Journal of Financial Econometrics, 2020, 1–34
doi: 10.1093/jjfinec/nbaa027
Article

Hedging Long-Term Liabilities*

Rogier Quaedvlieg[1] and Peter Schotman[2]
[1]Erasmus University Rotterdam and [2]Maastricht University

Address correspondence to Rogier Quaedvlieg, Department of Business Economics, Erasmus School of Economics, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands, or e-mail: quaedvlieg@ese.eur.nl.

Received November 7, 2018; revised July 16, 2020; editorial decision July 17, 2020; accepted July 23, 2020

Abstract

Pension funds and life insurers face interest rate risk arising from the duration mismatch of their assets and liabilities. With the aim of hedging long-term liabilities, we estimate variations of a Nelson–Siegel model using swap returns with maturities up to 50 years. We consider versions with three and five factors, as well as constant and time-varying factor loadings. We find that we need either five factors or time-varying factor loadings in the three-factor model to accommodate the long end of the yield curve. The resulting factor hedge portfolios perform poorly due to strong multicollinearity of the factor loadings in the long end, and are easily beaten by a robust, near Mean-Squared-Error- optimal, hedging strategy that concentrates its weight on the longest available liquid bond.

Key words: factor models, risk management, term structure

JEL classification: G12, C32, C53, C58

Pension funds and (life-)insurance companies face high interest rate risk stemming from the duration mismatch between their assets and liabilities. Liabilities are often long-dated, whereas the duration of liquid fixed-income assets in the market is much shorter. For example, for German life insurers, Domanski, Shin, and Sushko (2017) estimate the duration of liabilities at over 25 years.[1] Funds seek to hedge this risk, at least partially. Hedging the interest rate is far from straightforward due to the sheer volume of liabilities. In Europe, the European Central Bank estimates the amount to be over €8 trillion in 2016.[2] Without a similar volume in long-maturity liquid assets, a simple immunization strategy is impossible.

* Part of this research was financially supported by a grant from the Global Risk Institute.

1 In the Netherlands, a report of the Dutch Central Bank (DNB Bulletin 2013) shows that pension funds have liabilities that can be as long as 80 years.

2 ECB, Pension funds and insurance companies online, September 2016, Table *I.I*. www.ecb.europa.eu/press/pdf/icpf/icpf16q2.pdf

A natural alternative is to use a factor model for bond returns and construct a hedge portfolio with the same factor exposure as the liabilities. We estimate several variations and extensions of the factor model using bond returns with maturities up to 50 years. Based on the factor loadings we then construct factor-mimicking portfolios consisting of liquid bonds with maturities up to 20 years. From the factor-mimicking portfolios we design a hedge portfolio to eliminate the factor risk in the longer maturities. Our analysis is in the tradition of Litterman and Scheinkman (1991).

For our hedging purposes, the Nelson–Siegel (NS) model is an appealing candidate factor model.[3] Nelson and Siegel (1987) proposed it as a purely statistical model with good empirical fit of the yield curve, but its theoretical foundation within the general class of affine models (Dai and Singleton 2002) has since been established (Krippner 2015). The three factors in NS model have an intuitive interpretation as level, slope, and curvature. The level factor is also consistent with duration hedging, and therefore nests the most often used approach for interest rate hedging in practice. Diebold, Ji, and Li (2006a) proposed "generalized duration hedging," which amounts to hedging all three NS factors.

Importantly, the shape of the factor loadings in the NS model is governed by a single parameter λ, which determines both the steepness of the slope factor loadings and the convexity of the curvature factor loadings. Correspondingly, implied factor loadings for long maturities can easily be inferred, even if no data are available. Subsequent work by Christensen, Lopez, and Mussche (2019) similarly uses this feature of NS models to extrapolate yield curves and find that the model performs reasonably well in this context.

Our emphasis on very long maturities differs from much of the empirical term structure literature, which usually focuses on maturities between 0 and 10 years. For our data on euro swap rates with maturities ranging from 1 to 50 years, the three NS factors explain more than 95% of the variation in daily bond returns. Again, the NS model is an appealing candidate for modeling very long rates. Contrary to many affine models, the very long rate converges to a level close to the observed long-term interest rate at time t and not to an overall historical constant. Nevertheless, the basic NS model performs far from excellent in hedging the interest rate risk at very long maturities. Neither the duration factor on its own nor the full three-factor NS model can outperform a simple leveraged position in a 20-year bond. This non-diversified fixed-income portfolio, targeting a single maturity, provides more out-of-sample risk reduction than the NS factor model. Much of the article aims at explaining this result.

We investigate three different explanations for the NS model's relative poor performance. First, we consider additional factors, beyond the standard three NS factors, to fit the very long end of the yield curve. Second, we study the effect of estimation noise on portfolio weights. Third, we consider time variation in the factor loadings. All three approaches appear relevant and add new insights to the literature on NS models.

Our first explanation revolves around additional factors. Christensen, Diebold, and Rudebusch (2009) have previously proposed a five-factor NS model for modeling maturities up to 30 years. Dubecq and Gourieroux (2011) also find that a three-factor NS model does not suffice for long-term bonds with maturities longer than 10 years. In related work, Almeida et al. (2018) and Faria and Almeida (2018) propose "segmented" term structure

3 See Diebold and Rudebusch (2013) for a textbook analysis of the econometric implementation and empirical evidence for the NS model.

models that allow different dynamics for different parts of the curve, such as short-term and long-term yields. For our euro swap rates, the five-factor estimates result in two sets of distinct slope and curvature factors: one with a small λ to fit the very long end of the term structure, and the other with a large value for λ accounting for the variation at the short end. The five-factor model fits the data significantly better than the standard three-factor model. However, its hedging performance is the worst by a large margin.

The poor hedging performance of the five-factor model, despite its good in-sample fit, calls for further analysis, which leads to our second explanation. The hedge portfolio is nothing but a prediction model, which aims at predicting the return on very long maturity bonds using returns on liquid bonds as predictors. As in every prediction model, there is a tradeoff between bias and variance. From this perspective, adding more factors reduces the bias but increases the variance. On inspecting bias and variance for the five-factor model, we find that it has a huge variance. The factor returns for the "low λ" factors in the five-factor model are relevant for longer maturities, but they are hard to estimate from the liquid returns. As a result, the five-factor hedge portfolio becomes very erratic. The three-factor model has some bias, but much smaller estimation error. Generally, simpler models will have even less variance, but more bias. We derive the portfolio with the optimal bias/variance trade-off under squared error loss which turns out to be very similar to our naive benchmark hedge portfolio.

Finally, we investigate to what extent allowing for time-varying factor loadings increases the performance of the NS-based hedge portfolios relative to the simple duration-matched position. A three-factor model with time-varying factor loadings can mimic a low and high-λ regime. This interpretation of a five-factor model as a proxy for a three-factor model with time-varying factor loadings, was similarly used in Koopman, Mallee, and Van der Wel (2010), who modeled time variation in λ using a Kalman filter. We model the time-varying shape parameter using a Dynamic Conditional Score (DCS) specification. DCS is a modeling principle, which proposes to update parameters in the direction of their likelihood gradient for a new observation. Creal, Koopman, and Lucas (2013) explore the principle for a variety of models. For the NS shape parameter λ, the DCS principle finds a linear combination of the current residuals as a direction to improve the fit of the model in the next period. We estimate the optimal adjustment parameter to obtain a smoothly adapting time series of λ_t's.

For our European data, we find considerable time variation in the shape parameter λ and a much improved fit. Factor loadings have dramatically changed after the start of the financial crisis. To a large extent, the filtered paths of λ_t can be characterized by low and high λ "regimes," whose average level corresponds to the two estimates we obtain in the five-factor version, further confirming the link between the specifications. Importantly, however, the time-varying parameter three-factor version does not suffer from identification difficulties plaguing the performance of the five-factor model. The time variation has strong effects on the allocation within the hedge portfolios, and leads to improved hedging performance. Regardless, the performance still falls short of the naive hedging strategy using just the 20-year bond.

With this conclusion, we have come almost full circle. Starting from a factor model, we end up with a hedging strategy that is like immunization: invest in the asset that is most similar to the liabilities. Importantly, this hedging strategy is consistent with multifactor empirical term structure models, either a five-factor model, or a three-factor model with time-varying loadings.

The remainder of the article consists of two main parts. In the first part, Sections 1 and 2, we describe the hedging problem and introduce our adaptations to the NS model to make it suitable for this purpose. The second part, Sections 3 to 5 present our empirical findings. The first presents the data. The next two sections analyze the hedging performance of the models with constant and time-varying factor loadings respectively. Finally, Section 6 describes how to extend the analysis to hedging portfolios of long-dated liabilities and Section 7 concludes.

1 Hedging Long-Dated Liabilities

1.1 Hedge Portfolios

Suppose at time t a fund has a liability to pay one euro at time $T_0 = t + \tau_0$ in the distant future. If long-dated fixed income instruments exist, the fund could hedge this position by buying discount bonds with maturity τ_0. For long maturities, $\tau_0 > 50$ years, such instruments typically do not exist, while for medium-term maturities, with $20 < \tau_0 < 50$ years, the market is insufficiently liquid. Therefore, the fund needs to hedge its liability using instruments with shorter maturities $\tau_i < \tau_0$.

The most common hedge with fixed income securities is a duration hedge. The duration hedge is a bond portfolio with a duration equal to the maturity of the liability. For a portfolio consisting of n discount bonds with maturities τ_i, the duration of a portfolio with weights w_i is just $D = \sum_i w_i \tau_i$, a weighted average of the maturities of the discount bonds. Since the liability duration is longer than all of the traded instruments, the duration hedged position inevitably involves leverage, that is, some weights must be negative. The popularity of duration hedging derives from the property that duration measures the relative change in the value of the portfolio with respect to a small parallel movement of the yield curve. Duration hedging applies to changes over short intervals. For longer intervals, the nonlinearity of the price–yield relation introduces second-order convexity terms.

In our empirical work, we will consider a 50-year liability and assume that 20 years is the longest liquid maturity. The 20-year maturity corresponds with the assumed last liquid point in the European Insurance and Occupational Pensions Authority (EIOPA) regulations for the euro fixed income market.[4] We take the 50-year maturity as our target maturity since shorter maturities in the range 20–50 years will be easier to hedge, as they are closer to the available liquid maturities. For maturities longer than 50 years, we do not have data to evaluate the performance. In practice, there will be a portfolio of liabilities at different maturities, which we will discuss in Section 6.

As the simplest possible duration hedge, we consider a "naive" hedge that is long in a 20-year bond and short in the instantaneous riskfree asset, with weights equal to + 2.5 and –1.5, respectively. Its duration is $D = 2.5 \times 20 - 1.5 \times 0 = 50$ years. The hedge is robust in the sense that it does not require any estimation of interest rate dynamics and volatility. We call it naive because it ignores many empirical regularities. It implicitly assumes that

4 EIOPA (European Insurance and Occupational Pensions Authority) is an EU agency for supervision and regulation of the P&I sector.

changes in the 20-year and 50-year interest rates are perfectly correlated and have equal volatility. Both are crude approximations to the empirical data.

For a general portfolio of (discount) bonds, the difference between the asset portfolio A_t and the liability L_t evolves as

$$\begin{aligned}\frac{\mathrm{d}A}{A}-\frac{\mathrm{d}L}{L}&=\sum_i w_i\frac{\mathrm{d}P_i}{P_i}+\Big(1-\sum w_i\Big)y^f\,\mathrm{d}t-\frac{\mathrm{d}P_0}{P_0}\\&=\sum_i w_i\Big(\frac{\mathrm{d}P_i}{P_i}-y^f\,\mathrm{d}t\Big)-\Big(\frac{\mathrm{d}P_0}{P_0}-y^f\,\mathrm{d}t\Big),\end{aligned} \tag{1}$$

where $y^f=y_t(0)$ as the instantaneous riskfree rate, and where P_i is a shorthand notation for $P_t(\tau_i)$, the price at time t for a discount bond with maturity τ_i. To define a hedge we assume that excess returns are generated by

$$\frac{\mathrm{d}P_i}{P_i}-y^f\,\mathrm{d}t=\mu_i\mathrm{d}t+B_i\mathrm{d}W,\quad (i=0,\ldots,N), \tag{2}$$

with μ_i the expected excess return, $\mathrm{d}W$ a k-vector of Brownian motions, and B_i a row vector of exposures to the risk factors. Both μ_i, B_i, as well as the riskfree rate y^f may be time-varying depending on a vector of state variables. A perfect hedge can be constructed if we can find weights w_i such that

$$\sum_i w_iB_i=B_0. \tag{3}$$

Diebold, Ji, and Li (2006a) use the term "generalized duration" for the liability loadings B_0. Since such a portfolio is instantaneously riskfree, absence of arbitrage implies that it must also have a zero expected excess return. Standard asset pricing then implies that the drifts in Equation (2) must satisfy $\mu_i=B_i\nu$ for a k vector of risk prices ν. This leads to the factor model

$$\frac{\mathrm{d}P_i}{P_i}-y^f\,\mathrm{d}t=B_i(\nu\mathrm{d}t+\mathrm{d}W),\quad (i=0,\ldots,N). \tag{4}$$

Due to the interpretation of the model within an asset pricing framework, hedge portfolios based on the factor structure have an economic meaning and are not attempting to exploit some hidden and perhaps spurious arbitrage opportunity.

If $k>n$, and B_i are linearly independent, it will generally be impossible to construct a perfect hedge. Much of the empirical term structure literature assumes that bond prices are governed by $k=3$ factors. We would then only need $n=3$ instruments (plus the risk-free rate) for a perfect hedge. When more instruments are available, any choice that satisfies Equation (3) is equivalent assuming that the factor model is exact. In practice, model and measurement errors imply deviations from a strict factor model. In particular, the long-term liability may not be fully consistent with the model due to the reduced liquidity at the long end of the market. In the econometric model, we will allow model and measurement errors and measure returns over discrete time intervals.

1.2 Factor Hedge Portfolios

For an econometric model we define returns over discrete time intervals of length h equal to one day. Let $p_t(\tau)=-\tau y_t(\tau)$ denote log price of a discount bond with yield $y_t(\tau)$. In particular, $p_t(h) = -hy_t(h)$ is minus the return on the risk-free asset. Excess returns of a τ-maturity bond over the risk-free rate for a period of length h are defined as

$$r_{t+h}(\tau) = p_{t+h}(\tau) - p_t(\tau + h) + p_t(h). \tag{5}$$

The factor model (4) refers to simple returns. The log-returns defined in Equation (5) will be convenient later on, when we specify the NS model. The average logarithmic and simple returns differ by a term $\frac{1}{2}\mathrm{Var}_t[r_{t+h}(\tau)]$, which is small for a one-day horizon. We correct for it in the empirical work.

Based on a preliminary analysis, and the extant literature, we know that the volatility of bond returns increases almost linearly with maturity. The definition $p_t(\tau) = -\tau y_t(\tau)$ also suggests that returns are proportional to maturity if yields move in parallel. In our econometric model, we therefore scale the returns by their maturity, before adding an error term, that is, we define the scaled excess returns $\rho_t(\tau) = r_t(\tau)/\tau$ and consider the factor model

$$\begin{pmatrix} \rho_{0t} \\ \rho_t \end{pmatrix} = \begin{pmatrix} b_0 \\ b \end{pmatrix} f_t + \begin{pmatrix} \epsilon_{0t} \\ \epsilon_t \end{pmatrix}, \tag{6}$$

where the $(n \times k)$ matrix b contains the factor loadings for the liquid maturities, and b_0 is the $(1 \times k)$ vector of factor loadings for the target maturity. Rows of b are denoted b_i, and are defined as $b_i = B_i/\tau_i$ to match the scaling of the excess returns. The residuals ϵ_t and ϵ_{0t} contain idiosyncratic risk with diagonal covariance matrix $\sigma^2 I$.

For the remainder of this section, we assume that the factor loadings and number of factors are known. Working with known factor loadings is common in studies that use the NS model or when loadings are obtained from Principal Components Analysis (PCA) based on historic time series data for returns. An example of the latter is the seminal Litterman and Scheinkman (1991) study on hedging interest rate risk using a three-factor model with factor loadings estimated by PCA. Section 2 deals with the estimation of b and k.

Because of the errors ϵ_{0t} and ϵ_t, we cannot define a perfect hedge anymore. In the factor model with errors we also do not have redundant assets when $n > k$. Since the liability return ρ_0 is exposed to the same factors as the traded instruments ρ, an investor who holds a portfolio w with the same factor exposure B_0, that is $w'B = B_0$, will have hedged all factor risk. To fully define w we consider factor-mimicking portfolios as in, for example, Litterman and Scheinkman (1991). A factor-mimicking portfolio is a portfolio of liquid instruments with returns that best fit a factor.

To obtain the factor-mimicking portfolios, we consider Equation (6) as a linear regression, where b are the regressors and f_t the parameters. Estimating f_t by OLS, we obtain the mimicking portfolio excess returns $(b'b)^{-1}b'\rho_t$. The hedge portfolio predicts the out-of-sample liability return using the cross-section of "liquid" excess returns through the linear predictor

$$\hat{\rho}_{0t} = b_0(b'b)^{-1}b'\rho_t \equiv g'\rho_t. \tag{7}$$

The n-vector g determines the weights of the hedge portfolio. Since OLS is the best linear unbiased estimator, the estimator g minimizes the variance of the hedging error subject to the unbiasedness constraint $b_0 = g'b$. An equivalent way to derive the same portfolio is by minimizing the idiosyncratic risk $\sigma^2 g'g$ subject to the factor constraint, analogous to the setup in Diebold, Ji, and Li (2006a).

When factor loadings $b(\tau_i)$ change over time, as a function of information at time $t-h$, the hedge portfolio needs to be rebalanced to take into account the time-varying loadings. We then find the hedge portfolio by using time-varying b_t and b_{0t}.

Since the regression uses scaled returns, they need to be unscaled to obtain hedge portfolio excess returns

$$\hat{r}_{0t} = w' r_t, \tag{8}$$

with weights $w_i = g_i \tau_0 / \tau_i$. As both r_{0t} and r_t are excess returns, the portfolio weights imply an investment of $1 - \iota' w$ in the risk-free asset. As in Diebold, Ji, and Li (2006a), the portfolio w has the same generalized duration B_0 as the liability.

The hedge portfolio derived in Equation (7) aims at completely eliminating the factor risk and therefore does not depend on any time-series properties of the factors f_t. In essence, we followed Diebold and Li (2006) (and earlier literature) and have treated the factors f_t as time fixed effects. The fixed effects make our analysis robust against misspecification in the time-series process of f_t.

If we are willing to make assumptions on f_t, we may be able to find a better hedge portfolio. For this, we relax the constraint that the hedge portfolio has the same "generalized duration" as the liability. As before, let g be the scaled hedge portfolio, and let w be the unscaled portfolio weights. We choose w (or equivalently g) to minimize the Mean Squared Error (MSE) of the hedge return,

$$\min_w \mathrm{E}[(r_{0t} - w' r_t)^2] = \tau_0^2 \min_g \mathrm{E}[(\rho_{0t} - g' \rho_t)^2]. \tag{9}$$

The hedging error

$$\hat{\epsilon}_{0t} = \rho_{0t} - g' \rho_t = \epsilon_{0t} - g' \epsilon_t + (b_0 - g' b) f_t, \tag{10}$$

has three components: (i) the unhedgeable idiosyncratic error ϵ_{0t}; (ii) idiosyncratic noise in the cross-section of returns $g' \epsilon_t$, and (iii) a bias depending on how well the factor hedge portfolio immunizes the factor exposure. Given the additional assumption that we know or can estimate $\Omega = \mathrm{E}[f_t f_t']$, the squared error has expectation

$$\mathrm{E}[\hat{\epsilon}_{0t}^2] = \sigma^2 (1 + g' g) + (b_0 - g' b) \Omega (b_0 - g' b)'. \tag{11}$$

Minimizing with respect to g gives the optimal predictor

$$g = (b \Omega b' + \sigma^2 I)^{-1} b \Omega b_0'. \tag{12}$$

The expression reduces to the factor portfolio hedge (7) when the unbiasedness constraint $b_0 = g' b$ is imposed to be binding in Equation (9). One way to interpret Equation (12) is as a ridge regression with σ^2 as a regularization parameter to break the perfect collinearity among the explanatory variables b. When Ω and σ^2 vary over time the hedge portfolio also becomes time-varying, even if factor loadings remain stable.

In constructing the factor portfolios we pretend that the errors are cross-sectionally uncorrelated. That should be a reasonable approximation for residuals from a factor model that explains most of the common variation, but it cannot be literally true for term structure models. Errors on adjacent maturities must be closely related because of the smoothness of the yield curve and the factor loadings. When increasing the number of maturities n, the cross-sectional correlation will become stronger and stronger. In principle, we can include many maturities by interpolating on various segments of the yield curve, but this will just increase the cross-sectional correlation in the errors without adding new information. Our data contains $n=13$ real data points from the swap market in the maturity range 1–20 years, from which all other yields, forward rates, and returns are interpolated. Since n is thus necessarily small, overfitting becomes a potential problem with the cross-sectional regressions. For a k-factor model, we need to obtain k unknown parameters f_t every period from a small number of cross-sectional observations. Overfitting will be a central theme when we discuss the hedging results in detail in Section 4.

2 Estimating Factor Loadings

Crucial to the construction of hedge portfolios are the factor loadings and the number of factors. One way to determine both is by PCA. In PCA the factor loadings are fully unrestricted. The term structure literature, however, considers various parsimonious models for factor loadings as a function of maturity τ. A discount bond with maturity τ at time t with factor loading $b(\tau)$ will have maturity $\tau - h$ and factor loading $b(\tau - h)$ at time $t+h$. No-arbitrage then implies that loadings must be a smooth function of τ (on top of the drift conditions on μ_i in Equation (2) discussed in Section 1.1). The particular functional form depends on assumptions about the factor dynamics. Some popular choices are the loadings in the affine class (Duffie and Kan 1996; Dai and Singleton 2000; Duffee 2002). The NS factor loadings are a special case of a three-factor Gaussian affine model.

We first discuss the specification of the NS model, and then proceed with estimation details for both PCA and NS. We finally add time variation in the NS factor loadings.

2.1 The NS model for returns

Our version of the NS model differs slightly from most applications in the literature. Diebold and Rudebusch (2013) specify the NS model for continuously compounded discount yields $y_t(\tau)$ as

$$y_t(\tau) = b(\tau)F_t, \tag{13}$$

where F_t is a vector of three factors that are commonly interpreted as level, slope, and curvature. This interpretation is related to the factor loadings

$$b(\tau) = \left(1 \quad \frac{1-e^{-\lambda\tau}}{\lambda\tau} \quad \frac{1-(1+\lambda\tau)e^{-\lambda\tau}}{\lambda\tau}\right), \tag{14}$$

which depend on the shape parameter $\lambda > 0$. The first component has constant factor loadings of 1. As such, it influences short- and long-term yields equally, and can be considered as the level. The second component monotonically decreases in τ, and is therefore called the slope. The decay of the factor loadings is faster, the larger the shape parameter λ. For large λ the slope factor therefore mostly affects short-term rates. Factor loadings for the third factor are hump-shaped, moving from zero at the very short end to a maximum at $\tau = 1.8/\lambda$ years after which the factor loadings gradually fall back to zero. The third factor is therefore referred to as curvature.

Diebold and Li (2006) proposed the dynamic NS model for forecasting future interest rates. For hedging, our interest is not in time-series forecasting, but in cross-sectional prediction of changes in long maturity rates conditional on changes in more liquid shorter rates. We therefore transform the NS model to a model for excess returns. The NS model implies that log prices are given by $p_t(\tau) = -B(\tau)F_t$ with $B(\tau) = \tau b(\tau)$. For the excess returns $r_{t+h}(\tau)$ defined in Equation (5), we then obtain

$$\begin{aligned} r_{t+h}(\tau) &= -B(\tau)F_{t+h} + (B(\tau+h) - B(h))F_t \\ &= B(\tau)f_{t+h}, \end{aligned} \tag{15}$$

where $f_{t+h} = -(F_{t+h} - \mathcal{K}F_t)$ and

$$\mathcal{K} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mathrm{e}^{-\lambda h} & \lambda h \mathrm{e}^{-\lambda h} \\ 0 & 0 & \mathrm{e}^{-\lambda h} \end{pmatrix}. \tag{16}$$

Equation (15) holds because the NS factor loadings satisfy the linear difference equation

$$B(\tau + h) - B(h) = B(\tau)\mathcal{K}. \tag{17}$$

Dividing by τ, we obtain the scaled excess returns $\rho_t(\tau)$, which then have factor loadings $b(\tau) = B(\tau)/\tau$ that are identical to the NS loadings in Equation (13). Adding idiosyncratic noise to Equation (15) leads to the following factor model, formalizing the use of Equation (6) in Section 1.1,

$$r_t(\tau) = B(\tau)f_t + e_t(\tau), \tag{18}$$

where $r_t(\tau)$ is the excess return (return minus risk-free rate) on a discount bond with maturity τ, f_t are the three NS factors, $e_t(\tau)$ is idiosyncratic noise, and $B(\tau)$ are factor loadings that depend on a single "shape" parameter λ, which is the same for all maturities. The same λ also provides the factor loadings of the liabilities.

When h is small (daily observations) and τ is large (years), the scaled excess returns are approximately the negative of first differences in yields (plus a small spread term to preserve the mean). Compared to the usual NS model for yield levels, we have thus removed the unit root component from the data by first differencing. In different settings, estimation of term structure models on excess returns rather than yields has previously been advocated by Bams and Schotman (2003), Adrian, Crump, and Moench (2013), Goliński and Spencer (2017), and Bauer (2018).

Christensen, Diebold, and Rudebusch (2011) show that the NS model can be made consistent with no-arbitrage conditions. Under a no-arbitrage interpretation of the model, $\mathcal{K}$ represents the risk-neutral factor dynamics. Based on this, Christensen, Diebold, and Rudebusch (2011, p. 7) derive a maturity-specific constant term $a(\tau)$ to be added to Equation (13) to make it arbitrage-free. The adjustment term for the excess returns follows by taking the expressions for $a(\tau)$ for yields and applying transformation (5). The adjustment amounts to subtracting the conditional variance $\frac{1}{2}\mathrm{Var}_t[r_{t+h}(\tau)]$.

Our aim is to model the very long end of the term structure with maturities up to 50 years. For the standard NS model, Diebold and Li (2006) find that the curvature factor loadings have a maximum at slightly less than three years. In that case, maturities longer than 10 years are essentially only affected by the level factor. For a better fit at longer maturities, that is, longer than 10 years, the model is therefore often extended, either by a time-varying λ_t (see Section 2.3) or by a second curvature factor. The latter is known as the Svensson model. As noted by Christensen, Diebold, and Rudebusch (2009) and Krippner (2015), a problem with the Svensson model is that it cannot be made arbitrage-free. For this new slope and curvature factors must be added in pairs, as in Christensen, Diebold, and Rudebusch (2009). With a second pair of slope and curvature factors, we obtain a five-factor extended NS model that belongs to the class of Gaussian affine term structure models. For the five-factor model, the factor loadings become

$$b(\tau) = \left(1 \quad \frac{1-\mathrm{e}^{-\lambda_1\tau}}{\lambda_1\tau} \quad \frac{1-(1+\lambda_1\tau)\mathrm{e}^{-\lambda_1\tau}}{\lambda_1\tau} \quad \frac{1-\mathrm{e}^{-\lambda_2\tau}}{\lambda_2\tau} \quad \frac{1-(1+\lambda_2\tau)\mathrm{e}^{-\lambda_2\tau}}{\lambda_2\tau}\right), \tag{19}$$

with $\lambda = (\lambda_1\ \lambda_2)'$ now having two elements. When λ_2 is much smaller than λ_1, the second pair of factors aims at fitting the very long end of the yield curve. Dubecq and Gourieroux (2011) find that ultra-long yields need a second hump; the second λ can accommodate this.

2.2 Estimation

Estimation of λ in the NS model has not gained much attention. Often, it is set to a constant value, without any estimation (e.g., Diebold and Li 2006; Yu and Zivot 2011). However, λ is a crucial parameter for constructing a hedge portfolio, as it governs the exposure to the different factors across maturities. We write the factor model in vector notation as

$$\boldsymbol{R}_t = \boldsymbol{b} f_t + \boldsymbol{\varepsilon}_t, \tag{20}$$

where $\boldsymbol{R}_t$ is a N-vector containing the scaled excess returns $\rho_t(\tau_i)$ for maturities τ_i $(i = 1, \ldots, N)$ and $\boldsymbol{b}$ is the $(N \times k)$ matrix of factor loadings. The notation here uses $\boldsymbol{b}$ instead of b, and N instead of n as in Equation (6), to indicate that the number of included maturities may be different. For hedging we have n liquid maturities and consider hedging the longest maturity $\tau_0 = 50$, that is, $N = n + 1$. For estimation we add a few intermediate maturities in the range $20 - 50$ years to emphasize a good fit at the very long end of the term structure. In most cases we have $n = 13$ and $N = 17$.

For forecasting purposes, studies such as Diebold, Rudebusch, and Aruoba (2006b) and Koopman, Mallee, and Van der Wel (2010) complete the model by specifying a time series model for the factors F_t and thus implicitly f_t. Since our purpose in this article is the cross-sectional prediction of $r_t(\tau_0)$ conditional on liquid returns $r_t(\tau_i)$, we return to the original approach in Diebold and Li (2006) and treat the factors as time fixed effects. Under the fixed effects assumption, the factor model (20) applies both to NS as well as PCA. The only difference is the structure on the factor loadings.

We estimate the parameters $\boldsymbol{b}$, σ^2 and f_t $(t = 1, \ldots, T)$ by quasi-maximum likelihood. Conditional on $\boldsymbol{b}$ the factors f_t are estimated by cross-sectional regressions of returns on factor loadings as $\hat{f}_t = (\boldsymbol{b}'\boldsymbol{b})^{-1}\boldsymbol{b}'\boldsymbol{R}_t$. After concentrating with respect to f_t and σ^2, we obtain the concentrated log-likelihood

$$\ell^* = -\frac{NT}{2}\ln(\hat{\sigma}_k^2), \tag{21}$$

$$\hat{\sigma}_k^2 = \frac{1}{N}\left(\text{tr}(\boldsymbol{S}_\rho) - \text{tr}(\boldsymbol{Q}_k)\right), \tag{22}$$

where $\boldsymbol{Q}_k = \boldsymbol{b}'\boldsymbol{S}_\rho\boldsymbol{b}(\boldsymbol{b}\prime\boldsymbol{b})^{-1}$ depends on the sample second moment matrix of the scaled returns $\boldsymbol{S}_\rho = \frac{1}{T}\sum_t \boldsymbol{R}_t\boldsymbol{R}_t'$. For both the three-factor as well as five-factor NS model, the estimate for λ follows from maximizing (21) with respect to λ, which is equivalent to maximizing $\text{tr}(\boldsymbol{Q}_k)$. For PCA maximization of (21) leads to factor loadings $\boldsymbol{b}$ that are the eigenvectors of $\boldsymbol{S}_\rho$ corresponding to its k largest eigenvalues κ_i. In both cases, the idiosyncratic variance for a k-factor model is given by $\hat{\sigma}_k^2$, which for PCA reduces to $\hat{\sigma}_k^2 = \frac{1}{N}\sum_{j=k+1}^N \kappa_j$.[5] For standard errors on the NS parameter estimates, we cluster on maturities to be robust against cross-sectional residual correlation. Details are provided in Appendix C.

We use PCA to explore the data evidence on the number of factors and the fit of the NS factor loadings. The NS model is nested within PCA, but since the principal components (PCs) are only defined up to a rotation matrix, the NS and PCA factor loadings cannot be directly compared. We therefore project the NS factor loadings $\boldsymbol{b}_{NS}$ on the PCA factor loadings $\boldsymbol{b}_{PCA}$ such that $\hat{\boldsymbol{b}}_{PCA} = \boldsymbol{b}_{NS}(\boldsymbol{b}_{NS}'\boldsymbol{b}_{NS})^{-1}\boldsymbol{b}_{NS}'\boldsymbol{b}_{PCA}$.

Finally, for the "optimal" hedge portfolio (12), we need the factor second moments $\text{E}[f_tf_t']$. Starting from the identity $\hat{f}_t = f_t + (\hat{f}_t - f_t)$, we estimate Ω from the sample second moment matrix for the estimated factors $\hat{f}_t$, adjusted for estimation error, as

$$\Omega = \hat{\text{E}}[f_tf_t'] = \frac{1}{T}\sum_t \hat{f}_t\hat{f}_t' - \hat{\sigma}^2(\boldsymbol{b}\prime\boldsymbol{b})^{-1}. \tag{23}$$

For PCA, this reduces to the diagonal matrix with the k largest eigenvalues.

2.3 Time-varying factor loadings

We introduce time-varying factor loadings in the NS model for excess returns by allowing the λ parameter to change over time. When λ is time-varying, we specify a factor model for excess returns with the NS factor structure and time-varying factor loadings,

$$\rho_{t+h}(\tau) = b_t(\tau)f_{t+h} + \epsilon_{t+h}(\tau) \tag{24}$$

The model still has three NS factors. Also, given λ_t, the hedge portfolios remain as in Equation (7) in Section 1.1 but now replacing the constant b by the time-varying b_t.

When factor loadings are constant, the return factor model is implied by the model for yield levels, and the choice between modeling returns or yield levels does not make much of a difference for estimating the factor loadings and constructing hedge portfolios. But with time-varying λ_t, the two specifications differ. Most importantly, the return specification maintains a factor structure for the returns. The specification allows interpretation of the model within the class of Heath, Jarrow, and Morton (1992)arbitrage-free term structure

5 For a model in yield levels, both λ and the PCA loadings are estimated by replacing S_ρ in Equation (21) by the second moment matrix of yield levels S_y. When the model is correctly specified, that is, when the model errors are orthogonal to the factors, both provide consistent estimates for the loadings, even though they exploit different moment conditions. Under misspecification, for example due to time-varying λ_t, the two can be very different. For our data we do get similar λ's when we estimate them from the full sample moment matrices. They differ, however, on subsamples. This is one reason we consider time-varying λ_t.

models. This economic structure on the portfolios helps to avoid erratic weights depending on some spurious noisy correlation in the data, and provides an additional motivation for our specification in returns rather than levels.

We explicitly model time variation in the shape parameter. Few studies attempt to estimate λ allowing for time variation. Hevia et al. (2015) allow for variation through a two-state Markov switching model. In a model for yield levels, Koopman, Mallee, and Van der Wel (2010) specify a state space model that they linearize with respect to λ_t, treating λ_t as a fourth factor. We model the dynamics of λ_t by means of a DCS model. The general specification of DCS models is discussed in Creal, Koopman, and Lucas (2013). Theoretical results are established in, amongst others, Blasques, Koopman, and Lucas (2015). We apply the DCS principle as a natural way to update parameters over time.

In the class of DCS models, the dynamics of parameters are driven by the score of the likelihood with respect to that parameter. To derive a DCS model for λ_t, we consider the log of the conditional observation density of scaled returns $\ell(\boldsymbol{\rho}_t|\lambda)$, and let $z_t = \frac{\partial \ell_t}{\partial \lambda}$ be the score with respect to λ evaluated at λ_t. Then the DCS model for the shape parameter is defined as

$$\lambda_{t+h} = \overline{\lambda}(1 - \phi_1) + \phi_1 \lambda_t + \phi_2 s_t, \tag{25}$$

where $s_t = S_t z_t$ is the score multiplied by 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 likelihood function. 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 choice of S_t adds flexibility in how the score updates the parameter. Creal, Koopman, and Lucas (2013) discuss several options of the form $S_t = \mathcal{I}_t^{-a}$, where $\mathcal{I}_t$ is the information matrix. In a univariate model the choice of a is not that important. Natural choices for a are 1 to do a Newton step on the log-likelihood, and 0 to do a steepest descent step. As long as the second-order derivatives are relatively constant, the parameter ϕ_2 can adjust the scaling of the gradient. We choose $a = \frac{1}{2}$. In Appendix B, we find explicit formulas for the scores and the information matrix.

Since the "innovation" in the DCS model, s_t, a function of σ_t^2, it seems important to allow for GARCH effects in the idiosyncratic risks $\epsilon_t(\tau)$. When we allow for heteroskedasticity in $\boldsymbol{\varepsilon}_t$, the score is down-weighted in periods of high volatility, reducing the impact of large shocks in such periods, which seems valuable given the nature of our sample period. Many studies, such as Bianchi, Mumtaz, and Surico (2009), Koopman, Mallee, and Van der Wel (2010), and Hautsch and Ou (2012) allow for time variation in the variances of the innovations to NS factors. We allow for heteroskedasticity in the innovations to excess returns in a similar manner, using a single common GARCH process to drive the time variation in idiosyncratic risk. In order to have a parsimonious specification we assume a standard GARCH(1,1):

$$\sigma_{t+h}^2 = \overline{\sigma}^2(1 - \alpha - \beta) + \alpha \sigma_t^2 + \beta \frac{\boldsymbol{\varepsilon}_t' \boldsymbol{\varepsilon}_t}{N}. \tag{26}$$

At this point it is interesting to note that GARCH is itself a DCS updating formula for the error variance (Creal, Koopman, and Lucas 2013). With time-varying σ_t^2 we have two self-explanatory additional models, the NS-GARCH and NS-DCS-GARCH.

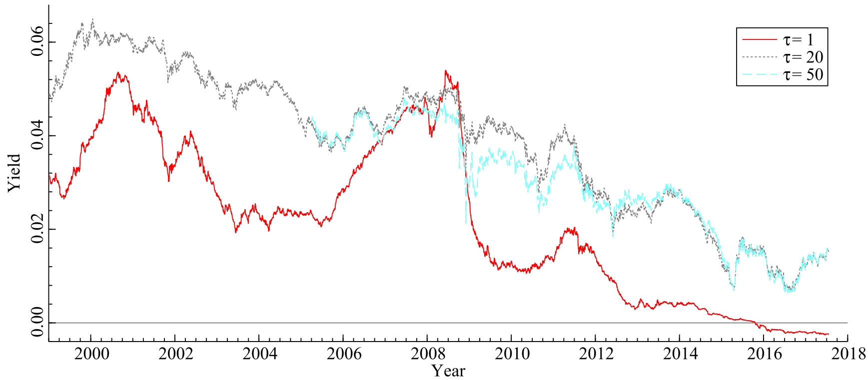

Figure 1 Yield time series. This figure plots the time series of yields derived from euroswap data for maturities of 1, 20, and 50 years.

3 Data

We construct daily yield curves using euro swap rates with maturities 1–10, 12, 15, 20, 25, 30, 40, and 50 years. Data are available from January 1999, except for the 40- and 50-year rates, which are available from April 2005 onward. Our sample ends on July 20, 2017. Appendix A provides the data sources and details of the transformation from swap rates to excess returns.

We will consider two different sets of maturities. The first set contains the full cross-section of all seventeen maturities; the second is the subset with only the maturities up until and including 20 years. The swaps with maturities longer than 20 years are generally regarded as less liquid. EIOPA, for example, considers 20 years to be the so-called "last liquid point" of the market.[6]

We plot a snapshot of the data in the form of the time-series of yields for maturities of 1, 20, and 50 years in Figure 1. The yields are obviously correlated but far from perfect. The 50-year maturity starts in April 2005 and tends to track the 20-year yield closely. However, it deviates significantly in the turbulent period started by the financial crisis up to the European sovereign debt crisis.

Table 1 reports descriptive statistics of the scaled excess returns $\rho_t(\tau) = r_t(\tau)/\tau$. The average excess returns are all positive, reflecting both the normal upward-sloping term structure as well as the capital gains from the decline in yield levels over the sample period. The scaled returns appear to have some remaining cross-sectional heteroskedasticity, as the standard deviation still increases in maturity. When the curvature loadings peak at three - years, the slope and curvature factors can account for the upward-sloping volatility between one and three years. The slight continuing increase in the long end, beyond the 10-year maturity, is harder to explain. This will cause some problems for a standard NS model, since very long rates will mostly be affected by the level factor. To accommodate the high

6 See the EIOPA publications CEIOPS (2010) and EIOPA (2014a, 2014b).

Table 1. Summary statistics

	Since 2005		Full sample	
τ	Mean	Variance	Mean	Variance
1	0.252	4.816	0.210	6.358
2	0.187	10.298	0.166	12.598
3	0.173	14.234	0.156	16.057
4	0.166	13.645	0.149	15.899
5	0.162	14.948	0.145	16.790
6	0.159	15.355	0.142	16.680
7	0.156	15.136	0.139	16.150
8	0.153	15.619	0.136	16.314
9	0.150	16.284	0.133	16.744
10	0.147	16.828	0.130	16.996
12	0.141	17.923	0.125	17.618
15	0.133	18.987	0.118	18.122
20	0.122	20.466	0.109	19.205
25	0.115	21.451	0.103	19.859
30	0.109	21.970	0.098	20.125
40	0.103	23.180		
50	0.101	23.886		

Notes: The table provides descriptive statistics on daily data for maturity scaled excess returns $\rho_t(\tau)$ over the period January 1999 to July 2017. The 40- and 50-year returns start in April 2005. Returns have been multiplied by 10^4.

Table 2. Principal components

Factors	All since 2005			Liquid since 1999		
	R^2	σ^2	c_{50}^2	R^2	σ^2	c_{50}^2
0	0.0	16.790		0.0	15.831	
1	0.831	2.834	0.780	0.875	1.987	0.605
2	0.940	1.012	0.144	0.960	0.630	0.266
3	0.963	0.616	0.044	0.973	0.421	0.061
4	0.972	0.477	0.002	0.980	0.314	0.001
5	0.977	0.385	$< 10^{-4}$	0.985	0.230	0.002

Notes: Entries are based on the eigenvalues κ_j of the second-moment matrix $\frac{1}{T}\sum_t \boldsymbol{R}_t\boldsymbol{R}_t'$ of scaled excess returns. Cumulative explained variance by k factors is defined as $R^2 = \sum_{j=1}^k \kappa_j / \sum_{j=1}^N \kappa_j$. Residual variance is defined as $\sigma^2 = \frac{1}{N}\sum_{j=k+1}^N \kappa_j$. The longest maturities are only available since April 2005, resulting in $(T, N) = (3210, 17)$. The full sample with only liquid maturities since 1999 has $(T, N) = (4838, 13)$. The column c_{50}^2 reports the squared correlation between the hedge target $\rho(50)$ and each of the PCs. For the liquid sample, the correlation c_{50} is based on the overlapping sample since 2005.

volatility at the very long end, we must either add additional factors or allow for selected periods with a very small λ.

An important part of our discussion involves the number of factors. Unrestricted PCs provide a first indication of the factor structure. Table 2 shows the percentage of variance explained by the first five PCs. The bulk of the variation is already explained by the first factor. Contrary to PCs extracted from yield level data, the dominance of the first PC in scaled returns is not due to a (near) unit root, but due to the strong correlation in yield changes. In both subsamples (all maturities since 2005, and full-time series for the liquid maturities only), the first three PCs explain around 97% of the variance. From this perspective there is little incremental value from a five-factor model. Our hedge target $\rho_t(50)$ is also strongly correlated with the first three PCs. PCs 4 and 5 appear almost uncorrelated and negligible. Even if PCs are estimated on data including the excess returns with maturities longer than 20 years, the additional factors still do not add anything to explaining the 50-year return. The additional factors may become more important once we restrict the factor loadings by the NS specification.

4 Model Estimates and Factor Hedging

4.1 Estimation Results

In order to implement the hedge portfolios described in Section 1.1, we require estimates of the factor loadings. Our primary focus is on both the five- and original three-factor versions of the NS model, described in Equations (14) and (19). Table 3 presents in-sample parameter estimates of the two specifications for both the full and the "liquid" cross-section.

First, consider the three-factor estimates. There is a large difference in λ between the two cross-sections. With a constant λ and just the liquid bonds, we replicate the common result in the literature: the value $\lambda = 0.53$ corresponds to a peak in the curvature at a maturity around three years. In this case, slope and curvature factors hardly have an effect on very long maturities. Adding the longer maturities in the estimation leads to the much lower value $\lambda = 0.25$, which puts the maximum of the curvature factor between seven and eight-years. With such low λ, slope and curvature factors have a substantial impact on the very long maturities. It appears that the ultra-long maturities are more than just the level factor.

This is confirmed by the parameter estimates of the five-factor model. As expected, there are two distinct λs, one small and one with the usual value around 0.5. The small λ for the second pair of slope and curvature factors implies a peak at 22 years. These factors therefore have strong loadings at very long maturities, and can potentially provide a link between the shorter maturities and the very long maturities. Estimating the five-factor model on the liquid subsample results in poor estimates for the λs. The two λs are almost equal, creating a huge multicollinearity problem in estimating the factors. Without the ultra-long maturities, the five-factor model appears under-identified.

Compared to the three-factor model, the five-factor model reduces the residual variance from 0.74 to 0.43. This is a large improvement in the in-sample fit, but it does come at the cost of many additional parameters with a risk of overfitting. Given λ, the k-factor model fits a cross-section of seventeen maturities using k parameters for each time period. With k factors and T time-series observations, we have kT parameters (plus the two λs) to fit the NT data points.

The loadings of the three-factor NS model are close to the unrestricted factor loadings obtained by PCA. Figure 2 shows the PCA factor loadings from the first three factors along with the implied loadings from the NS model. For both data samples, the NS factor loadings almost perfectly fit the loadings of the first two PCs. For the third factor, unrestricted PCA loadings show a little more pronounced curvature than implied by the NS model.

Table 3. Parameter estimates: NS

	NS five-factor		NS three-factor	
	All	Liquid	All	Liquid
λ_1	0.0864	0.4969	0.2514	0.5293
	(0.0098)	(0.0103)	(0.0386)	(0.0306)
λ_2	0.5196	0.5117		
	(0.0419)	(0.0195)		
σ^2	0.4328	0.3244	0.7377	0.5225
	(0.0529)	(0.0446)	(0.0767)	(0.0711)
$\mathcal{L}$	196723	156434	182173	146505

Notes: This table provides parameter estimates, with standard errors in brackets, for the three- and five-factor NS model with constant shape parameters. We report estimates 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 20 years. σ^2 is multiplied by 10^4. The maximized value of the log-likelihood is denoted by $\mathcal{L}$. Details on the standard errors are provided in Appendix C.

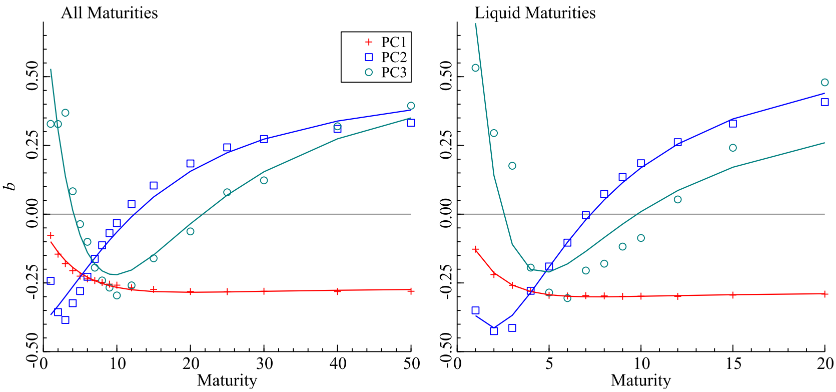

Figure 2. Factor loadings. The figure shows the unrestricted PC factor loadings (symbols) and the fitted NS curves (solid lines). The NS curves have been projected on the PCA loadings. The left panel uses all maturities since 2005; the right panel the liquid maturities for the longer sample since 1999.

4.2 Hedging Portfolios

We start our discussion of the hedging performance with a look at what these model estimates imply for the hedge portfolios described in Section 1.2. We use the full cross-section point estimates to determine the factor portfolio weights based on Equation (7) as well as the MSE-optimal weights given in Equation (12). Optimal weights are computed using factor loadings b from the five-factor model. For Ω we use the full sample estimate adjusted for estimation error of Equation (23). The resulting weights are plotted in Figure 3.

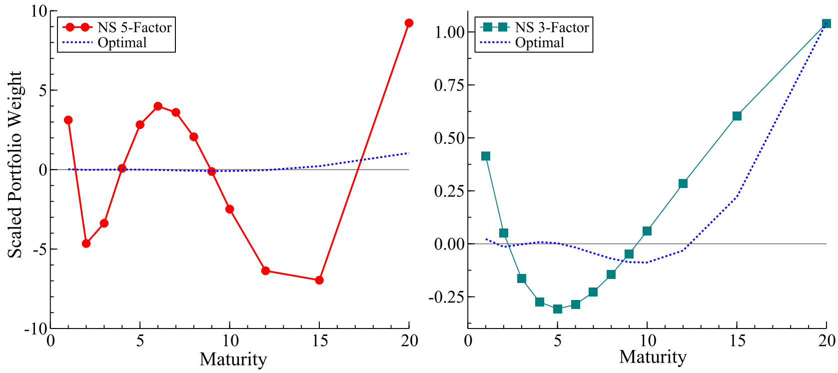

Figure 3. Hedge portfolio weights. This figure plots the maturity-scaled weights, $g(\tau)$, for factor mimicking portfolios based on five- and three-factor NS models. Both panels also show the optimal weights based on the five-factor model. Note the different scaling in the left and right panels. The dashed line with optimal weights is the same in both panels.

Comparing the five- and three-factor portfolios in the left- and right-hand panels clearly shows how the five-factor portfolio attempts to reduce the exposure to a second slope and curvature factor catered to the longer maturities. This results in more erratic portfolio weights across maturities, and more extreme portfolio weights in general. The MSE-optimal portfolio weights are much smoother and much less extreme than the factor portfolio weights for the five-factor model. The right-hand graph is even more revealing. It shows the same optimal weights, but this time in the same frame as the weights for the factor-mimicking portfolio for the three-factor NS model (with constant $\lambda = 0.25$). The weights implied by the factor-mimicking portfolio for the three-factor model are much closer to the MSE-optimal weights. The weights are of the same magnitude, especially in the segment with maturities of 10 years and longer. Most striking, however, is how close the optimized weights resemble the naive portfolio. The scaled weight at 20-year maturity is almost equal to one, both for the optimized as well as the three-factor model. Moreover, the optimal weights are very close to zero for all other maturities, except a small weight of 0.2 at the 15-year maturity.

The optimal portfolio strongly depends on Ω. In deriving the optimal weights for Figure 3, we used the full-sample Ω using all days and all maturities. Estimates of Ω from smaller subsamples can be noisy. Estimation error and time variation in Ω will deteriorate the empirical performance of the optimized hedge portfolio. Moreover, the optimized portfolio is only optimal given the correct specification of the factor model. In contrast, the factor hedge portfolios do not depend on Ω, and the naive portfolio is even more robust, since it does not rely on any parameter estimates at all.

4.3 Hedging Performance

We next turn to the empirical hedge performance of the different models. For the NS models, we consider the standard factor-mimicking portfolios based on parameter estimates from both the full and liquid cross-section. In addition to the five- and three-factor versions, we consider a one-factor version which only contains the level factor. Similar to the naive strategy, this results in a duration hedge. However, the one-factor hedge portfolio is a diversified portfolio of discount bonds across all maturities. We complete the NS-factor hedging results with the MSE-optimal portfolio based on full cross-section five-factor estimates. In addition to the NS-based portfolios, we form portfolios based on unrestricted factor loadings derived from PCA, and evaluate the naive portfolio which simply leverages up the last liquid maturity.

We compare models using rolling-window parameter estimates from the previous 1500 days. At a particular day t we then have factor loadings (b_t, b_{0t}). The break in data availability of 40- and 50-year rates in April 2005 serves as a natural starting point for the first out-of-sample forecast. In the period before the long-maturity data becomes available, we use the data up to the 30-year maturity for parameter estimation purposes. Correspondingly, evaluation is always based on observed returns, but the lack of their availability does have an impact on the estimates of λ and Ω early in our sample.

As our measure for hedge performance we use the average and standard deviation of the hedging errors $r_{0t} - B_{0t}\hat{f}_t$, which are jointly summarized by their Root Mean Squared Error (RMSE). We report averages across the full-sample as well as three subsamples. The first subsample runs to December 2007. For this period, the rolling windows used for estimation did not consistently have access to observed rates beyond 30 years. The second subsample encompasses the global financial crisis and lasts until December 2009. The final subsample contains the remainder. Apart from the descriptives on the hedging error we report some descriptives on the portfolio composition: the total weight in risky assets $\text{TW}_t = \iota' w_t$, the portfolio concentration $\text{CO}_t = \sqrt{w_t' w_t}$, and the portfolio turnover $\text{TO}_t = \sum_i |w_{it} - w_{i,t-h} \text{e}^{r_{i,t-h} - \hat{r}_{0,t-h}}|$.

Table 4 reports the hedging errors of the different models. The main result from the table is that the naive hedging strategy beats all alternative (model-based) strategies by quite some margin. The leveraged investment in the 20-year bond has the lowest RMSE, while at the same time it has very low turnover, implies less leverage and, despite investing in just a single bond, it is less concentrated than all other hedge strategies.

A comparison of the performance of the various other hedging strategies reveals some additional important differences. Despite the improvements in fit, the five-factor model performs much worse than the three-factor models and is even worse than not hedging at all. In addition to the poor hedging errors, the portfolio positions and the resulting turnover are extreme. Only considering a single factor in turn decreases hedging performance, suggesting that the bias–variance tradeoff in terms of a number of factors and the resulting multicollinearity is minimized at three factors. The relative ordering of the factor models is mostly stable over time.

The "Optimal" column shows that hedging performance can indeed be improved by dropping the unbiasedness constraint and considering the MSE-optimal portfolio of Equation (12). While the full-sample results are worse than even the one-factor model's results, the portfolio improves overall the factor-based portfolios from the second subsample onward. These results further corroborate our earlier discussion on the importance, and difficulty, of estimating Ω. The early sample does not consistently have access to the 40- and 50-year maturities, which are vital for the estimation of the five-factor model. The resulting poor estimates lead to similarly poor hedging performance. Interestingly, even in the latter two subsamples, the optimal portfolio does not improve over the naive portfolio in terms of RMSE.

Finally, we compare the parameteric three-factor NS models with PCA. Differences between PCA and NS are solely the result of the nonlinear restrictions that the NS

Table 4. Empirical hedging results

	NS five-factor		NS three-factor		NS one-factor	Optimal	PCA	Naive	No hedge
	All	Liquid	All	Liquid					
Panel A: Full sample (April 2005 to July 2017)									
Bias	−0.06	0.06	−0.01	−0.02	−0.03	0.00	−0.04	−0.01	0.05
Std dev	6.90	3.00	1.25	1.15	1.57	1.68	1.21	1.02	2.44
RMSE	6.90	3.00	1.25	1.15	1.57	1.68	1.21	1.02	2.45
TW	33.24	2.08	10.76	4.32	12.03	10.28	19.26	2.50	0.00
CO	157.76	12.28	18.48	10.82	4.81	34.28	19.64	2.50	0.00
TO	21.89	29.11	0.64	0.42	0.14	61.94	0.78	0.02	0.00
Panel B: Subsample hedging performance									
	April 2005 to December 2007								
Bias	0.02	0.02	0.01	0.01	0.05	0.00	0.03	0.02	−0.01
Std dev	0.57	0.56	0.53	0.52	0.81	2.85	0.50	0.50	1.57
RMSE	0.57	0.56	0.53	0.52	0.81	2.85	0.51	0.50	1.57
	January 2008 to December 2009								
Bias	−0.09	0.04	0.04	0.04	−0.09	0.06	−0.11	0.01	0.10
Std dev	9.48	1.70	1.90	1.94	2.73	1.74	1.78	1.68	3.48
RMSE	9.48	1.70	1.90	1.94	2.74	1.74	1.78	1.68	3.48
	January 2010 to July 2017								
Bias	−0.08	0.08	−0.03	−0.04	−0.04	−0.02	−0.05	−0.03	0.06
Std dev	7.32	3.71	1.22	1.03	1.34	0.93	1.20	0.93	2.37
RMSE	7.32	3.71	1.22	1.03	1.34	0.93	1.20	0.93	2.37

Notes: The table reports the hedge performance for the factor hedge portfolios implied by different versions of the NS model designed to hedge a 50-year liability using liquid bonds with maturities up to 20 years. Performance is measured by the average, standard deviation, and RMSE of the prediction error $r_{0t} - B_{0t}\hat{f}_t$. Estimates for the NS λ parameter in the five- and three-factor NS models, as well as the factor loadings based on PCA are based on a rolling sample of 1500 past days. For the NS model, the Liquid column is based on parameter estimates using maturities up to 20 years, the All and Optimal columns use parameter estimates from the full cross-section. The “No Hedge” benchmark refers to the model that predicts zero excess returns for all t. As a summary of the portfolio composition the table reports averages of the statistics $\mathrm{TW}_t = \iota' w_t$, $\mathrm{CO}_t = \sqrt{w_t' w_t}$, and $\mathrm{TO}_t = \sum_i |w_{it} - w_{i,t-h} e^{r_{i,t-h} - \hat{r}_{0,t-h}}|$.

model imposes on factor loadings. The hedging performance of PCA appears very similar to the two versions of the NS three -factor model, consistent with the similarity of the shape of the factor loadings shown earlier in Figure 2. Comparing the three -factor PCA with the three -factor NS, both estimated on the full cross-section, the constraints imposed by NS model seemingly hurt its ability to hedge long-term liabilities. In contrast, the NS three -factor model estimated on the liquid maturities only, improves over PCA. The NS model naturally predicts factor loadings for all maturities, regardless of whether they are included in estimation, while the PCA can only produce factor loadings for included maturities. PCA does, however, perform better during the financial crisis period, which suggests that the NS model may be too restrictive in times of market turmoil.

4.4 Understanding the hedging performance

The poor performance of the model-based hedging strategies is related to the relatively large number of cross-sectional parameters that need to be estimated with a limited amount of observations. At the estimation stage we could estimate the factor model using the full cross-section of seventeen maturities, including the four very long ones. For the hedging problem, we need to estimate five parameters using just thirteen observations for the five-factor version. Moreover, the omitted out-of-sample observations are the most informative for the second slope and curvature factors.

Below we demonstrate that the empirical ranking of models we observe is exactly what we should expect when the estimated five-factor model is the true Data Generating Process (DGP). We conduct the following experiment. Assuming that the five-factor model is the DGP, we analyze the prediction errors of alternative models. The analysis examines the tradeoff between bias and variance, which is standard in the forecasting literature, but nevertheless enlightening for the current hedging problem.

Let (b, b_0) be the factor loadings of the five-factor model. The prediction errors of this model are

$$\hat{\epsilon}_{0t} = \epsilon_{0t} - b_0(b\prime b)^{-1}b'\epsilon_t. \tag{27}$$

Assuming that the factor model (6) is correct, the predictor is unbiased and has MSE (equal to the variance),

$$\text{MSE} \equiv \text{E}[\hat{\epsilon}_{0t}^2] = \sigma^2(1 + b_0(b\prime b)^{-1}b_0'). \tag{28}$$

The first term is the irreducible variance from the idiosyncratic risk ϵ_{0t}, which will be shared by all prediction models. The second term arises from the variance of the estimated factors. For the five-factor model with λ estimated from the full cross-section (see Table 3), we find that it is disturbingly large,

$$D_0 = b_0(b\prime b)^{-1}b_0' = 264. \tag{29}$$

This means that all the variance comes from estimation error of the factors. The statistic D_0 is related to the *leverage* of the data point b_0. In regression analysis, *leverage* is defined as the diagonal of the hat-matrix $G = X(X\prime X)^{-1}X'$, where X is the regression design matrix, in our case $X' = (b'\, b_0')$. For observation b_0 we have $G_0 = D_0/(1 + D_0) = 0.996$. The term leverage for G_0 refers to the effect that ρ_0 has on its own fitted value $\hat{\rho}_0$, that is, $G_0 = \partial\hat{\rho}_0/\partial\rho_0$. The fitted value moves one-for-one with the actual data point, and is hardly affected by the liquid part of the cross-section. In other words, the out-of-sample predictor b_0 is an outlier with respect to the in-sample regressors b. The large value for D_0 is the accumulation of two problems. The first problem is multicollinearity. The condition number (ratio of largest to smallest eigenvalue) of $b'b$ equals 6×10^5. With maturities up to 20 years, two of the five factors are almost redundant. The second problem is the extrapolation. With the estimated λ the out-of-sample factor loadings b_0 are very different from the in-sample factor loadings b. In econometric terms: the second pair of slope and curvature factors is not identified in a sample with only shorter term maturities.

Next consider an alternative, potentially misspecified, model which uses the $(n \times m)$ matrix of factor loadings c and the $(1 \times m)$-vector of out-of-sample loadings c_0. This model predicts

$$\hat{\rho}_{0t} = c_0(c\prime c)^{-1}c'\rho_t, \tag{30}$$

and has prediction error

$$\hat{\epsilon}_{0t} = \epsilon_{0t} - c_0(c\prime c)^{-1}c'\epsilon_t + (b_0 - \hat{b}_0)f_t, \tag{31}$$

where $\hat{b}_0 = c_0(c\prime c)^{-1}c'b$ is the projection of the out-of-sample misspecified loadings on the true factor loadings. The extra term reflects the bias of the prediction due to misspecification. Conditional on the factors f_t the expected squared prediction error is

$$\text{MSE} = \sigma^2(1 + c_0(c\prime c)^{-1}c^{*\prime}) + (b_0 - \hat{b}_0)f_tf_t'(b_0 - \hat{b}_0)'. \tag{32}$$

In the misspecified model, there is a tradeoff between variance and bias. Letting c be the factor loadings for the three-factor model, its leverage $D_{0c} = c_0(c\prime c)^{-1}c_0' = 2$ is much smaller than D_0 for the five-factor model. In the three-factor model, the out-of-sample loadings c_0 are much more connected to the in-sample loadings c. The three-factor model will thus have much lower variance at the cost of some misspecification bias, which we will examine shortly.

In the three-factor model, we are still extrapolating the slope and curvature factors, and therefore c_0 is still outside the range of the design c. This problem will get worse for smaller λ. If λ is small at around 0.1, a value we will observe during certain periods in the next section, leverage increases sharply, and $D_{0c} = 12$. In such a period, the longest maturities are thus more disconnected from the liquid shorter maturities and therefore hedging is more difficult.

The difference with the variance of the five-factor model suggests, however, that the three-factor model may easily do better than a five-factor model, even if the five-factor model is the true DGP. In the extreme case, setting $c_0 = 0$, the liability is not hedged at all, and the variance due to estimation error is equal to zero.

The relative performance depends on the magnitude of the bias. For this we need to make assumptions on the factor second moments $\text{E}[f_tf_t']$. As a realistic setting we assume that the five factors have a second-moment matrix equal to the sample second-moment matrix for the estimated factors with the complete panel of all time-series and cross-section observations as given in Equation (23). Being based on the full cross-section, the factor estimates in Equation (23) are much more precise than the estimates in Equation (7) which only uses the "in-sample" liquid maturities. Comparing diagonal elements of $(b\prime b)^{-1}$ and $(\boldsymbol{b}\prime\boldsymbol{b})^{-1}$ it appears that the former is about 200 times as big for the level factor and 50 as big for the second slope factor. Adding the additional four series with the long maturities $(25, 30, 40, 50)$ years is thus essential in estimating the second level and slope factors.

We can now evaluate the hedging performance of alternative factor models under the assumption that the five-factor model is the DGP and Ω is the true factor second-moment matrix. Table 5 reports the results. As for the empirical data, the worst performance by far comes from the five-factor model, which in this case, however, is correctly specified. Despite being unbiased it has an enormous variance due to the extreme multicollinearity in the factor loadings. Not hedging at all exhibits the second-worst performance, albeit already with an RMSE that is less than half that of the correct five-factor model. All other hedge portfolios are better. The three-factor NS model is better than the single factor

Table 5. Theoretical hedging results

	All	Liquid	All	Liquid	NS one-factor	Optimal	Naive	No hedge
	NS five-factor		NS three-factor					
Bias	0	0.88	1.21	1.09	1.54	0.74	0.76	2.41
Std dev	5.36	0.82	0.58	0.43	0.34	0.43	0.43	0.33
RMSE	5.36	1.21	1.34	1.17	1.58	0.86	0.88	2.44

Notes: The table reports out-of-sample hedge results for hedging the excess returns on a 50-year discount bond using liquid instruments with maturities up to 20 years. All results are calculated assuming that the five-factor model is the DGP. Parameters for all models are as reported in Table 3. The "duration" model refers to a single-factor model with only the level factor. The "No Hedge" benchmark refers to the model that predicts zero excess returns for all t. For the multifactor NS models, the columns differ in the maturities used for model estimation. Performance statistics are the square roots of $\text{Bias}^2 = (b_0 - \hat{b}_0)\Omega(b_0 - \hat{b}_0)'$, $\text{Std dev}^2 = \sigma^2(1 + c_0(c\prime c)^{-1}c^{*\prime})$ and $\text{RMSE}^2 = \text{Bias}^2 + \text{Std dev}^2$, all multiplied by the maturity $\tau_0 = 50$ of the hedge target.

duration model. Duration hedging has a large bias, but since it just needs the average (maturity scaled) return on the liquid bonds, it does not gather much estimation noise.

A result that is harder to explain, both empirically as well as conditional on the five-factor DGP, is the reasonable good performance of models where the λ parameter is estimated using only the liquid maturities. These models obviously cannot capture the behavior of the very long maturities. They therefore have a fairly large bias. Nevertheless, the predictors have low variance. Both the three-factor as well as the five-factor NS model have fairly large λ estimates. When λ is large, very long-term returns are mainly driven by the level factor, and hence the slope and curvature factor hardly affect the 50-year return. The predicted exposure at τ_0 is essentially duration plus a small adjustment for slope and curvature. Paradoxically, this turns out to be a reasonably good predictor, given that the true (second) slope and curvature factors are not identifiable from the liquid subsample.

The MSE-optimal strategy, unsurprisingly, has the lowest RMSE. However, the naive strategy is remarkably close, without the need for, and hazards of, model estimates. Even in the later parts of our sample, where Ω can be estimated more precisely, the potential gains of the optimal strategy with respect to the naive strategy are too small to overcome the inherent difficulty of estimating the model.

Hedging errors under the five-factor DGP in Table 5 are smaller than the empirical hedging errors for the empirical data in Table 4, implying that there is still some remaining misspecification. The hedging strategies also leave a substantial part of the interest rate risk in the 50-year liability. Without hedging the RMSE of the liability equals 2.45. Both the naive as well as the three-factor NS models reduce this to the range 1.0–1.25, implying that about 50% of the risk is hedged. So even if a fund would decide to fully hedge this long-term interest rate risk, it would effectively only reduce it by half. In the next section, we investigate whether time variation in factor loadings can fill the gap.

5 Hedging with Time-Varying Factor Loadings

5.1 Estimation Results

The previous sections highlighted that term structure data with long maturities can benefit from more than three factors, but that the inherent correlation between yields, which only becomes stronger on long maturities, prevents accurate estimation of more than three factors. The five-factor model can alternatively be interpreted as a three-factor model with time-varying factor loadings. This same interpretation was put forth in the term structure context by Koopman, Mallee, and Van der Wel (2010), and more generally in for instance Breitung and Eickmeier (2011). In this section, we achieve time variation in factor loadings through the DCS model introduced in Section 2.3, and investigate whether allowing for time variation in the loadings sufficiently increases hedging performance to get close, or surpass, the naive strategy.[7]

We summarize the estimation results of the NS models with time-varying factor loadings and time-varying volatility in Figure 4.[8] The top panel presents the paths using all maturities, and the bottom panel presents the results that use the liquid maturities only. The most striking differences are between the cross-sections, not between the models. Until the fall of 2008, λ is stable around 0.5, irrespective of the included maturities. In this period a NS model with constant λ fits the data across all maturities. After the start of the financial crisis the paths of λ_t strongly diverge depending on whether all bonds or only the liquid bonds are used. Using all maturities, λ_t drops to values around 0.15 and stays at that level until 2011, while the liquid-only estimate has a smaller drop and rises to values close to one over the same period. After 2013 the full cross-section remains relatively stable around the 0.25 mark, while the liquid cross-section displays another turbulent period in 2015. The time path complements the results from the constant λ models: after 2008 the factor structure of the very long-term bonds requires more than just a level factor. Finally, the effect of allowing for GARCH effects is limited. Its impact becomes most apparent for the NS and NS-GARCH models in the top panel, which strongly differ, with the NS-GARCH model down-weighting the second half of the sample due to higher volatility.

An interesting corollary from the figure is that the two distinct five-factor λ estimates are very similar to the two "regimes" filtered out by the NS-DCS, as illustrated in predominantly the top panel of Figure 4. This confirms the notion that the five-factor model can be interpreted as a (potentially over-parameterized) mixture of the three-factor NS-DCS, where at different times either of the λs drives the dominant slope and curve factors. This also immediately shows the redundancy of a potential five-factor NS-DCS, as the time variation in the three-factor NS-DCS is picked up in a constant five-factor version through the factor returns f_t.

We further illustrate the distinction between the choice of three or five factors, as well as constant or time-varying factor loadings in Figure 5. Recall that the hedge-portfolios are essentially cross-sectional predictions of the long-maturity return. The figure plots the fitted returns from the constant λ, three- and five-factor models, as well as the three-factor DCS model. We also include the naive prediction, which is that all standardized returns are equal to the 20-year standardized return.

7 We have alternatively considered both static NS models and PCA estimated on shorter rolling windows of just 250 observations. These models' hedging performance tends to be worse than either the static, or the DCS models based on 1500 rolling windows observations.

8 A complete table with parameter estimates is deferred to Appendix D. Although a likelihood ratio test should be interpreted with caution in light of the strong cross-sectional correlation in residuals, the likelihoods overwhelmingly point toward evidence in favor of both time-varying factor loadings and volatility.

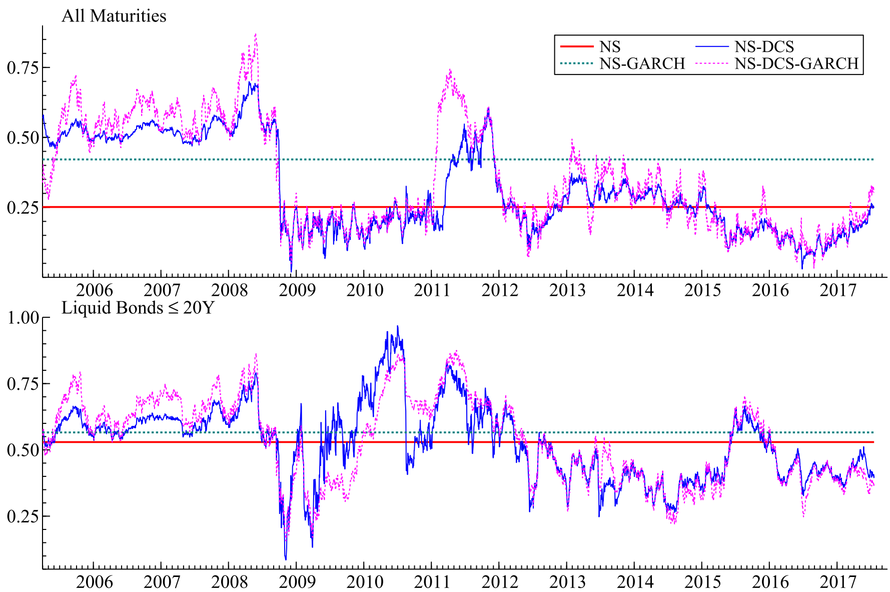

Figure 4. Time series of λ_t. This figure plots in-sample filtered path of λ_t for the four different models. The top panel provides the estimates based on maturities up to 50 years, while the bottom panel only uses maturities up to the last liquid point of 20 years.

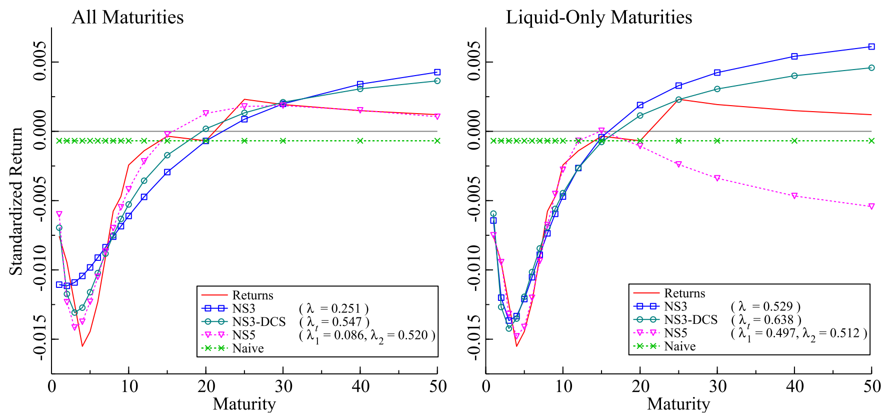

Figure 5. NS and NS-DCS fit on October 12, 2011. The two figures plot the maturity-standardized returns ρ_t against the models' fitted values $b_t f_t$ for the constant λ NS model, as well as the dynamic score-driven NS' λ_t.

The left-hand panel presents an in-sample illustration. It uses all maturities, both for estimation of the parameters and for fitting the return curve. The right-hand panel is cross-sectionally out-of-sample. It uses the liquid maturities up to 20 years only, both for estimation of the model parameters and for fitting the return curve. Both panels use the full-sample parameter estimates associated with Figure 4.

We have chosen a representative day at the end of 2011. During this period, all four time-varying parameter models in Figure 4 largely agree on the level of λ_t, but their estimates do differ from the constant NS models. On this particular day, the NS-DCS estimates $\lambda_t = 0.547$ versus 0.251 of the constant λ model, and is close to the higher of the two lambdas of the five-factor model. On the liquid only cross-section the three models' estimates of λ are close.

First consider the left-hand panel. It is clear that the three-factor NS model on standardized returns is able to capture the cross-sectional return variation well, provided the estimated factor loadings are accurate. While the NS-DCS almost perfectly captures the returns across maturities, the constant λ estimate provides a poor fit, clearly missing the point where the curve factor should be at its maximum. Both three factor models fail to capture the concavity of the returns on the long maturity. The constant λ five factor uses the first set of slope and curvature factors to fit the first 10 years of data, and the second set to fit the concavity in the long end, which cannot be accommodated by just three factors.

The right-hand panel presents the cross-sectional prediction. Note that this is the setting typically encountered in practice; liabilities of pension funds reach up to 80 years, where no instruments are available. Estimating the factor loadings on the liquid maturities only ensures that all models perfectly fit the maturities up to 15 years. On this particular day the 20-year maturity return already displays some concavity, such that the two three-factor models fail to track the full curve up to twenty years, and the five-factor model still offers benefits. However, the extrapolation to long-maturity returns of all the models is extremely poor. The flattening out of the return curve is either under- or over-estimated. We find similar patterns on most other days, although the discrepancy between the predictions and the actual return differs in both sign and magnitude.

The reasonable performance of the naive strategy on this day is not exceptional. The flattening out of the curve at long maturities is a stylized fact in the data, and the NS tend to extrapolate to too extreme values. The naive strategy's prediction is closer to the 50-year maturity return than any of the three models on 45% of the days, and closer than at least one model on 88% of the days.

5.2 Hedging performance

We now turn to the actual hedging performance of the NS models with time-varying factor loadings. Table 6 presents the hedge results of the various dynamic three-factor NS specifications along with the naive strategy. By and large the dynamic specifications of NS (the DCS versions) are very similar to the models with constant λ, whose results were presented in Table 4. During the turbulent financial crisis period in 2008–2009, the dynamic versions, estimated on the full cross-section, perform even slightly worse. As shown in Figure 4, this is the period when λ_t is small, which leads to poor identification of the factors and large hedging variance. So especially when the time variation is most relevant, it is difficult to

Table 6. Empirical hedging results: Time-varying parameters

	Nelson-Siegel: All				Nelson-Siegel: Liquid Only				Other	
	NS	NS-DCS	NS-GARCH	NS-DCS-GARCH	NS	NS-DCS	NS-GARCH	NS-DCS-GARCH	Naive	No Hedge
Panel A: Full Sample (April 2005 to July 2017)										
Bias	−0.01	−0.03	−0.01	−0.03	−0.01	0.00	−0.02	−0.01	−0.01	0.05
StDev	1.25	1.49	1.17	1.38	1.15	1.14	1.16	1.13	1.02	2.44
RMSE	1.25	1.49	1.17	1.39	1.15	1.14	1.16	1.13	1.02	2.44
TW	10.75	13.81	6.56	14.77	4.32	5.95	4.04	5.88	2.50	0.00
CO	18.48	21.67	13.50	22.68	10.82	12.64	10.43	12.62	2.50	0.00
TO	0.64	6.11	0.53	4.91	0.42	5.06	0.42	2.63	0.02	0.00
Panel B: Subsample Hedging Performance										
	April 2005 to December 2007									
Bias	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	−0.01
StDev	0.53	0.53	0.52	0.52	0.52	0.52	0.52	0.52	0.50	1.57
RMSE	0.53	0.53	0.52	0.53	0.52	0.52	0.52	0.52	0.50	1.57
	January 2008 to December 2009									
Bias	0.04	−0.12	0.04	−0.04	0.04	0.11	0.04	0.05	0.01	0.10
StDev	1.90	2.71	1.95	2.05	1.94	1.86	1.97	1.86	1.68	3.48
RMSE	1.90	2.72	1.95	2.05	1.94	1.87	1.97	1.86	1.68	3.48
	January 2010 to July 2017									
Bias	−0.03	−0.02	−0.04	−0.05	−0.04	−0.03	−0.04	−0.03	−0.03	0.06
StDev	1.22	1.24	1.07	1.39	1.03	1.04	1.03	1.03	0.93	2.37
RMSE	1.22	1.24	1.07	1.39	1.03	1.04	1.03	1.03	0.93	2.37

Notes: The table reports the hedge performance for the factor hedge portfolios implied by different versions of the NS. The NS models differ according to the specification for λ_t (DCS = time-varying), the inclusion of GARCH effects, and the maturities used for estimation. Performance is measured by the average, standard deviation, and RMSE of the prediction error $r_{0t} - B_{0t}\hat{f}_t$. Estimates for the NS models are based on a rolling sample of 1500 past days. The "No Hedge" benchmark refers to the model that predicts zero excess returns for all t. As a summary of the portfolio composition the table reports averages of the statistics $\text{TW}_t = \imath' w_t$, $\text{CO}_t = \sqrt{w_t' w_t}$, and $\text{TO}_t = \sum_i |w_{it} - w_{i,t-h} e^{r_{i,t-h} - \hat{r}_{0,t-h}}|$.

exploit. In this period long rates behave different from the level factor that is present in shorter maturities. The DCS models try to capture this with a low λ_t, but the behavior of very long rates is difficult to predict using the shorter term instruments. The DCS model estimated on the limited cross-section does much better in this period. These models also adapt to allow for lower than average λ_t, but as Figure 4 has shown, the estimated λ_t's remain relatively large, reducing the variance of the out-of-sample estimates.

The improvements in hedging performance gained from allowing for time-varying factor loadings are still dwarfed by the the naive hedging approach. Moreover, the increased performance of the NS-DCS models comes at the cost of higher turnover and more extreme portfolios, while the naive strategy by construction has low turnover and moderate portfolio positions.

6 Hedging a Portfolio of Liabilities

In practice, merely hedging a single 50-year liability is an extreme scenario. In this section, we consider the factor and optimal hedge portfolios for a realistic range of liabilities that a pension fund may face. In particular, we use a snapshot of the liabilities of two anonymized pension funds obtained from the Dutch Central Bank. The first is a "gray" pension fund which has both short term liabilities for retirees, and long-dated liabilities for the active participants. The second is a "green" fund whose average liability duration is longer.

Determining the hedge portfolios for a portfolio of liabilities is straightforward within the sphere of factor models. Let J be the number of liability maturities, with maturities $\tau_{L,j}$. The (scaled) liability weights are standardized to sum to 1 and gathered in the vector w_L. The factor exposure of the portfolio of liabilities is then simply

$$b_{Lt} = \sum_{j=1}^{J} w_{L,j} b_t(\tau_{L,j}). \tag{33}$$

The factor loadings of the liquid maturities available for portfolio construction are $b_t \equiv b_t(\tau)$ as before. The factor-hedge portfolio then follows as

$$g_{Lt} = b_t'(b_t'b_t)^{-1} b_{t,L}^*. \tag{34}$$

Similarly, the MSE-optimal portfolio is of the form

$$g_{Lt}^{Opt} = (b_t \Omega b_t' + \sigma^2 I)^{-1} b_t \Omega b_{Lt}'. \tag{35}$$

We compute the factor portfolio weights for each of the pension funds over time using the λ_t forecasts from the three-factor NS model estimated on the liquid maturities only. The MSE-optimal portfolios are based on the five-factor NS forecasts along with rolling estimates of Ω and σ.

Figure 6 plots the median and 95% quantiles over time of the portfolio allocations for both the factor and optimal allocations, for the "gray" and the "green" fund. The general structure of the portfolios is in line with those in Figure 3, but the magnitude differs. The portfolio of liabilities with differing maturities results in slightly less extreme positions than the 50-year liability only. Indeed, the average portfolio concentration measure, also reported in Table 4, decreases to 8.6 and 13.2 for the factor and optimal portfolios of the gray fund. The "green" fund's portfolio is more extreme, since it has to leverage up the last liquid point to cover the maturities beyond the last liquid point. For the "green" fund, 78% of liabilities has maturity over 20 years, while this is only 55% for the "gray" fund. This results in average portfolio concentrations of 11.0 and 24.8 for the factor and optimal portfolios, respectively. Interestingly the optimal portfolio is still remarkably close to the naive portfolio. Given the average duration of 23.9 and 33.1 for the gray and green fund, the naive duration hedge would suggest a scaled position of 0.48 and 0.66 in the 20-year liability, which is remarkably close to the position put forth by the optimal portfolio.

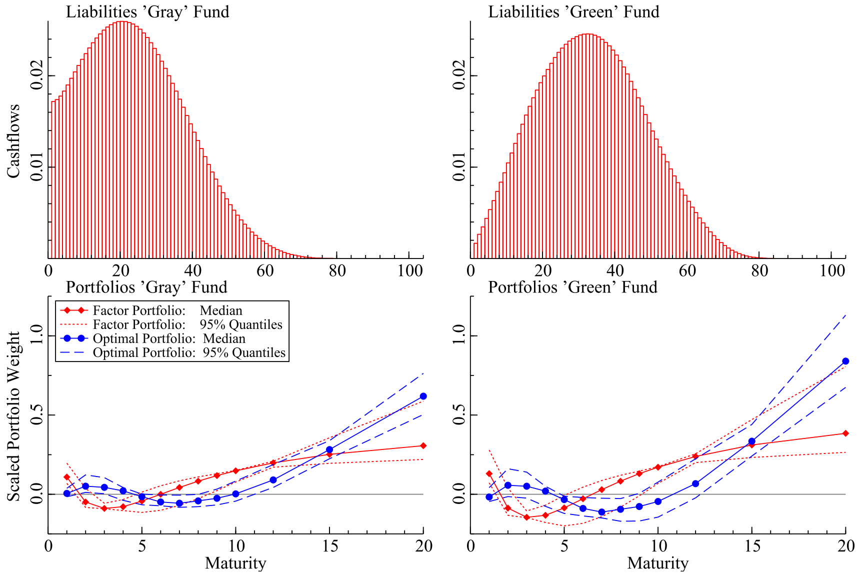

Figure 6. Empirical distribution of hedge portfolio weights for pension funds. The graph shows the median of the time series of scaled weights $h(\tau_i)$ implied by the DSC-NS model estimated on the liquid maturities only for two portfolios of liabilities. The top row presents the percentage weights in each maturity of liabilities. The second row presents the hedge weights, along with empirical 5 and 95% percentiles of the time series distribution of the weights.

7 Conclusion

We consider the problem of risk management of long-dated liabilities. Yields are typically assumed to follow a factor model, which translates in a similar factor structure on returns. The standard approach to hedging risk would be to construct a factor-mimicking portfolio with factor loadings equal to those of the liability. The hedging performance is therefore invariably linked to a model's ability to accurately estimate these factor loadings. In our candidate NS model, these loadings are driven by a single parameter λ. We consider two alternative extensions to the standard NS model for our purpose: a version with two additional factors with a corresponding second parameter, and a version with time-varying factor loadings λ_t. All strategies linked to the factor models are empirically beaten by a simple naive hedge of an appropriately levered position in the last liquidity.

We show empirically and analytically that strong multicollinearity in the factor loadings makes the empirical implementation of a factor-mimicking strategy ineffective. Even though in-sample fit of the five-factor model is far better than the three-factor model's, strong multicollinearity in the loadings results in prohibitive estimation error of factors. As a result, the five-factor hedged portfolio actually has almost three times higher RMSE than the completely unhedged liability. The three-factor model, constant loading model, while inevitably misspecified, has far superior hedging performance. Moreover, since the standard three-factor NS model is severely misspecified for long maturities, the parameters of the model are better estimated using assets up to 20 years, even if the objective is to hedge long-dated liabilities beyond those 20 years.

Balancing the tradeoff between misspecification bias and estimation error, we consider an RMSE-optimal portfolio, which uses the covariance matrix of the factors in addition to the loadings. This optimal portfolio is remarkably close to the "naive" approach. The naive approach does not require model estimation, and works surprisingly well, both theoretically and empirically, where it reduces the RMSE by almost 59% relative to the unhedged position.

We bridge the gap between the three- and five-factor model by proposing a three-factor model with time-variation in the factor loadings. We show that time variation in the parameter is large. We document that the parameter is relatively stable at the beginning of the sample which starts in 1999, but becomes highly volatile during the financial crisis of 2008, and remains more volatile after. The five-factor model has a low and a high value of λ, such that similar time variation in the loadings can effectively be achieved by dynamics in the factors themselves. The time-varying parameter models increase the hedging performance of the three-factor NS model, but are still easily beaten by the robust naive approach.

The poor performance of the factor models naturally opens the door for alternative approaches to construct a hedge for long-term liabilities.

Appendix A. Construction of Yield Curve

Daily data on euro swap interest rates are from Datastream using codes ICEIBsY, where s indicates the maturity in years. Available swap maturities are $1,\ldots,10$, 12, 15, 20, 25 30, 40, and 50 years. Daily data are available from January 1999, except for the 40- and 50-year rates, which are only available from March 2005. Euro swap rates are quoted on an annual bond 30/360 basis against 6 months Euribor.[9] To construct discount rates and excess returns we use Euribor rates with maturities 1 week (EIBOR1W) and 6 months (EIBOR6M), also from Datastream.

From the swap rates, we construct discount bond prices. Let S_1 be the one-year swap rate and $P_{1/2}$ the discount factor implied by the Euribor rate with maturity 6 months. From the one-year swap rate and the shorter term Euribor rate, the one-year discount factor P_1 is found by setting the present value of the fixed leg equal to one, that is, P_1 solves

$$1 = \frac{1}{2} S_1 \left(P_{1/2} + P_1 \right) + P_1. \tag{A.1}$$

The next swap rate is S_2, which should satisfy

$$1 = \frac{1}{2} S_2 (P_{1/2} + P_1 + P_{3/2} + P_2) + P_2. \tag{A.2}$$

We know $P_{1/2}$ and P_1, but not $P_{3/2}$. Therefore we need to make an assumption. One way to obtain $P_{3/2}$ is to first construct an artificial $S_{3/2}$ by interpolation. Then the 18-month discount factor is found as

$$P_{3/2} = \frac{1 - \frac{1}{2} S_{3/2} (P_{1/2} + P_1)}{1 + \frac{1}{2} S_{3/2}}. \tag{A.3}$$

9 On April 1, 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.

From here we can compute P_2. Doing the interpolation for all swap rates at half-year intervals, we construct the entire yield curve from the recursion

$$P_{s/2} = \frac{1 - \frac{1}{2} S_{s/2} \sum_{i=1}^{s-1} P_{i/2}}{1 + \frac{1}{2} S_{s/2}}. \tag{A.4}$$

The swap rates are interpolated using a local cubic interpolation to obtain rates with semi-annual maturity intervals. Except for the 3/2-year maturity, and the very long ones beyond 40 years, interpolation is based on four adjacent observed swap rates, two with maturities less than $j/2$ and two with maturities larger than $j/2$. For the very short and very long maturities, the interpolation is based on one maturity on one side, and three maturities on the other.

From the discount bond prices bootstrapped this way we compute the returns $r_t(\tau)$ from log prices using the definition in Equation (5). As risk-free, we subtract the one-week Euribor rate as a short-term yield, scaled to a daily interval by $p_t(h) = -y_t(1w)/252$.

Appendix B. Scores and Information

For the DCS updating we need the derivatives of the log likelihood with respect to λ. To avoid cluttered notation we only provide explicit formulas for the case with constant σ^2. The log-likelihood for a single time series observation is

$$\ell_t(\lambda, f_t) = -\frac{1}{2}\left(N\ln\sigma^2 + \frac{\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t}{\sigma^2}\right), \tag{B.1}$$

where $\boldsymbol{\varepsilon}_t = \boldsymbol{R}_t - \boldsymbol{b}f_t$, and $\boldsymbol{b}$ depends on λ. Maximizing with respect to f_t gives the factor portfolio $\hat{f}_t = (\boldsymbol{b}'\boldsymbol{b})^{-1}\boldsymbol{b}'\boldsymbol{R}_t$ and the concentrated log-likelihood for observation t as

$$\ell_t^*(\lambda) = -\frac{1}{2}\left(N\ln\sigma^2 + \frac{\boldsymbol{R}_t'\boldsymbol{M}\boldsymbol{R}_t}{\sigma^2}\right), \tag{B.2}$$

with $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{b}(\boldsymbol{b}'\boldsymbol{b})^{-1}\boldsymbol{b}'$. The scores are

$$z_t = \frac{\partial \ell_t^*}{\partial \lambda} = -\frac{1}{2\sigma^2}\boldsymbol{R}_t'\frac{\partial \boldsymbol{M}}{\partial \lambda}\boldsymbol{R}_t. \tag{B.3}$$

Let $\boldsymbol{c} = \frac{\partial \boldsymbol{b}}{\partial \lambda}$. Then we have

$$\frac{\partial \boldsymbol{M}}{\partial \lambda} = -\boldsymbol{c}(\boldsymbol{b}'\boldsymbol{b})^{-1}\boldsymbol{b}' - \boldsymbol{b}(\boldsymbol{b}'\boldsymbol{b})^{-1}\boldsymbol{c}' + \boldsymbol{b}(\boldsymbol{b}'\boldsymbol{b})^{-1}(\boldsymbol{b}'\boldsymbol{c} + \boldsymbol{c}'\boldsymbol{b})(\boldsymbol{b}'\boldsymbol{b})^{-1}\boldsymbol{b}' \tag{B.4}$$

and, using $\hat{f}_t = (\boldsymbol{b}'\boldsymbol{b})^{-1}\boldsymbol{b}'\boldsymbol{R}_t$ and $\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{R}_t - \boldsymbol{b}\hat{f}_t$,

$$\begin{aligned}\boldsymbol{R}_t'\frac{\partial \boldsymbol{M}}{\partial \lambda}\boldsymbol{R}_t &= -\boldsymbol{R}_t'\boldsymbol{c}\hat{f}_t - \hat{f}_t'\boldsymbol{c}'\boldsymbol{R}_t + \hat{f}_t'(\boldsymbol{b}'\boldsymbol{c} + \boldsymbol{c}'\boldsymbol{b})\hat{f}_t \\ &= -2\hat{\boldsymbol{\varepsilon}}_t'\left(\boldsymbol{c}\hat{f}_t\right).\end{aligned} \tag{B.5}$$

Therefore we obtain the scores

$$z_t = \frac{1}{\sigma^2}\hat{\boldsymbol{\varepsilon}}_t'\left(\boldsymbol{c}\hat{f}_t\right). \tag{B.6}$$

The expression has the natural implication that the residuals should be orthogonal to the derivative of the fitted values.

For the information matrix we need the second-order derivatives. At each time t we have, defining $\boldsymbol{d} = \frac{\partial \boldsymbol{c}}{\partial \lambda}$,

$$\frac{\partial^2 \ell_t^*}{\partial \lambda^2} = \frac{1}{\sigma^2}\left(\hat{\boldsymbol{\varepsilon}}_t' \boldsymbol{d}\hat{f}_t - \left(\boldsymbol{c}\hat{f}_t\right)'\left(\boldsymbol{c}\hat{f}_t\right) + \left(\boldsymbol{c}'\hat{\boldsymbol{\varepsilon}}_t - \boldsymbol{b}'\boldsymbol{c}\hat{f}_t\right)'(\boldsymbol{b}\prime\boldsymbol{b})^{-1}\left(\boldsymbol{c}'\hat{\boldsymbol{\varepsilon}}_t - \boldsymbol{b}'\boldsymbol{c}\hat{f}_t\right)\right). \tag{B.7}$$

An exact value for the information, the expected value of the negative second-order derivative, is not available. As an approximation, and in order to guarantee that it is positive for all λ, we assume that the cross-sectional averages $\boldsymbol{c}'\hat{\boldsymbol{\varepsilon}}_t$ and $\boldsymbol{d}'\hat{\boldsymbol{\varepsilon}}_t$ are close to zero, such that

$$\mathcal{I}_t = \frac{1}{\sigma^2}\left(\boldsymbol{c}\hat{f}_t\right)'\boldsymbol{M}\left(\boldsymbol{c}\hat{f}_t\right). \tag{B.8}$$

Appendix C. Standard Errors

In essence we have a panel data model that we estimate by quasi-maximum likelihood, assuming normality and treating the errors as cross-sectionally uncorrelated. Since the latter is a potential source of misspecification, we compute standard errors that are robust to omitted cross-sectional correlation. This can be done by only using the time-series variation in the scores.

To compute standard errors, either in the model with constant λ or in the DCS model, we evaluate the scores $\hat{z}_t$ at the QML estimates. The variance is estimated from using the Newey–West formula

$$S = \frac{1}{T}\sum_t \sum_{j=-J}^{J}\left(1 - \frac{j}{J+1}\right)\hat{z}_t \hat{z}_{t+j}'. \tag{C.1}$$

For the standard error of $\hat{\lambda}$ we further need to linearize the total gradient $\sum_t z_t$ around its true value λ. An explicit formula for the second order derivatives in the constant λ model is given in Equation (B.7). For the general case we can use the chain rule to take derivatives for λ_t with respect to $\overline{\lambda}$, ϕ_1 and ϕ_2 following the DCS specification. Summing over t then gives the negative Hessian H. Given the Hessian and the covariance matrix of the scores, the asymptotic variance of the parameters is

$$V = \frac{1}{T}H^{-1}SH^{-1}. \tag{C.2}$$

When σ^2 is unknown it can be replaced by the consistent estimator,

$$\hat{\sigma}^2 = \frac{1}{NT}\sum_t \boldsymbol{R}_t{}'\boldsymbol{M}\boldsymbol{R}_t, \tag{C.3}$$

in both S and H, since it is asymptotically independent of λ. Computing standard errors for the general case with a GARCH process for σ_t^2 is a straightforward extension. For the more complicated models computation of standard errors proceeds in the same way. Again, we use the robust formula and only consider the time-series variation of the scores. The scores z_t and Hessian have additional elements for the GARCH parameters $\overline{\sigma}^2$, α_1 and α_2.

Appendix D. Parameter Estimates NS-DCS-GARCH Models

Table D.1. Parameter estimates: Three-factor NS

	All maturities				Liquid bonds $\leq 20Y$			
	NS	NS DCS	NS GARCH	NS DCS GARCH	NS	NS DCS	NS GARCH	NS DCS GARCH
$\lambda/\bar{\lambda}$	0.2514	0.2887	0.4215	0.3911	0.5293	0.4743	0.5659	0.4691
	(0.0386)	(0.2950)	(0.0523)	(0.0523)	(0.0306)	(0.0138)	(0.0192)	(0.0040)
ϕ_1		0.9944		0.9938		0.9932		0.9954
		(0.0194)		(0.0003)		(0.0003)		(0.0003)
ϕ_2		0.0064		0.0083		0.0098		0.0087
		(0.0004)		(0.0003)		(0.0006)		(0.0001)
$\sigma^2/\bar{\sigma}^2$	0.7377	0.7000	1.6451	0.9599	0.5225	0.5151	19.624	0.0491
	(0.0767)	(0.1858)	(0.7621)	(0.1898)	(0.0600)	(0.0711)	(6.8642)	(0.0090)
α			0.6900	0.5299			0.7567	0.7218
			(0.1297)	(0.1760)			(0.0582)	(0.0145)
β			0.2979	0.3981			0.2429	0.1547
			(0.1180)	(0.0746)			(0.0568)	(0.0485)
$\mathcal{L}$	182173	183602	197720	200584	146505	146805	161485	161968

Notes: This table provides parameter estimates, with standard errors in brackets, for four different three-factor NS models. We report estimates 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 20 years. σ^2 and $\bar{\sigma}^2$ are multiplied by 10^4. The maximized value of the log-likelihood is denoted by $\mathcal{L}$. Details on the standard errors are provided in Appendix C.

References

Adrian, T., R. K. Crump, and E. Moench. 2013. Pricing the Term Structure with Linear Regressions. *Journal of Financial Economics* 110: 110–138.

Almeida, C., K. Ardison, D. Kubudi, A. Simonsen, and J. Vicente. 2018. Forecasting Bond Yields with Segmented Term Structure Models. *Journal of Financial Econometrics* 16: 1–33.

Bams, D., and P. Schotman. 2003. Direct Estimation of the Risk Neutral Factor Dynamics of Gaussian Term Structure Models. *Journal of Econometrics* 117: 179–206.

Bauer, M. D. 2018. Restrictions on Risk Prices in Dynamic Term Structure Models. *Journal of Business & Economic Statistics* 36: 196–211.

Bianchi, F., H. Mumtaz, and P. Surico. 2009. The Great Moderation of the Term Structure of UK Interest Rates. *Journal of Monetary Economics* 56: 856–871.

Blasques, F., S. J. Koopman, and A. Lucas. 2015. Information-Theoretic Optimality of Observation-Driven Time Series Models for Continuous Responses. *Biometrika* 102: 325–343.

Breitung, J., and S. Eickmeier. 2011. Testing for Structural Breaks in Dynamic Factor Models. *Journal of Econometrics* 163: 71–84.

Christensen, J. H., J. A. Lopez, and P. L. Mussche. 2019. "Extrapolating Long-Maturity Bond Yields for Financial Risk Measurement." *Federal Reserve Bank of San Francisco* Working paper 2018-09, available at https://doi.org/10.24148/wp2018-09

Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch. 2009. An Arbitrage-Free Generalized Nelson–Siegel Term Structure Model. *Economic Journal* 12: 33–64.

Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch. 2011. The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models. *Journal of Econometrics* 164: 4–20.

Committee of European Insurance and Occupational Pensions Supervisors. 2010. QIS5: Risk Free Interest Rates – Extrapolation Method, eiopa.europa.eu. Available at http://wayback.archive-it.org/org-1495/20170316005345/http://archive.eiopa.europa.eu/fileadmin/tx_dam/files/consultations/QIS/QIS5/ceiops-paper-extrapolation-risk-free-rates_en-20100802.pdf, last accessed July 15, 2020.

Creal, D., S. J. Koopman, and A. Lucas. 2013. Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics* 28: 777–795.

Dai, Q., and K. J. Singleton. 2000. Specification Analysis of Affine Term Structure Models. *The Journal of Finance* 55: 1943–1978.

Dai, Q., and K. J. Singleton. 2002. Expectation Puzzles, Time-Varying Risk Premia, and Dynamic Models of the Term Structure. *Journal of Financial Economics* 63: 415–441.

Diebold, F. X., L. Ji, and C. Li. 2006a. "A Three-Factor Yield Curve Model: Non-Affine Structure, Systematic Risk Sources and Generalized Duration." In L. Klein (ed.), *Long-Run Growth and Short-Run Stabilization: Essays in Memory of Albert Ando*, pp. 240–274. Cheltenham: Edward Elgar.

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

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

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

DNB Bulletin. 2013. Pension Sector Hedges Half the Interest Rate Risk. Available at http://www.dnb.nl/en/news/news-and-archive/dnbulletin-2013/dnb295970.jsp, last accessed July 15, 2020.

Domanski, D., H. Shin, and V. Sushko. 2017. The Hunt for Duration: Not Waving but Drowning?. *IMF Economic Review* 65: 113–130.

Dubecq, S., and C. Gourieroux. 2011. "An Analysis of the Ultra Long-Term Yields." SSRN Working Paper 1943535, available at 10.2139/ssrn.1943535

Duffee, G. R. 2002. Term Premia and Interest Rate Forecasts in Affine Models. *The Journal of Finance* 57: 405–443.

Duffie, D., and R. Kan. 1996. A Yield Factor Model of Interest Rates. *Mathematical Finance* 6: 379–406.

European Insurance and Occupational Pensions Authority. 2014a. Consultation Paper on a Technical Document Regarding the Risk-free Interest Rate Term Structure, Report EIOPA-CP-14/042.

European Insurance and Occupational Pensions Authority. 2014b. Technical Specification for the Preparatory Phase (Part I), Technical Report EIOPA 14/209.

Faria, A., and C. Almeida. 2018. A Hybrid Spline-Based Parametric Model for the Yield Curve. *Journal of Economic Dynamics and Control* 86: 72–94.

Goliński, A., and P. Spencer. 2017. The Advantages of Using Excess Returns to Model the Term Structure. *Journal of Financial Economics* 125: 163–181.

Hautsch, N., and Y. Ou. 2012. Analyzing Interest Rate Risk: Stochastic Volatility in the Term Structure of Government Bond Yields. *Journal of Banking & Finance* 36: 2988–3007.

Heath, D., R. Jarrow, and A. Morton. 1992. Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claim Valuation. *Econometrica* 60: 77–105.

Hevia, C., M. Gonzalez-Rozada, M. Sola, and F. Spagnolo. 2015. Estimating and Forecasting the Yield Curve Using a Markov Switching Dynamic Nelson and Siegel Model. *Journal of Applied Econometrics* 30: 987–1009.

Koopman, S., M. Mallee, and M. Van der Wel. 2010. Analyzing the Term Structure of Interest Rates Using the Dynamic Nelson - Siegel Model with Time-Varying Parameters. *Journal of Business & Economic Statistics* 28: 329–343.

Krippner, L. 2015. A Theoretical Foundation for the Nelson-Siegel Class of Yield Curve Models. *Journal of Applied Econometrics* 30: 97–118.

Litterman, R., and J. Scheinkman. 1991. Common Factors Affecting Bond Returns. *The Journal of Fixed Income* 1: 54–61.

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

Yu, W. C., and E. Zivot. 2011. Forecasting the Term Structures of Treasury and Corporate Yields Using Dynamic Nelson-Siegel Models. *International Journal of Forecasting* 27: 579–591.