A Quantitative Comparison of Yield Curve Models in the MINT Economies

Kelly Ayliffe[1] and Tomáš Rubín[2]

[1] kellyayliffe@gmail.com
[2] tomas.rubin@gmail.com

Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne, Switzerland

This report originates from the semester project written by Kelly Ayliffe and supervised by Tomáš Rubín at Ecole Polytechnique Fédérale de Lausanne.

Abstract

A yield curve is a line plotting bond yields (i.e. interest rates) as a function of their maturity date (their "expiration date"). When on a national scale, the yield curve represents the underlying interest rate structure of a country's economy. It is fundamental in many areas of finance as it helps understand the expected evolution of an asset's value, and therefore serves as a basis for pricing various assets and their derivatives, valuations of capital, determination of risk and more. Furthermore, yield curves are widely considered as a general indicator of a country's overall economic health.

During this study, we conducted a series of quantitative comparisons of various yield curve models, specifically on the MINT (Mexico, Indonesia, Nigeria, Turkey) economies. We implemented the one-step and two-step Dynamic Nelson-Siegel methods as well as traditional time-series models such as vector autoregressions with single and multiple lags and autoregression, using the random walk as a benchmark for comparison.

The results confirm what was perhaps to be expected: forecasting yield curves is no easy tasks, and in some cases the random walk cannot be outperformed by the selected methods. There are, however, other cases where the presented methods thrive. In these situations, the Dynamic Nelson-Siegel methods in particular, both one-step and two-step, proved to be considerably superior to the random walk.

Contents

1 Introduction **3**
- 1.1 An Economic Approach . . . 3
- 1.2 A Mathematical Approach . . . 7

2 The Dynamic Nelson-Siegel Model **10**
- 2.1 The Static Nelson-Siegel Model . . . 10
- 2.2 The Dynamic Approach . . . 15

3 Contending models **18**
- 3.1 Random Walk . . . 18
- 3.2 Autoregression . . . 19
- 3.3 Vector Autoregression . . . 19
- 3.4 Comparing Forecasting Accuracy . . . 20

4 Quantitative Data Analysis **23**
- 4.1 The Datasets . . . 23
- 4.2 Implementation . . . 29
- 4.3 The Results . . . 31
- 4.4 Results Analysis . . . 41
- 4.5 Conclusion . . . 50

A Estimation and Prediction in State-Space Models **52**
- A.1 General Linear State-Space Models . . . 52
- A.3 The Kalman Filter . . . 53

Bibliography **60**

Chapter 1

Introduction

The aim of this chapter is to clarify the meaning and the importance of yield curves, and therefore why it is so useful to forecast them. This chapter will then explore why the MINT (Mexico, Indonesia, Nigeria, Turkey) economies in particular were chosen to conduct this study.

Furthermore, it will include basic interest rate theory as well as a mathematical foundation of state-space models, which will be used in subsequently introduced concepts.

1.1 An Economic Approach

The inversion of the US yield curve in March 2019 lead to many headlines.

> The US bond yield curve has inverted. — CNBC (Srivastava, 2019)
>
> The U.S. Yield Curve Just Inverted. That's Huge. — Bloomberg (Chappatta, 2018)
>
> YIELD CURVE INVERTS: Recession indicator flashes red for first time since 2005 — Yahoo Finance (Cheung, 2019)

Those of us who are not financial analysts may not understand the hype around this event. Why is something like this so important? What does it influence? Even, what is a yield curve?

The following subsection aims to explain, in a non-mathematical way, the basics of yield curves and their importance.

1.1.1 Yield Curve Questions

What is a yield curve?

There is another question that needs answering first: what is a bond?

Companies or governments issue bonds to the market in order to fund their businesses or activities. By buying a bond, which is actually a contract, the bondholder lends the amount he pays for the contract to the entity, which promises to pay back the face value at maturity (all this is specified in the terms of the contract). In addition, some bonds can also pay a fraction of their value at pre-specified dates between the time of purchase and the maturity. These partial reimbursements are known as coupons. A bond with no coupons is called a zero-coupon bond. In this case, the price paid for the bond directly reflects the interest rate (i.e. the percentage of the full price that equal the profit) to be paid to the lender.

A yield curve is a line plotting bond yields (i.e. interest rates) as a function of their maturity dates. An example can be found in Figure 1.1. Usually, the bonds that are considered are zero-coupon government bonds. The corresponding yield curve can be viewed as a summary of the country's interest rates and an indicator of overall economical health. It is used as a benchmark for pricing many financial assets and their derivatives as well as valuations and risk determination.

Traditionally, a yield curve can take one of three shapes:

1. Normal: an upwards sloping curve, i.e. the bonds with long-term maturities have higher interest rates than the ones with short-term maturities.
2. Flat: all bonds have similar interest rates.
3. Inverted: a downward sloping curve, the bonds with short-term maturities have higher interest rates than the ones with long-term maturities.

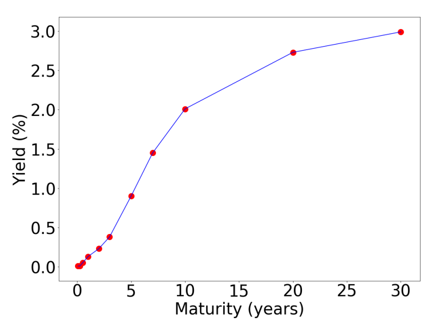


Figure 1.1: Plot of the linear interpolation of the US Treasury yields on 1st November, 2011. The measured yields are marked in red. The x-axis represents the maturities (in years) and the y-axis represents the zero-coupon bond yields (in percentage points).

The answers to the questions that follow were inspired by Goldstein and Romer (2019) and Garcia and Vanek Smith (2019).

How is it determined?

Take the country's public lending and borrowing scene — the bond market. It operates, like any other market, according to classical supply and demand dynamics: if there are many borrowers and not many lenders, there will be high interest rates. Conversely, less borrowers and more lenders lowers interest rates. One can therefore imagine that the market needs to find the "magical" interest rate at which the amounts people are willing to lend are in balance with the amounts people are looking to borrow. This is called the neutral/natural interest rate.

The country's central bank, which issues the yield, is basically tied to this neutral interest rate. A move away from this neutral level would have severe consequences.

If the central bank decided to set their official interest rate below the neutral rate, borrowers would flock to it, borrow much more than needed and would subsequently overspend with the probable outcome of rising inflation.

In the opposite scenario of interest rates being set significantly above the neutral rate, nobody could afford to borrow and the risk would be an economic slowdown.

Why is everybody so preoccupied with inverted yield curves?

It is a common assumption — at least in US treasury bonds — that an inverted yield curve leads to a recession.

This is again simply due to classic supply and demand dynamics. As seen previously, bond issuers increase yields if there are not enough lenders in order to make the purchase of such a bond more attractive. An increase in short term yields is therefore seen as being catalysed by a reluctance to invest in the short term economy. Conversely, one can establish similar reasoning regarding the link between a decrease in long term yields and a preference to invest in the long-term. Put in a nutshell, an inverted yield curve represents general misgivings about the short term economy, as opposed to the long term economy.

However, many economists remain sceptical. Even though yield curve inversion is synonymous with an impending recession, there are too few observations to develop a conclusive theory around the consequences of such an inversion. This subject will be of particular interest in the coming months as the yield curve inverted in March 2019 and became normal again in November 2019. As a result, the next few months may disprove rather than confirm current theories around these relationships.

Can a yield be negative?

Yes, it can. Even the data we will be working with contains negative yields.

At first glance, this seems completely counterintuitive. Why would somebody pay for a bond which at maturity will actually pay back *less* than the original price? Why not just "keep the money under the mattress" instead?

It is a fact that in an ageing society, people save more, simply because they need to accumulate a larger amount of capital to fund a longer retirement. This implies that the supply side of the bond market is increasing — there is more money to lend out. On the demand side, companies are borrowing less: physical goods cost less, and many of today's giants are heavily technology-based, meaning that, in comparison to more traditional industrial companies, their physical asset base is significantly reduced. This combination is driving interest rates down, to the extent that they are becoming negative.

Supply and demand is not choking out these negative rates because even though it can be acceptable to keep a few hundred dollars in cash, it is not safe to hold a lot more. Physically storing cash costs money — for example, one might need access to a vault or to hire a security team. Interest rates can therefore become slightly negative, just because people are still prepared to pay a small amount for keeping their money safe instead of "buying the vault" themselves. Another solution could be to store assets in a bank, which the bank basically considers as a loan. Doing this results in credit risk, the risk associated with a borrower failing to repay. If the credit risk is considered to be lower for a trustworthy government (famous examples include Switzerland or Germany), then investing with negative yield in this government's bonds could be preferred.

1.1.2 The MINT Economies

MINT is an acronym for Mexico, Indonesia, Nigeria and Turkey. Popularised in 2013 by Jim O'Neill, who also coined the BRICS (Brazil, Russia, India, China, South Africa) acronym, they are considered to be the next compelling emerging markets to invest in. Why? O'Neill claims that

> "Mexico, Indonesia, Nigeria and Turkey all have very favorable demographics for at least the next 20 years, and their economic prospects are interesting." (O'Neill, 2013)

Moreover, they are a subset of the "Next Eleven", 11 countries deemed by Goldman Sachs in 2005 to have the highest potential of becoming the world's most important economies of the 21st century.

Furthermore, as pointed out by Mexican Foreign Minister Jose Antonio Meade Kuribrena on BBC News (2014), their advantageous geographical positions represent some of the major power positions for global trade in years to come.

For example, Mexico's situation between Latin America and the US is a perfect liaison between two continents, while Turkey has one foot in Europe and one foot in Asia. Indonesia is an economy close to China but finds itself in the Middle of South-East Asia. Nigeria is perhaps not an obvious contender, but as Africa's most populous country, one can imagine how it might disproportionately benefit from a strongly growing pan-African economy and the implementation of multiple trade programmes on the continent.

1.2 A Mathematical Approach

1.2.1 Economic Fundamentals

> A dollar today is worth more than a dollar tomorrow.

This is the financial principle of the time-value of money and the one that underpins the constructs developed in this paper. Simply put, it indicates that if a dollar is kept for a day, the same dollar loses part of its earning potential over this period.

Interest Rates

This section serves to define terms and quantities of basic interest-rate theory upon which we will base further chapters.

Roughly speaking, interest is money charged to a borrower for the use of a lender's money over a period of time. If an invested amount P_0 accumulates an interest i per year, and we denote $P_{\mathrm{FV}}(n)$ as the future value of the investment at time n years, then

$$P_{\mathrm{FV}}(n) = P_0(1+i)^n. \tag{1.1}$$

In practice, the term "annual nominal interest rate", denoted here by $i^{(m)}$, often arises. It represents the "as stated" rate without adjustment for the full effect of compounding m-thly. This term is used when the compounding frequency (which is m below) is not the specified one. For t in years, we then have

$$P_{\mathrm{FV}}(t) = P_0\left(1+\frac{i^{(m)}}{m}\right)^{mt}. \tag{1.2}$$

If i is the continuous compounding rate (i.e. $m \to \infty$), Equation (1.2) can be extended to

$$P_{\mathrm{FV}}(t) = P_0 e^{ti}. \tag{1.3}$$

Yield Curves

The definitions and constructs here are taken from Diebold and Rudebusch (2013), Chapter 1, Section 2.

Definition 1.3. *The discount curve $P(\tau)$ is defined as the price of a \$1 face value zero-coupon bond with maturity τ.*

Note $P(\tau)$ denotes the *present* value of \$1 dollar at time τ, i.e. the amount that needs to be invested at time 0 (P_0 in the previous section) in order to have $P_{\mathrm{FV}}(\tau) = 1$.

Definition 1.4. *The forward rate curve is defined as*

$$f(\tau) = \frac{-P'(\tau)}{P(\tau)}, \tag{1.4}$$

which we can interpret as the instantaneous profit (proportionally to the price).

Definition 1.5. *The yield curve, plotting the yield of a bond as a function of its maturity τ, is defined as*

$$y(\tau) = \frac{1}{\tau} \int_0^\tau f(u) \mathrm{d}u. \tag{1.5}$$

Note The yield can therefore be seen as "the average of the instantaneous profits".

These three curves are ultimately inter-related. The knowledge of one provides the basis to all of them, and therefore, without loss of generality, one can choose to work with any one of them. Definitions 1.4 and 1.5 furthermore imply

$$P(\tau) = e^{-\tau y(\tau)}, \tag{1.6}$$

i.e. that $y(\tau)$ is the continuously compounded yield for a \$1 face value zero-coupon bond with present value $P(\tau)$.

In particular, $P(\tau)$ satisfies

$$\lim_{\tau \to \infty} P(\tau) = 0, \quad P(0) = 1, \tag{1.7}$$

which fits economic theory: the price of a bond with an infinite maturity is 0 and the price of a \$1 bond of maturity 0 is equal to it's face value of \$1.

Yield curve construction

In practice, the yields are unobservable. The actual bond prices are however available for certain maturities. If a discount (or forward rate) curve is determined, the yield curve easily follows (with Equation (1.6)).

The discount curve is determined by interpolation or approximation of the available prices. Two families can be distinguished: the spline and the parametric methods. The latter is also known as the parsimonious family and provide a good solution when high precision is not paramount. If this is necessary, the former may be a better choice.

There are many publications about both. Acclaimed methods include

- *for the spline family:* McCulloch (1971), who uses polynomial splines (whose long-term oscillation is not a desirable property) and Vasicek and Fong (1982), who use exponential splines instead. See Diebold and Rudebusch (2013), Section 1.2 for more details.
- *for the parametric family:* Nelson and Siegel (1987) and Svensson (1994) (an extension of Nelson-Siegel).

We will not expand on these but will discuss the Nelson-Siegel approach in depth later.

Chapter 2

The Dynamic Nelson-Siegel Model

2.1 The Static Nelson-Siegel Model

2.1.1 Definition

The static Nelson-Siegel model smoothly fits the yield curve at a *fixed* time $t \in \{1, \ldots, T\}$ as

$$y(\tau) = \beta_1 + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right), \tag{2.1}$$

where $(\beta_1, \beta_2, \beta_3, \lambda)$ are parameters to be determined.

Even though such a specific model may seem arbitrary, it actually presents a number of elegant properties:

It fits economic theory, namely two important constraints:

1. $P(0) = 1$: the price of a bond with maturity 0 must be $1.
2. $\lim_{\tau\to\infty} P(\tau) = 0$: as the maturity becomes infinitely long, the price must be zero.

The yield converges as $\tau \to \infty$, and $\lim_{\tau\to\infty} y(\tau) = \beta_1$. This will in fact be defined further on as the level of the yield curve.

It only has four parameters and therefore is sparse, making it easier to understand, less prone to over-fitting and smoother (which is desirable as yield tends to be a very smooth function of maturity). Furthermore, contrary to other models, the number of parameters does not increase when the number of maturities does.

It is (a priori) flexible as it can be increasing, decreasing, U-shaped and more depending on the parameter choice. However, it cannot have more than one optimum. Even though this is traditionally not thought of as an important restriction because regular markets such as the US do not have multiple optima, we will see that the markets we are considering do. Therefore, the Nelson-Siegel model presents a noticeable limit in our setting.

2.1.2 Interpretation

A maturity-dependent interpretation

Each β_i influences different parts of the curve:

β_2 scales $(1-e^{-\lambda\tau})(\lambda\tau)^{-1}$, a function that will start at one but rapidly decrease towards 0. As it will mostly affect only short-term yields, β_2 is known as "the short-term factor".

β_3 scales $(1-e^{-\lambda\tau})(\lambda\tau)^{-1}-e^{-\lambda\tau}$, which increases from zero (not affecting short maturities), peaks and then rapidly decreases again. It is therefore known as "the medium-term factor".

β_1 is a constant and the only term that does not decay to zero when the maturity increases, which is why it is "the long-term factor".

A plot of each function can be found in Figure 2.1.

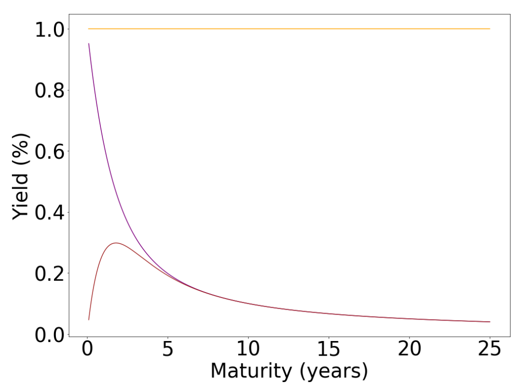


Figure 2.1: Plot of each factor loading as a function of the maturity (with $\lambda = 1$). The purple curve represents short-term factor loading, the brown curve the medium-term factor loading and the orange curve the long-term factor loading.

A shape interpretation

Diebold and Li (2006) are the first to note the influences $(\beta_1, \beta_2, \beta_3)$ have on the shape of the yield curve: β_1 is the level, β_2 is the slope, and β_3 is the curvature.

Figures 2.2 – 2.4 illustrate the influence each of the three factors has on the curve.

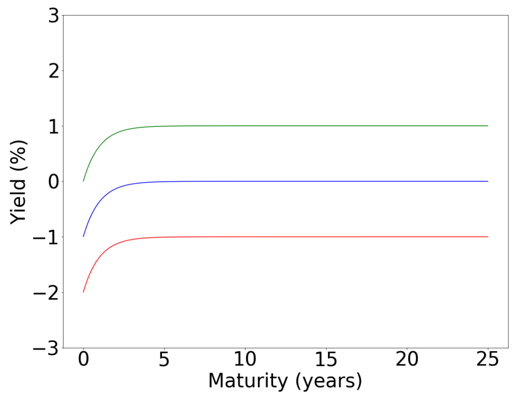


Figure 2.2: A change in level. We plot three NS curves at a given time with $(\lambda, \beta_2, \beta_3) = (1, -1, 1)$ as a function of the maturity. The blue curve is the NS parametrisation with $\beta_1 = 0$, the green curve with $\beta_1 = 1$, the red curve with $\beta_1 = -1$.

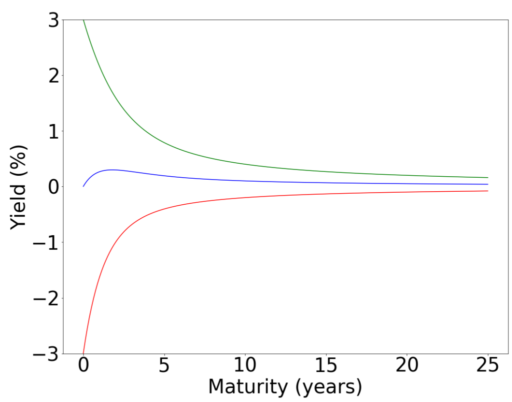


Figure 2.3: A change in slope. We plot three NS curves at a given time with $(\lambda, \beta_1, \beta_3) = (1, 0, 1)$ as a function of the maturity. The blue curve is the NS parametrisation with $\beta_2 = 0$, the green curve with $\beta_2 = 3$, the red curve with $\beta_1 = -3$.

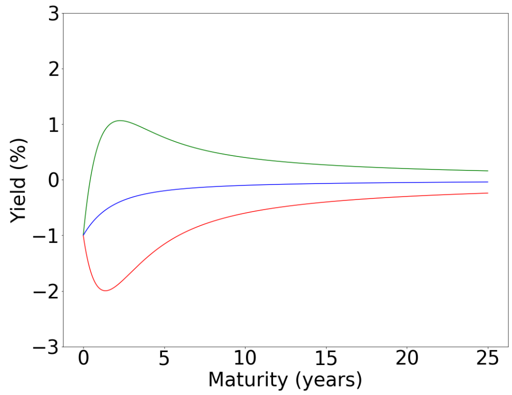


Figure 2.4: A change in curvature. We plot three NS curves at a given time with $(\lambda, \beta_1, \beta_2) = (1, 0, -1)$ as a function of the maturity. The blue curve is the NS parametrisation with $\beta_3 = 0$, the green curve with $\beta_3 = 5$, the red curve with $\beta_3 = -5$.

Note Even though we will refer to the second factor as the slope, we can see in Figure 2.3 that it actually represents the "negative slope" in early maturities, i.e. a positive β_2 equates to a decreasing curve, whereas a negative β_2 leads to an increasing curve.

With these interpretations in mind, the notations are changed to

$$y(\tau) = l + s\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\right) + c\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right). \tag{2.2}$$

Interpretation of λ

Figure 2.5 illustrates the effect the parameter λ has on the NS curve parametrisation.

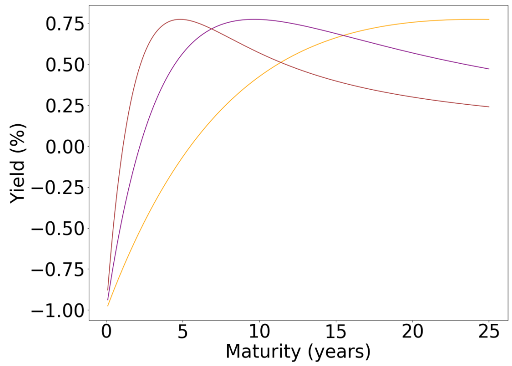


Figure 2.5: We plot three NS curves at a given time with $(\beta_1, \beta_2, \beta_3) = (0, -1, 4)$ as a function of the maturity. The orange curve is the NS parametrisation with $\lambda = 0.1$, the purple curve with $\lambda = 5.24$, the brown curve with $\lambda = 0.5$.

We can see from the graphs in Figure 2.5 that λ is the parameter determining the location of maximal curvature. We will discuss how this is used in the estimation of the model in Subsection 2.1.3.

2.1.3 Estimation

Given a vector of observed yields $y = (y(\tau_1), \ldots, y(\tau_p))'$, the parameter estimation becomes simple when λ is known. It is sufficient to conduct a least-squares regression of y on $\left(1, (1 - e^{-\lambda\tau_i})(\lambda\tau_i)^{-1}, (1 - e^{-\lambda\tau_i})(\lambda\tau_i)^{-1} - e^{-\lambda\tau_i}\right)_{i=1}^{p}$.

Diebold and Li (2006) determines λ by observing that it determines the location of maximal curvature.

Let $\tau_{\max} \in \{\tau_1, \ldots, \tau_p\}$ be a reasonable maturity for the maximum to be attained (can be observed numerically in the coordinates of y). This directly leads to finding a viable λ for our model by finding

$$\hat{\lambda} = \underset{\lambda \in \mathbb{R}}{\arg\max} \left(\frac{1 - e^{-\lambda\tau_{\max}}}{\lambda\tau_{\max}} - e^{-\lambda\tau_{\max}} \right), \tag{2.3}$$

i.e. setting the term multiplied by c_t to be maximal at $\tau_{\max}$ (l_t does not affect the maximum and we assume s_t does not significantly affect the Nelson-Siegel yield curve when the maturity is significantly non-zero).

A common practice (see for example Diebold and Li (2006) or Diebold and Rudebusch (2013)) is to consider that the theoretical maximal curvature is attained at $\tau = 30$ months, and therefore sets $\hat{\lambda} = 0.0609$. When working in cases that do not necessarily follow the traditional structure, one could also estimate an appropriate λ with a non-linear fitting method. A possible procedure, following a least-squares approach, is described in the following:

For each time $t = 1 \dots T$, choose $\hat{\lambda}_t$ satisfying the minimum of

$$\text{OLS}_t^{\text{NS}}(\lambda) = \min_{(l,s,c)\in\mathbb{R}^3} \sum_{i=1}^{p} \|y_t^*(\tau_i) - y_t(\tau_i, l, s, c)\|_2^2,$$

where $y_t^*(\tau_i)$ denotes the true yield with maturity τ_i at time t and $y_t(\tau_i, l, s, c)$ denotes the Nelson-Siegel parametrisation in Equation (2.2).

For robustness and because λ is assumed fixed in time, we take the mean of $\hat{\lambda}_t, t = 1, \dots, T$. The final the final estimation of λ is therefore

$$\hat{\lambda} = \frac{1}{T} \sum_{t=1}^{T} \left(\operatorname*{argmin}_{\lambda\in\mathbb{R}} \text{OLS}_t^{\text{NS}}(\lambda) \right).$$

We will hereafter assume that λ is known and fixed.

2.2 The Dynamic Approach

2.2.1 Definition

Even though the model above is proven to fairly accurately describe the yield curve at one given moment, it lacks in one crucial aspect: time dependency.

The Dynamic Nelson-Siegel (DNS) model therefore considers that for a fixed t, the curve would be fitted with the static Nelson-Siegel model, but that (l, s, c) would vary in time, i.e.

$$y_t(\tau) = l_t + s_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + c_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right),\ t = 1, \dots, T. \tag{2.4}$$

Note The parameter λ is considered to be an inherent property of the model and therefore does not vary with t.

Note The nature of (l_t, s_t, c_t) changes here — we now consider them as time-series, and not parameters.

For the theoretical model to explain the empirical data $y_t(\tau_i),\ i = 1 \dots p$, we introduce a stochastic error ϵ_t.

Equation (2.4) becomes

$$y_t = \Lambda f_t + \epsilon_t, \tag{2.5}$$

where

$$y_t = \begin{pmatrix} y_t(\tau_1) \\ \vdots \\ y_t(\tau_p) \end{pmatrix}, \quad f_t = \begin{pmatrix} l_t \\ s_t \\ c_t \end{pmatrix}, \quad \epsilon_t = \begin{pmatrix} \epsilon_t(\tau_p) \\ \vdots \\ \epsilon_t(\tau_p) \end{pmatrix}$$

and

$$\Lambda = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_p}}{\lambda\tau_p} & \frac{1-e^{-\lambda\tau_p}}{\lambda\tau_p} - e^{-\lambda\tau_p} \end{pmatrix}.$$

2.2.2 Estimation

The theoretical model above is sufficient to explain a cross-section of the data but the values f_t and a model for its evolution in time are still unknown. In this subsection, we will explore two different methods to estimate f_t and understand how it evolves.

Two-Step DNS

We could consider the data as a p-dimensional time-series (see the vector autoregression method described in Chapter 3). However, the strength of the Nelson-Siegel model is that it reduces into only three factors, (l_t, s_t, c_t), irrespectively of the number of maturities. Two-step DNS is a simple and numerically stable method to fit this three-variable model to the data at hand.

Step One: a static fitting

Step One fits the static Nelson-Siegel model on each time period $t = 1 \dots T$. As stated previously, this can be done using a simple least squares regression that outputs $\{(\hat{l}_t, \hat{s}_t, \hat{c}_t)\}_{t=1}^T$.

Step Two: introducing dynamics

Consider now a three-dimensional time series $\{l_t, s_t, c_t\}$, of which we have a sample from Step One. A suitable model can then be found using methods such as three dimensional (Gaussian) vector autoregression (VAR),

$$f_t = A + \sum_{i=1}^{q} B_i f_{t-i} + \epsilon_t, \tag{2.6}$$

where $q \in \mathbb{Z}_+$ is the lag of the process, $A \in \mathbb{R}^3$ is the offset, $B_i \in \mathbb{R}^{3\times3}$ are the matrices of VAR coefficients and $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \Sigma),\ t = 1, \dots, T$. The estimation of the parameters is realised by the Yule-Walker equations.

The forecast of the set of yields at time $T + h$, $\hat{y}_{T+h}$ is then simply

$$\hat{y}_{T+h} = \Lambda \hat{f}_{T+h}, \tag{2.7}$$

where $\hat{f}_{T+h}$ is the estimation of f_{T+h} with the VAR model that was estimated using Equation (2.6).

One-Step DNS

If the Dynamic Nelson-Siegel model is assumed to follow a state-space representation

$$\begin{aligned} y_t &= \Lambda f_t + \epsilon_t, \\ f_t &= T f_{t-1} + c + \eta_t, \\ f_1 &\overset{i.i.d.}{\sim} \mathcal{N}(a_1, P_1), \end{aligned} \tag{2.8}$$

where $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, H)$ and $\eta_t \overset{i.i.d.}{\sim} \mathcal{N}(0, Q)$ then we can also apply the Kalman Filter (or Smoother) in the frame described in Subsection A.1 to estimate the unobservable factors $f_t, t = 1, \cdots, T$ and the Expectation-Maximisation algorithm to estimate T, c, H, Q, a_1 and P_1. The framework of the state-space models including Kalman Filter and Expectation-Maximisation algorithms is explained in the appendix to this report.

The forecast of the set of yields at time $T + h$, $\hat{y}_{T+h}$, is then simply

$$\hat{y}_{T+h} = \Lambda \hat{f}_{T+h}, \tag{2.9}$$

where $\hat{f}_{T+h}$ is the estimation of f_{T+h} using the VAR component of the State-Space model in Equation (2.8) with the parameters estimated using the EM algorithm.

Chapter 3

Contending models

This chapter details the models that will be compared in the quantitative study conducted in this paper, which will follow similar specifications to Brechtken (2008) and Caldeira et al. (2019).

3.1 Random Walk

The random walk is the simplest of the models and will serve as a benchmark for evaluating how well other models perform. For τ the maturity corresponding to the yield $y(\tau)$, the random walk supposes that

$$y_{t+1}(\tau) = y_t(\tau) + \epsilon_t(\tau), \quad t = 1, \dots, T, \tag{3.1}$$

where $\epsilon_t(\tau) \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2(\tau))$.

If $\{y_t(\tau)\}_{t=1}^{T}$ is the available set of yields at maturity τ, the forecast at time $T + h$ is simply

$$\hat{y}_{T+h}(\tau) = y_T(\tau). \tag{3.2}$$

The value we predict at time $T + h$ is just the latest known value, i.e. the value at time T.

The random walk model is a common benchmark in literature for comparing yield curve models (see for example Caldeira et al. (2019), Brechtken (2008) or Mönch (2006)) because it is considered non-trivial to consistently outperform.

3.2 Autoregression

Autoregression (AR) is a traditional time-series modelling method. It considers that the available data $\{y_t(\tau_i)\}_{t=1}^T, i = 1 \dots p$ are p separate time-series. A new model is estimated for each maturity and the yields are forecast separately, and therefore this method completely overlooks any dynamics between yields with different maturities.

Recall that a (Gaussian) AR process with lag order $q \in \mathbb{Z}_+$, usually denoted as AR(q), is defined as

$$y_t(\tau_i) = \alpha + \sum_{j=1}^{q} \beta_j y_{t-j}(\tau_i) + \epsilon_t(\tau_i) \quad t = 1, \dots, T, \tag{3.3}$$

where $\epsilon_t(\tau_i) \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma^2(\tau_i)\right), \alpha \in \mathbb{R}, \beta_j \in \mathbb{R},\ j = 1, \dots, q$.

If the parameters α, β_j are unknown and we can produce suitable estimators[1], say $\hat{\alpha}, \hat{\beta}_j$, for them, the one-step ahead forecast of the yield at maturity τ_i, $\hat{y}_{T+1}(\tau_i)$, is given by

$$\hat{y}_{T+1}(\tau_i) = \hat{\alpha} + \sum_{j=0}^{q-1} \hat{\beta}_j y_{T-j}(\tau_i). \tag{3.4}$$

In the context of yield curve modelling, the lag q is usually chosen to be unity. This means that the forecast for time $T + h$ is simply

$$\hat{y}_{T+h}(\tau_i) = (1 + \hat{\beta} + \hat{\beta}^2 + \dots + \hat{\beta}^{h-1})\hat{\alpha} + \hat{\beta}^h y_T(\tau_i).$$

The estimation of the parameters is realised by the Yule-Walker equations.

3.3 Vector Autoregression

Contrary to the methods previously cited in this chapter, vector autoregression (VAR) takes into account dependencies between maturities as well as dependencies through time to model the available data $\{y_t\}_{t=1}^T$ where $y_t \in \mathbb{R}^p$. A (Gaussian) vector autoregression process with lag order $q \in \mathbb{Z}_+$, usually denoted as VAR(q), is defined as

$$y_t = A + \sum_{j=1}^{q} B_j y_{t-j} + \epsilon_t, \quad t = 1, \dots, T, \tag{3.5}$$

[1] An easy way in this case to produce suitable estimators for $\hat{\alpha}, \hat{\beta}_j$ is to use maximum likelihood estimation, which is equivalent to an ordinary least-squares regression here, on the available data. Python's method `AR.fit()` fits the unconditional maximum likelihood.

where $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \Sigma)$ for some covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$, $A \in \mathbb{R}^p, B_j \in \mathbb{R}^{p \times p}$, $j = 1, \ldots, q$.

If the parameters A, B_j are unknown and we can produce suitable estimators[2], say $\hat{A}, \hat{B}_j$, for them, the one-step ahead forecast, $\hat{y}_{T+1}$, is given by

$$\hat{y}_{T+1} = \hat{A} + \sum_{j=0}^{q-1} \hat{B}_j y_{T-j}. \tag{3.6}$$

Like in autoregression, yield curve models with vector autoregression only consider a single lag. In this context, the forecast at time $T + h$ is

$$\hat{y}_{T+h} = (1 + \hat{B} + \hat{B}^2 + \cdots + \hat{B}^{h-1})\hat{A} + \hat{B}^h y_T.$$

Once again, the estimation of the parameters is realised by the Yule-Walker equations.

3.4 Comparing Forecasting Accuracy

The purpose of this study is to determine how well the various models described in the previous sections forecast MINT yield curves. This section describes various methods to achieve this. Recall that this paper will always use the random walk as a benchmark.

Root Mean Square Error Ratios

Definition 3.5. *Let $\{\hat{y}_t(\tau)\}_{t=1}^T$ be the forecasts of $\{y_t(\tau)\}_{t=1}^T$, the actual yields at maturity τ. The Root Mean Square Error (RMSE) of these forecasts is defined as*

$$\text{RMSE}(\tau) = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(y_t(\tau) - \hat{y}_t(\tau)\right)^2}. \tag{3.7}$$

The RMSE assesses the performance of a model without regard to a benchmark. When comparison is needed, the ratio of RMSEs corresponding to two different forecasting schemes can be used to measure relative performance. This is, for example, used in Brechtken (2008). However, it is hard to find significance levels of over-/underperformance.

In order to justify significant over-/underperformance, the popular Diebold Mariano, and Harvey, Leybourne and Newbold, tests will be preferred.

[2] for example an OLS fitting, like in Python's method `VAR.fit()`.

The Diebold Mariano Test

The Diebold Mariano Test (Diebold and Mariano, 1995) is a common method to compare accuracy of forecasts.

Let $\{y_t\}_{t=1}^T$ be the actual values that we want to forecast and $\{\hat{y}_{1t}\}_{t=1}^T, \{\hat{y}_{2t}\}_{t=1}^T$ be two forecasts of these values we want to compare. The forecast errors are therefore

$$e_{it} = \hat{y}_{it} - y_t, \quad i = 1, 2. \tag{3.8}$$

The loss differential between two forecasts is defined as

$$d_t = g(e_{1t}) - g(e_{2t}), \tag{3.9}$$

where $g(\cdot)$ is the chosen loss function (in our case, the squared loss).

Definition 3.6. *Two forecasts are said to have equal accuracy if*

$$\mathbb{E}\left(d_t\right) \ = 0.$$

Note We suppose that $\mathbb{E}\left(d_t\right)$ does not depend on t.

Comparing accuracy of two forecasts can therefore be done by testing

$$H_0 : \mathbb{E}\left(d_t\right) = 0$$

and declaring that $\{\hat{y}_{1t}\}_{t=1}^T$ are more/less accurate than $\{\hat{y}_{2t}\}_{t=1}^T$ if H_0 is rejected because $\mathbb{E}\left(d_t\right)$ is actually negative/positive respectively.

Diebold and Mariano (1995) state that

$$\sqrt{T}(\bar{d} - \mu) \longrightarrow \mathcal{N}\left(0, 2\pi f_d(0)\right), \tag{3.10}$$

where

$\bar{d} = \frac{1}{T}\sum_{t=1}^T d_t$ is the sample mean of the loss differential,

$\mu = \mathbb{E}\left(d_t\right)$ is the actual mean of d_t,

$f_d(0) = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} \gamma_d(k)$ is the spectral density of d at frequency 0,

$\gamma_d(k) = \mathbb{E}\left[\left(d_t - \mu\right)\left(d_{t-k} - \mu\right)\right]$ is the autocovariance of d at lag $k \leq t$.

Definition 3.7. *The Diebold Mariano test statistic is defined as*

$$\mathrm{DM} = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{T}}}, \tag{3.11}$$

where $\hat{f}_d(0)$ is a consistent estimator of $f_d(0)$.

In practice, it is sufficient to use

$$\hat{f}_d(0) = \frac{1}{2\pi}\left(\hat{\gamma}_d(0) + 2\sum_{k=1}^{M}\hat{\gamma}_d(k)\right), \tag{3.12}$$

where $M = T^{1/3}$ and $\hat{\gamma}_d(k) = \frac{1}{T}\sum_{i=k+1}^{T}\left(d_t - \bar{d}\right)\left(d_{t-k} - \bar{d}\right)$.

The result in (3.10) implies that, for a large sample of forecasts, DM approximately follows the $\mathcal{N}(0,1)$ distribution and we can therefore base a statistical test for $H_0 : \mathbb{E}(d_t) = 0$ on this distribution.

The Harvey, Leybourne and Newbold Test (1997)

Diebold and Mariano (1995) show through simulations that the normal approximation can be very poor, as it rejects the null hypothesis too often. Harvey et al. (1997) propose a correction to this.

Definition 3.8. *Let h be the number of steps ahead that $\{\hat{y}_{1t}\}_{t=1}^{T}, \{\hat{y}_{2t}\}_{t=1}^{T}$ are forecast. The Harvey, Leybourne and Newbold Test statistic is defined as*

$$\mathrm{HLN} = \sqrt{\frac{T + 1 - 2h + h(h+1)}{T}}\mathrm{DM}. \tag{3.13}$$

Harvey et al. (1997) claim that HLN approximately follows a t_{T-1}, where t_{T-1} denotes the Student Distribution with $T-1$ degrees of freedom. It is then easy to construct a test with the HLN statistic that compares accuracy of different forecasts. This is the method of choice for comparison in this paper.

Chapter 4

Quantitative Data Analysis

4.1 The Datasets

The MINT data was sourced from the Reuters Database at CEDIF - UNIL. The United States data was taken from the U.S. Department of the Treasury Resource Center.

This paper assesses the performance of previously mentioned models on the set of daily zero-coupon government bond yields in MINT countries (Mexico, Indonesia, Nigeria, Turkey) from January 1st 2011 to October 31st 2019. The methods are also applied to the United States Treasury data in order to understand how the methods perform in a more traditional setting.

The maturities, which vary from country to country, are specified in the monthly plots of the datasets below (see Figures 4.1 – 4.5). Each figure represents yield curves in three dimensions evolving over time. It plots beginning-of-the-month yields of the country mentioned in the caption. The maturities (in years) are specified on the x-axis, the years of measurement on the y-axis (each plot spans from November 2011 – October 2019) and the yield (in percentage points) on the z-axis. Furthermore, descriptive statistics per country were included in order to give some numerical context for the data (see Tables 4.1 – 4.5). The statistics that were computed are specified in the header of each table, and are all rounded to two decimal points.

Note that to account for missing data on non-business days, the missing values are taken to be equal to the last known value (e.g. the weekend's yields are equal to Friday's).

United States

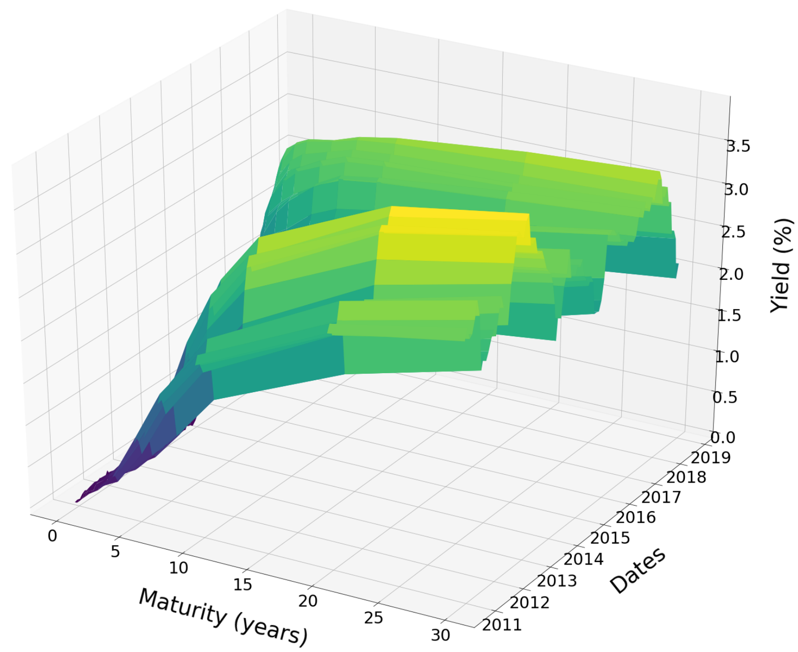


Figure 4.1: Monthly plot of the United States yield curve data. The structure of this plot can be found in Section 4.1.

Maturity	Mean	St. D.	Minimum	Maximum	$\rho(1)$	$\rho(12)$
1 month	0.62	0.82	0.00	2.51	0.67	0.45
3 months	0.66	0.84	0.00	2.49	0.70	0.49
6 months	0.75	0.86	0.02	2.58	0.73	0.52
1 year	0.84	0.86	0.09	2.74	0.74	0.54
2 years	1.03	0.80	0.20	2.98	0.64	0.45
3 years	1.22	0.74	0.28	3.05	0.54	0.35
5 years	1.62	0.61	0.56	3.09	0.36	0.17
7 years	1.95	0.52	0.91	3.18	0.25	0.05
10 years	2.26	0.43	1.37	3.24	0.17	-0.02
20 years	2.71	0.40	1.69	3.72	0.14	-0.02
30 years	2.98	0.38	1.94	3.96	0.12	0.01

Table 4.1: Descriptive Statistics - United States Data.

Mexico

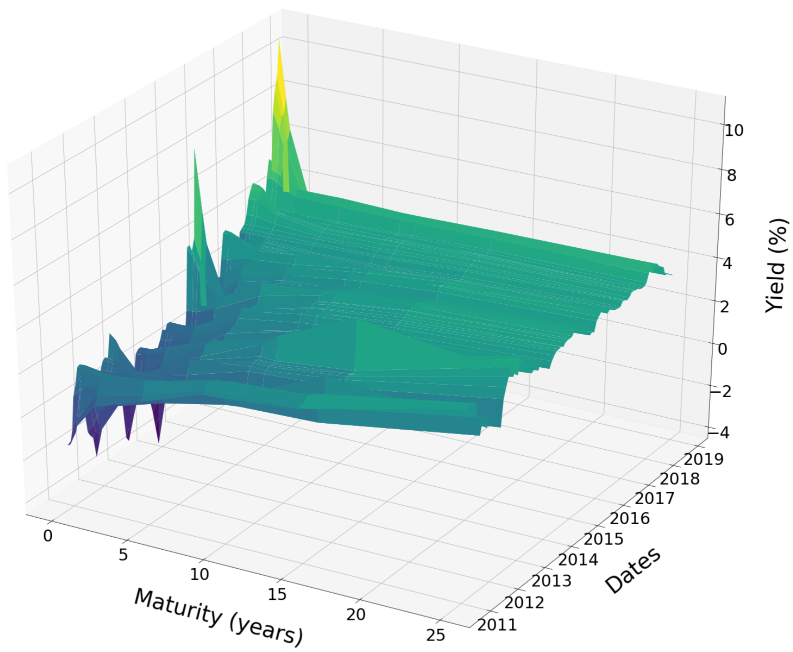


Figure 4.2: Monthly plot of the Mexico yield curve data. The structure of this plot can be found in Section 4.1.

Maturity	Mean	St. D.	Minimum	Maximum	$\rho(1)$	$\rho(12)$
1 month	1.52	2.59	-4.46	11.78	0.86	0.59
6 months	1.59	2.20	-2.89	8.93	0.90	0.66
9 months	1.65	1.95	-1,98	7.27	0.93	0.71
2 years	1.91	1.36	-0.35	4.99	0.98	0.80
5 years	2.48	0.99	0.39	4.72	0.97	0.74
9 years	2.90	0.76	1.00	4.68	0.96	0.57
10 years	2.96	0.75	1.02	4.66	0.96	0.58
15 years	3.24	0.67	1.26	4.90	0.92	0.53
25 years	3.85	0.47	2.12	5.13	0.83	-0.50

Table 4.2: Descriptive Statistics - Mexico Data.

Indonesia

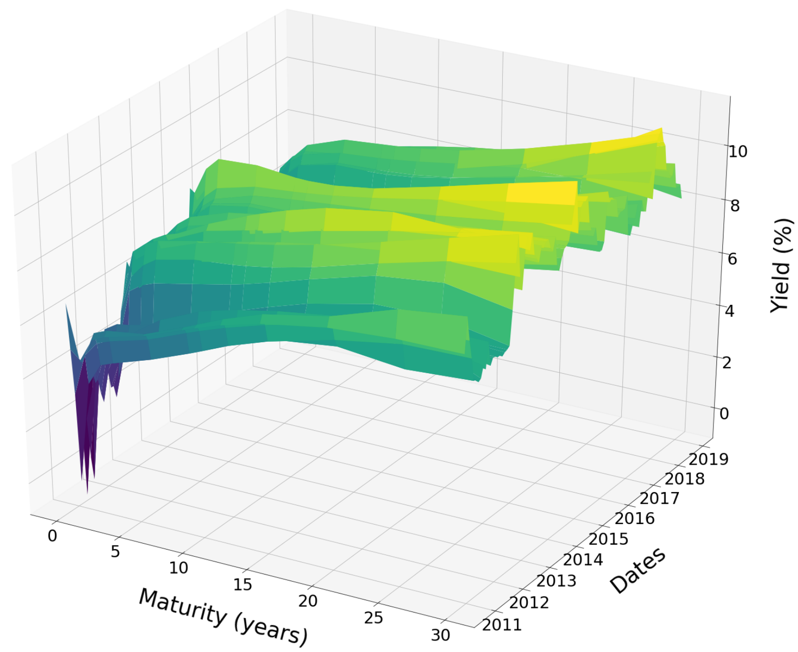


Figure 4.3: Monthly plot of the Indonesia yield curve data. The structure of this plot can be found in Section 4.1.

Maturity	Mean	St. D.	Minimum	Maximum	$\rho(1)$	$\rho(12)$
1 month	5.07	1.56	-1.88	10.59	0.84	0.30
1 year	6.08	1.05	3.27	8.17	0.94	0.29
2 years	6.57	1.09	4.03	9.32	0.94	0.26
3 years	6.86	1.15	4.27	9.86	0.94	0.24
6 years	7.26	1.12	4.73	10.06	0.93	0.16
9 years	7.51	0.99	5.30	9.99	0.92	0.08
10 years	7.60	0.96	5.46	9.98	0.91	0.06
12 years	7.79	0.98	5.77	10.00	0.90	0.03
15 years	8.10	0.84	6.23	10.26	0.89	-0.01
20 years	8.60	0.84	6.69	10.75	0.90	-0.03
25 years	8.95	0.94	6.77	11.18	0.89	0.00
30 years	9.12	1.19	6.51	11.94	0.85	0.09

Table 4.3: Descriptive Statistics - Indonesia Data.

Nigeria

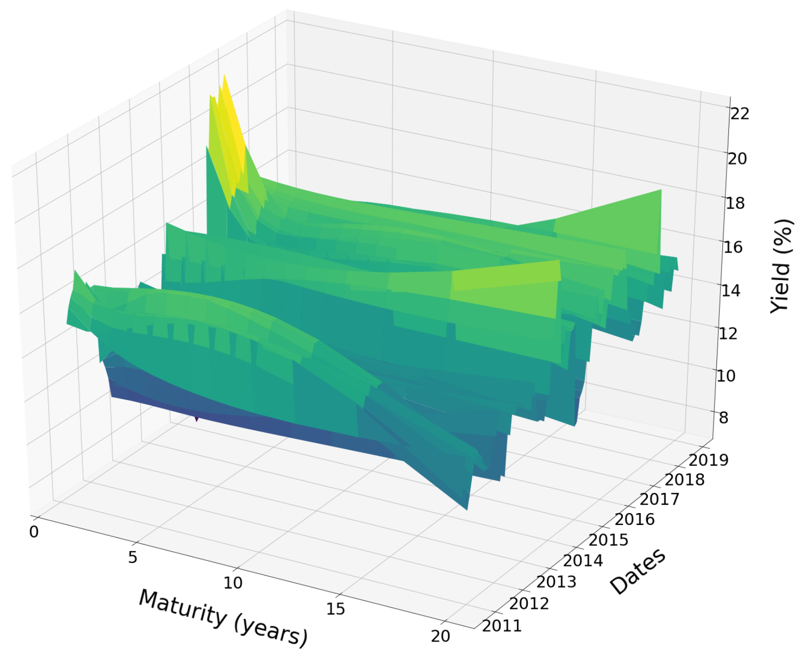


Figure 4.4: Monthly plot of the Nigeria yield curve data. The structure of this plot can be found in Section 4.1.

Maturity	Mean	St. D.	Minimum	Maximum	$\rho(1)$	$\rho(12)$
1 year	14.77	2.99	5.40	8.17	0.88	-0.32
2 years	14.53	2.07	7.87	9.32	0.88	-0.33
3 years	14.46	1.73	9.12	9.86	0.88	-0.29
4 years	14.48	1.61	9.88	9.32	0.87	-0.25
5 years	14.52	1.57	10.33	9.32	0.86	-0.22
6 years	14.57	1.55	10.51	10.06	0.85	-0.20
7 years	14.62	1.55	10.59	9.32	0.85	-0.19
8 years	14.67	1.54	10.67	9.32	0.84	-0.18
9 years	14.70	1.53	10.70	9.99	0.84	-0.17
10 years	14.72	1.52	10.70	9.98	0.84	-0.16
12 years	14.73	1.49	10.76	10.00	0.84	-0.14
15 years	14.67	1.44	10.94	10.26	0.84	-0.08
20 years	14.42	1.70	8.75	10.75	0.75	-0.00

Table 4.4: Descriptive Statistics - Nigeria Data.

Turkey

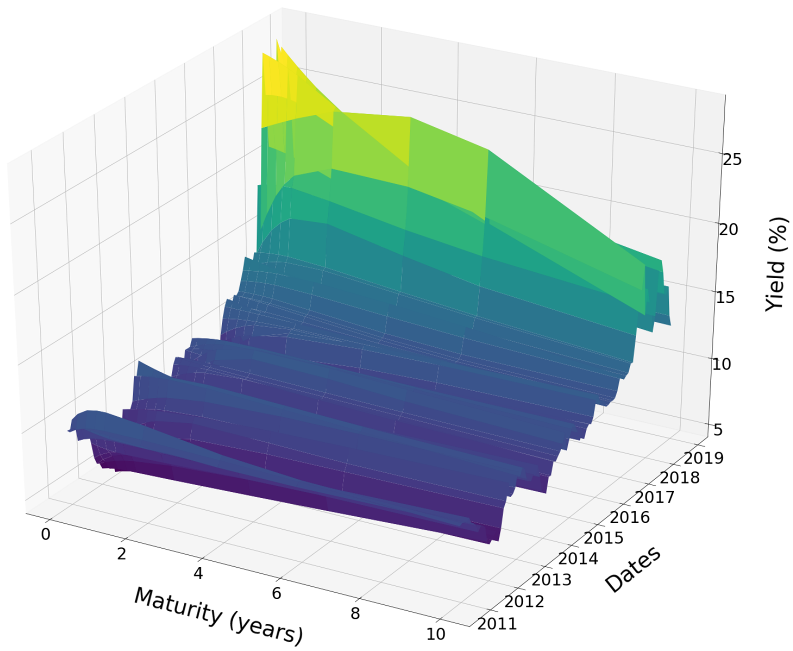


Figure 4.5: Monthly plot of the Turkey yield curve data. The structure of this plot can be found in Section 4.1.

Maturity	Mean	St. D.	Minimum	Maximum	$\rho(1)$	$\rho(12)$
1 month	11.31	5.31	3.71	30.44	0.94	0.38
3 months	11.42	5.18	3.92	29.09	0.96	0.40
6 months	11.53	5.05	4.21	28.45	0.97	0.42
9 months	11.59	4.96	4.37	27.92	0.97	0.43
1 year	11.62	4.88	4.50	27.78	0.97	0.43
2 years	11.58	4.60	4.93	28.84	0.96	0.44
4 years	11.35	4.07	5.49	27.67	0.96	0.45
6 years	11.13	3.60	5.88	25.34	0.95	0.47
10 years	10.81	2.81	5.77	19.70	0.95	0.50

Table 4.5: Descriptive Statistics - Turkey Data.

4.2 Implementation

All methods in this paper have been implemented in Python (Spyder 3.3.6 - Anaconda).

The `statsmodels` and `pykalman` packages were particularly useful in the coding process as they have implementations of the main methods used, such as ordinary least-squares regression, autoregression, vector autoregression and the Kalman Filter with Expectation Maximisation.

As in Caldeira et al. (2019), we use the rolling window method described in Algorithm 1 to determine the accuracy of each method using the random walk model as a benchmark.

Notations Let

- $M = \{\tau_1, \ldots, \tau_p\}$ be the set of maturities,
- $\mathcal{D} = \{y_1, \ldots, y_T\}$ be the data set at hand, where $y_t \in \mathbb{R}^n$ is the vector of yields at time $t \in \{1, \ldots T\}$,
- $\texttt{method} \in \{\texttt{RW}, \texttt{DNS+VAR(1)}, \texttt{DNS + KF}, \texttt{AR(1)}, \texttt{VAR(1)}\}$ be the model used (see following section for description of each model),
- $i \in \{7, 30, 90, 180, 360\}$ [1] be the horizon (in days) we want to forecast at.

The rolling window process is described in Algorithm 1.

Algorithm 1 Rolling Window

Input: ($\texttt{method}$, i) ▷ described in **Notations**
Output: matrix of all predictions i days ahead with method $\texttt{method}$
procedure Rolling Window($\texttt{method}, i$)
  $\hat{y}_t^{\texttt{method}} \in \mathbb{R}^{(T-1080)\times p}$
  for $t = 1, \ldots, T - 1080$ **do**
    $W_t \leftarrow \{y_t, \cdots, y_{t+719}\}$
    $model \leftarrow$ estimate_model $(W_t, \texttt{method})$
    $\hat{y}_t^{\texttt{method}} \leftarrow$ forecast$(model, i)$

  return $\hat{y}^{\texttt{method}}$
procedure Estimate_Model($W_t, \texttt{method}$)
  return model estimated using $\texttt{method}$ on W_t
procedure forecast($model, i$)
  return forecast using $model$, i days in advance

We then compare $\{\hat{y}_t^{\texttt{method}}\}_{t=1}^{T-1080}$ with $\{\hat{y}_t^{\texttt{RW}}\}_{t=1}^{T-1080}$ (from Algorithm 1) using the Harvey, Leybourne, Newbold test to determine if `method` outperforms `RW`.

[1] The 360-day calendar is adopted, which is common in financial markets. One year consists of 12 months of 30 days each.

Model Notation and Description

Recall that `method` $\in \{$`RW`, `DNS+VAR(1)`, `DNS + KF`, `AR(1)`, `VAR(1)`$\}$. This section clarifies to which method each acronym refers to. The following notations will also be used in the results tables and their analysis.

- `RW`: random walk.
- `DNS+VAR(1)`: two-step Dynamic Nelson-Siegel method using a VAR(1) process to model the DNS factors obtained by an ordinary least squares linear regression.
- `DNS+KF`: one-step Dynamic Nelson-Siegel using the Kalman Smoother to estimate the states, and the Expectation Maximisation algorithm to estimate the transition matrix, transition offset, transition covariance, observation offset, observation covariance and initial state mean and covariance. The observation matrix and the observation offset are fixed as per the Nelson-Siegel scheme. The selection matrix is supposed to be the identity matrix. Recall that the initialisation of the parameters to be estimated in the EM algorithm is essential. We initialised
  - the transition matrix and transition offset as the transition matrix and transition offset estimated in `DNS+VAR(1)` (see description above),
  - the transition covariance as $0.1 \times I_3$ (where I_3 denotes the three-dimensional identity matrix),
  - the observation covariance as $0.1 \times I_p$ where p is the dimension of the maturity vector,
  - the initial state mean as the OLS fit of the first observation in the window,
  - the initial state covariance as $0.1 \times I_3$.
- `AR(1)`: separate AR(1) processes on $\{(y_t)_i\}_t$ for $i = 1, \ldots, p$, similarily to Caldeira et al. (2019).
- `VAR(1)`: VAR(1) process on $\{y_t\}_t$, similarly to Caldeira et al. (2019).

Note The methods `DNS+VAR(1)` and `DNS+KF` require the choice of a parameter λ. The implemented selection of λ is the non-linear least-squares estimation described in Subsection 2.1.3, using a grid search with step 0.001 in $[0.001, 2]$ to find

$$\hat{\lambda}_t = \underset{\lambda \in \mathbb{R}}{\operatorname{argmin}} \ \mathrm{OLS}_t^{\mathrm{NS}}(\lambda), \quad t = 1, \ldots, T. \tag{4.1}$$

4.3 The Results

The results of the implementation discussed in Section 4.2 can be found in Tables 4.7 – 4.11.

The first column lists the RMSE for the random walk, which we will use as our benchmark. Each column after that presents the RMSE ratio for the method listed in the header (rounded to two decimal points) in the prediction scheme described above the table. A ratio greater than unity indicates that the method fails to outperform the random walk. The RMSE ratio is italicised and in bold / italicised if the HLN test declares that the method outperforms the random walk at a 5%, 10% level respectively. The ratio being replaced by $\times\ \times\ \times$ indicates that it is superior to 50. This arises as some methods produce defective forecasts due to nonstationarity of the dataset. In this case, we do not apply the HLN test either as it is of no use in these extremes.

A summary of the results obtained with the non-linear least squares computation of λ can be found in Table 4.6. Note that the estimation could have generated a value outside $[0.001, 2]$. We do not take this into account as values in these bounds already fit the data well, and most estimated values are in these bounds. Furthermore, the theory around the Nelson-Siegel framework was not developed with more extreme values in mind — recall Diebold and Rudebusch (2013) chooses $\lambda = 0.0609$ — and the model is robust to the choice of λ.

United States			
Mean	St. D.	Min.	Max.
0.496	0.291	≤ 0.001	1.796

Mexico				Indonesia			
Mean	St. D.	Min.	Max.	Mean	St. D.	Min.	Max.
0.518	0.440	≤ 0.001	≥ 1.999	0.299	0.378	≤ 0.001	≥ 1.999
Nigeria				**Turkey**			
Mean	St. D.	Min.	Max.	Mean	St. D.	Min.	Max.
0.355	0.447	≤ 0.001	≥ 1.999	0.783	0.529	≤ 0.001	≥ 1.999

Table 4.6: Descriptive statistics of the λ computation per country, rounded to three decimal points

Results of the Quantitative Comparison of the Selected Methods

United States

1 week horizon					
Maturity	`RW`	`DNS+VAR(1)`	`DNS+KF`	`AR(1)`	`VAR(1)`
1 month	0.05	1.41	2.10	1.28	1.86
3 months	0.04	1.42	1.42	1.31	2.18
6 months	0.04	2.44	1.98	1.28	2.21
1 year	0.04	2.66	1.08	1.23	2.04
2 years	0.06	1.64	1.38	1.17	1.68
3 years	0.08	1.15	1.06	1.10	1.45
5 years	0.10	1.27	1.25	1.05	1.29
7 years	0.10	1.38	1.34	1.04	1.27
10 years	0.10	1.30	1.26	1.04	1.28
20 years	0.10	1.36	1.23	1.04	1.31
30 years	0.09	1.42	1.47	1.04	1.28
1 month horizon					
Maturity	`RW`	`DNS+VAR(1)`	`DNS+KF`	`AR(1)`	`VAR(1)`
1 month	0.08	1.44	1.36	1.89	2.39
3 months	0.08	1.94	1.01	1.80	2.62
6 months	0.09	2.37	1.17	1.66	2.49
1 year	0.10	2.36	1.23	1.62	2.39
2 years	0.12	1.75	1.10	1.54	2.00
3 years	0.15	1.38	1.01	1.32	1.67
5 years	0.18	1.31	1.08	1.16	1.42
7 years	0.19	1.31	1.10	1.13	1.34
10 years	0.18	1.25	1.04	1.13	1.34
20 years	0.18	1.27	1.04	1.12	1.32
30 years	0.17	1.21	1.22	1.11	1.27
3 month horizon					
Maturity	`RW`	`DNS+VAR(1)`	`DNS+KF`	`AR(1)`	`VAR(1)`
1 month	0.18	1.80	0.93	2.14	2.25
3 months	0.18	1.98	***0.81***	1.95	2.29
6 months	0.19	2.23	***0.89***	1.90	2.28
1 year	0.19	2.28	*0.94*	1.95	2.31
2 years	0.22	1.87	0.97	1.84	1.98
3 years	0.24	1.59	0.99	1.58	1.73
5 years	0.27	1.39	1.06	1.32	1.47
7 years	0.29	1.28	1.06	1.22	1.31
10 years	0.29	1.13	0.98	1.16	1.20
20 years	0.31	1.13	1.05	1.10	1.13
30 years	0.29	1.07	1.25	1.05	1.05

6 month horizon					
Maturity	`RW`	`DNS+VAR(1)`	`DNS+KF`	`AR(1)`	`VAR(1)`
1 month	0.33	1.70	***0.84***	1.93	1.90
3 months	0.32	1.82	***0.84***	1.84	1.95
6 months	0.33	1.94	***0.90***	1.80	1.94
1 year	0.34	1.93	***0.96***	1.81	1.92
2 years	0.37	1.67	1.05	1.69	1.69
3 years	0.39	1.46	1.10	1.50	1.51
5 years	0.40	1.26	1.16	1.27	1.32
7 years	0.41	1.13	1.13	1.14	1.16
10 years	0.42	*0.96*	1.07	1.06	1.05
20 years	0.43	0.98	1.17	1.00	1.01
30 years	0.40	***0.95***	1.43	***0.94***	***0.94***
12 month horizon					
Maturity	`RW`	`DNS+VAR(1)`	`DNS+KF`	`AR(1)`	`VAR(1)`
1 month	0.60	1.48	1.20	1.57	1.55
3 months	0.60	1.51	1.26	1.54	1.55
6 months	0.62	1.54	1.43	1.50	1.52
1 year	0.64	1.47	1.45	1.44	1.45
2 years	0.66	1.31	1.61	1.32	1.30
3 years	0.65	1.20	1.69	1.22	1.21
5 years	0.62	1.07	1.73	1.07	1.08
7 years	0.60	***0.97***	1.69	***0.97***	***0.98***
10 years	0.60	***0.87***	1.60	***0.92***	***0.92***
20 years	0.59	***0.91***	1.74	***0.92***	***0.93***
30 years	0.53	***0.90***	2.14	***0.89***	***0.90***

Table 4.7: Rolling window results - United States. The structure of the table is explained in Section 4.3.

Mexico

Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 week horizon					
1 month	0.78	0.96	1.01	0.99	0.94
6 months	0.54	1.00	1.05	1.00	0.97
9 months	0.40	1.04	1.10	1.01	1.00
2 years	0.18	1.68	1.77	1.10	1.26
5 years	0.14	1.91	1.60	1.05	1.20
9 years	0.13	1.26	1.20	1.07	1.20
10 years	0.14	1.21	1.26	1.10	1.19
15 years	0.20	1.08	1.16	1.12	1.02
25 years	0.13	2.34	2.18	0.99	1.05
1 month horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 month	1.31	0.92	0.99	0.98	0.90
6 months	0.95	0.97	1.02	1.03	0.95
9 months	0.75	0.99	1.05	1.05	1.00
2 years	0.33	1.36	1.29	1.35	1.45
5 years	0.24	1.67	1.36	1.23	1.40
9 years	0.20	1.26	1.11	1.34	1.48
10 years	0.21	1.16	1.07	1.45	1.48
15 years	0.24	1.03	1.04	1.60	1.37
25 years	0.20	1.83	1.68	1.01	1.09
3 month horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 month	2.11	***0.85***	***0.89***	***0.90***	***0.85***
6 months	1.63	***0.88***	***0.90***	0.96	***0.86***
9 months	1.33	***0.92***	***0.92***	1.01	***0.93***
2 years	0.57	1.37	1.13	1.42	1.42
5 years	0.37	1.64	1.30	1.43	1.47
9 years	0.30	1.30	1.10	1.48	1.53
10 years	0.29	1.22	1.05	1.62	1.59
15 years	0.30	1.08	0.99	1.64	1.50
25 years	0.32	1.34	1.25	***0.79***	***0.85***

6 month horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 month	2.91	***0.83***	***0.84***	***0.86***	***0.83***
6 months	2.25	***0.86***	***0.83***	***0.91***	***0.87***
9 months	1.81	0.91	0.84	0.97	0.92
2 years	0.71	1.50	1.11	1.56	1.53
5 years	0.46	1.69	1.32	1.56	1.53
9 years	0.38	1.25	1.09	1.47	1.45
10 years	0.37	1.19	1.06	1.57	1.52
15 years	0.36	1.08	1.04	1.52	1.46
25 years	0.41	1.03	1.02	***0.58***	***0.64***
12 month horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 month	2.21	1.21	***0.95***	1.23	1.22
6 months	1.69	1.33	1.00	1.36	1.35
9 months	1.40	1.44	1.05	1.48	1.46
2 years	0.89	1.64	1.21	1.65	1.64
5 years	0.58	1.76	1.41	1.63	1.61
9 years	0.45	1.37	1.18	1.55	1.53
10 years	0.45	1.27	1.11	1.58	1.55
15 years	0.40	1.19	1.12	1.59	1.56
25 years	0.42	1.11	1.09	***0.68***	***0.67***

Table 4.8: Rolling window results - Mexico. The structure of the table is explained in Section 4.3.

Indonesia

1 week horizon					
Maturity	`RW`	`DNS+VAR(1)`	`DNS+KF`	`AR(1)`	`VAR(1)`
1 month	0.54	1.20	1.26	1.31	0.94
1 year	0.19	1.41	1.14	1.10	1.09
2 years	0.18	2.32	2.00	1.05	1.15
3 years	0.19	2.37	2.11	1.03	1.13
6 years	0.21	1.15	1.12	1.02	1.10
9 years	0.22	1.40	1.49	1.01	1.09
10 years	0.22	1.60	1.67	1.01	1.08
12 years	0.22	1.83	1.85	1.01	1.08
15 years	0.22	1.77	1.71	1.00	1.08
20 years	0.20	1.18	1.05	1.01	1.08
25 years	0.27	1.17	1.21	1.01	0.94
30 years	0.52	1.13	1.18	0.98	***0.86***
1 month horizon					
Maturity	`RW`	`DNS+VAR(1)`	`DNS+KF`	`AR(1)`	`VAR(1)`
1 month	0.66	1.09	1.31	1.76	0.94
1 year	0.34	1.36	1.26	1.37	1.22
2 years	0.36	1.60	1.36	1.16	1.23
3 years	0.39	1.59	1.38	1.11	1.21
6 years	0.40	1.11	1.05	1.07	1.17
9 years	0.40	1.14	1.19	1.05	1.13
10 years	0.40	1.21	1.26	1.05	1.12
12 years	0.41	1.29	1.33	1.03	1.09
15 years	0.40	1.24	1.26	1.02	1.07
20 years	0.39	1.02	1.02	1.02	1.04
25 years	0.45	1.04	1.05	1.09	0.97
30 years	0.73	1.00	1.00	1.09	***0.95***
3 month horizon					
Maturity	`RW`	`DNS+VAR(1)`	`DNS+KF`	`AR(1)`	`VAR(1)`
1 month	0.72	1.11	1.29	1.69	1.05
1 year	0.56	1.23	1.22	1.38	1.20
2 years	0.64	1.29	1.20	1.16	1.20
3 years	0.69	1.27	1.18	1.11	1.19
6 years	0.69	1.08	1.07	1.07	1.15
9 years	0.66	1.08	1.15	1.05	1.11
10 years	0.65	1.10	1.19	1.05	1.10
12 years	0.64	1.13	1.22	1.02	1.07
15 years	0.62	1.09	1.18	0.98	1.02
20 years	0.62	*0.94*	1.05	0.96	0.96
25 years	0.71	***0.92***	1.02	1.02	***0.90***
30 years	0.98	***0.91***	0.99	1.03	***0.83***

6 month horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 month	0.82	1.09	1.24	1.47	1.09
1 year	0.79	1.12	1.21	1.21	1.10
2 years	0.93	1.09	1.13	1.04	1.07
3 years	1.00	1.06	1.09	1.00	1.05
6 years	0.99	***0.96***	1.04	0.98	1.03
9 years	0.91	0.97	1.10	***0.96***	1.00
10 years	0.89	0.99	1.12	***0.96***	0.99
12 years	0.86	1.01	1.14	***0.94***	***0.96***
15 years	0.84	0.97	1.09	***0.90***	***0.92***
20 years	0.85	***0.85***	***0.91***	***0.86***	***0.85***
25 years	0.96	***0.82***	***0.85***	***0.85***	***0.81***
30 years	1.23	***0.83***	***0.86***	***0.89***	***0.79***
12 month horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 month	0.93	1.15	1.63	1.35	1.19
1 year	0.93	1.14	1.46	1.16	1.09
2 years	1.10	1.05	1.28	1.04	1.03
3 years	1.19	0.99	1.20	0.99	1.01
6 years	1.17	***0.89***	1.15	***0.94***	***0.96***
9 years	1.08	***0.92***	1.21	***0.91***	***0.93***
10 years	1.05	***0.94***	1.24	***0.91***	***0.92***
12 years	1.00	***0.95***	1.26	***0.89***	***0.89***
15 years	0.95	***0.92***	1.24	***0.86***	***0.86***
20 years	0.93	***0.79***	1.09	***0.80***	***0.79***
25 years	1.02	***0.76***	0.99	***0.79***	***0.76***
30 years	1.27	***0.78***	***0.93***	***0.84***	***0.78***

Table 4.9: Rolling window results - Indonesia. The structure of the table is explained in Section 4.3.

Nigeria

1 week horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 year	0.91	1.29	1.20	1.04	1.36
2 years	0.61	1.11	1.06	1.01	1.33
3 years	0.50	1.18	1.11	1.00	1.22
4 years	0.46	1.14	1.09	1.00	1.14
5 years	0.44	1.06	1.07	1.00	1.10
6 years	0.43	1.02	1.08	1.00	1.07
7 years	0.42	1.02	1.10	1.00	1.06
8 years	0.42	1.04	1.11	1.00	1.06
9 years	0.43	1.06	1.11	1.00	1.07
10 years	0.43	1.07	1.10	1.01	1.09
12 years	0.43	1.09	1.08	1.01	1.13
15 years	0.46	1.07	1.05	1.03	1.17
20 years	0.77	1.04	1.06	1.05	1.01
1 month horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 year	1.70	1.21	1.12	1.12	1.33
2 years	1.17	1.06	1.03	1.04	1.27
3 years	0.96	1.04	1.04	1.01	1.20
4 years	0.89	1.04	1.05	1.01	1.15
5 years	0.85	1.03	1.06	1.00	1.12
6 years	0.82	1.02	1.07	1.00	1.10
7 years	0.80	1.02	1.08	1.01	1.10
8 years	0.79	1.02	1.09	1.02	1.10
9 years	0.78	1.01	1.08	1.03	1.11
10 years	0.79	1.01	1.06	1.03	1.11
12 years	0.81	0.99	1.00	1.04	1.11
15 years	0.89	0.95	0.92	1.07	1.07
20 years	1.19	0.93	*0.86*	1.17	0.98
3 month horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 year	3.14	1.04	1.07	1.01	1.10
2 years	2.21	0.97	1.07	***0.94***	1.03
3 years	1.88	***0.91***	1.06	0.92	0.97
4 years	1.74	***0.89***	1.06	***0.90***	*0.94*
5 years	1.65	***0.88***	1.06	***0.90***	***0.93***
6 years	1.59	***0.89***	1.07	***0.90***	***0.93***
7 years	1.54	***0.90***	1.08	***0.91***	*0.94*
8 years	1.51	***0.91***	1.08	***0.92***	*0.95*
9 years	1.50	***0.91***	1.08	***0.92***	0.96
10 years	1.49	***0.92***	1.07	***0.93***	0.97
12 years	1.50	***0.92***	1.05	***0.94***	0.97
15 years	1.56	***0.91***	1.01	*0.95*	0.96
20 years	1.84	***0.84***	***0.92***	***0.93***	***0.88***

6 month horizon					
Maturity	`RW`	`DNS+VAR(1)`	`DNS+KF`	`AR(1)`	`VAR(1)`
1 year	4.46	***0.88***	1.06	***0.88***	***0.90***
2 years	3.05	***0.84***	1.11	***0.83***	***0.85***
3 years	2.49	***0.80***	1.10	***0.81***	***0.81***
4 years	2.24	***0.78***	1.09	***0.81***	***0.80***
5 years	2.09	***0.78***	1.08	***0.81***	***0.80***
6 years	2.00	***0.79***	1.08	***0.81***	***0.81***
7 years	1.95	***0.80***	1.08	***0.82***	***0.81***
8 years	1.94	***0.80***	1.07	***0.82***	***0.82***
9 years	1.94	***0.80***	1.07	***0.82***	***0.82***
10 years	1.96	***0.81***	1.06	***0.82***	***0.82***
12 years	2.02	***0.80***	1.05	***0.82***	***0.82***
15 years	2.11	***0.80***	1.04	***0.80***	***0.81***
20 years	2.33	***0.77***	1.02	***0.76***	***0.78***
12 month horizon					
Maturity	`RW`	`DNS+VAR(1)`	`DNS+KF`	`AR(1)`	`VAR(1)`
1 year	5.56	***0.70***	1.18	***0.71***	***0.71***
2 years	3.70	***0.67***	1.32	***0.67***	***0.66***
3 years	2.88	***0.65***	1.39	***0.64***	***0.64***
4 years	2.50	***0.64***	1.41	***0.64***	***0.63***
5 years	2.31	***0.64***	1.41	***0.65***	***0.64***
6 years	2.21	***0.65***	1.40	***0.66***	***0.66***
7 years	2.17	***0.67***	1.39	***0.67***	***0.67***
8 years	2.16	***0.68***	1.38	***0.68***	***0.68***
9 years	2.18	***0.69***	1.37	***0.69***	***0.69***
10 years	2.21	***0.70***	1.35	***0.70***	***0.70***
12 years	2.28	***0.71***	1.32	***0.70***	***0.70***
15 years	2.38	***0.71***	1.29	***0.70***	***0.70***
20 years	2.60	***0.71***	1.21	***0.71***	***0.71***

Table 4.10: Rolling window results - Nigeria. The structure of the table is explained in Section 4.3.

Turkey

1 week horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 month	1.09	1.04	1.02	1.17	1.35
3 months	0.85	1.12	0.98	1.17	1.65
6 months	0.62	1.29	0.94	1.17	2.22
9 months	0.52	1.42	1.00	1.15	2.67
1 year	0.51	1.46	1.12	1.12	2.80
2 years	0.57	1.43	1.25	1.09	2.69
4 years	0.58	1.62	1.04	1.11	2.59
6 years	0.57	1.61	0.98	1.10	2.39
10 years	0.48	1.29	1.45	1.05	1.79
1 month horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 month	1.88	1.09	0.85	1.42	1.41
3 months	1.62	1.18	0.84	1.41	1.59
6 months	1.38	1.32	0.95	1.35	1.83
9 months	1.28	1.41	1.15	1.28	1.98
1 year	1.25	1.44	1.33	1.24	2.04
2 years	1.28	1.42	1.62	1.24	2.00
4 years	1.34	1.36	1.49	1.20	1.80
6 years	1.26	1.32	1.40	1.17	1.70
10 years	0.82	1.27	1.94	1.24	1.62
3 month horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 month	2.54	1.32	1.47	1.61	1.37
3 months	2.41	1.35	2.28	1.59	1.42
6 months	2.35	1.35	3.34	1.47	1.43
9 months	2.39	1.33	4.06	1.33	1.41
1 year	2.46	1.29	4.52	1.24	1.38
2 years	2.63	1.20	5.14	1.17	1.28
4 years	2.63	1.12	4.70	1.11	1.18
6 years	2.35	1.12	4.33	1.10	1.17
10 years	1.42	1.25	5.17	1.29	1.34

6 month horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 month	3.49	1.34	12.70	1.48	1.35
3 months	3.33	1.37	21.75	1.49	1.38
6 months	3.27	1.35	32.88	1.43	1.36
9 months	3.30	1.32	40.51	1.34	1.33
1 year	3.38	1.27	45.44	1.25	1.28
2 years	3.54	1.16	52.65	1.14	1.17
4 years	3.44	1.07	48.77	1.06	1.08
6 years	3.07	1.06	44.28	1.05	1.06
10 years	1.98	1.18	48.41	1.21	1.19
12 month horizon					
Maturity	RW	DNS+VAR(1)	DNS+KF	AR(1)	VAR(1)
1 month	5.60	1.13	× × ×	1.16	1.14
3 months	5.43	1.14	× × ×	1.16	1.14
6 months	5.28	1.14	× × ×	1.16	1.14
9 months	5.20	1.13	× × ×	1.14	1.13
1 year	5.16	1.12	× × ×	1.12	1.12
2 years	4.97	1.08	× × ×	1.08	1.08
4 years	4.45	1.05	× × ×	1.06	1.06
6 years	3.85	1.06	× × ×	1.07	1.06
10 years	2.64	1.17	× × ×	1.18	1.17

Table 4.11: Rolling window results - Turkey. The structure of the table is explained in Section 4.3.

4.4 Results Analysis

This section analyses the performance of the presented methods with respect to three main axes: the horizon, the maturity, and, most importantly, the method.

United States

We conducted the analysis on the United States yield curve data to serve as a benchmark for expected performance. The following subsection analyses Table 4.7.

Horizon analysis

The random walk proves to be virtually impossible to beat in 1 week and 12 month ahead forecasts. Different interpretations can be made in each case.

In the former case, it is obvious that the performance of the random walk is likely to increase when the horizon becomes shorter — this is reflected in the RMSEs. As mentioned previously, it is already non-trivial to beat the random walk. The 1 week ahead scheme is therefore almost impossible to beat. Diebold and Li (2006) reaches the same conclusion.

In the latter case, all methods at all maturities failing to beat the benchmark is not due to the random walk's strength but rather to the weakness of the other models. For example, when the US yield curve exhibits significant oscillations, long-term forecasts for two-step DNS explode. This is reflected in the high ratios for 12 month horizons. Even though Diebold and Li (2006) presents considerably better results, it is important to note that they studied US Treasury data recorded between January 1985 and December 2000. Discrepancies between the two results are therefore likely to have been caused by the high variations present in our data that did not appear in previous decades.

The methods seem to perform better at medium horizons (1 month – 6 months). Equal performance/outperformance is undeniable particularly in 1 month and 3 month ahead forecasts as the HLN test indicates that all methods — `AR(1)` aside — do not significantly fail to outperform the benchmark for many of the maturities.

Maturity analysis

Significant outperformance is achieved in short maturities ($\leq$ 6 months) for medium horizons for most methods (again, it is not the case for `AR(1)`), and sometimes also in long maturities ($\geq$ 10 years), but virtually never in medium maturities (1 – 7 years). A plot of yields with various maturities can be found in Figure 4.6. It doesn't seem impossible for models to extract a pattern in the oscillations of long-term maturity yields or understand the smooth short-term maturity yields. The medium-term maturity yields, however, seem much harder to understand to the human eye and this is reflected in the results of the quantitative comparison of methods. This is a very loose interpretation. Developing a solid theory as to why this happens would call for in-depth analysis.

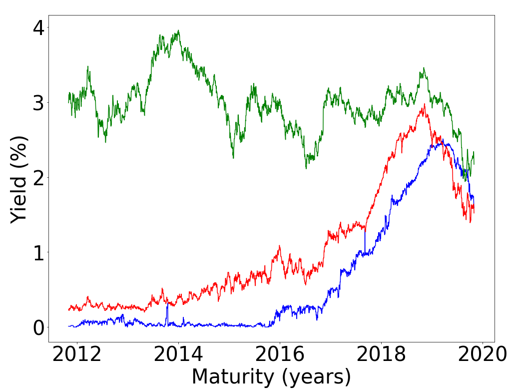


Figure 4.6: Plots of the US yields' evolution over time (in percentage points) at various fixed maturities. The blue curve represents yields with 1 month maturity, the red curve 5 years, and the green curve 25 years.

Method analysis

Firstly, it is very important to note that here, even though the implementation of `DNS+VAR(p)` enables a lag p larger than unity, it always chooses $p = 1$. This is particularly interesting as yield curve models rarely consider a larger lag (see for example Diebold and Rudebusch (2013)) and if models are developed around US data, this is understandable.

The DNS models perform reasonably in medium horizons. It is reasonable to acknowledge that `DNS+KF` is the best performer of the three methods, especially in the 6 and 12 month horizon cases. These methods are, however, neck and neck with `VAR(1)`, which performs very well, particularly in medium horizons. It is also the only method that manages to significantly outperform the random walk in long maturities (see the 3 month horizon).

We mentioned previously that ratios in the two-step DNS methods, and sometimes in `AR(1)`, skyrocket. This is because they involve (vector) autoregressions. Usually, (vector) autoregression processes require stationary data as if not the forecasts are likely to diverge. Therefore, when the rolling window is on the first side of a peak and we are computing long-term forecasts, the ratios explode on the other side. In the two-step DNS case, shocking values are mostly present in yields with low and medium maturity. Figure 4.7 shows that both factors that govern yields with low and medium maturities are nonstationary. This could explain the particularly abnormal ratios for yields with lower maturities.

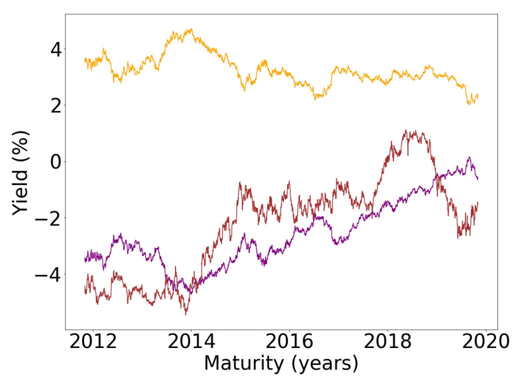


Figure 4.7: Plots of the OLS fitted factors' evolution over time (in percentage points). The orange curve is the level, the purple curve is the slope and the brown curve is the curvature.

Unlike the two-step DNS methods, `AR(1)` and `VAR(1)` seem immune to such blatant divergences. Figure 4.6 shows that the yields with lower maturities only start to exhibit a significant upward trends at a later stage. The fact that the extracted Nelson-Siegel factors display trends early on could explain why the ratios are significantly higher than the other autoregression based models.

The Kalman filter, however, does not pose stationarity restrictions and this is reflected in the RMSE ratios. Even though it does not outperform the random walk (when forecasting 12 months ahead), it is still not drastically wrong like the two-step methods.

Mexico

The following subsection analyses Table 4.8.

Horizon analysis

The methods seem to perform slightly better in extreme horizons than in the US case.

In the 1 week horizon case, the HLN test shows that all methods match the benchmark for low maturities — even `AR(1)`. This is better than expected if we compare to the US results. As the Mexican yields are slightly more volatile than the US yields, this is not necessarily surprising.

In the 12 month horizon case, different methods exhibit different forecasting qualities but the presence of outperformance of the random walk contrasts with the US case.

Like in the US case, the medium-length horizons are the ones that exhibit more equal performance to or outperformance of the random walk.

Maturity analysis

Significant outperformance is mostly achieved in short maturities ($\leq$ 9 months), and the DNS methods even perform at least as well as the random walk for all horizons in the shortest maturities. As the yields with short maturities are very volatile (see Figure 4.8), drastic outperformance is to be expected. Even though such a consistent outperformance cannot be observed for longer maturities ($\geq$ 9 years) – which is normal as they are a lot more stable than yields with short maturities (again, see Figure 4.8) – some matching of performance on the DNS methods can be noted. However, if forecasting the yields with the longest maturity is the aim, one should prefer `VAR(1)` as it always surpasses the other methods and considerably outperforms the random walk for a 25-year maturity at long horizons.

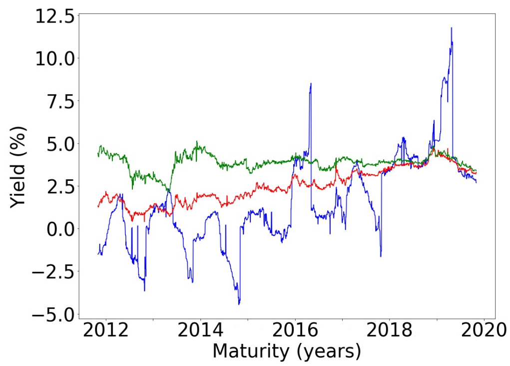


Figure 4.8: Plots of the Mexican yields' evolution over time (in percentage points) at various fixed maturities. The blue curve represents yields with 1 month maturity, the red curve 5 years, and the green curve 25 years.

Again, all methods noticeably fail to match the performance of the random walk at medium maturities (2 – 5 years) for all horizons.

Method analysis

The Dynamic Nelson-Siegel models and the vector autoregression seem to be particularly favourable methods for this dataset, especially in the low maturities, where they exhibit significant outperformance of the random walk.

There seems to be little difference between the DNS methods and therefore, in this case, one might prefer to work with `DNS+VAR(1)` for simplicity. The use of `VAR(1)` could be justified in some select cases, for example when forecasting yields with 25-years maturity at long horizons.

As seen previously, `AR(1)` does not count as a serious contender, even though it can be noted that its' performance in the low maturities sometimes equals that of the random walk. Figure 4.8 shows that the yields with the longest maturities in particular exhibit clear nonstationarity, which is why we have skyrocketing ratios. Note that from now on we will not comment on the performance of `AR(1)` as it is overall very poor. We will only mention it in passing if the results show otherwise in specific cases.

Indonesia

The following subsection analyses Table 4.9.

Horizon analysis

The observations that can be made are similar to the US case, with two noticeable exceptions: `VAR(1)` and `DNS+KF` significantly outperform the random walk for yields with 30 years maturity in the 1 week and 12 month horizons respectively.

Maturity analysis

Contrary to the first two datasets, the methods prove to be more effective in longer maturities ($\geq$ 20 years). Forecasting short and medium maturity yields is visibly a weak spot for the methods considered, which only match the random walk's performance in a few select cases. This is not without reason: the higher-than-average random walk RMSEs can be used as an indicator as to where the methods might perform similarily or outperform the benchmark. Figure 4.9 also showcases why the random walk does not perform as well in long maturities.

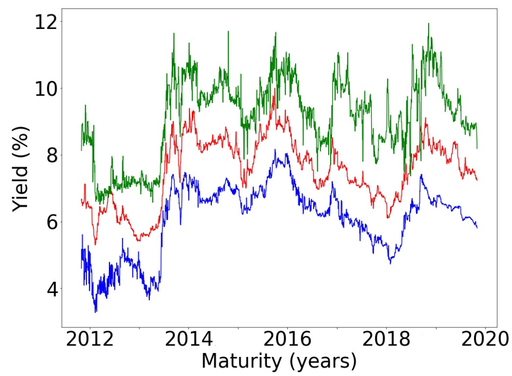


Figure 4.9: Plots of the Indonesian yields' evolution over time (in percentage points) at various fixed maturities. The blue curve represents yields with 1 year maturity, the red curve 9 years, and the green curve 30 years.

Method analysis

In this context, and in contrast with the US, it could be interesting to look at more than 1 lag for the two-step DNS methods. Indeed, choosing `DNS+VAR(p)` over `DNS+VAR(1)` is questionable, as sporadic indicators of (marginal) superior performance of the former appear. However, at this stage, to declare significant outperformance would call for further examination.

Nonetheless, there is a case to be made for superiority of `DNS+KF`, especially when the horizon lengthens, as it matches the random walk's performance more often and the outperformance is clearer. It is also the only method to outperform the benchmark in the 12 month horizon.

It is important to note that `VAR(1)` matches the results for `DNS+KF` in most cases. It is also the only method up until now to outperform the benchmark at the 1 week horizon. The choice between the two methods is again not clear-cut — it essentially depends on the forecasting horizon.

Nigeria

The following subsection analyses Table 4.10.

Horizon analysis

In this case, the methods fare better as horizons lengthen. The Nigerian yields oscillate much more than in the previous cases, which explains why the random walk's performance worsens for longer horizons.

The consistent outperformance of the two-step DNS methods in the 6 and 12 month horizons could be surprising when compared with the previous countries' results if it were not for the clear difference in the random walk's RMSEs.

Maturity analysis

There is less of a pattern to extract than in previous examples. This can be explained by Figure 4.10, where we can see that yields with short, medium and long maturities mostly oscillate in a similar manner.

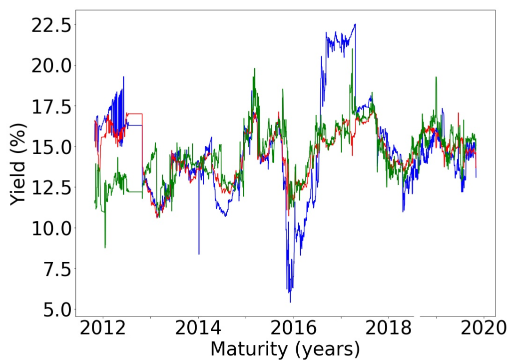


Figure 4.10: Plots of the Nigerian yields' evolution over time (in percentage points) at various fixed maturities. The blue curve represents yields with 1 year maturity, the red curve 7 years, and the green curve 20 years.

Note The Nigerian yields flat-lining between July 17$^{\text{th}}$ and October 29$^{\text{th}}$ 2012 is not without meaning. The country experienced significant floods during this period (NASA Earth Observatory, 2012), probably preventing the updates of yield curve data.

Method analysis

There are clear overall winners in this case: the two-step DNS methods. They outperform consistently and significantly more than any other method, including `DNS+KF`, which was a good contender in previous cases. The plot in Figure 4.10 indicates that the data here presents more stationarity and this could be one of the reasons two-step DNS methods fare this well in comparison to the one-step. That being said, there is definite presence of diverging estimations in `VAR(1)`. The factor extraction of the Nelson-Siegel model is therefore a great asset in this case.

Without further examination, one could not declare either one of the two-step methods to be superior. At this stage, `DNS+VAR(1)` should be chosen for simplicity. Nevertheless, both seem to be very good choices in the set of models we consider here.

Turkey

The following subsection analyses Table 4.11.

Horizon analysis

The results show that all methods struggle in this case. If promising ratios appear, they appear when forecasting short-term. The HLN test, however, still rejects any significant outperformance of the random walk, whatever the forecasting scheme.

Obvious divergence appears with all methods, as ratios shoot up when the horizon lengthens.

Maturity analysis

The divergences that appear are quickly understood when looking at Figure 4.11.

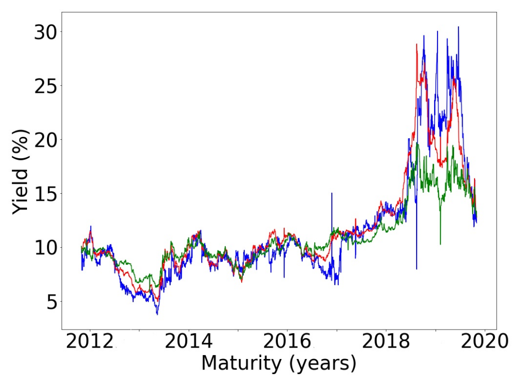


Figure 4.11: Plots of the Turkish yields' evolution over time (in percentage points) at various fixed maturities. The blue curve represents yields with 1 year maturity, the red curve 7 years, and the green curve 20 years.

As can be seen, all yields, be it with short, medium or long maturities, suddenly increase in mid-2018 and become extremely volatile.

Method analysis

There is not much more to add here as it is clear that all methods fail by quite a margin to outperform the random walk. This does not mean that Turkish yields have always been, or will always be, impossible to model with these methods. It simply indicates that the task is near impossible in a transition period such as the one we have selected.

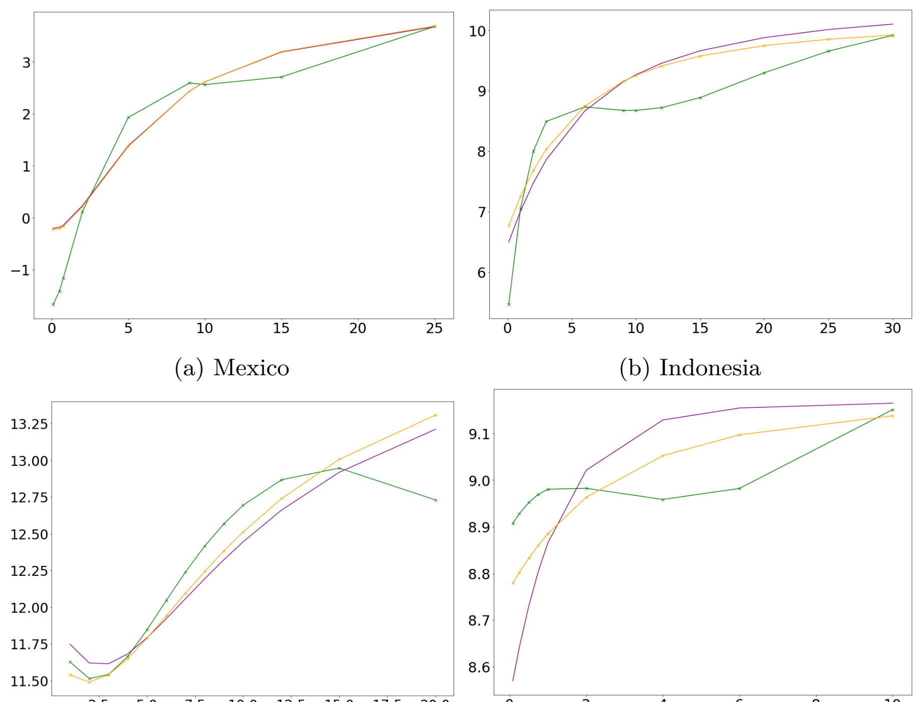
Figure 4.11 showcases the considerable change in regime which is responsible for divergence of the long-term forecasts. Even the Kalman Filter cannot cater for the shift. Incidentally, `AR(1)` is the least affected. This is perhaps explained by the fact that the model does not display as many inter-dependencies as others, thus moderating growth.

General Observations

The following subsection airs some final remarks that apply to all the cases above.

Firstly, it is important to note that determining λ is not a straightforward task. As can be seen in Table 4.6, our non-linear least-squares method is not selecting anything close to $\lambda = 0.0609$, the theoretical value proposed by Nelson-Siegel. Even though the Nelson-Siegel parametrisation is robust to the λ selection, the variance in our method remains very high (see Table 4.6). In order to better capture these fluctuations, a further study could consider comparing this fixed scheme with one where λ is a time-varying parameter.

Secondly, we must take into account one of the big limitation of the Nelson-Siegel parametrisation: the constraint of a single optimum. Traditional yield curves may not present multiple optima, but the ones we consider do. The plots in Figure 4.12 show the noticeable failure of DNS methods to properly fit the curve, even when forecasting only 1 week in advance.



(a) Mexico (b) Indonesia

(c) Nigeria (d) Turkey

Figure 4.12: Plots at selected cross-sections of 1 week ahead yield forecasts (in percentage points) generated by the implementation described in Section 4.2 against their maturity (in years). The true yield curve is in green, the forecasts using `DNS+VAR(1)` are in purple and the forecasts using `DNS+KF` are in orange. The country they originate from can be found in the caption of each figure.

4.5 Conclusion

Firstly, as can be seen in the results, forecasting yield curves is no easy task and in some cases the random walk cannot be outperformed by the methods we compared. In fact, when forecasting a short period of time ahead, it is almost certain that nothing will outperform the random walk. Diebold and Li (2006) showed that classical time-series methods and the Dynamic Nelson-Siegel model produced reasonable forecasts prior to the year 2000 in the US market. As we analysed the US data in a period of economic instability (due to the aftermath of the financial crisis and more), the methods perform poorly in comparison to previous studies, hinting that unpredictability of the yield curves is worsening. As MINT economies are still emerging, there is more instability in their yield curves as well. This induces noticeable failures to forecast.

Choosing one overall superior method for all countries is impossible. We can, however, draw meaningful conclusions as to which methods should be used in which situations.

The Turkish case is a prime example of a significant challenge to forecast yields. The plots of the original data show a considerable change in regime towards the end of the time-frame we studied. This made the data almost impossible to model. While studying the effect of external economic factors on yield curve forecasting is out of the scope of this project, it must not be forgotten that the yield curve values are driven by external economic factors. In this case, it is perhaps wise to study how the yield curve depend on macroeconomic variables or on other yield curves, instead of latent factors, in order to obtain better forecasts. Diebold et al. (2006) and Prasanna and Sowmya (2016) give two examples of what can be done. The former studies the effect of macroeconomic variables on the US yield curve while the latter studies the effect the US yield curve has on the yield curve in India.

The use of vector autoregression can be justified in some select cases (e.g. Mexican yields with long maturities at long horizons) but it generally does not outperform the DNS methods. Moreover, contrary to DNS methods, the number of parameters varies with the number of maturities considered. For these reasons, we do not recommend using a simple vector autoregression unless absolutely necessary. The use of a simple autoregression is obviously not encouraged either.

Yields that exhibit good stationarity (e.g. Nigerian yields) should be modelled with two-step DNS. The sparsity of the Nelson-Siegel parametrisation coupled with the qualities of vector autoregression modelling in stationary cases make it an unbeatable method. There seems to be no significant improvements when using more than one lag, and therefore we recommend using the `DNS+VAR(1)` model in this case.

Nonstationarity (e.g. Indonesian yields with long maturities) is well captured by the Kalman Filter. The fact that it is not constrained to stationary cases therefore makes `DNS+KF` an overall strong and safe method to use. It usually performs as well or better than two-step DNS methods and therefore, if stationarity is to doubt, we recommend the one-step DNS method `DNS+KF`.

Finally, although not studied in similar papers to this one (e.g. Brechtken (2008) or Caldeira et al. (2019)), another possibility to relax the stationarity constraints would be to consider some nonstationary classical time-series models such as vector ARIMA (Autoregressive Integrated Moving Average).

Appendix A

Estimation and Prediction in State-Space Models

A.1 General Linear State-Space Models

The definitions and constructs in this subsection are taken from Durbin and Koopman (2012), Chapter 4.

The State-Space setting is useful to model many different scenarios, as it can model a situation where y_t can be observed but α_t cannot.

Definition A.2. *For $t = 1, \ldots, T$ the general linear Gaussian State-Space model is a process of the form*

$$\begin{aligned} y_t &= Z_t \alpha_t + d_t + \epsilon_t, \\ \alpha_t &= T_t \alpha_{t-1} + c_t + R_t \eta_t, \\ \alpha_1 &\sim \mathcal{N}(a_1, P_1), \end{aligned} \tag{A.1}$$

where $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, H_t)$, $\eta_t \overset{i.i.d.}{\sim} \mathcal{N}(0, Q_t)$[1].

The first equation in (A.1) *is called the space equation. We will refer to*

y_t	$\in \mathbb{R}^p$	*as the observation vector,*
Z_t	$\in \mathbb{R}^{p \times m}$	*as the observation matrix,*
d_t	$\in \mathbb{R}^p$	*as the observation offset,*
H_t	$\in \mathbb{R}^{p \times p}$	*as the observation covariance (symmetric positive definite).*

[1] $\mathcal{N}(\mu, \Sigma)$ denotes the normal distribution with mean μ and covariance matrix Σ

The second equation in (A.1) *is called the state/transition equation. We will refer to*

α_t	$\in \mathbb{R}^m$	*as the (unobserved) state vector,*
T_t	$\in \mathbb{R}^{m\times m}$	*as the transition matrix,*
c_t	$\in \mathbb{R}^m$	*as the transition offset,*
R_t	$\in \mathbb{R}^{m\times r}$	*as the selection matrix,*
Q_t	$\in \mathbb{R}^{r\times r}$	*as the transition covariance (symmetric positive definite).*

The inital state α_1 follows an m-dimensional normal distribution with mean a_1 and covariance P_1.

Note For simplicity, it is possible to define $\eta_t^* = R_t\eta_t$, implying $\eta_t^* \overset{i.i.d.}{\sim} \mathcal{N}(0, Q_t^*)$ where $Q_t^* = R_tQ_tR_t'$ and R_t' denotes the transpose of matrix R_t. In practice R_t is often the identity matrix. Durbin and Koopman (2012) prefers to keep the notation in Definition A.2 as it is possible that $r < m$, which causes η_t^* to be singular while η_t stays non-singular. As mentioned, R_t is often called the selection matrix, The reason for this is that its' columns are a subset of the identity matrix's columns.

The General Linear Gaussian State-Space model is therefore composed of a VAR(1) process (described in Section 3.3) for y_t, and a regression of y_t on α_t for a fixed time t. Structures like these mean that various useful tools have been developed to work with State-Space models. The Kalman Filter in particular is practical and popular to work with.

A.3 The Kalman Filter

A.3.1 Derivation of the Recursion

The Kalman filter is a recursive estimator of the unobservable states $\{\alpha_t\}_{t=1}^T$. The following section describes the derivation of this algorithm. For simplicity we will assume in this derivation that $c_t = 0$, $d_t = 0\ \forall t$, and that a_1 and P_1 are known. The construction bases itself on Lemma (A.4).

Lemma A.4. *Let X, Y be jointly normally distributed random vectors such that*

$$\mathbb{E}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \quad \mathrm{Var}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \Sigma_{X,X} & \Sigma_{X,Y} \\ \Sigma_{X,Y}' & \Sigma_{Y,Y} \end{pmatrix},$$

where $\Sigma_{Y,Y}$ is non-singular. Then

$$X \mid Y \sim \mathcal{N}\left(\mu_X + \Sigma_{X,Y}\Sigma_{Y,Y}^{-1}\left(Y - \mu_y\right), \Sigma_{X,X} - \Sigma_{X,Y}\Sigma_{Y,Y}^{-1}\Sigma_{X,Y}'\right). \tag{A.2}$$

Proof. Let $Z = X - \Sigma_{X,Y}\Sigma_{Y,Y}^{-1}(Y - \mu_Y)$. Then

$$\begin{aligned}\mathbb{E}(Z) &= \mu_X, &\text{(A.3)}\\ \mathrm{Var}(Z) &= \mathbb{E}(ZZ') - \mu_X\mu_X' \\ &= \mathrm{Var}(X) - 2\Sigma_{X,Y}\Sigma_{Y,Y}^{-1}\Sigma_{X,Y}' + \mathrm{Var}\left(\Sigma_{X,Y}\Sigma_{Y,Y}^{-1}Y\right) \\ &= \Sigma_{X,X} - \Sigma_{X,Y}\Sigma_{Y,Y}^{-1}\Sigma_{X,Y}'. &\text{(A.4)}\end{aligned}$$

Furthermore,

$$\begin{aligned}\mathrm{Cov}(Y,Z) &= \mathbb{E}\left[(Y-\mu_Y)(Z-\mu_X)'\right] \\ &= \Sigma_{X,Y}' - \mathbb{E}\left[(Y-\mu_Y)(Y'-\mu_Y')\Sigma_{Y,Y}^{-1}\Sigma_{X,Y}'\right] \\ &= 0.\end{aligned}$$

Z is a linear transformation of X and Y, and therefore is also normal. Because uncorrelated Gaussians are independent, $Z \mid Y \sim \mathcal{N}(\mu_Z, \Sigma_Z)$ where μ_Z and Σ_Z are given in (A.3) and (A.4). The definition of Z implies that the conditional distribution of X given Y is also normal, with mean and variance as stated in Lemma A.4. □

Let $Y_{t-1} = \{y_1, \ldots, y_{t-1}\}$ be the set of previous observations. We suppose that Y_0 indicates that there are no previous observations. Let

$$\begin{aligned}a_{t|t} &= \mathbb{E}(\alpha_t \mid Y_t), & a_{t+1} &= \mathbb{E}(\alpha_{t+1} \mid Y_t), \\ P_{t|t} &= \mathrm{Var}(\alpha_t \mid Y_t), & P_{t+1} &= \mathrm{Var}(\alpha_{t+1} \mid Y_t).\end{aligned}$$

Because we assumed that the error terms ϵ_t, η_t were normally distributed, α_t, y_t and Y_t are also normally distributed and therefore

$$\alpha_t \mid Y_t \sim \mathcal{N}\left(a_{t|t}, P_{t|t}\right), \tag{A.5}$$

$$\alpha_{t+1} \mid Y_t \sim \mathcal{N}\left(a_{t+1}, P_{t+1}\right). \tag{A.6}$$

The aim from this point on is therefore to understand the recursion structures for $a_{t|t}, a_{t+1}, P_{t|t}, P_{t+1}$. The states α_t can then be estimated with the computed a_t, where P_t is an accuracy indicator of this estimation.

Let

$$v_t = y_t - \mathbb{E}(y_t \mid Y_{t-1}).$$

The definition of the space equation yields

$$v_t = y_t - \mathbb{E}\left(Z_t \alpha_t + \epsilon_t \mid Y_{t-1}\right) = y_t - Z_t a_t.$$

Notice that in particular for $s = 1, \ldots, t-1$

$$\begin{aligned}
\mathbb{E}\left(v_t \mid Y_{t-1}\right) &= \mathbb{E}\left(Z_t \alpha_t + \epsilon_t - Z_t a_t \mid Y_{t-1}\right) = 0 \qquad \implies \mathbb{E}\left(v_t\right) = 0, && \text{(A.7)}\\
\operatorname{Cov}\left(y_s, v_t\right) &= \mathbb{E}\left(y_s v_t'\right) - \mathbb{E}\left(y_s\right)\mathbb{E}\left(v_t\right)' \\
&= \mathbb{E}\left(y_s y_t' - y_s \mathbb{E}\left(y_t \mid Y_{t-1}\right)'\right) \qquad = 0. && \text{(A.8)}
\end{aligned}$$

Because if Y_t is known then so are Y_{t-1} and v_t,

$$\begin{aligned}
a_{t|t} &= \mathbb{E}\left(\alpha_t \mid Y_t\right) = \mathbb{E}\left(\alpha_t \mid Y_{t-1}, v_t\right), && \text{(A.9)}\\
a_{t+1} &= \mathbb{E}\left(\alpha_{t+1} \mid Y_t\right) = \mathbb{E}\left(\alpha_{t+1} \mid Y_{t-1}, v_t\right). && \text{(A.10)}
\end{aligned}$$

With Equation (A.9), the application of Lemma A.4 to the joint distribution of α_t and v_t given Y_{t-1}, coupled with (A.7) and (A.8) gives

$$a_{t|t} = \mathbb{E}\left(\alpha_t \mid Y_{t-1}\right) + \operatorname{Cov}\left(\alpha_t, v_t\right)\left[\operatorname{Var}\left(v_t\right)\right]^{-1} v_t. \tag{A.11}$$

Note The variance and covariance in Equation A.11 refer to the *joint* distribution of α_t and v_t given Y_{t-1}.

We can simplify Equation A.11 to

$$a_{t|t} = a_t + P_t Z_t F_t^{-1} v_t, \tag{A.12}$$

where $F_t = \operatorname{Var}\left(v_t \mid Y_{t-1}\right) = Z_t P_t Z_t' + H_t$ is assumed to be non-singular [2]. This result can be derived by substituting the definition of a_t into (A.11) and noticing that

$$\begin{aligned}
\operatorname{Cov}\left(\alpha_t, v_t\right) &= \mathbb{E}\left[\alpha_t \left(Z_t \alpha_t + \epsilon_t - Z_t a_t\right)' \mid Y_{t-1}\right]\\
&= \mathbb{E}\left[\alpha_t \left(\alpha_t - a_t\right) Z_t' \mid Y_{t-1}\right] = P_t Z_t'
\end{aligned}$$

and

$$\operatorname{Var}\left(v_t \mid Y_{t-1}\right) = \operatorname{Var}\left(Z_t \alpha_t + \epsilon_t - Z_t a_t \mid Y_{t-1}\right) = Z_t P_t Z_t' + H_t.$$

A similar reasoning can be applied to derive the recursion of $P_{t|t}$:

[2] This is usually a reasonable assumption. However, if it is not the case, the procedure to relax this assumption can be found in Durbin and Koopman (2012), Section 6.4.

$$
\begin{aligned}
P_{t|t} &= \mathrm{Var}\left(\alpha_t \mid Y_t\right) = \mathrm{Var}\left(\alpha_t \mid Y_{t-1}, v_t\right) \\
&= \mathrm{Var}\left(\alpha_t \mid Y_{t-1}\right) - \mathrm{Cov}\left(\alpha_t, v_t\right)\left[\mathrm{Var}\left(v_t\right)\right]^{-1} \mathrm{Cov}\left(\alpha_t, v_t\right)' \\
&= P_t - P_t Z_t F_t^{-1} Z_t P_t. \qquad \text{(A.13)}
\end{aligned}
$$

It now remains to develop a_{t+1} and P_{t+1}:

$$
\begin{aligned}
a_{t+1} &= \mathbb{E}\left(\alpha_{t+1} \mid Y_t\right) = \mathbb{E}\left(T_t \alpha_t + R_t \eta_t \mid Y_t\right) \\
&= T_t \mathbb{E}\left(\alpha_t \mid Y_t\right) \\
&= T_t a_{t|t}, \qquad \text{(A.14)} \\
P_{t+1} &= \mathrm{Var}\left(\alpha_{t+1} \mid Y_t\right) = \mathrm{Var}\left(T_t \alpha_t + R_t \eta_t \mid Y_t\right) \\
&= T_t \mathrm{Var}\left(\alpha_t \mid Y_t\right) T_t' + R_t Q_t R_t' \\
&\overset{\text{(A.13)}}{=} T_t P_t \left(T_t - K_t Z_t\right)' + R_t Q_t R_t', \qquad \text{(A.15)}
\end{aligned}
$$

where $K_t = T_t P_t Z_t F_t^{-1}$. Note that $a_{t+1} = T_t a_t + K_t v_t$. For this reason, K_t is known as the Kalman gain.

The full recursion can be summarised as

$$
\begin{aligned}
v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t' + H_t, \\
a_{t|t} &= a_t + P_t Z_t' F_t^{-1} v_t, & P_{t|t} &= P_t - P_t Z_t' F_t^{-1} Z_t P_t, \qquad \text{(A.16)} \\
a_{t+1} &= T_t a_{t|t}, & P_{t+1} &= T_t P_t \left(T_t - K_t Z_t\right)' R_t Q_t R_t'.
\end{aligned}
$$

This is what is known as the Kalman Filter recursion. As mentioned previously, (A.16) assumes $c_t = 0$, $d_t = 0$. We can, however, adapt all of the above reasoning to cater for non-zero offsets, yielding

$$
\begin{aligned}
v_t &= y_t - Z_t a_t - d_t, & F_t &= Z_t P_t Z_t' + H_t, \\
a_{t|t} &= a_t + P_t Z_t' F_t^{-1} v_t, & P_{t|t} &= P_t - P_t Z_t' F_t^{-1} Z_t P_t, \qquad \text{(A.17)} \\
a_{t+1} &= T_t a_{t|t} + c_t, & P_{t+1} &= T_t P_t \left(T_t - K_t Z_t\right)' R_t Q_t R_t'.
\end{aligned}
$$

Note If it is assumed that $Z_t = Z, T_t = T, H_t = H, Q_t = Q \ \forall t$, then P_{t+1} will converge to $\bar{P}$ satisfying

$$
\bar{P} = T\bar{P}T' - T\bar{P}Z'\bar{F}^{-1}Z\bar{P}T' + RQR'. \qquad \text{(A.18)}
$$

By finding $\bar{P}$, we can therefore forego the computation of $F_t, K_t, P_{t|t}, P_{t+1}$.

Note The Kalman Filter does not use the data optimally as it only uses past states in each recursion. Another recursion, the Kalman Smoother, also establishes recursions but instead of basing itself only on Y_t at step t, it uses the full dataset Y_T in all steps. We will not derive the recursion here, however Durbin and Koopman (2012) do in Chapter 4, Section 4.4.

A.4.1 The Expectation Maximisation Algorithm

The following method is taken from Durbin and Koopman (2012), Chapter 7, Subsection 7.3.4. and Dehaene and Obozinski (2019).

Up until now, we have assumed that the structure of the state-space model was established, i.e. $Z_t, T_t, c_t, d_t, H_t, Q_t$ were known. However, it is possible that they are not. This is the case in particular in this paper. The Expectation Maximisation Algorithm provides a method to remedy this by iteratively maximising the log-likelihood. It should be noted that the algorithm converges to a *local* maximum of the log-likelihood and therefore good parameter initialisation is essential.

General Derivation of the Expectation-Maximisation Algorithm

Let ψ be the vector of parameters we want to estimate (in the State-Space model scenario it could, for example, be all entries in Z_t, T_t). The aim of the algorithm is to maximise

$$l(\psi) = \log p\left(Y_n \mid \psi\right), \tag{A.19}$$

which can often be a complicated, non-convex optimisation problem.

The following constructions will be used to build the Expectation-Maximisation Algorithm.

Let $\alpha \in \{\alpha_i\}_{i \in I}$ be an unobservable latent factor with distribution $q(\cdot)$ for I a certain set. Then

$$\begin{aligned} l(\psi) &= \log p\left(Y_n \mid \psi\right) \\ &= \log \left[\sum_{\alpha_i} p\left(Y_n, \alpha_i \mid \psi\right)\right] \\ &= \log \left[\sum_{\alpha_i} q\left(\alpha_i\right) \frac{p\left(Y_n, \alpha_i \mid \psi\right)}{q\left(\alpha_i\right)}\right] \\ &= \log \mathbb{E}_q \left[\frac{p\left(Y_n, \alpha \mid \psi\right)}{q(\alpha)}\right] \\ &\geq \mathbb{E}_q \left[\log \frac{p\left(Y_n, \alpha \mid \psi\right)}{q(\alpha)}\right] =: \mathcal{L}(q, \psi), \end{aligned} \tag{A.20}$$

where $\mathbb{E}_q$ denotes the expectation taken with respect to probability distribution $q(\cdot)$ and (A.20) is obtained by applying Jensen's Inequality, as the logarithm is a concave function.

Definition A.5. *The Kullback-Leibler divergence between two probability functions in the same probability space, $p(\cdot)$ and $q(\cdot)$, is defined as*

$$\mathrm{KL}\left(p \parallel q\right) = \int_{-\infty}^{\infty} p(x)\, \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x.$$

The Kullback-Leibler divergence is a measure of how different one probability distribution is from another. It is also sometimes known as the relative entropy. An interesting property that will be used in this context is

$$\mathrm{KL}\left(p \parallel q\right) \geq 0 \text{ with equality if and only if } p = q. \tag{A.21}$$

Note that

$$\mathcal{L}\left(q, \psi\right) = \log\, p\left(Y_n \mid \psi\right) - \mathrm{KL}\left(q \parallel p\left(. \mid Y_n, \psi\right)\right), \tag{A.22}$$

and therefore if $\tilde{q}(\alpha) = p\left(\alpha \mid Y_n, \tilde{\psi}\right)$ for a certain fixed $\tilde{\psi}$, then $\mathcal{L}(\tilde{q}, \psi)$ is a lower bound for $l\left(\psi\right)$ which touches $l\left(\psi\right)$ at $\psi = \tilde{\psi}$.

With these observations in mind, the Expectation-Maximisation Algorithm initialises $\tilde{\psi}$ at ψ_0 chosen by the user before iterating the E- and M-steps (described below) until convergence of $\tilde{\psi}$.

The E-step: Computing the Expectation

Let $\tilde{q}(\alpha) = p\left(\alpha \mid Y_n, \tilde{\psi}\right)$. We then define

$$\tilde{g}(\psi) = \mathcal{L}\left(\tilde{q}, \psi\right). \tag{A.23}$$

The M-Step: Maximising the Expectation

Note that

$$\tilde{g}(\psi) = \mathbb{E}_q\left[\log\, \frac{p\left(Y_n, \alpha \mid \psi\right)}{\tilde{q}\left(\alpha\right)}\right] \tag{A.24}$$

$$= \mathbb{E}_{\tilde{q}}\left[\log\, \tilde{q}\left(Y_n, \alpha \mid \psi\right)\right] - c\left(Y_n, \alpha\right), \tag{A.25}$$

where $c\left(Y_n, \alpha\right)$ does not depend on ψ. Equation (A.25) implies that maximising $\tilde{g}$ is equivalent to finding

$$\tilde{\psi}_{new} = \underset{\psi \in \mathbb{R}}{\operatorname{argmax}}\; \mathbb{E}_{\tilde{q}}\left[\log\, p\left(Y_n, \alpha \mid \psi\right)\right]. \tag{A.26}$$

A graphic representation of the EM algorithm can be found in Figure A.1. We can see that the second step will choose a parameter $\tilde{\psi}_{new}$ very close to the true parameter ψ^*.

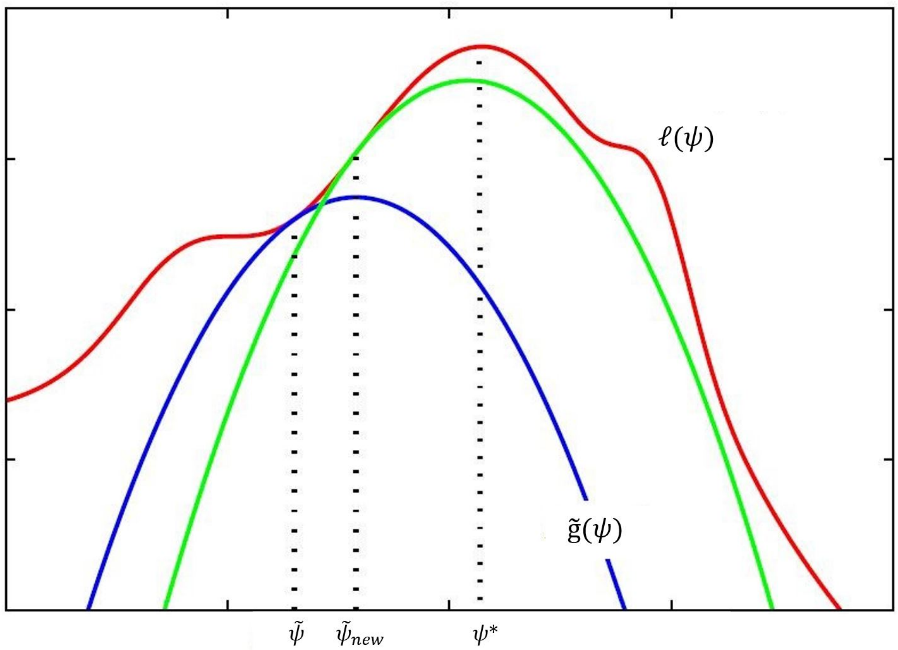


Figure A.1: Image taken from Misra (2019). Two steps of the EM algorithm when ψ (x-axis) is one-dimensional. The red curve is the true log-likelihood and ψ^* is the true parameter. The blue and green curves represent the E-Step of the first and second iterations respectively. The M-Step in the first iteration finds $\tilde{\psi}_{new}$.

It can be shown that

1. the EM algorithm never decreases the likelihood,
2. maxima in $\mathcal{L}$ correspond to maxima in l.

For these reasons, the EM algorithm is a good estimator for the Maximum Likelihood Estimator. It is important, however, to remember that the algorithm converges to a local maximum. Therefore, good parameter initialisation is key.

Derivation of the log-likelihood in State-Space Models

The algorithm derived above is the general EM algorithm. It now remains to adapt it to the State-Space framework by expliciting $l(\psi)$.

Note that

$$
\begin{aligned}
l(\psi) &= \log p\left(Y_n \mid \psi\right) \\
&= \log \left[p(y_1) \prod_{t=2}^{n} \frac{p\left(Y_t \mid \psi\right)}{p(Y_{t-1} \mid \psi)}\right] \\
&= \sum_{t=1}^{n} \log p\left(y_t \mid Y_{t-1}, \psi\right),
\end{aligned}
$$

where recall that $Y_t = \{y_1, \ldots, y_t\}$ and $p\left(y_1 \mid Y_0\right) = p(y_1)$.

If $\alpha_1 \sim \mathcal{N}\left(a_1, P_1\right)$ where a_1 and P_1 are known, then $\mathbb{E}\left(y_t \mid Y_{t-1}\right) = Z_t a_t$ and $\mathrm{Var}\left(y_t \mid Y_{t-1}\right) = F_t$, where a_t and F_t are calculated by the Kalman Filter using the parameter ψ. It then follows that

$$
y_t \mid Y_{t-1} \sim \mathcal{N}\left(Z_t a_t, F_t\right). \tag{A.27}
$$

This implies that

$$
l(\psi) = -\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{n} \left(\log|F_t| + v_t' F_t^{-1} v_t\right),
$$

where recall that n is the number of observations, $v_t = y_t - Z_t a_t$, p is the dimension of y_t and $|F_t|$ denotes the determinant of F_t.

The derivation of the log-likelihood when a_1 and P_1 are unknown can be found in Durbin and Koopman (2012).

Bibliography

BBC News. The Mint countries: Next economic giants? *BBC News*, 2014. https://www.bbc.co.uk/news/magazine-25548060, [Accessed: 2019-12-20].

G. L. Brechtken. The Dynamics of Interest Rates in the Czech Republic, Hungary and Poland: A Vector Autoregressive Latent Yield and Macro Factor Approach. Master's thesis, Universität Wien, 2008.

J. Caldeira, R. Gupta, T. Suleman, and H. S. Torrent. Forecasting the Term Structure of Interest Rates of the BRICS: Evidence from a Nonparametric Functional Data Analysis. Technical report, Department of Economics, University of Pretoria, 2019.

B. Chappatta. The U.S. Yield Curve Just Inverted. That's Huge., 2018. https://www.bloomberg.com/opinion/articles/2018-12-03/u-s-yield-curve-just-inverted-that-s-huge [Accessed: 2019-12-20].

B. Cheung. YIELD CURVE INVERTS: Recession indicator flashes red for first time since 2005, 2019. https://finance.yahoo.com/news/yield-curve-inverts-for-first-time-since-2007-102034083.html [Accessed: 2019-12-20].

G. Dehaene and G. Obozinski. Clustering, K-means, Gaussian mixture model and EM, 2019. *MATH412 - Statistical Machine Learning*, Ecole Polytechnique Fédérale de Lausanne.

F. X. Diebold and C. Li. Forecasting the New Term Structure of Government Bond Yields. *Journal of Econometrics, 130, 337-364.*, 130(2):337–364, 2006.

F. X. Diebold and R. S. Mariano. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263, 1995.

F. X. Diebold and G. D. Rudebusch. *Yield Curve Modeling and Forecasting, the Dynamic Nelson-Siegel Approach.* Princeton University Press, 2013. ISBN EM000007435973.

F. X. Diebold, G. D. Rudebusch, and S. B. Aruoba. The Macroeconomy and the Yield Curve: a Dynamic Latent Factor Approach. *Journal of Econometrics*, 1(131):309–338, 2006.

J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2012. ISBN EM000007117045.

C. Garcia and S. Vanek Smith. Episode 934: Two Yield Curve Indicators, Aug. 2019. https://www.npr.org/2019/08/21/753185863/episode-934-two-yield-curve-indicators [Accessed: 2019-1].

J. Goldstein and K. Romer. Episode 940: Interest Rates... Why So Negative?, 2019. https://www.npr.org/2019/09/20/762748958/episode-940-interest-rates-why-so-negative [Accessed: 2019-11-28].

D. Harvey, S. Leybourne, and P. Newbold. Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting*, 13(2):281–291, 1997.

J. McCulloch. Measuring the Term Structure of Interest Rates. *Journal of Business*, 44(1):19–31, 1971.

R. Misra. Inference using EM algorithm, 2019. https://towardsdatascience.com/inference-using-em-algorithm-d71cccb647bc [Accessed: 2019-12-20].

E. Mönch. Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach. Technical report, Humboldt University Berlin, 2006.

NASA Earth Observatory. Flooding in Nigeria. *NASA Earth Observatory*, 2012. https://earthobservatory.nasa.gov/images/79404/flooding-in-nigeria, [Accessed: 2020-01-08].

C. R. Nelson and A. F. Siegel. Parsimonious Modeling of Yield Curves. *The Journal of Business*, 60(4):473–489, 1987.

J. O'Neill. Jim O'Neill: From BRIC to MINT, 2013. https://www.business-standard.com/article/opinion/jim-o-neill-from-bric-to-mint-113111301201_1.html [Accessed: 2019-12-20].

K. Prasanna and S. Sowmya. Yield Curve in India and its Interactions with the US Bond Market. Technical report, Indian Institute of Technology Madras, Adyar, Chennai, 2016.

S. Srivastava. The US bond yield curve has inverted. Here's what it means, 2019. https://www.cnbc.com/2019/03/25/the-us-bond-yield-curve-has-inverted-heres-what-it-means.html [Accessed: 2019-12-20].

L. E. . Svensson. Estimating and Interpreting Forward Interest Rates. *NPER Working Paper Series*, 1(4871), 1994.

O. A. Vasicek and H. G. Fong. Term Structure Modeling Using Exponential Splines. *Journal of Finance*, 37(2):177–188, 1982.