# Direction-of-Change Yield Curves Forecasts with DNS model and time-varying volatility

Werley Cordeiro[a], João F. Caldeira[b], André A. P. Santos[c]

[a] *Graduate School of Economics – Federal University of Santa Catarina*
[b] *Department of Economics – Federal University of Santa Catarina*
[c] *University of Edinburgh Business School*

**Abstract**

This paper assesses the direction-of-change forecasts based on conditional variance from the Dynamic Nelson-Siegel model. Although the literature focuses on forecasting the level of yield curves, which is a difficult task, we propose forecasts for the direction-of-change of the yield curve returns. The results suggest that models with information of skewness and kurtosis of returns outperform the benchmark model, mainly in long maturities and short horizons of forecasts.

*Keywords:* Direction-of-change forecasts, Yield curves, Nelson-Siegel model, Time-varying volatility, Kalman filter

## 1. Introduction

The importance of forecasts of economic and financial indicators, such as interest rates, inflation expectations, economic activity, and asset returns, has guided governments' decision-making in conducting monetary policy and investors in portfolio allocation. A well-known time series studied for academics, central bankers, and investors is the yield curves or the term structure of interest rates. The yield curve is a graphical representation of interest rates traded on market with different maturities, this curves often unveils agents market expectation on the short-term and long-term of interest rates. In this sense, the literature on yield curves proposes non-arbitrage models, equilibrium models, and factor models to capture information that is not directly observable in the market. Researchers, notably Nelson & Siegel (1987) (henceforth NS), Svensson (1994) and Diebold & Li (2006), investigated and developed parametric interpolation models to fitting and forecasting the yield curve.

The reinterpretation of Diebold & Li (2006) of the NS model, however, explored the model's performance dynamically as a tool for forecasting the yield curve in such a way that the factors of the NS model were interpreted as level (long-term of interest rate), slope (short-term), and curvature (mid-term). Subsequently, Diebold *et al.* (2006) adds backward-looking macroeconomic

variables to the Factor-Augmented DNS model and analyzes its relation with the yield curve's behavior. Fernandes & Vieira (2019) employ a factor-augmented dynamic NS (FADNS) model to predict the yield curve in the US that relies on a large data set of mostly forward-looking macroeconomic variables. The FADNS with forward-looking macro factors significantly improves interest rate forecasts in out-of-sample analysis relative to many extant models in the literature. Vieira *et al.* (2017) suggest the same improvements in forecasting Brazilian yield curve using forward-looking variables.

Alternatively, Koopman *et al.* (2010) developed two major extensions of the DNS model to fit the yield curve, the first one is with time-varying loading factor ($\lambda_t$), since $\lambda_t$ estimation indicates the average maturity of bonds, although the standard procedure in the literature is to fixed $\lambda$. The second is an extension with time-varying volatility, since sometimes the yield curves is more volatile, sometimes is less volatile, a common stylized fact. As well as Diebold *et al.* (2006) and Koopman *et al.* (2010), Yu & Zivot (2011) provides forecasts of the yield curves to the US government bonds in a single step using the Kalman filter, an algorithm to estimate latent factors. Koopman *et al.* (2010) improved the estimation method by introducing the yield curves errors as a common latent factor into state-space framework to be filtered, as well as the loading factor with the Extended Kalman filter. Koopman *et al.* (2010) prefer to preserve the Kalman filter's elegance instead of using other methods such as the Bayesian method to estimate this time-varying parameters.

The accuracy of yield curve forecasts, not only of magnitude but specifically of the direction of change in financial assets, emerges as a research source in the prediction field. Greer (2003), for instance, after conducting tests on the directional accuracy of long-term interest rates forecasts published on *The Wall Street Journal*, suggests these predictions, in general, could be performed with the same precision as flipping a coin, even if there was a combination of the best predictions (see Greer, 2005).

On the other hand, authors such as Christoffersen & Diebold (2006) found that stock returns' volatility forecasts generated a significant forecastability of positive return probability. In other words, they suggest volatility predictability produces predictability in the direction of stock returns. They also suggest it would be more likely to find forecastability in intermediate return horizons such as monthly frequency. The authors argued the results are significant to academic studies and by market practitioners, who usually use *timing* strategies linked to volatility movements (see Rattray & Balasubramanian, 2003). Christoffersen & Diebold (2006) and Christoffersen *et al.* (2006) were the first to provide a rigorous investigation between the conditional volatility dynamics and the positive return probability in financial assets forecasts.

Our goal is to explore the literature gap on forecasting the yield curves using conditional means and conditional volatility forecasts as inputs to predict the direction of return in the fixed income. The main question is, will conditional mean and conditional volatility forecasts, as input to predict

the probability of positive returns, outperform the benchmark model? We think so. Sections two and three present the model to estimate and forecast the yields. The section four present the models to forecast direction-of-change of returns and section six presents data set and results.

## 2. The Dynamic Nelson-Siegel Model

The factors model for the yield curve can represent the forms usually associated with the yield curve, that is, monotonic, curved, and S. Diebold & Li (2006) modified the Nelson & Siegel (1987) model by incorporating time-varying factors in the following way[1]

$$y_{i,t}(\tau_i) = \beta_{1,t} + \beta_{2,t}\left(\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i}\right) + \beta_{3,t}\left(\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}\right), \tag{1}$$

where $y_t$ denotes the yields at time t and $\tau_i$ the maturity of the bond to $\{t\}_{i=1}^{T}$ and $\{i\}_{i=1}^{N}$, respectively. The parameter $\lambda$ determines the exponential decay rate, i.e., small $\lambda$ values results in slow decay and can better fit the curve for longer maturities; on the other hand, large $\lambda$ values produce rapid decay and can better fit the curve for shorter maturities.

In the Diebold & Li (2006) model, $\lambda$ is kept fixed while parameters $\beta_{1,t}$, $\beta_{2,t}$, $\beta_{3,t}$ are estimated by ordinary least squares for each period. Cross-section estimates can be obtained whenever there are sufficient interest rates for different maturities in time. In Diebold *et al.* (2006), on the other hand, $\lambda$ is estimated.

The DNS is the benchmark model, since it is referential for the other extensions in terms of results. In the following Section, I presented the models through the state-space representation.

### *2.1. The Dynamics of the Latent Factors*

Diebold *et al.* (2006) advanced by proposing that the NS model framework can be represented as a state-space model by treating vector $\boldsymbol{\beta_t} = (\beta_{1,t}, \beta_{2,t}, \beta_{3,t})'$ as a latent vector. The model equation can be written as follows

$$\begin{bmatrix} y_t(\tau_1) \\ \vdots \\ y_t(\tau_N) \end{bmatrix} = \begin{bmatrix} 1 & x_{1,2} & x_{1,3} \\ \vdots & \vdots & \vdots \\ 1 & x_{N,2} & x_{N,3} \end{bmatrix} \begin{bmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{bmatrix} + \begin{bmatrix} \varepsilon_t(\tau_1) \\ \vdots \\ \varepsilon_t(\tau_N) \end{bmatrix}, \tag{2}$$

[1]The equation (1) corresponds to equation (2) in the paper of Nelson & Siegel (1987). According to Diebold *et al.* (2006), the following notations are adopted: $\tau$ for maturity instead of $m$, and the loading parameter $\lambda$ equal to $\frac{1}{\tau}$.

where

$$x_{i,2} = \frac{1 - z_i}{\lambda \tau_i}, \quad x_{i,3} = \frac{1 - z_i}{\lambda \tau_i} - z_i,$$

$$z_i = \exp(-\lambda \tau_i).$$

The observation equation in (2) relates the observed interest rates of the $i = 1, \ldots, N$ maturities with the latent factors $\boldsymbol{\beta}_t$.

The vector autoregressive of order 1 of the factors that govern the dynamics of the state equation, is defined as follows

$$\begin{bmatrix} \beta_{1,t+1} \\ \beta_{2,t+1} \\ \beta_{3,t+1} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \phi_{1,3} \\ \phi_{2,1} & \phi_{2,2} & \phi_{2,3} \\ \phi_{3,1} & \phi_{3,2} & \phi_{3,3} \end{bmatrix} \begin{bmatrix} \beta_{1,t} - \mu_1 \\ \beta_{2,t} - \mu_2 \\ \beta_{3,t} - \mu_3 \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \end{bmatrix}, \tag{3}$$

in matrix notation, the equations (2) and (3) can be rewritten as follows

$$\mathbf{y}_t = \boldsymbol{\Lambda}(\lambda)\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, \tag{4}$$

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\boldsymbol{\beta}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \tag{5}$$

where $y_t$ is a vector $N \times 1$, $\boldsymbol{\Lambda}(\lambda)$ is a loading matrix $N \times 3$, $\boldsymbol{\Phi}$ is a VAR(1) parameters matrix $3 \times 1$, $\boldsymbol{\beta}_t$ and $\boldsymbol{\mu}$ are vectors $3 \times 1$. We assumed that $\boldsymbol{\eta}_t$ and $\boldsymbol{\varepsilon}_t$ are orthogonal to each other.

The variance matrix of the observation errors $\boldsymbol{\Sigma}_\varepsilon$ is diagonal. This assumption implies that interest rate deviations for different maturities are not correlated, which facilitates model estimation by reducing the number of parameters. On the other hand, the assumption that the state errors variance matrix $\boldsymbol{\Sigma}_\eta$ is unrestricted allows the shocks in the three factors to be correlated.

We have the representation of the DNS model in the state-space form. In this study, we used the Kalman filter, this algorithm is a recursive procedure to calculate the optimal estimator of the state vector at time $t$, based on the available information at time $t$, and make forecasts for the state vector at $t + 1$ based on $t$. In the following sections, we introduced factor-augmented Nelson-Siegel model.

### *2.2. Time-Varying Volatility*

In the DNS model, I assume that volatility is constant, which may be a flexible assumption since yield curves are related to trading in the financial markets, so volatility changes in these markets may occur over time; in general, heteroscedasticity is a constant problem in economics, especially in finance. The Kalman filter, on the other hand, can not solve this problem, that is, the

filter works under the hypothesis that the variance and covariance matrix is constant, or at least known. Assuming the GARCH structure, the array is unknown; in other words, it is time-varying.

The contribution in this Section and the following Sections is, therefore, to modify the Kalman filter and Extended Kalman filter to incorporate the GARCH approach, so that parameter estimation and volatility estimation are performed in a single step. The DNS model class, therefore, has a common volatility component that is modeled by a univariate GARCH process according to Harvey *et al.* (1992) and Koopman *et al.* (2010). The error vector, in the equation (4), is decomposed as follows

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Gamma}_\varepsilon \varepsilon_t^\star + \varepsilon_t^+, \tag{6}$$

where $\boldsymbol{\Gamma}_\varepsilon$ and $\boldsymbol{\varepsilon}_t^+$ are defined as a vector of weights and an error vector of dimensions N×1, respectively, and $\varepsilon_t^\star$ a scalar error factor. The error components are independent of each other as follows

$$\varepsilon_t^\star \sim \mathcal{N}(0, h_t), \quad \boldsymbol{\varepsilon}_t^+ \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon^+), \quad t = 1, ...T, \tag{7}$$

where $\boldsymbol{\Sigma}_\varepsilon^+$ is a diagonal matrix and $h_t$ is the variance specified as a GARCH process, according to Bollerslev (1986). In this case, I have the following

$$h_{t+1} = \gamma_0 + \gamma_1 \varepsilon_t^{\star 2} + \gamma_2 h_t, \quad t = 1, ..., T, \tag{8}$$

and the estimated parameters have the constraints $\gamma_0 > 0$, $0 < \gamma_1 < 0$, $0 < \gamma_2 < 0$, $h_1 = \gamma_0(1 - \gamma_1 - \gamma_2)^{-1}$ and $(\gamma_1 + \gamma_2) < 1$. The vector of weights $\Gamma_\varepsilon$ can be normalized to avoid identification problems, such that $\Gamma_\varepsilon' \Gamma_\varepsilon = 1$, however, this restriction can be replaced by $\gamma_0$ fixed at $1 \times 10^{-4}$. This last restriction, therefore, I used in estimation. The variance matrix of $\varepsilon_t$ in (8) is time-varying as follows

$$\boldsymbol{\Sigma}_\varepsilon(h_t) = h_t \Gamma_\varepsilon \Gamma_\varepsilon' + \boldsymbol{\Sigma}_\varepsilon^+, \tag{9}$$

where it depends on a single factor described by the GARCH process in (8). The unknown parameters in the GARCH specification, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \Gamma_\varepsilon')'$, are grouped in the parameter vector $\boldsymbol{\theta}$.

The state equation in (5) has one more unobservable component, i.e., $\varepsilon_t^\star$ is now calculated as a latent state. The state-space representation of the observation equation (4) and the state equation

(5) have some modifications as follows

$$\mathbf{y}_t = \underbrace{\begin{bmatrix}\boldsymbol{\Lambda}(\lambda) & \boldsymbol{\Gamma}_\varepsilon\end{bmatrix}}_{\boldsymbol{\Lambda}^\star(\boldsymbol{\lambda})} \underbrace{\begin{bmatrix}\boldsymbol{\beta}_t \\ \varepsilon_t^\star\end{bmatrix}}_{\boldsymbol{\beta}_t^\star} + \varepsilon_t^+, \quad \varepsilon_t^+ \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\varepsilon^+), \tag{10}$$

$$\underbrace{\begin{bmatrix}\boldsymbol{\beta}_{t+1} \\ \varepsilon_{t+1}^\star\end{bmatrix}}_{\boldsymbol{\beta}_{t+1}^\star} = \underbrace{\begin{bmatrix}(I_j - \Phi_j)\boldsymbol{\mu} \\ 0\end{bmatrix}}_{\boldsymbol{\mu}^\star} \underbrace{\begin{bmatrix}\boldsymbol{\Phi}_j & \mathbf{0}_{j\times 1} \\ \mathbf{0}_{1\times j} & 0\end{bmatrix}}_{\boldsymbol{\Phi}^\star} \underbrace{\begin{bmatrix}\boldsymbol{\beta}_t \\ \varepsilon_t^\star\end{bmatrix}}_{\boldsymbol{\beta}_t^\star} + \underbrace{\begin{bmatrix}\boldsymbol{\eta}_t \\ \varepsilon_{t+1}^\star\end{bmatrix}}_{\boldsymbol{\eta}_t^\star}, \tag{11}$$

$$\underbrace{\begin{bmatrix}\boldsymbol{\eta}_t \\ \varepsilon_{t+1}^\star\end{bmatrix}}_{\boldsymbol{\eta}_t^\star} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0} \\ \mathbf{0}\end{bmatrix}, \underbrace{\begin{bmatrix}\boldsymbol{\Sigma}_\eta & \mathbf{0}_{j\times 1} \\ \mathbf{0}_{1\times j} & h_{t+1}\end{bmatrix}}_{\boldsymbol{\Sigma}_\eta^\star}\right), \tag{12}$$

to $t = 1, ..., T$ e $j = 1, 2, 3$ refer to the DNS-GARCH model. Since $h_{t+1}$ in (8) is a function to its past values and unobserved values of $\varepsilon_t^\star$, it is not possible to calculate the values required for $h_{t+1}$ in time t. Harvey *et al.* (1992) propose to replace the square of the error term in (8) by their expected value. Therefore, $h_{t+1}$ can be replaced by its estimate based on observations $y_1, ..., y_t$ as follows

$$\hat{h}_{t+1|t} = \gamma_0 + \gamma_1 \mathbb{E}[\varepsilon_t^{\star 2}|Y_t] + \gamma_2 \hat{h}_{t|t-1}, \quad t = 1, ..., T, \tag{13}$$

in which the expected value can be calculated by the recursions of the Kalman filter using the increased state vector with $\varepsilon_t^\star$ filtered in the last element of vector $\mathbf{b}_{t|t}$, in the equation (17). The expected value follows

$$\mathbb{E}[\varepsilon_t^{\star 2}|Y_t] = \hat{\varepsilon}_{t|t}^{\star 2} + B_{t|t}^{\varepsilon}, \tag{14}$$

where $\hat{\varepsilon}_{t|t}$ is the filtered estimate of $\varepsilon_t$, and $B_{t|t}^{\varepsilon}$ is the variance of $\varepsilon_t$, which are computed for all states during recursions of the Kalman filter, given the observations until period t. Because of the substitution of $\hat{h}_{t|t-1}$ in $h_{t+1}$ in the equation (12), which in the filter is inserted into the $(j, j)$ element of the $\boldsymbol{\Sigma}_\eta^\star$ matrix in the equation (18), the filter and likelihood estimates are sub-optimal (see Harvey *et al.* (1992) for more details).

Therefore the (unconditional) time-varying variance matrix of $y_t$ is $\boldsymbol{\Lambda}^\star(\lambda)\boldsymbol{\Sigma}_{\boldsymbol{\beta}}{}^\star\boldsymbol{\Lambda}^\star(\lambda)' + \boldsymbol{\Sigma}_\varepsilon^\star(h_t)$, where $\boldsymbol{\Sigma}_\beta^\star$ is the solution of $\boldsymbol{\Sigma}_\beta^\star - \boldsymbol{\Phi}^\star\boldsymbol{\Sigma}_\beta^\star\boldsymbol{\Phi}^{\star\prime} = \boldsymbol{\Sigma}_\eta^\star$. To estimate the DNS-GARCH-Macro model, the equation (10), (11), and (12) are increased as appropriate. In the following Section, we presented the procedure for the estimation.

## 3. Estimation Based on the Kalman Filter

The equations model of the (4) and (5) is linear and Gaussian. The estimation is therefore based on the Kalman filter. This algorithm is a recursive procedure that uses the data information at time t to construct state estimates at time $t+1$. The proposed estimation method combines the filter with the Maximum Likelihood (ML) estimation.

The procedure for calculating latent values and unknown parameters is recursive; that is, I started the process by making an initial assumption about the unknown parameters $\boldsymbol{\theta}_1$ to execute the algorithm. The prediction error vector is calculated, $\mathbf{v}_t$, and the prediction error matrix, $\mathbf{F}_t$, in the equations (15) and (16), respectively, to analyze the log-likelihood in equation (21). Updating the state vector $\mathbf{b}_{t|t}$ and the variance matrix $\mathbf{B}_{t|t}$, in the equations (17) and (18), is done in the filtering step in t given the set of information up to $t$. Therefore, consider the model defined in (4) and (5), and define $\mathbf{b}_{t|s}$ as minimum mean squared error linear estimators of $\boldsymbol{\beta}_t$ given $y_t$, ..., $y_s$ to $s = t-1$, $t$ in the following recursion

$$\mathbf{v}_t = \mathbf{y}_t - \boldsymbol{\Lambda}(\boldsymbol{\lambda})\mathbf{b}_{t|t-1}, \tag{15}$$

$$\mathbf{F}_t = \boldsymbol{\Lambda}(\boldsymbol{\lambda})\mathbf{B}_{t|t-1}\boldsymbol{\Lambda}(\boldsymbol{\lambda})' + \boldsymbol{\Sigma}_{\varepsilon}, \tag{16}$$

$$\mathbf{b}_{t|t} = \mathbf{b}_{t|t-1} + \mathbf{B}_{t|t-1}\boldsymbol{\Lambda}(\boldsymbol{\lambda})'\mathbf{F}_t^{-1}\mathbf{v}_t, \tag{17}$$

$$\mathbf{B}_{t|t} = \mathbf{B}_{t|t-1} - \mathbf{B}_{t|t-1}\boldsymbol{\Lambda}(\boldsymbol{\lambda})'\mathbf{F}_t^{-1}\boldsymbol{\Lambda}(\boldsymbol{\lambda})\mathbf{B}_{t|t-1}, \tag{18}$$

$$\mathbf{b}_{t+1|t} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{b}_{t|t} - \boldsymbol{\mu}), \tag{19}$$

$$\mathbf{B}_{t+1|t} = \boldsymbol{\Phi}\mathbf{B}_{t|t}\boldsymbol{\Phi}' + \boldsymbol{\Sigma}_{\eta}, \tag{20}$$

where the parameters in the coefficient matrix of the VAR, $\boldsymbol{\Phi}$, the matrices of variances $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$, the mean vector $\boldsymbol{\mu}$ and the parameter $\lambda$ are are treated as unknown coefficients and grouped into the parameter vector $\boldsymbol{\theta}$, as previously mentioned. The forecast of $\mathbf{b}_{t+1|t}$ and $\mathbf{B}_{t+1|t}$, that is, a step forward is calculated in the filter prediction step in equations (19) and (20). The results of the prediction error vector, $\mathbf{v}_t$, and the prediction error matrix, $\mathbf{F}_t$, re again used as inputs into the log-likelihood function so that the estimate can be conducted to obtain new estimates of the unknown parameters $\theta_2$. These steps are then iterated until the parameters values of the $\boldsymbol{\theta}_{\boldsymbol{MV}}$ are found for which the log-likelihood function is maximized.

The estimation of $\boldsymbol{\theta}$ is based on the numerical maximization of the log-likelihood function via the prediction error decomposition, see Harvey (1989). Therefore, log-likelihood follows by form

$$\log L(Y_n) = -\frac{NT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}(\log|\mathbf{F}_t| + \mathbf{v}_t'\mathbf{F}_t^{-1}\mathbf{v}_t), \tag{21}$$

where $\mathbf{v}_t$ and $\mathbf{F}_t$ are calculated recursively by the Kalman filter (15) to (20) for a given set of $\boldsymbol{\theta}$, such

that log $L(Y_n)$ is computed using the filter result. The calculations required for implementation were made through the R language maintained by the R Core Team (2018), and the minimization of the log-likelihood function was obtained by the `nlminb` optimization function.

The initial parameters were calculated in the estimation in two steps according to Diebold & Li (2006), namely, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_\varepsilon$ diagonal matrices, $\boldsymbol{\Sigma}_\eta$ upper triangular matrix, and $\boldsymbol{\Phi}$ VAR parameters matrix. According to Koopman *et al.* (2010) and Christensen *et al.* (2010), $\mathbf{b}_{1|0}$ and $\boldsymbol{\Sigma}_\beta$ of the model can be calculated as follows, according to the distribution of $\beta_{j,1}$, given by

$$\beta_1 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\beta), \tag{22}$$

in which the unconditional covariance matrix of the state vector, $\boldsymbol{\Sigma}_\beta$, can be started as follows

$$\begin{aligned}
\boldsymbol{\Sigma}_\beta - \boldsymbol{\Phi}\boldsymbol{\Sigma}_\beta\boldsymbol{\Phi}' &= \boldsymbol{\Sigma}_\eta, \\
\operatorname{vec}(\boldsymbol{\Sigma}_\beta) - \operatorname{vec}(\boldsymbol{\Phi}\boldsymbol{\Sigma}_\beta\boldsymbol{\Phi}') &= \operatorname{vec}(\boldsymbol{\Sigma}_\eta), \\
\mathbf{I}_{j^2} \cdot \operatorname{vec}(\boldsymbol{\Sigma}_\beta) - (\boldsymbol{\Phi} \otimes \boldsymbol{\Phi}) \cdot \operatorname{vec}(\boldsymbol{\Sigma}_\beta) &= \operatorname{vec}(\boldsymbol{\Sigma}_\eta), \\
[\mathbf{I}_{j^2} - (\boldsymbol{\Phi} \otimes \boldsymbol{\Phi})] \cdot \operatorname{vec}(\boldsymbol{\Sigma}_\beta) &= \operatorname{vec}(\boldsymbol{\Sigma}_\eta), \\
\operatorname{vec}(\boldsymbol{\Sigma}_\beta) &= [\mathbf{I}_{j^2} - (\boldsymbol{\Phi} \otimes \boldsymbol{\Phi})]^{-1} \cdot \operatorname{vec}(\boldsymbol{\Sigma}_\eta),
\end{aligned} \tag{23}$$

then, under the assumption of a stationary process, the initial value of the state vector is equal to the unconditional mean, $\boldsymbol{\beta}_1 = \boldsymbol{\mu}$, and the initial value of the unconditional covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ is equal to (23).

## 4. Direction-of-Change Models

Let $R_t$ be a series of returns and $\Omega_t$ be the information set available at time $t$. $\Pr[R_t > 0]$ is the probability of a positive return at time $t$. The conditional mean and variance are denoted, respectively, as $\mu_{t+1|t} = \mathrm{E}[R_{t+1} \mid \Omega_t]$ and $\sigma^2_{t+1|t} = \operatorname{Var}[R_{t+1} \mid \Omega_t]$. The return series is said to display conditional mean predictability if $\mu_{t+1|t}$ varies with $\Omega_t$; conditional variance predictability is defined similarly. If $\Pr[R_t > 0]$ exhibits conditional dependence, i.e., $\Pr[R_{t+1} > 0 \mid \Omega_t]$ varies with $\Omega_t$, then we say the return series is sign predictable (or the price series is direction-of-change predictable).

Suppose $\mu_{t+1|t} = \mu$ for all $t$ and $\sigma^2_{t+1|t}$ varies with $t$ in a predictable manner. Denoting $\mathcal{D}(\mu, \sigma^2)$ as a generic distribution dependent only on its mean $\mu$ and variance $\sigma^2$, assume

$$R_{t+1} \mid \Omega_t \sim D\left(\mu, \sigma^2_{t+1|t}\right).$$

Then the conditional probability of positive return is

$$\begin{aligned}\Pr\left(R_{t+1} > 0 \mid \Omega_t\right) &= 1 - \Pr\left(R_{t+1} \leq 0 \mid \Omega_t\right), \\ &= 1 - \Pr\left(\frac{R_{t+1} - \mu}{\sigma_{t+1|t}} \leq \frac{-\mu}{\sigma_{t+1|t}}\right), \\ &= 1 - F\left(\frac{-\mu}{\sigma_{t+1|t}}\right), \end{aligned} \tag{24}$$

where $F$ is the distribution function of the "standardized" return $(R_{t+1|t} - \mu)/\sigma_{t+1|t})$. If the conditional volatility is predictable, then the sign of the return is predictable even if the conditional mean is unpredictable, provided $\mu \neq 0$. Note also that if the distribution is asymmetric, then the sign can be predictable even if the mean is zero: time-varying skewness can be driving sign prediction in this case.

Interaction between volatility and higher-ordered conditional moments can similarly affect the potency of conditional volatility as a predictor of return signs. We follow Christoffersen & Diebold (2006) and use

$$\Pr\left(R_{t+1} > 0 \mid \Omega_t\right) = 1 - F\left(\frac{-\mu_{t+1|t}}{\sigma_{t+1|t}}\right) \tag{25}$$

to explore the sign predictability of one-, -two, three-, and six-month returns in yield curves[2]. We also use an extended version of Equation (25) that explicitly considers the interaction between volatility and higher-ordered conditional moments.

### *4.1. Baseline model*

As Christoffersen *et al.* (2006), we evaluate the forecasting performance of two sets of forecasts and compare their performance against forecasts from a baseline model. Our baseline forecasts are generated using the empirical cumulative distribution function (cdf) of the $R_t$ using data from the beginning of our sample period right up to the time the forecast is made, i.e., at period $k$, we compute

$$\widehat{\Pr}\left(R_{k+1|k} > 0\right) = \frac{1}{k}\sum_{t=1}^{k} I\left(R_t > 0\right), \tag{26}$$

where $I(\cdot)$ is the indicator function.

[2]where $\hat{\mu}_{t+h}^{(\tau_i)} = \tau_i y_t^{\tau_i} - (\tau_i - h)\,\hat{y}_{t+h}^{\tau_i - h}$ is a return forecast for the bond with maturity $\tau_i$ in time $t$ and $\Sigma_{\hat{\mu}_{t+h}} = \tau'\tau \otimes \Sigma_{y_{t+h}}$ is their conditional covariance matrix.

*4.2. Non-parametric model*

Our first forecasting model makes direct use of Equation (26). Using all available data at time $k$, we first regress $R_t$ on a constant, $\log(\hat{\sigma}_t)$, and $[\log(\hat{\sigma}_t)]^2$, and compute

$$\hat{\mu}_t = \hat{\beta}_0 + \hat{\beta}_1 \log(\hat{\sigma}_t) + \hat{\beta}_2 [\log(\hat{\sigma}_t)]^2, \quad t = 1, \ldots, k \tag{27}$$

where $\hat{\sigma}_t$ is the square root of (actual, not forecasted) realized volatility. The period $k+1$ forecast is then generated by

$$\begin{aligned} \widehat{\Pr}(R_{k+1|k} > 0) &= 1 - \hat{F}\left(-\frac{\hat{\mu}_{k+1|k}}{\hat{\sigma}_{k+1|k}}\right), \\ &= 1 - \frac{1}{k}\sum_{t=1}^{k} I\left(\frac{R_t - \hat{\mu}_t}{\hat{\sigma}_t} \leq \frac{\hat{\mu}_{k+1|k}}{\hat{\sigma}_{k+1|k}}\right), \end{aligned} \tag{28}$$

i.e., $\hat{F}$ is the empirical cdf of $(R_t - \hat{\mu}_t)/\hat{\sigma}_t$. The forecasts of conditional mean $\hat{\mu}_{k+1|k}$ and conditional variance $\hat{\sigma}_{k+1|k}$ are from all DNS models and extensions.

*4.3. Extended model*

The second model is an extension of Equation (25) and explicitly considers the interaction between volatility, skewness and kurtosis. This is done by using the Gramâ€“Charlier expansion:

$$\begin{aligned} 1 - F\left(\frac{-\mu_{t+1|t}}{\sigma_{t+1|t}}\right) \approx 1 - \Phi\left(\frac{-\mu_{t+1|t}}{\sigma_{t+1|t}}\right) + \Phi\left(\frac{-\mu_{t+1|t}}{\sigma_{t+1|t}}\right)\left[\frac{\gamma_{3,t+1|t}}{3!}\left(\frac{\mu^2_{t+1|t}}{\sigma^2_{t+1|t}} - 1\right)\right. \\ \left. + \frac{\gamma_{4,t+1|t}}{4!}\left(\frac{\mu^3_{t+1|t}}{\sigma^3_{t+1|t}} + \frac{3\mu_{t+1|t}}{\sigma_{t+1|t}}\right)\right], \end{aligned}$$

where $\Phi(.)$ is the distribution function of a standard normal, and $\gamma_3$ and $\gamma_4$ are, respectively, the skewness and excess kurtosis, with the usual notation for conditioning on $\Omega_t$. This equation can be rewritten as

$$1 - F(-\mu_{t+1|t}x_{t+1}) \approx 1 - \Phi(-\mu_{t+1|t})(\beta_{0t} + \beta_{1t}x_{t+1} + \beta_{2t}x^2_{t+1} + \beta_{3t}x^3_{t+1}),$$

with $\beta_{0t}$=1+$\gamma_{3,t-1|t}/6$, $\beta_{1t}$=-$\gamma_{4,t+1|t}\mu_{t+1|t}/8$, $\beta_{2t}$=-$\gamma_{3,t+1|t}\mu^2_{t+1|t}/6$ e $\beta_{3t}$=$\gamma_{4,t+1|t}\mu^3_{t+1|t}/24$, where for notational convenience, we denote $x_{t+1}$=$1/\sigma_{t+1|t}$.

Whether $\mu_{t+1|t}$ is small, as in the case of short investment horizons, then $\beta_{2t}$ and $\beta_{3t}$ cac be

safely be ignored, resulting in

$$1 - F(-\mu_{t+1|t}x_{t+1}) \approx 1 - \Phi(-\mu_{t+1|t})(\beta_{0t} + \beta_{1t}x_{t+1}).$$

Thus, conditional skewness affects sign predictability through $\beta_{0t}$, and conditional kurtosis affects sign predictability through $\beta_{1t}$. When there is no conditional dynamics in skewness and kurtosis, the above equation is reduced to

$$1 - F(-\mu_{t+1|t}x_{t+1}) \approx 1 - \Phi(-\mu_{t+1|t}x_{t+1})(\beta + \beta x_{t+1}), \tag{29}$$

for some time-invariant quantities $\beta_0$ and $\beta_1$.

We use Equation (29) as our second model for sign prediction, i.e., we generate forecasts of the probability of positive returns as

$$\widehat{Pr}(R_{t+1|t} > 0 = x_{t+1}) \approx 1 - \Phi(-\widehat{\mu}_{t+1|t}\hat{x}_{t+1})(\widehat{\beta}_0 + \widehat{\beta}_1\widehat{x}_{t+1}), \tag{30}$$

where $\hat{x}_{t+1|t} = 1/\hat{\sigma}_{t+1|t}$, and where $\hat{\mu}_{t+1|t}$ and $\hat{\sigma}_{t=t+1|t}$ are as defined earlier. We refer to these as forecasts from the "extended" model. The parameters $\beta_0$ and $\beta_1$ are estimated by regressing $1 - I(R_t > 0)$ on $\Phi(-\hat{\mu}_t\hat{x}_t)$ and $\Phi(-\hat{\mu}_t\hat{x}_t)\hat{x}_t$ for $t = 1, \ldots, \mathrm{k}$. Although we have not explicitly placed any constraints on this model to require $\Phi(-\hat{\mu}_t\hat{x}_t)\left(\hat{\beta}_0 + \hat{\beta}_1\hat{x}_t\right)$ to lie between 0 and 1, we ensure it by applying the logistic function in extended model results to forecasts lie between 0 and 1.

## 5. Forecast Evaluation

We perform out-of-sample comparison of the forecast performance of equations (28), (30), and the logistic function[3] for the sign of return. Both are compared against baseline forecasts [Equation (25)]. This is done for one-, two-, -three and six-month returns. We assess the performance of the forecasting models using Brier scores:

[3]We also add the logistic function as Christoffersen & Diebold (2006) did: $F(x) = \frac{\exp(x)}{1+\exp(x)}$ where $x = \frac{\mu_{t+\hat{1}|t}}{\hat{\sigma}_{t+1|t}}$.

$$\text{Brier(Sq)} = \frac{1}{T-k}\sum_{t=k}^{T} 2\left(\widehat{\Pr}\left(R_{t+1|t} > 0\right) - z_{t+1}\right)^2,$$

$$\text{Brier(Abs)} = \frac{1}{T-k}\sum_{t=k}^{T} \left|\widehat{\Pr}\left(R_{t+1|t} > 0\right) - z_{t+1}\right|,$$

where $z_{t+1} = I(R_{t+1} > 0)$. The latter is the traditional Brier score for evaluating the performance of probability forecasts and is analogous to the usual RMSFE. A score of 0 for Brier(Sq) occurs when perfect forecasts are made: where at each period, correct probability forecasts of 0 or 1 are made. The worst score is 2 and occurs if at each period probability forecasts of 0 or 1 are made but turn out to be wrong each time.

Note that if we follow the usual convention where a correct probability forecast of $I(R_{t+1} > 0)$ is 1 that is greater than 0.5, then correct forecasts will have an individual Brier(Sq) score between 0 and 0.5, whereas incorrect forecasts have individual scores between 0.5 and 2. A few incorrect forecasts can therefore dominate a majority of correct forecasts. For this reason, we only consider a modified version of the Brier score, which we call Brier(Abs). Like Brier(Sq), the best possible score for Brier(Abs) is 0. The worst score is 1. Here correct forecasts have individual scores between 0 and 0.5, whereas incorrect forecasts carry scores between 0.5 and 1.

In the following sections, we introduced the data set and empirical findings

## 6. Data and Results

### *6.1. Data*

The data set consists of monthly closing prices observed for yields of future DI contracts. Based on the observed rates for the available maturities, the data were converted to fixed maturities of 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 36, 48, and 60 months, through interpolations using cubic splines. The database contains the maturities with the highest liquidity for January 2004 through December 2021 (T = 216 observations) and represents the most liquid DI contracts negotiated during the analyzed period. We assess the performance of the model by splitting the sample into two parts: the first one includes 108 observations used to estimate the parameters. The second part is used to analyze the performance out-of-sample of bond portfolios obtained from the model, with 108 observations.

Table 1 displays the descriptive statistics for the Brazilian interest rate curve. For each of the 14 time series, we report average, standard deviation, minimum, maximum, and the last three columns contain sample autocorrelations at displacements of 1, 6, and 12 months. Descriptive statistics presented in Table 1 seem to confirm key stylized facts about yield curves: the sample

Table 1: Summary Statistics

| Maturity | Mean | Std.dev. | Min. | Max. | $\rho_1$ | $\rho_6$ | $\rho_{12}$ | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 10.58 | 4.205 | 1.925 | 19.81 | 0.992 | 0.870 | 0.634 | -0.001 | 2.616 |
| 6 | 10.63 | 4.172 | 1.916 | 19.78 | 0.991 | 0.866 | 0.626 | -0.012 | 2.603 |
| 9 | 10.68 | 4.132 | 1.943 | 19.62 | 0.990 | 0.860 | 0.617 | -0.021 | 2.584 |
| 12 | 10.74 | 4.083 | 2.035 | 19.53 | 0.988 | 0.855 | 0.611 | -0.034 | 2.562 |
| 15 | 10.80 | 4.026 | 2.133 | 19.30 | 0.987 | 0.851 | 0.607 | -0.044 | 2.541 |
| 18 | 10.87 | 3.962 | 2.305 | 19.16 | 0.986 | 0.847 | 0.606 | -0.050 | 2.527 |
| 21 | 10.95 | 3.899 | 2.457 | 19.17 | 0.984 | 0.844 | 0.605 | -0.056 | 2.527 |
| 24 | 11.02 | 3.831 | 2.615 | 19.40 | 0.983 | 0.841 | 0.604 | -0.059 | 2.534 |
| 30 | 11.16 | 3.699 | 2.977 | 19.86 | 0.981 | 0.834 | 0.602 | -0.049 | 2.569 |
| 36 | 11.29 | 3.578 | 3.333 | 20.27 | 0.978 | 0.827 | 0.597 | -0.026 | 2.624 |
| 48 | 11.50 | 3.386 | 3.966 | 21.10 | 0.972 | 0.808 | 0.583 | 0.059 | 2.770 |
| 60 (Levell) | 11.66 | 3.246 | 4.600 | 21.84 | 0.968 | 0.792 | 0.567 | 0.159 | 2.963 |
| Slope | 1.078 | 1.824 | -3.498 | 5.752 | 0.920 | 0.505 | 0.115 | -0.223 | 2.658 |
| Curvature | -0.196 | 1.130 | -2.130 | 3.936 | 0.889 | 0.380 | -0.060 | 0.577 | 3.253 |

NOTE: The table reports summary statistics for Brazil yield curves over the period 2004-2021. We examine monthly data, constructed using the spline method. For each maturity we show mean, standard deviation, skewness, raw kurtosis, minimum, maximum, and three auto-correlations coefficients, $\hat{\rho}_1$, $\hat{\rho}_6$, $\hat{\rho}_{12}$. Also the table reports proxy estimates for level, slope, and curvature of the yield curve. The proxies are defined as follows: for level, the highest maturity bond (60 months); for slope, the difference between the bond of 60 months and the bond of 3 months; and for curvature, two times the bond of 18 months minus the sum of bond of 3 months and bond of 60 months.

average curve is upward sloping and concave, volatility is decreasing with maturity, autocorrelations are very high and decreasing with maturity. Also, there is a high persistence in the yields: the first-order autocorrelation for all maturities is above 0.889 for each maturity.

Figure 1 presents a three-dimensional plot of the data set and illustrates how yield levels and spreads vary substantially throughout the sample. Although the yield series change heavily over time for each of the maturities, a strong common pattern in the 14 series over time is apparent. The sample contains 204 monthly observations with maturities of $\tau = 3$, 6, 9, 12, 15, 18, 21 24, 27, 30, 36, 39, 48, and 60 months.

Figure 1: Brazil Yield Curves


NOTE: Monthly Brazil yield curves from January 2004 through December 2022.

*6.2. Results*

This subsection present absolute Brier score results and the cumulative of Brier score. Table 2 presents out-of-sample comparison of the forecast performance of equations (28), (30), and the logistic function for the sign of return. Both are compared against baseline forecasts [Equation (25)]. This is done for one-, two-, -three and six-month returns. We assess the performance of the forecasting models using Brier scores. The worst score is 1. Here correct forecasts have individual scores between 0 and 0.5, whereas incorrect forecasts carry scores between 0.5 and 1. The Logistic, Non-parametric and Extended are relative to Baseline, so values below 1 outperform the benchmark model. The results suggest that models with information of skewness and kurtosis of returns outperform the benchmark model, mainly in long maturities and shorts horizons of forecasts.

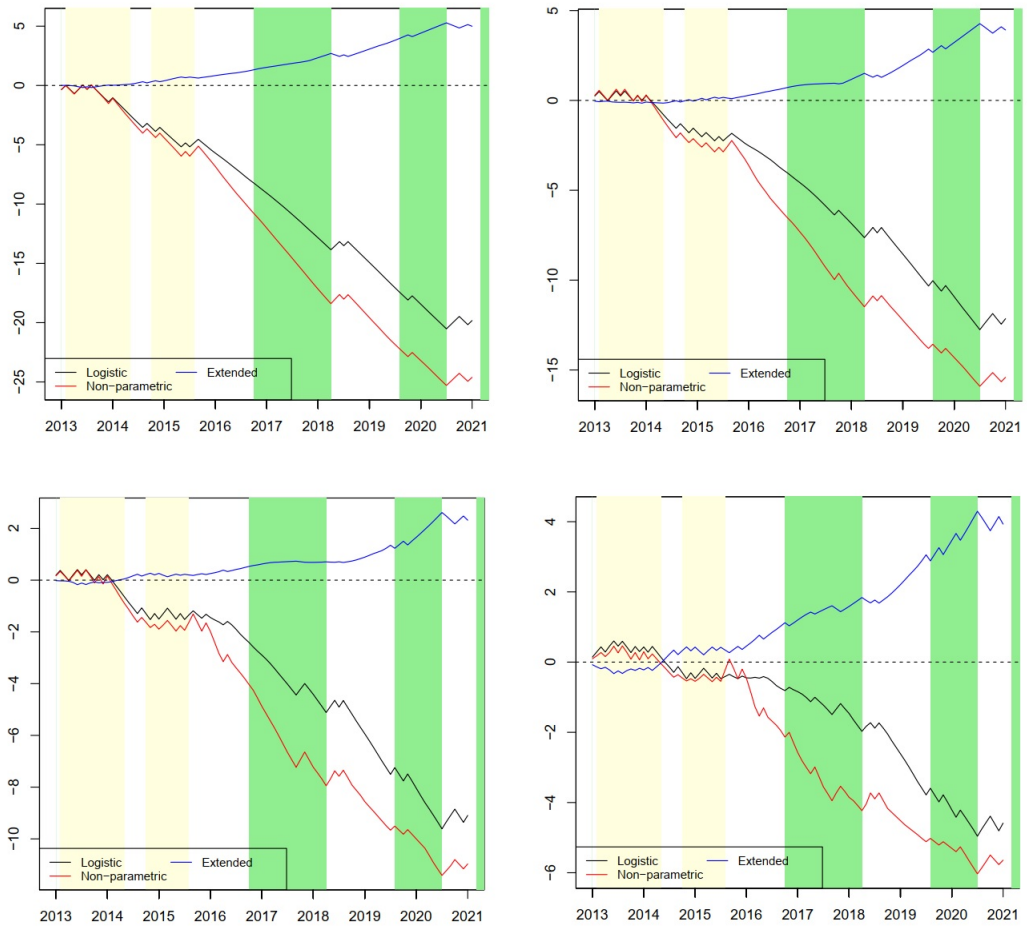
Figures 2 to 4 present the cumulative absolute brier score relative to the benchmark model to uncover the path of all models throughout the out-of-sample period. Hence, the competitor model outperforms the benchmark for values above zero. The top left chart presents results of maturity of 24 months, the top right present results of maturity of 36 months, the bottom left presents results of maturity of 48 months, bottom right presents results of maturity of 60 months. Yellows bar represent periods of rise of the Selic rate, green otherwise, and white is without move. In all months horizons forecasts, the results suggest all models have stable performance against the benchmark model when the Selic rate is rising, however it is not a rule.

Table 2: Absolute Brier Score Results

| Model | Maturity | | | | |
|---|---|---|---|---|---|
| | 18 | 24 | 36 | 48 | 60 |
| Panel (A): 1-month ahead forecasts | | | | | |
| Baseline | 0.1883885 | 0.2809550 | 0.3601736 | 0.3946826 | 0.4440586 |
| Logistic | 2.5670228 | 1.7273311 | 1.3477305 | 1.2375145 | 1.1064536 |
| Non-parametric | 2.9377880 | 1.9030514 | 1.4410967 | 1.2864999 | 1.1309873 |
| Extended | **0.8035142** | **0.8168572** | **0.8878698** | **0.9394830** | **0.9087392** |
| Panel (B): 2-months ahead forecasts | | | | | |
| Baseline | 0.0729246 | 0.1695062 | 0.3069995 | 0.4085730 | 0.4188823 |
| Logistic | 6.3564996 | 2.7673317 | 1.5399332 | 1.1697695 | 1.1436392 |
| Non-parametric | 8.3050595 | 3.5435339 | 1.8713326 | 1.3418327 | 1.2839378 |
| Extended | 2.6936524 | 1.1500709 | **0.9407575** | **0.9056502** | **0.9419552** |
| Panel (C): 3-months ahead forecasts | | | | | |
| Baseline | 0.0206186 | 0.1365165 | 0.2552360 | 0.3107976 | 0.3632122 |
| Logistic | 21.5196225 | 3.2943579 | 1.7891313 | 1.4772629 | 1.2826838 |
| Non-parametric | 30.6842105 | 4.5683625 | 2.3439626 | 1.8823324 | 1.5557215 |
| Extended | 13.5057761 | 1.4180008 | 1.0574732 | 1.0309401 | **0.9978042** |
| Panel (D): 6-months ahead forecasts | | | | | |
| Baseline | 0.0309278 | 0.0618557 | 0.2306023 | 0.2966902 | 0.3453066 |
| Logistic | 12.7649533 | 6.4039673 | 1.7954306 | 1.4407242 | 1.2482505 |
| Non-parametric | 21.8385965 | 10.7052632 | 2.7435294 | 2.0395026 | 1.7275299 |
| Extended | 9.1578898 | 4.8100035 | 1.2474831 | 1.1854429 | 1.1203636 |

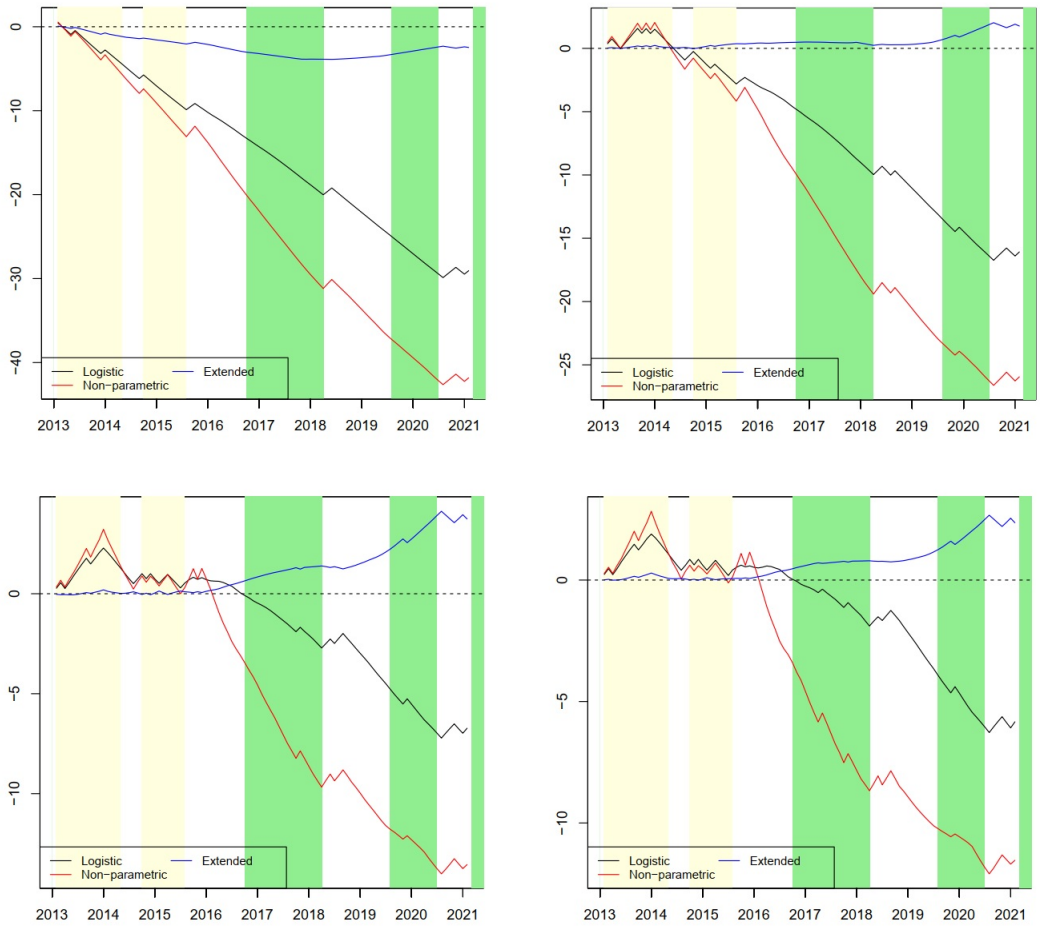
NOTE: We perform out-of-sample comparison of the forecast performance of equations (28), (30), and the logistic function for the sign of return. Both are compared against baseline forecasts [Equation (25)]. This is done for one-, two-, -three and six-month returns. We assess the performance of the forecasting models using Brier scores. The worst score is 1. Here correct forecasts have individual scores between 0 and 0.5, whereas incorrect forecasts carry scores between 0.5 and 1. The Logistic, Non-parametric and Extended are relative to Baseline, so values below 1 outperform the benchmark model.

Figure 2: Cumulative Absolute Brier Scores: 1-month ahead



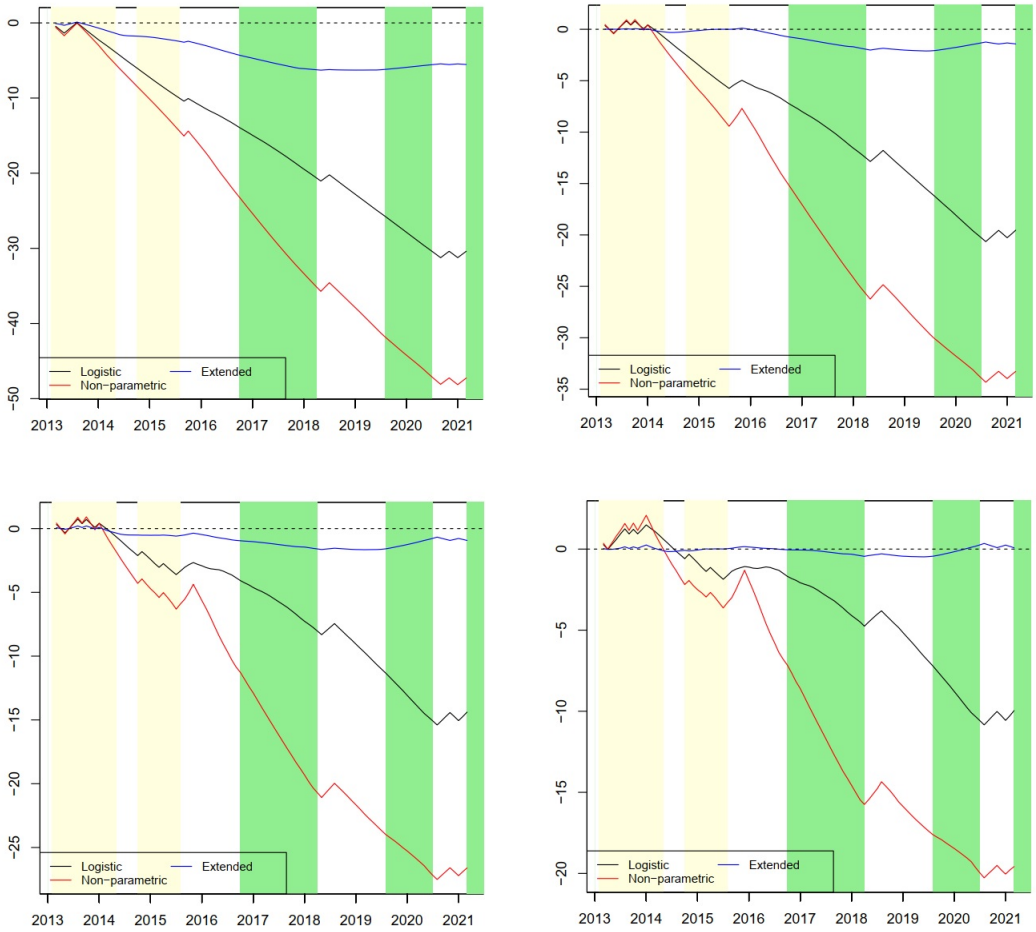

NOTE: We use the cumulative absolute brier score relative to the benchmark model to uncover the path of all models throughout the out-of-sample period. Hence, the competitor model outperforms the benchmark for values above zero. The top left chart presents results of maturity of 24 months, the top right present results of maturity of 36 months, the bottom left presents results of maturity of 48 months, bottom right presents results of maturity of 60 months. Yellows bar represent periods of rise of the Selic rate, green otherwise, and white is without move.

Figure 3: Cumulative Absolute Brier Scores: 2-months ahead




NOTE: We use the cumulative absolute brier score relative to the benchmark model to uncover the path of all models throughout the out-of-sample period. Hence, the competitor model outperforms the benchmark for values above zero. The top left chart presents results of maturity of 24 months, the top right present results of maturity of 36 months, the bottom left presents results of maturity of 48 months, bottom right presents results of maturity of 60 months. Yellows bar represent periods of rise of the Selic rate, green otherwise, and white is without move.

Figure 4: Cumulative Absolute Brier Scores: 3-months ahead




NOTE: We use the cumulative absolute brier score relative to the benchmark model to uncover the path of all models throughout the out-of-sample period. Hence, the competitor model outperforms the benchmark for values above zero. The top left chart presents results of maturity of 24 months, the top right present results of maturity of 36 months, the bottom left presents results of maturity of 48 months, bottom right presents results of maturity of 60 months. Yellows bar represent periods of rise of the Selic rate, green otherwise, and white is without move.

## 7. Concluding remarks

The accuracy of yield curve forecasts, not only of magnitude but specifically of the direction of change in financial assets, emerges as a research source in the prediction field. Our goal is to explore the literature gap on forecasting the yield curves using conditional means and conditional volatility forecasts as inputs to predict the direction of return in the fixed income. This paper assesses the direction-of-change forecasts based on conditional variance from the Dynamic Nelson-

Siegel model. Although the literature focuses on forecasting the level of yield curves, which is a difficult task, we propose forecasts for the direction-of-change of the yield curve returns. The results suggest that models with information of skewness and kurtosis of returns outperform the benchmark model, mainly in long maturities and short horizons of forecasts. Also, In all months horizons forecasts, the results suggest all models have stable performance against the benchmark model when the Selic rate is rising, however it is not a rule.

**References**


Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, **31**(3), 307–327.

Christensen, Bent Jesper, van der Wel, Michel, *et al.* 2010. An asset pricing approach to testing general term structure models including Heath-Jarrow-Morton specifications and affine subclasses. *Department of Economics and Business Economics, Aarhus University*.

Christoffersen, Peter, Diebold, Francis X, Mariano, Roberto S, Tay, Anthony S, & Tse, Yiu Kuen. 2006. Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence.

Christoffersen, Peter F, & Diebold, Francis X. 2006. Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, **52**(8), 1273–1287.

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

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

Fernandes, Marcelo, & Vieira, Fausto. 2019. A dynamic Nelson–Siegel model with forward-looking macroeconomic factors for the yield curve in the US. *Journal of Economic Dynamics and Control*, **106**, 103720.

Greer, Mark. 2003. Directional accuracy tests of long-term interest rate forecasts. *International Journal of Forecasting*, **19**(2), 291–298.

Greer, Mark R. 2005. Combination forecasting for directional accuracy: An application to survey interest rate forecasts. *Journal of Applied Statistics*, **32**(6), 607–615.

Harvey, Andrew C. 1989. *Forecasting, Structural Time Series Models and the Kalman filter*.

Harvey, Andrew C, Ruiz, Esther, & Sentana, Enrique. 1992. Unobserved component time series models with ARCH disturbances.

Koopman, Siem Jan, Mallee, Max IP, & Van der Wel, Michel. 2010. Analyzing the term structure of interest rates using the dynamic Nelson–Siegel model with time-varying parameters. *Journal of Business & Economic Statistics*, **28**(3), 329–343.

Nelson, Charles R, & Siegel, Andrew F. 1987. Parsimonious modeling of yield curves. *q Journal of Business*, **60**(4), 473–489.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rattray, Sandy, & Balasubramanian, Venkatesh. 2003. The new VIX as a market signal: It still works. *Equity Derivatives Strategy: Options & Volatility*.

Svensson, Lars EO. 1994. *Estimating and interpreting forward interest rates: Sweden 1992-1994*. Tech. rept. National Bureau of Economic Research.

Vieira, Fausto, Fernandes, Marcelo, & Chague, Fernando. 2017. Forecasting the Brazilian yield curve using forward-looking variables. *International Journal of Forecasting*, **33**(1), 121–131.

Yu, Wei-Choun, & Zivot, Eric. 2011. Forecasting the term structures of Treasury and corporate yields using dynamic Nelson-Siegel models. *International Journal of Forecasting*, **27**(2), 579–591.