Abstract

This paper obtains the forecasts of Colombian macroeconomic variables and the yield curve by jointly modeling their dynamics. For this purpose, I use unrestricted Bayesian Vector Auto Regressive (VAR) models and the no-arbitrage state-space representation developed by Ang and Piazzesi [2003]. Both the Bayesian VAR and the no-arbitrage representations are used to estimate closed economy, small open economy and macro-laten t factor models. The parameters of the models are estimated with Bayesian techniques for different horizons using the predictive likelihood function. Monthly data between 2006-2012 of the inflation, the overnight-interbank interest rate, an economic activity indicator, the 10-year treasury rate and the 5-year CDS was used. The main finding is that the out-of-sample forecasts of the interbank overnight interest rate and the inflation consistently improve when the yield curve is incorporated. Moreover, the models that impose the no-arbitrage restriction consistently out-perform the unrestricted VARs. On the other hand, the model with the best performance in terms of both the RMSE and the standard deviation of the forecasts incorporates closed-economy variables and the short-term yield. Adding longer-term yields and small open economy variables does not appear to improve further the forecasts.

Keywords: Yield Curve, Macro-Latent Factor Model, Forecasting, Bayesian Estimation, Predictive Likelihood.

JEL Codes: E43, E52, G12
1 Introduction

Recent literature in macro-finance has provided evidence that the term-structure of interest rates is an important source of information on market expectations and macroeconomic shocks (Ang and Piazzesi [2003], Diebold et al. [2006], Andreasen [2008], Zagaglia [2009] and Tristani and Amisano [2010]). For instance, Evans and Marshall [2007] found that monetary policy shocks have a significant impact on the slope of the yield curve.

Reinhart et al. [2015] highlight that the yield curve has information on an implicit disaster probability on financial markets, which should be taken into account in the long-term policy making process. Bernanke, in his speech “Reflections on the Yield Curve and Monetary Policy” in 2006, was one of the first policy-makers to mention the relevance of the information contained in the yield curve for policy analysis. Additional evidence on the relevance of the relationship between macroeconomic variables and the yield curve lies on the fact that several Central Banks that have reached the zero-lower bound on their policy interest rates have started targeting the term-structure of interest rates, in order to provide further stimulus to the economies and boost consumption. This has led to unconventional monetary policy actions such as the Quantitative Easing programs around the globe.

The main purpose of this thesis is to obtain forecasts for different horizons between 1 and 12 months of the interbank overnight interest rate, the inflation, a monthly activity index and yields of different maturities, by jointly modeling the dynamics of the macroeconomic variables and the yield curve. Additionally, I aim to improve the forecast of the macroeconomic variables and the yield curve when compared to simple Bayesian Vector Auto Regressive (VAR) models that do not model macroeconomic and financial variables jointly. Intuitively, adding more recent information by incorporating the yield curve and other financial variables should improve the forecast of macroeconomic variables, which have a time-lag. Even though the joint-dynamics of asset prices, macro and latent factors happen simultaneously in the models, missing data techniques are used (with the Kalman smoother) to incorporate the new available information during the forecasting process. The models are developed in the context of a small open economy, specifically in Colombia. The joint dynamic of Colombian macroeconomic variables and the Colombian yield curve has gained relevance as local financial markets have deepened during the last years. As Colombia has opened to external capital flows and the secondary market of public bonds has developed during the last decade, the yield curve has increased its sensibility to the international markets (Guarin et al. [2014] present evidence on this issue).

The relationship between the yield curve and macroeconomic variables in Colombia has been previously analyzed. Guarin and Murcia [2015] state that macroeconomic variables significantly contribute to explain the dynamics of Colombian bond yields. They highlight that domestic macroeconomic shocks (especially monetary policy shocks) explain most of the variance of the short-term yields and that the movements on the medium-term and long-term yields are mainly explained by the international interest rate adjusted by local risk. This is supported by the evidence from Espinosa et al. [2015] who found that the term premia of Colombian public bonds responds permanently to changes in the United States term premium. However, jointly modeling the yield curve and macroeconomic variables has not been used as a tool to improve the forecasts of this set of variables in Colombia. Vela [2013] compared different forecasting models for the yield curves of Latin American countries and found that the Colombian yield curve was the most difficult to forecast; for most of the time horizons, he
could not find a suitable forecasting model.

In this paper, I use Bayesian VARs and the no-arbitrage state-space representation developed by Ang and Piazzesi [2003] to jointly model macroeconomic and financial variables. Both representations are used to estimate closed economy, small open economy and macro-laten factor models. The performance of each model is compared in terms of the RMSE and the standard deviation of the forecasts relative to the following benchmarks: a Bayesian VAR with only macroeconomic variables and a Bayesian VAR with only selected zero-coupon rates. To estimate the parameters of the models, I use monthly data between 2006 and 2012 of macroeconomic variables (activity index, inflation and the interbank overnight interest rate) and financial variables (yield curve, 5-years CDS and 10 year treasury rate). For the variables that have a daily frequency, a monthly average is computed.

The parameters of the models are estimated using Bayesian techniques. An MCMC-Metropolis Hastings algorithm is implemented to obtain the posterior distribution of the parameters of each model. Given that the forecasts are obtained for periods up to 12 months, the predictive likelihood function for different horizons is used in this part of the process. Afterwards, the Kalman smoother is used to recover the macroeconomic data for the last month of data available. By treating the macroeconomic variables in the last month of the available information as a missing data, the forecast process in the current month is replicated because only the financial information is up-to-date, while macroeconomic information is lagged. After calculating the 1-month forecast with the Kalman smoother, the Kalman filter is iterated forward to obtain the forecast for the next time-horizons.

The main finding is that incorporating the yield curve consistently improves the forecast for the interbank interest rate and the inflation. Furthermore, imposing the no-arbitrage restriction significantly improves the out-of-sample performance of the forecasts when compared to simple Bayesian VAR models. The best model to forecast the macroeconomic variables incorporated the no-arbitrage restriction, closed economy variables and the short-term yield. Adding longer-term yields and small open economy variables does not appear to improve the forecasts of the macroeconomic aggregates. On the other hand, only a few models consistently improve the forecast for the 12 and 60 months zero-coupon rates while no model could consistently improve the forecast for the 120 months yield.

The remainder of the paper is organized as follows: Section 2 shows the relevant literature review. Section 3 introduces the macro-latent factor model under the no-arbitrage condition. Section 4 goes into details of the data set and the preliminary evidence on the relationship between the yield curve and macroeconomic variables. Section 5 summarizes the Bayesian techniques implemented to estimate the parameters of the models and presents the models that are estimated in the process. In Section 6, I present the methodology used to obtain the forecasts for the different periods and make a comparison of the forecast accuracy of the estimated models. Finally, Section 7 brings some conclusions.

2 Literature Review

Different frameworks have been developed on the study of the relationship between macroeconomic variables and the yield curve. One of them is a financial affine model approach in which no-arbitrage restrictions are imposed and the yields are modeled as a function of latent factors, as in Duffie and Kan [1996]. Afterwards, literature has developed a macro-finance approach in which macroeconomic variables are incorporated in the modeling of the yield curve.
- The yield curve and macroeconomic variables: jointly modeling

Ang and Piazzesi [2003] combined the no-arbitrage restrictions with the Taylor rule and included macroeconomic variables. They showed that macroeconomic factors explain up to 85% of the variance in the bond yields. However, unobservable factors still accounted for most of the variations in the long-term yields. Additionally, Cortes and Ramos, 2008 highlight two advantages of the no-arbitrage assumption: first, the models that use the no-arbitrage assumption span the whole yield curve (and not only the specific yields that are introduced in the model), and second, the movements of the yields are free from arbitrage. Duffee [2011] highlights that the no-arbitrage restrictions are useful to infer the cross-sectional properties of the yields with high precision.

Another method to study the relationship between macroeconomic aggregates and the term structure of interest rates is the one used by Diebold et al. [2006]. They estimate the yield curve by using unobservable factors and macroeconomic data. They found strong evidence on the effects of macroeconomic variables on the yield curve variations but they also highlight they found weaker evidence on the reverse relationship.

Recent literature also models the dynamics of the macroeconomic variables and latent factors as a general equilibrium model. Andreasen [2008] and Rudebusch and Wu [2008] reproduce the dynamics of the yield curve along with the macro variables using a Neo-Keynesian model. Cortes and Ramos, 2008 extend the models to a small open economy. Finally, Zagaglia [2009] introduces micro-founded Dynamic Stochastic General Equilibrium (DSGE) models for the jointly modeling of the yield curve and macroeconomic variables.

Literature has found a big difficulty in explaining the relationship between macroeconomic variables and the yield curve on emerging markets. Djurankovic [2014] found that the Indonesian yield curve had a large amount of variation that is not spanned by macroeconomic variables, and only the level factor can explain part of the variation of macroeconomic variables. Besides, Lange [2014] proved that a strong bilateral relationship exists between Canadian yield curve and U.S. yield curve factors (which were treated as exogenous variables), providing evidence that the yield curve of small open economies may be closely linked to the U.S. yield curve and international factors. This is supported by Guarin and Murcia [2015], who found that the risk premium and the foreign interest rate determine the medium-term and long-term Colombian yields.

- The yield curve and macroeconomic variables: forecasting

The usefulness of the yield curve to forecast macroeconomic variables has been addressed by Saar and Yagil [2015] who proved by making linear regressions that the government and the corporate yield curve can predict future economic growth and stock market trends in Europe, Britain and Asia. Abdymomunov [2013] showed that the U.S. yield curve can be used as a predictor for the output by incorporating the real GDP in the Nelson and Siegel's dynamic yield curve model. Additionally, Huse [2011] found that models with macroeconomic variables out-perform latent factor models in terms of the accuracy of the forecasts of the yield curve, by letting the factors of the model to be driven only by observable variables. However, in line with the evidence on emerging markets, which was mentioned previously, Kaya [2013] could not out-perform a random walk model for an horizon longer than 3 months when comparing the forecast of Turkey’s yield curve. In this paper, the yield curve was modeled as a function of three latent variables.

Concerning the affine models, Dai and Singleton [2000] argue that three factor affine models do not produce adequate forecasts. This conclusion is supported by Duffee [2002] that found
that affine equilibrium models led to poor forecasts of the yield curve. Furthermore, Ang and Piazzesi [2003] improved the out-of-sample forecasts of the yield curve by jointly imposing no-arbitrage restrictions and the Taylor rule (their evidence is supported by the findings of Duffee [2011]). Additionally, Ang et al. [2006] show that the out-of-sample forecast of the GDP improves when the no-arbitrage restriction is used.; they also state that the short term yield has more predictive power on macroeconomic variables than any other yield and that models that incorporate macroeconomic variables out-perform models that use only unobservable factors. However, Joslin et al. [2011] and Carriero and Giacomini [2011] found that no-arbitrage restrictions do not improve the out-of-sample forecasts of the bond yields. Finally, Diebold and Li [2006] used three time varying factors (level, slope and curvature) along with the Nelson and Siegel model to forecast the yield curve. The factors were forecasted with AR models and the forecasts of the yield curve had a good fit on the data.

3 Model

This section shows the state-space representation of a macro-latent factor model as Guarin and Murcia [2015]. The section also introduces the framework for pricing bonds under the no-arbitrage condition in the same way as Guarin and Murcia [2015]. Finally, I introduce the models that are analyzed through the thesis.

3.1 State Space Representation

Consider the state-space representation

\[ Z_t = C + D X_t + \eta_t \]  

(1)

\[ X_t - \mu = \Phi (X_{t-1} - \mu) + \Sigma \varepsilon_t \]  

(2)

for \( t = 1, \ldots, T \).

The measurement equation (1) describes a vector of observable variables \( Z_t \) as an affine function of a state vector \( X_t \). The observation vector \( Z_t = \begin{bmatrix} y_t' & m_t' \end{bmatrix}' \) where \( y_t = [y_{t,n_1}, y_{t,n_2}, \ldots, y_{t,N}]' \) denotes the bond yields vector for selected maturities \( n_1, n_2, \ldots, n_N \), and \( m_t = [m_1, m_2, \ldots, m_J]' \) stands for the set of \( J \) macroeconomic variables.

Bond yields are assumed to be observed with a measurement Gaussian error \( \eta_t \). The elements of this error are assumed to be uncorrelated. Therefore, \( R \), which is the variance-covariance matrix of the measurement equation, is a diagonal matrix.

The measurement equation (1) can be written as

\[
\begin{bmatrix} y_t' \\ m_t' \end{bmatrix} = \begin{bmatrix} A \\ 0 \end{bmatrix} + \begin{bmatrix} B_f' & B^m \end{bmatrix} \begin{bmatrix} f_t \\ m_t \end{bmatrix} + \begin{bmatrix} \eta_t^y \\ 0 \end{bmatrix} \]  

(3)

for \( t = 1, \ldots, T \). As described in this representation, measurement error is only different than zero for the zero-coupon yields. The values of \( A_n \) and \( B_n = [B^f\ B^m] \) impose the no-arbitrage restriction and will be detailed in Section 3.3.

The transition equation (2) describes the dynamics of the state vector \( X_t - \mu \) as a VAR(1) process. The state vector is composed of observable and unobservable variables, \( X_t = [X_t^o, X_t^u] \); \( \epsilon_t \) denotes a Gaussian error vector that has a variance matrix \( Q \). Finally, the total number of errors to be estimated (taking into account both the measurement and
transition equations) should be greater than or equal to the number of observable variables in order to avoid stochastic singularity problems.

3.2 Bond Pricing Under No-Arbitrage Condition

Following Ang and Piazzesi [2003], $X_t$ is a vector of state variables of dimension $k \times 1$. The state variables have an underlying dynamic; they may be observable or unobservable variables and. $X_t$ may be written as $X_t = [X^o_t, X^u_t]$ where $X^o_t$ stands for the observable variables and $X^u_t$ for the unobservable variables. For the purpose of this thesis, the unobservable variables only correspond to latent factors.

$X_t$ is supposed to follow the following VAR(1) process:

$$X_t = \Phi X_{t-1} + \Sigma \varepsilon_t$$

(4)

where $\Phi$ is a $k \times k$ matrix which I will refer as the transition matrix and $\varepsilon_t \overset{iid}{\sim} N(0_{k \times 1}, I_{k \times k})$ is a $k \times 1$ vector of Gaussian errors.

Next, the one-period short rate, $i_{1,t}$, is defined as an affine function of the state variables $X_t$:

$$i_{1,t} = \delta_0 + \delta_1 X_t$$

(5)

where $\delta_0$ is a scalar and $\delta_1$ is a $k \times 1$ coefficient vector. Note that if $\delta_1$ is restrained to be a function of only contemporary macroeconomic variables.

Under the no-arbitrage restriction, a risk-neutral measure $Q$ is guaranteed so that the price of the bonds that do not pay coupons at time $t+1$ satisfy:

$$V_t = E^Q_t (\exp(-i_{1,t} V_{t+1})$$

(6)

After Ang and Piazzesi [2003] the Radon-Nikodym derivative converts the risk-neutral measure to the data-generating measure and follows a log-normal process:

$$\xi_{t+1} = \xi_t \exp \left( -\frac{1}{2} \lambda_t \Sigma \lambda_t - \lambda_t \varepsilon_{t+1} \right),$$

(7)

where $\lambda_t$ is the time-varying market price of risk and is parametrized as an affine function:

$$\lambda_t = \lambda_0 + \lambda_1 X_t$$

(8)

Now, following Ang and Piazzesi [2003] for any random variable $Z_{t+1}$,

$$E^Q_t(Z_{t+1}) = \frac{E_t(\xi_{t+1} Z_{t+1})}{\xi_t}$$

(9)

and based on this the pricing kernel $m_{t+1}$ is defined:

$$m_{t+1} = \exp(-i_{1,t}) \xi_{t+1}/\xi_t$$

(10)

Based on Ang and Piazzesi [2003], any asset should satisfy that:

$$E_t(m_{t+1} R_{t+1}) = 1,$$

(11)
and the bond prices may be calculated recursively with the following equation:

\[ P_t^n = E_t \{ m_{t+1} P_{t+1}^{n-1} \} \]  

where \( P_t^n \) denotes the price of the bond at time \( t \) which has a maturity \( n \) periods ahead.

Ang and Piazzesi [2003] state that the bond prices are an affine function of the state variables and can be expressed as in the following equation:

\[ P_t^n = \exp \left( \tilde{A}_n + \tilde{B}_n' X_t \right) \]  

where \( \tilde{A}_n \) is a scalar and \( \tilde{B}_n \) is a \( k \times 1 \) vector. Ang and Piazzesi [2003] also show that \( \tilde{A}_n \) and \( \tilde{B}_n \) may be calculated recursively with \( A_1 = -\delta_0 \) and \( B_1 = -\delta_1 \) in the following way:

\[ \tilde{A}_{n+1} = \tilde{A}_n + \tilde{B}_{n-1} \Sigma \lambda_0 + \frac{1}{2} \tilde{B}_{n-1} \Sigma \tilde{B}_{n-1} - \delta_0 \]  
\[ \tilde{B}_{n+1} = \tilde{B}_n (\Phi - \Sigma \lambda_1) - \delta_1 \]  

Finally, defining the continuously compounded yields as \( y_t^n = -\frac{\log P_t^n}{n} \), they can be computed as

\[ y_t^n = A_n + B_n X_t, \quad A_n = -\frac{\tilde{A}_n}{n}, \quad B_n = -\frac{\tilde{B}_n}{n} \]  

4 Data

The data set consists of the monthly average of daily observations between 2006 and 2015 of the continuously compounded Colombian Yield Government bonds and financial variables, as well as macroeconomic aggregates. The zero coupon bond yields are constructed using the methodology from Nelson and Siegel [1987]. I use the bond yields with maturities to 12, 60 and 120 months.

Figure 1 shows the historical behavior of zero coupon bond yields. The bond yields increased between 2006 and 2008 mainly due to the increase of the policy rate and the international risk-perception. After 2008, the bond yields present a negative trend due to several reasons such as: the deepening of the secondary market of the public bonds, sound macroeconomic fundamentals and regulatory issues that increased the portfolio inflows (mainly to government bonds), the increasing demand from foreign investors that was mainly driven by the boost in the international liquidity and the investment grade that the credit rating agencies have assigned the Colombian government bonds since 2010. Table 1 presents some descriptive statistics of the yield curve. The zero-coupon yields are highly persistent. On average, as expected, the yield curve tends to slope upwards, with the volatility reducing as the maturity increases. The volatility ratios, also decrease as the maturity of the yields increase.

The domestic macroeconomic variables used are: inflation, an activity index and the overnight interbank lending rate (IBR) (Figure 2). These variables are in line with the literature shown above (Cortes and Ramos [2008], Ang and Piazzesi [2003], Diebold et al. [2006], Diebold and Li [2006]). The Departamento Administrativo Nacional de Estadística (DANE) provides the inflation and the activity index while the Central Bank of Colombia provides the overnight interbank lending rate.
Figure 1. Yield Curve

This figure shows the term structure of the interest rates for Colombia between 2006 and 2012. Panel left presents the term structure in terms of the maturity and trading dates. Panel right illustrates the dynamics of the interest rates over time.

Table 1. Descriptive Monthly Statistics of Colombian Government Bond Yields

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5.761</td>
<td>2.341</td>
<td>0.985</td>
<td>0.085</td>
<td>1.025</td>
</tr>
<tr>
<td>6</td>
<td>5.009</td>
<td>2.230</td>
<td>0.984</td>
<td>0.155</td>
<td>1.005</td>
</tr>
<tr>
<td>12</td>
<td>6.090</td>
<td>2.364</td>
<td>0.977</td>
<td>0.503</td>
<td>1.015</td>
</tr>
<tr>
<td>24</td>
<td>6.711</td>
<td>2.178</td>
<td>0.963</td>
<td>1.112</td>
<td>0.977</td>
</tr>
<tr>
<td>48</td>
<td>7.603</td>
<td>2.075</td>
<td>0.961</td>
<td>1.097</td>
<td>0.977</td>
</tr>
<tr>
<td>60</td>
<td>7.901</td>
<td>2.027</td>
<td>0.956</td>
<td>2.295</td>
<td>0.909</td>
</tr>
<tr>
<td>84</td>
<td>8.290</td>
<td>1.980</td>
<td>0.951</td>
<td>2.684</td>
<td>0.843</td>
</tr>
<tr>
<td>120</td>
<td>8.479</td>
<td>1.710</td>
<td>0.951</td>
<td>2.873</td>
<td>0.767</td>
</tr>
</tbody>
</table>

This table reports descriptive statistics of bond yields. Statistics are monthly (non-annualized). Yields are annualized. Spreads and volatility ratios are computed relative to the one-month comparable interest rate. Autocorrelations are first-order autocorrelations. The data period is January 2006 - December 2015.

The inflation rate constantly increased between 2006 and 2008, which caused the Central Bank to raise the policy rate in this period (consequently, the interbank overnight interest rate also increased). After 2008, the annual inflation rate moved between the 2%-4% band, which is the Central Bank’s target range, until the recent trend in 2015 when it has reached levels above 4%.

For the monthly economic activity index I use an indicator named “Indice de seguimiento a la economía” provided by the Departamento Administrativo Nacional de Estadística (DANE). The index is calculated by aggregating indicators from the mining, industrial, utilities and agricultural industries in order to fit as closely as possible the quarterly GDP. This indicator shows a negative trend between 2006 and 2009 (mainly driven by the international financial crisis) and a positive trend after 2009 that reversed in 2012. This reversal was probably driven by the policy rate increases between 2010 and 2012. Afterwards this indicator presented a positive trend (in 2013 and 2014) before the decrease in the price of oil and in the terms of trade affected Colombian economy and caused growth expectations to decline.

The interbank lending rate is closely linked to the policy rate. It presents a positive trend between 2006 and 2008 before falling significantly between 2008-2009 due to the policy response to the international financial crisis. Afterwards, this variable presented one complete cycle between 2010 and 2014. The international variables are the long-term interest rate and the sovereign risk premium (5-year CDS premium); the source for these variables is Bloomberg.
Figure 2. Macroeconomic and Financial Variables

This figure shows the data used between 2006 and 2015. The inflation, the activity index, the interbank overnight interest rate and the 10-year treasury rate are shown in percentages. The 5-year CDs series is shown in basic points.

Table 2. Statistics of Economic Variables and its Correlations with Bond Yields

<table>
<thead>
<tr>
<th>Macro</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Min</th>
<th>Max</th>
<th>AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EA</td>
<td>4.530</td>
<td>1.800</td>
<td>-0.110</td>
<td>7.560</td>
<td>0.971</td>
</tr>
<tr>
<td>π</td>
<td>3.900</td>
<td>1.630</td>
<td>1.760</td>
<td>7.940</td>
<td>0.982</td>
</tr>
<tr>
<td>i</td>
<td>5.300</td>
<td>2.310</td>
<td>2.070</td>
<td>10.010</td>
<td>0.989</td>
</tr>
<tr>
<td>Foreign</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i*</td>
<td>3.200</td>
<td>1.030</td>
<td>1.600</td>
<td>5.230</td>
<td>0.964</td>
</tr>
<tr>
<td>CDS</td>
<td>1.880</td>
<td>0.640</td>
<td>1.130</td>
<td>4.060</td>
<td>0.895</td>
</tr>
</tbody>
</table>

| Maturity (M) | Correlation Yields with |     |     |     |
|              | EA | π | i | i* | CDS |
| 6            | 0.152 | 0.007 | 0.308 | 0.571 | 0.448 |
| 12           | 0.191 | 0.393 | 0.037 | 0.607 | 0.428 |
| 24           | 0.213 | 0.860 | 0.947 | 0.680 | 0.398 |
| 48           | 0.150 | 0.795 | 0.826 | 0.749 | 0.434 |
| 60           | 0.114 | 0.767 | 0.821 | 0.763 | 0.442 |
| 84           | 0.065 | 0.740 | 0.790 | 0.754 | 0.436 |
| 120          | 0.029 | 0.783 | 0.794 | 0.702 | 0.512 |

This table reports descriptive statistics of both domestic and foreign covariates used in this research. We also report the correlation of the variables with bond yields data. "AR" refers to first-order autocorrelation. Economic activity, inflation, short-term interest rate, external long-term interest rate, and foreign risk premium interest rate are denoted by EA, π, i, i*, CDS respectively. Statistics are monthly (non-annualized). Yields are annualized. The data period is January 2006 - December 2015.

As shown in Figure 4, the 10 year treasury bill rate had a decreasing trend from 2008 to 2012, given the expansionary monetary policy maintained by the FED. This rate has a positive trend since 2013 due to the end of the Quantitative Easing programs and the speculation on a possible increase of the fed-funds rate.

The 5 years CDS have been relatively stable after the 2008 peak, which is related to financial crisis. However, during 2015, the 5 year-term CDS have reached their highest levels since 2009, mainly due to the uncertainty concerning the effect of the oil price on Colombia’s economy.

Table 2 shows some descriptive statistics of the variables used in the exercises. As shown in the table, macroeconomic variables are highly persistent. The domestic variables show more persistence than foreign variables (this may be related to the financial nature of the foreign variables used in the document). Additionally, yields with longer maturities have lower correlation coefficients with domestic macroeconomic data. Concerning the relationship of the yields with foreign variables, the inverse relationship is found: the yields with the longest maturities have the highest correlation coefficients with the international interest rate and the CDS. Finally, a strong relationship was found between the short term yields and both the interbank overnight interest rate and inflation.
Algorithm 1 Metropolis Hastings Algorithm

Step 0: Set $i = 0$. Initialize with a set of parameters, $\theta^i$. Calculate $L(Y^T | \theta)$ and $p(\theta)$. Set $i=1$.

Step 1: Get a proposal draw $\theta^*$ from $p(\theta)$.

Step 2: Calculate $L(Y^T | \theta^*)$ and $p(\theta^*)$.

Step 3: Accept/Reject $\theta^*$: Draw $X_i \sim U(0,1)$. If $X_i \leq \frac{L(Y^T | \theta^*) p(\theta^*)}{L(Y^T | \theta) p(\theta)}$ accept the proposal and set $\theta^i = \theta^*$. If not, set $\theta^i = \theta^{i-1}$.

If $i < \text{max}_{\text{iterations}}$, set $i = i+1$. If not, the algorithm should stop.

Step 4: If $i > \text{(number of draws)} \times \text{burning rate}$, then save $\theta^i$.

Output: Draws from $\pi(\theta | Y^T)$.

5 Estimation

In this section, I describe the Bayesian estimation process of the parameters of the models. I provide details on the likelihood functions used in the estimation and an in-depth description of the models estimated.

5.1 Bayesian Estimation

The parameters of the models are estimated with Bayesian techniques based on Fernandez-Villaverde et al. [2005] and Fernandez-Villaverde and Rubio-Ramirez [2006]. In the Bayesian estimation, the data are used to update the prior information on the parameters to find the posterior distribution based on the next equation:

$$\pi(\theta | Y^T) \approx p(Y^T | \theta) p(\theta)$$

(17)

where $\theta$ are the parameters of the model and $Y^T$ is the complete set of data. The goal of the Bayesian estimation is to obtain a set of draws of the posterior distribution of the parameters: $\pi(\theta | Y^T)$. From the last equation, $p(\theta)$ refers to the prior distribution of the parameters, and $p(Y^T | \theta) = L(Y^T | \theta)$ denotes the likelihood distribution of the data given the model and the parameters.

To obtain the posterior distribution, I use the Metropolis-Hastings algorithm with 500 draws and a 60% burning rate. Based on Fernandez-Villaverde and Rubio-Ramirez [2006], the pseudo-code for the estimation algorithm is described on Algorithm 1.

Finally, a short discussion of the prior distributions used follows: For the market price of risk, I used a Gaussian distribution with zero mean and unit variance due to the assumptions of the model. As mentioned by Ang and Piazzesi [2003], the market price of risk is an affine assumption of a vector of state variables that are described by a VAR(1) model with Gaussian errors. For the transition matrix, I also used a Gaussian distribution with 0 mean and unit variance as a prior for the non-diagonal elements; as the elements of the diagonal are expected to be bigger than other elements of the matrix, I used a beta distribution for these elements. On the other hand, for the variance-covariance matrix of the state variables, I used an inverse wishart distribution in order to avoid singularity problems in the Kalman Filter algorithm. Finally, for the measurement errors of the yields I use a gamma distribution taking into account the work of Negro and Schorfheide [2008].
5.2 Likelihood Function

As the models assume Gaussian errors, I use the Kalman filter to evaluate the likelihood function as Rabanal and Rubio-Ramirez [2008]. The likelihood function is the following:

\[
\log l(Y^T | \theta) = K - \frac{1}{2} \sum_{i=1}^{T} \log | \Omega_{t|t-1} | - \frac{1}{2} \sum_{i=1}^{T} v_i'^{'} \Omega_{t|t-1} v_t \tag{18}
\]

where \( K \) is a constant and \( T \) refers to the number of periods and:

\[
\Omega_{t|t-1} = E \left( (z_t - z_{t|t-1})(z_t - z_{t|t-1})' | z^{t-1} \right) \tag{19}
\]

\[
v_t = z_t - E(z_{t|t-1}) \tag{20}
\]

However, different papers (Berger and Pericchi, Eklund and Karlsson, Kapetanios et al.) state that good fitting models are not necessarily the best out-of-sample models for forecasting. For this reason, I estimate the parameters of the models using different predictive likelihood functions. The \( h \)-step predictive likelihood function for state space models with Gaussian errors is described in Warne et al. [2013] and Warne et al. [2014]. It is the following extension of the traditional likelihood function:

\[
\log l(Y^{T+h} | Y^T, \theta) = K - \frac{1}{2} \sum_{i=1}^{T-h} \log | \Omega_{t+h|t} | - \frac{1}{2} \sum_{i=1}^{T-h} v_{t+h}^{'} \Omega_{t+h|t} v_{t+h} \tag{21}
\]

where

\[
\Omega_{t+h|t} = E \left( (z_{t+h} - z_{t+h|t})(z_{t+h} - z_{t+h|t})' | z^t \right) \tag{22}
\]

\[
v_{t+h} = z_{t+h} - E(z_{t+h|t}) \tag{23}
\]

The matrices from the last equations can be computed by making additional iterations in each step of the Kalman Filter. To compare the forecast for different horizons, the models are estimated using the traditional likelihood function and the 1, 3, 6 and 12 months predictive likelihood functions.

5.3 Especific Models

This section presents a general description of the models that are analyzed in this thesis.

Bayesian VARs

The state space representation for the Bayesian VARs is the following:

\[
Z_t = 0 + IX_t + 0 \tag{24}
\]

\[
X_t - \mu = \Phi (X_{t-1} - \mu) + \Sigma \varepsilon_t \tag{25}
\]

where \( X_t \) and \( Z_t \) are the vector of observed variables. In this representation, the yields and the observed macroeconomic variables do not have a measurement error.
No-Arbitrage Models - Latent Factors:
This model is composed of three yields (with maturities of 12, 60 and 120 months) and three latent factors. The dynamics of latent factors are supposed to be driven by independent AR(1) processes. The state space representation of this model is the following:

\[
y_t = A + B^f X_t + \eta^y_t
\]

\[
X_t - \mu = \begin{pmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{pmatrix} (X_{t-1} - \mu) + \Sigma \varepsilon_t
\]

where \( X_t = \{LF_1, LF_2, LF_3\} \), \( y_t = \{Y_{12}, Y_{60}, Y_{120}\} \), \( \eta^y_t = \{\eta_1, \eta_2, \eta_3\} \)

and \( \Sigma = \begin{pmatrix} q_{11} & 0 & 0 \\ 0 & q_{22} & 0 \\ 0 & 0 & q_{33} \end{pmatrix} \).

No-Arbitrage Models - Closed Economy and Small Open Economy:
As described previously, the general state space representation of these models is the following:

\[
\begin{bmatrix} y_t \\ m_t \end{bmatrix} = \begin{bmatrix} A \\ 0 \end{bmatrix} + \begin{bmatrix} B^f \\ B^m \end{bmatrix} \begin{bmatrix} f_t \\ m_t \end{bmatrix} + \begin{bmatrix} \eta^y_t \\ 0 \end{bmatrix}
\]

\[
X_t - \mu = \Phi (X_{t-1} - \mu) + \Sigma \varepsilon_t
\]

- The closed economy models contain the following set of variables: \( m_t = y, \pi, i \). \( X_t \) contains \( m_t \) and some models also include one latent factor.
- The small-open economy models contain the following set of variables: \( m_t = y, \pi, i, CDS, i^* \). \( X_t \) contains \( m_t \) and some models also include one latent factor.
- In the representation of the models, if a latent factor is incorporated, \( f_t = \{LF_i\} \) and \( B^f \neq 0 \). If there are no latent factors in a given model, then \( B^f \) and \( f_t \) are not included in the representation.
- The models incorporate one or more yields in the vector \( y_t \); the of \( \eta^y_t \) and its variance, adjusts accordingly. \( \Sigma \) and \( \Phi \) adjust to the size of the state vector.

Bayesian VARs
In Table 3, the name of each model estimated as an unrestricted Bayesian VAR can be found along with the variables it contains (which are marked with a ‘\(^*\)’). The first two models are a closed economy model and a yields-only model and constitute the benchmarks. Next, there is a closed economy model that adds the 12 months-yield and four variations of the small open economy model.
### Table 3. Bayesian VARs

<table>
<thead>
<tr>
<th>Model</th>
<th>A1</th>
<th>π</th>
<th>1*</th>
<th>CDS</th>
<th>Y_12</th>
<th>Y_60</th>
<th>Y_120</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closed [Benchmark]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yieldols [Benchmark]</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closed_Y12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small Open Economy 1 (SOE1)</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small Open Economy 2 (SOE2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small Open Economy 3 (SOE3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small Open Economy 4 (SOE4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table reports the variables that are used in each specific model. All of the models in the table are Bayesian VARs which do not impose the no-arbitrage restriction. The variable column named refers to the activity index.

### No-arbitrage Models

In Table 4, the name of each model estimated using the no-arbitrage condition can be found along with the variables it contains (which are marked with a ✓). The first model is the latent factor model (which was described previously), followed by three variations of the closed economy model and three variations of the small open economy models.

### Table 4. No-arbitrage Models

<table>
<thead>
<tr>
<th>Model</th>
<th>A1</th>
<th>π</th>
<th>1*</th>
<th>CDS</th>
<th>Y_12</th>
<th>Y_60</th>
<th>Y_120</th>
<th>LF_1</th>
<th>LF_2</th>
<th>LF_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent Factors (LF)</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closed_LF1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yield 12 - Closed Economy (12YC)</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Yields - Closed Economy (AYC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small Open Economy - Latent Factor 1 (SOE-L1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Yields - Small Open Economy 1 (AY_SOE1)</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Yields - Small Open Economy 1 (AY_SOE2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table reports the variables that are used in each specific model. All of the models in the table are Bayesian VARs which do not impose the no-arbitrage restriction. The variables named LF_1, LF_2 and LF_3 correspond to latent factors. The variable column named refers to the activity index.

### 6 Results

This section summarizes the results of the models in terms of the accuracy of their forecast. First, the methodology used to obtain the forecasts is shown. Then, the models are compared in terms of the forecast of macroeconomic variables and the yield curve. Finally, the model with the best performance is selected and its results are analyzed in detail.

#### 6.1 Methodology

The description of the process to compare the models is detailed in algorithm 2. The parameters of the models are estimated using the traditional likelihood function and the predictive likelihood function. Afterwards, the h-step forecasts are obtained and are compared in terms of the RMSE and the standard deviation for the correspondent time-horizons.

The process that obtains the h-step forecasts is detailed in algorithm 3. It begins by setting the macroeconomic data in the month of prediction as missing data. Then, it uses the Kalman smoother to recover this data; this consists on the one-month forecast. This way, the process takes into account that financial data is available up to the month of analysis, while macroeconomic data is lagged. After the missing data is recovered, the transition equation is iterated forward using the Kalman filter equations, in order to obtain the forecast of the state variables. Finally, using the prediction of the state variables and the observation equation, the forecast of the observed variables is obtained.
Algorithm 2 Model Comparison Process
For periods 2006-2012 and 2006-2015
for h=1,3,6 and 12
for every model i
    get the draws of $\pi_i(\theta | Y^T)$ using $Logl(Y^{T+h} | Y^T, \theta)$
    from $\pi_i(\theta | Y^T)$ obtain the h-step forecast of $Z_i^t$ since 2013
    compare every variable in $Z^t_i$ to the benchmark model in terms of the RMSE and the standard deviation.

Algorithm 3 Process to Obtain the h-step Forecasts
Input: Draws from $\pi_i(\theta | Y^T)$ for every model i, h and $Z^t_i$
Step 0: Start with $j=1$
Step 1: Set $m_T$ as missing data
Step 2: Take the jth draw of $\theta$ to calculate $z^j_t | t-1$ with the Kalman smoother to obtain the missing data.
Step 3: Iterate h-1 times the transition equation. Obtain $X^j_{t+h-1} | t-1$
Step 4: From $X^j_{t+h-1} | t-1$ and $\pi^j_i(\theta | Y^T)$ obtain $Z^j_{t+h-1} | t-1$.
Step 5: Save $Z^j_{t+h-1} | t-1$. If $i<200$, $i=i+1$, and go back to Step 1. If $i=200$, then stop and go to Step 6.
Step 6: Take the mean of $Z^j_{t+h-1} | t-1$.

6.2 Forecast of Macroeconomic Variables

Table 5, shows the out-of sample results. For every model, the accuracy of the forecast of every macroeconomic variable in the period 2013-2015 is compared relative to the benchmark in terms of the RMSE and the standard deviation (this comparison is made for different horizons). The estimation for each horizon was made with the correspondent predictive likelihood function (for instance the 3-month horizon columns correspond to estimations done using the 3 month-predictive likelihood function). Table 5 also shows the model that has the best forecast for each variable in every time-horizon, in terms of both the RMSE and the standard deviation.

One of the main finding is that the interest rate is consistently forecasted with a higher accuracy when the yield curve (and especially the short-term yield) is included. When the no-arbitrage restriction is imposed, the accuracy of the forecasting of the interest rate improves further and out-performs the benchmark in every model. When the Bayesian VARs are estimated, the forecast of the inflation and the activity index do not consistently out-perform the benchmark. However, when no-arbitrage restrictions are imposed, the accuracy of the forecasts of these two variables improves significantly (the forecasts improve especially for the case of the inflation). These results are related to the findings of Guarin and Murcia [2015] who show that the short-term interest rate and the inflation are closely related to the yield curve.

The results do not improve significantly when the CDS and the international interest rate are added. Apparently, the biggest contributions relative to the benchmark model are made when the 12-month yield is added to closed economy variables and when no-arbitrage restrictions are imposed. Besides, adding the 60 and 120 month yields does not seem to add value to the forecasts (when using unrestricted VARs, adding these variables deteriorates the accuracy of the models). This is also consistent with the findings of Guarin and Murcia [2015], who did not find a strong relationship between the long term yields and macroeconomic variables. Additionally, latent factors do not seem to add value to the out-of sample forecast of the macroeconomic variables, which is also consistent with the literature.
Most of the models under-performed relative to the benchmark in terms of the standard deviation of the forecasts; only the 12YC and the AY_SOE1 models out-performed the benchmark model in most of the cases. Apparently, adding more variables increases the variance of the forecasts. However, it is interesting that the no-arbitrage models consistently out-perform the Bayesian VARs in this item of comparison.

The 12YC model is the best model in terms of the RMSE and the standard deviation. The forecast for all of the macroeconomic variables in this model out-perform the benchmark for every horizon that was compared in terms of the RMSE. Additionally, it consistently out-performs the benchmark for horizons longer than 1 month in terms of the standard deviation. As shown in Table 5, this model consistently out-performs the other models in terms of the forecast of most of the variables, in different forecast horizons. Finally, the lack of accuracy of some models in the forecast of the activity index may be explained the series itself: every month, the historical data of this series is recalculated and smoothed.

Table 5. Out-of-sample Results for Macroeconomic Variables*

<table>
<thead>
<tr>
<th>Unrestricted VAR - Out of Sample</th>
<th>No-Arbitrage Models - Out of Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE (10^-3)</td>
<td></td>
</tr>
<tr>
<td>Closed (Benchmark)</td>
<td>Closed 12M</td>
</tr>
<tr>
<td>1M</td>
<td>12M</td>
</tr>
<tr>
<td>3M</td>
<td>3M</td>
</tr>
<tr>
<td>6M</td>
<td>6M</td>
</tr>
<tr>
<td>12M</td>
<td>12M</td>
</tr>
<tr>
<td>SOE1</td>
<td>SOE1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed (Benchmark)</td>
</tr>
<tr>
<td>1M</td>
<td>1M</td>
</tr>
<tr>
<td>3M</td>
<td>3M</td>
</tr>
<tr>
<td>6M</td>
<td>6M</td>
</tr>
<tr>
<td>12M</td>
<td>12M</td>
</tr>
<tr>
<td>SOE1</td>
<td>SOE1</td>
</tr>
</tbody>
</table>

| Stdv. (10^-3)                   |                                     |
|----------------------------------|                                     |
| Closed (Benchmark)               | Closed 12M                         |
| 1M                              | 12M                                |
| 3M                              | 3M                                 |
| 6M                              | 6M                                 |
| 12M                             | 12M                                |
| SOE1                            | SOE1                               |

This table reports the out-of-sample results of the forecasts of the macroeconomic variables for the different models and for the different horizons. Every horizon was estimated with the correspondent predictive likelihood function.

*The results are multiplied by $10^{-3}$.

** These models were estimated using the traditional likelihood function.
6.3 Forecast of the Yield Curve

In Table 6, the out-of-sample results for all the models that incorporate at least one yield are shown. For every model, the accuracy of the forecast of the yields in the period 2013-2015 is compared relative to the benchmark for different time horizons (3, 6 and 12 months). The comparison for the 1-month period is not shown, given that, as described previously, the first month of prediction treats the yields as observed data. As in Table 5, the estimation for each horizon was made with the correspondent predictive likelihood function and the model that has the best forecast for each variable in every time-horizon, in terms of both the RMSE and the standard deviation is shown.

The Bayesian VARs and the no-arbitrage models do not consistently out-perform the benchmark. However, the forecasts of the short-term yield have a relatively good performance in most of the no-arbitrage models. In line with the previous results, the models with the no-arbitrage restriction have a superior performance compared to the Bayesian VARs.

As in the previous analysis, incorporating small open economy variables and additional yields does not seem to improve the forecasts. However, in the case of the yields, the model with latent factors (Yi_FL) is one of the models with the best performance in terms of the RMSE. As in the macroeconomic variables analysis, the 12YC model is also chosen as the best model in terms of the RMSE and the standard deviation for the forecasting of the short-term yield.

Some of the factors that explain the recent dynamics of the yield curve, which were previously mentioned in the Data section, are not covered in these models. These factors may be imperfectly captured by additional variables such as credit flows to emerging economies, information on the allocation of foreign funds to Colombia’s domestic bond market not reflected in the CDS’s, other measures of international liquidity (such as short-term interest rates) and liquidity in domestic markets. Some of these additional variables were tested in models that were not finally included in this thesis and in the work done by Guarin and Murcia [2015]. The models that incorporated these variables did not out-perform the models described in this document. These variables consist of the VIX, the bid-ask spread and the monthly turnover/outstanding of the secondary market of the domestic public bonds (as liquidity indicators), and the exchange rate, that may also incorporate the impact of international shocks to the local economy. Besides, the EMBI data was not in order to avoid endogeneity effects. Other variables like the capital flows to emerging countries, the LIBOR-OIS spread and the federal funds rate could also be incorporated in the models. However, given the results, I believe that incorporating additional variables to the closed-economy models would not significantly improve the RMSE of the forecasts and would increment its standard deviation. I also believe that most of the impact of these variables on the Colombian yield curve is currently being included through the CDS.
Table 6. Out-of-sample Results for the Yield Curve*

<table>
<thead>
<tr>
<th></th>
<th>Unrestricted VAR - Out of Sample</th>
<th>No-Arbitrage Models - Out of Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3M 6M 12M 3M 6M 12M 3M 6M 12M</td>
<td>3M 6M 12M 3M 6M 12M 3M 6M 12M</td>
</tr>
<tr>
<td>Yields</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{12}</td>
<td>4.6 7.1 22.6 6.6 17.9 15.4 3.6 8.2 34.9</td>
<td>Y_{12} 1.4 9.2 9.4 1.8 1.7 19.9 4.0 6.6 14.6</td>
</tr>
<tr>
<td>Y_{60}</td>
<td>8.7 9.8 27.0 - - - - - -</td>
<td>Y_{60} - - - - - - - - - -</td>
</tr>
<tr>
<td>Y_{120}</td>
<td>10.4 9.5 28.4 - - - - - -</td>
<td>Y_{120} - - - - - - - - - -</td>
</tr>
</tbody>
</table>

*The results are multiplied by 10^3.

6.4 Analysis of the Closed_Y12 Model

Figure 3 shows the fitting of the best model for different likelihood functions. As can be seen, with the exception of the 12 months horizon, the models were able to fit the short-term yield very closely. Besides, Figure 4 shows that the problem with the fitting of the data in the 12-months model affects the forecast of the short-term yield and consequently of the interest rate in the 12-month horizon.
This figure shows the fitting of the selected model. From left to right, the results for the estimations of the short-term yield using the 1, 6 and 12 months predictive likelihood function are shown, respectively.

This figure shows the forecast of the selected model vs. the observed data. From up to bottom, the results are shown for the estimations using the 1, 6 and 12 months predictive likelihood function, respectively.
7 Conclusions

In this thesis, it is shown that the forecasts of Colombian macroeconomic variables significantly improve when the yield curve information is added. Intuitively, these results were expected, as the yield curve contains more up-to-date data. The variables that seem to be more related to the yield curve in terms of the forecast improvement are the interbank overnight interest rate and the annual inflation.

When no-arbitrage restrictions are imposed, the models consistently out-perform the unrestricted VARs. I believe this conclusion could lead to further research on the efficiency of the secondary public bond market. However, the standard deviation of the forecasts of the macroeconomic aggregates increased in most of the models when compared to the benchmark.

Adding the short-term yield and no-arbitrage restrictions were found to be the biggest contributions to the models in order to improve the forecasts of the macroeconomic variables. On the other hand, including small-open economy variables and longer term yields did not significantly improve the results, which is consistent with the evidence provided by Guarin and Murcia [2015].

In line with previous evidence, the models do not have a very good fit when predicting the yield curve; only some models could consistently out-perform the benchmark in terms of the accuracy of the forecasts of the short-term yield. This should be expected in an at least semi-strongly efficient market. According to the semi-strong efficiency hypothesis, current yields include all available public information on the evolution of macroeconomic variables (including market expectations); therefore, macroeconomic variables theoretically should not add value to the forecast of the yield curve. For future research, a general equilibrium model that incorporates market expectations may improve the forecast of both the yield curve and macroeconomic variables.

The accuracy of the forecasts of the short-term yield and the interbank interest rate significantly decreased for the 12-month horizon. For this reason I suggest using these models for the forecasting of the closed economy variables for periods no longer than 6 months.

Finally, the results are in line with the evidence provided by Granger-causality tests, which are described in the Annex. It was found that, between the yields that were analyzed, only the short-term yield appears to have a strong causality relationship with macroeconomic variables. Furthermore, local macroeconomic variables do not appear to have a strong causality over the yield curve (especially when taking into account the longer-term yields).
8 Annex

State-Space Representation of the Closed_y12 model

As described previously in Section 3.2, the state-space representation of the closed_y12 model is the following:

\[
\begin{bmatrix}
y_t \\
m_t
\end{bmatrix} = 
\begin{bmatrix}
A \\
0
\end{bmatrix} m_t + 
\begin{bmatrix}
B^m \\
I
\end{bmatrix} \eta^y_t + \begin{bmatrix}
\eta^y_t \\
0
\end{bmatrix}
\] (30)

\[X_t - \mu = \Phi (X_{t-1} - \mu) + \Sigma \varepsilon_t\] (31)

where \(X_t = m_t = \{y, \pi, i\}\), \(y_t = Y_{12}\) (which means that it is only composed of the 12-months yield).

The parameters that were estimated with the 1 month-predictive likelihood are the following:

\[
\Phi = 
\begin{bmatrix}
0.7839 & -0.2722 & 0.1479 \\
0.0374 & 0.8246 & 0.1093 \\
0.0576 & 0.773 & 0.9236
\end{bmatrix}
\]

\[
\mu = 
\begin{bmatrix}
0.0126 \\
0.0009 \\
-0.0013
\end{bmatrix}
\]

\[
\Sigma = 10^{-4} \times 
\begin{bmatrix}
0.8886 & 0.0689 & 0.0547 \\
0.0689 & 0.0985 & 0.0267 \\
0.0547 & 0.0267 & 0.0835
\end{bmatrix}
\]

\[var(\eta^y_t) = 10^{-4} \times 0.3113\]

\[A = 0.0016\]

\[B = \begin{bmatrix}
0.1688 \\
0.0950 \\
0.8588
\end{bmatrix}\]

Granger causality analysis

Table 7 shows the results of a Granger-casuality analysis taking the 2006-2015 monthly data. The bivariate tests, were made with a 0.01 alpha. In the first column of the Table, the dependent variables are shown. In the following columns, the variables which are tested to have an effect on the dependent variable are shown (if the tests reveal a causality, a check appears in the respective box). Apparently, the CDS and the foreign interest rate have a causality on the yields; between the macroeconomic variables, only the interest rate appears to have a causality on the shorter-term yields.

On the other hand, concerning the causality on the macroeconomic variables, the short-term yield appears to have a causality on both the interest rate and the inflation. Other variables such as a medium-term yield and international variables only have an effect on the interest rate (the long-term yield does not have an effect on any macroeconomic variable). This is in line with the results shown in Section 6 and with the literature evidence on Colombia.
### Table 7. Granger-causality results

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variables</th>
<th>$\pi$</th>
<th>$AI$</th>
<th>$i$</th>
<th>$CDS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{12}$</td>
<td>$Y_{60}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$Y_{120}$</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$\pi$</td>
<td>$Y_{12}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$AI$</td>
<td>$Y_{60}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$i$</td>
<td>$Y_{120}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
References


