Predicting Sovereign Credit Risk Using the Artificial Neural Network: an application to Jamaica

R. Brian Langrin†
Financial Stability Department
Research and Economic Programming Division
Bank of Jamaica

This Version: 25 October 2012

The recent deterioration in credit risk across sovereigns, as well as banks and corporates exposed to sovereign risk, has renewed the focus on prediction of sovereign probabilities of default or downgrades, which should be accurately captured in the credit risk models of financial institutions. The aim of this paper is to identify the systemic risk drivers relevant for the 'forward-looking' modeling of the dynamics of Government of Jamaica (GOJ) sovereign credit risk. Importantly, these systemic drivers would also impact the external credit ratings of banks operating in Jamaica as they face the same underlying economic risk factors as the sovereign. The paper uses 3-month lagged values of the CPI inflation rate, US-Jamaica currency exchange rate, the real Treasury bill rate, external debt to exports, net international reserves to imports, real effective exchange rate, terms of trade index, current account of the BOP, real GDP growth and the unemployment rate to predict the GOJ sovereign rating. Sensitivity analysis using the Artificial Neural Network methodology show that external debt to exports, NIR to imports, unemployment rate and the fiscal balance are the most important leading indicators of sovereign rating downgrades.

Keywords: Sovereign Default, Artificial Neural Networks, Macroeconomic Variables

JEL Classification: C45, G20, H63

† R. Brian Langrin, Chief Economist, Financial Stability Dept., Bank of Jamaica, Nethersole Place, P.O. Box 621, Kingston, Jamaica, W.I., Office: +1 (876) 967-1880, Fax: +1 (876) 967-4265, Email: brian.langrin@boj.org.jm. The views expressed in this paper are not necessarily those of the Bank of Jamaica.

1.0 Introduction

Underscored by widespread deterioration in sovereign credit risk across both mature and developing countries following the recent global recession, the realignment of sovereign credit risk weightings based on external credit ratings and internal credit scoring systems has grown as a key area of emphasis in the financial system. Moreover, the rationale of zero risk weight legacy treatment of debt issued in domestic currency by a high risk sovereign has recently been brought into question by the regulatory standard setting bodies such as the Bank of International Settlements (BIS) and the International Monetary Fund (IMF). These institutions argue that domestic sovereign debt holdings by banks, even in the case of highly rated sovereigns such as OECD countries, should now be subject to Basel II-determined application of non-zero risk weights to quantify credit risks.[1] In line with this view, there has been a concerted focus on enhancing the credit risk models of financial institutions in regards to prediction of sovereign probabilities of default or downgrades.

The recent proliferation of internal credit risk models is also largely influenced by the Basel Committee on Banking Supervision's (BCBS, 2006) requirement for banks to use sophisticated credit scoring models for risk-based capital allocation under the internal ratings-based (IRB) approach. Credit scoring models rely on historical data related to borrower ratings for credit risk prediction to automate the assessment of a financial institution's decisions to increase its exposure to a particular borrower as well as to determine specific terms on the exposure as a function of borrower risk. Credit risk is generally measured using four key components, the one-year probability of default per rating grade (PD), the loss given default (LGD), the exposure at default (EAD) and the effective maturity (M).[2] Expected loss for each exposure can be expressed as EL=PD*LGD*EAD. Risk-weight functions may then be used to produce capital requirements for the unexpected loss portion (standard deviation) of the loss distribution.

Regarding the application of risk weights on borrower exposures in the banking book under Basel 2 – Standardized Approach, it should be noted that a zero risk weight is allowed for bank exposures to AAA and AA-rated sovereigns. In addition, national discretion is permitted for the application of a lower or

---

[1] See Speech delivered by Hervé Hannoun, Deputy General Manager, BIS, 'Sovereign risk in bank regulation and supervision: Where do we stand?' (Financial Stability Institute High-Level Meeting Abu Dhabi, UAE, 26 October 2011) and IMF's Global Financial Stability Report (September 2011).

[2] PD is an indication of the unlikeliness of the borrower to pay derived from the internal rating system of a bank, LGD indicates the expected percentage of exposure the bank could lose if the borrower defaults and EAD is the outstanding loan amount plus expected future drawdowns in case the borrower defaults.

zero risk-weight to banks' exposures to their sovereign of incorporation that are denominated in domestic currency and funded in that currency.[3] However, these exceptions do not apply under the IRB Approach wherein a meaningful differentiation of risk is stipulated. For example, a bank's internal risk estimates of PDs and LGDs for corporate, bank and sovereign exposures can be converted into risk weights (RW) and capital charges, where the capital requirement is calculated as 10.0 per cent of the RW multiplied by the EAD (see Table 1).[4] The Basel IRB formula for capital requirement (K) and risk-weighted assets (RWA) is expressed as:

$$K = \left[LGD \times N\left[\frac{1}{\sqrt{1-R}} \times G(PD) + \sqrt{\frac{R}{1-R}} \times G(0.999) \right] - PD \times LGD \right] \times \left(\frac{1+(M-2.5)\times b(PD)}{1-1.5\times b(PD)} \right), \qquad [1]$$

$$R = 0.12 \times \frac{\left(1\text{-}EXP^{(-50\times PD}\right)}{\left(1\text{-}EXP^{-50}\right))} + 0.24 \times \left[\frac{1-\left(1-EXP^{(-50\times PD}\right)}{\left(1\text{-}EXP^{-50}\right))} \right], \qquad [2]$$

$$b(PD) = \left(0.11852 - 0.05478 \times \ln(PD)\right)^2, \qquad [3]$$

where,

R = asset correlation,[5]

N[x] = the cumulative distribution for a standard normal variable,

G[z] = the inverse cumulative distribution for a standard normal variable,

Ln = the natural logarithm,

b(PD) = the slope of the adjustment function

M = effective maturity,

and

RWA = K x 10 x EAD [4]

[3] See BCBS (2006).

[4] Source: BCBS assuming minimum CAR of 10.0 per cent, LGD of 45.0 per cent and maturity of 2.5 years.

[5] The asset correlations are dependent on the type of asset class as different borrowers display different degrees of dependency on the overall economy. The Basel-derived asset correlations of the capital requirements formula for SME and retail asset exposures are different (see BCBS, 2006). Note that these correlations reflect historical loss data from supervisory databases for the G10 countries.

Table 1. Example of Risk Weights and Capital Charges per PD under Basel 2

Probability of Default (%)	*Risk Weight (%)*	*Capital Charges (%)*
0.01	7.53	0.60
0.02	11.32	0.91
0.03	14.44	1.16
0.05	19.65	1.57
0.10	29.65	2.37
0.25	49.47	3.96
0.40	62.72	5.02
0.50	69.61	5.57
0.75	82.78	6.62
1.00	92.32	7.39
1.30	100.95	8.08
1.50	105.59	8.45
2.00	114.86	9.19
2.50	122.16	9.77
3.00	128.44	10.28
4.00	139.58	11.17
5.00	149.86	11.99
6.00	159.61	12.77
10.00	193.09	15.45
15.00	221.54	17.72
20.00	238.23	19.06

Regarding the cyclicality of the risk components, there exists substantial empirical evidence that PD and LGD are influenced by variations through the economic cycle as defaults typically multiply in times of deteriorated macroeconomic conditions (for example, see Fama, 1986, Wilson, 1997, Altman and Brady, 2001). The aim of this paper is to identify the systemic risk drivers relevant for the 'forward-looking' modeling of the dynamics of Government of Jamaica sovereign credit risk. Importantly, these systemic drivers would also impact the external credit ratings of banks operating in Jamaica as they face the same underlying economic risk factors as the sovereign.[6] In addition, this exercise will be useful not only for prediction purposes but it will effectively provide a set of indicators which Jamaica should focus on improving, given the adverse implications for the GOJ financing activities and the knock-on effects on the wider economy from CRA rating downgrades.

Consistent with Cantor and Packer (1996) and Haque et al (1996), this study examines the relationship between GOJ sovereign default risk and a set of key macroeconomic variables. The variables used in this

[6] See Standard & Poors (2011), 'Analytical Linkages Between Sovereign And Bank Ratings,' RatingsDirect on the Global Credit Portal.

paper cover inflation, exchange rate, real effective exchange rate, real Treasury bill rate, unemployment rate, Gross Domestic Product (GDP) growth, ratio of external debt to exports, ratio of net international reserves to imports, terms of trade, fiscal balance and current account balance. Estimation of the relative impact for each of these variables is carried out for the purposes of developing a robust forward-looking financial stability framework for credit risk. In terms of defining a comprehensive credit rating (CCR) measure of GOJ sovereign credit rating, numerical values were assigned to each alphanumeric foreign currency sovereign risk rating assigned by Standard and Poor's. Similar to Gande and Parsley (2010), the numbers range from 0 (Selected Default) to 21 (AAA) to obtain an explicit credit rating (ECR) (see Table 2). Then information on the credit outlook (COL), ranging from -0.5 to +0.5, is added to CCR to attain the CCR, that is, CCR = ECR + COL (see Table 3).

Table 2. Explicit Credit Rating

Sovereign Rating	*ECR*
AAA	21
AA+	20
AA	19
AA-	18
A+	17
A	16
A-	15
BBB+	14
BBB	13
BBB-	12
BB+	11
BB	10
BB-	9
B+	8
B	7
B-	6
CCC+	5
CCC	4
CCC-	3
CC	2
C	1
SD, D	0

Table 3. Credit Outlook

Outlook	*COL*
Positive	0.5
Stable	0
Negative	-0.5

2.0 ANN Motivation

The choice of statistical methodology is a critical decision for credit risk modeling. Pure statistical models have been widely used to estimate credit scoring models. These models are parametric approaches that relate observable borrower attributes to credit quality ratings or default events. Linear discriminant analysis (LDA) and logistic regression statistical techniques have been the usual benchmarks for building credit scoring models.

LDA, pioneered by Altman (1968), was the first method used in building credit scoring models. This technique forms a linear combination of scores from present and historical values of observable attributes for discriminating between defaulters and non-defaulters for a predetermined horizon. Fitting the discriminant function or 'scoring' function to these attributes is also necessary to define cut-off values, which is juxtaposed with the associated scores to separate borrowers according to their group classification. 'Posterior default probabilities' or probabilities of default conditional on the score value are then assigned by transforming the scoring function to a default model using Bayes' theorem. An important drawback of LDA, however, is its unrealistic assumption that the classes are normally distributed with equal covariance matrices which could severely bias the classification results (see Anderson and Rosenfeld, 1988).

Logistic regression (LR) is another common alternative to develop credit scoring models especially when predicting binary default events (see Ohlson, 1980). The LR model uses the cumulative logistic probability distribution to estimate odds ratios for each of the attribute values in the model. The logarithm of the odds ratio or logit produces a linear relationship to predict default events given the set of attributes. The logit model is expressed as:

$$P_i = \frac{1}{1+e^{-Y_i}}, \quad Y_i = \alpha + \beta X_i + \varepsilon_i \qquad [5]$$

where P_i is the conditional probability of default, Y_i represents the binary default variable, X_i is the i^{th} attribute and e is the base of natural logarithms. The weights for each attribute in Y_i is estimated using the likelihood function and comprises the product of all P_i's for the present and historical values of all defaulters times the product of all $(1-P_i)$ in the case of non-defaulters. The α and β coefficients are estimated by maximizing the likelihood function.

Nonparametric techniques have gained in popularity in recent years as dependable alternatives to LDA and logit models as these techniques are not subject to the restrictive parametric assumptions, which would threaten the reliability of estimates if violated (see Luther, 1998 and Zhang et al., 1999). These restrictive assumptions such as no multicollinearity or autocorrelation as well as Gaussian distributions are unsuited particularly in cases where the default variable and observable attributes exhibit complex non-linear relationships with skewed and leptokurtic distributions.[7] Although flexible form non-parametric techniques such as ANN models typically contain a relatively larger number of non-interpretable parameters, these models of pattern recognition have been shown to produce more accurate parameter estimates when compared with the pure statistical methods, especially in applications with complex datasets (see, for example, Salchenberger, Cinar, and Lash (1992), Coats and Fant (1993), Luther (1998), Huang, Dorsey, and Boose (1994), and Brockett et al. (1994), Lacher et al. (1995), West, Brockett and Golden (1997), Jain and Nag (1997), Etheridge et al. (2000), Wu et al. (2006)). This feature of ANN models is critical for the practical use of the IRB approach under Basel II (BCBS, 2005).

The application of ANN to predict default probabilities was motivated by desire of researchers to simulate the learning processes that take place in the biological brain and nervous system when reacting to changes in the system's internal and external environment. Specifically, an ANN is built up of a group of many artificial neurons (processing units or nodes) interacting in parallel with their individual memories (synapses), creating networks through weighted connections. The aim of this network is to transform the inputs into outputs through the recognition and comprehension of the behavioral patterns of the environmental changes, similar to their biological counterparts.

Neurons in the human brain function by processing information using its main components of a nucleus, an axon and subdivided dendrites (see Mc Cullock and Pitts, 1943). Each of the neurons in the ANN system is excited or inhibited by sending and receiving signals (spikes) through axons and dendrites, respectively, which extend from the cell body (soba) and connect to cell inputs through synapses. The dendrites transform the signals into specific outputs which are then transmitted through the axon to other neurons. Signals are either purely transmitted or altered by the synapses which varies the signal strength and also stores knowledge. Synaptic strength modification contributes to neural learning and

[7] Practical applications of ANN models include character and voice recognition, weather forecasting, bankruptcy prediction, customer credit scoring, fraud detection, financial price prediction, aerospace and robotics.

can be simulated in the ANN through the application of mathematical optimization techniques to derive the parameters of the network.

The main elements of the processing units for learning are the inputs, weights, summation function, transformation function and output. These processing units are organized in different ways to form the network's configuration. The basic configuration is a single neuron with a number of inputs and one output, termed a perceptron. A subgroup of processing units is termed a layer in the ANN, where the first layer is the input layer and the last layer is the output layer. However, there may be additional layers of units between the input and output layers, called hidden layers (see Figure 1). Several hidden layers may be positioned between the input (independent variables, in standard statistical terminology) and output layers (dependent variables). The ANN with one input layer, one or more hidden layers and an output layer is called the Multilayer Perceptron (MLP) (see, Rosenblatt, 1962).

The supervised iterative learning (training) algorithm of a perceptron, in the context of default prediction using an ANN consists of three phases. In the first phase, input layers receive the incoming stimuli. In the second phase, input values are multiplied with initial syntactic weights and all the multiplications are summed. An ANN is trained by adjusting the values of the weights between elements. In the final phase, the summed value is converted to output values using an activation (transfer) function and then compares these predicted values to a predetermined threshold. If the final value does not exceed that threshold, the node will not be triggered. The learning algorithm and the weight vector modification process may be achieved either using backpropogation algorithms or a feed forward learning process. Input and target samples are automatically divided into training, validation and test sets. If the backpropogation algorithm is chosen, the flow of information travels in both directions, because there are feedback connections. Optimization of the weights is made by backward propagation of the error during training phase. To improve the overall predictive accuracy and to minimize the network total root mean squared error (RMSE) between desired and predicted output, weight vectors are revised in the network. This process is continued through the training set until a minimum tolerable level of error (threshold limit) or a predetermined number of iterations is achieved to stop the iterations (epochs). In contrast, during the training phase of a static multi-layer feed forward learning process, the hidden neurons learn the pattern in the data and map the relationship between input and output pairs using a transfer function with information moving in only in a forward direction. The training phase continues as long as the network continues improving on the test set.

Following the three phase process of a perceptron in the first layer of the MLP, the neurons of the input layers forward the information to all neurons of the middle layers. Receiving units in the middle layers (hidden units) repeat the identical process, which are critical for ANNs models to capture the complex patterns (non-linear interrelationships) in the data between input and output layers (see Zhang et al., 1999). The process is repeated again by the output layer neurons. The topology of the network architecture distinction is an important factor for achieving successful ANNs. For most ANNs, one hidden layer is sufficient and introducing additional layers may lead to convergence to local minima instead of the global minimum. Note also that if an insufficient number of neurons are used in the hidden layer, the ANN will fail to capture nonlinearities in the data. On the other hand, if the number of neurons is excessive, the ANN may over fit the data resulting in poor out-of-sample results. A validation process must be conducted to ensure that over fitting does not occur (see Refenes, 1995).

It is worth repeating that ANNs have very beneficial features for modeling complex unstructured relationships without any restrictive assumption about the underlying correlation. However, this also serves as a shortcoming in that no economic interpretation can be applied to the values for connection weights. In addition, the number of connection weights to be modified is typically very large which contributes to a very lengthy training time.

The next section discusses the data to be used in the estimation of the ANN application for macroprudential surveillance in the Jamaican case. A more detailed explanation of the network's architecture is given in section 4.

3.0 Data Description and Analysis

The main aim of this study is to investigate the appropriate key macroeconomic variables affecting Jamaica's sovereign credit risk. The specific explanatory variables utilized include 3-month lagged values of the CPI inflation rate, US-Jamaica currency exchange rate, the real Government of Jamaica (GOJ) 180-day Treasury bill rate (Tbill), external debt to exports, net international reserves to imports, real effective exchange rate, terms of trade index, current account of the BOP, real GDP growth and the unemployment rate (UR). The sovereign rating series were derived from S&P ratings of GOJ Global bonds. The data set spans 128 months from May 2001 to December 2011. In terms of data preparation, all independent variables are converted to 12-month moving averages and then normalized to avoid disproportional measurement of variable contributions to the predicted ratings due to diverse dimensions and units of input (see Table 4 for descriptive statistics for model variables in moving averages) . The normalization process transforms all the converted independent variables in the training set, X_{it}, to have values between -1 and 1 as given by

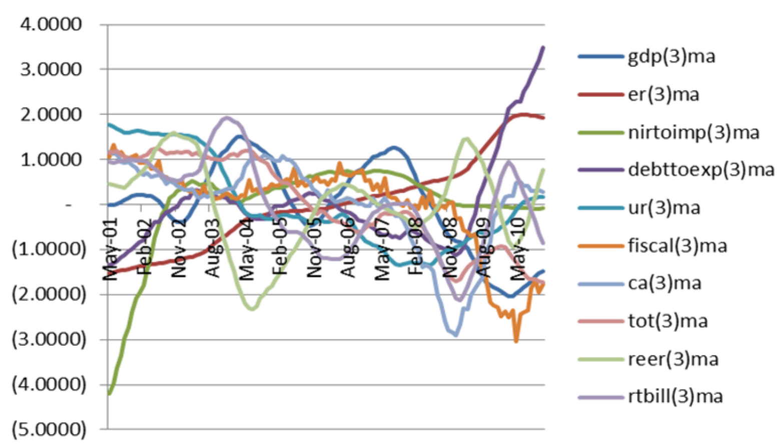
$$Z_{it} = \frac{\left(X_{it} - \mu_{i;}\right)}{\sigma_i} \qquad [6]$$

using the mean and standard deviation of X_{it}, denoted as μ_i and σ_i, respectively. The macroeconomic variables are transformed by applying the normalization process represented by equation (6) in order to avoid spurious contribution results (see Figure 1).

Table 4. Summary Statistics for Model Variables (Moving Averages)

	Real GDP	Exchange Rate	NIR/ Imports	Debt/ Exports	Unemploy-ment Rate	Fiscal Balance	Current Account Balance	Terms of Trade	Real Effective Exchange	Real Treasury Bill Rate	Sovereign Credit Rating
Mean	0.71	65.61	4.24	39.77	4.12	(3,759)	(100.99)	77.45	99.09	4.29	6.82
Variance	3.35	197.85	0.10	146.80	0.31	6,931,332	2,766.71	135.52	18.05	31.27	1.00
Std. Dev.	1.83	14.07	0.32	12.12	0.56	2,633	52.60	11.64	4.25	5.59	1.00
Skewness	(0.36)	0.23	(2.76)	1.35	0.50	(1)	(1.46)	(0.17)	(0.16)	0.12	(2.97)
Kurtosis	2.26	2.03	10.81	3.62	2.09	3	4.60	1.66	2.77	2.06	19.96
Median	0.85	63.83	4.31	36.26	3.96	(2,853)	(87.95)	77.12	99.26	4.16	7.00
Mean Abs. Dev.	1.47	11.40	0.19	9.02	0.46	2,067	37.94	10.04	3.32	4.75	0.60
Mode	1.22	85.89	4.22	38.24	5.01	(2,130)	(82.03)	90.84	100.02	1.64	7.00
Minimum	(2.96)	43.61	2.82	25.22	3.32	(11,381)	(260.66)	55.91	89.73	(6.57)	0.00
Maximum	3.61	89.27	4.50	69.30	5.14	(27)	(36.30)	92.32	107.80	15.37	8.00
Range	6.58	45.66	1.68	44.09	1.82	11,354	224.36	36.41	18.07	21.94	8.00
1st Quartile	(0.43)	54.79	4.22	31.43	3.72	(5,060)	(109.54)	67.07	96.77	(0.79)	6.80
3rd Quartile	2.24	73.69	4.41	39.23	4.35	(2,109)	(69.86)	90.16	102.08	8.81	7.00
Interquartile Range	2.66	18.91	0.19	7.80	0.63	2,951	39.68	23.09	5.30	9.59	0.20

Figure 1. Transformed macroeconomic variables after normalization



4. Methodology

4.1 ANN Architecture

The feedforward ANN typology (with no feedback information transfer) used in this paper consists of a MLP with three basic layers: the input layer, a hidden layer and the output layer. Technically, the network output $y_{h,j}$ for each node (neuron) j in the hidden layer h can be expressed:

$$y_{h,j} = f\left(w_{h,0} + \sum_{i=1}^{N_i} w_{h,ji} x_i \right) , \qquad [7]$$

where $w_{h,ji}$ is the weight which connects input node i to node j in the hidden layer weight matrix W_h , $w_{h,0}$ is the bias weight in the hidden layer associated with a vector of ones, x_i is the i^{th} element of the associated input vector $X=(x_1, x_2, \ldots, x_{Nj})^T$ and N_i is the number of nodes in the input layer, which is found empirically. The hidden layer activation function f is the nonlinear hyperbolic tangent function which is continuous and differentiable and produces an output value between -1 and 1 as defined by $f(u)=(1-e^{-bu})/(1+e^{-bu})$, where b is the slope parameter and u is the result of the weighted sum of the node inputs. All information in the input layer is fed-forward to the hidden layer with no feedback loop from output to input nodes.

Since the output of the hidden layer is an input in the output layer, the output of the network is computed as:

$$y_{o,k} = g\left[w_{o,0} + \sum_{j=1}^{N_j} f\left(w_{h,0} + \sum_{i=1}^{N_i} w_{h,ji} x_i \right) w_{o,kj} \right] . \qquad [8]$$

The identity function g(u) = u, for each u used in the output layer. In this case, the output of a neuron is simply the function of the linear combination of hidden unit’s activation.

The learning process determines adjustments to the set of weight values. Supervised learning for a training pattern p is assumed where each node k has a predefined threshold or target value ($t_{p,k}$) used to train the network. If the network output ($y_{p,o,k}$) does not exactly match the threshold, the error signal for the training process is given as

$$\varepsilon_{p,o,k} = t_{p,k} - y_{p,o,k} . \qquad [9]$$

The squared aggregate of these errors is minimized using the cost function J_p to determine the optimal solution where the computed outputs are within an acceptable tolerance of the target outputs with respect to the input units. The squared error for a training pattern *p* is given as

$$J_p(w) = \frac{1}{2}\sum_{k=1}^{N_o} \varepsilon^2_{p,o,k} \ . \qquad [10]$$

In the case of the overall training set of *p* patterns, the squared error is

$$J(w) = \sum_{p=1}^{P} J_p = (Y', w) \, . \qquad [11]$$

where *Y'* is the output vector and *w* is the weight vector. The optimal solution *w** must satisfy the condition J(w*) ≤ J(w) and the necessary condition for optimality is ΔJ(w) = (ƏE/Əw) = 0, where Δ is the gradient operator.

The modification of synaptic weight vector (Δ*w*) of the network is calculated after each presentation of a single pattern or at the end of an epoch. The weight update equation of the training algorithm is

$$w(t+1) = w(t) + \Delta w(t) = w(t) - \eta(t)\nabla(t)J(t) \, . \qquad [12]$$

where η is called the learning rate which defines the proportion of error (step size) for updating the weights and ∇ is the gradient vector.

The most common gradient descent technique to optimize the mean square error is the delta rule or back-propagation algorithm. However, this type of training algorithm will typically find local minima of the error function that are far from the global minimum and often lead to slow training. A second order optimization method called Conjugate Gradient (CG) which uses a numerical approximation for the second derivatives (Hessian matrix) is more powerful than the back-propagation algorithm in terms of efficiency and ability to find the global optimum. The CG method is employed in this study. The CG method chooses a suitable direction vector ***p*** to update the weight vector as

$$w(t+1) = w(t) + \eta(t)p(t) \qquad [13]$$

where, at the initial weight vector w_0

$$p_0(t) = -\nabla_0(t) \qquad [14]$$

At the minimum of the line search

$$\frac{\partial}{\partial \eta} J\left[w(t+1)+\eta(t)\, p(t)\right]=0 \qquad [15]$$

yields

$$\nabla(t+1)' \cdot P(t)=0\,. \qquad [16]$$

In order to reduce the likelihood of finding one of the multiple local minima rather than the global minimum, the CG algorithm is combined with a stochastic search method called simulated annealing (Kirkpatrick et al, 1983). This search technique introduces random noise *T* into weight update equation which is systematically decreased at a constant rate *d*. This optimization method presents the weights with the training data and allows a random change of search location on the error surface with probability given by the Boltzmann factor

$$P=\exp\left[\frac{-\left(J_1-J_2\right)}{dT}\right] \qquad [17]$$

which permits a more comprehensive search process.

4.2 ANN Model Estimation

Before, the ANN model is estimated the data set is partitioned into two samples, a validation or forecasting sample and an estimation sample. The forecasting sample selected covered the last 12 monthly observations for the overall sample, which is approximately 10.0 per cent of the data set. The estimation sample was further randomly subdivided into the training set (60.0 per cent of estimation sample) and testing set (remaining 40.0 per cent).

Regarding the ANN architecture using the multi-layer feed forward training process, the number of hidden layers and the number of neurons in the hidden layer(s) are determined based on the training data set. Note that the neurons in the input and output layers are the explanatory variables (11 neurons) and dependent variables (1 neuron), respectively. Once the network typology has been fixed, the training process to adjust weights is terminated to avoid ‘over-training’ when the RMSE reduces by less than 0.0001 or a maximum number of epochs is automatically determined by the software, whichever condition occur first.[8] Following the training process, the in-sample prediction performance of the network is assessed using the testing set of observations. Finally, after a stable matrix of weights is found, ex-post forecasts are conducted to assess the predictive power of the model by comparing with out-of-sample values for sovereign credit rating.

The CG- simulated annealing process for the feedforward network used in this study with the typology of one hidden layer (*h*) with four nodes and output layer (*o*) with one node is summarized in equations [18] and [19].

$$\left(\begin{bmatrix} \mathrm{gdp}_{\mathrm{t-3}} \\ \mathrm{debttoexp}_{\mathrm{t-3}} \\ \mathrm{er}_{\mathrm{t-3}} \\ \mathrm{ur}_{\mathrm{t-3}} \\ \mathrm{gojsovyield}_{\mathrm{t-3}} \\ \mathrm{nirtoimp}_{\mathrm{t-3}} \\ \mathrm{reer}_{\mathrm{t-3}} \\ \mathrm{tot}_{\mathrm{t-3}} \\ \mathrm{fiscal}_{\mathrm{t-3}} \\ \mathrm{rtbill}_{\mathrm{t-3}} \\ \mathrm{ca}_{\mathrm{t-3}} \end{bmatrix} \times \begin{bmatrix} w_{h,1,1} & w_{h,1,2} & w_{h,1,3} & w_{h,1,4} \\ w_{h,2,1} & w_{h,2,2} & w_{h,2,3} & w_{h,2,4} \\ w_{h,3,1} & w_{h,3,2} & w_{h,3,3} & w_{h,3,4} \\ w_{h,4,1} & w_{h,4,2} & w_{h,4,3} & w_{h,4,4} \\ w_{h,5,1} & w_{h,5,2} & w_{h,5,3} & w_{h,5,4} \\ w_{h,6,1} & w_{h,6,2} & w_{h,6,3} & w_{h,6,4} \\ w_{h,7,1} & w_{h,7,2} & w_{h,7,3} & w_{h,7,4} \\ w_{h,8,1} & w_{h,8,2} & w_{h,8,3} & w_{h,8,4} \\ w_{h,9,1} & w_{h,9,2} & w_{h,9,3} & w_{h,9,4} \\ w_{h,10,1} & w_{h,10,2} & w_{h,10,3} & w_{h,10,4} \\ w_{h,11,1} & w_{h,11,2} & w_{h,11,3} & w_{h11,4} \end{bmatrix}\right) + \begin{bmatrix} \alpha_{h,1} \\ \alpha_{h,2} \\ \alpha_{h,3} \\ \alpha_{h,4} \end{bmatrix} = f\left(\begin{bmatrix} y_{h,1} \\ y_{h,2} \\ y_{h,3} \\ y_{h,4} \end{bmatrix}\right) \qquad [18]$$

$$\left(f\left(\begin{bmatrix} y_{h,1} \\ y_{h,2} \\ y_{h,3} \\ y_{h,4} \end{bmatrix}\right) \times \begin{bmatrix} w_{o,1,1} \\ w_{o,2,1} \\ w_{o,3,1} \\ w_{o,4,1} \end{bmatrix}\right) + [\alpha_o] = g([y_{o,}]) \qquad [19]$$

[8] @Risk NeuralTools, Palisade Corporation.

5.0 Network Results

Various one-hidden-layer MLPs were estimated with the number of hidden nodes ranging between one and six (see Table 5). The MLP with a typology with one hidden layer with four nodes achieved the largest decline of the training RMSE (0.0076), with a correct classification rate of 98.6%, as well as the lowest test RMSE (0.2175), with a correct classification rate of 100.0% (see Figures 2 and 3). In addition, the results from a comparison of the ranges of RMSEs for testing set subdivisions of 10%, 20%, 30% and 40%, indicated that the use of 40% testing produced the most reliable estimates (see Table 6).

Table 5. Test RMSEs for various MLP Models

	RMS Error	Training Time in Minutes
MLP with 2 Nodes	0.59	0:18:00
MLP with 3 Nodes	0.90	0:21:00
MLP with 4 Nodes	0.22	0:17:00
MLP with 5 Nodes	1.02	0:19:00
MLP with 6 Nodes	0.78	0:27:00

Figure 2. Actual Vs. Predicted Ratings of Training Data

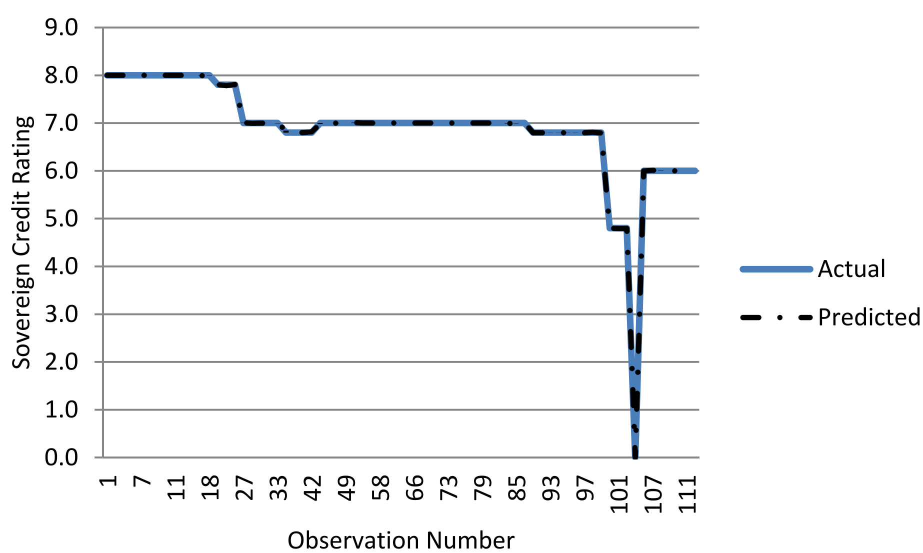


Figure 3. Actual Vs. Predicted Ratings of Testing Data

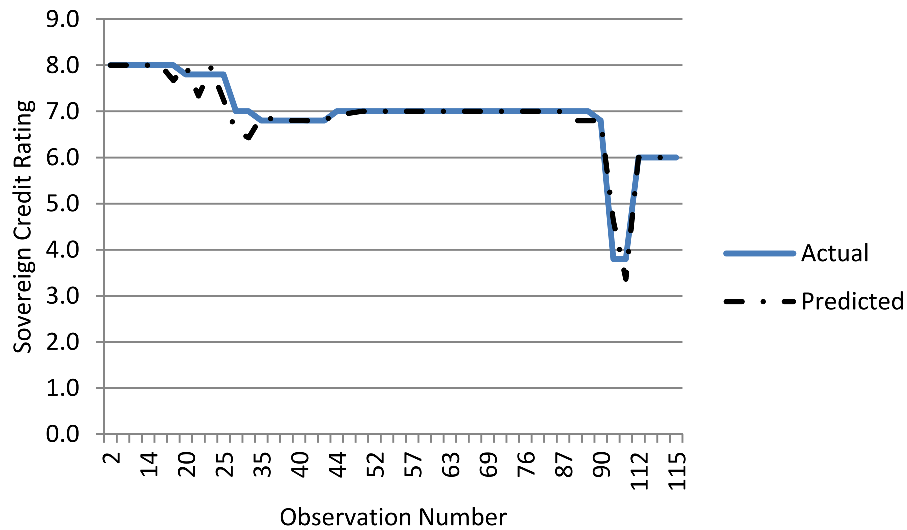
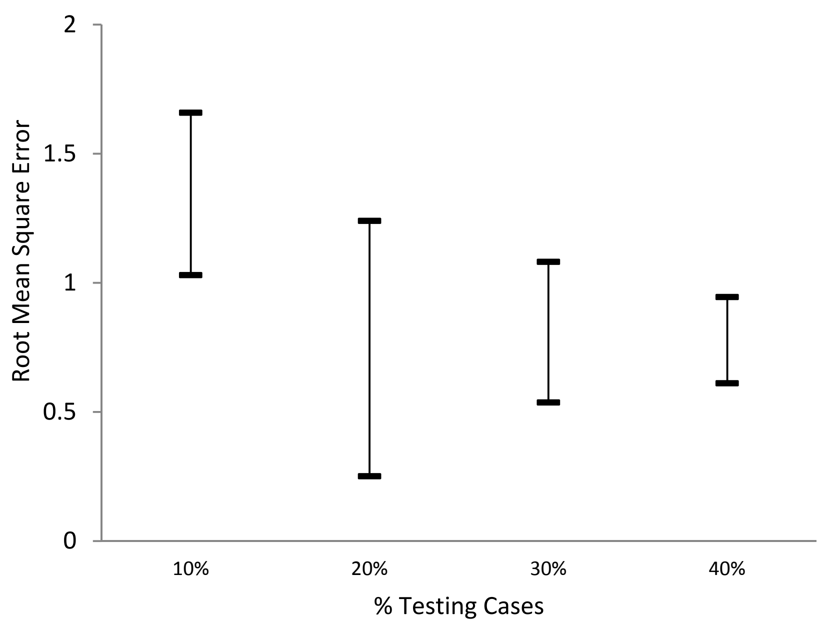


Table 6. Testing Sensitivity Analysis



Finally, sensitivity analysis was conducted on the ANN training data to provide information about relative significance of each independent variable. The results indicate that External Debt/Exports, NIR/Imports, Unemployment Rate and Fiscal Balance contribute 17.84%, 16.53%, 16.17% and 11.35% in explaining GOJ sovereign credit risk, respectively (see Table 7). The other independent variables such as

Current Account Balance, REER, Real Treasury Bill Rate, Exchange Rate, TOT and Real GDP accounted for lower contributions of 9.13%, 7.65%, 6.61%, 5.81%, 5.76% and 3.15%, respectively.

Table 7. Relative Impact Analysis

Variable	**Contribution**
External Debt/Exports	17.84%
NIR/Imports	16.53%
Unemployment Rate	16.17%
Fiscal Balance	11.35%
Current Account Balance	9.13%
Real Effective Exchange Rate	7.65%
Real GOJ Treasury Bill Rate	6.61%
Exchange Rate (US$/J$)	5.81%
Terms of Trade	5.76%
Real GDP	3.15%

The predictive power of the ANN model is investigated by forecasting using the out-of-sample independent variables for the period January 2011 to December 2011 and then comparing with actual S&P ratings. Out-of-sample performance results show that the model was able to accurately predict 9 out of the 12 ratings (see Table 8).

Table 8. Out-of-sample results

Out-of-sample period	Actual rating	Predicted rating
Jan-11	6.0	6.0
Feb-11	6.0	6.0
Mar-11	6.0	6.0
Apr-11	6.0	6.0
May-11	6.0	6.0
Jun-11	6.0	6.0
Jul-11	6.0	6.0
Aug-11	6.0	6.0
Sep-11	6.0	6.0
Oct-11	5.5	6.0
Nov-11	5.5	6.0
Dec-11	5.5	6.0

6.0 Summary and Concluding Remarks

The aim of this paper is to formulate an ANN model to predict the sovereign risk of the Jamaican Government. Neural networks offer an important advantage over traditional models of classification because they are able to capture complex empirical relationships. The estimated results of the ANN model show the importance of 3-month lagged, 12-month moving averages of CPI inflation rate, US-Jamaica currency exchange rate, the real GOJ 180-day Treasury bill rate, external debt to exports, net international reserves to imports, real effective exchange rate, terms of trade index, current account of the BOP, real GDP growth and the unemployment rate in determining GOJ sovereign risk. The overall prediction accuracy of the ANN model is 98.6% using training data and 100% for the testing observations.

The results also reveal that the ratio of external debt to exports (12-month moving average) is the most significant leading variable on GOJ sovereign risk rating (17.84%). The other macroeconomic variables contributing over 10% to GOJ default are NIR to imports (16.53%), unemployment rate (16.17%) and the fiscal balance (11.35%). Given the significant exposure of Jamaica's financial sector to GOJ sovereign default risk these variables should be monitored closely in the assessment of financial stability.

References

Altman, E. (1968), 'Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy,' *Journal of Finance*: 189–209.

Altman, E. and B. Brady (2001), 'Explaining the Correlation Between Default and Recovery Rates on Corporate Bonds,' NYU Salomon Center Working Paper.

Anderson J.A. and E. Rosenfeld (1988), *Neurocomputing: Foundations of Research*. Cambridge, MA: MIT Press.

Bank for International Settlements (2005), 'International Convergence of Capital Measurement and Capital Standards: A Revised Framework,' BIS Publication.

Brockett, P.L., W.W. Cooper, L.L. Golden and U. Pitaktong (1994), 'A Neural Network Method for Obtaining an Early Warning of Insurer Insolvency,' *The Journal of Risk and Insurance* 61 (3), pp. 402-424.

Cantor, R. and F. Packer (1996), 'Determinants and Impact of Sovereign Credit Ratings,' *Economic Policy Review,* Federal Reserve Bank of New York Research Paper Series Vol. 2, No. 2.

Coats, P.K and L.F. Fant (1993), 'Recognizing Financial Distress Patterns using a Neural Network Tool, *Financial Management*. Pp. 142-155.

Etheridge, H. L, Sriram, R.S, and Hsu, H.Y.K (2000), 'A Comparison of Selected Artifical Neural Networks That Help Auditors Evaluate Client Financial Viability,' *Decision Sciences*, vol.31 (2), pp. 531-550.

Fama E. (1986), 'Term Premiums and Default Premiums in Money Markets,' *Journal of Financial Economics*, *17*(1), pp. 175-96.

Gande, A. and D. Parsley (2010), 'Sovereign Credit Ratings, Transparency and International Portfolio Flows,' MPRA Paper No. 21118.

Hall, M.J.B., D. Muljawan, Suprayogic and L. Moorena (2009), 'Using the Artificial Neural Network to Assess Bank Credit Risk: a case study of Indonesia,' *Applied Financial Economics*, Vol. 19, pp. 1825–1846.

Haque, N.U., M. Kumar, N. Mark and D. Mathieson (1996), 'The Economic Contents of Indicators of Developing Country Creditworthiness,' *IMF Staff Papers*, 43. No.4, pp. 688-724.

Huang, C.S., R.E Dorsey and M.A. Boose (1994), 'Life Insurer Financial Distress Prediction: A Neural Network Model,' *Journal of Insurance Regulation* 13 (2), pp. 131-167.

International Monetary Fund (September 2011), *Global Financial Stability Report*. IMF, Washington DC.

Jain, B. A., and B. N. Nag (1997), 'Performance Evaluation of Neural Networks Decision Models,' *Journal of Management Information System*, 14, pp. 201-216.

Kirkpatrick S., C. D. Gelatt and M. P. Vecchi (1983), ‘Optimization by Simulated Annealing,’ *Science*, New Series, Vol. 220, No. 4598, pp. 671-680.

McCulloch W., and W. Pitts (1943), ‘A Logical Calculus of Ideas Immanent in Nervous Activity,’ *Bulletin of Mathematical Biophysics*, Vol. 5, (1-2), pp. 99-115.

Lacher, R.C., P.K. Coats, S.C. Sharma and L.F. Fant (1995), ‘A Neural Network for Classifying the Financial Health of a Firm,’ *European Journal of Operations Research* 85, pp. 53±65.

Luther, R. K. (1998), 'An Artificial Neural Network Approach to Predicting the Outcome of Chapter 11 Bankruptcy,' *The Journal of Business and Economic Studies*, vol. 4, no. 1, pp. 57-73.

Ohlson, J.A. (1980), ‘Financial Ratios and the Probabilistic Prediction of Bankruptcy,’ *Journal of Accounting Research*, Volume 18, Number 1, 109-31.

Panahian, H. (2011), ‘Using Artificial Neural Network for Predicting Default Risk in Banking Systems (A Comparative Study between Indonesian and Iranian Banking System),’ *American Journal of Scientific Research*, Issue 30, pp. 36-53.

Pesaran, M.H., T. Schuermann, B. Treutler and S.M. Weiner (2003), ‘Credit Risk and Macroeconomic Dynamics,’ Social Science Research Network.

Refenes, A.P. (1995). *Neural Networks in the Capital Markets*. John Wiley & Sons.

Rosenblatt, F. (1962), *Principles of Neurodynamics; Perceptrons and the Theory of Brain Mechanisms*. Washington: Spartan Books.

Salchengerger, L.M, E.M. Cinar and N.A. Lash (1992), ‘Neural Networks: A New Tool for Predicting Thrift Failures,’ *Decision Sciences* 23 (4), pp. 899-916.

Standard & Poors (2011), ‘Analytical Linkages Between Sovereign And Bank Ratings,’ RatingsDirect on the Global Credit Portal.

West, P.M., P.L. Brockett and L.L Golden (1997), ‘A Comparative Analysis of Neural Networks and Statistical Models for Predicting Consumer Choice,’ *Marketing Science* 16, pp. 370 – 391.

Wilson, T. (1997), ‘Portfolio Credit Risk,’ Social Science Research Network.

Wu, A., W.W. Hsieh and B. Tang (2006), ‘Neural Network Forecasts of the Tropical Pacific Sea Surface Temperatures,’ *Neural Networks*, Vol. 19, pp. 145-154.

Zhang, E., M.Y. Hu, B.E. Patuwo, D.C. Indro (1999), ‘Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analysis,’ *European Journal of Operational Research*, 116, pp. 16-32.