FUNDAMENTALS OF COMPUTATIONAL MATHEMATICS  
Prof. Manuel Aprile

Prof. Francesco Marchetti  
University of Padova

**Project: Yield Curve Model Calibration**

**Ibrahim Uali**

**Riccardo Caruso**

**Enrico Berto**

**Elisa De Colle**

February 16, 2024

# THE NELSON-SIEGEL MODEL:

The Nielson Siegel model is a model used for the prediction of yield curves, developed by Charles Nelson and Andrew Siegel in 1987 at the University of Washington.

This model aims to predict the movements of the yield curve basing itself on various terms to maturity, such as: flat, hump and S shapes.

The formulated model is the following:

With equal to yield at time t, while , and are parameters associated to the long-run yield level, slope of the yield and curvature of the yield respectively.

The last parameter, , is the one influencing the rate at which yields converge to the long-term level.

This model is widely used by economists because it has several characteristics that tend to make it a versatile and efficient model. First of all, it respects the restrictions imposed by the economic and financial theory, moreover its approximation avoids in-sample overfitting, such as that it leads to an increase in forecasting capacity.

In addition, versatility refers to the fact that it can take any yield curve form that has been empirically observed in the market.

For these reasons the model is extensively used by central banks and monetary policy makers, while fixed-income portfolio managers use it in order to immunize their portfolios.

Academic literature abounds with instances where researchers have derived significant conclusions by employing the Nielson-Siegel model: Diebold and Li (2006) benchmarked the Nielson-Siegel against other models in the term structure forecasts, the finding was that, especially for longer forecast horizons, it performs very well. Coroneo, Nyhlom and Vidava-Koleva (2008) test to which degree the Nielson-Siegel model approximate an arbitrage-free model.

Although there exist also other models used to estimate the term structure of interest rates, most of them have been proven to have undesirable economic properties, moreover they have been seen to be black-box models, so they are not easy to be interpreted and understood by humans.

## ESTIMATION DIFFICULTIES: THE PROBLEMS OF OPTIMIZATION AND MULTICOLLINEARITY

The estimation problems addressed to this model refer to his lack of linearity and to the presence of multicollinearity between variables.

Since the parameter that causes non – linearity in the model is λ, or the so-called ‘shape parameter’, many estimation approaches are based on the condition that λ is fixed, in order to linearize the model and perform the estimation by using Ordinary Least Squares.

We are going to investigate this approach more in the detail later, for now we just state that the main problem of this model is that it can become heavily collinear depending on the estimate or fixed shape parameter, λ.

The issue of multicollinearity is that, when dealing with it, the estimation of parameters performed by using least squares can lead to unstable and highly variable coefficient estimates.

Since the Nielson-Siegel model is highly non-linear, we need to estimate its parameters by using non - linear optimization techniques, like for example the Gradient Descent, always considering that this model tends to be very sensitive to the starting values used in the optimization. This issue refers to what is called the “Optimization problem”.

In order to investigate further these problems, we consider a more general representation of the model:

A number of mathematical equations

Description automatically generated

Here we divided the model in three main components, r0, r1 and r2 that represent respectively the level, the slope, and the curvature of the spot curve.

Deeping further in the behavior of the model, we can see that when the time to maturity grows to infinity, the slope and the curvature components goes to zero and the long-term spot rate converges to a constant level of interest, that is .

is the spread, it measures the slope of the term structure, a negative value represents an upward.

The degree of curvature is determined by , which is the rate at which the slope and curvature components decay to zero.

Λ here determines the location of the maximum/minimum value of the curvature component, it determines both the shape of the curvature component and the hump/through of the term structure.

When dealing with Ordinary Least Squares estimation, the multicollinearity problem refers to the high correlation between the slope (r1) and the curvature (r2) components, this correlation heavily depends on the choice of λ, that is usually set to be 1.37 or 3, other than that, it depends also on the choice of the time to maturity vector.

To summarize, Gilli et al. (2010) identifies three major problems with estimating this model by non-linear methods or OLS:

* The optimization problem is non-convex, showing multiple local optima.
* The model is badly conditioned for a certain range of parameters, the estimates are unstable before small price movements.
* The value of the decay factor (λ) directly affects the correlation between the loadings of the slope and curvature factors, the problem here could be in the forecasting use of the model.

# EXTENDED NELSON-SIEGEL FUNCTION or SVENSSON MODEL

In order to extend the Nielson – Siegel model and make it possible to capture economics change in case of market crisis, Svensson added an extra term to the original model, allowing for a second hump in the yield curve.

This fifth parameter is able to capture a greater variety of yield curve shapes, that can fit in a good way also yields in time of monetary crisis.

The model is computed as follows:

The parameters present in the original model have the same interpretation that they had in the first one, while and determine position and magnitude of a possible second hump in the yield curve.

# THE PROJECT:

## DATA COLLECTION:

We started by collecting the historical data of various European bonds, we selected four time to maturity: one, two, three, five and ten years, in order to cover the whole maturity range and distribute the data along the whole period.

We chose not to select bonds with very short maturities, because it is possible that they don’t show sufficient liquidity.

The data selected are both from European and non-European countries, to be more precise we chose bonds from four countries: United States, Portugal, Germany and South Corea.

The criteria followed to choose these countries is the necessity to have bonds coming from different market economies, in order to study the performances of the Nelson-Siegel model in various economic situations and markets.

United states bonds are believed to be one of the safest investments, moreover they are highly liquid and considered close to risk-free, since they are backed by the U.S. government’s; given their stability and reliability they tend to be very popular in the financial market.

German bonds have kind of the same characteristics of the US ones in terms of stability and reliability in the Eurozone, they have a high credit rating, AAA, that states for a very low risk of default.

Given their reliability they are often used as a benchmark in the bond market, their yields are often compared to those of other European bonds, they serve as a reference point for assessing credit risk in the Eurozone, this characteristic make these bonds highly liquid.

Portugal bonds, on the other hands are less liquid than the German one, even if they are relatively liquid, and they are considered to be less reliable due to the economic conditions of the country.

For what concern the choice of South Corean bonds, we selected them because they tend to show very interesting yield curves which tend to be unusual respect to the ones typical of developed economies, this may be due to the fact that the Korean currency is weak with respect to the US dollar.

The time frame we are referring to goes from the first of January, 2022 to the first of December, 2023.

## CREATION OF THE DATAFRAME:

We started with the creation of a global Data frame in order to be able to work directly on it, at first we imported on python the data downloaded from the website ‘investing.com’, since the data we found where divided by each maturity, at first we have six data frames for each country.

The data we found were provided jointly with other information that are not needed to perform the Nelson - Siegel model, in explicit market values of the bonds such as the open, high and low values of the price and the change in percentual. In order to get rid of them we applied the function *‘clear\_df’* which purpose is to drop useless columns and set the date of the month as index; please find this function in the script ‘fun.py’.

Once our data has been cleaned from the useless one, we find ourself with a data frame for each maturity for each country, with a column for the prices indexed by month.

In order to perform the Nelson-Siegel model we need a data frame for each date, which report the yield of the bonds with respect to the maturity. In order to obtain this, we developed the following function *‘join\_df\_date’:*

A close-up of a text

Description automatically generated

With this command we are able to join our data frames with a column for price and the other one for maturity for each month, the output is a list of data frame for each month, all joint together, one for every country.

For example, the data frame obtained for the German bonds referring to the date 1st of August 2023 is the following:

A screenshot of a computer

Description automatically generated

Now we want to add other columns to these data frames containing the predicted yields by mean of the Nelson-Siegel model, we implemented the function presented in a theoretical way in the previous paragraph in python, whit that function we can compute the predicted R(t).

A computer screen shot of a code

Description automatically generated

In this function we added also the possibility to predict the curve by means of the Nelson – Siegel – Svensson model as you can see in the ‘else’ statement, the advantages of this extended model are going to be exploited further in this paper.

The predictions that we are going to get with this curve won’t be any good but will change once we have optimized the function.

That is because this model relies on some parameters, that are:

* The time, for which the expected return has to be computed.
* The parameters insight the model, which are β0, β1, β2, τ. (β3 and τ2 for the Nelson-Siegel-Svensson).

The setting of the starting parameters is quietly important, especially for the optimization methods that we are going to use later in order to optimize our model. We underline how especially the Gradient Descent method tends to be very sensitive to the initial parameters, in such a way that if we set values of them that are not appropriate our optimization function will diverge to infinity.

We based our selection of appropriate values on insights from relevant academic papers, moreover our choice has been made after some trials of the behavior of the optimization methods with different starting values, we noticed that for example, by setting all the parameters to 0.01 the gradient descent method was diverging.

To kick off our exploration of optimal parameters, we adopted a randomized approach, selecting values from a range recommended by various academics, notably Gilli et al. (2010). This initial phase involved systematically testing different parameter combinations to identify the most promising configurations.

The ranges they suggested are:

* 0 β0 15
* -15 β1 30
* -30 β2 30
* -30 β3 30
* 0 τ 2.5
* τ2 5.5

We performed several trials with different combinations of values, the evidence that emerged from our empirical work, mostly by looking at the plots of the Nelson Siegel compared with the ones of the Historical data, was that values higher than 1 for the betas do not perform well, given that we decided to go on with our trials by employing lower values.

Moreover, the plots suggested to us that the value of β3 should be negative.

After more or less (tot) number of trials, we decided to select the following values, according to the higher performance they were showing for both the Gradient Descent and the Newton methods:

* β0 = 0.01
* β1 = 0.05
* β2 = 0.2
* β3 = 0.03
* τ, τ2 = 1

Added the Nelson Siegel model values to our data frames we have all the data we need in order to compute the sum of squared difference between observed and model predicted yields and optimize it according to the parameters.

The same has been made in order to add also the Nielson Siegel Svensson values.

## PARAMETER ESTIMATION:

The sum of squared difference between observed and model predicted yields is computed as follows:

This is the function to compute the sum of squared residuals for the Nelson-Siegel model, once again in the python code we also presented the version for the Nelson-Siegel-Svensson model.

A computer code with black text

Description automatically generated

Now we can proceed with the minimization of ƒ, adjusting the parameters via de Gradient Descent and the Newton methods.

In order to use these two methods, we imported them at the beginning, adjusting a bit at our convenience the one presented in class, during Laboratory 02; we imported *‘Lab02\_solutions’* as lb in order for our code to be more compact.

### GRADIENT-DESCENT METHOD:

We initiate the process with the Gradient Descent method, which operates on a continuous differentiable function, denoted as ƒ, aiming to minimize it.

The procedure relies on a descent direction, d, and a step size, α.

The descent direction is the gradient of f calculated in the point before the one we are trying to update, while α is found by performing approximate line search. The updated x is going to be the previous one plus alpha multiplied by the descent direction.

The procedure is going to stop either when the number of iterations reaches the maximum decided at the beginning or when the value of the last iteration is below a certain threshold decided at the beginning.

We established the maximum number of iterations to be 50.

The functions we are referring to are named in python as *‘gradient\_descent’* and *‘apx\_line\_search’* and they are stored in the script *‘Lab02\_solutions.py’.*

We ran a for loop to perform the gradient descent methods in each data frame for every country we are referring to. Once the optimization has been done, we collected the values of the parameters for every iteration in a data frame that we saved in Excell. For the complete code please refer to the script *‘Data.py’.*

By running the code, we obtained that the Gradient Descent method performed fifty iterations, which is the maximum number of iterations we decided at the beginning, clearly this state that there is an issue in the convergence of the method, which is converging slowly.

Also moving to the plots one can clearly notice that the observed yield curve and the predicted one are not following the same route.

The issue in the convergence of the model could be given by the fact that the initial parameters are not the optimal ones, given the sensibility of the model to the starting values, the step size alpha or the maximum number of iterations are too small.

To address this issue, we should explore various combinations of parameters.

### NEWTON METHOD:

The second method we implemented is the Newton method.

This procedure for functions optimization is similar to the Gradient Descent one, except that in this case the updating step is computed using second-order Taylor expansion. The ‘d’ we were referring about in the Gradient Descent method in this case is the opposite of the gradient at point xᵢ, while alpha is the inverse of the Hessian at point xᵢ.

In order for this method to work it is important to check that the Hessian is a Positive Definite matrix, otherwise the method is not applicable since the Hessian would not be invertible. In order to avoid this problem, usually a regularization term, given by λ I, is added to the matrix in order to make it positive definite.

The regularization term mentioned above has been called in the code ‘Damping factor’ and has been set to be 0.5.

Once again, to use this method, we implemented the code presented in laboratory hours, that we imported at the beginning of our script.

The function *‘newton\_method’* is looped across all the dataframes we created at the beginning for each country, in explicit we have twenty-three different dataframes for every country we selected, one for every month of our time frame.

At each iteration the Nelson Siegel parameters are optimized according to the Newton’s method, the number of iterations performed to achieve the optimal parameters are then returned by the *‘newton\_method’* function itself.

For example the number of iterations performed by Nelson Siegel in order to reach the optimal parameters for US bonds on the 1st of January, 2022 are 41.

We stored the values of every iteration in an Excel file, if it is of your interest, please find it in the zip file.

Once we have done that, in order to plot the curve predicted by the Nelson Siegel model with respect to the observed yield curve, we need to select the parameters found by the last iteration performed (that should be the optimal ones) and compute the expected returns by means of the Nelson Siegel model with those parameters.

The following piece of code is an example of this procedure, applied to US bonds:

A computer screen shot of a code

Description automatically generated

From this code we obtain the plots, here we present the example of US bonds for the 1st of January 2022.

A graph with numbers and lines

Description automatically generated

One can see that The Nelson Siegel curve closely mirrors the historical yields data, exhibiting a remarkably strong correspondence.

# COMPARISON OF GRADIENT DESCENT AND NEWTON METHODS

Based on the results we have presented thus far, it's evident that disparities exist between the outcomes obtained using the Gradient Descent method and those using Newton's method. These discrepancies extend beyond mere implementation nuances and likely stem from inherent divergences in the optimization strategies employed by each method.

First, we noticed that in the case of the Gradient Descent method convergence to the optimal values is way slower, while Newton always converges in less iterations. This difference is given by the fact that the former method doesn’t consider information about the curvature of the function, while the latter is able to capture them by employing the Hessian.

Newton's method has stronger constraints in terms of the differentiability of the function than Gradient descent. If the second derivative of the function is undefined in the function's root, then we can apply gradient descent on it but not Newton's method.

Moreover, the fact that the second derivative is required to apply the Newton method makes it more computationally expensive with respect to the Gradient Descent one, especially in the case of high-dimensional problems.

Newton and gradient descent method are gradient based algorithms, given that, a starting point is chosen and then the algorithm is forced to explore a small area around it. For this reason, other methods tend to outperform with respect to them, since we are trying to optimize a function with multiple local optima.

The gradient based method depends heavily on the starting points, so if the ones chosen are not the correct ones, this can lead to heavy errors in the optimization. However, from a financial perspective, employing the model in a financial frame allows a good initial guess of the parameters, leading to the possibility to make use of gradient-based methods in an efficient way.

Newton and gradient descent method are gradient based algorithms, given that, a starting point is chosen and then the algorithm is forced to explore a small area around it. For this reason, other methods tend to outperform with respect to them, since we are trying to optimize a function with multiple local optima.

Moreover, for what concern the Svensson model, proof have been found that gradient based methods are unable to exploit the extra term of the extended model, so they seem not to be efficient while studying this algorithm.

# OTHER APPROACHES IMPLEMENTED IN THE OPTIMIZATION OF THE MODEL:

We will see the most common approaches used to get an estimation of the model parameters.

We underline that, the approaches that are most used by analyst are, ordered by number of times used:

1. ORDINARY LEAST SQUARES
2. QUASI-NEWTON METHOD: BFGS algorithm
3. GRADIENT DESCENT METHOD

## ORDINARY LEAST SQUARES – RIDGE REGRESSION APPROACH:

Due to the convenience and the simplicity of linearizing the model, grid search or OLS approach are the most used estimation procedure to optimize the Nielson-Siegel model.

This approach is performed by fixing the parameter of the shape λ.

The main issues recurring in these approaches are that they can occur in the complication that the Nielson-Siegel model becomes heavily collinear, and this depends on the fixed shape parameter.

This problem is solved by performing the so-called ridge regression: in explicit a regularization term is added to the traditional least squares objective function of linear regression.

The regularization term is composed by a regularization parameter, which controls the strength of the regularization, multiplied by the sum of the squared values of the coefficients.

The aim of this regularization term is to help to prevent overfitting and to stabilize the model by providing a balance between a well fit to the data and the necessity of keeping coefficients far from being too large.

In this approach the non-linear problem is transformed into a linear one by fixing λ, that is the parameter responsible of the lack of linearity in the model; once the model has been linearized, in order to obtain parameters that ensures the best fit, they are estimated by OLS, conditional upon a grid of the fixed shape parameter.

The parameters are estimated by minimizing the sum of squared error, by using a grid search in order to determine the optimal λ, which is chosen according to the value of the R², and linear regression.

Once the optimal λ has been detected, there is the need to test the degree of multicollinearity of the two factors, if this degree is too high the results need to be re-estimated using ridge regression.

The main advantage of this method is that it does not depend on any starting values, while non-linear optimization techniques give estimates that are very sensitive to the starting value of the optimization.

On the other hand, as we stated before, in this case it must be kept in mind that the Nielson-Siegel model is very sensitive to the choice of the λ (which is usually set to be 1.37, Diebold and Li (2003) or 3, Fabozzi et al. (2006)). To understand the magnitude of this issue de Pooter (2007) studied that with different λ fixed, the remaining parameters can take extreme values.

Moreover, the correlation between the two parameters varies depending on the remaining maturity of the financial instruments chosen in the bootstrap.

# BIBLIOGRAFY:

Diebold, F. X., Li, C., and Yue, V. Z., “Global Yield Curve Dynamics and Interactions: A Generalized Nelson-Siegel Approach”, Journal of Econometrics, 146 (2008), 351-363

Fabozzi, F. J., Martellini, L. And Priaulet, P., “Predictability in the Shape of the Term Structure of Interest Rates”, Journal of Fixed Income (June, 2005) 40-53

Annaert J., Claes A.G.P., De Ceuster M.J.K., Zhang H., “Estimating Nielson-Siegel: A ridge regression approach”, Universiteit Antwerpen (2010)

P. Manousopoulos, M. Michalopoulos, “Comparison of non-linear optimization algorithms for yield curve estimation”, European Journal of Operational Research 192 (2009) 594–60

Wirz, Matt, and Saeedy, Alexander. “Bond Markets Forecast Long Financial Freeze for Russia.” Wall Street Journal, March 15, 2022.

Gilli et al., “Calibrating the Nelson-Siegel-Svensson model”, Comisef