Immunization with term structure dynamics*

Daniel Borup† Bent Jesper Christensen**

Jorge W. Hansen†† Torben B. Rasmussen‡

Abstract

We develop an immunization strategy that optimally exploits no-arbitrage in conjunction with time-series restrictions stemming from consistency between yield curve shapes and dynamic term structure models. This approach provides notable improvements in hedging performance compared to traditional and purely cross-sectional approaches based on factor analysis or a parametrized yield curve shape. We also introduce a new three-factor dynamic term structure model with level, slope and curvature varying stochastically, which leads to further improvements in hedging performance. A restricted version of this stochastic level, slope and curvature (SLSC) model belongs to the three-factor affine term structure model class.

Keywords: Immunization; interest rate model; consistency; factor models; no-arbitrage; generalized duration matching

JEL Classification: C32, C38, E43, G11, G12

This version: November 30, 2018

*We are grateful to Francis X. Diebold, Bob Jarrow, Christian M. Dahl and participants at the econometrics workshop at University of Pennsylvania and research seminars at Aarhus University for useful comments and suggestions, and to the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation, DNRF78), the Dale T. Mortensen Center, Aarhus University, and the Danish Social Science Research Council for research support. Some of this research was carried out while Borup and Hansen were visiting the Department of Economics, University of Pennsylvania, and Christensen was visiting the Department of Economics, Harvard University, and the generosity and hospitality of both departments are gratefully acknowledged.

†Aarhus University and CREATES. Email: dborup@econ.au.dk.

**Aarhus University and CREATES. Email: bjchristensen@econ.au.dk.

††Aarhus University and CREATES. Email: jh@econ.au.dk.

‡Aarhus University and CREATES. Email: trasmussen@econ.au.dk

I. Introduction

Traditional immunization strategies introduced by Redington (1952) and later revived in applications by Fisher and Weil (1971) amount to matching the basic bond duration measure of Macaulay (1938) across assets and liabilities. The subsequent generalized duration approach (e.g., Ingersoll (1983) and Nelson and Schaefer (1983)) relies implicitly or explicitly on a multivariate factor structure underlying market yields, matching each of a number of factor loadings or sensitivities across the immunization target and a hedge portfolio. This approach still enjoys frequent use (see e.g. Diebold, Ji, and Li (2006) for an application)[1] and is largely cross-sectional in nature, much as in the case of stock portfolio management. However, in the bond market case, when explicitly considering movements in the entire yield curve across calendar time, consistency (in the sense of Björk and Christensen (1999)) with arbitrage-free dynamic term structure models (DTSMs) places rather tight cross-restrictions on the admissible cross-sectional loading functions used in the construction of hedge portfolios and the dynamic behavior of market interest rates. Although ignoring these restrictions potentially leads to costly lack of precision in immunization strategies, they have not been explored in the literature. In this paper, we develop an approach that allows us to optimally exploiting term structure dynamics in hedging strategies, and we document the resulting improvement in performance.

The objective already in the early literature is to develop portfolio management, trading, and hedging strategies to allocate funds across a set of traded assets at each point in time. Typically, the focus has been on the contemporaneous correlation between yields and the construction of portfolio rules, while disregarding important dynamic aspects of term structure movements. Litterman and Scheinkman (1991) is a leading example of a cross-sectional approach using the classical statistical factor analysis as first stage in term structure hedging. Their analysis shows that three factors adequately describe the term structure and from their interpretation of the way in which factor loadings vary

[1] Other approaches to bond portfolio risk management has been proposed in the literature. Some notable contributions are the key-rate duration Ho (1992), contingent immunization Leibowitz and Weinberger (1981, 1982, 1983); Díaz, Gonzáez, Navarro, and Skinner (2009), M-square and M-vector models Chambers, Carleton, and McEnally (1988); Nawalkha and Chambers (1997); Nawalkha, Soto, and Zhang (2003), infinite-dimensional immunization Bueno-Guerrero, Moreno, and Navas (2015), and cross-shape risk Galluccio and Roncoroni (2006). For additional applications of some of above methods, see e.g. Carcano and Dall'O (2011) who extend models based on principal components, duration vectors and key rate duration to exploit model estimation error, Soto (2004) and Bravo and Silva (2006) who investigate the question of dimension of the models underlying a given risk management technique, and Agca (2005) who provides simulation-based evidence on the hedging performance of Heath-Jarrow-Morton model-based risk measures.

across instruments of different maturities, the three factors are labelled level, slope, and curvature, respectively. In fact, Nelson and Siegel (1987) interpret their parsimoniously parametrized yield curve shapes with monotonic and hump components in terms of short-, medium-, and long-term factors that correspond quite precisely to the Litterman and Scheinkman (1991) slope, curvature, and level factors, respectively. From this viewpoint, it would seem natural to impose the type of functional form restrictions considered in the Nelson and Siegel (1987) curve shape framework directly on the loadings estimated in the factor analysis as input into hedging and immunization strategies. Apparently, this would afford reduction in complexity through savings in degrees of freedom, and, under correct specification, efficiency gains in estimation and, potentially, hedging performance.

In this paper, we show that the approach of combining the factor analysis and associated hedging technology with the loading restrictions from level, slope, and curvature curve shapes may in fact be improved upon. The key to this result is that both the factor analysis structure and the parsimonious curve shape approach are entirely cross-sectional in nature and void of financial theory,[2] in particular the restrictions on the evolution through time of curve shapes stemming from no-arbitrage and consistency with DTSMs.[3] For example, the Nelson and Siegel (1987) (henceforth referred to as NS) yield curve shape is inadmissible, since arbitrage opportunities exist if the yield curve is in this class in each time period. After all, then, this functional form should not be imposed in estimation and hedging applications.

Our approach in this paper is to exploit the implications of consistent arbitrage-free interest rate dynamics on the generalized durations used in forming the hedge portfolio. We do this in three steps. First, we consider the effect of replacing the curve shape necessarily leading to arbitrage opportunities with one that possesses arbitrage-free dynamics. Second, we focus on the consistent arbitrage-free dynamics rather than on the curve shape in the calculation of the generalized durations. Third, we combine the two,

[2] We note that Krippner (2015) establish a theoretical foundation for the Nelson-Siegel class of yield curve models within the Gaussian affine term structure model class, seeking to legitimize the use of the former by the theoretical underpinnings of the latter. This foundation is, however, based on low-order approximations, whose empirical foundation is mixed. See also Coroneo, Nyholm, and Vidova-Koleva (2011).

[3] Joslin, Singleton, and Zhu (2011) conclude that in Gaussian DTSMs, no-arbitrage restrictions are not useful, *per se*, in forecasting the yield curve and attribute any empirical forecasting gain in these models to the combination structure imposed of both no-arbitrage and their restrictions on the distribution of yields under the objective probability measure. Our focus differs from theirs, since we evaluate the implication of no-arbitrage conditions and consistency on hedging performance of a bond portfolio using a wider set of models, not necessarily Gaussian DTSMs.

thus exploiting both the arbitrage-free dynamics and the yield curve shape consistent with this in deriving the hedge portfolio. These three steps are implemented both in a case with a single stochastic factor, and in a more general situation involving three stochastic factors, corresponding to level, slope, and curvature. The single-factor analysis combines a modified version of the NS curve shape consistent with no-arbitrage and the dynamic model of Hull and White (1990). The three-factor analysis is based on a new stochastic level-slope-curvature (SLSC) model unique to the present paper.

The empirical performance of existing and new approaches is investigated using weekly yield data from the Federal Reserve Bank (Fed) over the period 1983 through 2014. The targets for assessing hedging performance are a 5-year coupon bond and a portfolio consisting of 2-year, 5-year, and 10-year coupon bonds with positive and negative weights, all based on information on contractual terms from CRSP. Parameters are estimated in the yield data and used to calculate generalized durations. Portfolios of zero-coupon bonds are constructed on a monthly basis to match estimated generalized durations. The resulting ability to hedge one-month returns on the targets is recorded over the full sample period. From the empirical results, an improvement in hedging performance is achieved by replacing the NS curve shape by a modification that is not inconsistent with arbitrage-free interest rate dynamics. A further improvement is obtained by estimating parameters directly from the Hull and White (1990) dynamics consistent with the modified curve shape. However, performance decreases when combining the modified curve shape and consistent dynamics. This suggests that for hedging purposes, one-factor dynamics do not adequately capture term structure movements. We therefore introduce our new three-factor SLSC model. It specifies both the yield curve shape and the dynamics of this, and the two are consistent. In the special case where the SLSC model is restricted to the three-factor affine class in the sense of Duffie and Kan (1996), it reduces to the independent factor model considered by Christensen, Diebold, and Rudebusch (2011). We show that with this restriction imposed, the model provides almost no improvement in hedging performance relative to using basic NS curves. In contrast, relaxing the restriction and applying our new general SLSC curve shape leads to better performance than any of the previously considered approaches. As in the one-factor case a further improvement is obtained by using the dynamics instead of the curve shape in parameter estimation. Finally, the additional restrictions imposed by the fully consistent model, combining curve shape and dynamics, lead to comparable or in some cases improved hedging performance in the three-factor case. These results indicate that important

improvements in trading strategies may be achieved by using a general model that is dynamically consistent and involves three stochastic factors, reflecting the time-varying level, slope, and curvature features of the bond market.

The remainder of the paper is laid out as follows. In Section II, we present the basic hedging framework, starting with the underlying yield factor model. We present the main theorem on the portfolio that minimizes hedging error variance subject to generalized duration matching and value matching. Moreover, we describe the data and document the benchmark empirical performance of existing hedging approaches which do not exploit dynamics. These are duration matching and generalized duration matching based either on an unrestricted factor analysis or on an analysis imposing parsimonious parametric structure (specifically, the NS shape) on loadings. In Section III, we turn our attention to dynamically consistent arbitrage-free hedging. First, we modify the parametrization of the restricted loadings to ensure the existence of a DTSM consistent with the chosen shape and reassess hedging performance. Following this, we change the estimation procedure to exploit only the dynamics of the consistent model. We then combine the two, exploiting both the shape of the yield curve and the dynamics of the consistent model. Subsequently, we introduce our new arbitrage-free DTSM with level, slope, and curvature factors that all three vary stochastically through time. We apply the three approaches, curve shape, dynamics, and the combination, to this model. Lastly, we consider an alternative approach, relaxing the generalized duration matching constraint and instead using the dynamic model to trade off violations of the constraint against remaining hedging error variance. Section IV concludes. All proofs are contained in an Internet Appendix.

II. The basic hedging framework

We first set out notation for the basic yield factor model. With $y_t = \left(y_{t,\tau_1}, \ldots, y_{t,\tau_m}\right)'$ a vector of (continuously compounded, zero-coupon) yields at time t with terms to maturity $\tau_1, \ldots, \tau_m$, the classical factor analysis structure is

$$y_t = \mu + B f_t + \varepsilon_t, \tag{1}$$

where μ is an m-vector of mean yields, f_t is a k-vector of common, unobserved covariance-generating factors, $k < m$, B is an $m \times k$ matrix of sensitivities or factor loadings, and ε_t is

an m-vector of idiosyncratic error terms independent of f_t. Thus, the individual yield is

$$y_{t,\tau_i} = \mu_i + b_i' f_t + \varepsilon_{t,i}, \tag{2}$$

where b_i' is the i'th row of B. The approach is cross-sectional in that factors and error terms are treated as serially uncorrelated. Estimation on panel data $(y_1, \ldots, y_T)$ produces estimates $\widehat{B}$ and $\widehat{\Psi}$ of B and $\Psi = \mathrm{var}(\varepsilon_t)$, assumed diagonal. The predicted factor scores are $\widehat{f}_t = \widehat{F} y_t$, with scoring matrix given as $\widehat{F} = \left(\widehat{B}'\widehat{\Psi}^{-1}\widehat{B}\right)^{-1}\widehat{B}'\widehat{\Psi}^{-1}$, i.e., the weighted regression of current yields y_t on the columns B_j of B, $j = 1, \ldots, k$, the weights being the diagonal elements $\psi_1, \ldots, \psi_m$ of Ψ. In particular, $\widehat{f}_t$ is linear in current yields.

Given the objective of obtaining linear hedging rules, we suppose that the target claim to be hedged is a future payment τ_* periods hence, and no zero-coupon bond with term to maturity τ_* (the ideal hedging instrument) is available. The factor loadings b_* (a $k \times 1$ vector) of the target claim (precisely, of the yield to the missing ideal hedge) may be obtained by interpolation between the available maturities. With the yield data ordered in such a way that $\tau_1 < \tau_2 < \ldots < \tau_m$, simple linear interpolation would set $b_* = ((\tau_{i+1} - \tau_*) b_i + (\tau_* - \tau_i) b_{i+1})/(\tau_{i+1} - \tau_i)$, a k-vector constructed as a suitable convex combination of the adjacent loading vectors b_i and b_{i+1}, where $\tau_i < \tau_* < \tau_{i+1}$. If the future payment occurs before the shortest maturity $\tau_* < \tau_1$, we set $b_* = b_1$ equal to the first factor loading vector. Similarly, if the futures payment occurs after the longest maturity $\tau_m < \tau_*$, we set $b_* = b_m$. The required (fitted) yield on the hedge portfolio is then $b_*' \widehat{f}_t$. This is a linear function of the available yields, too, namely, $b_*' \widehat{F} y_t$. Nonetheless, this does not constitute a portfolio rule in itself, as the traded instruments are bonds, not yields. Litterman and Scheinkman (1991) considered excess returns rather than yields, thus obtaining linear hedging rules. We present a related approach, first fitting parameters in a yield factor model, then using the estimated parameters to form policies (portfolio weights) linear in returns. Thus, we turn to implications of the yield factor model for returns next.

A. Hedge portfolios

In order to acquire linear portfolio rules, we go via the relation between zero-coupon bond prices and yields, $p_{t,\tau_i} = \exp\left(-\tau_i y_{t,\tau_i}\right)$, and define returns $r_{t+1,i} = (p_{t+1,\tau_i} - p_{t,\tau_i+1})/p_{t,\tau_i+1}$. Over short horizons we can approximate this by the return on a constant maturity

zero-coupon bond and use log-linearization,

$$r_{t+1,i} \approx \log \frac{p_{t+1,\tau_i}}{p_{t,\tau_i}} = -\tau_i \Delta y_{t+1,\tau_i}, \tag{3}$$

where Δ is the first difference operator, i.e., $\Delta y_{t+1,\tau_i} = y_{t+1,\tau_i} - y_{t,\tau_i}$ is the change in the τ_i maturity zero-coupon bond yield.[4] Diebold et al. (2006) apply a similar return approximation.

Applying the factor model for yields (2) in (3) we can express the return by

$$r_{t+1,i} = -\tau_i \left(b_i' \Delta f_{t+1} + \Delta \varepsilon_{t+1,i} \right). \tag{4}$$

This is recognized in the simple case $k = 1$ with f_t a factor of unit loading $b_i = 1$ across all maturities τ_i as the well-known result that a one percentage point parallel shift up in the term structure is associated with a drop in bond prices by a percentage amount equal to duration, which is τ_i for the i'th zero-coupon bond. If the yields have different factor sensitivities, i.e., b_i is not constant across i, then (4) introduces the corresponding generalized duration measure $\tau_i b_i$, giving the percentage drop in the price of the i'th zero-coupon bond associated with a unit factor change. In case of multiple factors, $k > 1$, each entry in $\tau_i b_i$ similarly defines a generalized duration measure with respect to the associated factor, and the multivariate hedge is an immunization strategy with respect to each factor.

A.1. Hedging a simple claim

To immunize the future payment τ_* periods hence we consider investing in a portfolio comprised of the m zero-coupon bonds with times to maturity $(\tau_1, \ldots, \tau_m)$ for which the yield factor model (1) is estimated. It is useful to write returns of the instruments as the vectorized version of the return model (4),

$$r_{t+1} = -\mathcal{T}(B \Delta f_{t+1} + \Delta \varepsilon_{t+1}), \tag{5}$$

[4] For non-short horizons, a slope-adjustment (see e.g. Litterman and Scheinkman (1991)) may be added to improve upon the approximation. This adjustment will, however, be known at time t, which implies that the conditional variance of the hedge error in Theorem 1 is unaltered, and the optimal weights remain the same.

where $r_{t+1} = (r_{t+1,1}, \ldots, r_{t+1,m})'$ and $\mathcal{T} = \text{diag}(\tau_1, \ldots, \tau_m)$. At time t, to hedge the return on the target payment over the next period, r^*_{t+1}, we allocate relative proportions $w = (w_1, \ldots, w_m)'$ to the instruments. The return on the hedge portfolio is then

$$r^w_{t+1} = w' r_{t+1} = -w' \mathcal{T} (B \Delta f_{t+1} + \Delta \varepsilon_{t+1}), \tag{6}$$

where $B' \mathcal{T} w$ is the $k \times 1$ vector of generalized durations, controlled by the portfolio manager through the choice of w.

By analogy with the yield analysis of the previous subsection, the required return on the hedge portfolio to immunize the target against changes in factors is $-\tau_* b'_* \Delta \widehat{f}_{t+1}$. Here, the predicted factor change based on returns is $\Delta \widehat{f}_{t+1} = \widehat{F} \Delta y_{t+1} = -\widehat{F} \mathcal{T}^{-1} r_{t+1}$. The required hedge portfolio return is then linear in instrument returns, $\tau_* b'_* \widehat{F} \mathcal{T}^{-1} r_{t+1}$, and the portfolio with weights

$$\widetilde{w} = \tau_* \mathcal{T}^{-1} \widehat{F}' b_*, \tag{7}$$

matches the generalized durations $\tau_* b_*$ of the target. Among all portfolios doing so, $\widetilde{w}$ in (7) minimizes hedging error variance, as shown in Theorem 1 below. Note that estimated loadings $\widehat{B}$ and error variances $\widehat{\Psi}$ are those from the classical factor analysis applied to yields, based on (1), but that the resulting portfolio weights $\widetilde{w}$ are applied to returns, not yields. In (7), each row of $\widehat{F} \mathcal{T}^{-1}$ shows how to construct the factor-mimicking portfolio for the corresponding factor, and the generalized durations $\tau_* b_*$ of the target are used to combine the k factor-mimicking portfolios into the overall hedge portfolio.

A.2. Hedging of general payment streams

Suppose the target claim to be hedged at time t is a payment stream, say, a coupon bond, or an annuity, that promises payments c_h at future dates τ_h periods hence, $h = 1, \ldots, H$. Then the value of the claim is $v_* = \sum_{h=1}^H p_{t,\tau_h} c_h$, with $p_{t,\tau} = \exp(-\tau y_{t,\tau})$ the discount function, if necessary obtained by interpolation between observed yields. To hedge the payment stream, the idea is to hedge each payment according to the portfolio rule (7). The hedge portfolio for the target stream is then the combination of all these portfolios. Specifically, payment c_h is hedged by allocating the amount $p_{t,\tau_h} c_h$ across the m hedging instruments in the proportions indicated by the portfolio rule $\widetilde{w}_h = \tau_h \mathcal{T}^{-1} \widehat{F}' b_h$, with the k-vector of coupon loadings b_h obtained by interpolation as in the case of target loadings above. The overall strategy is to allocate the amount v_* across the instruments according to $\widetilde{w} = \sum_{h=1}^H p_{t,\tau_h} c_h \widetilde{w}_h / v_*$. This is equivalent to applying the rule (7) directly to the target

payment stream, assessing its generalized duration vector as

$$(\tau b)_* = \sum_{h=1}^{H} \frac{p_{t,\tau_h} c_h}{v_*} \tau_h b_h, \tag{8}$$

the weighted average of the individual payments' generalized duration vectors $\tau_h b_h$, each of dimension k, the weights being those of the individual payments' present values relative to the total claim. The result $(\tau b)_*$ is simply inserted in (7) to achieve generalized immunization of the target,

$$\widetilde{w} = \mathcal{T}^{-1} \widehat{F}' (\tau b)_* . \tag{9}$$

The immunization strategy (9) combines the hedging instruments to equate each of the k generalized durations of the portfolio to those of the target claim, $(\tau b)_*$, and, in addition, to minimize hedging error variance (see Theorem 1 below). In particular, generalized durations are $B' \mathcal{T} \widetilde{w} = (\tau b)_*$.

The expression (8) for the generalized duration vector of a payment stream facilitates not only the hedging of such a stream, but also the use of streams as hedging instruments. This includes hedging with coupon-bearing bonds instead of just zero-coupon bonds that may be unavailable. We demonstrate in the Internet Appendix how this can be conducted, but will focus in this paper on the hedging with zero-coupon bonds as instruments.

A.3. The optimal immunization portfolio

Value matching requires the value of the hedge portfolio to equal the value of the target, meaning that the portfolio should be fully invested in the sense of Nelson and Schaefer (1983). This is accomplished if the weights sum to one, $\widetilde{w}' \iota = 1$, where $\iota = (1, \ldots, 1)'$ and requires additional scaling of the generalized duration matching portfolio (9). An obvious choice would be to use the scaling factor $\left(\widetilde{w}' \iota\right)^{-1}$, but this would violate generalized duration matching. The following theorem shows how to adjust portfolio weights (9) to achieve value matching while maintaining generalized duration matching and minimizing hedging error variance.

Theorem 1. *The immunization portfolio w_* that minimizes total hedging error variance among all linear portfolio rules matching the generalized durations and value of the target claim,*

$$\min_{w} \operatorname{var}_t \left[r^*_{t+1} - w' r_{t+1} \right] \quad s.t. \quad B' \mathcal{T} w = (\tau b)_* \quad and \quad w' \iota = 1, \tag{10}$$

is given by the instruments weights

$$w_* = \widetilde{w} + \left(1 - \widetilde{w}'\iota\right)\frac{\Lambda\iota}{\iota'\Lambda\iota}, \tag{11}$$

where $\widetilde{w}$ are the weights (9) that solve the equivalent problem without the value matching constraint $w'\iota = 1$, and where $\Lambda = \mathcal{T}^{-1}\Psi^{-1}\left(\Psi - B\left(B'\Psi^{-1}B\right)^{-1}B'\right)\Psi^{-1}\mathcal{T}^{-1}$.

In this paper subscript t indicates the conditonal moments whenever applicable. The constraints in (10) are value matching, $w'\iota = 1$, and generalized duration matching, $B'\mathcal{T}w = (\tau b)_*$. Together, these would constitute a sufficient criterion in a complete market with only factor risk, cf. Harrison and Kreps (1979). In practice, in the incomplete market case, a perfect hedge is infeasible, so we look for the portfolio which generates a return as close as possible to the target return in the sense that it minimizes the hedging error variance subject to the two constraints. This criterion differs from that introduced by Ingersoll (1983) and frequently used by academics as well as practitioners (see, e.g., Diebold et al. (2006) for an application), where instead the sum of squared weights $w'w$ is minimized subject to the same two constraints. We implement this procedure below. Ingersoll suggests a diversification argument behind which lies the assumption that the idiosyncratic errors in instrument returns are of the same magnitude. Given the yield factor model on which generalized durations are based we know that idiosyncratic return errors must have variance $\Psi\mathcal{T}^2$. Therefore, once generalized durations are matched, minimization of the remaining hedging error amounts to minimization of $w'\Psi\mathcal{T}^2w$ which is done in Theorem 1. We also show in the Internet Appendix that Theorem 1 is in fact a special case of a further alternative approach that relaxes the generalized matching constraint. This allows some factor variance in the hedging error variance, striking a balance between minimizing factor variance and remaining idiosyncratic variance. The hedging performance of the models under consideration is seen in the Internet Appendix to be poorer than in the corresponding cases using Theorem 1, suggesting it is indeed important to remove all factor contribution to hedging error, rather than balancing factor versus idiosyncratic variance. The relative hedging performance among our considered models and procedures remains, nevertheless, leading us to only present results using Theorem 1 in the paper.

B. Portfolio targets and hedging instruments

We use data from the Fed's database of constant maturity zero-coupon yields on U.S. Treasury bills, notes, and bonds. The terms to maturity considered are 3, 6, 12, 24, 36, 60, 84, and 120 months. A weekly frequency data set is constructed by extracting Wednesday observations drawn from the Fed's daily database, rather than using their weekly database, which consists of weekly averages of daily data. Our sample period is the first week of 1983 through the last week of 2014 for a total of 1,670 observations in the time series dimension. Starting in 1983 avoids the Fed money supply targeting experiment of 1979 to 1982 (see Sanders and Unal (1988)). Table 1 shows means and standard deviations of the weekly yield data corresponding to the eight maturities. The means are monotonically increasing in maturity, from 4.12% to 5.98% (continuously compounded annualized yields) whereas the average volatilities or standard deviations exhibit a hump shape with a maximum of 3.09% at two years, and a low of 2.64% at ten years. We consider a one-month hedging period, from month-end to month-end. The weekly yield data are used to estimate model parameters, and the eight associated zero-coupon bonds are used as hedging instruments on the last trading day of each month. As this is not necessarily a Wednesday, the daily files are used again to get the correct zero-coupon bond prices when constructing the hedge portfolio.

Target assets for assessing hedging performance are constructed by drawing information on contractual terms (coupon dates and rates) from the CRSP Monthly Treasury files. We construct a monthly return series for each of two alternative target assets. The first is a 5-year coupon bond, and the second is a portfolio consisting of a long position in the same 5-year coupon bond and short positions in 2-year and 10-year coupon bonds. Using the individual 5-year coupon bond as target is similar to Diebold et al. (2006), whereas the specification of a target asset as a portfolio with short positions in the long and short ends follows Litterman and Scheinkman (1991). On the last trading day of each month, we select among all non-callable, non-convertible, non-flower bonds the issues with maturities closest to two, five, and ten years, subject to a liquidity requirement of at least $10 million in par value publicly outstanding. Portfolio weights[5] $(0,1,0)$ and $(-1,3,-1)$, respectively, are then assigned to construct the two target assets. As our hedging portfolios are always based on an estimation period of a minimum of four years, the hedging period starts four years later than the yield data, and our monthly target data span the period from January 1987 through December 2014 for a total of 335 months in

[5]Value weights, as opposed to numbers of certificates.

the time series.

For illustration, Figure 1 shows characteristics of the 5-year coupon bond that enters both target assets. The upper left panel shows the term to maturity for each selected bond in the time series. Most bonds are issued and mature on the 15th of the month and so in the figure are either 1/2 month above or below the 5-year target. The upper right panel shows the received coupon rates. They have been falling over the sample period, i.e., hedging performance is assessed on basis of the actual market development, rather than some fixed design. The lower left panel shows the resulting durations of the selected 5-year coupon bonds, which increase from below 4.0 to above 4.8 due to the drop in rates. The portfolio target further includes the 2-year and 10-year coupon bonds, and the corresponding coupon rates are shown in Figure 2, along with the resulting target durations.

For a fair comparison of methods, we set the prices of the target assets by using the Fed yields to value the bonds entering them, rather than using the CRSP prices directly. That is, the contractual terms are taken from CRSP, then priced using the eight zero-coupon yields on the last trading day of the month and linear interpolation. This produces a monthly series that should be a fair target for one-month ahead hedging using the corresponding eight zero-coupon bonds. The issue is that raw CRSP prices (bid-ask midpoints plus accrued interest) might reflect other factors not present in the Fed yields such as microstructure noise and, hence, thus do not constitute reasonable grounds for a comparison of methods. The differences between raw CRSP prices and the Fed valuations we use are shown in the lower right panel of Figure 1. Evidently, it would not be of interest to require the hedge portfolio to pick up this discrepancy which is seen to fluctuate in a 1% band.

We consider immunization of movements in the target over periods of one month. Thus, from our monthly target data, 1987:1-2014:12, we form $T = 334$ one-month returns for the 5-year coupon bonds and portfolio targets. The properties of these return series are shown in the first row of Table 3 for the coupon bond target, and first row of Table 4 for the portfolio target. The average one-month return on the 5-years coupon bond is 53bp or 0.53% (1 basis point (bp) equals 0.01%).[6] The standard deviation of the monthly returns is 128bp. For the portfolio target, the average return is 60bp and the standard deviation

[6]This is the unhedged return, and the column is labelled 'Bias' because average hedging errors are reported in the remainder of the table.

147bp.

C. Duration matching as benchmark

As a benchmark method we form a standard duration matching hedge portfolio. Doing so requires only two instruments, one to match duration and one to ensure that hedge portfolio weights sum to one. For the 5-year coupon bonds, target duration is always between three and five years, so we use the corresponding zero-coupon bonds. For the hedge portfolio we use the same instruments, occasionally substituting the 5-year with the 2-year zero-coupon bonds when target portfolio duration falls below three years, cf. Figure 2.

The performance of the duration hedge is shown in the second row of Tables 3 and 4 for the coupon bond and portfolio targets, respectively. The single coupon bond target duration matching achieves a bias (i.e., mean hedging error) of -6.83bp at a standard deviation of 7.57bp, which yields a root mean squared error (RMSE) of 10.19bp. A negative (positive) bias implies that the hedge portfolio generates a lower (higher) return than the target. The good performance of the hedge is due to the single coupon bond being dominated by the final payoff five years in the future. For the portfolio target, duration matching yields an average hedging error of -13.31bp at a standard deviation of 54.91bp, for an RMSE of 56.42bp. Evidently, the more complicated payment stream, loading negatively on the long and short ends, but with roughly the same duration as the coupon bond target, is an order of magnitude more difficult to hedge, and the duration matching approach may be too simplistic.

D. Generalized duration matching

To match generalized rather than basic durations, we start with a classical maximum likelihood factor analysis of the yield factor model (1). Details and estimation results can be found in the Internet Appendix. With estimated B and Ψ in hand, we may proceed to the actual hedging step, using the optimal hedge portfolio from Theorem 1. Again, this portfolio depends on the target to be hedged, and we consider the same coupon bond target and bond portfolio target as before. In each time period, the three-vector of generalized target durations is calculated from (8). This varies through time, primarily since both discount function and coupon rate vary. In addition, there is some variation in target duration, see Figures 1 and 2. Target generalized durations are next premultiplied by the constant $m \times 3$-matrix $\mathcal{T}^{-1}\widehat{F}'$ in (9) and the result used in Theorem 1 to get the

time-varying hedging weights. Hedging performance is documented in the third line of Tables 3 and 4. Similarly to Chambers et al. (1988) and Diebold et al. (2006) and since the portfolio in Theorem 1 minimizes hedging error variance we will in the following primarily compare the different models on the basis-point improvement in RMSE. If deemed large, we refer to this improvement as economic significant. It follows that generalized duration matching performs worse than basic duration matching for the single coupon bond target. Even though we are mainly interested in the RMSE, it is worth mentioning that the average hedging error decreases by matching generalized durations. Moreover, in the rolling estimation both the single and the portfolio coupon hedge portfolio generate a higher mean return than the immunization target, which is an appealing feature seen from the point of view of the portfolio manager. In case of the portfolio target, third line of Table 4, generalized duration matching does improve on basic duration matching by 14% in terms of RMSE, and also the average hedging error (bias) is smaller.

Full period calculations give the investor the benefit of hindsight in that the factor projection matrix $\widehat{F}$ that enters the hedging weights is based on a full period yield factor analysis, including data from time periods following those for which hedging performance is assessed. Similarly, the computation of target generalized durations uses output from the factor analysis. For a more relevant performance analysis corresponding to feasible investment strategies (in line with the duration matching strategy), we next consider rolling yield factor analyses, each based on data for the four-year period immediately prior to the hedging date.[7] Thus, in (9), both the matrix and the vector of target durations are now time-varying. Hedging performance based on the feasible strategy is about the same as for full period factor analysis in the single coupon bond target case, but dominates the latter in terms of both RMSE and average hedging error on the portfolio target. Since rolling estimation involves an out-of-sample hedging element, it is not necessarily given in advance that it dominates the full period estimation case. Thus, the empirical results are consistent with the importance of conditioning decisions on relevant information.

E. Flexible functional form calibration of loadings

Considering the simple level, slope, and curvature structure of the loadings in Figure 3, it is natural to smooth across several maturities by parametrizing loadings. This leads to a more general approach to the problem of hedging a claim with maturity τ_* (different from

[7] See e.g. Buraschi and Corielli (2005) for theoretical justifications of periodic recalibration of model parameters and initial conditions in no-arbitrage models from the perspective of a portfolio manager.

the τ_i's of the hedging instruments) by exploiting the functional form of the dependence of each factor loading on term to maturity. Thus, write $B_j(\tau)$ for the sensitivity to the j'th factor, viewed as a function of term to maturity τ. This function is now taken to be a smooth interpolation across all $i = 1,\ldots,m$ of the b_{ij} terms, for fixed j, given by a flexible parametric functional form adopted for this relation. This formalizes the visual plotting of B_j against τ_i in Figure 3. The i'th row of B now takes the form $b_i' = b(\tau_i)'$, where $b(\tau)' = (B_1(\tau),\ldots,B_k(\tau))$. Hence, using the functional form $b(\cdot)$ for the interpolation in (8) obtains the k-vector of generalized durations $(\tau b)_*$.

The parametrized form of the loading matrix B may be imposed in parameter estimation based on the yield data. The classical factor analysis has $mk - k(k-1)/2$ free parameters in B, e.g., 21 parameters in the three-factor model for eight yields, so there is ample room for exploiting parsimony and, under correct specification, efficiency gains through reduction in the number of parameters via functional form specifications.

Only a single parameter $a > 0$ enters the $B_j(\cdot)$ functions in case of the most popular parametrized functional form for cross-sectional yield curve calibration, the Nelson and Siegel (1987) (henceforth NS) curve shape, given by

$$y_{t,\tau} = f_{t,1} + f_{t,2}\left(\frac{1-e^{-a\tau}}{a\tau}\right) + f_{t,3}\left(\frac{1-e^{-a\tau}}{a\tau} - e^{-a\tau}\right). \tag{12}$$

Here, the loadings are $B_1(\tau) = 1$, i.e., flat, $B_2(\tau) = (1-e^{-a\tau})/a\tau$, downward sloping, and $B_3(\tau) = (1-e^{-a\tau})/a\tau - e^{-a\tau}$, hump-shaped, so the NS curve shape has the desired level, slope, and curvature feature. The parameter a governs the slope of $B_2(\cdot)$, as well as the curvature in the maturity direction in $B_3(\cdot)$.[8] The savings in degrees of freedom in B relative to the classical factor analysis is 20 in the case of three factors and eight yields. With parametrized loadings, though, it is not in general appropriate to take the factor variance as the identity matrix. Thus, along with parameters in B (a in the NS case), we also estimate $\Omega = \mathrm{var}(f_t)$ unrestricted, so there are $k(k+1)/2$ additional parameters, and the final saving relative to classical factor analysis is 14 with three factors and eight yields.

Diebold et al. (2006) confirms the good empirical fit of the NS yield curve, again reestimated monthly by OLS. Here, factors are treated as parameters and a is fixed across

[8] $\partial B_2(\tau)/\partial\tau = [(1+a\tau)e^{-a\tau} - 1]/(a\tau^2) < 0$ for all $\tau > 0$, and $\partial B_3(\tau)/\partial\tau = \left[\left(1+a\tau+a^2\tau^2\right)e^{-a\tau} - 1\right]/(a\tau^2)$ is positive for small τ and negative for large τ.

time at a value 0.0609 for τ measured in months, corresponding to $a = 0.731$ in our case with τ in annual terms. This value was chosen in another study by Diebold and Li (2006), who argued that it makes the hump in the third loading function in (12) occur at 30 months, thus striking an average between the two and three year maturities between which the yield curve hump is commonly observed. In fact, the 0.0609 value generates a maximum at 29.4 months whereas a maximum at 30 months requires $a = 0.717$ (or 0.0598 in monthly terms). Setting a at an exogenously prespecified value circumvents any empirical estimation of B whatsoever. As noted in Section 2, Diebold et al. (2006) do not minimize hedging error variance in their hedging strategy, and so do not need Ψ, either, so their hedge portfolio may be calculated without any preceding yield factor analysis.

Rather than fitting NS curves by cross-sectional OLS each period, we maximize the restricted factor analysis log-likelihood function based on the full panel of yields, imposing that B takes the functional form described above, depending only on a. Thus, the parameters estimated are (μ, a, Ψ, Ω), or 23 in total. Estimation results for models with parsimoniously parametrized loading functions appear in Table 6. The estimated a for the full period is 0.672, and precisely estimated, with a standard error (below the estimate) of 0.007. When inserted in B, this estimate of a generates the three NS loading functions exhibited in Figure 4. The hump in the third loading is at $\tau = 2.7$ years, somewhat larger than the values previously discussed, and consistent with the hump shape around the 12, 24, and 36 month entries in the unrestricted case (see Figure 3). Based on the estimated Ω, the correlation between the level and slope factors is -0.08, the correlation between the level and curvature factors is 0.61, and that between slope and curvature is 0.49.

The corresponding rolling four-year window estimation produces an average a of 0.851. Figure 5 shows the time series evolution of estimated a from the sequential factor analyses of four-year yield panels with pointwise 5% confidence bands. The figure also depicts flat lines indicating the positions of our full period estimate, and the value for a proposed by Diebold et al. (2006). The figure suggests that the parameter a should in fact not be fixed across time periods. The idiosyncratic standard deviations from the restricted factor analysis are very similar to those from the unrestricted factor analysis in the first row. This indicates that the model restricting loadings to be of the NS shape explains about as much of the variation in yields as the unrestricted model, although formally, the restricted model is rejected based on the log-likelihood values (cf. Table 5).

Again, given estimated B (now in terms of a) and Ψ, we turn to hedging performance.

Performance is somewhat worse in RMSE terms than in the unrestricted case for the portfolio target. This applies for both full period and rolling estimation. When hedging the single coupon bond, Table 3, performance is about the same for the full period estimation and better than in the unrestricted case for the rolling period. Hence, utilizing the structure of the loadings allows an out-of-sample improvement in the single coupon target case.

If a is kept fixed at the value $a = 0.731$ consistent with Diebold et al. (2006), but the remaining parameters are still estimated, the results in line seven of each table are obtained. When hedging the single coupon bond, performance is worse than full period and rolling estimation. For the portfolio target, performance is better in terms of RMSE for all models considered so far.

If Ψ is replaced by $\mathcal{T}^{-2}$, which corresponds to the case where the variance-covariance matrix of the idiosyncratic errors in log-prices of the instruments is proportional to the identity matrix, the optimal portfolio coincides with that simply minimizing the sum of squared hedging weights subject to the same two constraints. This portfolio was considered by, among others, Ingersoll (1983). Performance in this case is 8.96bp in terms of RMSE in the rolling estimation for the single coupon bond compared to 10.90bp with estimated Ψ. Similarly, when hedging the portfolio target RMSE drops from 48.70bp to 23.88bp. What causes these somewhat surprising results is that the idiosyncratic variances ψ_i in the yield factor model are of similar order of magnitude across maturities, so that $\tau_i^2\psi_i$ is much smaller for short instruments than for long. Thus, the hedge portfolios become dominated by short instruments and empirically to an excessive extent. More diversification improves performance here, presumably because the factor model only holds to a certain degree of approximation.[9]

All in all, these results suggest that basic immunization through duration matching performs well for the coupon bond target, but that generalized duration matching based on an estimated yield factor model provides an improvement when hedging the more complex portfolio target. In addition, there is an indication of a possible further improvement by parametrizing the factor model in a parsimonious fashion.

[9] See Carcano and Dall'O (2011) for ways to potentially improve hedging performance by accounting for the exposure to (the variance of) modeling errors.

III. Dynamically consistent arbitrage-free hedging

In this section, we exploit restrictions from dynamic term structure theory in an attempt to improve the quality of the hedge. We start out by looking more closely at the approach of restricting the functional form of the loadings or factor sensitivities. The restriction of the loading functions in the three-factor model to NS shape from the previous section is a special case of this approach. The critical issue that arises is that adopting a parametrized functional form for the loadings also amounts to imposing a parametrized functional form for the yield curve as a function of term to maturity τ as is clear from (12). In the equation, the parameters on date t are $f_t = (f_{t,1}, f_{t,2}, f_{t,3})'$, and a. In other parametrizations of B, a similar functional form for the yield curve, $\tau \mapsto y_{t,\tau}$, with parameters f_t and those in B is imposed. Any such functional form severely restricts the set of possible arbitrage-free DTSMs that could be behind interest rate movements in the market and may even rule out that any such model exists. This is the main result of Björk and Christensen (1999). In particular, it follows from Björk and Christensen (1999) and Filipović (1999) that no term structure model would generate yield curves that are of the NS shape (12) in each period, no matter how f_t and a are allowed to move over time.

The central concept in this theory is that of consistency between a parametrization of the shape of the yield curve and a DTSM. Consistency of the two requires that if the initial yield curve $y_{0,\tau}$ is of the given shape, for some parameter vector, and interest rate changes are driven by the dynamic model in question, then each subsequent yield curve $y_{t,\tau}$, for $t > 0$, is also of the given shape, for some (typically other) parameter vector. Thus, for a given parametrization of the shape of the yield curve, as a function of term to maturity, there is a particular set of DTSMs consistent with this curve shape, in the sense that these models would generate curves of the shape each period. This set could be empty, as it is in the case of the NS yield curve shape. This means that a minimum requirement on the chosen parametrization of the loadings it should generate a yield curve shape for which at least one consistent term structure model exists.

In the following, we use these insights to improve hedging performance. Specifically, we first modify the parametrization of the restricted loadings to ensure that there exists a term structure model consistent with the chosen shape. We reestimate the yield factor model with the modified loadings, and assess the resulting hedging performance. This is the route pursued in Section A. Next, having identified a term structure model consistent with the type of curves that empirically provide good cross-sectional yield curve fits, we

consider in Section B an alternative estimation procedure exploiting the dynamics of this model, as opposed to the yield curve shape. In Section C, we then combine the two, exploiting both the shape of the yield curve and the dynamics of the consistent model. Because the consistent model only involves a single stochastic factor, we introduce in Section D a new arbitrage-free DTSM with level, slope, and curvature factors that all three vary stochastically through time. We apply all three approaches (curve shape, dynamics, and the combination) to this model.

A. Dynamically consistent curve shape

The finding that the NS curve shape provides good empirical fits in purely cross-sectional yield curve calibration suggests aiming for an augmentation that retains the desirable empirical properties, in particular the level, slope, and curvature features, while achieving consistency with some arbitrage-free term structure model. Such an augmentation is provided by the modified curve shape given by

$$y_{t,\tau} = f_{t,1} + f_{t,2}\left(\frac{1-e^{-a\tau}}{a\tau}\right) + f_{t,3}\left(\frac{1-e^{-a\tau}}{a\tau} - e^{-a\tau}\right) + f_{t,4}\left(\frac{1-e^{-2a\tau}}{2a\tau}\right). \tag{13}$$

This is simply the NS yield curve shape (12), augmented with an additional slope factor $f_{t,4}$, and with parameter $2a$ in the new sensitivity or loading function $B_4(\tau) = (1-\exp(-2a\tau))/(2a\tau)$, instead of a, as in $B_2(\tau)$. Thus, the augmentation does not lead to a loss of degrees of freedom, as the additional loading function only depends on the same unknown parameter a, but merely achieves consistency with arbitrage-free term structure modeling, as shown by Björk and Christensen (1999).

To relate the cross-sectional yield curve shapes to the underlying DTSMs, we move to continuous time, writing $y(t,\tau)$ for the continuously compounded zero-coupon yield at t, with term to maturity τ, i.e., $y(t,0)$ is the instantaneous short rate at t. If interest rate dynamics are generated by the extended Vasicek or Hull and White (1990) (henceforth HW) model

$$dy(t,0) = a(\theta(t) - y(t,0))dt + \sigma dW_t, \tag{14}$$

and if the yield curve at any arbitrary point in time is of the augmented NS (henceforth ANS) shape (13), for some f_t vector, then all subsequent yield curves are also of the ANS shape, with different f_t. Thus, the four factors in f_t suffice as state vector for the term structure dynamics. In this sense, the augmented NS curve shape is consistent with

dynamic term structure theory, namely, with a particular dynamic model (the HW model). Note that $\theta(\cdot)$ in (14) is the time-varying target for mean-reversion that constitutes Hull and White's extension of the Vasicek model, σ is the short rate volatility, and a is the rate of mean-reversion, which coincides with the parameter a in the loading functions of the ANS curve shape (13).

To enhance intuition we recast the HW model in the Heath, Jarrow, and Morton (1992) (henceforth HJM) framework as

$$dy(t,\tau) = \alpha(t,\tau)dt + \sigma\left(\frac{1-e^{-a\tau}}{a\tau}\right)dW_t, \tag{15}$$

where the drift $\alpha(t,\tau)$ under no arbitrage is determined as a function of the volatility function and a market price of risk. The point is that if at any arbitrary point in time t_0 the yield curve $y(t_0,\cdot)$ takes the ANS form, and the dynamics are given by (15), with the HJM no-arbitrage condition imposed on the drift, then the dynamics for $t > t_0$ may be written in the form

$$dy(t,\tau) = \sum_{j=1}^{4} B_j(\tau)df_{t,j}, \tag{16}$$

so as to emphasize the factor structure, with $B_j(\tau)$ the loading function on the j'th ANS factor and $df_{t,j}$ the dynamics of this. An explicit expression for the factor dynamics is given in the following subsection. Since a consistent dynamic model exists for the ANS curve shape, but not for ordinary NS curves, we now use the $B_j(\cdot)$ functions from the ANS curve (that is, we add the required $B_4(\cdot)$ function) when imposing structure on the loading matrix B in the factor analysis of yields.

The yield factor analysis now uses four factors, i.e., B is $m \times 4$. The results in the third row of Table 6 show that the full period estimate of a when B is restricted according to ANS is similar to the estimate using NS parametrization. In the rolling estimation, the average estimate is lower in the ANS parametrization than in NS, but produces an average a close to the ANS full period estimate. Plots of the loading functions in unrestricted four-factor and ANS analyses are exhibited in the left and right panels of Figure 6. The fourth unrestricted loading has two small humps, but explains very little of the variation in yields (communality or sum of squared loadings is very close to zero). The fourth restricted (augmented NS) loading corresponds to a second slope factor. From Table 5, idiosyncratic standard deviations are less in both unrestricted and restricted four-factor models than in the corresponding three-factor models, except for maturity seven years, and slightly

larger for the restricted than for the unrestricted case. From the log likelihood values, the difference between the three-factor and four-factor cases is significant, which suggests that the addition of a factor (as in the arbitrage-free consistency augmentation of the NS model) is in line with the information in the data.

The resulting hedging performance in the unrestricted and ANS models is documented in lines eight through eleven in Tables 3 and 4. For both the single coupon bond target and the portfolio target, performance is now better than in the cases previously considered. In both tables, the structure imposed on the loadings improves hedging performance, in spite of the dramatic drop in number of parameters. Rolling dominates full period estimation in restricted loadings for both targets and in unrestricted loadings for the single coupon bond target. Indeed, the drop to 26.84bp in RMSE for hedging the portfolio target using rolling estimation of the augmented NS loading structure represents a considerable improvement in performance relative to other approaches so far. The improvements in performance using ANS loadings are achieved in spite of the fact that the restricted model is formally rejected based on the log likelihood criterion, showing that statistical and financial criteria do not necessarily coincide. The results support the notions that parametrizing loadings in accordance with empirically well established yield curve shapes is beneficial for hedging purposes, and that the augmentation of the NS curve shape to achieve consistency with arbitrage-free dynamics is warranted.

B. Consistent yield dynamics

In this section, we develop a new approach to exploiting consistent arbitrage-free dynamics for hedging purposes without imposing a particular curve shape in the estimation step. Consider first a general term structure model in the HJM framework, with yield curve dynamics given by the infinite dimensional stochastic differential equation (SDE)

$$dy(t,\tau) = \alpha(t,\tau)dt + \sigma(t,\tau)'dW_t,$$

with drift $\alpha(t,\tau)$ and yield volatility function $\sigma(t,\tau)$. Writing d for the dimension of the driving Wiener process W_t, the dimension of $\sigma(t,\tau)$ is $d \times 1$. The no-arbitrage drift condition of HJM is

$$\alpha(t,\tau) = \frac{1}{\tau}\left[y(t,\tau) - y(t,0)\right] + \frac{\partial y}{\partial \tau}(t,\tau) + \frac{\tau}{2}\sigma(t,\tau)'\sigma(t,\tau) + \sigma(t,\tau)'\lambda_t, \tag{17}$$

where λ_t is the d-vector of market prices of risk. Here, we use the parametrization of Brace and Musiela (1994) where term to maturity τ rather than maturity date enters as a separate argument in $y(t,\tau)$. As HJM considered the alternative parametrization with maturity date $t+\tau$ instead of τ as a separate argument, and represented the term structure via instantaneous forward rates instead of the yield curve, we present a brief derivation of (17) in the Appendix. The slope term $\partial y(t,\tau)/\partial\tau$ is not present in the drift condition in the HJM parametrization with constant maturity date, and it represents a locally deterministic ageing effect also discussed by Litterman and Scheinkman (1991). The leading spread term in (17) is also not present in the standard HJM representation and appears because we consider yields rather than forward rates.

In this framework, we now focus on the consistency between the HW dynamics and the augmented NS curve shape. Hull and White introduced their extension of the Vasicek model through the forward rate volatility function $\sigma_r(t,\tau) = \sigma\exp(-a\tau)$, exponentially decaying for longer maturities. An integration from 0 to τ and division by τ produces the yield volatility $\sigma(t,\tau) = \sigma(1-\exp(-a\tau))/(a\tau)$ used in (15). This yield volatility is recognized as the short rate volatility σ from (14) times the loading on the second factor in the NS yield curve, $\sigma(t,\tau) = \sigma B_2(\tau)$. In this case, with $d=1$, the arbitrage-free drift (17) is

$$\alpha(t,\tau) = \frac{1}{\tau}\big[y(t,\tau)-y(t,0)\big] + \frac{\partial y}{\partial\tau}(t,\tau) + \sigma^2\frac{\tau}{2}B_2(\tau)^2 + \sigma B_2(\tau)\lambda_t. \tag{18}$$

Now rewrite $B_2(\tau)^2$ as $2(B_2(\tau)-B_4(\tau))/(a\tau)$, again with $B_4(\tau)$ from the ANS curve, collect terms, and obtain the SDE

$$\begin{aligned} dy(t,\tau) &= \left\{\frac{1}{\tau}\big[y(t,\tau)-y(t,0)\big] + \frac{\partial y}{\partial\tau}(t,\tau)\right\}dt \\ &+ B_2(\tau)\left\{\left[\frac{\sigma^2}{a}+\sigma\lambda_t\right]dt + \sigma dW_t\right\} + B_4(\tau)\left\{-\frac{\sigma^2}{a}dt\right\}. \end{aligned} \tag{19}$$

Thus, a necessary condition for a model of yield curves $y(t,\cdot)$ to be consistent with HW dynamics is that it include factors with loading functions $B_2(\cdot)$ and $B_4(\cdot)$, e.g., the original NS curve (12) without B_4 would not suffice, although only the factor associated with B_2 is stochastic.

To derive the yield curve shape from the stochastic differentials (19), an initial condition is needed. For now, it is useful to take this as an otherwise unrestricted initial yield curve $y(t,\cdot)$ given at time t. Assuming from here on a constant market price of risk λ, the

resulting solution for the level of the yield curve is stated in the following proposition.

Proposition 1. *For initial yield curve $\tau \mapsto y(t_0,\tau)$ and interest rates that follow the Hull-White model (15), subsequent yield curves at times $t_1, t_2, \ldots$, with $t_{n+1} - t_n = \Delta_n$, are given by*

$$y(t_{n+1},\tau) = \frac{1}{\tau}\left[(\tau+\Delta_n)\,y(t_n,\tau+\Delta_n) - \Delta_n y(t_n,\Delta_n)\right] + B_2(\tau)\widetilde{f}_{n+1,2} + B_4(\tau)\widetilde{f}_{n+1,4} \tag{20}$$

with factors

$$\begin{aligned}\widetilde{f}_{n+1,2} &= \left(\sigma^2/a + \sigma\lambda\right)\Delta_n B_2(\Delta_n) + \sigma\sqrt{\Delta_n B_4(\Delta_n)}\,z_{n+1},\\ \widetilde{f}_{n+1,4} &= -\left(\sigma^2/a\right)\Delta_n B_4(\Delta_n),\end{aligned}$$

and $z_1, z_2, \ldots$ a sequence of independent $N(0,1)$ variables.

The first term in the solution is deterministic (not only locally) and is recognized as the forward rate as of t_n on a τ-period loan to be made at t_{n+1}. This would give $y(t_n,\tau)$ if the expectations hypothesis predicted future spot rates from initial forward rates without error. The proposition shows the additional terms that in fact enter future yield curves. They contribute with a shape across the maturity dimension τ that is spanned by the second and fourth ANS loading functions, and the dynamics assign time-varying coefficients $\widetilde{f}$ to these.

With the yield curve from the proposition in hand, we are now in a position to design the empirical model that exploits the HW dynamics. As in (1), we consider panel data $\left(y_{t_0}, y_{t_1}, \ldots, y_{t_N}\right)$ on yields, with $y_{t_n} = \left(y_{t_n,\tau_1}, \ldots, y_{t_n,\tau_m}\right)'$. From (20), collecting yields on the left hand side, separating the stochastic term on the right hand side, and allowing for measurement error $\widetilde{\varepsilon}_{n+1,\tau_i}$ in the i'th yield at time t_{n+1} produces the factor model

$$\begin{aligned}\widetilde{y}_{t_{n+1},\tau_i} &= B_2(\tau_i)\widetilde{f}_{n+1,2} + B_4(\tau_i)\widetilde{f}_{n+1,4} + \widetilde{\varepsilon}_{n+1,\tau_i}\\ &= \mu_i(\eta) + C_i(\eta)\,z_{n+1} + \widetilde{\varepsilon}_{n+1,\tau_i},\end{aligned} \tag{21}$$

$i = 1, \ldots, m$, where $\eta = (a,\sigma,\lambda)$, $\{z_n\}_{n=1}^N$ is an i.i.d. $N(0,1)$ sequence based on increments to

W_t from (19), and with the definitions

$$\begin{aligned}
\widetilde{y}_{t_{n+1},\tau_i} &= y_{t_{n+1},\tau_i} - y_{t_n,\tau_i+\Delta_n} - \frac{\Delta_n}{\tau_i}\left(y_{t_n,\tau_i+\Delta_n} - y_{t_n,\Delta_n}\right), \\
\mu_i(\eta) &= \left(\sigma^2/a + \sigma\lambda\right)\Delta_n B_2(\Delta_n;a)B_2(\tau_i;a) - \left(\sigma^2/a\right)\Delta_n B_4(\Delta_n;a)B_4(\tau_i;a), \qquad (22) \\
C_i(\eta) &= \sigma B_2(\tau_i;a)\sqrt{\Delta_n B_4(\Delta_n;a)},
\end{aligned}$$

where the dependence of the loading functions B_j on the parameter a from η has been made explicit. Thus, in contrast to the yield factor model (1) for levels, our dynamically consistent arbitrage-free factor model is for a certain adjusted yield change $\widetilde{y}_{t_{n+1},\tau_i}$. Christensen and van der Wel (forthcoming) show that in the ideal case with complete data for all maturities we can think of $-\tau_i \widetilde{y}_{t_n,\tau_i}$ as the one-period excess return to the discount bond with maturity τ_i at time t_n. In order to form the vectors $\widetilde{y}_{t_n} = \left(\widetilde{y}_{t_n,\tau_1},\ldots,\widetilde{y}_{t_n,\tau_m}\right)'$, we employ the shortest maturity τ_1 corresponding to the 3-months bill yield as proxy for y_{t_n,Δ_n} (see Chapman, Long, and Pearson (1999)). Similarly, we use the closest available maturity corresponding to the 10-year note yield as proxy for $y_{t_n,10+\Delta_n}$. The remaining vectors are obtained by interpolating between the observed yield so that the panel data set on adjusted yield changes is $\left(\widetilde{y}_{t_1},\ldots,\widetilde{y}_{t_N}\right)$. The factor analysis is now again restricted, in the sense that μ_i and C_i are parametrized functions. In vector form, the consistent factor analysis model is

$$\widetilde{y}_{t_n} = \mu(\eta) + C(\eta) z_n + \widetilde{\varepsilon}_n, \qquad (23)$$

where $\mu(\eta)$ and $C(\eta)$ are $m \times 1$ vectors with i'th elements given by $\mu_i(\eta)$ and $C_i(\eta)$, respectively. Thus, $C(\eta)$ gives the loadings on the common covariance-generating factor z_n, and $\mu(\eta)$ is the mean of $\widetilde{y}_{t_n}$. It is assumed that the error terms $\widetilde{\varepsilon}_n$ are uncorrelated with z_n, with diagonal variance matrix $\mathrm{var}(\widetilde{\varepsilon}_n) = \widetilde{\Psi}$. Upon estimation, the estimate of a may be used to form the loading functions $B_j(\tau)$, $j = 1,\ldots,4$, from the ANS curve shape, and we implement the hedge described in Theorem 1.[10] The potential advantage of this hedge portfolio is that the restrictions from the consistent arbitrage-free term structure dynamics are imposed in estimating the parameters that enter the generalized durations, both of the target claim and of the hedging instruments.

The results from estimating the adjusted yield change model (23) appear in lines seven and eight of Table 6. The model accommodates unevenly spaced observations, but as our yield data consist of weekly observations we now set $\Delta_n = \Delta = 1/52$. In contrast

[10] We assume proportionality of the diagonal idiosyncratic error variance matrices Ψ and $\tilde{\Psi}$ from the yield factor model and the consistent adjusted yield change factor model, respectively.

with previous results, estimated a is now very close to zero. From (19), the stochastic factor in the HW model is the slope factor, but for a near zero, the slope loading $B_2(\tau)$ becomes flat, so the stochastic factor is a level factor after all. This is consistent with the previous findings that the level factor explains most of the variation in the data. In the present estimation of the adjusted yield change model, the volatility parameter σ and the market price of risk λ are estimated along with a. All three parameters are similar across full period and rolling estimation. The market price of risk is negative, corresponding to the negative relation between yields and bond prices. It is significant in the full period estimation, with a t-statistic of -3.3, although it has the largest standard errors of the three parameters in η. From lines nine and ten of Table 5, idiosyncratic error variances are similar in the HW model and an unrestricted classical one-factor analysis of the adjusted yield changes $\widetilde{y}$. Table 7 shows the estimated loadings in this unrestricted analysis, revealing an initial slope, then a flat structure for maturities two years and higher. Although the unrestricted model produces a higher full-period likelihood value than the restricted in Table 5, we may still investigate the usefulness of the restricted model for hedging purposes, in particular paying attention also to the important out-of-sample (rolling estimation) performance criterion.

The notion of arbitrage-free consistency is based on the HJM drift restriction (17). Table 8 shows the results of estimation with and without this restriction imposed. The first column corresponds to the restricted model already considered, i.e., the reported μ vector is determined by the remaining parameters $\eta = (a, \sigma, \lambda)$ as in (22). The second column leaves μ free. This introduces eight new mean parameters, but leaves the market price of risk λ unidentified, so the difference in degrees of freedom is seven. Both a and σ (labelled σ_1 in the table) are similar in the restricted and unrestricted estimations, but the μ vectors are very different. In particular, the restricted μ is nearly flat in the maturity dimension, because estimated a is close to zero. The upshot is that the arbitrage condition is strongly rejected (the LR test takes the value 56.7 compared to a critical value of 14.1 at the 5% level in the asymptotic χ^2 distribution on seven degrees of freedom). Again, the restricted model could nevertheless prove useful for hedging purposes.

Turning to hedging performance, the estimated a from the restricted model is inserted in the ANS loadings and these are used along with the idiosyncratic variances to form the hedge. Lines twelve and thirteen of Tables 3 and 4 show the results. Hedging performance is the best so far for both the single coupon bond and the portfolio target. The improvement relative to the previous hedging results for ANS loadings is considerable. Since only the

parameter estimates separate the two cases, the results document the value of imposing the consistency restrictions from arbitrage-free term structure dynamics in the estimation stage of the hedging process.

C. Consistent yield dynamics and curve shape

We now present a new framework imposing both the HJM no-arbitrage drift restriction on the yield dynamics, as in the previous section, and the consistent yield curve shape, as in Section A on ANS. The combination of both types of restrictions is accommodated in a state space model implemented using the Kalman filter.

The hedging approach of the previous subsection is based on the yield curve model (20) from Proposition 1, where an arbitrary initial yield curve $y(t_0,\cdot)$ is taken as given. As a result, subsequent yield curves $y(t_{n+1},\cdot)$ involve the second and fourth loading functions from the ANS curve shape, but also a term depending on the previous yield curve $y(t_n,\cdot)$, so it is not necessarily of ANS form. The reason is that the initial curve $y(t_0,\cdot)$ is unrestricted. This specification carries with it the potential advantage that the current observed yield curve may be used as initial curve, which is the idea in the proposed hedging strategy in Section B. The ability to incorporate as much of the information in current market data as possible is often assumed to be desirable. On the other hand, as already discussed, the ANS curve shape has other desirable empirical as well as theoretical properties, i.e., it exhibits level, slope, and curvature features, and is consistent with the HW model in the sense that the yield curves generated by the latter may take on ANS shape in every time period. This was the motivation for the focus on the HW model from the outset. On these grounds, it seems an appealing alternative to consider the initial curve $y(t_0,\cdot)$ to take the ANS form. It is exactly in this case that every subsequent yield curve $y(t_n,\cdot)$ is also in the ANS class such that the combined model is fully consistent. We now develop the alternative hedging approach following these ideas.

When the yield curve at time t_n is on the ANS form $y(t_n,\tau)=\sum_{j=1}^{4}B_j(\tau)f_{n,j}$, with loadings B_j defined as in (13), then the first term in (20), giving the dependence of the yield curve at time t_{n+1} on the previous curve at time t_n, takes the form

$$\begin{aligned}\frac{1}{\tau}\left[(\tau+\Delta_n)y(t_n,\tau+\Delta_n)-\Delta_n y(t_n,\Delta_n)\right] &= f_{n,1}B_1(\tau)+e^{-a\Delta_n}\left(f_{n,2}+a\Delta_n f_{n,3}\right)B_2(\tau)\\ &+ e^{-a\Delta_n}f_{n,3}B_3(\tau)+e^{-2a\Delta_n}f_{n,4}B_4(\tau)\\ &= B(\tau)H(\Delta_n)f_n, \qquad (24)\end{aligned}$$

where $B(\tau) = (B_1(\tau), \ldots, B_4(\tau))$. The first equality in (24) follows from the general theorems below with the 4×4 transition matrix H given by

$$H(u) = \begin{pmatrix} e^{-bu} & 0 & 0 & 0 \\ 0 & e^{-au} & aue^{-au} & 0 \\ 0 & 0 & e^{-au} & 0 \\ 0 & 0 & 0 & e^{-2au} \end{pmatrix}. \tag{25}$$

Thus, in the present case, this term also is recognized to be of ANS form which means that it is a function of term to maturity τ, it is spanned by the loadings $B_j(\tau)$. This makes the entire yield curve $y(t_{n+1}, \tau)$ in (20) to take the ANS form when the yield curve at t_n does. Consequently, starting from an initial yield curve at t_0 on ANS form with dynamics following the HW model all subsequent yield curves at times $t_1, t_2, \ldots$ will be on ANS form. What happens here is that the ANS curve shape includes the B_2 and B_4 terms that by Proposition 1 are necessary for consistency with the HW model, and that the full curve shape is maintained when shifting maturity as in the left hand side of (24). Other curve shapes consistent with HW could be obtained by supplementing the component involving B_2 and B_4 in different ways, subject to this invariance condition. The reported empirical relevance of the original NS curve shape leads us to focus on the minimal augmentation of this, which both includes B_2 and B_4 and satisfies the invariance condition. By necessity, a factor with loading B_4 must be added, and since the resulting ANS curve satisfies the invariance condition, this constitutes the required minimal augmentation.

The following proposition gives the resulting ANS form of the HW extended Vasicek model, including the specific factor dynamics.

Proposition 2. *For an initial yield curve of ANS form, $y(t_0, \tau) = B(\tau) f_0$, and interest rates that follow the HW model, (15), subsequent yield curves at times $t_1, t_2, \ldots$, with $t_{n+1} - t_n = \Delta_n$, are on augmented Nelson-Siegel form,*

$$y(t_{n+1}, \tau) = B(\tau) f_{n+1},$$

with factors

$$f_{n+1} = \overline{f} + H(\Delta_n)\left[f_n - \overline{f}\right] + v_{n+1}.$$

Here, H is the transition matrix given in (25), $v_1, v_2, \ldots$ is a sequence of independent $N(0, \Omega(\Delta_n))$ variables with the (2,2) entry, $\sigma^2(1 - e^{-2a\Delta_n})/(2a)$, the only non-zero element in

$\Omega(\Delta_n)$, *and* $\overline{f}$ *the long-run factor levels are given by*

$$\overline{f} = \left[\begin{array}{cccc} f_{0,1}, & \frac{\sigma}{a}\left(\frac{\sigma}{a}+\lambda\right), & 0, & -\frac{\sigma^2}{2a^2} \end{array} \right]'. \tag{26}$$

From the proposition, the first factor is constant, at the level measured from the initial curve at time t_0, and the third factor is exponentially decaying. The fourth factor reverts exponentially towards level $-\sigma^2/(2a^2)$, and the second factor, which is the only stochastic factor, mean reverts to $\frac{\sigma}{a}\left(\frac{\sigma}{a}+\lambda\right)$. In f_{n+1} the second and fourth entry of the 4×1 vector $[I-H(\Delta_n)]\overline{f}+v_{n+1}=\widetilde{f}_{n+1}$ may be recognized as $\widetilde{f}_{n+1,2}$ and $\widetilde{f}_{n+1,4}$ from Proposition 1, respectively, while the first and third entries of $\widetilde{f}_{n+1}$ are zero. The remaining term in f_{n+1} is picked up from the dependence (24) on the yield curve at t_n, i.e., $H(\Delta_n)f_n$. By inserting $\tau=0$, it is seen that the short rate is given by $y(t_n,0)=f_{n,1}+f_{n,2}+f_{n,4}$, or, alternatively, the stochastic factor is given in terms of the short rate of interest as $f_{n,2}=y(t_n,0)-f_{n,1}-f_{n,4}$.

Compared to the original yield factor model (1), of the form $y_t=Bf_t+\varepsilon_t$, there is not only more structure on the shape of the loading functions in the arbitrage-free consistent model in Proposition 2, but also on the factor dynamics. The generalization of the classical static factor analysis structure that allows for (first-order Markov) serial dependence in factors is simply the Kalman filter combined with maximization of the appropriate likelihood function based on the filtered innovations in data. Of course, the model (23) for adjusted yield changes in the previous subsection also has factor structure with consistency restrictions on loadings, but the factors z_n are serially independent, hence filtering is unnecessary.

To make the state space form of the model from Proposition 2 explicit, the state vector is the vector of factors $f_n=(f_{n,1},\ldots,f_{n.4})'$, so the state transition equation is given by

$$f_n=\Phi_0+\Phi_1 f_{n-1}+v_n,$$

with transition matrix $\Phi_1=H(\Delta_{n-1})$ and intercepts $\Phi_0=(I-\Phi_1)\overline{f}$. Here, Φ_0 depends on all three model parameters $\eta=(a,\sigma,\lambda)$, and Φ_1 only on a. Thus, the no-arbitrage consistency conditions place at least 17 nonlinear restrictions on the 20 coefficients in the transition equation. In fact, η is identified in (Φ_0,Φ_1), i.e., a is clearly identified from $\Phi_1=H(\Delta_{n-1})$ from (25), and then σ may be solved for from $\Phi_{0,4}$ using $\overline{f}$ from (26), and in turn λ from $\Phi_{0,2}$. Hence, the number of restrictions is exactly 17. As Ω would have 10 free parameters in the unrestricted Kalman filter, but here is a function of the same model

parameters η which are already identified in (Φ_0, Φ_1), this yields another 10 restrictions, for a total of 27 in the transition equation alone.

The Kalman filter measurement equation is

$$y_{t_n} = Bf_n + \varepsilon_n,$$

where the entries in the $m \times 4$ loading matrix B are given in terms of the augmented NS loading functions as $b_{ij} = B_j(\tau_i; a)$, and the measurement errors ε_n are assumed i.i.d. $N(0, \Psi)$, with Ψ diagonal. The measurement equation in the unrestricted filter with $k = 4$ factors has $4m - 10$ identified parameters in B (the general count with k factors is $mk - k(k+1)/2$, when Ω is unrestricted, since in this case the model is invariant to scaling and rotation of factors), but in the restricted model, all of these are nonlinear functions of η, which is identified in the transition equation. Comparing to the Kalman filter with Ψ diagonal (this is not an arbitrage condition), there are $4m - 10$ consistency constraints in the measurement equation.

We base the Kalman filter recursions on the Koopman, Shephard, and Doornik (1999) low storage algorithm, inserting the updating step in the prediction step to save on calculations, and because we modify the algorithm to the square-root case, a brief description is provided in the Internet Appendix. The modified recursions generate a sequence of yield vector innovations $\zeta_n = y_{t_n} - \mathbb{E}(y_{t_n}|Y_{n-1})$ for $Y_{n-1} = \left(y_{t_1}, \ldots, y_{t_{n-1}}\right)$, with associated prediction error variances $\Gamma_n = \text{var}(\zeta_n|Y_{n-1})$. The model parameters (η, Ψ) are estimated by maximizing the log-likelihood based on ζ_n i.i.d. $N(0, \Gamma_n)$. Compared to the estimates of (η, Ψ) from the previous subsection, the new estimates impose all available no-arbitrage consistency restrictions, a total of $4m + 17$ theory restrictions, on the factor dynamics.

Estimation results appear under the label 'HW and ANS' in rows 9 and 10 of Table 6. The estimated a value is between those from cross-sectional curve fitting and those exploiting dynamics. There is a great deal of variation in a in the rolling estimation, consistent with the notion that the required slope varies through time and sometimes makes the single stochastic factor more akin to a level factor, which is what apparently happens when focussing only on dynamics (rows 7 and 8 of the table). The parameters σ and λ are similar to those estimated purely from the dynamics, which makes sense, since they do not directly impact the cross-section. From Table 5 row 6, the idiosyncratic variances Ψ are now higher than in other models, except at the three-year maturity.

This is a result of imposing more structure, i.e., both in the cross-sectional and time series dimension simultaneously, and thereby getting a poorer fit, as also seen from the maximised likelihood value.

The restricted estimates of a and Ψ are now used in forming the hedge portfolio from Theorem 1. From rows fourteen and fifteen in Tables 3 and 4, performance is poorer than that exploiting only dynamics (previous two lines in tables), but about or better than when trying to exploit the cross-section in the unrestricted three- and four-factor models and the NS and ANS cases. One peculiarity is that the performance deteriorates in the rolling case for the portfolio target relative to the full period results.

D. A dynamically consistent stochastic level, slope, and curvature model

There are two different ways of counting 'factors' in a given model. While there are four factors, f_t, in the curve shape (13), there is only one source of randomness (one driving Wiener process) in the consistent dynamic HW model (14). When combining both the curve shape and the dynamics (the cross-sectional and time series dimensions), the resulting model described in Proposition 2 may therefore either be labelled a four-factor model or a one-factor model. The consistency issue relates to the fact that interest rate dynamics place restrictions on the manner in which the term structure can change shape over time, i.e., on the four 'factors,' and in the present model restricts three of them to be deterministic. The count of sources of randomness is therefore probably most meaningful once consistency is imposed, so that the combined model is described as a one-factor model in this case.

Since from the previous subsection hedging performance deteriorates when trying to exploit the consistency conditions in the HW model by imposing the ANS curve shape in each period, it appears that the resulting one-factor model may be too simplistic in practice. Here, we present the three steps from the previous subsections (consistent yield curve shapes, interest rate dynamics, and combination of both) for a new stochastic three-factor model. The model retains the level-slope-curvature interpretation of yield curve shapes, but is driven by three Wiener processes instead of just one.

As shown by Björk and Christensen (1999), there is an intimate relation between yield curve shapes and the volatility function of the dynamic model consistent with these shapes. Accordingly, we may directly specify the volatility vector for a three-factor (three Wiener processes) DTSM to exhibit level, slope, and curvature features in the respective

entries. We specify a forward volatility structure given by

$$\sigma_r(\tau)' = \left(\sigma_1, \sigma_2 e^{-a\tau}, \sigma_3 a\tau e^{-a\tau}\right). \tag{27}$$

The first stochastic factor affects forward rates of all maturities equally. Restricting $\sigma_r(\tau)$ to only consist of its first entry generates the Ho-Lee model by Ho and Lee (1986) (consistent with affine forward rate curves), and the second entry alone generates the HW model (consistent with ANS curves). Thus, the two first entries taken together serve to generate the two-factor model considered by Heath et al. (1992). With the first two entries capturing level and slope, we propose the third, $\sigma_3 a\tau e^{-a\tau}$, as a natural candidate for curvature, or hump. To avoid nonstationarity of the dynamic model, we note that the rate of exponential decline of the volatility function in the maturity direction corresponds to the rate of mean reversion of the short rate process, see (15)-(16), so we work with the slightly generalized forward volatility structure

$$\sigma_r(\tau)' = \left(\sigma_1 e^{-b\tau}, \sigma_2 e^{-a\tau}, \sigma_3 a\tau e^{-a\tau}\right), \tag{28}$$

with $b > 0$ a small, positive number to guarantee stationarity. For $b \downarrow 0$, the structure (27) is approached.

In line with the rest of the paper we switch to modelling in terms of yields, and thus the stochastic level, slope, and curvature (henceforth SLSC) dynamic model is described by the yield SDE

$$dy(t,\tau) = \alpha(t,\tau)dt + B_{1:3}(\tau)\ \mathrm{diag}(\sigma_1,\sigma_2,\sigma_3)dW_t, \tag{29}$$

and a constant market price of risk vector λ. In line with the recommendation by CDR, we have chosen an independent factor specification,[11] but an immediate generalization is obtained by replacing $\mathrm{diag}(\cdot)$ by a general variance-covariance matrix. The yield volatilities corresponding to (27) are obtained by taking averages over maturities, i.e. $B_{1:3}(\tau)$ is the 1×3 vector

$$B_{1:3}(\tau) = \left(\frac{1-e^{-b\tau}}{b\tau}, \quad \frac{1-e^{-a\tau}}{a\tau}, \quad \frac{1-e^{-a\tau}}{a\tau} - e^{-a\tau} \right). \tag{30}$$

Again, for $b \downarrow 0$ the first stochastic factor affects yields of all maturities equally, $B_1(\tau) = 1$, and the functions $B_{1:3}(\tau)$ are the same as those that span the NS yield curve shape in (12).

[11] CDR document superior out-of-sample performance of the independent-factor model relative to the correlated-factor model.

Since $B_{1:3}(\tau)$ captures the directions in which the stochastic terms affect the yield curve, the interest rate model will have both level, slope and curvature changing stochastically. As we already discussed, no arbitrage-free interest rate model is consistent with the NS yield curve shape. The next theorem derives a minimal set of additional functions of τ to be added to the NS curve shape in order to span curves consistent with the suggested SLSC model.

Theorem 2. *If interest rate dynamics follow the stochastic level, slope, and curvature model (29), then consistent yield curves must in addition to $B_{1:3}(\tau)$ in (30) necessarily include the functions*

$$\left(\frac{1-e^{-2b\tau}}{\tau}, \quad \frac{1-e^{-2a\tau}}{\tau}, \quad e^{-2a\tau}, \quad \tau e^{-2a\tau} \right). \tag{31}$$

Conversely, the yield curve family spanned by $B_{1:3}(\tau)$ and (31) is consistent with the stochastic level, slope, and curvature interest rate model.

Thus, any representation of yield curves consistent with the SLSC model must include the seven functions given in (30) and (31) or some linear transformation thereof.[12] In the remainder of the paper we choose for the sake of convenience to work with the following particular rotation of the additional yield curve functions,

$$B_{4:7}(\tau) = \left(\frac{1-e^{-2a\tau}}{2a\tau}, \quad \frac{1-e^{-2b\tau}}{2b\tau}, \quad \frac{1-e^{-2a\tau}}{2a\tau} - e^{-2a\tau}, \quad a\tau e^{-2a\tau} \right), \tag{32}$$

which clearly span the same set of curves as (31) do. The loading function B_4 is the same loading function added to the NS curve to make it consistent with HW interest rate dynamics, i.e., the ANS curve, and it captures that stochastic changes in the B_2 direction induce drift in this direction by the HJM condition. The loading function B_5 is equivalently the drift direction induced by the first stochastic factor. The drift direction induced by the stochastic curvature factor is $B_6 - B_7/2$, but a yield curve including this function and $B_{1:5}(\tau)$ implies that the two leading (spread and slope) drift terms in (17) are not spanned. In particular, the drift has B_6 and B_7 in proportions different from 1 to -1/2, and thus we need to include these two functions separately. The result that the yield curve family defined by $B(\tau) = B_{1:7}(\tau)$ is minimal among all families not restricting coefficients on the necessary τ-functions to specific constants is stated in the following corollary.

[12] For $b = 0$ the first function should be τ. The proof of this follows the same steps as that for $b > 0$ given in the Appendix.

Corollary 1. *The yield curve family spanned by $B(\tau)$ is the minimal consistent family on the form $g(\tau)\phi$ for general coefficients ϕ.*

Henceforth, we refer to the yield curve family determined by $B(\tau) = B_{1:7}(\tau)$ as the SLSC curve shape family, because it is naturally associated with the SLSC dynamic model in the sense of the corollary. The (smaller) minimal consistent family in which the $B_{4:7}$ functions enter the yield curve with fixed coefficients (as opposed to general coefficients, as in the corollary) is considered by Christensen et al. (2011) (CDR), and the relation to our more general SLSC curve shape is discussed further in section D.1.

We now derive how coefficients, or factors, on the loading functions $B(\tau)$ will change over time for the SLSC model starting from a general yield curve and an initial SLSC curve spanned by $B(\tau)$, respectively. First, in the case where the initial yield curve at time t_0, $y(t_0,\tau)$, has general shape and yield dynamics follow the SLSC model (29), a result similar to Proposition 1 is obtained. Thus, we can write each subsequent curve as a sum of the same transformation of the initial curve and a term spanned by $B(\tau)$.

Theorem 3. *For an initial yield curve $y(t_0,\tau)$ and interest rates that follow the stochastic level, slope, curvature model (29), subsequent yield curves at times $t_1, t_2, \ldots$ with $t_{n+1} - t_n = \Delta_n$ take the form*

$$y(t_{n+1},\tau) = \frac{1}{\tau}\left[(\tau+\Delta_n)\,y(t_n,\tau+\Delta_n) - \Delta_n y(t_n,\Delta_n)\right] + B(\tau)\,\widetilde{f}_{n+1} \tag{33}$$

for $B(\tau) = (B_{1:3}(\tau), B_{4:7}(\tau))$ from (30) and (32), and with factors

$$\widetilde{f}_{n+1} = (I - H(\Delta_n))\,\overline{f} + v_{n+1}$$

for $v_1, v_2, \ldots$ a sequence of serially independent $N(0,\Omega(\Delta_n))$ vectors. The transition matrix H is given by

$$H(u) = \begin{pmatrix} e^{-bu} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{-au} & aue^{-au} & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{-au} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{-2au} & 0 & 2aue^{-2au} & (1-au)2aue^{-2au} \\ 0 & 0 & 0 & 0 & e^{-2bu} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{-2au} & -2aue^{-2au} \\ 0 & 0 & 0 & 0 & 0 & 0 & e^{-2au} \end{pmatrix},$$

and the upper left 3×3 *submatrix of* Ω *is*

$$\Omega_{1:3}(u) = \begin{pmatrix} \sigma_1^2 u B_5(u) & 0 & 0 \\ 0 & \sigma_2^2 u B_4(u) + \sigma_3^2 u\,[B_6(u) - B_7(u)]/2 & \sigma_3^2 u B_6(u)/2 \\ 0 & \sigma_3^2 u B_6(u)/2 & \sigma_3^2 u B_4(u) \end{pmatrix},$$

while the remaining entries are zero. The long-run factor levels are

$$\overline{f} = \begin{bmatrix} \frac{\sigma_1}{b}\left(\frac{\sigma_1}{b} + \lambda_1\right), & \overline{f}_3 + \frac{\sigma_2}{a}\left(\frac{\sigma_2}{a} + \lambda_2\right), & \frac{\sigma_3}{a}\left(\frac{\sigma_3}{a} + \lambda_3\right), & -\frac{\sigma_2^2}{2a^2} - \frac{\sigma_3^2}{2a^2}, & -\frac{\sigma_1^2}{2b^2}, & -\frac{3\sigma_3^2}{4a^2}, & \frac{\sigma_3^2}{4a^2} \end{bmatrix}'.$$

When the initial yield curve has SLSC shape (i.e., is spanned by $B(\tau)$), all subsequent curves take SLSC shape, as well, by Theorem 2. The resulting factor dynamics are given in the following theorem that generalizes Proposition 2 to the three-dimensional stochastic case.

Theorem 4. *For an initial curve on the form* $y(t_0, \tau) = B(\tau) f_0$ *and interest rates that follow the stochastic level, slope, curvature model (29), subsequent yield curves at times* $t_1, t_2, \ldots$ *with* $t_{n+1} - t_n = \Delta_n$ *are on the form*

$$y(t_{n+1}, \tau) = B(\tau) f_{n+1} \tag{34}$$

with factors

$$f_{n+1} = \overline{f} + H(\Delta_n)\left[f_n - \overline{f}\right] + v_{n+1}, \tag{35}$$

where $v_1, v_2, \ldots$ *is a sequence of serially independent* $N(0, \Omega(\Delta_n))$ *vectors and* $\overline{f}$*,* H*, and* Ω *all are the same as in Theorem 3.*

In the yield curve at time t_{n+1}, the term $B(\tau)H(\Delta_n)f_n$ reflects the impact of the previous curve at t_n, which stems from the appropriate generalization of (24) from ANS to SLSC curve shape. The impact of the SLSC dynamic model is captured by the additional term $B(\tau)\left[(I - H(\Delta_n))\overline{f} + v_{n+1}\right]$, recognized as $B(\tau)\widetilde{f}_{n+1}$ from Theorem 3 for general yield curve shapes. This lends further credence to the observation that that consistency between curve shape and dynamic model requires that the curve shape not only accommodate the yield curve changes induced by the dynamic interest rate model, but also that it is invariant to the initial term in (33).

We estimate the SLSC model in three ways analogously to the model with a single stochastic factor and refer to these specifications as the SLSC curve, the SLSC dynamic, and the SLSC combined model, respectively. The SLSC curve model is estimated by performing a

restricted seven-factor analysis with loading functions parameterized by $B(\tau)$. Then for general curve shape we estimate the SLSC dynamic model for $\widetilde{y}_{t_n}$ given by (23) with z_n three-dimensional, $C = B_{1:3}M_{1:3}(\Delta_n)$ for $M_{1:3}M_{1:3}' = \Omega_{1:3}$ and $\mu = B(I - H(\Delta_n))\overline{f}$, where B is $m \times 7$ with typical row $B(\tau_i)$. Lastly, for SLSC yield curves spanned by $B(\tau)$ we estimate the SLSC combined model using the Kalman filter with measurement equation (34), adding idiosyncratic error terms with variances ψ_i to the yields, and transition equation (35). In estimations we set $b = 0.02$ in (30) and (32) to avoid problems with non-stationary factor dynamics. For this value the first factor's loading on the 10-year yield, the longest maturity in our empirical study, is $B_1(10) = 0.91$, and the factor approximately impacts yields equally over the relevant range.

From Table 6, if the parameter a is estimated in the SLSC curve model, it takes a value very similar to the corresponding estimates imposing ANS curve shape, but estimates somewhat larger are obtained in the estimations using the SLSC dynamic model alone. The estimated a value in the fully consistent model based on the combination of SLSC curve shape and dynamics is close to those of the SLSC curve model, whereas the rolling estimate is similar to those of the SLSC curve model, whereas the rolling estimate is similar to those of the SLSC dynamic model. These findings stand in contrast to those for the HW model and the combination of this with the consistent ANS curve shape, where low a values are obtained throughout, and causing the level-slope-curvature structure disappear in these models. The SLSC approach with three stochastic factors allows us to work in a fully consistent model with level-slope-curvature characteristics matching those suggested by basic cross-sectional calibrations. The rolling estimates of a in Figure 7 exhibit time series behavior similar to those from the NS model in Figure 5, except that the somewhat dramatic increase in a around 1988 (i.e., rolling windows around 1985-1988) in the NS case is not seen in the SLSC model.

Table 6 shows further that the volatilities σ_1 and σ_2 of the level and slope factors in the SLSC model are roughly equal, whereas the volatility σ_3 of the curvature factor is about twice as high. The left panel of Figure 8 shows that the volatilities of the level and slope factors in the model combining curve shape and dynamics are similar period by period and stable over the period from 1992 through 2005. Outside of this period, they take values twice as large. The volatility of the curvature factor is higher except in the first two years, and varies more through time, with two peaks around 2004 and 2011. The remaining columns of Table 6 show that not all market prices of factor risks are priced. The market price of slope risk, λ_2, takes values, both in absolute and relative terms, close to twice as

large as the two remaining factor risks. This follows also from the right panel of Figure 8, where we see that λ_2 in the SLSC combined model is typically largest in magnitude. It is negative during most of the sample period, but positive around 2005-2008 and after a certain point in 2012. The market price of level risk, λ_1, is relatively close to zero, varies less over time and is nearly always negative. The market price of curvature risk, λ_3, varies almost as much λ_2 through time, but switches sign often.

Figure 9 shows time series plots of the fitted values of the three stochastic factors in the SLSC model. The level factor is the smoothest of the three, with the slope factor and curvature being somewhat more volatile, and the curvature factor moving the fastest and changing sign most frequently.

From Table 5, the SLSC combined model produces idiosyncratic standard deviations that are comparable to those based on cross-sectional curve-fitting (restricted factor analysis), both in size and pattern across maturities, and much higher likelihood value, showing the importance of the dynamics. Similarly, the restricted three-factor analysis of the adjusted yield changes $\widetilde{y}$ produces idiosyncratic standard deviations very similar to those from a corresponding unrestricted classical analysis of $\widetilde{y}$.

The HJM drift restriction (17) is tested again in the SLSC dynamic model because arbitrage-free consistency relies on this. The third and fourth columns of Table 8 show results for the restricted and unrestricted models, respectively. Again, restricted μ is given in terms of the seven parameters $\eta = (a, \sigma, \lambda)$ as $B(I - H(\Delta_n))\overline{f}$. The parameters (a, σ) are very similar in the restricted and unrestricted estimations, and also the resulting μ vectors are now much closer, in contrast to the HW case (first two columns of table). This shows that the pattern of time-series averages of adjusted yield changes is well-matched by the particular SLSC curve given by coefficients (factor averages) $(I - H(\Delta_n))\overline{f}$ consistent with the absence of arbitrage opportunities. The LR-statistic takes the value 4.55, for a p-value of 47% in the asymptotic χ^2 distribution on five degrees of freedom (the restricted model introduces three market prices of risk and drops eight parameters in μ), i.e., the test fails to reject at all conventional levels.

From the hedging performance documented in Tables 3 and 4, the SLSC approach clearly dominates all others considered. For the single coupon bond target, roughly the same hedging performance is obtained from restricted factor analysis based on the SLSC curve shape, exploiting SLSC dynamics (restricted factor analysis of adjusted yield changes),

and a combination of the two, using the Kalman filter. The SLSC combined model indeed has the lowest RMSE at 6.08bp with the RMSE reduced by more than 25% compared to the HW model. For the coupon bond portfolio target, improved performance is also obtained by exploiting the SLSC dynamics. RMSE is 15.33bp based on cross-sectional SLSC curve shapes in the out-of-sample hedging rolling estimation case, compared to 12.37bp for the SLSC dynamic model, and 12.85bp when combining SLSC curve shape and dynamics. Of all other cases considered, the HW dynamic model without curve shape imposed is best, at 22.55bp (rolling estimation), about 1.5 times as high as the worst of the SLSC based methods, and nearly two times the RMSE of the SLSC combined curve shape and dynamics based method.

If the rolling a estimates are replaced by the fixed value $a = 0.731$, corresponding to the Diebold et al. (2006) specification, RMSE is 6.10bp for the single coupon target and 10.87bp for the portfolio target, but these constant estimates may be less relevant than those behind the feasible rolling window out-of-sample hedge portfolios. These results suggest the importance for practical hedging purposes of a model that is consistent, reflects the required level, slope, and curvature structure of the market, and is general enough that the arbitrage condition is not violated.

D.1. Relation to affine term structure models

In this section we briefly clarify the relation between our SLSC model and the class of affine term structure models of Duffie and Kan (1996). In particular, a restricted special case of our model, with the four deterministic factors fixed at their long-run levels, $f_{4:7}(t) \equiv \overline{f}_{4:7}$, coincides with the independent factor model considered by CDR. To see this, note that the SLSC yield curve family is not the minimal family consistent with the SLSC dynamic model. If $f_{4:7}(t)$ at some point t takes the value $\overline{f}_{4:7}$, then from (35) it remains constant at this level.[13] Thus, the set of yield curves on the form

$$B_{4:7}(\tau)\overline{f}_{4:7} + \phi_{1:3}B_{1:3}(\tau), \tag{36}$$

where the three coefficients $\phi_{1:3}$ can take any value, forms a consistent family. If the initial yield curve happens to take the form (36), then the resulting restricted special case of the SLSC model is a three-factor affine model.

[13] Equivalently, the factor values $f_{4:7}$ ensuring no drift in the $B_{4:7}$ directions solve the problem $\mathcal{V}(B_{4:7}(\tau)f_{4:7}) + B_{4:7}(\tau)C_{4:7} = 0$ in the notation of the proof of Theorem 2 in the Appendix, i.e., $f_{4:7} = {h_{4:7}}^{-1}C_{4:7} = \overline{f}_{4:7}$.

Corollary 2. *If the initial yield curve $y(t_0,\tau)$ is spanned by $B(\tau)$ with the deterministic factors at their long-run levels, $f_{4:7}(t_0) = \overline{f}_{4:7}$, then the SLSC model is an affine three-factor model, in particular an $A_0(3)$ model in the Dai and Singleton (2000) notation. The parameters of the short rate, $r_t = \delta_0 + \delta_X' X_t$, are*

$$\delta_0 = \frac{\sigma_1^2}{2b^2} + \frac{\sigma_2^2}{2a^2} + \frac{\sigma_3^2}{2a^2}, \qquad \delta_X' = (1,1,0),$$

and the dynamics under the risk-neutral probability measure, $\mathbb{Q}$, for the factors $X_t = f_{1:3}(t) - \overline{f}_{1:3}^{\mathbb{Q}}$ (imposing $\lambda = 0$ on $\overline{f}_{1:3}$ from Theorem 3 to get $\overline{f}_{1:3}^{\mathbb{Q}}$) are given by

$$dX_t = \kappa(\theta - X_t)\,dt + \Sigma\,dW_t,$$

with parameters

$$\kappa = \begin{pmatrix} b & 0 & 0 \\ 0 & a & -a \\ 0 & 0 & a \end{pmatrix}, \qquad \Sigma = \operatorname{diag}(\sigma_1,\sigma_2,\sigma_3),$$

and $\theta = 0$. The relation between yield curves and factors is

$$y(t,\tau) = -A(\tau) + B_{1:3}(\tau) X_t, \tag{37}$$

with

$$\begin{aligned} A(\tau) &= \sigma_1^2\left(-\frac{1}{b^3}\frac{1-e^{-b\tau}}{\tau} + \frac{1}{4b^3}\frac{1-e^{-2b\tau}}{\tau}\right) + \sigma_2^2\left(-\frac{1}{a^3}\frac{1-e^{-a\tau}}{\tau} + \frac{1}{4a^3}\frac{1-e^{-2a\tau}}{\tau}\right) \\ &+ \sigma_3^2\left(\frac{1}{a^2}e^{-a\tau} - \frac{1}{4a}\tau e^{-2a\tau} - \frac{3}{4a^2}e^{-2a\tau} - \frac{2}{a^3}\frac{1-e^{-a\tau}}{\tau} + \frac{5}{8a^3}\frac{1-e^{-2a\tau}}{\tau}\right). \end{aligned}$$

The particular independent factor affine three-factor model of CDR corresponds to the case $b \downarrow 0$.[14] Even if the directions $B_{4:7}$ are removed from the drift function, they still enter the yield curve, now through $A(\tau)$ with constant coefficients, and thus are required for consistent curve shapes, cf. Theorem 2. Since changes in the $B_{1:3}$ directions are stochastic, they can clearly not be removed, so the minimal consistent curve shape for the SLSC model is (37), or $-A(\tau) + B_{1:3}(\tau)\phi_{1:3}$, where $\phi_{1:3}$ can take any value.

[14] Strictly, the special case of the SLSC model considered in the present section is the model that results from applying the method of CDR with $\kappa_{11} > 0$ ($\varepsilon > 0$ in their notation), but in the limit $\kappa_{11} = 0$ (i.e., $\varepsilon = 0$), stable long-run factor levels are precluded, and the model is no longer affine (there is a discontinuity at zero).

We estimate the reduced SLSC model with curves restricted to the form (36) using the Kalman filter with measurement equation (37), still including an error term with variance Ψ, and transition equation

$$X_{t_{n+1}} = \overline{X} + H_{1:3}(\Delta)\left[X_{t_n} - \overline{X}\right] + v_{n+1}, \tag{38}$$

where the v_{n+1} are i.i.d. $N(0, \Omega_{1:3}(\Delta))$ with $\Omega_{1:3}$ from Theorem 3, and $\overline{X}$ the long-run or mean-reversion levels of factors under the objective probability measure,

$$\overline{X} = \kappa^{-1}\text{diag}(\sigma)\lambda = \left[\frac{\sigma_1}{b}\lambda_1, \frac{\sigma_2}{a}\lambda_2 + \frac{\sigma_3}{a}\lambda_3, \frac{\sigma_3}{a}\lambda_3\right]'. \tag{39}$$

From Table 6, both point estimates and standard errors of (a, σ, λ) are similar to those in the full SLSC model. Also Ψ is similar, but the affine model restrictions are clearly rejected based on the likelihood values in Table 5. In terms of hedging performance, Tables 3 and 4, the restricted affine SLSC model is on par with the cross-sectional NS approach for both hedging targets, and so very far from the unrestricted SLSC model.

E. Statistical performance evaluation

In order to evaluate the improvements in hedging performance from each model statistically, we construct a standard t-statistic along the lines of Diebold and Mariano (1995) and Giacomini and White (2006) with duration matching as fixed benchmark and each procedure as alternative. We also do a Model Confidence Set (MCS) procedure (Hansen, Lunde, and Nason, 2011) to compare the performance of all the hedging procedures under consideration. For a fixed significance level, α, the method identifies the MCS, from the set of competing procedures, which contains the best models with $1-\alpha$ probability. See the Internet Appendix for additional details on their implementation.

From Table 3 and 4 it stands out that the statistical performance evaluation echoes our findings from above that the unrestricted SLSC model is superior in hedging performance. For the single coupon bond target, only the HW model provides a statistically significant improvement over the duration matching approach, except from the SLSC curve, dynamic and combined approaches. When comparing all procedures against each other, only the unrestricted SLSC approaches enter the MCS. For the portfolio coupon bond target, all our examined procedures improve upon duration matching significantly, however only the unrestricted SLSC approaches are contained in the MCS.

IV. Conclusion

The market prices of all traded bonds depend on the yield curve, and the hedging of selected interest rate-sensitive claims using others as instruments is among the most important purposes of bond market trading. Hitherto, attempts to improve on the basic duration matching approach have involved generalized duration matching based either on a factor analysis or a parametrized yield curve shape. We show in this paper that improved empirical hedging performance is obtained by exploiting interest rate dynamics rather than either of these purely cross-sectional approaches. Asset pricing theory promises a further improvement because the absence of arbitrage opportunities imposes cross-restrictions on interest rate dynamics and the yield curve shape. This additional improvement in hedging performance is realized empirically once we introduce a new term structure model involving three stochastically varying factors corresponding to level, slope, and curvature. Of course, the level, slope, and curvature features of the bond market have been noted frequently in the purely cross-sectional literature, e.g., in the factor analysis of Litterman and Scheinkman (1991), and the parametrized yield curve shape of Nelson and Siegel (1987). The dynamic model of Hull and White (1990) also generates curves with these three features, but only the slope factor is stochastic. Our results suggest the importance for practical hedging purposes of a model that consistently combines dynamics and curve shape, and involves three genuinely stochastic time-varying factors representing level, slope, and curvature. Furthermore, we find that a particular restriction on our general SLSC model to the affine class which reproduces the independent factor model considered by Christensen et al. (2011) does not improve very much on empirical hedging performance relative to the purely cross-sectional approaches.

References

AGCA, S. (2005): "The performance of alternative interest rate risk measures and immunization strategies under a Heath-Jarrow-Morton framework," *Journal of Financial and Quantitative Analysis*, 40, 645–669.

ANDREWS, D. W. K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817–858.

BJÖRK, T. AND B. J. CHRISTENSEN (1999): "Interest rate dynamics and consistent forward rate curves," *Mathematical Finance*, 9, 323–348.

BRACE, A. AND M. MUSIELA (1994): "A multifactor gauss markov implementation of Heath, Jarrow, and Morton," *Mathematical Finance*, 4, 563 – 576.

BRAVO, J. M. V. AND C. M. P. D. SILVA (2006): "Immunization using a stochastic-process independent multi-factor risk: The Portuguese experiment," *Journal of Banking and Finance*, 30, 133–156.

BUENO-GUERRERO, A., M. MORENO, AND J. F. NAVAS (2015): "Immunization of bond portfolios: a new general framework," Working Paper.

BURASCHI, A. AND F. CORIELLI (2005): "Risk management implications of time-inconsistency: Model updating and recalibration of no-arbitrage models," *Journal of Banking and Finance*, 29, 2883–2907.

CARCANO, N. AND DALL'O (2011): "Alternative models for hedging yield curve risk: An empirical comparison," *Journal of Banking and Finance*, 35, 2991–3000.

CARRARO, C. (1988): "Square root kalman filter algorithms in econometrics," *Computer Science in Economics and Management*, 1, 41–51.

CHAMBERS, D. R., W. T. CARLETON, AND R. W. MCENALLY (1988): "Immunizing default-free bond portfolios with a duration vector," *Journal of Financial and Quantitative Analysis*, 23, 89–104.

CHAPMAN, D., J. LONG, AND N. PEARSON (1999): "Using proxies for the short rate: When are three months like an instant?" *The Review of Financial Studies*, 12, 763 – 806.

CHRISTENSEN, B. J. AND M. VAN DER WEL (forthcoming): "An Asset Pricing Approach to Testing General Term Structure Models," .

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.

CORONEO, L., K. NYHOLM, AND R. VIDOVA-KOLEVA (2011): "How arbitrage-free is the Nelson-Siegel model?" *Journal of Empirical Finance*, 18, 393–407.

COX, J. C., J. E. INGERSOLL, AND S. A. ROSS (1985): "A theory of the term structure of interest rates," *Econometrica*, 53, 385–408.

DAI, Q. AND K. SINGLETON (2000): "Specification analysis of affine term structure models," *The Journal of Finance*, 55, 1943–1978.

DÍAZ, A., M. D. L. O. GONZÁEZ, E. NAVARRO, AND F. S. SKINNER (2009): "An evaluation of contingent immunization," *Journal of Banking and Finance*, 33, 1874–1883.

DIEBOLD, F. X., L. JI, AND C. LI (2006): "A three-factor yield curve model: non-affine structures, systematic risk sources, and generalized duration," in *Long-run growth and shorth-run stabilization: essays in memory of Albert Ando*, ed. by L. R. Klein, Cheltenham, UK: Edward Elgar, 240–274.

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 S. MARIANO (1995): "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

DUFFIE, D. AND R. KAN (1996): "A yield-factor model of interest rates," *Mathematical Finance*, 6, 379–406.

FILIPOVIĆ, D. (1999): "A note on the Nelson-Siegel family," *Mathematical Finance*, 9, 349–359.

FISHER, L. AND R. L. WEIL (1971): "Coping with the risk of interest rate fluctuations: returns to bondholders from naive and optimal strategies," *The Journal of Business*, 44, 408–431.

GALLUCCIO, S. AND A. RONCORONI (2006): "A new measure of cross-sectional risk and its empirical implications for portfolio risk management," *Journal of Banking and Finance*, 30, 2387–2408.

GIACOMINI, R. AND H. WHITE (2006): "Tests of conditional predictive ability," *Econometrica*, 74, 1545–1578.

HANSEN, P. R., A. LUNDE, AND J. M. NASON (2011): "The model confidence set," *Econometrica*, 79, 453–497.

HARRISON, J. M. AND D. M. KREPS (1979): "Martingales and arbitrage in multiperiod securities markets," *Journal of Economic Theory*, 20, 381–408.

HEATH, D., R. JARROW, AND A. MORTON (1992): "Bond pricing and the term structure of interest rates," *Econometrica*, 60, 77–106.

HO, T. AND S. LEE (1986): "Term structure movement and pricing interest rate contingent claims," *The Journal of Finance*, 41, 1011–1029.

HO, T. S. Y. (1992): "Key rate durations: measure of interest rate risks," *The Journal of Fixed Income*, 2, 29–44.

HULL, J. AND A. WHITE (1990): "Pricing interest-rate-derivative securities," *The Review of Financial Studies*, 3, 573–592.

INGERSOLL, J. E. (1983): "Is immunization feasible? Evidence from the CRSP data," in *Innovation in bond portfolio management: duration analysis and immunization*, ed. by G. Bierwag, G. Kaufman, and A. Toevs, JAI Press Inc., Greenwich, CT, 163–184.

JOSLIN, S., K. J. SINGLETON, AND H. ZHU (2011): "A new perspective on gaussian dynamic term structure models," *Review of Financial Studies*, 24, 926–970.

KOOPMAN, S. J., N. SHEPHARD, AND J. A. DOORNIK (1999): "Statistical algorithms for models in state space using SsfPack 2.2," *The Econometrics Journal*, 2, 113–166.

KRIPPNER, L. (2015): "A theoretical foundation for the Nelson-Siegel class of yield curves models," *Journal of Applied Econometrics*, 30, 97–118.

LEIBOWITZ, M. AND A. WEINBERGER (1981): "The uses of contingent immunization," *Financial Analysts Journal*, 8, 51–55.

——— (1982): "Contingent immunization - part I: risk control procedures," *Financial Analysts Journal*, 36, 17–31.

——— (1983): "Contingent immunization - part II: problem areas," *Financial Analysts Journal*, 39, 35–50.

LITTERMAN, R. AND J. SCHEINKMAN (1991): "Common factors affecting bond returns," *The Journal of Fixed Income*, 54–61.

MACAULAY, F. R. (1938): "Some theoretical problems suggested by the movements of interest rates, bond yields and stock prices in the United States since 1856," *National Bureau of Economic Research*.

NAWALKHA, S. K. AND D. R. CHAMBERS (1997): "The M-vector model: derivation and testing of extensions to the M-squared model," *The Journal of Portfolio Management*, 23, 92–98.

NAWALKHA, S. K., G. M. SOTO, AND J. ZHANG (2003): "Generalized M-vector models for hedging interest rate risk," *Journal of Banking and Finance*, 27, 1581–1604.

NELSON, C. AND A. SIEGEL (1987): "Parsimonious modelling of yield curves," *The Journal of Business*, 60, 473–489.

NELSON, J. AND S. SCHAEFER (1983): "The dynamics of the term structure and alternative portfolio immunization strategies," in *Innovations in bond portfolio management: duration analysis and immunization*, ed. by G. Bierwag, G. Kaufman, and A. Toevs, JAI Press Inc., Greenwich, CT, 61–102.

REDINGTON, F. M. (1952): "Review of principles of life office evaluation," *Journal of the Institute of Actuaries*, 78, 286–315.

SANDERS, A. B. AND H. UNAL (1988): "On the intertemporal behavior of the short-term rate of interest," *Journal of Financial and Quantitative Analysis*, 23, 417–423.

SOTO, G. M. (2004): "Duration models and IRR management: A question of dimensions," *Journal of Banking and Finance*.

VASICEK, O. (1977): "An equilibrium characterization of the term structure," *Journal of Financial Economics*, 5, 177–188.

A. Tables and figures

Table 1: Mean and standard deviation for each of the eight data series of constant maturity zero-coupon bond yields.

	3 mns.	6 mns.	12 mns.	2 yrs.	3 yrs.	5 yrs.	7 yrs.	10 yrs.
Mean (%)	4.12	4.29	4.46	4.84	5.07	5.46	5.76	5.98
Std. Dev. (%)	2.89	2.96	3.02	3.09	3.04	2.89	2.78	2.64

Table 2: The three columns (in percentage) of the loading matrix B as depicted in Figure 3.

	B_1	B_2	B_3
3 mns.	2.828	-0.548	-0.138
6 mns.	2.917	-0.503	-0.067
12 mns.	2.991	-0.375	0.026
2 yrs.	3.081	-0.117	0.095
3 yrs.	3.034	0.052	0.087
5 yrs.	2.877	0.309	0.005
7 yrs.	2.748	0.443	-0.059
10 yrs.	2.577	0.545	-0.118

Table 3: The target is a 5-year coupon bond; the table shows statistics for different methods used to construct the hedging portfolio. The columns report the average hedging error, or bias, the standard deviation of hedging errors, the root mean squared error, and the mean absolute error. Results are in basis points (0.01%) per month. The first line is for the unhedged target return series. If a given model provides a statistically significant improvement over the duration matching approach (indicated by S) and/or is included in the MCS on a 5% significance level is reported in parentheses.

Hedging performance - single coupon bond

	Model	Bias	Std. dev.	RMSE	MAE
1	Target movement	53.15	128.41	138.80	110.09
2	Duration matching	-6.83	7.57	10.19	8.09
3	Unrestricted 3-factor Full period	1.64	11.39	11.49	8.86
4	Unrestricted 3-factor Rolling 4-year	2.19	11.58	11.76	9.06
5	Nelson-Siegel Full period	1.48	11.64	11.72	8.94
6	Nelson-Siegel Rolling 4-year	1.64	10.79	10.90	8.56
7	Nelson-Siegel a = 0.731	1.70	12.51	12.61	9.23
8	Unrestricted 4-factor Full period	2.58	10.38	10.68	8.27
9	Unrestricted 4-factor Rolling 4-year	2.29	10.22	10.46	7.89
10	Augmented NS Full period	2.61	9.82	10.14	7.78
11	Augmented NS Rolling 4-year	2.11	9.85	10.05	7.00
12	Hull-White Full period	1.96	8.28	8.49	6.63
13	Hull-White Rolling 4-year	2.02	8.79	9.00 (S,-)	6.81 (S,-)
14	HW and ANS Full period	2.03	9.18	9.39	7.13
15	HW and ANS Rolling 4-year	2.42	9.02	9.33	7.24
16	SLSC curve Full period	2.17	5.87	6.25	4.84
17	SLSC curve Rolling 4-year	2.13	5.76	6.13 (S,MCS)	4.74 (S,MCS)
18	SLSC dynamic Full period	2.10	5.73	6.10	4.70
19	SLSC dynamic Rolling 4-year	2.16	5.79	6.17 (S,MCS)	4.76 (S,MCS)
20	SLSC combined Full period	2.11	5.72	6.08	4.69
21	SLSC combined Rolling 4-year	1.98	6.05	6.36 (S,MCS)	4.92 (S,MCS)
22	SLSC restricted Full period	1.50	10.89	10.97	8.38
23	SLSC restricted Rolling 4-year	1.75	10.57	10.69	8.22

Table 4: The target is a portfolio of (2, 5, 10)-year coupon bonds in proportions (-1, 3, -1). Otherwise the format of the table is equivalent to Table 2. Results are in basis points (0.01%) per month.

Hedging performance - portfolio

	Model	Bias	Std. dev.	RMSE	MAE
1	Target movement	60.28	146.98	158.65	125.25
2	Duration matching	-13.31	54.91	56.42	41.32
3	Unrestricted 3-factor Full period	-3.57	48.14	48.20	34.53
4	Unrestricted 3-factor Rolling 4-year	0.21	44.87	44.80 (S,-)	32.90 (S,-)
5	Nelson-Siegel Full period	-4.72	49.80	49.95	35.94
6	Nelson-Siegel Rolling 4-year	-3.56	48.64	48.70 (S,-)	35.30 (S,-)
7	Nelson-Siegel a = 0.731	-1.86	35.02	35.02 (S,-)	26.62 (S,-)
8	Unrestricted 4-factor Full period	3.46	34.95	35.07	26.30
9	Unrestricted 4-factor Rolling 4-year	0.86	36.59	36.55 (S,-)	25.40 (S,-)
10	Augmented NS Full period	2.15	34.34	34.36	25.10
11	Augmented NS Rolling 4-year	1.37	26.84	26.84 (S,-)	18.73 (S,-)
12	Hull-White Full period	0.06	20.09	20.06	15.45
13	Hull-White Rolling 4-year	0.24	22.59	22.55 (S,-)	16.07 (S,-)
14	HW and ANS Full period	0.38	22.62	22.59	16.74
15	HW and ANS Rolling 4-year	2.20	29.67	29.71 (S,-)	20.92 (S,-)
16	SLSC curve Full period	0.90	13.53	13.54	10.43
17	SLSC curve Rolling 4-year	1.15	15.31	15.33 (S,MCS)	10.99 (S,-)
18	SLSC dynamic Full period	0.61	12.22	12.22	9.45
19	SLSC dynamic Rolling 4-year	0.77	12.36	12.37 (S,MCS)	9.56 (S,MCS)
20	SLSC combined Full period	0.73	12.66	12.66	9.79
21	SLSC combined Rolling 4-year	0.88	12.84	12.85 (S,MCS)	9.94 (S,-)
22	SLSC restricted Full period	-3.79	46.17	46.26	33.26
23	SLSC restricted Rolling 4-year	-2.78	44.83	44.85 (S,-)	32.95 (S,-)

Table 5: For the different estimated models the table shows estimated idiosyncratic standard deviations for each yield maturity. The reported figures are $\sqrt{\psi_i} \cdot 1000$. The two columns to the right show the maximum log-likelihood value and the number of free parameters for each model. The first eight models considered are for yield levels y, whereas the last four models are for aging-adjusted yield changes $\tilde{y}$.

Idiosyncratic standard deviations and log likelihood values

	Model	3 mns.	6 mns.	12 mns.	2 yrs.	3 yrs.	5 yrs.	7 yrs.	10 yrs.	log$L$	# params.
1	Unrestricted 3-factor	1.21	0.30	0.72	0.40	0.30	0.51	0.28	1.01	65,408	37
2	Nelson-Siegel 3-factor	1.26	0.30	0.81	0.44	0.30	0.53	0.28	1.08	64,928	23
3	Unrestricted 4-factor	0.98	0.30	0.44	0.34	0.30	0.46	0.30	0.65	66,136	42
4	Augmented NS 4-factor	1.07	0.30	0.62	0.36	0.30	0.45	0.30	0.84	65,469	27
5	SLSC curve 7-factor	1.01	0.30	0.57	0.34	0.30	0.45	0.28	0.76	65,710	45
6	HW and ANS 1-factor	6.84	5.59	4.08	1.58	0.30	2.47	3.50	4.20	57,441	11
7	SLSC combined 3-factor	1.44	0.30	0.81	0.56	0.31	0.54	0.45	0.97	74,310	15
8	SLSC restricted 3-factor	1.46	0.30	0.81	0.60	0.30	0.55	0.49	0.86	74,204	15
9	Hull-White 1-factor on $\tilde{y}$	1.27	0.94	0.68	0.39	0.28	0.30	0.45	0.58	77,792	11
10	Unrestricted 1-factor on $\tilde{y}$	1.19	0.94	0.74	0.51	0.38	0.14	0.32	0.44	78,579	24
11	SLSC dynamic 3-factor on $\tilde{y}$	0.63	0.29	0.32	0.21	0.20	0.19	0.16	0.23	81,883	15
12	Unrestricted 3-factor on $\tilde{y}$	0.66	0.25	0.32	0.22	0.20	0.21	0.11	0.26	82,748	37

Table 6: For full period estimation the table shows parameter estimates with standard errors below. For rolling four-year estimation the table reports the time-series mean.

	Model	a	σ_1	λ_1	σ_2	λ_2	σ_3	λ_3
1	Nelson-Siegel Full period	0.672 0.007						
2	Nelson-Siegel Rolling 4-year	0.851 -						
3	Augmented NS Full period	0.672 0.008						
4	Augmented NS Rolling 4-year	0.660 -						
5	SLSC curve Full period	0.687 0.007						
6	SLSC curve Rolling 4-year	0.656 -						
7	Hull-White Full period	-0.027 0.002	0.010 0.000	-0.655 0.201				
8	Hull-White Rolling 4-year	-0.047 -	0.010 -	-0.726 -				
9	HW and ANS Full period	0.119 0.001	0.014 0.000	-0.863 0.200				
10	HW and ANS Rolling 4-year	0.164 -	0.011 -	-0.938 -				
11	SLSC dynamic Full period	0.775 0.020	0.012 0.000	-0.337 0.203	0.013 0.000	-0.719 0.205	0.023 0.001	-0.567 0.218
12	SLSC dynamic Rolling 4-year	0.768 -	0.012 -	-0.312 -	0.012 -	-0.526 -	0.022 -	-0.260 -
13	SLSC combined Full period	0.671 0.007	0.011 0.000	-0.430 0.200	0.011 0.000	-0.628 0.200	0.022 0.001	-0.239 0.200
14	SLSC combined Rolling 4-year	0.759 -	0.012 -	-0.368 -	0.012 -	-0.823 -	0.022 -	-0.028 -
15	SLSC restricted Full period	0.676 0.008	0.010 0.000	-0.388 0.200	0.011 0.000	-0.788 0.200	0.022 0.001	-0.220 0.200
16	SLSC restricted Rolling 4-year	0.815 -	0.010 -	-0.433 -	0.012 -	-0.938 -	0.022 -	-0.116 -

Table 7: The transpose of the loading vector B in percent.

Maturity	3 mns.	6 mns.	12 mns.	2 yrs.	3 yrs.	5 yrs.	7 yrs.	10 yrs.
1st factor	0.067	0.087	0.104	0.131	0.142	0.151	0.147	0.137

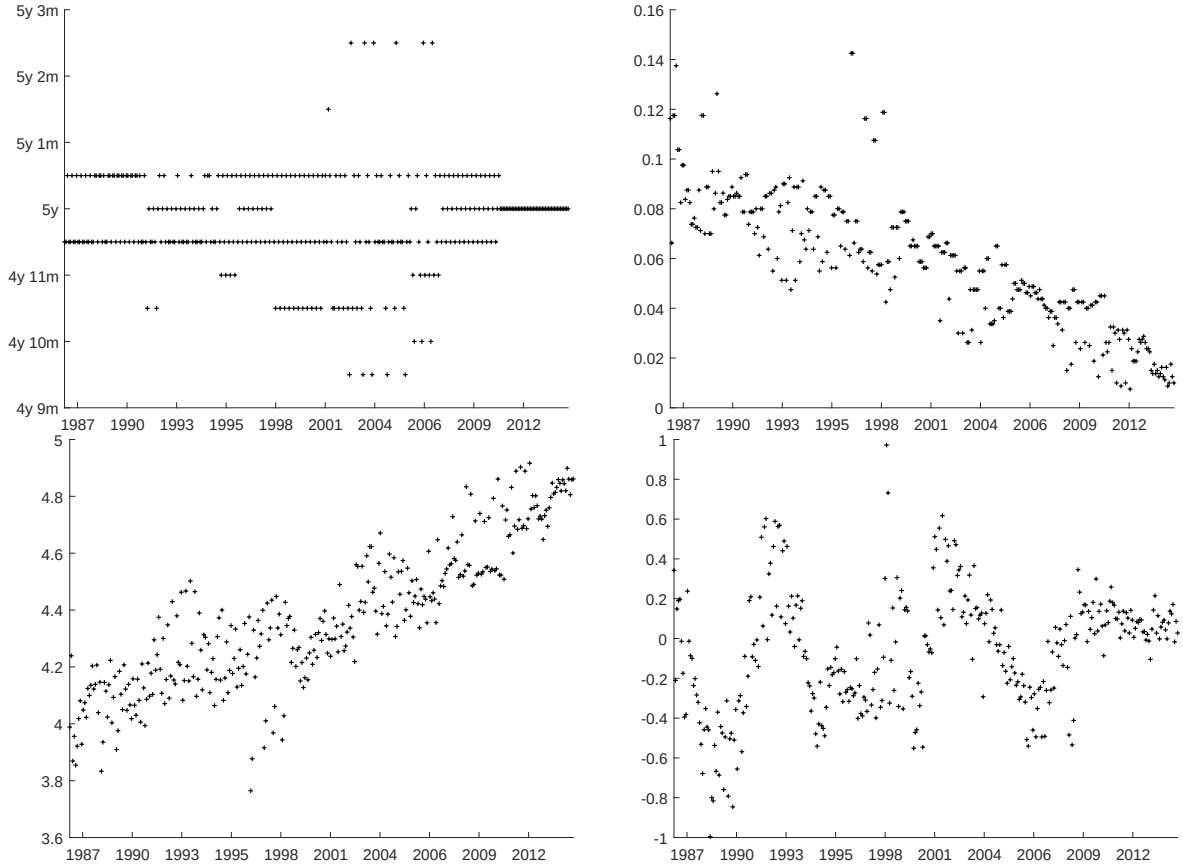
Table 8: Estimation of the Hull-White and the SLSC dynamic models, both with and without the HJM drift condition imposed. Figures for μ are in basis points, and μ is estimated freely when the drift is unrestricted, whereas with HJM drift it is calculated from parameters a, σ, and λ. The bottom of the table shows the test of restrictions imposed by the HJM drift condition using the likelihood ratio test.

	Hull-White HJM drift	Hull-White Unrestricted drift	SLSC dynamic HJM drift	SLSC dynamic Unrestricted drift
a	-0.027	-0.032	0.775	0.808
σ_1	0.010	0.009	0.013	0.013
λ_1	-0.752		-0.296	
σ_2			0.014	0.014
λ_2			-0.329	
σ_3			0.023	0.023
λ_3			-0.282	
μ_1	-1.379	-1.938	-1.637	-1.938
μ_2	-1.381	-1.859	-1.638	-1.859
μ_3	-1.386	-1.912	-1.609	-1.912
μ_4	-1.396	-1.662	-1.488	-1.662
μ_5	-1.406	-1.503	-1.350	-1.503
μ_6	-1.425	-1.313	-1.118	-1.313
μ_7	-1.443	-1.093	-0.956	-1.093
μ_8	-1.468	-0.849	-0.796	-0.849
$\log L$	77,792	77,820	81,883	81,885
# params.	11	18	15	20
LR		56.711		4.548
$\chi^2_{.95}$		14.067		11.070
p-val		0.000		0.473

Table 9: The transpose of the loading vector B in %.

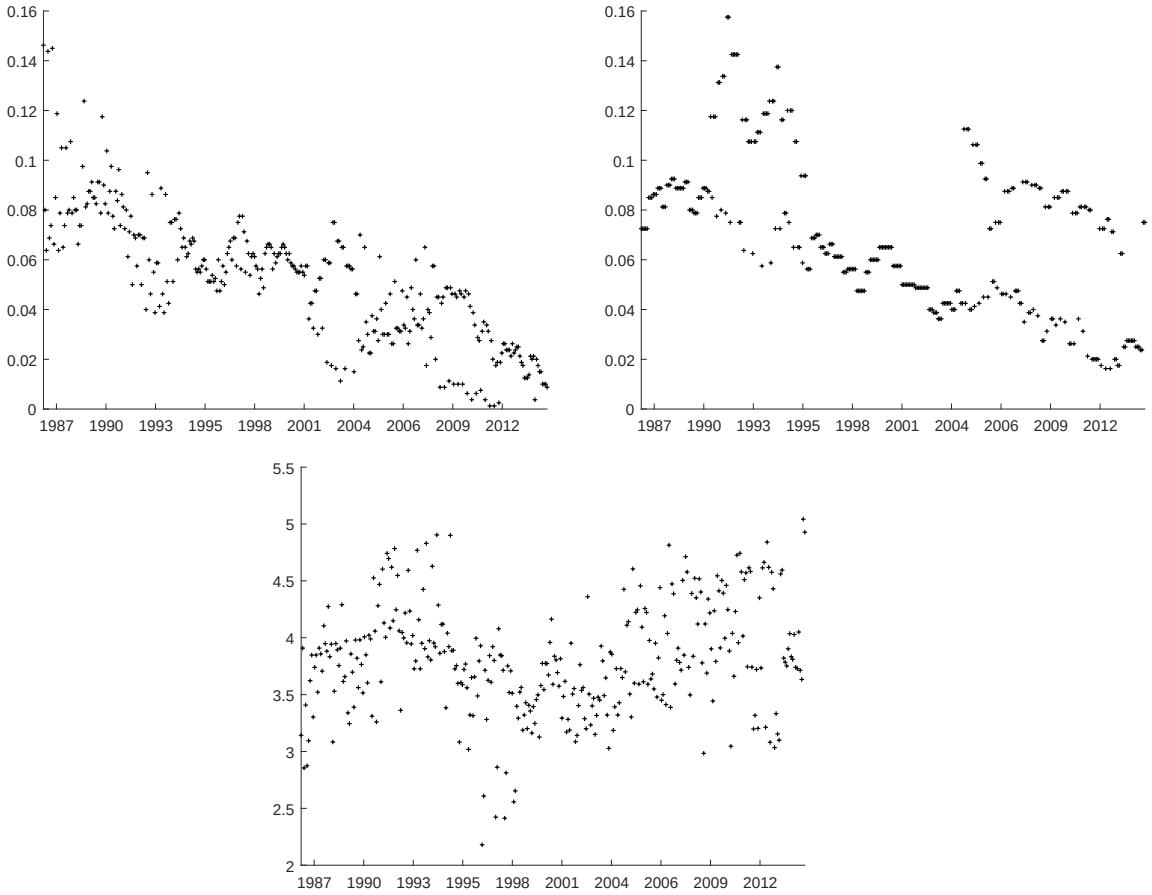
Maturity	3 mns.	6 mns.	12 mns.	2 yrs.	3 yrs.	5 yrs.	7 yrs.	10 yrs.
1st factor	0.071	0.091	0.106	0.132	0.142	0.150	0.148	0.138
2nd factor	0.041	0.032	0.002	-0.024	-0.022	-0.008	0.008	0.013
3rd factor	-0.088	-0.081	-0.062	-0.037	-0.021	0.006	0.024	0.029

Figure 1: Characteristics of 5-year target bond



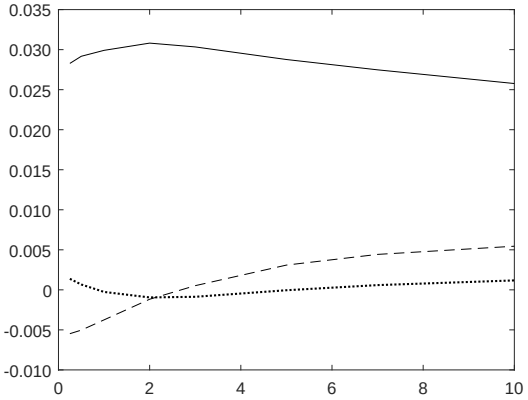
The panels in the upper row and the lower left panel show time to maturity, coupon rate, and duration, respectively, for the target 5-year bond to be hedged each month. These properties were retrieved from the CRSP monthly treasury file, selecting the coupon bond with maturity closest to five years, given a liquidity condition. The lower right panel shows the discrepancy between the actual price of the selected bonds as recorded in CRSP and the price that results from the FRB yield curve that was applied in the factor analysis. The discrepancy is shown in per cent that the FRB yield curve implied price exceeds the CRSP recorded price.

Figure 2: Characteristics of target portfolio



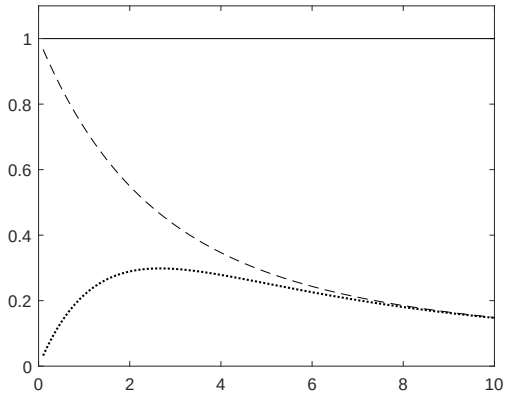
The panels in the first row show the coupon rate of the 2-year bond (left) and of the 10-year bond (right) in target portfolio. The panel in the second row plots the duration of the target portfolio which is the combination of the $(2, 5, 10)$-year bonds in proportions (-1, 3, -1).

Figure 3: Loadings in unrestricted 3-factor model



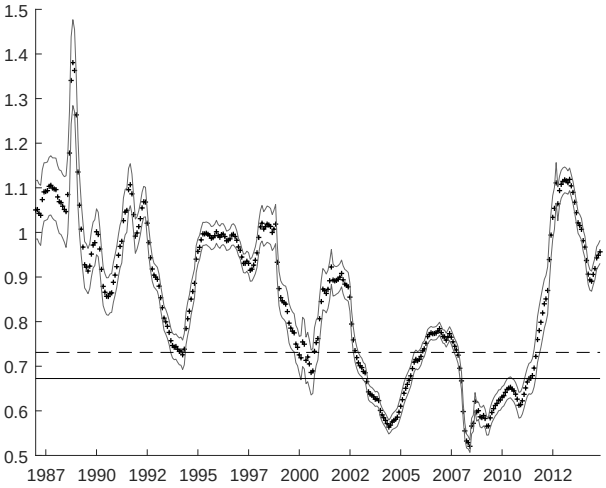
The three columns of the loading matrix B are plotted against maturity, and the values are given in percentage in Table 2.

Figure 4: Loadings in Nelson-Siegel Model



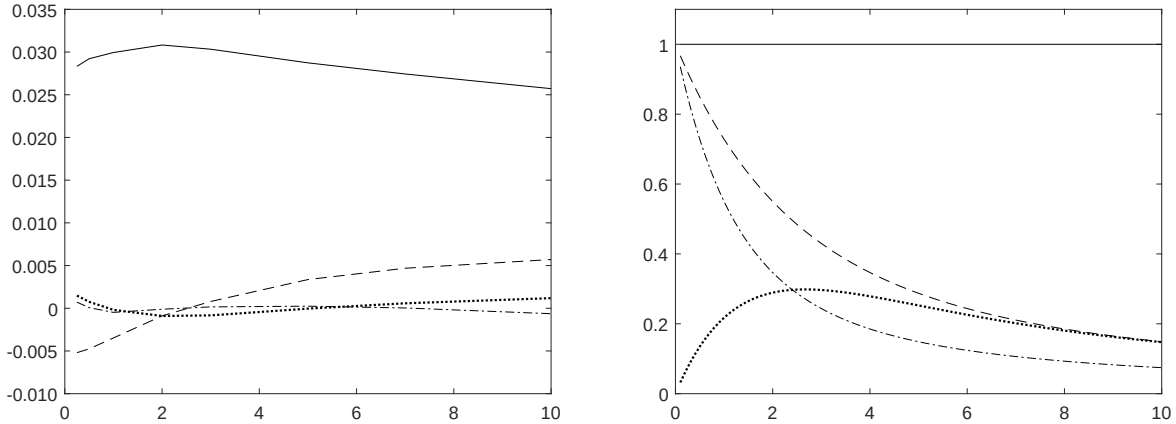
The three functions $B_j(\tau)$ in the NS model plotted against maturity τ.

Figure 5: Time series of a parameter in Nelson-Siegel Model



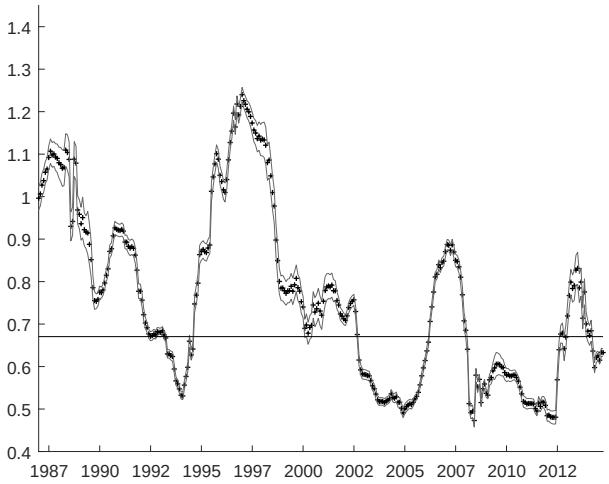
Time series of a in the rolling four-year estimations of the NS model with 95% confidence bands. The x-axis gives the end date of the four-year estimation window. The solid horizontal line indicates the full period estimate of a and the dashed line is the value chosen by Diebold et al. (2006).

Figure 6: Loadings in 4-factor model



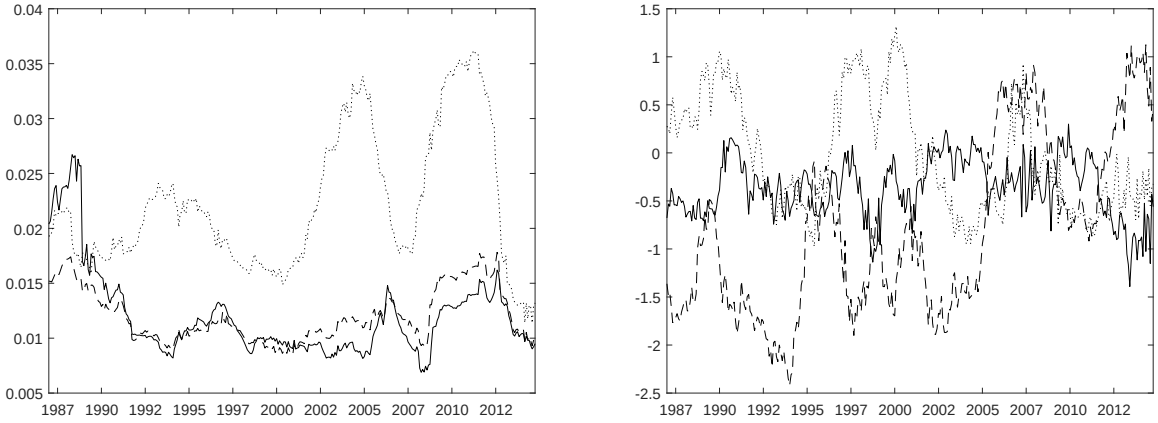
The panel to the left plots the columns B_j in the unrestricted 4-factor model against maturity and the panel to the right draws the four functions $B_j(\tau)$ from the augmented NS model.

Figure 7: Time series of a in the SLSC combined model



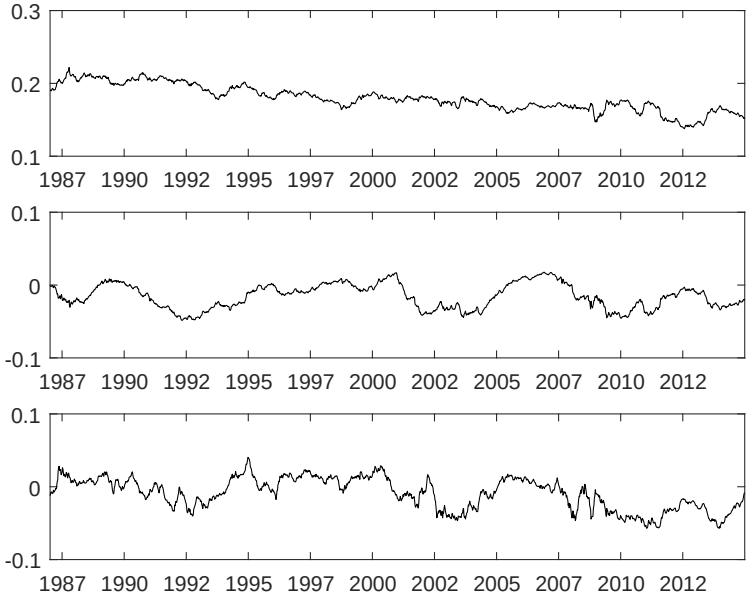
Time series of estimated a with 95% confidence bands in the rolling four-year estimation of the SLSC combined model. The x-axis gives the end date of the rolling windows. The solid horizontal line indicate the corresponding full period estimate of a.

Figure 8: Time series of σ and λ in the SLSC combined model



The left panel shows the time series estimates of the volatility parameters $(\sigma_1, \sigma_2, \sigma_3)$ in the four-year rolling window estimation of the combined SLSC model. For the same model the right panel shows the time series of rolling estimates of the market price of risk parameters $(\lambda_1, \lambda_2, \lambda_3)$. In each panel the parameter with index 1, 2, and 3 are the solid, the dashed, and the dotted line, respectively.

Figure 9: Time series of factors in the SLSC combined model



Time series of fitted factors in the SLSC combined model with the level factor, the slope factor, and the curvature factor in the upper, middle, and lower panel, respectively. The fitted factors are the smoothed Kalman filter estimates and thus use the full series of yield data, i.e., $\mathbb{E}(f_n \mid y_{t_0}, \ldots, y_{t_N})$.

Internet Appendix for
Immunization with term structure dynamics

Abstract

This Internet Appendix contains proofs of the theoretical results in the paper in Section A. Section B shows how to conduct immunization using coupon-bearing instruments. Section C details a classical maximum likelihood factor analysis of a yield factor model and presents general empirical results on our data set. Section D presents the state space form of consistent dynamic term structure models and the use of a square-root low-storage algorithm in Kalman filtering. Section E derives the immunization portfolio that minimizes the conditional mean squared hedging error and documents its performance among our models under consideration. Section F provides details on our implementation of the statistical performance evaluation.

A. Proofs

This section contains proofs of the theoretical results in the paper, as well as a number of auxiliary Lemmas and an additional Proposition 3.

Proof of Theorem 1. To find the portfolio weights that solve the minimization problem (10), we first need an expression for the hedge error variance. The generalized durations of the target stream are by (8) the vector $(\tau b)_*$, so analogously to (6) the return on the target stream is

$$r^*_{t+1} = -(\tau b)'_* \Delta f_{t+1} + \Delta \xi^*_{t+1}, \tag{A.1}$$

for ξ^*_t the idiosyncratic error in the target log-price, $\log p^*_t$. By substituting the return on an arbitrary linear portfolio of the instruments given in (6), the hedging error is

$$\begin{aligned} r^*_{t+1} - w' r_{t+1} &= -(\tau b)'_* \Delta f_{t+1} + \Delta \xi^*_{t+1} + w' \mathcal{T} (B \Delta f_{t+1} + \Delta \varepsilon_{t+1}) \\ &= \left[w' \mathcal{T} B - (\tau b)'_* \right] f_{t+1} + w' \mathcal{T} \varepsilon_{t+1} + \xi^*_{t+1} - w' \mathcal{T} y_t - \log p^*_t, \end{aligned} \tag{A.2}$$

where the second equality collects terms known at time t, i.e., $y_t = B f_t + \varepsilon_t$ and $\log p^*_t = -(\tau b)'_* f_t + \xi^*_t$. Given that the portfolio of instruments matches the generalized durations of the target, $B' \mathcal{T} w = (\tau b)_*$, the variance of the hedge error will be

$$\operatorname{var}_t \left(r^*_{t+1} - w' r_{t+1} \right) = w' \mathcal{T} \Psi \mathcal{T} w + \Psi_*, \tag{A.3}$$

for $\Psi_* = \operatorname{var}_t \left(\xi^*_{t+1} \right)$. Since the second component is not under the portfolio manager's control, minimization of the hedge error variance amounts to minimizing the first term in (A.3). The portfolio that matches generalized durations of the target for least hedge error variance therefore solves

$$\min_w w' \mathcal{T} \Psi \mathcal{T} w \quad \text{s. t.} \quad B' \mathcal{T} w = (\tau b)_* . \tag{A.4}$$

This is equivalent to the problem (A.5) in Lemma 5 below with $A = \mathcal{T} \Psi \mathcal{T}$, $g = 0$, $D = B' \mathcal{T}$, and $c = (\tau b)_*$. The solution is then by (A.6)

$$\widetilde{w} = \mathcal{T}^{-1} \Psi^{-1} B \left(B' \Psi^{-1} B \right)^{-1} (\tau b)_* = \mathcal{T}^{-1} F' (\tau b)_* ,$$

for the factor projection matrix $F = \left(B' \Psi^{-1} B \right)^{-1} B' \Psi^{-1}$. Therefore, the weights in (9) do indeed minimize hedge error variance as claimed. When the value matching constraint

$w'\iota = 1$ is added to the minimization problem (A.4) the correction of weights $\widetilde{w}$ can be found from (A.7) in Lemma 5 to be

$$w_* = \widetilde{w} + \left(1 - \widetilde{w}'\iota\right)\frac{\Lambda\iota}{\iota'\Lambda\iota},$$

for the matrix

$$\begin{aligned}\Lambda &= (\mathcal{T}\Psi\mathcal{T})^{-1} - (\mathcal{T}\Psi\mathcal{T})^{-1}\mathcal{T}B\left(B'\mathcal{T}(\mathcal{T}\Psi\mathcal{T})^{-1}\mathcal{T}B\right)^{-1}B'\mathcal{T}(\mathcal{T}\Psi\mathcal{T})^{-1} \\ &= \mathcal{T}^{-1}\Psi^{-1}\left(\Psi - B\left(B'\Psi^{-1}B\right)^{-1}B'\right)\Psi^{-1}\mathcal{T}^{-1}.\end{aligned}$$

□

Lemma 5. *Let w_u solve the unconstrained problem $\min_w (1/2)w'Aw - w'g$ for a symmetric matrix A and vector g. Then the solution to the constrained problem*

$$\min_w \frac{1}{2}w'Aw - w'g \qquad s.\ t. \qquad Dw = c \tag{A.5}$$

can be written as

$$w_c = w_u + A^{-1}D'\left(DA^{-1}D'\right)^{-1}(c - Dw_u). \tag{A.6}$$

When further adding the scaling constraint $w'\iota = 1$ to the problem (A.5), the solution is on the form

$$w_* = w_c + \frac{\Lambda\iota}{\iota'\Lambda\iota}\left(1 - w_c'\iota\right), \tag{A.7}$$

with $\Lambda = A^{-1} - A^{-1}D'\left(DA^{-1}D'\right)^{-1}DA^{-1}$.

Proof of Lemma 5. The unconstrained solution is $w_u = A^{-1}g$ and the lagrange function for (A.5) is

$$\mathcal{L} = \frac{1}{2}w'Aw - w'g - \lambda'(Dw - c).$$

This has the first order condition $0 = Aw_c - g - D'\lambda$, such that

$$w_c = A^{-1}\left(g + D'\lambda\right) = w_u + A^{-1}D'\lambda.$$

Substituting for w in the constraint gives $c = Dw_u + DA^{-1}D'\lambda$, which when λ is isolated and inserted in w_c gives the solution

$$w_c = w_u + A^{-1}D'\left(DA^{-1}D'\right)^{-1}(c - Dw_u). \tag{A.8}$$

When the constraint $w'\iota = 1$ is added to (A.5), the new solution can be found by substituting in the solution (A.8) D' with (D', ι) and c' with $(c', 1)$ to get

$$w_* = w_u + A^{-1}\begin{pmatrix} D' & \iota \end{pmatrix}\begin{bmatrix} DA^{-1}D' & DA^{-1}\iota \\ \iota' A^{-1}D' & \iota' A^{-1}\iota \end{bmatrix}^{-1}\begin{pmatrix} c - Dw_u \\ 1 - \iota' w_u \end{pmatrix}. \tag{A.9}$$

By the formula for the inverse of a block matrix

$$S^{-1} = \begin{bmatrix} DA^{-1}D' & DA^{-1}\iota \\ \iota' A^{-1}D' & \iota' A^{-1}\iota \end{bmatrix}^{-1} = \begin{bmatrix} \left(DA^{-1}D'\right)^{-1} + F\iota\iota' F'/\iota'\Lambda\iota & -F\iota/\iota'\Lambda\iota \\ -\iota' F'/\iota'\Lambda\iota & 1/\iota'\Lambda\iota \end{bmatrix}$$

for $F = \left(DA^{-1}D'\right)^{-1}DA^{-1}$. Now using that $w_c - w_u = F'(c - Dw_u)$ we get

$$S^{-1}\begin{pmatrix} c - Dw_u \\ 1 - \iota' w_u \end{pmatrix} = \begin{pmatrix} \left(DA^{-1}D'\right)^{-1}(c - Dw_u) + F\iota\iota'(w_c - w_u)/\iota'\Lambda\iota - F\iota\left(1 - \iota' w_u\right)/\iota'\Lambda\iota \\ -\iota'(w_c - w_u)/\iota'\Lambda\iota + \left(1 - \iota' w_u\right)/\iota'\Lambda\iota \end{pmatrix}.$$

Multiplying by $\begin{pmatrix} A^{-1}D' & A^{-1}\iota \end{pmatrix}$ from the left gives the last term in (A.9), such that

$$\begin{aligned} w_* &= w_u + F'(c - Dw_u) + \left(A^{-1}D'F - A^{-1}\right)\iota\iota'(w_c - w_u)/\iota'\Lambda\iota \\ &+ \left(A^{-1} - A^{-1}D'F\right)\iota\left(1 - \iota' w_u\right)/\iota'\Lambda\iota, \end{aligned}$$

and substituting $\Lambda = A^{-1} - A^{-1}D'F$ the solution is obtained,

$$\begin{aligned} w_* &= w_u + (w_c - w_u) - \frac{\Lambda\iota}{\iota'\Lambda\iota}\iota'(w_c - w_u) + \frac{\Lambda\iota}{\iota'\Lambda\iota}\left(1 - \iota' w_u\right) \\ &= w_c + \frac{\Lambda\iota}{\iota'\Lambda\iota}\left(1 - \iota' w_c\right). \end{aligned}$$

□

Proof of (17). Traded bonds have fixed time *of* maturity, $T = \tau + t$, not fixed time *to* maturity, hence write $P(t,T) = \exp(-(T-t)\,y(t,T-t))$ for the bond price, which by Itô's lemma has

$$\begin{aligned} \frac{dP(t,T)}{P(t,T)} &= -\tau dy(t,\tau) + \frac{1}{2}\tau^2\sigma(t,\tau)\sigma(t,\tau)'dt + y(t,\tau)dt + \tau\frac{\partial y}{\partial\tau}(t,\tau)dt \\ &= \left[y(t,\tau) + \tau\frac{\partial y}{\partial\tau}(t,\tau) + \frac{1}{2}\tau^2\sigma(t,\tau)\sigma(t,\tau)' - \tau\alpha(t,\tau)\right]dt - \tau\sigma(t,\tau)dW_t \\ &= \alpha_P(t,\tau)dt + \sigma_P(t,\tau)dW_t. \end{aligned}$$

By no-arbitrage there exist a market price of risk process $\lambda(t)$ such that the drift and volatility for the traded bond satisfy

$$\begin{aligned} r_t + \sigma_P(t,\tau)\lambda(t) &= \alpha_P(t,\tau) \\ r_t - \tau\sigma(t,\tau)\lambda(t) &= y(t,\tau) + \tau\frac{\partial y}{\partial \tau}(t,\tau) + \frac{1}{2}\tau^2\sigma(t,\tau)\sigma(t,\tau)' - \tau\alpha(t,\tau), \end{aligned}$$

and (17) follows. □

Proof of Proposition 1. This follows from Theorem 3 by setting $\sigma_1 = \sigma_3 = 0$. □

Proof of Proposition 2. In Theorem 4 set $\sigma_1 = \sigma_3 = f_{n,5} = f_{n,6} = f_{n,7} = 0$ and let $b \downarrow 0$. □

Proof of Theorem 2. $B_{1:3}(\tau)$ must be part of any yield curve shape consistent with the SLSC model since the yield curve changes stochastically in each of these directions by different Wiener processes. To see that the additional functions in (31) must be included as well, substitute the no-arbitrage HJM drift condition from (17) into the yield SDE to get

$$dy(t,\tau) = \left\{\frac{1}{\tau}\left[y(t,\tau) - y(t,0)\right] + \frac{\partial y}{\partial \tau}(t,\tau) + \widetilde{\alpha}(\tau)\right\}dt + \sigma(\tau)dW_t$$

with

$$\widetilde{\alpha}(\tau) = \frac{\tau}{2}\sigma(\tau)\sigma(\tau)' + \sigma(\tau)\lambda.$$

For the SLSC model which has $\sigma(\tau) = B_{1:3}(\tau)\,\mathrm{diag}(\sigma_1,\sigma_2,\sigma_3)$ the $\widetilde{\alpha}(\tau)$ term is

$$\begin{aligned} \widetilde{\alpha}(\tau) &= \frac{\tau}{2}B_{1:3}(\tau)\,\mathrm{diag}\left(\sigma_1^2,\sigma_2^2,\sigma_3^2\right)B_{1:3}(\tau)' + B_{1:3}(\tau)\,\mathrm{diag}(\sigma_1,\sigma_2,\sigma_3)\lambda \\ &= \sum_{j=1}^{3}\left[\sigma_j^2\frac{\tau}{2}B_j(\tau)^2 + \sigma_j B_j(\tau)\lambda_j\right]. \end{aligned} \tag{A.10}$$

Besides $B_{1:3}(\tau)$, the model therefore also produces drift in the directions $\tau/2B_j(\tau)^2$, for $j = 1,2,3$, which we now express in terms of $B_{1:3}(\tau)$ and the functions in (31).

First calculate that

$$\begin{aligned}\frac{\tau}{2}B_2(\tau)^2 &= \frac{\tau}{2}\left(\frac{1-e^{-a\tau}}{a\tau}\right)^2 = \frac{1-2e^{-a\tau}+e^{-2a\tau}}{2a^2\tau} = \frac{2(1-e^{-a\tau})-\left(1-e^{-2a\tau}\right)}{2a^2\tau} \\ &= \frac{1}{a}\frac{1-e^{-a\tau}}{a\tau} - \frac{1}{2a^2}\frac{1-e^{-2a\tau}}{\tau}, \end{aligned} \tag{A.11}$$

and similarly for B_1, replacing parameter a in B_2 by b,

$$\frac{\tau}{2}B_1(\tau)^2 = \frac{1}{b}\frac{1-e^{-b\tau}}{b\tau} - \frac{1}{2b^2}\frac{1-e^{-2b\tau}}{\tau}. \tag{A.12}$$

Finally, for B_3 we have

$$\begin{aligned}\frac{\tau}{2}B_3(\tau)^2 &= \frac{\tau}{2}\left(\frac{1-e^{-a\tau}}{a\tau} - e^{-a\tau}\right)^2 = \frac{1-2(1+a\tau)e^{-a\tau}+(1+a\tau)^2e^{-2a\tau}}{2a^2\tau} \\ &= \frac{1}{a}\left[\frac{1-e^{-a\tau}}{a\tau} - e^{-a\tau}\right] - \frac{1}{2a^2}\frac{1-e^{-2a\tau}}{\tau} + \frac{1}{a}e^{-2a\tau} + \frac{1}{2}\tau e^{-2a\tau}. \end{aligned} \tag{A.13}$$

Let $\widetilde{B}_{4:7}(\tau)$ be the vector of functions in (31), substitute these into (A.11)-(A.13), and collect terms in parentheses to obtain

$$\widetilde{\alpha}(\tau) = \left[B_{1:3}(\tau), \widetilde{B}_{4:7}(\tau)\right] D \tag{A.14}$$

for D the 7×1 vector of constants

$$D = \left(\frac{\sigma_1^2}{b}+\sigma_1\lambda_1, \quad \frac{\sigma_2^2}{a}+\sigma_2\lambda_2, \quad \frac{\sigma_3^2}{a}+\sigma_3\lambda_3, \quad -\frac{\sigma_1^2}{2b^2}, \quad -\frac{\sigma_2^2}{2a^2}-\frac{\sigma_3^2}{2a^2}, \quad \frac{\sigma_3^2}{a}, \quad \frac{\sigma_3^2}{2} \right)'.$$

Any function of τ that enters the drift term must also be a part of consistent curves. So to be sure that all functions in $\widetilde{B}_{4:7}$ are necessary we must check whether any of them can be cancelled by the remaining term in the drift, i.e.,

$$\frac{1}{\tau}\left[y(t,\tau)-y(t,0)\right] + \frac{\partial y}{\partial \tau}(t,\tau) = \frac{1}{\tau}\left[\frac{\partial}{\partial x}xg(x)\right]_{x=0}^{\tau} \equiv \mathcal{V}(y(t,\tau)).$$

The next step is thus to consider the implications on $y(t,\tau)$ for $\mathcal{V}(y(t,\tau))$ to include terms $-\widetilde{B}_j(\tau)D_j$, $j=4,\ldots,7$. Therefore consider the transformation $\mathcal{U}$,

$$\mathcal{U}(g(\tau)) = \frac{1}{\tau}\int_0^{\tau}[xg(\tau)]dx,$$

which is a one-sided inverse of $\mathcal{V}$ in the sense that $\mathcal{V}(\mathcal{U}(g(\tau))) = g(\tau)$. Conversely, we have that

$$\mathcal{U}(\mathcal{V}(y(t,\tau))) = y(t,\tau) - y(t,0), \tag{A.15}$$

so we can use $\mathcal{U}$ to derive the functions of τ that $y(t,\tau)$ must include for $\mathcal{V}(y(t,\tau))$ to cancel some of the terms $\widetilde{B}_j(\tau)D_j$, $j = 4,\dots,7$, in drift.

First calculate $\mathcal{U}$ of each of the $\widetilde{B}_j(\tau)$ functions. Thus

$$\mathcal{U}\left(\widetilde{B}_5(\tau)\right) = \frac{1}{\tau}\int_0^\tau \left[1 - e^{-2ax}\right]dx = 1 - \frac{1}{2a}\frac{1-e^{-2a\tau}}{\tau} = 1 - \frac{1}{2a}\widetilde{B}_5(\tau),$$

and similarly for $\widetilde{B}_4$ by substituting b for a,

$$\mathcal{U}\left(\widetilde{B}_4(\tau)\right) = 1 - \frac{1}{2b}\widetilde{B}_4(\tau).$$

For $\widetilde{B}_6$ we have

$$\mathcal{U}\left(\widetilde{B}_6(\tau)\right) = \frac{1}{\tau}\int_0^\tau xe^{-2ax}dx = \frac{1}{4a^2}\frac{1-e^{-2a\tau}}{\tau} - \frac{1}{2a}e^{-2a\tau} = \frac{1}{4a^2}\widetilde{B}_5(\tau) - \frac{1}{2a}\widetilde{B}_6(\tau),$$

while for $\widetilde{B}_7$

$$\begin{aligned}\mathcal{U}\left(\widetilde{B}_7(\tau)\right) &= \frac{1}{\tau}\int_0^\tau x^2e^{-2ax}dx = \frac{1}{4a^3}\frac{1-e^{-2a\tau}}{\tau} - \frac{1}{2a^2}e^{-2a\tau} - \frac{1}{2a}\tau e^{-2a\tau} \\ &= \frac{1}{4a^3}\widetilde{B}_5(\tau) - \frac{1}{2a^2}\widetilde{B}_6(\tau) - \frac{1}{2a}\widetilde{B}_7(\tau).\end{aligned}$$

If $\mathcal{V}(y(t,\tau))$ is to cancel out $\widetilde{B}_{4:7}(\tau)D_{4:7}$ in the drift, we may then by (A.15) conclude that $y(t,\tau) - y(t,0)$ will include the terms

$$\begin{aligned}\mathcal{U}\left(-\widetilde{B}_{4:7}(\tau)D_{4:7}\right) &= -\mathcal{U}\left(\widetilde{B}_{4:7}(\tau)\right)D_{4:7} \\ &= \widetilde{B}_{4:7}(\tau)\left(\ \tfrac{D_4}{2b},\ \ \tfrac{D_5}{2a} - \tfrac{D_6}{4a^2} - \tfrac{D_7}{4a^3},\ \ \tfrac{D_6}{2a} + \tfrac{D_7}{2a^2},\ \ \tfrac{D_7}{2a}\ \right)' - D_4 - D_5.\end{aligned}$$

This shows that it is only possible to cancel any of the $\widetilde{B}_{4:7}(\tau)$ functions in the drift if the function to be removed is already in the yield curve. To clarify this, the above result shows for instance that the drift in the $\widetilde{B}_7$ direction due to $\widetilde{\alpha}(\tau)$ only can be cancelled by $\mathcal{V}(y(t,\tau))$ if $y(t,\tau)$ includes the function $\widetilde{B}_7(\tau) = \tau e^{-2a\tau}$ exactly with the coefficient $D_7/2a = \sigma_3^2/4a$. Thus, even though there is no drift in the $\widetilde{B}_7$ direction, the yield curve must still include

$\widetilde{B}_7$ to be consistent. This holds for all the four functions, since D_j always loads on B_j, and we conclude that all functions $\widetilde{B}_{4:7}$ are necessarily part of any consistent yield curve family.

Next, we prove the converse result that the family $\mathcal{G}$ of all yield curves spanned by functions $\left(B_{1:3}, \widetilde{B}_{4:7}\right)$ is consistent with the SLSC interest rate model. Since $B_{4:7}$ in (32) is a linear transformation of $\widetilde{B}_{4:7}$, the same family is defined by $B(\tau) = [B_{1:3}(\tau), B_{4:7}(\tau)]$, the SLSC curve shape, and we choose to work with this rotation. By the definition of consistency, the family $\mathcal{G}$ is consistent with the SLSC dynamic model if starting from a curve in this family all subsequent curves produced by the interest rate model are also in $\mathcal{G}$. This will hold if both the drift and volatility functions of the SLSC model are spanned by $B(\tau)$ for all $y(t,\tau)$ in $\mathcal{G}$. For volatility, $\sigma(\tau) = B_{1:3}(\tau)\,\mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$, this is clearly the case, and that it also holds for $\widetilde{\alpha}(\tau)$ may be seen from (A.14). To explicitly specify $\widetilde{\alpha}(\tau)$ in terms of the new basis, we observe from (A.11)-(A.13) that

$$\begin{aligned}
\frac{\tau}{2} B_1(\tau)^2 &= \frac{1}{b}\left[B_1(\tau) - B_5(\tau)\right] \\
\frac{\tau}{2} B_2(\tau)^2 &= \frac{1}{a}\left[B_2(\tau) - B_4(\tau)\right] \\
\frac{\tau}{2} B_3(\tau)^2 &= \frac{1}{a}\left[B_3(\tau) - B_6(\tau) + B_7(\tau)/2\right],
\end{aligned}$$

such that we can write

$$\begin{aligned}
\widetilde{\alpha}(\tau) &= \left(\frac{\sigma_1^2}{b} + \sigma_1\lambda_1\right) B_1(\tau) + \left(\frac{\sigma_2^2}{a} + \sigma_2\lambda_2\right) B_2(\tau) + \left(\frac{\sigma_3^2}{a} + \sigma_3\lambda_3\right) B_3(\tau) \\
&- \frac{\sigma_2^2}{a} B_4(\tau) - \frac{\sigma_1^2}{b} B_5(\tau) - \frac{\sigma_3^2}{a} B_6(\tau) + \frac{\sigma_3^2}{2a} B_7(\tau) = B(\tau) C
\end{aligned} \tag{A.16}$$

for the 7×1 vector of constants

$$C = \left(\frac{\sigma_1^2}{b} + \sigma_1\lambda_1, \quad \frac{\sigma_2^2}{a} + \sigma_2\lambda_2, \quad \frac{\sigma_3^2}{a} + \sigma_3\lambda_3, \quad -\frac{\sigma_2^2}{a}, \quad -\frac{\sigma_1^2}{b}, \quad -\frac{\sigma_3^2}{a}, \quad \frac{\sigma_3^2}{2a} \right)'. \tag{A.17}$$

It is still left to check that $\mathcal{V}(y(t,\tau))$ is spanned by $B(\tau)$ for all $y(t,\tau)$ in $\mathcal{G}$, or equivalently, that the family of yield curves is closed with respect to the transformation $\mathcal{V}$.

First we calculate $\mathcal{V}\big(B_j(\tau)\big)$ for $j=1,\ldots,7$. For B_2

$$\begin{aligned}\mathcal{V}(B_2(\tau)) &= \frac{1}{\tau}\left[\frac{\partial}{\partial x}xB_2(x)\right]_{x=0}^{\tau} = \frac{1}{\tau}\left[\frac{\partial}{\partial x}\frac{1-e^{-ax}}{a}\right]_{x=0}^{\tau}\\ &= \frac{1}{\tau}\left[e^{-ax}\right]_{x=0}^{\tau} = \frac{e^{-a\tau}-1}{\tau} = -aB_2(\tau),\end{aligned}$$

and since the functional form is the same for B_1, B_4, and B_5 these give equivalent results when a is substituted with b, $2a$, and $2b$, respectively,

$$\mathcal{V}(B_1(\tau)) = -bB_1(\tau), \qquad \mathcal{V}(B_4(\tau)) = -2aB_4(\tau), \qquad \mathcal{V}(B_5(\tau)) = -2bB_5(\tau).$$

For B_3 we get

$$\begin{aligned}\mathcal{V}(B_3(\tau)) &= \mathcal{V}(B_2(\tau)) - \frac{1}{\tau}\left[\frac{\partial}{\partial x}xe^{-ax}\right]_{x=0}^{\tau} = \frac{e^{-a\tau}-1}{\tau} - \frac{1}{\tau}\left[e^{-ax}-axe^{-ax}\right]_{x=0}^{\tau}\\ &= ae^{-a\tau} = -a\left[\frac{1-e^{-a\tau}}{a\tau} - e^{-a\tau}\right] + \frac{1-e^{-a\tau}}{\tau} = -aB_3(\tau) + aB_2(\tau),\end{aligned}$$

and similarly for B_6 with a replaced by $2a$

$$\mathcal{V}(B_6(\tau)) = -2aB_6(\tau) + 2aB_4(\tau).$$

Finally, for B_7

$$\begin{aligned}\mathcal{V}(B_7(\tau)) &= \frac{1}{\tau}\Big[\frac{\partial}{\partial x}ax^2e^{-2ax}\Big]_{x=0}^{\tau} = \frac{1}{\tau}\Big[2axe^{-2ax} - 2a^2x^2e^{-2ax}\Big]_{x=0}^{\tau}\\ &= 2ae^{-2a\tau} - 2a^2\tau e^{-2a\tau} = \mathcal{V}(B_6(\tau)) - 2aB_7(\tau)\\ &= 2aB_4(\tau) - 2aB_6(\tau) - 2aB_7(\tau).\end{aligned}$$

Then, since any $y(t,\tau)\in\mathcal{G}$ can be written as $B(\tau)\phi$ for some 7×1 vector of coefficients ϕ, we may write

$$\mathcal{V}(y(t,\tau)) = \mathcal{V}(B(\tau))\phi = -B(\tau)h\phi \tag{A.18}$$

for the transformation matrix

$$h = \begin{pmatrix} b & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a & -a & 0 & 0 & 0 & 0 \\ 0 & 0 & a & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2a & 0 & -2a & -2a \\ 0 & 0 & 0 & 0 & 2b & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2a & 2a \\ 0 & 0 & 0 & 0 & 0 & 0 & 2a \end{pmatrix}. \tag{A.19}$$

We conclude that for any curve in the family $\mathcal{G}$ the changes, $dy(t,\tau)$, in the yield curve produced by the SLSC interest rate model are spanned by the functions $B(\tau)$ that define the family. Starting from a curve in $\mathcal{G}$ any subsequent curve produced by the SLSC dynamic model must, therefore, also be in the family. Therefore, the family $\mathcal{G}$ that defines the SLSC curve shape is consistent with the SLSC interest rate model. □

Proof of Corollary 1. From Theorem 2 all functions in $B(\tau)$ are necessary in any consistent curve shape. The family that they define can therefore only be reduced by either restricting some coefficients to fixed values or by combining some of the $B_i(\tau)$ functions such that they only are included in fixed combinations in the basis, e.g. $\left[x_i B_i(\tau) + x_j B_j(\tau)\right]$. The first of these options is ruled out by the statement in the corollary that the coefficients ϕ must be general, we thus need to check whether the second option is possible.

The different Wiener processes driving the changes in each of the $B_{1:3}(\tau)$ directions imply that these functions must enter the basis separately. With respect to the other four functions we must check whether any of them can be combined to define a smaller family $\mathcal{G}^*$ for which it still holds that $\mathcal{V}(y(t,\tau)) + \widetilde{\alpha}(\tau) \in \mathcal{G}^*$ for all $y(t,\tau) \in \mathcal{G}^*$. Since this must hold for all $y(t,\tau)$ in a family on the form $\phi B(\tau)$, it must in particular hold for $\phi = 0$, and thus for $\widetilde{\alpha}(\tau)$, i.e., $\widetilde{\alpha}(\tau) \in \mathcal{G}^*$. This implies that any candidate combination of the $B_{4:7}$ functions must be proportional to the coefficients in $\widetilde{\alpha}$ given by $C_{4:7}$, i.e., it must hold that $x_i/x_j = C_i/C_j$. Let h_{ij} be the 2×2 submatrix of rows and columns i and j in h corresponding to the two functions $\left[B_i, B_j\right] = B_{ij}$ that we consider combining in proportions $\left(x_i, x_j\right)' = x_{ij}$. Then $\mathcal{V}\left(B_{ij}(\tau) x_{ij}\right) = -B_{ij}(\tau) h_{ij} x_{ij}$ is spanned by the reduced family only if there exist a constant β for which this equals $-B_{ij}(\tau) x_{ij} \beta$. Therefore, it must hold that $h_{ij} x_{ij} = x_{ij} \beta$, and by proportionality to coefficients in $\widetilde{\alpha}$ also that $h_{ij} C_{ij} = C_{ij} \beta$. The resulting necessary

condition to combine any two B_i and B_j is thus that h_{ij} must have an eigenvector with entries in the same proportions as these functions enter $\widetilde{\alpha}$ with. This is not the case for any combination of the $B_{4:7}$ functions as can be seen from C and h in (A.17) and (A.19). Specifically, for B_5 in combination with any of the other functions, h_{ij} is diagonal with different diagonal values, so eigenvectors are $(s,0)$ or $(0,u)$. Any combination of the other three B_i's has h_{ij} upper triangular with identical diagonal element $2a$ and thus only the repeated eigenvector $(s,0)$. We conclude that none of the $B_{4:7}$ functions can be combined to reduce the basis and that the family is the minimal one with unrestricted coefficients on all τ functions entering the yield curve. □

Proof of Theorem 4. When the initial curve is on the form $y(t_0,\tau) = B(\tau) f_0$ and interest rates follow the SLSC dynamic model, then by Theorem 2 all subsequent curves are spanned by $B(\tau)$, so we may define coefficients or factors at each time t, $f(t)$, for which $y(t,\tau) = B(\tau) f(t)$. Then from (A.16) and (A.18) it follows that for all $t \geq t_0$

$$\begin{aligned} dy(t,\tau) &= \left[\mathcal{V}(y(t,\tau)) + \widetilde{\alpha}(\tau)\right]dt + \sigma(\tau)dW_t \\ &= \left[-B(\tau)hf(t) + B(\tau)C\right]dt + B(\tau)\Sigma dW_t \\ &= B(\tau)\left\{\left[C - hf(t)\right]dt + \Sigma dW_t\right\}, \end{aligned}$$

where we have written $\Sigma = \text{diag}(\sigma, 0_4)$ with 0_k a k-dimensional null vector. Since also $dy(t,\tau) = B(\tau)df(t)$, the movement of factors over time is

$$df(t) = \left[C - hf(t)\right]dt + \Sigma dW_t,$$

with initial factor vector $f(t_0) = f_0$. We can also write

$$df(t) = h\left[\overline{f} - f(t)\right]dt + \Sigma dW_t \tag{A.20}$$

if we define $\overline{f} = h^{-1}C$, the long-run factor levels, which in terms of basic parameters is

$$\overline{f} = \left[\ \tfrac{\sigma_1}{b}\left(\tfrac{\sigma_1}{b} + \lambda_1\right), \overline{f}_3 + \tfrac{\sigma_2}{a}\left(\tfrac{\sigma_2}{a} + \lambda_2\right), \tfrac{\sigma_3}{a}\left(\tfrac{\sigma_3}{a} + \lambda_3\right), -\tfrac{\sigma_2^2}{2a^2} - \tfrac{\sigma_3^2}{2a^2}, -\tfrac{\sigma_1^2}{2b^2}, -\tfrac{3\sigma_3^2}{4a^2}, \tfrac{\sigma_3^2}{4a^2}\ \right]'.$$

To find $y(t_{n+1},\tau)$ in terms of the factors at time t_n, we solve the factor SDE. First, observe that for the ansatz e^{ht} we get

$$d\left(e^{ht}f(t)\right) = e^{ht}hf(t)dt + e^{ht}df(t) = e^{ht}h\overline{f}dt + e^{ht}\Sigma dW_t.$$

Integrating from t_n to $t_{n+1} = t_n + \Delta_n$ and defining $f_n = f(t_n)$ produces

$$f_{n+1} = e^{-h\Delta_n} f_n + \int_0^{\Delta_n} e^{-hu} du \cdot h\overline{f} + \int_0^{\Delta_n} e^{-hu} \Sigma dW_{t_n+u}. \tag{A.21}$$

Define the function $H(u) = e^{-hu}$ which has $H'(u) = -H(u)h$. Then $\int H(u)du = -H(u)h^{-1}$, and in particular $\int_0^{\Delta_n} e^{-hu} du \cdot h = (I - H(\Delta_n))$. To calculate $H(u)$, let the diagonal of h be $d(h)$ and set $\widetilde{h} = h - d(h)$, the matrix with the off-diagonal elements of h. Then from (A.19) $\widetilde{h}$ is nilpotent of degree three and the only element of $\widetilde{h}^2$ different from zero is $[\widetilde{h}^2]_{4,7} = -4a^2$. Thus by rules of matrix exponentials

$$\begin{aligned} H(u) &= e^{-hu} = e^{-d(h)u} e^{-\widetilde{h}u} = e^{-d(h)u}\left(I - \widetilde{h}u + \widetilde{h}^2 u^2/2\right) \\ &= \begin{pmatrix} e^{-bu} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{-au} & aue^{-au} & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{-au} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{-2au} & 0 & 2aue^{-2au} & (1-au)2aue^{-2au} \\ 0 & 0 & 0 & 0 & e^{-2bu} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{-2au} & -2aue^{-2au} \\ 0 & 0 & 0 & 0 & 0 & 0 & e^{-2au} \end{pmatrix}. \end{aligned}$$

For the stochastic integral in (A.21) we can use that for a deterministic function $g(\cdot)$

$$\int_0^t g(u) dW_u \sim N\left(0, \int_0^t g(u) g(u)' du\right),$$

such that we can write $\int_0^{\Delta_n} H(u) \Sigma dW_{t_n+u} = v_{n+1}$, for $v_{n+1} \sim N(0, \Omega(\Delta_n))$. The upper 3×3 matrix of Ω is then found by

$$\begin{aligned} \Omega_{1:3}(\Delta_n) &= \int_0^{\Delta_n} H_{1:3}(u) \operatorname{diag}\left(\sigma_1^2, \sigma_2^2, \sigma_3^2\right) H_{1:3}(u)' du \\ &= \int_0^{\Delta_n} \begin{pmatrix} \sigma_1^2 e^{-2bu} & 0 & 0 \\ 0 & \sigma_2^2 e^{-2au} + \sigma_3^2 a^2 u^2 e^{-2au} & \sigma_3^2 aue^{-2au} \\ 0 & \sigma_3^2 aue^{-2au} & \sigma_3^2 e^{-2au} \end{pmatrix} du \\ &= \begin{pmatrix} \sigma_1^2 B_5(\Delta_n) & 0 & 0 \\ 0 & \sigma_2^2 B_4(\Delta_n) + \sigma_3^2 [B_6(\Delta_n) - B_7(\Delta_n)]/2 & \sigma_3^2 B_6(\Delta_n)/2 \\ 0 & \sigma_3^2 B_6(\Delta_n)/2 & \sigma_3^2 B_4(\Delta_n) \end{pmatrix} \Delta_n, \end{aligned}$$

while the remaining entries of Ω are zero. Substituting the results into (A.21), we conclude

that

$$f_{n+1} = H(\Delta_n) f_n + (I - H(\Delta_n))\overline{f} + v_{n+1}, \tag{A.22}$$

such that from an initial curve $y(t_n, \tau) = B(\tau) f_n$ the curve Δ_n later in time is $y(t_{n+1}, \tau) = B(\tau) f_{n+1}$. Further, since $v_{n+1} = \int_{t_n}^{t_{n+1}} H(u) \Sigma dW_u$, the sequence of error terms $v_1, v_2, \ldots$ are based on non-overlapping intervals of the Wiener processes W_t and they are thus serially independent. □

Proof of Theorem 3. Consider the special case with $y(t_n, \tau) = 0 = B(\tau)0$. Since this is still spanned by $B(\tau)$, we have by Theorem 4 that the curve at t_{n+1} is $y(t_{n+1}, \tau) = B(\tau)\widetilde{f}_{n+1}$, with factor dynamics obtained by substituting $\widetilde{f}_n = 0$ in (A.22), i.e.,

$$\widetilde{f}_{n+1} = (I - H(\Delta_n))\overline{f} + v_{n+1}. \tag{A.23}$$

From (A.24) in Proposition 3,

$$\begin{aligned} y(t_{n+1}, \tau) &= \int_0^{\Delta_n} \left\{ \frac{1}{\tau} \left[(x+u)\widetilde{\alpha}(x+u) \right]_{x=0}^{\tau} \right\} du \\ &+ \int_0^{\Delta_n} \left\{ \frac{1}{\tau} \left[(x+u)\sigma(x+u) \right]_{x=0}^{\tau} \right\} dW_{t_n - u} \\ &= B(\tau)\widetilde{f}_{n+1}. \end{aligned}$$

In the general solution to the yield SDE given by (A.24) the last two terms are not affected by changing the shape of the previous yield curve. Therefore for an arbitrary $y(t_n, \tau)$ we must have that the curve at t_{n+1} can be written as

$$y(t_{n+1}, \tau) = \frac{1}{\tau} \left[(\tau + \Delta_n) y(t_n, \tau + \Delta_n) - \Delta_n y(t_n, \Delta_n) \right] + B(\tau)\widetilde{f}_{n+1},$$

with $\widetilde{f}_{n+1}$ given by (A.23). □

Proposition 3 (Solution to yield SDE)**.** *Suppose the yield curve at time t_n has general shape, $y(t_n, \tau)$, that yield dynamics for $t \geq t_n$ are given by the SDE*

$$dy(t, \tau) = \alpha(t, \tau) dt + \sigma(\tau) dW_t,$$

i.e., with constant volatility function over time, and that the market price of risk is constant, λ. Then subsequent yield curves at time $t_{n+1} = t_n + \Delta_n$ can be written on the form

$$\begin{aligned} y(t_n+\Delta_n,\tau) &= \frac{1}{\tau}\left[(\tau+\Delta_n)y(t_n,\tau+\Delta_n)-\Delta_n y(t_n,\Delta_n)\right] \\ &+ \int_0^{\Delta_n}\left\{\frac{1}{\tau}\left[(x+u)\widetilde{\alpha}(x+u)\right]_{x=0}^{\tau}\right\}du \\ &+ \int_0^{\Delta_n}\left\{\frac{1}{\tau}\left[(x+u)\sigma(x+u)\right]_{x=0}^{\tau}\right\}dW_{t_{n+1}-u} \end{aligned} \tag{A.24}$$

with

$$\widetilde{\alpha}(\tau) = \frac{\tau}{2}\sigma(\tau)\sigma(\tau)' + \sigma(\tau)\lambda.$$

Proof of Proposition 3. By the HJM drift condition for yields under the objective probability measure in Brace and Musiela parametrization, (17), the drift term satisfies

$$\alpha(t,\tau) = \frac{1}{\tau}\left[y(t,\tau)-y(t,0)\right] + \frac{\partial y}{\partial \tau}(t,\tau) + \widetilde{\alpha}(\tau),$$

such that the yield curve SDE is

$$dy(t,\tau) = \left\{\frac{1}{\tau}\left[y(t,\tau)-y(t,0)\right] + \frac{\partial y}{\partial \tau}(t,\tau)\right\}dt + \widetilde{\alpha}(\tau)dt + \sigma(\tau)dW_t.$$

To solve this it is convenient to first transform to forward rates, i.e. multiply by τ and then differentiate with respect to τ to get

$$dr(t,\tau) = \frac{\partial r}{\partial \tau}(t,\tau)dt + \left\{\frac{\partial}{\partial \tau}\left[\tau\widetilde{\alpha}(\tau)\right]\right\}dt + \left\{\frac{\partial}{\partial \tau}\left[\tau\sigma(\tau)\right]\right\}dW_t. \tag{A.25}$$

Suppose $\tau = T - t$ then a differential that acts on both time and maturity relates to that acting only in the time direction by

$$d\left(r(t,T-t)\right) = dr(t,T-t) - \frac{\partial r}{\partial \tau}(t,T-t)dt.$$

Thus for $T = \tau + t_{n+1}$ and integrating from t_n to t_{n+1} we get

$$\begin{aligned} r(t_{n+1},\tau) &= r(t_n,\tau+t_{n+1}-t_n)+\int_{t_n}^{t_{n+1}}\left\{\frac{\partial}{\partial\tau}\left[(\tau+t_{n+1}-t)\widetilde{\alpha}(\tau+t_{n+1}-t)\right]\right\}dt \\ &+ \int_{t_n}^{t_{n+1}}\left\{\frac{\partial}{\partial\tau}\left[(\tau+t_{n+1}-t)\sigma(\tau+t_{n+1}-t)\right]\right\}dW_t \\ &= r(t_n,\tau+\Delta_n)+\int_0^{\Delta_n}\left\{\frac{\partial}{\partial\tau}\left[(\tau+u)\widetilde{\alpha}(\tau+u)\right]\right\}du \\ &+ \int_0^{\Delta_n}\left\{\frac{\partial}{\partial\tau}\left[(\tau+u)\sigma(\tau+u)\right]\right\}dW_{t_{n+1}-u}. \end{aligned}$$

To get back to yields integrate over maturity and divide by τ,

$$\begin{aligned} y(t_n+\Delta_n,\tau) &= \frac{1}{\tau}\int_0^\tau r(t_n,x+\Delta_n)dx+\frac{1}{\tau}\int_0^\tau\int_0^{\Delta_n}\left\{\frac{\partial}{\partial x}\left[(\tau+u)\widetilde{\alpha}(\tau+u)\right]\right\}dudx \\ &+ \frac{1}{\tau}\int_0^\tau\int_0^{\Delta_n}\left\{\frac{\partial}{\partial x}\left[(\tau+u)\sigma(\tau+u)\right]\right\}dW_{t_{n+1}-u}dx. \end{aligned}$$

The average of forward rates with maturity from Δ_n to $\tau+\Delta_n$ at time t_n is in terms of yields,

$$\begin{aligned} \frac{1}{\tau}\int_0^\tau r(t_n,x+\Delta_n)dx &= \frac{1}{\tau}\int_{\Delta_n}^{\tau+\Delta_n} r(t_n,x)dx \\ &= \frac{1}{\tau}\left[\int_0^{\tau+\Delta_n} r(t_n,x)dx-\int_0^{\Delta_n} r(t_n,x)dx\right] \\ &= \frac{1}{\tau}\left[(\tau+\Delta_n)y(t_n,\tau+\Delta_n)-\Delta_n y(t_n,\Delta_n)\right], \end{aligned}$$

and by changing the order of integration we get that the solution to the yield SDE is

$$\begin{aligned} y(t_n+\Delta_n,\tau) &= \frac{1}{\tau}\left[\left(\tau+\Delta_n\right)y(t_n,\tau+\Delta_n)-\Delta_n y(t_n,\Delta_n)\right] \\ &+ \int_0^{\Delta_n}\left\{\frac{1}{\tau}\left[(x+u)\widetilde{\alpha}(x+u)\right]_{x=0}^{\tau}\right\}du \\ &+ \int_0^{\Delta_n}\left\{\frac{1}{\tau}\left[(x+u)\sigma(x+u)\right]_{x=0}^{\tau}\right\}dW_{t_{n+1}-u} \end{aligned}$$

as stated in the proposition. □

Proof of Corollary 2. When $f_{4:7}(t_0) = \overline{f}_{4:7}$ we have by (35) in Theorem 4 that the factors $f_{4:7}$ are constant at this value. Therefore by (34) we can write all subsequent curves for

$t \geq t_0$ as

$$y(t,\tau) = B_{1:3}(\tau) f_{1:3}(t) + B_{4:7}(\tau) \overline{f}_{4:7},$$

and the diffusion for $f_{1:3}$ is obtained from the first three rows of f in (A.20) to be

$$df_{1:3}(t) = \kappa \left[\overline{f}_{1:3} - f_{1:3}(t) \right] dt + \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3) dW_t,$$

with

$$\kappa = h_{1:3} = \begin{pmatrix} b & 0 & 0 \\ 0 & a & -a \\ 0 & 0 & a \end{pmatrix}.$$

We would like to represent the model by factors that have zero mean under $\mathbb{Q}$, so we choose to rotate to the new factors $X_t \equiv f_{1:3}(t) - \overline{f}^{\mathbb{Q}}_{1:3}$, where $\overline{f}^{\mathbb{Q}}_{1:3}$ is $\overline{f}_{1:3}$ for $\lambda = 0$. Then

$$dX_t = \kappa \left(\overline{X} - X_t \right) dt + \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3) dW_t,$$

with

$$\overline{X} = \overline{f}_{1:3} - \overline{f}^{\mathbb{Q}}_{1:3} = \left[\frac{\sigma_1}{b}\lambda_1, \frac{\sigma_2}{a}\lambda_2 + \frac{\sigma_3}{a}\lambda_3, \frac{\sigma_3}{a}\lambda_3 \right],$$

and the yield curve in terms of X_t is

$$y(t,\tau) = B(\tau) \begin{bmatrix} \overline{f}^{\mathbb{Q}}_{1:3} \\ \overline{f}_{4:7} \end{bmatrix} + B_{1:3}(\tau) X_t \equiv -A(\tau) + B_{1:3}(\tau) X_t.$$

Here the function $A(\tau)$ in terms of basic parameters is

$$\begin{aligned}
A(\tau) &= -B(\tau) \left(\frac{\sigma_1^2}{b^2}, \frac{\sigma_2^2}{a^2} + \frac{\sigma_3^2}{a^2}, \frac{\sigma_3^2}{a^2}, -\frac{\sigma_2^2}{2a^2} - \frac{\sigma_3^2}{2a^2}, -\frac{\sigma_1^2}{2b^2}, -\frac{3\sigma_3^2}{4a^2}, \frac{\sigma_3^2}{4a^2} \right)' \\
&= -\frac{\sigma_1^2}{b^2} \frac{1-e^{-b\tau}}{b\tau} - \left(\frac{\sigma_2^2}{a^2} + \frac{\sigma_3^2}{a^2} + \frac{\sigma_3^2}{a^2} \right) \frac{1-e^{-a\tau}}{a\tau} + \frac{\sigma_3^2}{a^2} e^{-a\tau} \\
&+ \left(\frac{\sigma_2^2}{2a^2} + \frac{\sigma_3^2}{2a^2} + \frac{3\sigma_3^2}{4a^2} \right) \frac{1-e^{-2a\tau}}{2a\tau} + \frac{\sigma_1^2}{2b^2} \frac{1-e^{-2b\tau}}{2b\tau} - \frac{3\sigma_3^2}{4a^2} e^{-2a\tau} - \frac{\sigma_3^2}{4a^2} a\tau e^{-2a\tau} \\
&= \sigma_1^2 \left(-\frac{1}{b^3} \frac{1-e^{-b\tau}}{\tau} + \frac{1}{4b^3} \frac{1-e^{-2b\tau}}{\tau} \right) + \sigma_2^2 \left(-\frac{1}{a^3} \frac{1-e^{-a\tau}}{\tau} + \frac{1}{4a^3} \frac{1-e^{-2a\tau}}{\tau} \right) \\
&+ \sigma_3^2 \left(\frac{1}{a^2} e^{-a\tau} - \frac{1}{4a} \tau e^{-2a\tau} - \frac{3}{4a^2} e^{-2a\tau} - \frac{2}{a^3} \frac{1-e^{-a\tau}}{\tau} + \frac{5}{8a^3} \frac{1-e^{-2a\tau}}{\tau} \right).
\end{aligned}$$

The short rate as a function of X_t is obtained by letting $\tau \downarrow 0$ in (A.26). Since $\lim_{\tau \downarrow 0} B(\tau) = (1,1,0,1,1,0,0)$

we get that

$$\begin{aligned} r_t &= \lim_{\tau \downarrow 0} y(t,\tau) = \overline{f}_1^{\mathbb{Q}} + \overline{f}_2^{\mathbb{Q}} + \overline{f}_4 + \overline{f}_5 + X_{1t} + X_{2t} \\ &= \frac{\sigma_1^2}{2b^2} + \frac{\sigma_2^2}{2a^2} + \frac{\sigma_3^2}{2a^2} + X_{1t} + X_{2t} = \delta_0 + \delta_X' X_t, \end{aligned}$$

such that the values of δ_0 and δ_X given in the corollary are obtained. □

B. Hedging with coupon-bearing instruments

The expression (8) for the generalized duration vector of a payment stream facilitates not only the hedging of such a stream, but also the use of streams as hedging instruments. This includes hedging with coupon-bearing bonds instead of just zero-coupon bonds that may be unavailable. If there are L coupon-bearing hedging instruments, we use expression (8) to calculate the generalized duration vectors $(\tau b)_*$ and $(\tau b)_\ell$, $\ell = 1,\ldots,L$ of both the target and each of these instrument streams. The main difference compared to hedging with zero-coupon instruments is that the preceding yield factor analysis no longer is carried out on the instruments themselves. Instead, it is still applied to a balanced panel data set of yields not precisely corresponding to the set of available hedging instruments that may vary from period to period as bonds age, mature, etc.. Thus, b_h in (8) is interpolated based on output from the zero-coupon yield factor analysis, applied to data from the preceding period. Next, the construction of hedging weights on the coupon-bearing instruments may proceed in analogy with the case of zero-coupon hedging instruments. Thus, form the $L \times k$ matrix of generalized durations $\mathcal{B}$, with typical row $(\tau b)_\ell'$. This is the matrix that specializes to $\mathcal{T}B$ in the zero-coupon instrument case. The hedge portfolio for general instruments is then given by

$$\widetilde{w} = \Upsilon^{-1}\mathcal{B}(\mathcal{B}'\Upsilon^{-1}\mathcal{B})^{-1}(\tau b)_*. \tag{B.26}$$

Here, the $L \times L$ matrix Υ captures the idiosyncratic error variance of the coupon-bearing bond returns, which may be constructed by interpolation and weighted summation across coupons of elements of Ψ from the yield factor analysis. Again, (11) from the Theorem is used to obtain value matching (full investment), but now with $\Lambda = \Upsilon^{-1} - \Upsilon^{-1}\mathcal{B}(\mathcal{B}'\Upsilon^{-1}\mathcal{B})^{-1}\mathcal{B}'\Upsilon^{-1}$.

C. Classical maximum likelihood factor analysis

In the generalized duration matching procedure, we conduct a classical maximum likelihood factor analysis of the yield factor model (1). We assume $\text{var}(f_t) = I_k$, which is without loss of generality, since factor variances and covariances could be absorbed in the loadings B. For identification, the $k \times k$ matrix $B'\Psi^{-1}B$ is restricted to be diagonal. The reported rotation is that where the j'th factor explains the j'th most of the variation. A full period estimation with $k = 3$ factors produces the estimated loadings shown in Figure 3. As evidenced by the graphical depiction in the figure of each of the three columns B_j as a function of maturity, we recover the standard finding of flat, steep, and hump-shaped loading patterns, corresponding to the level, slope, and curvature factor structure also highlighted by Litterman and Scheinkman (1991) based on similar plots. To interpret the individual loadings b_{ij}, $i = 1,\ldots,m$, $j = 1,\ldots,k$, it ought to be noted that b_{ij}^2/σ_i^2 is the proportion of variance of yield y_{τ_i} explained by the j'th factor, with σ_i^2 the total variance from Table 1 for the i'th maturity. Taking averages across i, the level, slope, and curvature factors explain 97.83%, 2.02%, and 0.01%, respectively, of the variation in the yield data in the full period. Although these figures may make the curvature factor appear redundant, this is not so if judged by the log-likelihood values or standard information criteria of, say, one-, two- and three-factor models. The remaining 0.07% of variation not explained by the three factors is attributed to the idiosyncratic shocks ε_t in (1). The corresponding estimated variances Ψ are shown in the first line of Table 5, along with the maximized log-likelihood value and number of parameters. To avoid Heywood cases (the factors explaining more than the total variation for a given maturity, i.e., communality $\sum_{j=1}^{k} b_{ij}^2/\sigma_i^2$ exceeding unity), a lower bound of 10^{-4} is imposed on the unique variances, ψ_i/σ_i^2, for each maturity. All subsequent estimations apply this bound, too, thus avoiding the coincidence of factors with selected key yields, and guaranteeing some minimum amount of diversification in the hedging applications. From the first line of Table 5, the shortest and longest yields have the largest unexplained variation.[15] The seven-year idiosyncratic variance hit the lower bound.

[15] Besides the parameters in the variance-covariance structure, μ in (1) is estimated by the average yields reported in Table 1.

D. Square-root Kalman filter

The models that combine a dynamic interest rate model with a consistent curve shape can be written on the state space form

$$\begin{array}{rcll} \underset{m\times 1}{y_{t_n}} & = & \underset{m\times 1}{c} + \underset{m\times k}{B}\,\underset{k\times 1}{f_n} + \underset{m\times 1}{\varepsilon_n}, & \varepsilon_n \sim N(0,\Psi), \\ f_n & = & \underset{k\times 1}{\Phi_0} + \underset{k\times k}{\Phi_1} f_{n-1} + \underset{k\times 1}{v_n}, & v_n \sim N(0,\Omega). \end{array}$$

The observed yield data is $\left(y_{t_1},\ldots,y_{t_N}\right)$ and we write $Y_n = \left(y_{t_1},\ldots,y_{t_n}\right)$ for observations up to time t_n. Denote the factor estimate at t_n by $\mu_{n|n} = \mathbb{E}(f_n|Y_n)$ and the factor prediction by $\mu_{n+1|n} = \mathbb{E}(f_{n+1}|Y_n)$. The corresponding factor variance-covariance matrices are $\Sigma_{n|n} = \mathrm{var}(f_n|Y_n)$ and $\Sigma_{n+1|n} = \mathrm{var}(f_{n+1}|Y_n)$. We recall the low storage algorithm of Koopman et al. (1999). Start with an initial condition for the first factor vector given by $f_{1|0} \sim N(\mu_{1|0},\Sigma_{1|0})$, where $\mu_{1|0} = \overline{f}$ and $\Sigma_{1|0}$ solves $\Sigma_{1|0} = H(\Delta_n)\Sigma_{1|0}H(\Delta_n)' + \Omega(\Delta_n)$. The innovation in observing y_{t_n} is $\zeta_n = y_{t_n} - \left(c + B\mu_{n|n-1}\right)$ with variance-covariance matrix $\Gamma_n = B\Sigma_{n|n-1}B' + \Psi$. Then by the Kalman filter the update step is

$$\begin{aligned} \mu_{n|n} &= \mu_{n|n-1} + \Sigma_{n|n-1}B'\Gamma_n^{-1}\zeta_n, \\ \Sigma_{n|n} &= \Sigma_{n|n-1} - \Sigma_{n|n-1}B'\Gamma_n^{-1}B\Sigma_{n|n-1}, \end{aligned}$$

and the prediction step is

$$\mu_{n+1|n} = \Phi_0 + \Phi_1\mu_{n|n} = \Phi_0 + \Phi_1\mu_{n|n-1} + K_n\zeta_n, \tag{D.27}$$

$$\Sigma_{n+1|n} = \Phi_1\Sigma_{n|n}\Phi_1' + \Omega = \Phi_1\Sigma_{n|n-1}\Phi_1 + \Omega - K_n\Gamma_nK_n'. \tag{D.28}$$

Here the second equalities in both lines substitute the update step and set $K_n = \Phi_1\Sigma_{n|n-1}B'\Gamma_n^{-1}$. The contribution to log-likelihood of each new observation is

$$\log p\left(y_{t_n}|Y_{n-1}\right) = -\frac{m}{2}\log(2\pi) - \frac{1}{2}\log|\Gamma_n| - \frac{1}{2}\zeta_n'\Gamma_n^{-1}\zeta_n,$$

and the prediction-error decomposition of the log-likelihood function is therefore

$$\log L = \sum_{n=1}^{N}\log p\left(y_{t_n}|Y_{n-1}\right) = -\frac{mN}{2}\log(2\pi) - \frac{1}{2}\sum_{n=1}^{N}\left(\log|\Gamma_n| + \zeta_n'\Gamma_n^{-1}\zeta_n\right). \tag{D.29}$$

To calculate the log-likelihood function we only need the series of innovations ζ_n and Γ_n and these are calculated from $\mu_{n|n-1}$ and $\Sigma_{n|n-1}$ alone. The low storage algorithm only stores the series of ζ_n, Γ_n, and K_n and calculates $\mu_{n+1|n}$ and $\Sigma_{n+1|n}$ by (D.27) and (D.28) with $\mu_{n|n-1}$ and $\Sigma_{n|n-1}$ stored only from last period. Since this method minimizes the need for storage and avoids calculating the update step, it speeds up the filter which must be calculated many times in the optimization of (D.29) over parameters.

When searching for the parameters that maximizes (D.29), the matrix $\Sigma_{n+1|n}$ may fail to be positive semi-definite. This problem can be solved by using a square-root Kalman filter which runs the filter for $S_{n+1|n}$ with $\Sigma_{n+1|n} = S_{n+1|n}S'_{n+1|n}$ instead, as done by Carraro (1988) for the original Kalman filter. Here, we use the square-root version of the low storage algorithm. Thus we need to write the prediction step (D.28) in terms of $S_{n+1|n}$. First, write

$$\begin{aligned}
\Sigma_{n+1|n} &= \Phi_1\Sigma_{n|n-1}\Phi_1 + \Omega - K_nB\Sigma_{n|n-1}\Phi_1' \\
&= (\Phi_1 - K_nB)\Sigma_{n|n-1}\Phi_1 + \Omega \\
&= (\Phi_1 - K_nB)\Sigma_{n|n-1}(\Phi_1 - K_nB)' + (\Phi_1 - K_nB)\Sigma_{n|n-1}B'K_n' + \Omega \\
&= (\Phi_1 - K_nB)\Sigma_{n|n-1}(\Phi_1 - K_nB)' + K_n\Gamma_n^{-1}K_n' - K_n(\Gamma_n - \Psi)K_n' + \Omega \\
&= (\Phi_1 - K_nB)\Sigma_{n|n-1}(\Phi_1 - K_nB)' + K_n\Psi K_n' + \Omega.
\end{aligned}$$

Then defining $\Psi = NN'$ and $\Omega = MM'$ we can write

$$\Sigma_{n+1|n} = \left[(\Phi_1 - K_nB)S_{n|n-1}, K_nN, M\right]\begin{bmatrix} S'_{n|n-1}(\Phi_1 - K_nB)' \\ NK_n' \\ M \end{bmatrix} \equiv \widetilde{S}_{n+1|n}\widetilde{S}'_{n+1|n},$$

where $\widetilde{S}_{n+1|n}$ is a $k \times (2k+m)$ matrix. To find a $k \times k$ matrix that has the same product with its own transpose as $\widetilde{S}_{n+1|n}$ we can use the QR decomposition which writes a rectangular matrix as the product of an orthogonal matrix Q and an upper triangular matrix R. Thus calculate the QR decomposition for the transpose of $\widetilde{S}_{n+1|n}$

$$\widetilde{S}'_{n+1|n} = QR,$$

and then

$$\Sigma_{n+1|n} = \widetilde{S}_{n+1|n}\widetilde{S}'_{n+1|n} = R'QQR = R'R.$$

Therefore set $S_{n+1|n} = R'$ which is a lower triangular square matrix. Instead of $\Sigma_{n+1|n}$ the filter uses $S_{n+1|n}$ and then $\Sigma_{n+1|n}$ is positive semi-definite by construction.

E. Trading off hedging error bias and variance

The generalized duration matching approach removes factor risk from the hedging error variance, and the portfolio in Theorem 1 minimizes the remaining hedging error variance. A more general approach would be to relax the generalized duration matching constraint, thus allowing some factor variance in the hedging error variance, and striking a balance between minimizing factor variance and remaining idiosyncratic variance. When the hedge is constructed at time t, a plausible objective is the minimization of the conditional mean squared hedging error,

$$\mathbb{E}_t\left[\left(r^*_{t+1} - w' r_{t+1}\right)^2\right]. \tag{E.30}$$

This trades off hedging error bias and variance, whereas the generalized duration matching approach can be seen as minimizing hedging error variance given unbiasedness. To see this, it is shown as part of the proof of Theorem 1 in (A.2) that the hedging error is

$$r^*_{t+1} - w' r_{t+1} = \left(w'\mathcal{T}B - (\tau b)'_*\right)\Delta f_{t+1} + \left(w'\mathcal{T}\Delta\varepsilon_{t+1} + \Delta\xi^*_{t+1}\right).$$

Thus, when generalized durations are matched, $B'\mathcal{T}w = (\tau b)_*$, an unbiased hedge is achieved if $\mathbb{E}_t\left(\Delta\varepsilon_{t+1,\tau}\right) = 0$, for all τ, and $\mathbb{E}_t\left(\Delta\xi^*_{t+1}\right) = 0$, i.e., if the factor model is expected to fit the yield curve and the target at maturity of the hedge with the same errors as on the day the hedge is made.

The following theorem offers an alternative immunization strategy, based on the principle of the minimization of conditional mean squared hedging error (E.30). This approach requires a dynamic factor model, as it involves the conditional factor prediction, $\mathbb{E}_t[f_{t+1}] = \mu_{t+1|t}$, and conditional prediction error variance, $\mathrm{var}_t[f_{t+1}] = \Sigma_{t+1|t}$, to form the hedge portfolio.

Theorem 6. *The immunization portfolio $\widetilde{w}$ that minimizes the conditional mean squared hedging error among all linear portfolio rules,*

$$\min_w \mathbb{E}_t\left[\left(r^*_{t+1} - w' r_{t+1}\right)^2\right], \tag{E.31}$$

under the assumptions $\mathbb{E}_t\left(\Delta\varepsilon_{t+1,\tau}\right) = 0$ and $\mathbb{E}_t\left(\Delta\xi^*_{t+1}\right) = 0$, is given by the instruments

weights

$$\widetilde{w} = \mathcal{T}^{-1}\Psi^{-1}B\left[\left(\Sigma_{t+1|t} + \left(\mu_{t+1|t} - \mu_{t|t}\right)\left(\mu_{t+1|t} - \mu_{t|t}\right)'\right)^{-1} + B'\Psi^{-1}B\right]^{-1}(\tau b)_*. \tag{E.32}$$

Further imposing that weights sum to one, $w'\iota = 1$, adjusts the solution to

$$w_* = \widetilde{w} + \left(1 - \widetilde{w}'\iota\right)\frac{\Lambda\iota}{\iota'\Lambda\iota},$$

for

$$\Lambda = \mathbb{E}_t\left[r_{t+1}r_{t+1}'\right]^{-1} = \mathcal{T}^{-1}\left[B\left(\Sigma_{t+1|t} + \left(\mu_{t+1|t} - \mu_{t|t}\right)\left(\mu_{t+1|t} - \mu_{t|t}\right)'\right)B' + \Psi\right]^{-1}\mathcal{T}^{-1}.$$

Proof of Theorem 6. The weights that solve (E.31) are also the solution to

$$\min_w \frac{1}{2}w'\mathbb{E}_t\left[r_{t+1}r_{t+1}'\right]w - w'\mathbb{E}_t\left[r_{t+1}r_{t+1}^*\right]. \tag{E.33}$$

This is also known as the linear projection of r_{t+1}^* on r_{t+1}, and the solution is given by

$$\widetilde{w} = \mathbb{E}_t\left[r_{t+1}r_{t+1}'\right]^{-1}\mathbb{E}_t\left[r_{t+1}r_{t+1}^*\right]. \tag{E.34}$$

The conditional expected value of the instrument return vector (5) is

$$\mathbb{E}_t[r_{t+1}] = -\mathcal{T}B\left(\mu_{t+1|t} - \mu_{t|t}\right),$$

when $\mathbb{E}_t[\Delta\varepsilon_{t+1}] = 0$. Also rewriting (5) as

$$r_{t+1} = -\mathcal{T}\left(Bf_{t+1} + \varepsilon_{t+1} - y_t\right), \tag{E.35}$$

the conditional variance is given by

$$\mathrm{var}_t[r_{t+1}] = \mathcal{T}\left(B\Sigma_{t+1|t}B' + \Psi\right)\mathcal{T}.$$

Then

$$\begin{aligned}
\mathbb{E}_t\left[r_{t+1}r_{t+1}'\right] &= \mathrm{var}_t[r_{t+1}] + \mathbb{E}_t[r_{t+1}]\mathbb{E}_t[r_{t+1}]' \\
&= \mathcal{T}\left(B\Sigma_{t+1|t}B' + \Psi\right)\mathcal{T} + \mathcal{T}B\left(\mu_{t+1|t} - \mu_{t|t}\right)\left(\mu_{t+1|t} - \mu_{t|t}\right)'B'\mathcal{T} \\
&= \mathcal{T}\left[B\left(\Sigma_{t+1|t} + \left(\mu_{t+1|t} - \mu_{t|t}\right)\left(\mu_{t+1|t} - \mu_{t|t}\right)'\right)B' + \Psi\right]\mathcal{T}.
\end{aligned}$$

By (A.1) and for $\mathbb{E}_t\left[\Delta\xi^*_{t+1}\right]=0$, the conditional expected target return is

$$\mathbb{E}_t\left[r^*_{t+1}\right]=-(\tau b)'_*\left(\mu_{t+1|t}-\mu_{t|t}\right).$$

We can rewrite (A.1) as $r^*_{t+1}=-(\tau b)'_* f_{t+1}+\xi^*_{t+1}-\log p^*_t$, and then using (E.35) the conditional covariance is

$$\mathrm{cov}_t\left[r_{t+1},r^*_{t+1}\right]=\mathcal{T}B\Sigma_{t+1|t}(\tau b)_*.$$

This implies that

$$\begin{aligned}\mathbb{E}_t\left[r_{t+1}r^*_{t+1}\right] &= \mathrm{cov}_t\left[r_{t+1},r^*_{t+1}\right]+\mathbb{E}_t[r_{t+1}]\mathbb{E}_t\left[r^*_{t+1}\right]\\ &= \mathcal{T}B\Sigma_{t+1|t}(\tau b)_*+\mathcal{T}B\left(\mu_{t+1|t}-\mu_{t|t}\right)\left(\mu_{t+1|t}-\mu_{t|t}\right)'(\tau b)_*\\ &= \mathcal{T}B\left(\Sigma_{t+1|t}+\left(\mu_{t+1|t}-\mu_{t|t}\right)\left(\mu_{t+1|t}-\mu_{t|t}\right)'\right)(\tau b)_*,\end{aligned}$$

and the weights (E.34) are therefore

$$\begin{aligned}\widetilde{w} &= \mathcal{T}^{-1}\left[B\left(\Sigma_{t+1|t}+\left(\mu_{t+1|t}-\mu_{t|t}\right)\left(\mu_{t+1|t}-\mu_{t|t}\right)'\right)B'+\Psi\right]^{-1}\\ &\times\ B\left(\Sigma_{t+1|t}+\left(\mu_{t+1|t}-\mu_{t|t}\right)\left(\mu_{t+1|t}-\mu_{t|t}\right)'\right)(\tau b)_*\\ &= \mathcal{T}^{-1}\Psi^{-1}B\left[\left(\Sigma_{t+1|t}+\left(\mu_{t+1|t}-\mu_{t|t}\right)\left(\mu_{t+1|t}-\mu_{t|t}\right)'\right)^{-1}+B'\Psi^{-1}B\right]^{-1}(\tau b)_*,\end{aligned}$$

where the second equality applies the matrix inversion lemma.

Imposing that weights sum to one can be done by Lemma 5. The original problem (E.33) is unconstrained, and in the notation of (A.5) it sets $A=\mathbb{E}_t\left[r_{t+1}r'_{t+1}\right]$, $g=\mathbb{E}_t\left[r_{t+1}r^*_{t+1}\right]$, and $D=0$. The scaled weights are therefore by (A.7) given by

$$w^*=\widetilde{w}+\left(1-\widetilde{w}'\iota\right)\frac{\Lambda\iota}{\iota'\Lambda\iota}$$

with $\Lambda=A^{-1}=\mathbb{E}_t\left[r_{t+1}r'_{t+1}\right]^{-1}$. □

In the limiting case where the uncertainty about factors next period or the expected values of factors next period increase to infinity, the weights in (E.32) reduce to $\mathcal{T}^{-1}\Psi^{-1}B\left(B'\Psi^{-1}B\right)^{-1}(\tau b)_*$, this being exactly the hedge from Theorem 1. Thus, the hedge matching generalized durations and minimizing remaining variance is a special case of the general principle of minimizing conditional mean squared hedge error. It applies when the dynamic factor model is unavailable, or the portfolio manager is unwilling to fully exploit it. In these

cases the solution is to put full weight on removing the contribution of factors to hedging error by using as many degrees of freedom as there are factors to match generalized durations.

The hedging performance of the approach detailed in Theorem 6 is documented in Table 1 for the single coupon bond target and in Table 2 for the portfolio target. In each case, all the models estimated using the Kalman filter are considered, i.e., the combined models involving both curve shape and dynamics. This includes the HW-ANS model, the SLSC combined model, and the restricted affine special case of this. Line by line, performance is seen to be poorer than in the corresponding cases in Tables 3 and 4. This suggests that it is indeed important to remove all factor contribution to hedging error, which is the generalized duration approach, rather than balancing factor versus idiosyncratic variance. Still, the SLSC models dominate in terms of hedging performances relative to the competing approaches considered.

Table 1: Target is a 5-year coupon bond and hedge portfolios are constructed by the method in Theorem 6 that fully exploits the dynamic interest rate model. Otherwise the format of the table is equivalent to Table 3. Results are in basis points per month.

Hedging performance - exploiting dynamics - single bond

Model	Bias	Std. dev.	RMSE	MAE
Target Movement	53.15	128.41	138.80	110.09
HW and ANS Full period	0.70	28.35	28.32	20.25
HW and ANS Rolling 4-year	3.29	49.77	49.75	33.93
SLSC combined Full period	1.82	12.12	12.24	9.11
SLSC combined Rolling 4-year	1.74	13.08	13.18	9.89
SLSC restricted Full period	1.92	13.09	13.21	9.78
SLSC restricted Rolling 4-year	2.10	12.34	12.50	9.30

Table 2: Target is a portfolio of (2,5,10)-year coupon bonds in proportions (-1,3,-1) and hedge portfolios are constructed by the method in Theorem 6 that fully exploits the dynamic interest rate model. Otherwise the format of the table is equivalent to Table 4. Results are in basis points per month.

Hedging performance - exploiting dynamics - portfolio

Model	Bias	Std. dev.	RMSE	MAE
Target Movement	60.28	146.98	158.65	125.25
HW and ANS Full period	-2.07	47.94	47.91	35.46
HW and ANS Rolling 4-year	-1.23	69.94	69.85	49.85
SLSC combined Full period	-1.48	39.48	39.45	28.81
SLSC combined Rolling 4-year	-1.32	45.92	45.87	33.67
SLSC restricted Full period	-1.09	47.82	47.76	34.47
SLSC restricted Rolling 4-year	-2.07	49.58	49.55	35.77

F. Statistical performance evaluation

In order to evaluate the improvements in hedging performance from each model statistically, we construct a standard t-statistic along the lines of Diebold and Mariano (1995) and Giacomini and White (2006) with duration matching as fixed benchmark and each procedure as alternative. To that end, denote the loss differentials from the i'th model relative to the duration matching approach by $d_{i,t}$. The test of equal hedging performance can be conducted using

$$S = T^{1/2}\bar{d}\widehat{V}^{-1/2}, \tag{F.36}$$

where $\bar{d} = T^{-1}\sum_{t=1}^{T} d_{i,t}$ and $\hat{V}$ is an estimate of the long-run variance of the loss differentials. We employ in the test a HAC estimator and follow state-of-the art good practice by using the data-dependent bandwidth selection by Andrews (1991) based on an AR(1) approximation and a Bartlett kernel. Under regular assumptions, $S \xrightarrow{d} N(0,1)$.

We also do a Model Confidence Set (MCS) procedure (Hansen et al., 2011) to compare the performance of all the hedging procedures under consideration. For a fixed significance level, α, the procedure identifies the MCS, $\hat{M}^*_\alpha$, from the set of competing procedures, M_0, which contains the best models with $1-\alpha$ probability. The procedure is conducted recursively based on an equivalence test for any $M \subseteq M_0$ and an elimination rule, which identifies and removes a given model from M in case of rejection of the equivalence test. The equivalence test is based on pairwise comparisons using the t-statistic S_{ij} for all $i, j \in M$ and the range statistic

$$T_M = \max_{i,j \in M} \{|S_{ij}|\}, \tag{F.37}$$

where the eliminated model is identified by $\text{argmax}_{i \in M} \sup_{j \in M} \{S_{ij}\}$. Following Hansen et al. (2011), we implement the procedure using a block bootstrap and 10^4 replications. We only conduct these comparisons for the feasible investment strategies and use a significance level of $\alpha = 5\%$. For additional details, see above references.