Immunization with consistent term structure dynamics*

Daniel Borup† Bent Jesper Christensen** Jorge Wolfgang Hansen††

Abstract

We show that improved hedging of bond portfolios can be achieved by matching generalized durations that are parametrized according to a parsimonious yield curve shape which is dynamically consistent with a new term structure model with stochastic level, slope, and curvature factors. Performance deteriorates if matching basic durations, or generalized durations based on unrestricted factor models or dynamically inconsistent curve shapes. The dynamic consistency approach accommodates standard affine models as the special case in which locally deterministic factors are constant through time, corresponding to the intercept in the affine yield, but hedging performance deteriorates under this restriction.

Keywords: Hedging; generalized duration; bond portfolio; dynamic consistency; parsimonious yield curve

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

This version: January 12, 2023

*We are grateful to Martin Møller Andreasen, Christian M. Dahl, Francis X. Diebold, Damir Filipović, Esben Høg, Bob Jarrow, Frank de Jong, Andreas Schrimpf, Peter Spencer, seminar participants at University of Pennsylvania (2018) and Aarhus University (2021), and participants at the Sixth Annual Conference of the International Association for Applied Econometrics at University of Cyprus (2019), the Annual Joint European Economic Association and Econometric Society European Meetings at University of Copenhagen (2021), the 14th Annual Meeting of The Risk, Banking and Finance Society at University of Cagliari (2021), the 15th International Conference on Computational and Financial Econometrics at University of London (2021), and the Conference in Memory of Tomas Björk at the Swedish House of Finance (2022) for useful comments, and to Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation, DNRF78), the Dale T. Mortensen Centre, Aarhus University, and the Danish Social Science Research Council (grant number 2033-00137B) for research support. Support from the Danish Finance Institute (DFI) is gratefully acknowledged by Bent Jesper Christensen and Jorge Wolfgang Hansen.

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

**Corresponding author. Aarhus University, CREATES, the Dale T. Mortensen Centre, and the Danish Finance Institute. Address: Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. Email: bjchristensen@econ.au.dk, Phone: +4587165571.

††Aarhus University, CREATES, and the Danish Finance Institute. Email: jh@econ.au.dk.

1. Introduction

The management of interest rate risk is of crucial importance in the financial sector, and the success and failure of strategies have ramifications throughout the economy. Common fixed income hedging approaches are largely cross-sectional in nature, e.g., combining instruments of different maturities to match target durations, as in classical immunization. However, hedging performance depends on properties of returns, suggesting that dynamics should be accounted for. The term structure depends on multiple factors, and annihilating exposure to these simultaneously requires a generalized duration matching approach, involving the estimation of a host of parameters. Performance can potentially be enhanced by exploiting parsimony, simplifying the specification of loadings by relying on the level, slope, and curvature structure of yield curves noted by Litterman and Scheinkman (1991). However, such reduced parametrization of curve shapes is at risk of being at odds with dynamic term structure theory, according to which the yield curve simply cannot take on the restricted shape period after period. This calls the estimation of parsimoniously specified loadings and the resulting trading strategies into question.

In this paper, we embed the cross-sectional portfolio construction within a dynamically consistent modeling framework, and show that this leads to improved hedging performance, based on weekly yield data from the Federal Reserve (FED) over the period 1983 through 2019 for model estimation, CRSP data for returns to the hedging target, and a monthly rebalancing horizon. We build on the concept from Björk and Christensen (1999) of dynamic consistency between the shape of the yield curve and the stochastic process or dynamic term structure model (DTSM) driving it. According to this notion, a curve shape (a class of curves) is dynamically consistent with a DTSM if future yield curves belong to the class, given that the current curve does, and that dynamics are governed by the DTSM. In contrast, if a curve shape and a DTSM are dynamically inconsistent, then the dynamics will instantaneously drive the yield curve away from the class, even if it is currently of the given shape. The importance of dynamic consistency in hedging applications is that without it, cross-sectional portfolio construction relies in part on

information that is not relevant for future yield curves, and hence returns. Removal of such extraneous information facilitates parsimony and leads to forward-looking models that are more likely to apply out-of-sample.

The conditions for dynamic consistency between a curve shape and a DTSM are that the loadings from the curve shape, viewed as functions of maturity, span the yield drifts and volatilities from the DTSM. These conditions are distinct from the requirement of absence of arbitrage opportunities, which may therefore be imposed as an additional condition. The joint hypothesis of dynamic consistency and no arbitrage is equivalent to the condition that the loadings, besides yield volatilities, also span convexity and slope adjustments, i.e., average slope (yield spread, or carry) and local slope (or roll-down). In this case, the spanning coefficients in the yield volatility condition are given by the volatilities of the state variables, and those in the condition on slope adjustments and convexity by the state variable drifts. Inserting the first condition in the second produces a condition resembling the fundamental term structure partial differential equation (PDE), when viewed as an equation with the curve shape (loadings) as unknown, for given state drifts and volatilities. Our viewpoint is dual to this, i.e., for given curve shape, such as level, slope, and curvature, we look for spanning coefficients, taking instead the state drifts and volatilities as the unknowns when seeking a dynamically consistent DTSM. From this viewpoint, the relevant condition is an ordinary equation in spanning coefficients, rather than a PDE.

A prominent theory based on the PDE viewpoint is that of Duffie and Kan (1996) on affine term structure models (ATSMs). Here, by suitable affine specifications of the assumed state drifts and volatilities (local variances), the relevant PDE is reduced to an ordinary differential equation (ODE). Taking the short rate to be affine in the stochastic state variables generates an initial condition that leaves all yields affine throughout. In contrast, in the dynamic consistency approach, there is no requirement that the yield curve satisfies the given curve shape at all times. Once it does, it will continue to do so, i.e., the class of curves of the given shape is absorbing. Some of the state variables can be deterministic, although time-varying. If they reach their long-run levels, they

remain there, and the yield curve assumes standard affine shape, depending on calendar time only through the stochastic state variables. For other values of the deterministic state variables, their associated loading functions enter the yield curve with time-varying coefficients, hence representing a generalization of the maturity-specific intercept in the ATSMs. Furthermore, the dynamic consistency approach applies more generally, without affine restrictions on state drifts and volatilities, outside the factor model case, and for specifications not imposing the absence of arbitrage opportunities. We highlight all these cases, and show the incremental economic value of the dynamic consistency approach, without imposing the restrictions to the standard affine case.

For the hedging analysis, our starting point is a factor model for yields, governing both the hedging instruments, which we take to be a set of zero-coupon bonds of different maturities, and the target to be hedged. We derive the optimal portfolio that minimizes conditional hedging return error variance under generalized duration matching. We investigate the possibility that performance can be enhanced by adopting a parsimoniously parametrized curve shape that is dynamically consistent with a suitable arbitrage-free DTSM, i.e., with loadings spanning convexity and slope adjustments, as well as yield volatilities. Convexity can present a problem, because it involves maturity times a quadratic in volatility, and therefore in loadings. Spanning this through loadings can require augmenting the set of loading functions, and hence factors. Since volatility is already spanned, convexity is unaltered by such augmentation. However, slope adjustments change, and dynamic consistency requires that the new adjustments are spanned by the augmented loadings.

We exploit dynamic consistency by imposing increasing structure in the estimation of loadings over three stages. In the first, loadings are augmented as indicated, to span volatility and convexity corresponding to a suitable DTSM, as well as slope adjustments, and the factor model is estimated with the augmented loadings. This way, the dynamics are indirectly brought to bear on the estimation, via the shape of the loading functions. In the second stage, we model the dynamics directly. Based on the DTSM identified in the first stage, we specify a reduced factor model for slope-adjusted yield changes

(fixed-maturity yield changes less slope adjustments). As excess returns are negative slope-adjusted yield changes times maturities, modeling the latter brings in the dynamics directly, and allows testing the no-arbitrage condition, following Christensen and van der Wel (2019).[1] The reduction in the factor model stems from the fact that not all factors in the dynamically consistent DTSM need be stochastic. Since the yield curve may not yet have assumed the dynamically consistent shape, neither the current yield curve nor the slope adjustments are restricted by this shape in the second-stage approach. Instead, in the spirit of the general Heath, Jarrow, and Morton (1992) (henceforth HJM) approach of conditioning on an arbitrary initial (current) yield curve, the latter is taken directly from the data when computing the slope adjustments. In the third stage, if the current yield curve indeed assumes the dynamically consistent shape, then slope adjustments take on suitably restricted forms, too, and information is potentially lost by ignoring this. Jointly imposing the restrictions from the DTSM on the dynamics and from the dynamically consistent augmented curve shape on the current yield curve and the cross section of slope adjustments leads to a filtering approach based on consecutive yield curves along the dynamically consistent curve family.

Our reasons for requiring the absence of arbitrage opportunities, alongside dynamic consistency, are that it serves as a criterion for correct model specification and provides additional parsimony, as a limited number of market prices of risk are estimated, in place of unrestricted mean parameters. While Joslin, Singleton, and Zhu (2011) find that forecasts of factors are invariant to no-arbitrage restrictions in Gaussian DTSMs, our hedged positions do not depend on factor forecasts, as generalized duration matching removes factor exposure, so the gains in performance stem from reductions in model and parameter uncertainty.

The popular Nelson and Siegel (1987) (henceforth NS) curve shape is used extensively in the bond yield literature, and can be motivated by its level, slope, and curvature features. We consider this as our leading case of a curve shape imposed for parsimony. The NS factor loadings involve but a single parameter, which enters nonlinearly. We provide an

[1] Goliński and Spencer (2017) similarly consider excess returns rather than yields for model estimation.

example of a DTSM that is dynamically consistent with the NS curve shape and features time variation in both the three (linear) factors and the additional nonlinear parameter. However, we also show that the NS curve shape is dynamically inconsistent with all non-degenerate arbitrage-free DTSMs. Therefore, we augment the NS curve shape and show that the resulting augmented NS (henceforth ANS) curve shape is dynamically consistent with a certain arbitrage-free DTSM, which we label the ANS-extended Vasicek (1977) model. The dynamically consistent DTSM involves both a stochastic factor, associated with slope, and two locally deterministic factors (or state variables). When freezing the latter at their long-run levels, the DTSM reduces to standard affine (ATSM) form. We investigate the value of ANS in hedging, using the three-stage approach.

The ANS-extended Vasicek model involves but a single driving Wiener process. As it is widely believed that at least three stochastic factors govern term structure movements, we introduce a DTSM with three driving processes. Motivated by the observed level, slope, and curvature structure of yield curves, we set the volatility functions to be proportional to the three NS loading functions. We label this the stochastic level, slope, and curvature or SLSC model, and show that it is dynamically consistent with a curve shape involving seven loading functions and seven factors, of which four are locally deterministic. Again, when freezing the latter at their long-run levels, a standard ATSM emerges, in this case corresponding to the so-called arbitrage-free NS or AFNS model considered by Christensen, Diebold, and Rudebusch (2011) and Krippner (2015). This is a three-factor affine model, an $A_0(3)$ model in the Dai and Singleton (2000) classification.

In our empirical work, the target for assessing immunization performance is taken to be a portfolio consisting of two-year, five-year, and ten-year coupon bonds with positive and negative weights, based on CRSP data. Portfolios of zero-coupon bonds are constructed on a monthly basis to minimize conditional hedging error variance while matching generalized durations based on estimated loadings and idiosyncratic error variances, and the resulting ability to hedge one-month target returns is recorded. Estimation is performed based on weekly FED yield data using each of the methods outlined, i.e., unrestricted factor models for yields, parsimoniously restricted versions, augmented

versions requiring dynamic consistency with suitable DTSMs, reduced factor models for slope-adjusted yield changes based on the same DTSMs, and state space models fully exploiting both the relevant DTSMs and the dynamically consistent curve shapes (loadings). We present results both for full-period estimation and for rolling estimation using a four-year window of observations immediately preceding formation of the hedge. The rolling estimation mimics a feasible out-of-sample strategy.

As a simple benchmark, we consider traditional immunization by duration matching. All other methods involve estimation of loadings and idiosyncratic variances. We find that generalized duration matching based on an unrestricted three-factor model for yields offers only a modest gain over basic duration matching, in terms of root mean squared error (RMSE). Further, for a parsimoniously restricted version based on the NS curve shape, hedging performance deteriorates. Given that the NS curve shape is dynamically inconsistent with all non-degenerate arbitrage-free DTSMs, this suggests that the quest for gains from parsimony should proceed on a principled basis, exploiting dynamic consistency.

For our first-stage approach to dynamic consistency, imposing the parsimonious ANS curve shape leads to a gain in hedging performance relative to both NS and the corresponding unrestricted factor model. These results are broadly supportive of the importance of dynamic consistency. In the second stage, using a factor model for slope-adjusted yield changes, as opposed to yields, performance deteriorates when based on an unrestricted model, but improves when imposing the ANS-extended Vasicek model, hence reinforcing the importance of the dynamics. In the third stage, performance deteriorates, relative to the second-stage results. One possibility is that the DTSM (ANS-extended Vasicek) is correctly specified, but that the yield curve has not yet reached the dynamically consistent shape (ANS). In this case, the HJM approach of conditioning on an arbitrary initial (current) curve proves its value, with better results in the second stage than in the third. Another possibility is that the curve shape and DTSM are too restrictive, hence calling for a more flexible curve shape, dynamically consistent with a more general DTSM. Evidence pointing to the latter possibility is that the no-arbitrage condition is rejected

in the ANS-extended Vasicek model, which is particularly damaging, because it is the arbitrage-free version of this model that is dynamically consistent with ANS. Thus, for a more general specification, we consider the new SLSC model.

In the feasible rolling estimation case, hedging performance based on the SLSC model is stronger than that based on the other approaches considered, and the no-arbitrage condition is not rejected. Constructing standard pairwise model comparison t-statistics, following Diebold and Mariano (1995) and Giacomini and White (2006), we find that the improvement compared to our benchmark is significant. Moreover, using the Model Confidence Set (MCS) of Hansen, Lunde, and Nason (2011) to compare performance across all approaches considered, we find that the MCS includes both the second and third stage SLSC approaches, along with the second stage ANS-extended Vasicek approach. For the SLSC approaches, performance in the third stage is at least as strong as in the second. The results suggest that the more flexible SLSC curve shape better accommodates the current yield curve, and hence the cross section of slope adjustments, compared to ANS, and that the SLSC model with three stochastic factors generates value, relative to the ANS-extended Vasicek model with but one. Performance deteriorates when freezing the locally deterministic state variables at their long-run levels, thus reducing the model to the standard affine case (AFNS). Finally, instead of removing factor risk exposure completely from the hedged position, we consider an alternative approach that relaxes the generalized duration matching constraint and targets RMSE directly, thus trading off conditional expected hedging error (bias) against conditional variance, rather than minimizing the latter. However, the evidence favors the generalized duration approach. Parsimony is again the likely reason. The generalized duration matching strategies involve only estimated loadings and idiosyncratic variances, whereas those targeting RMSE involve all model parameters, including state transition coefficients, thus adding to estimation uncertainty.

Overall, the empirical results show that generalized duration matching by itself does not suffice for improving hedging performance, relative to traditional immunization. Neither does the combination with a flexible, parsimonious curve shape, motivated by prior

knowledge about relevant level, slope, and curvature yield curve shapes, and imposed on factor loadings in the estimation of generalized durations. Instead, performance is improved by requiring that the curve shape imposed on loadings be dynamically consistent with a suitable DTSM. The latter can involve further parsimony and lead to improved performance by allowing for arbitrary initial curve shape, in agreement with the general HJM approach. In our analysis, adopting a sufficiently rich DTSM such that the dynamically consistent curve shape is flexible enough to capture the current yield curve and imposing this shape throughout generates at least as strong performance. The results suggest that the shape of the yield curve is dynamically consistent with the stochastic process driving it, but is not at the long-run equilibrium, where it reduces to standard affine form. This indicates that the loadings associated with deterministic state variables do not enter yield curves in fixed proportions over time, and this insight can be exploited for immunization purposes.

Our work relates to a long tradition in finance. Redington (1952) introduced the traditional technique of matching the basic bond duration measure of Macaulay (1938) across assets and liabilities and coined the term immunization for this operation. This is an entirely cross-sectional and essentially single-factor strategy, with duration capturing the return sensitivity with respect to the target yield. Fisher and Weil (1971) documented the relatively strong empirical performance of this approach. Nelson and Schaefer (1983) acknowledge the possibility of a multivariate factor structure underlying market yields. Return sensitivities with respect to individual factors are calculated as negative yield sensitivities times maturity. The relevant immunization strategy matches the return sensitivity with respect to each factor of the hedging portfolio to that of the target and is fully invested, i.e., a value matching condition. This is achieved by considering $k+1$ hedging instruments in case of k factors, so there is no optimization. With more than $k+1$ instruments, Ingersoll (1983) seeks diversification by minimizing the sum of squared hedging weights, subject to the return sensitivity and value matching constraints. This approach only minimizes conditional hedging error variance if idiosyncratic return errors are homoskedastic, corresponding to idiosyncratic yield variances declining quadratically

in maturity, whereas we allow for an arbitrary term structure of idiosyncratic variances.

Litterman and Scheinkman (1991) is a leading example of a cross-sectional approach using the classical statistical factor analysis as the first stage in term structure hedging. They find that three factors, labeled level, steepness (or slope), and curvature, adequately describe the term structure. NS interpret their parsimoniously parametrized yield curve shapes with monotonic and hump components in terms of short-, medium-, and long-term factors, corresponding to the Litterman and Scheinkman (1991) slope, curvature, and level factors, respectively. This motivates imposing the NS curve shape on loadings in the factor analysis, for a reduction in complexity through savings in degrees of freedom, and potentially improved hedging performance. Willner (1996) calculates the functional form of the return sensitivities corresponding to the NS curve shape, labeling them level, slope, and curvature durations. Diebold, Ji, and Li (2006) refer to these as generalized durations and use them in an empirical hedging application based on the Ingersoll (1983) approach. Other applications of the Ingersoll (1983) approach include Chambers, Carleton, and McEnally (1988), Nawalkha, Soto, and Zhang (2003), Soto (2004), and Bravo and Silva (2006), who consider different restrictions on loadings.[2] Carcano and Dall'O (2011) extend the approach to allow for model error.

Some studies on hedging in other markets have instead used the full variance-covariance matrix. Campbell, Serfaty-De Medeiros, and Viceira (2010) and Opie and Riddiough (2020) use foreign currencies to hedge the exchange rate exposure in a given portfolio of stocks and bonds. However, adoption of a common factor structure is natural in fixed income markets. Some studies rely on closely related alternative specifications. Agca (2005) calculates sample standard deviations at different maturities and fits one-factor HJM volatility functions to these using cross-sectional regression. Galluccio and Roncoroni (2006) advocate targeting cross-shape risk, or factor volatilities, rather than cross-yield risk, or HJM volatilities, in hedging. Still, all these studies are cross-sectional

[2]While Macaulay (1938) duration is a weighted sum of raw terms to payments, the Chambers, Carleton, and McEnally (1988) loading functions involve powers, corresponding to polynomial loadings for zero coupon bonds. Nawalkha, Soto, and Zhang (2003) use powers in the differences between terms to payments and the planning horizon (term to target payment to be immunized). The NS loadings are exponential-polynomial functions.

in nature. The implications of dynamics for the curve shape are not exploited.[3]

The concept of dynamic consistency was introduced by Björk and Christensen (1999), and studied further by Filipović (1999), but has never been exploited in hedging. The analysis in Björk and Christensen (1999) was cast in terms of forward rates, and relied on the Stratonovich rather than the more familiar Itô stochastic calculus. We conduct our analysis at the level of yields to maturity because this allows easier interpretation and corresponds to how market prices are quoted, and we rely exclusively on the Itô calculus. We introduce and explain the dynamic consistency concept in detail, and provide the first complete proof from first principles of the fundamental result that the NS curve shape is dynamically inconsistent with all non-degenerate arbitrage-free DTSMs. The result was alluded to in Björk and Christensen (1999) and Filipović (1999), but never proved in detail.[4] The idea of incorporating a suitably augmented NS curve shape within an arbitrage-free DTSM is pursued by Christensen, Diebold, and Rudebusch (2011) in their AFNS model, and a related five-factor model is used by Quaedvlieg and Schotman (2020) in a hedging application, still imposing standard affine form. The dynamic consistency approach accommodates the generalization of the maturity-dependent intercept to a time-varying mixture of the four linearly independent loading functions associated with deterministic state variables in the SLSC model. We show empirically that this greatly enhances hedging performance.

The paper proceeds as follows. The basic hedging framework is presented in Section 2, and the optimal generalized duration matching portfolio is derived. Section 3 considers parsimonious curve shapes and dynamic consistency. The empirical strategy is introduced in Section 4. Section 5 presents the data, and Section 6 the empirical results. Section 7 concludes. The Appendix contains all proofs, as well as further details on implementation and additional empirical results.

[3] Díaz et al. (2009) consider a stop loss strategy, replacing passive generalized duration matching by active management for as long as interest rate forecasts are successful.

[4] The same is true for the textbook treatments, Christensen and Kiefer (2009), Filipović (2009), and Diebold and Rudebusch (2013).

2. The Hedging Framework

We consider the problem of hedging the return to some target asset, using m zero-coupon bonds as the available hedging instruments.[5] Writing r^*_{t+1} for the next period target return, $r_{t+1} = (r_{t+1,\tau_1}, \ldots, r_{t+1,\tau_m})'$ for the m-vector of returns to the zero-coupon bonds, with terms to maturity $\tau_1 < \ldots < \tau_m$, and w for the m-vector of hedging portfolio weights to be chosen at time t, the problem is of the type

$$\min_w \text{var}_t \left(r^*_{t+1} - w' r_{t+1} \right), \tag{1}$$

minimization of conditional hedging error variance as of time t. We focus on the case that the target is a bond portfolio, so (1) is a version of the classical immunization problem.

To relate returns to yields, write zero-coupon prices as $p_{t,\tau} = \exp(-\tau y_{t,\tau})$, with $y_{t,\tau}$ the continuously compounded yield at time t, for term to maturity τ. The log return $r_{t+1,\tau} = \log p_{t+1,\tau} - \log p_{t,\tau+1}$ is then given by

$$r_{t+1,\tau} = -\tau \Delta y_{t+1,\tau}, \tag{2}$$

with $\Delta y_{t+1,\tau} = y_{t+1,\tau} - y_{t,\tau}$ the constant maturity yield change, for purposes of conditional variance minimization, cf. (1). More precisely, the excess return above the risk-free rate (short yield) $y_{t,1}$ is

$$r_{t+1,\tau} - y_{t,1} = -\tau \tilde{y}_{t+1,\tau}, \tag{3}$$

with $\tilde{y}_{t+1,\tau}$ the slope-adjusted yield change,

$$\tilde{y}_{t+1,\tau} = \Delta y_{t+1,\tau} - \frac{y_{t,\tau+1} - y_{t,1}}{\tau} - \frac{y_{t,\tau+1} - y_{t,\tau}}{(\tau+1) - \tau}, \tag{4}$$

namely, the raw yield change adjusted for average slope (yield spread, or carry), as well as local slope (or roll-down) at τ.[6] From (3)-(4), as the short rate and both slope adjustments are read off the yield curve at t, only terms in the information set and thus

[5] The extension to coupon-bearing instruments is considered in Appendix A.3.

[6] Christensen and van der Wel (2019) derive (3)-(4) as an approximate excess return relation. Proof of the exact relation is given in Appendix A.1. The denominator in the local slope (last term in (4)) can differ from unity if time steps in calendar and maturity dimensions differ.

of zero conditional variance are suppressed in (2), which therefore suffices for conditional variance minimization.

Traditional immunization amounts to combining the hedging instruments in a portfolio with weighted average duration matching that of the target. Because the duration of a zero-coupon bond is maturity, application of (2) to $w'r_{t+1}$ shows that duration matching neutralizes returns if yields are common across maturities. As they are not, we consider a generalized duration matching strategy that minimizes residual or idiosyncratic risk (1) after removing exposure to common term structure factors by matching suitably defined generalized durations. To this end, we consider a factor model for yields.

Writing $y_t = (y_{t,\tau_1},\ldots,y_{t,\tau_m})'$ for the yields at time t to the zero-coupon bonds, the classical factor analysis structure is

$$y_t = \mu + Bf_t + \varepsilon_t\,, \tag{5}$$

with μ the m-vector of mean yields, f_t a k-vector of common, covariance-generating factors, $k < m$, with $k \times k$ variance-covariance matrix $\mathrm{var}(f_t) = \Sigma$, B an $m \times k$ matrix of factor loadings of rank k, and ε_t an m-vector of error terms, assumed independent of f_t and idiosyncratic, i.e., $\Psi = \mathrm{var}(\varepsilon_t)$ is diagonal.

Combining (2) and (5), returns are given by

$$r_{t+1} = -\mathcal{T}(B\Delta f_{t+1} + \Delta\varepsilon_{t+1})\,, \tag{6}$$

with $\mathcal{T} = \mathrm{diag}(\tau_1,\ldots,\tau_m)$ the $m \times m$ diagonal matrix with the maturities along the diagonal. Thus, the $m \times k$ matrix $\mathcal{T}B$, maturities times loadings, represents the return sensitivities or generalized durations of instruments with respect to factors. In particular, the generalized durations of the i^{th} zero-coupon bond are given by $\tau_i b_i$, maturity times b_i, the i^{th} row of B. The return to the hedge portfolio is then

$$w'r_{t+1} = -w'\mathcal{T}(B\Delta f_{t+1} + \Delta\varepsilon_{t+1})\,, \tag{7}$$

i.e., $w'\mathcal{T}B$ is the $1 \times k$ row vector of generalized durations for the portfolio.

The target return, r^*_{t+1}, is assumed to obey the factor model, too, with generalized durations given by the $1 \times k$ vector $(\tau b)_*$. Analogously to (6), $r^*_{t+1} = -(\tau b)_*\,\Delta f_{t+1} + \Delta\varepsilon^*_{t+1}$,

where $\Delta\varepsilon^*_{t+1}$ is an idiosyncratic target return error, with variance Ψ^*, a scalar, and uncorrelated with Δf_{t+1} and $\Delta\varepsilon_{t+1}$. Combining with (7), the hedging error is

$$r^*_{t+1} - w'r_{t+1} = \left(w'\mathcal{T}B - (\tau b)_*\right)\Delta f_{t+1} + \Delta\varepsilon^*_{t+1} + w'\mathcal{T}\Delta\varepsilon_{t+1}\,. \tag{8}$$

Generalized duration matching amounts to imposing $w'\mathcal{T}B = (\tau b)_*$ on the portfolio selection problem (1), thus removing all factor exposure from the hedged position. In this case, by (8), the conditional hedging error variance is

$$\mathrm{var}_t\left(r^*_{t+1} - w'r_{t+1}\right) = \Psi^* + w'\mathcal{T}\Psi\mathcal{T}w\,. \tag{9}$$

In a complete market with only factor risk, generalized duration matching would identify a perfect hedge. In practice, in the incomplete market case, (9) applies. Since the first term on the right side is outside the portfolio manager's control, we consider minimization of the second term. The approach of Ingersoll (1983) is instead to minimize $w'w$. Evidently, this approach is only variance minimizing if idiosyncratic yield variances Ψ are declining quadratically in maturities $\mathcal{T}$, whereas our approach applies for arbitrary $\Psi > 0$.

The following theorem provides the optimal portfolio, subject to generalized duration matching, both with and without the additional value matching (or full investment) constraint that weights sum to one, $w'\iota = 1$, with $\iota = (1,\ldots,1)'$.[7]

Theorem 1. *The immunization portfolio $\widetilde{w}$ that minimizes conditional hedging error variance subject to generalized duration matching,*

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

is given by

$$\widetilde{w} = \mathcal{T}^{-1}\Psi^{-1}B\left(B'\Psi^{-1}B\right)^{-1}(\tau b)'_*\,. \tag{11}$$

The immunization portfolio w^* that minimizes conditional hedging error variance subject to both generalized duration and value matching,

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

[7] All proofs are in Appendix A.

is given by

$$w^* = \widetilde{w} + \left(1 - \widetilde{w}'\iota\right) \frac{\Lambda\iota}{\iota'\Lambda\iota}, \tag{12}$$

with $\widetilde{w}$ *from* (11)*, and*

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

Since generalized duration matching removes all factor exposure, the weights in Theorem 1 do not depend on Σ.[8] Value matching requires adjustment of the portfolio $\widetilde{w}$ from (11), and hence (12). Simple scaling by $(\widetilde{w}'\iota)^{-1}$ would violate generalized duration matching, thus reintroducing factor exposure. Further, by the Theorem, the hedging portfolio only depends on the target through the generalized target durations $(\tau b)_*$. If the target takes the form of a single, known payment τ_* periods hence, then $(\tau b)_* = \tau_* b_*$, with b_* the $1 \times k$ vector of factor loadings of the yield to the hypothetical τ_*-period bond. More generally, if the target is a certain payment stream, then each payment has generalized durations of this form, and the stream has generalized durations given by the value-weighted average of these. Theorem 1 applies to such streams, too. In the applications, we fit the factor model only to the hedging instruments, then interpolate generalized durations of the target, thus accommodating situations without time series observations on target returns, e.g., company liability streams.[9]

3. Parsimonious Curve Shape and Dynamic Consistency

The unrestricted factor model (5) involves the estimation of a host of parameters, thus opening the door to possibly large estimation error. Writing $\text{var}(y_t) = \Upsilon = B\Sigma B' + \Psi$, the classical factor analysis takes $\Sigma = I_k$ and $B'\Psi^{-1}B$ diagonal for identification, so there are $mk - k(k-1)/2$ parameters in B, and m in Ψ. In a typical yield panel covering $m = 8$ maturities as in our application and with $k = 3$ factors, this amounts to 29 parameters in Υ, against $m(m+1)/2 = 36$ in the unrestricted variance-covariance matrix, i.e., not a great

[8] Replacing B, b by $B\Sigma^{1/2}$, $b\Sigma^{1/2}$ leaves (11)-(12) unaltered as $\Sigma^{1/2}$ cancels.

[9] Appendix A.3 provides details on interpolation and hedging in case the target and/or the instruments are general streams, e.g., coupon bonds.

reduction. We propose a number of increasingly structured approaches to incorporating restrictions from the shape of the yield curve on the loadings B.

3.1. Flexible parsimonious forms

In (5), the j^{th} column of B holds the sensitivities of yields to the j^{th} factor. It can be natural to model these by a known function of maturity and a small number of parameters, say, $\tau \mapsto B_j(\tau;\gamma)$, where $B_j(\,\cdot\,;\gamma)$ takes a simple form, and γ is estimated. In this case, the parsimoniously parametrized shape of the yield curve $\tau \mapsto y(t,\tau)$ at t is represented as

$$y(t,\tau) = \sum_{j=1}^{k} B_j(\tau;\gamma) f_{t,j} = B(\tau;\gamma) f_t\,, \tag{13}$$

where the functions $B_j(\,\cdot\,;\gamma)$ are linearly independent, and $B(\tau;\gamma)$ is $1\times k$. Imposing this structure on (5), we have $y_{t,\tau_i} = y(t,\tau_i)$, $B_{ij} = B_j(\tau_i;\gamma)$, and $\varepsilon_{t,i}$ is the measurement error. Consider, for example, the NS curve shape given by

$$y(t,\tau) = f_{t,1} + \left(\frac{1-e^{-a\tau}}{a\tau}\right) f_{t,2} + \left(\frac{1-e^{-a\tau}}{a\tau} - e^{-a\tau}\right) f_{t,3}\,, \tag{14}$$

for some parameter a. Thus, $k=3$, and in the representation (13), the loading functions are

$$\tilde{B}_{1:3}(\tau;a) = \begin{pmatrix} \tilde{B}_1 & \tilde{B}_2 & \tilde{B}_3 \end{pmatrix} = \begin{pmatrix} 1 & \dfrac{1-\mathrm{e}^{-a\tau}}{a\tau} & \dfrac{1-\mathrm{e}^{-a\tau}}{a\tau} - \mathrm{e}^{-a\tau} \end{pmatrix} \tag{15}$$

on the level, slope, and curvature factors, respectively, using tildes to indicate the specific as opposed to generic loading functions.[10] The generalized durations obtained by multiplying maturity τ on each loading function in (15) generates the level, slope, and curvature durations proposed by Willner (1996).

With a curve shape, such as that of NS, imposed on the loadings in (5), (μ,γ,Σ,Ψ) is estimated, rather than (μ,B,Ψ). In particular, Σ must be estimated if $B(\cdot)$ determines scaling and rotation, such as in the NS case. Still, with $g = \dim\gamma$, only $g+k$ parameters are estimated in $(B(\gamma),\Sigma)$ with uncorrelated factors, and $g+k(k+1)/2$ with correlated,

[10] The NS curve is frequently seen in the forward rate representation, which has different loading functions, with clear interpretation in terms of level, slope, and curvature. We show in Appendix A.4 that the yield representation (14) is equivalent to the well known forward rate representation, and that the level, slope, and curvature interpretation carries over to the yield case.

instead of the $mk - k(k-1)/2$ free parameters in B in the classical factor analysis. In the NS case, $B(\gamma) = \tilde{B}(a)$, where $\tilde{B}(a)$ has columns determined by (15), i.e., $\gamma = a$, and $g = 1$. This implies only $4+m$ variance-covariance parameters (a, Σ, Ψ) with uncorrelated factors, and $7+m$ with correlated. For $m = 8$ maturities, this corresponds to 12 or 15 parameters, compared to 29 in the general factor model, (B, Ψ), and 36 in the unrestricted case, hence providing considerable parsimony.

3.2. Intra-period yield curve movements and dynamic consistency

An important issue when restricting the yield curve to a parsimoniously parametrized shape, which we will elaborate on in this section, is that the restricted curve shape can lead to dynamic inconsistency. Assume that intra-period movements in yields between the discrete rebalancing dates $\bar{t}$ and $\bar{t}+1$ are governed by a continuous-time DTSM. We consider a general HJM specification,

$$\mathrm{d}y(t,\tau) = \alpha(t,\tau)\,\mathrm{d}t + \sigma(t,\tau)'\,\mathrm{d}W(t), \tag{16}$$

since this is explicitly a framework for the dynamics of the entire curve $\tau \mapsto y(t,\tau)$, and thus well suited for the study of curve shapes. In (16), $\alpha(t,\tau)$ is the instantaneous drift of $y(t,\tau)$, $\sigma(t,\tau)$ a d-vector of volatilities, and $W(t)$ a d-vector of standard Wiener processes, with α, σ adapted to $\{W(t)\}_t$. For each τ, (16) specifies the stochastic differential of the real-valued process $\{y(t,\tau)\}_t$, the constant-maturity yield. Alternatively, (16) can be viewed as a single equation of infinite dimension across τ, giving the differential of the process $\{y(t,\cdot)\}_t$ on the space of curves $\tau \mapsto c(\tau)$.[11] In this framework, a particular DTSM is identified with the choice of (α, σ).

Although there is some resemblance between (16) and the yield factor model (5), the one-period factor loadings B, which by Theorem 1 are needed for hedging in discrete time, cannot be identified with the continuous-time volatilities $\sigma(\cdot)$. At each instant, increments

[11] This notion can be formalized as in Björk and Christensen (1999) and Da Prato and Zabczyk (2014) by considering the curve space as a suitable Hilbert space, e.g., the space of differentiable curves equipped with the inner product $< c_1, c_2 >= \int_0^{\bar{\tau}} c_1(\tau)c_2(\tau)\mathrm{d}\tau$, for $\bar{\tau}$ the longest maturity (10 years in our empirical work), and imposing sufficient conditions on α and σ (local boundedness and Lipschitz continuity in c) to guarantee a solution. The specific models we work out satisfy these conditions.

to the function (yield curve) $y(t,\cdot)$ are added in the directions given by the two functions $\alpha(t,\cdot)$ and $\sigma(t,\cdot)$ in (16), and B in (5) must account for the cumulated impact over discrete intervals (between $\bar{t}$ and $\bar{t}+1$) of both of these effects. As we show, this implies that there are at least as many factors in the discrete-time f_t as in the continuous-time $W(t)$, i.e., $k \geq d$, but there are typically more, $k > d$. Thus, restricting B to be dynamically consistent with a particular DTSM can save not only on parameters, but also on number of driving processes. Conversely, for a given DTSM with d-dimensional $W(t)$, it may be necessary to consider more factors for dynamically consistent discrete-time hedging, i.e., hedging uses certain extended loadings B, of dimension $m \times k$, with $k \geq d$. The exact relation between the DTSM (in particular, $\sigma(\cdot)$) and B is subtle, and is the topic of the following dynamic consistency theory.

Consider a class of potential yield curves $Y(\tau,x)$, parametrized by $x \in \mathcal{X} \subseteq \mathbb{R}^k$, a suitable parameter (or state) space, i.e., the class is $\mathcal{Y} = \{Y(\cdot,x) | x \in \mathcal{X}\}$. Let $T_{\mathcal{Y}} = \inf_t\{t : \exists x \in \mathcal{X} \text{ s.t. } y(t,\cdot) = Y(\cdot,x)\}$ be the first hitting time for $\mathcal{Y}$ under (16).

Definition 1. *(a) Dynamic consistency between a DTSM (α,σ) and a class $\mathcal{Y}$ of yield curves means that if the yield curve dynamics are governed by (16), then $y(t,\cdot) \in \mathcal{Y}$, for $t \geq T_{\mathcal{Y}}$.*

(b) Strong dynamic consistency between (α,σ) and $\mathcal{Y}$ means that if the yield curve dynamics are governed by (16), then $y(t,\tau) = Y(\tau,x(t))$, for $t \geq T_{\mathcal{Y}}$, with

$$\mathrm{d}x(t) = \phi(t)\mathrm{d}t + \psi(t)'\mathrm{d}W(t), \tag{17}$$

for suitable ϕ, ψ such that (17) has a strong solution.

Thus, $T_{\mathcal{Y}}$ represents the first time the yield curve $y(t,\cdot)$ assumes a shape within the curve family $\mathcal{Y}$. By Definition 1.(a), under dynamic consistency, the curve remains in the family, once there, if movements in interest rates are described by (16). The class $\mathcal{Y}$ is invariant under the action (16). The yield curve for $t \geq T_{\mathcal{Y}}$ is $\tau \to Y(\tau,x(t))$, for suitable $x(t) \in \mathcal{X}$, which therefore serves as the relevant state variable. This obeys dynamics of the type (17) under mild regularity conditions that we henceforth assume (full rank of $\partial Y/\partial x'$ and invertibility of $x \to Y(\cdot,x)$ on $\mathcal{Y}$ suffice). Moreover, if (17) has a solution, so does (16), so

we henceforth use dynamic consistency and strong dynamic consistency interchangeably. The initial condition for (17) is $x(T_{\mathcal{Y}}) = x_{\mathcal{Y}}$, where $x_{\mathcal{Y}}$ satisfies $y(T_{\mathcal{Y}}, \tau) = Y(\tau, x_{\mathcal{Y}})$. Further, given $y(t,\tau) = Y(\tau,x)$ in $\mathcal{Y}$, for suitable x, we assume that the volatility $\sigma(t,\tau)$ takes the form $\sigma(\tau,x)$, for $t \geq T_{\mathcal{Y}}$, depending on time t only through the state variable $x = x(t)$, and similarly for the drift, $\alpha(\tau,x)$. In this case, imposing the same (Markov) condition on the coefficients in (17), $\phi(x)$ and $\psi(x)$, is without loss of generality, hence leading to the following technical assumption.

Assumption 1. *The SDE*

$$\mathrm{d}x(t) = \phi(x(t))\mathrm{d}t + \psi(x(t))'\mathrm{d}W(t) \tag{18}$$

with initial condition $x(T_{\mathcal{Y}}) = x_{\mathcal{Y}}$ has a strong solution.

In the HJM representation, α, σ must be adapted to $\{W(t)\}_t$. Under Assumption 1, they are.

If the curve shape $\mathcal{Y}$ is linear in x, i.e., $Y(\tau,x) = B(\tau)x$, with $B(\tau)$ a $1 \times k$ vector not depending on x, then we have the case (13), a factor model, and we write $\mathcal{B}$ for $\mathcal{Y}$. In case of the NS curve shape,

$$Y(\tau,x) = x_1 + x_2\left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau}\right) + x_3\left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau} - \mathrm{e}^{-a\tau}\right), \tag{19}$$

with fixed a, we obtain a factor model, $Y(\tau,x) = \tilde{B}_{1:3}(\tau)x$, with $x = (x_1,x_2,x_3)'$, loading functions from (15), and $\partial Y/\partial x' = \tilde{B}_{1:3}(\tau)$ not depending on x. In contrast, if a is considered part of the state variable x, i.e., $x = (x_1,x_2,x_3,a)'$, then

$$\frac{\partial Y}{\partial x'}(\tau,x) = \begin{pmatrix} \tilde{B}_{1:3}(\tau;a) & \frac{\partial Y}{\partial a} \end{pmatrix}, \tag{20}$$

with

$$\frac{\partial Y}{\partial a} = -\frac{1}{a}\left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau} - \mathrm{e}^{-a\tau}\right)(x_2+x_3) + \tau\mathrm{e}^{-a\tau}x_3 = -\frac{1}{a}\tilde{B}_3(\tau,a)(x_2+x_3) + \tau\mathrm{e}^{-a\tau}x_3\,. \tag{21}$$

Since $\partial Y/\partial a$ depends on x, (19) is not a factor model for time-varying a. Thus, whether or not NS is a factor model depends on whether a is fixed or allowed to vary over time.

For smooth movements in $x(t)$, $t \geq T_{\mathcal{Y}}$, dynamic consistency between (α,σ) and the

yield curve family $\mathcal{Y}$ would be equivalent to the conditions that $\alpha(\tau,x)$ and the d coordinate functions of $\sigma(\tau,x)$ all be spanned by the parameter derivatives $\partial Y(\cdot,x_t)/\partial x$ of the yield curve function. But smooth movements do not accommodate the Wiener processes in (16) and (18). Instead, we have the following.

Proposition 1. *Dynamic consistency between the DTSM* (α,σ) *and the yield curve family* $\mathcal{Y}$ *is equivalent to the existence of suitable* ϕ, ψ *satisfying Assumption 1 and the conditions*

$$\alpha(\tau,x) = \frac{\partial Y}{\partial x'}(\tau,x)\phi(x) + \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 Y}{\partial x \partial x'}(\tau,x)\psi(x)'\psi(x)\right), \tag{22}$$

$$\sigma(\tau,x)' = \frac{\partial Y}{\partial x'}(\tau,x)\psi(x)', \tag{23}$$

for all (τ,x)*, where* $\mathrm{tr}(\cdot)$ *is the matrix trace.*

Thus, the dynamic consistency conditions involve a drift condition and a volatility condition. Inclusion of the trace term in the former circumvents a switch from the Itô to the Stratonovich stochastic calculus invoked in Björk and Christensen (1999). It enters non-trivially for nonlinear curve shapes. Corollary A.5.1 provides an example of a DTSM (α,σ) that generates NS curves with time-varying coefficient $a(t)$ in the exponent. Thus, this is not a factor model, and the trace term is non-zero. This shows that neither nonlinear dependence on state variables nor NS curve shape precludes dynamic consistency.

Dynamic consistency is a property tying the cross-sectional curve shape to the dynamics and is distinct from the absence of arbitrage opportunities. To state the dynamic consistency condition under the additional no-arbitrage condition, we show in Appendix A.5 that the relevant arbitrage restriction on yield drifts is

$$\alpha(\tau,x) = \frac{1}{\tau}\left[Y(\tau,x) - Y(0,x)\right] + \frac{\partial Y}{\partial \tau}(\tau,x) + \frac{\tau}{2}\sigma(\tau,x)'\sigma(\tau,x) + \sigma(\tau,x)'\lambda(x). \tag{24}$$

The last two terms are convexity and risk compensation, based on suitable market prices of risk $\lambda(x)$. In addition, the no-arbitrage condition involves adjustments for average slope or yield spread (first term in (24)), as well as local slope of the yield curve (second term), like the slope-adjusted yield changes (4).[12] By (24), an arbitrage-free DTSM is identified

[12] HJM consider forward rates with maturity date T, $f(t,T)$, and derive the no-arbitrage condition $\alpha_f(t,T) = \sigma_f(t,T)'\int_t^T \sigma_f(t,u)\mathrm{d}u$ under the risk-neutral measure, subscripts f indicating forward rates. In

with (λ, σ). The conditions for dynamic consistency between an arbitrage-free DTSM and a yield curve family include the original volatility condition (23) and a condition equating the right sides of (22) and (24) (see Proposition A.7.1 in Appendix A.7).

If a curve family $\mathcal{Y}$ and a DTSM are not dynamically consistent, we say that they are dynamically inconsistent. In this case, even if the yield curve at some point in time belongs to $\mathcal{Y}$, it will deviate from this shape at some later point in time, i.e., it will leave $\mathcal{Y}$ if dynamics are governed by the given DTSM. The no-arbitrage condition (24) renders NS dynamically inconsistent with all non-degenerate DTSMs, i.e., models with $\sigma(\tau, x) \neq 0$.

Corollary 1. *The NS curve shape is dynamically inconsistent with all non-degenerate arbitrage-free DTSMs, whether a is fixed or not.*

The corollary calls any procedure relying on NS into question, including using NS curve shape for parsimony in hedging. It is in conflict with yield dynamics, and hence returns.

3.2.1. Factor models

It is useful to write out the previous conditions in the special case of a factor model $\mathcal{B}$, i.e., the curve shape is $Y(\tau, x) = B(\tau)x$, for given loadings $B(\cdot)$.

Corollary 2. *Dynamic consistency between the arbitrage-free DTSM (λ, σ) and the factor model $\mathcal{B}$ is equivalent to the existence of suitable ϕ, ψ satisfying Assumption 1 and*

$$\frac{1}{\tau}\left[B(\tau) - B(0)\right]x + \frac{\partial B}{\partial \tau}(\tau)x + \frac{\tau}{2}\sigma(\tau, x)'\sigma(\tau, x) + \sigma(\tau, x)'\lambda(x) = B(\tau)\phi(x), \tag{25}$$

$$\sigma(\tau, x)' = B(\tau)\psi(x)', \tag{26}$$

for all (τ, x).

Thus, for given curve shape Bx and DTSM (λ, σ), dynamic consistency investigations involve determining coefficients (ϕ, ψ), i.e., the drift and volatility of the state process (18), solving (25)-(26). If a solution exists, then B and (λ, σ) are dynamically consistent.[13]

(24), risk compensation appears because we consider the physical measure, convexity (in place of the term involving an integral in HJM) and yield spread because we consider yields, rather than forward rates, and local slope because our fixed term to maturity analysis avoids the bond aging effect noted by Litterman and Scheinkman (1991) (details are given in Appendix A.6).

[13] The corresponding conditions for an arbitrary (not necessarily arbitrage-free) DTSM (α, σ) take the form $\alpha = B\psi$, $\sigma' = B\psi'$ (see Corollary A.7.1).

Inserting (26) into (25) produces

$$\frac{1}{\tau}\big[B(\tau)-B(0)\big]x+\frac{\partial B}{\partial \tau}(\tau)x+\frac{\tau}{2}B(\tau)\psi(x)'\psi(x)B(\tau)'=B(\tau)(\phi(x)-\psi(x)'\lambda(x)). \tag{27}$$

This resembles the fundamental term structure PDE characterizing bond prices (here, the yield curve) for given (ϕ,ψ). However, as we solve for (ϕ,ψ), our viewpoint is dual to that in the PDE-based theories that assume a particular form of the state variable process (18), thus taking (ϕ,ψ) as the starting point. A prominent theory in the latter category is that on affine term structure models (ATSMs), in which the yield typically is written in the form $Y(\tau,\bar{x}(t))=\bar{A}(\tau)+\bar{B}(\tau)\bar{x}(t)$, where $\bar{B}(\tau)$ satisfies a Riccatti ODE. This is clearly the special case of the general form $Y(\tau,x)=B(\tau)x$ in which one of the state variables is constant, i.e., $x(t)=(\bar{x}(t)',1)'$, $B(\tau)=(\bar{B}(\tau),\bar{A}(\tau))$, $k=d+1$. The dynamic consistency approach accommodates the more general structure $Y(\tau,x(t))=\bar{B}(\tau)\bar{x}(t)+\bar{\bar{B}}(\tau)\bar{\bar{x}}(t)$, say, where $\bar{\bar{x}}(t)$ is a vector of locally deterministic but potentially time-varying state variables, with loadings $\bar{\bar{B}}(\tau)$. Thus, the time-invariant ATSM intercept $\bar{A}(\tau)$ is generalized to the time-varying form $\bar{\bar{B}}(\tau)\bar{\bar{x}}(t)$ under dynamic consistency, $x=(\bar{x}',\bar{\bar{x}}')'$, $\phi=(\bar{\phi}',\bar{\bar{\phi}}')'$, say, and $\psi'=(\bar{\psi},0)'$ is $k\times d$, where $\bar{\phi}$, $\bar{\psi}$ are the $d\times 1$ state drift and $d\times d$ state volatility in the ATSM, and $k\geq d+1$ – indeed, typically $k>d+1$, as we demonstrate.[14]

If ϕ is affine, $\phi(x)=\Phi(\theta-x)$, where the constant $k\times k$ matrix Φ satisfies stationarity conditions, then θ comprises the long-run means. Further, if the lower left $(k-d)\times d$ block of Φ vanishes, then $\bar{\bar{x}}$ remains at $\bar{\bar{\theta}}$, once there, $\bar{\bar{x}}(t)=\bar{\bar{\theta}}$, where $\theta=(\bar{\theta}',\bar{\bar{\theta}}')'$, because neither the drift nor the stochastic shocks to the locally stochastic state variables move $\bar{\bar{x}}(t)$. We provide examples in which standard affine models are obtained in the special case in which (i) the dynamically consistent curve shape is attained, and (ii) the locally deterministic state variables happen to be at their long-run levels, so that the affine model intercept is realized as $\bar{A}(\tau)=\bar{\bar{B}}(\tau)\bar{\bar{\theta}}$. However, the generalization of $\bar{A}(\tau)$ to the time-varying $\bar{\bar{B}}(\tau)\bar{\bar{x}}(t)$ preserves dynamic consistency, and we show in the empirical section that allowance for this feature greatly enhances hedging performance.

In ATSM theory, the affine curve shape applies throughout. This is due to the condition

[14] See Appendix A.8 for details on the ATSM case.

that the short rate be affine, $y(t,0) = \delta_0 + \delta_1' \bar{x}(t)$, say, depending on time t only through the stochastic state variables $\bar{x}(t)$. This generates an initial condition on the Riccatti ODE, and the resulting solution for $B = (\bar{B}, \bar{A})$ leaves all yields affine in $\bar{x}(t)$. In contrast, in the dynamic consistency approach, there is no condition that any yield need satisfy the curve shape condition $y(t,\tau) = B(\tau)x(t)$ at all times.

Thus, though the equations are formally similar, they are used differently in the dynamic consistency and affine term structure theories.[15] The former is more general and applies to factor models with time-varying deterministic yield components, factor models with non-affine state variable process (ϕ, $\psi'\psi$, $\psi'\lambda$ non-affine in (25)-(26)), models outside the factor model case (Propositions 1 and A.7.1, Corollaries 1 and A.5.1), and specifications not imposing the absence of arbitrage opportunities (Proposition 1, Corollary A.7.1). Of course, a given affine model, with specified $\bar{\phi}$, $\bar{\psi}$, λ, and $B = (\bar{B}, \bar{A})$ solving the resulting Riccatti equation, is dynamically consistent with the DTSM with HJM volatility $\sigma' = \bar{B}\bar{\psi}'$ and drift (25), but more general curve shapes $B = (\bar{B}, \bar{\bar{B}})$ can be dynamically consistent with the same (λ, σ).

Consider a given volatility function, σ. By (26), dynamic consistency with the factor model $\mathcal{B}$ requires that functions of τ in $\sigma(\tau,x)'$ should be included among the curve shapes represented by $B(\tau)$. By (27), under the additional no-arbitrage condition, more functions may be needed in $B(\tau)$. The original curve shapes in $B(\tau)$ on the right side of (27) must include all the derived curve shapes arising on the left side. This is because curve shapes, beside volatility itself, must account for the convexity this gives rise to, as well as slope adjustments. To see how these requirements can lead to dynamic inconsistency, consider again the NS case, with loadings (15). Depending on ψ, convexity in (27) can involve the terms $(\tau/2)\tilde{B}_1(\tau)^2 = \tau/2$,

$$\frac{\tau}{2}\tilde{B}_2(\tau)^2 = \frac{1}{a}\left(\tilde{B}_2(\tau) - \tilde{B}_4(\tau)\right), \tag{28}$$

[15] Given $B(\tau)$, $\sigma(\tau,x)$, and $\lambda(x)$, equations (25)-(26) (and hence (27)) form an ordinary linear-quadratic system in the (unknown) constants $\phi(x)$, $\psi(x)$, while for given functions $\phi(\cdot)$, $\psi(\cdot)$, $\lambda(\cdot)$ satisfying the ATSM conditions (see Appendix A.8), (27) is a system of linear-quadratic ODEs in the (unknown) functions $B(\cdot)$.

with

$$\tilde{B}_4(\tau) = \frac{1-\mathrm{e}^{-2a\tau}}{2a\tau}, \tag{29}$$

and

$$\frac{\tau}{2}\tilde{B}_3(\tau)^2 = \frac{1}{a}\left(\tilde{B}_2(\tau) - \tilde{B}_4(\tau)\right) + \frac{\tau}{2}\mathrm{e}^{-2a\tau} + \tau\left(\tilde{B}_2(\tau) - 2\tilde{B}_4(\tau)\right) \tag{30}$$

(the detailed derivations are in (A.7.7), (A.7.8) and (A.7.11)). If volatility (26) involves the second NS loading, $\tilde{B}_2(\tau)$, then convexity involves (28), and hence (29). As a function of τ, the latter is linearly independent of the NS loadings (15), and of the slope adjustments and remaining convexity terms based on these. Ultimately, this leads to violation of (27), and hence dynamic inconsistency, cf. Corollary 1 (the proof in Appendix A.7 covers all possible combinations of NS loadings in volatility (26)).

3.2.2. Dynamic consistency

Consider again the case that volatility is proportional to the second NS loading, $\tilde{B}_2(\tau)$, i.e.,

$$\sigma(\tau, x) = \psi_2(x)\left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau}\right), \tag{31}$$

with $\psi_2(x) \neq 0$. By (28), convexity is spanned by $\tilde{B}_2(\tau)$ itself and $\tilde{B}_4(\tau)$ from (29). This suggests augmenting the set of NS loadings with $\tilde{B}_4(\tau)$, to meet condition (25). Since volatility is already spanned, cf. (31), convexity is unaltered by this augmentation. Furthermore, in this case, the augmented slope adjustments are spanned by the augmented loadings, too, hence implying dynamic consistency under the no-arbitrage condition.

Proposition 2. *The augmented NS (ANS) curve shape given by loading functions*

$$\tilde{B}_{1:4}(\tau) = \begin{pmatrix} 1 & \frac{1-\mathrm{e}^{-a\tau}}{a\tau} & \frac{1-\mathrm{e}^{-a\tau}}{a\tau} - \mathrm{e}^{-a\tau} & \frac{1-\mathrm{e}^{-2a\tau}}{2a\tau} \end{pmatrix} \tag{32}$$

with fixed a is dynamically consistent with the arbitrage-free DTSM with drift

$$\begin{aligned} \alpha(\tau, x) = &\left(\frac{\psi_2(x)^2}{a} + \lambda(x)\psi_2(x) - a(x_2 - x_3)\right)\frac{1-\mathrm{e}^{-a\tau}}{a\tau} \\ &- ax_3\left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau} - \mathrm{e}^{-a\tau}\right) - \left(\frac{\psi_2(x)^2}{a} + 2ax_4\right)\frac{1-\mathrm{e}^{-2a\tau}}{2a\tau} \end{aligned} \tag{33}$$

and volatility (31)*, provided* $\phi(x) = \Phi(\theta - x)$ *given by*

$$\phi(x) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & a & -a & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & 2a \end{pmatrix} \left(\begin{pmatrix} 0 \\ \frac{\psi_2(x)^2}{a^2} + \frac{\lambda(x)\psi_2(x)}{a} \\ 0 \\ -\frac{\psi_2(x)^2}{2a^2} \end{pmatrix} - \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \right) \tag{34}$$

and $\psi(x) = (0, \psi_2(x), 0, 0)'$ *satisfy Assumption 1.*

If $\psi_2(x)^2$ and $\lambda(x)\psi_2(x)$ are affine in x, so is $\phi(x)$ in (34). In our empirical work, we take ψ, λ constant, so θ represents the long-run means. By (33)-(34), the reduced ANS (RANS) curve shape with loadings $\tilde{B}_{2:4}(\tau)$, i.e., dropping the level $\tilde{B}_1(\tau)$, suffices for dynamic consistency.[16]

Once the curve shape $y(t,\tau) = B(\tau)x(t)$ is attained, the short rate takes the form $y(t,0) = B(0)x(t)$. By l'Hôpital's rule, $\tilde{B}_2(0) = \tilde{B}_4(0) = 1$ (see the proof of Corollary 1), and $\tilde{B}_3(0) = 0$, so in the RANS case, $y(t,0) = x_2(t) + x_4(t)$. Thus, the RANS yield curve $\tilde{B}_{2:4}(\tau)x_{2:4}(t)$ is

$$\begin{aligned} y(t,\tau) &= \tilde{B}_2(\tau)y(t,0) + \tilde{B}_3(\tau)x_3(t) + \left(\tilde{B}_4(\tau) - \tilde{B}_2(\tau)\right)x_4(t) \\ &= \left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau}\right)y(t,0) + \left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau} - \mathrm{e}^{-a\tau}\right)x_3(t) + \left(\frac{1-\mathrm{e}^{-2a\tau}}{2a\tau} - \frac{1-\mathrm{e}^{-a\tau}}{a\tau}\right)x_4(t). \end{aligned} \tag{35}$$

The augmenting loading, $\tilde{B}_4(\tau)$, enters with non-zero coefficient unless $x_4 = 0$, a level it will instantaneously leave due to non-zero drift, $-\psi_2(x)^2/a < 0$, cf. (34).

Besides x_2, the stochastic state variable, x_3 and x_4 serve as additional state variables. Although locally deterministic, $\psi_3 = \psi_4 = 0$, they are time-varying. Thus, writing out (18) in the case from Proposition 2, we have $\mathrm{d}x_3(t) = -ax_3(t)\mathrm{d}t$, $\mathrm{d}x_4(t) = 2a(-\psi_2(x)^2/(2a^2) - x_4(t))\mathrm{d}t$, and

$$\mathrm{d}x_2(t) = a\left(\frac{\psi_2(x)^2}{a^2} + \lambda(x)\frac{\psi_2(x)}{a} - x_2(t) + x_3(t)\right)\mathrm{d}t + \psi_2(x)\mathrm{d}W_2(t). \tag{36}$$

[16] As for arbitrary (not necessarily arbitrage-free) DTSMs, ANS is dynamically consistent with the specification in Proposition A.9.1.

The short rate dynamics are given by adding $\mathrm{d}x_2(t)$ and $\mathrm{d}x_4(t)$,

$$\mathrm{d}y(t,0) = a\left(\lambda(x)\frac{\psi_2(x)}{a} - y(t,0) + x_3(t) - x_4(t)\right)\mathrm{d}t + \psi_2(x)\mathrm{d}W_2(t). \tag{37}$$

From (37), the rate of mean reversion, a, in the short rate coincides with the rate of decline of the volatility function (31). Inspection of the spanning condition reveals that this relation is due to the slope adjustment.[17] Further, in (37), the target for mean reversion in the short rate is moving with the additional locally deterministic state variables $\bar{\bar{x}}(t) = (x_3(t), x_4(t))'$.

In the square-root case, $\psi_2(x) = \sigma\sqrt{x_2}$, as in the Cox, Ingersoll, and Ross (1985) (CIR) model, x_4 is stochastic, as x_2 enters its dynamics, $\mathrm{d}x_4(t) = 2a(-\sigma^2 x_2(t)/(2a^2) - x_4(t))\mathrm{d}t$. In the homoskedastic case, $\psi_2(x) = \sigma$, the yield volatility function (31) is that from the Vasicek (1977) model. In this case, the solution for $x_4(t)$ is

$$x_4(t) = x_4(T_{\mathcal{Y}})\mathrm{e}^{-2a(t-T_{\mathcal{Y}})} - \frac{\sigma^2}{2a^2}\left(1 - \mathrm{e}^{-2a(t-T_{\mathcal{Y}})}\right), \tag{38}$$

and $x_3(t) = x_3(T_{\mathcal{Y}})\mathrm{e}^{-a(t-T_{\mathcal{Y}})}$, for $t > T_{\mathcal{Y}}$, with $T_{\mathcal{Y}}$ the first hitting time when the yield curve assumes RANS shape. Inserting these in (35) shows the evolution of the yield curve explicitly. Its variation via x_3, x_4 only ceases when these are at their long-run levels of 0 and $-\sigma^2/2a^2$. In this case, the yield curve takes on standard affine shape,

$$y(t,\tau) = \left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau}\right)y(t,0) + \left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau} - \frac{1-\mathrm{e}^{-2a\tau}}{2a\tau}\right)\frac{\sigma^2}{2a^2}, \tag{39}$$

depending on time t only through $y(t,0)$. This is now the sole state variable, and stochastic, i.e., $\bar{x}(t)$ can be taken as the short rate, and $y(t,\tau) = \bar{A}(\tau) + \tilde{B}_2(\tau)y(t,0)$, with $\bar{A}(\tau)$ the last term in (39), convexity divided by speed of mean reversion. The dynamic consistency approach accommodates the more general yield curve shape (35), with additional time-dependence through $\bar{\bar{x}}(t) = (x_3(t), x_4(t))'$. Thus, $\bar{A}(\tau)$ is generalized to the time-varying mixture of loadings $\bar{\bar{B}}(\tau)\bar{\bar{x}}(t)$ given by the last two terms in (35). The DTSM with $d = 1$ Wiener process drives the yield curve within the arbitrage-free dynamically consistent curve family with $k = 3$ state variables, only reducing to the standard affine case with $k = 1$ state variable (the short rate) when the locally deterministic state variables are at

[17] From (A.9.9), the term $-a(x_2 - x_3)$ in (36) is required in ϕ to span the slope adjustment.

their long-run levels.

It is worth noting that our ANS extension of the Vasicek (1977) model is different from the extension by Hull and White (1990). The latter replaced the long-run level for the short rate with a function of time, calibrated to fit the current yield curve. Such calibration would typically be non-parametric (e.g., a spline), and so not parsimonious. In contrast, our extension accommodates a fit to the current yield curve using the parsimonious ANS curve shape with loadings (32), and secures dynamic consistency with this shape.

3.3. A stochastic level, slope and curvature model

The arbitrage-free DTSM from Proposition 2, which is dynamically consistent with ANS curve shape, involves but a single Wiener process, W_2, driving the state variable x_2 according to (36). Motivated by the observed level, slope, and curvature structure of yield curves, we specify a DTSM with three driving Wiener processes and volatility functions proportional to the three NS loading functions (15). From (30), with $\tilde{B}_3(\tau)$ included in the volatility function, convexity involves the functions $\tilde{B}_5(\tau) = (\tau/2)\mathrm{e}^{-2a\tau}$ and $\tilde{B}_6(\tau) = \tau\left(2\tilde{B}_4(\tau) - \tilde{B}_2(\tau)\right)$. Therefore, these should be included as additional loadings for dynamic consistency under the no-arbitrage condition, cf. (27). Also, as noted already, the level factor loading $\tilde{B}_1(\tau)$ generates convexity $\tau/2$. Spanning this requires a linear (in maturity) loading function, i.e., diverging for long maturities, which is not realistic, hence calling reliance on $\tilde{B}_1(\tau)$ into question.[18] Further, $\tilde{B}_1(\tau)$ is the limit of $\tilde{B}_2(\tau)$ as $a \downarrow 0$, and as a reflects the rate of mean reversion for the associated state variable, cf. (36), the state variable (level factor) associated with $\tilde{B}_1(\tau)$ would exhibit a unit root, which is empirically unwarranted. Therefore, we henceforth (by slight abuse of notation) employ the modified loading function $\tilde{B}_1(\tau) = \left(1 - \mathrm{e}^{-b\tau}\right)/(b\tau)$ with small $b > 0$, and $0 < b < a$, in place of the constant specification. As $\tilde{B}_1(\tau)$ takes the same form as $\tilde{B}_2(\tau)$, with b replacing a, Proposition 2 and the related discussion shows that an additional loading function $\tilde{B}_7(\tau) = \left(1 - \mathrm{e}^{-2b\tau}\right)/(2b\tau)$ (corresponding to (29)) is required, too.

Of the resulting seven state variables, three are associated with the driving Wiener pro-

[18] Note also that $\tilde{B}_1(\tau)$ dropped out of the reduced ANS curve shape in relation to Proposition 2.

cesses, and four are locally deterministic. The model accommodates correlation between the second and third state variables, i.e.,

$$\psi(x) = \begin{pmatrix} \psi_{11}(x) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \psi_{22}(x) & \psi_{23}(x) & 0 & 0 & 0 & 0 \\ 0 & \psi_{32}(x) & \psi_{33}(x) & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{40}$$

The DTSM volatility function $\sigma(\tau,x)' = B(\tau)\psi(x)'$ from (26) in this case takes the form

$$\begin{aligned} \sigma(\tau,x)' &= \begin{pmatrix} \psi_{11}\tilde{B}_1(\tau) & \psi_{22}\tilde{B}_2(\tau) + \psi_{32}\tilde{B}_3(\tau) & \psi_{23}\tilde{B}_2(\tau) + \psi_{33}\tilde{B}_3(\tau) \end{pmatrix} \\ &= \left(\psi_{11}\left(\tfrac{1-e^{-b\tau}}{b\tau}\right) \quad \psi_{22}\left(\tfrac{1-e^{-a\tau}}{a\tau}\right) + \psi_{32}\left(\tfrac{1-e^{-a\tau}}{a\tau} - e^{-a\tau}\right) \quad \psi_{23}\left(\tfrac{1-e^{-a\tau}}{a\tau}\right) + \psi_{33}\left(\tfrac{1-e^{-a\tau}}{a\tau} - e^{-a\tau}\right)\right), \end{aligned} \tag{41}$$

suppressing x in ψ. Since for b small the first volatility function is associated with an approximate level factor, this is a stochastic level, slope, and curvature or SLSC specification.

For a more compact statement of the drift of the state variables x under the no-arbitrage condition in the following theorem, write $\omega(x) = \psi(x)'\psi(x) = \mathrm{diag}(\omega_{11}(x), \tilde{\omega}(x), 0_{4\times 4})$ for their 7×7 block-diagonal local variance matrix. By (40), $\omega_{11}(x) = \psi_{11}^2(x)$, and $\tilde{\omega}(x)$ is given by

$$\begin{pmatrix} \omega_{22}(x) & \omega_{23}(x) \\ \omega_{23}(x) & \omega_{33}(x) \end{pmatrix} = \begin{pmatrix} \psi_{22}^2(x) + \psi_{32}^2(x) & \psi_{22}(x)\psi_{23}(x) + \psi_{32}(x)\psi_{33}(x) \\ \psi_{22}(x)\psi_{23}(x) + \psi_{32}(x)\psi_{33}(x) & \psi_{23}^2(x) + \psi_{33}^2(x) \end{pmatrix}. \tag{42}$$

Theorem 2. *The SLSC curve shape given by loading functions*

$$\tilde{B}_{1:7}(\tau) = \left(\tfrac{1-e^{-b\tau}}{b\tau} \quad \tfrac{1-e^{-a\tau}}{a\tau} \quad \tfrac{1-e^{-a\tau}}{a\tau} - e^{-a\tau} \quad \tfrac{1-e^{-2a\tau}}{2a\tau} \quad \tfrac{\tau}{2}e^{-2a\tau} \quad \tfrac{1}{a}\left(e^{-a\tau} - e^{-2a\tau}\right) \quad \tfrac{1-e^{-2b\tau}}{2b\tau}\right) \tag{43}$$

with fixed a, b is dynamically consistent with the arbitrage-free DTSM with drift $\alpha(\tau,x) = \tilde{B}_{1:7}(\tau)\phi(x)$ and volatility function (41), provided $\psi(x)$ from (40) and $\phi(x) = \Phi(\theta - x)$ with

$$\Phi = \begin{pmatrix} b & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a & -a & 0 & -1 & 0 & 0 \\ 0 & 0 & a & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2a & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 2a & 0 & 0 \\ 0 & 0 & 0 & 0 & a & 2a & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2b \end{pmatrix} \tag{44}$$

and

$$\theta = \begin{pmatrix} \frac{1}{b^2}\psi_{11}^2 + \frac{1}{b}\lambda_1\psi_{11} \\ \frac{1}{4a^2}(4\omega_{22} + 7\omega_{33} + 10\omega_{23}) + \frac{1}{a}\left(\lambda_2\psi_{22} + (\lambda_2 + \lambda_3)\psi_{23} + \lambda_3\psi_{33}\right) \\ \frac{1}{4a^2}(\omega_{33} + 2\omega_{23}) + \frac{1}{a}\left(\lambda_2\psi_{23} + \lambda_3\psi_{33}\right) \\ -\frac{1}{4a^2}(2\omega_{22} + 5\omega_{33} + 6\omega_{23}) \\ \frac{1}{2a}\omega_{33} \\ -\frac{1}{4a}(3\omega_{33} + 2\omega_{23}) \\ -\frac{1}{2b^2}\psi_{11}^2 \end{pmatrix} \tag{45}$$

satisfy Assumption 1.

Of course, loadings can be combined differently, provided the span is maintained, e.g., $\tilde{B}_6(\tau)$ can be replaced by $\exp(-2a\tau)$ in (43), because $\exp(-a\tau)$ is spanned by other loadings, $\exp(-a\tau) = \tilde{B}_2(\tau) - \tilde{B}_3(\tau)$. An alternative, more elaborate expression for the drift of the state variables under the arbitrage condition, $\phi(x)$, can be obtained by substituting (42) in (45) (see (A.9.23)).[19]

For state-independent volatilities, i.e., $\omega_{ij}(x) = \omega_{ij}$, $i,j = 1,2,3$, the drifts of $\bar{\bar{x}}(t) = (x_4(t),\ldots,x_7(t))'$ are deterministic. Thus, the locally deterministic state variables are, indeed, deterministic, although in general time-varying, with long run levels $\bar{\bar{\theta}} = (\theta_4,\ldots,\theta_7)'$ given by the last four entries in (45). In this case, if the yield curve $y(t,\tau) = \tilde{B}_{1:7}(\tau)x_{1:7}(t)$ at any point in time assumes the special shape $y(t,\tau) = \tilde{B}_{1:3}(\tau)x_{1:3}(t) + \tilde{B}_{4:7}(\tau)\bar{\bar{\theta}}$, then it retains this shape, since the drifts and volatilities of the deterministic state variables vanish.

[19] Regarding arbitrary (not necessarily arbitrage-free) DTSMs, the SLSC curve shape given by (43) is dynamically consistent with the specification in Proposition A.9.2.

This restricted version of the SLSC model, with curve shape $y(t,\tau) = \bar{A}(\tau) + \tilde{B}_{1:3}(\tau)x_{1:3}(t)$, where $\bar{A}(\tau) = \tilde{B}_{4:7}(\tau)\bar{\bar{\theta}}$, corresponds closely to the AFNS model considered by Christensen, Diebold, and Rudebusch (2011) (see Appendix A.10).[20]

To summarize, the results from the present section suggest that if a parametrized curve shape is adopted for parsimony in hedging, then it should be dynamically consistent with a suitable DTSM. Otherwise, the optimal portfolio from Theorem 1 relies on information that is in conflict with interest rate dynamics, and hence returns.

4. Empirical Strategy

We exploit dynamic consistency in hedging by imposing increasing structure in the estimation of (B,Ψ) (required in Theorem 1) over three stages. Assume that initial estimates have been obtained from the unrestricted yield factor model (5), as well as from a restricted version imposing a parametric curve shape on B for parsimony. In the first stage, if the given curve shape is dynamically inconsistent, we use the results from Section 3 to augment it to achieve dynamic consistency with a suitable DTSM and impose the augmented curve shape on B instead. In the second stage, we exploit the consistent dynamics of the DTSM identified in the first stage using a reduced factor model for slope-adjusted yield changes involving only the stochastic factors, and accommodating the possibility that the current yield curve does not assume the dynamically consistent shape. Finally, if it does, then so do future yield curves under the given DTSM, and information is potentially lost by ignoring this. In the third stage, we fully exploit the dynamics by jointly imposing the DTSM and the restrictions on the curve shape, leading to a filtering approach along the dynamically consistent curve family. The second and third stage approaches are presented in more detail in the following.

[20] The independent factor AFNS model is obtained by restricting the correlation between the slope and curvature factors to zero, i.e., $\psi_{23} = \psi_{32} = 0$ in (40). Further, AFNS restricts $b = 0$, so $\tilde{B}_1(\tau) = 1$, convexity diverges in maturity, the first factor is an exact level factor exhibiting a unit root, i.e., there is no mean reversion or stable long-run level θ_1, and in fact no affine representation.

4.1. Reduced factor model for slope-adjusted yield changes

For the second stage, the model for the slope-adjusted yield changes (4) is obtained from an Euler approximation to (16), using the arbitrage condition (24) for $\alpha(t,\tau)$, and subtracting the slope adjustments on both sides, producing

$$\tilde{y}_{t+1,\tau} = \frac{\tau}{2}\sigma(x_t,\tau)'\sigma(x_t,\tau) + \sigma(x_t,\tau)'\lambda(x_t) + \sigma(x_t,\tau)'w_{t+1} + \upsilon_\tau + \varepsilon_{t+1,\tau}, \tag{46}$$

where $w_{t+1} = \int_t^{t+1} W_s ds$, υ_τ are maturity-specific pricing errors, and ε_{t+1} are zero-mean measurement errors. For a time-invariant volatility function $\sigma(x_t,\tau) = \sigma(\tau)$, specification (46) corresponds to that in Christensen and van der Wel (2019). This suggests that we can use a reduced factor analysis (or the Kalman filter, in case of dynamics in $\lambda_t = \lambda(x_t)$), with the $d \leq k$ factors given by w_{t+1}, to estimate $\sigma(\tau)$. The analysis in Christensen and van der Wel (2019) did not use parsimonious parametrization or curve shapes, but instead focused on estimation with and without the no-arbitrage restriction $\upsilon_\tau = 0$, and the test of this. Here, we impose the parametric restrictions from the curve shape $\mathcal{B}$ on the loadings in the reduced factor analysis.

Consider a curve shape (factor model) $\mathcal{B}$ and a DTSM (λ,σ) dynamically consistent with $\mathcal{B}$, with k factors in $\mathcal{B}$, $d \leq k$ Wiener processes in the DTSM, and coefficents (ϕ,ψ) in the state process (18). In (27), split $\phi(x)$ into the coefficients for spanning slope adjustments, and those for spanning convexity and risk compensation,

$$\frac{1}{\tau}\big[B(\tau) - B(0)\big]x + \frac{\partial B}{\partial \tau}(\tau)x = B(\tau)\phi_{sa}(x), \tag{47}$$

$$\frac{\tau}{2}B(\tau)\psi(x)'\psi(x)B(\tau)' + B(\tau)\psi(x)'\lambda(x) = B(\tau)\phi_{cr}(x), \tag{48}$$

with $\phi(x) = \phi_{sa}(x) + \phi_{cr}(x)$. From (47), it is clear that $\phi_{sa}(x)$ is linear in x, $\phi_{sa}(x) = -\Phi_{sa}x$, say. From (48), if $\psi(x)'\psi(x)$ and $\psi(x)'\lambda(x)$ are affine in x (see Section 3.2.1), so is $\phi_{cr}(x) = \phi_{cr} - \Phi_{cr}x$, say. In this case, $\phi(x) = \Phi(\theta - x)$, with $\Phi\theta = \phi_{cr}$, $\Phi = \Phi_{sa} + \Phi_{cr}$. If $\psi(x) = \psi$ (Gaussian state variables), then Φ_{cr} purely reflects state-dependence in $\lambda(x)$. Under the martingale measure ($\lambda(x) = 0$), or just constant market prices of risk ($\lambda(x) = \lambda$), we have $\Phi = \Phi_{sa}$, i.e., mean reversion corresponds to slope adjustment. Long-run levels θ in addition reflect convexity and risk compensation, whether or not ψ, λ depend on x. The

cross-sectional curve shape $Y(\tau,x) = B(\tau)x$ only involves the parameters in B, so these are parameters under the martingale measure. The market prices of risk under the physical measure enter with the dynamics via $\phi_{cr}(x)$. In our empirical work, we focus on the case without state dependence in ψ and λ, so $\Phi_{cr} = 0$, i.e., market prices of risk enter θ, not Φ.[21]

Inserting (26) and (48) in (46), with $\phi_{cr}(x) = \Phi\theta$, it follows that

$$\tilde{y}_{t+1,\tau} = B(\tau)\left(\Phi\theta + \psi' w_{t+1}\right) + \upsilon_\tau + \varepsilon_{t+1,\tau}\,. \tag{49}$$

The resulting model can be estimated with and without the no-arbitrage restriction $\upsilon_\tau = 0$ imposed. Thus, shifting the factor analysis from the level of yields to slope-adjusted yield changes allows reducing the number of factors from k in (5) to d (note that ψ' in (49) is $k \times d$, with $k \geq d$), and testing the no-arbitrage condition.[22]

For the ANS-extended Vasicek model, using B, Φ, θ, and ψ from Proposition 2 in (49) produces

$$\tilde{y}_{t+1,\tau} = \left(\frac{\psi_2^2}{a} + \psi_2\lambda\right)\tilde{B}_2(\tau) - \frac{\psi_2^2}{a}\tilde{B}_4(\tau) + \psi_2\tilde{B}_2(\tau)w_{t+1} + \upsilon_\tau + \varepsilon_{t+1,\tau}\,, \tag{50}$$

a single-factor representation with loadings restricted to $\psi_2\tilde{B}_2(\tau)$, and means in addition to loadings (volatilities) times market price of risk involving restricted convexity.[23] The specification (50) does not involve the term $\phi_{sa}(x) = -\Phi_{sa}x$ associated with parametrized slope adjustments, cf. (47), since the slope-adjusted yield changes $\tilde{y}_{t+1,\tau}$ from (4) are modeled, i.e., the current yield curve and the cross section of slope adjustments are taken from the data. The estimated a is used to construct the remaining loadings $\tilde{B}_3(\tau)$ not present in (50), but needed in Theorem 1. Similarly, for the SLSC model, B is given by (43), Φ by (44), θ by (45), and ψ by (40), so (49) is estimated with these specifications and $d = 3$ factors in w_{t+1}. We set $b = 0.02$ to avoid problems with diverging convexity and non-stationary factor dynamics. For this value, the first loading on the 10-year yield,

[21] In the ANS-extended Vasicek and SLSC models, Φ in (34) and (44) represent Φ_{sa} in the general case, also corresponding to $\Phi = \Phi_{sa} + \Phi_{cr}$ under the stated conditions, because the last term vanishes.

[22] From the discussion, with affine state dependence in $\psi'\psi$ or $\psi'\lambda$, the term $\Phi\theta$ in (49) is expanded to $\Phi\theta - \Phi_{cr}x_t$, and estimation relies on the Kalman filter (extended, if ψ depends on x) rather than factor analysis.

[23] In our empirical work, we use the exact discrete-time version of (50) (see Appendix F.2 for the derivation in the third-stage filtering case).

the longest in our data, is $\tilde{B}_1(10) = 0.91$, so the first factor loads nearly evenly over the relevant range.

4.2. Filtering along the dynamically consistent curve family

The measurement equation in the third-stage filtering approach in the factor model case is given by

$$y(t,\tau) = B(\tau)x(t) + \varepsilon(t,\tau), \tag{51}$$

for the available maturities τ. The state transition equation is a discretized version of (18). For dynamic consistency between the curve shape $B(\tau)$ and a suitable DTSM (λ,σ), the coefficients (ϕ,ψ) in (18) are determined by (25)-(26).[24] The resulting state space model can be estimated by maximum likelihood based on the (extended) Kalman filter. In case of affine state drift $\phi(x) = \Phi(\theta - x)$ (e.g., if (λ,σ) do not depend on x, or if (ϕ,ψ) satisfy ATSM restrictions), a simple Euler discretization of the transition equation is

$$x(t+1) = \Phi\theta + (I_k - \Phi)x(t) + \psi' w(t+1), \tag{52}$$

which allows running the linear Kalman filter.[25]

The state space model (B,ϕ,ψ) is a realization of the DTSM (λ,σ), with measurement equation (51) resembling the factor model (5), but now with serially dependent $x(t)$, which is required for dynamic consistency. Nevertheless, Theorem 1 continues to apply, because generalized duration matching removes all exposure to $x(t)$, so there are no gains to exploiting the dynamics to forecast $x(t)$.[26] The potential gain in hedging performance stems from more appropriate model specification and thus improved estimation of the parsimonious B.

In this third-stage approach, the shape B is imposed on the curve throughout (up to

[24] Without the no-arbitrage condition, (ϕ,ψ) instead satisfy (A.7.37) and (26). Outside the factor model case, the measurement equation is $y(t,\tau) = Y(\tau,x(t)) + \varepsilon(t,\tau)$, and (ϕ,ψ) satisfy (A.7.1) and (23) (or (22)-(23), without the no-arbitrage condition).

[25] In our empirical work, we base the Kalman filter recursions on the exact discrete-time transition equation (derived in Appendix F.2), using the Koopman, Shephard, and Doornik (1999) low storage algorithm, with the updating step inserted in the prediction step to save on calculations, and modified to the square-root case (see Appendix F.3).

[26] See also Appendix D.2 on this point.

measurement error, cf. (51)), and the consistent (ϕ,ψ) are imposed on the state dynamics. For comparison, in the first stage, the curve shape B is imposed, but not the dynamics (ϕ,ψ). In the second stage, involving the switch from yields $y_{t+1,\tau}$ to the slope-adjusted yield changes $\tilde{y}_{t+1,\tau}$ from (4), the state dynamics are exploited via $(\Phi\theta,\psi)$ in (49), and B is imposed on the factor loadings, but not on the full yield curve, as the current (time t) yield curve and slope adjustments on the left side of (49) are taken directly from the data. This is in the spirit of the HJM approach of conditioning on the initial (current) yield curve. The third-stage approach combines the first and second stages by jointly imposing the curve shape and the consistent dynamics.[27]

In the ANS-extended Vasicek case, it follows from Proposition 2 that the RANS loadings $\tilde{B}_{2:4}$ suffice for dynamic consistency. Thus, the measurement equation (51) takes the form

$$y(t,\tau)=\tilde{B}_2(\tau)x_2(t)+\tilde{B}_3(\tau)x_3(t)+\tilde{B}_4(\tau)x_4(t)+\varepsilon(t,\tau). \tag{53}$$

While the second-stage model (50) includes $\tilde{B}_4(\tau)$, to span convexity based on yield volatility proportional to $\tilde{B}_2(\tau)$, the third-stage model (53) accommodates $\tilde{B}_3(\tau)$, as well, in the dynamically consistent curve shape. With $\phi(x)$ given by the last three coordinates of (34), and $\psi=(\psi_2,0,0)'$, the transition equation (52) is

$$\begin{pmatrix} x_2(t+1)\\ x_3(t+1)\\ x_4(t+1)\end{pmatrix}=\begin{pmatrix} \frac{\psi_2^2}{a}+\lambda\psi_2\\ 0\\ -\frac{\psi_2^2}{a}\end{pmatrix}+\begin{pmatrix} 1-a & a & 0\\ 0 & 1-a & 0\\ 0 & 0 & 1-2a\end{pmatrix}\begin{pmatrix} x_2(t)\\ x_3(t)\\ x_4(t)\end{pmatrix}+\begin{pmatrix} \psi_2 w_2(t+1)\\ 0\\ 0\end{pmatrix}. \tag{54}$$

If the first hitting time, T_y, when the yield curve assumed the dynamically consistent ANS shape and the associated values x_y of the state variables at that time were all known with certainty, they could be used to initialize the filter. Because they are in fact uncertain, we use an uninformed prior on $x(0)$, and allow a transition shock of small but fixed size. Similarly, filtering along the SLSC curve family (43), which is dynamically

[27] Relative to the second-stage approach, the third stage restricts the current curve and the slope adjustments according to the dynamically consistent shape, cf. (25) and (A.7.1). Typically, the second stage involves as many parameters as the third, because $B(\tau)(\Phi\theta,\psi')$ in (49) generically includes all the parameters from (B,λ,σ) that enter (B,Φ,θ,ψ'), and this is the case in the models we implement.

consistent with the arbitrage-free DTSM with volatility function (41), is carried out based on the linear filter (51)-(52), with Φ, θ, and ψ from (44), (45), and (40), respectively, and $k = 7$ state variables, of which four deterministic and $d = 3$ stochastic.

5. Data

We use data from the Federal Reserve Board's (FED's) database of constant maturity zero-coupon yields on U.S. Treasury bills, notes, and bonds. The terms to maturity considered are 0.25, 0.5, 1, 2, 3, 5, 7, and 10 years. 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 runs from the first week of 1983 through the last week of 2019, for a total of 1,930 observations in the time series dimension. Starting in 1983 avoids the FED money supply targeting experiment from 1979 to 1982 (see Sanders and Unal (1988)).

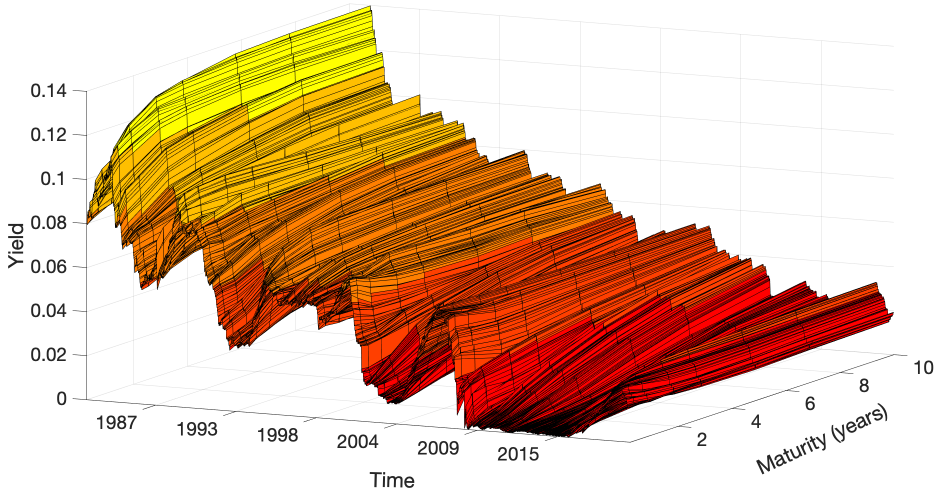
Table 1 shows means and standard deviations of the weekly data on the continuously compounded annualized yields corresponding to the eight maturities. The term structure of interest rates is upward sloping on average, with means monotonically increasing from 3.71% to 5.48%. The term structure of volatilities or standard deviations exhibits a hump shape, with a maximum of 3.10% at two years, and a low of 2.77% at ten years. Figure 1 presents a three-dimensional view of the evolution of yield curves through calendar time, revealing upward sloping, downward sloping, and hump shapes.

Table 1: Summary statistics
Mean and standard deviation for each of the eight weekly constant maturity zero-coupon yield series from January, 1983, through December, 2019.

	3 mos.	6 mos.	12 mos.	2 yrs.	3 yrs.	5 yrs.	7 yrs.	10 yrs.
Mean (%)	3.71	3.87	4.04	4.39	4.61	4.98	5.27	5.48
Std. Dev. (%)	2.90	2.97	3.02	3.10	3.07	2.96	2.88	2.77

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

Figure 1: Yield curves



This figure provides a three-dimensional view of the weekly yield data from January, 1983, through December, 2019, at maturities 0.25, 0.5, 1, 2, 3, 5, 7, and 10 years.

not necessarily a Wednesday, the daily files are used again to get the correct zero-coupon bond prices when constructing the hedge portfolio.

As target asset for assessing hedging performance, we consider a portfolio consisting of a long position in a five-year coupon bond and short positions in two-year and ten-year coupon bonds. This specification with short positions in the long and short ends follows Litterman and Scheinkman (1991).[28] The monthly return series is constructed by drawing information on prices and contractual terms from the CRSP Monthly Treasury files. On the last trading day of each month, we select among all non-callable and non-flower bonds the issues with maturities closest to two, five, and ten years.[29] Portfolio weights $(-1, 3, -1)$ are then assigned to construct the target asset. As our hedging portfolios are always based on an estimation period of at least 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, 2019. Properties of the resulting $T = 395$ monthly target returns are

[28] We also considered a single (five-year) coupon bond target, as in Diebold, Ji, and Li (2006). Overall conclusions were similar.

[29] Treasuries are non-convertible. Flower bonds were issued until 1965, with the last outstanding issues maturing in 1998. Callable bonds and notes were issued until 1985, but many of these subsequently repurchased by the Treasury and reissued as non-callable, although on a discretionary basis, without sinking fund provision. We remove all flower and convertible issues and apply a liquidity requirement of at least $10 million in par value publicly outstanding. Figure B.1 shows the evolution over time in the resulting number of Treasuries we consider.

shown in the first row of Table 2. The average return is 49bp, or 0.49%, and the standard deviation 149bp.[30]

6. Empirical Results

We first present the results on empirical hedging performance. This is followed by a discussion of statistical fit and test of the no-arbitrage condition.

6.1. Hedging Performance

Results appear in Table 2. Line 1 reports statistics for the unhedged target return. Each subsequent line reports bias (mean hedging error), standard deviation, root mean squared error (RMSE), and mean absolute error (MAE) for a given strategy. Following Chambers, Carleton, and McEnally (1988), Diebold, Ji, and Li (2006), and others, we will mainly focus on RMSE in the exposition.

The immunization performance of the traditional duration matching strategy is summarized in line 2.[31] It yields a bias (average return to hedged position) of 2.84bp. RMSE is large relative to unhedged variation, more than one third, indicating that immunization by simple duration matching is too simplistic.

All other strategies in Table 2 involve generalized duration matching based on estimation of B and Ψ, cf. Theorem 1. Results are shown both for full-period estimation and for an out-of-sample (OoS) experiment with rolling estimation over the four-year period prior to forming the hedge. While full period calculations provide the artificial investor with the benefit of hindsight, the OoS analysis mimics a feasible strategy, in line with basic duration matching. Although the models involve constant parameters, updating these is in the spirit of the HJM approach of conditioning on current information (see also Buraschi and Corielli (2005)). Restricting the estimation window to four years is done in order to reduce conditioning on obsolete information.

[30] This is the unhedged return. The column is labeled 'Bias' because average hedging errors are reported in the remainder of the table. More details on the construction and properties of the target asset are provided in Appendix B.

[31] Details on basic and generalized duration matching portfolio construction are given in Appendix C.

Table 2: Hedging performance

The target is a portfolio of $(2,5,10)$-year coupon bonds in the proportions $(-1,3,-1)$. Statistics in line 1 are for the unhedged target return, and in the remainder of the table for hedging errors from each of the methods considered for construction of the hedge portfolio with value matching from Theorem 1. The columns report the average (or bias), standard deviation, root mean squared error, and mean absolute error. Results are in basis points (0.01%) per month. An S indicates that a given method provides a statistically significant improvement over traditional duration matching at the 5% level, and MCS that a method is included in the Model Confidence Set at 5% (only conducted for the rolling strategies).

	Model	Bias	Std. dev.	RMSE	MAE
1	Target movement	49.20	149.44	157.33	122.63
2	Duration matching	2.84	65.90	65.96	48.72
3	Unrestricted 3-factor Full period	-0.94	56.39	56.40	41.04
4	Unrestricted 3-factor Rolling 4-year	-1.23	54.56	54.57 (S,-)	39.55 (S,-)
5	Nelson-Siegel Full period	-1.34	57.07	57.08	41.36
6	Nelson-Siegel Rolling 4-year	-0.79	58.63	58.63 (S,-)	41.93 (S,-)
7	Unrestricted 4-factor Full period	-0.68	41.34	41.34	31.61
8	Unrestricted 4-factor Rolling 4-year	-2.11	47.54	47.58 (S,-)	34.75 (S,-)
9	Augmented NS Full period	-1.58	44.56	44.58	33.62
10	Augmented NS Rolling 4-year	-2.93	39.39	39.50 (S,-)	28.21 (S,MCS)
11	Unrestricted 1-factor, $\tilde{y}$ Full period	0.80	64.71	64.71	47.15
12	Unrestricted 1-factor, $\tilde{y}$ Rolling 4-year	1.14	60.96	60.97 (S,-)	44.27 (S,-)
13	ANS-extended Vasicek, $\tilde{y}$ Full period	-3.69	33.27	33.47	24.75
14	ANS-extended Vasicek, $\tilde{y}$ Rolling 4-year	-4.77	36.80	37.11 (S,MCS)	27.95 (S,MSC)
15	ANS-extended Vasicek, filter Full period	-3.26	66.49	66.57	48.84
16	ANS-extended Vasicek, filter Rolling 4-year	-3.08	64.59	64.66	48.22
17	ANS-extended Vasicek, restricted Full period	-3.49	62.84	62.94	46.02
18	ANS-extended Vasicek, restricted Rolling 4-year	-2.88	67.46	67.52	51.00
19	Unrestricted 3-factor, $\tilde{y}$ Full period	-2.43	55.10	55.16	39.10
20	Unrestricted 3-factor, $\tilde{y}$ Rolling 4-year	-2.52	58.59	58.64 (S,-)	41.48 (S,-)
21	SLSC, $\tilde{y}$ Full period	-4.61	34.07	34.38	24.78
22	SLSC, $\tilde{y}$ Rolling 4-year	-4.58	34.00	34.31 (S,MCS)	24.66 (S,MCS)
23	SLSC, filter Full period	-4.59	34.06	34.36	24.86
24	SLSC, filter Rolling 4-year	-4.46	33.46	33.75 (S,MCS)	24.33 (S,MCS)
25	SLSC, restricted Full period	-1.62	54.13	54.16	39.53
26	SLSC, restricted Rolling 4-year	-1.06	55.60	55.61 (S,-)	41.08 (S,-)

From the results in lines 3-4, generalized duration matching based on the unrestricted three-factor model (5) offers only a modest gain, relative to basic duration matching. From lines 5-6, the parsimony achieved by imposing the NS curve shape (15) on B comes at the expense of deteriorating hedging performance, relative to the unrestricted case. This suggests that the search for gains from parsimony should proceed on a principled basis, requiring dynamic consistency of the curve shape imposed. Indeed, by Corollary 1, the NS curve shape is dynamically inconsistent with all non-degenerate arbitrage-free DTSMs.

We exploit dynamic consistency in the three stages outlined in Section 4. Results from the first stage appear in lines 7-10. Generalized duration matching using four factors instead of three improves performance. Moreover, the results are broadly supportive of the importance of dynamic consistency, as imposing the parsimonious ANS curve shape (32) improves performance relative to NS and, at least in the feasible (rolling estimation) case, also relative to the unrestricted four-factor factor model.

Results from the second stage (Section 4.1) appear in lines 11-14. First, yields y are replaced by slope-adjusted yield changes $\tilde{y}$ from (4) in an unrestricted reduced (single-factor) version of (5). Lines 11-12 show that performance deteriorates in this case. However, imposing the ANS-extended Vasicek structure and the no-arbitrage condition, leading to specification (50), generates the best performance so far, lines 13-14. Here, dynamic consistency provides parsimony in terms of both parameters and number of factors, and the second-stage approach dominates the first stage in this case.

In the third stage (Section 4.2), the dynamically consistent shape is imposed on the current curve and the slope adjustments, as well, leading to the filter (53)-(54).[32] Lines 15-16 show that hedging results deteriorate, to the level of basic duration matching, indicating that the yield curve is not of ANS shape at the time of portfolio formation. Restricting the curve shape further to the standard affine case (39) by freezing the deterministic state variables at their long-run levels $\bar{\bar{\theta}} = (0, -\psi_2^2/2a^2)'$ does not alter this conclusion, lines 17-18. One possibility is that the DTSM (ANS-extended Vasicek) is correctly specified, but that the yield curve has not yet reached the dynamically consistent

[32] We use the exact discrete-time version, see Appendices F.2-F.3.

shape (ANS), so results are better in the second stage (lines 13-14) than in the third (lines 15-18). Another possibility is that ANS is too restrictive, hence calling for a more flexible curve shape that is dynamically consistent with a more general DTSM.

For a more general specification, we turn to the new SLSC model from Section 3.3. Results for three-factor models for slope-adjusted yield changes appear in lines 19-22. Imposing the SLSC DTSM structure from (49), lines 21-22, improves performance relative to the unrestricted three-factor model for $\tilde{y}$, lines 19-20, thus mirroring the improvement from imposing ANS-extended Vasicek in the one-factor case, lines 11-14. Results from third stage filtering based on (51)-(52) (imposing the SLSC curve shape on the current yield curve and slope adjustments, too) appear in lines 23-24. In contrast to the ANS-extended Vasicek case, lines 13-16, performance in the third stage in the SLSC case, lines 23-24, is at least as strong as in the second stage, lines 21-22. This is consistent with the notion that the SLSC curve shape (43) from Theorem 2 better accommodates the current yield curve, and hence the slope adjustments, compared to ANS from (32). Indeed, for feasible (OoS rolling) estimations, the third stage SLSC model proves to be the strongest performing model. Thus, the SLSC model with three stochastic factors generates value relative to extended Vasicek with but one. Further, from lines 25-26, performance deteriorates when freezing the deterministic state variables at their long-run levels $\bar{\bar{\theta}}$, thus restricting the model to the standard affine case (here, AFNS). The latter does not offer any improvement over the classical three-factor yield model, lines 3-4.

For the SLSC model, rolling estimation generates better variation measures than full-period estimation in both the second and third stages. Since rolling involves an OoS element, it is not given in advance that it should dominate full-period estimation in these performance metrics. Thus, the results indicate the importance of conditioning on non-obsolete information.

As a robustness check, we present in Appendix D.1 the hedging results when using the FED yields to set the prices of the bonds entering the target assets, rather than using the CRSP recorded prices directly (bid-ask midpoints plus accrued interest). While the approach is not applicable in practice, it allows us to evaluate the hedging performance

in the absence of frictions, microstructure noise, and other features present in the raw CRSP prices. The results show that the performance of all models improves, compared to that based on raw target returns in Table 2, with the largest improvement seen in the SLSC models, which clearly dominate all other approaches. Also, for all second and third stage models, the hedging biases present for the raw target returns are reduced when using the cleaned returns.

As a further robustness check, presented in Appendix D.2, we consider again the CRSP data for the target asset, but replace conditional hedging error variance minimization by an RMSE criterion. In addition, we relax the generalized duration matching constraint. This allows trading off hedging error bias and variance, as well as admitting some factor exposure, if this reduces the criterion. From the results, both ANS-extended Vasicek and SLSC specifications produce higher RMSE than the third stage SLSC filtering specification in Table 2, lines 23-24, even though the alternative strategies target RMSE, rather than conditional hedging error variance. Thus, the evidence is that it pays off to remove factor exposure, i.e., perform generalized duration matching, and target remaining idiosyncratic variance, rather than trading this off against average hedging error. Parsimony is again the likely reason. The strategies from Theorem 1 (used in Table 2) involve only estimated B and Ψ, whereas those in the generalized case depend on parameters from the factor dynamics, hence increasing exposure to estimation uncertainty.

Overall, the empirical results show that generalized as opposed to traditional duration matching does not by itself secure a noteworthy improvement in immunization performance. Improvements can be achieved through parsimonious restrictions according to a parametrized yield curve shape, but this must be dynamically consistent with a suitable DTSM. Based on the latter, strong hedges are obtained for both one- and three-factor models, by allowing for an arbitrary initial curve. However, the best feasible OoS hedge is obtained by imposing the SLSC curve shape on current yields, and the consistent SLSC dynamics on the factors. If affine curve shape is imposed, hedging performance deteriorates. This is evidence that the deterministic factors move through time and, hence, the associated loading functions ($\tilde{B}_{3:4}(\tau)$ and $\tilde{B}_{4:7}(\tau)$ in the one- and three-factor cases,

respectively) do not enter yield curves in fixed proportions over time.

6.2. Statistical comparison of hedging performance

For a statistical assessment of the improvements in performance, we construct a standard t-statistic based on loss differentials, following Diebold and Mariano (1995) and Giacomini and White (2006), with basic duration matching as benchmark. In addition, we implement the Model Confidence Set (MCS) procedure of Hansen, Lunde, and Nason (2011) to compare performance across all approaches considered. Given a significance level, α, the MCS identifies the subset of approaches containing the best approaches with probability $1-\alpha$.[33] The comparisons are conducted for the feasible (OoS) strategies, using $\alpha = 5\%$. In Table 2, an S below a performance measure (RMSE or MAE) indicates that the approach improves significantly over duration matching, and MCS that it is included in the Model Confidence Set. The improvements in performance relative to duration matching are statistically significant throughout, except for the ANS-extended Vasicek third stage filtering approaches. The best approaches (the MCS) are the second-stage specifications, both the ANS-extended Vasicek and SLSC based, along with the third-stage SLSC specification. This is consistent with the notion that if the dynamically consistent curve shape is imposed on the slope-adjustments, it must be sufficiently flexible.[34]

6.3. Statistical Fit

Here, we consider the fit of the various models behind the strategies in Table 2. Table 3 reports the estimated (full-period) idiosyncratic standard deviations $\sqrt{\Psi_i} \cdot 1000$ for each model, along with the maximized log-likelihood value, number of parameters, and standard information criteria, AIC and BIC.[35] For comparison, results for simple one- and two-factor versions of (5) are reported in lines 1-2. For the yield factor models (5) in lines 1-6, besides the parameters in the variance-covariance structure $\Upsilon = B\Sigma B' + \Psi$ (with $\Sigma = I_k$ in the unrestricted cases, lines 1-3 and 5), estimates of the means μ are required,

[33] More details on the t-test and the MCS procedure are provided in Appendix E.
[34] Based on the MAE criterion, the four-factor yield model with ANS loadings imposed is in the MCS, too.
[35] More details on estimation methods are provided in Appendix F.

Table 3: Idiosyncratic standard deviations, log likelihood values, and information criteria

Columns 3-10 of this table show the estimated idiosyncratic standard deviation for each maturity and each model estimated. Reported figures are $\sqrt{\Psi_i} \cdot 1000$. The last four columns show the maximized log-likelihood value, the number of parameters, AIC, and BIC. Models for slope-adjusted yield changes are indicated with a $\tilde{y}$, and the remaining models are for yield levels y.

Line	Model	3 mos.	6 mos.	12 mos.	2 yrs.	3 yrs.	5 yrs.	7 yrs.	10 yrs.	$\log L$	# params.	AIC	BIC
1	Unrestricted 1-factor	6.20	5.52	4.24	1.75	0.31	2.85	4.40	5.80	59158	24	-118268	-118084
2	Unrestricted 2-factor	2.37	1.29	0.30	0.93	0.89	0.30	0.90	1.90	71351	31	-142640	-142403
3	Unrestricted 3-factor	1.15	0.30	0.68	0.38	0.31	0.49	0.29	0.98	75873	37	-151672	-151389
4	Nelson-Siegel 3-factor	1.18	0.30	0.78	0.42	0.31	0.50	0.29	1.03	75372	23	-150698	-150522
5	Unrestricted 4-factor	0.93	0.30	0.45	0.31	0.31	0.43	0.29	0.61	76725	42	-153366	-153045
6	Augmented NS 4-factor	0.98	0.30	0.58	0.36	0.31	0.42	0.29	0.80	75985	27	-151915	-151709
7	Unrestricted 1-factor, $\tilde{y}$	1.13	0.89	0.71	0.49	0.37	0.13	0.31	0.43	91510	24	-182998	-182914
8	ANS-extended Vasicek 1-factor, $\tilde{y}$	1.23	0.92	0.68	0.41	0.30	0.26	0.38	0.52	90539	11	-181030	-180846
9	ANS-ext. Vasicek, filter 3-factor	6.01	4.81	3.16	0.61	1.72	4.34	5.85	7.24	64075	11	-128129	-128044
10	ANS-ext. Vasicek, restr. 1-factor	7.07	5.99	4.42	1.68	0.70	3.08	4.65	6.13	63989	11	-127956	-127872
11	Unrestricted 3-factor, $\tilde{y}$	0.63	0.23	0.32	0.21	0.18	0.20	0.10	0.25	96348	37	-192666	-192551
12	SLSC 3-factor, $\tilde{y}$	0.61	0.27	0.32	0.22	0.18	0.19	0.16	0.21	95306	16	-190538	-190255
13	SLSC, filter 7-factor	1.36	0.30	0.77	0.52	0.31	0.51	0.45	0.96	86538	16	-173046	-172931
14	SLSC, restricted 3-factor	1.39	0.30	0.76	0.57	0.31	0.53	0.46	0.85	86401	16	-172771	-172657

and these are given by the average yields from Table 1. In the remaining models, the no-arbitrage condition is imposed on the means. The table shows that unexplained variation is generally largest at the shortest and longest maturities. To avoid Heywood cases (factors explaining more than total variation for a given maturity, see Appendix F.1), a lower bound of $10^{-4}\vartheta_i^2$ is imposed on Ψ_i, for each maturity, in all models, with ϑ_i^2 the total variance at maturity τ_i.[36] The six months, three years, and seven years idiosyncratic variances hit this lower bound in the three- and four-factor yield models, lines 3-6.

From Table 3, lines 2-3, the LR test of two against three factors in the unrestricted yield factor model takes the value $2(75,873-71,351)=9,044$, for a p-value of 0.00 in the asymptotic χ^2-distribution on $37-31=6$ degrees of freedom. This indicates that three factors are required, which is consistent with the information criteria. From lines 3-4, the NS restriction on the three-factor model is rejected on all criteria. From line 5, the unrestricted four-factor model is preferred over the previous models. The ANS-restricted version, line 6, is preferred over both NS, line 4, and the unrestricted three-factor model, line 3. This is consistent with the hedging results in Table 2, lines 3-6 and 9-10. In contrast, while the ANS-restricted factor model also provides a gain in OoS hedging performance relative to the unrestricted four-factor model (Table 2, line 10 against 8), it is statistically rejected in favor of the latter (Table 3, lines 5-6).[37]

The fact that economic and statistical criteria do not necessarily coincide is not confined to the first-stage approach, i.e., imposing ANS on the yield factor model. In the second stage (Section 4.1), while the unrestricted single-factor model for slope-adjusted yield changes $\widetilde{y}_{t+1}$ is preferred over the parsimoniously restricted ANS-extended Vasicek version (50) with factor w_{t+1} on statistical grounds (Table 3, lines 7-8), the restricted model provides a gain in hedging performance (Table 2, lines 11-14). Line 12 in Table 3 shows that the SLSC three-factor model for $\tilde{y}$ produces lower idiosyncratic standard deviations and better information criteria than the ANS-extended Vasicek and unrestricted single-factor models, lines 7-8, and the LR test rejects ANS-extended Vasicek in favor of SLSC

[36] For the yield models, ϑ_i are the standard deviations given in Table 1.

[37] Idiosyncratic standard deviations are slightly larger for the restricted models in lines 4 and 6 than for the corresponding unrestricted models in lines 3 and 5, and smaller with four factors (lines 5-6) than with three (lines 3-4), consistent with the LR tests.

(line 8 against 12). Still, the discrepancy between economic and statistical criteria carries over. The SLSC structure (line 12) is rejected in favor of the unrestricted three-factor model for $\tilde{y}$ (line 11), but hedging performance is far stronger under the SLSC restrictions (Table 2, lines 19-22). Thus, parametrizing loadings in accordance with the empirically well established parsimonious level-slope-curvature pattern pays off in financial terms, in spite of statistical rejection, provided dynamic consistency of the curve shape (ANS or SLSC) is respected in the implementation.

The remaining results in Table 3 relate to third-stage filtering (Section 4.2). In the ANS-extended Vasicek case, line 9, the idiosyncratic standard deviations are high, relative to those in other models. Hedging performance deteriorates, too (Table 2, lines 15-16), so second-stage results dominate third-stage results in terms of both statistical fit and hedging performance. Imposing ATSM restrictions leads to further deterioration in statistical terms (Table 3, line 10).

For the SLSC model, filtering (Table 3, line 13) generates idiosyncratic standard deviations that are smaller than those for ANS-extended Vasicek filtering (line 9) and comparable to those based on three- and four-factor analysis of yields (lines 3-6). However, information criteria are far better, reinforcing the importance of the dynamics. The LR test rejects the reduction to ANS-extended Vasicek (line 9 against 13), and hedging performance is strongest in the SLSC case, too (Table 2, lines 23-24 against 15-16). Thus, SLSC is preferred according to both economic and statistical criteria. Further, both fit and hedging performance deteriorate when ATSM restrictions are imposed (Table 3, line 14; Table 2, lines 25-26).

Estimates of the curve shape parameter a and market prices of risk λ for models with parsimoniously parametrized loadings appear in Table 4. From line 1, the NS estimate of a for the full period is 0.679, and quite precise, with a standard error of 0.005. When inserted in $B = \tilde{B}_{1:3}(a)$ from (15), the estimated a generates the three NS loading functions shown in the right exhibit of Figure F.1 in Appendix F.1. Although capturing a level-slope-curvature pattern similar to that of the unrestricted three-factor model in the left exhibit (up to scale) in a parsimonious fashion, the NS restrictions do not improve

Table 4: Estimates of a and market prices of risk

This table shows full-period estimates of a and market prices of risk, with standard errors below estimates, as well as time-series averages of four-year rolling estimates, for NS and generalizations. Models for slope-adjusted yield changes are indicated with a $\tilde{y}$, and the remaining models are for yield levels y.

Line	Model	a	λ_1	λ_2	λ_3
1	Nelson-Siegel Full period	0.679 0.005			
2	Nelson-Siegel Rolling 4-year	0.838			
3	Augmented NS Full period	0.686 0.006			
4	Augmented NS Rolling 4-year	0.612			
5	ANS-extended Vasicek, $\tilde{y}$ Full period	-0.033 0.002		-0.623 0.165	
6	ANS-extended Vasicek, $\tilde{y}$ Rolling 4-year	-0.070		-0.780	
7	ANS-extended Vasicek, filter Full period	0.034 0.001		-0.073 0.002	
8	ANS-extended Vasicek, filter Rolling 4-year	0.214		-0.047	
9	ANS-extended Vasicek, restr. Full period	0.033 0.001		-0.070 0.002	
10	ANS-extended Vasicek, restr. Rolling 4-year	0.201		-0.029	
11	SLSC, $\tilde{y}$ Full period	0.773 0.015	-0.327 0.167	-0.369 0.167	-0.314 0.178
12	SLSC, $\tilde{y}$ Rolling 4-year	0.737	-0.308	-0.331	-0.294
13	SLSC, filter Full period	0.656 0.005	-0.410 0.162	-0.540 0.165	-0.205 0.159
14	SLSC, filter Rolling 4-year	0.758	-0.373	-0.647	-0.007
15	SLSC, restricted Full period	0.669 0.005	-0.389 0.158	-0.732 0.157	-0.151 0.156
16	SLSC, restricted Rolling 4-year	0.791	-0.425	-0.761	-0.112

hedging performance (Table 2, lines 5-6), and are rejected statistically (Table 3, line 4). This factor model approach to NS is an alternative to the cross-sectional regressions of Diebold, Ji, and Li (2006), who fit the NS curve (14) each month by OLS, treating the factors $f_{t,j}$ as regression coefficients, and fixing a throughout at a value 0.0609 to position the maximum (or hump) in the third loading function at the 30 months maturity around which the yield curve hump is commonly observed.[38] Setting a at an externally prespecified value implies that B is not estimated from data at all. Further, as Diebold,

[38]The value 0.0609 is for τ measured in months and corresponds to $a = 0.731$ in our case, with τ in annual terms, i.e., outside the confidence band around our estimate of 0.679 at conventional levels. We find that $a = 0.731$ generates a maximum at 29.4 months, whereas a maximum at 30 months would require $a = 0.717$ (or 0.0598 in monthly terms). At the empirical estimate, $a = 0.679$, the hump is at $\tau = 31.8$ months.

Ji, and Li (2006) do not minimize hedging error variance, $w'\mathcal{T}\Psi\mathcal{T}w$, cf. (9), but instead follow Ingersoll (1983) and minimize $w'w$, the sum of squared hedging weights, Ψ is not needed, either, and the resulting hedge does not utilize any information from empirical (regression or factor) analysis of the data.

From Table 4, line 2, rolling estimation produces an average a of 0.838. Figure G.1 (Appendix G) shows that there is considerable variation in estimated a over time. One possibility is to consider NS with time-varying a, which is dynamically consistent with a DTSM of the type in Corollary A.5.1 (Appendix A). However, by Corollary 1, NS curve shape implies arbitrage opportunities, whether a is constant or time-varying.

Estimates of a from the first-stage approach to dynamic consistency, imposing ANS curve shape on loadings, rather than NS, are shown in lines 3-4. The full-period ANS and NS estimates are similar, and the rolling ANS estimate is closer to these than to rolling NS, consistent with ANS being less misspecified than NS. The loading functions in the unrestricted four-factor and ANS analyses are shown in Figure G.2 (Appendix G). The fourth unrestricted loading has two small humps, whereas the fourth restricted loading (the augmentation (29)) corresponds to a second (steeper) slope factor.

For the second-stage approach, results from (50) are reported in lines 5-6. Estimated a in the ANS-extended Vasicek for $\tilde{y}$ is now very close to zero.[39] The factor is associated with the slope loading $\tilde{B}_2(\tau;a)$, but for a close to zero, it is nearly flat, so the stochastic factor is essentially a level factor. The loadings in the unrestricted single-factor model for $\tilde{y}$ exhibit an initial slope, then a flat structure for maturities three years and longer (see Table G.1 in Appendix G). Thus, the discrepancy between economic and statistical criteria, with the ANS restrictions formally rejected (Table 3, lines 7-8), but improving hedging performance (Table 2, lines 11-14), is seen to be related to the importance of the level factor in hedging. This explains why the discrepancy disappears in the SLSC model, with both level and slope stochastic. Further, the estimated market price of risk λ in (50) is negative, at -0.62 over the full period, corresponding to the negative relation between

[39] In the reported results, we relax the stationarity condition $a > 0$ to highlight the information about the specification available in the data. The stationarity condition is satisfied in the SLSC model and in all models for yields y.

yields and bond prices, and significant, with a t-statistic of -3.78.

Results of third-stage filtering for the ANS-extended Vasicek model appear in lines 7-10. Although estimates of a are between those from cross-sectional curve fitting or factor analysis, lines 1-4, and those based on slope-adjusted yield changes, lines 5-6, consistent with the notion that the Kalman filter combines cross-sectional and intertemporal information, they remain small in magnitude, at 0.22 or less. In contrast, for the SLSC model, estimates of a based on both slope-adjusted yields, lines 11-12, and filtering, lines 13-16, are large in magnitude, at 0.65 or higher. The results confirm the need for three stochastic factors. With only one included, it is fit to level in the second and third stages ($a \approx 0$).

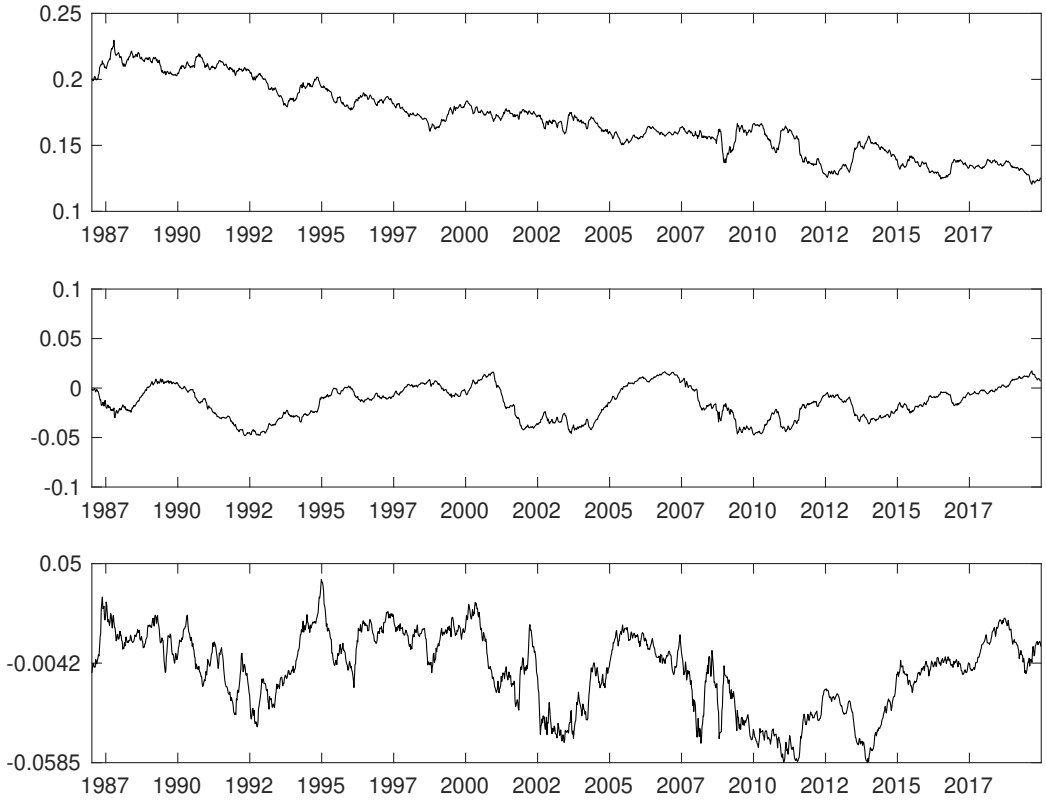
Figure 2 shows the evolution through time of the filtered stochastic factors. Level is the smoothest, and curvature the most volatile, changing sign most frequently. Thus, the NS interpretation of level, slope, and curvature as long-, short-, and medium-term factors is replaced by a long-, medium-, and short-term understanding of the three.[40]

Correspondingly, from the upper exhibit of Figure 3, the rolling estimates of the volatilities ψ of level and slope are similar period by period and relatively stable over time, whereas the volatility of curvature is higher and more variable, with peaks around 2005 and 2011. Nevertheless, on average, curvature risk is unpriced, as seen from the last column of Table 4. Slope gets the largest market price, λ_2, and both level and slope are significantly priced over the full period. Finally, the lower exhibit of Figure 3 shows that the risk prices switch signs over time. In particular, medium-term or slope risk switches from a regime of large negative prices before the financial crisis to one of predominantly positive prices after the crisis.

Figure 4 offers a visualization of the decomposition of the fitted yield curve into its separate stochastic and deterministic components, $y(t,\tau) = y_s(t,\tau) + y_d(t,\tau)$, say, and the evolution of this through calendar time in the rolling estimation case. The upper left exhibit shows the stochastic portion, $y_s(t,\tau) = \tilde{B}_{1:3}(\tau)x_{1:3}(t)$, of the general SLSC model, and the upper right exhibit the associated deterministic component, $y_d(t,\tau) = \tilde{B}_{4:7}(\tau)x_{4:7}(t)$. The lower exhibits show the corresponding terms for the restricted (AFNS) version, where

[40] The correlation between slope and curvature is negligible, estimated at 0.0128 over the full period by filtering, and 0.0124 in the slope-adjusted yield change analysis.

Figure 2: SLSC factors filtered along dynamically consistent curve family



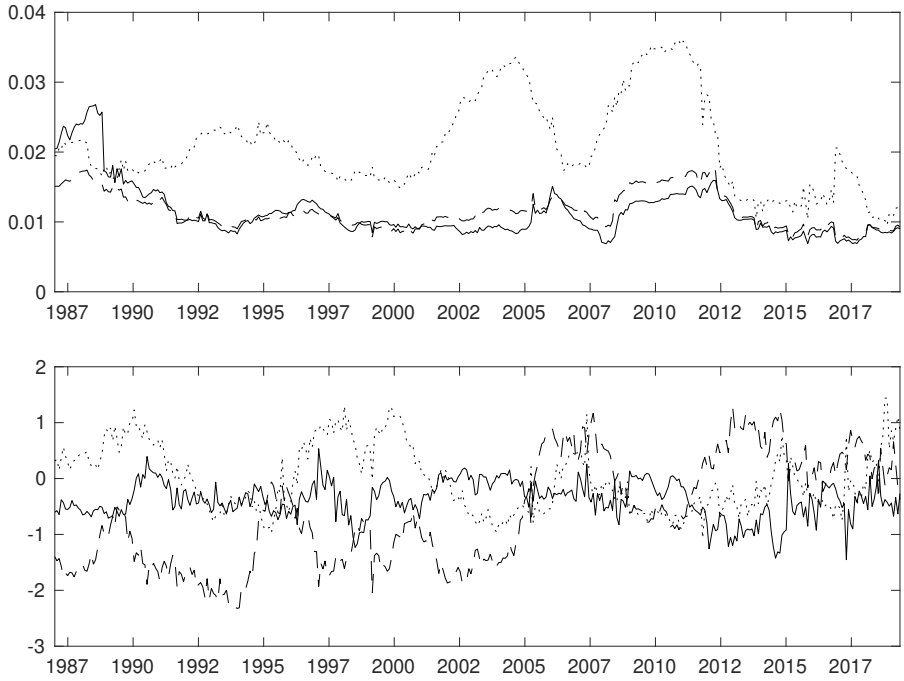
This figure shows the time series evolution of fitted factors (state variables) from filtering along the dynamically consistent curve family in the SLSC model, calculated using the full sample and the Kalman smoother, $\mathbb{E}(x(t) \mid y_1, \ldots, y_T)$, with the level, slope, and curvature factor in the upper, middle, and lower exhibit, respectively.

there is no factor dependence in the deterministic part, i.e., time-variation in $y_d(t,\tau) = \bar{A}(\tau)$ is solely due to changing parameter estimates. For example, a dip in the stochastic part of the curve between 2009 and 2015 is countered by a raised $\bar{A}(\tau)$ function. In the SLSC model (top exhibits), the stochastic and deterministic components of the curve move more freely, separately from each other, due to the presence of factors in $y_d(t,\tau)$. The results on statistical fit and hedging performance suggest that this added flexibility matters for yield curves, and that this can be exploited by investors.

6.4. The No-Arbitrage Condition

The reported second- and third-stage results are obtained with the no-arbitrage condition imposed. We also consider testing this condition. Results of second-stage estimation with and without $v_\tau = 0$ imposed on (49) are shown in Table 5. The first two columns refer to

Figure 3: Estimated volatilities and market prices of risk in SLSC model



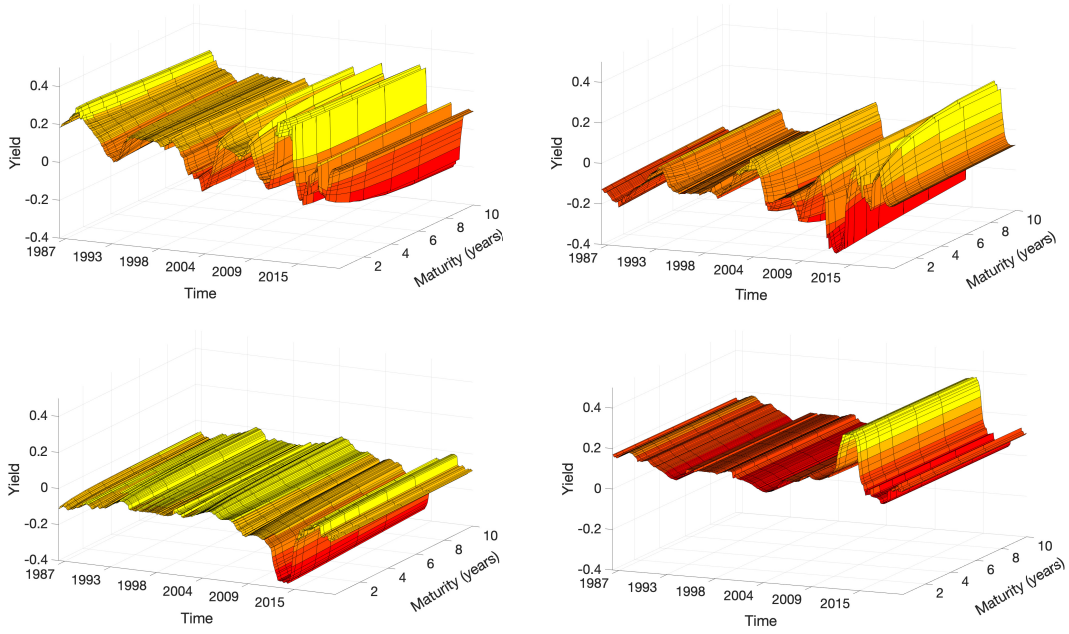
The upper exhibit shows the rolling estimates of the volatility parameters $(\psi_{11}, \psi_{22}, \psi_{33})$ from filtering in the SLSC model. The lower exhibit shows the rolling estimates of the market prices of risk $(\lambda_1, \lambda_2, \lambda_3)$. In each exhibit, the parameter indexed by 1, 2, and 3 is represented by the solid, dashed, and dotted line, respectively.

the ANS-extended Vasicek model. In the first column, the mean of $\tilde{y}$, denoted $\tilde{\mu}$, is restricted to $(\psi_2^2/a + \psi_2\lambda)\tilde{B}_2(\tau;a) - (\psi_2^2/a)\tilde{B}_4(\tau;a)$, cf. (50). The second column leaves $\tilde{\mu}$ free, hence introducing eight mean parameters in place of λ, so the difference in degrees of freedom is seven.[41] While a is similar in the two estimations, $\tilde{\mu}$ is upward-sloping when unrestricted, but nearly flat under no arbitrage, because $a \approx 0$, and the condition is rejected (the LR test takes the value 50.1, compared to a critical value of 14.1 at the 5% level in the χ_7^2-distribution). This confirms the need for further specification searches, and hence the SLSC approach.

The last two columns of Table 5 show results of estimation of the SLSC model with and without the no-arbitrage condition imposed, implying $\tilde{\mu} = \tilde{B}_{1:7}(\tau)\Phi\theta$ in the former case, cf. (49). The two resulting $\tilde{\mu}$ vectors are similar, i.e., the slope is now accommodated. The LR test gets a p-value of 21% in the asymptotic χ_5^2-distribution (the restriction introduces three market prices of risk and drops eight parameters in $\tilde{\mu}$). Thus, the test fails to reject

[41] Under the alternative, a is identified from the volatility structure, only, so this is a straight Vasicek (1977) specification, without extension.

Figure 4: Decomposition of fitted yield curves in the unrestricted and restricted SLSC model



This figure provides three-dimensional views of the stochastic, $y_s(t,\tau)$, and deterministic parts, $y_d(t,\tau)$, from filtering along the dynamically consistent yield curve families, $y(t,\tau) = y_s(t,\tau) + y_d(t,\tau)$, in the rolling (unrestricted) SLSC model and the rolling restricted SLSC model, respectively. The stochastic part in the SLSC model is given by $y_s(t,\tau) = \tilde{B}_{1:3}(\tau)x_{1:3}(t)$ (upper left exhibit), and the deterministic part is given by $y_d(t,\tau) = \tilde{B}_{4:7}(\tau)x_{4:7}(t)$ (upper right exhibit). The stochastic part in the restricted SLSC model is given by $y_s(t,\tau) = \tilde{B}_{1:3}(\tau)x_{1:3}(t)$ (lower left exhibit), and the deterministic part is given by $y_d(t,\tau) = \bar{A}(\tau)$ (lower right exhibit). The fitted factors used to compute $y_s(t,\tau)$ and $y_d(t,\tau)$ are the final fitted values, $x(t) = x(T)$, from each four-year rolling estimation window, $[1,T]$.

at all conventional levels for the SLSC model.

7. Conclusion

We consider generalized duration matching, i.e., removal of factor exposure, as an alternative to traditional immunization. Optimal hedging weights depend on factor loadings and idiosyncratic variances. However, the empirical results indicate that generalized duration matching by itself does not generate a noteworthy improvement in hedging performance, not even when imposing a flexible, parsimonious yield curve shape on loadings. Instead, performance can be enhanced by exploiting dynamic consistency, thus preventing that the hedging strategy relies on information that is in conflict with interest rate dynamics, and hence returns.

Our first empirical approach is to restrict the loadings in a factor model for yields

Table 5: Test of no-arbitrage condition

This table reports results from estimation of the ANS-extended Vasicek and SLSC models for slope-adjusted yield changes, with and without the no-arbitrage condition, $v_\tau = 0$ in (46), imposed. Reported means $\tilde{\mu}$ are in basis points, estimated freely in the unrestricted case, and as functions of the parameters a, ψ, and λ under no arbitrage. The bottom portion of the table shows the value of the maximized log likelihood function, number of parameters, information criteria, and the LR test of the restrictions implied by the no-arbitrage condition.

	ANS-extended Vasicek, $\tilde{y}$ No arbitrage	Vasicek, $\tilde{y}$ Unrestricted	SLSC, $\tilde{y}$ No arbitrage	SLSC, $\tilde{y}$ Unrestricted
a	-0.033	-0.035	0.773	0.751
λ_1			-0.327	
λ_2	-0.623		-0.369	
λ_3			-0.314	
$\tilde{\mu}_1$	-1.070	-1.720	-1.754	-1.720
$\tilde{\mu}_2$	-1.072	-1.651	-1.757	-1.651
$\tilde{\mu}_3$	-1.077	-1.694	-1.728	-1.694
$\tilde{\mu}_4$	-1.087	-1.481	-1.603	-1.481
$\tilde{\mu}_5$	-1.097	-1.355	-1.457	-1.355
$\tilde{\mu}_6$	-1.116	-1.204	-1.210	-1.204
$\tilde{\mu}_7$	-1.134	-1.005	-1.039	-1.005
$\tilde{\mu}_8$	-1.159	-0.780	-0.870	-0.780
logL	90539	90564	95306	95309
# params.	11	18	15	20
AIC	-181056	-181092	-190582	-190579
BIC	-180972	-180954	-190467	-190426
LR	50.088		7.193	
$\chi^2_{0.95}$	14.067		11.070	
p-value	0.000		0.207	

according to a parsimonious curve shape that is dynamically consistent with a suitable dynamic term structure model. The second is to estimate the same restricted loadings in a factor model for returns (or slope-adjusted yield changes), rather than yields, without imposing that the yield curve satisfies a particular shape (e.g., affine, or Nelson-Siegel) at all times. The third approach combines the first two. Thus, the restrictions from the curve shape are imposed on both the cross section and the dynamics, leading to a filtering approach for consecutive yield curves along the dynamically consistent curve family. In our application, the second approach generates stronger hedging performance gains than the first. At least as strong performance is achieved in the third approach by adopting a sufficiently rich dynamic term structure model, in particular, the new SLSC model, such that the dynamically consistent curve shape is flexible enough to capture the current yield curve and associated slope adjustments. This indicates that the yield curve is dynamically

consistent with the stochastic process driving it. The SLSC model generates the strongest hedging performance of all specifications considered, and passes the test of absence of arbitrage opportunities. The results show the importance for practical immunization purposes of an approach that is both dynamically consistent and parsimonious, yet sufficiently general to accommodate the level, slope, and curvature structure of the market. On the other hand, performance deteriorates when the restrictions reducing the general specifications to standard affine form are imposed, or when trading off bias and variance and admitting factor exposure.

Our work paves the way to a number of natural future extensions. One is state-dependent market prices of risk. In this case, our second-stage factor analysis of the slope-adjusted yield changes is replaced by a Kalman filter approach, following Christensen and van der Wel (2019). Second, some of the driving latent processes can be replaced by observable macro series, as in the latter study, but now imposing restrictions on loadings, for parsimony and dynamic consistency. A further generalization would be to allow for state-dependence in the volatilities in the transition equation, hence requiring an extended filter in both the second and third stages. Our theory covers these cases.

Far from mundane, the traditional topic of fixed income immunization has proved related to advanced geometry. It is somewhat eerie that Nelson and Siegel without any particular justification wrote down a curve shape including some of the essential features required for consistency with dynamic term structure models. Thus, the same coefficient $-a$ on maturity was used in both exponents in the proposed functional form. Without this common coefficient restriction, there would be no hope for dynamic consistency with the mean-reverting homoskedastic model. With the restriction, dynamic consistency is achieved by adding one more term to the curve shape, with double coefficient $-2a$ in the exponent. This corresponds to the loading on a new deterministic state variable. The approach is general, and extends to models with multiple deterministic and stochastic state variables. The evidence indicates that the loadings on deterministic state variables do not enter yield curves in fixed proportions over time, hence reinforcing the value of the dynamic consistency approach.

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 (3):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 (4):323–348.

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.

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.

Campbell, J. Y., K. Serfaty-De Medeiros, and L. M. Viceira. 2010. "Global currency hedging." *The Journal of Finance* 65 (1):87–121.

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 (1):89–104.

Cheridito, P., D. Filipović, and R. L. Kimmel. 2007. "Market price of risk specifications for affine models: Theory and evidence." *Journal of Financial Economics* 83 (1):123–170.

Christensen, B. J. and N. M. Kiefer. 2009. *Economic Modeling and Inference*. Princeton: Princeton University Press.

Christensen, B. J. and M. van der Wel. 2019. "An asset pricing approach to testing general term structure models." *Journal of Financial Economics* 134 (1):165–191.

Christensen, Jens 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 (1):4 – 20.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. 1985. "A theory of the term structure of interest rates." *Econometrica* 53:385–408.

Da Prato, G. and J. Zabczyk. 2014. *Stochastic equations in infinite dimensions*. Cambridge University Press.

Dai, Q. and K. Singleton. 2000. "Specification analysis of affine term structure models." *The Journal of Finance* 55 (5):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*, edited 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.

Diebold, F. X. and G. D. Rudebusch. 2013. *Yield Curve Modeling and Forecasting*. Princeton: Princeton University Press.

Duffie, D. and R. Kan. 1996. "A yield-factor model of interest rates." *Mathematical Finance* 6 (4):379–406.

Filipović, D. 1999. "A note on the Nelson-Siegel family." *Mathematical Finance* 9 (4):349–359.

Filipović, D. 2009. *Term-Structure Models: A Graduate Course*. Berlin: Springer.

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 (4):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.

Goliński, A. and P. Spencer. 2017. "The advantages of using excess returns to model the term structure." *Journal of Financial Economics* 125:163–181.

Hansen, P. R., A. Lunde, and J. M. Nason. 2011. "The model confidence set." *Econometrica* 79:453–497.

Heath, D., R. Jarrow, and A. Morton. 1992. "Bond pricing and the term structure of interest rates." *Econometrica* 60:77–106.

Hull, J. and A. White. 1990. "Pricing interest-rate-derivative securities." *The Review of Financial Studies* 3 (4):573–592.

Ingersoll, J. E. 1983. "Is immunization feasible? Evidence from the CRSP data." In *Innovation in bond portfolio management: Duration analysis and immunization*, edited 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 (3):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.

Litterman, R. and J. Scheinkman. 1991. "Common factors affecting bond returns." *The Journal of Fixed Income* 1:54–61.

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., 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*, edited by G. Bierwag, G. Kaufman, and A. Toevs. JAI Press Inc., Greenwich, CT, 61–102.

Opie, W. and S. J. Riddiough. 2020. "Global currency hedging with common risk factors." *Journal of Financial economics* 136 (3):780–805.

Quaedvlieg, R. and P. Schotman. 2020. "Hedging Long-Term Liabilities." *Journal of Financial Econometrics* .

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 (4):417–423.

Soto, G. M. 2004. "Duration models and IRR management: A question of dimensions." *Journal of Banking and Finance* 28:1089–1110.

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

Willner, R. 1996. "A new tool for portfolio managers: Level, slope, and curvature durations." *The Journal of Fixed Income* 6 (1):48–59.

Appendix to

Immunization with consistent term structure dynamics

Daniel Borup[†] Bent Jesper Christensen[**] Jorge W. Hansen[††]

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

[**]Corresponding author. Aarhus University, CREATES, the Dale T. Mortensen Centre, and the Danish Finance Institute. Email: bjchristensen@econ.au.dk.

[††]Aarhus University, CREATES, and the Danish Finance Institute. Email: jh@econ.au.dk.

Content (Appendix)

A Proofs **2**

A.1 Proof of (3)-(4) as an exact relation 2

A.2 Proof of Theorem 1 2

A.3 Interpolation and hedging for general payment streams 4

A.4 The NS yield curve representation 6

A.5 Proofs of Proposition 1, Corollary A.5.1, and (24) 9

A.6 Fixed term to maturity yields 11

A.7 Proofs of Proposition A.7.1 and Corollaries 1, A.7.1, and 2 14

A.8 Affine models 24

A.9 Proofs of Propositions 2, A.9.1, A.9.2, and Theorem 2 25

A.10 Relation between SLSC and AFNS 31

A.11 Proof of Theorem D.1.1 33

B Target Asset **35**

C Basic and Generalized Duration Matching **36**

D Robustness Checks **37**

D.1 Alternative Target Asset 37

D.2 Trading Off Hedging Error Bias and Variance 38

E Statistical Comparison of Hedging Performance **41**

F Estimation **42**

F.1 Factor models 42

F.2 Exact state transition for dynamically consistent specifications 44

F.3 A low-storage square-root filter 47

G Additional Empirical Results **49**

A. Proofs

This Appendix provides the proofs of propositions, theorems, and related results in the paper.

A.1. Proof of (3)-(4) *as an exact relation*

The log excess return is given by

$$r_{t+1,\tau} - y_{t,1} = -\tau \tilde{y}_{t+1,\tau}\,,$$

with $\tilde{y}_{t+1,\tau}$ the slope-adjusted yield change from (4).

Proof of (3).

$$\begin{aligned}
r_{t+1,\tau} - y_{t,1} &= \log p_{t+1,\tau} - \log p_{t,\tau+1} - y_{t,1} \\
&= -\tau y_{t+1,\tau} + (\tau+1) y_{t,\tau+1} - y_{t,1} \\
&= -\tau \left(y_{t+1,\tau} - y_{t,\tau}\right) - \tau y_{t,\tau} + (\tau+1) y_{t,\tau+1} - y_{t,1} \\
&= -\tau \Delta y_{t+1,\tau} + \tau \left(y_{t,\tau+1} - y_{t,\tau}\right) + y_{t,\tau+1} - y_{t,1} \\
&= -\tau \left(\Delta y_{t+1,\tau} - \frac{y_{t,\tau+1} - y_{t,1}}{\tau} - \left(y_{t,\tau+1} - y_{t,\tau}\right)\right) \\
&= -\tau \tilde{y}_{t+1,\tau}\,,
\end{aligned}$$

which is (3). □

A.2. Proof of Theorem 1

Proof of Theorem 1. By (9) and (10), under generalized duration matching, the optimal portfolio solves

$$\min_{w} w' \mathcal{T} \Psi \mathcal{T} w \quad \text{s.t.} \quad w' \mathcal{T} B = (\tau b)_* \,. \tag{A.2.1}$$

This is equivalent to the problem (A.2.2) in Lemma A.2.1 below, with $A = \mathcal{T}\Psi\mathcal{T}$, $g = 0$, $D = B'\mathcal{T}$, and $c = (\tau b)'_*$. Further, in the Lemma, $w_u = A^{-1} g = 0$ because $g = 0$. Thus, by (A.2.3), the solution to (A.2.1) is

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

which is (11) in the Theorem. When the value matching constraint $w'\iota = 1$ is added to the minimization problem (A.2.1), the corrected weights are found from (A.2.4) in Lemma A.2.1,

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

with

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

thus confirming (12). □

Lemma A.2.1. *For a symmetric, positive definite matrix A, and a conformable vector g, let $w_u = A^{-1}g$ be the solution to the unconstrained problem $\min_w (1/2)w'Aw - w'g$. If D has linearly independent rows, then the solution to the constrained problem*

$$\min_w \frac{1}{2}w'Aw - w'g \qquad s.t. \qquad Dw = c \tag{A.2.2}$$

is given by

$$w_c = w_u + A^{-1}D'\left(DA^{-1}D'\right)^{-1}(c - Dw_u). \tag{A.2.3}$$

When further adding the scaling constraint $w'\iota = 1$ to the problem (A.2.2), then the solution is

$$w^* = w_c + \left(1 - w_c'\iota\right)\frac{\Lambda\iota}{\iota'\Lambda\iota}, \tag{A.2.4}$$

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

Proof of Lemma A.2.1. The Lagrangian for (A.2.2) is

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

where ζ contains the Lagrange multipliers. The first order conditions are $Aw_c - g - D'\zeta = 0$, such that

$$w_c = A^{-1}\left(g + D'\zeta\right) = w_u + A^{-1}D'\zeta. \tag{A.2.5}$$

Substituting w_c for w in the constraint $Dw = c$ yields $Dw_u + DA^{-1}D'\zeta = c$, so $\zeta = \left(DA^{-1}D'\right)^{-1}$

$(c - Dw_u)$. Substitution in (A.2.5) gives the solution

$$w_c = w_u + A^{-1}D'\left(DA^{-1}D'\right)^{-1}(c - Dw_u)\,, \tag{A.2.6}$$

which is (A.2.3). When the constraint $w'\iota = 1$ is added to (A.2.2), the new solution can be found by substituting $(D', \iota)'$ for D and $(c', 1)'$ for c in the solution (A.2.6), yielding

$$w^* = w_u + A^{-1}\begin{pmatrix} D' & \iota \end{pmatrix}\begin{pmatrix} DA^{-1}D' & DA^{-1}\iota \\ \iota'A^{-1}D' & \iota'A^{-1}\iota \end{pmatrix}^{-1}\begin{pmatrix} c - Dw_u \\ 1 - \iota'w_u \end{pmatrix}. \tag{A.2.7}$$

By the formula for the inverse of a partitioned matrix,

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

for $F = \left(DA^{-1}D'\right)^{-1}DA^{-1}$, where the Schur complement of $DA^{-1}D'$ in S is $\iota'A^{-1}\iota - \iota'A^{-1}D'$ $\left(DA^{-1}D'\right)^{-1}DA^{-1}\iota = \iota'\Lambda\iota$. Using that $w_c - w_u = F'(c - Dw_u)$, we get

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

Multiplication from the left by $\begin{pmatrix} A^{-1}D' & A^{-1}\iota \end{pmatrix}$ produces the last term in (A.2.7), such that

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

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

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

which is (A.2.4). □

A.3. Interpolation and hedging for general payment streams

First, consider the case that the 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 $1 \times k$ vector) of the target claim (precisely, of the yield to the missing ideal hedge) may be obtained by interpolation between the maturities of

the hedging instruments,

$$b_* = \frac{(\tau_{i+1} - \tau_*) b_i + (\tau_* - \tau_i) b_{i+1}}{\tau_{i+1} - \tau_i} = s(\tau_*) B\,, \tag{A.3.1}$$

with $\tau_i < \tau_* < \tau_{i+1}$, i.e., $s(\tau_*)$ is $1 \times m$, selecting and weighting the appropriate loadings corresponding to maturities adjacent to τ_*. If the future payment occurs before the shortest maturity, $\tau_* < \tau_1$, we set $b_* = b_1$, the loadings for the shortest instrument. Similarly, if $\tau_m < \tau_*$, we set $b_* = b_m$ (other extrapolation schemes could be used). Target generalized durations are now set to $(\tau b)_* = \tau_* b_*$.

Suppose next that the target to be hedged at time t is a stream of 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, obtained by interpolation between observed yields, $y_{t,\tau} = s(\tau) y_t$. By Theorem 1, 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 (11), i.e., $\widetilde{w}_h = \mathcal{T}^{-1} \Psi^{-1} B \left(B' \Psi^{-1} B \right)^{-1} \tau_h b_h'$, with the k-vector of loadings b_h obtained by interpolation as in (A.3.1), $b_h = s(\tau_h) B$. 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 (11) directly to the target payment stream, assessing its generalized duration vector as

$$(\tau b)_* = \sum_{h=1}^{H} \frac{p_{t,\tau_h} c_h}{v_*} \tau_h b_h = \sum_{h=1}^{H} v_{*h} \tau_h b_h, \tag{A.3.2}$$

the value-weighted average of the generalized duration vectors $\tau_h b_h$ of the individual payments, each of dimension $1 \times k$, with value weights $v_{*h} = p_{t,\tau_h} c_h / v_*$. This is a portfolio-of-portfolios argument. Thus, Theorem 1 applies directly to general target payment streams. If value matching is desired, (12) is applied to $\widetilde{w}$.

Finally, the expression (A.3.2) 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 bonds. If there are M coupon-bearing hedging instruments, we use expression (A.3.2) to calculate the generalized duration vectors $(\tau b)_*$ and $(\tau b)_\ell$, $\ell = 1, \ldots, M$ of both the target and each of these instrument

streams. Thus,

$$(\tau b)_\ell = \sum_{h=1}^{H} v_{\ell h}\tau_h b_h = \sum_{h=1}^{H} v_{\ell h}\tau_h s(\tau_h)B = \beta_\ell S(\tau)B\,, \tag{A.3.3}$$

where $v_{\ell h}$ contains the value weights for instrument ℓ, $S(\tau)$ is $H \times m$ with typical row $s(\tau_h)$, and the vector $\beta_\ell = \{v_{\ell h}\tau_h\}_h$ of value weights times maturities is $1 \times H$. Suppose the preceding yield factor analysis for estimation of B and Ψ is still applied to a balanced panel of m zero-coupon bonds. The main difference is that the hedging instruments are now different, and may vary from period to period as bonds age, mature, etc. Thus, the relevant $M \times k$ matrix of generalized durations Ξ has typical row $(\tau b)_\ell$ from (A.3.3). This Ξ is the matrix that specializes to $\mathcal{T}B$ in the zero-coupon instrument case. Further, from the construction (see (6)), the $M \times M$ idiosyncratic (non-factor related) error variance matrix for the coupon bond returns is $\Theta = \Pi\Psi\Pi'$, where Π is $M \times m$ with typical row $\beta_\ell S(\tau)$. With B and Ψ estimated in the zero-coupon yield factor analysis, applied to data from the preceding periods, and all other required variables given by contractual terms and interpolation, hedging is based on a direct generalization of Theorem 1. The hedging portfolio for general instruments is given by

$$\widetilde{w} = \Theta^{-1}\Xi\left(\Xi'\Theta^{-1}\Xi\right)^{-1}(\tau b)'_*\,.$$

Again, (12) is used to obtain value matching (full investment), now with

$$\Lambda = \Theta^{-1} - \Theta^{-1}\Xi\left(\Xi'\Theta^{-1}\Xi\right)^{-1}\Xi'\Theta^{-1}.$$

In case of more hedging instruments than zero-coupon bonds, $M > m$, Θ^{-1} is understood as a (Moore-Penrose) generalized inverse.

A.4. The NS yield curve representation

The NS yield curve parametrization takes the form (14), i.e., the loading functions $\tilde{B}_j(\tau)$, $j = 1,2,3$, are those given in (15). Throughout, we focus on the case $a > 0$, since (15) is undefined for $a = 0$, and for $a < 0$ diverges for large maturities. By direct differentiation, we have the following results.

Lemma A.4.1. *For $a, \tau > 0$, the slopes of the NS loading functions (15) in the maturity*

direction are

$$\frac{\partial}{\partial\tau}\tilde{B}_1(\tau) = 0\,, \tag{A.4.1}$$

$$\frac{\partial}{\partial\tau}\tilde{B}_2(\tau) = -\frac{1}{\tau}\tilde{B}_3(\tau)\,, \tag{A.4.2}$$

$$\frac{\partial}{\partial\tau}\tilde{B}_3(\tau) = -\frac{1}{\tau}\tilde{B}_3(\tau) + a\mathrm{e}^{-a\tau}\,. \tag{A.4.3}$$

It is noted that the slope (A.4.3) of the third loading function is unambiguously greater than that of the second, in (A.4.2).

The loading functions $\tilde{B}_j(\tau)$ are associated with the fixed term yield parametrization of the term structure, $y(t,\tau)$. Alternative parametrizations include those by the fixed maturity date spot yields $s(t,T)$, with T the maturity date, the instantaneous forward rates $f(t,T)$, or the fixed term to maturity forward rates $F(t,\tau) = f(t,t+\tau)$. The relations between these follow from writing the price at t of the zero-coupon bond maturing at $T > t$ as

$$p(t,T) = \mathrm{e}^{-(T-t)\,y(t,T-t)} = \mathrm{e}^{-(T-t)\,s(t,T)} = \mathrm{e}^{-\int_t^T f(t,u)\,\mathrm{d}u} = \mathrm{e}^{-\int_0^{T-t} F(t,v)\,\mathrm{d}v}\,. \tag{A.4.4}$$

In the latter parametrization, by fixed term forward rates, the NS curve is the well known

$$F(t,\tau) = f_{t,1} + f_{t,2}e^{-a\tau} + f_{t,3}a\tau e^{-a\tau}\,. \tag{A.4.5}$$

It is evident that $f_{t,1}$, $f_{t,2}$, and $f_{t,3}$ are level, slope, and curvature factors in this parametrization, since the respective loadings are constant, exponentially declining for $a > 0$, and hump shaped, i.e., the linear component dominates the third loading function for short maturities τ, and the exponential for long. It is less obvious that $f_{t,j}$ are level, slope, and curvature factors in the yield parametrization (14), too. To be sure, we first clarify that the representations are equivalent.

Corollary A.4.1. *The NS yield curve* (14) *and forward rate curve* (A.4.5) *are equivalent representations of the term structure.*

Proof of Corollary A.4.1 . By (A.4.4), with term to maturity $\tau = T - t$, we have $\tau y(t,\tau) = \int_0^\tau F(t,v)\,\mathrm{d}v$. Differentiation with respect to τ produces

$$F(t,\tau) = y(t,\tau) + \tau\frac{\partial y(t,\tau)}{\partial\tau}\,. \tag{A.4.6}$$

In the NS case, inserting $y(t,\tau) = \tilde{B}_{1:3}(\tau) f_t$,

$$F(t,\tau) = \left(\tilde{B}_{1:3}(\tau) + \tau \frac{\partial \tilde{B}_{1:3}(\tau)}{\partial \tau} \right) f_t \,. \tag{A.4.7}$$

The right side is evaluated separately for each of the three loading functions $\tilde{B}_{1:3}(\tau)$ from (15). For the second, $\tilde{B}_2(\tau)$, using (A.4.2), the component multiplying $f_{t,2}$ on the right side of (A.4.7) is

$$\tilde{B}_2(\tau) + \tau \left(-\frac{1}{\tau} \tilde{B}_3(\tau) \right) = \tilde{B}_2(\tau) - \tilde{B}_3(\tau) = e^{-a\tau} \,,$$

confirming that the second loading function in the (fixed term) yield parametrization, $\tilde{B}_2(\tau)$, transforms to the second loading in the (fixed term) forward rate parametrization (A.4.5). Similarly, for $\tilde{B}_3(\tau)$, and using (A.4.3),

$$\tilde{B}_3(\tau) + \tau \left(-\frac{1}{\tau} \tilde{B}_3(\tau) + a \mathrm{e}^{-a\tau} \right) = a\tau e^{-a\tau} \,, \tag{A.4.8}$$

i.e., the third loading function in the yield and forward rate parametrizations correspond, too. Finally, the first (level) loading is flat, cf. (A.4.1), hence common. □

The qualitative interpretation of the factors in the NS yield parametrization in terms of level, slope, and curvature corresponds to that in the equivalent forward rate representation, as we show in the next corollary, where we further collect some results from the analysis.

Corollary A.4.2. *For $a, \tau > 0$, the factors $f_{t,1}$, $f_{t,2}$, and $f_{t,3}$ in the NS yield curve (14) are level, slope, and curvature factors, i.e., the respective loading functions are flat, downward sloping, and hump shaped. All three loading functions are positive. The third loading function is smaller than the second, but has greater slope.*

Proof of Corollary A.4.2 . The first is obvious, $f_{t,1}$ is a level factor, since the associated loading function is flat, cf. (A.4.1). Next, write the third loading function as

$$\tilde{B}_3(\tau) = \frac{1 - \mathrm{e}^{-a\tau}}{a\tau} - \mathrm{e}^{-a\tau} = \frac{1 - (1 + a\tau)\mathrm{e}^{-a\tau}}{a\tau} \,. \tag{A.4.9}$$

Recall that the exponential function $x \to \mathrm{e}^x$ is strictly convex, so its graph lies above its tangent at $x = 0$, given by the first order Taylor approximation, which is $1 + x$. Thus, $\mathrm{e}^x > 1 + x$ for $x > 0$, so $(1 + x)\mathrm{e}^{-x} < 1$. Using this with $x = a\tau$ in (A.4.9) shows that $\tilde{B}_3(\tau) > 0$.

From (14), $\tilde{B}_2(\tau) = \tilde{B}_3(\tau) + e^{-a\tau} > \tilde{B}_3(\tau) > 0$, so all three loading functions are positive. From (A.4.2), since $\tilde{B}_3(\tau) > 0$, $\tilde{B}_2(\tau)$ is downward sloping, so $f_{t,2}$ is a slope factor. Comparing (A.4.2) and (A.4.3), the slope of $\tilde{B}_3(\tau)$ exceeds that of $\tilde{B}_2(\tau)$ by $ae^{-a\tau} > 0$. It remains to show that $f_{t,3}$ is a curvature factor. Use (A.4.3) and (A.4.9) to write the derivative of the third loading function as

$$\frac{\partial}{\partial\tau}\tilde{B}_3(\tau) = -\frac{1}{\tau}\left(\frac{1-(1+a\tau)e^{-a\tau}}{a\tau}\right) + ae^{-a\tau} = \frac{(1+a\tau+a^2\tau^2)e^{-a\tau}-1}{a\tau^2}. \tag{A.4.10}$$

Signing this involves comparing $e^{a\tau}$ to $1+a\tau+a^2\tau^2$, rather than to $1+a\tau$, as in (A.4.9), so convexity no longer suffices. For large τ, the exponential does dominate, so (A.4.10) is negative. On the other hand, for $\tau > 0$ sufficiently near 0, the quadratic dominates, and (A.4.10) is strictly positive. This is seen by applying l'Hôpital's rule twice to (A.4.10), producing a ratio $a/2 > 0$ at $\tau = 0$. Thus, $\tilde{B}_3(\tau)$ has an interior maximum, or a hump. □

By the corollary, $f_{t,1}$ is a level factor, and changes in this induces parallel shifts in the NS yield curve. Movement in $f_{t,2}$ changes the slope, and $f_{t,3}$ governs curvature, just as in the forward rate parametrization.

A.5. Proofs of Proposition 1, Corollary A.5.1, and (24)

Proof of Proposition 1. Under the stated assumptions, dynamic consistency implies that if the evolution of $y(t,\tau)$ is governed by (α,σ), as in (16), then $y(t,\tau) = Y(\tau,x(t))$, for $t \geq T_y$, with $x(t)$ governed by (18), and the latter has a strong solution. Applying Itô's lemma to $y(t,\tau) = Y(\tau,x(t))$ produces (22)-(23). Conversely, given ϕ, ψ such that (22)-(23) hold, and (18) has a strong solution, consider $y(t,\tau)$ governed by (α,σ), as in (16), for $t \geq T_y$. By (22), (23) and (18), $y(t,\tau)$ is represented in the form $Y(\tau,x(t))$. □

Corollary A.5.1. *Let $\phi(x(t),a(t))$, a 3×1 vector, $\phi_4(x(t),a(t))$, a scalar, and $\psi(x(t),a(t))$, a $d\times 3$ matrix,* $\text{rank}(\psi) = d = \dim(W(t))$*, be such that (18) along with* $da(t) = \phi_4(x(t),a(t))dt$ *has a solution $(x(t),a(t))$, but otherwise arbitrary. Then the NS curve shape (19) is dynami-*

cally consistent with the DTSM (16) *with*

$$\begin{aligned}\alpha(\tau,x(t),a(t)) &= \phi_4(x(t),a(t))\left[\tau e^{-a(t)\tau}x_3(t) - \frac{1}{a(t)}\tilde{B}_3(\tau,a(t))(x_2(t)+x_3(t))\right] \\ &\quad + \tilde{B}(\tau,a(t))\phi(x(t),a(t)), \qquad \text{(A.5.1)} \\ \sigma(\tau,x(t),a(t))' &= \tilde{B}(\tau,a(t))\psi(x(t),a(t)).\end{aligned}$$

The corollary provides a non-trivial example of a DTSM (α,σ) that generates NS curves. In the special case $\phi_4 = 0$ and ϕ, ψ constant, the discrete time version is the DNS model considered by Diebold and Li (2006), who emphasize the constant $a(t) = a$. The model in Corollary A.5.1 is more general, and for non-zero ϕ_4, the coefficient $a(t)$ in the exponent in NS is not constant. Thus, this is not a factor model.

Proof of Corollary A.5.1. For the coefficients ϕ, ψ in Proposition 1, consider those from the corollary, namely, $(\phi',\phi_4)'$ and $(\psi,0)$, appending a fourth column of zeroes to the latter. In this case, the trace term vanishes in (22), since nonlinearity of $Y(\tau,x)$ only enters via the locally deterministic $a(t)$, not x_i, $i = 1,2,3$. Inserting (20)-(21) for $\partial Y/\partial x'$ in (22)-(23) produces the drift and volatility in (A.5.1). The result follows from Proposition 1. □

Proof of (24). Writing the price of the zero-coupon bond trading at t and maturing at T as $p(t,T) = G(t,y(t,T-t))$, with $G(t,y) = \exp(-(T-t)y)$, we have $\partial G/\partial t = yp$, $\partial G/\partial y = -(T-t)p$, and $\partial^2 G/\partial y^2 = (T-t)^2 p$. By the chain rule, $dG(t,y(t,T-t))/dt = \partial G/\partial t - \partial G/\partial y \cdot \partial y/\partial \tau$, with $\tau = T-t$. Combining, we have $dG(t,y(t,T-t))/dt = yp + \tau p\partial y/\partial\tau$. By Itô's lemma,

$$\begin{aligned}dp(t,T) &= \left(y(t,\tau) + \tau\frac{\partial y}{\partial\tau}(t,\tau) - \tau a(t,\tau) + \frac{\tau^2}{2}\sigma(t,\tau)'\sigma(t,\tau)\right)p(t,T)dt - \tau p(t,T)\sigma(t,\tau)'dW_t \\ &= p(t,T)\alpha_p(t,T)dt + p(t,T)\sigma_p(t,T)'dW_t, \qquad \text{(A.5.2)}\end{aligned}$$

where $\alpha_p(t,T)$ and $\sigma_p(t,T)$ are the expected return and return volatility of the bond. By no arbitrage, we have

$$\alpha_p(t,T) = y(t,0) + \sigma_p(t,T)'\lambda_t, \qquad \text{(A.5.3)}$$

for suitable market prices of risk λ_t. Inserting the expressions for $\alpha_p(t,T)$ and $\sigma_p(t,T)$ from (A.5.2) in (A.5.3) and solving, the no-arbitrage condition on the yield drift is

$$\alpha(t,\tau) = \frac{1}{\tau}\left[y(t,\tau) - y(t,0)\right] + \frac{\partial y}{\partial\tau}(t,\tau) + \frac{\tau}{2}\sigma(t,\tau)'\sigma(t,\tau) + \sigma(t,\tau)'\lambda_t. \qquad \text{(A.5.4)}$$

For $y(t,\tau) = Y(\tau, x)$ and $\lambda_t = \lambda(x)$, this is (24). □

A.6. Fixed term to maturity yields

We verify the claims in footnote 12 in the main text, namely, that (a) the yield spread enters (24) because we consider yields, as opposed to forward rates (as in HJM), (b) for the same reason, convexity in (24) replaces the term involving an integral in HJM, (c) the local slope enters (24) because our fixed term to maturity analysis avoids the bond aging effect, and (d) our no-arbitrage condition (24) is consistent with that in HJM, with risk compensation appearing under the physical measure. To this end, we derive the relations between drifts and volatilities of fixed term to maturity yields, $y(t,\tau)$, fixed maturity date spot yields, $s(t,T)$, and instantaneous forward rates, $f(t,T)$. In the main text, we consider an HJM framework for the fixed term to maturity yields. Available panel data sets are of the fixed term to maturity type. Equation (16) is restated as

$$\mathrm{d}y(t,\tau) = \alpha_y(t,\tau)\,\mathrm{d}t + \sigma_y(t,\tau)'\,\mathrm{d}W_t\,, \tag{A.6.1}$$

with subscript y on the drift and volatility functions highlighting that these are for the fixed term to maturity yield specification.

From (A.4.4), we have $(T-t)y(t,T-t) = (T-t)s(t,T) = \int_t^T f(t,u)\mathrm{d}u$. Differentiating with respect to T, we have in analogy with (A.4.6) that

$$f(t,T) = y(t,T-t) + (T-t)\frac{\partial y}{\partial \tau}(t,T-t) = s(t,T) + (T-t)\frac{\partial s}{\partial T}(t,T)\,. \tag{A.6.2}$$

The analysis in HJM is cast in terms of forward rates,

$$\mathrm{d}f(t,T) = \alpha_f(t,T)\,\mathrm{d}t + \sigma_f(t,T)'\,\mathrm{d}W_t\,. \tag{A.6.3}$$

Clearly, α_y and σ_y in (A.6.1) differ from α_f and σ_f in (A.6.3), both because the former coefficients are for yields, as opposed to forward rates, and because we consider fixed terms to maturity τ in (A.6.1), but fixed maturity dates T in (A.6.3), i.e., $T-t$ shrinks as t increases in the latter case. The first of these two differences arises for fixed maturity date spot yields, too. To isolate this effect, use (A.4.4) to write

$$s(t,T) = \frac{1}{T-t}\int_t^T f(t,u)\,\mathrm{d}u\,. \tag{A.6.4}$$

By letting $T \downarrow t$, we have $s(t,t) = f(t,t)$, the short spot and forward rates coincide. Since t enters in three places on the right side of (A.6.4), there are three terms in the stochastic differential. Using Leibniz' rule for the second term, we get

$$\begin{aligned} \mathrm{d}s(t,T) &= \frac{1}{(T-t)^2}\int_t^T f(t,u)\,\mathrm{d}u\,\mathrm{d}t - \frac{1}{T-t}f(t,t)\,\mathrm{d}t + \frac{1}{T-t}\int_t^T \mathrm{d}f(t,u)\,\mathrm{d}u \\ &= \frac{1}{T-t}\left(s(t,T)-s(t,t)\right)\mathrm{d}t + \frac{1}{T-t}\int_t^T \alpha_f(t,u)\,\mathrm{d}u\,\mathrm{d}t + \frac{1}{T-t}\int_t^T \sigma_f(t,u)\,\mathrm{d}u\,\mathrm{d}W_t\,, \end{aligned}$$

where the second equality follows from (A.6.4), (A.6.3), $s(t,t) = f(t,t)$, and Fubini. Thus, in the representation

$$\mathrm{d}s\,(t,T) = \alpha_s\,(t,T)\,\mathrm{d}t + \sigma_s\,(t,T)'\,\mathrm{d}W_t\,,$$

the spot and forward drifts and volatilities are related as

$$\begin{aligned} \alpha_s(t,T) &= \frac{s(t,T)-s(t,t)}{T-t} + \frac{1}{T-t}\int_t^T \alpha_f(t,u)\,\mathrm{d}u\,, \\ \sigma_s(t,T) &= \frac{1}{T-t}\int_t^T \sigma_f(t,u)\,\mathrm{d}u\,. \end{aligned} \tag{A.6.5}$$

By (A.6.4), moving from forward rates to spot yields clearly involves an integration, and by (A.6.5), the drift is in addition adjusted for the average slope, or yield spread. This verifies (a).

For fixed term yields $y(t,\tau)$, there is an additional adjustment. To see this, use (A.4.4) for fixed term $\tau = T-t$ to write

$$y(t,\tau) = \frac{1}{\tau}\int_t^{t+\tau} f(t,u)\,\mathrm{d}u\,. \tag{A.6.6}$$

Again, as in (A.6.4), t enters three times on the right side of (A.6.6). Differentiating first in the upper limit of the integral, then the lower, and then the integrand, we have

$$\begin{aligned} \mathrm{d}y(t,\tau) &= \frac{1}{\tau}f(t,t+\tau)\,\mathrm{d}t - \frac{1}{\tau}f(t,t)\,\mathrm{d}t + \frac{1}{\tau}\int_t^{t+\tau} \mathrm{d}f(t,u)\,\mathrm{d}u \\ &= \frac{1}{\tau}\left(y(t,\tau) + \tau\,\frac{\partial y}{\partial \tau}(t,\tau)\right)\mathrm{d}t - \frac{1}{\tau}y(t,0)\,\mathrm{d}t + \frac{1}{\tau}\int_t^{t+\tau} \mathrm{d}f(t,u)\,\mathrm{d}u \\ &= \frac{1}{\tau}\left(y(t,\tau)-y(t,0)\right)\mathrm{d}t + \frac{\partial y}{\partial \tau}(t,\tau)\,\mathrm{d}t + \frac{1}{\tau}\int_t^{t+\tau} \alpha_f(t,u)\,\mathrm{d}u\,\mathrm{d}t + \frac{1}{\tau}\int_t^{t+\tau} \sigma_f(t,u)\,\mathrm{d}u\,\mathrm{d}W_t\,, \end{aligned}$$

where the second equality follows from (A.6.2). Thus, the fixed term yield and forward

drifts and volatilities are related as

$$\begin{aligned}\alpha_y(t,\tau) &= \frac{1}{\tau}(y(t,\tau)-y(t,0)) + \frac{\partial y}{\partial \tau}(t,\tau) + \frac{1}{\tau}\int_t^{t+\tau} \alpha_f(t,u)\,\mathrm{d}u\,,\\ \sigma_y(t,\tau) &= \frac{1}{\tau}\int_t^{t+\tau} \sigma_f(t,u)\,\mathrm{d}u\,.\end{aligned} \tag{A.6.7}$$

In addition to the adjustment for the average slope or yield spread, shared with that for fixed maturity date yields in (A.6.5), the move to fixed term yields involves a further adjustment in the drift, by the local slope of the yield curve, $\partial y/\partial\tau$. The reason that this is not present in (A.6.5) is the bond aging effect noted by Litterman and Scheinkman (1991), i.e., as t increases, the bond $p(t,T)$ becomes shorter (it ages). Hence, so does $s(t,T)$, but not $y(t,\tau)$ in the fixed term to maturity panel. This verifies (c).

Next, we show that the no-arbitrage yield drift condition (24) is consistent with that given in HJM for forward rates under the risk-neutral measure. Writing $\alpha(t,\tau)=\alpha_y(t,\tau)$, $\sigma(t,\tau)=\sigma_y(t,\tau)$ in the no-arbitrage condition (A.5.4) and comparing with (A.6.7), we have

$$\frac{1}{T-t}\int_t^T \alpha_f(t,u)\,\mathrm{d}u = \frac{\tau}{2}\sigma_y(t,\tau)'\sigma_y(t,\tau) + \sigma_y(t,\tau)'\lambda_t\,, \tag{A.6.8}$$

with $\tau = T-t$. Under the risk-neutral measure, $\lambda_t = 0$. In this case, isolating the HJM forward rate drift from (A.6.8),

$$\begin{aligned}\alpha_f(t,T) &= \frac{\mathrm{d}}{\mathrm{d}T}\int_t^T \alpha_f(t,u)\,\mathrm{d}u = \frac{\mathrm{d}}{\mathrm{d}T}\left(\frac{(T-t)^2}{2}\sigma_y(t,T-t)'\sigma_y(t,T-t)\right)\\
&= (T-t)\sigma_y(t,T-t)'\sigma_y(t,T-t) + (T-t)^2\sigma_y(t,T-t)'\frac{\mathrm{d}}{\mathrm{d}T}\sigma_y(t,T-t)\\
&= (T-t)\sigma_y(t,T-t)'\sigma_y(t,T-t)\\
&\quad + (T-t)^2\sigma_y(t,T-t)'\frac{\mathrm{d}}{\mathrm{d}T}\left(\frac{1}{T-t}\int_t^T \sigma_f(t,u)\,\mathrm{d}u\right)\\
&= (T-t)\sigma_y(t,T-t)'\sigma_y(t,T-t)\\
&\quad + (T-t)^2\sigma_y(t,T-t)'\left(-\frac{1}{(T-t)^2}\int_t^T \sigma_f(t,u)\,\mathrm{d}u + \frac{1}{T-t}\sigma_f(t,T)\right)\\
&= (T-t)\sigma_y(t,T-t)'\sigma_y(t,T-t)\\
&\quad + (T-t)^2\sigma_y(t,T-t)'\left(-\frac{1}{T-t}\sigma_y(t,T-t) + \frac{1}{T-t}\sigma_f(t,T)\right)\\
&= (T-t)^2\sigma_y(t,T-t)'\left(\frac{1}{T-t}\sigma_f(t,T)\right)\\
&= (T-t)\int_t^T \sigma_f(t,u)'\,\mathrm{d}u\left(\frac{1}{T-t}\sigma_f(t,T)\right)\end{aligned}$$

$$= \sigma_f(t,T)' \int_t^T \sigma_f(t,u)\,\mathrm{d}u\,, \tag{A.6.9}$$

using (A.6.7) in the third, fifth, and seventh equality. This reproduces the HJM no-arbitrage forward rate drift condition under the risk-neutral measure, cf. Footnote 12, which is therefore consistent with the no-arbitrage yield drift condition (A.5.4), and hence (24). This verifies (d).

Finally, as noted in relation to (A.6.4), moving from forward rates to yields involves an integration, and this is why both yield drifts (fixed maturity date, (A.6.5), and fixed term, (A.6.7)) involve the integrated forward rate drift. From (A.6.9), under the risk-neutral measure, the forward rate drift is exactly the term involving an integral in the HJM forward rate condition, and from (A.6.8), upon integration in the yield case, this term is simply convexity. This completes (b).

A.7. Proofs of Proposition A.7.1 and Corollaries 1, A.7.1, and 2

Proposition A.7.1. *Dynamic consistency between the arbitrage-free DTSM (λ,σ) and the yield curve family $\mathcal{Y}$ is equivalent to the existence of suitable ϕ, ψ satisfying Assumption 1, condition (23), and*

$$\begin{aligned} &\frac{1}{\tau}\left[Y(\tau,x) - Y(0,x)\right] + \frac{\partial Y}{\partial \tau}(\tau,x) + \frac{\tau}{2}\sigma(\tau,x)'\sigma(\tau,x) + \sigma(\tau,x)'\lambda(x) \\ &\qquad = \frac{\partial Y}{\partial x'}(\tau,x)\phi(x) + \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 Y}{\partial x \partial x'}(\tau,x)\psi(x)'\psi(x)\right), \end{aligned} \tag{A.7.1}$$

for all (τ,x).

Proof of Proposition A.7.1. This follows from Proposition 1 by inserting the no-arbitrage condition (24) for the drift in (22). □

Proof of Corollary 1. We first prove the result for NS with fixed a. Subsequently, we extend the proof to include a among the time-varying state variables. Thus, suppose first that an arbitrage-free DTSM (λ,σ) is dynamically consistent with NS with fixed a. We show that this leads to the condition $\sigma = 0$, hence implying that all arbitrage-free DTSMs that are non-degenerate, i.e., with $\sigma \neq 0$, are dynamically inconsistent with the NS curve

shape with fixed a.

For NS with fixed a, the state vector is $x = (x_1, x_2, x_3)'$, and we have the factor model $Y(\tau, x) = \tilde{B}(\tau)x$, with $\tilde{B}(\tau) = \tilde{B}_{1:3}(\tau; a)$ from (15). By Proposition A.7.1, dynamic consistency under the no-arbitrage condition requires the conditions (23) and (A.7.1), with $\partial Y/\partial x' = \tilde{B}$. In (A.7.1), the trace term vanishes due to the factor structure, $\partial^2 Y/\partial x \partial x' = \partial \tilde{B}/\partial x' = 0$. Condition (23) requires

$$\sigma(\tau, x)' = \tilde{B}(\tau)\psi(x)', \tag{A.7.2}$$

with $\psi(x)$ $d \times 3$. This implies that convexity takes the form $\frac{\tau}{2}\tilde{B}(\tau)\psi(x)'\psi(x)\tilde{B}(\tau)'$. Inserting this in (A.7.1), together with (A.7.2) for $\sigma(\tau, x)'$, average yield spread $\frac{1}{\tau}\big[Y(\tau) - Y(0)\big] = \frac{1}{\tau}\big[\tilde{B}(\tau) - \tilde{B}(0)\big]x$, local slope adjustment $\partial Y/\partial \tau = (\partial \tilde{B}/\partial \tau)x$, and dropping the trace term, (A.7.1) reduces to

$$\frac{1}{\tau}\big[\tilde{B}(\tau) - \tilde{B}(0)\big]x + \frac{\partial \tilde{B}}{\partial \tau}(\tau)x + \frac{\tau}{2}\tilde{B}(\tau)\psi(x)'\psi(x)\tilde{B}(\tau)' = \tilde{B}(\tau)\big(\phi(x) - \psi(x)'\lambda(x)\big), \tag{A.7.3}$$

with $\phi(x)$ a 3×1 vector and $\lambda(x)$ a $d \times 1$ vector. Condition (A.7.3) requires that convexity, when viewed as a function of maturity τ, be spanned by the loadings (15) and the functions in the slope adjustments. To compute $\tilde{B}(0)$ as a limit for $\tau \downarrow 0$, we apply l'Hôpital's rule to $\tilde{B}_2(\tau) = (1 - e^{-a\tau})/a\tau$. Differentiating in numerator and denominator separately produces $e^{-a\tau}$, so $\tilde{B}_2(0) = 1$. Using this to find $\tilde{B}_3(0)$, too, we have $\tilde{B}(0) = (1, 1, 0)$. Thus, in (A.7.3), the contributions from $\tilde{B}$ to the yield spreads are

$$\begin{aligned}\frac{1}{\tau}\big[\tilde{B}(\tau) - \tilde{B}(0)\big] &= \frac{1}{\tau}\begin{pmatrix} \tilde{B}_1(\tau) - 1 & \tilde{B}_2(\tau) - 1 & \tilde{B}_3(\tau) - 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & \frac{1 - e^{-a\tau}}{a\tau^2} - \frac{1}{\tau} & \frac{1 - e^{-a\tau}}{a\tau^2} - \frac{e^{-a\tau}}{\tau} \end{pmatrix}. \end{aligned} \tag{A.7.4}$$

The contributions from $\tilde{B}$ to the local slope adjustments in (A.7.3) are obtained from (15) and Lemma A.4.1,

$$\begin{aligned}\frac{\partial \tilde{B}}{\partial \tau}(\tau) &= \begin{pmatrix} 0 & -\frac{1}{\tau}\tilde{B}_3(\tau) & -\frac{1}{\tau}\tilde{B}_3(\tau) + a e^{-a\tau} \end{pmatrix} \\ &= \begin{pmatrix} 0 & \frac{e^{-a\tau}}{\tau} - \frac{1 - e^{-a\tau}}{a\tau^2} & \frac{e^{-a\tau}}{\tau} - \frac{1 - e^{-a\tau}}{a\tau^2} + a e^{-a\tau} \end{pmatrix}. \end{aligned} \tag{A.7.5}$$

Write $\psi_i(x)$ for the i^{th} column in $\psi(x)$, a $d \times 1$ vector, $\omega_{ij}(x) = \psi_i(x)'\psi_j(x)$, and $\omega(x) =$

$\{\omega_{ij}(x)\}_{i,j}$ for $\psi(x)'\psi(x)$, a 3×3 matrix. Convexity $\frac{\tau}{2}\tilde{B}(\tau)\psi(x)'\psi(x)\tilde{B}(\tau)'$ in (A.7.3) is then

$$\frac{\tau}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\omega_{ij}(x)\tilde{B}_i(\tau)\tilde{B}_j(\tau)=\frac{\tau}{2}\left(\omega_{11}(x)+\omega_{22}(x)\tilde{B}_2(\tau)^2+\omega_{33}(x)\tilde{B}_3(\tau)^2\right) + \tau\left(\omega_{12}(x)\tilde{B}_2(\tau)+\omega_{13}(x)\tilde{B}_3(\tau)+\omega_{23}(x)\tilde{B}_2(\tau)\tilde{B}_3(\tau)\right). \tag{A.7.6}$$

Because we will use the calculations repeatedly, we present them explicitly here. First, in (A.7.6),

$$\tilde{B}_2(\tau)^2=\left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau}\right)^2=\frac{1-2\mathrm{e}^{-a\tau}+\mathrm{e}^{-2a\tau}}{a^2\tau^2}=\frac{2(1-\mathrm{e}^{-a\tau})-\left(1-\mathrm{e}^{-2a\tau}\right)}{a^2\tau^2}=\frac{2}{a\tau}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right), \tag{A.7.7}$$

hence introducing the function

$$\tilde{B}_4(\tau)=\frac{1-\mathrm{e}^{-2a\tau}}{2a\tau}, \tag{A.7.8}$$

which will play an important role in the analysis. Next, from (15) we clearly have

$$\tilde{B}_3(\tau)=\tilde{B}_2(\tau)-\mathrm{e}^{-a\tau}, \tag{A.7.9}$$

so in (A.7.6) we will need the product

$$\tilde{B}_2(\tau)\mathrm{e}^{-a\tau}=\left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau}\right)\mathrm{e}^{-a\tau}=\frac{\mathrm{e}^{-a\tau}-\mathrm{e}^{-2a\tau}}{a\tau}=\frac{2\left(1-\mathrm{e}^{-2a\tau}\right)}{2a\tau}-\frac{(1-\mathrm{e}^{-a\tau})}{a\tau}=2\tilde{B}_4(\tau)-\tilde{B}_2(\tau). \tag{A.7.10}$$

Thus, using (A.7.9) and the calculations (A.7.7) and (A.7.10) for the next term in convexity (A.7.6),

$$\tilde{B}_3(\tau)^2=\frac{2}{a\tau}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)+\mathrm{e}^{-2a\tau}-2\left(2\tilde{B}_4(\tau)-\tilde{B}_2(\tau)\right). \tag{A.7.11}$$

Using (A.7.7) and (A.7.10), an alternate version of (A.7.11) is

$$\tilde{B}_3(\tau)^2=\frac{2(1-\mathrm{e}^{-a\tau})-\left(1-\mathrm{e}^{-2a\tau}\right)}{a^2\tau^2}+\mathrm{e}^{-2a\tau}-\frac{2\left(\mathrm{e}^{-a\tau}-\mathrm{e}^{-2a\tau}\right)}{a\tau}. \tag{A.7.12}$$

For the last term in (A.7.6), we need but a portion of (A.7.11) (or (A.7.12)), i.e., using (A.7.7) and (A.7.10) again,

$$\begin{aligned}\tilde{B}_2(\tau)\tilde{B}_3(\tau)&=\tilde{B}_2(\tau)\left(\tilde{B}_2(\tau)-\mathrm{e}^{-a\tau}\right)\\&=\frac{2}{a\tau}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)-\left(2\tilde{B}_4(\tau)-\tilde{B}_2(\tau)\right)\end{aligned}$$

$$= \frac{2(1-e^{-a\tau}) - \left(1-e^{-2a\tau}\right)}{a^2\tau^2} - \frac{e^{-a\tau} - e^{-2a\tau}}{a\tau}. \tag{A.7.13}$$

Combining (15), (A.7.6), (A.7.7), (A.7.12), and (A.7.13), convexity in (A.7.3) is

$$\begin{aligned}\frac{\tau}{2}\tilde{B}(\tau)\psi(x)'\psi(x)\tilde{B}(\tau)' = {}& \omega_{11}(x)\frac{\tau}{2} + \frac{\omega_{22}(x)}{a}\left(\frac{1-e^{-a\tau}}{a\tau} - \frac{1-e^{-2a\tau}}{2a\tau}\right) \\ & + \omega_{33}(x)\left[\frac{1}{a}\left(\frac{1-e^{-a\tau}}{a\tau} - \frac{1-e^{-2a\tau}}{2a\tau}\right) + \frac{\tau}{2}e^{-2a\tau} - \frac{1}{a}\left(e^{-a\tau} - e^{-2a\tau}\right)\right] \\ & + \omega_{12}(x)\frac{1-e^{-a\tau}}{a} + \omega_{13}(x)\left(\frac{1-e^{-a\tau}}{a} - \tau e^{-a\tau}\right) \\ & + \omega_{23}(x)\left[\frac{2}{a}\left(\frac{1-e^{-a\tau}}{a\tau} - \frac{1-e^{-2a\tau}}{2a\tau}\right) - \frac{1}{a}\left(e^{-a\tau} - e^{-2a\tau}\right)\right]. \end{aligned} \tag{A.7.14}$$

For dynamic consistency, the spanning condition (A.7.3) requires that convexity (A.7.14), as a function of maturity τ, be spanned by the loadings $\tilde{B}(\tau)$ from (15) and the functions in the slope adjustments (A.7.4)-(A.7.5). The spanning coefficents are x, the state variables, on the slope adjustments, and $\phi(x) - \psi(x)'\lambda(x)$, risk-adjusted state drifts, on the loadings.

The first term in convexity (A.7.14), multiplying ω_{11}, is the linear, $\tau/2$. Evidently, this is linearly independent of the functions of τ in the remaining convexity terms in (A.7.14), in the loadings (15), and in the slope adjustments (A.7.4)-(A.7.5), since these are all spanned by exponential functions, the reciprocal, $1/\tau$, the constant function, products of these, and the linear-exponential functions $\tau e^{-a\tau}$ and $\tau e^{-2a\tau}/2$ appearing in (A.7.14) (recall $a > 0$). This implies $\omega_{11}(x) = 0$, so $\psi_1(x) = 0$. Thus, the first state variable is locally deterministic, and $\omega_{1j}(x) = \psi_1(x)'\psi_j(x) = 0$, $j = 2,3$, too. Only the terms multiplying $\omega_{22}(x)$, $\omega_{33}(x)$, and $\omega_{23}(x)$ remain in convexity (A.7.14), which therefore simplifies to

$$\begin{aligned}\frac{\tau}{2}\tilde{B}(\tau)\psi(x)'\psi(x)\tilde{B}(\tau)' = {}& \frac{\omega_{22}(x)}{a}\left(\frac{1-e^{-a\tau}}{a\tau} - \frac{1-e^{-2a\tau}}{2a\tau}\right) \\ & + \omega_{33}(x)\left[\frac{1}{a}\left(\frac{1-e^{-a\tau}}{a\tau} - \frac{1-e^{-2a\tau}}{2a\tau}\right) + \frac{\tau}{2}e^{-2a\tau} - \frac{1}{a}\left(e^{-a\tau} - e^{-2a\tau}\right)\right] \\ & + \omega_{23}(x)\left[\frac{2}{a}\left(\frac{1-e^{-a\tau}}{a\tau} - \frac{1-e^{-2a\tau}}{2a\tau}\right) - \frac{1}{a}\left(e^{-a\tau} - e^{-2a\tau}\right)\right]. \end{aligned}$$

Using (A.7.7), this is equivalently expressed as

$$\begin{aligned}\frac{\tau}{2}\tilde{B}(\tau)\psi(x)'\psi(x)\tilde{B}(\tau)' = {}& \frac{\omega_{22}(x) + \omega_{33}(x) + 2\omega_{23}(x)}{a}\left(\tilde{B}_2(\tau) - \tilde{B}_4(\tau)\right) + \omega_{33}(x)\frac{\tau}{2}e^{-2a\tau} \\ & - \frac{\omega_{23}(x) + \omega_{33}(x)}{a}\left(e^{-a\tau} - e^{-2a\tau}\right), \end{aligned} \tag{A.7.15}$$

thus involving $\tilde{B}_4(\tau)$ from (A.7.8). As a function of τ, this is clearly linearly independent of the NS loading functions (15). Because it is linearly independent of the slope adjustments (A.7.4)-(A.7.5) and the remaining terms in convexity (A.7.15), too, (A.7.3) implies the condition

$$\omega_{22}(x) + \omega_{33}(x) + 2\omega_{23}(x) = 0\,. \tag{A.7.16}$$

Thus, convexity (A.7.15) reduces to

$$\frac{\tau}{2}\tilde{B}(\tau)\psi(x)'\psi(x)\tilde{B}(\tau)' = \omega_{33}(x)\frac{\tau}{2}e^{-2a\tau} - \frac{\omega_{33}(x)+\omega_{23}(x)}{a}\left(e^{-a\tau} - e^{-2a\tau}\right). \tag{A.7.17}$$

The functions $\tau e^{-2a\tau}/2$ and $e^{-2a\tau}$ are linearly independent of each other, and of those in the loadings (15), the slope adjustments (A.7.4)-(A.7.5), and the remaining function $e^{-a\tau}$ in (A.7.17), so avoiding dependence on these requires the conditions $\omega_{33}(x) = 0$ and $\omega_{33}(x) + \omega_{23}(x) = 0$, and hence $\omega_{23}(x) = 0$, too. From (A.7.16), $\omega_{22}(x) = -\omega_{33}(x) - 2\omega_{23}(x) = 0$. Because $\omega_{jj}(x) = \psi_j(x)'\psi_j(x)$, we have $\psi_j(x) = 0$, $j = 2, 3$, and we had $\psi_1(x) = 0$ from earlier, so $\psi(x) = 0$. By (A.7.2), dynamic consistency requires $\sigma(\tau, x)' = B(\tau)\psi(x)'$, so $\sigma(\tau, x) = 0$. If the arbitrage-free DTSM is dynamically consistent with NS with fixed a, then it is degenerate. We conclude that NS with fixed a is dynamically inconsistent with all non-degenerate arbitrage-free DTSMs.

Next, we show how how each of the steps in the preceding argument is extended in case a is considered an additional state variable. By Proposition 1, dynamic consistency under the no-arbitrage condition requires the conditions (23) and (A.7.1), with $x = (x_1, x_2, x_3)'$ expanded to $x_a = (x', a)'$. In this case, $\psi(x_a)$ is $d \times 4$, and $\partial Y / \partial x_a'$ is no longer simply given by $\tilde{B}_{1:3}(\tau)$ from (15), but is instead expanded to (20), including the additional component $\partial Y / \partial a$ from (21). Since this depends on x, we no longer have a factor model. First, write the curve shape (19) as

$$Y(\tau, x_a) = \tilde{B}_{1:3}(\tau; a)x = x_1 + \tilde{B}_2(\tau; a)x_2 + \tilde{B}_3(\tau; a)x_3\,, \tag{A.7.18}$$

with $\tilde{B}_{1:3}(\tau; a)$ from (15) depending on a. The slope adjustments $\frac{1}{\tau}\left[Y(\tau, x_a) - Y(0, x_a)\right]$ and $\frac{\partial Y}{\partial \tau}(\tau, x_a)$ in (A.7.1) are computed for given x_a, and so are still obtained by multiplying

each of (A.7.4) and (A.7.5) by $x=(x_1,x_2,x_3)'$. For $\partial Y/\partial a$, we need the derivative

$$\frac{\partial \tilde{B}_2(\tau;a)}{\partial a}=\frac{\partial}{\partial a}\left(\frac{1-e^{-a\tau}}{a\tau}\right)=-\frac{1}{a}\left(\frac{1-e^{-a\tau}}{a\tau}-\mathrm{e}^{-a\tau}\right)=-\frac{1}{a}\tilde{B}_3(\tau;a), \tag{A.7.19}$$

highlighting that the third NS loading function is (proportional to) the derivative of the second with respect to a. Using (A.7.9) and (A.7.19), we immediately get

$$\frac{\partial \tilde{B}_3(\tau;a)}{\partial a}=\frac{\partial \tilde{B}_2(\tau;a)}{\partial a}+\tau\mathrm{e}^{-a\tau}=-\frac{1}{a}\left(\frac{1-e^{-a\tau}}{a\tau}-\mathrm{e}^{-a\tau}\right)+\tau\mathrm{e}^{-a\tau}=-\frac{1}{a}\tilde{B}_3(\tau;a)+\tau\mathrm{e}^{-a\tau}. \tag{A.7.20}$$

Combination of (A.7.18)-(A.7.20) produces

$$\frac{\partial Y}{\partial a}=-\frac{1}{a}\tilde{B}_3(\tau;a)(x_2+x_3)+\tau\mathrm{e}^{-a\tau}x_3, \tag{A.7.21}$$

thereby verifying (21).

The trace term in (A.7.1) involves the second derivative of Y with respect to a. By direct differentiation of (A.7.21), using (A.7.20),

$$\begin{aligned}\frac{\partial^2 Y}{\partial a^2}&=\frac{1}{a^2}\tilde{B}_3(\tau;a)(x_2+x_3)-\frac{1}{a}\left(-\frac{1}{a}\tilde{B}_3(\tau;a)+\tau\mathrm{e}^{-a\tau}\right)(x_2+x_3)-\tau^2e^{-a\tau}x_3\\&=\frac{1}{a^2}\left(2\tilde{B}_3(\tau;a)-a\tau\mathrm{e}^{-a\tau}\right)(x_2+x_3)-\tau^2e^{-a\tau}x_3.\end{aligned}$$

The trace term in (A.7.1) further involves the cross-derivatives with respect to a and either x_2 or x_3, and they are given by (A.7.19) and (A.7.20), respectively. Since $\partial^2 Y/\partial x_1\partial a$ and terms of the type $\partial^2 Y/\partial x_i\partial x_j$ vanish, there are no further terms in the trace in (A.7.1), which therefore takes the form

$$\begin{aligned}\frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 Y}{\partial x_a\partial x_a'}(\tau,x_a)\psi(x_a)'\psi(x_a)\right)&=\omega_{44}(x_a)\left[\frac{1}{a^2}\left(2\tilde{B}_3(\tau;a)-a\tau\mathrm{e}^{-a\tau}\right)(x_2+x_3)-\tau^2e^{-a\tau}x_3\right]\\&\quad-\frac{2(\omega_{24}(x_a)+\omega_{34}(x_a))}{a}\tilde{B}_3(\tau;a)+2\omega_{34}(x_a)\tau\mathrm{e}^{-a\tau}.\end{aligned} \tag{A.7.22}$$

Condition (23) requires

$$\sigma(\tau,x_a)'=\tilde{B}_{1:3}(\tau)\psi_{1:3}(x_a)'+\frac{\partial Y}{\partial a}(\tau,x_a)\psi_4(x_a)', \tag{A.7.23}$$

with the $d\times 3$ matrix $\psi_{1:3}(x)$ given by the first three columns of $\psi(x_a)$. Thus, convexity in (A.7.1) is (A.7.14) plus the terms involving $\partial Y/\partial a$. One term involves the square of

(A.7.21),

$$\frac{\tau}{2}\left(\frac{\partial Y}{\partial a}\right)^2 = \frac{\tau}{2}\left(-\frac{1}{a}\tilde{B}_3(\tau;a)(x_2+x_3)+\tau e^{-a\tau}x_3\right)^2. \tag{A.7.24}$$

Using (A.7.11)-(A.7.12) allows writing (A.7.24) as

$$\begin{aligned}\frac{\tau}{2}\left(\frac{\partial Y}{\partial a}\right)^2 = \frac{1}{a^2}\left[\frac{1}{a}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)+\frac{\tau}{2}e^{-2a\tau}-\frac{1}{a}\left(e^{-a\tau}-e^{-2a\tau}\right)\right](x_2+x_3)^2 \\ +\frac{\tau^3}{2}e^{-2a\tau}x_3^2-\frac{\tau^2}{a}\tilde{B}_3(\tau)e^{-a\tau}(x_2+x_3)x_3.\end{aligned} \tag{A.7.25}$$

Convexity further involves the three cross-products between (A.7.21) and the loadings in (15), multiplied by τ,

$$\tau\tilde{B}_1(\tau)\frac{\partial Y}{\partial a} = -\frac{\tau}{a}\tilde{B}_3(\tau)(x_2+x_3)+\tau^2e^{-a\tau}x_3, \tag{A.7.26}$$

$$\tau\tilde{B}_2(\tau)\frac{\partial Y}{\partial a} = -\frac{\tau}{a}\tilde{B}_2(\tau)\tilde{B}_3(\tau)(x_2+x_3)+\tau^2\tilde{B}_2(\tau)e^{-a\tau}x_3, \tag{A.7.27}$$

$$\tau\tilde{B}_3(\tau)\frac{\partial Y}{\partial a} = -\frac{\tau}{a}\tilde{B}_3(\tau)^2(x_2+x_3)+\tau^2\tilde{B}_3(\tau)e^{-a\tau}x_3. \tag{A.7.28}$$

Rewriting (A.7.27) using (A.7.13) yields

$$\tau\tilde{B}_2(\tau)\frac{\partial Y}{\partial a} = -\frac{\tau}{a}\left[\frac{2}{a\tau}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)-\left(2\tilde{B}_4(\tau)-\tilde{B}_2(\tau)\right)\right](x_2+x_3)+\tau^2\tilde{B}_2(\tau)e^{-a\tau}x_3. \tag{A.7.29}$$

Similarly, rewriting (A.7.28) using (A.7.11) yields

$$\tau\tilde{B}_3(\tau)\frac{\partial Y}{\partial a} = -\frac{\tau}{a}\left[\frac{2}{a\tau}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)+e^{-2a\tau}-2\left(2\tilde{B}_4(\tau)-\tilde{B}_2(\tau)\right)\right](x_2+x_3)+\tau^2\tilde{B}_3(\tau)e^{-a\tau}x_3. \tag{A.7.30}$$

Writing (A.7.14) in terms of the NS loadings, using (A.7.15), and extending with (A.7.25), (A.7.26), (A.7.29) and (A.7.30) yields convexity

$$\begin{aligned}\frac{\tau}{2}\frac{\partial Y}{\partial x_a'}\psi(x_a)'\psi(x_a)\frac{\partial Y}{\partial x_a} &= \omega_{11}(x_a)\frac{\tau}{2}+\frac{\omega_{22}(x_a)+\omega_{33}(x_a)+2\omega_{23}(x_a)}{a}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right) \\ &+\omega_{33}(x_a)\frac{\tau}{2}e^{-2a\tau}+\omega_{12}(x_a)\tau\tilde{B}_2(x)+\omega_{13}(x_a)\tau\tilde{B}_3(x)-\frac{\omega_{23}(x_a)+\omega_{33}(x_a)}{a}\left(e^{-a\tau}-e^{-2a\tau}\right) \\ &+\omega_{44}(x_a)\frac{1}{a^2}\left[\frac{1}{a}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)+\frac{\tau}{2}e^{-2a\tau}-\frac{1}{a}\left(e^{-a\tau}-e^{-2a\tau}\right)\right](x_2+x_3)^2 \\ &+\omega_{44}(x_a)\left[\frac{\tau^3}{2}e^{-2a\tau}x_3^2-\frac{\tau^2}{a}\tilde{B}_3(\tau)e^{-a\tau}(x_2+x_3)x_3\right] \\ &+\omega_{14}(x_a)\left[-\frac{\tau}{a}\tilde{B}_3(\tau)(x_2+x_3)+\tau^2e^{-a\tau}x_3\right] \\ &+\omega_{24}(x_a)\left\{-\frac{\tau}{a}\left[\frac{2}{a\tau}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)-\left(2\tilde{B}_4(\tau)-\tilde{B}_2(\tau)\right)\right](x_2+x_3)+\tau^2\tilde{B}_2(\tau)e^{-a\tau}x_3\right\}\end{aligned}$$

$$-\omega_{34}(x_a)\frac{\tau}{a}\left[\frac{2}{a\tau}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)+\mathrm{e}^{-2a\tau}-2\left(2\tilde{B}_4(\tau)-\tilde{B}_2(\tau)\right)\right](x_2+x_3).$$

$$+\omega_{34}(x_a)\tau^2\tilde{B}_3(\tau)\mathrm{e}^{-a\tau}x_3\,. \tag{A.7.31}$$

The steps analyzing (A.7.14) as convexity in (A.7.3) in the case of fixed a are now extended to (A.7.31) as convexity in (A.7.1), thus accommodating time-varying a. For dynamic consistency, (A.7.1) requires that convexity (A.7.31), as a function of maturity τ, be spanned by the loadings (15), $\partial Y/\partial a$ from (A.7.21), the functions in the slope adjustments (A.7.4)-(A.7.5), and the trace (A.7.22). The first term in (A.7.31), multiplying $\omega_{11}(x_a)$, is the linear, $\tau/2$. This is linearly independent of the functions of τ in the remaining convexity terms in (A.7.31), and in (15), (A.7.21), (A.7.4)-(A.7.5), and (A.7.22), since these are all spanned by functions involving exponentials, the reciprocal, $1/\tau$, the constant function, products of these, the functions $\tau\mathrm{e}^{-a\tau}$ and $\tau\mathrm{e}^{-2a\tau}/2$, and those involving τ^2 and τ^3 in (A.7.31). To be sure, the argument has already been made for the case with constant a, which has the same loadings (15) and slope adjustments, so attention can be restricted to the new terms in the time-varying a case, i.e., $\partial Y/\partial a$ from (A.7.21), the trace (A.7.22), and the terms in convexity (A.7.31) involving $\omega_{j4}(x_a)$, $j=1,\ldots,4$. In two cases, these new terms do involve τ multiplied by $\tilde{B}_j(\tau;a)$, $j=2$, 3, 4, without involving further exponentials, but never τ alone (or multiplied by $\tilde{B}_1(\tau;a)$, the constant). Thus, using (A.7.10), we have in (A.7.31) that

$$\frac{\tau}{a}\left(2\tilde{B}_4(\tau;a)-\tilde{B}_2(\tau;a)\right)=\frac{\mathrm{e}^{-a\tau}-\mathrm{e}^{-2a\tau}}{a^2}\,, \tag{A.7.32}$$

whereas

$$\tau\tilde{B}_3(\tau;a)=\frac{1-\mathrm{e}^{-a\tau}}{a}-\tau\mathrm{e}^{-a\tau}\,.$$

The constant function appears, but never the linear, $\tau/2$ (recall that $a>0$). This implies that, again, $\omega_{11}(x_a)=0$, so $\psi_1(x_a)=0$. The first state variable is locally deterministic, and

$\omega_{1j}(x_a) = \psi_1((x_a))'\psi_j((x_a)) = 0$, $j = 1,\ldots,4$. Convexity (A.7.31) simplifies to

$$\begin{aligned}
\frac{\tau}{2}\frac{\partial Y}{\partial x_a'}\psi(x_a)'\psi(x_a)\frac{\partial Y}{\partial x_a} &= \frac{\omega_{22}(x_a)+\omega_{33}(x_a)+2\omega_{23}(x_a)}{a}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)\\
&+\omega_{33}(x_a)\frac{\tau}{2}\mathrm{e}^{-2a\tau}-\frac{\omega_{23}(x_a)+\omega_{33}(x_a)}{a}\left(\mathrm{e}^{-a\tau}-\mathrm{e}^{-2a\tau}\right)\\
&+\omega_{44}(x_a)\frac{1}{a^2}\left[\frac{1}{a}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)+\frac{\tau}{2}\mathrm{e}^{-2a\tau}-\frac{1}{a}\left(\mathrm{e}^{-a\tau}-\mathrm{e}^{-2a\tau}\right)\right](x_2+x_3)^2\\
&+\omega_{44}(x_a)\left[\frac{\tau^3}{2}\mathrm{e}^{-2a\tau}x_3^2-\frac{\tau^2}{a}\tilde{B}_3(\tau)\mathrm{e}^{-a\tau}(x_2+x_3)x_3\right]\\
&+\omega_{24}(x_a)\left\{-\frac{\tau}{a}\left[\frac{2}{a\tau}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)-\left(2\tilde{B}_4(\tau)-\tilde{B}_2(\tau)\right)\right](x_2+x_3)+\tau^2\tilde{B}_2(\tau)\mathrm{e}^{-a\tau}x_3\right\}\\
&-\omega_{34}(x_a)\frac{\tau}{a}\left[\frac{2}{a\tau}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)+\mathrm{e}^{-2a\tau}-2\left(2\tilde{B}_4(\tau)-\tilde{B}_2(\tau)\right)\right](x_2+x_3)\\
&+\omega_{34}(x_a)\tau^2\tilde{B}_3(\tau)\mathrm{e}^{-a\tau}x_3\,. \qquad \text{(A.7.33)}
\end{aligned}$$

The first function of τ in convexity, $\tilde{B}_2(\tau;a)-\tilde{B}_4(\tau;a)$, especially the portion $\tilde{B}_4(\tau;a)$ from (A.7.8), is not in any other term in (A.7.1), i.e., neither in remaining convexity (A.7.33), nor in (15), (A.7.21), (A.7.4)-(A.7.5), (A.7.22), and it is linearly independent of all other terms ($\tilde{B}_4$ appears in the terms involving $\omega_{24}(x_a)$ and $\omega_{34}(x_a)$ in (A.7.33), too, but there, it is multiplied by τ). It follows that the first term in convexity vanishes, yielding the condition

$$\omega_{22}(x_a)+\omega_{33}(x_a)+\frac{\omega_{44}(x_a)(x_2+x_3)^2}{a^2}+2\omega_{23}(x_a)-\frac{2(\omega_{24}(x_a)+\omega_{34}(x_a))(x_2+x_3)}{a}=0\,. \quad \text{(A.7.34)}$$

Thus, convexity (A.7.33) reduces to

$$\begin{aligned}
\frac{\tau}{2}\frac{\partial Y}{\partial x_a'}(\tau,x_a)\psi(x)'\psi(x)\frac{\partial Y}{\partial x_a}(\tau,x_a) &= \omega_{33}(x_a)\frac{\tau}{2}\mathrm{e}^{-2a\tau}-\frac{\omega_{23}(x_a)+\omega_{33}(x_a)}{a}\left(\mathrm{e}^{-a\tau}-\mathrm{e}^{-2a\tau}\right)\\
&+\omega_{44}(x_a)\frac{1}{a^2}\left[\frac{\tau}{2}\mathrm{e}^{-2a\tau}-\frac{1}{a}\left(\mathrm{e}^{-a\tau}-\mathrm{e}^{-2a\tau}\right)\right](x_2+x_3)^2\\
&+\omega_{44}(x_a)\left[\frac{\tau^3}{2}\mathrm{e}^{-2a\tau}x_3^2-\frac{\tau^2}{a}\tilde{B}_3(\tau;a)\mathrm{e}^{-a\tau}(x_2+x_3)x_3\right]\\
&+\omega_{24}(x_a)\left\{\frac{\tau}{a}\left(2\tilde{B}_4(\tau;a)-\tilde{B}_2(\tau;a)\right)(x_2+x_3)+\tau^2\tilde{B}_2(\tau;a)\mathrm{e}^{-a\tau}x_3\right\} \qquad \text{(A.7.35)}\\
&+\omega_{34}(x_a)\left\{\frac{\tau}{a}\left[-\mathrm{e}^{-2a\tau}+2\left(2\tilde{B}_4(\tau;a)-\tilde{B}_2(\tau;a)\right)\right](x_2+x_3)+\tau^2\tilde{B}_3(\tau;a)\mathrm{e}^{-a\tau}x_3\right\}.
\end{aligned}$$

The function $\tau^3\mathrm{e}^{-2a\tau}$ in the term involving $\omega_{44}(x_a)$ is not in any of the remaining terms in (A.7.1), either, and it is linearly independent of these. It follows that either (i) $\omega_{44}(x_a)=0$,

or (ii) $x_3 = 0$. We consider these two cases separately.

In case (i), if $\omega_{44}(x_a) = 0$, then $\psi_4(x_a) = 0$, so $\omega_{j4}(x_a) = 0$, $j = 1,\ldots,4$, because $\omega_{ij}(x_a) = \psi_i(x_a)'\psi_j(x_a)$. Thus, convexity (A.7.35) reduces to (A.7.17), and as before, $\tau e^{-2a\tau}$ is not spanned, this time because it does not appear in the additional trace terms (A.7.22), either. It follows that $\omega_{33}(x_a) = 0$, so $\psi_3(x_a) = 0$, and therefore $\omega_{23}(x_a) = 0$. From (A.7.34), $\omega_{22}(x_a) = 0$, and $\omega(x_a) = 0$, because $\omega_{jj}(x_a) = 0$, $j = 1,\ldots,4$.

In case (ii), $x_3 = 0$, if $\psi_3(x_a) \neq 0$, then $x_3 \neq 0$ the next instant, and case (i) applies. If $\psi_3(x_a) = 0$, then x_3 is locally deterministic, so $\omega_{j3}(x_a) = 0$, $j = 1,\ldots,4$, and convexity (A.7.35) reduces to

$$\begin{aligned}\frac{\tau}{2}\frac{\partial Y}{\partial x_a'}(\tau,x_a)\psi(x)'\psi(x)\frac{\partial Y}{\partial x_a}(\tau,x_a) = \omega_{44}(x_a)\frac{1}{a^2}\left[\frac{\tau}{2}e^{-2a\tau} - \frac{1}{a}\left(e^{-a\tau} - e^{-2a\tau}\right)\right]x_2^2 \\ + \omega_{24}(x_a)\frac{\tau}{a}\left(2\tilde{B}_4(\tau) - \tilde{B}_2(\tau)\right)x_2\,.\end{aligned} \tag{A.7.36}$$

The function $e^{-2a\tau}$ is neither in remaining convexity, nor in (15), (A.7.21), (A.7.4)-(A.7.5), (A.7.22), and it is linearly independent of these. It follows that either (a) $\omega_{44}(x_a) = 0$, or (b) $x_2 = 0$. If (a) $\omega_{44}(x_a) = 0$, then $\omega_{j4}(x_a) = 0$, $j = 1,\ldots,4$. From (A.7.34), $\omega_{22}(x_a) = 0$, and $\omega(x_a) = 0$, because $\omega_{jj}(x_a) = 0$, $j = 1,\ldots,4$. If (b) $x_2 = 0$, then convexity (A.7.36) vanishes. The trace term (A.7.22) in (A.7.1) involves the function $\omega_{44}\tau^2 e^{-a\tau}$. Since this is not in (15), (A.7.21), (A.7.4)-(A.7.5), and it is linearly independent of these, it follows that $\omega_{44}(x_a) = 0$. Again, this implies $\omega_{j4}(x_a) = 0$, $j = 1,\ldots,4$, so from (A.7.34), $\omega_{22}(x_a) = 0$, and $\omega(x_a) = 0$, because $\omega_{jj}(x_a) = 0$, $j = 1,\ldots,4$.

Thus, we have have $\psi(x_a)'\psi(x_a) = \omega(x_a) = 0$, both in case (i) and (ii), and these are exhaustive, so $\psi(x_a) = 0$. By (A.7.23), dynamic consistency requires $\sigma(\tau,x_a)' = \left(\tilde{B}_{1:3}(\tau;a), \partial Y/\partial a\right)\psi(x_a)'$, so $\sigma(\tau,x_a) = 0$. If the arbitrage-free DTSM is dynamically consistent with NS, then it is degenerate. We conclude that the NS curve shape is dynamically inconsistent with all non-degenerate arbitrage-free DTSMs, whether a is fixed or time-varying. □

Corollary A.7.1. *Dynamic consistency between the DTSM* (a,σ) *and the factor model* $\mathcal{B}$ *is*

equivalent to the existence of suitable ϕ, ψ satisfying Assumption 1 and the conditions

$$\alpha(\tau,x) = B(\tau)\phi(x), \tag{A.7.37}$$

$$\sigma(\tau,x)' = B(\tau)\psi(x)', \tag{A.7.38}$$

for all (τ,x).

Proof of Corollary A.7.1. This follows from Proposition 1 for $Y(\tau,x) = B(\tau)x$. □

Proof of Corollary 2. This follows from Proposition A.7.1 for $Y(\tau,x) = B(\tau)x$. □

A.8. Affine models

In affine term structure models (ATSMs), the yield is written in the form $Y(\tau,\bar{x}(t)) = \bar{A}(\tau) + \bar{B}(\tau)\bar{x}(t)$, where $\bar{B}(\tau)$ satisfies a Riccatti ODE. This is clearly the special case of the general form $Y(\tau,x) = B(\tau)x$ in which one of the state variables is constant, i.e., $x(t) = (\bar{x}(t)',1)'$, $B(\tau) = (\bar{B}(\tau),\bar{A}(\tau))$, $k = d+1$. In this case, $(\bar{\phi},\bar{\psi})$ pertain to the d non-constant state variables, only, with $\bar{\psi}$ $d \times d$ and invertible. Viewing (27) as an equation in the unknown $B(\cdot)$, assuming elementwise affine forms (in $\bar{x}$) for $\bar{\phi}$ and $\bar{\psi}'\bar{\psi}$, and setting $\lambda = 0$ reduces the equation to the linear-quadratic Riccatti ODE, which under an initial condition giving the short rate as an affine function of $\bar{x}$ determines the solution $B = (\bar{B},\bar{A})$.[42] The state drift is affine under both the physical and the martingale measure under the additional assumption that $\lambda(x) = \bar{\psi}(x)\lambda_0 + (\bar{\psi}(x)^{-1})'(\lambda_1 + \Lambda_1 x)$, where λ_0, λ_1 are $d \times 1$ and Λ_1 $d \times d$, so that $\bar{\psi}'\lambda$ is affine, see Cheridito, Filipović, and Kimmel (2007).[43] The dynamic consistency approach accommodates the more general structure $Y(\tau,x(t)) = \bar{B}(\tau)\bar{x}(t) + \bar{\bar{B}}(\tau)\bar{\bar{x}}(t)$, say, where $\bar{\bar{x}}(t)$ is a vector of locally deterministic but potentially time-varying state variables, with loadings $\bar{\bar{B}}(\tau)$. Thus, the time-invariant ATSM intercept $\bar{A}(\tau)$ is generalized to the time-varying form $\bar{\bar{B}}(\tau)\bar{\bar{x}}(t)$ under dynamic consistency, $x = (\bar{x}',\bar{\bar{x}}')'$, $\phi = (\bar{\phi}',\bar{\bar{\phi}}')'$, say, and $\psi' = (\bar{\psi},0)'$ is $k \times d$, where $k \geq d$ – indeed, typically $k > d+1$, as we demonstrate.

[42] Since $\lambda = 0$, $\bar{\phi}$ is the state drift under the martingale measure in this case, $\bar{\phi} = \bar{\phi}^Q$, say.
[43] With λ non-zero, $\bar{\phi}$ is the drift under the physical measure, so $\bar{\phi}^Q = \bar{\phi}^P - \bar{\psi}'\lambda$.

A.9. Proofs of Propositions 2, A.9.1, A.9.2, and Theorem 2

Proof of Proposition 2. By Corollary 2, conditions (25)-(26) must be verified. Inserting (31) for $\sigma(\tau,x)$, $\psi(x)=(0,\psi_2(x),0,0)'$, and $\tilde{B}_{1:4}(\tau)$ from (32) for $B(\tau)$ in (26) verifies the condition. Next, by (31), the volatility function takes the form $\sigma(\tau,x)=\psi_2(x)\tilde{B}_2(\tau)$, and using (A.7.7), convexity in (25) is

$$\begin{aligned}\frac{\tau}{2}\sigma(\tau,x)'\sigma(\tau,x) &= \frac{\tau}{2}\psi_2(x)^2\tilde{B}_2(\tau)^2\\ &= \frac{\psi_2(x)^2}{a}\left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau}-\frac{1-\mathrm{e}^{-2a\tau}}{2a\tau}\right)\\ &= \frac{\psi_2(x)^2}{a}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)\\ &= \tilde{B}_{1:4}(\tau)\left(0,\frac{\psi_2(x)^2}{a},0,-\frac{\psi_2(x)^2}{a}\right)'. \end{aligned} \tag{A.9.1}$$

This is spanned in (25) by the loadings from $B(\tau)=\tilde{B}_{1:4}(\tau)$ from (32). From (A.9.1), the spanning coefficients are

$$\left(0,\frac{\psi_2(x)^2}{a},0,-\frac{\psi_2(x)^2}{a}\right)', \tag{A.9.2}$$

which are therefore part of $\phi(x)$ in (25). Similarly, in (25), we have

$$\sigma(\tau,x)'\lambda(x)=\psi_2(x)\tilde{B}_2(\tau)\lambda_2(x)=\tilde{B}_{1:4}(\tau)(0,\psi_2(x)\lambda_2(x),0,0)', \tag{A.9.3}$$

so

$$(0,\psi_2(x)\lambda_2(x),0,0)' \tag{A.9.4}$$

is another term in $\phi(x)$ in (25). It remains to check that the resulting slope adjustments are spanned by the loadings in $\tilde{B}_{1:4}$ in (25). This is done easily, because the fourth loading in (32) is of the same type as the second, with $2a$ replacing a. Thus, when (15) is augmented to (32), then the average yield spreads (A.7.4) are expanded to

$$\begin{aligned}\frac{1}{\tau}\left[\tilde{B}_{1:4}(\tau)-\tilde{B}_{1:4}(0)\right] &= \frac{1}{\tau}\begin{pmatrix}\tilde{B}_1(\tau)-1 & \tilde{B}_2(\tau)-1 & \tilde{B}_3(\tau)-0 & \tilde{B}_4(\tau)-1\end{pmatrix}\\ &= \begin{pmatrix}0 & \frac{1-\mathrm{e}^{-a\tau}}{a\tau^2}-\frac{1}{\tau} & \frac{1-\mathrm{e}^{-a\tau}}{a\tau^2}-\frac{\mathrm{e}^{-a\tau}}{\tau} & \frac{1-\mathrm{e}^{-2a\tau}}{2a\tau^2}-\frac{1}{\tau}\end{pmatrix}, \end{aligned} \tag{A.9.5}$$

and the local slope adjustments in (A.7.5) are expanded to

$$\frac{\partial \tilde{B}_{1:4}}{\partial \tau}(\tau) = \begin{pmatrix} 0 & \frac{e^{-a\tau}}{\tau} - \frac{1-e^{-a\tau}}{a\tau^2} & \frac{e^{-a\tau}}{\tau} - \frac{1-e^{-a\tau}}{a\tau^2} + ae^{-a\tau} & \frac{e^{-2a\tau}}{\tau} - \frac{1-e^{-2a\tau}}{2a\tau^2} \end{pmatrix}. \tag{A.9.6}$$

The total slope adjustment in (25) is therefore

$$\begin{aligned} \left(\frac{1}{\tau}\left[\tilde{B}_{1:4}(\tau) - \tilde{B}_{1:4}(0)\right] + \frac{\partial \tilde{B}_{1:4}}{\partial \tau}(\tau)\right)x &= -\frac{1-e^{-a\tau}}{\tau}x_2 + ae^{-a\tau}x_3 - \frac{1-e^{-2a\tau}}{\tau}x_4 \\ &= -a\tilde{B}_2(\tau)x_2 + ae^{-a\tau}x_3 - 2a\tilde{B}_4(\tau)x_4\,. \end{aligned} \tag{A.9.7}$$

Using the relation

$$e^{-a\tau} = \tilde{B}_2(\tau) - \tilde{B}_3(\tau), \tag{A.9.8}$$

total slope slope adjustment is written as

$$\begin{aligned} \left(\frac{1}{\tau}\left[\tilde{B}_{1:4}(\tau) - \tilde{B}_{1:4}(0)\right] + \frac{\partial \tilde{B}_{1:4}}{\partial \tau}(\tau)\right)x &= -a\tilde{B}_2(\tau)x_2 + a\left(\tilde{B}_2(\tau) - \tilde{B}_3(\tau)\right)x_3 - 2a\tilde{B}_4(\tau)x_4 \\ &= \tilde{B}_{1:4}(\tau)(0, -a(x_2 - x_3), -ax_3, -2ax_4)' \end{aligned}$$

This is spanned in (25), with coefficients

$$(0, -a(x_2 - x_3), -ax_3, -2ax_4)'\,. \tag{A.9.9}$$

Total spanning coefficients on $\tilde{B}_{1:4}$ in (25) for convexity, risk compensation, and slope adjustment are therefore obtained by adding (A.9.2), (A.9.4), and (A.9.9), i.e.,

$$\phi(x) = (0, \frac{\psi_2(x)^2}{a} + \psi_2(x)\lambda_2(x) - a(x_2 - x_3), -ax_3, -\frac{\psi_2(x)^2}{a} - 2ax_4)'. \tag{A.9.10}$$

Given spanning, dynamic consistency of ANS with the arbitrage-free DTSM with market prices of risk $\lambda(x)$ and volatility (31) follows from Corollary 2. The expression (33) for the drift follows by inserting (A.9.10) in $\alpha(\tau, x) = \tilde{B}_{1:4}(\tau)\phi(x)$. □

Proposition A.9.1. *The augmented NS (ANS) curve shape given by loading functions*

$$\tilde{B}_{1:4}(\tau) = \begin{pmatrix} 1 & \frac{1-e^{-a\tau}}{a\tau} & \frac{1-e^{-a\tau}}{a\tau} - e^{-a\tau} & \frac{1-e^{-2a\tau}}{2a\tau} \end{pmatrix} \tag{A.9.11}$$

with fixed a is dynamically consistent with the DTSM with drift

$$\alpha(\tau, x) = \phi_1(x) + \phi_2(x)\frac{1-e^{-a\tau}}{a\tau} + \phi_3(x)\left(\frac{1-e^{-a\tau}}{a\tau} - e^{-a\tau}\right) + \phi_4(x)\frac{1-e^{-2a\tau}}{2a\tau} \tag{A.9.12}$$

and volatility function (31)*, provided* $\phi(x) = (\phi_1(x), \ldots, \phi_4(x))'$ *and* $\psi(x) = (0, \psi_2(x), 0, 0)'$

satisfy Assumption 1.

Proof of Proposition A.9.1. By Corollary A.7.1, conditions (A.7.37)-(A.7.38) must be verified. Inserting (A.9.12) for $\alpha(\tau,x)$ and $\tilde{B}_{1:4}(\tau)$ from (A.9.11) for $B(\tau)$ in (A.7.37) verifies the condition. Inserting (31) for $\sigma(\tau,x)$, $\psi(x)=(0,\psi_2(x),0,0)'$, and $\tilde{B}_{1:4}(\tau)$ from (32) for $B(\tau)$ in (A.7.38) verifies the condition. The conclusion follows. □

Proof of Theorem 2. By Corollary 2, conditions (25)-(26) must be verified for $B(\tau)=\tilde{B}_{1:7}(\tau)$ from (43) and $\psi(x)$ from (40). Inserting these in the condition on $\sigma(\tau,x)$ in (26) produces (41). By slight abuse of notation, we write $\tilde{\tilde{B}}_1=\left(1-\mathrm{e}^{-b\tau}\right)/(b\tau)$ for the first loading function in $\tilde{B}_{1:7}$. Using (A.7.7), (A.7.11), (A.7.13), and the definition of $\tilde{B}_7(\tau)$ from (43), convexity in (25) is

$$\begin{aligned}
\frac{\tau}{2}\sigma(\tau,x)'\sigma(\tau,x) &= \frac{\tau}{2}\left(\psi_1(x)^2\tilde{\tilde{B}}_1(\tau)^2+\psi_2(x)^2\tilde{B}_2(\tau)^2+\psi_3(x)^2\tilde{B}_3(\tau)^2+2\omega_{23}\tilde{B}_2(\tau)\tilde{B}_3(\tau)\right)\\
&= \frac{\psi_1(x)^2}{b}\left(\frac{1-\mathrm{e}^{-b\tau}}{b\tau}-\frac{1-\mathrm{e}^{-2b\tau}}{2b\tau}\right)+\frac{\psi_2(x)^2}{a}\left(\frac{1-\mathrm{e}^{-a\tau}}{a\tau}-\frac{1-\mathrm{e}^{-2a\tau}}{2a\tau}\right)\\
&+\psi_3(x)^2\left(\frac{1}{a}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)+\frac{\tau}{2}\mathrm{e}^{-2a\tau}-\tau\left(2\tilde{B}_4(\tau)-\tilde{B}_2(\tau)\right)\right)\\
&+2\omega_{23}\left(\frac{1}{a}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)-\frac{\tau}{2}\left(2\tilde{B}_4(\tau)-\tilde{B}_2(\tau)\right)\right)\\
&= \frac{\psi_1(x)^2}{b}\left(\tilde{\tilde{B}}_1(\tau)-\tilde{B}_7(\tau)\right)+\frac{\psi_2(x)^2+\psi_3(x)^2+2\omega_{23}}{a}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)\\
&+\psi_3(x)^2\frac{\tau}{2}\mathrm{e}^{-2a\tau}-\left(\psi_3(x)^2+\omega_{23}\right)\tau\left(2\tilde{B}_4(\tau)-\tilde{B}_2(\tau)\right).
\end{aligned}$$

Using the definitions of $\tilde{B}_5(\tau)$ and $\tilde{B}_6(\tau)$ from (43), we have that $\tau\left(2\tilde{B}_4(\tau)-\tilde{B}_2(\tau)\right)=\tilde{B}_6(\tau)$, by (A.7.10), and that convexity is written as

$$\begin{aligned}
\frac{\tau}{2}\sigma(\tau,x)'\sigma(\tau,x) &= \frac{\psi_1(x)^2}{b}\left(\tilde{\tilde{B}}_1(\tau)-\tilde{B}_7(\tau)\right)+\frac{\psi_2(x)^2+\psi_3(x)^2+2\omega_{23}}{a}\left(\tilde{B}_2(\tau)-\tilde{B}_4(\tau)\right)\\
&+\psi_3(x)^2\tilde{B}_5(\tau)-\left(\psi_3(x)^2+\omega_{23}\right)\tilde{B}_6(\tau).
\end{aligned}$$

This is spanned in (25) by the loadings from $B(\tau)=\tilde{B}_{1:7}(\tau)$ from (43). The spanning coefficients are

$$\left(\frac{\psi_1(x)^2}{b},\frac{\psi_2(x)^2+\psi_3(x)^2+2\omega_{23}}{a},0,-\frac{\psi_2(x)^2+\psi_3(x)^2+2\omega_{23}}{a},\psi_3(x)^2,-\psi_3(x)^2-\omega_{23},-\frac{\psi_1(x)^2}{b}\right)', \tag{A.9.13}$$

which are therefore part of $\phi(x)$ in (25). Similarly, in (25), we have

$$\sigma(\tau,x)'\lambda(x) = \tilde{B}_{1:7}(\tau)\psi(x)'\lambda(x) = \tilde{B}_{1:7}(\tau)\begin{pmatrix} \psi_{11}(x)\lambda_1(x) \\ \psi_{22}(x)\lambda_2(x)+\psi_{23}(x)\lambda_3(x) \\ \psi_{23}(x)\lambda_2(x)+\psi_{33}(x)\lambda_3(x) \\ 0_4 \end{pmatrix},$$

with 0_4 a four-vector of zeroes, so

$$\left(\psi_{11}(x)\lambda_1(x), \psi_{22}(x)\lambda_2(x)+\psi_{23}(x)\lambda_3(x), \psi_{23}(x)\lambda_2(x)+\psi_{33}(x)\lambda_3(x), 0_4'\right)' \tag{A.9.14}$$

is another term in $\phi(x)$ in (25). It remains to check that the resulting slope adjustments are spanned in (25). The corresponding analysis for $\tilde{B}_{1:4}(\tau)$ from (32) was done in (A.9.5)-(A.9.7). This is extended to $\tilde{B}_{1:7}(\tau)$ from (43), now using $\tilde{\tilde{B}}_1(\tau) = \left(1-\mathrm{e}^{-b\tau}\right)/(b\tau)$ for the first loading function, which is of the same type as the second, but with a replaced by b. Thus, when (15) is augmented to (43), then the average yield spreads (A.7.4) are expanded to $\left(\tilde{B}_{1:7}(\tau)-\tilde{B}_{1:7}(0)\right)/\tau$ given by the functions

$$\frac{1-\mathrm{e}^{-b\tau}}{b\tau^2}-\frac{1}{\tau}, \frac{1-\mathrm{e}^{-a\tau}}{a\tau^2}-\frac{1}{\tau}, \frac{1-\mathrm{e}^{-a\tau}}{a\tau^2}-\frac{\mathrm{e}^{-a\tau}}{\tau}, \frac{1-\mathrm{e}^{-2a\tau}}{2a\tau^2}-\frac{1}{\tau},$$

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

The corresponding local slope adjustments are given by the functions

$$\frac{\mathrm{e}^{-b\tau}}{\tau}-\frac{1-\mathrm{e}^{-b\tau}}{b\tau^2}, \frac{\mathrm{e}^{-a\tau}}{\tau}-\frac{1-\mathrm{e}^{-a\tau}}{a\tau^2}, \frac{\mathrm{e}^{-a\tau}}{\tau}-\frac{1-\mathrm{e}^{-a\tau}}{a\tau^2}+a\mathrm{e}^{-a\tau}, \frac{\mathrm{e}^{-2a\tau}}{\tau}-\frac{1-\mathrm{e}^{-2a\tau}}{2a\tau^2},$$

$$\frac{1}{2}\mathrm{e}^{-2a\tau}-a\tau\mathrm{e}^{-2a\tau}, -\mathrm{e}^{-a\tau}+2\mathrm{e}^{-2a\tau}, \frac{\mathrm{e}^{-2b\tau}}{\tau}-\frac{1-\mathrm{e}^{-2b\tau}}{2b\tau^2}.$$

The total slope adjustment in (25) is therefore

$$\begin{aligned}\left(\frac{1}{\tau}\left[\tilde{B}(\tau)-\tilde{B}(0)\right]+\frac{\partial\tilde{B}}{\partial\tau}(\tau)\right)x &= -\frac{1-\mathrm{e}^{-b\tau}}{\tau}x_1-\frac{1-\mathrm{e}^{-a\tau}}{\tau}x_2+a\mathrm{e}^{-a\tau}x_3-\frac{1-\mathrm{e}^{-2a\tau}}{\tau}x_4 \\ &+\left(\mathrm{e}^{-2a\tau}-a\tau\mathrm{e}^{-2a\tau}\right)x_5+\left[\frac{1}{a\tau}\left(\mathrm{e}^{-a\tau}-\mathrm{e}^{-2a\tau}\right)+\left(-\mathrm{e}^{-a\tau}+2\mathrm{e}^{-2a\tau}\right)\right]x_6-\frac{1-\mathrm{e}^{-2b\tau}}{\tau}x_7.\end{aligned} \tag{A.9.15}$$

We must show that this is indeed spanned by the functions in $\tilde{B}_{1:7}(\tau)$. First off, the functions of τ multiplying x_1, x_2, x_4, and x_7 in (A.9.15) are proportional to $\tilde{B}_1(\tau)$, $\tilde{B}_2(\tau)$,

$\tilde{B}_4(\tau)$, and $\tilde{B}_7(\tau)$, respectively, so they are spanned, with coefficients read off directly as

$$(-bx_1, -ax_2, 0, -2ax_4, 0, 0, -2bx_7)', \tag{A.9.16}$$

which is therefore another term in $\phi(x)$ in (25). Next, by (A.9.8), we have $e^{-a\tau} = \tilde{B}_2(\tau) - \tilde{B}_3(\tau)$, so the term involving x_3 in (A.9.15) is $a\left(\tilde{B}_2(\tau) - \tilde{B}_3(\tau)\right)x_3$, and hence spanned, with coefficients

$$(0, ax_3, -ax_3, 0, 0, 0, 0)', \tag{A.9.17}$$

also a term in $\phi(x)$ in (25). Further, we have

$$e^{-2a\tau} = -a\tilde{B}_6(\tau) + e^{-a\tau} = -a\tilde{B}_6(\tau) + \tilde{B}_2(\tau) - \tilde{B}_3(\tau), \tag{A.9.18}$$

using (A.9.8) in the second equality. Thus, the function involving x_5 in (A.9.15) is $(-a\tilde{B}_6(\tau) + \tilde{B}_2(\tau) - \tilde{B}_3(\tau) - 2a\tilde{B}_5(\tau))x_5$, and hence spanned, with coefficients

$$(0, x_5, -x_5, 0, -2ax_5, -ax_5, 0)', \tag{A.9.19}$$

again a term in $\phi(x)$ in (25). Finally, using (A.7.10), (A.9.8) and (A.9.18), the function multiplying x_6 in (A.9.15) is

$$\begin{aligned}\frac{1}{a\tau}\left(e^{-a\tau} - e^{-2a\tau}\right) - e^{-a\tau} + 2e^{-2a\tau} &= \left(2\tilde{B}_4(\tau) - \tilde{B}_2(\tau)\right) - \left(\tilde{B}_2(\tau) - \tilde{B}_3(\tau)\right) \\ &\quad + 2\left(-a\tilde{B}_6(\tau) + \tilde{B}_2(\tau) - \tilde{B}_3(\tau)\right) \\ &= -\tilde{B}_3(\tau) + 2\tilde{B}_4(\tau) - 2a\tilde{B}_6(\tau),\end{aligned} \tag{A.9.20}$$

so the term in (A.9.15) given by x_6 multiplied by (A.9.20) is spanned, with coefficients

$$(0, 0, -x_6, 2x_6, 0, -2ax_6, 0)'. \tag{A.9.21}$$

The total spanning coefficients $\phi(x)$ on $\tilde{B}_{1:7}(\tau)$ in (25) are given by the sum of the coefficients for convexity, (A.9.13), risk compensation, (A.9.14), and slope adjustments, (A.9.16), (A.9.17), (A.9.19), and (A.9.21). Thus,

$$\phi(x) = \begin{pmatrix} \frac{1}{b}\psi_{11}(x)^2 + \psi_{11}(x)\lambda_1(x) - bx_1 \\ \frac{1}{a}(\omega_{22}(x) + \omega_{33}(x) + 2\omega_{23}(x)) + \psi_{22}(x)\lambda_2(x) + \psi_{23}(x)\lambda_3(x) - ax_2 + ax_3 + x_5 \\ \psi_{23}(x)\lambda_2(x) + \psi_{33}(x)\lambda_3(x) - ax_3 - x_5 - x_6 \\ -\frac{1}{a}(\omega_{22}(x) + \omega_{33}(x) + 2\omega_{23}(x)) - 2ax_4 + 2x_6 \\ \omega_{33}(x) - 2ax_5 \\ -\omega_{33}(x) - \omega_{23} - ax_5 - 2ax_6 \\ -\frac{1}{b}\psi_{11}(x)^2 - 2bx_7 \end{pmatrix}, \quad \text{(A.9.22)}$$

yielding the expression for $\phi(x)$ in Theorem 2. The conclusion follows from Corollary 2. □

To restate $\phi(x)$ from (A.9.22) with the local variance matrix $\omega(x)$ of the state variables expressed in terms of the volatilities $\psi(x)$, recall from the proof of Corollary 1 (in Appendix A.7) that the notation is $\omega(x) = \psi(x)'\psi(x)$, a 3×3 matrix, with $\omega_{ij}(x) = \psi_i(x)'\psi_j(x)$, and $\psi_i(x)$ for the i^{th} column in $\psi(x)$. From (40),

$$\psi_1(x) = \begin{pmatrix} \psi_{11}(x) \\ 0 \\ 0 \end{pmatrix}, \quad \psi_2(x) = \begin{pmatrix} 0 \\ \psi_{22}(x) \\ \psi_{32}(x) \end{pmatrix}, \quad \psi_3(x) = \begin{pmatrix} 0 \\ \psi_{23}(x) \\ \psi_{33}(x) \end{pmatrix}.$$

It follows that

$$\begin{aligned} \omega_{22}(x) &= \psi_{22}^2(x) + \psi_{32}^2(x), \\ \omega_{33}(x) &= \psi_{23}^2(x) + \psi_{33}^2(x), \\ \omega_{23}(x) &= \psi_{22}(x)\psi_{23}(x) + \psi_{32}(x)\psi_{33}(x), \end{aligned}$$

which is (42). Therefore,

$$\begin{aligned} \omega_{22}(x) + \omega_{33}(x) + 2\omega_{23}(x) &= \psi_{22}^2(x) + \psi_{32}^2(x) + \psi_{23}^2(x) + \psi_{33}^2(x) + 2\psi_{22}(x)\psi_{23}(x) + 2\psi_{32}(x)\psi_{33}(x), \\ \omega_{33}(x) + \omega_{23}(x) &= \psi_{23}^2(x) + \psi_{33}^2(x) + \psi_{22}(x)\psi_{23}(x) + \psi_{32}(x)\psi_{33}(x). \end{aligned}$$

Using these expressions (A.9.22) produces

$$\phi(x) = \begin{pmatrix} \frac{1}{b}\psi_{11}(x)^2 + \psi_{11}(x)\lambda_1(x) - bx_1 \\ \frac{1}{a}\left(\psi_{22}^2(x) + \psi_{32}^2(x) + \psi_{23}^2(x) + \psi_{33}^2(x) + 2\psi_{22}(x)\psi_{23}(x) + 2\psi_{32}(x)\psi_{33}(x)\right) \\ +\psi_{22}(x)\lambda_2(x) + \psi_{23}(x)\lambda_3(x) - ax_2 + ax_3 + x_5 \\ \psi_{23}(x)\lambda_2(x) + \psi_{33}(x)\lambda_3(x) - ax_3 - x_5 - x_6 \\ -\frac{1}{a}\left(\psi_{22}^2(x) + \psi_{32}^2(x) + \psi_{23}^2(x) + \psi_{33}^2(x) + 2\psi_{22}(x)\psi_{23}(x) + 2\psi_{32}(x)\psi_{33}(x)\right) - 2ax_4 + 2x_6 \\ \psi_{23}^2(x) + \psi_{33}^2(x) - 2ax_5 \\ -\psi_{23}^2(x) - \psi_{33}^2(x) - \psi_{22}(x)\psi_{23}(x) - \psi_{32}(x)\psi_{33}(x) - ax_5 - 2ax_6 \\ -\frac{1}{b}\psi_{11}(x)^2 - 2bx_7 \end{pmatrix} \tag{A.9.23}$$

as an alternative to the expression for $\phi(x)$ in Theorem 2.

Proposition A.9.2. *The SLSC curve shape given by loading functions*

$$\tilde{B}_{1:7}(\tau) = \left(\frac{1-e^{-b\tau}}{b\tau} \quad \frac{1-e^{-a\tau}}{a\tau} \quad \frac{1-e^{-a\tau}}{a\tau} - e^{-a\tau} \quad \frac{1-e^{-2a\tau}}{2a\tau} \quad \frac{\tau}{2}e^{-2a\tau} \quad \frac{1}{a}\left(e^{-a\tau} - e^{-2a\tau}\right) \quad \frac{1-e^{-2b\tau}}{2b\tau}\right)$$

with fixed a, b is dynamically consistent with the DTSM with drift $\alpha(\tau,x) = \tilde{B}_{1:7}(\tau)\phi(x)$ and volatility function (41), provided $\phi(x) = (\phi_1(x),\ldots,\phi_7(x))'$ and $\psi(x)$ from (40) satisfy Assumption 1.

Proof of Proposition A.9.2. By Corollary A.7.1, conditions (A.7.37)-(A.7.38) must be verified for $B(\tau) = \tilde{B}_{1:7}(\tau)$ from (43) and $\psi(x)$ from (40). Inserting $\tilde{B}_{1:7}(\tau)$ for $B(\tau)$ in (A.7.37) produces the condition $\alpha(\tau,x) = \tilde{B}_{1:7}(\tau)\phi(x)$ from the Proposition. Inserting $\tilde{B}_{1:7}(\tau)$ and $\psi(x)$ from (40) in the condition on $\sigma(\tau,x)$ in (26) produces (41). The conclusion follows. □

A.10. Relation between SLSC and AFNS

The SLSC model has HJM volatility function (41) which by Theorem 2 is dynamically consistent with the SLSC curve shape (43) under the no-arbitrage condition, with state dynamics

$$dx(t) = \Phi(\theta - x(t))dt + \psi(x(t))'dW(t),$$

where Φ, θ, and $\psi(x)$ are given by (44), (45), and (40), respectively. For state-independent volatilities, i.e., $\psi_{ij}(x) = \psi_{ij}$ (or $\omega_{ij}(x) = \omega_{ij}$), $i,j = 1,2,3$, the locally deterministic state variables, $\bar{\bar{x}}(t) = x_{4:7}(t) = (x_4(t),\ldots,x_7(t))'$, are deterministic, with long-run levels $\bar{\bar{\theta}} = \theta_{4:7} = (\theta_4,\ldots,\theta_7)'$. Hence, if $x_{4:7}(t')$ at some point t' takes the value $\bar{\bar{\theta}}$, then it remains constant

at this level. In this case, since the last four rows of $(\theta - x(t))$ vanish, the stochastic state variables $x_{1:3}(t)$ have drift not depending on the deterministic state variables, so they satisfy

$$\mathrm{d}x_{1:3}(t) = \Phi_{1:3}(\theta_{1:3} - x_{1:3}(t))\mathrm{d}t + \psi' \mathrm{d}W_t\,,$$

with $\Phi_{1:3}$ the upper left 3×3 submatrix of Φ. As in Christensen, Diebold, and Rudebusch (2011), we translate to factors, $\tilde{x}_t$, of zero mean under $\mathbb{Q}$, i.e., $\tilde{x}_t \equiv x_{1:3}(t) - \theta^{\mathbb{Q}}_{1:3}$, where $\theta^{\mathbb{Q}}_{1:3}$ is $\theta_{1:3}$ with $\lambda_{1:3} = 0$ imposed. Thus, using (45), we have

$$\mathrm{d}\tilde{x}_t = \Phi_{1:3}(\tilde{\theta} - \tilde{x}_t)\mathrm{d}t + \psi' \mathrm{d}W_t\,,$$

with

$$\tilde{\theta} = \theta_{1:3} - \theta^{\mathbb{Q}}_{1:3} = \begin{pmatrix} \frac{1}{b}\lambda_1\psi_{11} \\ \frac{1}{a}\left(\lambda_2\psi_{22} + (\lambda_2+\lambda_3)\psi_{23} + \lambda_3\psi_{33}\right) \\ \frac{1}{a}\left(\lambda_2\psi_{23} + \lambda_3\psi_{33}\right) \end{pmatrix}.$$

For $t \geq t'$, all subsequent yield curves assume the shape

$$y(t,\tau) = \tilde{B}_{1:3}(\tau)x_{1:3}(t) + \tilde{B}_{4:7}(\tau)\bar{\tilde{\theta}} = \tilde{B}_{1:3}(\tau)\tilde{x}_t + \tilde{B}(\tau)\begin{pmatrix} \theta^{\mathbb{Q}}_{1:3} \\ \bar{\tilde{\theta}} \end{pmatrix} = \tilde{B}_{1:3}(\tau)\tilde{x}_t - \tilde{A}(\tau),$$

with the function $\tilde{A}(\tau)$ given by

$$\begin{aligned}
-\tilde{A}(\tau) = \tilde{B}(\tau)\Bigg[& \frac{1}{b^2}\psi_{11}^2 \quad \frac{1}{4a^2}(4\omega_{22}+7\omega_{33}+10\omega_{23}) \quad \frac{1}{4a^2}(\omega_{33}+2\omega_{23}) \\
& -\frac{1}{4a^2}(2\omega_{22}+5\omega_{33}+6\omega_{23}) \quad \frac{1}{2a}\omega_{33} \quad -\frac{1}{4a}(3\omega_{33}+2\omega_{23}) \quad -\frac{1}{2b^2}\psi_{11}^2\Bigg]' \\
& = \psi_{11}^2\left(\frac{1}{b^3}\frac{1-e^{-b\tau}}{\tau}\right) + \omega_{22}\left(\frac{1}{a^3}\frac{1-e^{-a\tau}}{\tau}\right) + \omega_{33}\left(\frac{7}{4a^3}\frac{1-e^{-a\tau}}{\tau}\right) + \omega_{23}\left(\frac{10}{4a^3}\frac{1-e^{-a\tau}}{\tau}\right) \\
& + \omega_{33}\left(\frac{1}{4a^3}\frac{1-e^{-a\tau}}{\tau} - \frac{1}{4a^2}e^{-a\tau}\right) + \omega_{23}\left(\frac{1}{2a^3}\frac{1-e^{-a\tau}}{\tau} - \frac{1}{2a^2}e^{-a\tau}\right) + \omega_{22}\left(-\frac{1}{4a^3}\frac{1-e^{-2a\tau}}{\tau}\right) \\
& + \omega_{33}\left(-\frac{5}{8a^3}\frac{1-e^{-2a\tau}}{\tau}\right) + \omega_{23}\left(-\frac{3}{4a^3}\frac{1-e^{-2a\tau}}{\tau}\right) + \omega_{33}\left(\frac{\tau}{4a}e^{-2a\tau}\right) \\
& + \omega_{33}\left(\frac{3}{4a^2}e^{-2a\tau} - \frac{3}{4a^2}e^{-a\tau}\right) + \omega_{23}\left(\frac{1}{2a^2}e^{-2a\tau} - \frac{1}{2a^2}e^{-a\tau}\right) + \psi_{11}^2\left(\frac{1}{4b^3}\frac{1-e^{-2b\tau}}{\tau}\right) \\
& = \psi_{11}^2\left(\frac{1}{b^3}\frac{1-e^{-b\tau}}{\tau} - \frac{1}{4b^3}\frac{1-e^{-2b\tau}}{\tau}\right) + \omega_{22}\left(\frac{1}{a^3}\frac{1-e^{-a\tau}}{\tau} - \frac{1}{4a^3}\frac{1-e^{-2a\tau}}{\tau}\right)
\end{aligned}$$

$$+\omega_{33}\left(\frac{2}{a^3}\frac{1-e^{-a\tau}}{\tau}-\frac{5}{8a^3}\frac{1-e^{-2a\tau}}{\tau}-\frac{1}{a^2}e^{-a\tau}+\frac{1}{4a}\tau e^{-2a\tau}+\frac{3}{4a^2}e^{-2a\tau}\right)$$
$$+\omega_{23}\left(\frac{3}{a^3}\frac{1-e^{-a\tau}}{\tau}-\frac{1}{a^2}e^{-a\tau}-\frac{3}{4a^3}\frac{1-e^{-2a\tau}}{\tau}+\frac{1}{2a^2}e^{-2a\tau}\right).$$

This corresponds closely to the AFNS yield curve shape. More precisely, it corresponds to the model that would result from carrying out the program from Footnote 6 of Christensen, Diebold, and Rudebusch (2011), modifying the mean-reversion matrix $K^{\mathbb{Q}}$ (their notation) to include a sufficiently small $\varepsilon > 0$ in the upper left corner. In any case, the general SLSC specification accommodates time-varying deterministic state variables, $\bar{\bar{x}}(t) \neq \bar{\bar{\theta}}$, and thus yield curves outside the AFNS class.

Finally, for uncorrelated state variables, as in the independent-factor AFNS model, i.e., $\omega_{22} = \psi_{22}^2$, $\omega_{33} = \psi_{33}^2$, $\omega_{23} = 0$, the yield curve under the affine restriction ($\bar{\bar{x}}(t) = \bar{\bar{\theta}}$) reduces to $y(t,\tau) = \tilde{B}_{1:3}(\tau)\tilde{x}_t - \tilde{A}(\tau)$ with

$$-\tilde{A}(\tau) = \psi_{11}^2\left(\frac{1}{b^3}\frac{1-e^{-b\tau}}{\tau}-\frac{1}{4b^3}\frac{1-e^{-2b\tau}}{\tau}\right)+\psi_{22}^2\left(\frac{1}{a^3}\frac{1-e^{-a\tau}}{\tau}-\frac{1}{4a^3}\frac{1-e^{-2a\tau}}{\tau}\right)$$
$$+\psi_{33}^2\left(\frac{2}{a^3}\frac{1-e^{-a\tau}}{\tau}-\frac{5}{8a^3}\frac{1-e^{-2a\tau}}{\tau}-\frac{1}{a^2}e^{-a\tau}+\frac{1}{4a}\tau e^{-2a\tau}+\frac{3}{4a^2}e^{-2a\tau}\right).$$

A.11. Proof of Theorem D.1.1

The statement of Theorem D.1.1 is in Appendix D.2, on trading off hedging error bias and variance.

Proof of Theorem D.1.1. Writing out the objective function from (D.2.2), we have

$$\mathbb{E}_t\left[\left(r_{t+1}^* - w' r_{t+1}\right)^2\right] = \mathbb{E}_t\left[\left(r_{t+1}^*\right)^2\right] + w'\mathbb{E}_t\left(r_{t+1}r_{t+1}'\right)w - 2w'\mathbb{E}_t\left(r_{t+1}r_{t+1}^*\right). \tag{A.11.1}$$

The solution to the first order conditions with respect to w is the linear projection of r_{t+1}^* on r_{t+1},

$$\widetilde{w} = \left[\mathbb{E}_t\left(r_{t+1}r_{t+1}'\right)\right]^{-1}\mathbb{E}_t\left(r_{t+1}r_{t+1}^*\right). \tag{A.11.2}$$

The conditional expected returns of the hedging instruments (6) are

$$\mathbb{E}_t\left[r_{t+1}\right] = -\mathcal{T}B\left(\mu_{t+1|t} - \mu_{t|t}\right) \tag{A.11.3}$$

when $\mathbb{E}_t[\Delta\varepsilon_{t+1}] = 0$. Using (5) to rewrite (6) as

$$r_{t+1} = -\mathcal{T}\left(Bf_{t+1} + \varepsilon_{t+1} - \left(y_t - \mu\right)\right), \tag{A.11.4}$$

the conditional variance is given by

$$\text{var}_t[r_{t+1}] = \mathcal{T}\left(B\Sigma_{t+1|t}B' + \Psi\right)\mathcal{T}. \tag{A.11.5}$$

From (A.11.3) and (A.11.5), the second moment is

$$\begin{aligned}
\mathbb{E}_t\left[r_{t+1}r_{t+1}'\right] &= \text{var}_t[r_{t+1}] + \mathbb{E}_t[r_{t+1}]\mathbb{E}_t[r_{t+1}]' \\
&= \mathcal{T}\left(B\Sigma_{t+1|t}B' + \Psi\right)\mathcal{T} + \mathcal{T}B\left(\mu_{t+1|t} - \mu_{t|t}\right)\left(\mu_{t+1|t} - \mu_{t|t}\right)'B'\mathcal{T} \\
&= \mathcal{T}\left[B\left(\Sigma_{t+1|t} + \left(\mu_{t+1|t} - \mu_{t|t}\right)\left(\mu_{t+1|t} - \mu_{t|t}\right)'\right)B' + \Psi\right]\mathcal{T}.
\end{aligned} \tag{A.11.6}$$

Similarly, the conditional expected target return is

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

when $\mathbb{E}_t\left[\Delta\varepsilon_{t+1}^*\right] = 0$. Combining with (A.11.4), the conditional covariance is

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

This implies that

$$\begin{aligned}
\mathbb{E}_t\left[r_{t+1}r_{t+1}^*\right] &= \text{cov}_t\left[r_{t+1}, r_{t+1}^*\right] + \mathbb{E}_t[r_{t+1}]\mathbb{E}_t\left[r_{t+1}^*\right] \\
&= \mathcal{T}B\Sigma_{t+1|t}(\tau b)_* + \mathcal{T}B\left(\mu_{t+1|t} - \mu_{t|t}\right)\left(\mu_{t+1|t} - \mu_{t|t}\right)'(\tau b)_* \\
&= \mathcal{T}B\left(\Sigma_{t+1|t} + \left(\mu_{t+1|t} - \mu_{t|t}\right)\left(\mu_{t+1|t} - \mu_{t|t}\right)'\right)(\tau b)_*.
\end{aligned} \tag{A.11.7}$$

Inserting (A.11.6) and (A.11.7) in (A.11.2),

$$\begin{aligned}
\widetilde{w} = \mathcal{T}^{-1}&\left[B\left(\Sigma_{t+1|t} + \left(\mu_{t+1|t} - \mu_{t|t}\right)\left(\mu_{t+1|t} - \mu_{t|t}\right)'\right)B' + \Psi\right]^{-1} \\
&\times B\left(\Sigma_{t+1|t} + \left(\mu_{t+1|t} - \mu_{t|t}\right)\left(\mu_{t+1|t} - \mu_{t|t}\right)'\right)(\tau b)_*.
\end{aligned} \tag{A.11.8}$$

From Woodbury's lemma,

$$(A + USV)^{-1} = A^{-1} - A^{-1}U\left(S^{-1} + VA^{-1}U\right)^{-1}VA^{-1}$$

for conformable matrices such that A, S, and $S^{-1}+VA^{-1}U$ are invertible. Thus,

$$\begin{aligned}(A+USV)^{-1}US &= A^{-1}US - A^{-1}U\left(S^{-1}+VA^{-1}U\right)^{-1}VA^{-1}US \\ &= A^{-1}U\left[I-\left(S^{-1}+VA^{-1}U\right)^{-1}VA^{-1}U\right]S \\ &= A^{-1}U\left(S^{-1}+VA^{-1}U\right)^{-1}\left(S^{-1}+VA^{-1}U-VA^{-1}U\right)S \\ &= A^{-1}U\left(S^{-1}+VA^{-1}U\right)^{-1}.\end{aligned}$$

Application of this to (A.11.8) with $A=\Psi$, $U=B$, $S=\Sigma_{t+1|t}+\left(\mu_{t+1|t}-\mu_{t|t}\right)\left(\mu_{t+1|t}-\mu_{t|t}\right)'$, and $V=B'$ yields

$$\widetilde{w}=\mathcal{T}^{-1}\Psi^{-1}B\left[\left(\Sigma_{t+1|t}+\left(\mu_{t+1|t}-\mu_{t|t}\right)\left(\mu_{t+1|t}-\mu_{t|t}\right)'\right)^{-1}+B'\Psi^{-1}B\right]^{-1}(\tau b)_*\,, \tag{A.11.9}$$

which is (D.2.3).

Imposing that weights sum to unity is done using Lemma 1. The original problem (D.2.2) is unconstrained. From (A.11.1), in the notation of (A.2.2), $A=2\mathbb{E}_t\left[r_{t+1}r_{t+1}'\right]$, $g=2\mathbb{E}_t\left[r_{t+1}r_{t+1}^*\right]$, and $D=0$, the latter due to relaxation of generalized duration matching. By (A.2.4), the optimal portfolio under value matching is

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

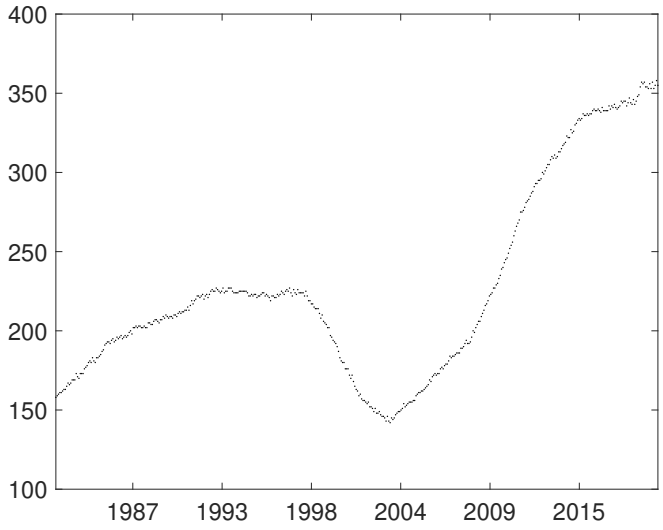
with $\widetilde{w}$ from (A.11.9), and $\Lambda_t=A^{-1}=\left[\mathbb{E}_t\left(r_{t+1}r_{t+1}'\right)\right]^{-1}/2$. Inserting the expression (A.11.6) for $\mathbb{E}_t\left(r_{t+1}r_{t+1}'\right)$ produces Λ_t from the Theorem. □

B. Target Asset

On the last trading day of each month, we select among all non-callable and non-flower bonds the issues with maturities closest to two, five, and ten years. Figure B.1 shows the evolution over time in the resulting number of Treasuries used in the construction of our target asset.

Figure B.2 shows characteristics of the five-year coupon bonds. The upper left exhibit shows the term to maturity for each selected bond in the time series. Since many bonds are issued and mature close to the 15$^{\text{th}}$ of the month, many maturities are a half month above or below five years. The upper right exhibit shows that the received coupon rates

Figure B.1: Number of Treasuries



This figure shows the number of Treasuries considered in the empirical application in each month from January, 1983, through December, 2019.

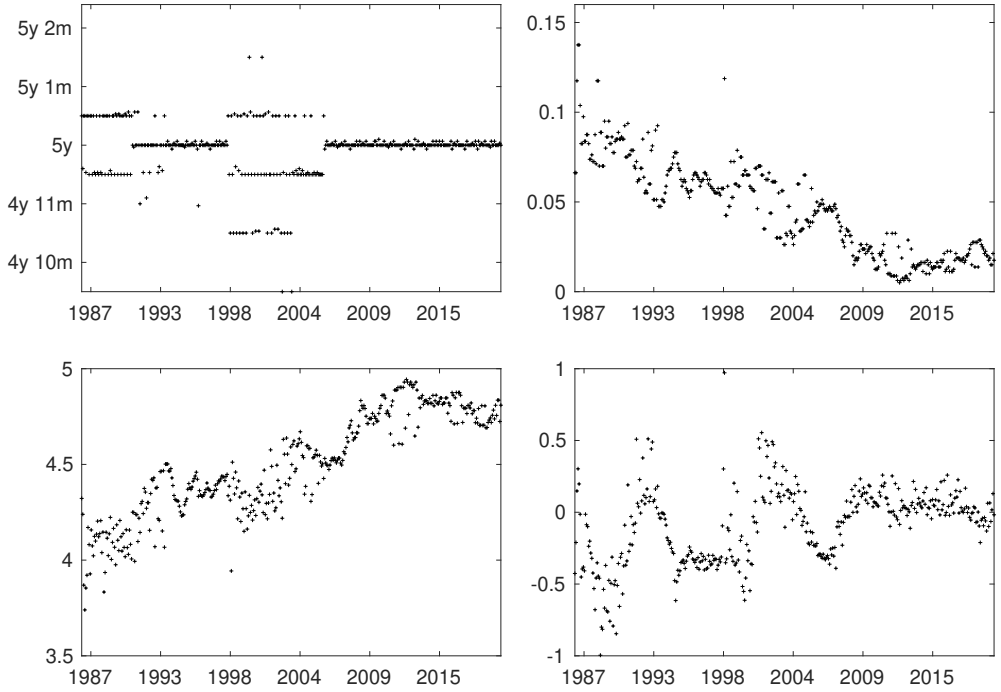
decrease over the sample period.

The lower left exhibit shows the resulting durations of the selected five-year coupon bonds, which increase from below 4.0 to above 4.9 due to the drop in rates. The target portfolio further includes the two-year and ten-year coupon bonds, and the corresponding coupon rates are shown in Figure B.3, along with the resulting target durations.

C. Basic and Generalized Duration Matching

Basic duration matching (line 2 of Table 2) requires only two instruments in order to simultaneously match duration and ensure that hedge portfolio weights sum to one. The target portfolio usually has duration between three and five years, so we use the three- and five-year zero-coupon bonds for immunization, except that we replace the five-year with the two-year zero-coupon bond when portfolio duration falls short of three years, cf. Figure B.3. Next, to match generalized rather than basic durations (lines 3 through 26 of Table 2), loadings B and idiosyncratic variances Ψ are estimated in the relevant model. Based on B, the k-vector $(\tau b)_*$ of target generalized durations is calculated by interpolation (see Appendix A.3) and used in Theorem 1, together with estimated B and Ψ, to construct the hedging weights. As discount functions, coupon rates, and target durations (Figures B.2 and B.3) vary over time, so does $(\tau b)_*$ (see (A.3.2)), and therefore

Figure B.2: Characteristics of five-year coupon bond



This figure shows characteristics of the five-year coupon bonds used in the hedging target. The upper exhibits show time to maturity (left) and coupon rate (right), and the lower left exhibit shows duration. The characteristics were retrieved from the CRSP monthly Treasury files, each month selecting the coupon bond with maturity closest to five years, given a liquidity condition. The lower right exhibit shows the percentage excess of the price of the selected bond as implied by the FED yield curve above the CRSP recorded price. The monthly data span the period from January, 1987, through December, 2019.

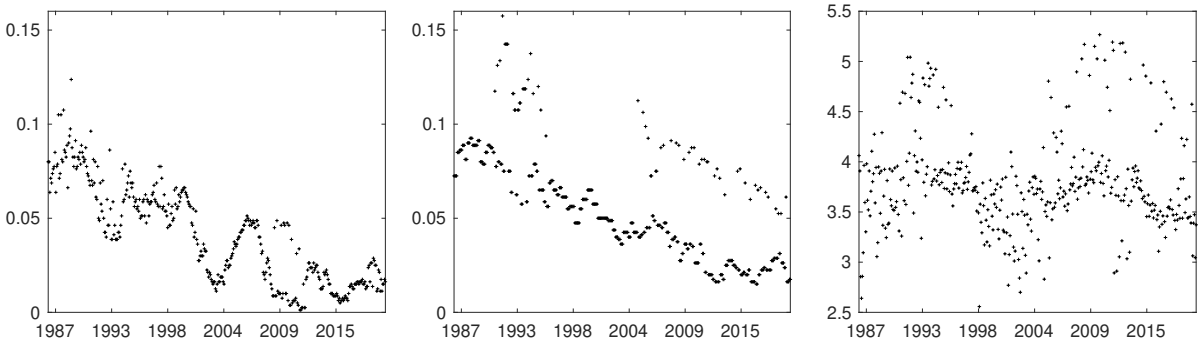
hedging weights. Further time-variation in weights is induced via estimated B and Ψ in the rolling case.

D. Robustness Checks

D.1. Alternative Target Asset

In the main text, target asset prices are based on CRSP data. The issue arises that raw CRSP prices (bid-ask midpoints plus accrued interest) might reflect frictions, microstructure noise, and other features not present in the FED yields used for model estimation. For comparison, we also used the FED yields to set the prices of the bonds entering the target asset, rather than using the CRSP recorded 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

Figure B.3: Characteristics of portfolio target



This figure shows the coupon rates of the two-year bonds (left exhibit) and the ten-year bonds (center exhibit) in the target portfolio, which is the combination of the $(2,5,10)$-year coupon bonds in proportions $(-1,3,-1)$. The right exhibit shows the durations of the target portfolio. The characteristics were retrieved from the CRSP monthly Treasury files, each month selecting the coupon bonds with maturities closest to two, five, and ten years, given a liquidity condition. The monthly data span the period from January, 1987, through December, 2019.

we use as an alternative target for one-month ahead hedging using the corresponding eight zero-coupon bonds. The differences between raw CRSP prices and the FED valuations fluctuate within a 1% band, shown in the lower right exhibit of Figure B.2. The possibility exists that comparisons of methods depend on which are better at picking up these discrepancies, likely stemming from different noise in CRSP prices and FED yields. Results on hedging performance when pricing the target asset based on the FED yields, to be compared with the results in Table 2, based on CRSP data for the target, are shown in Table D.1. The performance of all models improves, compared to that based on raw target returns, with the largest improvement seen in the SLSC models, which clearly dominate all other approaches.

D.2. Trading Off Hedging Error Bias and Variance

Our focus has been on minimization of conditional hedging error variance (1). The generalized duration matching approach immunizes factor exposure, and the portfolio in Theorem 1 minimizes remaining hedging error variance (9). An alternative to (1) is the minimization of conditional mean squared hedging error,

$$\min_{w} \mathbb{E}_t \left[\left(r^*_{t+1} - w' r_{t+1} \right)^2 \right], \tag{D.2.1}$$

Table D.1: Hedging performance using FED yields to price target

The target is a portfolio of $(2,5,10)$-year coupon bonds in the proportions $(-1,3,-1)$. The prices used to compute the target return are set using the FED yields. Statistics in line 1 are for the unhedged target return, and in the remainder of the table for hedging errors from each of the methods considered for construction of the hedge portfolio with value matching from Theorem 1. The columns report the average (or bias), standard deviation, root mean squared error, and mean absolute error. Results are in basis points (0.01%) per month. An S indicates that a given method provides a statistically significant improvement over traditional duration matching at the 5% level, and MCS that a method is included in the Model Confidence Set at 5% (only conducted for the rolling strategies).

	Model	Bias	Std. dev.	RMSE	MAE
1	Target movement	53.84	143.77	153.52	119.94
2	Duration matching	8.03	55.60	56.18	40.73
3	Unrestricted 3-factor Full period	4.31	47.34	47.53	34.45
4	Unrestricted 3-factor Rolling 4-year	4.03	46.62	46.79 (S,-)	34.13 (S,-)
5	Nelson-Siegel Full period	3.91	48.13	48.29	34.86
6	Nelson-Siegel Rolling 4-year	4.42	49.74	49.94 (S,-)	36.09 (S,-)
7	Unrestricted 4-factor Full period	4.61	28.77	29.13	22.56
8	Unrestricted 4-factor Rolling 4-year	3.21	36.34	36.48 (S,-)	24.38 (S,-)
9	Augmented NS Full period	3.69	29.94	30.17	22.67
10	Augmented NS Rolling 4-year	2.41	26.28	26.39 (S,-)	16.89 (S,-)
11	Unrestricted 1-factor, $\tilde{y}$ Full period	6.07	55.18	55.52	39.27
12	Unrestricted 1-factor, $\tilde{y}$ Rolling 4-year	6.38	51.90	52.29	37.16 (S,-)
13	ANS-extended Vasicek, $\tilde{y}$ Full period	1.64	15.87	15.96	12.06
14	ANS-extended Vasicek, $\tilde{y}$ Rolling 4-year	0.56	22.38	22.39 (S,-)	15.28 (S,-)
15	ANS-extended Vasicek, filter Full period	2.06	62.62	62.65	45.06
16	ANS-extended Vasicek, filter Rolling 4-year	2.21	56.64	56.68	40.08
17	ANS-extended Vasicek, restricted Full period	1.79	56.59	56.62	40.54
18	ANS-extended Vasicek, restricted Rolling 4-year	2.41	58.76	58.81	42.21
19	Unrestricted 3-factor, $\tilde{y}$ Full period	2.87	45.53	45.62	32.55
20	Unrestricted 3-factor, $\tilde{y}$ Rolling 4-year	2.76	50.01	50.09 (S,-)	34.78 (S,-)
21	SLSC, $\tilde{y}$ Full period	0.72	8.71	8.74	6.88
22	SLSC, $\tilde{y}$ Rolling 4-year	0.74	9.08	9.11 (S,MCS)	7.12 (S,MCS)
23	SLSC, filter Full period	0.75	9.49	9.52	7.36
24	SLSC, filter Rolling 4-year	0.86	8.31	8.36 (S,MCS)	6.40 (S,MCS)
25	SLSC, restricted Full period	3.64	44.85	45.00	32.59
26	SLSC, restricted Rolling 4-year	4.18	45.72	45.91 (S,-)	33.22 (S,-)

thus trading off conditional hedging error bias and variance, corresponding to the RMSE criterion from the performance evaluation. For zero conditional mean hedging errors, the alternative objectives (1) and (D.2.1) coincide. From (8), under generalized duration matching, $w'\mathcal{T}B = (\tau b)_*$, the hedging errors are of conditional mean zero if the idiosyncratic errors are, i.e., if $\mathbb{E}_t\left(\Delta\varepsilon_{t+1,\tau}\right) = 0$, for all τ, and $\mathbb{E}_t\left(\Delta\varepsilon^*_{t+1}\right) = 0$. As a further generalization, we relax the generalized duration matching constraint, thus admitting some factor risk in the hedged position. Minimizing the objective (D.2.1) without the generalized duration matching constraint, the optimal strategy strikes a balance between minimizing factor and idiosyncratic variance (by relaxation of the constraint), while trading off the total against resulting bias (by using the RMSE criterion for the optimization).

The following theorem, supplementing Theorem 1, provides the optimal strategy under the zero conditional mean idiosyncratic error condition. It involves the conditional factor prediction, $\mu_{t|t} = \mathbb{E}_t(f_t)$, based on yield data through t, the one step ahead prediction, $\mu_{t+1|t} = \mathbb{E}_t(f_{t+1})$, and the conditional variance, $\Sigma_{t+1|t} = \mathrm{var}_t(f_{t+1})$, all from the Kalman filter.[44]

Theorem D.2.1. *The immunization portfolio $\widetilde{w}$ that minimizes conditional mean squared hedging error*

$$\min_w \mathbb{E}_t\left[\left(r^*_{t+1} - w'r_{t+1}\right)^2\right] \tag{D.2.2}$$

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

$$\widetilde{w} = \mathcal{T}^{-1}\Psi^{-1}B\left[\left(\Sigma_{t+1|t} + \left(\mu_{t+1|t} - \mu_{t|t}\right)\left(\mu_{t+1|t} - \mu_{t|t}\right)'\right)^{-1} + B'\Psi^{-1}B\right]^{-1}(\tau b)_* \,. \tag{D.2.3}$$

Further imposing value matching, $w'\iota = 1$, changes the optimal portfolio to

$$w^* = \widetilde{w} + \left(1 - \widetilde{w}'\iota\right)\frac{\Lambda_t\iota}{\iota'\Lambda_t\iota}, \tag{D.2.4}$$

with $\widetilde{w}$ from (D.2.3), and

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

In general, if factors are predictable, then this can be exploited in hedging, as shown in Theorem D.2.1. Predictable factors arise under dynamic consistency, leading to the

[44] The proof is in Appendix A.11.

filtering approach from Section 4.2. In the special cases of either (i) no factor exposure in hedging errors due to generalized duration matching, or (ii) excessive uncertainty about future factors ($\Sigma_{t+1|t}$ tending to infinity, so no predictability after all), the prediction components drop out, i.e., the strategy (D.2.3) reduces to (11), and (D.2.4) to (12).

The performance of the hedging approach from Theorem D.2.1 is documented in Table D.2. All the previous (third stage) specifications estimated using the Kalman filter are considered. Throughout, performance is poorer in Table D.2 than in the corresponding cases in Table 2. This includes performance according to the RMSE criterion, although the objective (D.2.2) is targeting this. Interestingly, within Table D.2, there is no penalty to imposing the affine restriction. Still, the evidence is that it pays off to remove factor exposure, i.e., perform generalized duration matching, and target remaining idiosyncratic variance, Table 2, rather than trading this off against average hedging error. Parsimony is again the likely reason. The strategies from Theorem 1 involve only estimated B and Ψ, whereas those from Theorem D.2.1 involve all model parameters, via the output from the Kalman filter.

E. Statistical Comparison of Hedging Performance

For the analysis in Section 6.2, the loss differentials from the i^{th} approach relative to the benchmark (denoted by b) are

$$d_{i,t} = \left(r_t^* - w_{i,t-1}' r_t\right)^2 - \left(r_t^* - w_{b,t-1}' r_t\right)^2$$

in the MSE case, with $|\cdot|$ replacing $(\cdot)^2$ for MAE. Following Diebold and Mariano (1995) and Giacomini and White (2006), the test of equal hedging performance is conducted using $S_{i,b} = T^{1/2}\bar{d}\hat{V}^{-1/2}$, where $\bar{d} = T^{-1}\sum_{t=1}^{T} d_{i,t}$, and $\hat{V}$ is a HAC estimate of the long-run variance, using the data-dependent bandwidth selection of Andrews (1991) based on an AR(1) approximation and a Bartlett kernel.

For the Model Confidence Set (MCS) procedure of Hansen, Lunde, and Nason (2011), denote by M_0 the set of all competing approaches. The procedure is conducted recursively based on an equivalence test for arbitrary $M \subseteq M_0$ and an elimination rule which identifies

and removes an approach from M in case the equivalence test rejects. The equivalence test is based on pairwise comparisons using the t-statistic $S_{i,j}$, for all $i, j \in M$, and the range statistic $T_M = \max_{i,j \in M} \{|S_{i,j}|\}$. If the test rejects, the k^{th} approach is eliminated from M, where $k = \text{argmax}_{i \in M} \sup_{j \in M} \{S_{i,j}\}$. Following Hansen, Lunde, and Nason (2011), we implement the procedure using a block bootstrap and 10^4 replications.

Table D.2: Hedging performance under RMSE objective

The target is a portfolio of $(2, 5, 10)$-year coupon bonds in the proportions $(-1, 3, -1)$. The first line shows statistics for the unhedged target return, and the remainder of the table for hedging errors from each of the methods considered for construction of the hedge portfolio under the root mean squared error (RMSE) objective with value matching from Theorem D.1.1. The columns report the average (or bias), standard deviation, RMSE, and mean absolute error. Results are in basis points (0.01%) per month. An S indicates that a given method provides a statistically significant improvement over traditional duration matching at the 5% level.

Model	Bias	Std. dev.	RMSE	MAE
Target Movement	49.20	149.44	157.33	122.63
Duration matching	2.84	65.90	65.96	48.72
ANS-extended Vasicek, filter Full period	3.01	72.98	73.05	53.43
ANS-extended Vasicek, filter Rolling 4-year	2.55	82.78	82.82	59.73
ANS-extended Vasicek, restricted Full period	2.59	65.79	65.84	48.32
ANS-extended Vasicek, restricted Rolling 4-year	1.80	83.32	83.34	60.49
SLSC, filter Full period	-0.32	59.66	59.66	43.08
SLSC, filter Rolling 4-year	-0.49	58.49	58.50 (S)	41.88
SLSC, restricted Full period	-0.03	59.63	59.63	43.10
SLSC, restricted Rolling 4-year	0.23	60.75	60.75	43.82

F. Estimation

F.1. Factor models

The implementation of each strategy considered in Table 2 involves some choices, and except duration matching, each includes an estimation step. First, the classical factor analysis is applied to (5). Maximization of the factor analysis log likelihood function over the full period produces the idiosyncratic standard deviations $\sqrt{\Psi_i} \cdot 1000$ shown in

lines 1, 2, 3, and 5 of Table 3, along with the maximized log-likelihood value, number of parameters, and standard information criteria, AIC and BIC. For the model with $k = 3$ factors, the loadings $B_{ij} = B_j(\tau_i)$ are given in percent in Table F.1, and a visualization of the three loading functions $B_j(\cdot)$, $j = 1, 2, 3$ is given in the left exhibit of Figure F.1. The level, slope, and curvature pattern highlighted by Litterman and Scheinkman (1991) is evident. In the reported rotation, factor j explains the j^{th} highest proportion of total variation. Specifically, with B_{ij}^2/ϑ_i^2 the proportion of the total variance ϑ_i^2 of y_{τ_i} from Table 1 explained by factor j, and $\frac{1}{m}\sum_{i=1}^{m} B_{ij}^2/\vartheta_i^2$ the average proportion across maturities, the latter is j^{th} highest for factor j. To avoid Heywood cases (factors explaining more than total variation for a given maturity, i.e., communality $\sum_{j=1}^{k} B_{ij}^2/\vartheta_i^2$ exceeding unity), a lower bound of $10^{-4}\vartheta_i^2$ is imposed on Ψ_i, for each maturity τ_i, in all models.

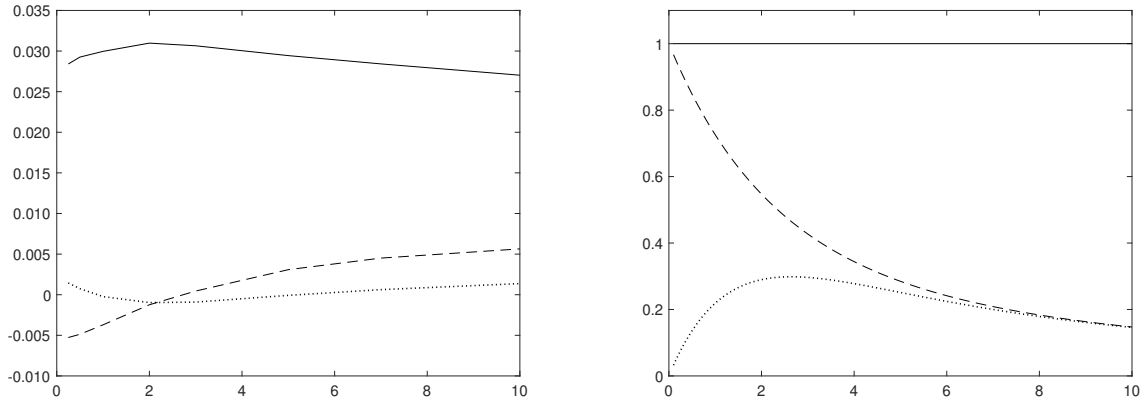
Table F.1: Loading functions in unrestricted three-factor model
This table shows the loading functions $B_j(\tau)$, $j = 1, 2, 3$ given by the columns (in percent) of the loading matrix B in the unrestricted three-factor model, and displayed graphically in Figure F.1, left exhibit.

τ	$B_1(\tau)$	$B_2(\tau)$	$B_3(\tau)$
3 mos.	2.84	0.53	-0.14
6 mos.	2.93	0.49	-0.07
12 mos.	3.00	0.37	0.02
2 yrs.	3.10	0.12	0.10
3 yrs.	3.06	-0.05	0.09
5 yrs.	2.94	-0.31	0.01
7 yrs.	2.84	-0.45	-0.06
10 yrs.	2.70	-0.56	-0.14

For the restricted models, corresponding to parsimonious yield curve shape, lines 4 and 6 of Table 3, the factor analysis log likelihood function is maximized subject to the restriction that B takes NS or ANS form, depending only on a, and with $k = 3$ and 4, respectively. Thus, the parameters estimated are (μ, a, Σ, Ψ). Although Σ does not enter the hedging weights, cf. Section 3.1, estimation of the restricted models allows for correlated factors. For NS, from the estimated Σ (not reported), the correlation between the level and slope factors is -0.08, between level and curvature 0.61, and between slope and curvature 0.49.

For the models for $\tilde{y}$, lines 7-8 and 11-12 of Table 3, yields are replaced by slope-

Figure F.1: Loading functions in three-factor models



The left exhibit shows the loadings $B_j(\tau)$, $j = 1, 2, 3$, in the unrestricted three-factor model, as functions of maturity, τ. Numerical values (in percent) are given in Table F.1. The right exhibit shows the NS loading functions $\tilde{B}_j(\tau)$, $j = 1, 2, 3$, from (15).

adjusted yield changes in the factor analysis log likelihood function. The second-stage models (Section 4.1), i.e., lines 8 and 12 of Table 3, are restricted according to (49).

The remaining models in Table 3, lines 9-10 and 13-14, relate to third-stage Kalman filtering (Section 4.2). In the estimations, rather than the Euler discretization of the transition equation (18) (see (52) and (54)), we use the exact discretization derived in Appendix F.2, and a low-storage square-root filter (see Appendix F.3).

F.2. Exact state transition for dynamically consistent specifications

Here, we derive the exact discrete-time state process for the third-stage approach from Section 4.2. We focus on the case of affine drift and state-independent volatility in (18), i.e.,

$$\mathrm{d}x(t) = \Phi(\theta - x(t))\mathrm{d}t + \psi' \mathrm{d}W(t).$$

To solve the SDE, recall that for the ansatz $\mathrm{e}^{\Phi t}$ we have

$$\mathrm{d}\left(\mathrm{e}^{\Phi t} x(t)\right) = \mathrm{e}^{\Phi t}\Phi x(t)\,\mathrm{d}t + \mathrm{e}^{\Phi t}\mathrm{d}x(t) = \mathrm{e}^{\Phi t}\Phi\theta dt + \mathrm{e}^{\Phi t}\psi' \mathrm{d}W_t\,.$$

Integrating from t to $t + \Delta$ produces

$$x(t+\Delta) = \mathrm{e}^{-\Phi\Delta}x(t) + \int_0^\Delta \mathrm{e}^{-\Phi u}\mathrm{d}u \cdot \Phi\theta + \int_0^\Delta \mathrm{e}^{-\Phi u}\psi' \mathrm{d}W(t+u)\,. \tag{F.2.1}$$

Define the function $H(u) = \mathrm{e}^{-\Phi u}$, which has $H'(u) = -H(u)\Phi$. Then $\int H(u)\mathrm{d}u = -H(u)\Phi^{-1}$, so $\int_0^\Delta \mathrm{e}^{-\Phi u}\mathrm{d}u \cdot \Phi = (I - H(\Delta))$. It follows that the exact discrete-time state process is

$$x(t+\Delta) = H(\Delta)x(t) + (I - H(\Delta))\theta + v(t+\Delta), \tag{F.2.2}$$

with $v(t+\Delta) = \int_0^\Delta H(u)\psi'\mathrm{d}W(t+u)$. For data with time increment Δ between observations, the terms $v(t+\Delta)$ are based on increments to the driving Wiener processes over non-overlapping intervals, so they are serially independent innovations. Further, by linearity and the Itô isometry, $v(t+\Delta) \sim N_k(0, \Omega(\Delta))$, with $\Omega(\Delta) = \int_0^\Delta H(u)\psi'\psi H(u)'\,\mathrm{d}u$.

The SLSC model has HJM volatility function (41), which by Theorem 2 is dynamically consistent with the SLSC curve shape (43) under the no-arbitrage condition, with state dynamics where Φ, θ, and ψ are given by (44), (45), and (40), respectively. To calculate $H(u)$ for Φ from (44), let $d(\Phi)$ be the diagonal of Φ, and $\widetilde{\Phi} = \Phi - d(\Phi)$ the matrix containing the off-diagonal elements. Then $\widetilde{\Phi}$ is nilpotent of degree four, and the only non-zero entry in $\widetilde{\Phi}^3$ is $[\widetilde{\Phi}^3]_{2,5} = -a^2$. By the rules of matrix exponentials, we find that

$$H(u) = \mathrm{e}^{-\Phi u} = \mathrm{e}^{-d(\Phi)u}\mathrm{e}^{-\widetilde{\Phi}u} = \mathrm{e}^{-d(\Phi)u}\left(I - \widetilde{\Phi}u + \widetilde{\Phi}^2u^2/2 - \widetilde{\Phi}^3u^3/6\right)$$

$$= \begin{pmatrix}
\mathrm{e}^{-bu} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \mathrm{e}^{-au} & au\mathrm{e}^{-au} & 0 & (a^2u^3/6 - au^2/2 + u)\mathrm{e}^{-2au} & -au^2\mathrm{e}^{-2au}/2 & 0 \\
0 & 0 & \mathrm{e}^{-au} & 0 & (au^2/2 - u)\mathrm{e}^{-2au} & -u\mathrm{e}^{-2au} & 0 \\
0 & 0 & 0 & \mathrm{e}^{-2au} & -au^2\mathrm{e}^{-2au} & 2u\mathrm{e}^{-2au} & 0 \\
0 & 0 & 0 & 0 & \mathrm{e}^{-2au} & 0 & 0 \\
0 & 0 & 0 & 0 & -au\mathrm{e}^{-2au} & \mathrm{e}^{-2au} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \mathrm{e}^{-2bu}
\end{pmatrix}.$$

This is used in (F.2.2). For $\Omega = \Omega(\Delta)$, the innovation variance, note that the lower left 4×3 submatrix of $H(u)$ is zero. For ψ' from (40), the lower 4×4 submatrix of ψ' is zero, as well. It follows that the last four rows of $H(u)\psi'$ are zero, and therefore only the upper left 3×3 submatrix $\Omega_{1:3}(\Delta)$ of $\Omega(\Delta)$ is non-zero. This corresponds to the stochastic state variables. Thus, the exact state transition in the SLSC model, which we use in our empirical work, in place of the Euler discretization (52), is given by (F.2.2), with this $\Omega(\Delta)$ as the variance-covariance matrix of the discrete-time transition shocks, and $H(\cdot)$ given

above.

As an example, in case of uncorrelated state variables (slope and curvature), $\psi_{23} = \psi_{32} = 0$,

$$\begin{aligned}
\Omega_{1:3}(\Delta) &= \int_0^\Delta H_{1:3}(u)\,\mathrm{diag}\left(\psi_{11}^2, \psi_{22}^2, \psi_{33}^2\right) H_{1:3}(u)'\,\mathrm{d}u \\
&= \int_0^\Delta \begin{pmatrix} \psi_{11}^2 e^{-2bu} & 0 & 0 \\ 0 & \psi_{22}^2 e^{-2au} + \psi_{33}^2 a^2 u^2 e^{-2au} & \psi_{33}^2 a u e^{-2au} \\ 0 & \psi_{33}^2 a u e^{-2au} & \psi_{33}^2 e^{-2au} \end{pmatrix} \mathrm{d}u \\
&= \begin{pmatrix} \psi_{11}^2 \frac{1-e^{-2b\Delta}}{2b\Delta} & 0 & 0 \\ 0 & \psi_{22}^2 \frac{1-e^{-2a\Delta}}{2a\Delta} + \psi_{33}^2 \left[\frac{1-e^{-2a\Delta}}{2a\Delta} - e^{-2a\Delta} - a\Delta e^{-2a\Delta}\right]/2 & \psi_{33}^2 \left[\frac{1-e^{-2a\Delta}}{2a\Delta} - e^{-2a\Delta}\right]/2 \\ 0 & \psi_{33}^2 \left[\frac{1-e^{-2a\Delta}}{2a\Delta} - e^{-2a\Delta}\right]/2 & \psi_{33}^2 \frac{1-e^{-2a\Delta}}{2a\Delta} \end{pmatrix} \Delta \\
&= \begin{pmatrix} \psi_{11}^2 B_7(\Delta) & 0 & 0 \\ 0 & \psi_{22}^2 B_4(\Delta) + \psi_{33}^2 [B_4(\Delta)/2 - B_5(\Delta)(1/\Delta + a)] & \psi_{33}^2 (B_4(\Delta)/2 - B_5(\Delta)/\Delta) \\ 0 & \psi_{33}^2 (B_4(\Delta)/2 - B_5(\Delta)/\Delta) & \psi_{33}^2 B_4(\Delta) \end{pmatrix} \Delta,
\end{aligned}$$

with $H_{1:3}(u)$ the upper left 3×3 submatrix of $H(u)$. The remaining entries of $\Omega(\Delta)$ are zero. Further, in (F.2.2), $H(\cdot)$ is unchanged, whereas long-run means simplify, because by (42) we have $\omega_{23} = 0$, i.e.,

$$\theta = \begin{pmatrix} \frac{1}{b^2}\psi_{11}^2 + \frac{1}{b}\lambda_1\psi_{11} \\ \frac{1}{4a^2}(4\omega_{22} + 7\omega_{33}) + \frac{1}{a}\left(\lambda_2\psi_{22} + \lambda_3\psi_{33}\right) \\ \frac{1}{4a^2}\omega_{33} + \frac{1}{a}\lambda_3\psi_{33} \\ -\frac{1}{4a^2}(2\omega_{22} + 5\omega_{33}) \\ \frac{1}{2a}\omega_{33} \\ -\frac{3}{4a}\omega_{33} \\ -\frac{1}{2b^2}\psi_{11}^2 \end{pmatrix} \tag{F.2.3}$$

replaces (45) in the uncorrelated state variable case.

The exact discrete-time state process for the ANS-extended Vasicek model, which we use in place of the Euler discretization (54) in our empirical work, is obtained by writing $\psi_{22} = \psi_2$, $\lambda_2 = \lambda$, and setting all of x_1, $x_{5:7}$, λ_1, λ_3, b, and the remaining coefficients in the ψ-matrix equal to zero. This leaves a transition equation of the same form as (F.2.2), but

with

$$\theta = \begin{pmatrix} 0 & \frac{\psi_2^2}{a^2} + \frac{\lambda\psi_2}{a} & 0 & -\frac{\psi_2^2}{2a^2} \end{pmatrix}',$$

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

and the only non-zero element of the innovation variance matrix $\Omega(\Delta)$ that corresponding to slope, $\Omega_{22}(\Delta) = \psi_2^2(1-\mathrm{e}^{-2a\Delta})/(2a)$.

F.3. A low-storage square-root filter

For the third-stage approach from Section 4.2, we base the Kalman filter recursions on the exact discrete-time transition equation (derived in Appendix F.2), using the Koopman, Shephard, and Doornik (1999) low storage algorithm, with the updating step inserted in the prediction step to save on calculations, and modified to the square-root case. The modified recursions generate a sequence of yield vector innovations or prediction errors $\zeta_t = y_t - \mathbb{E}(y_t|Y_{t-1})$, with $Y_{t-1} = (y_1, \ldots, y_{t-1})$, and associated prediction error variances $\Gamma_t = \mathrm{var}(\zeta_t|Y_{t-1})$. The parameters are estimated by maximizing the log-likelihood based on ζ_t i.i.d. $N(0, \Gamma_t)$. The third-stage specifications can be written in the state space form

$$\underset{m\times 1}{y_t} = \underset{m\times 1}{c} + \underset{m\times k}{B}\,\underset{k\times 1}{x_t} + \underset{m\times 1}{\varepsilon_t}, \qquad \varepsilon_t \sim N(0, \Psi),$$

$$x_t = \underset{k\times 1}{\Phi_0} + \underset{k\times k}{\Phi_1} x_{t-1} + \underset{k\times 1}{v_t}, \qquad v_t \sim N(0, \Omega).$$

For example, corresponding to (F.2.2), we have $\Phi_1 = H(1)$, $\Phi_0 = (I - H(1))\theta$, and $\Omega = \Omega(1)$. The optimal portfolio from Theorem 1 depends on B and Ψ. The observed yield data are $(y_1, \ldots, y_T)$, and we write $Y_t = (y_1, \ldots, y_t)$ for observations up to time t. Denote the filtered state at t by $\mu_{t|t} = \mathbb{E}(x_t|Y_t)$, and the one step ahead prediction by $\mu_{t+1|t} = \mathbb{E}(x_{t+1}|Y_t)$. The associated conditional variance-covariance matrices are $\Sigma_{t|t} = \mathrm{var}(x_t|Y_t)$ and $\Sigma_{t+1|t} = \mathrm{var}(x_{t+1}|Y_t)$. We start with an initial condition for the first factor vector given by $x_{1|0} \sim N(\mu_{1|0}, \Sigma_{1|0})$, where $\mu_{1|0} = \bar{x}$ and $\Sigma_{1|0}$ solves $\Sigma_{1|0} = \Phi_1\Sigma_{1|0}\Phi_1' + \Omega$. The innovation

in the observation y_t is the prediction error $\zeta_t = y_t - \left(c + B\mu_{t|t-1}\right)$, with variance-covariance matrix $\Gamma_t = B\Sigma_{t|t-1}B' + \Psi$. The Kalman filter prediction step is

$$\mu_{t+1|t} = \Phi_0 + \Phi_1\mu_{t|t},$$
$$\Sigma_{t+1|t} = \Phi_1\Sigma_{t|t}\Phi_1' + \Omega,$$

and the update step is

$$\mu_{t|t} = \mu_{t|t-1} + \Sigma_{t|t-1}B'\Gamma_t^{-1}\zeta_t,$$
$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}B'\Gamma_t^{-1}B\Sigma_{t|t-1}.$$

The low storage filter is implemented by substituting the update step in the prediction step, and so iterates only on $\mu_{t+1|t}$ and $\Sigma_{t+1|t}$ in

$$\mu_{t+1|t} = \Phi_0 + \Phi_1\mu_{t|t-1} + K_t\zeta_t, \tag{F.3.1}$$
$$\Sigma_{t+1|t} = \Phi_1\Sigma_{t|t-1}\Phi_1 + \Omega - K_t\Gamma_tK_t', \tag{F.3.2}$$

using the Kalman gain $K_t = \Phi_1\Sigma_{t|t-1}B'\Gamma_t^{-1}$. The contribution to log-likelihood from each new observation is

$$\log p\left(y_{t_t}|Y_{t-1}\right) = -\frac{m}{2}\log(2\pi) - \frac{1}{2}\log|\Gamma_t| - \frac{1}{2}\zeta_t'\Gamma_t^{-1}\zeta_t,$$

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

$$\log L = \sum_{t=1}^{T}\log p\left(y_{t_t}|Y_{t-1}\right) = -\frac{mT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log|\Gamma_t| + \zeta_t'\Gamma_t^{-1}\zeta_t\right). \tag{F.3.3}$$

This is constructed recursively, with only $\mu_{t|t-1}$ and $\Sigma_{t|t-1}$ stored from the most recent period, calculating ζ_t, Γ_t, and K_t from these and the new observation y_t, and then $\mu_{t+1|t}$ and $\Sigma_{t+1|t}$ by (F.3.1) and (F.3.2). By minimizing storage need and circumventing the update step, the algorithm speeds up the filter, which must be run many times in the iterative maximization of (F.3.3) over parameters.

In the iterations towards the maximum of (F.3.3), the matrix $\Sigma_{t+1|t}$ may fail to be positive semi-definite. We solve this problem using a square-root filter, i.e., running the low-storage filter for $S_{t+1|t}$ satisfying $\Sigma_{t+1|t} = S_{t+1|t}S_{t+1|t}'$ instead, as done by Carraro (1988) for the original Kalman filter. To this end, we rewrite the prediction step (F.3.2) in

terms of $S_{t+1|t}$. First, write

$$\begin{aligned}
\Sigma_{t+1|t} &= \Phi_1 \Sigma_{t|t-1} \Phi_1 + \Omega - K_t B \Sigma_{t|t-1} \Phi_1' \\
&= (\Phi_1 - K_t B) \Sigma_{t|t-1} \Phi_1 + \Omega \\
&= (\Phi_1 - K_t B) \Sigma_{t|t-1} (\Phi_1 - K_t B)' + (\Phi_1 - K_t B) \Sigma_{t|t-1} B' K_t' + \Omega \\
&= (\Phi_1 - K_t B) \Sigma_{t|t-1} (\Phi_1 - K_t B)' + K_t \Gamma_t^{-1} K_t' - K_t (\Gamma_t - \Psi) K_t' + \Omega \\
&= (\Phi_1 - K_t B) \Sigma_{t|t-1} (\Phi_1 - K_t B)' + K_t \Psi K_t' + \Omega .
\end{aligned}$$

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

$$\Sigma_{t+1|t} = \left[(\Phi_1 - K_t B) S_{t|t-1}, K_t N, M \right] \begin{bmatrix} S_{t|t-1}' (\Phi_1 - K_t B)' \\ N K_t' \\ M \end{bmatrix} \equiv \widetilde{S}_{t+1|t} \widetilde{S}_{t+1|t}' ,$$

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

$$\widetilde{S}_{t+1|t}' = Q_{t+1|t} R_{t+1|t} ,$$

so that

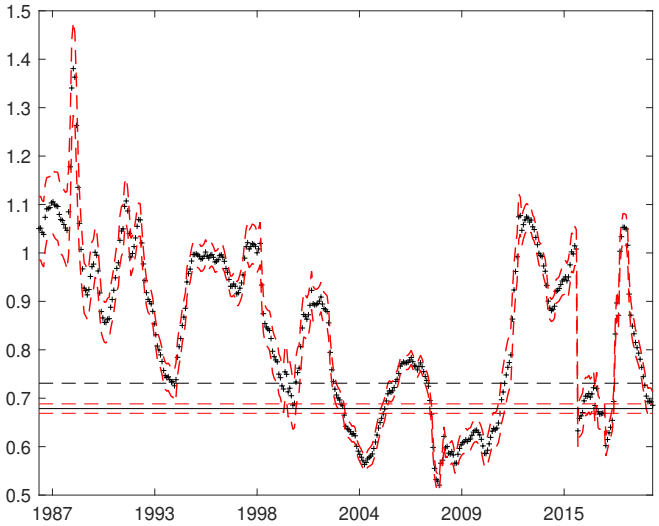
$$\Sigma_{t+1|t} = \widetilde{S}_{t+1|t} \widetilde{S}_{t+1|t}' = R_{t+1|t}' Q_{t+1|t} Q_{t+1|t} R_{t+1|t} = R_{t+1|t}' R_{t+1|t} .$$

Therefore, set $S_{t+1|t} = R_{t+1|t}'$, which is a lower triangular square matrix. Instead of $\Sigma_{t+1|t}$, the filter uses $S_{t+1|t}$, and the resulting $\Sigma_{t+1|t}$ is positive semi-definite by construction.

G. Additional Empirical Results

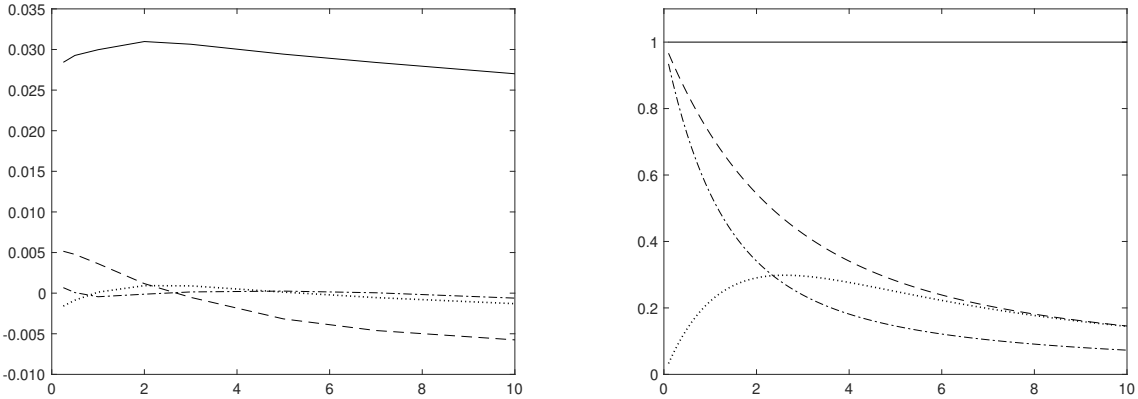
This appendix provides some additional empirical results supplementing those in Section 6.

Figure G.1: Time series evolution of estimated a in Nelson-Siegel



This figure shows the rolling four-year NS estimates of a, with 95% confidence bands in red. The solid horizontal line indicates the full period NS estimate, with 95% confidence band in red, and the dashed black line the Diebold, Ji, and Li (2006) value.

Figure G.2: Loading functions in four-factor models



The left exhibit shows the loadings $B_j(\tau)$, $j = 1,2,3,4$, in the unrestricted four factor model, as functions of maturity, τ. The right exhibit shows the ANS loading functions $\tilde{B}_j(\tau)$, $j = 1,2,3,4$, from (32).

Table G.1: Loadings in unrestricted single-factor model for $\tilde{y}$

This table shows the loadings $B(\tau)$ (in percent) as function of maturity τ in the unrestricted single-factor model for slope-adjusted yield changes $\tilde{y}$.

τ	3 mns.	6 mns.	12 mns.	2 yrs.	3 yrs.	5 yrs.	7 yrs.	10 yrs.
$B(\tau)$	0.062	0.081	0.097	0.124	0.136	0.144	0.141	0.132