UNIVERSIDAD DEL ROSARIO

Forecasting the Term Structure of Interest Rates: Macro-Economic and Co-integration Analysis for Colombia Data

Master's Thesis

Author:
Pierre Ricardo Acuña Trujillo

Bogotá D.C.
2019

UNIVERSIDAD DEL ROSARIO

Forecasting the Term Structure of Interest Rates: Macro-Economic and Co-integration Analysis for Colombia Data

Master's Thesis

Author:
Pierre Ricardo Acuña Trujillo

Tutors:
Cristhian Andrés Rodríguez Revilla
Carlos Alberto Castro Iragorri

Program:
Master Quantitative Finance

Bogotá D.C.
2019

CONTENTS

ABSTRACT

Esta tesis pretende pronosticar la estructura de términos de la tasa de interés para Colombia, tomando el procedimiento de dos pasos implementado por (Diebold & Li (2006)); sin embargo, propone alternativas para el enfoque del factor dinámico en la segunda etapa, dependiendo de la estructura presentada por los datos, se agrega un análisis de estacionariedad y cointegración para corregir cualquier especificación erronea, además, para tener más control en los choques externos al modelo, se incorporan algunas variables macroeconómicas relacionadas con la estructura de términos como (Ang y Piazzesi (2003), Ang et al. (2006)).

This thesis takes the two step procedure implemented by (Diebold & Li (2006)) in which yields at different maturities are anticipated in time; however it proposes alternatives for the dynamic factor approach in the second stage depending in the structure presented by the data, for example, due to poor forecast using the VAR-1 of factors in levels (Rodriguez Revilla et al. (2016)) a stationarity and co-integration analysis are added to correct any miss specification, besides, in order to have more control in shocks coming from outside the model, some macro-economic variables correlated with the term structure are incorporated as (Ang & Piazzesi (2003), Ang et al. (2006)). Results show that having more information placed in exogenous variables does not improve any scenario in which these same observations are ignored, instead, it can diminish the quality of forecast; On the other hand, when any miss specification coming from non-stationary factors are corrected, some improvements can be achieved in particular for medium and long term bonds these corrections show equal performance of a random walk. **.**

KEYWORDS

Forecast, Term structure, Stationary, Co-integration, VAR models, Curve Nelson and Siegel model.

1 INTRODUCTION

Forecasting the term structure of interest rate has become an important objective for central banks and investors; the former are more concerned to manage monetary policy to establish the target inflation and economic stability using macro-economic variables that are correlated with the level of interest rates. The latter will pursue arbitrage strategies exploiting the bias between forward rates and expected future short rates.

A significant amount of literature has been developed to understand the stochastic behavior of the term structure. At first, researchers focused in exploiting the form of short rates and how this can be unbiased predictors of forward rates (Fama & Bliss (1987), Vasicek (1977), Cox et al. (1981)); however, empirical results were not completely optimal. Initially, improvements were obtained through equilibrium models, principal component analysis and the use polynomial. This thesis will focus efforts taking of some previous approximation by some authors; for example, (Nelson & Siegel (1987)) proposed a parsimonious exponential representation that models the yield curve, with good results fitting the cross section data through time without the need of imposing arbitrage opportunities; other source comes with (Knez et al. (1994)) that found the importance of having at least three factors for explaining the variations of the term structure in more than 95 percentage. Then, the work will implement and exponential polynomial model that use three factors to explain and anticipate the dynamic across the surface from the yield curve.

The estimation procedure is based in (Diebold & Li (2006)) that proposed a twostep procedure in which the first step allows to identify the three factors that build the curve, then in a second step a VAR-1 dynamic model is performed in order to describe the future style of the curve. Nonetheless, in this work we go beyond and implement a VAR-1 with the factors in difference and a vector error correction method VECM due to some results that prove the non-stationary pattern of the factors and the chance to have cointegration relations among them (Hall et al. (1992), Bowsher & Meeks (2008)). This thesis also pretends to exploit exogenous information beyond the factors captured by the polynomial, therefore, it uses macro-economic information that has a decent degree of correlation with the factors and can help to improve the factor dynamic and consecutively the forecast exercise (Diebold et al. (2006), Ang & Piazzesi (2003), Ang et al. (2006)).

There are some documents for Colombian case that propose different estimation methods of the yield curve (Arango et al. [2003], Vásquez and Melo [2005]), with respect to the investigations that relate the term structure to macroeconomic fundamentals, subsequently the non-structural relationships between the estimated factors of the curve and some macroeconomic variables was analyzed. Melo and Castro [2010] provides bidirectional feedback between the factors of the curve and the macroeconomic variables in an unrestricted model of non-arbitrage with structural specification of the Colombian economy, showing results that indicate that there is a bidirectional relationship between the factors of the curve of yields and the macro variables incorporated in the model. However, there is a stronger causality of the macroeconomic variables towards the curve than in the opposite direction.

Then, the thesis has 2 objectives. First, to test the importance of macro variables to improve the dynamic factor approach. Second, verify considering the non-stationary relationships, the forecast can be improved against a random walk. The results show that incorporating macro-economic variables, the forecasting is reduced, but also, that correcting any miss specification coming from the non-stationary of the factors, the anticipation of futures yields can be ameliorate. The thesis is divided in four main parts; In section (2) the polynomial model is stated, but also and explanation of how yields can recover the forward rates is presented; The model estimation is exposed trough one combination of ordinary least squares (OLS) and variations of dynamic factor approach represented in VAR-VECM specification; In section (3) some test of stationarity and co-integration analysis are presented, then a final frame of results describe in summary the out of sample root mean square error (RMSE) across all specification and maturities for three different horizons of one, six and twelve months ahead. Finally, in section (4) conclusions are manifested.

2 MODEL PROCEDURE

2.1 NELSON AND SIEGEL MODEL

(Nelson & Siegel (1987)) proposed an exponential polynomial in which three latent (unobserved) factors known as level, slope, and curvature can capture most of variation through the term structure in a successful and parsimonious way.

The model instead of trying to capture the stochastic behavior of an unknown future spot (short) rate, directly try to use interpolation on the forward rate with the purpose to extrapolate the behavior avoiding inclusion of variables that at the end can cause problems fitting the real pattern of rates, then, it proposes a polynomial given by equation (1).

$$f_t(\tau) = \beta_t^1 + \beta_t^2 \exp(-\lambda\tau) + \beta_t^3 \exp(-\lambda\tau), \quad \lambda_t = \lambda \tag{1}$$

The previous equation describes the behavior of forward rates generating a family of curves that can take monotonic, humped or S shapes depending of the values adopted by β_t^i. In the original equation λ depends on time, however, in our representation we fix the value into a constant in order to solve the equation linearly with lower dimension of parameters. As our interest is to test the model using observed data (details in section 3), instead of working with the forward equation we should adopt a transformation given by the yield representation; thus, we pursue de definition that in equilibrium yield rates are an average of all forward rates contained in the period for which the bond is active (until maturity), details are provided in appendix A.1.

$$y_t(\tau) = \frac{1}{\tau}\int_0^\tau f_t(s)ds \qquad (2)$$

$$y_t(\tau) = \beta_t^1 + \beta_t^2\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_t^3\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) \qquad (3)$$

$$y_t(\tau) = L_t + S_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + C_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$$

As the yields rate are dependent of forward rates, all the shapes adopted by the these curves are also possible to be represented in the yield curve, then, equation (3) is a linear structure which make possible to capture three unobserved factors that some researcher associate to level slope and curvature $\{\beta_t^1 = L_t, \beta_t^2 = S_t, \beta_t^3 = C_t\}$.

2.2 ESTIMATION AND FORECAST PROCESS

The estimation and forecast process is based in (Diebold & Li (2006)) for which a twostep procedure is proposed. In the initial stage, an OLS regression is stated for the yield equation; As we are only interested in estimate the factors, this exercise assumes as given the value of λ that according to (Nelson & Siegel (1987)) is the value that maximizes the curvature at a maturity level, then, in order to make estimation easier, the value used by (Diebold & Li (2006)) in which λ = 0.0609 is adopted.

$$\begin{bmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ y_t(\tau_3) \\ \dots \\ \dots \\ y_t(\tau_N) \end{bmatrix} = \begin{bmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ 1 & \frac{1-e^{-\lambda\tau_3}}{\lambda\tau_3} & \frac{1-e^{-\lambda\tau_3}}{\lambda\tau_3} - e^{-\lambda\tau_3} \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{bmatrix} \begin{bmatrix} \beta_t^1 \\ \beta_t^2 \\ \beta_t^3 \end{bmatrix} + \begin{bmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \epsilon_t(\tau_3) \\ \dots \\ \dots \\ \epsilon_t(\tau_N) \end{bmatrix}$$

$$Y_t = X\beta_t \ + \ \epsilon_t \tag{4}$$

The previous representation is performed in order to fit in each period of time the best cross section across all maturities composing the yield curve. Then, it is assumed that all markets composing each maturity are independent among them. This assumption allows us to find endogenously the unobserved betas known as level slope and curvature that fit the yield curve in each period of time (t). However, as our purpose is focused in forecast the yield, it is required a second step that could describe the time series dynamic of the curve, then, (Diebold & Li (2006)) stated a VAR-1 approach for the yield curve in which the betas that fit the curve are put them into a time series process that is named as dynamic factor approach.

$$\begin{bmatrix} \beta_t^1 \\ \beta_t^2 \\ \beta_t^3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} a_{1\,1} & a_{1\,2} & a_{1\,3} \\ a_{2\,1} & a_{2\,2} & a_{2\,3} \\ a_{3\,1} & a_{3\,2} & a_{3\,3} \end{bmatrix} \begin{bmatrix} \beta_{t-1}^1 \\ \beta_{t-1}^2 \\ \beta_{t-1}^3 \end{bmatrix} + \begin{bmatrix} v_t^1 \\ v_t^2 \\ v_t^3 \end{bmatrix}$$

$$\beta_t = B \ + \ A\beta_{t-1} + v_t \tag{5}$$

(Rodriguez Revilla et al. (2016)) Proved that for Colombia data, results of forecasting using a VAR-1 of betas where quite weak against a random walk, then, this thesis wants to pursue the same analysis but making some additions and improvements developed by other authors studding the term structure. First, based in (Hall et al. (1992), Bowsher & Meeks (2008)) in which the term structure presents some co-integration relationships, this work tests if the betas estimated in the first step are stationary or not, because given the case in which they were not the VAR-1 approach would be useless due to spurious results of coefficients $\{B,A\}$.

The work presents two alternatives for the VAR-1 approach given both a non-stationarity (augmented Dickey-Fuller test: A.2) and the Johansen co-integration test (A.3); If the test has zero co-integration the procedure will follow a VAR-1 approach in differences, that means, instead to work with the factors in levels the dynamic will be represented with differences assuming a cost in which the long-run relationships cannot be identified; On the other hand, if the test presents one or more co-integration relationships the procedure will be based in a vector error correction model of lag one VECM-1 (Appendix A.4).

$$\Delta\beta_t = B \ + \ A\Delta\beta_{t-1} + \upsilon_t \tag{6}$$

$$\Delta\beta_t = \mu + \psi D_t + \Pi_1 B_{t-1} + \Gamma\Delta\beta_{t-1} + \epsilon_t \tag{7}$$

This thesis also pretends to test how exogenous information can affect the dynamic of yield curve, for that purpose it bases the approach in (Estrella & Hardouvelis (1991), Ang & Piazzesi (2003), Ang et al. (2006),Diebold et al. (2006)) who showed how patterns of correlation between the term structure and macro-economic activity as GDP, inflation and unemployment can improve future expectations in yield curve; Thus, the last step to the dynamic factor analysis is to introduce exogenous macro variables that have correlation with the term structure with the purpose to exploit causality in the forecast of yields. According to this addition, equations (5), (6) and (7) are restructured in a way that vectors contain both unobserved endogenous factors beta (β) and main observed macroeconomic variables (M).

$$\begin{bmatrix} \beta_t^1 \\ \beta_t^2 \\ \beta_t^3 \\ M_{(K_X 1)} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_{(K_x 1)} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{1K\ (1_X K)} \\ a_{21} & a_{22} & a_{23} & a_{2K\ (1_X K)} \\ a_{31} & a_{32} & a_{33} & a_{3K\ (1_X K)} \\ a_{K1\ (K_X 1)} & a_{K2\ (K_X 1)} & a_{K3\ (K_X 1)} & a_{KK\ (K_X K)} \end{bmatrix} \begin{bmatrix} \beta_{t-1}^1 \\ \beta_{t-1}^2 \\ \beta_{t-1}^3 \\ M_{t-1\ (K_x 1)} \end{bmatrix} + \begin{bmatrix} \upsilon_t^1 \\ \upsilon_t^2 \\ \upsilon_t^3 \\ \upsilon_{t\ (K_x 1)}^K \end{bmatrix}$$

$$Z_t = B \ + \ AZ_{t-1} + \upsilon_t \tag{8}$$

$$\Delta Z_t = B \ + \ A\Delta Z_{t-1} + \upsilon_t \tag{9}$$

$$\Delta Z_t = \mu + \psi D_t + \Pi_1 Z_{t-1} + \Gamma\Delta Z_{t-1} + \epsilon_t \tag{10}$$

Equations (5) to (10) are different alternatives for the dynamic factor approach, Thus, in the forecasting exercise efforts are focused in anticipate the expected value of the unobserved betas

(β) that endogenously explain the future value of the yield curve and consecutively the term structure of interest rates.

$$\hat{Y_{t+h}} = E[Y_{t+h|t}] = X\hat{\beta_{t+h}} = XE[\beta_{t+h|t}] \quad (11)$$

Equation (11) Describes for any h the process of anticipation on the yield at any maturity that compose the yield curve, thus, equation (11) allows to extrapolate the term structure of interest rates using the polynomial form. However, as in (Diebold & Li (2006)) we need to test all the previous specification against a simple non parametric approach known as random walk in which the best h future magnitude is the present value take by the yield.

$$\hat{Y_{t+h}} = Y_t \quad (12)$$

The process for testing all specifications is based in an out of sample exercise in which a fix length estimation rolling window is implemented through an evaluation period of time in which observed data is available to contrast how closer are forecast values of realizations. The loss function used by this work is the root mean square error (RMSE) which measure the distance in magnitude without paying attention if the deviations are due to over or under expectation.

$$RMSE = \sqrt{(\hat{Y_{t+h}} - Y_{t+h})^2} \quad (13)$$

In order to present results, we define some specification in abbreviations. For example, equation (5) is represented as (NS), (6) as (NS-Stationary), (8) as (NS-M), (10) as (NS-M-Stationary) and finally (12) as (RW). Some specifications as (7) and (9) are not abbreviated because they are not implemented due to rejection given by the data (further details in section 3); therefore, the exercise has in total 5 specifications competing for finding the lower RMSE. For the out sample test it recreates different scenarios under three horizons of one, six and twelve months ahead representing short, middle and long term forecast.

3 DATA AND RESULTS

The term structure of interest rates is also linked to the study of macroeconomic fundamentals, given the average rate at which bond rates can be the expected values of the average interest rates. In the short term in the future, one might think that the yield curve contains information about agents' expectations of an economy (Piazzesi (2002)). Similarly, the yield curve could contain information on the transmission of monetary policy and information on fiscal policy decisions (Cochrane (2001)), which makes the study of relationships between the dynamics of the term structure of interest rates and macroeconomic variables.

It should be noted that the work carried out on the subject usually associates the movements of the yield curve with variables of economic activity, prices and monetary policy interest rates. According to (Wu, 2003) and then exposed by (Arango & Florez (2005)) for the Colombian case it is usual to relate the reaction of the short part of the yield curve with the actions of the monetary authority and, to some extent, with the real and financial conditions of the economy. On the other hand, the middle section has a high informative capacity of inflation expectations in the medium term and, therefore, the objectives of the monetary authority. The long tranche is influenced by the economy's potential growth perceptions, long-term inflation expectations and the credibility of the central bank in its fight against inflation. The above given that in Colombia the policy focuses on achieving a low and stable inflation target and maximum sustainable growth; thus, the implicit definition of the policy actions lies in the reaction of the inflation and growth variables primarily, making the macroeconomic variables can be decisive and fundamental in explaining the movements of the yield curve.

In this sense, it is necessary to have indicators that allow the time structure to be based on the dynamics of these macroeconomic variables, which are considered fundamental for the returns. Indicators of economic activity A large range of leading indicators of economic activity have been built around the world for this purpose, because PIB is generally reported quarterly and with several months of lag, it is highly desirable to have alternative indicators of economic activity. the most frequent and timely added economic activity, which anticipates the behavior of the economic cycle. In particular, because this information is essential given the lags with which policy measures are transmitted to the economy, in this sense the IMACO (Kamil, Pulido & Torres (2010)) has other desirable predictive properties: (i) a high correlation advanced with the economic cycle, (ii) that anticipates its break points without giving false signals, and (iii) that minimizes forecast errors on PIB growth.

The exchange rate (TRM), the central bank responds to movements in the exchange rate as long as those movements affect the inflation forecasts and the product gap. And the answer is to modify its intervention rate or intervention in the foreign exchange market, or both. This approach provides greater stability or volatility in exchange rates, therefore, it is expected that term structure will be directly affected by this policy of a monetary nature.

The Consumer Price Index (IPC) is the main instrument for quantifying inflation, since it represents the variation of the general price level of goods and services consumed by households; and, it allows to measure the evolution of the value of money against a market basket.

Given the dynamism of economic activities, the indicator of the financial PIB (PIBFIN) that measures the value of economic activity focuses only on the explanation of economic activity from financial services.

The data implemented are the yields corresponding to zero coupon bonds of maturity one, five and ten years for Colombia that come from the central bank (Banco de la República); Frequency is expressed in monthly terms, then, the complete sample starts in (2005:01) to (2015:08) given as a result a time series of length equal to (129) observation for the three maturities than composed the yield curve. On the other hand, macro-economic variables are obtained in the same period at the same frequency from the department of economic research within Citibank Colombia; however, some sources are cited to come from the central bank (Banco de la República) as well.

Macro-economic variables are observed and exogenous from the betas that describe the behavior of yield rates, in total the thesis sets a number of variables K equal to six, starting from IMACO index that describes the economic activity of the country, then, TRM describes the exchange rate dollar peso, IPC does the same fitting inflation from consumers, PIBFIN is focused only in the explanation of economic activity coming from the financial services, DESEMPLEO is the unemployment rate and finally INTRATE is the central bank rate driving the monetary policy of the country. Next table shows correlations across unobserved beta factors estimated from the yields at first stage together with the observed factors that comes from external sources to the yield equation.

	B1	B2	B3	IMACO	TRM	IPC	PIBFIN	DESEMPLEO	INTRATE
B1	1.00	-1.00	-0.99	0.22	0.18	0.26	0.35	-0.43	0.47
B2	-1.00	1.00	0.99	-0.18	-0.17	-0.30	-0.39	0.45	-0.46
B3	-0.99	0.99	1.00	-0.25	-0.13	-0.28	-0.37	0.44	-0.46
IMACO	0.22	-0.18	-0.25	1.00	0.05	-0.53	-0.44	0.11	-0.14
TRM	0.18	-0.17	-0.13	0.05	1.00	-0.26	-0.19	0.08	-0.15
IPC	0.26	-0.30	-0.28	-0.53	-0.26	1.00	0.98	-0.62	0.32
PIBFIN	0.35	-0.39	-0.37	-0.44	-0.19	0.98	1.00	-0.68	0.33
DESEMPLEO	-0.43	0.45	0.44	0.11	0.08	-0.62	-0.68	1.00	-0.36
INTRATE	0.47	-0.46	-0.46	-0.14	-0.15	0.32	0.33	-0.36	1.00

Table 1: Table of Correlations

We can see from the previous table that all macro variables considered in this exercise show a degree of correlation with the unobserved beta factors, being TRM the weakest variable and INTRATE the strongest that co-move with the level, slope and curvature from the term structure. Thus, dimension of vector Z is adjusted to three plus six that is equal to nine variables.

Initially the exercise will pursue the analysis without macro variables as in (Diebold & Li (2006)); thus, the first model is based in a VAR-1 (NS) representation for the beta (β) dynamic, however, we perform in each one of the series a stationary test trough an augmented Dickey-Fuller procedure; results in table 2 show that level, slope and curvature fail to reject the null hypothesis of unit root, then, the VAR-1 approach is not a correct dynamic specification for the betas (β); In order to find a correct specification is necessary to check if the variables have a co-integration relationship that allows to run a VECM; nonetheless, results of the second test in table 3 underline that there is no co-integration (r=0) among the unobserved factors which makes incorrect to use either a VAR-1 or a VECM approach for the structure of betas.

ADF Test	Dickey-Fuller	P-Value
B1	-2.32	0.44
B2	-2.40	0.41
B3	-2.26	0.47

Table 2: Dickey-Fuller Test

	Statistic	10pct	5pct	1pct
r <= 2 |	1.38	7.52	9.24	12.97
r <= 1 |	11.09	13.75	15.67	20.20
r = 0 |	20.79	**19.77**	22.00	26.81

Table 3: Johansen Test

As the level, slope and curvature are integrated variables of order one I(1) but also do not have co-integration relationship, the best strategy is to adopt a VAR-1 model with the factors in differences assuming a cost of losing information in the long-run; Thus, equation (6) is used as (NS-Stationary) and the variables in differences are modeled, however, the goal must be adjusted with variables in levels using an invert transformation in which knowing the predicted difference and the last observation we can recover or measure the future value of the variable in levels ($\beta_{t+1} = \Delta\beta_{t+1} + \beta_t$) and consecutively the predicted yield can be identified. Despite table presents results at different levels of confidence the criteria is based at 5 percentage giving as a result the no rejection of null hypothesis (r=0).

The next step is to incorporate in an exogenously way the macro-economic variables that are correlated with the unobserved factors; then, the work runs a VAR-1 model of three unobserved factors plus (K=6) observed variables as (NS-M); Nonetheless, it is known from table 3 that only the factors are non-stationary, then, the same analysis is applied to the exogenous variables finding in table 4 the same conclusions which make to reject (NSM) specification, because all the variables that are contained in Z present non-stationary patterns; Hence, a Johansen co-integration test is developed taking in consideration jointly all nine variables, in consequence table 5 allows us to identify that the null hypothesis (r=4) at 10 percentage is not rejected and co-integration relationship exist.

ADF Test	Dickey-Fuller	P-Value
IMACO	-2.30	0.45
TRM	0.05	0.99
IPC	-1.52	0.78
PIBFIN	-0.77	0.96
DESEMPLEO	-2.65	0.31
INTRATE	-3.22	0.09

Table 4: Dickey-Fuller Test: Macro

	Statistic	10pct	5pct	1pct
r <= 8 |	4.03	7.52	9.24	12.97
r <= 7 |	6.74	13.75	15.67	20.20
r <= 6 |	15.01	19.77	22.00	26.81
r <= 5 |	20.49	25.56	28.14	33.24
r <= 4 |	35.21	31.66	**34.40**	39.79
r <= 3 |	39.65	37.45	40.30	46.82
r <= 2 |	57.35	43.25	46.45	51.91
r <= 1 |	73.87	48.91	52.00	57.95
r = 0 |	88.95	54.35	57.42	63.71

Table 5: Johansen Test: Macro

As (*r* 6= 0) a VECM is possible to apply into the data Z, then, a model with intercept and without deterministic trend is specified; the way to estimate the model is through a long-run form of lag one that is denoted (NS-M-Stationary); Even though at five percentages the co-integration relationship is five instead of four, this work prefers to set (r=4) in order to have less variables in a linear combination that can produce a stationary process. For further research we can play with the degree of co-integration and specification (intercept plus trend), (only trend) or (nothing), besides, the way of estimation can be changed from long-run to transitory.

3.1 OUT OF SAMPLE FORECAST

The forecast exercise is implemented in an out of sample scenario in which a fix estimation window is set with length equal to (62) observations, meaning that the evaluation window for testing models trough the RMSE is set to start at (2010:01) until August of two-thousand-fifteen (2015:08); Results are an average across sixty-seven (67) , sixty-two (62) and fifty-six (56) square root errors corresponding to horizon (h) of one, six and twelve months ahead respectively. As our data only has three maturities of one, five and ten years, the average for cross-section is also presented with the goal to check in which horizons at different maturities some specification can defeat or equalize the performance of a random walk.

	Maturity	NS	NS-Stationary	NS-M	NS-M-Stationary	RW
	1Y	0.1859	0.1896	0.2055	0.2293	0.1681
h=1	5Y	0.3175	0.2678	0.3498	0.3156	0.2678
	10Y	0.3241	0.2723	0.3525	0.3395	0.2739
	Average	**0.2759**	**0.2433**	**0.3026**	**0.2948**	**0.2366**
	1Y	0.4795	0.4935	0.6469	0.6664	0.4898
h=6	5Y	0.9328	0.7538	0.9178	0.7895	0.7647
	10Y	1.0016	0.8068	1.0064	0.8718	0.8076
	Average	**0.8046**	**0.6847**	**0.8571**	**0.7759**	**0.6874**
	1Y	0.7119	0.7640	0.8092	0.9790	0.7455
h=12	5Y	1.1480	0.8194	1.2563	1.1019	0.8221
	10Y	1.2075	0.9698	1.4396	1.1936	0.9684
	Average	**1.0225**	**0.8510**	**1.1684**	**1.0915**	**0.8454**

Table 6: Out of Sample RMSE

It is clear that in average for all horizons excepting (h=6) the random walk is unbeatable, Nonetheless, brief conclusions should be read carefully, for example, table 6 presents a detailed comparisons across all specifications; Firstly, for middle and long term bonds (5Y, 10Y) the correction of non-stationarity only in the unobserved factors (NS-Stationarity) can improve a VAR-1 (NS) specification, but also it can defeat or equalize the performance of a random walk for eighty-three percentage of cases (83%). Secondly, when exogenous information is considered and placed in a dynamic factor approach results are even worse than a simple specification (NS). Therefore, it is conclusive that correcting miss specification coming from non-stationarity is helpful in terms of forecasting, but also when external information is considered in an exogenous representation results are destroyed even when non-stationarity is taken in consideration.

4 CONCLUSIONS

In this thesis the research implemented by (Diebold & Li (2006)) is adjusted into Colombia data. Besides, the work goes beyond incorporating the possibility of non-stationarity structure in the variables that composed the dynamic factor approach and the chance to have more information through exogenous macro-economic variables as well. Similarly, to (Rodriguez (2016)) the testing exercise found that initially Colombia data reject the specification implemented by (Diebold & Li (2006)), however, it also found that the chances to increase performance through the dynamic factor approach are possible when only the endogenous variables are considered in a non-stationarity approach. On a second stage, the exercise also found that having more information not necessary implies better performance in forecasting at least not when the additional information is introduced in an exogenous structure for the yields; then, despite some papers as (Ang & Piazzesi (2003), Ang et al. (2006)) showed some gains when macro variables are induced, for Colombia this seems to not be true completely, at least for the period of time in which the sample is analyzed.

For further research it will be interesting to test how results can change if additional information coming from macro-economic variables are placed endogenously to the yield equation instead of just an exogenous structure in the dynamic approach. Besides, it will be quite enrichment to play with different specification in the VECM that can improve forecast for endogenous variables explaining the yields at different maturities.

REFERENCES

Ang, A., & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary economics, 50(4), 745–787.

Ang, A., Piazzesi, M., & Wei, M. (2006). What does the yield curve tell us about gdp growth? Journal of econometrics, 131(1-2), 359–403.

Arango, L. E., & Flórez, A. (2005). Tramo corto de la curva de rendimientos, cambio de régimen inflacionario y expectativas de inflación en Colombia. Borradores de Economía Banco de la República.

Backus, D., Foresi, S., Mozumdar, A., & Wu, L. (2001). Predictable changes in yields and forward rates. Journal of Financial Economics, 59(3), 281–311.

Bautista, R., A. Riascos, Y N. Suárez (2007): "La aplicación de un modelo de factores a las curvas de rendimiento del mercado de deuda pública colombiano," Galeras de Administración 14, Universidad de los Andes, Facultad de Administración.

Bowsher, C. G., & Meeks, R. (2008). Stationarity and the term structure of interest rates: a characterisation of stationary and unit root yield curves.

Cochrane, J. H. (2001). Long term debt and optimal policy in the fiscal theory of the price. Econometrica, 69-116

Cox, J. C., Ingersoll Jr, J. E., & Ross, S. A. (1981). The relation between forward prices and futures prices. Journal of Financial Economics, 9(4), 321–346.

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. Journal of econometrics, 130(2), 337–364.

Diebold, F. X., Rudebusch, G. D., & Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. Journal of econometrics, 131(1-2), 309–338.

Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. The journal of Finance, 46(2), 555–576.

Fama, E. F. (1976). Forward rates as predictors of future spot rates. Journal of Financial Economics, 3(4), 361–377.

Fama, E. F., & Bliss, R. R. (1987). The information in long-maturity forward rates. The American Economic Review, 680–692.

Hall, A. D., Anderson, H. M., & Granger, C. W. (1992). A cointegration analysis of treasury bill yields. The Review of Economics and Statistics, 116–126.

Kamil H., & Pulido, J., & Torres, L., (2010). "El "IMACO": un índice mensual líder de la actividad económica en Colombia," Borradores de Economía 609, Banco de la Republica de Colombia.

Knez, P. J., Litterman, R., & Scheinkman, J. (1994). Explorations into factors explaining money market returns. The Journal of Finance, 49(5), 1861–1882.

Melo, L. F., & Castro, G. A. (2010). Relación entre variables macro y la curva de rendimientos. Borradores de Economía Banco de la República.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. Journal of business, 473–489.

Piazzesi, M. (2002). Affine Term Structure Models. En Y. A.-S. Hansen, Handbook of Financial Econometrics. (págs. 692-766). Amsterdam: North Holland.

Rodriguez Revilla, C. A., et al. (2016). Estimating and forecasting the term structure of interest rates: Us and Colombia analysis (Unpublished doctoral dissertation). Universidad del Rosario.

Shiller, R. J., Campbell, J. Y., Schoenholtz, K. L., & Weiss, L. (1983). Forward rates and future policy: Interpreting the term structure of interest rates. Brookings Papers on Economic Activity, 1983(1), 173–223.

Vasicek, O. (1977). An equilibrium characterization of the term structure. Journal of financial economics, 5(2), 177–188.

Vásquez, D. M., Y L. F. Melo (2005): “Estimación de la estructura a plazos de las tasas de interés en Colombia por medio del método de funciones B-spline cúbicas,” Revista de Economía del Rosario, 8(1), 1–23.

Wu, T. (2003). What Makes the Yield Curve Move? Economic Letter Federal Reserve Bank of San Francisco.