Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

Score-Driven Nelson Siegel

Hedging Long-Term Liabilities

Rogier Quaedvlieg and Peter Schotman

DP 02/2016-019

Score-Driven Nelson Siegel: Hedging Long-Term Liabilities.

This Version: February 25, 2016

Rogier Quaedvlieg[a,*], Peter Schotman[a]

[a]*Department of Finance, Maastricht University, Netherlands*

Abstract

Due to its affine structure the Nelson-Siegel model for yield curves can be transformed to a factor model for excess bond returns. Hedging interest rate risk in this framework amounts to eliminating the factor exposure and minimizing the residual risk. Fitting the model directly on excess returns with constant factor loadings leads to large hedging errors caused by substantial and persistent time-variation in the shape parameter of the Nelson-Siegel factor loadings. To capture this variation we develop a Dynamic Conditional Score (DCS) model for the shape parameter. This dynamic model offers superior hedging performance and reduces the hedging error standard deviation by almost 50% during the financial crisis. Much of the improvement is due to the model for the shape parameter with some further reduction achieved by a GARCH model for the residual risk.

Keywords: Dynamic Conditional Score, Risk Management, Term Structure.

1. Introduction

Hedging interest rate risk requires a model of how prices of bonds with different maturities react to shocks. There is a large literature that fits and models the time series of the term structure of bonds, returns and their yields. As the aim is to forecast a large cross-section of maturities, the literature imposes a factor structure on the yield curve to reduce the dimensionality of the problem. Most popular is the class of affine term structure models (Duffie and Kan, 1996; Dai and Singleton, 2000; Duffee, 2002), which is both tractable from an econometric perspective as well as consistent with no-arbitrage conditions from finance theory.

Nelson and Siegel (1987) propose a purely statistical model which has good empirical fit. The model has received quite some attention since Diebold et al. (2006) who extend the

[☆]This research was financially supported by a grant of the Global Risk Institute. We thank participants in the Duke Financial Econometrics Seminar.

[*]Corresponding author: Department of Finance, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands.

Email addresses: r.quaedvlieg@maastrichtuniversity.nl (Rogier Quaedvlieg), p.schotman@maastrichtuniversity.nl (Peter Schotman)

standard model to a dynamic version, which describes the yield curve over time as a three factor model. The three factors are the well-known level, slope and curve factors, whose exact shape is determined by a single parameter λ.[1] As a further extension Christensen et al. (2011) showed that a simple additive constant term makes the model arbitrage free and a a member of the affine class.

In this paper we use the Nelson-Siegel model for the purpose of risk management and specifically to hedge the risk of a long-term liability using shorter-term bonds. This empirical problem is of great interest to institutions like pension funds, who often have difficulty hedging long-term obligations. To this purpose, we transform the Nelson-Siegel to a factor model on excess bond returns, and obtain a model of the impact of shocks on bond prices. Since the Nelson Siegel model imposes a factor structure, hedging amounts to forming a portfolio with opposite factor exposure to the liability, and as such, for a successful hedge, a correct specification of the factor structure is vital. In the Nelson-Siegel model, the factor structure is fully controlled by the single λ parameter. This parameter has mostly been disregarded, and is often not even estimated, but set to a value chosen by the users. We focus on the λ and propose a Dynamic Conditional Score version of the model which provides an intuitive way to model time-variation in λ. Crucially, if λ is time-varying, the true factor exposure cannot be adequately hedged. We know of two other papers that allow for time-variation of λ. Koopman et al. (2010) estimate time-variation using an extended Kalman filter, while Creal et al. (2008) allow for time-variation as a function of the time-varying factor exposures.

Additionally, we allow for heteroskedasticity in the model. Bianchi et al. (2009) model time-variation in the variances of the factors. Here, the factor loadings for the term structure are also used as weights for the volatility in the term structure. Koopman et al. (2010) assume a state-space model where the first principal component of the Nelson-Siegel error follows a GARCH(1,1) process. The same construction is used for the volatility of the measurement error of the state variables. That means that both the innovation and the measurement error are driven by a single GARCH(1,1) process. Hautsch and Ou (2012) propose a stochastic volatility specification of the Nelson-Siegel model. Similar to Koopman et al. (2010) we use a single GARCH model that drives volatility dynamics for all maturities.

In our model the time-variation in λ and the variance are interconnected as the covariance matrix directly influences the updates of λ. We find considerable time-variation in λ and show the importance of the heteroskedasticity. Importantly, when not taking into account the time-variation of the variance, the factor structure breaks down during the financial crisis of 2008: λ converges to zero, effectively making it a two-factor model. Taking into account the volatility, λ actually increases to levels not seen before. If not modelled, the large variation in returns simply obscures the factor structure. The relation between the factor loadings and

[1] For a textbook description, see Diebold and Rudebusch (2013).

volatility in our model prevents the factor structure from breaking down.

In an out-of-sample hedging exercise we show that the factor exposure of the hedge portfolio constructed using the Nelson Siegel with time-varying λ and volatility is reduced by almost 50% compared to the standard Nelson-Siegel model where both are constant. The gains mainly stem from the time-variation in λ, but also allowing for heteroskedasticity improves the hedging results further.

The remainder of the paper is organized as follows. In Section 2 we recap the Nelson-Siegel model and introduce how it can be used for hedging. Section 3 introduced the Dynamic Conditional Score version of the NS model with time-varying λ. Section 4 gives in-sample estimation results of the various models, which are used in an out-of-sample hedging exercise in Section 5.

2. Nelson Siegel model

Denote by $y_t(\tau)$ the yields at time t for maturity τ. At any time, the yield curve is some smooth function representing the yields as a function of maturity τ. Nelson and Siegel (1987) (NS) provide a parsimonious factor description of these yields. The Diebold and Li (2006) formulation can be written as

$$y_t(\tau) = B(\tau)\tilde{f}_t, \tag{1}$$

where

$$B(\tau) = \begin{pmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \end{pmatrix}' \tag{2}$$

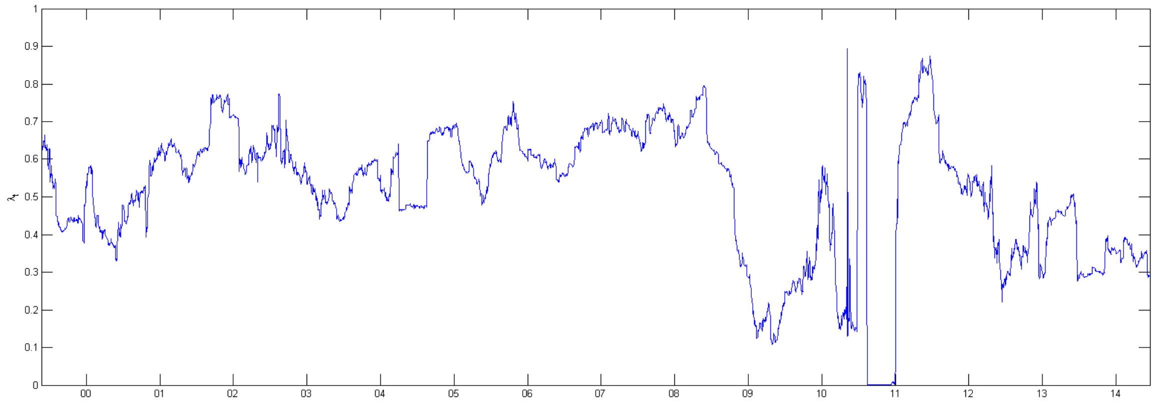
depends on the shape parameter $\lambda > 0$. The NS model is a three-factor model where the three factors are often interpreted as level, slope and curve respectively. The first component has constant factor loadings of 1. As such, it influences short and long term yields equally, and can be considered as an overall level curve. The second component converges to one as $\tau \downarrow 0$ and converges to zero as $\tau \to \infty$. This component mostly affects short-term rates. The third factor converges to zero as $\tau \downarrow 0$ and $\tau \to \infty$, but is concave in τ. This component therefore mainly influences the medium-term rates.

Given a value for λ the factors are usually estimated by the cross-sectional regression model

$$y_t(\tau) = \sum_{i=1}^{3} B_i(\tau)\tilde{f}_{i,t} + \epsilon_t(\tau), \tag{3}$$

for $t = 1, \ldots, T$, where $B_i(\tau)$ denotes the i-th column of $B(\tau)$. At each t the data consist of yields for K different maturities, τ_j, $(j = 1, ..., K)$. The estimated time series of $\tilde{f}_t$ have empirically been shown to be strongly correlated over time, and are deemed forecastable. As such, Diebold and Li (2006) estimate the cross-sections of f_t and treat these estimates as time

Figure 1: Time-variation in λ



Note: This graph plots the rolling-window NLS estimates of λ using 100 days of data.

series to forecast future values of f_t, and thus the yield-curve, using a VAR model. Diebold et al. (2006) and Koopman et al. (2010) put the model into a state-space framework, and estimate the dynamics using a Kalman filter. For hedging purposes, the dynamics of $\tilde{f}_t$ are less important, since the aim is to obtain a portfolio that is orthogonal to the factor loadings, and hence not exposed to the factors.

In contrast to the factors $\tilde{f}_t$, λ has not gained much attention. Generally, it is set to a constant value to maximize the curve at a certain maturity, without any estimation. However, λ is a crucial parameter, as it governs the exponential decay of the second and third term, and as such exposure to the different factors across maturities. Small values of λ produce slow decay, and vice versa. We are aware of only two efforts of modeling time-variation in λ. Koopman et al. (2010) allow λ to be time-varying in a state-space framework and Creal et al. (2008) let it evolve as a linear combination of the f_t.

In this paper the time-variation of λ is of main interest. As a preliminary indication of the importance of modelling λ as a time-varying parameter we estimate the parameter on a rolling-window of 100 days using NLS, plotted in Figure 1. The data is described in detail in Section 4. The figure shows there is great variability in λ_t with typical values ranging between 0.1 and 0.8, whereas a full sample estimate yields $\lambda = 0.3834$. The factor structure appears to largely break down during the financial crisis. Not adequately modelling this leads to suboptimal fit where hedge portfolios will have unintended factor exposure.

2.1. Factor Hedging with NS

Our main empirical focus is on hedging long term liabilities. Suppose the yield curve can be described by the Nelson-Siegel model of Equation (1). The NS model implies discount factors of the form

$$\delta_t(\tau) = \exp(-\tau B(\tau)\tilde{f}_t). \tag{4}$$

These discount factors can be used for the valuation of a stream of liabilities with cash-flows $Z_t(\tau)$,

$$L_t = \int_0^{\mathcal{T}} Z_t(\tau)\delta_t(\tau)\,\mathrm{d}\tau. \tag{5}$$

The exposure of the liabilities with respect to the factors is given by

$$\frac{\partial L_t}{\partial \tilde{f}_t} = -\int_0^{\mathcal{T}} w_t(\tau)\tau B(\tau)\,\mathrm{d}\tau \equiv -D_t, \tag{6}$$

where the weight of each of the liability cash-flows is given by

$$w_t(\tau) = \frac{Z_t(\tau)\delta_t(\tau)}{L_t}. \tag{7}$$

As $B_1(\tau) = 1$, the first element in D_t is the duration of the liabilities. Hedging the level factor in Nelson-Siegel is therefore equivalent to duration hedging. This result was similarly derived in Diebold et al. (2004), and following them, we refer to D_t as the generalized duration of the liabilities. The second and third element are in this case exposures to the slope and curvature factors.

In practice hedging focuses on the first factor, ignoring the slope and curve factor, despite wide acceptance of the presence of at least three factors in interest rate curves. In our sample, the first factor explains about 82% of the variation in returns. With a second factor a total of 93% of variation is explained, and a third factor raises this to over 96%. Addressing only the first factor therefore inherently limits the hedging potential. Portfolios can easily be formed to hedge exposure to all three factors. Given that the factors are accurate, the risk of the hedge portfolio should be significantly reduced.

2.2. Estimation on yields versus returns

Since we are interested in reducing the risk of a hedge portfolio, ultimately we are interested in returns on the portfolio, not in changes in yield. We therefore propose to model factor structure of returns instead of yields. Importantly, excess returns have the same factor structure as yields.

Log-prices have the structure

$$p_t(\tau) = -\tau y_t(\tau) = -b(\tau)\tilde{f}_t, \tag{8}$$

where $b(\tau) = \tau B(\tau)$. Excess returns over a period of length h are defined as

$$\begin{aligned} r_{t+h}(\tau) &= p_{t+h}(\tau) - p_t(\tau + h) + p_t(h) \\ &= -b(\tau)\tilde{f}_{t+h} + (b(\tau + h) - b(h))\tilde{f}_t. \end{aligned} \tag{9}$$

Using the property

$$b(\tau+h)-b(h)=b(\tau)\begin{pmatrix}1 & 0 & 0\\ 0 & \mathrm{e}^{-\lambda h} & \lambda h\,\mathrm{e}^{-\lambda h}\\ 0 & 0 & \mathrm{e}^{-\lambda h}\end{pmatrix}$$
$$\equiv b(\tau)A, \tag{10}$$

where A does not depend on maturity τ, excess returns can be written as

$$\begin{aligned}r_{t+h}(\tau) &= -b(\tau)(\tilde{f}_{t+h}-A\tilde{f}_t)\\ &= b(\tau)f_t,\end{aligned} \tag{11}$$

with $f_t = -(\tilde{f}_{t+h} - A\tilde{f}_t)$. As such, returns standardized by their maturity have the same factor structure as yields

$$\rho_{t+h} \equiv \frac{r_{t+h}(\tau)}{\tau} = B(\tau)f_t. \tag{12}$$

The standardization by τ has the additional benefit of taking away a large part of the cross-sectional heteroskedasticity, since volatility of bond returns increases with maturity. We estimate all our models on these standardized returns. The fact that excess returns and yields have the same factor structure, is not just a coincidence for the NS model, but a general property of all affine term structure models. Although the transformation leaves the factor structure unchanged, it will affect the properties of the error term ϵ_t of Equation (3). When estimated on yield levels, the errors are usually strongly autocorrelated. However, the errors of the model applied to returns will most likely not exhibit autocorrelation.

With constant factor loadings a hedging problem would reduce to hedging the factor risk. If the liabilities have a factor exposure B^*, then a bond portfolio with the same factor exposure should be a perfect hedge. The need for a factor model does not, however, imply that factor loadings must be constant. Suppose the parameters λ in the factor loadings follow a stochastic process λ_t. Then the hedge portfolio should also consider changes in the factor structure. We can locally hedge using the current factors B_t, but the hedge needs to be rebalanced if factor loadings change. If they are different the return on the hedge portfolio could be exposed to factor risk.

$$\begin{aligned}r_{t+h}(\tau) &= -b_{t+h}(\tau)\tilde{f}_{t+h} - b_t(\tau)A_t\tilde{f}_t\\ &= b_t(\tau)f_{t+h} - (b_{t+h}(\tau)-b_t(\tau))\tilde{f}_{t+h}.\end{aligned} \tag{13}$$

A portfolio that is orthogonal to $b_t(\tau)$ is exposed to the change in the factor structure, except when it is still orthogonal to $b_{t+h}(\tau)$. The hedge portfolio can be updated every period depending on a change in $b_t(\tau)$. Still the hedging problem becomes far more difficult, since the time-varying factor loadings will be estimated with less precision than fixed factor loadings. For fixed factor loadings, estimation accuracy will be close to perfect with a long enough sample of reasonably frequent observations. With time-varying factor loadings estimation error becomes important. Estimation error implies that the hedge will be imperfect, because the estimated factor loadings differ from the true factor loadings, which exposes the portfolio to factor risk.

Finally, time-variation in factor loadings could also be a sign that the factor model does not have a sufficient number of factors. Suppose the factor loadings depend on a time-varying parameter λ_t

$$b_t(\tau) = b(\lambda_t, \tau). \tag{14}$$

Then we can linearise the model for excess returns as

$$\begin{aligned} r_{t+h}(\tau) &= b(\lambda_t, \tau) f_{t+h} \\ &= \tilde{b}(\tau) f_{t+h} + C(\tau)(\lambda_t - \bar{\lambda}) f_{t+h}, \end{aligned} \tag{15}$$

where C is the derivative of b_t with respect to λ_t evaluated at $\bar{\lambda}$. This is again a factor model, but it has constant factor loadings and additional factors $(\lambda_t - \bar{\lambda}) f_{t+h}$. In the Nelson-Siegel model with just a single parameter λ_t and three-factors, the linearisation would lead to a model with six factors, since all elements in f_{t+h} need to be scaled by $\lambda_t - \bar{\lambda}$. It seems much more parsimonious to model the variation in λ_t directly, instead of specifying a larger factor model.

3. DCS-NS Model

We propose a version of the Nelson-Siegel model with time-variation in λ_t, in which its dynamics are governed by a Dynamic Conditional Score (DCS) model. DCS models were first proposed in their most general form in Creal et al. (2013). The class of models is now often used empirically, and theoretical results are derived in, amongst others, Blasques et al. (2014a) and Blasques et al. (2014b).

In the class of DCS models, the dynamics of parameters are driven by the score of the likelihood with respect to that parameter and provide a natural way to update a wide variety of (latent) parameters, when a functional forms for their dynamics are not directly apparent. The DCS principle turns out to nest a lot of the models commonly used in time-series analysis. For instance, when assuming a gaussian likelihood, dynamics in the mean are given by an ARMA model, and the dynamics of volatility are given by the well-known GARCH model, which we will also use to allow for time-varying volatility.

Let r_t and ρ_t be K-vectors of (standardized) excess returns with different maturities. Consider the conditional observation density of returns $p(r_t|\lambda_t)$, and let ∇_t the score with respect to λ_t, $\nabla_t = \frac{\partial \log p(r_t|\lambda_t)}{\partial \lambda_t}$. Then the DCS model is defined as

$$\lambda_{t+1} = \phi_0(1-\phi_1) + \phi_1\lambda_t + \phi_2 s_t, \tag{16}$$

where $s_t = S_t\nabla_t$ is the score times an appropriate scaling function. Time-variation of the parameter is driven by the scaled score of the parameter and as such the dynamics of the parameters are linked to the shape of the density. Intuitively, when the score is negative the likelihood is improved when the parameter is decreased, and the DCS updates the parameter in that direction. The NS-DCS model nests a model of constant λ.

The choice of S_t delivers flexibility in how the score ∇_t updates the parameter. A natural candidate is a form of scaling that depends on the variance of the score: $S_t = \mathcal{I}_{t|t-1}^{-1}$, where $\mathcal{I}_{t|t-1} = E_{t-1}\left[\nabla_t\nabla_t'\right]$. Another option to consider is $S_t = \mathcal{J}_{t|t-1}$, where $\mathcal{J}_{t|t-1}'\mathcal{J}_{t|t-1} = \mathcal{I}_{t|t-1}^{-1}$, such that it is the square root matrix of the inverse information matrix, or no scaling at all, S_t being the identity matrix. We choose $S_t = \mathcal{J}_{t|t-1}$.

We use the DCS model to introduce time-variation in λ_t in the NS model. Let B_t denote a $(K\times 3)$ time-varying matrix with as rows time-varying versions of $B(\tau)$, where the t subscript relates to time-variation in λ_t. We consider the following specification, which we call NS-DCS:

$$\begin{aligned} \rho_t &= B_t f_t + \epsilon_t \\ \lambda_t &= \phi_0(1-\phi_1) + \phi_1\lambda_{t-1} + \phi_2 s_{t-1}, \end{aligned} \tag{17}$$

where we assume $\epsilon_t \sim N(0,\Sigma_t)$. The $(K\times K)$ error covariance matrix will be specified below. For this model we have $\nabla_t = \epsilon_t'\Sigma_t^{-1}G_t f_t$ and $\mathcal{I}_{t|t-1} = f_t'G_t'\Sigma_t^{-1}G_t f_t$, where $G_t \equiv \frac{\partial B_t}{\partial \lambda_t}$. The log-likelihood function is

$$\mathcal{L}_N = -\frac{1}{2}\ln|\Sigma_t| - \frac{1}{2}\epsilon_t'\Sigma_t^{-1}\epsilon_t, \tag{18}$$

Importantly, we do not model the time-variation in factors f_t. We simply define $f_t = (B_t'B_t)^{-1}B_t'\rho_t$. The reason for this is two-fold. First, as note before the f_t have unit roots, and as such the optimal forecast is the random walk forecast. Second, for hedging and forecasting purposes, B_t is of interest and not f_t. By concentrating f_t out of the likelihood we do not have to consider the dependence between λ_t and f_t in the observation density and in the specification of the DCS model. This greatly simplifies the model, and makes it far easier to estimate. Moreover, it makes our analysis robust against any misspecification in the time-series process of f_t.

Finally, we want to discuss the choice for a DCS model. An alternative way to model time-variation in the latent parameter is estimation by Kalman Filter of a state space model. The two are fundamentally different, and their difference is similar of Stochastic Volatility models versus GARCH-type models, where DCS models are similar to GARCH. We prefer the DCS method for a number of reasons. The main reason is that in the DCS model, λ_t

is known conditional on $\mathcal{F}_{t-1}$ as it is a deterministic function. As such, the likelihood is known analytically, making estimation straightforward. For a state-space model, λ_t is not deterministic conditional on the past, it remains unobserved. Evaluation of the log-likelihood therefore requires integrating over the path space of λ_t, and is much more computationally demanding.

Second, recent results by Blasques et al. (2014b) show that only parameter updates based on the score always reduce local Kullback-Leibler divergence between the true conditional density and the model implied conditional density. Under some regularity conditions on the smoothness of the model, it leads to considerable improvements in global KB divergence as well. DCS models are thus optimal from an information-theoretic perspective.

3.1. Time-varying volatility

The 'innovation' s_t is a function of Σ_t, the covariance matrix of ϵ_t. Keeping the covariance matrix constant or time-varying is similar to OLS versus GLS. When we allow for dynamics in Σ_t, the score is down-weighted more in high volatility times, reducing the impact of large idiosyncratic errors. As such we consider two versions of the DCS. In the base case we consider $\Sigma = \sigma^2 I_K$ as a constant. In the more general model we allow for a common GARCH-type process. In order to have a parsimonious specification we let $\Sigma_t = \sigma_t^2 I_K$, where

$$\sigma_t^2 = \omega + \alpha\sigma_{t-1}^2 + \beta\Gamma'\epsilon_{t-1}'\epsilon_{t-1}\Gamma, \tag{19}$$

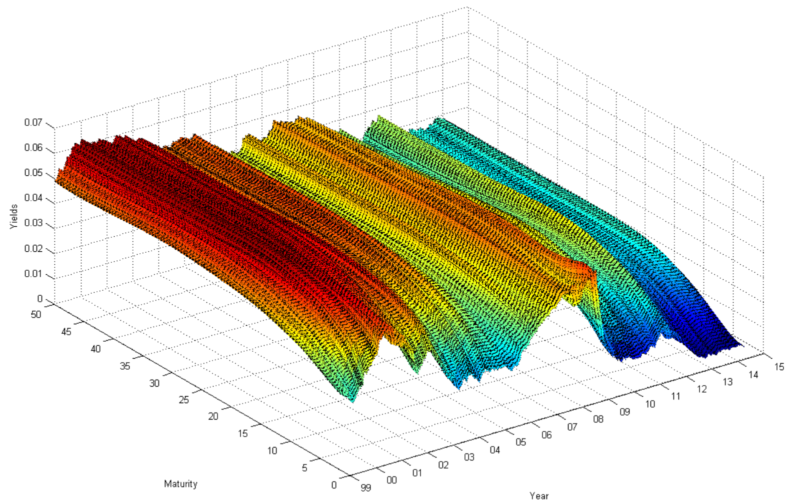
and Γ is a loading vector of the cross-sectional dimension such that $\Gamma'\Gamma = 1$. The loading vector may be estimated or, as we do, set to equal weights for each maturity, i.e. $\Gamma = \iota/\sqrt{K}$. The volatility dynamics of all maturities are thus governed by a single GARCH process. Any full multivariate model would quickly suffer from the curse of dimensionality as the cross-sectional dimension typically exceeds ten or fifteen. This leads to two self-explanatory additional models, the NS-GARCH and NS-DCS-GARCH.

4. Empirical Analysis

4.1. Data

We obtain daily data on euro swap interest rates with maturities 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30, 40 and 50 years from Datastream. Data is available from January 1999, except for the 40 and 50 years rates, which are available form March 2005 onwards. All rates are ask rates quoted at the close of London business. Euro swap rates are quoted on an annual bond 30/360 basis against 6 months Euribor with the exception of the 1 year rate which is quoted against 3 months Euribor. The main advantage of using swaps is that they are available and relatively liquid until long maturities, which allows us to verify the

Figure 2: Yield Curves



Note: This figure plots the yield curve of our euroswap data evolving over time for the different maturities.

accuracy of the model for long-term liabilities.[2] The series are interpolated using a local cubic interpolation to obtain rates with annual intervals between maturities. We then use a bootstrap method to construct the yield curve. The resulting panel of yields is plotted in Figure 2.

The next step is to compute returns. We first obtain the price from the yields as

$$P_t(\tau) = e^{-\tau y_t(\tau)} \tag{20}$$

and obtain returns as 100 times the log-difference over time. We subtract the risk-free rate defined as $r_t^f = y_t(1)/252$, the shortest-term yield.

In the remainder of the paper we will estimate the model on two different samples. First we use the full cross-section of seventeen maturities, and second we use a limited cross-section, with only the maturities up until twenty years. There are good reasons for excluding maturities longer than twenty years. The most important reason is the general assumption that the so-called 'last liquid point' of the market rates is at the twenty year maturity.[3] The prices of bonds with maturities greater than this are severely affected by their limited liquidity, which significantly impacts the yield curve and therefore estimates of the NS-model.

[2]On April 1st 2004 there is almost surely a data error for the 6Y rate, which is exactly equal to the 5Y rate. This observation distorts all estimations and is replaced with an interpolated value.

[3]This is the point assumed by the European Insurance and Occupational Pensions Authority (EIOPA)

Table 1: In-Sample Parameter Estimates

	Full Cross-Section				Limited Cross-Section			
	NS	NS-DCS	NS-GARCH	NS-DCS-GARCH	NS	NS-DCS	NS-GARCH	NS-DCS-GARCH
λ	0.3834		0.5502		0.5742		0.6372	
	(0.0446)		(0.0204)		(0.0228)		(0.0233)	
ϕ_0		0.4108		0.4161		0.5799		0.6531
		(0.0262)		(0.0221)		(0.0130)		(0.0278)
ϕ_1		0.9888		0.9859		0.9659		0.9614
		(0.0955)		(0.0759)		(0.0759)		(0.0716)
ϕ_2		0.0345		0.0208		0.0288		0.0145
		(0.0188)		(0.0102)		(0.0082)		(0.0045)
$\sigma^2(\times 10^5)$	6.7973	6.3411			5.0114	4.9419		
	(0.6929)	(0.6809)			(0.5587)	(0.5349)		
$\omega(\times 10^7)$			6.1422	6.5799			7.0818	5.0953
			(8.3806)	(5.2184)			(6.2002)	(4.8337)
α			0.8737	0.8760			0.8794	0.8616
			(0.1053)	(0.0841)			(0.1067)	(0.0612)
β			0.1226	0.1193			0.1191	0.1346
			(0.0987)	(0.0755)			(0.0935)	(0.0602)
LL$(\times 10^{-5})$	2.2720	2.2954	2.4945	2.5029	1.8158	1.8194	1.9748	1.9998

Note: This table provides in-sample parameter estimates for the four different models, both on the full cross-section of all seventeen maturities, as well as the model estimated on the limited cross-section which only includes bonds up until maturity of twenty years. Newey-West standard errors (22 lags) in brackets.

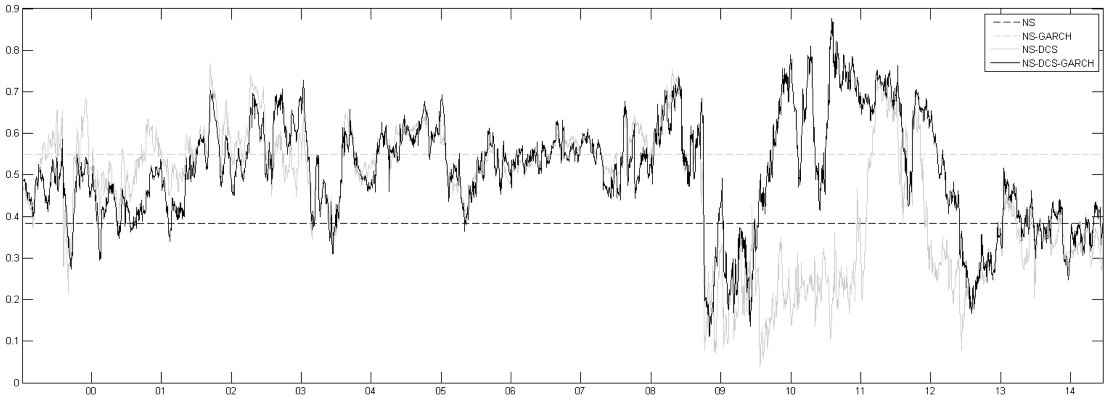
4.2. In-sample Results

Estimation results are reported in Table 1. The results of the full and limited cross-section are qualitatively similar, apart from the limited cross-section having higher estimates of λ, which puts more exposure on short maturities. On the full cross-section the standard NS model the shape parameter is estimated at $\lambda = 0.3834$, putting the maximum of the curve factor between 4 and 5 years. Due to the presence of heteroskedasticity, the NS-GARCH has a far larger estimate, 0.5502, which maxes the curve out between 3 and 4 years. A quick look at the rolling window estimates of Figure 1 explains the difference between the constant and GARCH volatility models, as λ is lowest during the financial crisis when volatility was high.

Next consider the DCS models. The DCS models improve over their standard counterparts: both the NS-DCS and NS-DCS-GARCH have a much larger value for the log-likelihood. The λ's are highly persistent with autoregressive coefficient over 0.98. Importantly, the coefficient of the score innovation, ϕ_2, is significantly different from zero. The models therefore produce significant dynamics in λ.

To illustrate the dynamics of the DCS model we plot the estimated λ_t in Figure 3 for the different models on the full cross-section. The main series of interest are the λ_t estimates of

Figure 3: Time Series of λ_t for various models.



the NS-DCS and NS-DCS-GARCH. Interestingly, up until the second half of 2008, volatility does not play a large role and the fitted λ_t of both models are very similar, and more over, variation is limited. However, during the crisis period, the two models completely diverge. The NS-DCS' estimates of λ go down, with an absolute minimum at 0.04, which effectively reduces the model to two factors. Interestingly, the λ_t implied by the NS-DCS-GARCH actually increases to levels not seen in the entirety of the first half of our sample. After most of the volatility has died down, the two DCS models converge again.

5. Hedging long-term liabilities

The aim of this section is to use the various forms of the Nelson-Siegel model introduced in this paper for a hedging exercise. We use the four models highlighted in the previous section, and contrast their results with the hedging performance of simple duration hedging, in which we only hedge the level factor.

5.1. Setup

Let w_τ denote the weight in bond with maturity τ. We consider the following problem at any point in time t:

$$w_{50} = -1 \tag{21a}$$

$$\sum_i w_i = 0 \tag{21b}$$

$$\sum_i w_i \tau_i B_k(\tau_i) = 0 \quad \forall k = 1, 2, 3 \tag{21c}$$

$$\min w' \mathcal{T}' \Sigma_t \mathcal{T} w, \tag{21d}$$

where $\mathcal{T}$ is the vector of maturities τ_i. Line a shows a short position in the long-maturity bond. Line b requires the portfolio to be self-financing, and line c implies zero factor exposure of the portfolio. If the factor loadings are estimated correctly, that is, λ_t is equal to the true λ_t, the only remaining risk is the idiosyncratic risk of the individual bonds. Finally, line d chooses the portfolio with minimum idiosyncratic risk out of the infinite set of portfolios satisfying the first three constraints. Note that τ re-enters in the last two equations, as we now look at the true returns, not those standardized by τ.

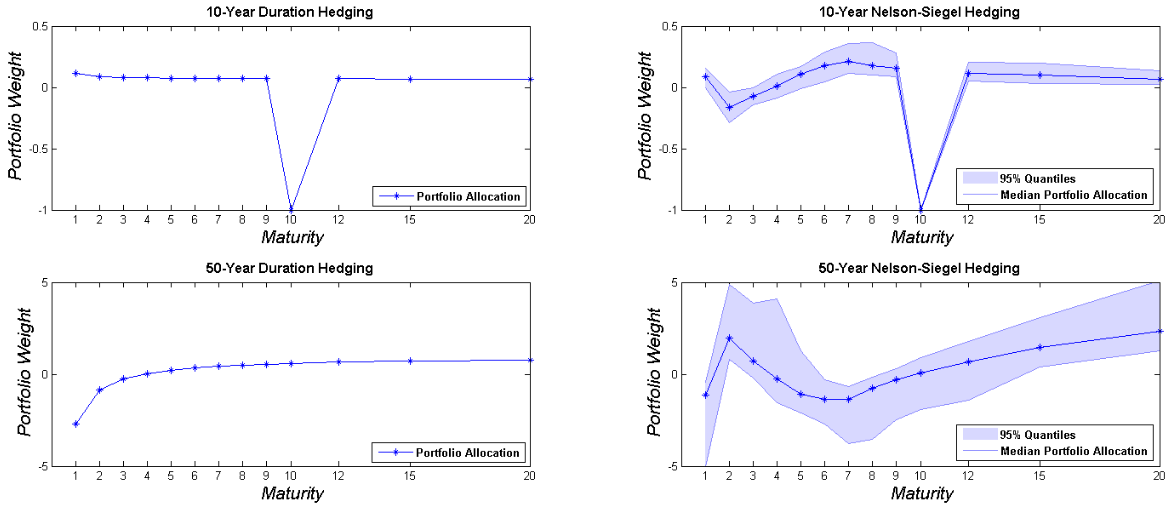
The hedging of long-term liabilities is of great interest for practitioners such as pension funds, but also a great illustration of the qualities and importance of accurately modeling the factor structure. Hedging long-term liabilities is a much more intricate problem then short to medium term liabilities. To hedge a ten-year bond, taking an opposite position in a number of bonds with close maturity will work quite well as their factor loadings are similar to the ten year bond. For long-term bonds, such as the fifty year bond, no liquid bonds with close maturity exist, making this infeasible.

Hence, the 50-year bond needs to be hedged using large positions in bonds with much shorter maturities. In terms of the Nelson-Siegel model, the 50-year maturity bond essentially only loads on the level factor. However, the slope and curve factor play a crucial role in hedging the long-term liability, since the portfolio of bonds used to hedge the long-position do load significantly on the slope and curve factor. These shorter bonds have to be chosen carefully, since the overall hedge portfolio also needs to be neutral to these factors.

To illustrate the three-factor hedge for short and long-term liabilities we plot the empirical distribution of the ten- and fifty-year hedge portfolios in Figure 4. The top graphs depict the weights in the ten-year hedge portfolio for both duration hedging and the portfolios that hedge all three NS factors, where λ_t is derived form the NCS-DS model. The portfolios do not differ much, as the factor loadings change gradually over the neighboring maturities, and neutralizing the slope and level factor requires a marginal change in the portfolios. Moreover, there is very little time-variation as changes in λ only affect the second and third factor, which have limited impact on portfolio composition.

On the other hand, the dynamics in the 50-year bond hedge are far more extreme. First the duration hedging portfolio takes a large positive position in the bond with closest maturity (20 in this case), but spreads out the weight over different maturities to diversify away as much of the idiosyncratic risk as possible. The three-factor hedge for the 50-year maturity shows that portfolio composition is in fact largely determined by the need to neutralize the second and third factor. Moreover, there is a lot more time-variation in the portfolio composition, as the changing lambda mainly only affects the slope and curve factors, which play a limited role for short-term hedges.

Figure 4: Hedging Long-Term Liabilities



Note: The graphs depict the importance of hedging all factors for long-term hedging. The top graphs show the portfolio allocation for hedging a bond with 10-year maturity, and the bottom graphs show the portfolio allocation for hedging a 50-year bond. The two graphs on the left show Duration hedging and the right graphs show the results of hedging all three factors using λ_t from the NCS-DS model. For the three-factor hedge we show the empirical distribution of the portfolio weights over time.

5.2. Hedging Results

We set up two out-of-sample forecasting exercises. In both we re-estimate the parameters of the model every Friday using a rolling window of 1000 observations, and use these parameters to make daily forecasts of λ_t. These forecasts are used to solve the system of equations of the minimum risk hedge portfolio given by Equations (26a-d). However, in one set we use the full cross-section of the 17 maturities including the 50 year maturity itself. In the second exercise we use only the 13 maturities up until and including 20 years. This is an economically important restriction, as bonds with longer maturities are not as liquid, and the market is simply not deep enough to accommodate the needs of pension funds. Here the 50 year maturity is out-of-sample both in terms of the time-series and the cross-sectional dimensional. In this case we do not estimate the full covariance matrix which makes minimizing (21d) infeasible. Instead, we minimize the variance of the portfolio of the first 13 maturities, with their combined weight standardized to one.

We report the mean absolute hedging error (MAE) as well as the variance of the hedging error (Var). The better the hedge, the smaller these two quantities will be. We also report the mean hedging error (Mean). If the mean is positive, the hedging portfolio is earning a higher average return than the target asset which is the more desirable of deviations from zero. The results are reported in Table 2. We report results for the full out-of-sample period, as well as three sub-periods, corresponding to the periods before, during and after the financial crisis. The top half shows the results for the hedging portfolios constructed using λ forecasts based on the sample including the maturities over twenty years, whereas the bottom panel is based on forecasts using only data with maturities up to twenty years.

The results for the full and partial cross-section are qualitatively similar, and models on the full cross-section perform slightly better. All models have average return close to zero, so none of the models has systematic bias. Therefore, what matters for hedging purposes is the variance of the hedging errors. The MAE and portfolio return variance show differences across the different models. For instance, looking at the full cross-section on the full sample, we see that the variance of hedging errors is 0.58 for duration hedging, 0.37 for the standard Nelson-Siegel and 0.26 for the DCS version. To put things into perspective, the actual variance of the 50-year maturity return is 5.30 over this period. Duration hedging therefore takes out 89% of the variance and the NS improves by hedging 93% of the variance. Importantly, the NS-DCS has additional non-trivial improvements, and hedges 95% of the variance. Because of the poor performance of standard duration hedging, we will not discuss it further.

For the standard NS, the inclusion of time-varying volatility does not necessarily lead to superior performance. The NS-GARCH performs better in the early sample, whereas the NS model performs better in the latter part of the sample. This is in line with the value of λ implied by our DCS models in Figure 3. There the DCS-NS shows that λ is high in early sample, and decreases during the crises. The NS-GARCH estimate of λ is typically higher than the NS estimate, and therefore more accurate in the beginning of the sample.

The DCS models improve on the standard models without exception. Overall, the NS-DCS-(GARCH) lower the MAE and variance by approximately 30% compared to the NS(-GARCH). This is clearly an economically significant improvement. The NS-DCS-GARCH performs slightly better in the early sample and during the financial crisis, while the NS-DCS beats the time-varying volatility variant in the final part of our sample. A possible explanation for the latter result is that yields have hit the zero-lower bound in recent years, and variance is artificially low.

In the first part of the sample DCS improvements are only about 20%, while during the crisis improvements are 40% to 50%, and they are still about 30% after the crisis. As would be expected, variation in λ is greater in highly dynamic markets, and the DCS models offer the greatest improvements over the constant λ models. However, even in tranquil times the DCS offers improvements over the constant model.

Although the in-sample likelihood improvement of time-varying volatility is large, for our empirical application the gains are small. The reverse is true for the time-variation in λ, where the DCS model offers smaller likelihood improvements, but the empirical applications shows that benefits can be very large.

Table 2: Daily hedging error descriptives

	Duration Hedging	NS	NS-DCS	NS-GARCH	NS-DCS-GARCH
			Full Cross-Section		
			Full Sample		
Mean	-0.0024	0.0007	-0.0021	0.0004	0.0020
MAE	0.4053	0.3091	0.2009	0.2964	0.1997
Var	0.5876	0.3691	0.2570	0.3366	0.2215
			2003-2007		
Mean	0.0061	0.0039	0.0039	0.0035	0.0086
MAE	0.1796	0.1595	0.1259	0.1537	0.1211
Var	0.1577	0.1654	0.1279	0.1563	0.1064
			2008-2009		
Mean	0.0165	0.0097	-0.0062	0.0140	0.0049
MAE	0.8106	0.5531	0.3673	0.5829	0.2927
Var	1.7106	0.8066	0.5052	0.8873	0.4167
			2010-2014		
Mean	-0.0207	-0.0070	-0.0070	-0.0093	-0.0068
MAE	0.4817	0.3708	0.2652	0.3830	0.3035
Var	0.5770	0.4067	0.2848	0.4262	0.3736
			Limited Cross-Section		
			Full Sample		
Mean	0.0031	0.0004	0.0042	0.0003	0.0019
MAE	0.4284	0.3174	0.2165	0.3193	0.2333
Var	0.6234	0.3892	0.3281	0.3947	0.3217
			2003-2007		
Mean	0.0056	0.0035	0.0036	0.0034	0.0085
MAE	0.1976	0.1529	0.1238	0.1511	0.1282
Var	0.1899	0.1553	0.1250	0.1533	0.1047
			2008-2009		
Mean	0.0211	0.0145	0.0031	0.0154	0.0116
MAE	0.8618	0.5867	0.3646	0.5944	0.3204
Var	1.8961	0.8992	0.6519	0.9245	0.5141
			2010-2014		
Mean	-0.0196	-0.0095	0.0054	-0.0101	-0.0100
MAE	0.5247	0.3847	0.2913	0.3883	0.3163
Var	0.6110	0.4288	0.3516	0.4341	0.3967

Note: This table gives descriptives of the hedging performance of the different models in sub-periods. The Mean is the average hedging error, the MAE is the mean absolute hedging error, and the Var is the variance of the hedge portfolio returns. The top panel shows results when the models are estimated on the full cross-section of maturities, and all are used to hedge the 50-year maturity, the bottom panel shows the results when optimization occurs over assets with a maximum maturity of 20 years.

6. Conclusion

We have proposed a Dynamic Conditional Score version of the Nelson-Siegel in which time-variation of the shape-parameter λ is modeled. The parameter is of great importance as it completely specifies the factor loadings of bonds at different maturities. We show that the time-variation in the parameter is large, and assuming it constant is too limiting. In-sample estimation shows the parameter varies between 0.1 and 0.9, and even within a year the parameter has a range of up to 0.6. We document that the parameter is relatively stable in the beginning of the sample, but becomes highly volatile during the financial crisis.

In our empirical application, we use the DCS version of the NS model to hedge long-term liabilities using euro-swap data. We find that allowing for time-variation in λ leads to significant improvements in hedging performance, greatly reducing the mean absolute hedging error and hedge portfolio variance. Specifically, hedging error variance is reduced by an additional 50% compared to the NS model's improvement over duration hedging. The variation in the shape parameter is of far greater importance than taking into account time-varying volatility, which only offers small improvements, both in the constant and DCS λ setting.

Bianchi, F., Mumtaz, H., Surico, P., 2009. The great moderation of the term structure of UK interest rates. Journal of Monetary Economics 56 (6), 856–871.

Blasques, F., Koopman, S. J., Lucas, A., 2014a. Maximum likelihood estimation for generalized autoregressive score models. Tech. rep., Tinbergen Institute.

Blasques, F., Koopman, S. J., Lucas, A., 2014b. Optimal formulations for nonlinear autoregressive processes.

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.

Creal, D., Koopman, S. J., Lucas, A., 2008. A general framework for observation driven time-varying parameter models. Tech. rep., Tinbergen Institute Discussion Paper.

Creal, D., Koopman, S. J., Lucas, A., 2013. Generalized autoregressive score models with applications. Journal of Applied Econometrics 28 (5), 777–795.

Dai, Q., Singleton, K. J., 2000. Specification analysis of affine term structure models. The Journal of Finance 55 (5), 1943–1978.

Diebold, F. X., Ji, L., Li, C., 2004. A three-factor yield curve model: non-affine structure, systematic risk sources, and generalized duration. Working Paper.

Diebold, F. X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of econometrics 130 (2), 337–364.

Diebold, F. X., Rudebusch, G. D., 2013. Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach. Princeton University Press.

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

Duffee, G. R., 2002. Term premia and interest rate forecasts in affine models. The Journal of Finance 57 (1), 405–443.

Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical finance 6 (4), 379–406.

Hautsch, N., Ou, Y., 2012. Analyzing interest rate risk: Stochastic volatility in the term structure of government bond yields. Journal of Banking & Finance 36 (11), 2988–3007.

Koopman, S. J., Mallee, M. I., Van der Wel, M., 2010. Analyzing the term structure of interest rates using the dynamic nelson–siegel model with time-varying parameters. Journal of Business & Economic Statistics 28 (3), 329–343.

Nelson, C. R., Siegel, A. F., 1987. Parsimonious modeling of yield curves. Journal of business, 473–489.

Appendix A. Derivations of the Score and Information

In this section we show the derivations of the score and information which are used for the DCS model in (17). Let $\epsilon_t = y_t - B_t f_t$. Assume $\epsilon_t \sim N(0, \Sigma_t)$. Then the likelihood is

$$\mathcal{L}_N = -\frac{1}{2}\ln|\Sigma_t| - \frac{1}{2}\epsilon_t'\Sigma_t^{-1}\epsilon_t. \tag{A.1}$$

Defining $G_t \equiv \frac{\partial B_t}{\partial \lambda_t} = \begin{pmatrix} 0 & \frac{-b_{t,3}}{\lambda_t} & \tau b_{t,2} - (\tau + \frac{1}{\lambda_t})b_{t,3} \end{pmatrix}$ and $H_t \equiv \frac{\partial^2 B_t}{\partial \lambda_t^2}$ the scores are

$$\begin{aligned} \nabla_t(\lambda_t) = \frac{\partial \mathcal{L}_N}{\partial \lambda_t} &= r_t'\Sigma^{-1}G_t f_t - f'B_t'\Sigma^{-1}G_t f_t \\ &= \epsilon_t \Sigma^{-1} G_t f_t. \end{aligned} \tag{A.2}$$

The second order derivatives are

$$\begin{aligned} \frac{\partial^2 \mathcal{L}_N}{\partial \lambda_t^2} &= r_t\Sigma^{-1}H_t f_t - f_t'G_t'\Sigma^{-1}G_t f_t - f_t'B_t'\Sigma^{-1}H_t f_t \\ &= \epsilon_t\Sigma^{-1}H_t f_t - f_t'G_t\Sigma^{-1}G_t f_t \end{aligned} \tag{A.3}$$

from which we find the information matrix as

$$\begin{aligned} \mathcal{I}_{t|t-1}(\lambda_t) &= -\mathbb{E}\left(\frac{\partial^2 \mathcal{L}_N}{\partial \lambda_t^2}\right) \\ &= f_t'G_t'\Sigma^{-1}G_t f_t. \end{aligned} \tag{A.4}$$

and thus

$$\mathcal{J}_{t|t-1}(\lambda_t) = \sqrt{\frac{1}{f_t' G_t' \Sigma^{-1} G_t f_t}}. \tag{A.5}$$