

Escuela de Economía y Finanzas

# Documentos de trabajo

## Economía y Finanzas

Centro de Investigación  
Económicas y Financieras

No. 15-23  
2015

### Affine term structure models: forecasting the Colombian yield curve

*Velásquez-Giraldo, Mateo; Restrepo-Tobón, Diego A.*

# Affine term structure models: forecasting the Colombian yield curve

Velásquez-Giraldo, Mateo\*

*Mathematical Engineering, EAFIT University  
Medellín, Colombia*

Restrepo-Tobón, Diego A.†

*Department of Finance, EAFIT University  
Medellín, Colombia*

## Abstract

Superior modeling of the yield curve is useful for asset pricing, financial planning, and risk management. In this article, we estimate five affine term structure models using daily Colombian data. We find that a three-factor model outperforms the other models in one and five days ahead forecasts. The model's factors closely mimic empirical proxies for the level, the slope, and the curvature of the Colombian yield curve.

**Keywords:** Term structure, Forecasting, Interest rates, Multifactor models.

**JEL classification numbers:** G12, G17.

## Resumen

Modelar de forma superior la curva de rendimientos es útil para valoración de activos, planeación financiera y administración de riesgos. En este artículo se estiman cinco modelos afines de la estructura a plazos colombiana usando datos diarios. Se encuentra que un modelo de tres factores tiene un desempeño superior a los demás modelos para pronósticos intra-muestrales y para pronósticos (fuera de la muestra) con horizontes de uno y cinco días. Los factores del modelo se asemejan a sus contrapartes empíricas del nivel, la pendiente y la curvatura de la curva de rendimientos colombiana.

**Palabras clave:** Estructura a plazo, Pronóstico, Tasas de interés, Modelos multifactoriales.

**Números de clasificación JEL:** G12, G17.

---

\*Email adress: [mvelas26@eafit.edu.co](mailto:mvelas26@eafit.edu.co); Corresponding author

†Email adress: [drestri16@eafit.edu.co](mailto:drestri16@eafit.edu.co)

# Affine term structure models: forecasting the Colombian yield curve.

## Contents

<b>I</b>	<b>Introduction</b>	<b>3</b>
<b>II</b>	<b>Literature review</b>	<b>4</b>
<b>III</b>	<b>Affine term structure models</b>	<b>6</b>
<b>IV</b>	<b>Methodology</b>	<b>8</b>
A	The loglikelihood function . . . . .	8
B	Optimization procedure . . . . .	9
C	Forecasts . . . . .	10
<b>V</b>	<b>Data</b>	<b>11</b>
<b>VI</b>	<b>Empirical results</b>	<b>11</b>
A	In-sample fit . . . . .	12
B	Out-sample forecasts . . . . .	14
C	Robustness tests . . . . .	18
<b>VII</b>	<b>Conclusions</b>	<b>30</b>
<b>Appendix A</b>	<b>Plots of the in-sample fit for the <math>A_0(1)</math>, <math>A_1(1)</math>, <math>A_0(2)</math> and <math>A_1(2)</math> models</b>	<b>33</b>
<b>Appendix B</b>	<b>Differential evolution implementation in Matlab®</b>	<b>37</b>

# I. Introduction

The term structure of interest rates (or yield curve for short) plays a central role in an economy. Current yields have useful information for forecasting future short yields and, potentially, real economy activity, inflation, and other key economic variables, Piazzesi (2010). Market participants use these forecasts for pricing financial assets, taking investment decisions, and managing financial risks. Central banks use them to inform monetary policy. Consumers use them to make saving and consumption decisions. Thus, superior modeling and forecasting of the yield curve serve policymakers in evaluating past, current, and future economic conditions and help market participants and consumers in taking better financial decisions.

In this paper, we model and forecast the Colombian daily yield curve using non-arbitrage affine term structure models (ATSMs). The affine term structure modeling framework dominates the theoretical and empirical literature on term structure models, Piazzesi (2010). To our knowledge, this is the first study to test the in-sample fit and the out-sample forecasting capabilities of ATSMs using Colombian data, which is a necessary step to determine the usefulness of ATSMs.

ATSMs specify the risk-neutral evolution of some unobservable factors responsible for the dynamics of the yield curve by making the yields of different maturities an affine (linear) function of those factors. These models provide a flexible structure for examining the dynamics of zero-coupon bond yields by ruling out arbitrage opportunities. They consistently link the cross-sectional and the time-series properties of the yield curve, Piazzesi (2010). The no-arbitrage conditions, in turn, improve the efficiency of the estimates.

Using daily data from 2002 to 2015, we estimate a battery of ATSMs following the estimation methods presented in Ait-Sahalia and Kimmel (2010). We find that a three-factor Gaussian model fits the data and forecasts the daily Colombian yield curve remarkably well. As Dai and Singleton (2000) point out, Gaussian models are fully flexible regarding the signs and magnitudes of conditional and unconditional correlations of the underlying factors but at the cost of assuming constant conditional variances.<sup>1</sup> We find that allowing for conditional heteroscedasticity in the analyzed ATSMs has little effects on the accuracy of the forecasts for our sample data and severely complicate the maximization of the log-likelihood functions.

In our daily data, the three-factor ATSM outperforms the one- and two-factor ATSM. Average yields' root mean squared errors (RMSE) for in-sample, one, and five days ahead forecasts are below twenty basis points. This makes the three-factor ATSM specially appealing for pricing financial instruments, taking investment decisions, and managing financial risk. It also offers the possibility of investigating the yield curve dynamics at higher frequencies than those traditionally used in the literature.

We show that the factors of the estimated three-factor ATSM closely mimic three widely used empirical proxies for the level, the slope, and the curvature of the yield curve. In particular, the third factor has a strong correlation with an average of short- and long-run zero coupon yields. The second factor is highly correlated with the slope of the yield curve

---

<sup>1</sup>Gaussian models allow nominal interest rate to take on negative values. However, this is a limitation of ATSMs in general. They cannot accommodate simultaneously unrestricted correlations among the underlying factors and positive interest rates, Dai and Singleton (2000).

computed as the difference between ten-year and three-month interest rates. The first factor is highly correlated with a curvature empirical proxy computed as twice the four-year yield minus the slope of the yield curve. In addition, principal component analyses produce similar looking level, slope, and curvature factors as those for the three-factors ATSM. Our results are robust to the choice of estimation periods, the use of unsmoothed zero coupon bond yields to estimate the models, and the use of lower frequency data (e.g. weekly and monthly).

The article is organized as follows. Section II reviews the previous literature. Section III presents the ATSMs. Section IV explains the methodology we use for estimation and forecasts. Section V describes the data. Section VI presents and discusses the main empirical results. We conclude in Section VII.

## II. Literature review

Before ATSMs, the dominant framework for explaining the term structure of interest rate was the expectation hypothesis, according to which expected returns are constant through time, Campbell (1986). The liquidity preference and the preferred habitat theories of the term structure of interest rates can be seen as extension of the expectation hypothesis making additional predictions regarding term premiums as a function of the time to maturity of zero coupon bonds. Most tests of the expectation hypothesis reject the existence of constant risk premiums, Campbell and Shiller (1991) and Fama and Bliss (1987). Modeling time-varying risk premiums is at the heart of ATSMs.

ATSMs date back to the work of Vasicek (1977) and Cox et al. (1985). Duffie and Kan (1996) study ATSMs in detail and show how yields for every maturity can be represented as affine functions of some unobserved factors—latent variables. Dai and Singleton (2000) define a canonical representation of ATSMs according to which an  $A_m(N)$ -ATSM includes  $N$  factors,  $m$  of which affect the conditional volatility of the other factors. For instance, an  $A_0(3)$ -ATSM implies that three homoscedastic latent factors explain the dynamics of the yield curve. Likewise, an  $A_1(3)$ -ATSM implies that three latent factors explain the dynamics of the yield curve and that one of the factors determines the conditional volatility of all of them.

Ang and Piazzesi (2003) estimate a discrete Gaussian ATSM incorporating two observable macroeconomic factors: inflation and real activity. Their study finds that macroeconomic factors highly determine the movements of the short and middle ends of the curve, while latent factors are more influential over long yields. They also show, using out-of-sample forecast comparisons, that no-arbitrage cross-equation restrictions make ATSMs more accurate than unrestricted VARs, and that the inclusion of macroeconomic factors further improves their performance. We contribute to the Colombian literature by estimating a benchmark model without macroeconomic factors as a natural step for future research incorporating them.

Estimating ATSMs is challenging. In the literature, different methods exist (e.g. quasi maximum likelihood, Kalman filtering, simulation, and method of moments, among others). Duan and Simonato (1999) propose a state-space representation of ATSMs and approximate the conditional mean and variance under the assumption that the diffusion process is Gaussian. They estimate the latent factors using the Kalman filter method. This allows them

to evaluate the likelihood function (*quasi likelihood* function for non gaussian models) and estimate the parameters of various ATSMs.

Brandt and He (2002) argue that the *quasi maximum likelihood* method is skewed for multi-factor models. They present a correction for the *quasi-likelihood* function, which converges to the real likelihood function and is obtained by simulation. This method reduces the skewness and variability of the estimated parameters, but it is computationally intensive.

More recently, Ait-Sahalia and Kimmel (2010) propose a new method to estimate ATSMs. They use closed-form approximations of the loglikelihood functions for the state variables following the methods presented in Ait-Sahalia (2008). They find that their proposed method generates superior parameter estimates.

ATSMs have been used for several purposes. For instance, Singleton and Umantsev (2002) price options on coupon bonds and swaptions using ATSMs. Ho et al. (2014) propose a method for pricing inflation-indexed derivatives based on ATSMs. Duffee (2002) uses ATSMs to analyze the behavior of expected excess returns. Durham (2006) uses ATSMs to model observed and unobserved components of nominal U.S treasury curves and estimate inflation risk premiums.

Another branch of models currently being used for representing the yield-curve are the dynamic models derived from the Nelson-Siegel equation, originally proposed by Nelson and Siegel (1987) as a curve-fitting tool. Diebold and Li (2006) reparametrize the original model to depend on three dynamic factors associated with *level*, *slope*, and *curvature*. They produce forecasts by fitting autoregressive models to these factors and compare them to several other models. While the proposed model performs poorly for short forecast horizons (1 month), results improve as the horizon grows.

These models are further analyzed by Diebold et al. (2006), who include macroeconomic factors in their regressions and allow for correlated dynamic factors. They use variance decompositions to assess the effects of latent and macroeconomic factors on different ends of the curve, finding a greater influence of the observable factors in short yields.

In a more recent study, Christensen et al. (2011) adjust the dynamic Nelson-Siegel models (both in their independent and correlated factors versions) to be arbitrage-free. They derive a model similar to traditional ATSMs but in which the coefficients of the affine functions that describe yields match the Nelson-Siegel factor loadings. Their results show that the correlated specification has a better in-sample performance, while the simpler uncorrelated factors model produces better out-of-sample forecasts. They also suggest that no-arbitrage restrictions improve the forecast accuracy of the models.

This kind of dynamic Nelson-Siegel models has been widely accepted amongst practitioners. It has paralleled the development of ATSMs as dynamic models of the yield curve. The main difference between both types of models hinges on the no-arbitrage restrictions, but recent studies, such as Christensen et al. (2011), have brought them closer together. A complete review of dynamic Nelson-Siegel models can be found in Diebold and Rudebusch (2013).

For Colombia, most studies model the yield curve focusing on interpolation and curve-fitting of cross-sectional data. The Nelson-Siegel model and cubic splines are the main methodologies used. Few studies use dynamic models, most of them adopting the dynamic Nelson-Siegel framework. Melo-Velandia and Castro-Lancheros (2010) use the methodology in Diebold et al. (2006) to relate monthly Colombian yield data to macroeconomic factors.

They find a strong relation between the model's factors (level, slope and curvature) and the inter-bank rate, inflation, GDP gap, and the EMBI. They show, using Granger causality tests, that macroeconomic variables affect the yield curve factors. Maldonado-Castaño et al. (2014) use the reparametrization of the Nelson-Siegel model presented in Diebold and Li (2006) and apply the Kalman filter to estimate and forecast its factors.

Restrepo-Tobón and Botero-Ramírez (2008) calibrate one-factor arbitrage-free interest rate models to daily Colombian yield curves. Their study concludes that this type of models can closely represent the Colombian term structure of interest rates. To our knowledge, no study has extended this results to multi-factor models and out-of-sample forecasts. We intend to examine these generalizations and forecast capabilities using arbitrage-free ATSMs with daily Colombian data.

### III. Affine term structure models

We denote the yield of a zero-coupon bond with maturity  $\tau$  by  $\gamma_\tau$ . ATSMs assume that the short interest rate is an affine function<sup>2</sup> of a state vector  $X(t)$  of  $N$  underlying factors, which can be observable (macroeconomic variables) or latent, Piazzesi (2010).

$$r = \lim_{\tau \rightarrow 0} \gamma_\tau = \delta_0 + \delta_1^\top X(t) \quad (1)$$

with  $\delta_0 \in \mathbb{R}$  and  $\delta_1 \in \mathbb{R}^N$ .

The state vector is assumed to follow an affine diffusion process under the risk-neutral measure  $Q$ :

$$dX(t) = \tilde{\kappa} \left( \tilde{\Theta} - X(t) \right) dt + \Sigma \sqrt{S(t)} d\tilde{W}_t \quad (2)$$

where  $\tilde{\kappa}, \Sigma \in \mathbb{R}^{N \times N}$ ,  $\tilde{\Theta} \in \mathbb{R}^N$ ,  $\tilde{W}$  is an  $N$ -dimensional independent brownian motion and  $S(t)$  is a  $N \times N$  diagonal matrix with entries:

$$[S(t)]_{i,i} = \alpha_i + \beta_i^\top X(t) \quad (3)$$

with  $\alpha_i \in \mathbb{R}$  and  $\beta_i \in \mathbb{R}^N$ .

The market price of risk  $\Lambda(X) \in \mathbb{R}^N$  is also specified in order to obtain the physical dynamics. Following the literature, we assume that  $\Lambda(t) = \sqrt{S(t)}\lambda$ , where  $\lambda$  is a vector of constants (See Dai and Singleton (2000)). Thus, the state process is also affine under the physical measure  $P$ , Duffie and Kan (1996):

$$dX(t) = \kappa(\Theta - X(t))dt + \Sigma \sqrt{S(t)} dW_t \quad (4)$$

Under this structure, Duffie and Kan (1996) show that the yield for any maturity  $\tau$  can be obtained as an affine function of the state vector:

$$\gamma_\tau(t) = A(\tau) + B(\tau)^\top X(t) \quad (5)$$

---

<sup>2</sup>A function  $F : \mathbb{R}^N \rightarrow \mathbb{R}^M$  is said to be affine if  $F(X) = A + B * X$  for some vector  $A$  and matrix  $B$ .

The coefficients  $B(\tau)$  and  $A(\tau)$  are the solution to the following system of differential equations:

$$\begin{aligned} a'(\tau) &= -\delta_0 + b(\tau)^\top \tilde{\kappa} \tilde{\Theta} + \frac{1}{2} \sum_{i=1}^N [b(\tau)^\top \Sigma]_i^2 \alpha_i \\ b'(\tau) &= -\delta_1 - \tilde{\kappa}^\top b(\tau) + \frac{1}{2} \sum_{i=1}^N [b(\tau)^\top \Sigma]_i^2 \beta_i \end{aligned} \quad (6)$$

with  $a(0) = 0$ ,  $b(0) = \vec{0}$ ,  $A(\tau) = -a(\tau)/\tau$  and  $B(\tau) = -b(\tau)/\tau$ . These equations from imposing no-arbitrage restrictions, Duffie and Kan (1996).

Dai and Singleton (2000) propose a canonical representation in which  $\Sigma$  is an identity matrix. We adopt their notation and representations in this article. As an example the following equations specify the physical dynamics of three-factor models:

The  $A_3(3)$  model:

$$dX(t) = \begin{bmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \left( \begin{bmatrix} \theta_{11} \\ \theta_{21} \\ \theta_{31} \end{bmatrix} - X(t) \right) dt + \begin{bmatrix} \sqrt{X_1(t)} & 0 & 0 \\ 0 & \sqrt{X_2(t)} & 0 \\ 0 & 0 & \sqrt{X_3(t)} \end{bmatrix} dW_t \quad (7)$$

The  $A_2(3)$  model:

$$\begin{aligned} dX(t) &= \begin{bmatrix} \kappa_{11} & \kappa_{12} & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \left( \begin{bmatrix} \theta_{11} \\ \theta_{21} \\ 0 \end{bmatrix} - X(t) \right) dt \\ &+ \begin{bmatrix} \sqrt{X_1(t)} & 0 & 0 \\ 0 & \sqrt{X_2(t)} & 0 \\ 0 & 0 & \sqrt{1 + \beta_{31}X_1(t) + \beta_{32}X_2(t)} \end{bmatrix} dW_t \end{aligned} \quad (8)$$

The  $A_1(3)$  model:

$$\begin{aligned} dX(t) &= \begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \left( \begin{bmatrix} \theta_{11} \\ 0 \\ 0 \end{bmatrix} - X(t) \right) dt \\ &+ \begin{bmatrix} \sqrt{X_1(t)} & 0 & 0 \\ 0 & \sqrt{1 + \beta_{21}X_2(t)} & 0 \\ 0 & 0 & \sqrt{1 + \beta_{31}X_1(t)} \end{bmatrix} dW_t \end{aligned} \quad (9)$$



The  $A_0(3)$  model:

$$dX(t) = - \begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} X(t)dt + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} dW_t \quad (10)$$

See Ait-Sahalia and Kimmel (2010) for a full specification of all the models mentioned in this article.

## IV. Methodology

To estimate the models, we follow the technique for maximum likelihood estimation of ATSMs proposed in Ait-Sahalia and Kimmel (2010) and Ait-Sahalia (2008). They approximate the loglikelihood function of ATSMs by a series of highly accurate expansions for the conditional distributions of the state processes. The resulting density expansion from this approach is in closed form. To estimate the parameter vector  $\theta$  of an  $A_m(N)$  model,  $N$  yields in a panel data of bond yields are assumed to be observed without error. All others yields are assumed to be observed with independent Gaussian errors.

### A. The loglikelihood function

For a given parameter vector  $\theta$ , Eq 6 gives  $A(\tau)$  and  $B(\tau)$  for all  $\tau$ .<sup>3</sup> We then form a vector  $\Gamma_0(\theta)$  ( $N \times 1$ ) whose elements are  $A(\tau)$  and a matrix  $\Gamma(\theta)$  ( $N \times N$ ) whose columns are  $B(\tau)$  for all  $\tau$  (maturities) observed without error. The same is done for maturities observed with errors, obtaining  $\Gamma_{e0}(\theta)$  and  $\Gamma_e(\theta)$ .

A vector containing yields observed without errors  $\gamma_{ne}(t)$  can then be expressed as

$$\gamma_{ne}(t) = \Gamma_0(\theta) + \Gamma(\theta)^\top X(t) \quad (11)$$

using Eq 5. Eq 11 is a linear system with  $N$  equations and  $N$  unknown variables, which allows us to obtain time series of the values of each state variable in  $X(t)$ .

Using these estimated state variables, we calculate the estimated yields observed with error using Eq 5.

As Ait-Sahalia and Kimmel (2010), we denote by  $p_X(\Delta, x|x_0; \theta)$  the conditional density of  $X(t + \Delta) = x$  given  $X(t) = x_0$  and  $p_\gamma(\delta, \gamma|\gamma_0; \theta)$  the transition function of the vector of yields observed without errors. It follows from 11 that:

$$p_\gamma(\Delta, \gamma|\gamma_0, \theta) = \det |(\Gamma(\theta)^\top)^{-1}| p_X \left( \Delta, \hat{X}(t + \Delta) | \hat{X}(t); \theta \right) \quad (12)$$

Where  $\hat{X}(t) = (\Gamma(\theta)^\top)^{-1} (\gamma_{ne}(t) - \Gamma_0(\theta))$ . As the conditional density of the state process is not known in closed form most models, we use the approximations introduced in Ait-

<sup>3</sup>We solve this ODE system using the *ode45(...)* solver in Matlab®R2013A.

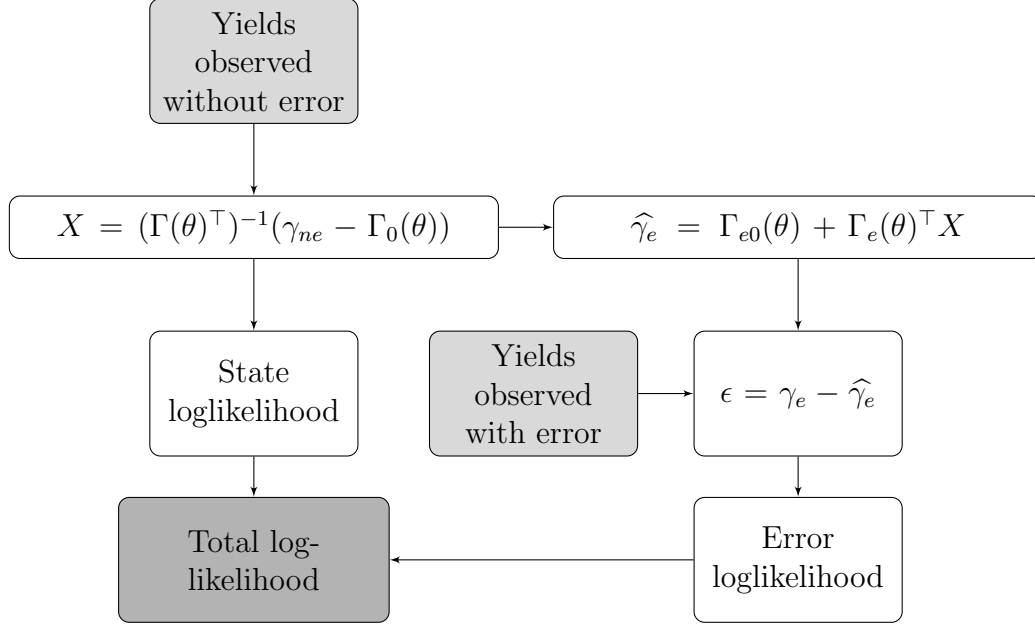


Figure 1: Obtaining the loglikelihood functions. Authors' elaboration based on Ait-Sahalia and Kimmel (2010).

Sahalia (2008). The loglikelihood of yields observed without errors at times  $t_0, \dots, t_n$  with constant time steps of  $\Delta$  is then:

$$\ell(\theta) = \frac{1}{n} \sum_{i=1}^n \ln(p_{\gamma}(\Delta, \gamma_{ne}(t_i) | \gamma_{ne}(t_{i-1}); \theta)) \quad (13)$$

Estimates for yields observed with error are obtained using  $\hat{X}(t)$  and Eq 5. Errors are obtained by comparing the estimates with their observed values and their loglikelihood is computed using the Gaussian density function. The final loglikelihood function, which we maximize to obtain parameter estimates, is computed by adding the loglikelihood of observation errors with the loglikelihood of yields observed without error. Fig 1 summarizes this process.

## B. Optimization procedure

The loglikelihood functions of most ATSMs present multiple local maxima. This makes traditional optimization methods (e.g. gradient based methods) unreliable. In addition, the domain of feasible parameters is restricted for various models. This makes the estimation procedure a difficult task. We turn to heuristic algorithms to maximize the loglikelihood function. In particular, we use the ‘differential evolution’ heuristic approach (Storn and Price (1997)) given its ability to search for optimal parameter values in a continuous space. The algorithm works as follows:

Solutions (sets of parameter values) are treated as vectors. We represent the  $j$ -th solution of a population  $P$  as  $P(j)$ , and its  $i$ -th parameter value as  $P(j)_i$ . ‘Evolution’ is recreated by comparing individuals (solutions) from an initial population with new ones and preserving

**Data:** # Generations: $ng$ , Population size: $np$ ,  $F \in [0, 2]$ ,  $CR \in [0, 1]$   
 $P1 \leftarrow$  Random initial population  
**for**  $i = 1$  **to**  $ng$  **do**  
     $P0 \leftarrow P1$   
    **for**  $j = 1$  **to**  $np$  **do**  
         $\{a, b, c\} \leftarrow$  random individuals from  $P0$   
         $V \leftarrow a + F * (b - c)$   
        **for**  $k = 1$  **to**  $\#Params$  **do**  
            **if**  $rand \leq CR$  **then**  
                 $U_k \leftarrow V_k$   
            **else**  
                 $U_k \leftarrow P0(j)_k$   
            **end**  
        **end**  
        **if**  $fobj(U) \leq P0(j)$  **then**  
             $P1(j) \leftarrow U$   
        **end**  
    **end**  
**end**

**Algorithm 1:** Differential evolution.

the best. New individuals are generated as a linear combination of individuals from the initial population. Before being compared with the initial individual, they can ‘mutate’ by changing some of their parameter values with a given probability. This “evolutionary” process is repeated over numerous generation and the best (according to a given objective function) individual from the last population is taken as the final solution.<sup>4</sup>

### C. Forecasts

ATSMs have mostly been used for monthly and yearly studies of the term structure of interest rates. We intend to evaluate their performance in forecasting the Colombian yield curve at higher frequencies. Short-horizon forecasts of yield movements can aid investors and portfolio managers in building strategies and managing risk. For instance, accurate forecasts could give trading signals or be used to compute measures such as the *value at risk* of a portfolio.

With the estimated parameters, we find the ‘true’ state value for the last ‘known’ date (determined by the horizon) using Eq 5. We then simulate 10,000 state trajectories from this date up to the desired forecast date using Euler’s numeric scheme on Eq 4. For each trajectory, we use Eq 5 to obtain yields for all maturities (observed with and without error). We take our forecasted value as the mean of the 10,000 simulated points for every maturity.

We also report 95% simulated confidence intervals for the forecasted yields. Limits are obtained by computing the 2.5 and 97.5 percentile points of the 10,000 simulated values for

<sup>4</sup>The differential evolution’s pseudo-code is presented in Algorithm 1 and an implementation in Matlab®R2013A is available in Appendix B.

Maturity (years)	0.25	0.50	1.00	2.00	4.00	5.00	7.00	8.00	9.00	10.00
Minimum	1.44	2.02	2.91	3.67	4.08	4.27	4.54	4.65	4.75	4.84
First quartile	3.97	4.10	4.54	5.19	6.22	6.55	6.96	7.10	7.19	7.29
Median	5.23	5.39	5.76	6.60	8.03	8.48	8.85	8.93	8.98	9.02
Third quartile	7.54	8.12	8.99	9.66	10.68	10.98	11.55	11.69	11.84	11.90
Maximum	11.44	11.66	12.88	14.52	16.75	17.10	17.20	17.34	17.44	17.53
Mean	5.76	6.09	6.67	7.56	8.61	8.93	9.33	9.46	9.55	9.63
Variance	4.59	4.82	5.42	6.49	7.58	7.81	8.02	8.06	8.09	8.11
Standard deviation	2.14	2.20	2.33	2.55	2.75	2.79	2.83	2.84	2.84	2.85
Excess kurtosis	-1.07	-1.31	-1.39	-1.04	-0.73	-0.72	-0.74	-0.75	-0.75	-0.75

Table 1: Summary statistics for the Nelson-Siegel data set. The total number of observations is 3051.

every yield. We repeat this procedure for all days in the validation sample.

## V. Data

We use daily zero-coupon yields estimated using the Nelson-Siegel model obtained from *Infoval* between 1/08/2002 and 03/02/2015. The sample includes 3,051 trading days. We use the first 2,000 observations for estimation, reserving the rest for out-of-sample forecasts and validations. Summary statistics for this dataset are presented in Table 1.

We consider maturities of 0.25, 0.5, 1, 2, 4, 5, 7, 8, 9 and 10 years. As pointed out above, for every  $A_m(N)$  model, we assume that the  $N$  yields observed without errors are:

- 3 months yield for one-factor models.
- 3 months and 4 years yields for two-factor models.
- 3 months, 4 years and 10 years yields for three-factor models.

The remaining yields are assumed to be observed with independent Gaussian errors.

In order to check the robustness of the models, we also work with a data set of bootstrapped zero rates obtained from Bloomberg®. These zero coupon yields are unsmooth counterparts of the Nelson-Siegel yields. We use daily observations from 29/04/2005 to 22/05/2015. The first 1,700 observations are used for estimation. Table 2 presents summary statistics for the bootstrapped yield data set.

## VI. Empirical results

We estimate nine ATSMs ranging from one to three factors. Following the notation in Dai and Singleton (2000), we consider the following models:  $A_M(N)$  with  $N, M \in \{1, 2, 3\}$  and  $M \leq N$ .

Maturity (years)	0.25	0.5	1	2	4	5	7	8	9	10
Minimum	2.81	3.00	3.37	3.77	4.12	4.25	4.67	4.78	4.86	4.97
First quartile	3.93	3.99	4.44	5.10	6.14	6.42	6.95	7.11	7.18	7.28
Median	4.77	4.92	5.23	6.01	7.23	7.69	8.33	8.37	8.40	8.41
Third quartile	7.42	7.57	8.06	8.66	9.33	9.72	9.80	9.84	9.82	9.82
Maximum	10.16	10.51	11.07	11.98	12.71	12.93	13.12	13.11	13.10	13.13
Mean	5.67	5.83	6.17	6.79	7.69	8.03	8.45	8.56	8.60	8.62
Variance	4.67	4.71	4.77	4.42	4.02	3.98	3.65	3.44	3.22	3.06
Standard deviation	2.16	2.17	2.18	2.10	2.00	1.99	1.91	1.85	1.79	1.75
Excess kurtosis	-0.94	-0.89	-0.84	-0.75	-0.70	-0.80	-0.66	-0.58	-0.48	-0.41

Table 2: Summary statistics for the bootstrapped data set. The total number of observations is 2625.

The estimation of the models  $A_2(2)$ ,  $A_1(3)$ ,  $A_2(3)$  and  $A_3(3)$  does not converge using our data. This can be attributed to the higher complexity that the feasible solution space acquires as conditional volatility is introduced. We tested numerous optimization procedures which either found no feasible solutions or stagnated on initial parameter values. Out of the tested methodologies, Differential Evolution performed best.

We report results for the models  $A_0(1)$ ,  $A_1(1)$ ,  $A_0(2)$ ,  $A_1(2)$  and  $A_0(3)$ .

### A. *In-sample fit*

The RMSE for the in-sample fit of each model and maturity is presented in Table 3. The parameter values used for the Differential Evolution heuristic are presented in Table 4. A plot of the in-sample fit for the model with the lowest RMSE is presented in this section (all others can be found in Appendix A)

Maturity (Years)	RMSE (basis points)									
	$A_0(1)$		$A_1(1)$		$A_0(2)$		$A_1(2)$		$A_0(3)$	
	NS	BS	NS	BS	NS	BS	NS	BS	NS	BS
0,25	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,5	34,7293	14,3485	33,5475	13,5516	13,3769	18,1574	14,3687	16,4851	9,8554	17,2051
1	84,4390	34,1240	81,5798	30,6891	26,2842	33,4925	25,2418	32,6466	18,3131	31,5518
2	139,5312	54,1757	135,2037	50,1033	25,8577	32,7213	27,4021	31,8013	16,0763	31,8009
4	183,5011	78,8338	178,2805	76,2121	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
5	194,6745	86,6065	189,2136	84,6585	13,4699	30,9987	15,2345	29,7403	4,8226	22,5503
7	211,2276	92,6067	204,9971	93,3573	37,5874	56,8845	38,5727	56,1768	7,3120	29,2402
8	218,4159	90,3075	211,4492	90,2334	48,4840	57,7654	48,7305	57,1240	5,8909	24,5676
9	225,3108	90,6555	217,3645	86,9753	58,8277	57,9986	60,1145	56,3779	3,3835	17,0939
10	232,0046	98,9167	222,8885	89,4189	68,7309	61,6099	74,0213	57,7069	0,0000	0,0000

Table 3: In-sample RMSE between modeled and observed yields. Errors are reported in basis points. Columns marked with **NS** correspond to the Nelson-Siegel yields (2,000 observations) and **BS** indicates bootstrapped rates (1,700 observations).

Parameter	Value
$ng$	2000
$np$	200
$F$	0.4
$CR$	0.9

Table 4: Differential Evolution parameter values used for estimation. The algorithm was stopped if the standard deviation between the objective functions of all solutions in a population was less than 0.1.

One factor models (Figs 11, 12) have an acceptable fit for short yields ( $\tau < 1$ ). They fail to model longer yields as their approximations are replications of the 3 months yield (with different intercepts and amplitudes). Conditional heteroscedasticity (in model  $A_1(1)$ ) reduces the RMSE for all maturities. Including a second factor improves the in-sample fit. The particular characteristics of each yield are captured closely (Fig 13). The  $A_1(2)$  model (Fig 14) has a greater RMSE than the  $A_0(2)$  for most maturities in the Nelson-Siegel dataset.

Figure 2 when compared with Figures 11 - 14 depicts how the the  $A_0(3)$  model outperforms all other models based on its ability to fit the data. Table 3 shows that the in-sample RMSEs for this model are below 18,31 basis points for all maturities. Short and middle-maturity yields obtain higher in-sample errors because of their more platykurtic distributions (Tables 1 and 2). For long-term maturities, the RMSEs are below 7,31 basis points. Thus, based on the RMSE criterion, the  $A_0(3)$  model is the best model, which implies that a three-factor homoscedastic structure best describes the evolution of the Colombian yield curve, out of the models taken into account.

## B. Out-sample forecasts

We conduct out of sample forecasts for one and five days. Tables 5 and 6 report the RMSEs between the mean forecast (of the 10,000 simulated trajectories) and the observed values. The  $A_0(3)$  model outperforms the other models considered in forecasting out-sample yields. RMSEs for in-sample, one and five days ahead forecasts are below twenty basis points. Again, shorter yields (0.5, 1 and 2 years) present higher errors, while longer maturities are more stable and easier to forecast.

The  $A_0(2)$  model performs marginally better than the  $A_0(3)$  forecasting some of the shorter maturities. However, RMSEs for longer yields ( $\tau > 5$ ) are considerably bigger for the  $A_0(2)$ .

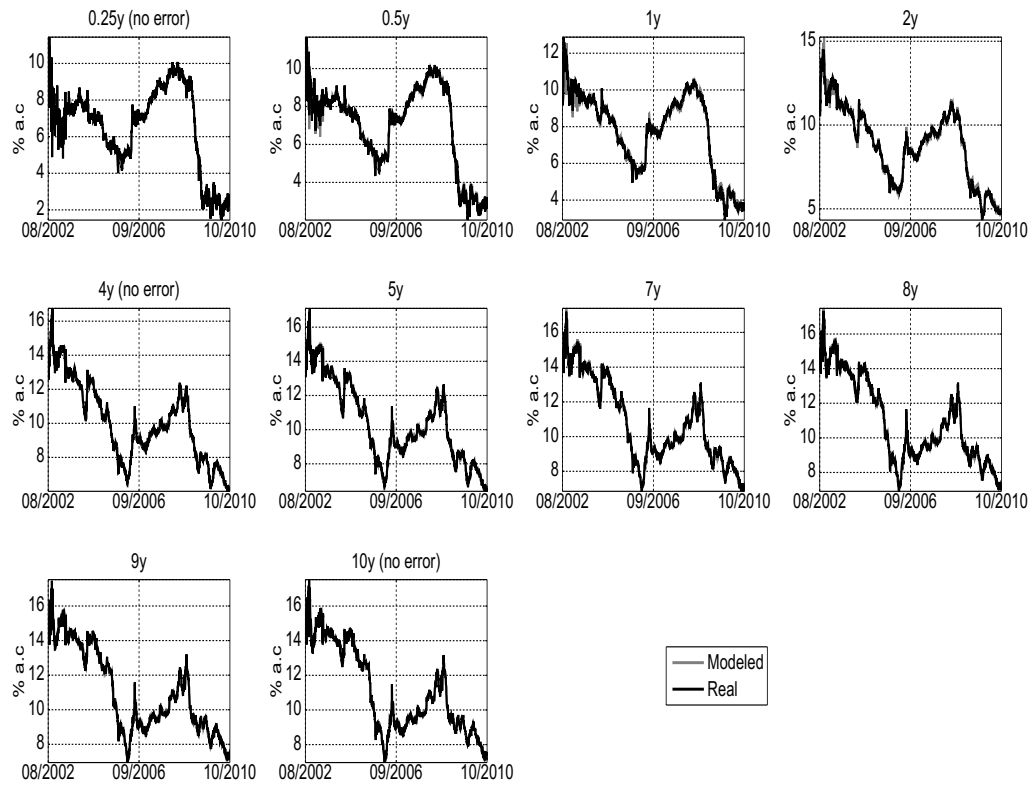


Figure 2:  $A_0(3)$  In-sample fit (Nelson-Siegel data).



Maturity (Years)	RMSE (basis points)									
	$A_0(1)$		$A_1(1)$		$A_0(2)$		$A_1(2)$		$A_0(3)$	
	NS	BS	NS	BS	NS	BS	NS	BS	NS	BS
0,25	8,1601	5,8908	7,9698	5,8018	8,2444	5,8321	91,5271	1531,2320	7,9759	5,8237
0,5	37,9014	15,1482	36,4622	14,3558	11,4499	12,4465	76,9352	1481,7873	12,0973	14,1969
1	100,5005	41,7525	95,2800	40,7645	17,0386	21,3431	74,3880	1410,7824	17,5101	25,1460
2	191,4617	80,9970	182,8747	81,5551	13,4894	21,6089	78,0349	1282,2888	14,2782	24,5432
4	299,7667	143,2359	279,1309	144,1055	6,1066	4,9791	50,3135	1100,0893	6,0965	5,0062
5	333,5070	167,0939	304,3713	165,5840	8,2769	13,2989	37,2752	1029,5393	6,8395	10,1583
7	378,8010	199,4466	333,1596	189,3652	16,1299	29,6337	30,2056	890,7205	7,5100	17,0006
8	394,2433	213,9765	341,5084	198,2714	21,1197	35,2649	34,7320	830,5686	7,2349	11,5639
9	406,5176	229,8556	347,5867	208,2402	26,7962	45,0049	43,2201	779,4247	6,7211	12,5920
10	416,4146	245,3361	352,1222	217,6140	33,0581	55,3252	55,2536	732,6228	6,4229	6,0447

Table 5: One-day out-sample forecast mean RMSE. Errors are reported in basis points. Columns marked with **NS** correspond to the Nelson-Siegel yields (1050 observations) and **BS** indicates bootstrapped rates (924 observations).

Maturity (Years)	RMSE (basis points)									
	$A_0(1)$		$A_1(1)$		$A_0(2)$		$A_1(2)$		$A_0(3)$	
	NS	BS	NS	BS	NS	BS	NS	BS	NS	BS
0,25	15,1925	12,2025	12,4830	11,2027	16,1403	11,5939	428,0572	7642,6494	12,7942	11,5112
0,5	33,1407	20,1558	39,0481	17,4795	13,5451	15,6078	392,5900	7452,2251	14,9358	17,0529
1	95,5089	45,5055	97,3026	41,9420	15,8807	22,8493	354,2918	7112,1345	19,3979	26,6267
2	187,4598	84,2295	184,7206	82,1965	14,8798	23,4962	310,8762	6492,8519	17,9351	27,5118
4	297,0978	145,9109	280,6704	144,5052	14,9246	14,0779	235,1709	5490,7906	14,8562	14,2994
5	331,2655	169,5456	305,7547	165,9627	16,4029	19,6076	207,0488	5077,5200	15,4784	15,9843
7	377,1450	201,5111	334,2807	189,6990	21,1581	33,4648	175,7343	4360,6013	15,8685	21,7465
8	392,7905	215,9600	342,5253	198,6720	24,9186	38,6090	170,9744	4056,5068	15,7456	17,6619
9	405,2278	231,7549	348,5144	208,6865	29,6757	47,8283	172,1168	3786,4570	15,5817	18,2408
10	415,2573	247,0460	352,9735	217,9718	35,2785	57,9955	178,1510	3542,6739	15,6207	16,2657

Table 6: Five-day out-sample forecast mean RMSE. Errors are reported in basis points. Columns marked with **NS** correspond to the Nelson-Siegel yields (1046 observations) and **BS** indicates bootstrapped rates (920 observations).

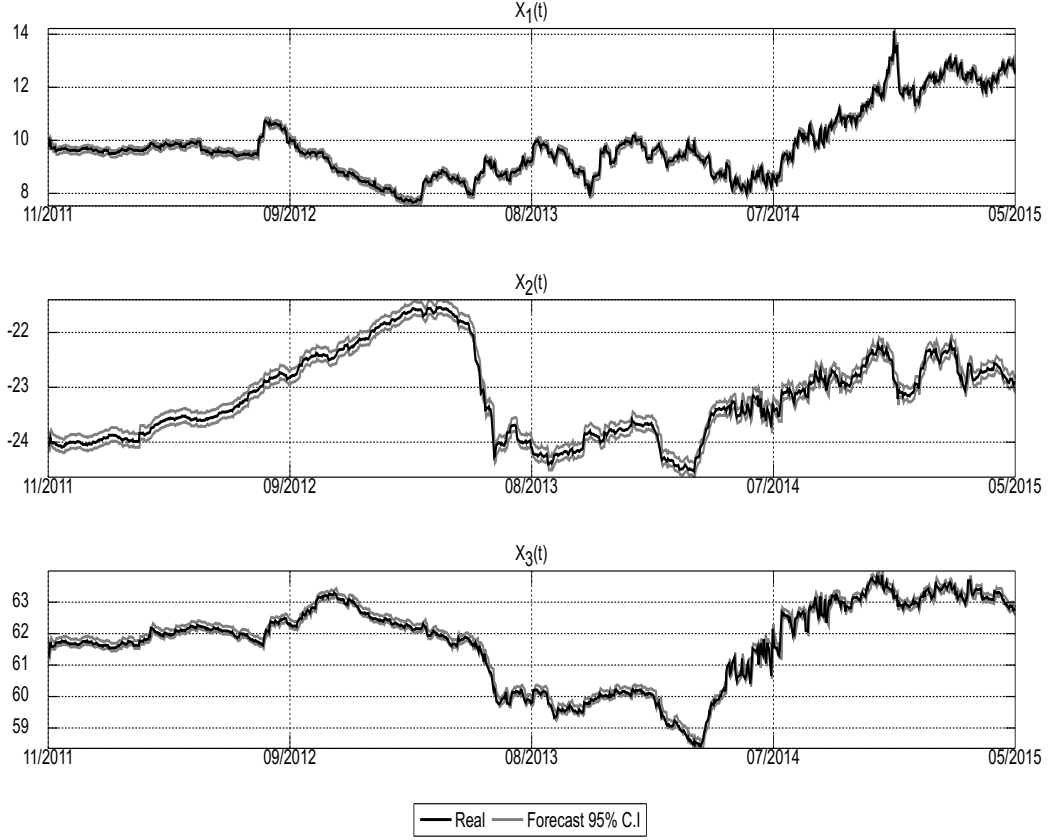


Figure 3:  $A_0(3)$  Out-sample one-day state forecast confidence intervals (Nelson-Siegel data).

Apart from the  $A_0(2)$  and  $A_0(3)$ , the models aren't good at forecasting the specified maturities. One-factor models fall short when adjusting a high number of yields. On the other hand the  $A_1(2)$  model presents very high errors, which might be due to complications in the estimation procedure, introduced by conditional heteroscedasticity.

Figs 3-6 show simulated confidence intervals for states and yields forecasts. Consistently with previous results, short yields have wider confidence intervals and sometimes deviate from them. Longer maturities have narrower intervals and follow them more consistently.

### C. Robustness tests

All the results discussed earlier are based on a data set of yields extracted from market data using the Nelson-Siegel method. These yield curves are smoother than what is normally observed in the market. In order to test the robustness of our results, we also use unsmoothed data obtained from zero coupon yields constructed using the bootstrap method, also known as the unsmoothed Fama-Bliss method (Fama and Bliss (1987)), which iteratively builds the discount rate function by computing the forward rates necessary to price successively longer

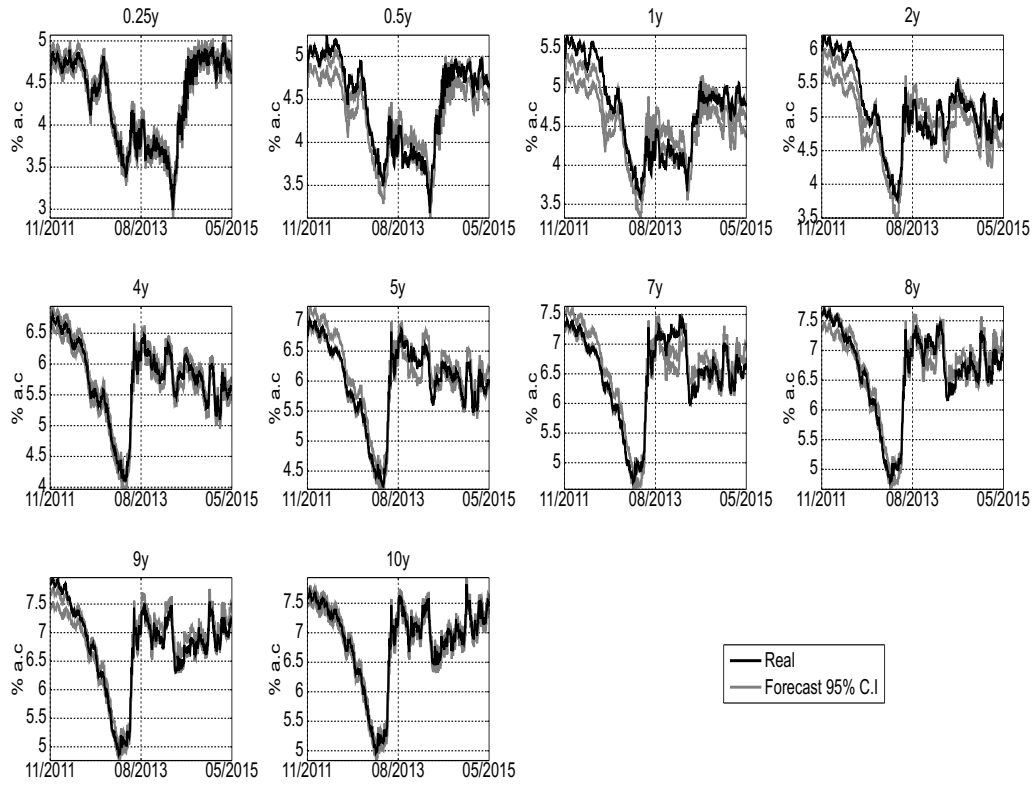


Figure 4:  $A_0(3)$  Out-sample one-day yield forecast confidence intervals (Nelson-Siegel data).

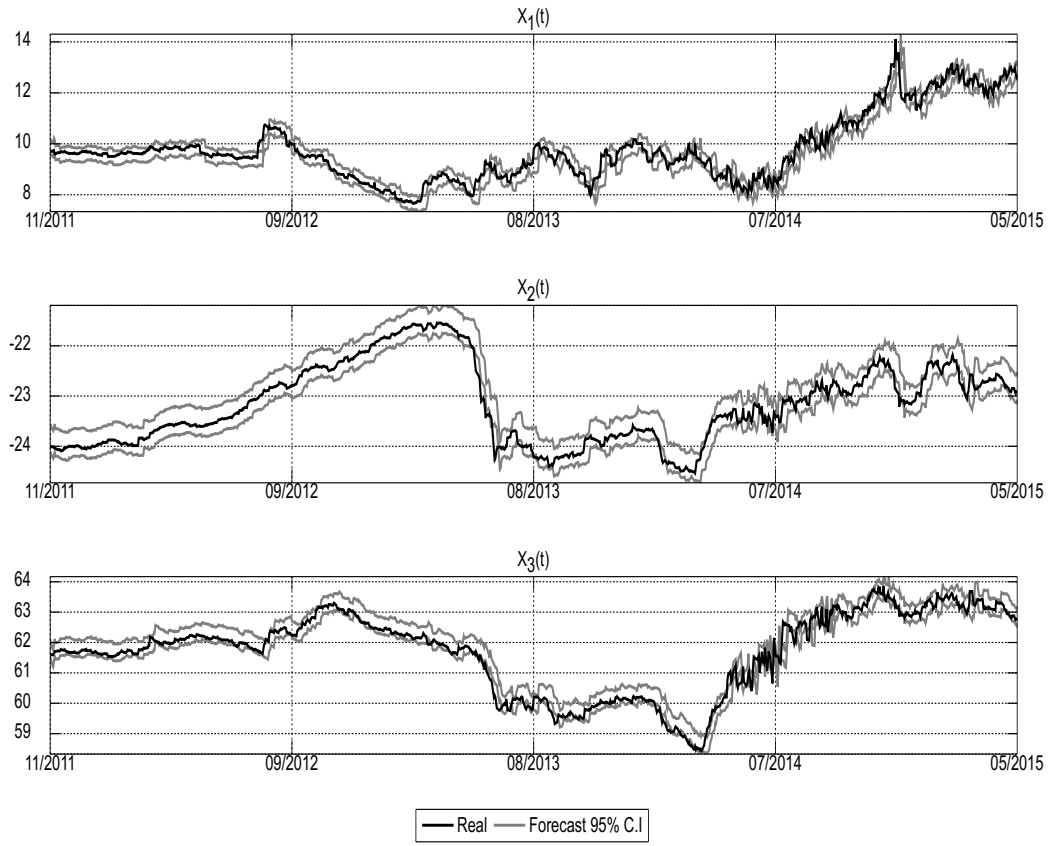


Figure 5:  $A_0(3)$  Out-of-sample five-day state forecast confidence intervals (Nelson-Siegel data).

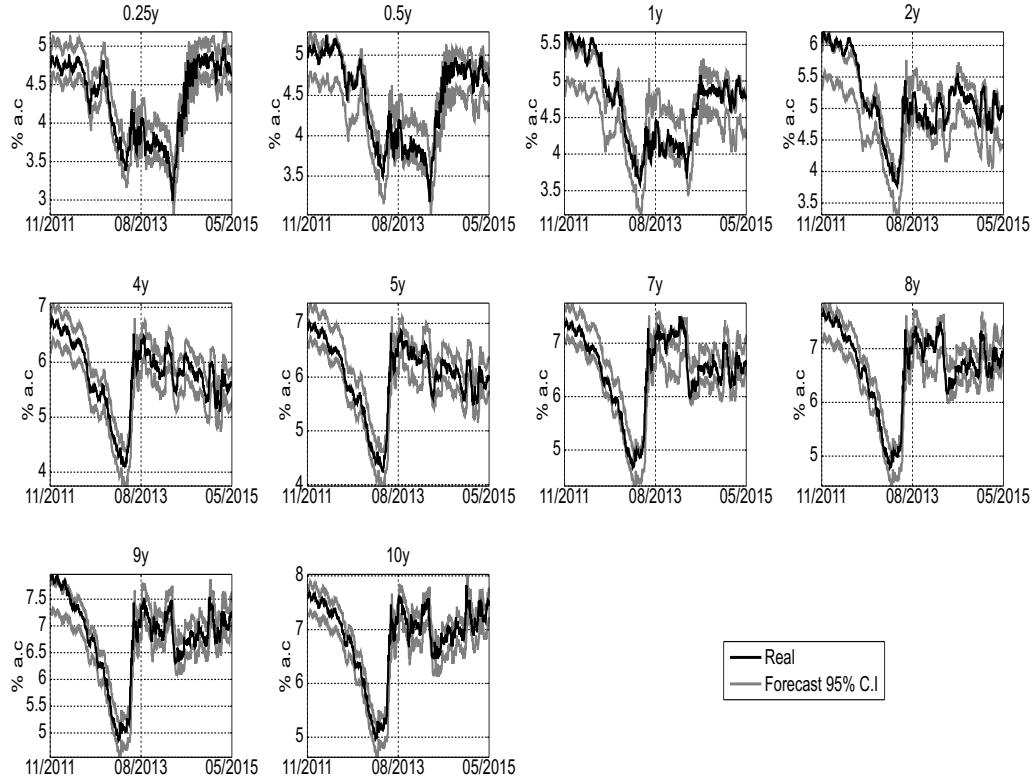


Figure 6:  $A_0(3)$  Out-of-sample five-day yield forecast confidence intervals (Nelson-Siegel data).

Maturity (Years)	1-day forecast RMSE (Basis points)				5-days forecast RMSE (Basis points)			
	Nelson-Siegel		Bootstrapped		Nelson-Siegel		Bootstrapped	
	$A_0(3)$	Random walk	$A_0(3)$	Random walk	$A_0(3)$	Random walk	$A_0(3)$	Random walk
0,25	7,9759	7,9753	5,8237	5,8115	12,7942	12,5073	11,5112	11,2754
0,50	12,0973	7,2085	14,1969	5,1796	14,9358	11,3416	17,0529	10,7431
1,00	17,5101	6,3573	25,1460	4,3725	19,3979	10,5281	26,6267	10,1987
2,00	14,2782	5,3922	24,5432	4,8386	17,9351	11,9459	27,5118	12,3527
4,00	6,0965	6,0908	5,0062	4,9523	14,8562	14,7843	14,2994	13,7999
5,00	6,8395	6,1126	10,1583	5,0093	15,4784	15,0885	15,9843	13,8300
7,00	7,5100	6,0007	17,0006	5,2088	15,8685	15,1914	21,7465	14,2843
8,00	7,2349	6,0505	11,5639	5,3522	15,7456	15,2614	17,6619	14,4649
9,00	6,7211	6,1963	12,5920	5,5129	15,5817	15,3958	18,2408	14,7716
10,00	6,4229	6,4222	6,0447	5,9155	15,6207	15,5942	16,2657	15,0288

Table 7: Comparison of out-of-sample forecast RMSE's between the  $A_0(3)$  model and a random walk benchmark.

maturity bonds.<sup>5</sup>

With the bootstrapped yields, our results change little (Tables 3-6). However, both the in- and out-sample forecasting errors tend to be slightly higher. The  $A_0(3)$  model still outperforms the other models based on in- and out-sample forecasts for all maturities. However, the  $A_0(2)$  model obtains lower errors than the  $A_0(3)$  in forecasting shorter maturity yields. Similar to the Nelson-Siegel yields, the  $A_1(1)$  model has lower RMSEs than the  $A_0(1)$ . The big errors in the  $A_1(2)$  model also persist.

In order to evaluate the  $A_0(3)$  model's forecast capabilities, we compare its out-of-sample results with a random walk benchmark. We chose this simple model as a benchmark, because it has remained hard to beat by models in the literature. For instance, Ang and Piazzesi (2003) find that it outperforms unrestricted VARs, and only manage to slightly beat it (although not for all maturities) using an arbitrage-free model with macroeconomic factors. Duffee (2002) also documents random walks beating ATSMs' forecast capabilities. The 1-day and 5-days out-of-sample forecast RMSE comparisons are presented in Table 7.

Consistently with the literature, the  $A_0(3)$  forecasts are outperformed by random walks. While the shorter end of the curve presents bigger differences in its forecast RMSEs, yields with maturities over four years obtain RMSEs less than two basis points over the random walk benchmark for both 1-day and 5-days ahead forecasts with the Nelson-Siegel data set.

<sup>5</sup>There are other methods to compute zero coupon yields from market data. For instance, the Smoothed Fama-Bliss method makes these discount rates smoothed by fitting a function to the "unsmoothed" rates. The McCulloch method uses a cubic spline with an implicit smoothness penalty. The Fisher-Nychka-Zervos method employs a cubic spline to the forward rate function. The Nelson-Siegel method uses an exponential function for the discount rate function and applies it directly to bond prices. Bliss (1996) shows the Unsmoothed Fama-Bliss method is the most appropriate.

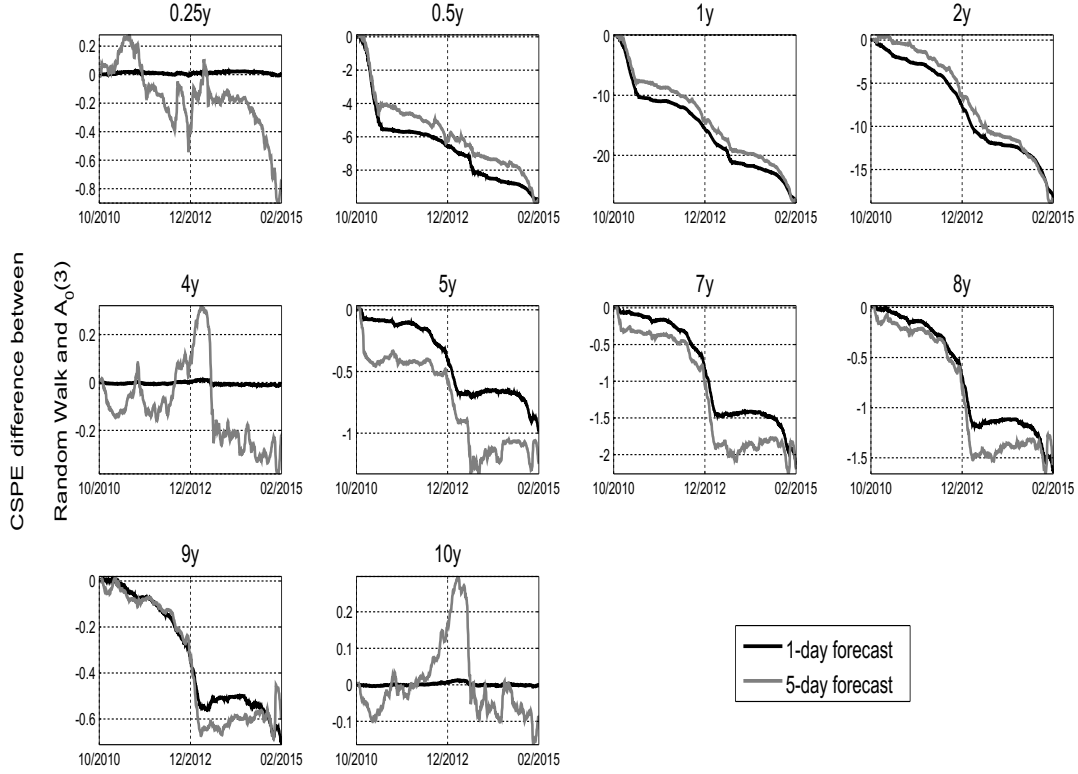


Figure 7: Cumulative squared prediction error difference between a random walk and the  $A_0(3)$  model. These results are obtained using the Nelson-Siegel data set.

The higher forecast errors for short yields can be attributed to the lack of macroeconomic factors in our setting. The literature shows macroeconomic factors help forecasting short and mid-maturity yields (Ang and Piazzesi, 2003).

To further evaluate the performance of the  $A_0(3)$  model against a random walk over time and maturity, we analyze the *cumulative squared prediction error* (CSPE), a metric introduced in Welch and Goyal (2008). The CSPE compares the forecasting performance of a model against a benchmark over time. Calling  $\gamma_i$  the measured value of a yield observed at time  $i$  and  $\hat{\gamma}_i^{RW}$ ,  $\hat{\gamma}_i^{A_0(3)}$  its forecasted values by the random walk and  $A_0(3)$  models respectively, we compute the CSPE as follows:

$$CSPE_t = \sum_{i=1}^t \left[ \left( \hat{\gamma}_i^{RW} - \gamma_i \right)^2 - \left( \hat{\gamma}_i^{A_0(3)} - \gamma_i \right)^2 \right] \quad (14)$$

A positive slope in this metric means the  $A_0(3)$  model outperforms the random walk over a period of time. We obtain the CSPE for all of the maturities over the entire validation sample. Fig 7 presents the CSPE obtained for 1-day and 5-day forecast horizons using the  $A_0(3)$  model estimated with daily observations.



Maturity (Years)	RMSE (basis points)		
	In-sample $A_0(3)$ fit	Out-of-sample 1-week forecast	
		$A_0(3)$	Random walk
0,25	0,0000	23,7651	12,3279
0,50	10,3195	23,3404	11,3276
1,00	18,4792	25,4479	10,3468
2,00	15,5717	23,7895	11,9871
4,00	0,0000	16,9017	15,1659
5,00	4,6319	16,5392	15,6178
7,00	7,1523	16,4781	15,8800
8,00	5,8183	16,3442	15,9748
9,00	3,3527	16,3501	16,1189
10,00	0,0000	16,8046	16,3293

Table 8: Modeled and forecast RSME for the  $A_0(3)$  model estimated with weekly observations. 500 observations were used for estimation and 124 for forecast validations.

Coherently with the results from Table 7, all the CSPEs from Fig 7 have a negative value at the end of the validation sample, indicating that the random walk model had a better accuracy overall. However, various interesting results can be drawn from the behavior of the CSPE metric over time and maturity. For 1-day forecasts, CSPEs are very close to zero for the set of maturities which were assumed to be observed without error for the estimation of the model (0.25, 4 and 10 years). This is not the case for the rest of maturities, which have a steady decline in their CSPEs through most of the sample. This marked difference made by the assumption of an observation error in the estimation procedure repeats itself in the 5-day forecasts: the  $A_0(3)$  is more accurate when forecasting yields observed without error, even surpassing the random walk model for a long period of time around the end of 2012.

When analyzing the effect of maturity on the CSPE, the difference between short and long yields is highlighted again. Yields with maturities inferior to four years have a much steeper descend in their CSPE than longer yields. This causes them to have significantly lower final values for both 1-day and 5-days forecasts. Longer yields' CSPEs fall at a much slower rate and even stagnate after 2012, indicating the  $A_0(3)$  and the random walk are comparably accurate from 2013 to 2015.

In order to check the robustness of our results to the frequency of observation, we reestimate the  $A_0(3)$  model using weekly and monthly observations from the Nelson-Siegel dataset. We follow the same estimation and forecast methodologies that were used for daily observations. Table 8 reports weekly results and Table 9 reports monthly results. Forecasts are obtained with a one-period ahead horizon (one week and one month respectively).

Maturity (Years)	RMSE (basis points)		
	In-sample $A_0(3)$ fit	Out-of-sample 1-month forecast	
		$A_0(3)$	Random walk
0,25	0,0000	37,7977	20,3857
0,50	10,7852	34,2667	19,7754
1,00	19,9255	34,2831	19,7233
2,00	17,4827	34,3616	23,9683
4,00	0,0000	33,1740	31,3470
5,00	5,0450	33,9651	32,9028
7,00	7,6208	34,5065	33,9756
8,00	6,0756	34,3494	34,0358
9,00	3,5034	34,1842	33,9807
10,00	0,0000	34,2838	33,8912

Table 9: Modeled and forecast RSME for the  $A_0(3)$  model estimated with monthly observations. 100 observations were used for estimation and 49 for forecast validations.

Forecast RMSEs are below 26 basis points for the weekly estimations and below 38 basis points for the monthly estimations. As was the case with daily estimations, short term yields have higher errors in the weekly results, but this is not the case with the monthly forecasts.

We also compute CSPEs for both weekly and monthly forecasts. These results are presented in Fig 8. The distinction between yields assumed to be observed with and without error during estimation seems to lose relevance at these lower frequencies. For instance, the four year yield’s CSPEs resemble those from the five year yield very closely, which wasn’t the case for daily observations (see Fig 7). Maturity is the differentiating factor in Fig 8, with short yields yet again having a poor performance, but yields with maturities over four years showing better accuracy with the  $A_0(3)$ . One month forecasts’ CSPEs also show a better behavior than weekly ones, suggesting the lower frequency facilitates forecasting.

Although most studies focus their analyses and forecast tests on lower frequencies, we compare the out-of-sample RMSEs from our article with a few results from literature.

- Duffee (2002) reports RMSEs ranging from 28 to 52 basis points when forecasting U.S yields with maturities up to 10 years with a 3-month forecast horizon using various ATSMs and “essentially affine” term structure models, which are a more flexible variation of ATSMs.
- Ang and Piazzesi (2003) also forecast U.S yields with maturities up to 5 years and obtain RMSEs from 18 to 30 basis points making out-of-sample 1-month forecasts

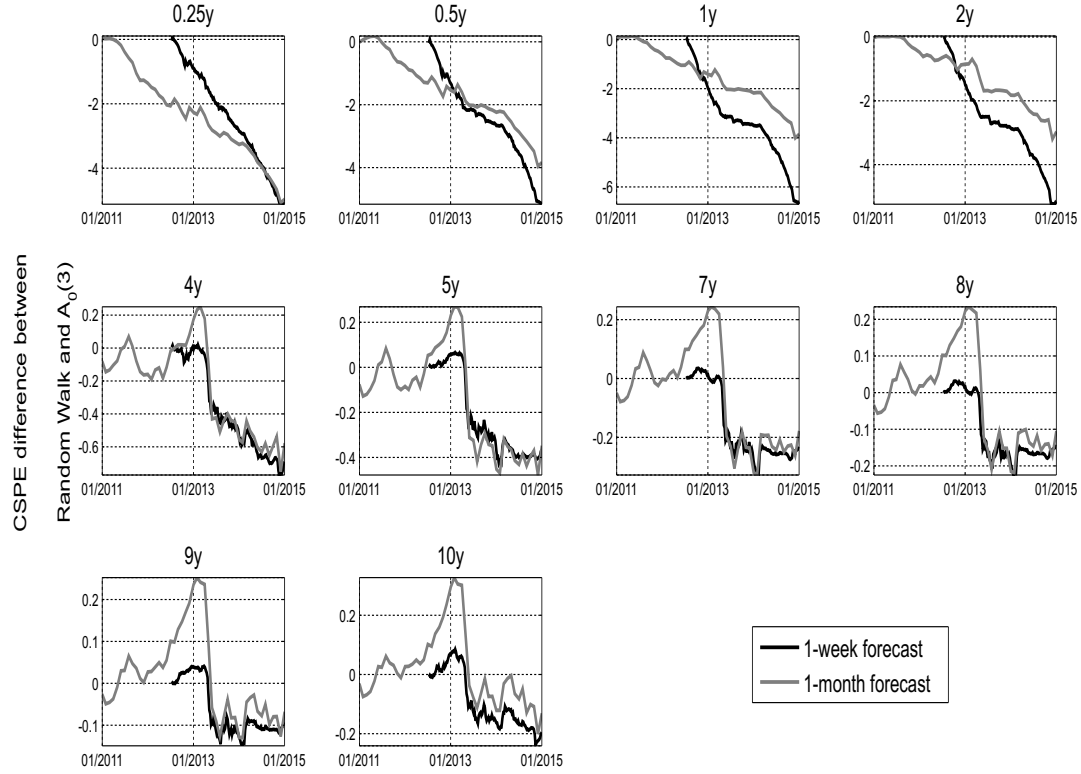


Figure 8: Cumulative squared prediction error difference between a random walk and the  $A_0(3)$  model.  $A_0(3)$  forecasts are obtained using parameters estimated from weekly observations for weekly forecasts and monthly observations for monthly forecasts. These results are obtained using the Nelson-Siegel data set.

	Level	Slope	Curvature
$X_1$	0,7198	-0,7027	0,3667
$X_2$	0,8928	-0,5102	0,7095
$X_3$	-0,9929	0,1166	-0,2949

Table 10: Correlation coefficients between the  $A_0(3)$  factors and the proxies for empirical factors.

and updating estimations at every observation. They manage to lower their errors by including macroeconomic factors in their model.

- Maldonado-Castaño et al. (2014) use a dynamic Nelson-Siegel model, estimated by kalman filtering, to produce one-day yield forecasts in the Colombian market. They achieve RMSEs from 21 to 57 basis points for yields with maturities of 3 months, 3 years and 13 years.

Overall, our forecast RMSEs compare well against these results. Our monthly errors are in the range delimited by Duffee (2002) (which should be taken as an upper bound, considering the longer forecast horizon). One-month RMSEs from Ang and Piazzesi (2003) are lower but close to our results. Finally, our one-day forecasts achieve lower errors than those reported by Maldonado-Castaño et al. (2014) for Colombian data.

We also compare the modeled latent factors from the  $A_0(3)$  model to three empirical proxies for the *level*, the *slope*, and the *curvature* of the yield curve. Following Diebold and Li (2006), we take the following proxies for the empirical factors:

- Level:  $(\gamma_{0.25} + \gamma_4 + \gamma_{10})/3$
- Slope:  $\gamma_{0.25} - \gamma_{10}$
- Curvature:  $2\gamma_4 - \gamma_{10} - \gamma_{0.25}$

We compare these proxies with the modeled latent factors from the  $A_0(3)$  model with the Nelson-Siegel dataset. Results for the unsmoothed yields data set are similar. Table 10 presents correlation coefficients between the estimated  $A_0(3)$ -ATSM's latent factors and the empirical proxies for the level, the slope, and the curvature of the yield curve. The three estimated factors have high correlations with the level of the yield curve. The empirical proxies for the slope and the curvature show a high correlations with the first and second estimated factor, respectively. Fig 9 depicts these relations.

We also compared the three  $A_0(3)$  estimated factors with the first three principal components of the yield curve. Table 11 presents their corresponding correlation coefficients. Fig 10 depicts their relations. As with the empirical proxies, there is a strong correlation between the estimated factors and the three first principal components which together explain close to 96% of the yield curve variance.

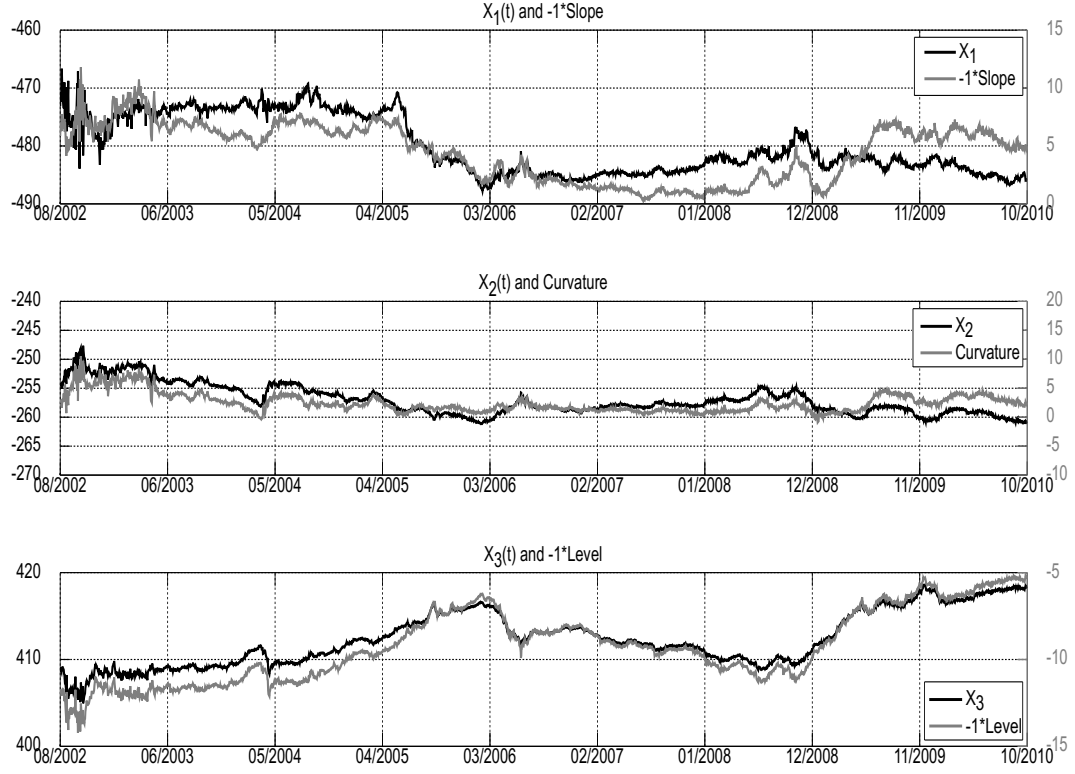


Figure 9: Empirical factor comparison with estimated latent factors for the  $A_0(3)$  model with the Nelson-Siegel data set.

	PC1	PC1	PC3
$X_1$	0,7554	-0,4945	-0,4201
$X_2$	0,9332	-0,2361	0,2683
$X_3$	-0,9807	-0,1919	-0,0267

Table 11: Correlation coefficients between the  $A_0(3)$  factors and the principal components of the yield data.

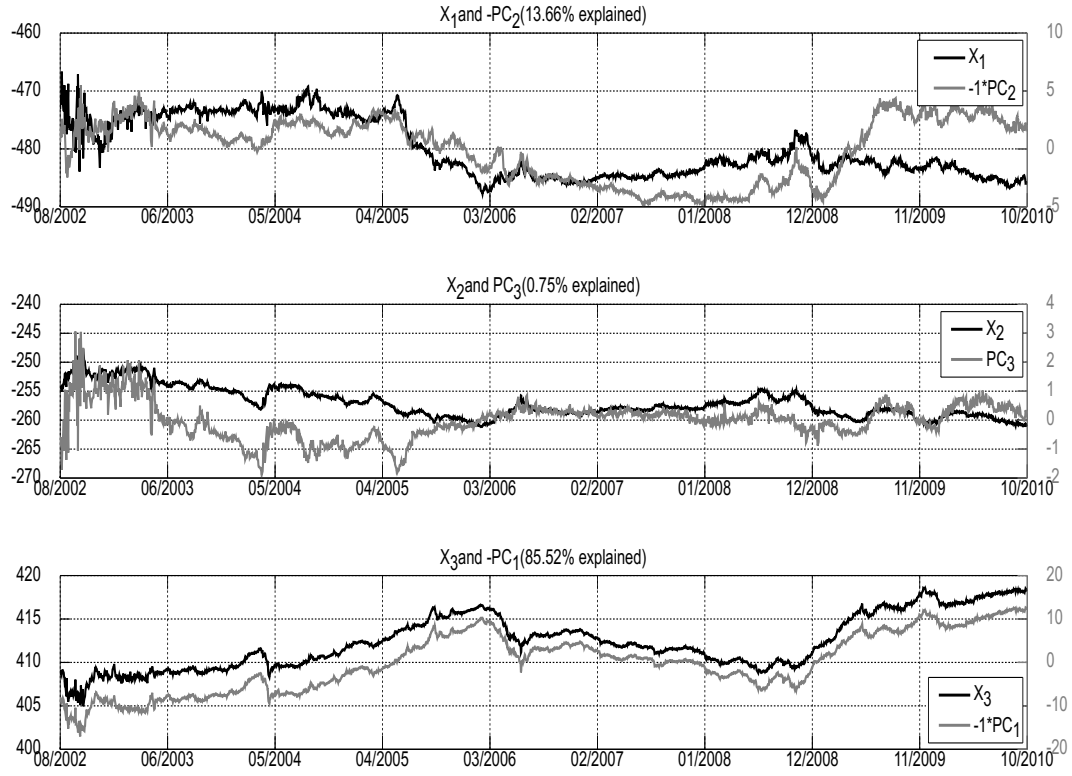


Figure 10: Comparison of the first three principal components of the Nelson-Siegel yields with estimated latent factors for the  $A_0(3)$  model.

## VII. Conclusions

We estimate five ATSMs using daily Colombian data. To our knowledge, this is the first paper applying ATSMs to the Colombian bond market. Our main empirical results indicate that a homoscedastic three-factor ATSM fits the data remarkably well. Our results hold under a series of robustness tests using an alternative data set and data observed at lower frequencies (e.g. weekly and monthly).

We find that the three estimated factors for the ATSM closely mimic the behavior of three empirical proxies for the level, the slope, and the curvature of the yield curve, which according to principal components analyses, account for about 96% of its variance. One- and two-factor ATSMs are unable to describe the behavior of the Colombian yield curve and should not be used in practice.

Our forecasts results have similar errors to those reported in past works and are close to the random walk benchmark. We encounter bigger forecast errors in the short end of the curve, which are highly influenced by macroeconomic factors. Our results can serve as a benchmark for future research in which observable macroeconomic variables could be used. We think, this is a natural next step in our research.

We find evidence supporting the assumption of normally distributed measurement errors on some of the yields disturbs forecasts for said maturities at high frequencies. We hypothesize that as our current estimation methodology is only concerned with the distribution of these errors, but not their minimization, other methodologies could yield better forecast results. These errors become less significant as the forecast horizon grows possibly due to the increment of variability between observations. This may explain why the existing literature, which focuses on lower frequency analyses, has not yet found this assumption to be problematic.

Therefore, the estimation of ATSMs applying different methodologies from the one used in this paper is also a necessary future development of our research. Because of its complexity, the literature on estimation of ATSMs is growing quickly. It is important to assess whether our results and choice of model hold using novel estimation procedures and to find less computationally intensive routines which can ease the use of ATSMs in the Colombian market.

Our work opens the possibility for future research on the relation between macroeconomic factors and the behavior of the yield curve using ATSMs. A promising area for further work is the identification of empirical macroeconomic factors not spanned by the yield curve and that can potentially be useful in forecasting it.

## References

- Ait-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *The Annals of Statistics*, 36(2):906–937.
- Ait-Sahalia, Y. and Kimmel, R. L. (2010). Estimating affine multifactor term structure models using closed-form likelihood expansions. *Journal of Financial Economics*, 98(1):113–144.

- Ang, A. and 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.
- Bliss, R. R. (1996). Testing term structure estimation methods. Technical report, Working paper, Federal Reserve Bank of Atlanta.
- Brandt, M. W. and He, P. (2002). Simulated likelihood estimation of affine term structure models from panel data. Technical report, working paper, Wharton.
- Campbell, J. Y. (1986). A defense of traditional hypotheses about the term structure of interest rates. *The Journal of Finance*, 41(1):183–193.
- Campbell, J. Y. and Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird’s eye view. *The Review of Economic Studies*, 58(3):495–514.
- Christensen, J. H., Diebold, F. X., and Rudebusch, G. D. (2011). The affine arbitrage-free class of nelson–siegel term structure models. *Journal of Econometrics*, 164(1):4–20.
- Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica: Journal of the Econometric Society*, pages 363–384.
- Dai, Q. and Singleton, K. J. (2000). Specification analysis of affine term structure models. *The Journal of Finance*, 55(5):1943–1978.
- Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of econometrics*, 130(2):337–364.
- Diebold, F. X. and Rudebusch, G. D. (2013). *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. Princeton University Press.
- Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of econometrics*, 131(1):309–338.
- Duan, J.-C. and Simonato, J.-G. (1999). Estimating and testing exponential-affine term structure models by Kalman filter. *Review of Quantitative Finance and Accounting*, 13(2):111–135.
- Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, 57(1):405–443.
- Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, 6(4):379–406.
- Durham, J. B. (2006). An estimate of the inflation risk premium using a three-factor affine term structure model. *Federal reserve: Finance and economics discussion series*.
- Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, 77(4):pp. 680–692.



- Ho, H.-W., Huang, H. H., and Yildirim, Y. (2014). Affine model of inflation-indexed derivatives and inflation risk premium. *European Journal of Operational Research*, 235(1):159–169.
- Maldonado-Castaño, R., Zapata-Rueda, N., and Pantoja-Robayo, J. O. (2014). Dynamic estimation of an interest rate structure in Colombia. Empirical analysis using the Kalman filter. *Journal of Economics Finance and Administrative Science*, 19(37):70–77.
- Melo-Velandia, L. F. and Castro-Lancheros, G. A. (2010). Relación entre variables macro y la curva de rendimientos. *Borradores de economía*, (605).
- Nelson, C. R. and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of business*, pages 473–489.
- Piazzesi, M. (2010). Affine term structure models. *Handbook of financial econometrics*, 1:691–766.
- Restrepo-Tobón, D. A. and Botero-Ramírez, J. C. (2008). Modelos unifactoriales de tipos de interés: aplicación al mercado colombiano. *Cuadernos de administración*, 21(36).
- Singleton, K. J. and Umantsev, L. (2002). Pricing coupon-bond options and swaptions in affine term structure models. *Mathematical Finance*, 12(4):427–446.
- Storn, R. and Price, K. (1997). Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2):177–188.
- Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

**A. Plots of the in-sample fit for the  $A_0(1)$ ,  $A_1(1)$ ,  $A_0(2)$  and  $A_1(2)$  models**

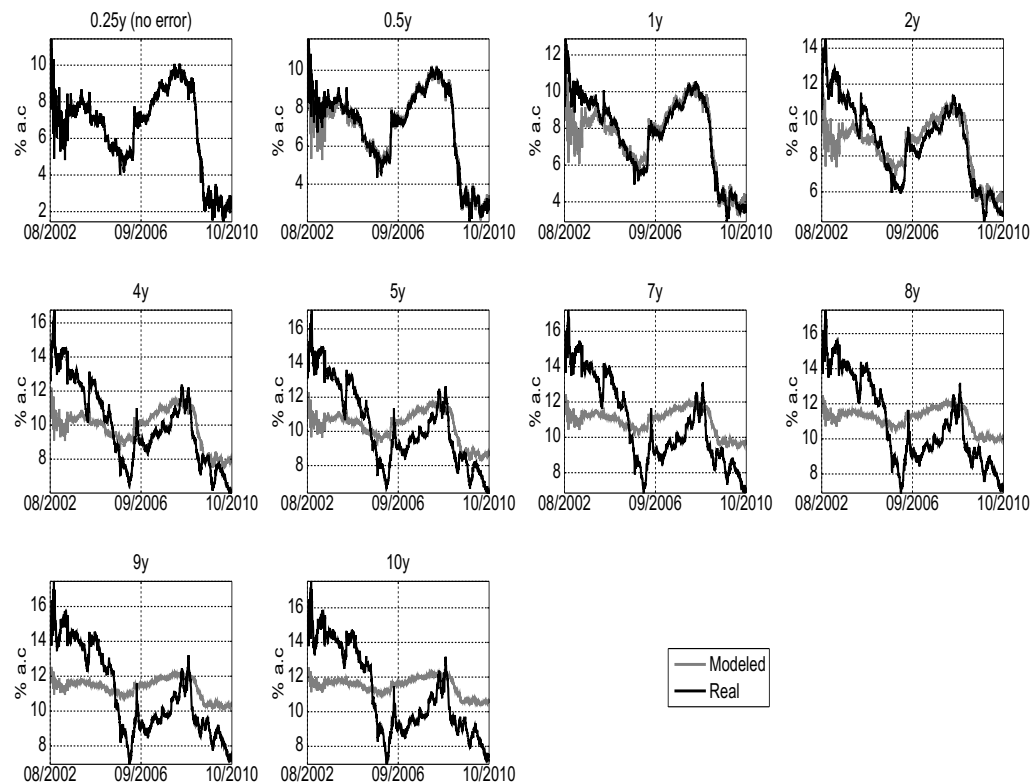


Figure 11:  $A_0(1)$  In-sample fit (Nelson-Siegel data).

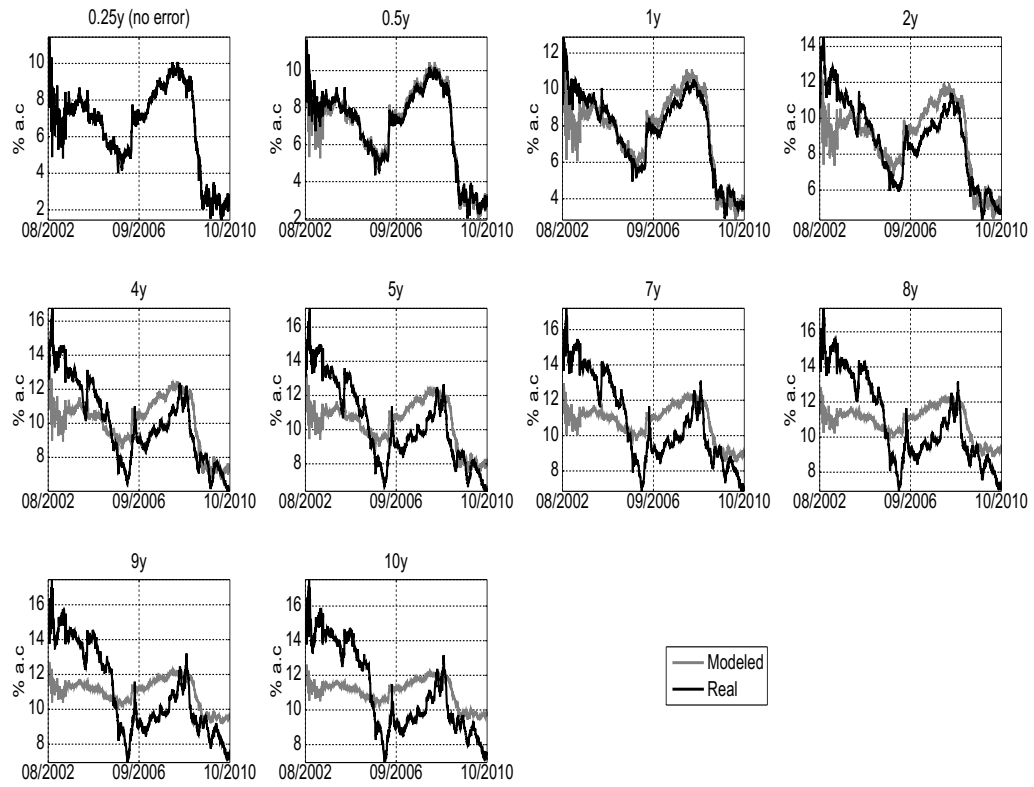


Figure 12:  $A_1(1)$  In-sample fit (Nelson-Siegel data).

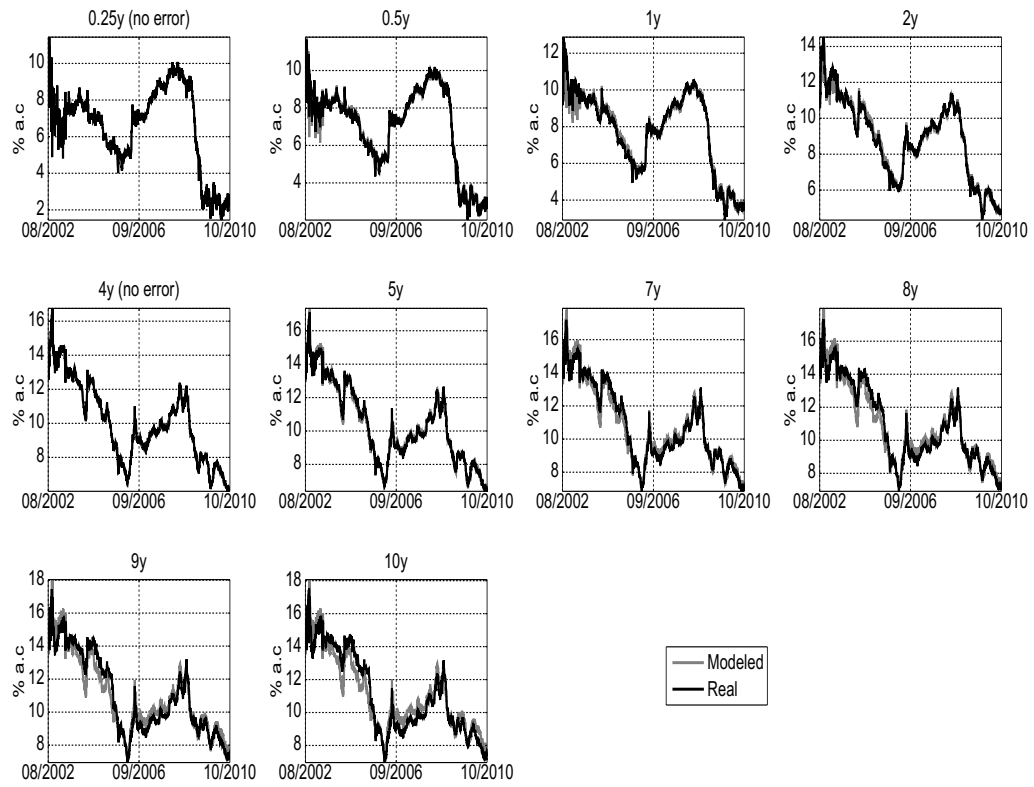


Figure 13:  $A_0(2)$  In-sample fit (Nelson-Siegel data).

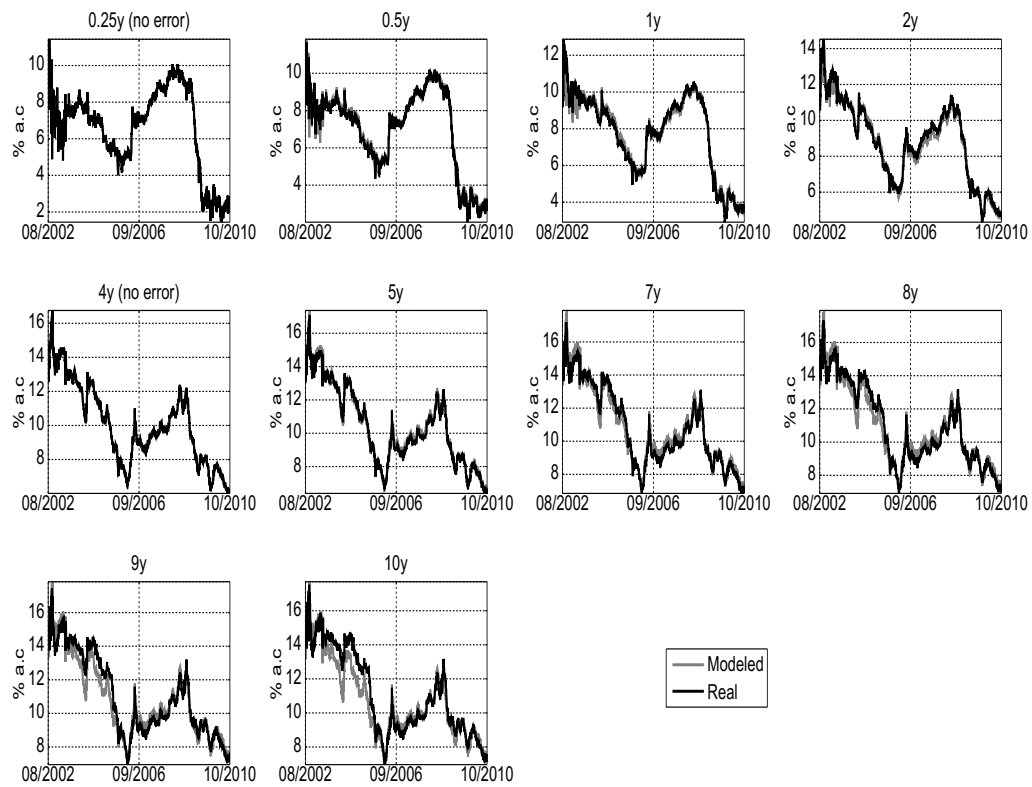


Figure 14:  $A_1(2)$  In-sample fit (Nelson-Siegel data).

## B. Differential evolution implementation in Matlab®

```
function output = DEestimationParTol(fobj,lowPar,uppPar,np,ng,f,cr, tol)
% Mateo Velásquez-Giraldo (mvelas26@eafit.edu.co)
% Universidad EAFIT, Medellín, Colombia.
% June 2015
% Tested with Matlab(R) R2013A.

% Parameters:

% -fobj: objective function which evaluates the "goodness" of a set of
%       parameters. It will be MINIMIZED.
% -lowpar: vector with the lower bounds for the parameters in the
%         initial population.
% -uppPar: vector with the upper bounds for the parameters in the
%         initial population.
% -np: number of solutions in every population.
% -ng: number of generations to be simulated.
% -f: differential weight.
% -cr: crossover probability.
% -tol: tolerance for the standard deviation of the objective functions
%       in the population.

%% Create initial population
numpar = length(lowPar);
p1 = zeros(numpar,np);
for i = 1:numpar;
    p1(i,:) = unifrnd(lowPar(i),uppPar(i),np,1);
end

% Keep objective functions in a vector
objP1 = zeros(1,np);
for i = 1:np
    objP1(i) = fobj(p1(:,i));
end

%% Display:
disp('Diferential evolution:');
disp('*****');
disp('  Generation    Best fobj    S.D Obj');
disp('*****');

bestP1 = min(objP1);
variation=std(objP1);
info = [0,bestP1,variation];
disp(num2str(info));

%% Evolutive cycle
k=0;
while k<=ng && (variation>tol || isnan(variation))
    k=k+1;
    p0 = p1;
    objP0 = objP1;

    % Generate solutions
    parfor i = 1:np
        pU = zeros(numpar,1);
        % Random selection of three solutions
        samp = randsample([1:i-1,i+1:np],3);
        % Differential step
        pV = p0(:,samp(1))+f*(p0(:,samp(2))-p0(:,samp(3)));
        % Crossing
        R = randi(numpar);
        for j = 1:numpar;
            if rand < cr || j == R
                pU(j) = pV(j);
            else
```

```

        pU(j) = p0(j,i);
    end
    end
    % "Goodness" comparison with predecessor
    objSon = fobj(pU(:));
    if(objSon < objP0(i))
        % If the new individual is better, it replaces its
        % predecessor.
        p1(:,i) = pU(:);
        objP1(i) = objSon;
    end
    end
    % Display iteration results
    bestP1 = min(objP1);
    variation = std(real(objP1));
    info = [k,bestP1,variation];
    disp(num2str(info));
end

%% Find the best solution in the final population
I = find(objP1==bestP1);
best = p1(:,I(1));
output = best;
end

```