An Arbitrage-free Dynamic Nelson-Siegel Model

Master Thesis

Part-Time Master in Finance
Goethe Business School
Johann Wolfgang Goethe-Universität Frankfurt

Submitted to:
Asst. Prof. Dr. Christoph Hambel
Tilburg University

Submitted by: Dr. Jens Hesse

Submission Date: 28.03.2024

Abstract

We consider an arbitrage-free dynamic generalized Nelson-Siegel model of interest rates, and investigate its theoretical properties, including the applicability for pricing interest rate derivatives, and its empirical performance when calibrated, using a Kalman filter maximum likelihood estimation, on recent US treasury data.

Contents

List of Figures iv

List of Tables v

List of Symbols vi

List of Abbreviations x

1. Introduction 1

2. The interest rate modeling problem 3

2.1. Motivation . . . 3

2.2. Landscape of models . . . 3

2.2.1. Point-in-time yield curve models . . . 3

2.2.2. Stochastic models . . . 4

2.3. Arbitrage and martingales . . . 5

3. Dynamic Generalized Nelson-Siegel Model 7

3.1. Dynamic Nelson-Siegel Model . . . 7

3.2. Estimation . . . 9

3.3. Arbitrage in the Dynamic Nelson-Siegel Model . . . 11

3.4. Arbitrage-free Dynamic Nelson-Siegel Model . . . 12

3.4.1. Duffie-Kan Framework . . . 12

3.4.2. Model Specification . . . 13

3.5. Generalization . . . 15

3.6. Derivative Pricing . . . 16

3.6.1. Path-independent Derivatives . . . 16

3.6.2. European Zero Coupon Bond Options . . . 19

3.6.3. Pricing by Monte Carlo Simulation . . . 20

4. Empirical evaluation of the various model variants on recent data 21

4.1. Methodology . . . 21

4.1.1. Details regarding the numerical optimization . . . 22

4.2. Data . . . 23

4.3. Results and discussion . . . 23

5. Conclusion 36

Bibliography 37

A. Implementation Details 40

List of Figures

3.1. Plots of $\frac{1-\exp(-\lambda_1 T)}{\lambda_1 T}$ (blue and on the top) and $\frac{1-\exp(-\lambda_1 T)}{\lambda_1 T} - \exp(-\lambda_1 T)$ (orange and on the bottom) as a function of T for the values $\lambda_1 = 0.3, 0.7308, 1.2$ (where $0.7308 = 12 \cdot 0.0609$ is the preferred value of Diebold and Li (2006)): Small values of λ_1 allow for better fit at long maturities (and high values at short maturities). 7

4.1. Summary statistics of the Liu and Wu (2021) yields, and a plot of the yields at three selected tenors through time. 23
4.2. Plot of the function $\lambda \mapsto T_{\max}(\lambda)$ described in the text. 24
4.3. The a priori yield errors produced by the Kalman filter for the optimal parameters with $(p, AF?) = (3, \text{yes})$ on top, and $(p, AF?) = (3, \text{no})$ on bottom, on two different calibration time frames (namely, 85–22 on the left and 15–22 on the right). 29
4.4. The a priori yield errors produced by the Kalman filter for the optimal parameters with $(p, AF?) = (5, \text{yes})$ on top, and $(p, AF?) = (5, \text{no})$ on bottom, on two different calibration time frames (namely, 85–22 on the left and 15–22 on the right). 30
4.5. The estimated state vectors produced by the Kalman filter for the optimal parameters with $(p, AF?) = (5, \text{yes})$ on top, and $(p, AF?) = (5, \text{no})$ on bottom, on two different calibration time frames (namely, 85–22 on the left and 15–22 on the right). 31
4.6. Yield adjustment term $-\frac{C(0,\tau)}{\tau}$ (linearly interpolated) for the optimal parameters with $p = 3$ on top, and $p = 5$ on bottom, on two different calibration time frames (namely, 85–22 on the left and 15–22 on the right). 32

List of Tables

4.1. Overview of the data used for calibration. 24
4.2. Log-likelihood and optimal λ_1, λ_2 values. The precise time frames are as described in table 4.1. The second and third columns indicate the type of model. 25
4.3. A posteriori root mean square errors, in basis points, of yields produced by the Kalman filter on the optimal parameters. 26
4.4. A priori root mean square errors, in basis points, of yields produced by the Kalman filter on the optimal parameters. 27
4.5. A posteriori mean errors, in basis points, of yields produced by the Kalman filter on the optimal parameters. 28
4.6. Estimated $K^{\mathbb{P}}$ values. 33
4.7. Estimated Σ values. 34
4.8. Estimated $\theta^{\mathbb{P}}$ values. 35

List of Symbols

$\cdot^{\mathsf{T}}$	matrix transposition	6, 8–14, 16–20
A	transition matrix in state vector autoregression	8–11, 22
B^{MM}	by $B^{\mathrm{MM}}(t)$ we denote the value of the cash account at time t (the money-market numéraire)	5, 6, 18, 20
B	vector-valued function in Duffie and Kan (1996) price formula equation (3.27); in the special cases we most care about, it contains (generalized) Nelson-Siegel factor loadings as in equations (3.32) and (3.38)	13–20
B	(generalized) Nelson-Siegel coefficient matrix for some given maturities, cf. equation (3.9)	9–11, 40–42
C	scalar-valued function in Duffie and Kan (1996) price formula equation (3.27); in the special cases we most care about, it is the yield adjustment term as in equations (3.33) and (3.39)	iv, 13–20, 32, 40
F	conditional covariance matrix of the one-step prediction error v	10, 11
H	covariance matrix of random noise in yield observation given state vector	9–11
H	Hessian matrix	17, 18
I	by I_n we denote the $n \times n$ identity matrix	9, 22
K	matrix encoding dynamics of state vector evolution in stochastic differential equation	v, 8, 10, 13, 14, 16–19, 21, 22, 26, 33, 41, 42
K	strike price of an option	17, 19
N	number of tenors considered in estimation	9
P	covariance matrix of X	9, 10
P	by $P(t,T)$ we denote the price at time t of a zero-coupon bond maturing at time T	vii, ix, 5, 6, 13, 16–19

Q	covariance matrix of random noise in state vector autoregression	8–11
T_{E}	time of exercise	viii, 17–20
T_{M}	time of maturity	ix, 17–19
V	by $V(t,T)$ we denote the value at time t of a zero-coupon bond maturing at time T discounted to time 0; obtained by discounting $P(t,T)$	5
W	standard Wiener process, sometimes we add a superscript to indicate the measure under which this applies; note that we use $W(t)$ and W_t interchangeably	8, 11–15, 17–19, 21
X	state vector of the interest rate model; note that we use $X(t)$ and X_t interchangeably	vi, ix, 7–21
Φ	cumulative distribution function of the standard normal distribution	19
Σ	matrix encoding volatility of state vector evolution in stochastic differential equation	v, viii, 8, 10, 13–19, 21, 22, 26, 34, 42
β_0	coefficient in the Nelson-Siegel(-Svensson) model: the long run level of interest rates	3, 4
β_1	coefficient in the Nelson-Siegel(-Svensson) model: the short-term component	3, 4
β_2	coefficient in the Nelson-Siegel(-Svensson) model: the medium-term component	3, 4
β_3	coefficient in the Nelson-Siegel-Svensson model: Svensson's second medium-term component	4
$\mathcal{LN}$	log-normal distribution; $X \sim \mathcal{LN}(\mu,\sigma)$ means $\log(X) \sim \mathcal{N}(\mu,\sigma)$	19
$\mathcal{N}$	by $\mathcal{N}(\mu,\Sigma)$ we denote the normal distribution with mean μ and covariance matrix Σ; in the univariate case we specify the standard deviation instead of the covariance matrix	vii, 8–10, 19
λ_1	coefficient in the Nelson-Siegel(-Svensson) model: the (first) decay factor	iv, v, 3, 4, 7, 9–11, 13–15, 21, 24, 25, 41, 42

λ_2	coefficient in the Nelson-Siegel-Svensson model: decay factor in the Svensson term	v, 4, 15, 21, 24, 25, 41, 42
$\mathbb{E}$	expected value; $\mathbb{E}^{\mu}[X] = \int X \,\mathrm{d}\mu$ with μ a probability measure; $\mathbb{E}[\cdots \mid \cdots]$ denotes conditional expectation	viii, 5, 6, 10, 11, 20
$\mathbb{P}$	the real-world (also called physical or natural) measure	v, 5, 8, 21, 22, 26, 33, 35, 36, 41, 42
$\mathbb{Q}$	a risk-neutral measure (*the* risk-neutral measure under condition of a complete market)	5, 6, 8, 13–21
$\mathscr{F}$	the time filtration $\{\mathscr{F}_t\}_t$ of the measure space; in the cases we consider it is always the filtration generated by the Wiener process (which can be higher-dimensional) supplying the randomness	5, 6, 19, 20
∇	gradient (as a row vector(-valued function))	17, 18
ν	a normally distributed stochastic error term	8, 9
Re	real part of a complex number	22
Var	variance; $\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$	9, 10
tr	trace of a matrix	17, 18
vec	by $\mathrm{vec}(A)$ we denote the vector consisting of the stacked columns of the matrix A; put differently, vec is the isomorphism $M_n(\mathbb{R}) \xrightarrow{\sim} \mathbb{R}^{n^2}$ induced by the standard choice of basis	9, 10
$\otimes$	Kronecker product of two matrices	9, 10
σ	an entry of the matrix Σ	15, 34, 41
σ	spectrum of a matrix; the set of all eigenvalues	viii, 22
τ	a maturity	iv, 9, 32
θ	drift term	v, 8–11, 13–16, 21, 26, 35, 42
$\tilde{g}$	by $\tilde{g}(x)$ we denote the pay-off of the derivative in consideration at exercise time T_{E}	17, 18, 20
ε	a normally distributed stochastic error term	9, 11
ϱ	spectral radius of a matrix: $\varrho(A) = \max_{\lambda \in \sigma(A)} \lvert\lambda\rvert$	22
$\mathbb{1}$	indicator function	12

f	by $f(t,x)$ we denote the bond price of the zero coupon bond of maturity T_{M} at time t in state x; see definition in equation (3.43)	8, 17
g	by $g(t,x)$ we denote the value (at time t) of the derivative in consideration at time t in state x	17, 18, 20
h	size of time step in estimation	8–11, 22
m	number of observations minus one	10, 11, 22
p	dimension of the state vector	iv, 21, 22, 24–35, 41, 42
r	the short rate	6, 13–15, 18
v	one-step prediction error	vi, 10, 11
x	expected value of X	9–11
y	zero coupon bond yield (or short: zero yield), $y(t,T) = -\frac{1}{T-t}\log P(t,T)$	3, 4, 7, 9–11, 14, 16
almost surely	true except on a set contained in a set of measure zero	5, 6, 12

List of Abbreviations

AF	arbitrage-free	iv, 25–31, 33–35
AR(1)	autoregressive model of order 1	8
VAR(1)	vector autoregressive model of order 1	8, 41

1. Introduction

Developing models of interest rate and its term structure serves at least two important purposes: (a) they form an important input into policy making (cf. European Central Bank, 2008, box 1), and (b) they are needed to price and manage the risk of interest rate sensitive financial products, like bond options.

Whereas a model developed for the first purpose models the term structure of interest rates, a model developed for the second purpose models the stochastic, and ideally arbitrage-free, evolution of the term structure of interest rates.

One might infer that therefore the first modeling problem is just a special case of the second. In fact however, not having to worry about arbitrage opportunities provides more liberty in solving the first problem. Indeed, one solution, that of Nelson and Siegel (1987), is just to fit a certain parametric functional form, which from empirical evidence is able to capture a large variety of term structures well, to the data. While it is relatively straightforward to take this idea, and interpret the parameters as stochastic processes with some evolution law, there is no reason to assume that such a model will be arbitrage-free (and as we shall see (section 3.3) indeed it is not).

For a long time, this has meant that while there have been models with good performance in empirically matching yield curves, theoretically rigorous arbitrage-free dynamic models tended not have such performance (cf. Christensen, Diebold, and Rudebusch, 2011, p. 4).

Christensen, Diebold, and Rudebusch (2009, 2011) then developed an arbitrage-free dynamic Nelson-Siegel model (and later a generalized version of it), which essentially follows the above-described procedure of making dynamic the Nelson-Siegel model, with the extra twist of adding a certain non-stochastic “correction term” to the Nelson-Siegel functional form.

In this thesis, we review their construction, show how to employ it for derivative pricing, and evaluate its performance on recent data, comparing also to versions of the model without correction term.

The interest rate environment has changed dramatically since Christensen, Diebold, and Rudebusch (2009, 2011) estimated and evaluated their models in the original papers. As The Economist puts it, “[h]igh inflation, not seen in the rich world since the 1980s, is back, which in turn has brought to an end ten years of near-zero interest rates” (The Economist, 2022a). In 2022, “[t]he Federal Reserve has tightened more quickly than at any time since the 1980s, and other central banks have been dragged along behind” (The Economist, 2022b).

We provide some general background in modeling interest rates in chapter 2, before moving on to the family of dynamic generalized Nelson-Siegel models in chapter 3. There we will introduce the model (sections 3.1, 3.4 and 3.5), explain how to estimate the parameters using a Kalman filter maximum likelihood estimation (section 3.2), explain why the "naive" version of the model has arbitrage opportunities (section 3.3), and discuss the pricing of derivatives (section 3.6). chapter 4 forms the empirical part, where we discuss the results of the estimation of the model on recent data. It is supplemented by appendix A, which contains some more technical details. Finally, chapter 5 provides the conclusion.

2. The interest rate modeling problem

2.1. Motivation

Interest rates have a term structure, where yields of different maturities encode information about expected future yields. This relationship is straightforward under the risk-neutral measure, which is characterized by the non-existence of risk premia under it. And though the precise nature of the relationship in the real world is less clear, this still makes the term structure of interest rates an important monetary policy indicator for central banks and others, cf. European Central Bank, 2008, box 1. For this reason, there is an interest in mathematical models that can capture and explain the term structure and in particular allow for meaningful interpolation and extrapolation.

Another angle to look at interest rates stems from their appearance as price factors and risk factors in the valuation and risk measurement of financial instruments. Most strikingly, this concerns bond derivatives like bond options and interest rate derivatives like interest rate swaps and swaptions, but really every financial instrument since there always is the problem of discounting future values to the present using an appropriate interest rate. What is needed here is a model of the future stochastic evolution of interest rates, i.e., of the interest rate curve.

2.2. Landscape of models

Let us give a rough overview of the models typically used to address the twin modeling problem, fit of the initial (current) yield curve and stochastic evolution.

2.2.1. Point-in-time yield curve models

For the first aspect, the seminal work of Nelson and Siegel (1987) reviews the earlier history of this problem and suggests a simple but powerful parametric yield curve model, namely

$$y(T) = y(0,T) = \beta_0 + \beta_1 \left(\frac{1 - \exp(-\lambda_1 T)}{\lambda_1 T} \right) + \beta_2 \left(\frac{1 - \exp(-\lambda_1 T)}{\lambda_1 T} - \exp(-\lambda_1 T) \right), \tag{2.1}$$

with model parameters (which need to be estimated appropriately)

- β_0 the long run level of interest rates,
- β_1 the short-term component,
- β_2 the medium-term component,
- λ_1 the decay factor.

Making the model more interpretable, β_0, β_1 and β_2 can also be interpreted as the level, slope and curvature, respectively, of the yield curve, cf. Diebold and Li, 2006.

The work of Svensson (1995) extends this model by one additional term (with two additional model parameters, β_3 and λ_2) to improve flexibility and fit:

$$\begin{aligned} y(T) = \beta_0 + \beta_1 \left(\frac{1 - \exp(-\lambda_1 T)}{\lambda_1 T} \right) &+ \beta_2 \left(\frac{1 - \exp(-\lambda_1 T)}{\lambda_1 T} - \exp(-\lambda_1 T) \right) \\ &+ \beta_3 \left(\frac{1 - \exp(-\lambda_2 T)}{\lambda_2 T} - \exp(-\lambda_2 T) \right). \end{aligned} \tag{2.2}$$

However, let us briefly remark that these improvements don't come for free: Wahlstrøm, Paraschiv, and Schürle (2021) show that parameter stability suffers compared to the simpler Nelson-Siegel model, mostly due to confounding effects between the curvature factors—with the additional flexibility often being superfluous.

There are other models as well, like the "Super-Bell" model described by Bolder and Stréliski (1999) (that paper also contains a detailed discussion around estimating Nelson-Siegel-Svensson parameters) and non-parametric approaches, in particular linear or spline interpolation. A more comprehensive survey and reasoning by the European Central Bank for why it uses the Svensson model can be found in Nymand-Andersen, 2018. The method of Nelson-Siegel, with or without Svensson's extension, is also popular with other central banks, see BIS Monetary and Economic Department, 2005. In this thesis, we will not consider competing models.

2.2.2. Stochastic models

There is a wide array of models—with different strengths, weaknesses and scopes of application—and accompanying literature for the stochastic evolution of interest rates; see for example Brigo and Mercurio, 2006, and, less daunting in terms of size, Cairns, 2004.

Cheyette (2002) gives a taxonomy: First of all, models vary in the number of random factors (sources of randomness) they incorporate to drive the dynamic: one-factor versus multi-factor.

Then there is a distinction between models with an exogenous term structure and those with an endogenous term structure. In the latter case, the model assumptions derived by equilibrium conditions (see e.g. Vasicek, 1977) or otherwise imply some initial yield curve, which however typically won't align with the actual observed yield curve, see also Cairns, 2004, chapter 5. By contrast, a model with exogenous term structure matches a prescribed initial yield curve.

A third distinction is between short rate models and term structure models. Short rate models model the evolution of the short rate directly, that is, the (annualized) spot rate for an infinitesimally short maturity. All other rates are derived from it when using a short rate model. The alternative is to model the dynamics of the entire term structure directly.

Another important property a model may or may not have is that of being Markovian. This by definition is the case if the future evolution of bond prices only depends on the current values of the stochastic state variables, cf. Cairns, 2004, section 5.2. It is a desirable property, since it makes valuation more tractable, cf. Cheyette, 2002, pp. 8–9.

This thesis deals with Markovian multi-factor term structure models with arbitrage (section 3.1), and with Markovian multi-factor short rate models without arbitrage (section 3.4 and on)—more on arbitrage in the following section.

2.3. Arbitrage and martingales

An absolutely crucial characteristic, the absence of which makes an interest rate model practically useless for many pricing and risk management purposes, is that of being arbitrage-free. That is, going with the security prices implied by the model, it ought to be impossible to construct a risk-free self-financing trading strategy with zero net present value and strictly positive expected future value.

Very closely related to the condition of no-arbitrage is the condition of the existence of an equivalent martingale measure, which for our purposes we can express as follows: the interest rate model is such that there is a measure $\mathbb{Q}$ equivalent to the real-world measure $\mathbb{P}$ such that the discounted (to time 0) bond price processes (the main output of the model) $V(t,T) = \frac{P(t,T)}{B^{\mathrm{MM}}(t)}$ are martingales under $\mathbb{Q}$, i.e., we have the martingale pricing formula

$$V(t,T) = \frac{P(t,T)}{B^{\mathrm{MM}}(t)} = \mathbb{E}^{\mathbb{Q}}\left[V(s,T) \mid \mathscr{F}_t\right] = \mathbb{E}^{\mathbb{Q}}\left[\frac{P(s,T)}{B^{\mathrm{MM}}(s)} \,\middle|\, \mathscr{F}_t\right], \quad t \leq s \leq T, \tag{2.3}$$

almost surely (with respect to $\mathbb{Q}$ or, equivalently, $\mathbb{P}$), where $\mathscr{F}$ is the time filtration.

It is rather straightforward that the existence of an equivalent martingale measure implies no arbitrage, and the First Fundamental Theorem of Asset Pricing states that the converse also is true in a discrete setting (Harrison & Pliska, 1981, theorem 2.7) and "essentially" also in general, though the general formulation (Delbaen & Schachermayer, 1994, theorem 1.1) requires to be careful with some technicalities.

Assuming (as is practically always done) a continuously compounded money market account,

$$\mathrm{d}B^{\mathrm{MM}}(t) = r(t)\, B^{\mathrm{MM}}(t)\,\mathrm{d}t, \quad B^{\mathrm{MM}}(0) = 1, \tag{2.4}$$

where r is $\mathscr{F}$-adapted (which essentially means that $r(t)$ is known at time t) and sample-continuous (i.e., the paths are continuous almost surely), and therefore

$$B^{\mathrm{MM}}(t) = \exp\left(\int_0^t r(s)\mathrm{d}s\right), \tag{2.5}$$

equation (2.3) implies (with $s = T$, and using $P(T,T) = 1$ as well as the $\mathscr{F}_t$-measurability of $B^{\mathrm{MM}}(t)$)

$$\begin{aligned} P(t,T) &= \exp\left(\int_0^t r(s)\,\mathrm{d}s\right) \mathbb{E}^{\mathbb{Q}}\left[P(T,T)\,\exp\left(-\int_0^{\mathsf{T}} r(s)\,\mathrm{d}s\right) \mid \mathscr{F}_t\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^{\mathsf{T}} r(s)\,\mathrm{d}s\right) \mid \mathscr{F}_t\right] \end{aligned} \tag{2.6}$$

almost surely.

Put differently, with the money-market account B^{MM} as above as numéraire, providing an arbitrage-free[1] interest rate model is tantamount to providing an evolution of the short rate.

[1]Strictly speaking, “interest rate model for which an equivalent martingale measure exists”; see the technicalities inherent in the First Fundamental Theorem of Asset Pricing alluded to above.

3. Dynamic Generalized Nelson-Siegel Model

3.1. Dynamic Nelson-Siegel Model

Our object of desire, following Diebold and Li (2006), is a dynamic model of interest rates, usable to for example price derivatives, which has yield curves of the functional form and with the flexibility of the Nelson-Siegel (later also a generalization of Nelson-Siegel-Svensson, see section 3.5) model. That is, we would like to have a (stochastic) process $X = (X_1, X_2, X_3)$ such that

$$\begin{aligned} y(t,T) = X_1(t) + X_2(t) &\left(\frac{1-\exp(-\lambda_1(T-t))}{\lambda_1(T-t)}\right) \\ + X_3(t) &\left(\frac{1-\exp(-\lambda_1(T-t))}{\lambda_1(T-t)} - \exp(-\lambda_1(T-t))\right) \end{aligned} \tag{3.1}$$

(cf. equation (2.1)). Note that the decay λ_1 is left as a constant model parameter. This simplification goes back to Nelson and Siegel (1987, p. 475), and is motivated by Diebold and Li (2006, p. 341); λ_1 mostly determines where on the maturity axis the factor loadings of $X_2(t)$ and $X_3(t)$ make their contributions, cf. figure 3.1. We will return to this point at a later time, cf. figure 4.2.

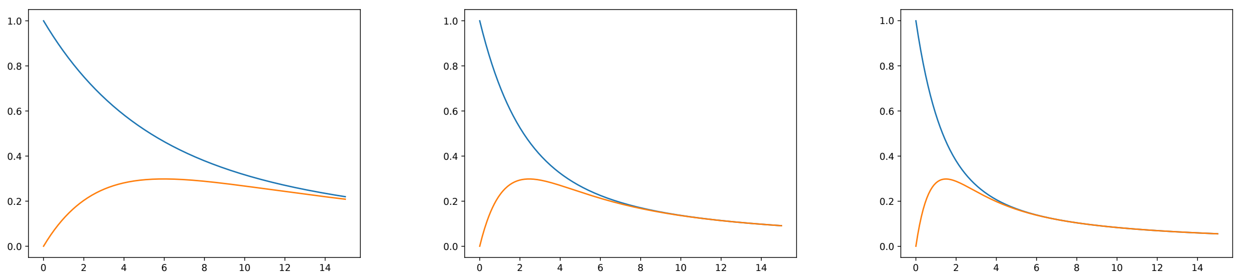

Figure 3.1: Plots of $\frac{1-\exp(-\lambda_1 T)}{\lambda_1 T}$ (blue and on the top) and $\frac{1-\exp(-\lambda_1 T)}{\lambda_1 T} - \exp(-\lambda_1 T)$ (orange and on the bottom) as a function of T for the values $\lambda_1 = 0.3, 0.7308, 1.2$ (where $0.7308 = 12 \cdot 0.0609$ is the preferred value of Diebold and Li (2006)): Small values of λ_1 allow for better fit at long maturities (and high values at short maturities).

As for the evolution of X, Diebold and Li (2006) strictly speaking do not posit a stochastic dynamics, but rather are interested in the h-month ahead forecasting problem. They consider as their preferred approach an AR(1) dynamics,

$$X(t+h) = d + AX(t), \tag{3.2}$$

where d is a vector in three dimensions and A a diagonal 3×3 matrix, and θ and A are obtained by regressing $X(t)$ on $X(t-h)$. Diebold and Li (2006, p. 358) also list and evaluate numerous competing approaches, including a VAR(1) dynamics—i.e., the same as the AR(1) dynamics except that the restriction that A be diagonal is dropped.

This suggests a stochastic dynamics (see also Diebold, Rudebusch, and Aruoba, 2006) of the form

$$X(t) - \theta = A(X(t-h) - \theta) + \nu_t \tag{3.3}$$

in discrete time with drift θ, A a 3×3 matrix, and $\nu_t \sim \mathcal{N}(0, Q)$ for a 3×3 covariance matrix Q. Again, Diebold and Li (2006) would favor (cf. Christensen, Diebold, and Rudebusch, 2011, section 2.4) imposing diagonality constraints on A and Q.

Now consider an Ornstein-Uhlenbeck process

$$\mathrm{d}X(t) = K(\theta - X(t))\mathrm{d}t + \Sigma \mathrm{d}W(t). \tag{3.4}$$

It is well-known that this stochastic differential equation is solved by

$$X(t) = \exp(-hK)X(t-h) + (1 - \exp(-hK))\theta + \int_{t-h}^{t} \exp((s-t)K)\Sigma \mathrm{d}W(s). \tag{3.5}$$

Since $\int_{t-h}^{t} f(s)\mathrm{d}W(s) \sim \mathcal{N}\left(0, \int_{t-h}^{t} f(s)f(s)^\top \mathrm{d}s\right)$ for all square-integrable deterministic f,

$$\int_{t-h}^{t} \exp((s-t)K)\Sigma \mathrm{d}W(s) \sim \mathcal{N}\left(0, \int_{t-h}^{t} \exp((s-t)K)\Sigma\Sigma^\top \exp((s-t)K)^\top \mathrm{d}s\right), \tag{3.6}$$

so that we are exactly in the situation of equation (3.3) with

$$\begin{aligned} A &= \exp(-hK), \\ Q &= \int_{t-h}^{t} \exp((s-t)K)\Sigma\Sigma^\top \exp((s-t)K)^\top \mathrm{d}s = \int_{0}^{h} \exp(-sK)\Sigma\Sigma^\top \exp(-sK)^\top \mathrm{d}s. \end{aligned} \tag{3.7}$$

Equations (3.1) and (3.4) form the dynamic Nelson-Siegel model—though of course in order to fully specify a model, one still has to specify various parameters and also state whether these are supposed to be $\mathbb{P}$- or $\mathbb{Q}$-dynamics.

3.2. Estimation

To discuss parameter estimation, let us denote by $y_t = (y(t,\tau_1),\ldots,y(t,\tau_N))$ the vector of yields observed at time t for some fixed maturities $\tau_1,\ldots,\tau_N$, and, introducing an error term also in the yield observation, cast the model as

$$\begin{aligned} y_t &= BX(t) + \varepsilon_t, \\ X(t) &= (1-A)\theta + AX(t-h) + \nu_t, \end{aligned} \tag{3.8}$$

where

$$B = \begin{pmatrix} 1 & \dfrac{1-\exp(-\lambda_1\tau_1)}{\lambda_1\tau_1} & \dfrac{1-\exp(-\lambda_1\tau_1)}{\lambda_1\tau_1} - \exp(-\lambda_1\tau_1) \\ \vdots & \vdots & \vdots \\ 1 & \dfrac{1-\exp(-\lambda_1\tau_N)}{\lambda_1\tau_N} & \dfrac{1-\exp(-\lambda_1\tau_N)}{\lambda_1\tau_N} - \exp(-\lambda_1\tau_N) \end{pmatrix} \tag{3.9}$$

is a $N\times 3$ matrix, and for the error terms we have

$$\begin{pmatrix} \varepsilon_t \\ \nu_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0_N \\ 0_3 \end{pmatrix}, \begin{pmatrix} H & 0 \\ 0 & Q \end{pmatrix}\right). \tag{3.10}$$

Here $\{X(0),\varepsilon_0,\varepsilon_h,\ldots,\nu_0,\nu_h,\ldots\}$ are assumed to be mutually independent.

This constitutes a linear Gaussian state space model, cf. Durbin and Koopman, 2012, chapter 3.

For estimating this model and subsequent similar models, there is the *two-step approach* used by Diebold and Li (2006): starting with observed yields, first make point-in-time estimations (simple ordinary least squares regressions) $\left(\hat{X}(t)\right)_t$, then another ordinary least squares regression to find the optimal A,θ,Q. The parameter λ_1 is assumed as given and not estimated under this approach.

Alternatively, one can follow a *one-step approach*, as in Diebold, Rudebusch, and Aruoba, 2006, section 2.1 (cf. also De Pooter, 2007, section 4), by using the Kalman filter (cf. Durbin and Koopman, 2012, chapter 4.3): for fixed parameters $\theta, A, H, Q, \lambda_1$ we obtain optimal factor estimates and prediction errors from the Kalman filter and then choose parameters so as to maximize the likelihood (cf. Durbin and Koopman, 2012, chapter 7.2) of the observations.

Let us walk through this in some more detail. Following Diebold, Rudebusch, and Aruoba (2006) and Christensen, Diebold, and Rudebusch (2011), we initialize with the unconditional mean $x(0) = \theta$ and unconditional variance. For the variance $P(0) = \mathrm{Var}[X(t)]$, assuming stationarity of the state vector (ensured if all eigenvalues of A are smaller than 1 in absolute value), we have $P(0) = AP(0)A^{\mathsf{T}} + Q$ from equation (3.8), which we can solve analytically by solving the linear equation

$$(I_{3^2} - A\otimes A)\,\mathrm{vec}(P(0)) = \mathrm{vec}(Q), \tag{3.11}$$

where $\otimes$ is the Kronecker product and vec the embedding that stacks the matrix columns, cf. Durbin and Koopman, 2012, section 5.6.2. In our setting it can also be expressed as

$$P(0) = \int_0^\infty \exp(-sK)\Sigma\Sigma^\top \exp(-sK)^\top \mathrm{d}s. \tag{3.12}$$

So we have $X(0) \sim \mathcal{N}(x(0), P(0))$ for some known $x(0), P(0)$. Define

$$\begin{aligned} x(t \mid t) &= \mathbb{E}[X(t) \mid y_0, \dots, y_t], \\ x(t+h) = x(t+h \mid t) &= \mathbb{E}(X(t+h) \mid y_0, \dots, y_{th}), \\ P(t \mid t) &= \mathrm{Var}[X(t) \mid y_0, \dots, y_{th}], \\ P(t+h) = P(t+h \mid t) &= \mathrm{Var}[X(t+h) \mid y_0, \dots, y_{th}], \\ v_t &= y_t - \mathbb{E}[y_t \mid y_0, \dots, y_{(t-1)h}] = y_t - Bx(t), \\ F_t &= \mathrm{Var}[v_t \mid y_0, \dots, y_{(t-1)h}] = BP(t)B^\top + H \end{aligned} \tag{3.13}$$

Then the *updating step* of the Kalman filter is given by

$$\begin{aligned} x(t \mid t) &= x(t) + P(t)B^\top F_t^{-1} v_t, \\ P(t \mid t) &= P(t) - P(t)B^\top F_t^{-1} BP(t). \end{aligned} \tag{3.14}$$

Referring the reader again to Durbin and Koopman, 2012, chapter 4.3 for more detail on this, let us point out that here we assume F_t to be non-singular, which is unproblematic in practice.

Observe also that while v_t measures the a priori (forecast) error, we can also measure the a posteriori error

$$v(t \mid t) = y_t - Bx(t \mid t). \tag{3.15}$$

The *prediction step* is given by

$$\begin{aligned} x(t+h) &= Ax(t \mid t) + (1-A)\theta, \\ P(t+h) &= AP(t \mid t)A^\top + Q. \end{aligned} \tag{3.16}$$

So, with chosen fixed parameters $\theta, A, H, Q, \lambda_1$ and with $x(0), P(0)$ initialized as described above, equations (3.14) and (3.16) allow us to compute optimal (knowing the information available up to the given time t) parameter estimates. Importantly, we also get a formula for the log-likelihood, given our chosen parameters $\theta, A, H, Q, \lambda_1$, of the observations $y_0, \dots, y_{mh}$ we made. Namely, with p denoting probability density (of multivariate normal distributions in this case), we have

$$\begin{aligned}\log L(\theta, A, H, Q, \lambda_1)(y_0, \ldots, y_{mh}) &= \log p(y_0) + \sum_{t=1}^{m} \log p(y_{th} \mid y_0, \ldots, y_{(t-1)h}) \\ &= -\frac{3(T+1)}{2} \log 2\pi - \frac{1}{2} \sum_{t=0}^{m} \left(\log \det F_{th} + v_{th}^{\top} F_{th}^{-1} v_{th} \right).\end{aligned} \tag{3.17}$$

Later, we will also need to consider the modification of the estimation problem having a deterministic mean adjustment term in the equation defining y_t, i.e., modify the first equation in (3.8) to read

$$y_t = BX(t) + C + \varepsilon_t. \tag{3.18}$$

As explained in Durbin and Koopman (2012, section 4.3.3), the only change in the formulas above that results from this is in the equation defining v_t, which becomes

$$v_t = y_t - \mathbb{E}[y_t \mid y_0, \ldots, y_{(t-1)h}] = y_t - Bx(t) - C. \tag{3.19}$$

3.3. Arbitrage in the Dynamic Nelson-Siegel Model

Filipović (1999) demonstrated that there can be no non-trivial arbitrage-free dynamic Nelson-Siegel model. We briefly recapitulate his argument to give a flavor of it.

He denotes by F the Nelson-Siegel forward rate functional (note that equation (2.1) displays the form of Nelson-Siegel zero rate, not the forward rate)

$$F(x, z) = z_1 + z_2 \exp(-z_4 x) + z_3 x \exp(-z_4 x) \tag{3.20}$$

and then the question is whether there exists a stochastic process $Z = (Z_t)_{0 \le t < \infty}$ such that having the instantaneous forward rate at time t for date $t + \Delta t$ defined to be $F(\Delta t, Z_t)$ yields an arbitrage-free interest rate model in the sense of section 2.3. Assume an Itô process

$$\mathrm{d}Z_t = b_t \mathrm{d}t + \sigma_t \mathrm{d}W_t \tag{3.21}$$

where the Wiener process in d dimensions is specified with respect to the measure under which the discounted bond prices implied by $F(\cdot, Z)$ are martingales. In that case, Filipović (1999, proposition 3.2) shows that the partial differential equation

$$\begin{aligned}\frac{\partial}{\partial x}F(x,Z) = &\sum_{i=1}^{4} b^i \frac{\partial}{\partial z_i}F(x,Z)\\ &+\frac{1}{2}\sum_{i,j=1}^{4} a^{ij}\Big(\frac{\partial^2}{\partial z_i \partial z_j}F(x,Z) - \frac{\partial}{\partial z_i}F(x,Z)\int_0^x \frac{\partial}{\partial z_j}F(\eta,Z)\mathrm{d}\eta\\ &\qquad - \frac{\partial}{\partial z_j}F(x,Z)\int_0^x \frac{\partial}{\partial z_i}F(\eta,Z)\mathrm{d}\eta\Big)\end{aligned} \tag{3.22}$$

is satisfied almost surely, where $\left(a^{ij}\right)_{1\le i,j\le 4}$ is the matrix $\sigma\sigma^{\top}$. From this, Filipović (1999, theorem 4.1) derives that Z must be of the form

$$\begin{aligned} Z_t^1 &= Z_0^1,\\ Z_t^2 &= Z_0^2\exp(-Z_0^4 t) + Z_0^3\, t\exp(-Z_0^4\, t),\\ Z_t^3 &= Z_0^3\exp(-Z_0^4\, t),\\ Z_t^4 &= Z_0^4 + \left(\int_0^t b_s^4 \mathrm{d}s + \sum_{j=1}^{d}\int_0^t \sigma_s^{4j}\mathrm{d}W_s^j\right)\mathbb{1}_{\{Z_0^2=Z_0^3=0\}}, \end{aligned} \tag{3.23}$$

which implies $F(x,Z_t) = F(t+x,Z_0)$, i.e., the model is quasi-deterministic in that all randomness is constrained to time zero.

Finally, Filipović (1999, proposition 5.1) goes on to show that this result also hold if one no longer assumes, as we did at the outset, that Z is specified with respect to the martingale measure, but only with respect to a measure equivalent to it. In other words, there is no non-trivial arbitrage-free dynamic Nelson-Siegel model; “trivial” meaning “quasi-deterministic” here.

3.4. Arbitrage-free Dynamic Nelson-Siegel Model

We have just (in section 3.3) gone through the argument for why there is no non-trivial dynamic arbitrage-free interest rate model that produces Nelson-Siegel type yield curves on the nose. It is possible however, if we relax our requirements in such a way that it suffices for the yield curves to *almost* follow a Nelson-Siegel formula. This is what was achieved by Christensen, Diebold, and Rudebusch (2011) within the framework of Duffie and Kan (1996).

3.4.1. Duffie-Kan Framework

Following Christensen, Diebold, and Rudebusch (2011, section 2.2) without going into all of the technical details, in the framework of Duffie and Kan (1996), the state variable X_t is a Markov process in n-dimensional space solving the stochastic differential equation

$$\mathrm{d}X_t = K^{\mathbb{Q}}(t)\left(\theta^{\mathbb{Q}}(t) - X_t\right)\mathrm{d}t + \Sigma(t)D(X_t, t)\mathrm{d}W_t^{\mathbb{Q}}, \tag{3.24}$$

where (with $\cdot$ denoting the scalar product in n-dimensional space) $D(X_t, t)$ is diagonal of the form

$$\begin{pmatrix} \sqrt{\gamma^1(t) + \delta^1(t) \cdot X_t} & & \\ & \ddots & \\ & & \sqrt{\gamma^n(t) + \delta^n(t) \cdot X_t}, \end{pmatrix} \tag{3.25}$$

and the short rate r is of the form

$$r(t) = \rho_0(t) + \rho_1(t) \cdot X_t. \tag{3.26}$$

This constitutes an affine model (which, as an aside, is not true of the dynamic Nelson-Siegel model(s) previously described; see Diebold, Ji, and Li, 2006, section 2) in the sense that bond prices are of the form

$$P(t, T) = \exp(B(t, T)^{\top} X_t + C(t, T)), \tag{3.27}$$

where $B(t, T)$ and $C(t, T)$ are the solutions of the system of ordinary differential equations

$$\begin{aligned} \frac{\mathrm{d}B(t,T)}{\mathrm{d}t} &= \rho_1(t) + K^{\mathbb{Q}}(t)^{\top} B(t,T) - \frac{1}{2}\sum_{j=1}^{n}\left(\Sigma(t)^{\top} B(t,T) B(t,T)^{\top} \Sigma(t)\right)_{j,j} \delta^j(t), \\ \frac{\mathrm{d}C(t,T)}{\mathrm{d}t} &= \rho_0(t) - B(t,T)^{\top} K^{\mathbb{Q}}(t)\theta^{\mathbb{Q}}(t) - \frac{1}{2}\sum_{j=1}^{n}\left(\Sigma(t)^{\top} B(t,T) B(t,T)^{\top} \Sigma(t)\right)_{j,j} \gamma^j(t), \end{aligned} \tag{3.28}$$

with boundary conditions $B(T, T) = 0$, $C(T, T) = 0$.

From the discussion in section 2.3, it follows that Duffie-Kan models are arbitrage-free (simply by virtue of being based around the short rate).

3.4.2. Model Specification

Christensen, Diebold, and Rudebusch (2011) consider the following particular model specification within the Duffie and Kan (1996) framework:

$$n = 3, \quad K^{\mathbb{Q}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda_1 & -\lambda_1 \\ 0 & 0 & \lambda_1 \end{pmatrix}, \quad \rho_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \rho_0 = 0, \quad \delta^j = 0, \quad \gamma^j = 1 \tag{3.29}$$

for all $j = 1, 2, 3$, with a constant parameter $\lambda_1 > 0$. The mean adjustment term $\theta^{\mathbb{Q}}$ and the volatility Σ are also kept constant through time.

Therefore, the short rate is of the form

$$r(t) = X_1(t) + X_2(t), \tag{3.30}$$

with X obeying the stochastic differential equation

$$\mathrm{d}X(t) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda_1 & -\lambda_1 \\ 0 & 0 & \lambda_1 \end{pmatrix} \left(\theta^{\mathbb{Q}} - X(t)\right) \mathrm{d}t + \Sigma \mathrm{d}W^{\mathbb{Q}}(t). \tag{3.31}$$

The salient point now is the connection to the Nelson-Siegel model (equation (2.1)): Christensen, Diebold, and Rudebusch (2011, proposition 1) show that, given these parameters, the ordinary differential equations (3.28) have the solutions

$$\begin{aligned} B_1(t,T) &= -(T-t), \\ B_2(t,T) &= -\frac{1-\exp(-\lambda_1(T-t))}{\lambda_1}, \\ B_3(t,T) &= (T-t)\exp(-\lambda_1(T-t)) - \frac{1-\exp(-\lambda_1(T-t))}{\lambda_1}, \end{aligned} \tag{3.32}$$

and

$$\begin{aligned} C(t,T) &= \left(K^{\mathbb{Q}}\theta^{\mathbb{Q}}\right)_2 \int_t^T B_2(s,T)\mathrm{d}s + \left(K^{\mathbb{Q}}\theta^{\mathbb{Q}}\right)_3 \int_t^T B_3(s,T)\mathrm{d}s \\ &\quad + \frac{1}{2}\sum_{j=1}^{3}\int_t^T \left(\Sigma^\top B(s,T) B(s,T)^\top \Sigma\right)_{j,j} \mathrm{d}s, \end{aligned} \tag{3.33}$$

leading to zero yields of the form

$$\begin{aligned} y(t,T) &= X_1(t) + \frac{1-\exp(-\lambda_1(T-t))}{\lambda_1(T-t)} X_2(t) \\ &\quad + \left(\frac{1-\exp(-\lambda_1(T-t))}{\lambda_1(T-t)} - \exp(-\lambda_1(T-t))\right) X_3(t) - \frac{C(t,T)}{T-t}, \end{aligned} \tag{3.34}$$

i.e., “Nelson-Siegel with a correction term”, sensibly called *yield adjustment term* by Christensen, Diebold, and Rudebusch (2011), of $-\frac{C(t,T)}{T-t}$.

Christensen, Diebold, and Rudebusch (2009, 2011) choose to fix $\theta^{\mathbb{Q}} = 0$ as an identifying restriction, and we shall follow them in that from now on. This in particular simplifies the expression for $C(t,T)$ in equation (3.33), and Christensen, Diebold, and Rudebusch (2011, section 2.3) also give a somewhat lengthy analytical expression (using only polynomials and exponentials in t, T, λ_1 and the entries of

Σ) for $C(t,T)$. From this expression they can also derive that there is no loss in generality in further assuming that Σ is of the form

$$\Sigma = \begin{pmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}. \tag{3.35}$$

3.5. Generalization

Having outlined the theory behind the arbitrage-free dynamic Nelson-Siegel model, let us now turn to the arbitrage-free dynamic *generalized* Nelson-Siegel model of Christensen, Diebold, and Rudebusch (2009, proposition 3.1). It is a straightforward generalization given by the same authors as in the simpler Nelson-Siegel case:

In that model the short rate is defined by

$$r(t) = X_1(t) + X_2(t) + X_3(t), \tag{3.36}$$

and the state variables $X = (X_1, \dots, X_5)$ follow the stochastic differential equation

$$\mathrm{d}X(t) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & -\lambda_1 & 0 \\ 0 & 0 & \lambda_2 & 0 & -\lambda_2 \\ 0 & 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix} \left(\theta^{\mathbb{Q}} - X(t)\right) \mathrm{d}t + \Sigma \mathrm{d}W^{\mathbb{Q}}(t), \tag{3.37}$$

where $\lambda_1 > \lambda_2 > 0$.

With

$$\begin{aligned}
B_1(t,T) &= -(T-t), \\
B_2(t,T) &= -\frac{1-\exp(-\lambda_1(T-t))}{\lambda_1}, \\
B_3(t,T) &= -\frac{1-\exp(-\lambda_2(T-t))}{\lambda_2}, \\
B_4(t,T) &= (T-t)\exp(-\lambda_1(T-t)) - \frac{1-\exp(-\lambda_1(T-t))}{\lambda_1}, \\
B_5(t,T) &= (T-t)\exp(-\lambda_2(T-t)) - \frac{1-\exp(-\lambda_2(T-t))}{\lambda_2},
\end{aligned} \tag{3.38}$$

and

$$\begin{aligned} C(t,T) = \left(K^{\mathbb{Q}}\theta^{\mathbb{Q}}\right)_2 \int_t^T B_2(s,T)\mathrm{d}s + \left(K^{\mathbb{Q}}\theta^{\mathbb{Q}}\right)_3 \int_t^T B_3(s,T)\mathrm{d}s \\ + \left(K^{\mathbb{Q}}\theta^{\mathbb{Q}}\right)_4 \int_t^T B_4(s,T)\mathrm{d}s + \left(K^{\mathbb{Q}}\theta^{\mathbb{Q}}\right)_5 \int_t^T B_5(s,T)\mathrm{d}s \\ + \frac{1}{2}\sum_{j=1}^{5} \int_t^T \left(\Sigma^\top B(s,T) B(s,T)^\top \Sigma\right)_{j,j} \mathrm{d}s, \end{aligned} \tag{3.39}$$

zero coupon bond prices are then given by

$$\begin{aligned} P(t,T) = \exp(B_1(t,T)X_1(t) + B_2(t,T)X_2(t) + B_3(t,T)X_3(t) \\ + B_4(t,T)X_4(t) + B_5(t,T)X_5(t) + C(t,T)) \end{aligned} \tag{3.40}$$

and yields therefore by

$$\begin{aligned} y(t,T) = X_1(t) - \frac{B_2(t,T)}{T-t}X_2(t) - \frac{B_3(t,T)}{T-t}X_3(t) - \frac{B_4(t,T)}{T-t}X_4(t) - \frac{B_5(t,T)}{T-t}X_5(t) \\ - \frac{C(t,T)}{T-t}, \end{aligned} \tag{3.41}$$

which is "a generalization of Nelson-Siegel-Svensson with a correction term" (cf. equation (2.2)). The remarks concerning assuming $\theta^{\mathbb{Q}} = 0$, the existence of an analytical formula for the yield adjustment term $C(t,T)$ and assuming a triangular shape of Σ hold as in section 3.4.2.

Note that Christensen, Diebold, and Rudebusch (2009) argue that for technical reasons it is not possible to generalize the 3-factor arbitrage-free dynamic Nelson-Siegel model to a 4-factor arbitrage-free dynamic Nelson-Siegel-*Svensson* model; instead one has to pass directly to the 5-factor generalization given above.

In the case of the Nelson-Siegel model, we discussed (in section 3.3) the necessity of a correction term in order to obtain a non-trivial arbitrage-free dynamic version of the model. Regarding the analogous question in the case of Nelson-Siegel-Svensson model, we point to Filipović, 2009, proposition 9.5. Notably, in this case it is *not* an outright non-existence result: a certain specification of the Hull-White extended Vasicek model achieves the goal without correction term.

3.6. Derivative Pricing

3.6.1. Path-independent Derivatives

In this section we consider the pricing of general path-independent derivatives under the arbitrage-free dynamic generalized Nelson-Siegel model, following Duffie and Kan (1994, section 5) and the text book by Baxter and Rennie (1996, section 3.8 (in a slightly different and simpler setting)).

Denote by T_{E} the exercise time. For the derivatives in consideration, the value (at time t) of the derivative is a function $g(t,x)$ of the time $t \leq T_{\mathrm{E}}$ and the state x. In particular, $g(T_{\mathrm{E}},x) =: \tilde{g}(x)$ is the pay-off. For example, for a European call option at strike price K on the zero coupon bond with maturity T_{M}, we have

$$\tilde{g}(x) = g(T_{\mathrm{E}},x) = \max(0, f(T_{\mathrm{E}},x) - K), \tag{3.42}$$

where

$$f(t,x) := f_{T_{\mathrm{M}}}(t,x) := \exp(B(t,T_{\mathrm{M}})^{\mathsf{T}} x + C(t,T_{\mathrm{M}})), \tag{3.43}$$

so that $P(t,T_{\mathrm{M}}) = f(t,X(t))$ by equation (3.40).

Note that one cannot express every derivative in such a way, for example it is not possible for Asian options due to the dependence on the price history.

The basic approach for pricing now is to leverage the martingale representation theorem, which guarantees the existence of a replicating strategy, and Itô's lemma, which allows us to compute the derivative dynamics. Combining the two, we can express the derivative price as the solution to a certain, specific partial differential equation.

To start, we apply Itô's lemma to the stochastic differential equation (3.37) and the zero coupon bond pricing formula (3.40). Thus, we obtain

$$\begin{aligned} \mathrm{d}f(t,X(t)) &= \frac{\partial f}{\partial t}\mathrm{d}t + \nabla_{X(t)} f \mathrm{d}X(t) + \frac{1}{2}(\mathrm{d}X(t))^{\mathsf{T}}(H_{X(t)}f)\mathrm{d}X(t) \\ &= \left(\frac{\partial f}{\partial t} + (\nabla_{X(t)}f)(-K^{\mathbb{Q}}X(t)) + \frac{1}{2}\mathrm{tr}(\Sigma^{\mathsf{T}}(H_{X(t)}f)\Sigma)\right)\mathrm{d}t + (\nabla_{X(t)}f)\Sigma\mathrm{d}W^{\mathbb{Q}}(t). \end{aligned} \tag{3.44}$$

We have

$$\begin{aligned} \nabla f &= f(t,x)\begin{pmatrix} \frac{\partial}{\partial t}B(t,T_{\mathrm{M}})^{\mathsf{T}}x + \frac{\partial}{\partial t}C(t,T_{\mathrm{M}}) \\ B(t,T_{\mathrm{M}}) \end{pmatrix}^{\mathsf{T}}, \\ H_x f &= B(t,T_{\mathrm{M}})\nabla_x f = f(t,x)B(t,T_{\mathrm{M}})B(t,T_{\mathrm{M}})^{\mathsf{T}}, \end{aligned} \tag{3.45}$$

and

$$\begin{aligned} \frac{\partial}{\partial t}B(t,T) &= \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \end{pmatrix}^{\mathsf{T}} + (K^{\mathbb{Q}})^{\mathsf{T}}B(t,T), \\ \frac{\partial}{\partial t}C(t,T) &= -\frac{1}{2}\sum_{j=1}^{5}\left(\Sigma^{\mathsf{T}}B(t,T)B(t,T)^{\mathsf{T}}\Sigma\right)_{j,j}. \end{aligned} \tag{3.46}$$

Plugging equations (3.45) and (3.46) into equation (3.44), we obtain

$$\begin{aligned}
\mathrm{d}P(t,T_{\mathrm{M}}) &= P(t,T_{\mathrm{M}})\Big(\big(\frac{\partial}{\partial t}B(t,T_{\mathrm{M}})^{\top}X(t)+\frac{\partial}{\partial t}C(t,T_{\mathrm{M}})\big)\mathrm{d}t+B(t,T_{\mathrm{M}})^{\top}\mathrm{d}X(t)\\
&\qquad+\frac{1}{2}(\mathrm{d}X(t))^{\top}B(t,T_{\mathrm{M}})B(t,T_{\mathrm{M}})^{\top}\mathrm{d}X(t)\Big)\\
&= \Big(P(t,T_{\mathrm{M}})\left(\begin{pmatrix}1 & 1 & 1 & 0 & 0\end{pmatrix}+B(t,T_{\mathrm{M}})^{\top}K^{\mathbb{Q}}\right)X(t)\\
&\qquad+P(t,T_{\mathrm{M}})\frac{\partial}{\partial t}C(t,T_{\mathrm{M}})+B(t,T_{\mathrm{M}})^{\top}(-K^{\mathbb{Q}}X(t))\\
&\qquad+\frac{1}{2}P(t,T_{\mathrm{M}})\mathrm{tr}(\Sigma^{\top}B(t,T_{\mathrm{M}})B(t,T_{\mathrm{M}})^{\top}\Sigma)\Big)\mathrm{d}t\\
&\quad+P(t,T_{\mathrm{M}})B(t,T_{\mathrm{M}})^{\top}\Sigma\mathrm{d}W^{\mathbb{Q}}(t)\\
&= P(t,T_{\mathrm{M}})\big(r(t)\mathrm{d}t+B(t,T_{\mathrm{M}})^{\top}\Sigma\mathrm{d}W^{\mathbb{Q}}(t)\big).
\end{aligned} \tag{3.47}$$

Next, consider the replicating strategy for g, which exists by the martingale representation theorem:

$$g(t,X(t))=\phi_t P(t,T_{\mathrm{M}})+\psi_t B^{\mathrm{MM}}(t). \tag{3.48}$$

Then, by the self-financing property,

$$\begin{aligned}
\mathrm{d}g(t,X(t)) &= \phi_t\mathrm{d}P(t,T_{\mathrm{M}})+\psi_t\mathrm{d}B^{\mathrm{MM}}(t)\\
&= \phi_t\mathrm{d}P(t,T_{\mathrm{M}})+\psi_t r(t)\,B^{\mathrm{MM}}(t)\mathrm{d}t,
\end{aligned} \tag{3.49}$$

so by comparing the coefficients of $\mathrm{d}W^{\mathbb{Q}}(t)$ and $\mathrm{d}t$ between equation (3.49) (together with equation (3.47)) and equation (3.44) (applied to g instead of f), we obtain $\phi_t P(t,T_{\mathrm{M}})B(t,T_{\mathrm{M}})^{\top}=\nabla_{X(t)}g$ and

$$\begin{aligned}
r(t)g(t,X(t)) &= \phi_t P(t,T_{\mathrm{M}})r(t)+\psi_t r(t)B^{\mathrm{MM}}(t)\\
&= \frac{\partial g}{\partial t}-(\nabla_{X(t)}g)K^{\mathbb{Q}}X(t)+\frac{1}{2}\mathrm{tr}\left(\Sigma^{\top}(H_{X(t)}g)\Sigma\right).
\end{aligned} \tag{3.50}$$

So pricing the derivative means solving the partial differential differential equation (3.50) in the variables $t, x=(X_1(t),\ldots,X_5(t))$ (note that $r(t)$ and $P(t,T_{\mathrm{M}})$ are also functions of t,x) with boundary condition $g(T_{\mathrm{E}},x)=\tilde{g}(x)$.

The article by Duffie and Kan (1994, section 5) also contains some words on numerical solution, and indicates that an analytical solution appears unlikely to be feasible in our setting in the present generality. That being said, an approach of Duffie, Pan, and Singleton (2000, section 1.3) puts the pricing problem in a form where only *ordinary* differential equations and Fourier inversion are involved.

3.6.2. European Zero Coupon Bond Options

For European zero coupon bond options (we will consider call options only for notational simplicity—mutatis mutandis everything also works out for put options), following Jamshidian (1989) and the text book by Brigo and Mercurio (2006, section 2.6), we can be more explicit. We will use the forward measure and the log-normality of bond prices under the model we consider to derive a Black formula for the option price.

As already remarked in equations (3.5) and (3.6), the stochastic differential equation (3.37) is solved by

$$X(T_{\mathrm{E}}) = \exp(-T_{\mathrm{E}}K^{\mathbb{Q}})X(t) + \int_t^{T_{\mathrm{E}}} \exp((s-T_{\mathrm{E}})K^{\mathbb{Q}})\Sigma \mathrm{d}W^{\mathbb{Q}}(s), \tag{3.51}$$

and this implies for the conditional distribution

$$X(T_{\mathrm{E}}) \mid \mathscr{F}_t \sim \mathcal{N}\left(\exp(-T_{\mathrm{E}}K^{\mathbb{Q}})X(t), \int_t^{T_{\mathrm{E}}} \exp((s-T_{\mathrm{E}})K^{\mathbb{Q}})\Sigma\Sigma^{\top} \exp((s-T_{\mathrm{E}})K^{\mathbb{Q}})^{\top} \mathrm{d}s\right) \tag{3.52}$$

under the risk-neutral measure, and by Girsanov's theorem likewise, except with a different mean (which shall not be of interest to us), under the T_{E}-forward measure (i.e., the equivalent martingale measure when using the zero coupon bond with maturity T_{E} as numéraire).

Hence by equation (3.40),

$$\begin{aligned} P(T_{\mathrm{E}},T_{\mathrm{M}}) \mid \mathscr{F}_t \sim \mathcal{LN}\big(&B(T_{\mathrm{E}},T_{\mathrm{M}})^{\top} \exp(-T_{\mathrm{E}}K^{\mathbb{Q}})X(t) + C(T_{\mathrm{E}},T_{\mathrm{M}}), \\ &\int_t^{T_{\mathrm{E}}} \left\|B(T_{\mathrm{E}},T_{\mathrm{M}})^{\top} \exp((s-T_{\mathrm{E}})K^{\mathbb{Q}})\Sigma\right\|_2^2 \mathrm{d}s\big) \end{aligned} \tag{3.53}$$

under the risk-neutral measure, and again we have the same log-normal variance under the T_{E}-forward measure.

Exactly as in the article by Jamshidian (1989), we can now price options using a Black formula: The call option price is

$$c(t,T_{\mathrm{E}},T_{\mathrm{M}}) = P(t,T_{\mathrm{M}})\Phi(d_+) - P(t,T_{\mathrm{E}})K\Phi(d_-), \tag{3.54}$$

where

$$d_{\pm} = \frac{\log(F/K)}{v} \pm \frac{v}{2} \tag{3.55}$$

with

$$v^2 = \int_t^{T_{\mathrm{E}}} \left\|B(T_{\mathrm{E}},T_{\mathrm{M}})^{\top} \exp((s-T_{\mathrm{E}})K^{\mathbb{Q}})\Sigma\right\|_2^2 \mathrm{d}s \quad \text{and} \quad F = \frac{P(t,T_{\mathrm{M}})}{P(t,T_{\mathrm{E}})}. \tag{3.56}$$

Observe that there is a hedging strategy implicit in this formula, namely to hold (at time t) $\Phi(d_+)$ zero coupon bonds of maturity T_{M} (long position) and $-K\Phi(d_-)$ zero coupon bonds of maturity T_{E} (short position).

3.6.3. Pricing by Monte Carlo Simulation

We return to the more general framework of section 3.6.1. Let us briefly remark that equation (3.52) also is very interesting for the Monte Carlo approach to derivative pricing, where one approximates the expected value

$$\frac{g(t, X(t))}{B^{\mathrm{MM}}(t)} = \mathbb{E}^{\mathbb{Q}}\left[\frac{\tilde{g}(X(T_{\mathrm{E}}))}{B^{\mathrm{MM}}(T_{\mathrm{E}})} \,\middle|\, \mathscr{F}_t\right] = \mathbb{E}^{\mathbb{Q}}\left[\tilde{g}(X(T_{\mathrm{E}})) \exp(B(0, T_{\mathrm{E}})^{\top} X(T_{\mathrm{E}}) + C(0, T_{\mathrm{E}})) \,\middle|\, \mathscr{F}_t\right], \quad (3.57)$$

using the law of large numbers, as the average of

$$\left(\tilde{g}(z) \exp(B(0, T_{\mathrm{E}})^{\top} z + C(0, T_{\mathrm{E}}))\right)_{z \in S}, \quad (3.58)$$

where S is a set of (pseudo-)randomly drawn samples of the multivariate normal distribution with location parameters as specified in equation (3.52).

4. Empirical evaluation of the various model variants on recent data

4.1. Methodology

We shall evaluate the empirical performance of four classes of models: (1) dynamic Nelson-Siegel (without yield adjustment term), (2) arbitrage-free dynamic Nelson-Siegel, (3) generalized dynamic Nelson-Siegel (without yield adjustment term), and (4) arbitrage-free generalized dynamic Nelson-Siegel.

Let us denote by p the dimension of the state space, i.e., $p = 3$ in cases 1 and 2, and $p = 5$ in cases 3 and 4.

Observe that the dynamics for the arbitrage-free models were specified with respect to the risk-neutral measure $\mathbb{Q}$ in equations (3.31) and (3.37), and for the models having arbitrage we only gave a very general specification in equation (3.4).

However, following Christensen, Diebold, and Rudebusch (2009, 2011), the plan is to estimate the models on historical time series of zero yields at various tenors, i.e., observations under the *physical* measure $\mathbb{P}$. Again following Christensen, Diebold, and Rudebusch (2009, 2011), we assume that we have an essentially affine risk premium resulting in affine dynamics also under $\mathbb{P}$. Thus,

$$\begin{aligned} \mathrm{d}X(t) &= K^{\mathbb{Q}}\left(\theta^{\mathbb{Q}} - X(t)\right)\mathrm{d}t + \Sigma \mathrm{d}W^{\mathbb{Q}}(t), \\ \mathrm{d}X(t) &= K^{\mathbb{P}}\left(\theta^{\mathbb{P}} - X(t)\right)\mathrm{d}t + \Sigma \mathrm{d}W^{\mathbb{P}}(t) \end{aligned} \tag{4.1}$$

(note that by Girsanov's theorem we don't need to put a superscript indicating the measure on Σ—it's independent of the measure). In the models without yield adjustment term (and with arbitrage), we will only estimate the $\mathbb{P}$-dynamics. In the arbitrage-free models, we made the identifying restriction $\theta^{\mathbb{Q}} = 0$ (cf. section 3.4.2), and with that we can estimate all parameters for the $\mathbb{Q}$-dynamics from the $\mathbb{P}$-dynamics: recall that $K^{\mathbb{Q}}$ depends only on λ_1, λ_2, and these parameters also figure in the formulas for the zero yield, and therefore are within the scope of estimation under $\mathbb{P}$.

In practical terms, we now have the problem that $K^{\mathbb{P}}$, $\theta^{\mathbb{P}}$ might be anything and for Σ we also so far only have the simplification that we may assume it to be lower triangular (cf. equation (3.35)). This makes parameter optimization expensive, potentially numerically unstable, and the results hard to interpret. Therefore, again following Christensen, Diebold, and Rudebusch (2009) we further impose that $K^{\mathbb{P}}$ and Σ be diagonal (assumption of independent factors under $\mathbb{P}$).

We will use the Kalman filter to estimate the models, as described in section 3.2. Note that that discussion immediately generalizes from the dynamic Nelson-Siegel case to all the other cases; in particular we explained how to incorporate the yield adjustment term in equations (3.18) and (3.19), and generalizing from $p = 3$ to $p = 5$ is straightforward.

4.1.1. Details regarding the numerical optimization

To implement the optimization (maximization of the loglikelihood (3.17)), we use Python code, and in particular the L-BFGS-B numerical optimization method (limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm with bounds) as implemented in SciPy (Virtanen et al., 2020).

Two details regarding the optimization procedure appear worthwhile to discuss at this point; we go into some further technical details in appendix A.

Firstly, like De Pooter (2007, appendix A), we exclude the first few ($\min(10, \lfloor 5\%(m+1) \rfloor)$, to be precise) observations in the calculation of the loglikelihood (3.17); this is in order to allow the Kalman filter some time to "attune to the data" without penalty.

Secondly, as mentioned in section 3.2, to ensure stationarity of the state vector under $\mathbb{P}$, we put in place the constraint that all eigenvalues of $A = \exp(-hK^{\mathbb{P}})$ be smaller than 1 in absolute value. Note that for the spectra we have $\sigma(\exp(-hK^{\mathbb{P}})) = \exp(-h\sigma(K^{\mathbb{P}}))$, and $\exp(\{z \in \mathbb{C} \mid \mathrm{Re}(z) < 0\}) = \{z \in \mathbb{C} \mid |z| < 1\}$, so that we can also phrase the condition differently: all the eigenvalues of $K^{\mathbb{P}}$ need to have strictly positive real part. Denoting by ϱ the spectral radius function, we can make this constraint amenable for numerical optimization by observing that

$$\xi \colon M \mapsto M + \log(1 + \exp(-\varrho(M)))I \tag{4.2}$$

establishes a smooth diffeomorphism between the manifold of all square matrices of a given size and those where all eigenvalues have strictly positive real part. Consequently, instead of

$$\begin{gathered} \max \text{objective}(\dots, K^{\mathbb{P}}, \dots) \\ \text{such that all the eigenvalues of } K^{\mathbb{P}} \text{ have positive real part} \end{gathered} \tag{4.3}$$

we can equivalently solve

$$\max \text{objective}(\dots, \xi(\tilde{K}), \dots) \tag{4.4}$$

where there is no constraint on $\tilde{K}$ (see also Kim and Nelson, 1999, p. 15).

4.2. Data

The input we require is a time series of zero coupon yields at a grid of certain fixed maturities. The maturities we used are 1, 2, 3, 5, 6, 8, 10, 12, 15, 20, 25 years. Since there usually is no bond with remaining maturity of exactly, say, 5 years on a given date, these yields need to synthesized from market observable data. As described in European Central Bank (2008), the European Central Bank for example fits a Nelson-Siegel-Svensson model every day and publishes the yields and parameters. In principle we could use this data, but it is already making too many model assumptions; it appears unwise to use input data that (if only point-in-time) follows a Nelson-Siegel-Svensson model exactly in order to evaluate dynamic versions of such a model. What we need are raw yield curves with few model assumptions. There is classical work on this by, among others, Fama and Bliss (1987) and Gürkaynak, Sack, and Wright (2007), and more recently Liu and Wu (2021).

We use the data (and therefore the method) of Liu and Wu (2021), which is available from https://sites.google.com/view/jingcynthiawu/yield-data and partially shown in figure 4.1.

	count	mean	std	min	25%	50%	75%	max
tenor								
1	446	0.034	0.026	0.001	0.007	0.033	0.055	0.096
2	446	0.036	0.026	0.001	0.009	0.037	0.058	0.096
3	446	0.038	0.026	0.001	0.014	0.039	0.059	0.095
5	446	0.042	0.025	0.002	0.018	0.042	0.062	0.093
6	446	0.044	0.024	0.003	0.021	0.042	0.064	0.095
8	446	0.046	0.024	0.004	0.024	0.044	0.065	0.095
10	446	0.048	0.023	0.005	0.027	0.046	0.066	0.097
12	446	0.050	0.023	0.006	0.029	0.048	0.067	0.099
15	446	0.052	0.023	0.007	0.030	0.051	0.069	0.102
20	446	0.053	0.022	0.010	0.032	0.052	0.071	0.102
25	446	0.053	0.021	0.013	0.033	0.052	0.071	0.098

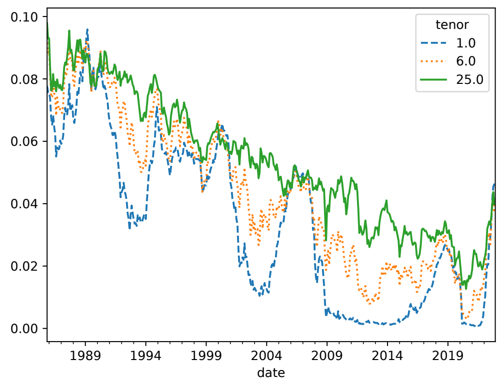

Figure 4.1: Summary statistics of the Liu and Wu (2021) yields, and a plot of the yields at three selected tenors through time.

We run the calibration on different time frames, as shown in table 4.1.

4.3. Results and discussion

We obtain the parameter estimates given in tables 4.2 and 4.6 to 4.8, and root mean square errors for the yields at the various tenors given in tables 4.3 and 4.4. Observe that the order of magnitude of the (a posteriori) errors, of around 10 basis points in aggregate and up to 45 basis points on individual tenors, is in line with the results of Christensen, Diebold, and Rudebusch (2009, 2011). The a priori

data source	frequency	time frame start	time frame end
Liu and Wu (2021)	monthly	2015-02-01	2019-12-01
Liu and Wu (2021)	monthly	2015-02-01	2022-12-01
Liu and Wu (2021)	monthly	2020-02-01	2022-12-01
Liu and Wu (2021)	monthly	2021-02-01	2022-12-01
Liu and Wu (2021)	monthly	1985-11-01	2006-12-01
Liu and Wu (2021)	monthly	1985-11-01	2008-12-01
Liu and Wu (2021)	monthly	1985-11-01	2009-12-01
Liu and Wu (2021)	monthly	1985-11-01	2019-12-01
Liu and Wu (2021)	monthly	1985-11-01	2022-12-01

Table 4.1.: Overview of the data used for calibration.

error estimates of table 4.4 can be meaningfully compared to the one-month ahead forecasting errors of Christensen, Diebold, and Rudebusch (2011, Table 7), and they are in line as well in terms of the magnitude of around 30–50 basis points. Table 4.5, which displays the mean errors, indicates that the models with only $p = 3$ factors tend to make systematic errors at certain maturities, in particular at the long end of the curve, to attain overall fit. Again, Christensen, Diebold, and Rudebusch (2011, Table 3) reported similar results.

In terms of parameter estimates, to get some feeling for which parts of the final curve are influenced by the various factors, let us consider for which $T = T_{\max}(\lambda)$ the functions $T \mapsto \frac{1-\exp(-\lambda T)}{\lambda T} - \exp(-\lambda T)$, where λ is either λ_1 or λ_2 achieve their maximum value (cf. figure 3.1). This point on the curve, as a function of λ, is shown in figure 4.2. In particular, for the full time frame of 1985–2022, for $p = 3$, the

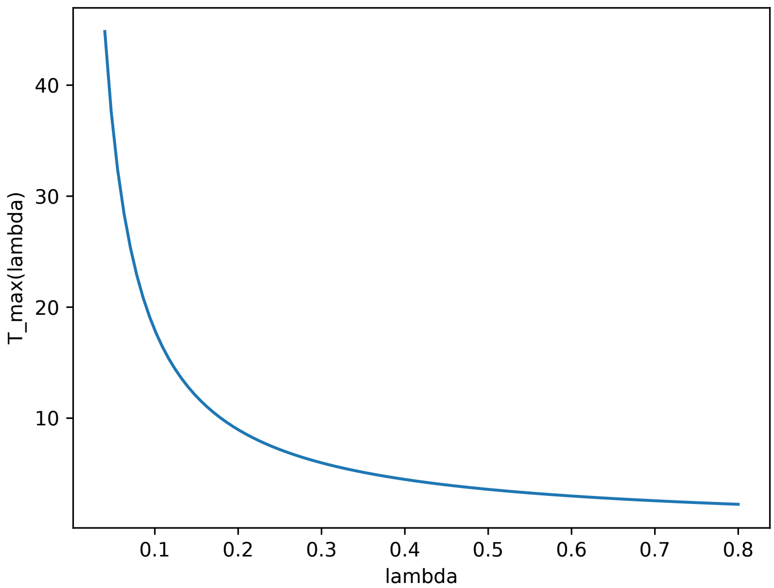

Figure 4.2: Plot of the function $\lambda \mapsto T_{\max}(\lambda)$ described in the text.

maximum is attained at around 4 years in the version without yield adjustment term, and at around

time frame	p	AF?	log-likelihood	λ_1	λ_2
85–06	3	no	16831.1	0.3139	—
85–06	3	yes	16651.7	0.2599	—
85–06	5	no	16861.7	0.5643	0.0882
85–06	5	yes	17110.0	0.7037	0.1032
85–09	3	no	18586.0	0.4239	—
85–09	3	yes	18453.8	0.1835	—
85–09	5	no	18786.2	0.5582	0.1033
85–09	5	yes	18783.4	0.4604	0.0107
85–19	3	no	26767.3	0.4387	—
85–19	3	yes	26287.1	0.2130	—
85–19	5	no	27114.4	0.6595	0.1036
85–19	5	yes	27165.7	0.7112	0.1210
15–19	3	no	4298.3	0.3601	—
15–19	3	yes	4251.9	0.3053	—
15–19	5	no	4382.5	0.3617	0.0936
15–19	5	yes	4384.8	0.3536	0.0897
85–22	3	no	29061.6	0.4353	—
85–22	3	yes	28576.7	0.1740	—
85–22	5	no	29538.2	0.6895	0.1090
85–22	5	yes	29608.2	0.7098	0.1175
15–22	3	no	6483.1	0.5559	—
15–22	3	yes	6358.4	0.4873	—
15–22	5	no	6623.4	0.5796	0.1119
15–22	5	yes	6621.6	0.5822	0.0528
20–22	3	no	2269.2	0.1372	—
20–22	3	yes	2211.7	0.1476	—
20–22	5	no	2399.4	0.5564	0.2075
20–22	5	yes	2393.6	0.6074	0.0544
21–22	3	no	1448.7	0.1763	—
21–22	3	yes	1487.1	0.1246	—
21–22	5	no	1541.2	0.6473	0.2036
21–22	5	yes	1537.2	0.5581	0.1941

Table 4.2.: Log-likelihood and optimal λ_1, λ_2 values. The precise time frames are as described in table 4.1. The second and third columns indicate the type of model.

10 years for the version with yield adjustment term. For the same time frame and $p = 5$, it is almost the same whether the yield adjustment term is included or not, at around 2.5 years for λ_1 and 15–16.5 years for λ_2. Once more confirming Christensen, Diebold, and Rudebusch (2009), this shows that the factors associated with λ_1 and λ_2, respectively, control quite different parts of the yield curve being output. We observe the same behavior also on all of the other calibration time frames.

We also note that λ_1, λ_2 vary materially as the time frame varies, in particular for the $p = 3$ models—for example, we get $\lambda_1 \approx 0.14$ in both the models with and without yield adjustment term for the 2020–2022 time frame, but get around 0.49 and 0.56, respectively, for the 2015–2022 time frame. Also, it is *not* the case for every time frame that the λ_1, λ_2 are approximately independent of whether the yield adjustment term is included or not. It is not clear however, whether these observations should be interpreted as evidence for the appropriateness of modeling time-dependent λ_1, λ_2, or whether they should be taken to indicate that some of the calibration time frames are just too short too obtain stable results. Similar remarks apply to $K^{\mathbb{P}}$, Σ, and $\theta^{\mathbb{P}}$.

time frame	p	AF?	1	2	3	5	10	15	20	25	total
85–06	3	no	14.24	0.00	4.41	3.27	4.12	10.52	8.89	19.79	8.88
85–06	3	yes	17.22	2.98	3.55	3.30	5.73	10.34	8.14	13.22	8.30
85–06	5	no	5.05	2.57	1.52	2.26	4.49	8.15	6.27	2.04	4.23
85–06	5	yes	0.87	2.97	1.30	2.29	4.04	6.92	4.31	1.54	3.37
85–09	3	no	13.00	0.00	4.45	2.72	3.84	13.91	21.48	33.97	13.71
85–09	3	yes	22.38	2.76	6.46	9.38	6.75	9.50	1.81	10.26	9.70
85–09	5	no	3.90	2.72	3.04	3.88	5.44	6.73	4.42	3.53	4.34
85–09	5	yes	11.35	0.14	4.04	2.51	4.66	8.58	3.92	6.65	5.74
85–19	3	no	13.37	0.31	4.24	2.27	3.57	11.60	18.92	30.80	12.41
85–19	3	yes	20.55	2.80	5.70	7.14	7.22	8.95	4.73	14.59	9.57
85–19	5	no	2.08	2.68	1.98	3.67	4.47	6.67	5.78	2.58	3.96
85–19	5	yes	1.68	2.54	2.16	3.33	3.86	7.26	6.67	3.08	4.09
15–19	3	no	9.08	0.00	3.06	0.57	1.29	2.11	7.97	22.87	7.92
15–19	3	yes	8.66	0.00	2.87	0.78	1.34	1.94	11.01	30.87	10.31
15–19	5	no	8.20	0.27	2.36	0.23	1.46	1.60	2.91	4.93	3.22
15–19	5	yes	8.57	0.04	2.38	0.08	1.47	1.70	3.13	4.64	3.30
85–22	3	no	13.27	0.90	4.18	2.08	4.18	10.34	19.35	30.14	12.24
85–22	3	yes	23.15	4.18	5.60	8.18	6.26	10.49	4.50	12.80	10.11
85–22	5	no	2.11	2.44	2.19	3.23	3.62	7.73	8.21	2.58	4.31
85–22	5	yes	1.62	2.54	2.14	3.16	3.28	7.73	8.05	2.62	4.26
15–22	3	no	11.29	0.02	2.77	0.01	1.85	4.87	25.08	32.45	12.99
15–22	3	yes	11.52	0.00	2.85	0.00	2.07	6.27	31.26	47.39	17.63
15–22	5	no	11.77	0.13	2.12	0.61	0.76	2.59	12.27	1.76	5.33
15–22	5	yes	11.42	0.17	2.29	0.52	0.74	2.57	12.46	2.40	5.33
20–22	3	no	36.58	10.92	1.54	6.13	0.02	5.88	16.65	3.60	13.25
20–22	3	yes	36.79	10.78	2.61	7.37	0.51	8.72	15.05	4.07	13.54
20–22	5	no	15.20	0.52	2.05	2.08	0.63	1.56	16.90	6.46	7.27
20–22	5	yes	12.60	0.04	2.15	1.42	0.09	2.44	19.22	4.14	7.22
21–22	3	no	44.17	13.13	0.02	6.11	0.00	1.61	29.48	20.10	17.92
21–22	3	yes	41.38	14.59	3.60	3.31	0.00	1.55	19.99	13.57	15.24
21–22	5	no	11.10	2.23	1.03	1.75	0.00	2.84	21.31	3.20	7.53
21–22	5	yes	20.15	0.50	2.53	2.52	1.46	1.11	18.82	5.85	8.66

Table 4.3.: A posteriori root mean square errors, in basis points, of yields produced by the Kalman filter on the optimal parameters.

Comparing the models' errors and loglikelihoods, we can observe that, broadly speaking, having more degrees of freedom, i.e., $p = 5$ rather than 3, improves model performance noticeably, and that the effect of including or not including the yield adjustment term is smaller and less consistent. For example, on the 1985–2006 time frame, for $p = 3$ the root mean square errors for the versions with and without arbitrage looked approximately the same, while on the 1985–2009 time frame (including the Great Financial Crisis), the model without yield adjustment term performs materially worse. For $p = 5$, the model with yield adjustment term was a bit better before, and a bit worse than the competitor model without yield adjustment term after inclusion of the Great Financial Crisis. All of this could be taken to be an argument in favor of including the yield adjustment term, since it has the advantage of

time frame	p	AF?	1	2	3	5	10	15	20	25	total
85–06	3	no	27.29	30.08	31.20	30.62	28.27	28.69	26.34	30.38	29.12
85–06	3	yes	31.22	33.32	33.32	32.22	30.41	30.44	27.74	28.46	30.93
85–06	5	no	28.67	33.66	33.46	32.31	29.92	29.34	26.72	25.12	30.20
85–06	5	yes	29.71	33.82	33.39	32.27	29.88	29.07	26.43	25.10	30.24
85–09	3	no	26.68	31.09	32.27	31.58	28.96	32.20	34.99	42.85	32.09
85–09	3	yes	31.61	34.56	35.92	35.49	30.11	30.40	27.95	29.30	31.96
85–09	5	no	31.26	35.28	34.82	33.82	30.48	30.38	28.27	27.05	31.51
85–09	5	yes	29.51	34.07	34.36	33.47	30.21	30.74	28.13	27.41	31.14
85–19	3	no	27.39	32.36	33.85	33.05	30.61	34.46	35.80	39.06	32.99
85–19	3	yes	26.77	32.51	35.31	34.77	30.25	32.56	30.73	29.47	31.70
85–19	5	no	30.09	32.93	33.51	33.54	30.36	32.07	31.10	27.41	31.49
85–19	5	yes	30.11	32.87	33.53	33.52	30.31	32.41	31.58	27.58	31.57
15–19	3	no	14.64	14.92	16.87	18.32	18.55	18.41	18.99	27.94	18.89
15–19	3	yes	11.67	14.44	16.90	18.29	18.64	18.74	21.17	35.22	20.00
15–19	5	no	11.39	14.35	16.54	17.96	18.43	18.33	17.61	16.89	17.12
15–19	5	yes	11.56	14.32	16.53	18.03	18.51	18.44	17.66	17.07	17.20
85–22	3	no	30.08	36.44	38.49	37.61	32.91	35.87	37.19	39.16	35.71
85–22	3	yes	28.26	32.54	35.41	35.16	30.71	33.72	30.75	28.55	32.09
85–22	5	no	30.45	32.93	33.59	33.56	30.74	33.20	31.99	27.73	31.85
85–22	5	yes	30.63	33.02	33.70	33.68	30.74	33.33	32.13	27.95	31.97
15–22	3	no	24.06	24.56	25.54	24.87	23.34	24.03	35.27	38.91	27.06
15–22	3	yes	22.60	24.33	25.62	24.86	23.30	24.56	40.23	52.40	29.54
15–22	5	no	24.68	23.55	24.39	24.09	22.92	23.15	26.96	20.54	23.74
15–22	5	yes	25.04	23.65	24.41	24.23	23.06	23.16	27.16	20.58	23.87
20–22	3	no	26.06	26.12	31.03	31.37	26.60	26.99	33.12	22.86	28.30
20–22	3	yes	28.11	27.57	32.13	32.18	26.58	27.71	32.61	24.01	29.01
20–22	5	no	33.11	30.57	30.84	29.13	26.70	26.29	32.65	21.87	28.67
20–22	5	yes	33.89	30.06	30.16	29.05	26.24	26.07	34.17	22.44	28.72
21–22	3	no	34.01	28.25	33.48	34.03	30.37	29.72	42.37	31.87	32.87
21–22	3	yes	51.77	65.74	70.92	66.93	55.72	50.46	54.47	40.07	58.15
21–22	5	no	30.12	30.80	32.79	31.44	28.46	27.01	33.63	22.86	29.72
21–22	5	yes	34.55	33.54	35.31	33.30	30.59	29.61	35.59	25.23	32.11

Table 4.4.: A priori root mean square errors, in basis points, of yields produced by the Kalman filter on the optimal parameters.

making the model arbitrage-free, which clearly is desirable if it doesn't harm the fit materially.

In figures 4.3 and 4.4 we plot the errors through time that we obtain calibrating (a) on the whole available time frame, and (b) on the time frame starting 2015, respectively.

In all cases with time frame (a) we can see that the models produced the biggest errors in the Great Financial Crisis around 2009, though the most recent observations also create errors at least coming close. Also on the shorter time frame (b), which does not contain the financial crisis, the errors increase towards the end of the available period. Nevertheless, the models do not struggle extraordinarily with the most recent observations. Also, perhaps with the exclusion of the very short 21–22 window, the parameter estimation and the models themselves still appear to produce sane results when calibrated only on shorter time frames; there are no parameter or error blow-ups.

In figure 4.5, we show the state vector estimates coming out of the Kalman filter for the optimal

time frame	p	AF?	1	2	3	5	10	15	20	25	total
85–06	3	no	-5.16	-0.00	1.83	1.13	-2.39	7.25	2.78	-12.79	-0.70
85–06	3	yes	-9.16	-1.41	1.75	1.52	-4.05	5.76	5.06	-3.45	-0.64
85–06	5	no	2.12	-0.71	-0.36	0.92	-3.08	4.35	4.32	-0.63	0.52
85–06	5	yes	0.06	-0.09	-0.12	0.30	-1.90	2.57	-0.45	-0.03	0.01
85–09	3	no	-4.78	-0.00	1.80	0.64	-1.38	6.88	0.93	-15.43	-0.93
85–09	3	yes	-8.22	-1.01	2.88	3.62	-2.76	1.93	0.45	-2.66	-0.39
85–09	5	no	1.51	-1.24	-0.10	1.60	-2.97	1.80	1.78	-1.11	0.09
85–09	5	yes	1.01	-0.02	0.18	0.33	-1.25	2.49	-0.73	0.38	0.31
85–19	3	no	-3.06	-0.01	0.90	0.41	-1.20	4.52	-0.43	-10.53	-0.80
85–19	3	yes	-2.33	-0.57	0.78	1.71	-2.04	1.18	0.14	1.36	0.06
85–19	5	no	0.50	-0.64	-0.05	1.09	-1.94	2.08	1.09	-0.46	0.17
85–19	5	yes	-0.08	0.19	-0.18	0.08	-0.75	0.96	-2.13	0.35	-0.05
15–19	3	no	4.33	0.00	-2.33	-0.26	-0.08	-0.43	6.61	21.86	2.93
15–19	3	yes	3.68	0.00	-2.10	-0.16	-0.52	0.25	10.26	30.19	3.95
15–19	5	no	-0.98	0.09	-0.86	-0.05	-0.63	-0.21	-0.06	-0.00	-0.15
15–19	5	yes	-1.30	0.01	-0.84	-0.02	-0.64	-0.23	-0.06	0.01	-0.18
85–22	3	no	-3.08	-0.03	0.74	0.30	-1.47	4.03	1.47	-7.73	-0.53
85–22	3	yes	-3.06	-0.89	0.64	2.05	-1.80	0.33	0.32	0.68	-0.04
85–22	5	no	0.41	-0.44	-0.12	0.79	-1.69	2.36	2.50	-0.54	0.34
85–22	5	yes	-0.14	0.38	-0.30	-0.15	-0.60	0.25	-1.90	0.20	-0.11
15–22	3	no	-2.99	0.01	-0.68	-0.00	-0.69	3.00	19.01	30.20	4.51
15–22	3	yes	-3.04	0.00	-0.56	-0.00	-1.21	4.81	26.37	45.83	6.66
15–22	5	no	-7.54	0.08	-0.07	-0.35	-0.07	-0.33	4.24	-0.00	-0.23
15–22	5	yes	-6.11	0.10	-0.26	-0.27	-0.12	-0.00	4.98	-0.16	-0.01
20–22	3	no	-7.82	-2.33	-0.85	3.06	0.00	-4.44	11.32	1.94	0.68
20–22	3	yes	-13.72	-4.78	-1.00	4.96	-0.08	-7.74	7.64	1.45	-0.45
20–22	5	no	-11.70	0.30	-0.09	-1.70	0.27	0.50	13.55	-1.11	0.09
20–22	5	yes	-7.48	0.03	-0.57	-0.79	0.01	1.27	15.70	-0.71	0.87
21–22	3	no	-17.60	-4.76	-0.01	4.89	-0.00	0.63	27.17	17.22	3.29
21–22	3	yes	10.92	5.62	0.03	-1.90	0.00	0.35	17.23	-1.89	2.79
21–22	5	no	-6.63	1.39	-0.10	-1.09	0.00	0.46	19.27	-0.05	1.30
21–22	5	yes	-16.14	0.23	0.19	-2.20	0.65	0.16	16.16	-0.41	-0.10

Table 4.5.: A posteriori mean errors, in basis points, of yields produced by the Kalman filter on the optimal parameters.

parameters for $p = 5$. Again, the Great Financial Crisis is clearly visible in the data as period where the state vector quickly undergoes large movements. We also can confirm the observation from Christensen, Diebold, and Rudebusch (2009) that the level factor is the most persistent factor, whereas the slope and in particular curvature factors exhibit a higher degree of volatility. Interestingly, the level factor is much higher in the models with yield adjustment term, which makes sense, because that adjustment is a down adjustment of the yields, as can be seen in figure 4.6.

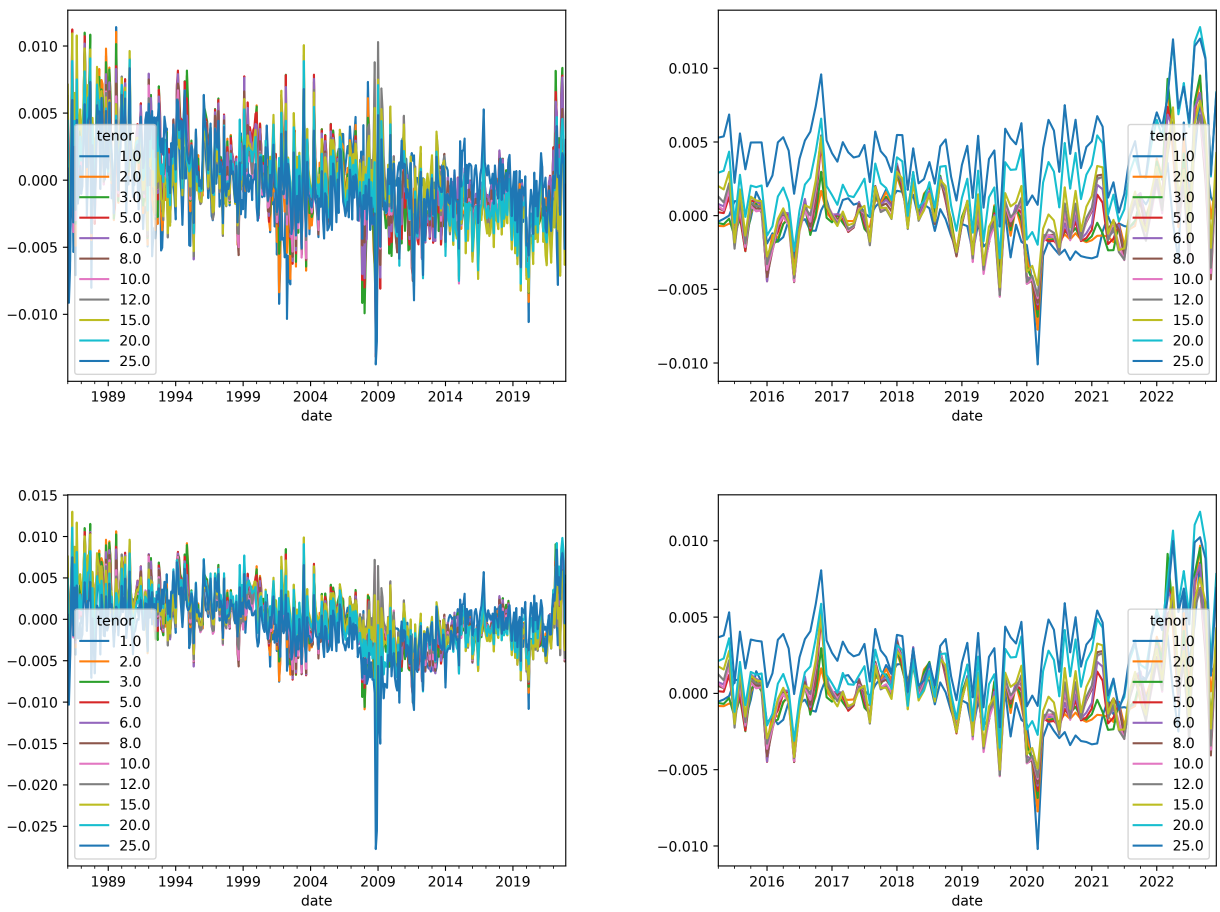

Figure 4.3: The a priori yield errors produced by the Kalman filter for the optimal parameters with $(p, AF?) = (3, \text{yes})$ on top, and $(p, AF?) = (3, \text{no})$ on bottom, on two different calibration time frames (namely, 85–22 on the left and 15–22 on the right).

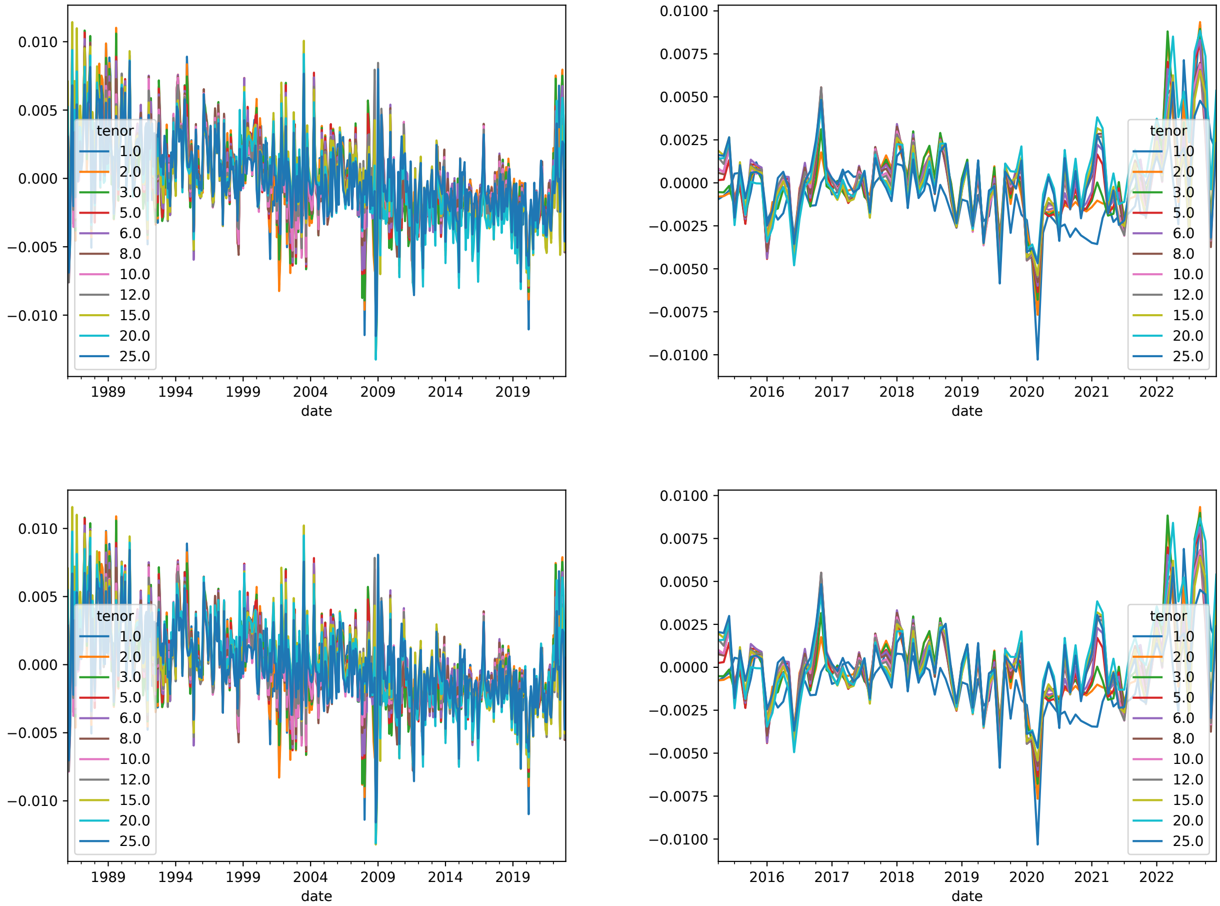

Figure 4.4: The a priori yield errors produced by the Kalman filter for the optimal parameters with $(p, AF?) = (5, \text{yes})$ on top, and $(p, AF?) = (5, \text{no})$ on bottom, on two different calibration time frames (namely, 85–22 on the left and 15–22 on the right).

Figure 4.5: The estimated state vectors produced by the Kalman filter for the optimal parameters with $(p, AF?) =$ (5, yes) on top, and $(p, AF?) =$ (5, no) on bottom, on two different calibration time frames (namely, 85–22 on the left and 15–22 on the right).

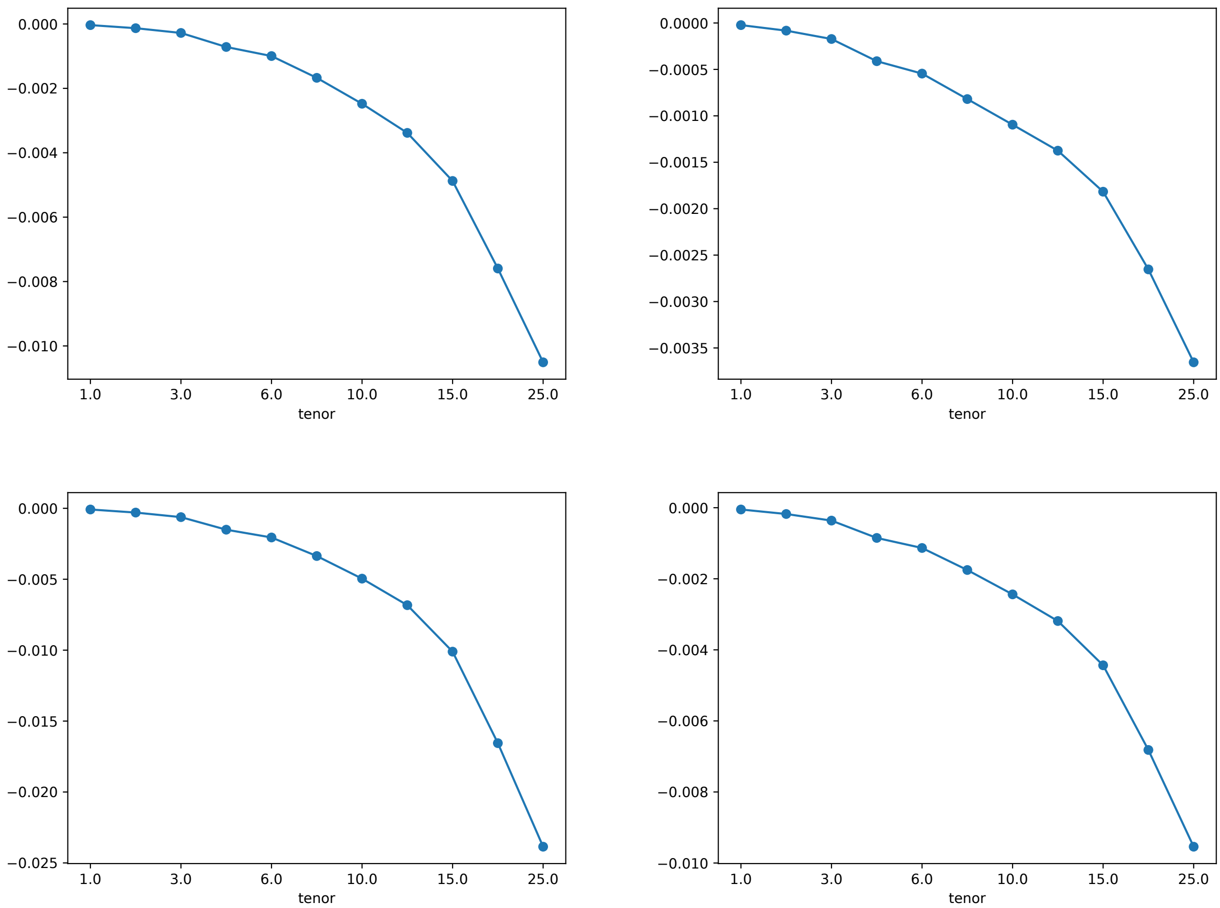

Figure 4.6: Yield adjustment term $-\frac{C(0,\tau)}{\tau}$ (linearly interpolated) for the optimal parameters with $p = 3$ on top, and $p = 5$ on bottom, on two different calibration time frames (namely, 85–22 on the left and 15–22 on the right).

time frame	p	AF?	K_{level}	K_{slope1}	K_{slope2}	$K_{\text{curvature1}}$	$K_{\text{curvature2}}$
85–06	3	no	0.12	0.14	—	1.49	—
85–06	3	yes	0.95	0.95	—	1.20	—
85–06	5	no	0.98	0.98	0.98	1.08	0.98
85–06	5	yes	0.98	0.98	0.98	1.02	0.98
85–09	3	no	0.23	0.13	—	1.00	—
85–09	3	yes	0.98	0.98	—	0.98	—
85–09	5	no	0.99	0.99	0.99	1.02	0.99
85–09	5	yes	2.18	0.94	0.94	0.94	0.94
85–19	3	no	0.96	0.96	—	0.96	—
85–19	3	yes	0.98	0.98	—	0.99	—
85–19	5	no	0.99	0.99	0.99	0.99	0.99
85–19	5	yes	0.98	0.98	0.98	1.00	0.98
15–19	3	no	1.86	0.32	—	1.80	—
15–19	3	yes	0.86	0.86	—	1.95	—
15–19	5	no	2.23	0.94	0.94	1.71	1.55
15–19	5	yes	1.80	0.93	0.93	1.78	2.09
85–22	3	no	0.91	0.91	—	3.62	—
85–22	3	yes	0.96	0.96	—	0.96	—
85–22	5	no	0.97	0.97	0.97	0.98	0.97
85–22	5	yes	0.98	0.98	0.98	0.98	0.98
15–22	3	no	0.95	0.95	—	1.54	—
15–22	3	yes	0.81	0.81	—	1.67	—
15–22	5	no	0.97	0.97	0.97	0.97	1.41
15–22	5	yes	0.99	0.98	1.03	0.98	0.98
20–22	3	no	0.95	0.95	—	0.95	—
20–22	3	yes	0.97	1.03	—	0.97	—
20–22	5	no	0.98	0.98	1.47	0.98	0.98
20–22	5	yes	1.71	0.94	1.75	0.94	0.94
21–22	3	no	0.83	0.99	—	0.83	—
21–22	3	yes	0.14	54.14	—	0.14	—
21–22	5	no	0.10	0.10	29.15	1.32	1.86
21–22	5	yes	0.92	0.92	2.51	1.32	0.92

Table 4.6.: Estimated $K^{\mathbb{P}}$ values.

time frame	p	AF?	σ_{level}	σ_{slope1}	σ_{slope2}	$\sigma_{\text{curvature1}}$	$\sigma_{\text{curvature2}}$
85–06	3	no	0.0071	0.0118	—	0.0242	—
85–06	3	yes	0.0051	0.0135	—	0.0189	—
85–06	5	no	0.0111	0.0146	0.0106	0.0267	0.0296
85–06	5	yes	0.0121	0.0149	0.0122	0.0263	0.0303
85–09	3	no	0.0091	0.0137	—	0.0252	—
85–09	3	yes	0.0080	0.0146	—	0.0210	—
85–09	5	no	0.0119	0.0146	0.0127	0.0275	0.0328
85–09	5	yes	0.0012	0.0158	0.0107	0.0267	0.0515
85–19	3	no	0.0110	0.0144	—	0.0251	—
85–19	3	yes	0.0073	0.0146	—	0.0213	—
85–19	5	no	0.0122	0.0170	0.0140	0.0267	0.0358
85–19	5	yes	0.0093	0.0170	0.0149	0.0254	0.0323
15–19	3	no	0.0068	0.0076	—	0.0162	—
15–19	3	yes	0.0039	0.0072	—	0.0183	—
15–19	5	no	0.0009	0.0041	0.0033	0.0168	0.0175
15–19	5	yes	0.0002	0.0039	0.0036	0.0177	0.0178
85–22	3	no	0.0105	0.0147	—	0.0327	—
85–22	3	yes	0.0070	0.0135	—	0.0228	—
85–22	5	no	0.0119	0.0169	0.0138	0.0268	0.0369
85–22	5	yes	0.0097	0.0174	0.0145	0.0275	0.0330
15–22	3	no	0.0085	0.0128	—	0.0232	—
15–22	3	yes	0.0050	0.0120	—	0.0221	—
15–22	5	no	0.0067	0.0145	0.0096	0.0239	0.0165
15–22	5	yes	0.0041	0.0150	0.0111	0.0227	0.0146
20–22	3	no	0.0094	0.0115	—	0.0290	—
20–22	3	yes	0.0049	0.0144	—	0.0182	—
20–22	5	no	0.0093	0.0122	0.0009	0.0290	0.0209
20–22	5	yes	0.0008	0.0200	0.0141	0.0319	0.0244
21–22	3	no	0.0117	0.0127	—	0.0303	—
21–22	3	yes	0.0106	0.0927	—	0.0282	—
21–22	5	no	0.0077	0.0112	0.0000	0.0340	0.0269
21–22	5	yes	0.0086	0.0129	0.0001	0.0313	0.0235

Table 4.7.: Estimated Σ values.

time frame	p	AF?	θ_{level}	θ_{slope1}	θ_{slope2}	$\theta_{\text{curvature1}}$	$\theta_{\text{curvature2}}$
85–06	3	no	0.0602	-0.0150	—	-0.0045	—
85–06	3	yes	0.0782	-0.0301	—	-0.0106	—
85–06	5	no	0.0449	0.0128	-0.0129	0.0194	0.0831
85–06	5	yes	0.1663	-0.0271	-0.0929	-0.0106	-0.1314
85–09	3	no	0.0668	-0.0370	—	-0.0152	—
85–09	3	yes	0.0886	-0.0462	—	-0.0163	—
85–09	5	no	0.0425	0.0243	-0.0277	0.0222	0.1005
85–09	5	yes	-0.5616	-0.0099	0.6120	0.0015	0.9149
85–19	3	no	0.0585	-0.0265	—	-0.0198	—
85–19	3	yes	0.0747	-0.0437	—	-0.0192	—
85–19	5	no	0.0434	0.0072	-0.0200	0.0052	0.0594
85–19	5	yes	0.1235	-0.0225	-0.0681	-0.0172	-0.0868
15–19	3	no	0.0287	-0.0022	—	-0.0072	—
15–19	3	yes	0.0316	-0.0154	—	-0.0095	—
15–19	5	no	0.0679	-0.1038	0.0537	-0.0574	-0.1449
15–19	5	yes	0.0912	-0.1160	0.0427	-0.0627	-0.1884
85–22	3	no	0.0559	-0.0251	—	-0.0187	—
85–22	3	yes	0.0766	-0.0456	—	-0.0182	—
85–22	5	no	0.0455	0.0058	-0.0213	0.0037	0.0464
85–22	5	yes	0.1291	-0.0257	-0.0725	-0.0176	-0.1046
15–22	3	no	0.0264	-0.0058	—	-0.0134	—
15–22	3	yes	0.0292	-0.0073	—	-0.0163	—
15–22	5	no	0.0492	-0.0321	0.0051	-0.0334	-0.0656
15–22	5	yes	0.1547	-0.0177	-0.1159	-0.0238	-0.1731
20–22	3	no	0.0411	-0.0164	—	-0.0172	—
20–22	3	yes	0.0578	-0.0326	—	-0.0396	—
20–22	5	no	0.0445	-0.1526	0.1357	-0.0738	-0.1149
20–22	5	yes	0.1804	-0.0133	-0.1437	-0.0128	-0.1874
21–22	3	no	0.0439	0.0040	—	-0.0317	—
21–22	3	yes	0.5269	-0.3895	—	-0.5861	—
21–22	5	no	0.1270	0.0666	0.0673	-0.0337	-0.0797
21–22	5	yes	0.0810	-0.2331	0.2006	-0.1106	-0.2344

Table 4.8.: Estimated $\theta^{\mathbb{P}}$ values.

5. Conclusion

In summary, the conclusions of Christensen, Diebold, and Rudebusch (2009, 2011) regarding the performance of the model they put forward still hold up. The Great Financial Crisis remains as the period, where judging from the yield prediction errors and unusual state variable values, the models struggle most, cf. figures 4.3 and 4.5. The general magnitude of the errors is in line with the results of Christensen, Diebold, and Rudebusch (2009, 2011).

Possible and interesting extensions of our empirical part would be

(a) to estimate, in addition to the parameters themselves, also parameter standard deviations as Christensen, Diebold, and Rudebusch (2009, 2011) do; see also Durbin and Koopman (2012, section 7.3.6);

(b) to investigate the performance when using other data sources, in particular for currencies other than US dollar,

(c) to go beyond ($\mathbb{P}$-)*independent* factor models,

(d) to follow a one-step approach in the sense of Andreasen, Christensen, and Rudebusch (2019) and Pancost (2021): instead of first constructing raw yield curves as described in section 4.2 and then doing parameter optimization using a fixed grid of maturities, one can also forego the first step by measuring bond prices (instead of yields) in an approach using an *extended* Kalman filter.

On the more theoretical side, we showed (section 3.6.2) how the arbitrage-free (generalized) Nelson-Siegel model leads to a simple closed price formula for European options on zero-coupon bonds. This could be extended to coupon-bearing options using what has become known as Jamshidian's trick (Jamshidian, 1989).

As for possible future research, there are many interesting directions left to explore around the arbitrage-free dynamic generalized Nelson-Siegel model, like incorporating non-constant volatility or regime-switching approaches.

Bibliography

Andreasen, M. M., Christensen, J. H., & Rudebusch, G. D. (2019). Term structure analysis with big data: One-step estimation using bond prices [Big Data in Dynamic Predictive Econometric Modeling]. *Journal of Econometrics*, *212*(1), 26–46. https://doi.org/10.1016/j.jeconom.2019.04.019

Baxter, M., & Rennie, A. (1996). *Financial calculus: An introduction to derivative pricing*. Cambridge University Press. https://doi.org/10.1017/CBO9780511806636

BIS Monetary and Economic Department. (2005). *Zero-coupon yield curves estimated by central banks* (BIS Paper No. 25). Bank for International Settlements. https://www.bis.org/publ/bppdf/bispap25.pdf

Bolder, D., & Stréliski, D. (1999). *Yield Curve Modelling at the Bank of Canada* (Technical Report No. 84). Bank of Canada. https://doi.org/10.2139/ssrn.1082845

Brigo, D., & Mercurio, F. (2006). *Interest rate models — theory and practice. with smile, inflation and credit* (2nd ed.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-34604-3

Cairns, A. J. G. (2004). *Interest Rate Models. An Introduction*. Princeton University Press. https://doi.org/10.1515/9780691187426

Cheyette, O. (2002). Interest rate models. In F. J. Fabozzi (Ed.), *Interest Rate, Term Structure, and valuation modeling* (pp. 3–25). John Wiley & Sons, Inc. https://catalogimages.wiley.com/images/db/pdf/E0471220949.01.pdf

Christensen, J. H., Diebold, F. X., & Rudebusch, G. D. (2009). An arbitrage-free generalized Nelson-Siegel term structure model. *The Econometrics Journal*, *12*(3), C33–C64. https://doi.org/10.1111/j.1368-423X.2008.00267.x

Christensen, J. H., Diebold, F. X., & Rudebusch, G. D. (2011). The affine arbitrage-free class of Nelson-Siegel term structure models. *Journal of Econometrics*, *164*(1), 4–20. https://doi.org/10.1016/j.jeconom.2011.02.011

De Pooter, M. (2007). *Examining the Nelson-Siegel class of term structure models: In-sample fit versus out-of-sample forecasting performance* (Discussion Paper 07-043/4). Tinbergen Institute. https://doi.org/10.2139/ssrn.992748

Delbaen, F., & Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, *300*(1), 463–520. https://doi.org/10.1007/bf01450498

Diebold, F. X., Ji, L., & Li, C. (2006). A three-factor yield curve model: Non-affine structure, systematic risk sources and generalized duration. In L. R. Klein (Ed.), *Long-run growth and short-run stabilization* (pp. 240–274). Edward Elgar Publishing. https://doi.org/10.4337/9781781950500.00014

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, *130*(2), 337–364. https://doi.org/10.1016/j.jeconom.2005.03.005

Diebold, F. X., Rudebusch, G. D., & Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, *131*(1-2), 309–338. https://ideas.repec.org/a/eee/econom/v131y2006i1-2p309-338.html

Duffie, D., & Kan, R. (1994). Multi-factor term structure models. *Philosophical Transactions: Physical Sciences and Engineering*, *347*(1684), 577–586. https://doi.org/10.1098/rsta.1994.0067

Duffie, D., & Kan, R. (1996). A Yield-Factor Model of Interest Rates. *Mathematical Finance*, *6*(4), 379–406. https://doi.org/10.1111/j.1467-9965.1996.tb00123.x

Duffie, D., Pan, J., & Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, *68*(6), 1343–1376. https://doi.org/10.1111/1468-0262.00164

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (2nd ed.). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199641178.001.0001

European Central Bank. (2008). The new Euro area yield curves. *ECB Monthly Bulletin*, 95–103. https://www.ecb.europa.eu/pub/pdf/other/pp95-103_mb200802en.pdf

Fama, E. F., & Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, *77*, 680–692. http://www.jstor.org/stable/1814539

Filipović, D. (1999). A note on the Nelson-Siegel family. *Mathematical Finance*, *9*(4), 349–359. https://doi.org/10.1111/1467-9965.00073

Filipović, D. (2009). *Term-structure models: A graduate course*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-68015-4

Gürkaynak, R. S., Sack, B., & Wright, J. H. (2007). The U.S. Treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, *54*(8), 2291–2304. https://doi.org/10.1016/j.jmoneco.2007.06.029

Harrison, J., & Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, *11*(3), 215–260. https://doi.org/https://doi.org/10.1016/0304-4149(81)90026-0

Jamshidian, F. (1989). An exact bond option formula. *The Journal of Finance*, *44*(1), 205–209. https://doi.org/10.1111/j.1540-6261.1989.tb02413.x

Kim, C.-J., & Nelson, C. R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. The MIT Press. https://doi.org/10.7551/mitpress/6444.001.0001

Liu, Y., & Wu, J. C. (2021). Reconstructing the yield curve. *Journal of Financial Economics*, *142*(3), 1395–1425. https://doi.org/10.1016/j.jfineco.2021.05.059

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, *60*(4), 473. https://doi.org/10.1086/296409

Nymand-Andersen, P. (2018). Yield curve modelling and a conceptual framework for estimating yield curves: evidence from the European Central Bank's yield curves. *ECB Statistics Paper Series*, *27*. https://www.ecb.europa.eu/pub/pdf/scpsps/ecb.sps27.en.pdf

Pancost, N. A. (2021). Zero-Coupon Yields and the Cross-Section of Bond Prices. *The Review of Asset Pricing Studies*, *11*(2), 209–268. https://doi.org/10.1093/rapstu/raab002

Svensson, L. E. (1995). Estimating Forward Interest Rates with the Extended Nelson & Siegel Method. *Sveriges Riksbank Quarterly Review*, *1995*(3), 23–26. https://larseosvensson.se/papers/95QRabs

The Economist. (2022a). Investing in an era of higher interest rates and scarcer capital. Retrieved January 26, 2024, from https://www.economist.com/leaders/2022/12/08/investing-in-an-era-of-higher-interest-rates-and-scarcer-capital

The Economist. (2022b). Rising interest rates and inflation have upended investing. Retrieved January 26, 2024, from https://www.economist.com/briefing/2022/12/08/rising-interest-rates-and-inflation-have-upended-investing

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, *5*(2), 177–188. https://doi.org/10.1016/0304-405X(77)90016-2

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wahlstrøm, R. R., Paraschiv, F., & Schürle, M. (2021). A Comparative Analysis of Parsimonious Yield Curve Models with Focus on the Nelson-Siegel, Svensson and Bliss Versions. *Computational Economics*, *59*(3), 967–1004. https://doi.org/10.1007/s10614-021-10113-w

Appendix A.

Implementation Details

As already mentioned in section 4.1.1, the code is implemented in Python. It can be found under https://github.com/snejens/arbfree_dynamic_nelson_siegel.

Running the code

To run the code, one will first need to change the definition of the `BASE_PATH` variable in the file `arbfree_dyn_ns/config.py`. Apart from that, one needs to download the data (`LW_monthly.xlsx`) from the source indicated in section 4.2 and put that Excel file into a new subfolder `data` under the `BASE_PATH` configured previously.

Structure of the code

We describe the structure of the code:

- The main part is the Python module `arbfree_dyn_ns`, containing:
  - `config.py`: houses some general configuration, like paths (see section above) and the definition of the grid of maturities used during optimization,
  - `generate_testdata.py`: generates completely synthetic yields (see also section below),
  - `kalman.py`: implements the Kalman filter,
  - `lw_data.py`: reads in and processes the data described in section 4.2,
  - `main.py`: starts optimization runs and saves the results of that to the disk; invoked as a command-line script,
  - `nss.py`: stands for "Nelson-Siegel-Svensson"; implements the formulas for the generalized Nelson-Siegel coefficient matrix B and for the yield adjustment C,

  - `optimization.py`: the optimization code; the objective function that is defined here transforms the arguments (to encode the stationarity constraint on $K^{\mathbb{P}}$—cf. the discussion in section 4.1.1—and also the constraints $\lambda_1 > \lambda_2 > 0$, and $\sigma_{ii} > 0$ for all i) and invokes the Kalman filter from `kalman.py` and extracts the loglikelihood from its return value,
  - `packing.py`: auxiliary functions to format input consisting of a sequence of matrices, vectors and the like as a single flat array to feed into the generic numerical optimization method, and to undo such a transformation,
  - `utils.py`: contains various small auxiliary functions used elsewhere in the module,
  - `vector_autoregression_process_simulation.py`: simulates a VAR(1) and the associated yields under a (potentially arbitrage-free) dynamic (potentially generalized) Nelson-Siegel specification,
  - `buba_data.py`, `curves.py`: these files are ultimately not used; they relate to processing another data source, which was not pursued to the end.

- The shell script `runner.sh` shows how `main.py`, which has been explained above, can be invoked.

- The `notebooks` folder contains Jupyter notebooks, in particular

  - `read-results.ipynb`: here the results saved to the disk by `main.py` are read in and analyzed, in particular it contains the code for generating most of the tables and plots in this thesis,
  - `Test-B-C.ipynb`: tests that the (generalized) Nelson-Siegel coefficient matrix B is consistent between $p = 3$ and $p = 5$ (see section below), and reproduces the yield adjustment term one sees in Christensen, Diebold, and Rudebusch (2011, Figure 1).

Implementation tests

The following tests were used to verify the implementation's correctness:

- reproduction of the yield adjustment term as shown in Christensen, Diebold, and Rudebusch (2009, 2011) given the parameters from there,

- running the optimization on generated data that exactly follows an arbitrage-free generalized Nelson-Siegel model plus some white noise reproduces approximately the parameters that were input,

- check that the (generalized) Nelson-Siegel coefficient matrix B for $p = 3$ is consistent with B for $p = 5$ in that the former can be obtained by striking the third and fifth columns of the latter,
- check that interchanging the "slope 1" and "curvature 1" components of $K^{\mathbb{P}}, \theta^{\mathbb{P}}, \Sigma$ as well as λ_1 with the "slope 2" and "curvature 2" components of $K^{\mathbb{P}}, \theta^{\mathbb{P}}, \Sigma$ as well as λ_2, respectively, does not change the outcome (search for "interchange" in `optimization.py` to find this test).