest banks experienced difficulties.	1993	4.0

[15]Information taken from Kaminsky (2006) crises chronology.

[16]Frydl (1999).

[17]Taken from Reinhart & Rogoff (2009).

the value one on the preceeding 24 months from the starting month of the banking crises and zero otherwise. There are 1478 observations in the final dataset, which of 13.4 % are crisis episodes and 86.6 % normal periods of the economy. The beginning of a banking crisis is determined by an event. The information of these events was taken from financial newspapers and articles in economic journals. There are two types of events that are accounted as banking crises by Kaminsky (2006): (1) Bank runs that lead to the closure, merging, or takeover by the public sector of one or more financial institutions and (2) If there are no runs, the closure, merging, takeover, or large-scale government assistance of an important financial institution (or group of institutions), that marks the start of a string of similar outcomes for other financial institutions. All four Scandinavian crises are labelled by the second definition of a banking crises.

The main facts of these crises can be seen in table 1. The background and causes of the four crises can be described by "bad luck, bad policies and bad banking" according to Koskenkyla & Pensala (1992). Bad luck refers to the recession that affected the exports of the Nordic countries. Especially the collapse of eastern exports for Finland and the drop of oil prices for Norway. Bad policies attributes to the financial liberalization and the insufficient monetary policies of that time. The bad banking surfaced as careless lending and investments. Credit scoring was minimal and bank personnel were not educated enough. Overall the attitude towards risk was indifferent in general.

The choice of explanatory variables was mainly driven by data availability and the seminal paper of Demirguc-Kunt & Detragiache (1998). Some variables were also chosen

Table 2: Mean comparison of the indicators between normal and crises periods

Indicator	Normal period			Banking crises period				
	Obs	Mean	std	Obs	Mean	std	t	p-value
GDP Growth %	1280	0.09	0.28	198	0.08	0.18	0.866	0.380
Inflation %	1280	0.53	0.62	198	0.24	0.40	8.954	0.000
Germany interest rate % per year	1280	6.04	2.75	198	7.72	1.84	-11.075	0.000
U.S Iinterest rate % per year	1280	7.53	3.31	198	5.48	2.31	10.891	0.000
Stock price index	1280	88.04	150.36	198	76.58	32.14	2.395	0.017
Stock price index growth %	1280	1.08	5.95	198	1.50	6.29	-0.886	0.377
Stock price index deviation from trend	1280	0.52	40.41	198	-0.34	12.11	0.607	0.544
M2/Reserves	1280	58.09	30.63	198	41.60	13.90	12.617	0.000
Domestic credit/GDP	1280	0.90	1.159	198	1.41	1.70	-4.096	0.000
Domestic credit/GDP growth %	1280	0.79	4.28	198	-0.08	2.79	3.764	0.000
Exchange rate deviation from trend	1280	0.00	0.31	198	0.02	0.51	-0.533	0.594
Exchange rate change %	1280	0.11	2.77	198	0.33	0.04	-0.725	0.469
Growth of exports %	1280	1.80	13.99	198	1.11	10.83	0.794	0.428
Growth of terms of trade %	1280	0.03	1.67	198	-0.01	0.02	0.260	0.795
Domestic credit growth %	1280	1.01	4.29	198	0.01	2.81	4.298	0.000

from Reinhart & Rogoff (2009). 15 indicator candidates was chosen from the database and they can be labelled into macroeconomic and financial variables. Some of these variables were taken from the original database as they were, but mainly deviations from trend and growth variables were constructed by ourselves. The theory behind majority of the chosen indicators is introduced in Demirguc-Kunt & Detragiache (1998) and Reinhart & Rogoff (2009). For example real GDP growth and terms of trade affect the banks share of nonperforming loans. Domestic credit growth and domestic credit to GDP can be seen as proxies for financial liberalization. Inflation can be associated with macroeconomic mismanagement.

4 Results

4.1 Classification ability comparison

The classification ability of the two methods is compared with the case where all four countries are pooled together into a single dataset. Then 25% of the observations are randomly removed to be in the independent test set. The remaining 75% of the data is used to estimate the models and to choose the optimal amount of complexity with 10-fold cross validation. The random selection of the observations to the test set is done in a way that the same share of crises periods is in both sets. This allows a fair comparison of the methods, because even if the countries implement some heterogeneity the models have the chance to learn the relations between the variables from each country. This case can be used to compare the performance of the two methods, but it can not really answer the

Table 3: AUROC of different models for banking crises classification

	In-sample	Validation	Out-of-Sample	Validation difference with logit	p-value
Logit	0.886	0.879	0.861		
ANN1	1.000	0.995	0.985	0.116	0.000
ANN2	1.000	0.989	0.985	0.110	0.000

question were the Scandinavian crises predictable. When this kind of training scheme is used, the homogeneity assumption is not so crucial.

The results for the case, where the dependent variable[18] is the binary banking crises variable can be seen in table 3. In this case all observations are included. We can see from the results that none of these models performed poorly in this classification task, but the superiority of the artificial neural network is quite clear. Both ANN variable specifications[19] fit the data perfectly with an AUROC of 1. This result would usually mean some overfitting and bad out-of-sample results, but the validation results are very similar - near one. The difference of artificial neural networks and final logit models[20] AUROC for each of the 10 cross validation folds can be seen in figure 3. Artificial neural network is better in all folds and the difference is on average 11%-units. This result is also statistically significant. The results for the other ANN are slightly better, but almost the same. This is in line with the fact that the backpropagation algorithm tends to shrink the weights of insignificant variables to zero. Finally the test results give almost the same results as the validation phase for all three models. The artificial neural networks has about 10%-units better AUROC than the final logit model. This same result can be seen in figure 4, where the artificial neural network has second order stochastic dominance (SOSD) over the logit model. For specificity over 20% the stochastic dominance is of first order (FOSD). As an example of an individual threshold prediction result, the sensitivity and specificity for the best threshold c for both models is also shown in the figures. Only 2 % of the banking crises periods is not classified correctly by the ANN compared with logit models 24.5 %. The difference in classifying the normal periods of the economy is about 7.8%-units for the benefit of ANN. The artificial neural network seems to be significantly

[18]We also calculated the results for different dependent variables and for a dataset where after crises observations were excluded to account for possible endogeneity problems. These results can be seen in table 6. In all cases both ANN models had significantly better results for training, validation and test sets. When after crises observations are excluded, the difference between ANN and logit AUROC is smaller, but still significant.

[19]ANN1 - same explanatory variables as in the final logit model, ANN2 - all 15 original explanantory variables.

[20]The indicators in the final logit model can be seen in table 7. High inflation, U.S real interest rate, stock price index, M2/reserves and domestic credit growth seem to raise the probability of a banking crises in a Scandinavian country in this time period. The decline of Germanys interest rate, ratio of domestic gredith to GDP and its growth, exports and the deviations of stock price index and exchange rate from HP-filtered trend also raise the probability of the crises.

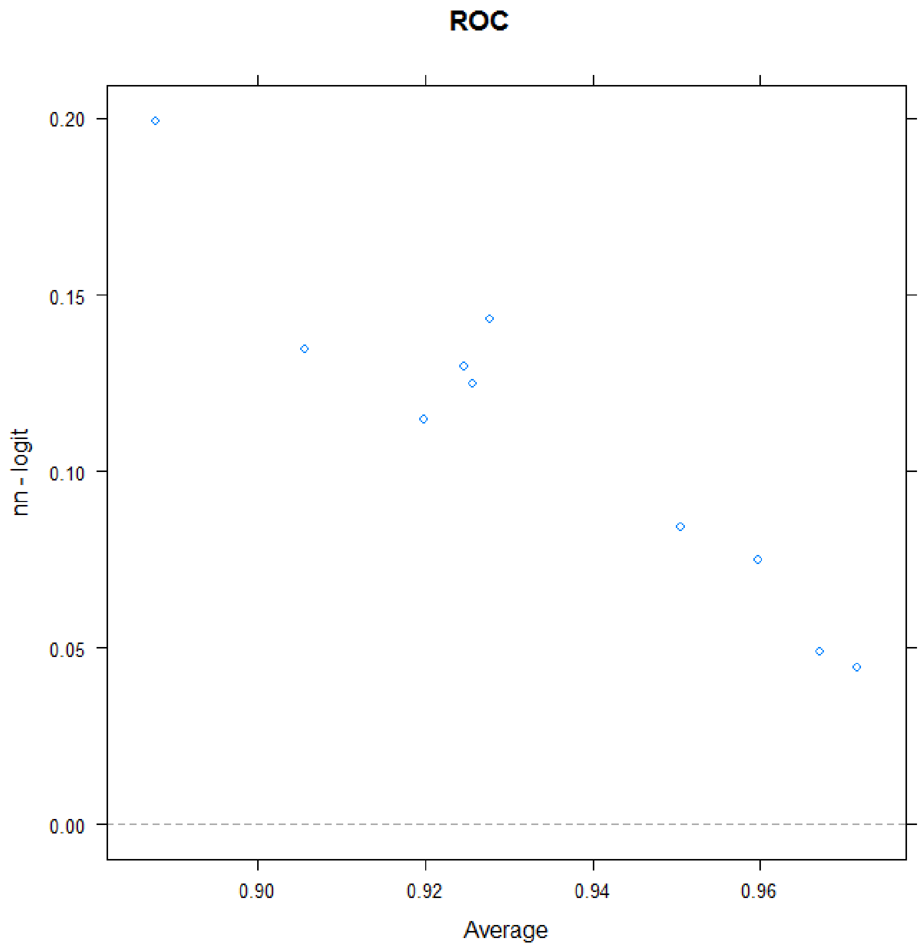
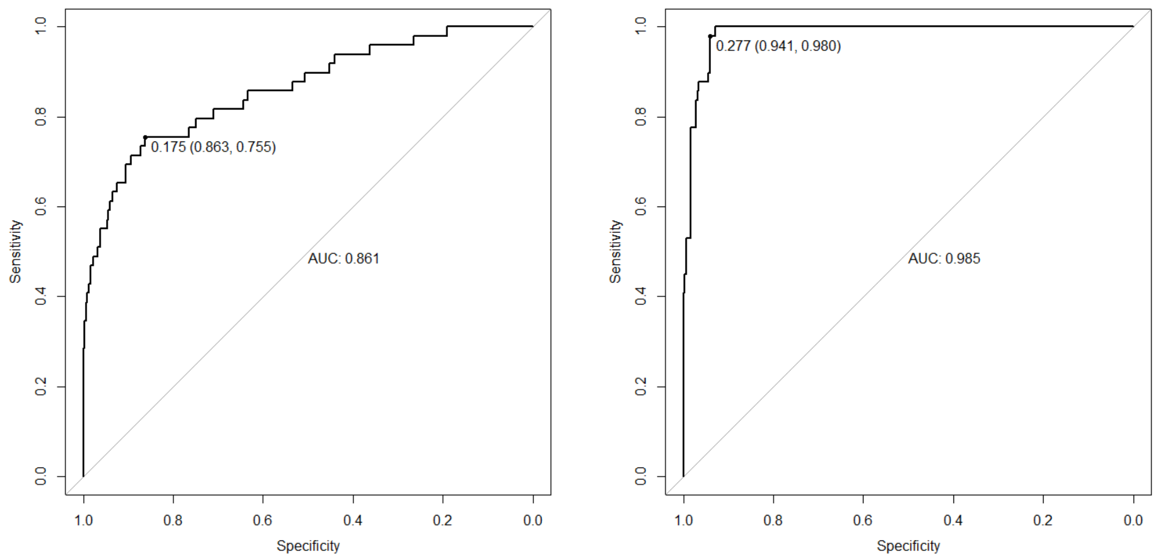


Figure 3: Difference of ANN2 and Logit AUROC of 10 cross validation resamples for banking crisis classification problem.

better in classification of the Scandinavian banking crises periods and normal periods of the economy compared with the logit model. This would strongly imply that the decision boundary of this banking crises classification problem is nonlinear. The artificial neural network was given the opportunity to choose the number of hidden neurons to be from one to 30 in the cross validation phase. ANN with no hidden layer would be equivalent to a logit model, which can solve only lineary separable classification problems. The nonlinearity can be argued by the fact that the model performed better in every dataset compared with the logit model.

These results do not imply that the perfect prediction of all four Scandinavian countries was possible before they actually occurred in the late 1980s and early 1990s. Firstly we use data from actual crises that we are predicting and the dataset has observations even after the actual crises. This kind of information was not obviously available at that time. Secondly the dependent variable here is a binary banking crisis dummy, that would - even with a perfect classifier - give the signal of a crises in the same month as it is happening. The crises signal should be given as soon as possible so there would be time to take some precautionary measures to prevent it. Thirdly the use of after crises obser-



(a) LOGIT (b) ANN2

Figure 4: ROC-curves for the test set of banking crises classification problem.

vations could raise some endogeneity problems, because the banking crises would affect the explanatory variables that are used to predict the actual crises. In the light of these facts, this current set up of the classification models is not suitable for building an early warning system. These results simply tell us that the artificial neural network can learn to classify the crises periods and the normal periods of the economy nearly perfectly via nonlinear functions of the explanatory variables, if the sufficient amount of information is given to the model. The logit models linear combination of the explanatory variables does significantly worse and does not achieve perfect classification. The results for the early warning system models, where these issues are solved are presented in the next subsection.

4.2 EWS out-of-sample prediction of crises

To build a proper early warning system we set the dependent variable to be equal to one in 24 months preceding the banking crises. This way we can classify the pre-crisis periods and can possibly predict the banking crises quite precisely depending on the classification ability of the model. This relies on the assumption that the explanatory variables act differently in pre-crises periods and normal periods of the economy. A dataset where after crises observations are excluded is used to remove the endogeneity problem. To be able to refer to pure out-of-sample prediction of the Scandinavian banking crises, we use the Norwegian and Danish banking crises to estimate and validate the models and

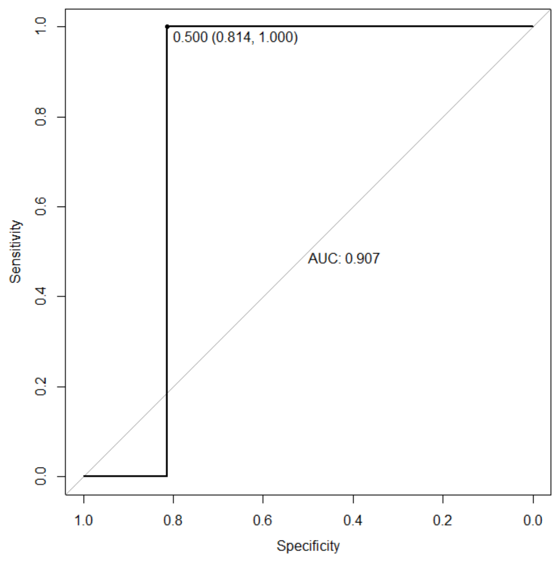
Table 4: AUROC for out-of-sample prediction of Finnish and Swedish crises when after crises obs are excluded

	Sweden								
	PRE12			PRE18			PRE24		
	Train	Validate	Test	Train	Validate	Test	Train	Validate	Test
Logit	1.000	0.952	0.934	1.000	0.983	0.926	1.000	0.969	0.907
ANN1	1.000	1.000	**0.993**	1.000	1.000	0.995	1.000	1.000	<u>***1.000***</u>
ANN2	1.000	1.000	**0.993**	1.000	1.000	**0.997**	1.000	1.000	0.993
	Finland								
	PRE12			PRE18			PRE24		
	Train	Validate	Test	Train	Validate	Test	Train	Validate	Test
Logit	1.000	0.952	0.606	1.000	0.983	0.611	1.000	0.969	0.611
ANN1	1.000	1.000	0.751	1.000	1.000	0.745	1.000	1.000	0.793
ANN2	1.000	1.000	**0.755**	1.000	1.000	**0.753**	1.000	1.000	<u>***0.967***</u>

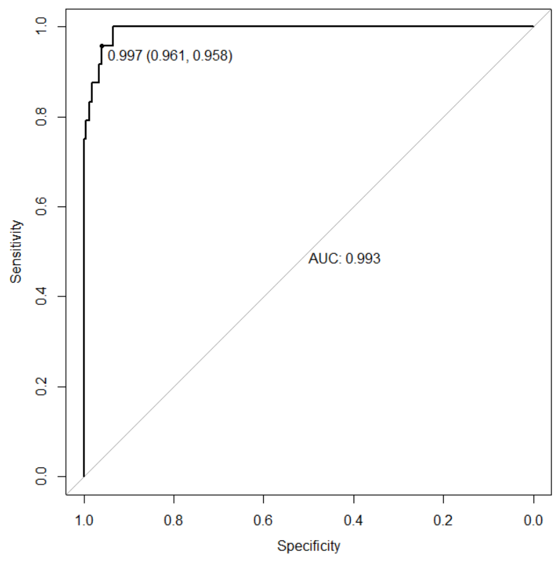
test them with the Finnish and Swedish Crises[21]. This way we can say, if these crises were predictable before they occurred given the information of that time. This approach relies more than the first case on the assumption that the crises were as homogeneous as possible in the four countries.

The results in table 4 indicate that the answer might be yes, atleast for the Swedish crises. Both logit and ANN give pretty promising results. This time the ANN with the same variables as in the logit model chose only one hidden neuron in the cross validation phase. The logit model gave predicted probabilities of only zero and one. This is because the maximum likelihood estimation found a perfect linear separation with a linear combination of the explanatory variables in the training data. In other words one or more variables have a threshold t, which can divide the pre-crises and the normal periods. E.g. all pre-crises observations have export growth under 2% and all normal periods above it. This will put $X_i\beta$ to ∞ or $-\infty$, so that p_i is one or zero. This situation is called Hauck-Donner phenomenon and there is no procedure to prevent this. The maximum likelihood estimation is just doing what it is told to do. Simply this problem just implies that logit regression is not possibly needed for the classification problem, but a simple threshold rule for a linear function would be enough. The artificial neural network with the same variables as in the final logit does not fit the predicted probabilities to zero and one, which might be due to the weight decay parameter that penalizes larger weights and prevents overfitting. This could be the solution of preventing the H-D phenomenon in general.

[21]This approach was also used by Frank & Schmied (2003) in predicting currency crises in Brazil and Russia.



(a) LOGIT



(b) ANN2

Figure 5: ROC-curves for out-of-sample prediction of Swedish crises

From the results we can see that the ANN with one hidden neuron and weight decay parameter of 0.1, achieves perfect classification in every dataset, even with the test set of Sweden. The model predicts all of the 24 pre-crises periods correctly out-of-sample and gives no false alarms. The basic logit that suffers from H-D phenomenon does not predict bad with an AUROC of 0.907. However the ANN with 24 hidden neurons and all original 15 indicators predicts the Swedish crises almost perfectly with an AUROC of 0.993. The ROC-curves can be seen in figure 5. The ANN again stochastically dominates the logit model, when specificity is over 80%. The ANN probability for pre-crises period in figure 6 seems to spike two times before the first true pre-crises period, but then gradually rises from 1985 until it reaches one when the pre-crises period begins and remains there until the banking crises occurs. This kind of early warning systems are very promising, because it gives no false alarms and gives consistent signals when the crises is getting closer.

The same models do not predict the Finnish banking crises so well. Logit model has close to no predictive information at all with an AUROC of 0.611. The logit model would have been useless as an EWS for the Finnish crises as it signalled a crises consistently 14 years too early[22]. However the ANN with all original variables does almost as well as in the Swedish crises prediction with an AUROC of 0.967. The difference of the prediction ability of the model is quite dramatic and can be seen in figure 7. The ANN first order stochastically dominates the logit model from specificity of 23.3% to 93.4%.

[22]See figure 9b.

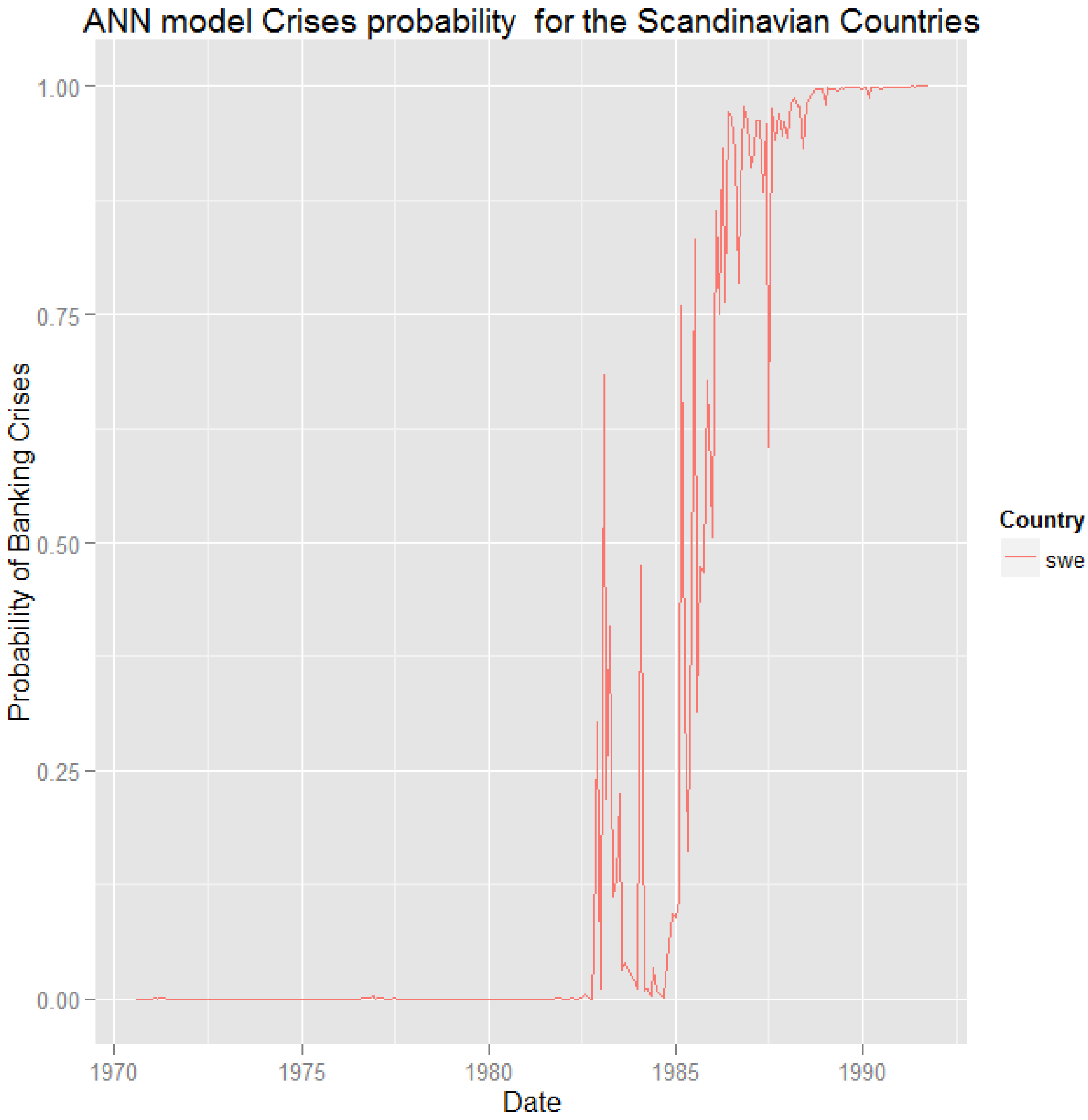
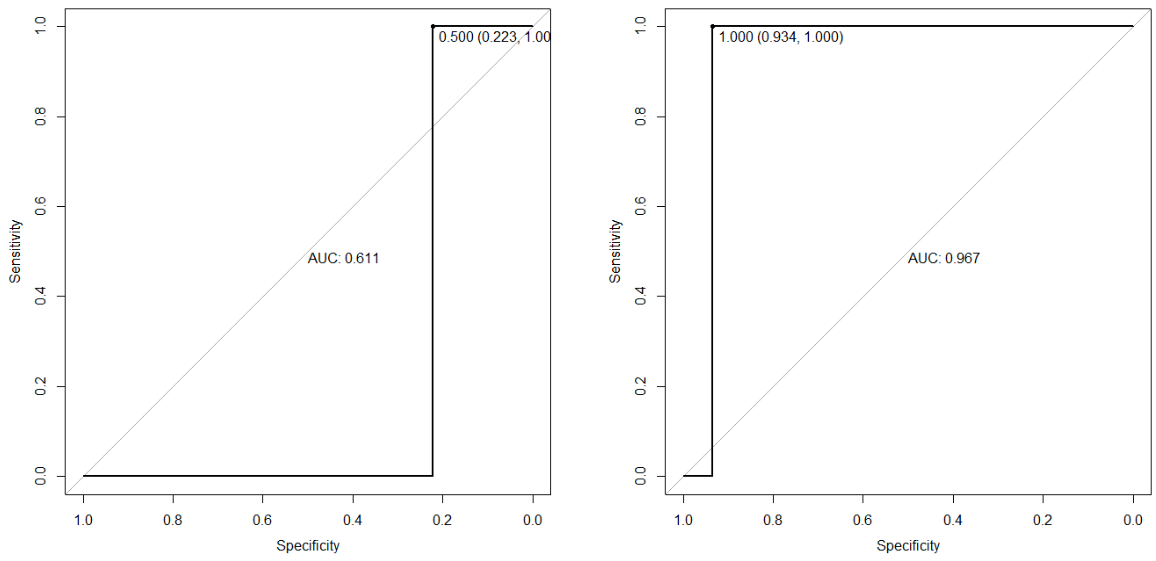


Figure 6: Banking crises (PRE24) probability for Sweden given by ANN2 out-of-sample



(a) LOGIT

(b) ANN with all

Figure 7: ROC-curves for out-of-sample prediction of Finnish crises

5 Conclusions

This paper has filled a cap in the early warning system literature by constructing an artificial neural network based model to predict banking crises in Scandinavia. The results suggest that ANN based models are highly promising tools for banking crises prediction and they clearly outperform the usual models like the logit regression. This furthermore suggests that the decision boundary that classifies the pre-crisis periods and the normal periods in the crisis indicator space is nonlinear.

The advantage of estimating early warning models for as homogeneous region as possible such as the Scandinavian countries and using cross validation to tackle the poor generalization ability usually disturbing these kind of models allows us to predict the Swedish crises perfectly with 24 months of constant signals and no false alarms before the crises occurred in November of 1991. The banking crises of Finland in September 1991 was also predicted[23] by the artificial neural network based early warning system, but it gave also 6.6 % of false alarms. It can be stated in the light of the results of this paper that the Finnish and Swedish banking crises were predictable out-of-sample given the information from the Danish and Norwegian crises, if an ANN based EWS would have been used at that time.

The usual way of accessing and comparing the prediction ability of EWS models in the literature with a single arbitrary threshold for the crises probability to signal a crises was replaced with the area under the ROC-curve that gives a more general and sufficient way to state the superiority of one model over another. The area under this curve states

[23]crises signalled in all 24 months prior to the banking crises beginning.

the probability of a randomly chosen banking crises period to be ranked higher by the model than a randomly chosen normal period of the economy. This statistic was 99.3% and 96.7% for the artificial neural networks out-of-sample prediction for the Swedish and Finnish data. Although there are several problems to be solved in constructing a comprehensive and reliable early warning system for banking crises such as the decision of the time horizon for crises signalling and the arbitrary choice of the crises signal threshold, this paper shows some significant gains from utilizing artificial neural networks in this prediction task.

References

Berg, A. — Pattillo, C. (1999) Predicting currency crises: The indicators approach and an alternative. *Journal of International Money and Finance*, vol. 18, 561–586.

Berge, T. J. — Jorda, O. (2011) Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics, American Economic Association*, vol. 3 (2), 246–277.

Bishop, C. (1995) *Neural Network for Pattern Recognition*. Oxford University Press.

Caprio, C. — Klingebiel, D. (1996) Bank insolvencies:cross-country experience. *World Bank Policy Recearch Working Paper*, (1620).

Davis, E. P. — Karim, D. (2008) Comparing early warning systems for banking crises. *Journal of Financial Stability*, vol. 4 (2), 89–120.

Davis, P. E., Karim, D. — Liandze, I. (2011) Should multivariate early warning systems for banking crises pool across regions? *Review of World Economics*, vol. 147 (4), 693–716.

Demirguc-Kunt, A. — Detragiache, E. (1998) The determinants of banking crises in developing and developed countries. *IMF Staff Papers*, vol. 45 (1), 81–109.

Fioramanti, M. (2008) Predicting sovereign debt crises with artficial neural network: a comparative approach. *Journal of Financial Stability*, vol. 4 (2), 149–164.

Frank, R. — Schmied, A. (2003) Predicting currency crises contagion from east asia to russia and brazil: an artificial neural network approach. *AMCB Working Paper*, (2).

Frydl, E. J. (1999) The length and cost of banking crises. *IMF Working Paper*, (30).

Hastie, T., Tibshirani, R. — Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hoggart, G. — Saporta, V. (2001) Costs of banking system instability: some empirical evidence. *Financial Stability Review*.

Hsieh, F. — Turnbull, B. W. (1996) Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Annals of Statistics*, vol. 24 (1), 25–40.

Kaminsky, G. (2006) Currency crises: Are they all the same? *Journal of International Money and Finance*, vol. 25, 503–527.

Kaminsky, G., Lizondo, S. — Reinhart, C. M. (1998) Leading indicators of currency crises. *IMF Staff Paper*, vol. 45, 1–48.

Kaminsky, G. — Reinhart, C. M. (1999) The twin crises: the causes of banking and balance of payments problems. *American Economic Review*, vol. 89 (3), 473–500.

Koskenkyla, H. — Pensala, J. (1992) Pohjoismaiden pankkikriisi ja viranomaisten tukitoimenpiteet. *The Finnish Economic Journal*, vol. 88 (4), 431–447.

Lo Duca, M. — Peltonen, T. (2012) Assessing systemic risk ans predictiong systemic events. *Journal of Bank Finance(forthcoming)*.

Manasse, P., Roubini, N. — Schimmelpfennig, A. (2003) Predicting sovereign debt crises. *IMF Working Paper*, (221).

Nag, A. — Mitra, A. (1999) Neural nnetwork and early warning indicators of currency crisis. *Reserve Bank India Occas Pap*, vol. 20 (2), 183–222.

Reinhart, C. M. — Rogoff, K. (2008b) Is the 2007 u.s subprime crisis so different? an international historical comparison. *American Economic Review*, vol. 98 (2), 339–344.

Reinhart, C. M. — Rogoff, K. (2009) *This time is different: eight centuries of financial folly*. Princeton University Press.

Rojas, R. (1996) *Neural Networks-A systematic Introduction*. Springer-Verlag.

Sarlin, P. (2012) On biologically inspired predictions of the global financial crisis. *Neural Computing and Applications*, (24), 663–673.

A Tables and Figures

Table 5: Descriptive statistics for the indicators candidates

Variable	Obs	Mean	std	Min	Max
GDP Growth%	1478	0.09	0.27	-0.63	0.56
Inflation%	1478	0.49	0.60	-2.55	4.34
Germany interest rate% pre year	1478	6.26	2.71	2.51	14.57
U.S interest rate% per year	1478	7.26	3.27	1.24	19.1
Stock price index	1478	86.51	140.47	3.47	1246.57
Stock price index growth%	1478	1.13	6.00	-34.15	29.87
Stock price index deviation from trend	1478	0.40	37.87	-243.74	482.06
M2/Reserves	1478	55.88	29.49	16.75	223.48
Domestic credit/GDP	1478	0.967	1.257	0.093	5.221
Domestic credit/GDP growth	1478	0.67	4.12	-10.12	123.21
Exchange rate deviation from trend	1478	0.004	0.346	-1.45	1.96
Exchange rate change%	1478	0.14	3.00	-10.46	18.69
Growth of exports%	1478	1.70	13.60	-50.01	85.82
Growth of terms of trade%	1478	0.02	1.76	-11.37	18.19
Domestic credit growth%	1478	0.88	4.13	-10.00	123.76

Table 6: AUROC for classification of different pre-crises configurations when all countries in training set

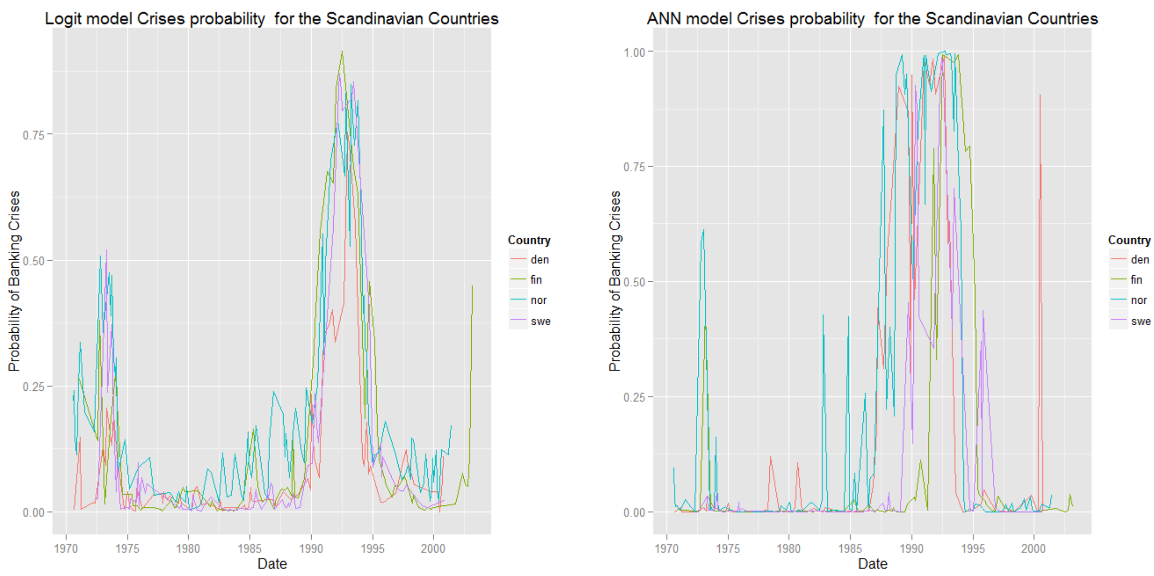
	With all observations								
	PRE12			PRE18			PRE24		
	Train	Validate	Test	Train	Validate	Test	Train	Validate	Test
LOGIT	0.859	0.836	0.748	0.842	0.811	0.782	0.757	0.693	0.723
ANN1	0.995	0.972	0.930	1.000	0.979	0.848	0.990	0.961	**0.924**
ANN2	1.000	0.963	***0.986***	1.000	0.973	**0.980**	0.999	0.974	0.899
	After crises observations excluded								
	PRE12			PRE18			PRE24		
	Train	Validate	Test	Train	Validate	Test	Train	Validate	Test
Logit	0.983	0.957	0.912	0.979	0.969	0.911	0.978	0.965	0.966
ANN1	1.000	0.997	0.964	1.000	0.998	***0.996***	0.998	0.996	**0.984**
ANN2	1.000	0.991	**0.977**	1.000	0.996	0.985	0.999	0.994	**0.984**

Table 7: Regression table of best logit model for classification of crises periods when all countries in training set

Indicator	Estimate	Std.Error	z value	Pr(>—z—)
(Intercept)	-0.309	0.539	-0.572	0.567
INF	96.130	24.280	3.960	0.000 ***
GIR	-0.456	0.049	-9.313	0.000 ***
WIR	0.400	0.048	8.394	0.000 ***
SP	0.004	0.002	2.588	0.010 **
SPDEV	-0.010	0.005	-1.778	0.075 .
M2RES	0.040	0.007	6.036	0.000 ***
DCGDP	-1.438e-10	8.418e-11	-1.709	0.088 .
DCGDPG	-107.7	35.05	-3.074	0.002 **
ERUSDEV	-0.854	0.298	-2.862	0.004 **
EXPOG	-1.366	0.844	-1.619	0.105
DCG	114.9	34.73	3.308	0.001 ***
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1				
AIC: 586.06		ROC: 0.879		

Table 8: AUROC for out-of-sample predictions of Finnish and Swedish crises when all observations are used

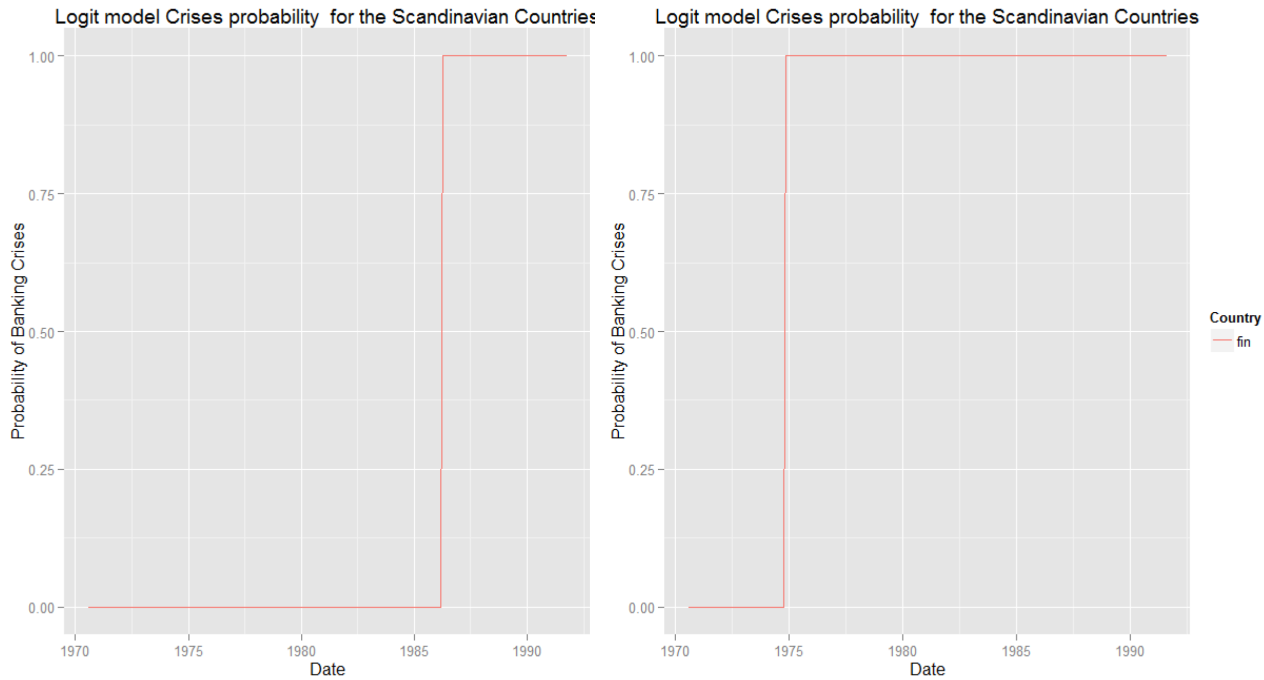
	Finland											
	Crises			PRE12			PRE18			PRE24		
	Train	Validate	Test	Train	Validate	Test	Train	Validate	Test	Train	Validate	Test
Logit	0.959	0.949	0.809	0.997	0.940	0.613	0.976	0.968	0.757	0.959	0.937	**0.789**
ANN1	1.000	0.999	**0.868**	1.000	0.998	***0.915***	0.999	1.000	0.893	0.997	0.994	0.572
ANN2	1.000	0.999	0.728	1.000	0.997	0.876	1.000	0.999	**0.898**	1.000	0.997	0.487
	Sweden											
	Crises			PRE12			PRE18			PRE24		
	Train	Validate	Test	Train	Validate	Test	Train	Validate	Test	Train	Validate	Test
Logit	0.959	0.949	**0.979**	0.997	0.940	0.757	0.976	0.968	**0.665**	0.959	0.937	**0.573**
ANN1	1.000	0.999	0.825	1.000	0.998	0.712	0.999	1.000	0.620	0.997	0.994	0.507
ANN2	1.000	0.999	0.896	1.000	0.997	***0.772***	1.000	0.999	0.599	1.000	0.997	0.508



(a) Crises probability Logit case1

(b) Crises probability ANN1 case1

Figure 8: Probability figures when all countries in training set



(a) Banking crises probability(pre24) for Sweden given by Logit out-of-sample

(b) Banking crises(PRE24) probability for Finland given by logit out-of-sample

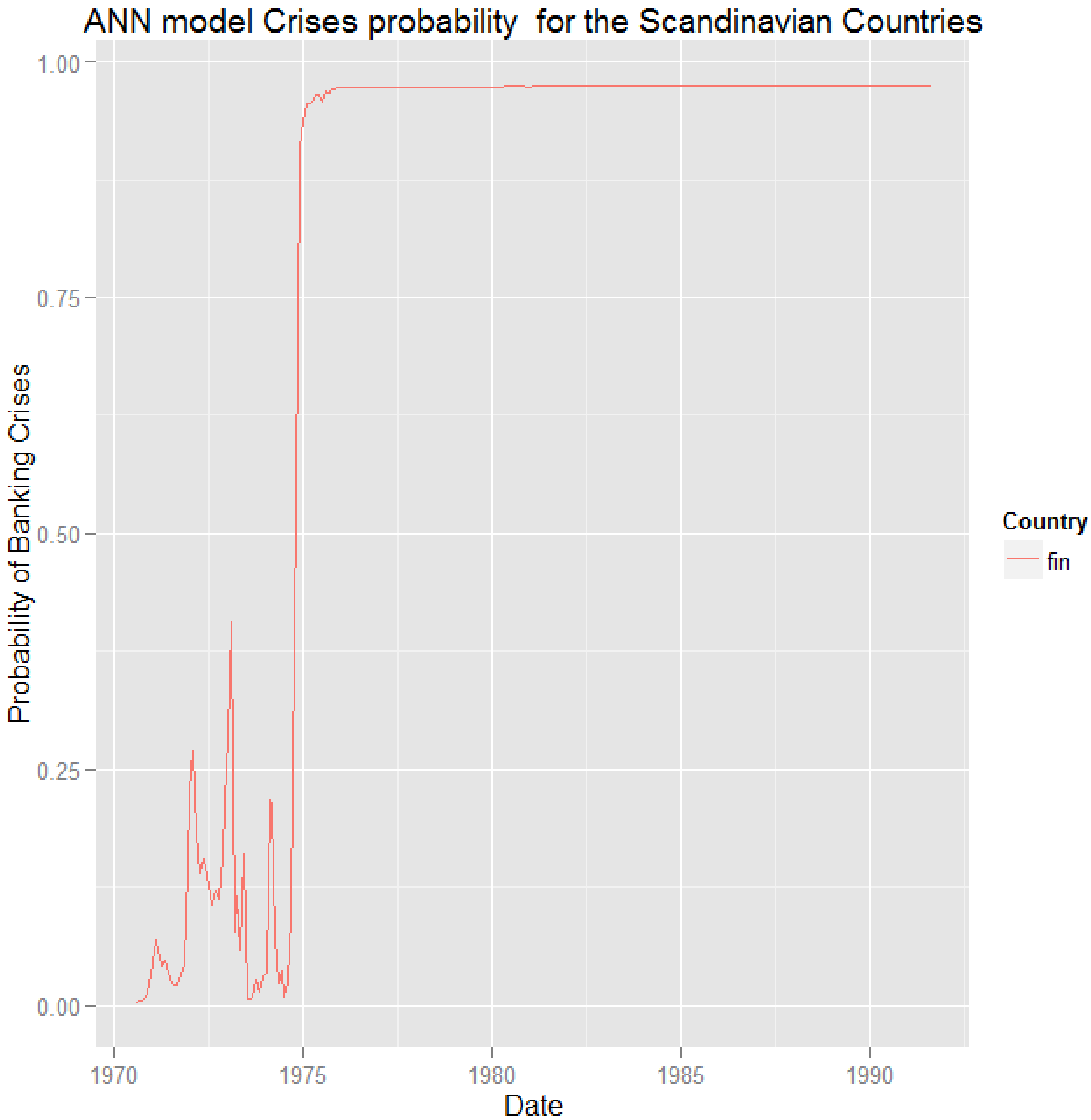


Figure 10: Banking crises(PRE24) probability for Finland given by ANN2 out-of-sample

The **Aboa Centre for Economics (ACE)** is a joint initiative of the economics departments of the Turku School of Economics at the University of Turku and the School of Business and Economics at Åbo Akademi University. ACE was founded in 1998. The aim of the Centre is to coordinate research and education related to economics.

Contact information: Aboa Centre for Economics, Department of Economics, Rehtorinpellonkatu 3, FI-20500 Turku, Finland.

www.ace-economics.fi

ISSN 1796-3133