Forecasting Latin-American yield curves: An artificial neural network approach

Por: Daniel Vela

Borradores de ECONOMÍA

Núm. 761
2013

BANCO DE LA REPÚBLICA COLOMBIA

Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia

Forecasting Latin-American yield curves: An artificial neural network approach*

Daniel Vela†
Banco de la República (Colombia)

Abstract

This document explores the predictive power of the yield curves in Latin America (Colombia, Mexico, Peru and Chile) taking into account the factors set by the specifications of Nelson & Siegel and Svensson. Several forecasting methodologies are contrasted: an autoregressive model, a vector autoregressive model, artificial neural networks on each individual factor, and artificial neural networks on all factors that explain the yield curve. The out-of-sample performance of the fitting models improves with the neural networks in the one-month-ahead forecast along all studied yield curves. Moreover, the three factor model developed by Nelson & Siegel proves to be the best choice for out-of-sample forecasting. Finally, the success of the cross variable interaction strongly depends on the selected yield curve.

Keywords: Term structure of interest rates, Nelson & Siegel, Svensson, out-of-sample forecast, Artificial Neural Networks

JEL classification: C32, C45, E43, G17

---

* I am grateful to Marco Ruíz who provided valuable comments and suggestions. The opinions and statements are the sole responsibility of the author and do not necessarily represent neither those of the Banco de la República nor of its Board of Directors. Any remaining errors are my own.

† Specialized Professional, Research and Analysis Group, Foreign Reserves Department, Banco de la República, dvelabar@banrep.gov.co.

1. Introduction

The term structure of interest rates defines the relationship between the interest rates and the time to maturity. Consequently, it has inspired practitioners and researchers to find different approaches to forecast it, in order to achieve an investment gain. Dolan (1999) classified the yield curve models in three different types: i) Stochastic and arbitrage free, ii) Principal Components, and iii) Fundamental. The most successful models are the ones that focus in latent factors that determined the term structure. Litterman & Scheinkman (1991) established that the dynamic of the term structure of interest rates can be depicted with three principal components[3].

Dolan (1999) was the first to link these principal components to the functional form of the term structure given by Nelson & Siegel (1987) (hereafter NS), by determining that the three factors resemble the level, slope and curvature (or butterfly as describe by the author) of the yield curve. Later on, Diebold and Li (2006) determined that the future dynamic of the term structure of interest rates could be described by using a statistical or parametric model that forecasted the factors, which are restructured into the yield curve with the NS model. Subsequent authors, such as De Pooter (2007), Cziráky (2007) or De Rezende & Ferreira (2011), studied the forecasting ability of other functional forms that depart from the NS model, like the Svensson model (1994), Bliss model (1997) or Björk & Christensen (1999).

Almost all the literature that uses the Diebold & Li methodology to forecast the term structure of interest rates focuses exclusively on the yield curve of United States. Few authors, like Bolder (2006) or De Rezende & Ferreira (2011), evaluated the methodology in different markets. Therefore, one of the contributions of this paper is that it tests this methodology in Latin-American yield curves (Colombia, Mexico, Chile and Peru). Additionally, the paper also studies the ability to forecast the latent factors of the yield curve by using feedforward artificial neural network (ANN), that review each factor individually or that include the interaction between the state variables. These types of ANN are compared with the performance of an autoregressive and vector autoregressive models. The functional forms used to determine the factors to forecast are the NS model and the Svensson expansion (hereafter NSS).

This paper is divided in six chapters. The first one is this introduction. The second presents the data of the yield curves of the analyzed countries. The third describes the modeling framework given by the NS and NSS models. The fourth explains the estimation methods used to forecast the latent factors of the yield curve. The fifth presents the main results of the different approaches and compares the root mean squared error (RMSE) of the resulting out-of-sample yield curves. Finally, the sixth makes some concluding remarks.

2. Data

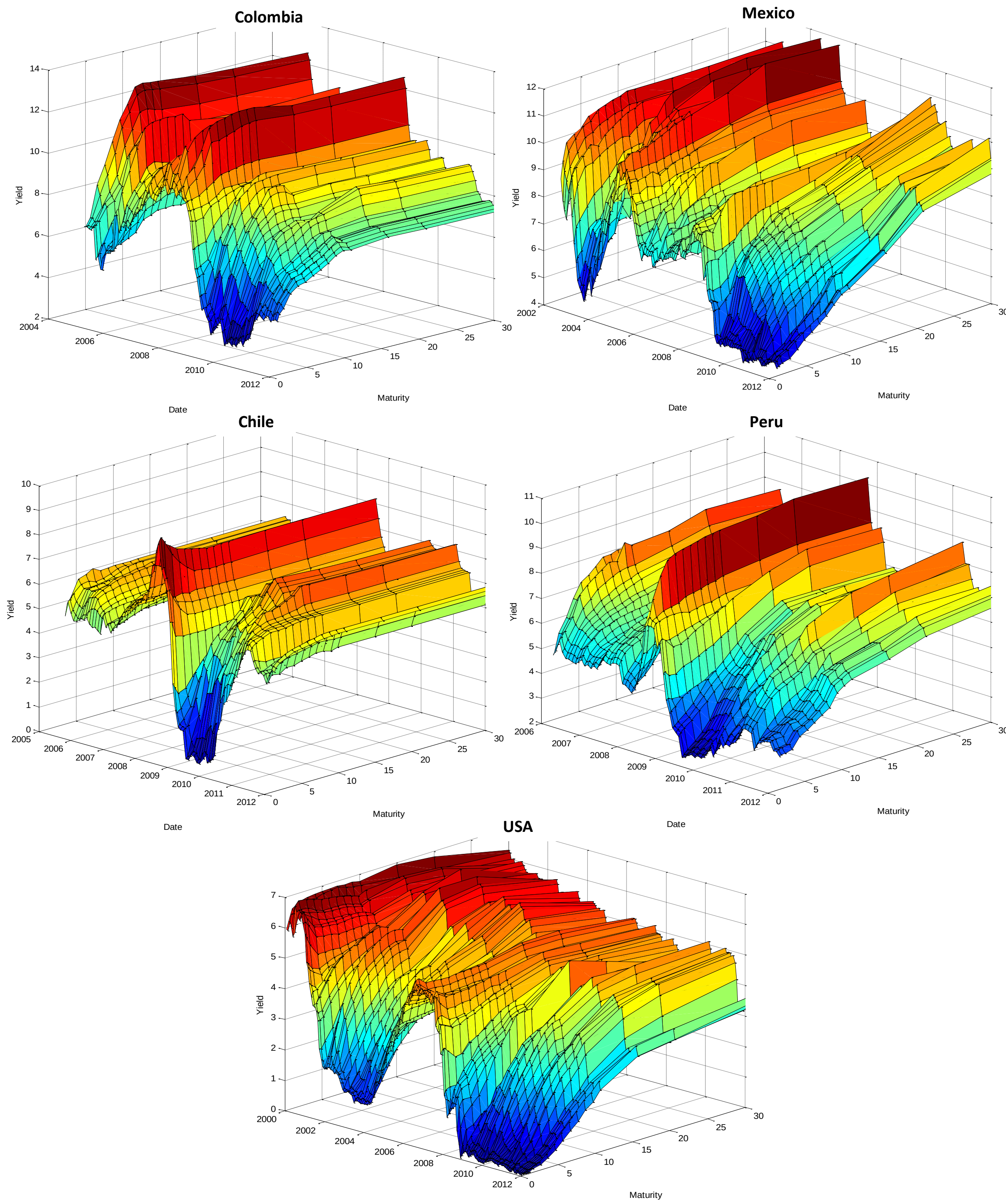
As mentioned before, one of the main objectives of this document is to analyze Diebold & Li framework in the context of Latin American yield curves. Due to accessibility of information, the selected countries

---

[3] With a principal component analysis (PCA) one can find orthogonal factors that explain in a statistical sense the variance of the changes in interest rates.

are Colombia, Mexico, Chile and Peru. The yield curve of United States is also analyzed to review the results of the ANN estimation model against others proposed in the literature.

Figure 1: Yield Curves



Source: Bloomberg

When studying this framework, the literature presents different vertices that represent the zero-coupon yield curve accordingly to the liquidity conditions of the markets and the availability of information. For instance, De Rezende & Ferreira (2011) use vertices from 1 month up to 60 month for the yield curve of Brazil, whilst Fabozzi, Martellini & Priaulet (2005) use vertices from 3 months up to 30 years for the zero-coupon yield curve of United States and De Pooter (2007) uses vertices from 3 months up to 10 years for the same yield curve. For comparative purposes the vertices chosen in this document are: 0.25, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 and 30 years.

Figure 1 presents the historical yield curves for the selected countries. The availability of information varies depending on the country, from 77 months for Peru to 152 months for United States. Therefore, to ensure sufficient data to estimate the forecasting models, one and a half years are left to be used for the evaluation of the out-of-sample forecast.

3. Modeling Framework

The main goal of Nelson & Siegel (1987) was to extract the zero-coupon and forward interest rates curves from a collection of coupon bond prices to find a continuous functional form that describes the term structure of interest rates. They suggested the following form of instantaneous forward rates up to maturity m:

$$f(m;b) = \beta_0 + \beta_1 e^{\left(-\frac{m}{\tau}\right)} + \beta_2 \frac{m}{\tau} e^{\left(-\frac{m}{\tau}\right)} \qquad [1]$$

$$b = (\beta_0, \beta_1, \beta_2, \tau)$$

Likewise, the zero coupon curve at a particular point is given by:

$$z(m;b) = \beta_0 + \beta_1 \left(\frac{1 - e^{\left(-\frac{m}{\tau}\right)}}{\frac{m}{\tau}} \right) + \beta_2 \left(\frac{1 - e^{\left(-\frac{m}{\tau}\right)}}{\frac{m}{\tau}} - e^{\left(-\frac{m}{\tau}\right)} \right) \qquad [2]$$

This is the starting point for Diebold & Li (2006), by also using the research of Litterman & Scheinkman (1991); they concluded that the coefficients β_0, β_1, β_2 represented the latent factors of the yield curve, level, slope and curvature. The parameter τ, or $\lambda = {}^{1}/_{\tau}$, is fixed by the authors, consequently the equation is reduced to a linear combination of three functions with coefficients that represent the latent factors. By fitting the yield curve in a time period, the authors develop a time series for the level, slope and curvature, which they used to forecast the factors with an estimation method. Bolder (2006) explained that the main disadvantage of this approach is that it does not has a theoretical model foundation, as seen with least successful forecasting models such as theoretical affine term-structure models, which incorporate a notion of risk premia.

Furthermore, other authors review the forecasting ability of supplementary yield curve functional forms that include more factors. For instance, Cziráky (2007) analyzed the inclusion of the Svensson fourth term in the NS equation, whilst De Pooter (2007) evaluated an adjusted version of the NSS model, the Bliss model and the model of Björk & Christensen. Additionally from the listed models, De Rezende & Ferreira (2011) proposed a five factor equation. Almost all the authors reported a better in-sample fitting with the enhanced NS models, considering that the extra factors provide more flexibility. Although, the models with a larger amount of factors do not necessary have to perform well out-of-sample because of the risk of overfitting. In this document the NSS functional form is studied.

The contribution of Svensson (1994) was to extend the NS model by including an additional term that allowed a second hump shape by including to additional parameters β_3 and τ_2. Therefore, the form of instantaneous forward rates and the zero coupon rates up to maturity m are given by:

$$f(m;b) = \beta_0 + \beta_1 e^{\left(-\frac{m}{\tau_1}\right)} + \beta_2 \frac{m}{\tau_1} e^{\left(-\frac{m}{\tau_1}\right)} + \beta_3 \frac{m}{\tau_2} e^{\left(-\frac{m}{\tau_2}\right)} \qquad [3]$$

$$b = (\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2)$$

$$z(m;b) = \beta_0 + \beta_1 \left(\frac{1 - e^{\left(-\frac{m}{\tau_1}\right)}}{\frac{m}{\tau_1}} \right) + \beta_2 \left(\frac{1 - e^{\left(-\frac{m}{\tau_1}\right)}}{\frac{m}{\tau_1}} - e^{\left(-\frac{m}{\tau_1}\right)} \right) + \beta_3 \left(\frac{1 - e^{\left(-\frac{m}{\tau_2}\right)}}{\frac{m}{\tau_2}} - e^{\left(-\frac{m}{\tau_2}\right)} \right) \qquad [4]$$

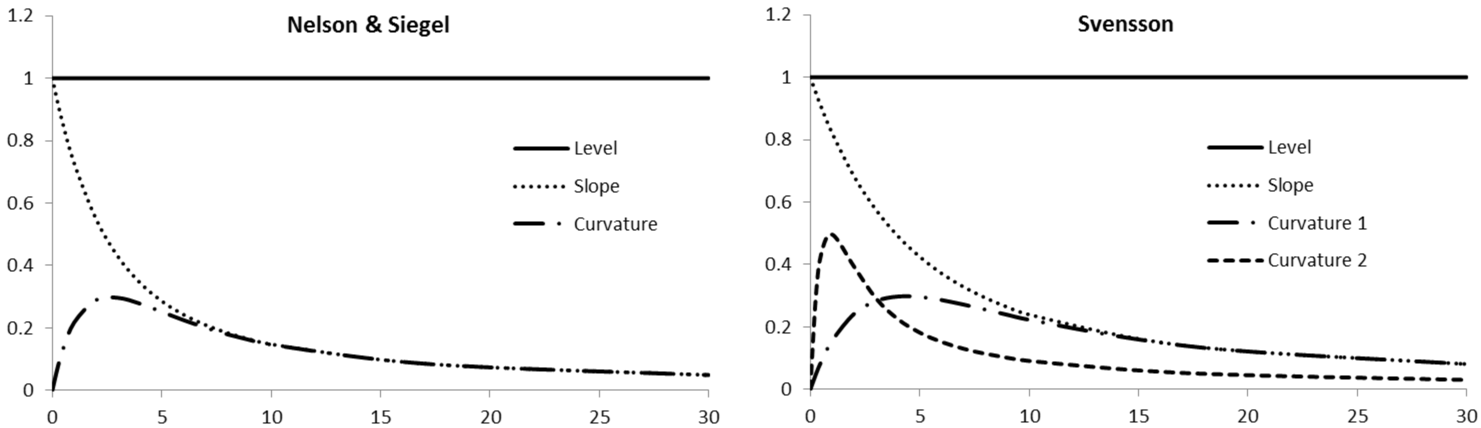
These two parameters provide a better goodness-of-fit, given that the model incorporates a second curvature along the vertex scores. This document follows adjusted version given by De Pooter (2007), in order to avoid multicolinearity problems between the two curvature components, which may occurred during the estimation process. Therefore the zero coupon curve at a particular point is given by:

$$z(m;b) = \beta_0 + \beta_1 \left(\frac{1 - e^{\left(-\frac{m}{\tau_1}\right)}}{\frac{m}{\tau_1}} \right) + \beta_2 \left(\frac{1 - e^{\left(-\frac{m}{\tau_1}\right)}}{\frac{m}{\tau_1}} - e^{\left(-\frac{m}{\tau_1}\right)} \right) + \beta_3 \left(\frac{1 - e^{\left(-\frac{m}{\tau_2}\right)}}{\frac{m}{\tau_2}} - e^{\left(-\frac{2m}{\tau_2}\right)} \right) \qquad [5]$$

The parameter λ, or the parameters λ_1 and λ_2 in the NSS model, defines the decaying shape of the factor loadings, as shown in figure 2[4]. By determining the value of the decay parameter, one defines the weight of each of the latent factors. As presented by De Pooter (2007), Bolder (2006) or Dolan (1999) the level is the long-term component as it is the one constant throughout all the maturities. The investment results over long maturities are dominated by the change in the level of yields, considering that the slope and curvature effects are almost imperceptible. The slope is the short-term component as it starts with a high value but decays at an exponential rate given by λ. The curvature is the medium-term component it starts at a low value and increases for the medium term maturities but then decays to zero in the long-term maturities. The peak of this hump is reached in the value of τ.

[4] The figure shows on the left side the factor loadings given by the NS model for a yield curve assuming a fixed value of $\tau = 1.48\ years$. On the right side it depicts the factor loadings given by the NSS model for a yield curve assuming a value of $\tau_1 = 2.44\ years$ and $\tau_2 = 0.91\ years$.

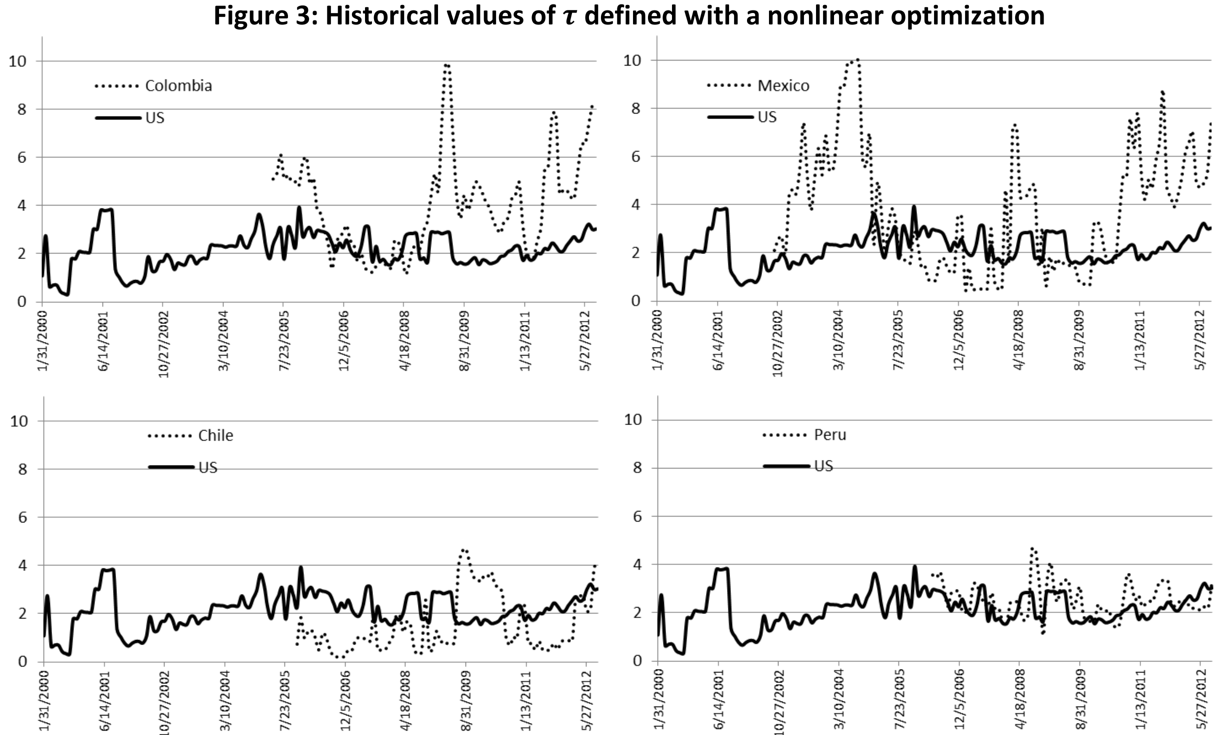
Figure 2: Factor Loadings



Source: Author's calculations

Consequently a change in β_0 creates a parallel shifts up or down, a positive value to β_1 will lead to a steeping of the zero coupon curve, while a negative value will produce a flattening or a downward sloping yield curve. Positive values of β_2 increase the curvature and negative values decrease the curvature. With the NSS model one can make the same analysis, as the three first factors are the same. The fourth factor resembles the third (right sided graph in figure 2) as it is a second hump in the term structure, although it is not related with the slope as it has an independent parameter λ_2. Likewise, positive values of β_3 increase the second curvature and negative values decrease this curvature.

Figure 3: Historical values of τ defined with a nonlinear optimization



Source: Author's calculations

Considering that λ is a nonlinear parameter and numerically unstable, almost all the authors fixed the value of τ, Fabozzi, Martellini & Priaulet (2005) used a value of 3 years, and Diebold & Li (2006) fixed a value of 1.3684 years. Gilli, Große, & Schumann (2010) reported numerical difficulties when calibrating the model, particularly they established two problems. The first one is the optimization issue that results from a not convex problem with multiple local minima. The second one is a collinearity problem, as called by the authors, resulting from the highly correlated factor loadings depending on the value of τ. Consequently the authors advised using differential evolution techniques[5] in order to produce an accurate estimate of the parameters. This approach is tested in this document, but the results were very unstable having an adverse effect in the out of sample estimation process. Figure 3 depicts the changes of τ in the NS model accordingly to the result of the nonlinear optimization. While the parameter of US does not change abruptly throughout the time series, the selected values for the yield curves of Colombia, Mexico and Chile present unexpected jumps that negatively affect the estimation process for the forecast models.

Therefore the value of τ is fixed following the approach of De Rezende & Ferreira (2011), which determined the decay factor by choosing the value that minimizes the average of the Root Mean Squared Error (RMSE) computed for each period m, considering the actual yield curve and the estimates with NS. The optimal value is chosen from a set $\Omega = \left\{l + {}^{r}\!/_{500} * i\right\}_{i=1}^{500}$, where l defines the minimum value given for τ in the estimation with differential evolution heuristic in all the observations n available, and r delineates the range of the possible values of τ. The parameters β of the functional form are estimated by ordinary least squares. Consequently the value of τ will be given by:

$$\hat{\tau} = \underset{\tau \epsilon \Omega}{\mathrm{arg\_min}} \left\{ \frac{1}{N} \sum_{n=1}^{N} \sqrt{\frac{1}{M} \sum_{m=1}^{M} \left(y_m(t_n) - \hat{y}_m(t_n, \lambda, \beta_n)\right)^2} \right\} \qquad [6]$$

In the case of NSS the problem is enhanced to solve for τ_1 and τ_2:

$$(\hat{\tau}_1, \hat{\tau}_2) = \underset{(\tau_1, \tau_2) \epsilon \Lambda}{\mathrm{arg\_min}} \left\{ \frac{1}{N} \sum_{n=1}^{N} \sqrt{\frac{1}{M} \sum_{m=1}^{M} \left(y_m(t_n) - \hat{y}_m(t_n, \lambda_1, \lambda_2, \beta_n)\right)^2} \right\} \qquad [7]$$

where $\Lambda = \{(\tau_1, \tau_2) | \tau_1 \epsilon\, \Omega, \tau_2 \epsilon\, \Omega\}$

The resulting values are shown in Table 1. The value of τ for United States in the NS model is consistent with the one selected by Diebold & Li. For the NS model, small values of τ imply a fast exponential decay that gives more weight to the short term maturities, disregarding the middle and long term maturities. Whilst a large value of τ, produces a slow decay that gives more relevance to the middle and long term

[5] Storn & Price (1997) describe differential evolution is a parallel direct search method which utilizes D-dimensional parameter vectors as a population for each generation G. The initial vector population is chosen randomly and should cover the entire parameter space. Differential evolution generates new parameter vectors by adding the weighted difference between two population vectors to a third vector.

maturities. Likewise, the curvature factor gives more weight to the vertices around the value of τ. Consequently, small values indicate that the historical yield curve has been more flat, while large value of τ, represent steeper historical yield curves. Thus the values shown in Table 1, clearly represent the data depicted in Figure 1. Where Chile present a flat yield curves in most of the cases, whilst Colombia and Mexico have steeper yield curves.

Table 1: Parameter τ optimal values

Country		COLOMBIA	MEXICO	CHILE	PERU	US
Nelson & Siegel	τ_1	3.56	4.70	0.62	2.27	1.48
Nelson & Siegel - Svensson	τ_1	2.80	4.92	2.44	6.69	3.97
	τ_2	2.24	4.36	0.91	6.13	3.39

Source: Author's calculations

Once a fixed value of τ has been selected, the NS or NSS equation is turned into a linear combination of the βparameters, one can estimate the coefficients with an ordinary least squares model.

$$\min_{\beta_0,\beta_1,\beta_2} \sum_{i=1}^{12} \left[z(t,m) - \beta_0 - \beta_1 \left(\frac{1 - e^{\left(-\frac{m}{\tau_1}\right)}}{\frac{m}{\tau_1}} \right) - \beta_2 \left(\frac{1 - e^{\left(-\frac{m}{\tau_1}\right)}}{\frac{m}{\tau_1}} - e^{\left(-\frac{m}{\tau_1}\right)} \right) \right]^2 \qquad [8]$$

Table 2 presents the statistics of the resulting coefficients β (level, slope and curvature(s)). The mean of the historical value of the level shows a higher historical rate for the Latin American countries compared to the rate of United States. The difference between the two values represents the default risk premium of the emerging countries; this is higher for Colombia, Mexico and Peru, whilst for Chile the difference is of just 100 basis points.

Moreover, the NS model and the NSS model present similar values for the mean of the level and the slope in all the markets, but the mean value of the curvature in the NS model is not always similar to any of the mean values of the curvatures shown in the NSS model. In the NS model the curvature is also the factor that presents the greater standard deviation for all the countries, in the NSS model the first curvature is the most unstable coefficient. Furthermore, the kurtosis and skewness of the time series of the latent factors are distant from a normal distribution assumption. Regarding the time series autocorrelations, the first sample autocorrelations are high, close to 1, for all the time series. The sixth sample autocorrelations are lower than the first sample autocorrelations, but higher than 0.5 for all the time series. The twelfth sample autocorrelations are low, closely to 0 (with the exception of US), indicating the absence of a seasonal yearly effect.

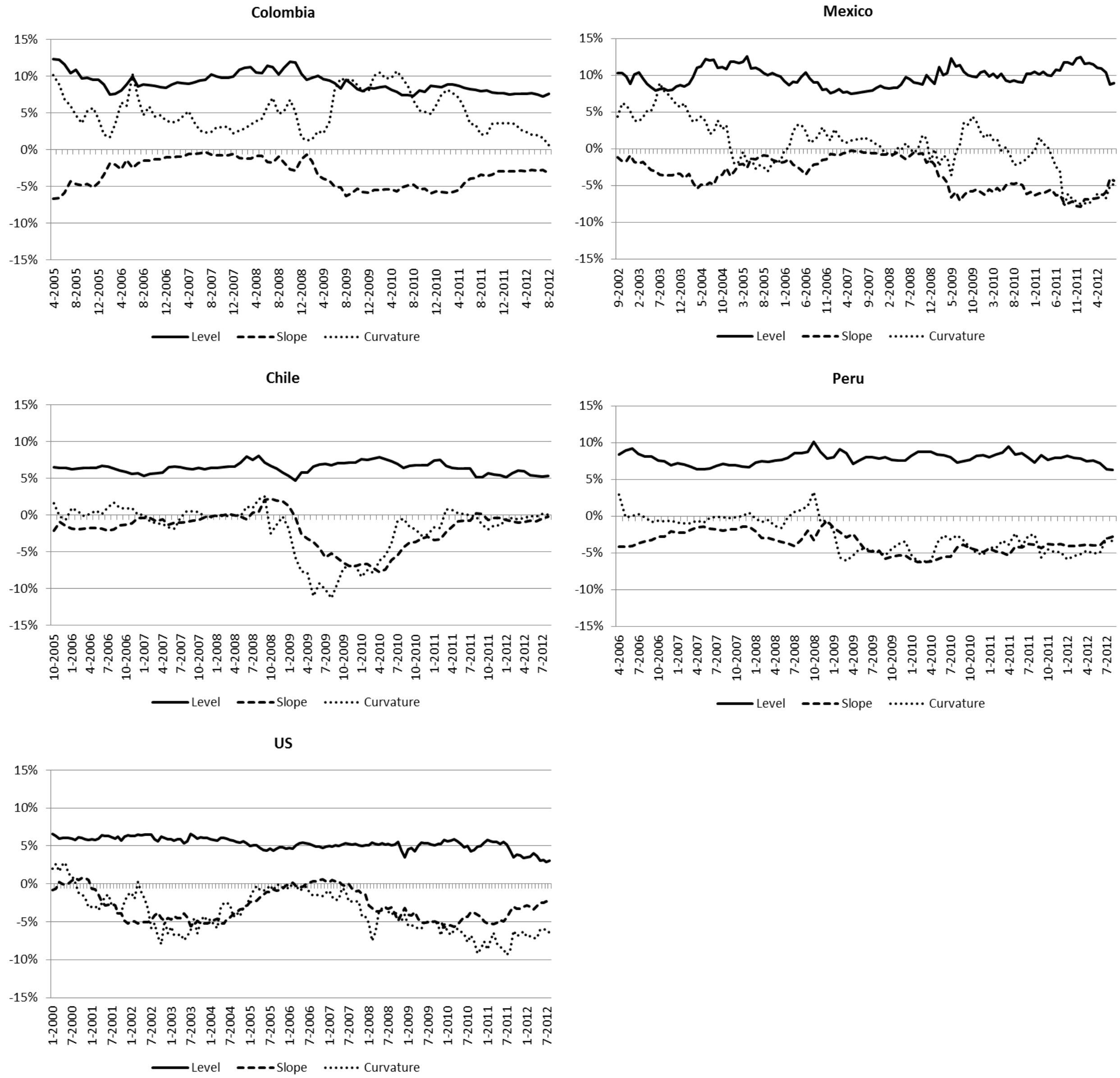
Table 2: Statistics of the latent factors of the yield curve

		Nelson & Siegel			Nelson & Siegel - Svensson			
		Level (β_1)	Slope (β_2)	Curvature (β_3)	Level (β_1)	Slope (β_2)	Curvature 1 (β_3)	Curvature 2 (β_4)
COLOMBIA	Mean	9.086%	-3.256%	5.111%	9.159%	-3.072%	5.939%	-1.655%
	Standard Deviation	1.279%	1.926%	2.666%	1.271%	1.550%	4.686%	2.857%
	Skewness	0.677	-0.082	0.518	0.555	-0.525	0.863	-0.782
	Kurtosis	-0.256	-1.433	-0.796	-0.543	-0.441	-0.316	-0.134
	Minimum Value	7.217%	-6.727%	0.552%	7.263%	-6.909%	0.053%	-9.499%
	Maximum Value	12.333%	-0.349%	10.669%	11.970%	-0.205%	17.423%	2.393%
	1^{st} sample Autocorrelation	0.912	0.956	0.869	0.914	0.937	0.927	0.929
	6^{th} sample Autocorrelation	0.659	0.782	0.256	0.616	0.639	0.627	0.638
	12^{th} sample Autocorrelation	0.238	0.398	0.180	0.241	0.426	0.245	0.147
MEXICO	Mean	9.802%	-3.409%	0.527%	9.802%	-3.411%	0.636%	0.036%
	Standard Deviation	1.331%	2.233%	3.610%	1.362%	2.392%	4.662%	1.681%
	Skewness	0.147	-0.246	-0.316	0.102	-0.425	0.168	0.085
	Kurtosis	-0.906	-1.294	0.043	-1.082	-1.197	-0.117	0.079
	Minimum Value	7.472%	-7.852%	-7.877%	7.477%	-8.094%	-9.545%	-4.370%
	Maximum Value	12.612%	-0.200%	8.966%	12.697%	0.140%	12.511%	5.150%
	1^{st} sample Autocorrelation	0.886	0.965	0.942	0.869	0.956	0.891	0.765
	6^{th} sample Autocorrelation	0.553	0.812	0.712	0.524	0.829	0.704	0.217
	12^{th} sample Autocorrelation	0.111	0.546	0.411	0.090	0.627	0.497	0.190
CHILE	Mean	6.388%	-1.942%	-1.821%	6.297%	-1.979%	1.607%	-0.041%
	Standard Deviation	0.734%	2.399%	3.332%	0.673%	2.211%	2.295%	1.909%
	Skewness	0.068	-0.892	-1.413	0.245	-0.773	0.938	-0.808
	Kurtosis	-0.381	0.092	0.928	0.172	-0.073	0.647	0.496
	Minimum Value	4.706%	-7.761%	-11.322%	4.906%	-7.213%	-2.896%	-5.368%
	Maximum Value	8.042%	2.206%	2.492%	8.213%	2.458%	7.277%	4.071%
	1^{st} sample Autocorrelation	0.867	0.969	0.933	0.847	0.963	0.948	0.876
	6^{th} sample Autocorrelation	0.239	0.671	0.610	0.189	0.658	0.636	0.454
	12^{th} sample Autocorrelation	0.026	0.104	0.124	-0.045	0.138	0.013	-0.106

PERU	**Mean**	7.852%	-3.595%	-2.543%	7.048%	-2.789%	5.477%	-1.476%
	Standard Deviation	0.772%	1.375%	2.374%	1.444%	1.068%	7.150%	3.164%
	Skewness	0.153	0.046	0.239	-0.603	0.149	0.878	-0.339
	Kurtosis	0.197	-0.754	-0.970	2.289	0.471	0.400	-0.417
	Minimum Value	6.301%	-6.224%	-6.134%	2.066%	-5.107%	-6.923%	-8.776%
	Maximum Value	10.174%	-0.563%	3.281%	11.568%	0.647%	26.297%	6.453%
	1st sample Autocorrelation	0.779	0.920	0.888	0.846	0.722	0.885	0.869
	6th sample Autocorrelation	0.283	0.582	0.547	0.339	0.373	0.520	0.529
	12th sample Autocorrelation	-0.099	0.262	0.511	-0.295	0.086	0.094	0.260
US	**Mean**	5.306%	-2.935%	-3.824%	5.336%	-3.076%	1.831%	-1.358%
	Standard Deviation	0.801%	2.031%	2.803%	0.637%	2.016%	2.292%	1.400%
	Skewness	-0.871	0.461	0.171	-0.589	0.604	0.220	0.246
	Kurtosis	0.665	-1.233	-0.811	-0.015	-1.180	-0.586	-0.649
	Minimum Value	2.904%	-5.763%	-9.323%	3.307%	-5.566%	-2.561%	-4.844%
	Maximum Value	6.580%	0.787%	2.778%	6.493%	0.788%	9.432%	2.073%
	1st sample Autocorrelation	0.925	0.978	0.940	0.881	0.980	0.867	0.862
	6th sample Autocorrelation	0.684	0.843	0.758	0.593	0.879	0.562	0.547
	12th sample Autocorrelation	0.480	0.525	0.580	0.451	0.599	0.351	0.367

Source: Author's calculations

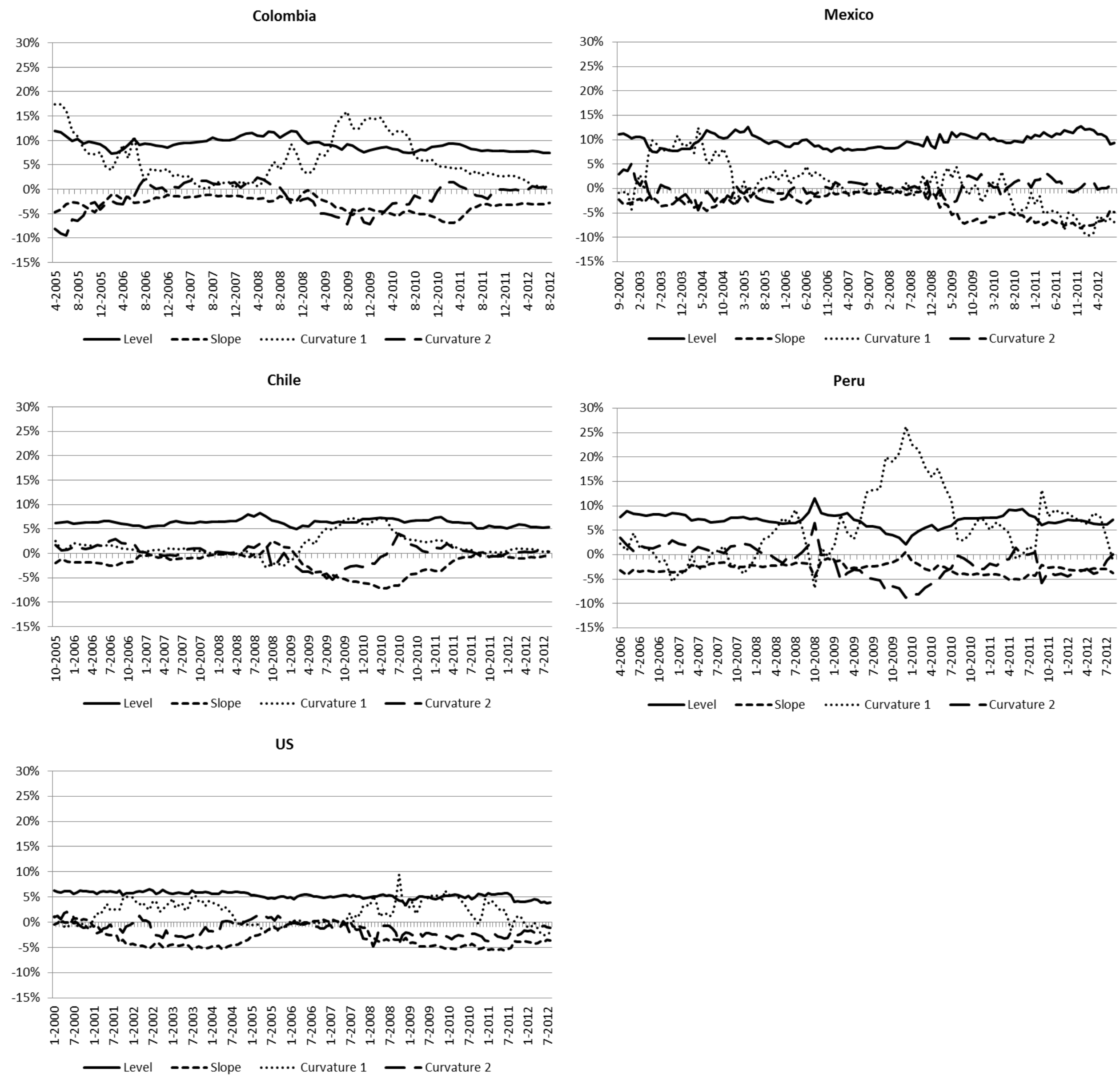
Figure 4: Time series of Nelson-Siegel factors



Source: Author's calculations

Figure 4 depicts the time series of the latent factors of the countries with the NS model following the fixed values of τ, as shown in table 1. As expected, from the information in table 2, the curvature is the most unstable factor in all the yield curves. Furthermore, the level is the most constant factor, although it does not seems to be stationary. A Dickey-Füller test of stationarity, indicate that all the time series have a unit root. After differentiating the series once, all the factors appear to be stationary. Figure 5 shows the time series of the latent factors of the NSS model, again the first curvature is the factors that is more unstable throughout the sample period. As in the NS model, all the factors are an $I(1)$ process. Some differentiated time series show little autocorrelation, indicating that the autoregressive models may not have good predictive power. Although, other differentiated series present higher autocorrelation that allow $AR(1)$ models to be fitted to the data. Consequently, the most adverse effect may occur in the out-sample performance of the vector autoregressive models, because they would include some time series behaving as a random walk that can alter the predictions.

Figure 5: Time series of Nelson-Siegel factors



Source: Author’s calculations

4. Forecasting Methods

By fixing the value of the parameters τ, the resulting yield curve in period t with the NS or NSS model can be described as a linear combination of the latent factors of the yield curve in the following representation:

$$Y_t = X_t\beta_t + \varepsilon_t \qquad [9]$$

where Y_t is a vector that contains every chosen vertex of the yield curve, X_t represents the matrix of the factor loadings for all the selected vertices, β_t is the vector of coefficient of the factors and ε_t is a vector of error with a normal standard distribution.

Consequently one can determine the future yield curve forecasting the value of the vector β_t. The forecasting power of two approaches is tested in this document: parametric models and ANN. Both methods are tested for each factor individually and taking into account the interaction of all the factors.

4.1. Parametric Models

Following Diebold & Li (2006) the parametric models chosen to forecast the level, slope and curvature(s) are an $AR(1)$ and a $VAR(1)$. With the $AR(1)$ one model is estimated for each factor, resulting in three for the NS functional form and four for the NSS functional form.

$$\hat{\beta}_{f,t} = \mu_f + \phi_f\hat{\beta}_{f,t-1} + \upsilon_{f,t} \qquad [10]$$

The historical information of each factor, excluding the data left for the forecast evaluation, is used to estimate the parameters μ_f and ϕ_f; $\upsilon_{f,t}$ represents the residual.

Alternatively, the $VAR(1)$ groups all the factors to take into account the interaction between all the states variables.

$$\beta_t = \mathrm{M} + \Phi\beta_{t-1} + \Upsilon_t \qquad [11]$$

Diebold & Li (2006) reported inferior results with the $VAR(1)$ forecasts compared to the ones given by the $AR(1)$. They give two reasons for this: i) the first one explains that this conclusion is consistent with the observed tendency in macroeconomics, where unrestricted $VARs$ tend to produce poor forecasts of economic variables even when there is important cross-variable interaction, this is a result of the in-sample over-fitting that occurs for the numerous parameters that need to be estimated. ii) The second one is that they found that the factors are not highly correlated; therefore the multivariate model result in as a set of univariate models. These reasons also explain why the authors avoided functional forms with more than three latent factors. The authors also reported outstanding results for the $AR(1)$ model in a twelve-month-ahead forecast, whilst the results at one month horizon forecast did not show a significant difference between the RMSE of the model and the one of the random walk.

Likewise, Bolder (2006) outlined that the forecasts of the NS model outperform exponential-spline and Fourier series models on a six month horizon, but the one and three-month ahead forecasts of the random walk always exceeded the other methodologies. Moreover, De Rezende & Ferreira (2011) studied the effect of a quantile autoregression approach with which they conclude that the Bliss functional form delivered the best results in a one month horizon, whilst the NSS-QAR model is the best option for the three months horizon. When the authors tested a five factor functional form it performs poorly, due to sample over-fitting. Furthermore, De Pooter (2007) established that more flexible models, as the NSS, improve the out-of-sample predictability, his adjusted version of the NSS model with an $AR(1)$ parametric model outperformed other choices in the six and twelve month horizons. Additionally, Cziráky (2007) concluded that the NS and the NSS models had poor forecasting performance around the points of non-parallel shifts.

4.2. Neural Networks

Artificial neural networks are an analogy of the organization of neurons in the human brain. Thus, are formed by incoming nodes or neurons, which are the equivalent of the stimuli received by the human brain. These are followed by hidden or intermediate nodes that process the information, which communicate with an output layer that produce the solution or response to the incoming information. As with the human brain, the ANN learns to perform its task by trial and error.

In constructing forecasts of a random variable $(\hat{\beta})$ ANN uses the information of past observations as follows:

$$\hat{\beta}_t = f\left(\beta_{t-1}, \beta_{t-2}, \dots, \beta_{t-p}, W\right) + \epsilon_t \qquad [12]$$

where f is the function that determines the structure and connections of the nodes and their respective weights W.

With ANN one can identify the non-linearity of the time series, as this tool can approximate almost every conceivable function, consequently ANN can be viewed as a notable option for forecasting the latent factors of the yield curves. Two different approaches are analyzed, the first one considers each time series separately, whilst the second analyzed all the factors that describe the yield curve.

In the literature, the opinions of the advantages of ANN over parametric models are divided, though more biased towards models that consider nonlinear components. The comparison made by Hill, O'Connor, & Remus (1996) concluded that ANN significantly exceeds the usual statistical models in assessing quarterly and monthly data. Other authors like Deo & Sridhar (1999)[6], Abdel-Aal (2008)[7], Zou,

[6] The main objective of the model developed by the authors is to forecast the height of the waves on the coast Yaman (India).
[7] The author sought the best tool to forecast the demand for energy.

Xia Yang, & Wang (2007)[8], and Misas, López & Querubín (2002)[9] obtained favorable results for ANN that establish their superiority over linear models.

However, Faraway & Chatfield (1998) criticized the adjustment of ANN to some of the univariate series used in the article "Time series Forecasting analysis and control" of Box, Jenkins & Reinsel, supporting the use of ARIMA models. Similarly Callen, Kwan, Yip, & Yuan (1996) reported that the linear models outperformed the ANN, when forecasting financial returns of the shares of the New York Stock Exchange.

According to McNelis (2005) a linear model may be a very imprecise approximation to the real world, but it gives very easy, quick, and exact solutions. In contrast, the ANN produces a precise approximation capturing nonlinear behavior, but it does not have exact, easy-to-obtain solutions. Furthermore, Hamzacebi, Akay, & Kutay (2009), when analyzing different series, concluded that the best tool for forecasting is heavily dependent on the analyzed time series.

The ANN developed in this document are feed-forward networks, which defines an organization where no node gives information to any of its predecessors. Therefore, the n lags will be in charge of activating the m hidden nodes (O) according to a previously determined activation function (g). These in turn activate the output layer nodes (S) with another function (f) to produce a result. Mathematically the salient nodes are modeled according to the following equation:

$$S_t = f\left(\gamma_0 + \sum_{i=1}^{m} \gamma_i\, g\left(\alpha_{i0} + \sum_{j=1}^{n} \alpha_{ij}\beta_j\right)\right) \qquad [13]$$

The inclusion of each outgoing node requires $m + 1$ additional parameters, given the hidden nodes. Therefore more outputs require more network nodes according to settings in the intermediate layer, without being affected by the neurons in the input layer. Furthermore, according to Kuan & Liu (1995) functions f and g can be chosen arbitrarily, although it is suggested a bounded g function. Most of the literature recommends a sigmoidal logistic function:

$$g(w) = \frac{1}{1 + e^{-w}} \qquad [14]$$

Once one has defined the network structure, the next step is the learning process of the ANN; this process is done by means of a minimization problem.

The process minimizes the sum of squared errors to find the optimal weights of the nodes in the middle and output layer:

---

[8] The authors conducted a comparison of ARIMA models and ANN while forecasting the price of wheat in the Chinese market.

[9] The authors used models to forecast the movements of the inflation in Colombia.

$$min_W \Psi(W) = min_W \sum_{t=1}^{T} \left(\beta_t - \widehat{\beta_t}\right)^2 \qquad [15]$$

In the previous equation W is a matrix formed by the vectors of the weights of the hidden and the output layers, Ψ is the loss function given by the sum of squared errors, T is the number of observations and $(\widehat{\beta_t})$ is given by the structure defined in equation $[12]$. Given that $\Psi(W)$ is a nonlinear function, iterative processes must be performed from point W_0 to find the global minimum. Generally this search is done by means of a base gradient, where $\Psi'(W)$ and $\Psi''(W)$ are updated with different values of W, until reaching the minimum value of Ψ. The more applied procedure, backpropagation, establishes that that the movement direction from the initial weight is given by the gradient ∇_0 and the learning rate ρ:

$$(\mathrm{W}_1 - \mathrm{W}_0) = -\rho_0 \nabla_0 \qquad [16]$$

An important caveat that should be noted is that the learning process[10] is not always the absolute best as it can stop at a local optimum. To improve the search of the global optimum is usually suggested the inclusion of a momentum that ensures that the learning rate facilitates the convergence of the weights. Although one must consider that a bad choice of these parameters may cause over-fitting problems.

Construction of the Neural Networks:

For the construction of the ANN this document follows the steps given by Kaastra & Boyd (1996): i) Variables selection; ii) Data Collection; iii) Data preprocessing; iv) Training, testing and validation sets; v) Neural networks paradigms; vi) Evaluation criteria; vii) Neural Network Training; viii) Implementation.

The selected variables in the neural networks constructed in this paper are given by number of lags that determined the behavior of the network. Considering that two approaches are used, the forecasts are explained by the past values of the same factor or/and the past values of the complementary factors. The selection process is explained in detail in the section that defines the number of layers and nodes. Moreover, the second step was explained in the second section of this article. For the third step, the data preprocessing is performed by normalizing the time series, in order to establish a range between -1 and 1. This will ensure a faster training, due to the avoidance of very high or very low values that difficult the convergence to the global optimum.

The fourth step defines how to divide the data to ensure the network learning. 60% of the information is used in the training, whilst 20% is used to validate the forecasts of the network. The training is given if the sum squared error of the validation data is reduced. The remaining 20% is used to test the effectiveness of forecasts with data that have never been used in the learning process.

---

[10] Other authors mention other forms of training. On the one hand McNelis (2005) suggests a stochastic search with genetic algorithms or simulated annealing. On the other hand Deo Naidu & Sridhar (1999) mentioned a cascade correlation algorithm. It should be noted that these methods are meta-heuristics that do not guarantee a global optimum.

The fifth step is one of the most important and yet the most ambiguous, considering the different points of view, either depending on the type of network or the time series used. At this stage one must define the structural form:

- Input Layer: One must ensure that the number of nodes corresponds to the number of lags that define the future observations. The number of units is defined by evaluating different intervals to find the one that provided a better evaluation criterion. The largest lag window considered for all the time series is the first to the twelfth.
- Hidden layers: Kaastra & Boyd (1996) showed that more than two middle layers produced unsatisfactory results. This paper uses only one hidden layer, as is proved empirically that an additional layer gave no additional benefits. Therefore the relevant parameter in this case is the number of nodes. Some authors mentioned that twice the incoming nodes plus one or five achieved adequate accuracy in the model, see Coakley & Brown (2000) or Figueiredo, Hall Barbosa, Da Cruz, Vellasco, Pacheco, & Contreras (2007). But Hamzacebi, Akay, & Kutay, (2009) emphasized that a significant magnitude of nodes leads to inadequate results in the optimization. Likewise, a large number of hidden units increase the probability that the parameters converge to a local optimum. Therefore this document follows Misas, López & Querubín (2002) approach, where the number of nodes is selected by iterating between different values, not exceeding the double of the maximum number of outgoing nodes.
- Output layer: This layer produces the forecasting results, therefore is closely related to the forecast horizon. Consequently, the number of nodes is defined by the forecasted periods.

Once the structural form of the neural network is defined one must determine the evaluation criteria that establish the accuracy of the forecasts and subsequently the best neural network for the time series. The most used measures are MAPE (mean absolute percentage error) and RMSE (root mean squared error). This paper follows the suggestion McNelis (2005), to use the Hannan-Quinn information criterion, which penalizes the model with more parameters (k) in a more strictly way than AIC, but not as severely as with Schwartz criteria.

$$hqif = \left\{ln\left(\sum_{t=1}^{N}\frac{\left(\beta_t-\hat{\beta}_t\right)^2}{N}\right)\right\}+\frac{k[ln\{ln(N)\}]}{N} \qquad [17]$$

The criteria used to compare the forecasts of the term structure of interest rates, given by the different methodologies is RMSE.

The seventh step defines the training process of the neural network. This document defines the following conditions:

- A sigmoidal logistic function that describes the nonlinear relationship between the input layer and the hidden layer. And a linear function that relates the intermediate layer and the output layer.
- A backpropagation methodology for the learning process.
- To prevent convergence to a local optimum a momentum is added in the learning algorithm.

- The initialization of the weights is given by the Nguyen-Widrow algorithm, where the weights are defined in an equitable manner to the entire space of the next layer. This technique main benefit is that it uses the information of almost all nodes in the layer and ensures faster training.
- As for the stopping criteria during the training, the following conditions should be reached:
  - 1000 iterations.
  - The sum of squared errors reaches a value of 1.00E-05.
  - The magnitude of the gradient is less than 1.00E-10.
  - The sum of quadratic errors of the training data is lower than the sum of quadratic errors of the validation data.

Finally implementation techniques are defined. As previously mentioned, several networks are constructed, for each time series and for each yield curve. The variable values are the incoming and hidden nodes. Therefore, a comparison between networks is made using the Hannan-Quinn information criterion, considering the forecasts of the in-sample data; this is done to choose the best network with which forecasts of the unused data are made. Another aspect to be considered is that neural networks with the same structure can produce different results because a heuristics is used in the training process. Therefore in this paper the parameters of the best neural network are estimated several times, and the chosen value is the average of 1000 simulations.

Table 3 presents the number of nodes in the input and hidden layers, given by the iterative process mentioned above. In most cases, the ANN of the whole set of factors (Set) requires more input nodes, or time series lags, than a neural network of an individual factor. The minimum value of nodes in an input layer is two, and the maximum value is twelve. In the hidden layer the lowest value is two, whilst the highest value is twenty. It is important to note that the output layer contains twelve nodes in all the constructed neural networks.

Table 3: Resulting nodes in the input and hidden layers

Country	Layer	Nelson & Siegel				Nelson & Siegel - Svensson				
		Level	Slope	Curvature	Set	Level	Slope	Curvature 1	Curvature 2	Set
Colombia	Input	8	12	6	11	8	8	11	11	4
	Hidden	4	5	14	19	3	16	20	2	17
Mexico	Input	10	4	11	11	9	5	9	12	12
	Hidden	4	12	13	20	13	9	13	9	12
Chile	Input	9	10	6	11	12	6	8	7	12
	Hidden	17	15	5	20	18	20	9	6	16
Peru	Input	11	5	4	10	9	8	12	8	10
	Hidden	11	14	18	20	15	19	4	4	18
US	Input	8	7	11	11	2	12	11	10	5
	Hidden	8	11	15	4	16	10	15	7	18

Source: Author’s calculations

5. Results

In this section the forecasting results of the entire zero-coupon curve are shown, in order to review the describing dynamics of the whole term structure of interest rates and not to focus in particular vertices. Two functional form of the term structure are considered: i) the NS model which defines that the level, the slope and the curvature are the factors that determine the yield curve; and ii) the NSS model, which form if equal to the NS model but enhanced with a fourth factor that consists of a second curvature. Additionally, four different forecasting methodologies are used: i) $AR(1)$ models for each individual factor; ii) $VAR(1)$ model that forecasts all factors to take into account their interaction; iii)Neural networks that forecast each factor individually ($ANN\ Ind.$); and iv)Neural Networks that forecast all the factors at the same time ($ANN\ All$). It is also important to note that the forecast of the parametric models require an iterative process to forecast higher horizons than one month, whilst the ANN forecasts the selected horizon in one trial, avoiding the sum of errors that can result in a recursively process. To check the forecasting performance of the models they are compared with the alternative of forecasting with the current value (random walk- RW), consequently each vertex m on the yield curve in time t will be given by its value at time $t-1$: $\hat{Y}_{m,t} = Y_{m,t-1}$.

Considering the short amount of data in some of the selected countries only a forecast up to six months is studied, by leaving the last one and a half year of each yield curve data for out-of-sample evaluation. The standard forecast error evaluation that is in in this document is the Root Mean Squared Error (RMSE), by averaging the RMSE of each vertex on the zero-coupon curves.

Table 4 presents the standard deviation of the RMSE disaggregated by the forecasting method, the countries and forecasting horizons. The results are diverse for every country. In Colombia the $NS - ANN\ Ind.$ forecasts present the lowest variation in four of the studied forecasting horizons, as it also happens in the case of Peru. Alternatively, in Mexico the $NSS - ANN\ All$ method present the smallest standard deviation in all time horizons with the exception of one month, where the $NS - ANN\ Ind.$ varies the least. In Chile and the US there is no particular forecasting tool that stands out for having the smallest variation in all the forecasting horizons, but is notable that the RW option has not the minimum standard deviation in any of the forecasting horizons.

Table 4: Standard Deviation of the RMSE of the different methodologies in the analyzed countries

		Nelson & Siegel				Nelson & Siegel - Svensson				RW
	Forecasting Horizon	ANN Ind.	ANN All	AR(1)	VAR(1)	ANN Ind.	ANN All	AR(1)	VAR(1)	
COLOMBIA	1 Month	0.056%	0.573%	0.133%	0.150%	0.062%	0.335%	0.120%	0.130%	0.131%
	2 Month	0.068%	0.627%	0.270%	0.141%	0.256%	0.388%	0.261%	0.190%	0.141%
	3 Month	0.056%	0.729%	0.288%	0.225%	0.199%	0.420%	0.284%	0.257%	0.139%
	4 Month	0.196%	0.797%	0.282%	0.248%	0.201%	0.405%	0.280%	0.267%	0.192%
	5 Month	0.221%	0.806%	0.270%	0.249%	0.189%	0.403%	0.269%	0.257%	0.209%
	6 Month	0.189%	0.818%	0.281%	0.262%	0.210%	0.413%	0.280%	0.262%	0.255%

	Forecasting Horizon	ANN Ind.	ANN All	AR(1)	VAR(1)	ANN Ind.	ANN All	AR(1)	VAR(1)	RW
MEXICO	1 Month	0.042%	0.112%	0.132%	0.365%	0.109%	0.090%	0.135%	0.205%	0.207%
	2 Month	0.147%	0.113%	0.186%	0.142%	0.237%	0.112%	0.215%	0.144%	0.247%
	3 Month	0.234%	0.120%	0.178%	0.240%	0.192%	0.099%	0.214%	0.231%	0.250%
	4 Month	0.266%	0.136%	0.178%	0.260%	0.242%	0.088%	0.224%	0.251%	0.290%
	5 Month	0.368%	0.144%	0.143%	0.242%	0.382%	0.099%	0.382%	0.236%	0.371%
	6 Month	0.477%	0.198%	0.148%	0.255%	0.394%	0.111%	0.252%	0.250%	0.411%
CHILE	Forecasting Horizon	ANN Ind.	ANN All	AR(1)	VAR(1)	ANN Ind.	ANN All	AR(1)	VAR(1)	RW
	1 Month	0.085%	0.135%	0.106%	0.279%	0.069%	0.247%	0.107%	0.292%	0.293%
	2 Month	0.143%	0.247%	0.105%	0.074%	0.125%	0.239%	0.107%	0.073%	0.240%
	3 Month	0.107%	0.321%	0.123%	0.183%	0.250%	0.213%	0.146%	0.191%	0.251%
	4 Month	0.170%	0.400%	0.161%	0.233%	0.300%	0.202%	0.190%	0.237%	0.250%
	5 Month	0.205%	0.457%	0.196%	0.257%	0.357%	0.219%	0.220%	0.258%	0.383%
	6 Month	0.193%	0.218%	0.175%	0.232%	0.395%	0.154%	0.194%	0.233%	0.290%
PERU	Forecasting Horizon	ANN Ind.	ANN All	AR(1)	VAR(1)	ANN Ind.	ANN All	AR(1)	VAR(1)	RW
	1 Month	0.059%	0.258%	0.105%	0.090%	0.343%	0.367%	0.102%	0.173%	0.173%
	2 Month	0.063%	0.366%	0.284%	0.297%	0.429%	0.315%	0.282%	0.300%	0.201%
	3 Month	0.153%	0.280%	0.301%	0.304%	0.596%	0.410%	0.292%	0.307%	0.206%
	4 Month	0.205%	0.424%	0.304%	0.293%	0.597%	0.338%	0.292%	0.297%	0.227%
	5 Month	0.274%	0.328%	0.288%	0.267%	0.382%	0.365%	0.272%	0.271%	0.249%
	6 Month	0.228%	0.295%	0.260%	0.227%	0.588%	0.323%	0.244%	0.231%	0.211%
US	Forecasting Horizon	ANN Ind.	ANN All	AR(1)	VAR(1)	ANN Ind.	ANN All	AR(1)	VAR(1)	RW
	1 Month	0.109%	0.038%	0.100%	0.164%	0.154%	0.156%	0.117%	0.276%	0.413%
	2 Month	0.195%	0.085%	0.172%	0.158%	0.257%	0.178%	0.181%	0.137%	0.266%
	3 Month	0.266%	0.123%	0.170%	0.112%	0.373%	0.159%	0.180%	0.108%	0.305%
	4 Month	0.279%	0.136%	0.167%	0.097%	0.401%	0.125%	0.178%	0.102%	0.331%
	5 Month	0.376%	0.132%	0.175%	0.109%	0.391%	0.109%	0.188%	0.116%	0.351%
	6 Month	0.480%	0.088%	0.187%	0.122%	0.450%	0.151%	0.202%	0.121%	0.271%

Source: Author’s calculations

Table 5 shows the mean RMSE broken down by the forecasting options, the studied counties and the forecasting horizons tested. The shaded frames indicate the smallest RMSE, the best forecasting option. Further, if the numbers are bold it represents that the null hypothesis of the Diebold–Mariano test is rejected with a 90% confidence. The null hypothesis is that the factor model forecast with the minimum measure RMSE and the random walk RMSE are the same.

In the case of Colombia the $NS - ANN\ Ind.$ forecast model stands out in the short-term horizons (from one to three months), but only the one-month-ahead prediction is significantly different from the result of the random walk. The RW produces the best forecasting results in the four and six months horizons, discrediting the functionality of the tested models. On the contrary, in the five months horizon the $NSS - ANN\ Ind.$ methodology produces a significantly better forecast than the other options. The A$NN\ All$ and the $VAR(1)$ are not reliable forecasting methodologies, indicating that the numerous parameters may generate in-sample over-fitting.

Concerning the forecasting results of the term structure of the interest rates in Mexico the $NS - ANN\ Ind.$ methodology has a significantly improved performance than any other option in the one month prediction, excepting the $NS - AR(1)$ which has a similar RMSE. Unlike with the yield curve of Colombia, the $NS - ANN\ All$ and the $NSS - ANN\ All$ forecasting results improve radically. The two-months-ahead RMSE is superior with the $NS - ANN\ All$, whilst the results at a four, five and six months horizons are better with the $NSS - ANN\ All$ at the ten percent level.

As for the results of the zero-coupon curve of Chile, the $NS - ANN\ Ind.$ methodology delivers the lowest RMSE for the short-term horizons (one to three months) and for the five months horizon. In the short-term horizons the mentioned tool exceeds the other option according to the Diebold–Mariano test. The $NSS - ANN\ All$ presents the best six-months-ahead forecasting result, but its RMSE is not significantly different from that of the RW. Moreover, in the four-months-ahead forecasting result neither forecasting methodology emerges as the best option.

The $NS - ANN\ Ind.$ is once again the forecasting methodology with the lowest RMSE for the short-term horizons in the results of the yield curve of Peru. This ranks it as the best model one, two and three months ahead at the ten percent level, although its RMSE in a time horizon of one month is very similar to that one of the $NS - AR(1)$. In the four months horizon $NS - ANN\ Ind.$ also shows the lowest RMSE, but it is not significantly different from the random walk option. Likewise, in the forecasts five and six months ahead there is not a methodology that exceeds the RW with a significant value.

Finally in the yield curve of the US, there are outstanding results with the $NS - ANN\ All, NS - AR(1), NSS - AR(1)$ and the $NSS - VAR(1)$; but the $NSS - ANN\ Ind.$ outperforms all of them in the one-month-ahead forecast. However, over all the forecasting horizons the $NS - ANN\ All$ is the best options, as it has the lowest RMSE for the forecasting horizons from two months to six months. In the two, five and six months ahead forecast it is significantly different from the random walk at a 90% confidence.

Table 5: Mean RMSE of the different methodologies in the analyzed countries

		Nelson & Siegel				Nelson & Siegel - Svensson				RW
	Forecasting Horizon	ANN Ind.	ANN All	AR(1)	VAR(1)	ANN Ind.	ANN All	AR(1)	VAR(1)	
COLOMBIA	1 Month	**0.214%**	1.399%	0.349%	0.748%	0.274%	1.061%	0.312%	0.277%	0.277%
	2 Month	0.281%	1.570%	0.585%	1.105%	0.483%	1.196%	0.516%	1.442%	0.307%
	3 Month	0.320%	1.717%	0.809%	1.467%	0.529%	1.413%	0.704%	1.695%	0.343%
	4 Month	0.543%	1.849%	1.031%	1.591%	0.506%	1.546%	0.903%	1.739%	0.381%
	5 Month	0.740%	1.956%	1.233%	1.663%	**0.394%**	1.714%	1.091%	1.748%	0.421%
	6 Month	0.853%	1.937%	1.396%	1.708%	0.446%	1.791%	1.246%	1.744%	0.434%
MEXICO	Forecasting Horizon	ANN Ind.	ANN All	AR(1)	VAR(1)	ANN Ind.	ANN All	AR(1)	VAR(1)	RW
	1 Month	**0.264%**	0.377%	0.278%	1.441%	0.275%	0.474%	0.281%	0.434%	0.434%
	2 Month	0.490%	**0.382%**	0.447%	1.503%	0.395%	0.455%	0.417%	1.507%	0.498%
	3 Month	0.776%	0.387%	0.495%	2.229%	0.408%	0.422%	0.463%	2.105%	0.485%
	4 Month	0.912%	0.436%	0.503%	2.585%	0.609%	**0.414%**	0.489%	2.398%	0.524%
	5 Month	1.020%	0.430%	0.487%	2.758%	1.220%	**0.411%**	1.220%	2.561%	0.607%
	6 Month	1.161%	0.448%	0.504%	2.819%	1.355%	**0.448%**	0.554%	2.639%	0.639%
CHILE	Forecasting Horizon	ANN Ind.	ANN All	AR(1)	VAR(1)	ANN Ind.	ANN All	AR(1)	VAR(1)	RW
	1 Month	**0.146%**	0.421%	0.255%	1.447%	0.214%	0.553%	0.263%	0.415%	0.416%
	2 Month	**0.253%**	0.593%	0.339%	0.536%	0.352%	0.514%	0.364%	0.536%	0.430%
	3 Month	**0.385%**	0.679%	0.399%	0.681%	0.561%	0.489%	0.444%	0.695%	0.416%
	4 Month	0.543%	0.663%	0.454%	0.767%	0.666%	0.484%	0.515%	0.779%	0.426%
	5 Month	0.408%	0.578%	0.494%	0.773%	0.719%	0.455%	0.569%	0.777%	0.413%
	6 Month	0.410%	0.543%	0.500%	0.720%	0.686%	0.398%	0.583%	0.718%	0.410%
PERU	Forecasting Horizon	ANN Ind.	ANN All	AR(1)	VAR(1)	ANN Ind.	ANN All	AR(1)	VAR(1)	RW
	1 Month	**0.235%**	0.491%	0.271%	0.672%	0.376%	1.247%	0.285%	0.404%	0.404%
	2 Month	**0.233%**	0.559%	0.554%	1.810%	0.702%	1.443%	0.609%	1.856%	0.454%
	3 Month	**0.446%**	0.537%	0.647%	2.147%	0.969%	1.418%	0.707%	2.197%	0.519%
	4 Month	0.521%	0.630%	0.741%	2.154%	1.052%	1.484%	0.798%	2.197%	0.565%
	5 Month	0.979%	0.594%	0.841%	2.085%	0.748%	1.589%	0.890%	2.116%	0.649%
	6 Month	1.116%	0.800%	0.939%	2.012%	0.959%	1.660%	0.975%	2.029%	0.716%
US	Forecasting Horizon	ANN Ind.	ANN All	AR(1)	VAR(1)	ANN Ind.	ANN All	AR(1)	VAR(1)	RW
	1 Month	0.757%	0.308%	0.319%	1.826%	**0.307%**	1.954%	0.324%	0.345%	0.563%
	2 Month	1.239%	**0.364%**	0.673%	3.080%	0.663%	1.954%	0.656%	3.119%	0.414%

	3 Month	1.539%	0.399%	0.830%	3.711%	0.790%	1.993%	0.793%	3.679%	0.447%
	4 Month	1.743%	0.424%	0.977%	3.985%	1.044%	2.016%	0.919%	3.907%	0.480%
	5 Month	2.115%	**0.349%**	1.101%	4.075%	1.191%	2.037%	1.020%	3.974%	0.485%
	6 Month	2.203%	**0.320%**	1.207%	4.074%	1.356%	2.053%	1.104%	3.965%	0.511%

Source: Author's calculations

6. Concluding Remarks

In this paper the out-of-sample forecasting performance of several Latin-American yield curves, as well as the United States yield curve are tested using the Diebold & Li process. This is done by establishing a Nelson-Siegel specification and the Svensson enhancement. These functional forms were complemented with various forecasting methods, including parametric models and artificial neural networks.

Overall in the one-month-ahead forecast, the neural networks showed the better out-of-sample performance in all the studied yield curves, by surpassing the results of the tested parametric models and being significantly different from the result of the random walk. Particularly the ANN show better results by forecasting each factor individually and then grouping them with a functional form. For the Latin American yield curves the best results are presented with the Nelson & Siegel model, whilst with the US is the Svensson enhancement that exceeds the other options.

In other forecasting time horizons the results are not conclusive, since in some cases some on the neural networks outperform the parametric models, but in other cases the results do not exceed the random walk. The results are heavily dependent on the studied yield curve. On the one hand the yield curve of Colombia is the most difficult to forecast as only in two time horizons was possible to find a suitable forecasting model. On the other hand, the yield curves of Mexico and US are the ones where the forecasting models with neural networks showed the better result. Further, these yield curves are the ones with more historical data. In the case of the yield curves of Peru and Chile, the results of the paper show that there is forecastability in the short-term horizon (up to three months).

Additionally in the analysis of the yield curve of Colombia, there are evident problems in the methodologies that require numerous parameter estimations, given the small cross variable interaction between the factors. Alternatively, in the case of Mexico and US some of the models with cross variable interaction presented the lowest RMSE. Moreover, the one month horizon is the one in which a forecasting methodology always outperformed the random walk option. In all the Latin-American yield curves the optimal forecasting methodology one month ahead are the neural networks that forecast the individually the factors of the Nelson & Siegel specification. Whilst in the US yield curve this horizon can be forecasted with neural networks and the Svensson specification.

In conclusion, there is not enough evidence that proves that with neural networks a better forecast is always going to be accomplished. But neither there is any evidence that demonstrates that the neural networks cannot exceed parametric models and the random walk option. Thus, in a practical approach this tool is recommended as another feasible option to forecast yield curves.

This research can be extended in many ways. Firstly, once more data is available, forecasts of longer time horizons can be studied. Especially considering that some authors, like Diebold & Li (2006) reported better results of parametric models in a one year horizon. Additionally, other extensions of the Nelson & Siegel specification can be analyzed, considering that the Svensson approached showed acceptable results. Among these functional forms are the ones suggested by: Björk & Christensen (1999), Bliss (1997) or De Rezende & Ferreira (2011).

Furthermore, other explanatory variables can be included in the neural networks to forecast the factors of the selected functional form. Fabozzi, Martellini, & Priaulet (2005) suggested variables related to interest rates, variables related to risk, variable related to relative cheapness of stock prices and a sentiment variable (a measure of imbalance between market volume on puts versus calls). Finally Latin-American yield curves may also be useful to test the hierarchical dynamic factor model for sets of a country yield curves suggested by Diebold, Li, & Yue (2008), in which country yields may depend on country factors, and country factors may depend on global factors.

References

Abdel-Aal, R. (2008). Univariate modeling and forecasting of monthly energy demand time series using abductive and neural networks. *Computers & Industrial Engineering*, 903-917.

Adamowski, J., & Sun, K. (2010). Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, 85–91.

Björk, T., & Christensen, B. (1999). Interest Rate Dynamics and Consistent Forward Rate Curves. *Mathematical Finance*, 323-348.

Bliss, R. R. (1997). Testing Term Structure Estimation Methods. *Advances in Futures and Options Research*, 197-231.

Bodyanskiy, Y., & Popov, S. (2006). Neural network approach to forecasting of quasiperiodic financial time series. *European Journal of Operational Research*, 1357–1366.

Bolder, D. J. (December 2006). *Modelling the Term-Structure Dynamics for Risk Management: A Practitioner's Perspective.* Ottawa: Bank of Canada Working Paper.

Bolder, D., & Gusba, S. (2002). *Exponentials, Polynomials, and Fourier Series: More Yield Curve Modelling at the Bank of Canada.* Ottawa: Bank of Canada Working Paper.

Callen, J., Kwan, C., Yip, P., & Yuan, Y. (1996). Neural network forecasting of quarterly accounting earnings. *International Journal of Forecasting*, 475-482.

Cheridito, P., Filipović, D., & Kimmel, R. L. (September 2005). *Market Price of Risk Specifications for Affine Models: Theory and Evidence.* Princeton University Working Paper.

Christensen, J., Diebold, F., & Rudebusch, G. (2008). *An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model.* San Francisco: Federal Reserve Bank of San Fransisco Working Paper Series.

Ciou, S.-J., Fung, K.-Z., & Chiang, K.-W. (2008). A comparison of the artificial neural network model and the theoretical model used for expressing the kinetics of electrophoretic deposition of YSZ on LSM. *Journal of Power Sources*, 338–344.

Coakley, J., & Brown, C. (2000). Artificial Neural Networks in Accounting and Finance: Modeling Issues. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 119-144.

Cziráky, D. (2007). Forecasting the Yield Curve with S-Plus. *Wilmott Magazine*, 2-9.

de Menezes, L. M., & Nikolaev, N. Y. (2006). Forecasting with genetically programmed polynomial neural networks. *International Journal of Forecasting*, 249– 265.

De Pooter, M. (2007). *Examining the Nelson-Siegel Class of Term Structure Models.* Amsterdam: Tinbergen Institute Discussion Papers, Tinbergen Institute.

De Rezende, R. B., & Ferreira, M. S. (2011). Modeling and forecasting the yield yurve by an extended Nelson-Siegel class of models: A quantile autoregression approach. *Journal of Forecasting*.

Deo, M., & Sridhar Naidu, C. (1999). Real time wave forecasting using neural networks. *Ocean Engineering*, 191-203.

Diebold, F. X., & Li, C. (2006). Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics*, 337-364.

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics 13*, 253-263.

Diebold, F., Li, C., & Yue, V. Z. (2008). Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach. *Journal of Econometrics*, 351-363.

Dolan, C. P. (1999). Forecasting the Yield Curve Shape: Evidence in Global Markets. *The Journal of Fixed Income*, 92-99.

Fabozzi, F. J., Martellini, L., & Priaulet, P. (June 2005). Predictability in the Shape of the Term Structure of Interest Rates. *The Journal of Fixed Income*, 40-53.

Faraway, J., & Chatfield, C. (1998). Time series forecasting with neural networks: a comparative study using the airline data. *Journal of the Royal Statistical Society*, 231-250.

Figueiredo, K., Hall Barbosa, C. R., Da Cruz, A., Vellasco, M., Pacheco, M., & Conteras, R. (2007). Neural Networks for Inflow Forecasting Using Precipitation Information. *LNAI*, 552–561.

Gabitsinashvili, G., Namgaladze, D., & Uvo, C. B. (2007). Evaluation of artificial neural network techniques for river flow forecasting. *Enviromental Engineering and Management Journal*, 37-43.

Gilli, M., Große, S., & Schumann, E. (2010). *Calibrating the Nelson-Siegel-Svensson model.* COMISEF Working Papers Series.

Hamzacebi, C., Akay, D., & Kutay, F. (2009). Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. *Expert Systems with Applications*, 3839–3844.

Hecht-Nielsen, R. (1988). Neurocomputing: picking the human brain. *IEEE Spectrum*, 36-41.

Hill, T., O'Connor, M., & Remus, W. (1996). Neural Network Models for Time Series Forecasts. *Management Science*.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, 359-366.

IEEE. (1987). Neural Networks, Part 1. *Expert Focus*, 10-15.

IEEE. (1988). Neurocomputing: picking the human brain. *IEEE Spectrum*, 36-41.

Islam, S. M., Al-Alawi, S. M., & Ellithy, K. A. (1995). Forecasting monthly electric load and energy for a fast growing utility using an artificial neural network. *Electric Power Systems Research*, 1-9.

Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 215-236.

Kahn, R., & Basu, A. (1994). *Neural Nets and Fixed Income Strategies.* Retrieved December 13, 2012, from Barra NEWSLETTER: http://www.barra.com/research/BarraPub/nnaf-n.aspx

Kuan, C.-M., & Liu, T. (1995). Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks. *Journal of Applied Econometrics*, 347-364.

Lachtermacher, G., & Fuller, J. (1995). Backpropagation in time-series forecasting. *Journal of Forecasting*, 381-393.

Litterman, R., & Scheinkman, J. (1991). Common Factors Affecting Bond Returns. *Journal of Fixed Income*, 54-61.

McNelis, P. D. (2005). *Neural Networks in Finance: Gaining Predictive Edge in the Market.* San Diego: Elsevier Inc.

Misas Arango, M., López Enciso, E., & Querubín Borrero, P. (2002). La inflación en Colombia: una aproximación desde las redes neuronales. *Borradores de Economía Banco de la República*.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious Modeling of Yield Curves. *Journal of Business*, 473-489.

Storn, R., & Price, K. (1997). Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, 341-359.

Svensson, L. (1994). Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994. *NBER working paper series No. 4871*.

Swamynathan, V. (2005). *Indian Sovereign Yield Curve using Nelson: Siegel-Svensson Model.* CCIL Publications.

Swanson, N. R., & White, H. (1995). A Model Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks. *Journal of Business and Economic Statistics*, 265-276.

The MIT Press. (1989). *Methods in Neuronal Modelling.* Massachusetts Institute of Technology.

Verona, F. B., & Ceraolo, M. (1998). Use of neural networks for customer tariff exploitation by means of short-term load forecasting. *Neurocomputing*, 135-149.

Willner, R. (June 1996). A new tool for Portfolio Managers: Level, Slope, and Curvature Durations. *The Journal of Fixed Income*, 48-59.

Zhang, G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 159 – 175.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with Artificial Neural Networks: The state of the art. *International Journal of Forecasting*, 35-62.

Zimmermann, H. (2002). *Yield Curve Forecasting by Error Correction Neural Networks and Partial Learning.* Bruges: ESANN'2002 proceedings - European Symposium on Artificial Neural Networks.

Zou, H. F., Xia, G. P., Yang, F. T., & Wang, H. Y. (2007). An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. *Neurocomputing*, 2913–2923.