ESSAYS ON DYNAMIC TERM STRUCTURE MODELS

By Jorge W. Hansen

A PhD thesis submitted to
School of Business and Social Sciences, Aarhus University,
in partial fulfilment of the requirements of the requirements of the PhD degree in
Economics and Business Economics

March 2020
This dissertation is the result of my studies from March 2016 to February 2020 at the Department of Economics and Business Economics at Aarhus University. First and foremost, I would like to express my sincere gratitude to Bent Jesper Christensen, who gave me the opportunity to accomplish this work by supervising my thesis and for allowing me to grow as a research scientist. Second, I would like to thank my co-supervisor Martin M. Andreasen, whose door was always open. I am also truly grateful to Aarhus University and the Center for Research in Econometric Analysis of Time Series (CREATES) for providing an excellent research environment and financial support for courses and conferences. I am especially grateful to Solveig Nygaard Sørensen and Pernille Vorsø Jachobsen for their invaluable assistance with administrative matters, their patience, and tireless effort to answer all my question and concerns.

During my studies, I had the pleasure of visiting the Department of Economics, University of Pennsylvania, USA; I would like to thank Francis X. Diebold for hosting me, for the intense discussions about climate econometrics, and for inviting me to present at the Econometrics seminar series. A special thank you goes to Priyanka who made my stay at Penn so enjoyable. I would also like to express my gratitude to Danmark-Amerika Fondet, Augustinus Fonden, Christian og Ottillia Brorsons Rejselegat, and Knud Højgaards Fond for providing financial support for my stay abroad.

I consider myself very lucky for having as wonderful a friend at work as Daniel. Having you around made my PhD feel less like driving behind the safety car, and more like a successful undercut. I promise you that our bi-annual Formula 1 trips will continue perpetually. I also want to thank Benjamin, Johan, and Simon with whom I shared office, concerns and laughs. A big thank also goes to my fellow PhD students and junior faculty for years of intense work, interesting conversations, and fun. Most responsible characters to thank in this context are Christian, Christoffer, Jonas, Niels, and Thomas.

I would not have been able to present this dissertation without the great proofreading support of Dennis. Your amusing comments always made me smile! Special thanks go to Bahne, Finn, Flemming, Jebbe, Malte, and Marvin. And to Rune for showing me
your finest $\theta$. Having such wonderful friends is of immense value. I sincerely hope, we have a long way to go together!

Big thanks go to my parents Astrid and Wolfgang for listening to, understanding, and supporting me at all times, and to my brothers Janes and Morten for unforgettable shared experiences. Lastly, a deep heartfelt thank you is deserved for Liv for her love and for her patient support during my graduate studies – I’m glad you’re there.

*Jorge W. Hansen*

*Aarhus, March 2020*
The pre-defence meeting was held on April 17, 2020. I am grateful to the members of the assessment committee consisting of Michel van der Wel (Erasmus University Rotterdam), Lasse Bork (Aalborg University), and Bezirgen Veliyev (Aarhus University) for their careful reading of the dissertation and their many insightful comments and suggestions. Some of the suggestions have been incorporated into the present version of the dissertation while others remain for future research.

*Jorge W. Hansen*
*Aarhus, May 2020*
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary</strong></td>
<td>vii</td>
</tr>
<tr>
<td><strong>Danish summary</strong></td>
<td>xi</td>
</tr>
<tr>
<td>1 Immunization with term structure dynamics</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Brief technical preview of the analysis</td>
<td>5</td>
</tr>
<tr>
<td>1.3 The basic immunization framework</td>
<td>6</td>
</tr>
<tr>
<td>1.4 Dynamically consistent arbitrage-free hedging</td>
<td>17</td>
</tr>
<tr>
<td>1.5 Conclusion</td>
<td>38</td>
</tr>
<tr>
<td>1.6 References</td>
<td>41</td>
</tr>
<tr>
<td>A.1 Appendix</td>
<td>45</td>
</tr>
<tr>
<td>2 Can linear-rational term structure models capture conditional volatility in the Treasury yield market?</td>
<td>79</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>80</td>
</tr>
<tr>
<td>2.2 The linear-rational square-root model</td>
<td>84</td>
</tr>
<tr>
<td>2.3 Estimation methodology</td>
<td>87</td>
</tr>
<tr>
<td>2.4 Empirical analysis</td>
<td>89</td>
</tr>
<tr>
<td>2.5 Predictive performance</td>
<td>100</td>
</tr>
<tr>
<td>2.6 Extensions</td>
<td>102</td>
</tr>
<tr>
<td>2.7 Robustness</td>
<td>105</td>
</tr>
<tr>
<td>2.8 Concluding remarks</td>
<td>106</td>
</tr>
<tr>
<td>2.9 References</td>
<td>107</td>
</tr>
<tr>
<td>A.1 Appendix</td>
<td>111</td>
</tr>
<tr>
<td>3 The impact of curve-fitting procedure on estimation and testing of term structure models</td>
<td>131</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>132</td>
</tr>
<tr>
<td>3.2 Modeling framework</td>
<td>135</td>
</tr>
<tr>
<td>3.3 Estimation of DTSMs</td>
<td>141</td>
</tr>
<tr>
<td>3.4 Test for no-arbitrage in DTSMs</td>
<td>145</td>
</tr>
<tr>
<td>3.5 Empirical application of DTSMs</td>
<td>147</td>
</tr>
<tr>
<td>3.6 Dynamically consistent term structure models</td>
<td>155</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>3.7</td>
<td>Test for no-arbitrage in DCTSMs</td>
</tr>
<tr>
<td>3.8</td>
<td>Empirical application of DCTSMs</td>
</tr>
<tr>
<td>3.9</td>
<td>Conclusion</td>
</tr>
<tr>
<td>3.10</td>
<td>References</td>
</tr>
<tr>
<td>A.1</td>
<td>Appendix</td>
</tr>
</tbody>
</table>
This thesis consists of three self-contained chapters on the term structure of interest rates. In particular, empirical properties of dynamic term structure models (DTSMs) and the impact of data preprocessing procedures on conclusions derived from using these models. The first chapter considers strategies to protect the value of a bond portfolio against interest rate changes. The second chapter shows how introducing factors in DTSMs that are not spanned by the yield curve helps to capture conditional volatility in the bond market. Lastly, the third chapter explores the impact of the choice of a curve fitting procedure in the bond data preprocessing step on the inference drawn from the received yield curve data.

Chapter 1: Immunization with term structure dynamics (joint with Daniel Borup, Bent Jesper Christensen, and Torben B. Rasmussen) shows that improved immunization performance of bond portfolios is obtained by exploiting dynamic term structure theory. Prior attempts to improve on the basic duration matching approach involve generalized duration matching based either on a factor analysis or a parametrized yield curve shape. The chapter shows how to improve upon this by exploiting interest rate dynamics rather than either of these purely cross-sectional approaches.

Asset pricing theory promises a further improvement because the absence of arbitrage opportunities imposes cross-restrictions on interest rate dynamics and the yield curve shape. This additional improvement in hedging performance is empirically realized once a new term structure model involving three stochastically varying factors corresponding to level, slope, and curvature is introduced. Furthermore, it is shown that a particular restriction on our general SLSC model reproduces the independent affine term structure model (ATSM) considered by Christensen, Diebold and Rudebusch (2011)). This model does, however, not improve very much on empirical hedging performance relative to the purely cross-sectional approaches. Finally, in a further hedging exercise, the results show the empirical evidence favors the generalized duration approach in that it is more important to remove the factor contribution to hedging error variance than to balance factor versus idiosyncratic contributions.

Chapter 2: Can linear-rational term structure models capture conditional volatility in the Treasury yield market? studies the question of whether the linear-rational square
The root (LRSQ) model is able to break the tension documented for the popular ATSMs from matching the conditional first and second moment of yields simultaneously. This failure of the ATSMs is commonly attributed to their fundamental property, according to which all yield volatility risk is spanned by the term structure factors. However, numerous papers in the literature find evidence of so-called unspanned stochastic volatility (USV) in the fixed-income market in the sense that sources of risk that determine prices of fixed-income derivatives cannot be hedged using bonds only, rendering the fixed-income market incomplete. In this chapter, a new class of models given by the LRSQ model developed by Filipović, Larsson and Trolle (2017) is considered, which among other appealing properties, is designed to incorporate USV. The LRSQ model is estimated based on a state space system in which the measurement equation is extended to include conditional yield volatility as extra measurement. Moreover, to account for the fact that volatility in the LRSQ model is nonlinear in the factors, the approach is based on the Unscented Kalman Filter (UKF) (developed by Julier and Uhlmann (1997)), which is designed to accommodate nonlinearities.

An empirical application to US Treasury bond yields shows that the model is able to break the tension noted for ATSMs. On a general level, the analysis reveals the importance of complementing cross-sectional yield data with informative second-moment data and highlights the importance of inducing USV in order to estimate conditional volatility of the yield curve adequately. The performance of various LRSQ model specifications is evaluated based on in-sample and out-of-sample exercises and the results show that the preferred model specification relies on three USV factors, which correlate strongly with the level and slope factor of conditional yield volatility.

Chapter 3: The impact of curve-fitting procedure on estimation and testing of term structure models (joint with Bent Jesper Christensen) explores the impact of the choice of a curve fitting procedure to a panel data set of coupon bond prices on the inferences on model structure and absence of arbitrage based on the received yield curve data. Empirical analysis of interest rates using panel data with both calendar time and term to maturity dimensions generally proceeds in two steps: First, a yield curve is fitted to prevailing market prices on the current calendar date using a flexible parameterized functional form. Secondly, the resulting time series of yield curves is used to address a substantive issue of interest, such as the number of factors driving the term structure through time, testing arbitrage restrictions, or distinguishing between parametric DTSMs. However, some of the typical questions examined in the second stage are raising suspicion that the choice of curve-fitting procedure in the first step may actually increase the risk of improper inference in the second step. An important tool in this context is the concept of consistency, i.e., the consistency between the functional form of the yield curve and the DTSM, as introduced by Björk and Christensen (1999).
Thus, this paper addresses the question of whether the use of a curve fitting procedure, which is inconsistent with the chosen parameterized DTSM leads to misspecification of model parameters, incorrect hypothesis testing, and the similar. The analysis is casted in the framework of Heath, Jarrow and Morton (1992) (henceforth HJM), and applies different curve fitting procedures to empirical data on prices of US Treasury bonds. It shows that the evidence favors at least two factors to exclude arbitrage-opportunities and three factors to describe the volatility structure but the exact number of factors needed as well as the shape of the volatility structure varies between curve fitting procedures. Moreover, it shows that the concept of consistency plays an important role in the conclusion drawn from the analysis.

References


Danish summary


Kapitel 1: Immunization with term structure dynamics (fælles med Daniel Borup, Bent Jesper Christensen og Torben B. Rasmussen) omhandler risikoafdækning af obligationsporteføljer ved at anvende dynamisk rentekurve teori. Tidligere forsøg på at forbedre den grundlæggende varighedsstrategi baserer på at matche det generaliserede varigheds mål baseret på enten faktoranalyse eller en parametreringsmethode. Kapitlet viser empirisk, at risikoafdækning forebygdes ved at anvende rentefitmetoden frem for en hvilken som helst tværsnitsmetode.

Risikoafdækningen forbedres yderligere ved at udnytte tværsnitsrestrktionerne mellem rentefitmetoden og rentekurven pålagt af fravær af arbitrage. Denne forbedring realiseres empirisk ved at introducere en ny dynamisk rentestrukturmodell (SLSC-modellen), der rummer tre stokastiske faktorer svarende til niveauet, hældningen og krumningen af rentekurven. Der vises, at SLSC-modellen under specielle restriktioner svarer til den affine rentestrukturmodell (ARSM) af Christensen et al. (2011). Igennem en yderligere afdækningsøvelse vises der empirisk, at det er vigtigere at fjerne faktorerens bidrag til afdækningsfejlers varians end at forsøge at afbalancere bidragene fra faktorerne med de idiosynkratiske bidrag.

Kapitel 2: Can linear-rational term structure models capture conditional volatility in the Treasury yield market? undersøger om linear-rational square root (LRSQ) modellerne er tilstrækkelig fleksibel til at matche nulkuponrentens betingede første og andet moment simultant. På trods af ARSMernes popularitet, er disse nemlig ikke i stand til at matche begge momenter simultant. I ARSMerne bliver denne begræns-


Litteratur


IMMUNIZATION WITH TERM STRUCTURE DYNAMICS

Daniel Borup
Aarhus University, CREATEs, and the Danish Finance Institute

Bent Jesper Christensen
Aarhus University, CREATEs, Dale T. Mortensen Centre, and the Danish Finance Institute

Jorge W. Hansen
Aarhus University and CREATEs

Torben B. Rasmussen
Aarhus University and CREATEs

Abstract

We show that improved hedging performance of bond portfolios is obtained by exploiting no-arbitrage and consistency restrictions on the interest rate dynamics rather than using purely cross-sectional approaches based on factor analysis or a parametrized yield curve shape. We also introduce a new term structure model involving three stochastically varying factors corresponding to level, slope, and curvature, which leads to a further improvement in hedging performance. A restricted version of this model belongs to the three-factor affine class. The evidence shows the importance of immunizing the factor contribution to hedging error variance contrary to balancing factor versus idiosyncratic contributions.
1.1 Introduction

We show that improved immunization performance of bond portfolios is obtained by exploiting dynamic term structure theory. Based on a classical statistical factor model for yields, we address that previous attempts in the literature to protect against interest rate risk are largely cross-sectional in nature and void of financial theory. In particular, we reduce model complexity by imposing restrictions on loadings and factors using dynamic term structure theory and document the resulting improvement in performance. The key to this result is that estimating a large number of parameters (as in an unrestricted factor model) may induce noise in the estimation and hence in the hedge portfolios. By means of this, we explore the potential efficiency gains by imposing parameter restrictions. In particular, the implications of consistent arbitrage-free interest rate dynamics on the generalized durations used in forming the hedge portfolios.\(^1\) By consistency we refer to the consistency between the functional form of the yield curve and a dynamic term structure model, as introduced by Björk and Christensen (1999). They give necessary and sufficient conditions for whether a dynamic term structure model, if started at an initial yield curve of a given functional form will continue to generate future term structures of the same general form.

Our approach is developed in steps of increasing complexity. First, we consider immunization based on a classical statistical factor model for yields. We obtain immunization formulas by estimating the factor model and minimize immunization error variance subject to generalized duration matching and value matching. Following this, we exploit the level, slope, and curvature (LSC) structure of the term structure. That is, in an attempt to reduce the model degrees of freedom, we use that the parsimoniously parametrized Nelson and Siegel (1987) (NS) yield curve shape with monotonic and hump components correspond quite precisely to the LSC structure. Thus, imposing this type of functional form restrictions directly on the loadings estimated in the factor analysis allows saving on parameters. Second, we turn our attention to dynamically consistent arbitrage-free immunization and note that not all such restrictions (as the NS restriction) on the loadings are consistent with a dynamic term structure model. By means of Björk and Christensen (1999) we modify the parametrization of the restricted loadings to ensure the existence of a dynamic term structure model consistent with the chosen shape and reassess immunization performance. In order to obtain an arbitrage-free yield curve model, we require, in

\(^1\)Joslin, Singleton and Zhu (2011) conclude that in Gaussian dynamic term structure models, no-arbitrage restrictions are not useful, per se, in forecasting the yield curve and attribute any empirical forecasting gain in these models to the combination structure imposed of both no-arbitrage and their restrictions on the distribution of yields under the objective probability measure. Our focus differs from theirs, since we evaluate the implication of no-arbitrage conditions and consistency on immunization performance of a bond portfolio using a wider set of models, not necessarily Gaussian dynamic term structure models.
1.1. **Introduction**

Particular, consistency with the dynamic arbitrage-free model of *Heath et al. (1992)* (HJM). Third, we note that even though the resulting restricted factor model has loadings consistent with the desired yield curve shape, this neither requires nor guarantees that the yield curve generated by the model actually takes on this shape. Yet, what it implies is that if the yield curve shape takes on this shape at any point in time, it will continue to do so thereafter. Furthermore, a state space model for yields (as opposed to returns) is obtained in this case, which may be estimated by the Kalman filter, and corresponds to the most parsimonious approach, imposing all implications of the financial theory. This set-up is ideally suited to cases where a particular optimal bond for hedging is illiquid. By exploiting a factor model for bond returns we can construct a hedge portfolio from a set of liquid bonds with the same factor sensitivity as the target portfolio.

These steps are implemented and tested empirically both in a case with a single stochastic factor, and in a more general situation involving three stochastic factors, corresponding to LSC. The single-factor analysis combines a modified version of the NS curve shape consistent with no-arbitrage and the dynamic model of HJM. The three-factor analysis is based on a new stochastic LSC (SLSC) model unique to the present paper, which in the special case where the SLSC model is restricted to the three-factor affine class in the sense of *Duffie and Kan (1996)* reduces to the independent factor model considered by *Christensen et al. (2011)*.

The empirical performance of our approaches is investigated using weekly yield data from the Federal Reserve over the period 1983 through 2014. The targets for assessing immunization performance are a five-year coupon bond and a more complex portfolio consisting of two-year, five-year, and ten-year coupon bonds with positive and negative weights. Portfolios of zero-coupon bonds are constructed on a monthly basis to match estimated generalized durations. The resulting ability to hedge one-month returns on the targets is recorded over the full sample period. The empirical results show that an improvement in immunization performance is achieved by replacing the NS curve shape by a modification that is consistent with arbitrage-free interest rate dynamics. We find further substantial improvements that are economically and statistically important by estimating parameters directly from the HJM dynamics consistent with the modified curve shape. However, performance decreases when combining the modified curve shape and consistent dynamics. This suggests that for immunization purposes, one-factor dynamics do not adequately capture term structure movements. We therefore introduce our new three-factor SLSC model, which leads to better performance than any of the previously considered approaches. These results thus indicate that important improvements in immunization strategies may be achieved by using a general model that is dynamically consistent and involves three stochastic factors, reflecting the time-varying LSC features of the bond market.
Our approach shows that combining the factor analysis and associated immunization technology of prior literature can be substantially improved upon. Traditional immunization strategies as introduced by Redington (1952) and later revived in applications by Fisher and Weil (1971) are entirely cross-sectional in nature and void of financial theory.\textsuperscript{2} Their strategies amount to matching the basic bond duration measure of Macaulay (1938) across assets and liabilities. This applies also to the subsequent generalized duration approach (e.g., Ingersoll (1983), Nelson and Schaefer (1983), and Diebold, Ji and Li (2006)), which rely implicitly or explicitly on a multivariate factor structure underlying market yields, matching each of a number of factor loadings or sensitivities across the immunization target and a immunzation portfolio.\textsuperscript{3} This approach still enjoys frequent use (see e.g. Quaedvlieg and Schotman (2019) for an application to long-term liabilities) and is largely cross-sectional in nature, much as in the case of stock portfolio management. However, when explicitly considering movements in the entire yield curve across calendar time, consistency with dynamic arbitrage-free term structure models places rather tight cross-restrictions on the admissible cross-sectional loading functions used in the construction of hedge portfolios and the dynamic behavior of market interest rates. Although ignoring these restrictions potentially leads to costly lack of precision in immunization strategies, they have not been explored in the literature.

The remainder of the paper is laid out as follows. In Section 1.2, we provide a brief technical preview of the paper. Next in Section 1.3, we present the basic immunization framework, starting with the underlying yield factor model. In Section 1.4, we turn our attention to dynamically consistent arbitrage-free immunization. Lastly, in Section 1.4.6, we consider an alternative approach, relaxing the generalized duration matching constraint and instead using the dynamic model to trade off violations of the constraint against remaining immunization error variance. Section 1.5 concludes. All proofs are contained in the Appendix.

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

\textsuperscript{3}Other approaches to bond portfolio risk management have been proposed in the literature. Some notable contributions are the key-rate duration Ho (1992), contingent immunization Leibowitz and Weinberger (1981, 1982, 1983); Díaz, González, Navarro and Skinner (2009), M-square and M-vector models Chambers, Carleton and McEnally (1988); Nawalkha and Chambers (1997); Nawalkha, Soto and Zhang (2003), infinite-dimensional immunization Bueno-Guerrero, Moreno and Navas (2015), and cross-shape risk Galluccio and Roncoroni (2006). For additional applications of some of above methods, see e.g. Carcano and Dall’O (2011) who extend models based on principal components, duration vectors and key rate duration to exploit model estimation error, Soto (2004) and Bravo and Silva (2006) who investigate the question of dimension of the models underlying a given risk management technique, and Agca (2005) who provides simulation-based evidence on the immunization performance of Heath-Jarrow-Morton model-based risk measures.
1.2 Brief technical preview of the analysis

To keep the overall strategy and plan for the analysis in mind, this section offers a brief technical preview of the paper. Starting from a classical statistical factor model for yields \( y \) given by (1.1), we obtain immunization formulas (from deriving the associated factor model for returns) as presented in the main theorem, i.e., Theorem 1 in Section 1.3.1.3, that minimizes immunization error variance. The theorem establishes a link between between the optimal portfolios and the parameters of the factor model. This corresponds to the case where the available instruments are the zero-coupon bonds given by the yields in the factor model. Similar formulas for the case where the instruments are coupon-bearing bonds are easily obtained as shown in the Appendix.

Too many parameters may induce noise in the estimation and hence in the immunization portfolios. Under correct specification it may therefore improve immunization to save on parameters by imposing restrictions on (1.1). Since the LSC structure is expected to hold and the NS factor model has a LSC structure, but depends only on a single parameter, it seems natural to impose this type of functional form restrictions directly on the loadings in (1.1). Doing this, equivalently to the unrestricted case, we can use the estimated restricted factor loadings as input into the immunization portfolios. We test the NS restriction on loadings by assessing their value in immunization.

Following this, we use dynamic term structure theory to improve the restrictions. By means of Björk and Christensen (1999) we compute restrictions on the loadings that make the yield factor model (1.1) consistent with a dynamic term structure model. Moreover, in order to achieve consistency with an arbitrage-free dynamic term structure model, we cast the framework into the HJM model (1.17). Since the HJM model addresses the stochastic differential \( dy \), we derive the dynamic version of (1.1) given by (1.16). For the HJM model and (1.16) to be mutually consistent, it requires cross restrictions on the loadings in (1.16). We find, in fact, that this implies restrictions on the factors, too. The general advantage of this approach is that no-arbitrage is directly imposed on the factor model, which is now for returns, not yields, while simultaneously imposing parsimony motivated by insights from the yield curve shape on the factor model. To ensure an arbitrage-free model, the model should be tested for arbitrage, before the model is used for immunization. To test for arbitrage, we transform to \( \tilde{y} \) from (1.23). The \( \tilde{y} \) are proportional to returns and follow a factor model similar to that for \( y \). This is convenient, since we need returns and thus \( \tilde{y} \), anyway for immunization.

While the factor model for returns has loadings consistent with the desired yield curve shape, this neither requires nor guarantees that the yield curve generated by the model actually takes on this shape. But if it does so at any point in time, it will
continue to do so thereafter (in this case, the slope-adjustment in $\tilde{y}$ is restricted by the given yield curve shape). If this final condition is imposed, along with the previous, then a state space model for yields (as opposed to returns) is obtained. This can be estimated by the Kalman filter. We compare the immunization performance resulting from calculating portfolios from this smallest (most restricted) set of model parameters to those relying on fewer model restrictions.

In sum, we consider immunization based on a factor model for yields. We then explore the possible gains from parsimony by imposing parameter restrictions. We do so in three steps. First, motivated by the observed LSC pattern in yields, we restrict loadings, and assess the potential benefits in immunization. Second, we note that not all such restrictions on loadings would be consistent with a dynamic term structure model, and, when imposing such consistency, additional restrictions on factors are implied. This yields a restricted factor model for returns, calculated by applying a slope-adjustment to yield changes. Third, if the LSC or other desired yield curve structure is imposed on this slope-adjustment, too, a state space model for yields (as opposed to returns) results. We compare immunization performance across the increasingly restricted (parsimonious) models.

1.3 The basic immunization framework

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

$$y_t = \mu + B f_t + \varepsilon_t, \quad (1.1)$$

where $\mu$ is an $m$-vector of mean yields, $f_t$ is a $k$-vector of common, unobserved covariance-generating factors, $k < m$, $B$ is an $m \times k$ matrix of sensitivities or factor loadings, and $\varepsilon_t$ is an $m$-vector of idiosyncratic error terms independent of $f_t$. Thus, the individual yield is

$$y_{t,\tau_i} = \mu_i + b_{i}' f_t + \varepsilon_{t,i}, \quad (1.2)$$

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

Given the objective of obtaining linear immunization rules, we suppose that the target claim to be hedged is a future payment $\tau_*$ periods hence, and no zero-coupon
bond with term to maturity \( \tau_* \) (the ideal hedging instrument) is available. The factor loadings \( b_* \) (a \( k \times 1 \) vector) of the target claim (precisely, of the yield to the missing ideal hedge) may be obtained by interpolation between the available maturities. With the yield data ordered in such a way that \( \tau_1 < \tau_2 < \ldots < \tau_m \), simple linear interpolation would set

\[
 b_* = ((\tau_{i+1} - \tau_*) b_i + (\tau_* - \tau_i) b_{i+1}) / (\tau_{i+1} - \tau_i),
\]

a \( k \)-vector constructed as a suitable convex combination of the adjacent loading vectors \( b_i \) and \( b_{i+1} \), where \( \tau_i < \tau_* < \tau_{i+1} \). If the future payment occurs before the shortest maturity \( \tau_* < \tau_1 \), we set \( b_* = b_1 \) equal to the first factor loading vector. Similarly, if the futures payment occurs after the longest maturity \( \tau_m < \tau_* \), we set \( b_* = b_m \). The required (fitted) yield on the hedge portfolio is then \( b_*' \hat{f}_t \). This is a linear function of the available yields, too, namely, \( b_*' \hat{F} y_t \). Nonetheless, this does not constitute a portfolio rule in itself, as the traded instruments are bonds, not yields. Litterman and Scheinkman (1991) considered excess returns rather than yields, thus obtaining linear hedging rules. We present a related approach, first fitting parameters in a yield factor model, then using the estimated parameters to form policies (portfolio weights) linear in returns. Thus, we turn to implications of the yield factor model for returns next.

### 1.3.1 Hedge portfolios

In order to acquire linear portfolio rules, we go via the relation between zero-coupon bond prices and yields, \( p_{t, \tau_i} = \exp(-\tau_i y_{t, \tau_i}) \), and define returns

\[
 r_{t+1, i} = (p_{t+1, \tau_i} - p_{t, \tau_i+1}) / p_{t, \tau_i+1}.
\]

Over short horizons we can approximate this by the return on a constant maturity zero-coupon bond and use log-linearization,

\[
 r_{t+1, i} \approx \log \frac{p_{t+1, \tau_i}}{p_{t, \tau_i}} = -\tau_i \Delta y_{t+1, \tau_i},
\]

where \( \Delta \) is the first difference operator, i.e., \( \Delta y_{t+1, \tau_i} = y_{t+1, \tau_i} - y_{t, \tau_i} \) is the change in the \( \tau_i \) maturity zero-coupon bond yield.\(^4\)

Applying the factor model for yields (1.2) in (1.3) we can express the return by

\[
 r_{t+1, i} = -\tau_i (b_i' \Delta f_{t+1} + \Delta \varepsilon_{t+1, i}).
\]

This is recognized in the simple case \( k = 1 \) with \( f_t \) a factor of unit loading \( b_1 = 1 \) across all maturities \( \tau_i \) as the well-known result that a one percentage point parallel shift up in the term structure is associated with a drop in bond prices by a percentage amount equal to duration, which is \( \tau_i \) for the \( i \)’th zero-coupon bond. If the yields have different factor sensitivities, i.e., \( b_i \) is not constant across \( i \), then (1.4) introduces

\(^4\)For non-short horizons, a slope-adjustment (see e.g. Litterman and Scheinkman (1991)) may be added to improve upon the approximation. This adjustment will, however, be known at time \( t \), which implies that the conditional variance of the hedge error in Theorem 1 is unaltered, and the optimal weights remain the same.
the corresponding generalized duration measure $\tau_i b_i$, giving the percentage drop in the price of the $i$'th zero-coupon bond associated with a unit factor change. In case of multiple factors, $k > 1$, each entry in $\tau_i b_i$ similarly defines a generalized duration measure with respect to the associated factor, and the multivariate hedge is an immunization strategy with respect to each factor.

### 1.3.1.1 Hedging a simple claim

To immunize the future payment $\tau_*$ periods hence we consider investing in a portfolio comprised of the $m$ zero-coupon bonds with times to maturity $(\tau_1, \ldots, \tau_m)$ for which the yield factor model (1.1) is estimated. It is useful to write returns of the instruments as the vectorized version of the return model (1.4),

$$r_{t+1} = -\mathcal{J}(B \Delta f_{t+1} + \Delta \varepsilon_{t+1}), \quad (1.5)$$

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

$$r^*_w = w' r_{t+1} = -w' \mathcal{J}(B \Delta f_{t+1} + \Delta \varepsilon_{t+1}), \quad (1.6)$$

where $B' \mathcal{J} w$ is the $k \times 1$ vector of generalized durations, controlled by the portfolio manager through the choice of $w$.

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

$$\tilde{w} = \tau_* \mathcal{J}^{-1} \hat{F}' b_* \quad (1.7)$$

matches the generalized durations $\tau_* b_*$ of the target. Among all portfolios doing so, $\tilde{w}$ in (1.7) minimizes hedging error variance, as shown in Theorem 1 below. Note that estimated loadings $\hat{B}$ and error variances $\hat{\Psi}$ are those from the classical factor analysis applied to yields, based on (1.1), but that the resulting portfolio weights $\tilde{w}$ are applied to returns, not yields. In (1.7), each row of $\hat{F} \mathcal{J}^{-1}$ shows how to construct the factor-mimicking portfolio for the corresponding factor, and the generalized durations $\tau_* b_*$ of the target are used to combine the $k$ factor-mimicking portfolios into the overall hedge portfolio.

### 1.3.1.2 Hedging of general payment streams

Suppose the target claim to be hedged at time $t$ is a payment stream, say, a coupon bond, or an annuity, that promises payments $c_h$ at future dates $\tau_h$ periods hence,
1.3. THE BASIC IMMUNIZATION FRAMEWORK

$h = 1, \ldots, H$. Then the value of the claim is $v_* = \sum_{h=1}^{H} p_{t, \tau h} c_h$, with $p_{t, \tau} = \exp(-\tau y_{t, \tau})$ the discount function, if necessary obtained by interpolation between observed yields. To hedge the payment stream, the idea is to hedge each payment according to the portfolio rule (1.7). The hedge portfolio for the target stream is then the combination of all these portfolios. Specifically, payment $c_h$ is hedged by allocating the amount $p_{t, \tau h} c_h$ across the $m$ hedging instruments in the proportions indicated by the portfolio rule $\tilde{w}_h = \tau_h^{T} \tilde{F}' b_h$, with the $k$-vector of coupon loadings $b_h$ obtained by interpolation as in the case of target loadings above. The overall strategy is to allocate the amount $v_*$ across the instruments according to $\tilde{w} = \sum_{h=1}^{H} p_{t, \tau h} c_h \tilde{w}_h / v_*$. This is equivalent to applying the rule (1.7) directly to the target payment stream, assessing its generalized duration vector as

$$\tau^* = \frac{\sum_{h=1}^{H} p_{t, \tau h} c_h \tau_h b_h}{v_*},$$

(1.8)

the weighted average of the individual payments’ generalized duration vectors $\tau_h b_h$, each of dimension $k$, the weights being those of the individual payments’ present values relative to the total claim. The result $(\tau^*)_*$ is simply inserted in (1.7) to achieve generalized immunization of the target,

$$\tilde{w} = \tilde{F}^{-1} (\tau^*).$$

(1.9)

The immunization strategy (1.9) combines the hedging instruments to equate each of the $k$ generalized durations of the portfolio to those of the target claim, $(\tau^*)_*$, and, in addition, to minimize hedging error variance (see Theorem 1 below). In particular, generalized durations are $B' T \tilde{w} = (\tau^*)_*$.

1.3.1.3 Minimizing hedging error variance

Value matching requires the value of the hedge portfolio to equal the value of the target, meaning that the portfolio should be fully invested in the sense of Nelson and Schaefer (1983). This is accomplished if the weights sum to one, $\tilde{w}' i = 1$, where $i = (1, \ldots, 1)'$ and requires additional scaling of the generalized duration matching portfolio (1.9). An obvious choice would be to use the scaling factor $(\tilde{w}' i)^{-1}$, but this would violate generalized duration matching. The following theorem shows how to adjust portfolio weights (1.9) to achieve value matching while maintaining generalized duration matching and minimizing hedging error variance.

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

$$\min_{w} \text{var}_t [r_{t+1}^* - w' r_{t+1}] \quad s.t. \quad B' T w = (\tau^*)_* \quad \text{and} \quad w' i = 1,$$

(1.10)
CHAPTER 1. IMMUNIZATION WITH TERM STRUCTURE DYNAMICS

is given by the instruments weights

\[ w_* = \tilde{w} + (1 - \tilde{w}' \iota) \frac{\Lambda \iota}{\iota' \Lambda \iota}, \]  

(1.11)

where \( \tilde{w} \) are the weights (1.9) that solve the equivalent problem without the value matching constraint \( w' \iota = 1 \), and where \( \Lambda = \mathcal{T}^{-1} \Psi^{-1} (\Psi - B (B' \Psi^{-1} B)^{-1} B') \Psi^{-1} \mathcal{T}^{-1} \).

In this paper subscript \( t \) indicates the conditional moments whenever applicable. The constraints in (1.10) are value matching, \( w' \iota = 1 \), and generalized duration matching, \( B' \mathcal{T} w = (\tau b)^*_r \). Together, these would constitute a sufficient criterion in a complete market with only factor risk, cf. Harrison and Kreps (1979). In practice, in the incomplete market case, a perfect hedge is infeasible, so we look for the portfolio which generates a return as close as possible to the target return in the sense that it minimizes the hedging error variance subject to the two constraints. This criterion differs from that introduced by Ingersoll (1983) and frequently used by academics as well as practitioners (see, e.g., Diebold et al. (2006) for an application), where instead the sum of squared weights \( w' w \) is minimized subject to the same two constraints. Ingersoll suggests a diversification argument behind which lies the assumption that the idiosyncratic errors in instrument returns are of the same magnitude. Given the yield factor model on which generalized durations are based we know that idiosyncratic return errors must have variance \( \Psi \mathcal{T}^2 \). Therefore, once generalized durations are matched, minimization of the remaining hedging error amounts to minimization of \( w' \Psi \mathcal{T}^2 w \) which is done in Theorem 1. Corresponding formulas for the case where the instruments are coupon-bearing bonds are easily obtained (see Appendix A.1.1).

1.3.2 Data

We use data from the Federal Reserve Board’s database of constant maturity zero-coupon yields on U.S. Treasury bills, notes, and bonds. The terms to maturity considered are 3, 6, 12, 24, 36, 60, 84, and 120 months. To mitigate day-of-week effects in the bond market (see Johnston, Kracaw and McConnell (1991)), a weekly frequency data set is constructed by extracting Wednesday observations drawn from the Fed’s daily database, rather than using their weekly database, which consists of weekly averages of daily data. Our sample period is the first week of 1983 through the last week of 2014 for a total of 1,670 observations in the time series dimension. Starting in 1983 avoids the Fed money supply targeting experiment of 1979 to 1982 (see Sanders and Unal (1988)). Table 1 shows means and standard deviations of the weekly yield data corresponding to the eight maturities. The means are monotonically increasing in maturity, from 4.12% to 5.98% (continuously compounded annualized yields) whereas the average volatilities or standard deviations exhibit a hump shape with a maximum of 3.09% at two years, and a low of 2.64% at ten years. We consider a one-month hedging period, from month-end to month-end, but for efficiency of the
model parameters we estimate model parameters in the weekly yield. The eight associated zero-coupon bonds are used as hedging instruments on the last trading day of each month. As this is not necessarily a Wednesday, the daily files are used again to get the correct zero-coupon bond prices when constructing the hedge portfolio.

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

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

For a fair comparison of methods, we set the prices of the target assets by using the Fed yields to value the bonds entering them, rather than using the CRSP prices directly. That is, the contractual terms are taken from CRSP, then priced using the eight zero-coupon yields on the last trading day of the month and linear interpolation. This produces a monthly series that should be a fair target for one-month ahead hedging using the corresponding eight zero-coupon bonds. The issue is that raw CRSP prices (bid-ask midpoints plus accrued interest) might reflect other factors not present

\(^5\)Value weights, as opposed to numbers of certificates.
in the Fed yields such as microstructure noise and, hence, thus do not constitute reasonable grounds for a comparison of methods. The differences between raw CRSP prices and the Fed valuations we use are shown in the lower right panel of Figure 1. Evidently, it would not be of interest to require the hedge portfolio to pick up this discrepancy which is seen to fluctuate in a 1% band.

We consider immunization of movements in the target over periods of one month. Thus, from our monthly target data, 1987:1-2014:12, we form \( T = 334 \) one-month returns for the five-year coupon bonds and portfolio targets. The properties of these return series are shown in the first row of Table 3 for the coupon bond target, and first row of Table 4 for the portfolio target. The average one-month return on the five-years coupon bond is 53bp or 0.53% (1 basis point (bp) equals 0.01%). The standard deviation of the monthly returns is 128bp. For the portfolio target, the average return is 60bp and the standard deviation 147bp.

### 1.3.3 Duration matching as benchmark

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

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

---

6This is the unhedged return, and the column is labelled 'Bias' because average hedging errors are reported in the remainder of the table.
1.3.4 Generalized duration matching

To match generalized rather than basic durations, we start with a classical maximum likelihood factor analysis of the yield factor model (1.1). We assume \( \text{var}(f_t) = I_k \), which is without loss of generality, since factor variances and covariances could be absorbed in the loadings \( B \). For identification, the \( k \times k \) matrix \( B'\Psi^{-1}B \) is restricted to be diagonal. The reported rotation is that where the \( j \)'th factor explains the \( j \)'th most of the variation. A full period estimation with \( k = 3 \) factors produces the estimated loadings shown in Figure 3. As evidenced by the graphical depiction in the figure of each of the three columns \( B_j \) as a function of maturity, we recover the standard finding of flat, steep, and hump-shaped loading patterns, corresponding to the level, slope, and curvature factor structure also highlighted by Litterman and Scheinkman (1991) based on similar plots. To interpret the individual loadings \( b_{ij} \), \( i = 1, \ldots, m \), \( j = 1, \ldots, k \), it ought to be noted that \( b_{ij}^2/\sigma_i^2 \) is the proportion of variance of yield \( y_{\tau_i} \) explained by the \( j \)'th factor, with \( \sigma_i^2 \) the total variance from Table 1 for the \( i \)'th maturity. Taking averages across \( i \), the level, slope, and curvature factors explain 97.83%, 2.02%, and 0.01%, respectively, of the variation in the yield data in the full period. Although these figures may make the curvature factor appear redundant, this is not so if judged by the log-likelihood values or standard information criteria of, say, one-, two- and three-factor models. The remaining 0.07% of variation not explained by the three factors is attributed to the idiosyncratic shocks \( \epsilon_t \) in (1.1). The corresponding estimated variances \( \Psi \) are shown in the first line of Table 5, along with the maximized log-likelihood value and number of parameters. To avoid Heywood cases (the factors explaining more than the total variation for a given maturity, i.e., communality \( \sum_{j=1}^{k} b_{ij}^2/\sigma_i^2 \) exceeding unity), a lower bound of \( 10^{-4} \) is imposed on the unique variances, \( \psi_i/\sigma_i^2 \), for each maturity. All subsequent estimations apply this bound, too, thus avoiding the coincidence of factors with selected key yields, and guaranteeing some minimum amount of diversification in the hedging applications. From the first line of Table 5, the shortest and longest yields have the largest unexplained variation.\(^7\) The seven-year idiosyncratic variance hit the lower bound.

With the estimated \( B \) and \( \Psi \) in hand, we may now proceed to the actual hedging step, using the optimal hedge portfolio from Theorem 1. Again, this portfolio depends on the target to be hedged, and we consider the same coupon bond target and bond portfolio target as before. In each time period, the three-vector of generalized target durations is calculated from (1.8). This varies through time, primarily since both discount function and coupon rate vary. In addition, there is some variation in target duration, see Figures 1 and 2. Target generalized durations are next premultiplied by the constant \( m \times 3 \)-matrix \( \Upsilon^{-1}\hat{F}' \) in (1.9) and the result used in Theorem 1 to get the time-varying hedging weights. Hedging performance is documented in the third

\(^7\)Besides the parameters in the variance-covariance structure, \( \mu \) in (1.1) is estimated by the average yields reported in Table 1.
line of Tables 3 and 4. As in Chambers et al. (1988) and Diebold et al. (2006) and since the portfolio in Theorem 1 minimizes hedging error variance we will in the following primarily compare the different models on the basis-point improvement in RMSE. If deemed large, we refer to this improvement as economic significant. It follows that generalized duration matching performs worse than basic duration matching for the single coupon bond target. Even though we are mainly interested in the RMSE, it is worth mentioning that the average hedging error decreases by matching generalized durations. Moreover, in the rolling estimation both the single and the portfolio coupon hedge portfolio generate a higher mean return than the immunization target, which is an appealing feature seen from the point of view of the portfolio manager. In case of the portfolio target, third line of Table 4, generalized duration matching does improve on basic duration matching by 14% in terms of RMSE, and also the average hedging error (bias) is smaller.

Full period calculations give the investor the benefit of hindsight in that the factor projection matrix \( \hat{F} \) that enters the hedging weights is based on a full period yield factor analysis, including data from time periods following those for which hedging performance is assessed. Similarly, the computation of target generalized durations uses output from the factor analysis. For a more relevant performance analysis corresponding to feasible investment strategies (in line with the duration matching strategy), we next consider rolling yield factor analyses, each based on data for the four-year period immediately prior to the hedging date.\(^8\) Thus, in (1.9), both the matrix and the vector of target durations are now time-varying. Hedging performance based on the feasible strategy is about the same as for full period factor analysis in the single coupon bond target case, but dominates the latter in terms of both RMSE and average hedging error on the portfolio target. Since rolling estimation involves an out-of-sample hedging element, it is not necessarily given in advance that it dominates the full period estimation case. Thus, the empirical results are consistent with the importance of conditioning decisions on relevant information.

### 1.3.5 Flexible functional form calibration of loadings

Considering the simple level, slope, and curvature structure of the loadings in Figure 3, it is natural to smooth across several maturities by parametrizing loadings. This leads to a more general approach to the problem of hedging a claim with maturity \( \tau^* \) (different from the \( \tau_i \)'s of the hedging instruments) by exploiting the functional form of the dependence of each factor loading on term to maturity. Thus, write \( B_j(\tau) \) for the sensitivity to the \( j \)'th factor, viewed as a function of term to maturity \( \tau \). This function is now taken to be a smooth interpolation across all \( i = 1, \ldots, m \) of the \( b_{ij} \) terms, for fixed \( j \), given by a flexible parametric functional form adopted for this

\(^8\)See e.g. Buraschi and Corielli (2005) for theoretical justifications of periodic recalibration of model parameters and initial conditions in no-arbitrage models from the perspective of a portfolio manager.
relation, This formalizes the visual plotting of $B_j$ against $\tau_i$ in Figure 3. The $i^{th}$ row of $B$ now takes the form $b'_i = b(\tau_i)'$, where $b(\tau)' = (B_1(\tau), \ldots, B_k(\tau))$. Hence, using the functional form $b(\cdot)$ for the interpolation in (1.8) obtains the $k$-vector of generalized durations $(\tau b)_+$. 

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

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

$$y_{t,\tau} = f_{t,1} + f_{t,2}(\frac{1-e^{-at}}{at}) + f_{t,3}(\frac{1-e^{-at}}{at} - e^{-at}).$$

(1.12)

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

Diebold et al. (2006) confirms the good empirical fit of the NS yield curve, again reestimated monthly by OLS. Here, factors are treated as parameters and $a$ is fixed across time at a value 0.0609 for $\tau$ measured in months, corresponding to $a = 0.731$ in our case with $\tau$ in annual terms. This value was chosen in another study by Diebold and Li (2006), who argued that it makes the hump in the third loading function in (1.12) occur at 30 months, thus striking an average between the two and three year maturities between which the yield curve hump is commonly observed. In fact, the 0.0609 value generates a maximum at 29.4 months whereas a maximum at 30 months requires $a = 0.717$ (or 0.0598 in monthly terms). Setting $a$ at an exogenously prespecified value circumvents any empirical estimation of $B$ whatsoever. As noted in Section 1.3.1.3, Diebold et al. (2006) do not minimize hedging error variance in

---

$^9\partial B_2(\tau)/\partial \tau = |(1 + at)e^{-at} - 1|/(at^2)^2 < 0$ for all $\tau > 0$, and $\partial B_3(\tau)/\partial \tau = |(1 + at + a^2\tau^2)e^{-at} - 1|/(at^2)$ is positive for small $\tau$ and negative for large $\tau$. 

---
their hedging strategy, and so do not need $\Psi$, either, so their hedge portfolio may be calculated without any preceding yield factor analysis.

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

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

Again, given estimated $B$ (now in terms of $a$) and $\Psi$, we turn to hedging performance. Performance is somewhat worse in RMSE terms than in the unrestricted case for the portfolio target. This applies for both full period and rolling estimation. When hedging the single coupon bond, Table 3, performance is about the same for the full period estimation and better than in the unrestricted case for the rolling period. Hence, utilizing the structure of the loadings allows an out-of-sample improvement in the single coupon target case.

If $a$ is kept fixed at the value $a = 0.731$ consistent with Diebold et al. (2006), but the remaining parameters are still estimated, the results in line seven of each table are obtained. When hedging the single coupon bond, performance is worse than full period and rolling estimation. For the portfolio target, performance is better in terms of RMSE for all models considered so far.
If $\Psi$ is replaced by $\mathcal{J}^{-2}$, which corresponds to the case where the variance-covariance matrix of the idiosyncratic errors in log-prices of the instruments is proportional to the identity matrix, the optimal portfolio coincides with that simply minimizing the sum of squared hedging weights subject to the same two constraints. This portfolio was considered by, among others, Ingersoll (1983). Performance in this case is 8.96bp in terms of RMSE in the rolling estimation for the single coupon bond compared to 10.90bp with estimated $\Psi$. Similarly, when hedging the portfolio target RMSE drops from 48.70bp to 23.88bp. What causes these somewhat surprising results is that the idiosyncratic variances $\psi_i$ in the yield factor model are of similar order of magnitude across maturities, so that $\tau_i^2 \psi_i$ is much smaller for short instruments than for long. Thus, the hedge portfolios become dominated by short instruments and empirically to an excessive extent. More diversification improves performance here, presumably because the factor model only holds to a certain degree of approximation. 10

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

1.4 Dynamically consistent arbitrage-free hedging

In this section, we exploit restrictions from dynamic term structure theory in an attempt to improve the quality of the hedge. We start out by looking more closely at the approach of restricting the functional form of the loadings or factor sensitivities. The restriction of the loading functions in the three-factor model to NS shape from the previous section is a special case of this approach. The critical issue that arises is that adopting a parametrized functional form for the loadings also amounts to imposing a parametrized functional form for the yield curve as a function of term to maturity $\tau$ as is clear from (1.12). In the equation, the parameters on date $t$ are $f_t = (f_{t,1}, f_{t,2}, f_{t,3})'$, and $a$. In other parametrizations of $B$, a similar functional form for the yield curve $\tau \rightarrow y_{t,\tau}$ with parameters $f_t$ and those in $B$ is imposed. Any such functional form severely restricts the set of possible arbitrage-free dynamic term structure models that could be behind interest rate movements in the market and may even rule out that any such model exists. This is the main result of Björk and Christensen (1999). In particular, it follows from Björk and Christensen (1999) and Filipović (1999) that no term structure model would generate yield curves that are of the NS shape (1.12) in each period, no matter how $f_t$ and $a$ are allowed to move over time.

10 See Carcano and Dall’O (2011) for ways to potentially improve hedging performance by accounting for the exposure to (the variance of) modeling errors.
The central concept in this theory is that of consistency between a parametrization of the shape of the yield curve and a dynamic term structure model. Consistency of the two requires that if the initial yield curve \( y_{0,\tau} \) is of the given shape, for some parameter vector, and interest rate changes are driven by the dynamic model in question, then each subsequent yield curve \( y_{t,\tau} \), for \( t > 0 \), is also of the given shape, for some (typically other) parameter vector. Thus, for a given parametrization of the shape of the yield curve, as a function of term to maturity, there is a particular set of dynamic term structure models consistent with this curve shape, in the sense that these models would generate curves of the shape each period. This set could be empty, as it is in the case of the NS yield curve shape. This means that a minimum requirement on the chosen parametrization of the loadings it should generate a yield curve shape for which at least one consistent term structure model exists.

In the following, we use these insights to improve hedging performance. First, we modify the parametrization of the restricted loadings to ensure that there exists a term structure model consistent with the chosen shape. We reestimate the yield factor model with the modified loadings, and assess the resulting hedging performance. This is the route pursued in Section 1.4.1. Next, having identified a term structure model consistent with the type of curves that empirically provide good cross-sectional yield curve fits, we consider in Section 1.4.2 an alternative estimation procedure exploiting the dynamics of this model, as opposed to the yield curve shape. In Section 1.4.3, we then combine the two, exploiting both the shape of the yield curve and the dynamics of the consistent model. Because the consistent model only involves a single stochastic factor, we introduce in Section 1.4.4 a new arbitrage-free dynamic term structure model with level, slope, and curvature factors that all three vary stochastically through time. We apply all three approaches (curve shape, dynamics, and the combination) to this model. Finally, in Section 1.4.6, we consider an alternative approach, relaxing the generalized duration matching constraint and using the dynamic model to trade off violation of the constraint against remaining hedging error variance.

### 1.4.1 Dynamically consistent curve shape

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

\[
y_{t,\tau} = f_{t,1} + f_{t,2} \left( \frac{1 - e^{-\alpha t}}{\alpha t} \right) + f_{t,3} \left( \frac{1 - e^{-\alpha t}}{\alpha t} - e^{-\alpha t} \right) + f_{t,4} \left( \frac{1 - e^{-2\alpha t}}{2\alpha t} \right). \tag{1.13}
\]

\[11\text{Here, examples of dynamic term structure models are those of Vasicek (1977) and Cox, Ingersoll and Ross (1985), and NS is an example of a parametrization of the shape of the yield curve.}\]
This is simply the NS yield curve shape (1.12), augmented with an additional slope factor $f_{t,A}$, and with parameter $2a$ in the new sensitivity or loading function $B_4(\tau) = (1 - \exp(-2a\tau))/(2a\tau)$, instead of $a$, as in $B_2(\tau)$. Thus, the augmentation does not lead to a loss of degrees of freedom, as the additional loading function only depends on the same unknown parameter $a$, but merely achieves consistency with arbitrage-free term structure modeling, as shown by Björk and Christensen (1999).

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

$$dy(t, 0) = a(\theta(t) - y(t, 0))dt + \sigma dW_t, \quad (1.14)$$

and if the yield curve at any arbitrary point in time is of the augmented NS (henceforth ANS) shape (1.13), for some $f_t$ vector, then all subsequent yield curves are also of the ANS shape, with different $f_t$. Thus, the four factors in $f_t$ suffice as state vector for the term structure dynamics. In this sense, the augmented NS curve shape is consistent with dynamic term structure theory, namely, with a particular dynamic model (the HW model). Note that $\theta(\cdot)$ in (1.14) is the time-varying target for mean-reversion that constitutes Hull and White’s extension of the Vasicek model, $\sigma$ is the short rate volatility, and $a$ is the rate of mean-reversion, which coincides with the parameter $a$ in the loading functions of the ANS curve shape (1.13).

To enhance intuition we recast the HW model in the Heath et al. (1992) (henceforth HJM) framework as

$$dy(t, \tau) = \alpha(t, \tau)dt + \sigma\left(\frac{1 - e^{-a\tau}}{a\tau}\right)dW_t, \quad (1.15)$$

where the drift $\alpha(t, \tau)$ under no arbitrage is determined as a function of the volatility function and a market price of risk. The point is that if at any arbitrary point in time $t_0$ the yield curve $y(t_0, \cdot)$ takes the ANS form, and the dynamics are given by (1.15), with the HJM no-arbitrage condition imposed on the drift, then the dynamics for $t > t_0$ may be written in the form

$$dy(t, \tau) = \sum_{j=1}^{4} B_j(\tau)df_{t,j}, \quad (1.16)$$

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

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

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

1.4.2 Consistent yield dynamics

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

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

(1.17)

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

$$\alpha(t, \tau) = \frac{1}{\tau} [y(t, \tau) - y(t, 0)] + \frac{\sigma^2}{2} \frac{\partial^2}{\partial \tau^2} \left[ \int_0^\tau \sigma^2(\tau, \tau') \sigma(t, \tau') \, d\tau' \right] + \frac{\sigma}{\tau} \lambda_t,$$

(1.18)

where $\lambda_t$ is the $d$-vector of market prices of risk. Here, we use the parametrization of Brace and Musiela (1994) where term to maturity $\tau$ rather than maturity date enters as a separate argument in $y(t, \tau)$. As HJM considered the alternative parametrization with maturity date $t + \tau$ instead of $\tau$ as a separate argument, and represented the term structure via instantaneous forward rates instead of the yield curve, we present a brief derivation of (1.18) in the Appendix. The slope term $\sigma(t, \tau) / \partial \tau$ is not present in the drift condition in the HJM parametrization with constant maturity date, and it represents a locally deterministic ageing effect also discussed by Litterman and Scheinkman (1991). The leading spread term in (1.18) is also not present in the standard HJM representation and appears because we consider yields rather than forward rates.

In this framework, we now focus on the consistency between the HW dynamics and the augmented NS curve shape. Hull and White introduced their extension of the Vasicek model through the forward rate volatility function $\sigma_f(t, \tau) = \sigma \exp(-a \tau)$, exponentially decaying for longer maturities. An integration from 0 to $\tau$ and division by $\tau$ produces the yield volatility $\sigma(t, \tau) = \sigma(1 - \exp(-a \tau)) / (a \tau)$ used in (1.15). This yield volatility is recognized as the short rate volatility $\sigma$ from (1.14) times the loading on the second factor in the NS yield curve, $\sigma(t, \tau) = \sigma B_2(\tau)$. In this case, with $d = 1$, the arbitrage-free drift (1.18) is

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

(1.19)

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

$$
\begin{align*}
\frac{dy(t, \tau)}{dt} & = \left\{ \frac{1}{\tau} (y(t, \tau) - y(t, 0)) + \frac{\sigma^2}{2} B_2(\tau) + \sigma \lambda_t \right\} dt + B_2(\tau) \left\{ - \frac{\sigma^2}{a} - dt \right\}.
\end{align*}
$$

(1.20)

Thus, a necessary condition for a model of yield curves $y(t, \cdot)$ to be consistent with HW dynamics is that it include factors with loading functions $B_2(\cdot)$ and $B_4(\cdot)$, e.g., the original NS curve (1.12) without $B_4$ would not suffice, although only the factor associated with $B_2$ is stochastic.
To derive the yield curve shape from the stochastic differentials (1.20), an initial condition is needed. For now, it is useful to take this as an otherwise unrestricted initial yield curve $y(t, \cdot)$ given at time $t$. Assuming from here on a constant market price of risk $\lambda$, the resulting solution for the level of the yield curve is stated in the following proposition.

**Proposition 2.** For initial yield curve $\tau \mapsto y(t_0, \tau)$ and interest rates that follow the Hull-White model (1.15), subsequent yield curves at times $t_1, t_2, \ldots$, with $t_{n+1} - t_n = \Delta_n$, are given by

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

(1.21)

with factors

$$\tilde{f}_{n+1, 2} = (\sigma^2 / a + \sigma \lambda) \Delta_n B_2(\Delta_n) + \sigma \sqrt{\Delta_n B_4(\Delta_n)} z_{n+1},$$

$$\tilde{f}_{n+1, 4} = -(\sigma^2 / a) \Delta_n B_4(\Delta_n),$$

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

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

With the yield curve from the proposition in hand, we are now in a position to design the empirical model that exploits the HW dynamics. As in (1.1), we consider panel data $(y_{i0}, y_{i1}, \ldots, y_{iN})$ on yields, with $y_{in} = (y_{in, \tau_1}, \ldots, y_{in, \tau_m})'$. From (1.21), collecting yields on the left hand side, separating the stochastic term on the right hand side, and allowing for measurement error $\tilde{\varepsilon}_{n+1, \tau_i}$ in the $i$'th yield at time $t_{n+1}$ produces the factor model

$$\tilde{y}_{t_{n+1}, \tau_i} = B_2(\tau_i) \tilde{f}_{n+1, 2} + B_4(\tau_i) \tilde{f}_{n+1, 4} + \tilde{\varepsilon}_{n+1, \tau_i}$$

(1.22)

$$i = 1, \ldots, m,$$

where $\eta = (a, \sigma, \lambda), \{z_n\}_{n=1}^N$ is an i.i.d. $N(0, 1)$ sequence based on increments to $W_t$ from (1.20), and with the definitions

$$\tilde{y}_{t_{n+1}, \tau_i} = y_{t_{n+1}, \tau_i} - y_{t_n, \tau_i + \Delta_n} - \frac{\Delta_n}{\tau_i} (y_{t_n, \tau_i + \Delta_n} - y_{t_n, \Delta_n}),$$

$$\mu_i(\eta) = (\sigma^2 / a + \sigma \lambda) \Delta_n B_2(\Delta_n; a) B_2(\tau_i; a) - (\sigma^2 / a) \Delta_n B_4(\Delta_n; a) B_4(\tau_i; a),$$

$$C_i(\eta) = \sigma B_2(\tau_i; a) \sqrt{\Delta_n B_4(\Delta_n; a)},$$

(1.23)
where the dependence of the loading functions $B_j$ on the parameter $a$ from $\eta$ has been made explicit. Thus, in contrast to the yield factor model (1.1) for levels, our dynamically consistent arbitrage-free factor model is for a certain adjusted yield change $\tilde{y}_{t_{n+1}, \tau_i}$. Christensen and van der Wel (2019) show that in the ideal case with complete data for all maturities we can think of $-\tau_i \tilde{y}_{t_n} + \tau_i$ as the one-period excess return to the discount bond with maturity $\tau_i$ at time $t_n$. In order to form the vectors $\tilde{y}_{t_n} = (\tilde{y}_{t_n, \tau_1}, \ldots, \tilde{y}_{t_n, \tau_m})'$, we employ the shortest maturity $\tau_1$ corresponding to the 3-months bill yield as proxy for $y_{t_n}$, see Chapman, Long and Pearson (1999). Similarly, we use the closest available maturity corresponding to the 10-year note yield as proxy for $y_{t_n,10}^{\Delta_n}$. The remaining vectors are obtained by interpolating between the observed yield so that the panel data set on adjusted yield changes is $(\tilde{y}_1, \ldots, \tilde{y}_N)$. The factor analysis is now again restricted, in the sense that $\mu_i$ and $C_i$ are parametrized functions. In vector form, the consistent factor analysis model is

$$\tilde{y}_{t_n} = \mu(\eta) + C(\eta)z_n + \tilde{\epsilon}_n, \quad (1.24)$$

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

The results from estimating the adjusted yield change model (1.24) appear in lines seven and eight of Table 6. The model accommodates unevenly spaced observations, but as our yield data consist of weekly observations we now set $\Delta_n = \Delta = 1/52$. In contrast with previous results, estimated $a$ is now very close to zero. From (1.20), the stochastic factor in the HW model is the slope factor, but for $a$ near zero, the slope loading $B_2(\tau)$ becomes flat, so the stochastic factor is a level factor after all. This is consistent with the previous findings that the level factor explains most of the variation in the data. In the present estimation of the adjusted yield change model, the volatility parameter $\sigma$ and the market price of risk $\lambda$ are estimated along with $a$. All three parameters are similar across full period and rolling estimation. The market price of risk is negative, corresponding to the negative relation between yields and bond prices. It is significant in the full period estimation, with a t-statistic of -3.3, although it has the largest standard errors of the three parameters in $\eta$. From lines

\[\text{We assume proportionality of the diagonal idiosyncratic error variance matrices } \Psi \text{ and } \tilde{\Psi} \text{ from the yield factor model and the consistent adjusted yield change factor model, respectively.}\]
nine and ten of Table 5, idiosyncratic error variances are similar in the HW model and an unrestricted classical one-factor analysis of the adjusted yield changes \( \tilde{y} \). Table 7 shows the estimated loadings in this unrestricted analysis, revealing an initial slope, then a flat structure for maturities two years and higher. Although the unrestricted model produces a higher full-period likelihood value than the restricted in Table 5, we may still investigate the usefulness of the restricted model for hedging purposes, in particular paying attention also to the important out-of-sample (rolling estimation) performance criterion.

The notion of arbitrage-free consistency is based on the HJM drift restriction (1.18). Table 8 shows the results of estimation with and without this restriction imposed. The first column corresponds to the restricted model already considered, i.e., the reported \( \mu \) vector is determined by the remaining parameters \( \eta = (a, \sigma, \lambda) \) as in (1.23). The second column leaves \( \mu \) free. This introduces eight new mean parameters, but leaves the market price of risk \( \lambda \) unidentified, so the difference in degrees of freedom is seven. Both \( a \) and \( \sigma \) (labelled \( \sigma_1 \) in the table) are similar in the restricted and unrestricted estimations, but the \( \mu \) vectors are very different. In particular, the restricted \( \mu \) is nearly flat in the maturity dimension, because estimated \( a \) is close to zero. The upshot is that the arbitrage condition is strongly rejected (the LR test takes the value 56.7 compared to a critical value of 14.1 at the 5% level in the asymptotic \( \chi^2 \) distribution on seven degrees of freedom). Again, the restricted model could nevertheless prove useful for hedging purposes.

Turning to hedging performance, the estimated \( a \) from the restricted model is inserted in the ANS loadings and these are used along with the idiosyncratic variances to form the hedge. Lines twelve and thirteen of Tables 3 and 4 show the results. Hedging performance is the best so far for both the single coupon bond and the portfolio target. The improvement relative to the previous hedging results for ANS loadings is considerable. Since only the parameter estimates separate the two cases, the results document the value of imposing the consistency restrictions from arbitrage-free term structure dynamics in the estimation stage of the hedging process.

### 1.4.3 Consistent yield dynamics and curve shape

We now present a new framework imposing both the HJM no-arbitrage drift restriction on the yield dynamics, as in the previous section, and the consistent yield curve shape, as in Section 1.4.1 on ANS. The combination of both types of restrictions is accommodated in a state space model implemented using the Kalman filter.

The hedging approach of the previous subsection is based on the yield curve model (1.21) from Proposition 2, where an arbitrary initial yield curve \( y(t_0, \cdot) \) is taken as given. As a result, subsequent yield curves \( y(t_{n+1}, \cdot) \) involve the second and fourth loading
functions from the ANS curve shape, but also a term depending on the previous yield curve \( y(t_n, \cdot) \), so it is not necessarily of ANS form. The reason is that the initial curve \( y(t_0, \cdot) \) is unrestricted. This specification carries with it the potential advantage that the current observed yield curve may be used as initial curve, which is the idea in the proposed hedging strategy in Section 1.4.2. The ability to incorporate as much of the information in current market data as possible is often assumed to be desirable. On the other hand, as already discussed, the ANS curve shape has other desirable empirical as well as theoretical properties, i.e., it exhibits level, slope, and curvature features, and is consistent with the HW model in the sense that the yield curves generated by the latter may take on ANS shape in every time period. This was the motivation for the focus on the HW model from the outset. On these grounds, it seems an appealing alternative to consider the initial curve \( y(t_0, \cdot) \) to take the ANS form. It is exactly in this case that every subsequent yield curve \( y(t_n, \cdot) \) is also in the ANS class such that the combined model is fully consistent. We now develop the alternative hedging approach following these ideas.

When the yield curve at time \( t_n \) is on the ANS form \( y(t_n, \tau) = \sum_{j=1}^{4} B_j(\tau) f_{n,j} \), with loadings \( B_j \) defined as in (1.13), then the first term in (1.21), giving the dependence of the yield curve at time \( t_{n+1} \) on the previous curve at time \( t_n \), takes the form

\[
\frac{1}{\tau} [(\tau + \Delta n) y(t_n, \tau + \Delta n) - \Delta n y(t_n, \Delta n)] = f_{n,1} B_1(\tau) + e^{-a\Delta n} (f_{n,2} + a\Delta n f_{n,3}) B_2(\tau) + e^{-a\Delta n} f_{n,3} B_3(\tau) + e^{-2a\Delta n} f_{n,4} B_4(\tau) = B(\tau) H(\Delta n) f_n,
\]

where \( B(\tau) = (B_1(\tau), \ldots, B_4(\tau)) \). The first equality in (1.25) follows from the general theorems below with the \( 4 \times 4 \) transition matrix \( H \) given by

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

Thus, in the present case, this term also is recognized to be of ANS form which means that it is a function of term to maturity \( \tau \), it is spanned by the loadings \( B_j(\tau) \). This makes the entire yield curve \( y(t_{n+1}, \tau) \) in (1.21) to take the ANS form when the yield curve at \( t_n \) does. Consequently, starting from an initial yield curve at \( t_0 \) on ANS form with dynamics following the HW model all subsequent yield curves at times \( t_1, t_2, \ldots \) will be on ANS form. What happens here is that the ANS curve shape includes the \( B_2 \) and \( B_4 \) terms that by Proposition 2 are necessary for consistency with the HW model, and that the full curve shape is maintained when shifting maturity as in the left hand side of (1.25). Other curve shapes consistent with HW could be obtained by supplementing the component involving \( B_2 \) and \( B_4 \) in different ways, subject to
this invariance condition. The reported empirical relevance of the original NS curve shape leads us to focus on the minimal augmentation of this, which both includes $B_2$ and $B_4$ and satisfies the invariance condition. By necessity, a factor with loading $B_4$ must be added, and since the resulting ANS curve satisfies the invariance condition, this constitutes the required minimal augmentation.

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

**Proposition 3.** For an initial yield curve on augmented Nelson-Siegel form, $y(t_0, \tau) = B(\tau) f_0$, and interest rates that follow the HW model, (1.15), subsequent yield curves at times $t_1, t_2, \ldots$, with $t_{n+1} - t_n = \Delta_n$, are on augmented Nelson-Siegel form, $y(t_{n+1}, \tau) = B(\tau) f_{n+1}$, with factors

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

Here, $H$ is the transition matrix given in (1.26), $v_1, v_2, \ldots$ is a sequence of independent $N(0, \Omega(\Delta_n))$ variables with the (2,2) entry, $\sigma^2 (1 - e^{-2a\Delta_n})/(2a)$, the only non-zero element in $\Omega(\Delta_n)$, and $\bar{f}$ the long-run factor levels are given by

$$\bar{f} = \left[ f_{0,1}, \frac{\sigma}{a} \left( \frac{\sigma}{a} + \lambda \right), 0, -\frac{\sigma^2}{2a^2} \right]^T.$$  

From the proposition, the first factor is constant, at the level measured from the initial curve at time $t_0$, and the third factor is exponentially decaying. The fourth factor reverts exponentially towards level $-\sigma^2/(2a^2)$, and the second factor, which is the only stochastic factor, mean reverts to $\sigma/(\sigma/a + \lambda)$. In $f_{n+1}$ the second and fourth entry of the $4 \times 1$ vector $[I - H(\Delta_n)] \bar{f} + v_{n+1} = \tilde{f}_{n+1}$ may be recognized as $\tilde{f}_{n+1,2}$ and $\tilde{f}_{n+1,4}$ from Proposition 2, respectively, while the first and third entries of $\tilde{f}_{n+1}$ are zero. The remaining term in $f_{n+1}$ is picked up from the dependence (1.25) on the yield curve at $t_n$, i.e., $H(\Delta_n) f_n$. By inserting $\tau = 0$, it is seen that the short rate is given by $y(t_n,0) = f_{n,1} + f_{n,2} + f_{n,4}$, or, alternatively, the stochastic factor is given in terms of the short rate of interest as $f_{n,2} = y(t_n,0) - f_{n,1} - f_{n,4}$.

Compared to the original yield factor model (1.1), of the form $y_t = B f_t + \varepsilon_t$, there is not only more structure on the shape of the loading functions in the arbitrage-free consistent model in Proposition 3, but also on the factor dynamics. The generalization of the classical static factor analysis structure that allows for (first-order Markov) serial dependence in factors is simply the Kalman filter combined with maximization of the appropriate likelihood function based on the filtered innovations in data. Of course, the model (1.24) for adjusted yield changes in the previous subsection also has factor structure with consistency restrictions on loadings, but the factors $z_n$ are serially independent, hence filtering is unnecessary.
To make the state space form of the model from Proposition 3 explicit, the state vector is the vector of factors $f_n = (f_{n,1}, \ldots, f_{n,4})'$, so the state transition equation is given by

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

with transition matrix $\Phi_1 = H(\Delta_{n-1})$ and intercepts $\Phi_0 = (I - \Phi_1)\bar{f}$. Here, $\Phi_0$ depends on all three model parameters $\eta = (a, \sigma, \lambda)$, and $\Phi_1$ only on $a$. Thus, the no-arbitrage consistency conditions place at least 17 nonlinear restrictions on the 20 coefficients in the transition equation. In fact, $\eta$ is identified in $(\Phi_0, \Phi_1)$, i.e., $a$ is clearly identified from $\Phi_1 = H(\Delta_{n-1})$ from (1.26), and then $\sigma$ may be solved for from $\Phi_{0,4}$ using $\bar{f}$ from (1.27), and in turn $\lambda$ from $\Phi_{0,2}$. Hence, the number of restrictions is exactly 17. As $\Omega$ would have 10 free parameters in the unrestricted Kalman filter, but here is a function of the same model parameters $\eta$ which are already identified in $(\Phi_0, \Phi_1)$, this yields another 10 restrictions, for a total of 27 in the transition equation alone.

The Kalman filter measurement equation is

$$y_n = B f_n + \epsilon_n,$$

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

We base the Kalman filter recursions on the Koopman, Shephard and Doornik (1999) low storage algorithm, inserting the updating step in the prediction step to save on calculations, and because we modify the algorithm to the square-root case, a brief description is provided in the Appendix. The modified recursions generate a sequence of yield vector innovations $\zeta_n = y_{tn} - \mathbb{E}(y_{tn} | Y_{n-1})$ for $Y_{n-1} = (y_{t1}, \ldots, y_{tn-1})$, with associated prediction error variances $\Gamma_n = \text{var}(\zeta_n | Y_{n-1})$. The model parameters $(\eta, \Psi)$ are estimated by maximizing the log-likelihood based on $\zeta_n$ i.i.d. $N(0, \Gamma_n)$. Compared to the estimates of $(\eta, \Psi)$ from the previous subsection, the new estimates impose additionally impose all available no-arbitrage consistency restrictions, a total of $4m + 17$ theory restrictions, on the factor dynamics.

Estimation results appear under the label ‘HW and ANS’ in rows 9 and 10 of Table 6. The estimated $a$ value is between those from cross-sectional curve fitting and those exploiting dynamics. There is a great deal of variation in $a$ in the rolling estimation, consistent with the notion that the required slope varies through time and sometimes
makes the single stochastic factor more akin to a level factor, which is what apparently happens when focussing only on dynamics (rows 7 and 8 of the table). The parameters $\sigma$ and $\lambda$ are similar to those estimated purely from the dynamics, which makes sense, since they do not directly impact the cross-section. From Table 5 row 6, the idiosyncratic variances $\Psi$ are now higher than in other models, except at the three-year maturity. This is a result of imposing more structure, i.e., both in the cross-sectional and time series dimension simultaneously, and thereby getting a poorer fit, as also seen from the maximised likelihood value.

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

1.4.4 A dynamically consistent stochastic level, slope, and curvature model

There are two different ways of counting ‘factors’ in a given model. While there are four factors, $f_t$, in the curve shape (1.13), there is only one source of randomness (one driving Wiener process) in the consistent dynamic HW model (1.14). When combining both the curve shape and the dynamics (the cross-sectional and time series dimensions), the resulting model described in Proposition 3 may therefore either be labelled a four-factor model or a one-factor model. The consistency issue relates to the fact that interest rate dynamics place restrictions on the manner in which the term structure can change shape over time, i.e., on the four ‘factors,’ and in the present model restricts three of them to be deterministic. The count of sources of randomness is therefore probably most meaningful once consistency is imposed, so that the combined model is described as a one-factor model in this case.

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

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

$$\sigma_f(\tau)' = (\sigma_1, \sigma_2 e^{-a\tau}, \sigma_3 a \tau e^{-a\tau}).$$

(1.28)

The first stochastic factor affects forward rates of all maturities equally. Restricting \(\sigma_f(\tau)\) to only consist of its first entry generates the Ho-Lee model by Ho and Lee (1986) (consistent with affine forward rate curves), and the second entry alone generates the HW model (consistent with ANS curves). Thus, the two first entries taken together serve to generate the two-factor model considered by Heath et al. (1992). With the first two entries capturing level and slope, we propose the third, \(\sigma_3 a \tau e^{-a\tau}\), as a natural candidate for curvature, or hump. To avoid nonstationarity of the dynamic model, we note that the rate of exponential decline of the volatility function in the maturity direction corresponds to the rate of mean reversion of the short rate process, see (1.15)-(1.16), so we work with the slightly generalized forward volatility structure

$$\sigma_f(\tau)' = (\sigma_1 e^{-b\tau}, \sigma_2 e^{-a\tau}, \sigma_3 a \tau e^{-a\tau}),$$

(1.29)

with \(b > 0\) a small, positive number to guarantee stationarity. For \(b \downarrow 0\), the structure (1.28) is approached.

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

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

(1.30)

and a constant market price of risk vector \(\lambda\). In line with the recommendation by CDR, we have chosen an independent factor specification, but an immediate generalization is obtained by replacing \(\text{diag}()\) by a general variance-covariance matrix. The yield volatilities corresponding to (1.28) are obtained by taking averages over maturities, i.e. \(B_{1:3}(\tau)\) is the \(1 \times 3\) vector

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

(1.31)

Again, for \(b \downarrow 0\) the first stochastic factor affects yields of all maturities equally, \(B_1(\tau) = 1\), and the functions \(B_{1:3}(\tau)\) are the same as those that span the NS yield curve shape in (1.12). Since \(B_{1:3}(\tau)\) captures the directions in which the stochastic terms affect the yield curve, the interest rate model will have both level, slope and curvature changing stochastically. As we already discussed, no arbitrage-free interest rate model is consistent with the NS yield curve shape. The next theorem derives a minimal set

---

13CDR document superior out-of-sample performance of the independent-factor model relative to the correlated-factor model.
of additional functions of τ to be added to the NS curve shape in order to span curves consistent with the suggested SLSC model.

**Theorem 4.** *If interest rate dynamics follow the stochastic level, slope, and curvature model (1.30), then consistent yield curves must in addition to \( B_{1:3}(\tau) \) in (1.31) necessarily include the functions

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

Conversely, the yield curve family spanned by \( B_{1:3}(\tau) \) and (1.32) is consistent with the stochastic level, slope, and curvature interest rate model.*

Thus, any representation of yield curves consistent with the SLSC model must include the seven functions given in (1.31) and (1.32) or some linear transformation thereof. In the remainder of the paper we choose for the sake of convenience to work with the following particular rotation of the additional yield curve functions,

\[
B_{4:7}(\tau) = \left( \frac{1-e^{-2a\tau}}{2a\tau}, \frac{1-e^{-2b\tau}}{2b\tau}, \frac{1-e^{-2a\tau}}{2a\tau} - e^{-2a\tau}, a\tau e^{-2a\tau} \right),
\]

which clearly span the same set of curves as (1.32) do. The loading function \( B_4 \) is the same loading function added to the NS curve to make it consistent with HW interest rate dynamics, i.e., the ANS curve, and it captures that stochastic changes in the \( B_2 \) direction induce drift in this direction by the HJM condition. The loading function \( B_5 \) is equivalently the drift direction induced by the first stochastic factor. The drift direction induced by the stochastic curvature factor is \( B_6 - B_7/2 \), but a yield curve including this function and \( B_{1:5}(\tau) \) implies that the two leading (spread and slope) drift terms in (1.18) are not spanned. In particular, the drift has \( B_6 \) and \( B_7 \) in proportions different from 1 to -1/2, and thus we need to include these two functions separately. The result that the yield curve family defined by \( B(\tau) = B_{1:7}(\tau) \) is minimal among all families not restricting coefficients on the necessary τ-functions to specific constants is stated in the following corollary.

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

Henceforth, we refer to the yield curve family determined by \( B(\tau) = B_{1:7}(\tau) \) as the SLSC curve shape family, because it is naturally associated with the SLSC dynamic model in the sense of the corollary. The (smaller) minimal consistent family in which the \( B_{4:7} \) functions enter the yield curve with fixed coefficients (as opposed to general coefficients, as in the corollary) is considered by Christensen et al. (2011) (CDR), and the relation to our more general SLSC curve shape is discussed further in section 1.4.4.1.

\[^{14}\text{For } b = 0 \text{ the first function should be } \tau. \text{ The proof of this follows the same steps as that for } b > 0 \text{ given in the Appendix.}\]
We now derive how coefficients, or factors, on the loading functions $B(\tau)$ will change over time for the SLSC model starting from a general yield curve and an initial SLSC curve spanned by $B(\tau)$, respectively. First, in the case where the initial yield curve at time $t_0$, $y(t_0, \tau)$, has general shape and yield dynamics follow the SLSC model (1.30), a result similar to Proposition 2 is obtained. Thus, we can write each subsequent curve as a sum of the same transformation of the initial curve and a term spanned by $B(\tau)$.

**Theorem 5.** For an initial yield curve $y(t_0, \tau)$ and interest rates that follow the stochastic level, slope, curvature model (1.30), subsequent yield curves at times $t_1, t_2, \ldots$ with $t_{n+1} - t_n = \Delta_n$ take the form

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

(1.34)

for $B(\tau) = (B_{1,3}(\tau), B_{4,7}(\tau))$ from (1.31) and (1.33), and with factors

$$f_{n+1} = (I - H(\Delta_n)) f_n + v_{n+1}$$

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

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

and the upper left $3 \times 3$ submatrix of $\Omega$ is

$$\Omega_{1,3}(u) = \begin{pmatrix}
s_1^2 u B_5(u) & 0 & 0 \\ 0 & s_2^2 u B_4(u) + s_3^2 u [B_6(u) - B_7(u)]/2 & s_3^2 u B_6(u) / 2 \\ 0 & s_3^2 u B_6(u) / 2 & s_3^2 u B_4(u) \end{pmatrix}$$

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

$$\bar{f} = \left[ \frac{\sigma_1}{b} \left( \frac{\sigma_1}{b} + \lambda_1 \right), \frac{\sigma_2}{a} \left( \frac{\sigma_2}{a} + \lambda_2 \right), \frac{\sigma_3}{a} \left( \frac{\sigma_3}{a} + \lambda_3 \right), -\frac{\sigma_1^2}{2a^2} - \frac{\sigma_2^2}{2a^2}, -\frac{\sigma_1^2}{2a^2}, -\frac{3\sigma_2^2}{4a^2}, -\frac{\sigma_2^3}{4a^2} \right]^\prime.$$
with factors
\[ f_{n+1} = \bar{f} + H(\Delta_n)[f_n - \bar{f}] + \nu_{n+1}, \]
where \( \nu_1, \nu_2, \ldots \) is a sequence of serially independent \( N(0, \Omega(\Delta_n)) \) vectors and \( \bar{f}, H, \) and \( \Omega \) all are the same as in Theorem 5.

In the yield curve at time \( t_{n+1} \), the term \( B(\tau)H(\Delta_n)f_n \) reflects the impact of the previous curve at \( t_n \), which stems from the appropriate generalization of (1.25) from ANS to SLSC curve shape. The impact of the SLSC dynamic model is captured by the additional term \( B(\tau)[(I - H(\Delta_n))\bar{f} + \nu_{n+1}] \), recognized as \( B(\tau)\tilde{f}_{n+1} \) from Theorem 5 for general yield curve shapes. This lends further credence to the observation that consistency between curve shape and dynamic model requires that the curve shape not only accommodate the yield curve changes induced by the dynamic interest rate model, but also that it is invariant to the initial term in (1.34).

We estimate the SLSC model in three ways analogously to the model with a single stochastic factor and refer to these specifications as the SLSC curve, the SLSC dynamic, and the SLSC combined model, respectively. The SLSC curve model is estimated by performing a restricted seven-factor analysis with loading functions parameterized by \( B(\tau) \). Then for general curve shape we estimate the SLSC dynamic model for \( \tilde{y}_n \) given by (1.24) with \( z_n \) three-dimensional, \( C = B_{1:3}M_{1:3}(\Delta_n) \) for \( M_{1:3}M_{1:3}' = \Omega_{1:3} \) and \( \mu = B(I - H(\Delta_n))\bar{f} \), where \( B \) is \( m \times 7 \) with typical row \( B(\tau_i) \). Lastly, for SLSC yield curves spanned by \( B(\tau) \) we estimate the SLSC combined model using the Kalman filter with measurement equation (1.35), adding idiosyncratic error terms with variances \( \psi_i \) to the yields, and transition equation (1.36). In estimations we set \( b = 0.02 \) in (1.31) and (1.33) to avoid problems with non-stationary factor dynamics. For this value the first factor’s loading on the 10-year yield, the longest maturity in our empirical study, is \( B_1(10) = 0.91 \), and the factor approximately impacts yields equally over the relevant range.

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

Table 6 shows further that the volatilities \( \sigma_1 \) and \( \sigma_2 \) of the level and slope factors in the SLSC model are roughly equal, whereas the volatility \( \sigma_3 \) of the curvature factor is about twice as high. The left panel of Figure 8 shows that the volatilities of the level and slope factors in the model combining curve shape and dynamics are similar period by period and stable over the period from 1992 through 2005. Outside of this period, they take values twice as large. The volatility of the curvature factor is higher except in the first two years, and varies more through time, with two peaks around 2004 and 2011. The remaining columns of Table 6 show that not all market prices of factor risks are priced. The market price of slope risk, \( \lambda_2 \), takes values, both in absolute and relative terms, close to twice as large as the two remaining factor risks. This follows also from the right panel of Figure 8, where we see that \( \lambda_2 \) in the SLSC combined model is typically largest in magnitude. It is negative during most of the sample period, but positive around 2005-2008 and after a certain point in 2012. The market price of level risk, \( \lambda_1 \), is relatively close to zero, varies less over time and is nearly always negative. The market price of curvature risk, \( \lambda_3 \), varies almost as much \( \lambda_2 \) through time, but switches sign often.

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

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

The HJM drift restriction (1.18) is tested again in the SLSC dynamic model because arbitrage-free consistency relies on this. The third and fourth columns of Table 8 show results for the restricted and unrestricted models, respectively. Again, restricted \( \mu \) is given in terms of the seven parameters \( \eta = (a, \sigma, \lambda) \) as \( B(I - H(\Delta_n))\tilde{f} \). The parameters \( (a, \sigma) \) are very similar in the restricted and unrestricted estimations, and also the resulting \( \mu \) vectors are now much closer, in contrast to the HW case (first two columns of table). This shows that the pattern of time-series averages of adjusted yield changes is well-matched by the particular SLSC curve given by coefficients (factor averages) \( (I - H(\Delta_n))\tilde{f} \) consistent with the absence of arbitrage opportunities. The LR-statistic
takes the value 4.55, for a p-value of 47% in the asymptotic $\chi^2$ distribution on five degrees of freedom (the restricted model introduces three market prices of risk and drops eight parameters in $\mu$), i.e., the test fails to reject at all conventional levels.

From the hedging performance documented in Tables 3 and 4, the SLSC approach clearly dominates all others considered. For the single coupon bond target, roughly the same hedging performance is obtained from restricted factor analysis based on the SLSC curve shape, exploiting SLSC dynamics (restricted factor analysis of adjusted yield changes), and a combination of the two, using the Kalman filter. The SLSC combined model indeed has the lowest RMSE at 6.08bp with the RMSE reduced by more than 25% compared to the HW model. For the coupon bond portfolio target, improved performance is also obtained by exploiting the SLSC dynamics. RMSE is 15.33bp based on cross-sectional SLSC curve shapes in the out-of-sample hedging rolling estimation case, compared to 12.37bp for the SLSC dynamic model, and 12.85bp when combining SLSC curve shape and dynamics. Of all other cases considered, the HW dynamic model without curve shape imposed is best, at 22.55bp (rolling estimation), about 1.5 times as high as the worst of the SLSC based methods, and nearly two times the RMSE of the SLSC combined curve shape and dynamics based method.

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

### 1.4.4.1 Relation to affine term structure models

In this section we briefly clarify the relation between our SLSC model and the class of affine term structure models of Duffie and Kan (1996). In particular, a restricted special case of our model, with the four deterministic factors fixed at their long-run levels, $f_{4:7}(t) \equiv \bar{f}_{4:7}$, coincides with the independent factor model considered by CDR. To see this, note that the SLSC yield curve family is not the minimal family consistent with the SLSC dynamic model. If $f_{4:7}(t)$ at some point $t$ takes the value $\bar{f}_{4:7}$, then from (1.36) it remains constant at this level.\(^{15}\) Thus, the set of yield curves on the form

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

\(^{15}\)Equivalently, the factor values $f_{4:7}$ ensuring no drift in the $B_{4:7}$ directions solve the problem $V(B_{4:7}(\tau) f_{4:7}) + B_{4:7}(\tau) C_{4:7} = 0$ in the notation of the proof of Theorem 4 in the Appendix, i.e., $f_{4:7} = h_{4:7}^{-1} C_{4:7} = \bar{f}_{4:7}$.\)
where the three coefficients $\phi_{1:3}$ can take any value, forms a consistent family. If the initial yield curve happens to take the form (1.37), then the resulting restricted special case of the SLSC model is a three-factor affine model.

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

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

and the dynamics under the risk-neutral probability measure, $Q$, for the factors $X_t = f_{1:3}(t) - \bar{f}_{1:3}$ (imposing $\lambda = 0$ on $\bar{f}_{1:3}$ from Theorem 5 to get $\bar{f}_{1:3}$) are given by

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

with parameters

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

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

$$y(t, \tau) = -A(\tau) + B_{1:3}(\tau)X_t,$$ (1.38)

with

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

The particular independent factor affine three-factor model of CDR corresponds to the case $b > 0$.\(^{16}\) Even if the directions $B_{4:7}$ are removed from the drift function, they still enter the yield curve, now through $A(\tau)$ with constant coefficients, and thus are required for consistent curve shapes, cf. Theorem 4. Since changes in the $B_{1:3}$ directions are stochastic, they can clearly not be removed, so the minimal consistent curve shape for the SLSC model is (1.38), or $-A(\tau) + B_{1:3}(\tau)\phi_{1:3}$, where $\phi_{1:3}$ can take any value.

\(^{16}\)Strictly, the special case of the SLSC model considered in the present section is the model that results from applying the method of CDR with $\kappa_{11} > 0$ ($\varepsilon > 0$ in their notation), but in the limit $\kappa_{11} = 0$ (i.e., $\varepsilon = 0$), stable long-run factor levels are precluded, and the model is no longer affine (there is a discontinuity at zero).
We estimate the reduced SLSC model with curves restricted to the form (1.37) using the Kalman filter with measurement equation (1.38), still including an error term with variance \( \Psi \), and transition equation
\[
X_{t+1} = X + H_{1:3}(\Delta)[X_t - \bar{X}] + v_{t+1},
\]
where the \( v_{t+1} \) are i.i.d. \( N(0, \Omega_{1:3}(\Delta)) \) with \( \Omega_{1:3} \) from Theorem 5, and \( \bar{X} \) the long-run or mean-reversion levels of factors under the objective probability measure,
\[
\bar{X} = \kappa^{-1} \text{diag}(\sigma) \lambda = \left[ \frac{\sigma_1}{b} \lambda_1, \frac{\sigma_2}{a} \lambda_2 + \frac{\sigma_3}{a} \lambda_3, \frac{\sigma_3}{a} \lambda_3 \right]'.
\]

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

### 1.4.5 Statistical performance evaluation

In order to evaluate the improvements in hedging performance from each model statistically, we construct a standard t-statistic along the lines of Diebold and Mariano (1995) and Giacomini and White (2006) with duration matching as fixed benchmark. To that end, denote the loss differentials from the \( i \)'th model relative to the duration matching approach by \( d_{i,t} \). The test of equal hedging performance can be conducted using
\[
S = T^{1/2} \bar{d} \tilde{V}^{-1/2},
\]
where \( \bar{d} = T^{-1} \sum_{t=1}^{T} d_{i,t} \) and \( \tilde{V} \) is an estimate of the long-run variance of the loss differentials. We employ in the following a HAC estimator and follow state-of-the art good practice by using the data-dependent bandwidth selection by Andrews (1991) based on an AR(1) approximation and a Bartlett kernel. We also do a Model Confidence Set (MCS) procedure (Hansen, Lunde and Nason, 2011) to compare the performance of all the hedging procedures under consideration. For a fixed significance level, \( \alpha \), the procedure identifies the MCS, \( \hat{M}^*_{\alpha} \), from the set of competing procedures, \( M_0 \), which contains the best models with \( 1 - \alpha \) probability. The procedure is conducted recursively based on an equivalence test for any \( M \subseteq M_0 \) and an elimination rule, which identifies and removes a given model from \( M \) in case of rejection of the equivalence test. The equivalence test is based on pairwise comparisons using the t-statistic \( S_{ij} \) for all \( i, j \in M \) and the range statistic \( T_M = \max_{i,j \in M} |S_{ij}| \), where the eliminated model is identified by \( \arg\max_{i \in M} \sup_{j \in M} |S_{ij}| \). Following Hansen et al. (2011), we implement the procedure using a block bootstrap and \( 10^4 \) replications. We only conduct these comparisons for the feasible investment strategies and use a significance level of \( \alpha = 5\% \).

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

1.4.6 Trading off hedging error bias and variance

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

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

(1.41)

This trades off hedging error bias and variance, whereas the generalized duration matching approach can be seen as minimizing hedging error variance given unbiasedness. To see this, it is shown as part of the proof of Theorem 1 in (A.1.3) that the hedging error is

$$r_{t+1}^* - w' r_{t+1} = (w' \mathcal{T} B - (\tau b)^*_t) \Delta f_{t+1} + (w' \mathcal{T} \Delta \varepsilon_{t+1} + \Delta \xi_{t+1}^*).$$

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

The following theorem offers an alternative immunization strategy, based on the principle of the minimization of conditional mean squared hedging error (1.41). This approach requires a dynamic factor model, as it involves the conditional factor prediction, $\mathbb{E}_t [f_{t+1}] = \mu_{t+1|t}$, and conditional prediction error variance, $\text{var}_t [f_{t+1}] = \Sigma_{t+1|t}$, to form the hedge portfolio.

Theorem 7. The immunization portfolio $\tilde{w}$ that minimizes the conditional mean squared hedging error among all linear portfolio rules,

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

(1.42)

under the assumptions $\mathbb{E}_t (\Delta \varepsilon_{t+1,t}) = 0$ and $\mathbb{E}_t (\Delta \xi_{t+1}^*) = 0$, is given by the instruments weights

$$\tilde{w} = \mathcal{J}^{-1} \Psi^{-1} B \left[ \Sigma_{t+1|t} + \left( \mu_{t+1|t} - \mu_{t|t} \right) \left( \mu_{t+1|t} - \mu_{t|t} \right)' \right]^{-1} + B' \Psi^{-1} B \right]^{-1} (\tau b)^*_t.$$

(1.43)
Further imposing that weights sum to one, \( w' \eta = 1 \), adjusts the solution to

\[
\tilde{w} = \bar{w} + (1 - \tilde{w}' \eta) \frac{\Lambda t}{\bar{t}' \Lambda \bar{t}}
\]

for

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

In the limiting case where the uncertainty about factors next period or the expected values of factors next period increase to infinity, the weights in (1.43) reduce to \( \mathcal{T}^{-1} \Psi^{-1} B (B' \Psi^{-1} B)^{-1} (\tau b)' \), this being exactly the hedge from Theorem 1. Thus, the hedge matching generalized durations and minimizing remaining variance is a special case of the general principle of minimizing conditional mean squared hedge error. It applies when the dynamic factor model is unavailable, or the portfolio manager is unwilling to fully exploit it. In these cases the solution is to put full weight on removing the contribution of factors to hedging error by using as many degrees of freedom as there are factors to match generalized durations.

The hedging performance of the approach detailed in Theorem 7 is documented in Table 10 for the single coupon bond target and in Table 11 for the portfolio target. In each case, all the models estimated using the Kalman filter are considered, i.e., the combined models involving both curve shape and dynamics. This includes the HW-ANS model, the SLSC combined model, and the restricted affine special case of this. Line by line, performance is seen to be poorer than in the corresponding cases in Tables 3 and 4. This suggests that it is indeed important to remove all factor contribution to hedging error, which is the generalized duration approach, rather than balancing factor versus idiosyncratic variance. Still, the SLSC models dominate in terms of hedging performances relative to the competing approaches considered.

### 1.5 Conclusion

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

We are grateful to Francis X. Diebold, Bob Jarrow, Andreas Schrimpf, Martin Møller Andreasen, Frank de Jong, Peter Spencer, Christian M. Dahl, and Esben Høg for useful comments and suggestions. We also thank participants at the Sixth Annual Conference of the International Association for Applied Econometrics (2019) at University of Cyprus, the econometrics workshop at University of Pennsylvania (2018) and the joint Econometrics-Finance seminar at Aarhus University (2017), and the 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 for research support.
1.6 References


1.6. References


A.1 Appendix

A.1.1 Hedging with coupon-bearing instruments

The expression (1.8) for the generalized duration vector of a payment stream facilitates not only the hedging of such a stream, but also the use of streams as hedging instruments. This includes hedging with coupon-bearing bonds instead of just zero-coupon bonds that may be unavailable. If there are \( L \) coupon-bearing hedging instruments, we use expression (1.8) to calculate the generalized duration vectors \((\tau b)_\star\) and \((\tau b)_\ell\), \(\ell = 1, \ldots, L\) of both the target and each of these instrument streams.

The main difference compared to hedging with zero-coupon instruments is that the preceding yield factor analysis no longer is carried out on the instruments themselves. Instead, it is still applied to a balanced panel data set of yields not precisely corresponding to the set of available hedging instruments that may vary from period to period as bonds age, mature, etc.. Thus, \( b_h \) in (1.8) is interpolated based on output from the zero-coupon yield factor analysis, applied to data from the preceding period. Next, the construction of hedging weights on the coupon-bearing instruments may proceed in analogy with the case of zero-coupon hedging instruments. Thus, form the \( L \times k \) matrix of generalized durations \( B \), with typical row \((\tau b)_\ell\). This is the matrix that specializes to \( JB \) in the zero-coupon instrument case. The hedge portfolio for general instruments is then given by

\[
\tilde{w} = Y^{-1}B'(B'Y^{-1}B)^{-1}(\tau b)_\star. \tag{A.1.1}
\]

Here, the \( L \times L \) matrix \( Y \) captures the idiosyncratic error variance of the coupon-bearing bond returns, which may be constructed by interpolation and weighted summation across coupons of elements of \( \Psi \) from the yield factor analysis. Again, (1.11) from the Theorem is used to obtain value matching (full investment), but now with \( \Lambda = Y^{-1} - Y^{-1}B'(B'Y^{-1}B)^{-1}B'Y^{-1} \).

A.1.2 Proofs

Proof of Theorem 1. To find the portfolio weights that solve the minimization problem (1.10), we first need an expression for the hedge error variance. The generalized durations of the target stream are by (1.8) the vector \((\tau b)_\star\), so analogously to (1.6) the return on the target stream is

\[
r_{t+1}^* = -(\tau b)^T \Delta f_{t+1} + \Delta \xi_{t+1}^*, \tag{A.1.2}
\]

for \( \xi_{t+1}^* \) the idiosyncratic error in the target log-price, \( \log p_t^* \). By substituting the return on an arbitrary linear portfolio of the instruments given in (1.6), the hedging error is

\[
r_{t+1}^* - w^T r_{t+1} = -(\tau b)^T \Delta f_{t+1} + \Delta \xi_{t+1}^* + w^T(\bm{B} \Delta f_{t+1} + \Delta \varepsilon_{t+1}) \tag{A.1.3}
\]

\[
= [w^T JB - (\tau b)^T]f_{t+1} + w^T \varepsilon_{t+1} + \xi_{t+1}^* - w^T y_t - \log p_t^*,
\]
where the second equality collects terms known at time $t$, i.e., $y_t = B f_t + \epsilon_t$ and $\log p_t^* = -(\tau b)^*_t f_t + \xi_t^*$. Given that the portfolio of instruments matches the generalized durations of the target, $B'\mathcal{T}w = (\tau b)_*$, the variance of the hedge error will be

$$\text{var}_t(r^*_t - w'r_{t+1}) = w'\mathcal{T}\Psi\mathcal{T}w + \Psi_*,$$

(A.1.4)

for $\Psi_* = \text{var}_t(\xi^{*}_t + 1)$. Since the second component is not under the portfolio manager’s control, minimization of the hedge error variance amounts to minimizing the first term in (A.1.4). The portfolio that matches generalized durations of the target for least hedge error variance therefore solves

$$\min_w w'\mathcal{T}\Psi\mathcal{T}w \quad \text{s. t.} \quad B'\mathcal{T}w = (\tau b)_*. \quad (A.1.5)$$

This is equivalent to the problem (A.1.6) in Lemma A below with $A = \mathcal{T}\Psi\mathcal{T}$, $g = 0$, $D = B'\mathcal{T}$, and $c = (\tau b)_*$. The solution is then by (A.1.7)

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

for the factor projection matrix $F = (B'\Psi^{-1}B)^{-1}B'\Psi^{-1}$. Therefore, the weights in (1.9) do indeed minimize hedge error variance as claimed. When the value matching constraint $w'\iota = 1$ is added to the minimization problem (A.1.5) the correction of weights $\tilde{w}$ can be found from (A.1.8) in Lemma A to be

$$w_* = \tilde{w} + (1 - \tilde{w}'\iota)\frac{\Lambda t}{t'\Lambda t},$$

for the matrix

$$\Lambda = \begin{pmatrix} \mathcal{T}\Psi\mathcal{T}^{-1} - (\mathcal{T}\Psi\mathcal{T})^{-1}\mathcal{T}B(B'\Psi^{-1}(\mathcal{T}\Psi\mathcal{T})^{-1}\mathcal{T}B)^{-1}B'\mathcal{T}(\mathcal{T}\Psi\mathcal{T})^{-1} \\ \mathcal{T}^{-1}\Psi^{-1}(\Psi - B(B'\Psi^{-1}B)^{-1}B')\Psi^{-1}\mathcal{T}^{-1} \end{pmatrix}.$$

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

$$\min_w \frac{1}{2}w'Aw - w'g \quad \text{s. t.} \quad Dw = c \quad (A.1.6)$$

can be written as

$$w_c = w_u + A^{-1}D'(DA^{-1}D')^{-1}(c - Dw_u). \quad (A.1.7)$$

When further adding the scaling constraint $w'\iota = 1$ to the problem (A.1.6), the solution is on the form

$$w_* = w_c + \frac{\Lambda t}{t'\Lambda t}(1 - w'_c\iota), \quad (A.1.8)$$

with $\Lambda = A^{-1} - A^{-1}D'(DA^{-1}D')^{-1}DA^{-1}$.
Proof of Lemma A. The unconstrained solution is \( w_u = A^{-1} g \) and the Lagrange function for (A.1.6) is
\[
\mathcal{L} = \frac{1}{2} w'A w - w' g - \lambda'(D w - c).
\]
This has the first order condition \( 0 = A w_c - g - D' \lambda \), such that
\[
w_c = A^{-1}(g + D' \lambda) = w_u + A^{-1} D' \lambda.
\]
Substituting for \( w \) in the constraint gives \( c = D w_u + D A^{-1} D' \lambda \), which when \( \lambda \) is isolated and inserted in \( w_c \) gives the solution
\[
w_c = w_u + A^{-1} D' (D A^{-1} D')^{-1} (c - D w_u).
\]
(A.1.9)

When the constraint \( w' t = 1 \) is added to (A.1.6), the new solution can be found by substituting in the solution (A.1.9) \( D' \) with \( (D', t) \) and \( c' \) with \( (c', 1) \) to get
\[
w_* = w_u + A^{-1}( D' t) \left[ \begin{array}{cc} D A^{-1} D' & D A^{-1} t \\ t' A^{-1} D' & t' A^{-1} t \end{array} \right]^{-1} \left[ \begin{array}{c} c - D w_u \\ 1 - t' w_u \end{array} \right].
\]
(A.1.10)

By the formula for the inverse of a block matrix
\[
S^{-1} = \left[ \begin{array}{cc} D A^{-1} D' & D A^{-1} t \\ t' A^{-1} D' & t' A^{-1} t \end{array} \right]^{-1} = \left[ \begin{array}{cc} (D A^{-1} D')^{-1} + F t'/t' \Lambda t & -F t'/t' \Lambda t \\ -t' F'/t' \Lambda t & 1/t' \Lambda t \end{array} \right]
\]
for \( F = (D A^{-1} D')^{-1} D A^{-1} \). Now using that \( w_c - w_u = F' (c - D w_u) \) we get
\[
S^{-1} \left[ \begin{array}{c} c - D w_u \\ 1 - t' w_u \end{array} \right] = \left[ \begin{array}{c} (D A^{-1} D')^{-1} (c - D w_u) + F t'(w_c - w_u)/t' \Lambda t - F(1 - t' w_u)/t' \Lambda t \\ -t' (w_c - w_u)/t' \Lambda t + (1 - t' w_u)/t' \Lambda t \end{array} \right].
\]
Multiplying by \( (A^{-1} D' \ A^{-1} t) \) from the left gives the last term in (A.1.10), such that
\[
w_* = w_u + F' (c - D w_u) + (A^{-1} D' F - A^{-1}) u' (w_c - w_u)/t' \Lambda t \\
+ (A^{-1} - A^{-1} D' F) t (1 - t' w_u)/t' \Lambda t,
\]
and substituting \( \Lambda = A^{-1} - A^{-1} D' F \) the solution is obtained,
\[
w_* = w_u + (w_c - w_u) \frac{\Lambda t}{t' \Lambda t} (w_c - w_u) + \frac{\Lambda t}{t' \Lambda t} (1 - t' w_u)
\]
\[
= w_c + \frac{\Lambda t}{t' \Lambda t} (1 - t' w_c).
\]
**Proof of (1.18).** Traded bonds have fixed time of maturity, $T = \tau + t$, not fixed time to maturity, hence write $P(t, T) = \exp(-(T - t) y(t, T - t))$ for the bond price, which by Itô’s lemma has

$$
\frac{dP(t, T)}{P(t, T)} = -\tau dy(t, \tau) + \frac{1}{2} \tau^2 \sigma(t, \tau) \sigma(t, \tau)' dt + y(t, \tau) dt + \tau \frac{\partial y}{\partial \tau}(t, \tau) \, dt
$$

$$
= \left[ y(t, \tau) + \tau \frac{\partial y}{\partial \tau}(t, \tau) + \frac{1}{2} \tau^2 \sigma(t, \tau) \sigma(t, \tau)' - \tau \alpha(t, \tau) \right] dt - \tau \sigma(t, \tau) dW_t
$$

$$
\alpha(t, \tau) = \alpha_p(t, \tau) dt + \sigma_p(t, \tau) dW_t.
$$

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

$$
r_t + \sigma_P(t, \tau) \lambda(t) = \alpha_p(t, \tau)
$$

$$
r_t - \tau \sigma(t, \tau) \lambda(t) = y(t, \tau) + \tau \frac{\partial y}{\partial \tau}(t, \tau) + \frac{1}{2} \tau^2 \sigma(t, \tau) \sigma(t, \tau)' - \tau \alpha(t, \tau),
$$

and (1.18) follows. \qed

**Proof of Proposition 2.** This follows from Theorem 5 by setting $\sigma_1 = \sigma_3 = 0$. \qed

**Proof of Proposition 3.** In Theorem 6 set $\sigma_1 = \sigma_3 = f_{n,5} = f_{n,6} = f_{n,7} = 0$ and let $b \mid 0$. \qed

**Proof of Theorem 4.** $B_{1:3}(\tau)$ must be part of any yield curve shape consistent with the SLSC model since the yield curve changes stochastically in each of these directions by different Wiener processes. To see that the additional functions in (1.32) must be included as well, substitute the no-arbitrage HJM drift condition from (1.18) into the yield SDE to get

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

with

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

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

$$
\tilde{\alpha}(\tau) = \frac{\tau}{2} B_{1:3}(\tau) \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2) B_{1:3}(\tau)' + B_{1:3}(\tau) \text{diag}(\sigma_1, \sigma_2, \sigma_3) \lambda
$$

$$
= \sum_{j=1}^{3} \left[ \sigma_j^2 \frac{\tau}{2} B_j(\tau)^2 + \sigma_j B_j(\tau) \lambda_j \right]. \quad (A.1.11)
$$
Besides $B_{1:3}(\tau)$, the model therefore also produces drift in the directions $\tau / 2B_j(\tau)^2$, for $j = 1, 2, 3$, which we now express in terms of $B_{1:3}(\tau)$ and the functions in (1.32).

First calculate that
\[
\frac{\tau}{2} B_2(\tau)^2 = \frac{\tau}{2} \left( \frac{1 - e^{-\alpha \tau}}{\alpha \tau} \right)^2 = \frac{1 - 2e^{-\alpha \tau} + e^{-2\alpha \tau}}{2a^2 \tau} = \frac{2(1 - e^{-\alpha \tau}) - (1 - e^{-2\alpha \tau})}{2a^2 \tau}
\]
\[
= \frac{1}{a^2} \frac{1 - e^{-\alpha \tau}}{\alpha \tau} - \frac{1}{2a^2} \frac{1 - e^{-2\alpha \tau}}{\tau},
\]
and similarly for $B_1$, replacing parameter $a$ in $B_2$ by $b$,
\[
\frac{\tau}{2} B_1(\tau)^2 = \frac{1 - e^{-bt}}{b \tau} - \frac{1 - e^{-2bt}}{2b^2 \tau}.
\]

Finally, for $B_3$ we have
\[
\frac{\tau}{2} B_3(\tau)^2 = \frac{\tau}{2} \left( \frac{1 - e^{-\alpha \tau}}{\alpha \tau} - e^{-\alpha \tau} \right)^2 = \frac{1 - 2(1 + \alpha \tau) e^{-\alpha \tau} + (1 + \alpha \tau)^2 e^{-2\alpha \tau}}{2a^2 \tau}
\]
\[
= \frac{1}{a} \left( \frac{1 - e^{-\alpha \tau}}{\alpha \tau} - e^{-\alpha \tau} \right) - \frac{1}{2a^2} \frac{1 - e^{-2\alpha \tau}}{\tau} + \frac{1}{a} e^{-2\alpha \tau} + \frac{1}{2} \alpha e^{-2\alpha \tau}.
\]

Let $\tilde{B}_{4:7}(\tau)$ be the vector of functions in (1.32), substitute these into (A.1.12)-(A.1.14), and collect terms in parentheses to obtain
\[
\tilde{\alpha}(\tau) = [B_{1:3}(\tau), \tilde{B}_{4:7}(\tau)]D
\]

for $D$ the $7 \times 1$ vector of constants
\[
D = \left( \begin{array}{c} \frac{\sigma_1^2}{b} + \sigma_1 \lambda_1, \frac{\sigma_2^2}{a} + \sigma_2 \lambda_2, \frac{\sigma_3^2}{a} + \sigma_3 \lambda_3, -\frac{\sigma_1^2}{2b^2}, -\frac{\sigma_2^2}{2a^2}, -\frac{\sigma_3^2}{a^2}, \frac{\sigma_4^2}{2} \end{array} \right)'.
\]

Any function of $\tau$ that enters the drift term must also be a part of consistent curves. So to be sure that all functions in $\tilde{B}_{4:7}$ are necessary we must check whether any of them can be cancelled by the remaining term in the drift, i.e.,
\[
\frac{1}{\tau} [y(t, \tau) - y(t, 0)] + \frac{dy}{d\tau}(t, \tau) = \frac{1}{\tau} \left[ \frac{\partial}{\partial x} xg(x) \right]_{x=0} \equiv \mathcal{V}(y(t, \tau)).
\]

The next step is thus to consider the implications on $y(t, \tau)$ for $\mathcal{V}(y(t, \tau))$ to include terms $-\tilde{B}_j(\tau) D_j$, $j = 4, \ldots, 7$. Therefore consider the transformation $\mathcal{U}$,
\[
\mathcal{U}(g(\tau)) = \frac{1}{\tau} \int_0^\tau [xg(\tau)] dx,
\]
which is a one-sided inverse of $\mathcal{V}$ in the sense that $\mathcal{V}(\mathcal{U}(g(\tau))) = g(\tau)$. Conversely, we have that
\[
\mathcal{U}(\mathcal{V}(y(t, \tau))) = y(t, \tau) - y(t, 0),
\]
(A.1.16)
so we can use \( \mathcal{U} \) to derive the functions of \( \tau \) that \( y(t, \tau) \) must include for \( \mathcal{V}(y(t, \tau)) \) to cancel some of the terms \( \tilde{B}_j(\tau) D_j, j = 4, \ldots, 7, \) in drift.

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

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

and similarly for \( \tilde{B}_4 \) by substituting \( b \) for \( a, \)

\[
\mathcal{U}(\tilde{B}_4(\tau)) = 1 - \frac{1}{2b} \tilde{B}_4(\tau).
\]

For \( \tilde{B}_6 \) we have

\[
\mathcal{U}(\tilde{B}_6(\tau)) = \frac{1}{\tau} \int_0^\tau x e^{-2ax} dx = \frac{1}{4a^2} \frac{1 - e^{-2a\tau}}{\tau} - \frac{1}{2a} e^{-2a\tau} = \frac{1}{4a^2} \tilde{B}_5(\tau) - \frac{1}{2a} \tilde{B}_6(\tau),
\]

while for \( \tilde{B}_7 \)

\[
\mathcal{U}(\tilde{B}_7(\tau)) = \frac{1}{\tau} \int_0^\tau x^2 e^{-2ax} dx = \frac{1}{4a^3} \frac{1 - e^{-2a\tau}}{\tau} - \frac{1}{2a^2} e^{-2a\tau} - \frac{1}{2a} \tau e^{-2a\tau} = \frac{1}{4a^3} \tilde{B}_5(\tau) - \frac{1}{2a^2} \tilde{B}_6(\tau) - \frac{1}{2a} \tilde{B}_7(\tau).
\]

If \( \mathcal{V}(y(t, \tau)) \) is to cancel out \( \tilde{B}_{4:7}(\tau) D_{4:7} \) in the drift, we may then by (A.1.16) conclude that \( y(t, \tau) - y(t, 0) \) will include the terms

\[
\mathcal{U}(-\tilde{B}_{4:7}(\tau) D_{4:7}) = -\mathcal{U}(\tilde{B}_{4:7}(\tau)) D_{4:7} = \tilde{B}_{4:7}(\tau) \left( \frac{D_4}{2a}, \frac{D_5}{2a}, \frac{D_6}{4a^2}, \frac{D_7}{4a^3}, \frac{D_8}{2a^2}, \frac{D_9}{2a^3} \right) - D_4 - D_5.
\]

This shows that it is only possible to cancel any of the \( \tilde{B}_{4:7}(\tau) \) functions in the drift if the function to be removed is already in the yield curve. To clarify this, the above result shows for instance that the drift in the \( \tilde{B}_7 \) direction due to \( \tilde{a} \) (\( \tau \)) only can be cancelled by \( \mathcal{V}(y(t, \tau)) \) if \( y(t, \tau) \) includes the function \( \tilde{B}_7(\tau) = \tau e^{-2a\tau} \) exactly with the coefficient \( D_7/2a = \sigma_3^2/4a. \) Thus, even though there is no drift in the \( \tilde{B}_7 \) direction, the yield curve must still include \( \tilde{B}_7 \) to be consistent. This holds for all the four functions, since \( D_j \) always loads on \( B_j, \) and we conclude that all functions \( \tilde{B}_{4:7} \) are necessarily part of any consistent yield curve family.

Next, we prove the converse result that the family \( \mathcal{G} \) of all yield curves spanned by functions \( (B_{1:3}, \tilde{B}_{4:7}) \) is consistent with the SLSC interest rate model. Since \( B_{4:7} \) in (1.33) is a linear transformation of \( \tilde{B}_{4:7} \), the same family is defined by \( B(\tau) = [B_{1:3}(\tau), B_{4:7}(\tau)] \), the SLSC curve shape, and we choose to work with this rotation. By the definition of consistency, the family \( \mathcal{G} \) is consistent with the SLSC dynamic model if starting from a curve in this family all subsequent curves produced by
the interest rate model are also in \( \mathcal{J} \). This will hold if both the drift and volatility functions of the SLSC model are spanned by \( B(\tau) \) for all \( y(t, \tau) \) in \( \mathcal{J} \). For volatility, \( \sigma(\tau) = B_{1:3}(\tau) \text{diag}(\sigma_1, \sigma_2, \sigma_3) \), this is clearly the case, and that it also holds for \( \tilde{\alpha}(\tau) \) may be seen from (A.1.15). To explicitly specify \( \tilde{\alpha}(\tau) \) in terms of the new basis, we observe from (A.1.12)-(A.1.14) that

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

such that we can write

\[
\tilde{\alpha}(\tau) = \left( \frac{\sigma_1^2}{b} + \sigma_1 \lambda_1 \right) B_1(\tau) + \left( \frac{\sigma_2^2}{a} + \sigma_2 \lambda_2 \right) B_2(\tau) + \left( \frac{\sigma_3^2}{a} + \sigma_3 \lambda_3 \right) B_3(\tau) - \frac{\sigma_1^2}{a} B_4(\tau) - \frac{\sigma_2^1}{b} B_5(\tau) - \frac{\sigma_3^2}{a} B_6(\tau) + \frac{\sigma_3^2}{2a} B_7(\tau) = B(\tau) C
\]

(A.1.17)

for the \( 7 \times 1 \) vector of constants

\[
C = \begin{pmatrix} \frac{\sigma_1^2}{b} + \sigma_1 \lambda_1, & \frac{\sigma_2^2}{a} + \sigma_2 \lambda_2, & \frac{\sigma_3^2}{a} + \sigma_3 \lambda_3, & -\frac{\sigma_1^2}{a}, & -\frac{\sigma_2^1}{b}, & -\frac{\sigma_3^2}{a}, & \frac{\sigma_3^2}{2a} \end{pmatrix}^T.
\]

(A.1.18)

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

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

\[
\mathbb{V}\{B_2(\tau)\} = \frac{1}{\tau} \left[ \frac{\partial}{\partial x} x B_2(x) \right]_{x=0}^\tau = \frac{1}{\tau} \left[ \frac{\partial}{\partial x} \frac{1 - e^{-ax}}{a} \right]_{x=0}^\tau
\]

\[
= \frac{1}{\tau} \left[ e^{-ax} \right]_{x=0}^\tau = \frac{e^{-\alpha \tau} - 1}{\tau} = -a B_2(\tau),
\]

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

\[
\mathbb{V}\{B_1(\tau)\} = -b B_1(\tau), \quad \mathbb{V}\{B_4(\tau)\} = -2a B_4(\tau), \quad \mathbb{V}\{B_5(\tau)\} = -2b B_5(\tau).
\]

For \( B_3 \) we get

\[
\mathbb{V}\{B_3(\tau)\} = \mathbb{V}\{B_2(\tau)\} - \frac{1}{\tau} \left[ \frac{\partial}{\partial x} x e^{-ax} \right]_{x=0}^\tau = \frac{e^{-\alpha \tau} - 1}{\tau} - \frac{1}{\tau} \left[ e^{-ax} - a x e^{-ax} \right]_{x=0}^\tau
\]

\[
= a e^{-\alpha \tau} = -a \left[ \frac{1 - e^{-\alpha \tau}}{\alpha \tau} - e^{-\alpha \tau} \right] + \frac{1 - e^{-\alpha \tau}}{\alpha \tau} = -a B_3(\tau) + a B_2(\tau),
\]
and similarly for $B_6$ with $a$ replaced by $2a$

$$\nabla(B_6(\tau)) = -2aB_6(\tau) + 2aB_4(\tau).$$

Finally, for $B_7$

$$\nabla(B_7(\tau)) = -2aB_6(\tau) + 2aB_4(\tau) - 2aB_7(\tau).$$

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

$$\nabla(y(t, \tau)) = \nabla(B(\tau)) \phi = -B(\tau) h\phi$$

for the transformation matrix

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

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

**Proof of Corollary 1.** From Theorem 4 all functions in $B(\tau)$ are necessary in any consistent curve shape. The family that they define can therefore only be reduced by either restricting some coefficients to fixed values or by combining some of the $B_i(\tau)$ functions such that they only are included in fixed combinations in the basis, e.g. $x_iB_i(\tau) + x_jB_j(\tau)$. The first of these options is ruled out by the statement in the corollary that the coefficients $\phi$ must be general, we thus need to check whether the second option is possible.

The different Wiener processes driving the changes in each of the $B_{1:3}(\tau)$ directions imply that these functions must enter the basis separately. With respect to the other four functions we must check whether any of them can be combined to define a smaller family $\mathcal{G}^*$ for which it still holds that $\nabla(y(t, \tau)) + \tilde{\alpha}(\tau) \in \mathcal{G}^*$ for all $y(t, \tau) \in \mathcal{G}^*$. 

Since this must hold for all \( y(t, \tau) \) in a family on the form \( \phi B(\tau) \), it must in particular hold for \( \phi = 0 \), and thus for \( \tilde{\alpha}(\tau) \), i.e., \( \tilde{\alpha}(\tau) \in \mathbb{S}^* \). This implies that any candidate combination of the \( B_{4,7} \) functions must be proportional to the coefficients in \( \tilde{\alpha} \) given by \( C_{4,7} \), i.e., it must hold that \( x_i/x_j = C_i/C_j \). Let \( h_{ij} \) be the 2 \times 2 submatrix of rows and columns \( i \) and \( j \) in \( h \) corresponding to the two functions \([B_i, B_j] = B_{ij}\) that we consider combining in proportions \( (x_i, x_j)' = x_{ij} \). Then \( \mathcal{V}(B_{ij}(\tau)x_{ij}) = -B_{ij}(\tau)h_{ij}x_{ij} \) is spanned by the reduced family only if there exist a constant \( \beta \) for which this equals \(-B_{ij}(\tau)x_{ij}\beta\). Therefore, it must hold that \( h_{ij}x_{ij} = x_{ij}\beta \), and by proportionality to coefficients in \( \tilde{\alpha} \) also that \( h_{ij}C_i = C_i\beta \). The resulting necessary condition to combine any two \( B_i \) and \( B_j \) is thus that \( h_{ij} \) must have an eigenvector with entries in the same proportions as these functions enter \( \tilde{\alpha} \) with. This is not the case for any combination of the \( B_{4,7} \) functions as can be seen from \( C \) and \( h \) in (A.1.18) and (A.1.20). Specifically, for \( B_5 \) in combination with any of the other functions, \( h_{ij} \) is diagonal with different diagonal values, so eigenvectors are \((s, 0)\) or \((0, u)\). Any combination of the other three \( B_i \)'s has \( h_{ij} \) upper triangular with identical diagonal element \( 2a \) and thus only the repeated eigenvector \((s, 0)\). We conclude that none of the \( B_{4,7} \) functions can be combined to reduce the basis and that the family is the minimal one with unrestricted coefficients on all \( \tau \) functions entering the yield curve.

**Proof of Theorem 6.** When the initial curve is on the form \( y(t_0, \tau) = B(\tau)f_0 \) and interest rates follow the SLSC dynamic model, then by Theorem 4 all subsequent curves are spanned by \( B(\tau) \), so we may define coefficients or factors at each time \( t, f(t) \), for which \( y(t, \tau) = B(\tau)f(t) \). Then from (A.1.17) and (A.1.19) it follows that for all \( t \geq t_0 \)

\[
\begin{align*}
dy(t, \tau) &= \mathcal{V} \left( y(t, \tau) \right) + \tilde{\alpha}(\tau) \, dt + \sigma(\tau) \, dW_t \\
&= \left[ -B(\tau)h f(t) + B(\tau)C \right] dt + B(\tau) \Sigma dW_t \\
&= B(\tau) \left\{ \left[ C - h f(t) \right] dt + \Sigma dW_t \right\},
\end{align*}
\]

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

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

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

\[
\begin{align*}
df(t) &= h \left[ \bar{f} - f(t) \right] dt + \Sigma dW_t, \\
&= \begin{bmatrix} a_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a_7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_9 \end{bmatrix} \mathbf{f}_0 \end{align*}
\]

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

\[
\bar{f} = \begin{bmatrix} \sigma_1 b + \lambda_1 & \sigma_2 a + \lambda_2 & \sigma_3 a + \lambda_3 & -\sigma_4 a^2 - \sigma_5 a^2 - \sigma_6 a^2 - \sigma_7 a^2 - \sigma_8 a^2 - \sigma_9 a^2 \end{bmatrix}.
\]
To find $y(t_{n+1}, \tau)$ in terms of the factors at time $t_n$, we solve the factor SDE. First, observe that for the ansatz $e^{ht}$ we get

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

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

$$f_{n+1} = e^{-h\Delta t} f_n + \int_0^{\Delta t} e^{-hu} du \cdot h f_n + \int_0^{\Delta t} e^{-hu} \Sigma dW_{t_n+u}.$$  \tag{A.1.22}

Define the function $H(u) = e^{-hu}$ which has $H'(u) = -H(u) h$. Then $\int H(u) d u = -H(u) h^{-1}$, and in particular $\int_0^{\Delta u} e^{-hu} du \cdot h = (I - H(\Delta u))$. To calculate $H(u)$, let the diagonal of $h$ be $d(h)$ and set $\tilde{h} = h - d(h)$, the matrix with the off-diagonal elements of $h$. Then from (A.1.20) $\tilde{h}$ is nilpotent of degree three and the only element of $\tilde{h}^2$ different from zero is $[\tilde{h}^2]_{4,7} = -4a^2$. Thus by rules of matrix exponents

$$H(u) = e^{-hu} = e^{-d(h)u} e^{-\tilde{h}u} = e^{-d(h)u} \left( I - \tilde{h}u + \tilde{h}^2 u^2 / 2 \right)$$

For the stochastic integral in (A.1.22) we can use that for a deterministic function $g(\cdot)$

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

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

$$\Omega_{1:3} (\Delta u) = \int_0^{\Delta u} H_{1:3}(u) \Sigma \left( \sigma_1^2, \sigma_2^2, \sigma_3^2 \right) H_{1:3}(u) \, du$$

while the remaining entries of $\Omega$ are zero. Substituting the results into (A.1.22), we conclude that

$$f_{n+1} = H(\Delta u) f_n + \left( I - H(\Delta u) \right) \tilde{f} + \nu_{n+1},$$  \tag{A.1.23}
such that from an initial curve \( y(t_n, \tau) = B(\tau) f_n \) the curve \( \Delta_n \) later in time is \( y(t_{n+1}, \tau) = B(\tau) f_{n+1} \). Further, since \( v_{n+1} = \int_{t_n}^{t_{n+1}} H(u) \Sigma dW_u \), the sequence of error terms \( v_1, v_2, \ldots \) are based on non-overlapping intervals of the Wiener processes \( W_t \) and they are thus serially independent.

**Proof of Theorem 5.** Consider the special case with \( y(t_n, \tau) = 0 = B(\tau) 0 \). Since this is still spanned by \( B(\tau) \), we have by Theorem 6 that the curve at \( t_{n+1} \) is \( y(t_{n+1}, \tau) = B(\tau) \tilde{f}_{n+1} \), with factor dynamics obtained by substituting \( \tilde{f}_n = 0 \) in (A.1.23), i.e.,

\[
\tilde{f}_{n+1} = \left(I - H(\Delta_n)\right) \tilde{f} + v_{n+1}.
\]

From (A.1.25) in Proposition 8,

\[
y(t_{n+1}, \tau) = \int_0^{\Delta_n} \left\{ \frac{1}{\tau} \left[ (x + u) \tilde{\alpha}(x + u) \right]_{x=0}^{\tau} \right\} du + \int_0^{\Delta_n} \left\{ \frac{1}{\tau} \left[ (x + u) \sigma(x + u) \right]_{x=0}^{\tau} \right\} dW_{t_{n+1}}.
\]

In the general solution to the yield SDE given by (A.1.25) the last two terms are not affected by changing the shape of the previous yield curve. Therefore for an arbitrary \( y(t_n, \tau) \) we must have that the curve at \( t_{n+1} \) can be written as

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

with \( \tilde{f}_{n+1} \) given by (A.1.24).

**Proposition 8 (Solution to yield SDE).** Suppose the yield curve at time \( t_n \) has general shape, \( y(t_n, \tau) \), that yield dynamics for \( t \geq t_n \) are given by the SDE

\[ dy(t, \tau) = a(t, \tau) \, dt + \sigma(t) \, dW_t, \]

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

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

with

\[ \tilde{\alpha}(\tau) = \frac{\tau}{2} \sigma(\tau) \sigma(\tau)' + \sigma(\tau) \lambda. \]
**Proof of Proposition 8.** By the HJM drift condition for yields under the objective probability measure in Brace and Musiela parametrization, (1.18), the drift term satisfies

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

such that the yield curve SDE is

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

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

\[ dr(t, \tau) = \frac{\partial r}{\partial \tau}(t, \tau) dt + \left\{ \frac{\partial}{\partial \tau} \left[ (\tau + t_n + \Delta_n - \tau) \tilde{\alpha}(\tau + t_n + \Delta_n - \tau) \right] \right\} dW_t. \quad (A.1.26) \]

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

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

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

\[
\begin{align*}
   r(t_{n+1}, \tau) &= r(t_n, \tau + t_{n+1} - t_n) + \int_{t_n}^{t_{n+1}} \left\{ \frac{\partial}{\partial \tau} \left[ (\tau + t_{n+1} - \tau) \tilde{\alpha}(\tau + t_{n+1} - \tau) \right] \right\} d\tau \\
   &\quad + \int_{t_n}^{t_{n+1}} \left\{ \frac{\partial}{\partial \tau} \left[ (\tau + t_{n+1} - \tau) \sigma(\tau + t_{n+1} - \tau) \right] \right\} dW_t \\
   &= r(t_n, \tau + \Delta_n) + \int_0^{\Delta_n} \left\{ \frac{\partial}{\partial \tau} \left[ (\tau + u) \tilde{\alpha}(\tau + u) \right] \right\} du \\
   &\quad + \int_0^{\Delta_n} \left\{ \frac{\partial}{\partial \tau} \left[ (\tau + u) \sigma(\tau + u) \right] \right\} dW_{t_{n+1}-u}.
\end{align*}
\]

To get back to yields integrate over maturity and divide by \( \tau \),

\[
\begin{align*}
y(t_n + \Delta_n, \tau) &= \frac{1}{\tau} \int_0^\tau r(t_n, x + \Delta_n) dx + \frac{1}{\tau} \int_0^\tau \int_0^{\Delta_n} \left\{ \frac{\partial}{\partial x} \left[ (\tau + u) \tilde{\alpha}(\tau + u) \right] \right\} du dx \\
   &\quad + \frac{1}{\tau} \int_0^\tau \int_0^{\Delta_n} \left\{ \frac{\partial}{\partial x} \left[ (\tau + u) \sigma(\tau + u) \right] \right\} dW_{t_{n+1}-u} dx.
\end{align*}
\]

The average of forward rates with maturity from \( \Delta_n \) to \( \tau + \Delta_n \) at time \( t_n \) is in terms of
yields,
\[ \frac{1}{\tau} \int_0^\tau r(t_n, x + \Delta_n) \, dx = \frac{1}{\tau} \int_0^{\tau + \Delta_n} r(t_n, x) \, dx \]
\[ = \frac{1}{\tau} \left[ \int_0^{\tau + \Delta_n} r(t_n, x) \, dx - \int_0^{\Delta_n} r(t_n, x) \, dx \right] \]
\[ = \frac{1}{\tau} \left[ (\tau + \Delta_n) y(t_n, \tau + \Delta_n) - \Delta_n y(t_n, \Delta_n) \right], \]
and by changing the order of integration we get that the solution to the yield SDE is
\[ y(t_n + \Delta, \tau) = \frac{1}{\tau} \left[ (\tau + \Delta) y(t_n, \tau + \Delta) - \Delta y(t_n, \Delta) \right] \]
\[ + \int_0^{\Delta_n} \left\{ \frac{1}{\tau} \left[ (x + u) \tilde{\alpha}(x + u) \right]_{x=0}^T \right\} \, du \]
\[ + \int_0^{\Delta_n} \left\{ \frac{1}{\tau} \left[ (x + u) \sigma(x + u) \right]_{x=0}^T \right\} \, dW_{t_n+1-u} \]
as stated in the proposition.

**Proof of Corollary 2.** When \( f_{4:7}(t_0) = \overline{f}_{4:7} \) we have by (1.36) in Theorem 6 that the factors \( f_{4:7} \) are constant at this value. Therefore by (1.35) we can write all subsequent curves for \( t \geq t_0 \) as
\[ y(t, \tau) = B_{1:3}(\tau) f_{1:3}(t) + B_{4:7}(\tau) \overline{f}_{4:7}, \]
and the diffusion for \( f_{1:3} \) is obtained from the first three rows of \( f \) in (A.1.21) to be
\[ df_{1:3}(t) = \kappa \left[ \overline{f}_{1:3} - f_{1:3}(t) \right] \, dt + \text{diag}(\sigma_1, \sigma_2, \sigma_3) \, dW_t, \]
with
\[ \kappa = h_{1:3} = \begin{pmatrix} b & 0 & 0 \\ 0 & a & -a \\ 0 & 0 & a \end{pmatrix}. \]
We would like to represent the model by factors that have zero mean under \( Q \), so we choose to rotate to the new factors \( X_t \equiv f_{1:3}(t) - \overline{f}_{1:3}^Q \), where \( \overline{f}_{1:3}^Q \) is \( \overline{f}_{1:3} \) for \( \lambda = 0 \). Then
\[ dX_t = \kappa \left( \overline{X} - X_t \right) \, dt + \text{diag}(\sigma_1, \sigma_2, \sigma_3) \, dW_t, \]
with
\[ \overline{X} = \overline{f}_{1:3}^Q - \overline{f}_{1:3} = \begin{bmatrix} \sigma_1/a \lambda_1 \lambda_2 + \sigma_3/a \lambda_3, \sigma_3/a \lambda_3 \end{bmatrix}, \]
and the yield curve in terms of \( X_t \) is
\[ y(t, \tau) = B(\tau) \left[ \overline{f}_{1:3}^Q \overline{f}_{4:7} \right] + B_{1:3}(\tau) X_t \equiv -A(\tau) + B_{1:3}(\tau) X_t. \]
Here the function $A(\tau)$ in terms of basic parameters is

$$A(\tau) = -B(\tau)\left(\frac{\sigma_1^2}{b^2} - \frac{\sigma_2^2}{2a^2} - \frac{\sigma_3^2}{4a^2} - \frac{\sigma_4^2}{2a^2} - \frac{\sigma_5^2}{2b^2} - \frac{3\sigma_6^2}{4a^2} \right)^t$$

$$\quad = -\frac{\sigma_1^2}{b^2} 1 - e^{-b\tau} - \left(\frac{\sigma_2^2}{a^2} + \frac{\sigma_3^2}{2a^2} + \frac{\sigma_4^2}{4a^2}\right) 1 - e^{-a\tau} + \frac{\sigma_3^2}{a^2} e^{-a\tau}$$

$$\quad + \left(\frac{\sigma_2^2}{2a^2} + \frac{3\sigma_3^2}{4a^2}\right) 1 - e^{-2a\tau} 1 - e^{-2b\tau} 1 - e^{-2b\tau} - \frac{3\sigma_3^2}{4a^2} e^{-2a\tau} - \frac{\sigma_2^2}{4a^2} e^{-2a\tau}$$

$$\quad = \sigma_1^2\left(-\frac{1}{b^2} 1 - e^{-b\tau} + \frac{1}{a^2} 1 - e^{-a\tau}\right) + \sigma_2^2\left(-\frac{1}{a^2} 1 - e^{-a\tau} + \frac{1}{a^2} 1 - e^{-a\tau}\right)$$

$$\quad + \sigma_3^2\left(\frac{1}{a^2} e^{-a\tau} - \frac{1}{4a^2} \tau e^{-2a\tau} - \frac{3}{4a^2} e^{-2a\tau} - \frac{2}{a^2} e^{-2a\tau} + \frac{5}{8a^2} e^{-2a\tau}\right).$$

The short rate as a function of $X_t$ is obtained by letting $\tau \downarrow 0$ in (A.1.27). Since $\lim_{\tau \downarrow 0} B(\tau) = (1,1,0,1,1,0,0)$ we get that

$$r_t = \lim_{\tau \downarrow 0} \eta(t,\tau) = f_0 + \tilde{f}_1 + \tilde{f}_2 + \tilde{f}_3 + \tilde{f}_4 + X_{1t} + X_{2t}$$

$$\quad = \sigma_1^2\left(\frac{1}{b^2} + \frac{\sigma_2^2}{2a^2} + \frac{\sigma_3^2}{2a^2}\right) + X_{1t} + X_{2t} = \delta_0 + \delta_X X_t,$$

such that the values of $\delta_0$ and $\delta_X$ given in the corollary are obtained. \hfill \Box

**Proof of Theorem 7.** The weights that solve (1.42) are also the solution to

$$\min_w \frac{1}{2} w' \mathbb{E}_t \left[r_{t+1} r_{t+1}'\right] w - w' \mathbb{E}_t \left[r_{t+1} r_{t+1}^*\right]. \quad (A.1.27)$$

This is also known as the linear projection of $r_{t+1}'$ on $r_{t+1}$, and the solution is given by

$$\tilde{w} = \mathbb{E}_t \left[r_{t+1} r_{t+1}'\right]^{-1} \mathbb{E}_t \left[r_{t+1} r_{t+1}^*\right]. \quad (A.1.28)$$

The conditional expected value of the instrument return vector (1.5) is

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

when $\mathbb{E}_t \left[\Delta \epsilon_{t+1}\right] = 0$. Also rewriting (1.5) as

$$r_{t+1} = -\mathcal{T} \left(B f_{t+1} + \epsilon_{t+1} - y_t\right), \quad (A.1.29)$$

the conditional variance is given by

$$\text{var}_t \left[r_{t+1}\right] = \mathcal{T} \left(B \Sigma_{t+1} B' + \Psi\right) \mathcal{T}.$$
Then
\[
\mathbb{E}_t \left[ r_{t+1|r_{t+1}} \right] = \var_t \left[ r_{t+1} \right] + \mathbb{E}_t \left[ r_{t+1} \right] \mathbb{E}_t \left[ r_{t+1}' \right] = \mathcal{J} \left( \mathbf{B} \mathbf{\Sigma}_{t+1|t} \mathbf{B}' + \Psi \right) + \mathcal{J} \mathbf{B} \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right) \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right)' \mathbf{B}' \mathcal{J} = \mathbf{J} \left( \mathbf{B} \left( \mathbf{\Sigma}_{t+1|t} + \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right) \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right)' \right) \mathbf{B}' + \Psi \right) \mathcal{J}.
\]

By (A.1.2) and for \( \mathbb{E}_t \left[ \Delta \xi_{t+1}^* \right] = 0 \), the conditional expected target return is
\[
\mathbb{E}_t \left[ r^*_{t+1} \right] = - (\mathbf{\tau} \mathbf{b})^*_* \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right).
\]

We can rewrite (A.1.2) as \( r^*_{t+1} = - (\mathbf{\tau} \mathbf{b})^*_* f_{t+1} + \xi_{t+1}^* - \log p_t^* \), and then using (A.1.29) the conditional covariance is
\[
\text{cov}_t \left[ r_{t+1}, r^*_{t+1} \right] = \mathcal{J} \mathbf{B} \mathbf{\Sigma}_{t+1|t} \left( \mathbf{\tau} \mathbf{b} \right)_*.
\]

This implies that
\[
\mathbb{E}_t \left[ r_{t+1|r_{t+1}} \right] = \text{cov}_t \left[ r_{t+1}, r^*_{t+1} \right] + \mathbb{E}_t \left[ r_{t+1} \right] \mathbb{E}_t \left[ r^*_{t+1} \right] = \mathcal{J} \mathbf{B} \mathbf{\Sigma}_{t+1|t} \left( \mathbf{\tau} \mathbf{b} \right)_* + \mathcal{J} \mathbf{B} \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right) \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right)' \left( \mathbf{\tau} \mathbf{b} \right)_*,
\]
and the weights (A.1.28) are therefore
\[
\tilde{\mathbf{w}} = \mathcal{J}^{-1} \left[ \mathbf{B} \left( \mathbf{\Sigma}_{t+1|t} + \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right) \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right)' \right) \mathbf{B}' + \Psi \right]^{-1} \times \mathbf{B} \left( \mathbf{\Sigma}_{t+1|t} + \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right) \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right)' \right) \left( \mathbf{\tau} \mathbf{b} \right)_* = \mathcal{J}^{-1} \mathbf{\Psi}^{-1} \mathbf{B} \left( \mathbf{\Sigma}_{t+1|t} + \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right) \left( \mathbf{\mu}_{t+1|t} - \mathbf{\mu}_t \mathbf{1} \right)' \right)^{-1} \mathcal{J}^{-1} \mathbf{\Psi}^{-1} \mathbf{B} \left( \mathbf{\tau} \mathbf{b} \right)_*,
\]
where the second equality applies the matrix inversion lemma.

Imposing that weights sum to one can be done by Lemma A. The original problem (A.1.27) is unconstrained, and in the notation of (A.1.6) it sets \( \mathbf{A} = \mathbb{E}_t \left[ r_{t+1|r_{t+1}} \right] \), \( \mathbf{g} = \mathbb{E}_t \left[ r_{t+1\mathbf{r}_{t+1}} \right] \), and \( \mathbf{D} = 0 \). The scaled weights are therefore by (A.1.8) given by
\[
\mathbf{w}^* = \tilde{\mathbf{w}} + \left( 1 - \tilde{\mathbf{w}}' \right) \frac{\mathbf{A} \mathbf{1}}{\tilde{\mathbf{w}}' \mathbf{A} \mathbf{1}}
\]
with \( \mathbf{A} = \mathbf{A}^{-1} = \mathbb{E}_t \left[ r_{t+1\mathbf{r}_{t+1}} \right] \). \( \square \)
Kalman Filter  The models that combine a dynamic interest rate model with a consistent curve shape can be written on the state space form

\[
\begin{align*}
y_{tn} & = \begin{bmatrix} c \\ m \times 1 \end{bmatrix} + \begin{bmatrix} B \\ m \times k \times 1 \end{bmatrix} f_n + \begin{bmatrix} \varepsilon_n \\ m \times 1 \end{bmatrix}, \\
f_n & = \begin{bmatrix} \Phi_0 + \Phi_1 f_{n-1} + \nu_n \\ k \times 1 \end{bmatrix}, \quad \nu_n \sim N(0, \Omega).
\end{align*}
\]

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

\[
\begin{align*}
\mu_{n|n} & = \mu_{n|n-1} + \Sigma_{n|n-1} B' \Gamma_n^{-1} \zeta_n, \\
\Sigma_{n|n} & = \Sigma_{n|n-1} - \Sigma_{n|n-1} B' \Gamma_n^{-1} B\Sigma_{n|n-1},
\end{align*}
\]

and the prediction step is

\[
\begin{align*}
\mu_{n+1|n} & = \Phi_0 + \Phi_1 \mu_{n|n} = \Phi_0 + \Phi_1 \mu_{n|n-1} + K_n \zeta_n, \\
\Sigma_{n+1|n} & = \Phi_1 \Sigma_{n|n} \Phi_1' + \Omega = \Phi_1 \Sigma_{n|n-1} \Phi_1 + \Omega - K_n \Gamma_n K_n'. \tag{A.1.30}
\end{align*}
\]

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

\[
\log p(y_{tn}|Y_{n-1}) = -\frac{m}{2} \log(2\pi) - \frac{1}{2} \log|\Gamma_n| - \frac{1}{2} \zeta_n' \Gamma_n^{-1} \zeta_n,
\]

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

\[
\log L = \sum_{n=1}^{N} \log p(y_{tn}|Y_{n-1}) = -\frac{mN}{2} \log(2\pi) - \frac{1}{2} \sum_{n=1}^{N} \left( \log|\Gamma_n| + \zeta_n' \Gamma_n^{-1} \zeta_n \right). \tag{A.1.32}
\]

To calculate the log-likelihood function we only need the series of innovations \( \zeta_n \) and \( \Gamma_n \) and these are calculated from \( \mu_{n|n-1} \) and \( \Sigma_{n|n-1} \) alone. The low storage algorithm only stores the series of \( \zeta_n, \Gamma_n, \text{ and } K_n \) and calculates \( \mu_{n+1|n} \) and \( \Sigma_{n+1|n} \) by (A.1.30) and (A.1.31) with \( \mu_{n|n-1} \) and \( \Sigma_{n|n-1} \) stored only from last period. Since this method minimizes the need for storage and avoids calculating the update step, it speeds up the filter which must be calculated many times in the optimization of (A.1.32) over parameters.
Square-Root Kalman Filter  When searching for the parameters that maximizes (A.1.32), the matrix $\Sigma_{n+1|n}$ may fail to be positive semi-definite. This problem can be solved by using a square-root Kalman filter which runs the filter for $S_{n+1|n}$ with $\Sigma_{n+1|n} = S_{n+1|n} S'_{n+1|n}$ instead, as done by Carraro (1988) for the original Kalman filter. Here, we use the square-root version of the low storage algorithm. Thus we need to write the prediction step (A.1.31) in terms of $S_{n+1|n}$. First, write

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

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

$$
\Sigma_{n+1|n} = \left[ (\Phi_1 - K_n B) S_{n|n-1}, K_n N, M \right] \left[ \begin{array}{c} S'_{n|n-1} (\Phi_1 - K_n B)' \\ N K_n' \\ M \end{array} \right] \equiv \tilde{S}_{n+1|n} \tilde{S}'_{n+1|n},
$$

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

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

and then

$$
\Sigma_{n+1|n} = \tilde{S}_{n+1|n} \tilde{S}'_{n+1|n} = R' Q R = R' R.
$$

Therefore set $S_{n+1|n} = R'$ which is a lower triangular square matrix. Instead of $\Sigma_{n+1|n}$ the filter uses $S_{n+1|n}$ and then $\Sigma_{n+1|n}$ is positive semi-definite by construction.
### A.2 Tables

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

<table>
<thead>
<tr>
<th></th>
<th>3 mns.</th>
<th>6 mns.</th>
<th>12 mns.</th>
<th>2 yrs.</th>
<th>3 yrs.</th>
<th>5 yrs.</th>
<th>7 yrs.</th>
<th>10 yrs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (%)</td>
<td>4.12</td>
<td>4.29</td>
<td>4.46</td>
<td>4.84</td>
<td>5.07</td>
<td>5.46</td>
<td>5.76</td>
<td>5.98</td>
</tr>
<tr>
<td>Std. Dev. (%)</td>
<td>2.89</td>
<td>2.96</td>
<td>3.02</td>
<td>3.09</td>
<td>3.04</td>
<td>2.89</td>
<td>2.78</td>
<td>2.64</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>$B_1$</th>
<th>$B_2$</th>
<th>$B_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 mns.</td>
<td>2.828</td>
<td>-0.548</td>
<td>-0.138</td>
</tr>
<tr>
<td>6 mns.</td>
<td>2.917</td>
<td>-0.503</td>
<td>-0.067</td>
</tr>
<tr>
<td>12 mns.</td>
<td>2.991</td>
<td>-0.375</td>
<td>0.026</td>
</tr>
<tr>
<td>2 yrs.</td>
<td>3.081</td>
<td>-0.117</td>
<td>0.095</td>
</tr>
<tr>
<td>3 yrs.</td>
<td>3.034</td>
<td>0.052</td>
<td>0.087</td>
</tr>
<tr>
<td>5 yrs.</td>
<td>2.877</td>
<td>0.309</td>
<td>0.005</td>
</tr>
<tr>
<td>7 yrs.</td>
<td>2.748</td>
<td>0.443</td>
<td>-0.059</td>
</tr>
<tr>
<td>10 yrs.</td>
<td>2.577</td>
<td>0.545</td>
<td>-0.118</td>
</tr>
</tbody>
</table>
Table 3: The target is a 5-year coupon bond; the table shows statistics for different methods used to construct the hedging portfolio. The columns report the average hedging error, or bias, the standard deviation of hedging errors, the root mean squared error, and the mean absolute error. Results are in basis points (0.01%) per month. The first line is for the unhedged target return series. If a given model provides a statistically significant improvement over the duration matching approach (indicated by S) and/or is included in the MCS on a 5% significance level is reported in parentheses.

<table>
<thead>
<tr>
<th>Model</th>
<th>Bias</th>
<th>Std. dev.</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Target movement</td>
<td>53.15</td>
<td>128.41</td>
<td>138.80</td>
<td>110.09</td>
</tr>
<tr>
<td>2 Duration matching</td>
<td>-6.83</td>
<td>7.57</td>
<td>10.19</td>
<td>8.09</td>
</tr>
<tr>
<td>3 Unrestricted 3-factor</td>
<td>1.64</td>
<td>11.39</td>
<td>11.49</td>
<td>8.86</td>
</tr>
<tr>
<td>4 Unrestricted 3-factor Rolling 4-year</td>
<td>2.19</td>
<td>11.58</td>
<td>11.76</td>
<td>9.06</td>
</tr>
<tr>
<td>5 Nelson-Siegel Full period</td>
<td>1.48</td>
<td>11.64</td>
<td>11.72</td>
<td>8.94</td>
</tr>
<tr>
<td>6 Nelson-Siegel Rolling 4-year</td>
<td>1.64</td>
<td>10.79</td>
<td>10.90</td>
<td>8.56</td>
</tr>
<tr>
<td>7 Nelson-Siegel a = 0.731</td>
<td>1.70</td>
<td>12.51</td>
<td>12.61</td>
<td>9.23</td>
</tr>
<tr>
<td>8 Unrestricted 4-factor Full period</td>
<td>2.58</td>
<td>10.38</td>
<td>10.68</td>
<td>8.27</td>
</tr>
<tr>
<td>9 Unrestricted 4-factor Rolling 4-year</td>
<td>2.29</td>
<td>10.22</td>
<td>10.46</td>
<td>7.89</td>
</tr>
<tr>
<td>10 Augmented NS Full period</td>
<td>2.61</td>
<td>9.82</td>
<td>10.14</td>
<td>7.78</td>
</tr>
<tr>
<td>11 Augmented NS Rolling 4-year</td>
<td>2.11</td>
<td>9.85</td>
<td>10.05</td>
<td>7.00</td>
</tr>
<tr>
<td>12 Hull-White Full period</td>
<td>1.96</td>
<td>8.28</td>
<td>8.49</td>
<td>6.63</td>
</tr>
<tr>
<td>13 Hull-White Rolling 4-year</td>
<td>2.02</td>
<td>8.79</td>
<td>9.00</td>
<td>6.81</td>
</tr>
<tr>
<td>14 HW and ANS Full period</td>
<td>2.03</td>
<td>9.18</td>
<td>9.39</td>
<td>7.13</td>
</tr>
<tr>
<td>15 HW and ANS Rolling 4-year</td>
<td>2.42</td>
<td>9.02</td>
<td>9.33</td>
<td>7.24</td>
</tr>
<tr>
<td>16 SLSC curve Full period</td>
<td>2.17</td>
<td>5.87</td>
<td>6.25</td>
<td>4.84</td>
</tr>
<tr>
<td>17 SLSC curve Rolling 4-year</td>
<td>2.13</td>
<td>5.76</td>
<td>6.13</td>
<td>4.74</td>
</tr>
<tr>
<td>18 SLSC dynamic Full period</td>
<td>2.10</td>
<td>5.73</td>
<td>6.10</td>
<td>4.70</td>
</tr>
<tr>
<td>19 SLSC dynamic Rolling 4-year</td>
<td>2.16</td>
<td>5.79</td>
<td>6.17</td>
<td>4.76</td>
</tr>
<tr>
<td>20 SLSC combined Full period</td>
<td>2.11</td>
<td>5.72</td>
<td>6.08</td>
<td>4.69</td>
</tr>
<tr>
<td>21 SLSC combined Rolling 4-year</td>
<td>1.98</td>
<td>6.05</td>
<td>6.36</td>
<td>4.92</td>
</tr>
<tr>
<td>22 SLSC restricted Full period</td>
<td>1.50</td>
<td>10.89</td>
<td>10.97</td>
<td>8.38</td>
</tr>
<tr>
<td>23 SLSC restricted Rolling 4-year</td>
<td>1.75</td>
<td>10.57</td>
<td>10.69</td>
<td>8.22</td>
</tr>
</tbody>
</table>
Table 4: The target is a portfolio of (2,5,10)-year coupon bonds in proportions (-1,3,-1). Otherwise the format of the table is equivalent to Table 2. Results are in basis points (0.01%) per month.

<table>
<thead>
<tr>
<th>Model</th>
<th>Bias</th>
<th>Std. dev.</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Target movement</td>
<td>60.28</td>
<td>146.98</td>
<td>158.65</td>
<td>125.25</td>
</tr>
<tr>
<td>2 Duration matching</td>
<td>-13.31</td>
<td>54.91</td>
<td>56.42</td>
<td>41.32</td>
</tr>
<tr>
<td>3 Unrestricted 3-factor</td>
<td>-3.57</td>
<td>48.14</td>
<td>48.20</td>
<td>34.53</td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Unrestricted 3-factor</td>
<td>0.21</td>
<td>44.87</td>
<td>44.80</td>
<td>32.90</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Nelson-Siegel</td>
<td>-4.72</td>
<td>49.80</td>
<td>49.95</td>
<td>35.94</td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Nelson-Siegel</td>
<td>-3.56</td>
<td>48.64</td>
<td>48.70</td>
<td>35.30</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Nelson-Siegel a = 0.731</td>
<td>-1.86</td>
<td>35.02</td>
<td>35.02</td>
<td>26.62</td>
</tr>
<tr>
<td>8 Unrestricted 4-factor</td>
<td>3.46</td>
<td>34.95</td>
<td>35.07</td>
<td>26.30</td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Unrestricted 4-factor</td>
<td>0.86</td>
<td>36.59</td>
<td>36.55</td>
<td>25.40</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Augmented NS</td>
<td>2.15</td>
<td>34.34</td>
<td>34.36</td>
<td>25.10</td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 Augmented NS</td>
<td>1.37</td>
<td>26.84</td>
<td>26.84</td>
<td>18.73</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 Hull-White</td>
<td>0.06</td>
<td>20.09</td>
<td>20.06</td>
<td>15.45</td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 Hull-White</td>
<td>0.24</td>
<td>22.59</td>
<td>22.55</td>
<td>16.07</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 HW and ANS</td>
<td>0.38</td>
<td>22.62</td>
<td>22.59</td>
<td>16.74</td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 HW and ANS</td>
<td>2.20</td>
<td>29.67</td>
<td>29.71</td>
<td>20.92</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 SLSC curve</td>
<td>0.90</td>
<td>13.53</td>
<td>13.54</td>
<td>10.43</td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 SLSC curve</td>
<td>1.15</td>
<td>15.31</td>
<td>15.33</td>
<td>10.99</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 SLSC dynamic</td>
<td>0.61</td>
<td>12.22</td>
<td>12.22</td>
<td>9.45</td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 SLSC dynamic</td>
<td>0.77</td>
<td>12.36</td>
<td>12.37</td>
<td>9.56</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 SLSC combined</td>
<td>0.73</td>
<td>12.66</td>
<td>12.66</td>
<td>9.79</td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 SLSC combined</td>
<td>0.88</td>
<td>12.84</td>
<td>12.85</td>
<td>9.94</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 SLSC restricted</td>
<td>-3.79</td>
<td>46.17</td>
<td>46.26</td>
<td>33.26</td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 SLSC restricted</td>
<td>-2.78</td>
<td>44.83</td>
<td>44.85</td>
<td>32.95</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5: For the different estimated models the table shows estimated idiosyncratic standard deviations for each yield maturity. The reported figures are $\sqrt{\psi_i} \cdot 1000$. The two columns to the right show the maximum log-likelihood value and the number of free parameters for each model. The first eight models considered are for yield levels $y$, whereas the last four models are for aging-adjusted yield changes $\tilde{y}$.

<table>
<thead>
<tr>
<th>Model</th>
<th>3 mns.</th>
<th>6 mns.</th>
<th>12 mns.</th>
<th>2 yrs.</th>
<th>3 yrs.</th>
<th>5 yrs.</th>
<th>7 yrs.</th>
<th>10 yrs.</th>
<th>log$L$</th>
<th># params.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Unrestricted</td>
<td>1.21</td>
<td>0.30</td>
<td>0.72</td>
<td>0.40</td>
<td>0.30</td>
<td>0.51</td>
<td>0.28</td>
<td>1.01</td>
<td>65,408</td>
<td>37</td>
</tr>
<tr>
<td>2 Nelson-Siegel</td>
<td>1.26</td>
<td>0.30</td>
<td>0.81</td>
<td>0.44</td>
<td>0.30</td>
<td>0.53</td>
<td>0.28</td>
<td>1.08</td>
<td>64,928</td>
<td>23</td>
</tr>
<tr>
<td>3 Unrestricted</td>
<td>0.98</td>
<td>0.30</td>
<td>0.44</td>
<td>0.34</td>
<td>0.30</td>
<td>0.46</td>
<td>0.30</td>
<td>0.65</td>
<td>66,136</td>
<td>42</td>
</tr>
<tr>
<td>4 Augmented NS</td>
<td>1.07</td>
<td>0.30</td>
<td>0.62</td>
<td>0.36</td>
<td>0.30</td>
<td>0.45</td>
<td>0.30</td>
<td>0.84</td>
<td>65,469</td>
<td>27</td>
</tr>
<tr>
<td>5 SLSC curve</td>
<td>1.01</td>
<td>0.30</td>
<td>0.57</td>
<td>0.34</td>
<td>0.30</td>
<td>0.45</td>
<td>0.28</td>
<td>0.76</td>
<td>65,710</td>
<td>45</td>
</tr>
<tr>
<td>6 HW and ANS</td>
<td>6.84</td>
<td>5.59</td>
<td>4.08</td>
<td>1.58</td>
<td>0.30</td>
<td>2.47</td>
<td>3.50</td>
<td>4.20</td>
<td>57,441</td>
<td>11</td>
</tr>
<tr>
<td>7 SLSC combined</td>
<td>1.44</td>
<td>0.30</td>
<td>0.81</td>
<td>0.56</td>
<td>0.31</td>
<td>0.54</td>
<td>0.45</td>
<td>0.97</td>
<td>74,310</td>
<td>15</td>
</tr>
<tr>
<td>8 SLSC restricted</td>
<td>1.46</td>
<td>0.30</td>
<td>0.81</td>
<td>0.60</td>
<td>0.30</td>
<td>0.55</td>
<td>0.49</td>
<td>0.86</td>
<td>74,204</td>
<td>15</td>
</tr>
<tr>
<td>9 Hull-White</td>
<td>1.27</td>
<td>0.94</td>
<td>0.68</td>
<td>0.39</td>
<td>0.28</td>
<td>0.30</td>
<td>0.45</td>
<td>0.58</td>
<td>77,792</td>
<td>11</td>
</tr>
<tr>
<td>10 Unrestricted</td>
<td>1.19</td>
<td>0.94</td>
<td>0.74</td>
<td>0.51</td>
<td>0.38</td>
<td>0.14</td>
<td>0.32</td>
<td>0.44</td>
<td>78,579</td>
<td>24</td>
</tr>
<tr>
<td>11 SLSC dynamic</td>
<td>0.63</td>
<td>0.29</td>
<td>0.32</td>
<td>0.21</td>
<td>0.20</td>
<td>0.19</td>
<td>0.16</td>
<td>0.23</td>
<td>81,883</td>
<td>15</td>
</tr>
<tr>
<td>12 Unrestricted</td>
<td>0.66</td>
<td>0.25</td>
<td>0.32</td>
<td>0.22</td>
<td>0.20</td>
<td>0.21</td>
<td>0.11</td>
<td>0.26</td>
<td>82,748</td>
<td>37</td>
</tr>
</tbody>
</table>
Table 6: For full period estimation the table shows parameter estimates with standard errors below. For rolling four-year estimation the table reports the time-series mean.

<table>
<thead>
<tr>
<th>Model</th>
<th>1st factor</th>
<th>2nd factor</th>
<th>3rd factor</th>
<th>4th factor</th>
<th>5th factor</th>
<th>6th factor</th>
<th>7th factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nelson-Siegel</td>
<td>0.672</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson-Siegel</td>
<td>0.851</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Augmented NS</td>
<td>0.672</td>
<td>0.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Augmented NS</td>
<td>0.660</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLSC curve</td>
<td>0.687</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLSC curve</td>
<td>0.656</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hull-White</td>
<td>-0.027</td>
<td>0.010</td>
<td>-0.655</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full period</td>
<td>0.002</td>
<td>0.000</td>
<td>0.201</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hull-White</td>
<td>-0.047</td>
<td>0.010</td>
<td>-0.726</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td>0.001</td>
<td>0.000</td>
<td>0.200</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HW and ANS</td>
<td>0.119</td>
<td>0.014</td>
<td>-0.863</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full period</td>
<td>0.001</td>
<td>0.000</td>
<td>0.200</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HW and ANS</td>
<td>0.164</td>
<td>0.011</td>
<td>-0.938</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLSC dynamic</td>
<td>0.775</td>
<td>0.012</td>
<td>-0.337</td>
<td>0.013</td>
<td>-0.719</td>
<td>0.023</td>
<td>-0.567</td>
</tr>
<tr>
<td>Full period</td>
<td>0.020</td>
<td>0.000</td>
<td>0.203</td>
<td>0.000</td>
<td>0.205</td>
<td>0.001</td>
<td>0.218</td>
</tr>
<tr>
<td>SLSC dynamic</td>
<td>0.768</td>
<td>0.012</td>
<td>-0.312</td>
<td>0.012</td>
<td>-0.526</td>
<td>0.022</td>
<td>-0.260</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLSC combined</td>
<td>0.671</td>
<td>0.011</td>
<td>-0.430</td>
<td>0.011</td>
<td>-0.628</td>
<td>0.022</td>
<td>-0.239</td>
</tr>
<tr>
<td>Full period</td>
<td>0.007</td>
<td>0.000</td>
<td>0.200</td>
<td>0.000</td>
<td>0.200</td>
<td>0.001</td>
<td>0.200</td>
</tr>
<tr>
<td>SLSC combined</td>
<td>0.759</td>
<td>0.012</td>
<td>-0.368</td>
<td>0.012</td>
<td>-0.823</td>
<td>0.022</td>
<td>-0.028</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLSC restricted</td>
<td>0.676</td>
<td>0.010</td>
<td>-0.388</td>
<td>0.011</td>
<td>-0.788</td>
<td>0.022</td>
<td>-0.220</td>
</tr>
<tr>
<td>Full period</td>
<td>0.008</td>
<td>0.000</td>
<td>0.200</td>
<td>0.000</td>
<td>0.200</td>
<td>0.001</td>
<td>0.200</td>
</tr>
<tr>
<td>SLSC restricted</td>
<td>0.815</td>
<td>0.010</td>
<td>-0.433</td>
<td>0.012</td>
<td>-0.938</td>
<td>0.022</td>
<td>-0.116</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Maturity</th>
<th>3 mns.</th>
<th>6 mns.</th>
<th>12 mns.</th>
<th>2 yrs.</th>
<th>3 yrs.</th>
<th>5 yrs.</th>
<th>7 yrs.</th>
<th>10 yrs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st factor</td>
<td>0.067</td>
<td>0.087</td>
<td>0.104</td>
<td>0.131</td>
<td>0.142</td>
<td>0.151</td>
<td>0.147</td>
<td>0.137</td>
</tr>
</tbody>
</table>
### Table 8: Estimation of the Hull-White and the SLSC dynamic models, both with and without the HJM drift condition imposed. Figures for $\mu$ are in basis points, and $\mu$ is estimated freely when the drift is unrestricted, whereas with HJM drift it is calculated from parameters $a$, $\sigma$, and $\lambda$. The bottom of the table shows the test of restrictions imposed by the HJM drift condition using the likelihood ratio test.

<table>
<thead>
<tr>
<th></th>
<th>Hull-White</th>
<th>Hull-White</th>
<th>SLSC dynamic</th>
<th>SLSC dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HJM drift</td>
<td>Unrestricted drift</td>
<td>HJM drift</td>
<td>Unrestricted drift</td>
</tr>
<tr>
<td>$a$</td>
<td>-0.027</td>
<td>-0.032</td>
<td>0.775</td>
<td>0.808</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.010</td>
<td>0.009</td>
<td>0.013</td>
<td>0.013</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.752</td>
<td>-0.296</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td></td>
<td></td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td></td>
<td></td>
<td>-0.329</td>
<td></td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td></td>
<td></td>
<td>0.023</td>
<td>0.023</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td></td>
<td>-0.282</td>
<td></td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-1.379</td>
<td>-1.938</td>
<td>-1.637</td>
<td>-1.938</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-1.381</td>
<td>-1.859</td>
<td>-1.638</td>
<td>-1.859</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-1.386</td>
<td>-1.912</td>
<td>-1.609</td>
<td>-1.912</td>
</tr>
<tr>
<td>$\mu_4$</td>
<td>-1.396</td>
<td>-1.662</td>
<td>-1.488</td>
<td>-1.662</td>
</tr>
<tr>
<td>$\mu_5$</td>
<td>-1.406</td>
<td>-1.503</td>
<td>-1.350</td>
<td>-1.503</td>
</tr>
<tr>
<td>$\mu_6$</td>
<td>-1.425</td>
<td>-1.313</td>
<td>-1.118</td>
<td>-1.313</td>
</tr>
<tr>
<td>$\mu_7$</td>
<td>-1.443</td>
<td>-1.093</td>
<td>-0.956</td>
<td>-1.093</td>
</tr>
<tr>
<td>$\mu_8$</td>
<td>-1.468</td>
<td>-0.849</td>
<td>-0.796</td>
<td>-0.849</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>log $L$</th>
<th>77,792</th>
<th>77,820</th>
<th>81,883</th>
<th>81,885</th>
</tr>
</thead>
<tbody>
<tr>
<td># params.</td>
<td>11</td>
<td>18</td>
<td>15</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>$LR$</td>
<td>56.711</td>
<td></td>
<td></td>
<td>4.548</td>
<td></td>
</tr>
<tr>
<td>$\chi^2_{95}$</td>
<td>14.067</td>
<td></td>
<td></td>
<td>11.070</td>
<td></td>
</tr>
<tr>
<td>p-val</td>
<td>0.000</td>
<td></td>
<td></td>
<td>0.473</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Maturity</th>
<th>3 mns.</th>
<th>6 mns.</th>
<th>12 mns.</th>
<th>2 yrs.</th>
<th>3 yrs.</th>
<th>5 yrs.</th>
<th>7 yrs.</th>
<th>10 yrs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st factor</td>
<td>0.071</td>
<td>0.091</td>
<td>0.106</td>
<td>0.132</td>
<td>0.142</td>
<td>0.150</td>
<td>0.148</td>
<td>0.138</td>
</tr>
<tr>
<td>2nd factor</td>
<td>0.041</td>
<td>0.032</td>
<td>0.002</td>
<td>-0.024</td>
<td>-0.022</td>
<td>-0.008</td>
<td>0.008</td>
<td>0.013</td>
</tr>
<tr>
<td>3rd factor</td>
<td>-0.088</td>
<td>-0.081</td>
<td>-0.062</td>
<td>-0.037</td>
<td>-0.021</td>
<td>0.006</td>
<td>0.024</td>
<td>0.029</td>
</tr>
</tbody>
</table>
Table 10: Target is a 5-year coupon bond and hedge portfolios are constructed by the method in Theorem 7 that fully exploits the dynamic interest rate model. Otherwise the format of the table is equivalent to Table 3. Results are in basis points per month.

<table>
<thead>
<tr>
<th>Model</th>
<th>Bias</th>
<th>Std. dev.</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Movement</td>
<td>53.15</td>
<td>128.41</td>
<td>138.80</td>
<td>110.09</td>
</tr>
<tr>
<td>HW and ANS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full period</td>
<td>0.70</td>
<td>28.35</td>
<td>28.32</td>
<td>20.25</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td>3.29</td>
<td>49.77</td>
<td>49.75</td>
<td>33.93</td>
</tr>
<tr>
<td>SLSC combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full period</td>
<td>1.82</td>
<td>12.12</td>
<td>12.24</td>
<td>9.11</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td>1.74</td>
<td>13.08</td>
<td>13.18</td>
<td>9.89</td>
</tr>
<tr>
<td>SLSC restricted</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full period</td>
<td>1.92</td>
<td>13.09</td>
<td>13.21</td>
<td>9.78</td>
</tr>
<tr>
<td>Rolling 4-year</td>
<td>2.10</td>
<td>12.34</td>
<td>12.50</td>
<td>9.30</td>
</tr>
</tbody>
</table>
Table 11: Target is a portfolio of (2, 5, 10)-year coupon bonds in proportions (-1, 3, -1) and hedge portfolios are constructed by the method in Theorem 7 that fully exploits the dynamic interest rate model. Otherwise the format of the table is equivalent to Table 4. Results are in basis points per month.

<table>
<thead>
<tr>
<th>Model</th>
<th>Bias</th>
<th>Std. dev.</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Movement</td>
<td>60.28</td>
<td>146.98</td>
<td>158.65</td>
<td>125.25</td>
</tr>
<tr>
<td>HW and ANS Full period</td>
<td>-2.07</td>
<td>47.94</td>
<td>47.91</td>
<td>35.46</td>
</tr>
<tr>
<td>HW and ANS Rolling 4-year</td>
<td>-1.23</td>
<td>69.94</td>
<td>69.85</td>
<td>49.85</td>
</tr>
<tr>
<td>SLSC combined Full period</td>
<td>-1.48</td>
<td>39.48</td>
<td>39.45</td>
<td>28.81</td>
</tr>
<tr>
<td>SLSC combined Rolling 4-year</td>
<td>-1.32</td>
<td>45.92</td>
<td>45.87</td>
<td>33.67</td>
</tr>
<tr>
<td>SLSC restricted Full period</td>
<td>-1.09</td>
<td>47.82</td>
<td>47.76</td>
<td>34.47</td>
</tr>
<tr>
<td>SLSC restricted Rolling 4-year</td>
<td>-2.07</td>
<td>49.58</td>
<td>49.55</td>
<td>35.77</td>
</tr>
</tbody>
</table>
The panels in the upper row and the lower left panel show time to maturity, coupon rate, and duration, respectively, for the target 5-year bond to be hedged each month. These properties were retrieved from the CRSP monthly treasury file, selecting the coupon bond with maturity closest to five years, given a liquidity condition. The lower right panel shows the discrepancy between the actual price of the selected bonds as recorded in CRSP and the price that results from the FRB yield curve that was applied in the factor analysis. The discrepancy is shown in per cent that the FRB yield curve implied price exceeds the CRSP recorded price.


**Figure 2: Characteristics of target portfolio**

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

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

Figure 4: Loadings in Nelson-Siegel Model

The three functions \( B_j(\tau) \) in the NS model plotted against maturity \( \tau \).
**Figure 5:** Time series of $a$ parameter in Nelson-Siegel Model

Time series of $a$ in the rolling four-year estimations of the NS model with 95% confidence bands. The x-axis gives the end date of the four-year estimation window. The solid horizontal line indicates the full period estimate of $a$ and the dashed line is the value chosen by Diebold et al. (2006).
Figure 6: Loadings in 4-factor model

The panel to the left plots the columns $B_j$ in the unrestricted 4-factor model against maturity and the panel to the right draws the four functions $B_j(\tau)$ from the augmented NS model.
Figure 7: Time series of \( a \) in the SLSC combined model

Time series of estimated \( a \) with 95% confidence bands in the rolling four-year estimation of the SLSC combined model. The x-axis gives the end date of the rolling windows. The solid horizontal line indicates the corresponding full period estimate of \( a \).
Figure 8: Time series of $\sigma$ and $\lambda$ in the SLSC combined model

The left panel shows the time series estimates of the volatility parameters ($\sigma_1, \sigma_2, \sigma_3$) in the four-year rolling window estimation of the combined SLSC model. For the same model the right panel shows the time series of rolling estimates of the market price of risk parameters ($\lambda_1, \lambda_2, \lambda_3$). In each panel the parameter with index 1, 2, and 3 are the solid, the dashed, and the dotted line, respectively.
**Figure 9:** Time series of factors in the SLSC combined model

Time series of fitted factors in the SLSC combined model with the level factor, the slope factor, and the curvature factor in the upper, middle, and lower panel, respectively. The fitted factors are the smoothed Kalman filter estimates and thus use the full series of yield data, i.e., $E(f_n | y_0, \ldots, y_N)$. 
CAN LINEAR-RATIONAL TERM STRUCTURE MODELS CAPTURE CONDITIONAL VOLATILITY IN THE TREASURY YIELD MARKET?

Jorge W. Hansen
Aarhus University and CREATEs

Abstract

We show that the class of linear-rational square-root (LRSQ) model is able to match the cross section of yields and the time variability of conditional yield volatility simultaneously. Models in this class are, in this regard, able to break the tension noted for the affine term structure models from matching the conditional first and second moments of yields. Using a panel data set of US Treasury yields and realized yield volatilities, we evaluate the performance of various LRSQ model specifications based on in-sample and out-of-sample exercises and find that the preferred specification relies on three unspanned stochastic volatility factors, which, correlate strongly with the level and slope factor of conditional yield volatility.
2.1 Introduction

Even though the affine term structure models (ATSMs)\(^1\) constitute the main building block within the literature on interest rate term structure modeling, the accomplishments of modeling bond yield volatility within this class are in several respects still insufficient. That is, their popularity due to a large number of appealing properties\(^2\) can not distract from the fact that their good fit to the conditional first moments of yields does not carry over to an adequate fit to the conditional second moment. For instance, Collin-Dufresne, Goldstein and Jones (2009) and Jacobs and Karoui (2009) find that standard three- and four-factor ATSMs are capable of fitting the cross-section of yields, but model-implied conditional volatility of yields appears to be virtually unrelated to model-free measures of conditional volatility. This is commonly attributed to the fundamental property in ATSMs according to which yields as well as their conditional volatilities are affine in the term structure factors. Conversely, this suggests that yield volatility risk is assumed to be spanned by the term structure factors, where, though, some of them have the dual role of driving both the cross-sectional dimension and time series properties of yields.\(^3\) However, the failure of ATSMs to capture conditional volatility demonstrates that the applied estimation criteria tend to assign more weight to the cross-section and shows that the aforementioned factors fail to play their dual role.

To address this issue, Collin-Dufresne and Goldstein (2002) and Collin-Dufresne et al. (2009) suggest parameter constraints in ATSMs that release some volatility factors from their dual role such that they become independent of the level of yields and are free to drive the time series dimension of yield volatility. This property is in accordance with numerous papers in the literature that find evidence of so-called unspanned stochastic volatility (USV) in the fixed-income market. That is, the exist sources of risk which determine prices of fixed-income derivatives but cannot be hedged using bonds only, rendering the fixed-income market incomplete. To name but a few, Andersen and Benzoni (2010) use high-frequency US Treasury yield data and find that the yield curve is unable to span bond yield volatility. Using interest rate surveys, yields, and their realized and implied volatilities from the US Treasury market, Cieslak and Povala (2016) find that almost all variation in Sharpe ratios of

\(^{1}\)Examples of ATSMs are the models of Vasicek (1977), Cox et al. (1985), and Hull and White (1990). For a characterization of ATSMs, see Duffie and Kan (1996), and for a full classification of the admissible ATSMs; see Dai and Singleton (2000).

\(^{2}\)To name a few, ATSMs provide analytically tractable expressions for bond prices and interest rates derivatives, they fit the cross-section of bond yields well (see Dai and Singleton (2000) and Brandt and Chapman (2002), among others), and there has been developed a number of efficient methods to estimate ATSMs. Moreover, Gaussian ATSMs match the conditional first and second moments of yields under both the physical probability measure \(\mathbb{P}\) and the risk-neural probability measure \(\mathbb{Q}\) successfully (see Dai and Singleton (2002)).

\(^{3}\)Collin-Dufresne et al. (2009), for example, note that those factors that drive stochastic volatility in popular three- and four-factor ATSMs also drive changes in yields.
bonds is attributable to variation in bond yield volatility. Heidari and Wu (2003) extract principal components (PCs) from swaption-implied volatility and find that they vary independently of the spanned term structure factors but are crucial in explaining the variation in volatility.

Collin-Dufresne et al. (2009) find that introducing USV in ATSMs tilts the estimation from fitting the cross-section of yields towards fitting time series properties of yields. Unfortunately, Joslin (2018) notes that the USV restrictions in commonly studied ATSMs place rather tight constraints on the factor process which prevent the models from producing the observed cross-section of yield volatility but also impacts the dynamics of the conditional mean.\(^4\) We will for this reason consider a new class of models given by the linear-rational term structure model developed by Filipović et al. (2017) (henceforth FLT), which among other appealing properties, such as being tractable, accommodating the zero lower bound of interest rates, and providing semi-analytical expressions for swaption prices, is designed to incorporate USV. USV is achieved by postulating a state price density which is linear in the spanned factor process, which, in turn, has linear drift driven by a multivariate martingale.

FLT show that this structure implies that bond prices do not depend on the martingale process and allows them to write components into the stochastic part of the spanned factor process, which, consequently, do not appear in the bond pricing formula and become unspanned by bond prices. We will particularly consider the linear-rational square-root (LRSQ\((m, n)\)) model within the linear-rational term structure model in which the factor process follows an \((m + n)\)-dimensional square root diffusion process, where \(m\) and \(n \leq m\) denote the number of spanned and unspanned factors respectively. Considering the square-root specification allows us to adopt methods from the large toolbox of affine term structure models. Moreover, FLT show that swaption pricing becomes particularly tractable in this case.

On the empirical side, the first goal of this paper is to examine the ability of the LRSQ\((m, n)\) model to capture conditional volatility in the US Treasury yield market. More precisely, we intend to investigate whether the LRSQ\((m, n)\) model is able to break the tension from matching the cross-sectional dimension and time series properties noted for ATSMs. This is, apart the study in FLT, the first extensive empirical application of the LRSQ\((m, n)\) model, and the first application to bond yields. Furthermore, as explained below, we infer the USV factors from realized variance of yields instead of swaption prices as in FLT. The second goal is to study the latent factor structure underlying conditional volatility. In particular, we want to analyze the USV factors and how they drive conditional volatility. This extends the results in Heidari and Wu (2003) who find that the PCs extracted from swaption-implied volatility can

\(^4\)As noted by Dai and Singleton (2002), there is a trade-off in fitting conditional means and volatilities in ATSMs. This is also the reason why Gaussian ATSMs are the dominate models for modeling bond risk premia.
be interpreted as volatility level, slope, and curvature factors respectively. However, since the PCs of Heidari and Wu (2003) summarize purely cross-sectional information in the unconditional second moment of yields, they differ from variance factors inferred from a fully specified dynamic term structure model as in the present case. The latter also takes the conditional second moment into account, i.e., time-series properties, and places more structure on the dynamic evolution of the variance factors through time. This constitutes a straightforward framework for out-of-sample exercises.

The fact that conditional volatility is unobserved makes standard estimation techniques such as linear Kalman filtering or methods based on inverting the latent factors from the term structure of interest rates infeasible. As noted by Bates (2006), this hampers parameter inference and inference of the historical time series of volatility. Efforts in the literature to address this issue are based on generalized method of moments (Joslin, 2018), spectral analysis (Bates (2006) and Thompson (2008)), and Bayesian methods with filtering via Markov chain Monte Carlo (Collin-Dufresne et al., 2009). We estimate the LRSQ($m, n$) model by the Unscented Kalman Filter (UKF) (developed by Julier and Uhlmann (1997)), which is computationally simpler than, for instance, Markov chain Monte Carlo. For this, we establish a state space system for the LRSQ($m, n$) model in which we extend the measurement equation to include conditional yield volatility as extra measurement. In this regard, we exploit the result in Cieslak and Povala (2016), who find that complementing yield data with informative second moment data is effective for adequately inferring yield volatility. To account for the fact that volatility in the LRSQ($m, n$) model is nonlinear in the factors, we base our approach on the UKF, which is designed to accommodate nonlinearities. The ability of our approach to infer the spanned and unspanned factors, we demonstrate through a Monte Carlo study.

We consider various specifications for the LRSQ($m, n$) model which differ in the number of USV factors but are all based on the three spanned factors. We fix the number of spanned factors to three to capture the well-established result in the literature that the yield curve follows an level, slope, and curvature structure (see, e.g., Litterman and Scheinkman (1991)). We evaluate their ability to capture conditional volatility by means of a variety of regression analyses. In-sample and out-of-sample, we regress conditional volatility forecast errors on both cross-sectional information of yields and realized variances. These analyses let us determine whether the LRSQ($m, n$) models miss cross-sectional information that is relevant for matching conditional yield volatility dynamics, and moreover the importance of inducing USV in this context. In addition, in order to also address the time series dimension of yield volatility, we follow the USV modeling literature and compare to model-free measures of conditional volatility by using EGARCH estimates from the actual yields to model-implied conditional volatility.
On a general level, the analysis reveals the importance of complementing cross-sectional yield data with informative second-moment data. The cross-section of yields carries some information about conditional volatility but the significant part stems from the conditional second moments of yields. This highlights the importance of inducing USV in order to adequately estimate conditional volatility of the yield curve. In particular, the sample correlations between model-implied and EGARCH estimates of conditional volatility are largest for those model specifications that exhibit USV. Importantly, this does not come at the expense of a poor fit to the cross-section of yields. The regression analyses reveal that the model with only a single USV factor misses cross-sectional information relevant for modeling yield volatility. This is consistent with the EGARCH-analysis, showing that short-term and long-term volatility behave differently. In this regard, the additional flexibility of the LRSQ(3, 2) model and LRSQ(3, 3) model appears to be helpful in capturing the conditional volatility dynamics, and they are able to incorporate most of the cross-sectional information across the entire maturity dimension. This superior in-sample performance translates into better out-of-sample predictions. The predictive performance of the LRSQ(3, 1) model to capture future conditional volatility proxies is inferior to the models with two and three USV factors. Finally, the analysis shows that the LRSQ(3, 0) model is capable of fitting the cross-section of yields but fails on capturing both in-sample and out-of-sample conditional volatility.

With respect to the factor structure, we find that the spanned factors align closely with the PCs of yields and can be interpreted as the level, slope, and curvature of the term structure of interest rates (see Litterman and Scheinkman (1991)). A principal component analysis (PCA) on the set of realized variance reveals that a similar level-slope-curvature structure underlies the second moment of yields. Considering the sample correlation between these PCs and the inferred USV factors, we find that the first USV factor can be interpreted as the level factor of conditional volatility, and the second USV factor has a moderate positive correlation with the slope factor.

The remainder of this paper is organized into six sections. In Section 2.2, we describe the LRSQ($m$, $n$) model. In Section 2.3, we outline the estimation procedure. Section 2.4 provides an empirical application of four LRSQ($m$, $n$) model specifications to US Treasury bond yield data. In Section 2.5, we investigate the model’s predictive performances. We conclude in Section 2.8 and provide technical details in several appendices.

---

5 Note that the USV phenomena states that volatility cannot be replicated by yields only, but does not rule out a link between yields and volatility completely.
2.2 The linear-rational square-root model

The LRSQ\((m, n)\) model belongs to the general class of linear-rational term structure models, in which, for \(0 \leq n \leq m\), the \((m + n)\)-dimensional factor process \(X_t\) can be separated into an \(m\)-dimensional spanned factor \(Z_t\) and an \(n\)-dimensional unspanned factor \(U_t\). The basic approach is to define \(X_t\) such that \(Z_t\) has linear drift, and to choose the specification for the state price density \(\zeta_t\) wisely which then jointly imply that zero-coupon bond prices depend only on the drift of \(Z_t\). Based on this result, and furthermore the assumption that the drift of \(Z_t\) does not depend on \(U_t\), \(U_t\) becomes unspanned by bond prices. In order to make the separation strict, conditions on the drift parameters of \(Z_t\) ensure that all components in \(Z_t\) are spanned by zero-coupon bond prices.

More formally, for an auxiliary probability measure \(\mathbb{A}\) that is equivalent to the physical measure \(\mathbb{P}\),\(^6\) \(X_t\) is assumed to follow an \((m + n)\)-dimensional square-root process given by

\[
dX_t = \beta(\theta_x - X_t)dt + \Sigma\sqrt{S_t}dW^A_t,
\]

with state space \(\mathbb{R}^{m+n}\), where \(\beta\) is an \((m + n) \times (m + n)\)-matrix, \(\theta_x\) is an \((m + n)\)-vector, \(\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_{m+n})\) for positive volatility parameters \(\sigma_1, \ldots, \sigma_{m+n}\), and \(S_t\) is an \((m + n)\)-diagonal matrix with elements \([S_t]_{ii} = \nu_i'X_t\), where \(\nu_i\) is a vector of zeros with a one on the \(i\)th position. Moreover, \(W^A_t\) is a \(d\)-dimensional Brownian motion under \(\mathbb{A}\). Then, the state price density \(\zeta^A_t\) with respect to \(\mathbb{A}\) is defined as a function of \(Z_t\)

\[
\zeta^A_t = \exp(-\alpha T)(\phi + \psi'Z_t),
\]

for \(\alpha, \phi \in \mathbb{R}\), and \(\psi \in \mathbb{R}^m\) such that \(\phi + \psi \top x\) for all \(x \in \mathbb{R}^m\). For \(\tau = T - t\), using basic valuation principles, the time-\(\tau\) price of a zero-coupon bond paying one dollar at maturity \(T \geq t\), \(B(t, T) = (\zeta^A_t)^{-1}\mathbb{E}_t[\zeta^A_T \cdot 1]\), can be shown to be given by the ratio of two linear functions of \(Z_t\)

\[
B(t, T) = F(\tau, z) = \exp(-\alpha \tau)\frac{\phi + \psi' \theta + \exp(-\tau \kappa)(z - \theta_z)}{\phi + \psi' z},
\]

hence the name linear rational. Accordingly, the short rate, \(r_t = -\partial_T \log B(t, T)|_{T=t}\), is given by

\[
r_t = \alpha + \frac{\psi' \kappa(Z_t + \theta_z)}{\phi + \psi' Z_t}.
\]

Specifying the model under \(\mathbb{A}\) allows us to incorporate a permanent component into the state price density,\(^7\) admitting more sophisticated bond risk premium dynamics.\(^8\)

---

\(^6\)We consider a model defined on a standard filtered probability space \((\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})\), where \(\mathbb{P}\) denotes the physical probability measure. The information set \(\mathcal{F}_t\) is the natural filtration generated by a \(d\)-dimensional Brownian motion \(W^P_t\).

\(^7\)Alvarez and Jermann (2005) and Hansen and Scheinkman (2009) show that a state price density can be decomposed into a transitory and a permanent component. See Equation (29) in FLT for such a decomposition of the state price density in the LRSQ\((m, n)\) model.

\(^8\)FLT show that the permanent component is required to fit observed risk premiums successfully.
The dynamics of $X_t$ under $\mathbb{P}$ can be derived from the Radon-Nikodym density process given by
\[
\mathbb{E}[\frac{d\mathbb{P}}{d\mathbb{A}}] = \exp\left(\int_0^t \delta_s dW_s^\mathbb{A} - \frac{1}{2} \int_0^t \|\delta_s\|^2 ds\right),
\] (2.2.4)
where the integrand $\delta_t = \sqrt{S_t} \delta$ is given in terms of an $(m+n)$-vector of constants $\delta$. This specification preserves the square-root structure of $X_t$ under $\mathbb{P}$,
\[
dX_t = \tilde{\beta}(\theta_t - X_t) dt + \Sigma \sqrt{S_t} dW_t^\mathbb{P},
\] (2.2.5)
where $W_t^\mathbb{P}$ is a $d$-dimensional Brownian motion under $\mathbb{P}$, $\tilde{\beta} = \beta - \Sigma \Phi, \tilde{\theta}_t = \tilde{\beta}^{-1} \beta \theta_x$, and the $i$th row of $\Phi$ is given by $\delta_i \nu_i$. For an $(m+n) \times (m+n)$-matrix $M$ given by
\[
M = \begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix}, \quad A = \begin{pmatrix} I_n \\ 0 \end{pmatrix},
\]
where $I_k$ denotes a $k$-dimensional identity matrix, a decomposition into an $m$-dimensional spanned factor process $Z_t$ and an $n$-dimensional unspanned factor process $U_t$ is obtained by premultiplying $X_t$ by $M$, i.e., $M X_t = (Z_t, U_t)$ with state space $M(\mathbb{R}^{m+n})$. In particular, $Z_{it} = X_{it} + X_{m+i,t}$ and $U_{it} = X_{m+i,t}$ for $1 \leq i \leq n$, and $Z_{ii} = X_{ii}$ for $n+1 \leq i \leq m$. Moreover, choosing the following specifications in (2.2.1) for $\beta$ and $\theta_x$
\[
\beta = M^{-1} \begin{pmatrix} \kappa & 0 \\ 0 & A^\prime \kappa A \end{pmatrix} M = \begin{pmatrix} \kappa A - AA^\prime \kappa A \\ 0 & A^\prime \kappa A \end{pmatrix},
\]
\[
\theta_x = M^{-1} \begin{pmatrix} \theta_z \\ \theta_u \end{pmatrix}, \quad \beta \theta_x = \beta M^{-1} \begin{pmatrix} \theta_z \\ \theta_u \end{pmatrix} = \begin{pmatrix} \kappa \theta_z - AA^\prime \kappa A \theta_u \\ A^\prime \kappa A \theta_u \end{pmatrix},
\]
for $\theta_z \in \mathbb{R}^{m}, \theta_u \in \mathbb{R}^{n}$, and $\kappa \in \mathbb{R}^{m \times m}$ the desired structure for the joint factor process $(Z_t, U_t)$ is obtained and given by
\[
dZ_t = \kappa(\theta_z - Z_t) dt + \sigma(Z_t, U_t) dW_t^\mathbb{A},
\]
\[
dU_t = A^\prime \kappa A(\theta_u - U_t) dt + \text{diag}(\sigma_{m+1} \sqrt{U_{1,t}} dW_{m+1,t}, \ldots, \sigma_{m+n} \sqrt{U_{n,t}} dW_{m+n,t}),
\] (2.2.6)
where $\sigma(Z, U) = (I_m, A) \text{diag}(\sigma_1 \sqrt{U_1}, \ldots, \sigma_{m+n} \sqrt{U_n})$. FLT show that a unique solution to (2.2.1) and (2.2.6) exists if $\sigma_1, \ldots, \sigma_{m+n} \geq 0$, the mean reversion levels $\theta_z$ and $\theta_u$ ensure that $\beta \theta_x \in \mathbb{R}^{m+n}$, and $\kappa$ with nonpositive off-diagonal elements satisfies
\[
\text{span}\left\{ \mathbb{I}, \kappa \mathbb{I}, \ldots, \kappa^{(m-1)/2} \mathbb{I} \right\} = \mathbb{R}^{m}.
\] (2.2.7)

For future reference, we note that the yield $y(\tau, z) = -1/\tau \ln F(\tau, z)$ on a zero-coupon bond with term-to-maturity $\tau$ is given by (cf. Appendix A.1.1)
\[
y(\tau, z) = \alpha - \frac{1}{\tau} \log \left( \frac{\phi + \psi'(\theta_z + \exp(-\tau \kappa)(z - \theta_z))}{\phi + \psi' z} \right).
\] (2.2.8)
Since the level of yields does not provide any information of unspanned factors by definition, we seek to infer the USV factors in the empirical part from second-moments measurements of yields. For this, we note that the model-implied instantaneous dispersion matrix of the time-\(t\) yield with time-of-maturity \(T\) is given by (cf. Appendix A.1.2)

\[
\sigma^y_{t,T} = \nu_{t,T}(Z_t, U_t) \frac{\nabla_z F(T - t, Z_t)}{F(T - t, Z_t)},
\]

where \(\nu_{t,T}(Z_t, U_t)\) is the dispersion function of \(B(t,T)\),

\[
\nu_{t,T}(Z_t, U_t) = \sigma(Z_t, U_t) \frac{\nabla_z F(T - t, Z_t)}{F(T - t, Z_t)},
\]

and \(\nabla_z F(t, z)\) denotes the gradient with respect to the variable \(z\) such that

\[
\nabla_z F(t, z) = \exp(-\alpha(T - t)) \left( \frac{\exp(-\kappa(T - t))\psi}{F(T - t, z)} - \exp(\alpha(T - t))\psi \right).
\]

Appropriate scaling of \(\zeta^A_t\) does not affect any model prices such that we for convenience can assume that \(\phi = 1\) and \(\psi = 1_m\), where \(1_m\) denotes an \(m\)-vector of ones. This scaling also ensures that the short rate is bounded from below.\(^9\)

We close this section by emphasizing that starting by postulating a model for the state price density in the LRSQ\((m, n)\) model differs from the approach in ATSMs, where the usual approach is to start from an affine diffusion for the short rate instead.\(^10\) A second difference between the LRSQ\((m, n)\) model and an ATSM exhibiting USV is the fact that the latter is obtained from a model that builds fundamentally on the spanning relation between yield volatility and term structure factors, i.e., a non-USV model which therefore inevitably results in reduced flexibility of the model. Joslin (2018) shows, for example, that inducing USV in an \(A_1(4)\) model\(^11\) imposes a 1:2:4 ratio restriction on some parameters in the mean-reversion matrix in order to cancel out convexity effects which is necessary to make bond prices independent of volatility. This, in turn, implies that some risk factors have constant volatility and prevent the model from generating realistic volatility dynamics. Even though introducing USV in the LRSQ\((m, n)\) model also requires parameter constraints, that is, one needs to ensure that the gradient of \(F(t, z)\) with respect to \(z\) is zero along some dimensions in the state space of \(Z_t\),\(^12\) the required constraints are less restrictive than those imposed in the ATSMs. There is, for instance, no need to cancel convexity effects exactly, and the LRSQ\((m, n)\) model allows for much richer dynamics for the USV factors.

\(^9\)FLT show that the extremal values of the short rate in this case are given by \(r_t^{\min} = \min S\) and \(r_t^{\max} = \max S\), where \(S = \{1'\kappa_1, \ldots, 1'\kappa_m\}\) and \(\kappa_i\) denotes the \(i\)th column vector of \(\kappa\).

\(^10\)Specifying the dynamics of the factor process under the physical measure is equivalent to specifying the dynamics of the state price density. Hence, strictly speaking, ATSMs start also by postulating a model for the state price density.

\(^11\)In the notation of Dai and Singleton (2000), an \(A_m(n)\) model is an \(n\)-factor model in which \(m\) factors affect the volatility matrix.

\(^12\)See Theorem 2 in FLT for more details.
2.3 Estimation methodology

Based on the assumption that yields are measured with errors, Kalman filtering is among the most frequently applied methods used to estimate term structure models. Since the Kalman filter recursion computes the conditional mean and variance of the state variable from the estimated factors instead of from the true unobserved factors, quasi-maximum likelihood in conjunction with Kalman-filtering yields inconsistent estimates (Duan and Simonato (1999)). However, de Jong (2000), Chen and Scott (2003), and Duffee and Stanton (2012) examine this issue through simulation studies and conclude that the bias is negligible in practice. Furthermore, Lund (1997) and de Jong (2000) show that Kalman filtering in models with factor-dependent covariances (as in the present case) induces an additional inconsistency. But again, Duffee and Stanton (2012) find that kalman-filtering performs well in models where all stochastic volatility is spanned by bond prices. As argued by Collin-Dufresne et al. (2009), this does not apply to models exhibiting USV since the level of yields in this case are insufficient to identify the factors, making the bias severe. Hence, to enable identification of the USV factors, we pursue the route of Cieslak and Povala (2016) and complement the measurement equation with realized yield variance data.

2.3.1 Quasi maximum likelihood estimation and the Unscented Kalman Filter

As we will discuss below, the measurement function in the state space representation of the LRSQ($m$, $n$) model is nonlinearly in the factors. We will therefore rely on the UKF, which is designed to handle those nonlinearities. The UKF was originally developed by Julier and Uhlmann (1997) and is based on the unscented transform, which is a method of approximating the nonlinear transformed mean and covariance matrix of a random variable by a set of so-called sigma points. However, in order to circumvent potential numerical instabilities from a negative semi-definiteness covariance matrix during the estimation, we will rely on the more robust square-root UKF of van der Merwe and Wan (2002) which runs the filter for the square-root matrices of the covariance matrices. Further details on the square-root UKF are provided in Appendix A.1.3.

The factor $X_t$ in the LRSQ($m$, $n$) model follows a square-root diffusion process which implies that its transition equation is unknown, i.e., exact maximum likelihood estimation is infeasible. Nevertheless, its conditional first and second moments under the physical measure $\mathbb{P}$ are known and given by

$$E^\mathbb{P}[X_T | \mathcal{F}_t] = (I_{m+n} - \exp(-\tilde{\beta} \Delta t))\tilde{\theta}_x + \exp(-\tilde{\beta} \Delta t)X_t,$$

$$V^\mathbb{P}[X_T | \mathcal{F}_t] = \int_0^{\Delta t} \exp(-\tilde{\beta}s)\Sigma\Sigma'\exp(-\tilde{\beta}'s)ds,$$
where $\Delta t = T - t$. It is therefore common practice to use quasi-maximum likelihood estimation instead. That is, one approximates the non-Gaussian transition density of $X_t$ by a Gaussian density with the same conditional mean and variance, in which case the transition equation of $X_t$ becomes

$$X_t = G(X_{t-1}; \Theta) + \eta_t, \quad \eta_t \sim N(0, Q_t),$$

(2.3.1)

where $\Theta$ is the vector of model parameters, $\Delta t$ is the time between observations, $G(X_{t-1}; \Theta)$ is the transition function given by

$$G(X_{t-1}; \Theta) = (I_{m+n} - \exp(-\tilde{\beta} \Delta t))\tilde{\theta}_x + \exp(-\tilde{\beta} \Delta t)X_{t-1},$$

(2.3.2)

and

$$Q_t = V^p \left[ X_T | \mathcal{F}_t \right] = \int_0^{\Delta t} \exp(-\tilde{\beta} s)\Sigma \Sigma' \exp(-\tilde{\beta}' s) ds,$$

(2.3.3)

which is computed using the analytical solution provided in Fisher and Gilles (1996).

The measurement equation entails a 14-dimensional observation vector where the first seven components consist of the yields of different maturities, and the remaining components are their realized variances

$$\begin{pmatrix} y_t \\ \nu_t \end{pmatrix} = \begin{pmatrix} h_1(Z_t) \\ h_2(Z_t, U_t) \end{pmatrix} + \begin{pmatrix} \sigma_y I_7 \\ \sigma_{\nu} I_7 \end{pmatrix} \epsilon_t, \quad \epsilon_t \sim N(0, I_{14}),$$

(2.3.4)

where $\epsilon_t$ is independent of $\eta_t$ and the nonlinear functions $h_1(z)$ and $h_2(z, u)$ are given by

$$h_1(z) = \alpha - \frac{1}{\tau} \ln \left( \frac{\phi + \psi(z + \exp(-\tau \kappa)(z - \theta z))}{\phi + \psi z} \right),$$

(2.3.5)

$$h_2(z, u) = \frac{1}{\tau^2} \nu_{t,T}(z, u) \nu_{t,T}(z, u),$$

(2.3.6)

with $\nu_{t,T}(z, u)$ given in (2.2.9). The UKF is initialized at the unconditional mean and variance of the augmented state variable under $\mathbb{P}$

$$x_{0|0} = \left[ \tilde{\theta}^\prime_x, 0 \right]^\prime, \quad S_{xx(0|0)} = \sqrt{Q}, \quad S_{\xi(0|0)} = I_N,$$

where

$$Q = \int_0^\infty \exp(-\tilde{\beta} s)\Sigma \Sigma' \exp(-\tilde{\beta}' s) ds.$$

Then for each $t \in \{1, \ldots, T\}$, the expected value of the augmented state variable using information up to time $t$ (the filtered state) and the covariance matrix of the filtering uncertainty are inferred using the unscented transformation as shown in Appendix A.1.3. Asymptotic standard errors are obtained from the Fisher information matrix. In particular, they are obtained as the square root of the diagonal of the inverse Hessian matrix approximated numerically by a central finite difference scheme.
2.3.2 Monte Carlo analysis

We evaluate the efficiency of the UKF estimation in a Monte Carlo analysis. In particular, we examine the efficiency to estimate the LRSQ(3, 1) and LRSQ(3, 2) models. To that end, we divide each month into 1,000 time intervals, \( \Delta = 1/12 \times 10^{-3} \) and set the hypothetical true parameter values to the estimated parameter values obtained from fitting the empirical analysis below. From an Euler discretization of (2.2.1),

\[
X_{t+\Delta} = \max\left( \beta_x (\theta - X_t) \Delta + \Sigma \sqrt{S_t} \varepsilon_{t+\Delta}, 10^{-8} \right),
\]

we generate 500 factor trajectories. Then, we extract the 1,000th factor value within each month and compute the associated monthly zero-coupon bond yields with maturities of 0.5, 1, 2, 3, 5, 7, and 10 years and their realized variances corresponding to 25 years of data. In untabulated results, we find no significant effect from increasing the number of simulation paths. The standard deviation of the measurement errors is taken to be equal to \( \sigma_y = 3.5 \) basis points and \( \sigma_{\varepsilon r} = 0.8 \) basis points, similar to our maximum likelihood estimates. Finally, we estimate both models on each sample. Panel A in Table 1 shows that there is no systematic bias in the estimation of the LRSQ(3, 1) model. All estimated mean values are close to their true value, and no difference is significant. The same conclusion can be drawn from the results of the LRSQ(3, 2) model shown in Panel B of Table 1, and we conclude that the UKF is able to infer the true parameter values reliably.

2.4 Empirical analysis

In this section, we are going to estimate different LRSQ\((m, n)\) model specifications and examine subsequently their ability to capture conditional volatility. In accordance with the well-established fact that three term structure factors are necessary to describe the cross-section of yields (see, e.g., Litterman and Scheinkman (1991)), we fix \( m = 3 \) in all LRSQ\((m, n)\) model specifications. Then, in order to examine the significance of inducing USV as well as the number of USV factors required to match conditional volatility, we examine the effect of gradually increasing \( n \) from zero to three, \( n = 0, 1, 2, 3 \).

2.4.1 Data and summary statistics

The most frequently used US Treasury zero-coupon bond yield data sets in the term structure literature are the Fama and Bliss (1987) data set and the Gürkaynak, Sack and Wright (2007) (GSW) data set. Since the former only consists of zero-coupon bond yields with maturities of up to five years, we use the GSW data set from August 1987 through December 2017 in the empirical application below for a total of 7,591
calendar dates. This data set is obtained by fitting the Nelson-Siegel-Svensson yield curve introduced by Svensson (1995) to observed US Treasury bills, notes, and bonds prices from CRSP. We consider maturities of 0.5, 1, 2, 3, 5, 7, and 10 years, and start the sample in August 1987 in order to avoid the high yield and high yield volatility regime that lasted until the late 1980s. Andersen, Bollerslev, Diebold and Labys (2001) show that realized variance under suitable conditions is an unbiased ex-post estimator of conditional variance. Thus, for extracting the USV factors, we approximate yield variance by monthly realized variance computed from the daily yields. Feldhütter, Heyerdahl-Larsen and Illeditsch (2016) examined the accuracy of realized volatility computed from the daily GSW data set empirically and documented that it matches option-implied volatility well. Denoting the time-\(t\) monthly average yield and realized variance with term-to-maturity \(\tau\) by \(y_t(\tau)\) and \(rv_t(\tau)\), we perform the empirical analysis on

\[
y_t(\tau) = \frac{1}{N_t} \sum_{i=1}^{N_t} y_{d,t}(\tau),
\]

\[
rv_t(\tau) = 12 \sum_{i=1}^{N_t} \left(y_{d,t}(\tau) - y_{d,t-1}(\tau)\right)^2,
\]

where \(N_t\) denotes the number of trading days in month \(t\) and \(y_{d,t}(\tau)\) is the yield with term-to-maturity \(\tau\) observed at the \(i\)th day in month \(t\). In the case where \(i-1=0\), the last observation from the previous month \(t-1\) is used.

Figure 1 plots yields (left column) of different maturities and their realized variances (right column) at the monthly frequency. Moreover, in Table 2, we present summary statistics for the yields (Panel A) and realized variances (Panel B). For each of the seven maturities, the sample mean and standard deviation as well as the minimum and maximum values are shown. On average, the yield curve slopes strictly upward, from 4.08% to 5.83%. The standard deviations show that the short end of the yield curve varies more over time than the long end. Similarly, realized variances increase with maturity, and its standard deviations decrease.

### 2.4.2 Parameter estimates

All models are estimated using data from August 1987 to December 2011, whereas data from 2012-2017 is saved for out-of-sample analysis. The estimation appears to be very sensitive to the applied starting points. To accommodate for this, we adopt the heuristic approach of Jacobs and Karoui (2009) and generate 10,000 starting

---


14 The yields are taken directly from the GSW data set, except from the 0.5-year yield which we computed from the Nelson-Siegel-Svensson yield curve by using the estimated parameter values provided in the GSW data set.
points from the multivariate Gaussian distribution with a meaningful mean vector and covariance matrix that satisfy the admissibility condition mentioned in Section 2.2. Then, we compute their associated log-likelihood values and use those 50 points as starting points for the final estimation that have the largest log-likelihood values. The robustness of the obtained maxima was checked by replicating this approach several times.

In all considered LRSQ(3, n) specifications, the estimated upper-triangular elements of κ as well as its lower left element appeared to be in the vicinity of zero. Hence, in an attempt to decrease the complexity of the models, and since the maximized log-likelihood values, fit to yields, and realized variances are largely unaffected by this, we reestimate all specifications with those elements constrained to zero. Moreover, as noted by Duffee (2011), highly volatile expected returns are hard to reconcile with risk-based explanations of expected excess returns. Model-implied Sharpe ratio can therefore be used as an informal model specification test. Since we initially obtain slightly high Sharpe ratios, but constraining the mean maximum conditional Sharpe to values below 1.5 during the estimation does not reduce the model fits, we reestimate all models with this constrain imposed (we study the model-implied Sharpe ratios more formally in in Section 2.6). The estimated parameter values are reported in Table 3; they are quite precisely estimated, of similar magnitude across model specifications. Furthermore, they satisfy all condition (2.2.7), which cf. Definition 3 in FLT guarantees that (2.2.1) has a unique solution and that all estimated model specifications are well-defined. The estimated parameter values of κ imply that the second term structure factor (we will be using spanned factors and term structure factors interchangeably) is the least persistent term structure factor with the first and third term structure factor being somewhat more smooth. For 1 ≤ i ≤ n, consistent across all specifications it holds that σ_i < σ_i+3, which implies that all USV factors affect the instantaneous bond return covariances (cf. Corollary 1 in FLT). Moreover, FLT show that α can be interpreted as the infinite-maturity zero-coupon bond yield. The estimated parameter values for α ranging between 7.43% and 7.66% seem thus reasonable.

### 2.4.3 Estimated factors

Figure 2, which plots the time-series of the fitted spanned and unspanned factors, shows that the latter in many cases are large relative to the former, providing a first indication of the relevance of inducing USV. The sample correlations of the term structure factors show that the LRSQ(3, n) models bypass an empirical model inconsistency recognized within the popular square-root $\mathbb{A}_m(n)$ models. That is, admissibility in the square-root $\mathbb{A}_m(n)$ models restricts the correlations among the term structure factors to be positive which stands in contrast to negative sample correlations obtained by e.g. Dai and Singleton (2000) from fitting square-root $\mathbb{A}_m(n)$
models to US Treasury yield data. In the present case, however, in accordance with the structure in (2.2.6), we find positive sample correlations among the three term structure factors. In the LRSQ(3,3) model, for example, the sample correlation between $Z_{1t}$ and $Z_{2t}$, $Z_{1t}$ and $Z_{3t}$, and $Z_{2t}$ and $Z_{3t}$ is 50%, 25%, and 43% respectively.\footnote{Not reported here, we also find positive correlation coefficients in the remaining models.} Here, we want to emphasize that even though the volatility function $\sigma(z,u)$ of $Z_t$ in (2.2.6) implies that the instantaneous correlation between the term structure factors is zero, the lower bi-diagonal structure of $\kappa$ with negative off-diagonal elements allows for a nonnegative unconditional correlation.

In general, the loadings of the first, second, and third PC obtained from performing a PCA on a set of actual yields imply that they are usually interpreted as the level, slope, and curvature of the yield curve and can be empirically proxied by the ten-year yield (level), the difference between the ten-year yield and the one-year yield (slope), and twice the five-year yield minus the sum of the one-year yield and ten-year yield (curvature). In order to check whether the fitted term structure factors allow for a similar interpretation, we note that normalizing $Z_t$ as

$$\tilde{Z}_t = \frac{Z_t}{\phi + \psi' Z_t}$$

and linearizing $y(\tau, \tilde{Z}_t)$ around $\theta_z/\phi + \psi' \theta_z$, yields become affine in the normalized term structure factors (cf. FLT)

$$y(\tau, \tilde{Z}_t) \approx \alpha - A(\tau) - \frac{B(\tau)}{\tau} \tilde{Z}_t,$$

where

$$A(\tau) = 1 + \psi' \left( I_m - \exp(-\kappa \tau) \right) \frac{\theta_z}{\phi},$$

$$B(\tau) = -\psi' \left( I_m - \exp(-\kappa \tau) \right) \left( I_m + \frac{\theta_z}{\phi} \psi' \right).$$

An appealing property of this linearization is the fact that the relationship between the normalized factors and yield variance becomes close to what PCA seeks to achieve.\footnote{See Joslin (2018) for further details on this observation within the ATSMs where yields also are affine in the term structure factors.} Indeed, computing the sample correlations between the normalized term structure factors and the empirical level, slope, and curvature factors, as defined above, shows that the former also allow for the level-slope-curvature interpretation. That is, within e.g. the LRSQ(3,1) model, the correlation between $\tilde{Z}_{3t}$ and the empirical level factor is 97%, the correlation between $\tilde{Z}_{1t}$ and the empirical slope factor is 88%, and the correlation between $\tilde{Z}_{2t}$ and the empirical curvature factor is 50%.

In order to study the USV factors, we plot in Figure 3 EGARCH(1,1) estimates of conditional volatility for selected maturities. The volatility dynamics of long-term
yields appear to be highly correlated and less volatile than short end yield volatility and suggest that a quickly mean-reverting factor underlies short-term movements, whereas a more persistent level factor drives the long end. Joslin (2018) shows that the failure of commonly applied three- and four-factor ATSMs exhibiting USV to capture this bivariate structure in the cross-section of yield volatility can be attributed to the strong restrictions on the mean-reversion rates of the risk factors implied by USV.

In the present case, in accordance with Figure 3, the half-life times π computed from the estimated mean-reversion matrix $A' \kappa A$ of the USV factors\(^\text{17}\) in the LRSQ(3,2) model show that the first USV factor $U_{1t}$ is very persistent ($\pi = 4.31$ years) and might thus drive long-term yield volatility, and $U_{2t}$ is quickly mean-revering ($\pi = 0.57$ years), driving short-term yield volatility. The eigenvalues of the mean-reversion matrix $A' \kappa A$ in the LRSQ(3,3) model imply that the half-life times of $U_{1t}$, $U_{2t}$, and $U_{3t}$ are equal to 3.23, 0.52, and 6.82 years respectively. Given the fact that the LRSQ(3,3) model dominates the in-sample fit analysis (as further elaborated on below), it implies that the long-term variation in Figure 3 may be better captured by two persistent factors. Conversely, and not surprisingly, the bivariate structure produces a tension in the LRSQ(3,1) model. The mean-reversion coefficient $A' \kappa A = 0.21$ of $U_{1t}$ corresponds to a half-life time of $\pi = 3.24$ years and implies that the single USV factor is too rigid to capture the large spikes of conditional volatility. This is consistent with the upper right panel of Figure 4 which plots the actual and fitted realized variances of the ten-year yield for the LRSQ(3,1) model, where it is evident that fitted realized variance is less volatile than actual realized variance. For completeness, we also note that the LRSQ(3,0) model-implied variances in the upper left panel of the figure are even smoother. The lower left and right panels of the figure show that the increased flexibilities of the LRSQ(3,2) model and LRSQ(3,3) model where more than one factor is free to match conditional volatility provide an advantage in fitting realized variance. Nevertheless, it applies also to these models that fitted realized variance is less volatile than actual realized variance, which, however, might as well be explained by the fact that the LRSQ(3, n) model in general seeks to match ex-ante expectations of variance, but the outliers are ex-ante unpredictable (see Andersen, Bollerslev and Meddahi (2005) for more details).

To better understand the role of the USV factors and how to interpret them, we conduct a PCA on the set of actual realized variances. This shows that most variation in realized variances can be explained by three PCs, with the individual PCs accounting for 73.12%, 19.67%, and 6.36% of the variation respectively. Moreover, it follows from Figure 5 which plots the PC loadings against maturity that the underlying factor structure resembles the level-slope-curvature-structure commonly documented for yields. That is, the first factor is flat in maturity dimension, the second factor loads

\(\text{17}\)The eigenvalues of the mean-reversion matrix determine the half-life time of shocks to the factors. An eigenvalue of $A$ corresponds to a half-life time of $\pi = \ln 2/\lambda$. 

Chapter 2. Conditional volatility in the Treasury yield market

strongest on short-term maturities and changes sign for longer maturities, and the third factor mostly loads on intermediate maturities.

In the LRSQ(3,2) model, consistent with our previous findings from Figure 3, $U_{1t}$ is most highly correlated with the variance level factor (62%). The interpretation of $U_{2t}$ is less clear, though, with only a moderate positive correlation with the variance slope factor (22%). Similarly, in the LRSQ(3,3) model, the correlation of $U_{1t}$ with the variance level factor is 51%, and the correlation of $U_{2t}$ with the variance slope factor is 21%. Interesting is, however, that $U_{3t}$, too, is most highly correlated with the variance level factor (43%) and acts thus as a second level factor. As expected, in the LRSQ(3,1) model, $U_{1t}$ is most highly correlated with the variance level factor (57%). When interpreting those values, we must to take into account that PCs in general are different from $U_{it}$ obtained from a fully specified LRSQ(3,n) model since PCs in contrast to $U_{it}$ summarize purely cross-sectional variance information. Furthermore, the relationship between the USV factors and yield variance is nonlinear in the LRSQ(3,n) model, and hence different from the relationship between PCs and yield variance in PCA. Despite of this, the reported correlation coefficients show that there is a significant link between the USV factors and the PCs in the present case. In Section 2.4.5, we test formally to which extent the LRSQ(3,n) models are able to infer all cross-section realized volatility information that is of statistical significance for modeling conditional volatility. There, we obtain positive results for most LRSQ(3,n) model specifications. The relatively low correlation with the variance slope factor, which loads mostly on short-term volatility, indicates that matching those dynamics puts a high bar on the models, though. We conclude that the factor structure underlying realized variance behaves in many ways similarly to the factor structure usually noted for yields with the first and second factor accounting for most variation, where $U_{1t}$ and $U_{3t}$ correlate most highly with the variance level factor, and $U_{2t}$ correlates most highly with the variance slope factor.

2.4.4 Model comparison

We now compare the different LRSQ(3,n) model specifications more formally. The mean errors and root mean-squared errors (RMSEs) between the actual and fitted yields for all maturities are depicted in Table 4 where the fitted yields are obtained from (2.2.8). In line with the commonly recognized three-factor structure of the yield curve all models provide a good yield fit and are, across models, within a few basis points. For instance, RMSEs for the 5-year yield range between five and eight basis points. The LRSQ(3,2) and LRSQ(3,3) model have largest RMSEs, which may be a result of the fact that we constrain the drift of $Z_t$ by increasing the number of USV factors, since $\kappa$ appears both in $\theta_z$ and $\theta_u$. Yet, the log-likelihood values reported in Table 3, which are increasing in the number of USV factors, show that these economically small differences are statistical unimportant. Thus, maximum
likelihood trades-off the constraint on $\kappa$ against a small amount of in-sample accuracy. We cannot test the model restrictions statistically since the different LRSQ(3, $n$) model specifications are not nested within each other. The relatively large differences in the log-likelihood values (for example, the log-likelihood value of the LRSQ(3, 1) model is 28,067 vs. 28,747 of the LRSQ(3, 2) model) suggest, though, that gradually increasing the number of USV factors is statistically significant. To check for evidence of overfitting, we study in Section 2.6 model-implied Sharpe ratios and find that all model-implied Sharpe ratios are within a reasonable range.

### 2.4.5 Fit to the conditional second moments of yields

In this section, we examine the ability of the different LRSQ(3, $n$) model specifications to capture conditional volatility. Since conditional volatility is unobserved, it raises a number of methodological decisions.

(i) The first decision concerns the approach used to infer conditional volatility of the actual yields.

(ii) The second decision relates to choice of method used to infer model-implied conditional volatility.

(iii) The third decision concerns the metric(s) used to evaluate the reliability of model-implied conditional volatility.

For (i), in order to obtain a model-free estimate of conditional volatility, one approach that is widely used is to estimate a univariate EGARCH(1, 1) model on the time series of the actual yields for a given maturity or to proxy conditional volatility by realized volatility. These estimates are then considered as the ‘true’ conditional volatilities. However, the model-free approach must be treated with care as it is only model-free to the extent that it is independent of the term structure model but nevertheless reliant on a conditional volatility model. We will, as elaborated on below, use both approaches.

For (ii), a common approach relies on the reprojection method of Gallant and Tauchen (1998) (see Dai and Singleton (2003) and Bikbov and Chernov (2011)). In general, this method is used to approximate the unknown conditional density function of some model of interest with the density function of an auxiliary model. In this regard, we follow the USV modeling literature and estimate the unknown conditional volatility by fitting an EGARCH(1, 1) model on the (in-sample) forecasts of yields computed from an extracted time series of fitted factor values.

Lastly, for (iii), we consider several metrics. First, we examine the in-sample conditional volatility fit computed as the difference between the actual and fitted realized volatilities. Subsequently, we evaluate the reliability of the fitted conditional volatility...
dynamics by means of multiple regression analyses involving model-free and model-dependent conditional volatility proxies. To that end, denote by \( \text{rv}_t^\text{vol}(\tau) \) and \( \text{rv}^\text{m,vol}_t(\tau) \) the actual and fitted realized volatilities and by \( \text{PC}_{i,t-1}^\text{y} \) the PCs of yields at time \( t - 1 \). Then we sequentially regress \( \text{rv}_t^\text{vol}(\tau) \) onto its model-implied counterpart and onto the lagged PCs, \(^{18}\) i.e

\[
\text{rv}_t^\text{vol}(\tau) = \alpha + \beta_m \text{rv}^\text{m,vol}_t(\tau) + \sum_{i=1}^{3} \beta_i \text{PC}_{i,t-1}^\text{y} + \varepsilon_t. \tag{2.4.3}
\]

This approach is related to the volatility forecasting regressions in Collin-Dufresne et al. (2009) who regress realized swap volatility onto lagged PCs of actual yields and onto in-sample realized volatility swap forecasts obtained from ATSMs. These regressions allow us to investigate the relationship between conditional volatility and the yield curve and the importance of inducing USV in this context. In addition, it allows us to examine whether the cross-section of yields or time series information of conditional volatility carries most information for modeling conditional volatility.

We will use two criteria to analyze the regressions. Firstly, we conduct a Wald test of the joint hypothesis that \( \alpha = 0 \) and \( \beta = 0 \) which we base on simple unadjusted standard errors since the residuals in the regression under the null are serially uncorrelated. Secondly, we consider the adjusted \( R^2 \) values where a higher value implies a stronger relation between volatility and the yield curve. In order to examine the accuracy of the model forecasts, we also regress \( \text{rv}_t^\text{vol}(\tau) \) onto \( \text{rv}^\text{m,vol}_t(\tau) \) only. A reliable proxy should be positively related to the observed measurements, so we consider the sign and statistical significance of the regression coefficient as well as the adjusted \( R^2 \) values.

As noted before, using an EGARCH model to measure conditional volatility serves as a useful alternative to realized volatility. Hence, as a point of reference, we also estimate an EGARCH(1, 1) model with an AR(1) specification for the conditional mean for the actual yields and include its conditional volatility estimates in the regression. In the following, we will refer to this regression as the volatility fit regression.

We also perform a regression analysis where we regress the in-sample fitting errors \( \text{rv}_t^\text{vol}(\tau) - \text{rv}^\text{m,vol}_t(\tau) \) onto the lagged PCs of realized volatility, i.e.,

\[
\text{rv}_t^\text{vol}(\tau) - \text{rv}^\text{m,vol}_t(\tau) = \alpha + \sum_{i=1}^{3} \beta_i \text{PC}_{i,t-1}^\text{rv} + \varepsilon_t, \tag{2.4.4}
\]

where \( \text{PC}_{i,t-1}^\text{rv} \) denotes the PCs of realized volatility at time \( t - 1 \). This regression allows us to test whether the LRSQ(3, n) models disregard relevant information for modeling volatility. It is a priori unclear whether a correctly specified model should be able to drive out the significance of the cross-sectional information of realized volatility.

\(^{18}\)The relationship between volatility and the term structure factors suggested by the regression involving the PCs only corresponds to the relationship suggested in the ATSMs.
2.4. Empirical Analysis

since the PCs in (2.4.3) in contrast to \( Z_t \) and \( U_t \), as noted above, summarize purely cross-sectional information. As before, we will test the null hypothesis that the model is correctly specified by means of a Wald test of zero-regression coefficients, i.e., \( \alpha = 0 \) and \( \beta = 0 \), based on simple unadjusted standard errors. We also include the in-sample fitting errors from estimating an EGARCH(1,1) model for the actual yields. We will henceforth refer to this regression as the volatility fitting error regression.

In order to also address the time series dimension of yield volatility, we follow the literature and examine lastly the quality of the model-implied conditional volatility by comparing EGARCH(1,1) estimates of conditional volatility obtained from actual yields to those obtained from fitted yields.

2.4.6 Fit to realized variance

Table 5 reports the mean errors and RMSEs (in basis points) between actual and fitted realized variances. The RMSEs are largest in the short end and decrease along the maturity dimension.\(^{19}\) For instance, the RMSEs of the LRSQ(3,2) model range from 0.62 to 0.97 basis points and are maturity-by-maturity consistently smaller than the RMSEs of the LRSQ(3,0) model. But as for yields, having three USV factors in the LRSQ(3,3) model decreases the fit to the data. The fits across all models are though within one basis point.

2.4.6.1 Volatility fit regressions

Next, we consider the results from the volatility fit regressions (2.4.3) summarized in Table 6. Starting with the regression involving the PCs only, the adjusted \( R^2 \) values in Panel A reveal that the explanatory power of the cross-section of yields for conditional volatility increases along the maturity dimension (an adjusted \( R^2 \) value of 4% for \( \tau = 3 \) against an adjusted \( R^2 \) value of 18% for \( \tau = 10 \)), but the \( p \)-values of the Wald test show that this link is insignificant at the 1% level for \( \tau = 1, 2 \). We note here that this does not contradict the hypothesis of USV, since USV states that volatility cannot be replicated by yields only, but does not rule out a link between yields and volatility completely. Panel C shows that the LRSQ(3,0) model proxies conditional volatility poorly. The LRSQ(3,0) model has some explanatory power in the very short end, but the estimated coefficients from the regressions involving only the model-implied proxies (Panel C.2) are insignificant for \( \tau = 2, 3, 5, 7 \), and the significance for \( \tau = 10 \) is driven out when complementing the regression with the PCs (Panel C.1).

This suggests that the missing USV in this specification prevents the model from exploiting second-moment data information systematically, and its fitted conditional volatility proxies are for most maturities unrelated to realized volatility or simply

\(^{19}\) The fit to realized yield variance is comparable to the fit typically found in the literature, see, e.g., Cieslak and Povala (2016).
spanned by the PCs. Panel D.2 shows that the intermediate maturities also put a high bar on the LRSQ(3, 1) model where some of the regression coefficients are insignificant, and the adjusted $R^2$ values are close to zero. However, the estimated regression coefficients for the remaining maturities are significantly positive, and the model provides incremental information beyond the PCs as measured by the difference between the adjusted $R^2$ values in Panel D.1 and those in Panel A. Panel E and F show that the regression coefficients in the models with at least two USV factors are in most cases significantly positive, and the adjusted $R^2$ values increase along the maturity dimension and double compared to the regressions in Panel A. This demonstrates the value of complementing yield data with second-moment measurements and that these models manage to exploit this in terms of larger adjusted $R^2$ values.

Moreover, both models are to a large extent able to drive out the significance of the level factor, but the slope and curvature factor remain statistically significant in the long end. Even though the adjusted $R^2$ values of the LRSQ(3, 2) model exceed the adjusted $R^2$ values of the LRSQ(3, 3) model for some maturities (for example, 27% vs. 26% for $\tau = 10$ in the regressions involving only the model-implied estimates), the adjusted $R^2$ values of the LRSQ(3, 3) model are more stable across the maturity spectrum. Also the fact that maturity-by-maturity the difference between the adjusted $R^2$ values in the first and second subpanel within each panel, which measures the incremental information provided by the PCs beyond the fitted conditional volatility proxies, is smallest for the LRSQ(3, 3) model demonstrates the superior performance of the model with three USV factors.

Comparing the results to the EGARCH(1, 1) model, Panel B shows that this model has a clear advantage over all LRSQ(3, n) models in the short end with the adjusted $R^2$ values for both regressions being more than twice as large than those of the LRSQ(3, 3) model. However, the advantage vanishes in the long end, and the negligible differences between the adjusted $R^2$ values for $\tau \leq 2$ in Panel B.1 and Panel B.2 show that close to all explanatory power for short-term conditional volatility stems from the conditional second moment of yields. This thus demonstrates further that short-end volatility behaves differently than long-end volatility. Moreover, it shows that the performance of the LRSQ(3, n) models peaks when both the yield curve and the conditional second moments of yields are valuable for modeling conditional volatility, in which case these models are able to match or in some cases able to outperform an univariate EGARCH(1, 1) model.

While not directly comparable to the results in Collin-Dufresne et al. (2009) because they use in-sample volatility forecasts, it is worth noting that their results for the short-end are in accordance with Table 6. Similarly, they find for the 0.5-year volatility that only those ATSMs which exhibit USV are able to add incremental information beyond the PCs to the regression, and the adjusted $R^2$ value of their affine four-factor $A_1(4)$ model with a single USV factor is of similar magnitude as adjusted $R^2$ value obtained
from the four-factor LRSQ(3, 1) model here. However, they find that the performance of their models deteriorates for longer maturities which stands in contrast to the good performance of the USV models in the present case.

### 2.4.6.2 Volatility fitting error regressions

The results from the volatility fitting error regression (2.4.4) are reported in Table 7.\(^{20}\) The main result from the table is that the USV models are correctly specified as none of the nulls of the joint test can be rejected. Moreover, Panel A shows that also most \( p \)-values for the LRSQ(3,0) model associated with the Wald test exceed the 5% level across the maturity dimension, but many individual regression coefficients are significantly different from zero. The regression coefficients of the level factor are in fact significant in all cases. Furthermore, the adjusted \( R^2 \) values are fairly large, especially in the long-end where they are equal to 20%. The results of the LRSQ(3, 1) model and LRSQ(3, 2) model (Panel B and C) show that inducing USV helps fit cross-sectional volatility. The adjusted \( R^2 \) values decrease when increasing the number of USV factors which range between 1% and 16% for the LRSQ(3, 1) model and between 2% and 12% for the LRSQ(3, 2) model. They are, however, larger than those of the LRSQ(3, 3) model in Panel D, which do not exceed 3%. Comparing with the EGARCH(1, 1) model, the results in Panel A show that the LRSQ(3, 3) model matches the performance of the EGARCH(1, 1) model for short-term volatility but clearly outperforms it for \( \tau \geq 3 \). The performance of the EGARCH(1, 1) model deteriorates significantly for long-term volatility, and the model appears to be incorrectly specified for \( \tau = 10 \). The missing cross-sectional link in the EGARCH(1, 1) model let is miss certain aspects of yield volatility, which constitutes a clear disadvantage compared to the LRSQ(3, \( n \)) model.

### 2.4.6.3 Conditional volatility implied by an EGARCH model

Following the usual practice in the literature, we also evaluate the ability of a model to capture conditional volatility based on comparing EGARCH(\( p, q \)) models (for \( 1 \leq p, q \leq 2 \)) estimates of conditional volatility of both actual and fitted yields.

The correlations between model-implied volatilities obtained from ATSMs and model-free volatilities obtained from, for instance, EGARCH models reported in the literature are usually relatively low (see e.g. Collin-Dufresne et al. (2009), Jacobs and Karoui (2009), and Christensen, Lopez and Rudebusch (2014)). They are below 80% and often close to zero or even negative.\(^{21}\) Whereas Jacobs and Karoui (2009) interpret

\(^{20}\)This approach is related to the cross-sectional regression in Cieslak and Povala (2016). They examine the cross-sectional fit to realized volatility obtained from high-frequency yield data in a single step and hence include the lagged PCs of both realized volatility and yields in the regression. They consider the fit of their ATSMs with six risk sources and for an \( A_2(4) \) model. The adjusted \( R^2 \) values for the former model vary between 0% and 3% across the maturity dimension and between 5% and 23% for the latter model.

\(^{21}\)See Table 4 in Bikbov and Chernov (2009) for an overview of selected results.
this as evidence that ATSMs (without USV) perform rather poorly at extracting conditional volatilities of yields. Collin-Dufresne et al. (2009) find that conditional volatility cannot be extracted from the cross-section of yields only and inducing USV in ATSMs, coupled with time-series information, improves the fit to short-rate volatility in this model class.\textsuperscript{22} The results in Table 8 show that the correlations coefficients in the present case are larger than those reported for ATSMs. Moreover, this is not limited to those models that exhibit USV only, as can be seen from the first column in the table which shows that even the correlation coefficients obtained from the LRSQ(3,0) model range consistently above 95%. This thus indicates that part of the poor performance usually documented for ATSMs might in fact be explained by only including first-moment data in the estimation. That is, including informative second-moment data tilts the estimation criteria successfully towards matching time-series properties of yields which was the motivation for including those from the outset.

In summary, the empirical in-sample results tell us the following: Firstly, it is important to include informative second-moment data to reliably infer conditional volatility of yields, that is, USV factors are required to capture the conditional second moment of yields. Secondly, short-term and long-term conditional volatilities behave differently, i.e., more than one factor is not need to capture conditional volatility across the maturity spectrum. Thirdly, the preferred model based on the in-sample analysis has three USV factors and is in many respects able to match or even outperform an EGARCH(1,1) model.

2.5 Predictive performance

The analysis has so far been purely in-sample. We turn now to the out-of-sample performances of the different LRSQ(3, n) models to capture conditional volatility and the role of USV in this context. For this sake, we consider the volatility fitting error regression (2.4.4) applied in Section 2.4 without the benefit of hindsight. Collin-Dufresne et al. (2009) point out the importance of USV for forecasting realized volatility within the canonical $A_m(n)$ models. They document a superior performance of the $A_{1}(4)$ model exhibiting USV over the $A_{1}(3)$ model without USV across the entire maturity dimension and an advantage over the $A_{1}(4)$ model without USV in the short end.

As input to the out-of-sample regressions, we focus on one-month-ahead forecasts, $h = 1/12$, of conditional volatility. To that end, at a given month $t$ with the estimated

\textsuperscript{22}Christensen et al. (2014) and Bikbov and Chernov (2009) review those (E)GARCH-based methods in simulation studies and find that they do not serve as reliable tests of the null hypothesis that volatility is spanned by the level of yields. Low estimated correlations coefficients or low adjusted $R^2$ values obtained from regressing model-free (E)GARCH-estimates onto model-implied (E)GARCH-estimates might not be sufficiently informative to reject spanned volatility due to the generated regressor problem and misspecifications. This implies that one needs to be cautious when interpreting the results in terms of presence of USV. We include this analysis despite its shortcomings, since it allows us to assess the degree to which the LRSQ(3, n) models are able to capture conditional volatility.
2.5. PREDICTIVE PERFORMANCE

With parameter values in hand, we firstly infer the current factor vector \( \hat{X}_t \). Then, we simulate 10,000 trajectories for the factor process between \( t \) and \( t + h \) using an Euler discretization scheme for (2.2.5). For each trajectory, we obtain the model forecasts for a maturity-\( \tau \) yield \( \hat{y}_{t+h}(\tau) \) made at time \( t \) for \( t + h \) from (2.2.8). Finally, we proxy conditional volatility by expected absolute one-month changes in yields \( E_t(|\Delta y(\tau)|) \).

We perform the out-of-sample forecasting exercise based on an expanding estimation window by adding step-by-step one month of data until the end of the sample period is reached. The first out-of-sample estimation window is August 1987 through December 2011, the second is August 1987 through January 2012, and the last is August 1987 through November 2017, which yields 72 out-of-sample months in the largest out-of-sample window. Before we proceed to the regression analysis, we also evaluate the forecast accuracy across the different model specifications by computing the root mean squared forecast error (RMSFE) of the \( h \)-step-ahead forecast error for term-to-maturity \( \tau \), \( \hat{e}_{t+h,t}(\tau) = \text{rv}_{t+h,t}(\tau) - E_t(|\Delta y(\tau)|) \). We assess the forecast accuracy of the different model specifications by means of the Diebold and Mariano (1995) test using the LRSQ(3,2) model as benchmark. Thus, for a given term-to-maturity \( \tau \), we denote by \( d_{i,t}(\tau) \) the time-\( t \) loss differential between the squared forecast error of the LRSQ(3,2) model and the \( i \)th model,

\[
d_{i,t}(\tau) = \left( \hat{e}_{t+h,t}^{(3,2)}(\tau) \right)^2 - \left( \hat{e}_{t+h,t}^i(\tau) \right)^2.
\]

The null hypothesis of equal forecast performance is tested by the test statistic

\[
S = T^{1/2} \tilde{d}(\tau) \tilde{V}(\tau)^{-1/2},
\]

where \( \tilde{d}(\tau) \) is the sample mean of the loss differentials \( \tilde{d}(\tau) = T^{-1} \sum_{t=1}^T d_{i,t}(\tau) \), and \( \tilde{V}(\tau) \) is an estimate of the long-run variance of the loss differential. We apply HAC standard errors in \( \tilde{V}(\tau) \) based on the data-dependent bandwidth selection by Andrews (1991) by means of an AR(1) approximation and a Bartlett kernel. Under regularity conditions, the test statistic is asymptotically \( N(0,1) \) distributed under the null hypothesis.

The RMSFEs for expected absolute changes in yields are depicted in Table 9, and whether the difference for a given maturity compared to the LRSQ(3,2) model is statistically significant at the 5% and 1% level is indicated by * and ** respectively. In general, the LRSQ(3,2) model is superior to all models at the short end, but the advantage vanishes at the longer end. The results in the first and second column show that the performance of the LRSQ(3,0) and LRSQ(3,1) model are to a large degree on par. In sum, this shows that the flexibility of two USV factors does translate into a better out-of-sample performance, but the additional flexibility of a third USV factor does not provide further gains.

Based on the conditional volatility forecasts, we now consider the out-of-sample
equivalent of the volatility fitting error regression (2.4.4) defined as

\[ \text{rv}_{t+h}^{\text{vol}}(\tau) - \mu_t(\Delta y(\tau)) = \alpha + \sum_{i=1}^{3} \beta_i \text{PC}_{t,i}^{\text{rv}} + \epsilon_{t+h}. \] (2.5.1)

In order to account for autocorrelation in the regression residuals due to overlapping data in the forecasts, we apply HAC standard errors based on the approach previously mentioned.

The regression results are summarized in Table 10. In general, the table shows that in the USV models the null hypotheses of the joint zero restriction on all regression coefficients can be rejected at the 5% significance level. The ten-year maturity in the LRSQ(3, 2) model is the only exception. However, some of the individual coefficients are statistically significant. The fact that \( \alpha \) is estimated to be negative in most cases may be a result of the low volatility regime in the last part of the sample (cf. Figure 1).

The second main result from the table is that two rather than three USV factors are needed to incorporate ex-ante cross-sectional volatility information adequately. The LRSQ(3, 3) model (Panel E) has the stablest adjusted \( R^2 \) values ranging between 6% and 15%. Moreover, except for the very short maturities, the adjusted \( R^2 \) values of the LRSQ(3, 2) model in Panel D are lower than those obtained for the LRSQ(3, 0) model in Panel B and the LRSQ(3, 1) model in Panel C. Again as a benchmark, in Panel A, we also include the results obtained from EGARCH(1, 1) forecasts. The adjusted \( R^2 \) values for the longest maturities are below 5% for this model and lower than those of the LRSQ(3, 3) model. But the adjusted \( R^2 \) values for the intermediate maturities are notably larger in which case the LRSQ(3, 3) model is clearly superior.

In sum, in accordance with the in-sample findings, USV volatility is important to adequately capture ex-ante volatility information across the maturity spectrum with the LRSQ(3, 3) model being the preferred specification. When combining the results with the forecast accuracy, the conclusion regarding the preferred number of USV factor is less clear cut since the Diebold-Mariano test is not able to differentiate between the LRSQ(3, 2) and LRSQ(3, 3) model.

## 2.6 Extensions

In this section, we present further results. First, following Duffee (2011), we check for overfitting by means of Sharpe ratios. Second, we study whether the LRSQ(3, \( n \)) model can match the conditional first of moments.

### 2.6.1 Overfitting diagnostic

As noted by Duffee (2011), model-implied Sharpe ratios can be used as an informal model specification test. Highly volatile expected returns are hard to reconcile with
risk-based explanations of expected excess returns and Duffee (2011) argues that large Sharpe ratios indicate overfitting. The term structure modeling literature has come to conflicting conclusions on this issue. Duffee (2011) and Joslin, Singleton and Zhu (2011) find unreasonable high maximum conditional Sharpe ratios in ATSMs with more than three factors. In contrast, Adrian, Crump and Moench (2013) find that maximal Sharpe ratios obtained from their five-factor do not exceed one. Similarly, FLT obtain maximum conditional Sharpe ratios between 0.64 and 1.40 for their LRSQ$(m, s)$ model specifications.

We compute model-implied maximum conditional Sharpe ratio along with population means of conditional Sharpe ratios of individual bonds. Following FLT, model-implied maximum conditional Sharpe ratio in the LRSQ$(m, n)$ are given by

$$sr_t^{max} = \sqrt{\lambda^p_t \lambda^p_t},$$

where the market price of risk with respect to $P$ is

$$\lambda^p_t = \lambda^A_t + \delta_t,$$

with the market price of risk with respect to $A$ provided in (A.1.1). The population conditional Sharpe ratios for individual bonds are computed from simulating data of length 50,000 years from each model. As noted in Section 2.4.2, since we initially obtained slightly high Sharpe ratios, but the model fits were largely unaffected from constraining the Sharpe ratios during the estimation, we reestimated all models with the mean maximal Sharpe ratio constrained. Table 13 shows that the Sharpe ratio constrain is non-binding and the resulting mean population conditional Sharpe ratios and mean maximum conditional Sharpe ratios are economically plausible. The Sharpe ratios of individual bonds are hump shaped across the maturity spectrum and range between 0.25 and 0.41. As a reference, Duffee (2011) finds that his three-factor ATSM has a yearly mean maximum conditional Sharpe ratio of 1.07. The mean maximum conditional Sharpe ratios of the present models are slightly larger (close to 1.3) but still within a reasonable range. Thus, the conditional Sharpe ratios of the different estimated LRSQ$(3, n)$ models do not suggest symptoms of overfitting.

### 2.6.2 Campbell-Shiller regressions

Given our findings for the conditional second moment of yields, a natural question to ask is whether LRSQ models can match the conditional first moment of yields. We examined the performance in fitting in-sample and out-of-sample yields, i.e., in matching conditional expectations of current and future yields. While a sophisticated analysis is beyond the scope of this paper, we briefly examine the two linear projections of yields (LPY) tests of Dai and Singleton (2002) which is a popular assessment of the ability of term structure models to match the conditional expectations of future yields.
The LPY tests are essentially a test of the expectations hypothesis (EH), saying that the yield on a \(\tau\)-maturity bond grows at the same rate as the term spread widens. The LPY(i) test examines whether a model can reproduce sample slope coefficients from the Campbell and Shiller (1991) (CS) regression; that is, a regression of yield changes onto the scaled yield spread. The LPY(ii) test asks whether the EH holds for risk-adjusted yields, i.e., actual yields adjusted for model-implied term premia. This is done by testing whether the loadings in the risk-adjusted CS regression equal a flat line at unity across the maturity spectrum.

Denoting the time-\(t\) yield on a zero-coupon bond with term-to-maturity \(m\) by \(y^j_t\), the CS regressions is given by

\[
y^j_{t+m} - y^j_t = \alpha^j_1 + \beta^j_1 \frac{m}{j-m} (y^j_t - y^m_t) + u^1_{t,j}, \quad u^1_{t,j} \sim \text{IID}(0, \text{var}(u^1_{t,j})), \quad (2.6.1)
\]

for \(j = m+1, m+2, \ldots, K\). Furthermore, the risk-adjusted CS regression is given by

\[
y^j_{t+m} - y^j_t - (c^j_{t+m} - c^j_t) + \frac{m}{j-m} \theta^j_{t-m} = \alpha^2_2 + \beta^2_2 \frac{m}{j-m} (y^j_t - y^m_t) + u^2_{t,j}, \quad u^2_{t,j} \sim \text{IID}(0, \text{var}(u^2_{t,j})), \quad (2.6.2)
\]

where \(c^j_t = y^j_t - \mathbb{E}_t^P[\int_t^j r_s \, ds]\) is the term premium in \(y^j_t\) and \(\theta^j_t = f^j_t - \mathbb{E}_t^P[r_{t+j} \, ds]\) is the term premium in the forward rate \(f_{t,j} = -\log(B(t, j+1)/B(t,j))\).

For the LPY(i) test, we simulate 2,000 samples of yields in the 1-year to 10-year maturity range at 1-year intervals of length 293 months (the same length as the length of actual data) from their distribution. For each sample, based on the LRSQ(3, \(n\)) model parameter estimates, we estimate (2.6.1) setting \(m = 1\) year, where, then, \(\beta^1_j\) is obtained as the mean of the individual estimates. While we use actual yields in the regression equation (2.6.2) in the LPY(ii) test (cf. CS), the term premiums are obtained by simulation using 2,000 paths of the short rate (1,000 of these are antithetic paths to reduce variance).

We display the results of the LPY(ii) test in the left panel of Figure 6. It shows that none of the models can match the downward sloping pattern of the sample slope coefficients, obtained using actual yields. Similarly, the right panel shows that all model specifications fail to match the LPY(ii) test, where for well-specified model-implied risk premiums the risk-adjusted loadings should be equal to one across the maturity dimension. These results are consistent with the findings for the related (square-root) \(A_3(3)\) model in Dai and Singleton (2002). They attribute the failure to too constrained risk premiums and show that the market price of risk specification in \(A_m(3)\) models has significant effect on the model’s abilities to match LPY. In particular, they show that a market prices of risk specification that depends on the risk factors directly and not only on factor volatilities (e.g. an essentially-affine market price of risk specification) can better match LPY. Following this recipe, working on the
market prices of risk specification in the LRSQ models (i.e., the integrand \( \delta_t \) in the Radon-Nikodym density process (2.2.4) may be helpful to match the conditional expectations of future yields.

## 2.7 Robustness

Given that conditional volatility is unobserved, it might be the case that using expected absolute one-month changes in yields as proxy drives the out-of-sample results obtained in Section 2.5. Thus, in order to test the robustness of the predictive performance of the LRSQ(3, \( n \)) models, we now review the out-of-sample volatility fit regression (2.5.1) and proxy conditional volatility by realized volatility obtained from the simulated distribution of \( \hat{y}_{t+h}(\tau) \). Here, we focus again on one-month-ahead forecasts of volatility and simulate 10,000 discrete trajectories for the factor process between \( t \) and \( t+h \) using the same approach as described in Section 2.5. Along each trajectory of \( \{ \hat{X}_i \}_{i=t}^{i=t+h} \), we obtain from (2.2.8) the daily model-implied yield forecast \( \hat{y}_{i,t}^{d}\) for the \( i \)th day in month \( t \) for \( i = 0, \ldots, N_t \), where \( N_t \) denotes the number of trading days of a given month \( t \). The time-\( t \) realized volatility forecast of the yield with term-to-maturity \( \tau \) is then given by

\[
rv_{t+h,t}^{vol,m}(\tau) = \left( \sum_{i=1}^{N_t} \left( \hat{y}_{i,t}^{d} - \hat{y}_{i-1,t}^{d} \right)^2 \right)^{1/2},
\]

where their respective mean values constitute the forecast values. Based on the same estimation windows and out-of-sample windows as described in Section 2.5, Table 11 reports the RMSFEs between \( rv_{t+h,t}^{vol}(\tau) \) and \( rv_{t+h,t}^{vol,m}(\tau) \), and Table 12 depicts the results from the out-of-sample volatility fit regression. The fact that the parameter estimates used to simulate the daily yields are obtained from data observed at a monthly frequency constitutes a weakness of this approach. However, the results that emerge from both tables confirm the conclusions based on expected absolute changes in yields from Section 2.5. Firstly, the forecasts of the LRSQ(3,0) model are significantly less accurate than the forecasts obtained from the USV models. Secondly, the LRSQ(3,2) model does neither have a consistent advantage over the LRSQ(3,1) model nor over the LRSQ(3,3) model in terms of a higher forecast accuracy. Thirdly, among the different LRSQ(3, \( n \)) model specifications, it is again the LRSQ(3,3) model that is able to incorporate most ex-ante cross-sectional variance information with adjusted \( R^2 \) values of similar magnitude than those obtained for expected absolute changes in yields.

The results that emerge from both tables confirm previous conclusions from Section 2.5 but less pronounced. Firstly, the forecasts of the LRSQ(3,0) model on average less accurate than the forecasts obtained from the LRSQ(3,2) model. Secondly, the LRSQ(3,2) model does not have a consistent advantage over the LRSQ(3,3) model but a slight advantage over the LRSQ(3,1) model in terms of higher forecast accuracy.
Thirdly, it is again the LRSQ(3,3) model that is able to incorporate most ex-ante cross-sectional variance information in terms of lowest adjusted $R^2$ values, which though are slightly larger than for expected absolute changes in yields.

### 2.8 Concluding remarks

We examined the ability of the linear-rational square-root (LRSQ) model developed by Filipović et al. (2017) to capture conditional volatility in the US Treasury bond yield market. We did this by means of a variety of in-sample and out-of-sample regression analyses as well a comparison between model-implied and model-free measures of conditional volatility. Their main implications are: First, it is important to augment the cross-section of yields with informative second-moment data in order to adequately estimate conditional volatility. In doing so, the Unscented Kalman Filter in conjunction with quasi maximum likelihood is able to infer the unspanned stochastic volatility (USV) factors – despite being computationally simpler than prior estimation methodologies of the USV term structure literature such as Markov chain Monte Carlo. Secondly, an application to empirical data showed that those LRSQ model specifications that exhibit USV are able to break the tension noted for the popular affine term structure models from matching the cross-section dimension and time-series properties of US Treasury bond yields simultaneously. In this regard, we documented a reasonable in-sample performance of the model with only a single USV factor. But, we found, the additional flexibility of the model specifications with two or three USV factors provides further gains for capturing conditional volatility. This superior in-sample performance did, however, not translate into more accurate out-of-sample forecasts compared to the simpler model with a single USV factor. Finally, the factor structure underlying conditional volatility resembles the level-slope-curvature embedded in yields.

### Acknowledgements

I am grateful to Bent Jesper Christensen, Damir Filipović, Martin M. Andreasen, Daniel Borup, Thomas M. Jensen, and the participants at research seminars at Aarhus University for helpful comments and suggestions, and to the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation, DNRF78), the Dale T. Mortensen Centre, Aarhus University, and the Danish Social Science Research Council for research support. Some of this research was performed while I was visiting the Department of Economics, University of Pennsylvania, and the generosity and hospitality of the department is gratefully acknowledged.


2.9 References


A.1 Appendix

A.1.1 Yield expression in the LRSQ model

The zero-coupon bond price formula (2.2.2) implies that zero-coupon bond yields
\[ y(\tau, z) = -\frac{1}{\tau} \log F(\tau, z) \]
are given by
\[
y(\tau, z) = -\frac{1}{\tau} \log \left( \exp \left( -\alpha \tau \right) \frac{\phi + \psi'(\theta z + \exp(-\tau \kappa)(z - \theta z))}{\phi + \psi' z} \right)
\]
\[
= a - \frac{1}{\tau} \log \left( \exp \left( -\alpha \right) \right)
\]
\[
= -\frac{1}{\tau} \log \left( \frac{\phi + \psi'(\theta z + \exp(-\tau \kappa)(z - \theta z))}{\phi + \psi' z} \right) .
\]

A.1.2 Instantaneous diffusion matrix of yields in the LRSQ model

FLT show that the dynamics of the zero-coupon bond price \( B(t, T) = F(T - t, z) \) with
term-of-maturity \( T \) at time \( t \) are given by
\[
\frac{dB(t, T)}{B(t, T)} = \left( r_t + \nu_{t,T}(Z_t, U_t)\lambda_t^A \right) dt + \nu_{t,T}(Z_t, U_t)'dW_t^A,
\]
where \( r_t \) is the short rate at time \( t \), \( \lambda_t^A \) is the market price of risk given by
\[
\lambda_t^A = -\frac{\sigma(Z_t, U_t)\psi}{\phi + \psi' Z_t},
\]
(A.1.1)
the dispersion function \( \nu_{t,T}(Z_t, U_t) \) is given by
\[
\nu_{t,T}(Z_t, U_t) = \frac{\sigma(Z_t, U_t)'\nabla_z F(T - t, Z_t)}{F(T - t, Z_t)},
\]
(A.1.2)
and
\[
\sigma(Z, U) = (I_m, A) \text{ diag} \left( \sigma_1 \sqrt{Z_1 - U_1}, \ldots, \sigma_{m+n} \sqrt{U_n} \right) .
\]
(A.1.3)
Since the function
\[
g(x, t) = -\frac{\log x}{t}
\]
applied to the zero-coupon bond price \( B(t, T) \) and term-to-maturity \( \tau = T - t \) corresponds to the yield expression
\[
g(B(t, T), T - t) = -\frac{\log B(t, T)}{T - t} = y(\tau, z).
\]
Itô’s lemma with \( g(x, t) \) allows us to derive the dynamics of \( y(t, T) \). That is, since
\[
\frac{\partial g(x, T - t)}{\partial T - t} = -\frac{1}{(T - t)^2} \log x, \quad \frac{\partial g(x, T - t)}{\partial x} = -\frac{1}{T - t} \frac{1}{x}, \quad \frac{\partial^2 g(x, T - t)}{\partial x^2} = \frac{1}{T - t} \frac{1}{x^2},
\]
it follows from Itô’s lemma that
\[ dy(t, z) = \mu^Y_{t,T} dt + \sigma^Y_{t,T} dW^\lambda_t, \]
where using that the drift of the zero-coupon bond price is
\[ \mu_{Bt} = r_t + \nu_t, \]
the drift \( \mu^Y_{t,T} \) and the volatility rate \( \sigma^Y_{t,T} \) of the yield process are given by (suppressing dependencies when there is no ambiguity)
\[ \mu^Y_{t,T} = -\frac{1}{(T-t)^2} \log B - \frac{1}{T-t}(r + \nu' \lambda^\lambda) B + \frac{1}{2 T-t} \nu'^2 B^2 - \frac{1}{B^2} \left( -\frac{1}{T-t} \log B - (r + \nu' \lambda^\lambda) + \frac{1}{2} \nu'^2 \right) \frac{1}{T-t}, \]
\[ \sigma^Y_{t,T} = -\frac{1}{B} \nu' \frac{1}{T-t} B = -\frac{\nu'}{T-t}. \]
This implies that the instantaneous diffusion matrix \( a^Y_{t,T} \) of the yield is
\[ a^Y_{t,T} = \sigma^Y_{t,T} \sigma^Y_{t,T}' = \nu_{t,T}(Z_t, U_t)' \nu_{t,T}(Z_t, U_t) \frac{1}{(T-t)^2}, \]
where \( \nu_{t,T}(Z_t, U_t) \) is given by (2.2.9). From Equation (A2) in FLT it follows that
\[ \nabla z F(T-t, z) = \exp(-\alpha(T-t)) \frac{\phi + \psi' z}{\phi + \psi' z} \left[ \exp(-\kappa'(T-t)) \psi - \exp(\alpha(T-t)) F(T-t, z) \psi \right] \]
so that
\[ \frac{\nabla z F(T-t, z)}{F(T-t, z)} = \exp(-\alpha(T-t)) \frac{\left( \exp(-\kappa'(T-t)) \psi \right)}{\phi + \psi' z} \frac{\left( \exp(\alpha(T-t)) \psi \right)}{F(T-t, z)} - \exp(\alpha(T-t)) \psi. \]

A.1.3 The Unscented Kalman Filter

The UKF was originally developed by Julier and Uhlmann (1997) and subsequently extended to a more robust square-root UKF by van der Merwe and Wan (2002). The advantage of the square-root UKF is that it circumvents potential numerical instability induced by a negative semidefinite covariance matrix in the recursion by running the filter for the square-root matrix of the covariance matrix. In what follows, for a given covariance matrix \( \Sigma_{ij} \) we denote by \( S_{ij} \) the square-root matrix of \( \Sigma_{ij} = S_{ij} S_{ij}' \).

The key ingredient of the UKF is the unscented transform which is a method to approximate the untransformed mean \( \mu_x \) and covariance matrix \( \Sigma_{xx} \) of a given \( N \)-dimensional random variable \( x \) by a set of so-called sigma points in terms of
\[ \mathcal{X}_0 = \mu_x, \]
\[ \mathcal{X}_i = \mu_x + (\sqrt{(N+\zeta) \Sigma_{xx}})_i, \quad i = 1, \ldots, N, \]
\[ \mathcal{X}_i = \mu_x - (\sqrt{(N+\zeta) \Sigma_{xx}})_i, \quad i = N+1, \ldots, 2N, \]
with weights

\[ w_0^m = \frac{\zeta}{N + \zeta}, \]
\[ w_0^c = \frac{\zeta}{N + \zeta} + (1 - \rho^2 + \gamma), \]
\[ w_i^m = w_i^c = \frac{1}{2(N + \zeta)}, \quad i = 1, \ldots, 2N, \]

where \( (\sqrt{(N + \zeta)\Sigma_{xx})_i} \) denotes the \( i \)th column of the matrix \( \sqrt{(N + \zeta)\Sigma_{xx}} \). For \( \rho, \alpha > 0 \) the scaling parameter \( \zeta = \rho^2(N + \alpha) - N \) determines the spread of the sigma points around the mean. The parameter \( \rho \) is thought to minimize higher-order effects, \( \alpha \) ensures positive semi-definiteness of the covariance matrix, and \( \gamma \) is related to the distribution of \( x \), where \( \gamma = 2 \) is optimal for the Gaussian distribution. The UKF uses the unscented transform to approximate the covariance matrix of the states and the observations \( \Sigma_{x|y(t|t-1)} \) and the covariance matrix of the observations \( \Sigma_{y|y(t|t-1)} \).

The UKF is applied to the \( 2N \)-dimensional state vector augmented by the state white noise \( \xi_t = \eta_t(\sqrt{Q_t})^{-1} \),

\[ x_a^t = \left[ x_t', \xi_t' \right]', \]

where \( Q_t \) is the covariance matrix given in (2.3.3). The recursion is initialized using the unconditional mean and variance

\[ x_{0|0} = \left[ E[x_t]' \right], \quad S_{xx(0|0)} = \sqrt{Q}, \quad S_{\xi\xi(0|0)} = I_N. \]

Consider period \( t - 1 \) and suppose that the state mean vector update \( x_{t-1|t-1} \) and the square-root matrix of its variance-covariance matrix \( S_{t-1|t-1} \) have been obtained. The \( 2N + 1 \) sigma points are computed as

\[ \mathcal{X}_{t-1|t-1} = \left[ x_{t-1|t-1}', x_{t-1|t-1}' + \sqrt{(N_a + \zeta)S_{t-1|t-1}}, x_{t-1|t-1}' - \sqrt{(N_a + \zeta)S_{t-1|t-1}} \right], \]

where \( S_{t-1|t-1} \) is a block matrix with upper right block given by \( S_{xx(t-1|t-1)} \) and lower left block given \( S_{\xi\xi(t-1|t-1)} \) and zeros otherwise. The prediction step is given by projecting the sigma points ahead through the state transition equation function given in (2.3.2),

\[ \mathcal{X}_{t|t-1} = G(\mathcal{X}_{t-1|t-1}; \Theta) + \sqrt{Q_t}\mathcal{X}_{t-1|t-1}. \]

The predicted state mean vector is computed as the weighted average of the transformed sigma points

\[ x_{t|t-1} = \sum_{i=0}^{2N} w_i^m \mathcal{X}_{i,t|t-1}. \]
The predicted conditional observation mean vector is computed from the observation function

\[
\mathcal{Y}_{t|t-1} = H \left( \mathcal{X}_{t|t-1}; \Theta \right),
\]

\[
y_{t|t-1} = \sum_{i=0}^{2N} w_i^m \mathcal{Y}_{i, t|t-1}.
\]

For the square-root step, qr \{ A \} is the short-hand notation for the orthogonal matrix \( Q \) of the QR-decomposition of a given matrix \( A = QR \). Furthermore, for \( R = \text{chol}(A) \) being the Cholesky factor of a given matrix \( A \) such that \( A = R'R \), we use the short-hand notation \( \text{chol}(R, x, \pm z) \) for the upper triangular Cholesky factor of \( A \pm zx'x \).

If \( x \) is a matrix and not a vector, then the update is performed using the columns of \( x \) sequentially. Based on those definitions the square-root step is 23

\[
S_{xx}(t|t-1) = \text{chol} \left[ \text{qr} \left( \sqrt{\sum_{i=0}^{2N} \mathcal{Y}_{i, t|t-1} - x_{t|t-1}} \right) \mathcal{X}_{t|t-1} - x_{t|t-1}, w_0^c \right],
\]

\[
S_{xy}(t|t-1) = \text{chol} \left[ \text{qr} \left( \sqrt{\sum_{i=0}^{2N} \mathcal{Y}_{i, t|t-1} - y_{t|t-1}} \right) \mathcal{X}_{t|t-1} - y_{t|t-1}, w_0^c \right],
\]

where \( U \) is the covariances matrix from (2.3.4). The cross-covariance matrix is

\[
\Sigma_{xy}(t|t-1) = \sum_{i=0}^{2N} w_i^c \left[ \mathcal{Y}_{i, t|t-1} - x_{t|t-1} \right] \left[ \mathcal{Y}_{i, t|t-1} - y_{t|t-1} \right]' .
\]

The time \( t \) update step is given by

\[
x_{t|t} = x_{t|t-1} + K_t (y_{t} - y_{t|t-1}),
\]

\[
\Sigma_{xx}(t|t) = \Sigma_{xx}(t|t-1) - K_t \Sigma_{xy}(t|t-1) K_t',
\]

\[
S_{xx}(t|t) = \text{chol} \left[ S_{xx}(t|t-1) K_t S_{xy}(t|t-1), -1 \right],
\]

where the Kalman gain \( K_t \) can be computed by means of the back substitution operator \( '/' \). 24

\[
K_t = \left( \Sigma_{xy}(t|t-1)/S_{yy}(t|t-1) \right) / S_{yy}(t|t-1).
\]

Finally, the log-likelihood function is given by

\[
\log \ell (y_1, \ldots, y_T; \Theta) = -\frac{1}{2} \sum_{t=1}^{T} \left( n \log (2\pi) + \log |\Sigma_{yy}(t|t-1)|
\right.

\[
+ \left( y_t - y_{t|t-1} \right) \Sigma_{yy}(t|t-1)^{-1} \left( y_t - y_{t|t-1} \right)' \right),
\]

23 This decomposing provides the Cholesky factor \( S_{xx}(t|t-1) \) of the predicted covariance matrices \( \Sigma_{xx}(t|t-1) = S_{xx}(t|t-1) S_{xx}(t|t-1)' \) and \( \Sigma_{yy}(t|t-1) = S_{yy}(t|t-1) S_{yy}(t|t-1)' \) and ensures positive semidefiniteness of the covariance matrices by construction.

24 If the Cholesky factor \( S \) of a given matrix \( A \) is known, then the least squares solution to the problem \( Ax = b \) can be computed efficiently using back substitution without the need for matrix inversion.
where $T$ denotes the length of the sample size and $n$ is the number of terms to maturities in $y_t$. The QML estimator $\hat{\Theta}_{QML}$ of $\Theta$ is given by the set of parameter values that maximizes the log-likelihood function, $\hat{\Theta}_{QML} = \arg \min_{\Theta} \log \ell (y_1, \ldots, y_T; \Theta)$. 
A.2 Tables

Table 1: This table summarizes the Monte Carlo results for the UKF parameter estimation of the LRSQ(3, 1) and LRSQ(3, 2) model based on 500 simulated data sets for monthly observations of yields with maturities 0.5, 1, 2, 3, 5, 7, and 10 years and their realized variances. Measurement errors with standard deviation of 3.5 basis points for yields and 0.8 basis points for realized variances was added to the observations. The \( t \)-value is for the significance of the difference in parameter estimates. Each data set is based on 300 observations corresponding to 25 years.

<table>
<thead>
<tr>
<th></th>
<th>True value</th>
<th>Median</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>( t )-value</th>
<th>True value</th>
<th>Median</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>( t )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: LRSQ(3,1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \kappa_{11} )</td>
<td>0.214</td>
<td>0.213</td>
<td>0.213</td>
<td>0.004</td>
<td>0.135</td>
<td>0.159</td>
<td>0.157</td>
<td>0.156</td>
<td>0.021</td>
<td>0.167</td>
</tr>
<tr>
<td>( \kappa_{22} )</td>
<td>1.153</td>
<td>1.153</td>
<td>1.154</td>
<td>0.013</td>
<td>-0.060</td>
<td>1.308</td>
<td>1.308</td>
<td>1.277</td>
<td>0.155</td>
<td>0.202</td>
</tr>
<tr>
<td>( \kappa_{33} )</td>
<td>0.101</td>
<td>0.101</td>
<td>0.102</td>
<td>0.002</td>
<td>-0.139</td>
<td>0.126</td>
<td>0.126</td>
<td>0.133</td>
<td>0.061</td>
<td>-0.112</td>
</tr>
<tr>
<td>( \kappa_{21} )</td>
<td>-0.310</td>
<td>-0.309</td>
<td>-0.309</td>
<td>0.004</td>
<td>-0.089</td>
<td>-0.257</td>
<td>-0.258</td>
<td>-0.289</td>
<td>0.146</td>
<td>0.215</td>
</tr>
<tr>
<td>( \kappa_{32} )</td>
<td>-1.259</td>
<td>-1.259</td>
<td>-1.261</td>
<td>0.015</td>
<td>0.157</td>
<td>-1.408</td>
<td>-1.410</td>
<td>-1.395</td>
<td>0.184</td>
<td>-0.068</td>
</tr>
<tr>
<td>( \theta^1_z )</td>
<td>0.356</td>
<td>0.356</td>
<td>0.358</td>
<td>0.009</td>
<td>-0.214</td>
<td>0.468</td>
<td>0.485</td>
<td>0.520</td>
<td>0.147</td>
<td>-0.352</td>
</tr>
<tr>
<td>( \theta^2_z )</td>
<td>0.394</td>
<td>0.394</td>
<td>0.394</td>
<td>0.024</td>
<td>0.008</td>
<td>0.522</td>
<td>0.528</td>
<td>0.620</td>
<td>0.324</td>
<td>-0.303</td>
</tr>
<tr>
<td>( \theta^3_z )</td>
<td>1.364</td>
<td>1.366</td>
<td>1.381</td>
<td>0.054</td>
<td>-0.307</td>
<td>1.225</td>
<td>1.225</td>
<td>1.257</td>
<td>0.378</td>
<td>-0.084</td>
</tr>
<tr>
<td>( \theta^{10^3}_u )</td>
<td>0.010</td>
<td>0.011</td>
<td>0.037</td>
<td>0.065</td>
<td>-0.556</td>
<td>0.010</td>
<td>0.010</td>
<td>0.016</td>
<td>0.019</td>
<td>-0.852</td>
</tr>
<tr>
<td>( \theta^2_u )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.033</td>
<td>0.023</td>
<td>0.032</td>
<td>0.030</td>
<td>0.029</td>
</tr>
<tr>
<td>( \sigma_1 )</td>
<td>0.295</td>
<td>0.296</td>
<td>0.316</td>
<td>0.058</td>
<td>-0.374</td>
<td>0.255</td>
<td>0.254</td>
<td>0.258</td>
<td>0.055</td>
<td>-0.057</td>
</tr>
<tr>
<td>( \sigma_2 )</td>
<td>0.350</td>
<td>0.350</td>
<td>0.350</td>
<td>0.006</td>
<td>-0.003</td>
<td>0.242</td>
<td>0.232</td>
<td>0.221</td>
<td>0.037</td>
<td>0.568</td>
</tr>
<tr>
<td>( \sigma_3 )</td>
<td>0.187</td>
<td>0.186</td>
<td>0.186</td>
<td>0.004</td>
<td>0.232</td>
<td>0.145</td>
<td>0.131</td>
<td>0.121</td>
<td>0.036</td>
<td>0.656</td>
</tr>
<tr>
<td>( \sigma_4 )</td>
<td>1.410</td>
<td>1.411</td>
<td>1.408</td>
<td>0.039</td>
<td>0.044</td>
<td>1.317</td>
<td>1.308</td>
<td>1.260</td>
<td>0.197</td>
<td>0.292</td>
</tr>
<tr>
<td>( \delta_5 )</td>
<td>1.117</td>
<td>1.117</td>
<td>1.116</td>
<td>0.128</td>
<td>0.010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \delta_1 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>0.639</td>
<td>0.636</td>
<td>0.555</td>
<td>0.399</td>
<td>0.212</td>
<td>0.075</td>
<td>0.075</td>
<td>0.076</td>
<td>0.008</td>
<td>-0.151</td>
</tr>
<tr>
<td>( \delta_3 )</td>
<td>-1.040</td>
<td>-1.040</td>
<td>-1.066</td>
<td>0.568</td>
<td>0.046</td>
<td>-0.216</td>
<td>-0.212</td>
<td>-0.190</td>
<td>0.750</td>
<td>-0.155</td>
</tr>
<tr>
<td>( \delta_4 )</td>
<td>-0.504</td>
<td>-0.504</td>
<td>-0.517</td>
<td>0.243</td>
<td>0.057</td>
<td>-0.837</td>
<td>-0.837</td>
<td>-0.998</td>
<td>0.871</td>
<td>0.186</td>
</tr>
<tr>
<td>( \delta_5 )</td>
<td>-0.209</td>
<td>-0.209</td>
<td>-0.276</td>
<td>0.302</td>
<td>0.221</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Table 2:** This table reports summary statistics of monthly average yields and realized variance (RV) (see (2.4.1) and (2.4.2) for an exact specification) over the (in-sample) period 1987-2011. Maturities are in years, yields are in percent, and realized variances are in basis points.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.</td>
</tr>
<tr>
<td>0.5</td>
<td>4.08</td>
<td>2.40</td>
</tr>
<tr>
<td>1</td>
<td>4.23</td>
<td>2.41</td>
</tr>
<tr>
<td>2</td>
<td>4.50</td>
<td>2.33</td>
</tr>
<tr>
<td>3</td>
<td>4.74</td>
<td>2.23</td>
</tr>
<tr>
<td>5</td>
<td>5.14</td>
<td>2.03</td>
</tr>
<tr>
<td>7</td>
<td>5.46</td>
<td>1.87</td>
</tr>
<tr>
<td>10</td>
<td>5.83</td>
<td>1.71</td>
</tr>
</tbody>
</table>
Table 3: This table reports the maximized log-likelihood value and the parameter estimates of four LRSQ(3, \(n\)) model specifications with standard errors in parenthesis.

<table>
<thead>
<tr>
<th>(\kappa_{11})</th>
<th>(\kappa_{22})</th>
<th>(\kappa_{33})</th>
<th>(\kappa_{21})</th>
<th>(\kappa_{32})</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRSQ(3,0) 0.2375</td>
<td>1.2204</td>
<td>0.0832</td>
<td>-0.3182</td>
<td>-1.3167</td>
</tr>
<tr>
<td>(0.0005)</td>
<td>(0.0027)</td>
<td>(0.0004)</td>
<td>(0.0014)</td>
<td>(0.0028)</td>
</tr>
<tr>
<td>LRSQ(3,1) 0.2138</td>
<td>1.1533</td>
<td>0.1014</td>
<td>-0.3096</td>
<td>-1.2590</td>
</tr>
<tr>
<td>(0.0003)</td>
<td>(0.0007)</td>
<td>(0.0003)</td>
<td>(0.0003)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>LRSQ(3,2) 0.1591</td>
<td>1.3078</td>
<td>0.1258</td>
<td>-0.2574</td>
<td>-1.4078</td>
</tr>
<tr>
<td>(0.0003)</td>
<td>(0.0010)</td>
<td>(0.0003)</td>
<td>(0.0006)</td>
<td>(0.0010)</td>
</tr>
<tr>
<td>LRSQ(3,3) 0.1389</td>
<td>1.4107</td>
<td>0.2234</td>
<td>-0.1568</td>
<td>-1.4693</td>
</tr>
<tr>
<td>(0.0004)</td>
<td>(0.0010)</td>
<td>(0.0005)</td>
<td>(0.0007)</td>
<td>(0.0010)</td>
</tr>
</tbody>
</table>

| \(\theta_1\) | \(\theta_2\) | \(\theta_3\) | \(\theta_4\) | \(\theta_5\) | \(\theta_6\) |
|----------------|----------------|----------------|----------------|----------------|
| LRSQ(3,0) 0.3225 | 0.3474 | 1.5033 | 0.3225 | 0.3474 | 1.5033 |
| (0.0009) | (0.0057) | (0.0056) | (0.0009) | (0.0057) | (0.0056) |
| LRSQ(3,1) 0.3557 | 0.3943 | 1.3640 | 0.3557 | 0.3943 | 1.3640 |
| (0.0004) | (0.0008) | (0.0012) | (0.0004) | (0.0008) | (0.0012) |
| LRSQ(3,2) 0.4685 | 0.5217 | 1.2255 | 0.4685 | 0.5217 | 1.2255 |
| (0.0004) | (0.0012) | (0.0032) | (0.0004) | (0.0012) | (0.0032) |
| LRSQ(3,3) 2.5458 | 0.6457 | 3.4069 | 0.6457 | 3.4069 | 0.6457 |
| (0.0006) | (0.0013) | (0.0033) | (0.0006) | (0.0013) | (0.0033) |

<table>
<thead>
<tr>
<th>(\sigma_1)</th>
<th>(\sigma_2)</th>
<th>(\sigma_3)</th>
<th>(\sigma_4)</th>
<th>(\sigma_5)</th>
<th>(\sigma_6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRSQ(3,0) 0.4905</td>
<td>0.3921</td>
<td>0.2024</td>
<td>0.4905</td>
<td>0.3921</td>
<td>0.2024</td>
</tr>
<tr>
<td>(0.0161)</td>
<td>(0.0082)</td>
<td>(0.0044)</td>
<td>(0.0161)</td>
<td>(0.0082)</td>
<td>(0.0044)</td>
</tr>
<tr>
<td>LRSQ(3,1) 0.2946</td>
<td>0.3500</td>
<td>0.1868</td>
<td>0.2946</td>
<td>0.3500</td>
<td>0.1868</td>
</tr>
<tr>
<td>(0.0034)</td>
<td>(0.0032)</td>
<td>(0.0026)</td>
<td>(0.0034)</td>
<td>(0.0032)</td>
<td>(0.0026)</td>
</tr>
<tr>
<td>LRSQ(3,2) 0.2551</td>
<td>0.2417</td>
<td>0.1453</td>
<td>0.2551</td>
<td>0.2417</td>
<td>0.1453</td>
</tr>
<tr>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td>(0.0008)</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td>(0.0008)</td>
</tr>
<tr>
<td>LRSQ(3,3) 0.5152</td>
<td>0.4457</td>
<td>0.1107</td>
<td>0.5152</td>
<td>0.4457</td>
<td>0.1107</td>
</tr>
<tr>
<td>(0.0012)</td>
<td>(0.0014)</td>
<td>(0.0027)</td>
<td>(0.0012)</td>
<td>(0.0014)</td>
<td>(0.0027)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(\delta_1)</th>
<th>(\delta_2)</th>
<th>(\delta_3)</th>
<th>(\delta_4)</th>
<th>(\delta_5)</th>
<th>(\delta_6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRSQ(3,0) 0</td>
<td>0.6386</td>
<td>-1.1589</td>
<td>0</td>
<td>0.6386</td>
<td>-1.1589</td>
</tr>
<tr>
<td>(0.1673)</td>
<td>(0.1820)</td>
<td>(0.1673)</td>
<td>(0.1820)</td>
<td>(0.1673)</td>
<td>(0.1820)</td>
</tr>
<tr>
<td>LRSQ(3,1) 0</td>
<td>0.6390</td>
<td>-1.0400</td>
<td>0.6390</td>
<td>-1.0400</td>
<td>0.6390</td>
</tr>
<tr>
<td>(0.0258)</td>
<td>(0.0296)</td>
<td>(0.0258)</td>
<td>(0.0296)</td>
<td>(0.0258)</td>
<td>(0.0296)</td>
</tr>
<tr>
<td>LRSQ(3,2) 0</td>
<td>0.0751</td>
<td>-0.2159</td>
<td>-0.0751</td>
<td>-0.2159</td>
<td>-0.0751</td>
</tr>
<tr>
<td>(0.0053)</td>
<td>(0.0043)</td>
<td>(0.0053)</td>
<td>(0.0043)</td>
<td>(0.0053)</td>
<td>(0.0043)</td>
</tr>
<tr>
<td>LRSQ(3,3) 0</td>
<td>0.0741</td>
<td>-0.0182</td>
<td>-0.0741</td>
<td>-0.0182</td>
<td>-0.0741</td>
</tr>
<tr>
<td>(0.0078)</td>
<td>(0.0109)</td>
<td>(0.0078)</td>
<td>(0.0109)</td>
<td>(0.0078)</td>
<td>(0.0109)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(\sigma_y \times 10^4)</th>
<th>(\sigma_{\nu \epsilon} \times 10^4)</th>
<th>(\alpha)</th>
<th>(\ell)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRSQ(3,0) 3.3755</td>
<td>0.8536</td>
<td>0.0766</td>
<td>27.841</td>
</tr>
<tr>
<td>(0.0173)</td>
<td>(0.0022)</td>
<td>(0.0173)</td>
<td>(0.0022)</td>
</tr>
<tr>
<td>LRSQ(3,1) 3.4619</td>
<td>0.6512</td>
<td>0.0760</td>
<td>28.067</td>
</tr>
<tr>
<td>(0.0037)</td>
<td>(0.0032)</td>
<td>(0.0037)</td>
<td>(0.0032)</td>
</tr>
<tr>
<td>LRSQ(3,2) 3.5826</td>
<td>0.0036</td>
<td>0.0743</td>
<td>28.747</td>
</tr>
<tr>
<td>(0.0163)</td>
<td>(0.0054)</td>
<td>(0.0163)</td>
<td>(0.0054)</td>
</tr>
<tr>
<td>LRSQ(3,3) 3.6694</td>
<td>0.0015</td>
<td>0.0753</td>
<td>29.243</td>
</tr>
<tr>
<td>(0.0092)</td>
<td>(0.0081)</td>
<td>(0.0092)</td>
<td>(0.0081)</td>
</tr>
</tbody>
</table>
Table 4: This table reports summary statistics for in-sample yield fits. For each model, the table reports the residual means and RMSEs in basis points for seven maturities in years.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>LRSQ(3,0) Mean</th>
<th>RMSE</th>
<th>LRSQ(3,1) Mean</th>
<th>RMSE</th>
<th>LRSQ(3,2) Mean</th>
<th>RMSE</th>
<th>LRSQ(3,3) Mean</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.28</td>
<td>8.42</td>
<td>-0.30</td>
<td>8.23</td>
<td>-1.74</td>
<td>9.10</td>
<td>-1.34</td>
<td>9.77</td>
</tr>
<tr>
<td>1</td>
<td>0.20</td>
<td>9.30</td>
<td>-0.51</td>
<td>9.11</td>
<td>-0.10</td>
<td>9.09</td>
<td>-0.40</td>
<td>9.19</td>
</tr>
<tr>
<td>2</td>
<td>-0.61</td>
<td>7.42</td>
<td>-1.23</td>
<td>7.25</td>
<td>0.98</td>
<td>6.79</td>
<td>-0.20</td>
<td>7.31</td>
</tr>
<tr>
<td>3</td>
<td>-1.55</td>
<td>6.49</td>
<td>-1.98</td>
<td>6.07</td>
<td>0.90</td>
<td>5.70</td>
<td>-0.81</td>
<td>7.47</td>
</tr>
<tr>
<td>5</td>
<td>-1.71</td>
<td>5.83</td>
<td>-1.91</td>
<td>5.36</td>
<td>1.09</td>
<td>5.27</td>
<td>-1.14</td>
<td>7.63</td>
</tr>
<tr>
<td>7</td>
<td>-0.70</td>
<td>4.74</td>
<td>-0.94</td>
<td>4.35</td>
<td>1.69</td>
<td>4.63</td>
<td>-0.56</td>
<td>6.07</td>
</tr>
<tr>
<td>10</td>
<td>-0.75</td>
<td>6.14</td>
<td>-1.24</td>
<td>5.88</td>
<td>0.72</td>
<td>6.37</td>
<td>-0.95</td>
<td>6.53</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.69</td>
<td>6.91</td>
<td>-1.16</td>
<td>6.61</td>
<td>0.51</td>
<td>6.71</td>
<td>-0.77</td>
<td>7.71</td>
</tr>
</tbody>
</table>

Table 5: This table reports summary statistics for in-sample realized variance fits. For each model, the table reports the residual means and RMSEs in basis points for seven maturities in years.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>LRSQ(3,0) Mean</th>
<th>RMSE</th>
<th>LRSQ(3,1) Mean</th>
<th>RMSE</th>
<th>LRSQ(3,2) Mean</th>
<th>RMSE</th>
<th>LRSQ(3,3) Mean</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>-0.22</td>
<td>1.07</td>
<td>-0.32</td>
<td>1.04</td>
<td>-0.10</td>
<td>0.97</td>
<td>-0.14</td>
<td>1.01</td>
</tr>
<tr>
<td>1</td>
<td>-0.11</td>
<td>0.81</td>
<td>-0.16</td>
<td>0.75</td>
<td>-0.04</td>
<td>0.73</td>
<td>-0.16</td>
<td>0.75</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
<td>0.85</td>
<td>0.05</td>
<td>0.82</td>
<td>0.01</td>
<td>0.80</td>
<td>-0.13</td>
<td>0.79</td>
</tr>
<tr>
<td>3</td>
<td>0.14</td>
<td>0.88</td>
<td>0.14</td>
<td>0.86</td>
<td>0.04</td>
<td>0.83</td>
<td>-0.09</td>
<td>0.81</td>
</tr>
<tr>
<td>5</td>
<td>0.21</td>
<td>0.82</td>
<td>0.21</td>
<td>0.79</td>
<td>0.12</td>
<td>0.70</td>
<td>-0.02</td>
<td>0.72</td>
</tr>
<tr>
<td>7</td>
<td>0.24</td>
<td>0.76</td>
<td>0.23</td>
<td>0.69</td>
<td>0.21</td>
<td>0.62</td>
<td>0.06</td>
<td>0.63</td>
</tr>
<tr>
<td>10</td>
<td>0.27</td>
<td>0.73</td>
<td>0.27</td>
<td>0.65</td>
<td>0.31</td>
<td>0.62</td>
<td>0.16</td>
<td>0.60</td>
</tr>
<tr>
<td>Mean</td>
<td>0.08</td>
<td>0.85</td>
<td>0.06</td>
<td>0.80</td>
<td>0.08</td>
<td>0.75</td>
<td>-0.05</td>
<td>0.76</td>
</tr>
</tbody>
</table>
Table 6: This table reports the results from the cross-sectional yield fit regression (2.4.3) for seven maturities in years. Panel A summarizes the results from the regression involving only the lagged PCs of yields. Panel B shows the results from the regression obtained from including both the PCs and the EGARCH(1, 1) estimates of conditional volatility. Panel C, D, E, and F show the results from the regression obtained from including both the PCs and one of the four model-implied realized volatility estimates respectively. Moreover, the bottom part of each panel shows the results from only including the model-implied realized volatility estimates. For each regression the adjusted $R^2$ values and the estimated regression coefficients are shown. Whether a regression coefficient is statistically insignificant at the 5% and 1% level is indicated by \(*\) and \(**\) respectively. The $p$-value is for the Wald test of zero regression parameters, $H_0: (\alpha, \beta) = 0$.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Panel A</th>
<th>Panel B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PC1</td>
<td>EGARCH(1,1)</td>
</tr>
<tr>
<td></td>
<td>Adj. $R^2$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>$\beta_1$</td>
</tr>
<tr>
<td>0.5</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>1</td>
<td>0.07</td>
<td>-0.06</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.09</td>
<td>-0.02</td>
</tr>
<tr>
<td>7</td>
<td>0.09</td>
<td>-0.02</td>
</tr>
<tr>
<td>10</td>
<td>0.09</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: $n = 0$</th>
<th>Panel D: $n = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj. $R^2$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$\beta_1$</td>
</tr>
<tr>
<td>0.02</td>
<td>0.19</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>0.45</td>
<td>0.83</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>$R^2$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>0.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel E: $n = 2$</th>
<th>Panel F: $n = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj. $R^2$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$\beta_1$</td>
</tr>
<tr>
<td>0.02</td>
<td>0.51</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>0.02</td>
<td>0.07</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>$R^2$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>0.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel G: $n = 4$</th>
<th>Panel H: $n = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj. $R^2$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$\beta_1$</td>
</tr>
<tr>
<td>0.02</td>
<td>0.10</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>$R^2$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>0.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Table 7: This table reports the results from the cross-sectional variance fit regression (2.4.4) for seven maturities in years. For each model, the adjusted $R^2$ values and the estimated regression coefficients are shown, with $\alpha$ multiplied by $10^9$. Panel A shows the results from the regression obtained from including EGARCH(1,1) estimates of conditional volatility. Whether a regression coefficient is statistically insignificant at the 5% and 1% level is indicated by * and ** respectively. The $p$-value is for the Wald test of zero regression parameters, $H_0 : (\alpha, \beta) = 0$.

| Maturity | 0.5 | 1 | 2 | 3 | 5 | 7 | 10 | 0.5 | 1 | 2 | 3 | 5 | 7 | 10 |
|----------|-----|---|---|---|---|---|----|-----|---|---|---|---|---|---|----|
| $\alpha$ | -0.87* | -0.66* | 1.24 | 2.13 | 2.70 | 3.05 | 3.44 | -3.25 | -1.61 | 0.50* | 1.45 | 2.16 | 2.41 | 2.74 |
| $\beta_1$ | -0.05* | -0.02* | 0.07 | 0.13 | 0.16 | 0.15 | 0.14 | 0.04* | 0.07 | 0.14 | 0.18 | 0.17 | 0.12 | 0.06 |
| $\beta_2$ | 0.05* | 0.04* | -0.04* | -0.04* | -0.13 | -0.13 | -0.16 | 0.05* | 0.05* | 0.01* | -0.03* | -0.07* | -0.08* | -0.07** |
| $\beta_3$ | -0.17* | -0.16* | -0.20* | -0.15* | -0.03* | 0.08* | 0.18 | -0.23* | -0.27 | -0.26 | -0.19* | -0.07* | 0.01* | 0.08 |
| $p$-value | 0.26 | 0.18 | 0.22 | 0.55 | 0.70 | 0.21 | 0.03 | 0.30 | 0.13 | 0.30 | 0.69 | 0.77 | 0.55 | 0.40 |
| Adj. $R^2$ | 0.81 | 0.01 | 0.04 | 0.10 | 0.17 | 0.21 | 0.23 |

Panel B: LRSQ(3,0)  

| $\alpha$ | -2.31 | -1.13** | 0.64* | 1.51 | 2.20 | 2.44 | 2.78 | -3.25 | -1.61 | 0.50* | 1.45 | 2.16 | 2.41 | 2.74 |
| $\beta_1$ | 0.12 | 0.13 | 0.17 | 0.19 | 0.18 | 0.17 | 0.14 | 0.04* | 0.07 | 0.14 | 0.18 | 0.17 | 0.12 | 0.06 |
| $\beta_2$ | -0.01* | 0.00* | -0.00* | -0.03* | -0.12 | -0.15 | -0.15* | 0.05* | 0.05* | 0.01* | -0.03* | -0.07* | -0.08* | -0.07** |
| $\beta_3$ | -0.19* | -0.25 | -0.27 | -0.26* | -0.05* | 0.08* | 0.21 | -0.23* | -0.27 | -0.26 | -0.19* | -0.07* | 0.01* | 0.08 |
| $p$-value | 0.58 | 0.25 | 0.29 | 0.63 | 0.57 | 0.14 | 0.01 | 0.30 | 0.13 | 0.30 | 0.69 | 0.77 | 0.55 | 0.40 |
| Adj. $R^2$ | 0.85 | 0.11 | 0.15 | 0.17 | 0.19 | 0.20 | 0.20 | 0.01 | 0.06 | 0.12 | 0.16 | 0.17 | 0.11 | 0.04 |

Panel C: LRSQ(3,1)  

| $\alpha$ | -0.84* | -0.38* | 0.23* | 0.65* | 1.52 | 2.29 | 3.21 | -3.18* | -1.51 | -1.18** | -0.81* | -0.06* | 0.62* | 1.62 |
| $\beta_1$ | 0.03* | 0.05* | 0.10 | 0.12 | 0.13 | 0.10 | 0.07 | 0.01* | 0.01* | 0.01* | 0.01* | 0.01* | 0.01* | 0.01* |
| $\beta_2$ | 0.06* | 0.12 | 0.15 | 0.12* | 0.03* | -0.04* | -0.10 | 0.09* | 0.09* | 0.07* | 0.03* | -0.03* | -0.06* | -0.09** |
| $\beta_3$ | -0.25* | -0.34 | -0.41 | -0.36 | -0.18* | -0.02* | 0.14* | -0.27* | -0.29 | -0.26* | -0.18* | -0.03* | 0.08* | 0.19 |
| $p$-value | 0.29 | 0.10 | 0.12 | 0.29 | 0.84 | 0.59 | 0.16 | 0.23 | 0.08 | 0.13 | 0.29 | 0.74 | 0.73 | 0.21 |
| Adj. $R^2$ | 0.82 | 0.09 | 0.12 | 0.12 | 0.10 | 0.08 | 0.07 | 0.02 | 0.03 | 0.02 | 0.09 | 0.01 | 0.00 | 0.03 |

Table 8: This table reports the unconditional sample correlation coefficients between the time-series of conditional volatility implied by EGARCH($p$, $q$) estimates of the actual yields and fitted yields for seven maturities in years.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>LRSQ(3,0)</th>
<th>LRSQ(3,1)</th>
<th>LRSQ(3,2)</th>
<th>LRSQ(3,3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.980</td>
<td>0.981</td>
<td>0.977</td>
<td>0.972</td>
</tr>
<tr>
<td>1</td>
<td>0.976</td>
<td>0.975</td>
<td>0.965</td>
<td>0.961</td>
</tr>
<tr>
<td>2</td>
<td>0.952</td>
<td>0.958</td>
<td>0.954</td>
<td>0.911</td>
</tr>
<tr>
<td>3</td>
<td>0.974</td>
<td>0.976</td>
<td>0.966</td>
<td>0.932</td>
</tr>
<tr>
<td>5</td>
<td>0.967</td>
<td>0.967</td>
<td>0.955</td>
<td>0.950</td>
</tr>
<tr>
<td>7</td>
<td>0.983</td>
<td>0.984</td>
<td>0.967</td>
<td>0.972</td>
</tr>
<tr>
<td>10</td>
<td>0.972</td>
<td>0.973</td>
<td>0.950</td>
<td>0.969</td>
</tr>
</tbody>
</table>
Table 9: This table reports the out-of-sample one-month ahead RMSFEs for seven maturities in years between actual and model-implied future conditional volatility. We use (2.4.2) to proxy actual conditional volatility and expected absolute changes in yields $E_t(|\Delta y(t)|)$ to obtain model-implied forecasts of conditional volatility. Whether the difference for a given maturity compared to the LRSQ(3, 2) model is statistically significant at the 5% and 1% level is indicated by * and ** respectively.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>LRSQ(3, 0)</th>
<th>LRSQ(3, 1)</th>
<th>LRSQ(3, 2)</th>
<th>LRSQ(3, 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>12.16**</td>
<td>12.60**</td>
<td>6.70</td>
<td>7.02</td>
</tr>
<tr>
<td>1</td>
<td>12.61**</td>
<td>12.79**</td>
<td>8.26</td>
<td>11.36**</td>
</tr>
<tr>
<td>2</td>
<td>8.12**</td>
<td>8.57**</td>
<td>6.95</td>
<td>10.96**</td>
</tr>
<tr>
<td>3</td>
<td>5.59**</td>
<td>5.87**</td>
<td>5.57</td>
<td>8.57**</td>
</tr>
<tr>
<td>5</td>
<td>6.59</td>
<td>5.81</td>
<td>6.05</td>
<td>5.48</td>
</tr>
<tr>
<td>7</td>
<td>8.20**</td>
<td>7.35**</td>
<td>7.91</td>
<td>4.89**</td>
</tr>
<tr>
<td>10</td>
<td>9.32**</td>
<td>8.94**</td>
<td>10.00</td>
<td>5.71**</td>
</tr>
<tr>
<td>Mean</td>
<td>8.94</td>
<td>8.85</td>
<td>7.35</td>
<td>7.71</td>
</tr>
</tbody>
</table>
Table 10: This table reports the results of the cross-sectional volatility forecast regression (2.5.1) for seven maturities in years. We use expected absolute changes in yields $E_t(|\Delta y(\tau)|)$ to proxy model-implied conditional volatility. For each model, the adjusted $R^2$ values and the estimated regression coefficients are shown, with $\alpha$ multiplied by $10^2$. Panel A shows the results from the regression obtained from including EGARCH(1, 1) model forecasts of conditional volatility. Whether a regression coefficient is statistically insignificant at the 5% and 1% level is indicated by * and ** respectively. The $p$-value is for the Wald test of zero regression parameters, $H_0 : (\alpha, \beta) = 0$.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>10</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>0.00*</td>
<td>-0.00*</td>
<td>0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>-0.00*</td>
<td>0.00*</td>
<td>-0.00*</td>
<td>0.00*</td>
<td>-0.00*</td>
<td>0.00*</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00*</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.46</td>
<td>0.60</td>
<td>0.11</td>
<td>0.24</td>
<td>0.54</td>
<td>0.66</td>
<td>0.04</td>
<td>0.06</td>
<td>0.15</td>
<td>0.23</td>
<td>0.16</td>
<td>0.07</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.06</td>
<td>0.06</td>
<td>0.15</td>
<td>0.23</td>
<td>0.11</td>
<td>0.16</td>
<td>0.07</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel A: EGARCH(1,1)
Table 11: This table reports the out-of-sample one-month ahead RMSFEs for seven maturities in years between actual and model-implied future conditional volatility. We use (2.4.2) to proxy actual conditional volatility and we proxy the model-implied forecasts of conditional volatility by realized yield volatility obtained from $r_{t+h,t}^{\text{vol,m}}(\tau) = \sum_{i=1}^{N_t} (\hat{y}_{d,t}^i(\tau) - \hat{y}_{d,t}^{i-1}(\tau))^2$, where $N_t$ denotes the number of trading days of month $t$, and yields are in all cases expressed in basis points. Whether the difference for a given maturity compared to the LRSQ(3,2) model is statistically significant at the 5% and 1% level is indicated by * and ** respectively.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>LRSQ(3,0)</th>
<th>LRSQ(3,1)</th>
<th>LRSQ(3,2)</th>
<th>LRSQ(3,3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>17.02**</td>
<td>17.93**</td>
<td>10.37</td>
<td>10.55</td>
</tr>
<tr>
<td>1</td>
<td>17.09**</td>
<td>17.79**</td>
<td>12.10</td>
<td>15.46**</td>
</tr>
<tr>
<td>2</td>
<td>12.00**</td>
<td>12.95**</td>
<td>10.81</td>
<td>15.56**</td>
</tr>
<tr>
<td>3</td>
<td>8.08**</td>
<td>9.26</td>
<td>8.77</td>
<td>13.10**</td>
</tr>
<tr>
<td>5</td>
<td>5.47*</td>
<td>5.82*</td>
<td>6.14</td>
<td>8.89**</td>
</tr>
<tr>
<td>7</td>
<td>5.83</td>
<td>5.46</td>
<td>5.96</td>
<td>6.44</td>
</tr>
<tr>
<td>10</td>
<td>6.55</td>
<td>6.38</td>
<td>7.42</td>
<td>4.91**</td>
</tr>
<tr>
<td>Mean</td>
<td>10.29</td>
<td>8.85</td>
<td>7.35</td>
<td>7.71</td>
</tr>
</tbody>
</table>
Table 12: This table reports the results of the cross-sectional volatility forecast regression (2.5.1) for seven maturities in years. We proxy model-implied conditional volatility by realized yield volatility obtained from \( r_{t+h,t}'(\tau) = \sum_{i=1}^{N_t} (y_{d,t}'(\tau) - \hat{y}_{d,t}(\tau))^2 \), where \( N_t \) denotes the number of trading days of month \( t \), and yields are in all cases expressed in basis points. For each model, the adjusted \( R^2 \) values and the estimated regression coefficients are shown, with \( \alpha \) multiplied by 10^2. Panel A shows the results from the regression obtained from including EGARCH(1, 1) model forecasts of conditional volatility. Whether a regression coefficient is statistically insignificant at the 5% and 1% level is indicated by * and ** respectively. The \( p \)-value is for the Wald test of zero regression parameters, \( H_0 : (\alpha, \beta) = 0 \).

<table>
<thead>
<tr>
<th>Maturity</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>10</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>-1.67</td>
<td>-1.69</td>
<td>-1.12</td>
<td>-0.62</td>
<td>-0.02</td>
<td>0.26</td>
<td>0.42</td>
<td>-1.77</td>
<td>-1.77</td>
<td>-1.25</td>
<td>-0.80</td>
<td>-0.24</td>
<td>0.08</td>
<td>0.33</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( p )-value</td>
<td>0.17</td>
<td>0.03</td>
<td>0.37</td>
<td>0.65</td>
<td>0.53</td>
<td>0.18</td>
<td>0.02</td>
<td>0.28</td>
<td>0.05</td>
<td>0.36</td>
<td>0.60</td>
<td>0.40</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>Adj. ( R^2 )</td>
<td>0.17</td>
<td>0.25</td>
<td>0.31</td>
<td>0.24</td>
<td>0.17</td>
<td>0.15</td>
<td>0.15</td>
<td>0.01</td>
<td>0.05</td>
<td>0.16</td>
<td>0.16</td>
<td>0.17</td>
<td>0.20</td>
<td>0.22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maturity</th>
<th>LRSQ(3, 0)</th>
<th>LRSQ(3, 1)</th>
<th>LRSQ(3, 2)</th>
<th>LRSQ(3, 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.349</td>
<td>0.376</td>
<td>0.257</td>
<td>0.319</td>
</tr>
<tr>
<td>1</td>
<td>0.356</td>
<td>0.386</td>
<td>0.335</td>
<td>0.389</td>
</tr>
<tr>
<td>2</td>
<td>0.341</td>
<td>0.368</td>
<td>0.368</td>
<td>0.409</td>
</tr>
<tr>
<td>3</td>
<td>0.328</td>
<td>0.350</td>
<td>0.384</td>
<td>0.410</td>
</tr>
<tr>
<td>5</td>
<td>0.319</td>
<td>0.338</td>
<td>0.393</td>
<td>0.404</td>
</tr>
<tr>
<td>7</td>
<td>0.313</td>
<td>0.328</td>
<td>0.397</td>
<td>0.396</td>
</tr>
<tr>
<td>10</td>
<td>0.309</td>
<td>0.321</td>
<td>0.399</td>
<td>0.386</td>
</tr>
<tr>
<td>Max</td>
<td>1.345</td>
<td>1.357</td>
<td>1.253</td>
<td>1.329</td>
</tr>
</tbody>
</table>

Table 13: This table reports the conditional yearly Sharpe ratios of different maturities in years based on simulated data of length 50,000 years from each model. The bottom part of the table shows the mean maximum conditional yearly Sharpe ratios, which are given as \( s_{r_t}^{max} = (\lambda_t^{\text{LHS}} \lambda_t^{\text{RHS}})^{1/2} \).

<table>
<thead>
<tr>
<th>Maturity</th>
<th>LRSQ(3, 0)</th>
<th>LRSQ(3, 1)</th>
<th>LRSQ(3, 2)</th>
<th>LRSQ(3, 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.349</td>
<td>0.376</td>
<td>0.257</td>
<td>0.319</td>
</tr>
<tr>
<td>1</td>
<td>0.356</td>
<td>0.386</td>
<td>0.335</td>
<td>0.389</td>
</tr>
<tr>
<td>2</td>
<td>0.341</td>
<td>0.368</td>
<td>0.368</td>
<td>0.409</td>
</tr>
<tr>
<td>3</td>
<td>0.328</td>
<td>0.350</td>
<td>0.384</td>
<td>0.410</td>
</tr>
<tr>
<td>5</td>
<td>0.319</td>
<td>0.338</td>
<td>0.393</td>
<td>0.404</td>
</tr>
<tr>
<td>7</td>
<td>0.313</td>
<td>0.328</td>
<td>0.397</td>
<td>0.396</td>
</tr>
<tr>
<td>10</td>
<td>0.309</td>
<td>0.321</td>
<td>0.399</td>
<td>0.386</td>
</tr>
<tr>
<td>Max</td>
<td>1.345</td>
<td>1.357</td>
<td>1.253</td>
<td>1.329</td>
</tr>
</tbody>
</table>
A.3 Figures

Figure 1: Time series of yields and realized variances

This figure plots the time series of yields (left column), and the realized yield variances (right column) (see (2.4.1) and (2.4.2) for an exact specification) over the period 1987-2017 of different maturities in years.
This figure shows in order from the top row to the bottom row the estimated factor values of the LRSQ(3, 0), LRSQ(3, 1), LRSQ(3, 2), and LRSQ(3, 3) model respectively. If part of the respective model the first column shows $Z_{1t}$ and $U_{1t}$, the second column shows $Z_{2t}$ and $U_{2t}$, and the third row shows $Z_{3t}$ and $U_{3t}$, where the solid line depicts the spanned factors and the dashed line depicts the unspanned factors.
Figure 3: EGARCH(1,1) conditional volatilities

This figure plots the EGARCH(1,1) estimates of conditional volatilities of the 0.5, 3, 7, and 10 years yields.
**Figure 4:** Fit to realized variance

This figure shows the fit of the LRSQ(3,0), LRSQ(3,1), LRSQ(3,2), and LRSQ(3,3) models to realized variance of the ten-year yield.
**Figure 5:** PC loadings of realized variances

This figure shows the PC loadings of realized variances plotted against maturity.

**Figure 6:** Campbell-Shiller loadings

This figure shows the Campbell-Shiller (CS) loadings (left figure) implied by the data and implied by the models as well as model-implied risk-adjusted CS loadings (right figure). The model-implied CS loadings are the mean loadings from running 2,000 CS regression on simulated samples of yields of length 293 months. The data-implied CS loadings are estimated using a sample from August 1987 to December 2011. The 95% confidence intervals of data-implied CS loadings are obtained from block bootstrapping with a block length of 256 months and 5,000 repetitions. The risk-adjusted CS loadings are obtained by running the risk-adjusted CS regression based on model-implied term premiums obtained from simulating 2,000 paths of the short rate.
CHAPTER 3

THE IMPACT OF CURVE-FITTING PROCEDURE ON ESTIMATION AND TESTING OF TERM STRUCTURE MODELS

Bent Jesper Christensen
Aarhus University, CREATES, Dale T. Mortensen Centre, and the Danish Finance Institute

Jorge W. Hansen
Aarhus University and CREATES

Abstract

We introduce a statistical invariance test for the consistency between the shape of the yield curve and the stochastic process driving interest rates. The analysis is cast in the Heath-Jarrow-Morton framework, and utilizes factor analysis of yield rates estimated from coupon bond data. The theoretical properties of the invariance tests are investigated, and an application to US Treasuries is offered.
3.1 Introduction

Empirical analysis of interest rates using panel data with both calendar time and term to maturity dimensions generally proceeds in two steps: First, a yield curve is fitted to the available bond price data for each trading date. Second, the resulting time series of yield curves is used to address a substantive issue of interest, such as the testing of a particular dynamic term structure model (DTSM) or testing for arbitrage opportunities. Examples of studies using such a two-step procedure include Duffee (2002) and Dai, Singleton and Yang (2007). It is usually assumed that splitting the analysis in a yield curve fitting step and a structural estimation and testing step is innocuous. However, some of the typical questions examined in the second stage of such an analysis are of a character raising suspicion that the choice of curve fitting procedure in the first step may actually increase the risk of improper inference in the second step. For example, an important issue in empirical analysis of interest rates is the number of factors driving the term structure through time. Early studies on this topic include Singleton (1980), who used classical factor analysis and concluded that two factors suffice for describing the comovements of interest rates. Later studies have typically found evidence of three (Litterman and Scheinkman, 1991) or four factors (Knez, Litterman and Scheinkman, 1994) to capture comovements. A pressing issue is whether this count reflects the true number of underlying factors or simply the number of significant or important parameters of the first step curve fit impounded in the data analyzed in the second step analysis. If a four-parameter functional form for the yield curve is fit to bond data on every trading date, this would seem to make a finding that a fixed number of factors (likely four) suffices for describing term structure movement much more likely in the second step panel data analysis. Other substantive issues, such as testing arbitrage restrictions or distinguishing between parametric DTSMs, suffer from the same problem. The general issue is to what extent inferences are determined by the particular curve fitting method adopted.

We cast our analysis in the framework of Heath et al. (1992) (henceforth HJM). This is a general and consistent framework for arbitrage-free interest rate modeling and pricing of interest rate contingent claims. The framework contains a large number of existing interest rate models as special cases. This includes the well-known one-factor short rate models of Cox et al. (1985), Ho and Lee (1986), Black and Karasinski (1991), Hull and White (1993), the two-factor model of Brennan and Schwartz (1979), and the Markovian volatility model by Ritchken and Sankarasubramanian (1995). These models are all in the affine term structure class of Duffie and Kan (1996). However, the non-affine alternatives, such as the constant elasticity of variance-type model studied by Chan, Karolyi, Longstaff and Sanders (1992) are also contained in the HJM framework.

The HJM approach is to model the stochastic evolution through time of the entire term structure of instantaneous forward rates, with an arbitrary initial term struc-
ture as starting point. The ability to accommodate an arbitrary initial term structure facilitates applications to derivative pricing, where conditioning on the current observed yield curve is important. To exclude arbitrage, HJM derive conditions on the drift function of the forward rates, restricting it to depend only on the forward rate volatility structure, the market price of interest rate risk under the historical (physical) measure, and on the volatility structure under an equivalent martingale measure. As a result, the dynamics are entirely determined by the initial yield curve, the volatility function under the martingale measure, and on these two in conjunction with the market price of risk under the historical measure. In practice, the initial yield curve is typically fit to prevailing market prices on the current calendar date using a flexible parameterized functional form. Volatility and risk price parameters may be estimated using both current and historical data, typically based on yield curve fits for each of a sequence of consecutive calendar dates. Thus, it is natural to ask about the impact of the choice of functional form for the individual yield curve fits. In particular, if a different functional form were used, would this lead to markedly different inference for the proper shape of the volatility function, risk prices, derivative prices, and the similar, or are these robust to the choice of curve fitting procedure?

An important tool for analyzing this question is the concept of consistency, i.e., the consistency between the functional form of the yield curve and a given DTSM, as introduced by Björk and Christensen (1999) (henceforth BC). They give necessary and sufficient conditions for whether a DTSM, if started at an initial curve of a given functional form (e.g., the yield curve representation introduced by Nelson and Siegel (1987) (henceforth NS)), will continue to generate future term structures of the same general form (with different parameters on each future date).

The general question which we address in the present paper is whether the use of a curve fitting procedure which does not go well with the chosen parameterized DTSM (in the sense that it is inconsistent with this) potentially leads to misspecification of model parameters, incorrect hypothesis testing, and the similar. We do this by applying different curve fitting procedures to empirical data on prices of US coupon bonds from January 1999 to December 2014. We use the computed yield curve to investigate the impact of the choice of curve fitting procedure on the received inferences.\(^1\) In particular, if the yield rate panel data is constructed by fitting curves to coupon bond prices using a procedure which generates yield curves consistent with a particular theoretical DTSM with a certain number of factors. Does this imply that the number of factors when applying classical factor analysis and similar factor counting techniques to the panel data equals that of the theoretical consistent model? Further-

\(^1\)An alternative would be to conduct a Monte Carlo study, simulating interest rates from an assumed model, and investigating the impact on inferences of applying different curve fitting schemes to the simulated data, including both consistent and inconsistent functional forms among the schemes studied. We prefer to focus attention on the actual empirical market data, and leave the simulation possibility for future research.
more, consistency of a given curve shape is in fact with a DTSM with the no-arbitrage
drift condition imposed. This raises the additional question of whether panel data
constructed using a given curve fitting procedure leave the test of the no-arbitrage
drift restriction biased in favor of the restriction, given that this in a sense already is
built into the fitted curve shapes.

Previously, very little empirical work has been done on the subject of consistency.
Angelini and Herzel (2002) calibrate ATM cap prices within a one-factor Hull and
White (1993) setup to yield curves fit to swap data using forward curves of various
functional forms. Matters such as parameter stability and pricing errors are discussed.
However, the paper does not provide a formalized testing framework, and many
questions are left unanswered. Our analysis is focused on the underlying DTSM,
rather than derivative prices. We consider panel data sets of yields obtained using
various different types of popular yield curve families, that is, different functional
forms of yield curve fits. If one of the curve shapes implemented is consistent with
the (unknown) true model, then significant differences in inferences indicate that
application of an inconsistent curve family (a misspecified calibration procedure)
yields erroneous inferences. More generally, differences in inferences indicate that
the choice of curve fitting (calibration) procedure matters in an important manner
for the term structure analysis of interest. We look at the DTSMs by Ho and Lee (1986),
Hull and White (1993) as well as the two-factor model by HJM, and extend the latter
to a three- and a four-factor model, respectively. While we find that consistency plays
an important role for the conclusion drawn from the analysis, the focus during the
estimation step is based on the DTSM only, without imposing a particular yield curve
shape. We will therefore also consider the dynamically consistent term structure
models (DCTSMs) of Borup, Christensen, Hansen and Rasmussen (2020) (henceforth
BCHR), which enable us to impose a particular consistent yield curve shape during
the estimation. Model estimation is done by maximum likelihood factor analysis. We
test for no-arbitrage using a likelihood-ratio (LR) test, comparing model fits based on
the HJM drift restriction against an unrestricted mean. In addition, we introduce a
computationally simpler Wald test, which only requires estimation of the unrestricted
model. These tests are closely related to the test of no-arbitrage in Christensen and
van der Wel (2019). They measure the deviation from the HJM drift restriction based
on a likelihood-ratio, and interpret a rejection of the test as an indication of either
presence of arbitrage opportunities, or, in terms of the efficient market hypothesis,
model misspecification. We add another dimension to this analysis by studying the
implications of the initial yield curve parameterization on the outcome of the test.

The results of our analysis are twofold. First, we find that while we need at least
two factors to exclude arbitrage-opportunities, three factors are generally required to
describe the volatility structure. Second, the choice of the yield curve family has signif-
icanat impact on the conclusions drawn from the analysis. Furthermore, consistency
appears to play an important role for the outcome of the test for no-arbitrage.

The remainder of the paper is organized as follows. In Section 3.2, we present general model setup, notation, problem formulation, and consistency conditions. In Section 3.3, we outline the estimation procedure of the DTSM, and in Section 3.4, we introduce the statistical test for no-arbitrage in the DTSMs. Section 3.5 presents an application of the DTSM to US Treasury bond data. We next discuss the DCTSM in Section 3.6. In Section 3.7, we introduce a test for no-arbitrage in the DCTSM, and in Section 3.8 we analyze the DCTSM empirically. We conclude in Section 3.9.

### 3.2 Modeling framework

In this section, we present theory of term structure models and consistent yield curves that forms the basis of the empirical analysis. In Section 3.2.1, the general theory is outlined, in Section 3.2.2 the concept of consistency between a yield curve family and a DTSM is reviewed, and in Section 3.2.3 specific parameterized DTSMs and associated consistent yield curve families are discussed.

#### 3.2.1 DTSMs and consistent yield curves

Our analysis is cast in the framework of HJM, where we use yields as the modeling object.\(^2\) The usual transformations allow equivalent restatements of models and results in terms of discount factors and forward rates. The instantaneous yield at time \(t\) for delivery at time \(t + \tau \leq T\) is defined by

\[
y(t, \tau) = -\frac{\log P(t, \tau)}{\tau},
\]

where \(P(t, \tau)\) is the time-\(t\) price of a deflated zero-coupon bond maturing at time \(t + \tau\). We use the parametrization by Musiela (1993) and parametrize in terms of time to maturity \(\tau\) rather than the maturity date \(t + \tau\). The dynamics of the instantaneous yields are assumed to be given as

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

for some (scalar) drift \(\alpha\) and a \(m \times 1\) dimensional volatility function \(\sigma\), where \(m\) denotes the dimension of the driving Wiener process \(W_t\). The no-arbitrage drift restriction of HJM is

\[
\alpha(t, \tau) = \frac{1}{\tau} (y(t, \tau) - y(t, 0)) + \frac{\partial y}{\partial \tau}(t, \tau) + \sigma(t, \tau)' \lambda(t) + \frac{\tau}{2} \sigma(t, \tau)' \sigma(t, \tau),
\]

\(^2\)We consider a model defined on a standard filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})\), where \(\mathbb{P}\) denotes the objective probability measure. The information set \(\mathcal{F}_t\) is the natural filtration generated by an \(m\)-dimensional Wiener process \(W_t\) which drives all interest rates until some fixed time \(T\), as usual augmented with the null sets.
where $\lambda(t)$ represents the $m$-dimensional market price of interest rate risk which is independent of $\tau$.\(^3\) We consider the case of constant and deterministic market price of risk $\lambda$ and volatility $\sigma(\tau)$, although $\sigma(\tau)$ may depend on time to maturity $\tau$. We can view (3.2.1) as an infinite-dimensional SDE governing the evolution of the yield curve $y_t = y(t, \cdot) = \{y(t, \tau) \mid t \leq \tau \leq T\}$ across calendar time, where $y_t$ is itself indexed by term to maturity $\tau$, running from current date $t$ through the final date $T$, for each $t$. In particular, the initial yield curve is given by $y_0 = \{y(0, \tau) \mid 0 \leq \tau \leq T\}$, and the distribution of all rates in $[0, T]$ is completely specified by the volatility function $\sigma$, the market price of risk $\lambda$, and the initial yield curve $y_0$. Thus, we can think of an interest rate model as a class of $(\sigma, \lambda)$-pairs, mapping each initial curve to the distribution of future interest rates. Indeed, the very reason for HJM to introduce this very general framework was to accommodate any arbitrary initial yield curve.

In practice, data on current market prices of traded interest rate sensitive instruments is used to fit an initial (i.e., current) yield curve. For the purposes of derivative pricing, hedging, and risk management the fitted curve is then used as initial curve for the future stochastic evolution of interest rates in the above modeling framework. More generally, any empirical analysis of DTSMs depends on the availability of interest rate data, and interest rates are typically read off yield curves that are fit on a day-to-day basis on each given date using current market prices in the fitting procedure. In both cases, the fitting of an empirical yield curve to current market prices is the important starting point. To analyze this a little further, suppose we are given data on current market prices of a number of possibly coupon-bearing bonds. Let $p_i$ be the observed market price of the $i$th bond, which promises $J_i$ payments $c_{i1}, \ldots, c_{iJ_i}$ at future dates $\tau_{i1}, \ldots, \tau_{iJ_i}$ periods hence. The fitted yield curve, $y(z, \tau)$, will in general be a function of term to maturity $\tau$ and some (possibly vector valued) parameter $z$ estimated as part of the fitting procedure. Thus, $y(z, \tau)$ should be thought of as the fitted (empirical) version of the yield curve, with $y(\cdot, \cdot)$ a suitable flexibly parameterized functional form. These are estimates of the current curves at the given calendar date, say $t$, so that the dependence on the calendar date is through the current estimate of the curve fitting parameter $z$. To be precise, the fitted yield curve is $y(\hat{z}, \tau)$, where

$$\hat{z} = \arg\min_z \sum_{i=1}^{N} w_i \left( p_i - \sum_{j=1}^{J_i} c_{ij} e^{-y(z, \tau) \tau_{ij}} \right)^2. \quad (3.2.3)$$

Here, $N$ denotes the number of bond prices and $w_1, \ldots, w_N$ are estimation weights typically based on bond specific information such as outstanding nominal amount or time to maturity. We follow the usual practice in the yield fitting literature and weight the calibration by inverse bond duration (see e.g. Gürkaynak et al. (2007)).\(^4\)

\(^3\)HJM considered the alternative parametrization with maturity date $t + \tau$ instead of $\tau$ as a separate argument, and represented the term structure via instantaneous forward rates instead of yields. A formal proof of (3.2.2) can be found in Christensen and van der Wel (2019).

\(^4\)By weighting prices by the inverse of the bond duration, we are, to a first order approximation, minimizing yield errors.
A popular yield curve parameterization is the NS yield curve given by

\[ y(z_t, \tau) = z_{1t} + z_{2t} \frac{1 - e^{-a_2 \tau}}{a_2 \tau} + z_{3t} \left( \frac{1 - e^{-a_2 \tau}}{a_2 \tau} - e^{-a_2 \tau} \right), \tag{3.2.4} \]

where \( z_{1t}, z_{2t}, z_{3t}, \) and \( a_2 \) are constants to be estimated. The NS loading functions given by

\[ B_1(\tau) = 1, \quad B_2(\tau) = \frac{(1 - e^{-a_2 \tau})}{(a_2 \tau)}, \quad B_3(\tau) = \frac{(1 - e^{-a_2 \tau})}{(a_2 \tau)} - e^{-a_2 \tau} \]

are flat, downward sloping, and hump shaped, respectively, motivated by the relevance of both monotonicity and hump-shape features in the observed term structures. The calibration procedure is typically repeated on a regular basis (e.g., daily or weekly), each time with a new number of bond prices \( N \), and a new data set \( \{p_i, J_i, w_i, c_{ij}, \tau_{ij}\}_{ij} \).

Thus, on each calendar date \( t \) where a calibration is performed, new fitted values \( \hat{z}_t \) are obtained, producing a time series \( \{\hat{z}_t\}_t \), and hence a sequence of fitted yield curves \( \{y(\hat{z}_t, \cdot)\}_t \). Note that these fitted curves are obtained in separate data sets and that in many applications only the current (i.e. most recent) fitted curve is used, e.g., as initial curve for modeling the future evolution of interest rates as a basis for interest rate derivative pricing.

### 3.2.2 Invariant manifolds

When using a fitted yield curve (or, equivalently, a fitted forward curve) as initial curve in applications to pricing, hedging, and so forth, a particular DTSM is adopted, in which case, one has to specify the form of the volatility function. A natural question is whether the choice of a functional parametric family has any effect on the estimated model parameters and gives rise to the following issue of consistency between the functional forms of the curve fits and the volatility function: Do the chosen functional forms go together, or is there a mismatch, in some sense? Specifically, if a particular DTSM is used for pricing, does this place certain restrictions on the shape of cross-sectional yield fit that should be performed? Conversely, if a particular functional form is used regularly in the periodical yield curve fits, does this have any bearing on the kind of DTSM that should subsequently be started off at the fitted initial curve? This is the consistency issue studied in BC. Given an initial yield curve and parameterized interest rate model, they study the question of whether the interest rate model will produce future yield curves belonging to the same family of yield curves (e.g. the Nelson-Siegel family) as the initial yield curve. In fact, BC study the consistency issue in terms of forward curves. Since the usual transformations allow equivalent restatements of models and results in terms of yield curves (and discount factors), we follow the common practice in the term structure literature and work with the yield curve in our empirical application.

Following BC, we spell out the consistency issue more explicitly. Let \( \mathcal{M} \) be an interest rate model, i.e., a set of \((\sigma, \lambda)\)-pairs, specifying the conditional distribution of all rates

\[ 5\text{Other desirable empirical properties of the NS yield curve are discussed in Diebold and Li (2006).} \]
given through (3.2.1), given any arbitrary initial curve \( y_0 \). We now ask which kinds of curves \( y(t, \tau) \) the model \( M \) will generate for an initial curve given by \( y(0, \tau) \) over the interval \( t \in (0, T) \)? More specifically, suppose the initial curve is selected from some parameterized function family \( \tau \mapsto y(z, \tau) \), where \( z \) is a parameter vector chosen from a parameter set \( Z \subseteq \mathbb{R}^d \), in which case, the curve family is given by \( \mathcal{Y} = \{ y(z, \cdot) : z \in Z \} \).

The interest rate model \( M \) and the curve family \( \mathcal{Y} \) are then said to be consistent if all subsequent curves \( y_t \) for \( t \in (0, \tau] \) belong to the same family \( \mathcal{Y} \) whenever the initial curve \( y_0 \) does so (of course, future curves may correspond to different parameter values \( z \) than \( y_0 \)). In practice, it is common to fit yield curves from parameterized classes, so the question of consistency with relevant DTSMs is natural.

BC provide necessary and sufficient conditions for mutual consistency of a given parameterized family of forward curves and the dynamics of a given interest rate model (see (A.1.4) and (A.1.5) in Appendix A.1.5). Based on these conditions the issue addressed in the present paper is: if panel data on interest rates is constructed by cross-sectional curve fitting on, say, a daily or weekly basis to coupon bonds and then used for estimation of DTSMs, then biases may conceivably occur if the curve fitting procedure imposes a shape that is not consistent with the volatility function, i.e., DTSM of interest. Hence, we investigate the empirical implications of consistency and inconsistency between the cross-sectional curve fits and the DTSMs used.

### 3.2.3 DTSMs and yields curves under consideration

Although the consistency conditions appear complex in the general version, the curve families consistent with several important DTSMs are quite manageable. We consider first the one-factor model developed by Ho and Lee (1986) (henceforth HL) with flat forward rate volatility function given by \( \sigma_f(t, \tau) = a_0 \). The relationship between yield and forward rates (see (A.1.3) in Appendix A.1.1) implies that the yield volatility function coincides with the forward rate volatility function \( \sigma(t, \tau) = a_0 \). Hence, the yield dynamics in the HL model are parameterized by a single volatility parameter, \( k_1 = 1 \), satisfying the SDE of the form

\[
d y(t, \tau) = \alpha(t, \tau) \, dt + a_0 \, dW_t^1,
\]

where \( W_t^1 \) is a one-dimensional Wiener process and \( a_0 > 0 \). The flat volatility term can be interpreted as a long-run factor, which shifts all maturity yields equally. Due to the simple constant volatility, bond prices are log-normally distributed. BC show that the HL model is inconsistent with the NS yield curve family. However, they introduce a class of restricted exponential-polynomial (REP) yield curve families,\(^6\) and show that the family of yield curves denoted by REP\((1, (1), (0))\) in their nomenclature given

\(^6\)For a definition of the REP yield curve family, we refer to Appendix A.1.2, and for further details on this particular curve family, we refer to BC.
3.2. Modeling Framework

by

\[ y(z_t, \tau) = z_{1t} + z_{2t} \frac{\tau}{2}, \]  

(3.2.5)

for varying values of the parameters \( z_1 \) and \( z_2 \), is consistent with the HL model.\(^7\)

Even though affine yield curves do not imply that forward curves and discount functions are affine, they are nonetheless rather special. A more general one-factor DTSRM is the one developed by Vasicek (1977) and extended by Hull and White (1993) (henceforth HW) with forward rate volatility function \( \sigma_f(t, \tau) = a_1 \exp(-a_2 \tau) \) exponentially decaying for longer maturities. This generates the yield volatility function \( \sigma(t, \tau) = a_1 (1 - \exp(-a_2 \tau))/a_2 \tau \) in which case yields are parameterized by two volatility parameters, \( k_2 = 2 \), and follow the SDE given by

\[ dy(t, \tau) = \alpha(t, \tau) dt + a_1 \frac{1 - e^{-a_2 \tau}}{a_2 \tau} dW_1^1, \]  

(3.2.6)

with \( a_1, a_2 > 0 \). This model has a volatility structure that is exponentially decaying in term to maturity and a mean reverting short rate \( r(t, 0) \) with mean reversion parameter \( a_2 \).\(^8\) BC show that the HW model is inconsistent with the NS yield curve family, but consistent with the family of yield curves denoted by \( \text{REP}(3, (0, 0, 0), (0, a_2, 2a_2)) \), i.e.,

\[ y(z_t, \tau) = z_{1t} + z_{2t} \frac{1 - e^{-a_2 \tau}}{a_2 \tau} + z_{3t} \frac{1 - e^{-2a_2 \tau}}{2a_2 \tau}, \]

where the \( z \)-parameters may vary across calendar dates. For a two-factor model, we consider the specification explored by HJM, which we will refer to as the HJM2 model. Here, the yield dynamics are parameterized by three, \( k_3 = 3 \), volatility parameters

\[ dy(t, \tau) = \alpha(t, \tau) dt + a_0 dW_1^1 + a_1 \frac{1 - e^{-a_2 \tau}}{a_2 \tau} dW_2^2, \]

where \( W_1^1 \) and \( W_2^2 \) are independent Wiener processes and \( a_0, a_1, a_2 > 0 \). The model has a flat and an exponentially decaying volatility term, thus encompassing both the HL and HW model. The HJM2 model is consistent with the \( \text{REP}(3, (1, 0, 0), (0, a_2, 2a_2)) \) yield curve family given by

\[ y(z_t, \tau) = z_{1t} + z_{2t} \frac{1}{2} \tau + z_{3t} \frac{1 - e^{-a_2 \tau}}{a_2 \tau} + z_{5t} \frac{1 - e^{-2a_2 \tau}}{2a_2 \tau}. \]

\(^7\)The spot rate in the \( \text{REP}(1, (1), (0)) \) yield curve family is determined by the limiting value \( \lim_{\tau \to 0} y_t(z, \tau) \). This implies that we can think of the parameter \( z_{1t} \) as the short rate.

\(^8\)Empirical studies have shown that negative values of the mean reversion coefficient \( a_2 \) are by no means uncommon. Amin and Morton (1994) estimate implied volatility of forward curves from Eurodollar futures and option data and find that the estimated value of the mean reversion coefficient \( a_2 \) is negative on average. They conclude that the forward rate volatility structure is humped and that Eurodollar data shows the upwards sloping side of the hump.
For comparison purposes we also introduce a three- and four-factor model with deterministic volatility functions parameterized by five, $k_4 = 5$, and seven, $k_5 = 7$, parameters, respectively. The starting point of both models is the flat and the exponentially decaying factor of the HJM2 model. We will therefore refer to the two models as the extended HJM three-factor (EHJM3) model and the extended HJM four-factor (EHJM4) model. The EHJM3 model has dynamics given by

$$dy(t, \tau) = \alpha(t, \tau)dt + a_0 dW^1_t + a_1 \frac{1 - e^{-a_2 \tau}}{a_2 \tau} dW^2_t + a_3 \frac{1 - e^{-a_4 \tau} (a_4 \tau + 1)}{a_4^2 \tau} dW^3_t,$$

and the EHJM4 model has dynamics given by

$$dy(t, \tau) = \alpha(t, \tau)dt + a_0 dW^1_t + a_1 \frac{1 - e^{-a_2 \tau}}{a_2 \tau} dW^2_t + a_3 \frac{1 - e^{-a_4 \tau} (a_4 \tau + 1)}{a_4^2 \tau} dW^3_t$$

$$+ a_5 \frac{1 - e^{-a_6 \tau} (a_6 \tau + 1)}{a_6^2 \tau} dW^4_t.$$

In both cases, $a_i > 0$ for $i = 1, \ldots, 6$, and $W^i$ are mutually independent Wiener processes for $i = 1, 2, 3, 4$. The first and second factor have the same interpretations as in the HJM2 model. The third and fourth factor are obtained from an exponentially decaying forward rate volatility function multiplied by a linear state dependent term $\sigma_f(t, \tau) = a_i \tau \exp(-a_j \tau)$, where the exponentially decaying term affects the short-term forward rates the most, and the linear term affects the long-term forward rates the most. We introduce this linear state dependency in order to fit empirical observed curvatures in the volatility structure.

Using the same argumentation as in BC, consistent exponential polynomial families for the EHJM3 and EHJM4 models can be found. We show in Appendix A.1.2 that the EHJM3 model is consistent with the REP(5, (1, 0, 0, 1, 2), $(a_2, 2a_2, a_4, 2a_4)$) (henceforth REP3) yield curve family and the EHJM4 model is consistent with the REP(7, (1, 0, 0, 1, 2, 1, 2), $(0, a_2, 2a_2, a_4, 2a_4, a_6, 2a_6)$) yield curve family. Since we consider the REP3 yield curve family in the empirical application below, we state it here for completeness

$$y(z_t, \tau) = z_{1t} + z_{2t} \frac{1 - e^{-a_2 \tau}}{a_2 \tau} + z_{3t} \frac{1 - e^{-2a_2 \tau}}{2a_2 \tau} + (z_{5t} a_4 + z_{6t}) \frac{1 - e^{-a_4 \tau}}{a_4^2 \tau}$$

$$+ (z_{7t} 2a_4^2 + z_{8t} a_4 + z_{9t}) \frac{1 - e^{-2a_4 \tau}}{4a_4^3 \tau} - ((z_{8t} + z_{9t}) a_4 + z_{9t}) e^{-2a_4 \tau}.$$

The HL, HW, and HJM2 models all belong to the computationally tractable class of HJM models introduced by Ritchken and Sankarasubramanian (1995) (RS). However, Mercurio and Moraleda (1998) point out that by restricting the volatility to belong to the RS class of HJM models, it is not possible to choose the parameters of the model
3.3. Estimation of DTSMs

in such a way that the volatility is humped and stationary. Hence, hump shaped volatility, stationarity, and the computationally tractability of RS cannot be fulfilled within one model. As we will not consider option pricing issues in this paper, we do not fear looking at a model outside the RS class. Furthermore, we stress that the three- and four-factor models in general are computationally demanding.

In order to analyze the consequences for empirical analysis of the choice of yield curve family implemented in the fitting procedure, we will consider various yield curve families in the empirical application below. In addition to the NS and REP3 yield curve families, we will consider the Augmented-Nelson-Siegel (ANS) yield curve family of BC as well as the Nelson-Siegel-Svensson (NSS) yield curve family introduced by Svensson (1995) defined below. Starting from a general parametric yield curve given by

\[ y(z_t, \tau) = z_{1t} + z_{2t} \frac{1 - e^{-a_2 \tau}}{a_2 \tau} + z_{3t} \left( \frac{1 - e^{-a_2 \tau}}{a_2 \tau} - e^{-a_2 \tau} \right) + z_{4t} \frac{1 - e^{-2a_2 \tau}}{2a_2 \tau} + z_{5t} \left( \frac{1 - e^{-a_4 \tau}}{a_4 \tau} - e^{-a_4 \tau} \right), \]

which reduces to the NS yield curve (3.2.4) for \( z_{4t} = z_{5t} = 0 \), we note that BC and Filipović (1999) find that no term structure model generates yield curves that belong to the NS yield curve family. However, by augmenting the NS yield curve family with an additional slope factor, BC prove consistency with the HW model. Thus, \( z_{5t} = 0 \) in (3.2.8) defines the ANS yield curve family. Next, the NSS yield curve family is obtained by setting \( z_{4t} = 0 \) in (3.2.8), which then, compared to the simpler NS yield curve family, allows for additional flexibility by adding a second curvature term.

### 3.3 Estimation of DTSMs

In order to apply the HJM framework to empirical data, we discretize the SDE given by (3.2.1) in calendar time \( t \) using the Euler approximation scheme as well as in term to maturity \( \tau \), producing the following discrete time yield dynamics

\[ y(t_i, \tau_j) - y(t_{i-1}, \tau_j) = \frac{y(t_{i-1}, \tau_j) - y(t_i, 0)}{\tau_j - \tau_0} \Delta_i + \frac{\partial y(t_{i-1}, \tau_j)}{\partial \tau_j} \Delta_i \]

\[ + \gamma_j \Delta_i + \sigma_j' z_i + \varepsilon_{ij}, \]

with a discrete version of the HJM no-arbitrage drift restriction

\[ \gamma_j \equiv \sigma_j' \lambda + \frac{\tau_j}{2} \sigma_j' \sigma_j, \quad j = 1, \ldots, M. \] (3.3.2)

In (3.3.1) the discretization uses the calendar dates \( t_1, t_2, \ldots \) with \( \Delta_i = t_{i+1} - t_i \) denoting the calendar time increment, where we assume equally spaced calendar time
intervals, $\Delta_i = \Delta$. For each calendar date we consider a fixed and finite set of terms to maturity $\{\tau_1, \tau_2, \ldots, \tau_M\}$. In the discrete time approximation, $\sigma_j(\tau) = \sigma_j$ is the $m$-dimensional volatility vector, $\lambda$ is the $m$-dimensional market price of risk vector, $\epsilon_{ij}$ is an error term, and we model $z_i$ as a normal increment.

Let now $\sigma$ be the $M \times m$-matrix, where the $j$th row vector is $\sigma_j'$, let further $\epsilon_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,M})'$ be the vector of noise terms, and $\gamma = (\gamma_1, \ldots, \gamma_M)'$ the vector of drift terms. Furthermore, consider the $M$-dimensional vector $\tilde{y}_i$ with the $j$th entry given by

$$\tilde{y}_{ij} \equiv y(t_i, \tau_j) - y(t_i, \tau_j) - y(t_i, 0)$$

which is the vector of yield changes adjusted for both average and local slope of the yield curve. Christensen and van der Wel (2019) show that in the ideal case with complete data for all maturities we can think of $-\tau_j \tilde{y}_{ij}$ as the one-period excess return to the discount bond with maturity $\tau_j$ at time $t_i$. Rewriting (3.3.1) we obtain

$$\tilde{y}_i = \gamma \Delta + \sigma z_i + \epsilon_i, \quad i = 1, \ldots, n,$$

where $n$ denotes the number of time series observations on the $M$ slope-adjusted yield changes. Since the volatility is time invariant and deterministic, we can estimate (3.3.4) by factor analysis. From this point of view, it is natural to write down the more general model without the no-arbitrage restriction imposed, that is

$$\tilde{y}_i = \mu + \sigma z_i + \epsilon_i, \quad i = 1, \ldots, n,$$

where $\mu$ is an $M$-vector of slope-adjusted mean yields. In the classical factor analysis the $\sigma$-matrix is allowed to vary freely. That is, if $\sigma$ is an $M \times m$-matrix we estimate $M \times m$ parameters. In the following, we refer to this volatility specification as the general $m$-factor model, or in short the GF$m$ model, and we will sometimes refer to the columns of $\sigma$ as factor loadings. As standard in the classical factor analysis we impose the following model assumptions

$$E(z) = 0_{m \times 1}, \quad \text{var}(z) = I_m,$$

$$E(\epsilon) = 0_{M \times 1}, \quad \text{var}(\epsilon) = \Psi = \text{diag}(\psi_{11}, \ldots, \psi_{MM}),$$

$$\text{cov}(z, \epsilon) = 0_{m \times M},$$

where $I_m$ denotes the $m$-dimensional identity matrix. The assumption (3.3.6) comes naturally due to the Gaussian model starting point, where $\text{var}(z) = I_m$ is without loss.

---

9 Our empirical work uses $\Delta$ equal to one week.

10 The assumption of deterministic volatility implies a Gaussian model. However, because of the use of Musiela parametrization, the Euler approximation (3.3.1) is no longer exact, which is why we need the error term.
of generality, since factor variance and covariances could be absorbed in the factor loadings $\sigma$. Assumption (3.3.7) implies orthogonal factors. In what follows, we refer to the diagonal elements $\psi_{ii}$ as unique variances as they explain the variability in $\tilde{y}_i$ not shared with other variables, and write

$$\text{var} (\tilde{y}) \equiv \Sigma \equiv \sigma \sigma' + \Psi$$

for the covariance matrix. Thus, we can estimate (3.3.5) under assumptions (3.3.6)-(3.3.8) by maximum likelihood factor analysis. We refer to this model as the unrestricted $m$-factor model and the model given by (3.3.4), where we impose the HJM drift restriction (3.3.2), as the restricted $m$-factor model.

The general log-likelihood function takes the form

$$\ell (\tilde{y}_1, \ldots, \tilde{y}_n; \theta) = -\frac{n}{2} \log |2\pi\Sigma| - \frac{1}{2} \sum_{i=1}^{n} (\tilde{y}_i - \mu)' \Sigma^{-1} (\tilde{y}_i - \mu)$$

$$= -\frac{n}{2} \log |2\pi\Sigma| - \frac{n}{2} \text{tr} \Sigma^{-1} S - \frac{n}{2} (\bar{y} - \mu)' \Sigma^{-1} (\bar{y} - \mu),$$

(3.3.9)

where $S$ is the sample covariance matrix and $\theta$ denotes the vector of model parameters. We consider the following versions of this log-likelihood function:

(i) Without further restrictions on $\mu$ and $\Sigma$, this is the classical factor analysis log-likelihood function. Here, the maximum likelihood estimate of $\mu$ is the (time series) sample average $\bar{y}$. This model can be estimated for different values of $m$ to find the appropriate number of common factors.

(ii) For any specification of the volatility matrix $\sigma$, we may impose the no-arbitrage restriction $\mu = \gamma \Delta$. As discussed in the next section, we may estimate the model with and without this restriction and test for no-arbitrage in a LR test. The estimation with the restriction imposed is no longer a special case of the classical factor analysis.

(iii) With or without the no-arbitrage restriction imposed, we may consider special parameterized functional forms for the volatility matrix corresponding to classical DTSMs.

In the empirical work below, we furthermore consider various versions of (i)-(iii). For all such analyses the starting point is a panel data set of yields, $\{y(t_i, x_j)\}_{ij}$. The focus of the present paper is then the following: What are the consequences for empirical analysis of DTSMs — such as moving through the steps (i)-(iii) — of the choice of curve fitting procedure implemented in order to build the data set $\{y(t_i, \tau_j)\}_{ij}$? For example, if a curve fitting procedure based on estimating four parameters on each calendar date is used to generate the initial data set, does it make it more likely that the classical factor analysis (i) chooses four factors? Does it make it more likely that a specially structured DTSM (a $\sigma$-specification) with four (rather than fewer) factors is
preferred? Does it make the test for no-arbitrage more or less likely to be rejected than when using alternative initial curve fitting procedures? In general, what is the relation between curve fitting procedure and resulting statistical inferences? We investigate this issue, and specifically the role of consistency between the fitted curve shape and the DTSM.

In the remainder of this section, we add some details on the statistical analysis to be conducted in steps (i)-(iii) above. For (i), the mean vector may be concentrated out of the classical factor analysis, since there are no cross-restrictions on $\mu$ and $\Sigma$ when the no-arbitrage condition is not imposed. This leaves the concentrated (partially maximized) log-likelihood function

$$
\ell_U(\tilde{y}_1, \ldots, \tilde{y}_n; \theta) = -\frac{n}{2} \log|2\pi \Sigma| - \frac{n}{2} \text{tr} \Sigma^{-1}S,
$$

(3.3.10)

to be maximized with respect to $\sigma$ and $\Psi$. The classical factor analysis is invariant to orthogonal rotation, i.e., if an $m$-factor model holds for (3.3.5), then it also holds if $\sigma$ is rotated by an orthogonal $m \times m$-matrix $A$ as $\sigma A(\sigma A)' = \sigma \sigma'$. This problem of indeterminacy is usually solved by restricting $\sigma' \Psi^{-1} \sigma$ to be diagonal, which reduces the number of free parameters in $\sigma$ from $Mm$ to $Mm - m(m - 1)/2$. These may be concentrated out of the log-likelihood function, too, for a given value of $\Psi$. Finally, the latter is estimated by maximizing the concentrated log-likelihood function with respect to these $M$ parameters, only. The restricted volatility structures that are considered in (iii) are also invariant to rotation, but in those models, the parameters are already sufficiently restricted such that they are identified without further specifying the rotation.

For (ii), imposing the no-arbitrage condition, we replace the sample mean $\mu$ with the HJM drift $\gamma \Delta$ of (3.3.2), in which case the log-likelihood function becomes

$$
\ell_R(\tilde{y}_1, \ldots, \tilde{y}_n; \theta) = -\frac{n}{2} \log|2\pi \Sigma| - \frac{1}{2} \sum_{i=1}^{n} (\tilde{y}_i - \gamma \Delta)' \Sigma^{-1} (\tilde{y}_i - \gamma \Delta)
$$

$$
= -\frac{n}{2} \log|2\pi \Sigma| - \frac{1}{2} \sum_{i=1}^{n} (\tilde{y}_i - \tilde{y})' \Sigma^{-1} (\tilde{y}_i - \tilde{y}) - \frac{1}{2} \sum_{i=1}^{n} (\tilde{y} - \gamma \Delta)' \Sigma^{-1} (\tilde{y} - \gamma \Delta)
$$

$$
= -\frac{n}{2} \log|2\pi \Sigma| - \frac{n}{2} \text{tr} \Sigma^{-1}S - \frac{n}{2} (\tilde{y} - \gamma \Delta)' \Sigma^{-1} (\tilde{y} - \gamma \Delta),
$$

(3.3.11)

which is maximized with respect to $(\sigma, \Psi, \lambda)$. For the case of a general volatility matrix $\sigma$, we still restrict $\sigma' \Psi^{-1} \sigma$ to be diagonal, and $(\sigma, \Psi, \lambda)$ are estimated iteratively. For the case of a parameterized volatility matrix, as in (iii), again $(\Psi, \lambda)$ and in this case the parameters in $\sigma$ are estimated jointly. However, we show in Appendix A.1.4 that the maximum likelihood estimator for $\lambda$ in this case is given by

$$
\hat{\lambda} = (\delta' \hat{\Sigma} \delta)^{-1} (\Delta^{-1} \delta' \hat{\Sigma}^{-1} \tilde{y} - \delta' \hat{\Sigma}^{-1} \delta),
$$
3.4 Test for no-arbitrage in DTSMs

where $s$ is the $M$-vector given by

$$
\hat{s} \equiv \left[ \frac{T_1}{2} (\hat{\sigma}_{11}^2 + \ldots + \hat{\sigma}_{1m}^2) \ldots , \frac{T_M}{2} (\hat{\sigma}_{M1}^2 + \ldots + \hat{\sigma}_{Mm}^2) \right]'.
$$

If $M = m$ such that $\sigma$ becomes a quadratic $M \times M$-matrix and if, in addition, $\sigma$ is invertible, this expression can be reduced to

$$
\hat{\lambda} = \hat{\sigma}^{-1} (\Delta^{-1} \bar{y} - \hat{s}),
$$

which can be substituted back into (3.3.11) such that the estimation problem can be treated in terms of $(\Psi, \sigma)$ only.

For (iii), when a parameterized $\sigma$ is introduced for the analysis of a structured volatility matrix, this is simply substituted into the log-likelihood function, with or without the no-arbitrage condition on the means imposed. For example, in the case of the HW model

$$
\sigma_j = a_1 \frac{1 - e^{-a_2 \tau_j}}{a_2 \tau_j}, \quad \sigma = \left( a_1 \frac{1 - e^{-a_2 \tau_1}}{a_2 \tau_1} , \ldots , a_1 \frac{1 - e^{-a_2 \tau_M}}{a_2 \tau_M} \right)',
$$

is substituted into the log-likelihood function, which, if the no-arbitrage condition is imposed, is maximized with respect to $M + 3$ parameters given by $(\Psi, \lambda, a_1, a_2)$, or, without this restriction, with respect to $2M + 2$ parameters given by $(\Psi, \mu, a_1, a_2)$. In the latter case, the sample average $\bar{y}$ remains the maximum likelihood estimate of $\mu$ despite the special structure on $\sigma$.

3.4 Test for no-arbitrage in DTSMs

In the following, we present two statistical tests for estimating the HJM drift restriction given by (3.3.2). In particular, we test the following hypothesis

$$
H_0 : \gamma_j \equiv a_j' \lambda + \frac{\tau_j}{2} a_j' \sigma_j a_j, \quad j = 1, \ldots, M.
$$

We test the hypothesis with a standard LR test and a Wald test. The details are presented in Section 3.4.1 and 3.4.2, respectively. The test is basically a test for no-arbitrage within the HJM model setup, and we will refer to it as a test for no-arbitrage.

3.4.1 The likelihood-ratio test

The standard LR test is done by estimating both the unrestricted and restricted $m$-factor model and comparing the log-likelihood values given by (3.3.10) and (3.3.11), respectively.

Assuming that the estimates of $\sigma$ and $\Psi$ are approximately the same for both models, we basically compare the difference between the sample mean, $\bar{y}$, and the estimate
of the HJM drift term $\gamma \Delta$.\footnote{We point out that $\gamma \Delta$ is the drift of the slope-adjusted yield changes $\tilde{y}$ and not the discrete time drift of the yields themselves. However, we will sometimes refer to $\gamma \Delta$ as the HJM drift or just the drift rather than the drift of $\tilde{y}$.} In other words, we test whether we can approximate the sample mean by the HJM drift term. The HJM drift term depends, in addition to the volatility estimate, on the estimated market price of risk. Hence, in advance we would expect the estimate of the market price of risk to affect the outcome of the test. Denoting the likelihood value of the restricted and unrestricted model by $L_R$ and $L_U$, respectively, the LR test is given by

$$LR = 2(\log L_U - \log L_R) = 2(\ell_U - \ell_R),$$

which is approximately $\chi^2$-distributed with $M - m$ degrees of freedom corresponding to $M$ parameters saved in $\mu$ by imposing the HJM drift restriction, whereas the $m$-dimensional market price of risk parameter $\lambda$ is a new parameter introduced in the restricted model.

### 3.4.2 The Wald test

As pointed out earlier, the restricted model is more cumbersome to estimate. In an attempt to reduce the computational work, we present an alternative test for no-arbitrage in which we only need the estimates $\hat{\sigma}$ and $\hat{\Psi}$ of the unrestricted model. The starting point of this test is to estimate $\mu$ by the sample mean $\bar{y}$. Let then $\tilde{y}_i$ be the $j^{th}$ entry in $\tilde{y}$, and form observations given by

$$z_j = \frac{\tilde{y}_j}{\Delta} - \frac{\tau_j}{2} \hat{\sigma}_j' \hat{\sigma}_j, \quad j = 1, \ldots, M.$$  

To estimate the market price of risk, we consider the following regression equation based on (3.3.2)

$$z_j = \hat{\sigma}_j' \lambda + \epsilon_j, \quad j = 1, \ldots, M. \quad (3.4.2)$$

We can estimate (3.4.2) by feasible generalized least squares, in which case the estimate of the market price of risk is

$$\hat{\lambda} = (\hat{\sigma}' \hat{\Sigma}^{-1} \hat{\sigma})^{-1} \hat{\sigma}' \hat{\Sigma}^{-1} z.$$  

The Wald test for no-arbitrage is then given by

$$W = n \epsilon' \hat{\Sigma}^{-1} \epsilon,$$

which is approximately $\chi^2$-distributed with $M - m$ degrees of freedom, and $\epsilon = (\hat{\epsilon}_1, \ldots, \hat{\epsilon}_M)'$ is the vector of residuals from the estimation (3.4.2).
3.5 Empirical application of DTSMs

In this section, we fit the different volatility structures to yield data. That is, we try to explain the correlation structure in that data. Then, we estimate the no-arbitrage mean restriction and test this restriction as described in Section 3.4. In doing this, we are basically fitting to the sample mean and testing whether the HJM model drift specification fits the data. In addition, for comparison purposes, we fit the general factor models to the data and test for absence of arbitrage within this model setup. Furthermore, we investigate, whether the manner of estimating the yield rate data has any impact on the estimation. In particular, we look at the effect of consistency between the yield curve family used for yield curve fitting and the interest rate model used for estimation.

3.5.1 Data and summary statistics

Our data consists of weekly US Treasury bills, notes, and bonds from CRSP. The sample covers the first week of 1999 through the last week of 2014 for a total of 835 calendar dates (weeks). We include all non-callable bonds and Treasury bills with maturities from three months up to forty years. The number of bonds on each calendar date (week) ranges from 135 to 323 (see Figure 1). We estimate the yield curve for each calendar date over the term to maturity interval from zero to thirty years. Specifically, on a given date, each observed bond price is written as the discounted sum of future coupon payments and face value. The discount function is written in terms of yields, which in turn are parameterized functions given by the chosen curve shape, and the sum of squared pricing errors is minimized with respect to parameters in a non-linear least squares application.

As input to the subsequent data analysis, we use yields with maturities of 0.25, 0.5, 1, 2, 3, 5, 7, 10, 15, 20, and 25 years. Thus, for each curve fitting procedure, these eleven points are read off the fitted yield curves for each calendar date, thus producing the required panel data set for further empirical investigation. Summary statistics for all selected curve fitting procedures are shown in Table 1. For each of the eleven maturities, the sample mean and standard deviation of the fitted yields are shown. All yield curves are, on average, strictly upward sloping from around 2% to 4.9%. Only, the additional flexibility in the REP3 shape allows the yield curve to bend slightly downwards in the long-end. The standard deviations show that the short-ends of the yield curves vary more over time than the long-ends.

In order to form the vector of slope-adjusted yield changes \( \hat{\gamma}_i \) given by (3.3.3), we use the yield rate with the shortest maturity \( \tau_1 \) as proxy for \( \gamma(t,0) \) (the three-months bill yield is a good proxy for the short rate, see Chapman et al. (1999)) whereby we lose one cross-sectional dimension. We use analytical expressions (we provide an expression for the analytical derivative of the most complex yield curve under consideration - the
REP3 yield curve family - in Appendix A.1.5) to calculate derivatives of the parametric curves. The sample mean and standard deviation of the resulting panel data variables are tabulated in Table 2. As for the raw yields, we see that the results differ by curve fitting method. Whereas the sample mean of the NS slope-adjusted yield changes is u-shaped, it is strictly upward sloping for the remaining yield curves. Considering standard deviations, we notice that the medium-term varies most. Comparing across curve fitting methods, we notice that the variation for NS, ANS, and NSS slope-adjusted yield changes is larger in the long-end than for the REP3 yield curve.

In Figure 2, we provide a three-dimensional plot of the raw yields and the slope-adjusted yield changes obtained by using the NS yield curve in the fitting step. The scale of the vertical axis differs between the two panels as the slope-adjusted yield changes are of smaller magnitude. The level factor dominates in both cases the slope and curvature factor in that the curves are flat in maturity dimension (we do not provide corresponding three-dimensional plots for the remaining yield curves since the visual difference between them is negligible).

Overall, this indicates differences between the yield curve estimation procedures in the resulting yield curves. In the following, we will investigate how these differences in the yield curve fits are reflected in the data analysis, i.e., we investigate the implications of these differences for inference on the number of factors, the shape of the volatility function, and tests for no-arbitrage.

3.5.2 Estimation of general factor model volatility by principal factor analysis

For each of the four data sets corresponding to the different curve fitting procedures, we first estimate the volatility of the unrestricted general factor models by principal factor analysis (PFA), which is a simpler type of analysis than the maximum likelihood approach presented in Section 3.3. PFA is basically principal component analysis, where the diagonal entries of the correlation matrix are replaced with the so-called communalities, representing the proportion of the variance of the individual variables to be explained by the factor analysis (for details, see Mardia, Kent and Bibby (1979)). The results of the PFA applied to the reduced correlation matrix are summarized in Table 3, which shows for NS data that the first factor accounts for 75.26% of the total variation, whereas two and three factors account for 93.72% and 98.90%, respectively. This result is similar to results found by others e.g. Litterman and Scheinkman (1991). The results for ANS data are similar to the NS case, but for

---

12 As factor analysis is scale invariant, we can apply factor analysis to the sample correlation matrix $R$ instead of the covariance matrix $S$.

13 In particular, Litterman and Scheinkman (1991) find that most variation in returns on all fixed income securities can be explained by three factors. They find that the first factor explains 89.5% of the total variance, and that the second factor explains 81% of the remaining variation in returns.
3.5. Empirical application of DTSMs

REP3 and NSS data the first factors explain less of the variation. This suggests that the NS-procedure, which builds in a small number of factors through the curve fitting procedure, potentially makes the data look more like the outcome of a process with only few driving factors.

The estimated volatility structure corresponding to the first three factors is shown in Figure 3. In existing literature, the first factor is usually found to be roughly flat and equally weighted across maturity corresponding to a parallel shift in the term structure. However, as it can be seen from Figure 3, the first factor is hump shaped. The second and third factor resemble the patterns found in existing literature more. The second factor is downwards sloping except in the very short- and long-end and changes signs from being positive in the short-end to negative in the long-end. This factor causes the term structure to tilt. The shape of the third factor is in contrast positive in the short- and long-end but negative in the middle. This ‘reversed hump shaped’ factor causes the term structure to flex. Rebonato (1996) interprets the first three factors as the average level of the yield curve, the slope of the yield curve, and the curvature of the yield curve.\(^{14}\)

3.5.3 Maximum likelihood estimation of factor models

Here, we estimate the volatility structure of the unrestricted factor models using maximum likelihood factor analysis. Figure 4-5 show the estimated volatility structures of the four data sets. In addition, they illustrate the differences in the volatility structure between the general and parameterized models. Overall, the factors of the estimated volatility structures across the general factor models of the four curve fitting procedures are almost indistinguishable and similar to those estimated by PFA. The level-slope-curvature structure is evident for all data sets, and the shape of the factors is generally more concave/convex than those estimated by PFA, especially for the one- and two-factor models. Increasing the number of factors to four results in an indistinguishable volatility structures between the two estimation methods.

Now we estimate the volatility structures of the parameterized models. Hence, we reduce the number of volatility parameters compared to the general models and impose parametric restrictions on the shape of the volatility structure, where we estimate all volatility parameters under a non-negativity constraint. The overall volatility structures are comparable across the four data sets, so there is no need to discuss them individually. Instead, we discuss the volatility structure along the lines of the NS data and explain briefly how the four data sets differ from another.

The NS data estimates appear in Panel A of Table 4. In the HW model the estimate of \(\alpha_2\) is zero. This implies a flat HW volatility structure, and, as a result, the estimated

\(^{14}\)Strictly speaking Rebonato (1996) addresses principal component analysis. However, the interpretation is the same for PFA.
volatility structures of the HL and HW models align. A possible explanation for this can be found in the upper-left of Figure 4, which illustrates the estimated volatility structures of the unrestricted one-factor models for the NS data. The estimated volatility of the GF1 model is hump shaped. Hence, the constant volatility structure of the HL model can be interpreted as an overall average of this humped factor. However, the exponentially decaying volatility of the HW model does not fit this hump shaped factor at all. Looking at the humped shape of the estimated volatility of the GF1 model, it seems reasonable that a constant volatility fits the data better than an exponentially decaying one. This explains why the estimated volatility of the HL and HW models become equivalent.

The upper-middle of Figure 4 illustrates the differences in the volatility structures of the GF2 and HJM2 models. We have changed the signs of some factors of the general factor model in order to visually compare the estimated volatility of the two models. Similar to the HL/HW model, the first hump shaped factor of the GF2 model is replaced by a flat first factor by construction. The second hump shaped factor of the GF2 model is replaced by an exponentially decaying second factor. The second factor of the HJM2 model is positive by construction, whereas the second factor of the GF2 model shifts signs and is positive in the short-end and negative in the long-end.

From the upper-right and lower-left of Panel A can be seen that the volatility structure of the EHJM3 model, in addition to the flat and exponential factors of the HJM2 model, has a third hump shaped factor. The first factor, i.e., the level factor, is slightly humped in the GF3 model and flat by construction for the EHJM3 model. The second factor, i.e., the slope factor, is decreasing except in the very short-end and changes signs for the GF3 model and is exponentially decaying and positive by construction for the EHJM3 model. The third factor, i.e., the curvature factor, is humped and positive by construction for the EHJM3 model, but changes sign and has a turned around s-shape for the GF3 model. This shape cannot be replicated by the EHJM3 model.

Lastly, the lower-middle and lower-left of Panel A illustrate the differences in the estimated volatility structures between the GF4 model and the EHJM4 model. Remarks similar to those made about the EHJM3 model can made about the first three factors. The fourth factors mimics the third factors, albeit less curly.

Figure 6 shows the differences in the unique variances for all models. The unique variances decrease with the numbers of factors. This is because, as we increase the number of factors, we increase the variability explained by the factors. Secondly,

\footnote{Changing the signs does not change the overall result, as \( \sigma \) enters only in the log-likelihood function as \( \sigma \sigma' \) except when we impose the no-arbitrage drift restriction, where changing the sign of \( \sigma \) will change the sign of the risk premium.}
the overall shape of the unique variances of the general factor models differs from those of the parameterized volatility models. The latter is humped in the middle with upward sloping tails, and the first is more or less decreasing with maturity. What is interesting, though, is that the unique variances of the general factor models and the parameterized volatility models is of the same size for a given number of factors, implying that we do not lose very much in using the parameterized models in terms of unique factors. This is in spite of the fact that the restrictions that the parametrization puts on the volatility; the parameterized volatility models do not capture the skews and humps of the volatility structures of general factor models. The difference between the models is therefore mostly due to the estimation of the volatility.

The general factor and parameterized models are nested, so we can test the parametric restrictions using a standard LR test. The classical factor analysis has $Mm - m(m - 1)/2$ free parameters in $\sigma$. In the parameterized models $\sigma$ is parameterized by only $k_i < Mm - m(m - 1)/2$ parameters. Thus, the difference in degrees of freedom between the general and parameterized, is $Mm - m(m - 1)/2 - k_i$. We report the log-likelihood function values and the LR test statistic for the NS data in Panel A of Table 6. In agreement with the visual inspection the estimated volatility structures differ significantly. We can conclude that the volatility structure of the parameterized models is not flexible enough to generate the humped shaped factors of the general factor models.

The estimated volatility structures for the ANS data set appear in Panel B of Figure 4, and the REP3 and NSS data sets are illustrated in Figure 5. Also here, the level-slope-curvature structure is clearly present in all cases. However, even though the overall volatility structures are comparable to the structures in the NS case, there are markedly differences between the individual factor loadings. To mention a few, we note that the loading on the second factor in the parameterized volatility models is considerably flatter in the REP3 data than for NSS data and the maturity at which it intersects the loading on the first factor varies between seven months and five years. Also the maturity at which the loading on the third factor achieves its maximum varies considerably between the data sets.

Thus far, we have looked at the fit to data in terms of the correlation. Recall that when estimating $\sigma$ and $\Psi$, we are basically trying to fit to the sample covariance matrix. However, another important issue is whether the HJM drift specification fits the sample mean $\bar{y}$. Essentially, in doing this we examine the existence of arbitrage opportunities within the model. Clearly, the no-arbitrage model property can be considered just as or more important than the fit to the correlation structure, which is addressed in the following section.
3.5.4 No-arbitrage drift restriction within parameterized models

In this section, we test for absence of arbitrage within the parameterized volatility models. We assume, for example, that the HL model holds and test the no-arbitrage drift restriction within this model. We apply both the LR and Wald test as described in Section 3.4.

To calculate the LR test statistic, we need to estimate the restricted models, imposing the no-arbitrage constraint on the drift. The volatility parameters of the restricted models are given in Table 5. Their values reveal that the difference between the volatility structures of the unrestricted models and the restricted models are very small. Based on these, we estimate the market price of risk by regression, which in turn is used in the Wald test statistic. For each of the four different curve fitting procedures, the LR test statistics, the Wald test statistics, and their test probabilities are given in Table 6. The two test statistics for the no-arbitrage test are of similar size — the LR test statistic in most cases being the smallest — and the conclusions drawn from them are the same. We stress that the LR test statistic should be smaller than the Wald test statistic by construction. This is due to the fact that by estimating the market price of risk by regression, we are actually maximizing over $\lambda$ after maximizing over $\sigma$ and $\Psi$ parameters. Whereas estimating the restricted factor model by maximum likelihood analysis, we maximize over $\lambda, \sigma$, and $\Psi$ parameters simultaneously. However, the size of the two test statistics and the conclusions drawn from them suggest that the Wald test provides a good approximation to the LR test. This is an important observation as the Wald test is much simpler to compute in that it requires only estimation of the unrestricted model. We stress that the Wald test is only to be seen as an alternative to and not as a replacement of the LR test.

For NS data, we see that the test is rejected only for the one-factor HL/HW model at a 5% significance level and is fails to reject the no-arbitrage drift restriction in the HJM2, EHJM3, and EHJM4 models. The same results are obtained for NSS data. As discussed in Section 3.2.2, the ANS family is consistent with the HW model, the REP3 family is the minimal consistent family of the EHJM3 model, and the HL, HW, and HJM2 models have minimal consistent families that are contained in this REP3 family. Consequently, the REP3 family is consistent with the HL, HW, HJM2, and EHJM3 models.

Whereas we for ANS data accept absence of arbitrage for the HJM2 and EHJM3 models, we only reject the test within the HL/HW model for REP3 data. This indeed confirms our prior argumentation for looking at the model and suggests that consistency of the model and the yield curve family matter. The HL and HW models are also consistent with the REP3 family, but the family is in no way close to the minimal consistent family of these models. In a way we are overfitting the models which could explain why the no-arbitrage hypothesis is rejected. Similar findings on consistent
non-minimal forward curve families can be found in Angelini and Herzel (2002). Also related is that Christensen and van der Wel (2019) test the HJM drift restrictions in a restricted factor model on unsmoothed Fama-Bliss yields. Imposing almost the same affine term structure restrictions on the volatility function with observable macro factors and time-varying market price of risk, they reject the HJM drift restrictions in most cases, which gives further credence to the importance of the curve fitting procedure implemented.

For each of the three curve fitting procedures, the estimated market prices of risk are given in Table 8. The unrestricted market price of risk is estimated by regression, and the restricted market price of risk is estimated by maximum likelihood analysis imposing the no-arbitrage restriction on the drift. Comparing the two estimates, we see that they are mostly negative, capturing the negative relation between yields and bond prices, and there are only small differences between the market prices of risk estimated by regression and by maximum likelihood.

The analysis conducted in Sections 3.5.2 and 3.5.3 showed that we need three factors to explain the volatility structure. However, this analysis considers only the fit to the correlation structure but not to the sample mean. Interesting is that in general, when taking the absence of arbitrage into account and hence fitting the sample mean by the HJM drift in addition to fitting the volatility structure to the correlation structure, less factors are needed to describe the data. Moreover, the exact number of factors needed depends on the chosen yield curve family. For NS, NSS, and REP3 data we need at least two factors to accept absence of arbitrage. Similarly, in case of ANS data the parameterized two- and three-factor models provide a good fit to the HJM drift, but the fit deteriorates significantly when using four factors.

3.5.5 No-arbitrage drift restriction within general factor models

Thus far our empirical analysis of the no-arbitrage drift restriction has been focused on the parameterized models only, i.e., models where we restrict the volatility in terms of shape and number of parameters. The results presented have varied between data sets and number of factors. Naturally, we would like to know if these results are model specific. To investigate this further, we test for absence of arbitrage in the general factor model. We apply both the LR and Wald test.

The test results of the LR and Wald test for no-arbitrage for NS data are shown in Table 9 for the general factor models. We see that the null hypothesis is accepted only in the four-factor model implying no-arbitrage. This differs from our findings in Section 3.5.4, where absence of arbitrage was rejected only in the case of a one-factor HL/HW model. This could indicate some kind of instability in estimating the general factor models, which could be caused by using an inconsistent yield curve family in the initial fitting step producing yield curves outside the chosen family. Evidently,
this instability is reduced by putting model specific restrictions on the shape of the volatility structure as in the parametric models. Here, we find a much better fit to the sample mean, and as a result, less factors are needed to accept absence of arbitrage. A similar conclusion can be drawn from the remaining three data sets, where the HJM drift restriction, in fact, is rejected for all general factor models (see Table 9).

Instability stemming from inconsistency is also observed by Angelini and Herzel (2002), who show that choosing a forward curve family, which is consistent with the interest model, has significant impact on the estimated model parameters. Their analysis is carried out on a one-factor HW model and the minimal consistent forward curve family of this particular model. Empirical yield curves are calibrated to swap data, and model parameters are calibrated to ATM cap prices. They show that the minimal consistent forward curve family results in more stable estimates, smaller pricing errors, and better forecasting capability compared to other forward curve families analyzed.

The results of the LR test for no-arbitrage are shown in Table 9. Additionally, the value of the log-likelihood function for the restricted and unrestricted models and the estimated market price of risk of the restricted models are shown. As the differences between the estimated volatility structure of the restricted and unrestricted models are almost indistinguishable, we do not include the restricted factor volatility structure in the paper. The test results and their implications are similar to the ones obtained when using the simpler Wald test. As previously, the LR test size is marginally smaller than the Wald test size suggesting that the Wald test is a good substitute for the LR test.

Additionally, comparing the left and the right panel of Table 7, we notice that the estimated market prices of risk of the restricted factor model differ only slightly from those estimated by regression.

### 3.5.6 Drift comparison

To further analyze the effect of adding more factors to the general factor model and the differences between the general factor models and the parameterized volatility type of models, we look at the ability of the HJM drift estimate to fit to the sample mean. Figure 7 illustrates the differences between the sample mean $\bar{y}$ and the HJM drift estimate $\hat{\gamma} \Delta$ of the general factor models and parameterized volatility models estimated by regression for NS data. We see that the more factors are included, the closer $\hat{\gamma} \Delta$ is to the sample mean.

Generally, we remark that the shape of the HJM drift estimates obtained from the parameterized volatility structures is less flexible than those estimated from the general factor models. This is not surprising, as parametrizations put restrictions on the shape of volatility structures, and the drift estimate is calculated from the
estimated volatility structures. Furthermore, none of the one-factor models is able to capture the shape of the sample mean, which is also not surprising, remembering that absence of arbitrage was rejected for both one-factor models. However, we remark that the drift estimate in the HL/HW model gives the best fit to the sample mean of the two.

For the two-factor models, the drift estimate in the HJM2 model is in general closer to the sample mean than the estimate of the general two-factor model, explaining why we accept absence of arbitrage within the HJM2 model and not in the general two-factor model. In effect, it is clear that even though the general factor model fits the correlation structure better than the HJM2 model, the latter model gives a better fit to the sample mean. The three-factor models provide a good fit to the HJM drift. Even though the general factor model fits the correlation structure better than the EHJM3 model, the two models fit the sample mean equally well. Lastly, the EHJM4 model drift estimate seems to fit the sample mean slightly worse than the drift estimate of the general four-factor model. Hence, the general factor model fits both the correlation structure and sample mean better.

Similar results apply for ANS, REP3, and NSS data. However, the fit to the sample mean for ANS data deteriorates in the EHJM4 model compared to the EHJM3 model, which explains why we reject no-arbitrage within the EHJM4 model.

Even though the estimated volatility of the general factor models fit the correlation structure better by construction, the parameterized models fit the sample means better in some cases. This is an interesting observation. Especially, since absence of arbitrage is accepted for a fewer number of factors for the parameterized volatility models.

3.6 Dynamically consistent term structure models

In the previous analysis, we have used conditions in BC to derive cross-restrictions on the loading functions in the yield curve in order to guarantee that a consistent DTSM exists. While the analysis demonstrated that consistency matters empirically, the focus in the estimation step has been on the yield dynamics without imposing a particular yield curve. That is, while the derived restricted factor models have loadings consistent with a desired yield curve, it neither requires nor guarantees that the yield curve generated by the model actually takes on this shape. We address this in this section by considering the approach of BCHR, who provide conditions, in addition to those of the previous analysis, that force the yield curve to adopt a particular consistent yield curve shape during the estimation. This procedure connects arbitrage-free DTSMs more tightly to popular yield curve models and combines desirable properties of both fields. In what follows, we refer to this class of models as the class of dynamically consistent term structure models (DCTSMs). The advantage of this procedure,
from a practical point of view, is if one trust in a particular yield curve, this approach allows to impose this curve during the estimation. This approach is related to the class of arbitrage-free NS models by Christensen, Diebold and Rudebusch (2011), who show how to tie the NS yield curve to an arbitrage-free DTSM.

We will consider one- and three-factor DCTSMs. They will be based on the HW dynamics (3.2.6) and the stochastic level, slope, and curvature (SLSC) dynamics introduced by BCHR. Along the lines of the analysis of the previous sections, we want to examine the relation between the yield curve implemented in the initial fitting step and its consequences for empirical analysis.

Seeking for a dynamically consistent one-factor model, it may seen inadequate to base the approach on the two-factor HW dynamics (3.2.6). Yet, BCHR show that for an initial yield curve \( y(z_{t_0}, \tau) \) and yields that follow the HW dynamics, subsequent yield curves at \( t > t_0 \) can be formulated in a consistent restricted one-factor model\(^1\) for the slope-adjusted yield changes (3.3.3) given by

\[
\tilde{y}_i = \mu + C z_i + \varepsilon_i, \quad i = 1, \ldots, n, \tag{3.6.1}
\]

where \( z_i \), as before, is a \( m \times 1 \) dimensional normal increments. We provide expressions for the \( M \times 1 \) dimensional mean vector \( \mu \) and \( M \times m \) dimensional loading vector \( C \) in Appendix A.1.6. We though want to emphasize that both \( \mu \) and \( C \) are spanned by the ANS yield curve. As before, we assume that the error terms \( \varepsilon_i \) are uncorrelated with \( z_i \) and have a diagonal variance matrix \( \Psi \). Thus, we can estimate the model by maximum likelihood factor estimation. We refer to this model as the dynamically consistent HW (DCHW) model.

BCHR suggest two initial conditions \( y(z_{t_0}, \tau) \) in the DCHW model. One can either use an unrestricted initial yield curve \( y(t_0, \tau) \) or a yield curve that takes the ANS yield curve shape at time \( t_0 \). The former has the potential advantage that an arbitrary current yield curve at time \( t_0 \) can be chosen as initial yield curve in order to incorporate all the market information available at the moment. In this case, the slope-adjustment in \( \tilde{y}_i \) depends on the previous yield curve and is unrestricted. In the latter case, BCHR show that the slope-adjustment in \( \tilde{y}_i \), too, adopts the ANS form, in which case the underlying yield curve in 3.6.1 takes on the ANS yield curve. This offers a consistency result similar to that in the DTSMs in Section 3.2 and enables us to investigate the role of consistency in the model proposed here; by comparing the empirical implications of the choice of curve fitting procedures implemented, i.e., a consistent or inconsistent yield curve, to build the yield panel data for inference.

For a three-factor model, we consider the SLSC model, which involves three stochastic

\(^1\)The second factor becomes deterministic but is important to achieve consistency.
3.7 Test for no-arbitrage in DCTSMs

In this section, we present a test for no-arbitrage within the DCTSMs. We recall for this that the factor dynamics in the DCTSMs rely on the no-arbitrage HJM drift restriction (3.2.2). Thus, we can test for no-arbitrage by testing the restricted mean $\mu_r$ within the DCHW and DCSLSC model, respectively, against an unrestricted alternative specified for testing purposes $\mu_u = \alpha$, where $\alpha = (\alpha_1, \ldots, \alpha_M)'$ is the sample mean.

The null-hypothesis $H_0 : \mu_r = \mu_u$ can be tested by a standard LR test. The test statistic is approximately $\chi^2$-distributed, and the difference in degrees of freedom between the restricted and unrestricted model is nine in the DCHW model and seven in the DCSLSC model, corresponding to $M = 10$ parameters saved in $\mu_r$ by imposing the no-arbitrage restriction, whereas the one-and three-dimensional market price of risk parameter $\lambda$ is a new parameter introduced in both models.

$\chi^2$-distributed, and the difference in degrees of freedom between the restricted and unrestricted model is nine in the DCHW model and seven in the DCSLSC model, corresponding to $M = 10$ parameters saved in $\mu_r$ by imposing the no-arbitrage restriction, whereas the one-and three-dimensional market price of risk parameter $\lambda$ is a new parameter introduced in both models.

17 The SLSC dynamics coincide with the EHJM3 model for $a_2 = a_4$ and $a_3 = a_3 a_2$. 
3.8 Empirical application of DCTSMs

We fit the DCHW and DCSLSC model to the four panel data sets of slope-adjusted yield changes. Along the lines of the empirical application of the DTSMs, we examine the effect of modifying the yield parameterization on the volatility structure and on the test for no-arbitrage. We consider the same panel data sets of slope-adjusted yield changes as described in Section 3.5, though we replace the REP3 panel data with a panel data set of slope-adjusted yield changes obtained by fitting the SLSC yield curve to bond prices. We discuss the similarity between the REP3 and SLSC yield curve in Appendix A.1.2 and consider it as unnecessary to include both sets in the analysis. The summary statistics for raw SLSC yields and slope-adjusted yield changes appear in Table 1-2, and the results of PFA to the reduced correlation matrix are reported in Table 3.

3.8.1 Maximum likelihood estimation of DCTSMs

The estimated restricted and unrestricted DCHW models (3.6.1) appear in Panel A of Table 10. Overall, the estimated values are homogeneous across the restricted and unrestricted models as well as across the four data sets. The estimated market prices of risks are negative in all cases and larger in magnitude than in the DTSMs. As it can been seen from Figure 11, which shows the differences in the volatility structure of the GF1 and DCHW models, the estimated low values for $a_2$ flatten the loadings of the stochastic slope factor in maturity dimension and cause them to act like a level factor, reflecting the previous observation that the level factor explains most of the variation in the data. Hence, we can again think of it as an overall average of the hump shaped volatility loading of the GF1 model, but it is slightly more upward sloping than the loadings of the parameterized one-factor DTSMs.

The estimated DCSLSC models are reported in Panel A of Table 11. As for the DCHW model, the estimated values are of similar magnitude across models and data sets. The curvature factor has the largest volatility given by the estimated value for $\tilde{a}_3$ being approximately twice as large as the estimated volatility parameters $a_0$ and $a_1$ for the level and slope factor, respectively. Next, not all market price of risks are significant. In fact, the market price of curvature risk $\lambda_3$ is insignificant in all cases, and the market price of slope risk $\lambda_2$ is only borderline significant. The market price of level risk $\lambda_1$ is largest in all cases. Lastly, the estimated values for $a_2$ are larger than in the DCHW models. Hence, as it can be seen from Figure 12, which illustrates the volatility structures of the DCSLSC models, the level-slope-curvature structure is clearly present for all data sets. Yet, whereas the volatility structure in the parameterized three-factor DTSMs between the different data sets were noticeably different, the differences are considerably smaller here. That is, the enhanced structure by imposing a yield curve shape in the model estimation step decreases the variability in the volatility structure.
Figure 13 shows the estimated unique variances in $\Psi$ for the restricted DCHW and DCSLSC models, respectively. The conclusions drawn from the parameterized DTSMs also apply here: adding more factors decreases the unique variances, they are humped shaped with upward sloping tails, and they are of the same size as the corresponding general factor models.

### 3.8.2 No-arbitrage drift restriction within DCTSMs

Here, we test for absence of arbitrage within the DCTSMs. Even though the estimated values for $a_1$ and $a_2$ are of similar magnitude in the restricted and unrestricted DCHW models for all data sets, it follows clearly from Figure 14, which illustrates the differences between the restricted $\mu_r$ and unrestricted means $\mu_u$, i.e., the sample means, that $\mu_r$ differs significantly from $\mu_u$ in all cases. The restricted means are flat in maturity dimension and cannot capture the upward sloping behavior of the unrestricted means. For each data set the LR test statistic for no-arbitrage and their test probabilities are given in Panel B of Table 10. In agreement with the visual inspection of the means as well as the results of the tests for no-arbitrage in the one-factor models in Section 3.5.4, we reject the no-arbitrage drift restriction at a 5% significance level in all cases for the DCHW models. This suggests that more factors are needed. Indeed, the additional flexibility of the DCSLSC model, where $\mu_r$ is given in terms of seven parameters, apparently enables the model to generate a better fit to $\mu_u$. Consequently, from Panel B of Table 11, showing the results from the LR test in the DCSLSC models, follows that the test statistic fails to reject the test only for NS and ANS data but not for SLSC and NSS data. The overall closer fit to the mean compared to the one-factor model as well as the superior fit for SLSC and NSS data can also be seen from Figure 14.

The analysis in this section showed that even though we imposed consistency directly on the factor models in the DCTSMs, thereby imposing more structure on the estimation procedure compared to Section 3.5, the conclusion drawn from the four data sets again differed significantly with respect to the outcome of the test for no-arbitrage. Moreover, since we accepted absence of arbitrage for the DCSLSC model based on slope-adjusted yield changes from the consistent SLSC yield curve, it also provided further evidence for the importance of consistency. However, it appears that the volatility structure in the DCTSMs is less sensitive to different initial choices of the yield curve parameterization in the fitting step compared to the analysis in Section 3.5.

### 3.9 Conclusion

We introduced a statistical test for the consistency between the shape of the yield curve and the stochastic process driving interest rates through time. We did this by
testing the no-arbitrage drift condition within the framework developed by Heath et al. (1992) using a standard LR test and a computationally simpler, though accurate, Wald test. We analyzed the affine DTSMs by Ho and Lee (1986), Hull and White (1993) and Heath et al. (1992) empirically and extended the two-factor model by Heath et al. (1992) to a three- and four-factor model by adding hump shaped factors. The empirical analysis was based on five different data sets all fitted to the same US Treasury bond data but derived by different yield curve fitting methods.

The model assumptions of the affine DTSMs considered place restrictions on the shape of the volatility factors. Application to empirical data showed that the estimated parameterized volatility structure of these models is significantly different from those of the general factor models for all data sets considered. The general factor models are much more flexible and give a better fit to the correlation structure than the parameterized volatility models. However, when taking the fit to the sample mean into account the parameterized volatility models proved to be empirically arbitrage-free in many cases and more stable. These properties outweigh the flexibility of the general factor models, and hence the parameterized models are to be preferred. The evidence favored at least two factors to exclude arbitrage-opportunities and three factors to describe the volatility structure but the exact number of factors needed as well as the shape of the volatility structure varied between data sets, i.e., curve fitting procedures.

Overall, it confirmed our initial argumentation that the yield curve matters for inference. We also addressed the concept of consistency in the sense of Björk and Christensen (1999) between a yield curve and a DTSM and demonstrated that consistency matters not only theoretically, but matters empirically. This finding guided us to the approach of BCHR which allowed us in a second empirical analysis to impose a consistent yield curve directly on the factor model. Its results confirmed our prior findings on the number of factors required to fit to the correlation structure and (no-arbitrage) mean. Moreover, it provided further evidence for the importance of the choice of yield curve family for the conclusion drawn from the analysis.

**Acknowledgements**

We are grateful to Svend Jakobsen for useful comments and suggestions. We also thank participants at the joint Econometrics-Finance seminar at Aarhus University (2017), and the 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 for research support. Some of this research was conducted while Hansen was visiting the Department of Economics, University of Pennsylvania, and the generosity and hospitality of the department is gratefully acknowledged.
3.10 References


A.1 Appendix

A.1.1 Bond market terminology

Here, we introduce some basic bond market terminology. Especially, we derive how the term structure of interest rates can be represented equivalently in terms of forwards and yields.

The relationship between the yield curve and forward curve is

\[ y(z, x) = \frac{1}{x} \int_0^x G(z, u) du. \] (A.1.1)

If now forward rates follow a SDE given by

\[ dG(t, x) = \alpha_f(t, x) dt + \sigma_f(t, x) dW_t, \] (A.1.2)

where \( \alpha_f \) and \( \sigma_f \) are the forward rate drift and volatility functions, \( t \) is calendar time, and \( x \) the fixed maturity date, we can derive yield rate dynamics from (A.1.2) by applying Leibnitz’ rule for stochastic integrals to (A.1.1), which gives

\[
\frac{dy(t, x)}{dt} = \frac{1}{x} \left[ \left( \int_t^x \alpha_f(t, s) ds \right) - G(t, t) \right] dt + \frac{1}{x} \left( \int_t^x \sigma_f(t, s) ds \right) dW_t
\]

\[
= \alpha(t, x) dt + \sigma(t, x) dW_t,
\]

where \( \alpha \) and \( \sigma \) are the yield rate drift and volatility functions. Hence, the yield rate volatility function in terms of the forward rate volatility function is

\[ \sigma(t, x) = \frac{1}{x} \int_t^x \sigma_f(t, s) ds. \] (A.1.3)

A.1.2 Consistent forward curve families for the EHJM3 and EHJM4 models

In the following we derive consistent forward curve families for the three- and four-factor models presented in Section 3.2.3. As in Björk and Christensen (1999) we consider the restricted exponential polynomial (REP) forward curve family which we define below. For further details on the properties of this particular forward curve family we refer to Björk and Christensen (1999).

**Definition 1.** Let \( K \in \mathbb{N}_0^K \) be given and consider a parameter vector \( z = (z_1, \ldots, z_K) \in \mathbb{R}_+^K \). The restricted forward curve manifold \( \text{REP}(K, n, z) \) is defined as the set of all curves of the form

\[ G(x) = \sum_{i=1}^K p_i(x) e^{-z_i x}, \]

where \( p_i \) is a polynomial of degree less than \( n_i \), \( \deg(p_i) \leq n_i \), for all \( i \).
Björk and Christensen (1999) provide necessary and sufficient conditions for mutual consistency of a given parameterized family of forward rate curves and the dynamics of a given interest rate model, which in the present setup with a model given by \((\sigma, \lambda)\) and a class of curves given by \(G\) are

\[
\frac{\partial}{\partial x} G(z, x) + \sigma(x) \left( \lambda + \int_0^x \sigma(s) \, ds \right) \in \text{Im} \left[ \frac{\partial}{\partial z} G(z, x) \right], \quad (A.1.4)
\]

\[
\sigma(x) \in \text{Im} \left[ \frac{\partial}{\partial z} G(z, x) \right], \quad (A.1.5)
\]

for all \(x\) and \(z\), where \(\partial G/\partial z\) denotes the Frechet derivative (Jacobian) of \(G\) at \(z\), and \(\text{Im}[\partial G/\partial z]\) denotes the image of the linear map \(\partial G/\partial z\). That is, a variable \(y\) belongs to this set if it can be written as a linear combination of the derivatives of \(G\), i.e., if \(y = \sum_i \eta_i \partial G/\partial z_i\), where \(z = (z_1, \ldots, z_d)\) and \(\eta_i\) are coefficients that may depend on the parameter point \(z\), but not on \((t, x)\). Condition (A.1.4) is called the consistent drift condition and (A.1.5) the consistent volatility condition (the latter is understood to apply to each of the \(m\) coordinates of the volatility function \(\sigma\)). These conditions were derived for the case \(\lambda = 0\) (that is, under the risk neutral or martingale measure \(Q\)) by Björk and Christensen (1999). For our empirical purposes, we need the relevant conditions under the objective or physical measure \(P\), and it follows that if condition (A.1.5) is met, then

\[
\sigma(x) \in \text{Im} \left[ \frac{\partial}{\partial z} G(z, x) \right],
\]

and condition (A.1.4) reduces to

\[
\frac{\partial}{\partial x} G(z, x) + \sigma(x) \int_0^x \sigma(s) \, ds \in \text{Im} \left[ \frac{\partial}{\partial z} G(z, x) \right]. \quad (A.1.6)
\]

Thus, the conditions under \(P\) and \(Q\) are equivalent.

When working with the REP forward curve family, the consistent drift condition (A.1.6) simplifies to

\[
\sigma(x) \int_0^x \sigma(s) \, ds \in \text{Im} \left[ \frac{\partial}{\partial z} G(z, x) \right], \quad (A.1.7)
\]

as \(\partial G(z, x)/\partial z \in \text{Im}[\partial G(z, x)/\partial z]\) when \(G = \text{REP}(K, n, \beta)\) as shown by Björk and Christensen (1999).

We now turn our attention to the EHJM3 model volatility structure given by

\[
\sigma = \left( a_0, a_1 e^{-a_2 x}, a_3 x e^{-a_4 x} \right).
\]
Using (A.1.7), the consistent drift condition is given by
\[
\sigma(x) \int_0^x \sigma(s) \, ds = a_0^2 x + \frac{a_1^2}{a_2} e^{-a_2 x} \left( 1 - e^{-a_2 x} \right) + \frac{a_3^2}{a_4} x e^{-a_4 x} \left( 1 - a_4 x e^{-a_4 x} - e^{-a_4 x} \right) = a_0^2 x + \frac{a_1^2}{a_2} e^{-a_2 x} - \frac{a_1^2}{a_2} e^{-2a_2 x} + \frac{a_3^2}{a_4} x e^{-a_4 x} - \left( \frac{a_3^2}{a_4} x^2 + \frac{a_3^2}{a_4} x \right) e^{-2a_4 x}
\]
\[\in \text{Im} \left[ \frac{\partial}{\partial z} G(z, x) \right], \quad (A.1.8)\]
and the consistency volatility condition
\[\left[ a_0, a_1 e^{-a_2 x}, a_3 x e^{-a_4 x} \right] \in \text{Im} \left[ \frac{\partial}{\partial z} G(z, x) \right]. \quad (A.1.9)\]
This leads us to consider the \( \text{REP}(5, (1, 0, 0, 1, 2), (0, a_2, 2a_2, a_4, 2a_4)) \) family given by
\[G_3(z, x) = z_1 + z_2 x + z_3 e^{-a_2 x} + z_4 e^{-2a_2 x} + \left( z_5 + z_6 x \right) e^{-a_4 x} + \left( z_7 + z_8 x + z_9 x^2 \right) e^{-2a_4 x}. \quad (A.1.10)\]
Taking derivative, it follows that
\[\frac{\partial}{\partial z} G_3(z, x) = \left[ 1, x, e^{-a_2 x}, e^{-2a_2 x}, e^{-a_4 x}, x e^{-a_4 x}, e^{-2a_4 x}, x^2 e^{-2a_4 x} \right].\]
As conditions (A.1.8) and (A.1.9) are fulfilled, we conclude that the EHJM3 model is consistent with the \( \text{REP}(5, (1, 0, 0, 1, 2), (0, a_2, 2a_2, a_4, 2a_4)) \) forward curve family. However, this family is not the minimal family consistent with the EHJM3 model. In fact, it is easy to see that we can set \( z_7 = 0 \) in (A.1.10) without losing consistency with the EHJM3 model. Thus, the minimal consistent family is given by
\[G_3^m(x, z) = z_1 + z_2 x + z_3 e^{-a_2 x} + z_4 e^{-2a_2 x} + (z_5 + z_6 x) e^{-a_4 x} + (z_8 x + z_9 x^2) e^{-2a_4 x}. \quad (A.1.11)\]
We have noted in Section 3.6 that the EHJM3 and SLSC dynamics coincide for \( a_2 = a_4 \) and \( a_3 = a_3 a_2 \). In addition, Borup et al. (2020) show that the SLSC forward curve family is the minimal family consistent with the SLSC dynamics. In order for this to be true, it must be the case that the forward curve family defined by (A.1.11) with the restriction \( a_2 = a_4 \) and the SLSC forward curve family differ only up to a linear transformation. Applying (A.1.1) on the SLSC yield curve spanned by the loading functions \( B(\tau) = [B_{1:3}(\tau), B_{4:7}(\tau)] \) with \( B_{1:3}(\tau) \) being the NS loading functions defined in (3.2.4) and \( B_{1:4}(\tau) \) defined in (1.32) the SLSC forward curve is
\[G_{\text{SLSC}}(z, x) = z_1 + z_2 e^{-a_2 x} + z_3 a_2 x e^{-a_2 x} + z_4 e^{-2a_2 x} + z_5 2x + z_6 2a_2 x e^{-2a_2 x} + z_7 (2a_2 x e^{-2a_2 x} - 2a_2^2 x^2 e^{-2a_2 x}) = z_1 + z_2 e^{-a_2 x} + z_3 a_2 x e^{-a_2 x} + z_4 e^{-2a_2 x} + z_5 2x + (z_6 + z_7) 2a_2 x e^{-2a_2 x} - z_7 2a_2^2 x^2 e^{-2a_2 x},\]
where the claim follows now immediately from this expression. Hence, the $G_{SLSC}$ forward curve and the restricted $G_3^m$ forward curve span the same set of term structure models.

We now take a look at the EHJM4 model volatility structure, which is given by

$$\sigma = (a_0, a_1 e^{-a_2 x}, a_3 x e^{-a_4 x}, a_5 x e^{-a_6 x}).$$

As before, we consider the consistent drift condition given by

$$\sigma(x) \int_0^x \sigma(s) ds = a_0^2 x + \frac{a_1^2}{a_2} e^{-a_2 x} (1 - e^{-a_2 x}) + \frac{a_3^2}{a_4} x e^{-a_4 x} (1 - a_4 x e^{-a_4 x} - e^{-a_4 x})$$

$$+ \frac{a_5^2}{a_6} x e^{-a_6 x} (1 - a_6 x e^{-a_6 x} - e^{-a_6 x})$$

$$= a_0^2 x + \frac{a_1^2}{a_2} e^{-a_2 x} - \frac{a_1^2}{a_2} e^{-2a_2 x} + \frac{a_3^2}{a_4} x e^{-a_4 x} - \left( \frac{a_3^2}{a_4} x + \frac{a_3^2}{a_4} x^2 \right) e^{-2a_4 x}$$

$$\in \text{Im} \left[ \frac{\partial}{\partial z} G(z, x) \right],$$

(A.1.12)

and the consistent volatility condition is

$$\left[ a_0, a_1 e^{-a_2 x}, a_3 x e^{-a_4 x} \right] \in \text{Im} \left[ \frac{\partial}{\partial z} G(z, x) \right].$$

(A.1.13)

This leads us to consider the $\text{REP}(7, (1, 0, 0, 1, 2, 1, 2), (0, a_2, 2a_2, a_4, 2a_4, a_6, 2a_6))$ family given by

$$G_4(z, x) = z_1 + z_2 x + z_3 e^{-a_2 x} + z_4 e^{-2a_2 x} + (z_5 + z_6 x) e^{-a_4 x} + (z_7 + z_8 x + z_9 x^2) e^{-2a_4 x}$$

$$+ (z_{10} + z_{11} x) e^{-a_6 x} + (z_{12} + z_{13} x + z_{14} x^2) e^{-2a_6 x}.$$

(A.1.14)

Taking derivatives, we have that

$$\frac{\partial}{\partial z} G_4(z, x) = \left[ 1, x, e^{-a_2 x}, e^{-2a_2 x}, e^{-a_4 x}, x e^{-a_4 x}, e^{-2a_4 x}, x e^{-2a_4 x}, x^2 e^{-2a_6 x}, e^{-a_6 x}, x e^{-a_6 x}, e^{-2a_6 x}, x e^{-2a_6 x}, x^2 e^{-2a_6 x} \right].$$

As conditions (A.1.12) and (A.1.13) are fulfilled, we conclude that the EHJM4 model is consistent with the $\text{REP}(7, (1, 0, 0, 1, 2, 1, 2), (0, a_2, 2a_2, a_4, 2a_4, a_6, 2a_6))$ forward curve family.

(A.1.15)

**A.1.3 Consistent yield curve families for the EHJM3 and EHJM4 models**

In order to derive the corresponding consistent yield curve families for the EHJM3 and EHJM4 models, we apply (A.1.1) on (A.1.10) and (A.1.14), respectively. This implies
we can write the third term in (A.1.15)

\[ y(z, x) = \frac{1}{x} \int_0^x G_3(z, u) du \]

\[ = z_1 + z_2 \frac{1}{2} x + z_3 \frac{1 - e^{-a_2 x}}{a_2 x} + z_4 \frac{1 - e^{-2a_2 x}}{2a_2 x} + (a_4 z_5 + z_6) \left( \frac{1 - e^{-a_4 x}}{a_4^2 x} - z_6 \frac{e^{-a_4 x}}{a_4} \right) + (2a_4^2 z_7 + a_4 z_8 + z_9) \frac{1 - e^{-2a_4 x}}{4a_4^3 x} - (a_4 (z_8 + z_9 x) + z_9) \frac{e^{-2a_4 x}}{2a_4^2}, \]

and similarly that the consistent EHJM3 model yield curve family is given by

\[ y(z, x) = \frac{1}{x} \int_0^x G_4(z, u) du \]

\[ = z_1 + z_2 \frac{1}{2} x + z_3 \frac{1 - e^{-a_2 x}}{a_2 x} + z_4 \frac{1 - e^{-2a_2 x}}{2a_2 x} + (a_4 z_5 + z_6) \left( \frac{1 - e^{-a_4 x}}{a_4^2 x} - z_6 \frac{e^{-a_4 x}}{a_4} \right) + (2a_4^2 z_7 + a_4 z_8 + z_9) \frac{1 - e^{-2a_4 x}}{4a_4^3 x} - (a_4 (z_8 + z_9 x) + z_9) \frac{e^{-2a_4 x}}{2a_4^2} + (a_6 z_{10} + z_{11}) \frac{1 - e^{-a_6 x}}{a_6^2 x} - z_{11} \frac{e^{-a_6 x}}{a_6} + (2a_6^2 z_{12} + a_6 z_{13} + z_{14}) \frac{1 - e^{-2a_6 x}}{4a_6^3 x} - (a_6 (z_{13} + z_{14} x) + z_{14}) \frac{e^{-2a_6 x}}{2a_6^2}. \]

### A.1.4 Maximum likelihood estimate for the market price of risk in restricted parameterized models

The log-likelihood function in the restricted parameterized models is given by

\[ \ell_R(\tilde{y}_1, \ldots, \tilde{y}_n; \theta) = -\frac{n}{2} \log |2\pi \Sigma| - \frac{n}{2} \text{tr} \Sigma^{-1} S - \frac{n}{2} (\tilde{y} - \gamma \Delta)' \Sigma^{-1} (\tilde{y} - \gamma \Delta). \quad (A.1.15) \]

As the first two terms do not depend on \( \lambda \), maximizing (A.1.15) with respect to \( \lambda \) corresponds to minimizing the third term only. Defining the \( M \)-vector \( s \) as

\[ s \equiv \left[ \frac{X_1}{2} (\delta_{11}^2 + \cdots + \delta_{1m}^2), \ldots, \frac{X_M}{2} (\delta_{M1}^2 + \cdots + \delta_{Mm}^2) \right]^t, \]

we can write the third term in (A.1.15), ignoring the factor \( n/2 \) in the minimization, as

\[ (\tilde{y} - \gamma \Delta)' \Sigma^{-1} (\tilde{y} - \gamma \Delta) = (\tilde{y} - (\sigma \lambda + s) \Delta)' \Sigma^{-1} (\tilde{y} - (\sigma \lambda + s) \Delta) \]

\[ = \tilde{y}' \Sigma^{-1} \tilde{y} - \tilde{y}' \Sigma^{-1} \sigma \lambda \Delta - \tilde{y}' \Sigma^{-1} s \Delta - \lambda' \sigma' \Sigma^{-1} \tilde{y} \Delta + \lambda' \sigma' \Sigma^{-1} \sigma \lambda \Delta^2 + \lambda' \sigma' \Sigma^{-1} s \Delta + \lambda' \sigma' \Sigma^{-1} \sigma \Delta^2 + \lambda' \sigma' \Sigma^{-1} s \Delta^2. \]

Differentiating with respect to \( \lambda \) we obtain the first order condition

\[ -\tilde{y}' \Sigma^{-1} \sigma \Delta - \tilde{y}' \Sigma^{-1} \sigma \Delta + 2 \lambda' \sigma' \Sigma^{-1} \sigma \Delta^2 + \lambda' \sigma' \Sigma^{-1} \sigma \Delta^2 + \lambda' \sigma' \Sigma^{-1} \sigma \Delta^2 = 0, \]
where we have used that \(\Sigma\) is a symmetric matrix. Solving for \(\lambda\) gives the maximum likelihood estimator for \(\lambda\) as
\[
\lambda = ((\hat{y}'\Sigma^{-1}\sigma\Delta - s\Sigma^{-1}\sigma\Delta^\top)(\sigma'\Sigma^{-1}\sigma)^{-1})\Delta^{-2} = (\sigma'\Sigma^{-1}\sigma)^{-1}(\Delta^{-1}\sigma'\Sigma^{-1}\hat{y} - \sigma'\Sigma^{-1}s).
\]
If \(M = m\) such that \(\sigma\) is a quadratic \(M \times M\)-matrix and if, in addition, \(\sigma\) invertible this expression can be reduced to
\[
\lambda = \sigma^{-1}(\Delta^{-1}\hat{y} - s).
\]

A.1.5 Yield curve derivative

The analytical derivative of the \(\text{RE}(5, (1, 0, 1, 2), (0, a_2, 2a_2, a_4, 2a_4))\) yield curve family given by (3.2.7) with respect to time to maturity \(s\) is
\[
\frac{\partial y}{\partial x}(z, x) = \frac{1}{2} z_2 + z_3 \frac{a_2 x - e^{a_2 x} + 1}{a_2 x^2} e^{-a_2 x} + z_4 \frac{2a_2 x - e^{2a_2 x} + 1}{2a_2 x^2} e^{-2a_2 x} - (a_4 z_5 + z_6) \frac{1 - e^{-a_4 x}}{a_4^2 x^2} + z_6 e^{-a_4 x} + \frac{a_4 z_5 + z_6}{a_4 x} e^{-a_4 x} - (2a_4^2 z_7 + a_4 z_8 + z_9) \frac{1 - e^{2a_4 x}}{4a_4^3 x^2} + (2a_4^2(x^2 z_9 + x z_8) + 2a_4 x z_9) \frac{e^{-2a_4 x}}{2a_4^2 x} + (2a_4^2 x^2 z_9 + x z_8) + 2a_4 x z_9) \frac{e^{-2a_4 x}}{4a_4^3 x^2} + (2a_4^2 z_7 + a_4 z_8 + z_9) \frac{e^{-2a_4 x}}{2a_4^2 x} - (2a_4^2(2x z_9 + z_8) + 2a_4 z_9) \frac{e^{-2a_4 x}}{4a_4^3 x}.
\]

A.1.6 Details on restricted factor model in DCHW model

The \(M \times 1\) dimensional mean vector \(\mu\) and \(M \times 1\) dimensional loading vector \(C\) in the restricted factor model (1.24) implied by the DCHW model have \(j\)th element given by
\[
\mu_j = \frac{a_j^2}{a_2} + a_1 \lambda_1 \Delta_i B_2(\Delta_i) B_2(\tau_j) - \frac{a_j^2}{a_2} \Delta_i B_4(\Delta_i) B_4(\tau_j), \quad (A.1.16)
\]
\[
C_j = a_1 B_2(\tau_j) \sqrt{\Delta_i B_4(\Delta_i)}, \quad (A.1.17)
\]
where \(B_2(\tau_j)\) and \(B_4(\tau_j)\) are the second and fourth loading function of the ANS yield curve shape and \(\Delta_i = t_{i+1} - t_i\).

A.1.7 Details on restricted factor model in DCSLSC model

The \(M \times 1\) dimensional mean \(\mu\) and \(M \times m\) dimensional loading \(C\) on the \(3 \times 1\) dimensional common covariance generating factor \(z_i\) in the DCSLSC model in (3.6.1) are given by
\[
\mu = B(\tau)(I - H(\Delta_i)) \bar{z},
\]
\[
C = B_{13}(\tau) K_{13}(\Delta_i),
\]
for $\Delta_i = t_{i+1} - t_i, K_{1:3}(\Delta_i)K_{1:3}(\Delta_i)' = \Omega_{1:3}(\Delta_i)$, where $B$ is $M \times 7$ with typical row given by the SLSC loading functions $B_{1:7}(\tau_j)$ and the long-run factor level $\tilde{z}$ is given by

$$\tilde{z} = \left[ \frac{a_0}{b} \left( \frac{a_0}{\tau} + \lambda_1 \right), \quad \tilde{z}_3 + \frac{a_1}{a_2} \left( \frac{a_1}{\tau} + \lambda_2 \right), \quad \frac{\tilde{a}_3}{\tilde{a}_2} \left( \frac{\tilde{a}_3}{\tau} + \lambda_3 \right), \quad -\frac{a_1^2}{2a_2^2} - \frac{\tilde{a}_2^2}{2a_2^2}, \quad -\frac{\tilde{a}_3^2}{2a_2^2}, \quad -\frac{3\tilde{a}_3^2}{4a_2^2}, \quad \frac{\tilde{a}_3^2}{4a_2^2} \right]' .$$

Furthermore, the transition matrix $H$ is given by

$$H(u) = \begin{pmatrix}
    e^{-bu} & 0 & 0 & 0 & 0 & 0 & 0 \\
    0 & e^{-a_2u} & a_2ue^{-a_2u} & 0 & 0 & 0 & 0 \\
    0 & 0 & e^{-a_2u} & 0 & 0 & 0 & 0 \\
    0 & 0 & 0 & e^{-2a_2u} & 0 & 0 & 0 \\
    0 & 0 & 0 & 0 & e^{-2bu} & 0 & 0 \\
    0 & 0 & 0 & 0 & 0 & e^{-2a_2u} & 0 \\
    0 & 0 & 0 & 0 & 0 & 0 & e^{-2a_2u}
\end{pmatrix} ,$$

and the upper left $3 \times 3$ submatrix of $\Omega$ is

$$\Omega_{1:3}(u) = \begin{pmatrix}
    a_1^2 uB_5(u) & 0 & 0 \\
    0 & a_3^2 uB_4(u) + a_2^2 u \left( B_6(u) - B_7(u) \right) / 2 & 0 \\
    0 & a_3^2 uB_6(u) / 2 & a_3^2 uB_4(u) / 2
\end{pmatrix} ,$$

while the remaining entries are zero.
A.2 Tables

Table 1: The table reports summary statistics of yields for selected yield curve shapes with weekly observations. Maturities are in years and yields are in percent.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>NS</th>
<th>ANS</th>
<th>REP3</th>
<th>NSS</th>
<th>SLSC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. dev.</td>
<td>Mean</td>
<td>Std. dev.</td>
<td>Mean</td>
</tr>
<tr>
<td>0.25</td>
<td>2.038</td>
<td>2.045</td>
<td>2.019</td>
<td>2.036</td>
<td>2.016</td>
</tr>
<tr>
<td>0.5</td>
<td>2.087</td>
<td>2.059</td>
<td>2.124</td>
<td>2.060</td>
<td>2.119</td>
</tr>
<tr>
<td>1</td>
<td>2.202</td>
<td>2.050</td>
<td>2.271</td>
<td>2.060</td>
<td>2.262</td>
</tr>
<tr>
<td>2</td>
<td>2.479</td>
<td>1.960</td>
<td>2.503</td>
<td>1.966</td>
<td>2.511</td>
</tr>
<tr>
<td>3</td>
<td>2.772</td>
<td>1.838</td>
<td>2.754</td>
<td>1.838</td>
<td>2.775</td>
</tr>
<tr>
<td>10</td>
<td>4.131</td>
<td>1.253</td>
<td>4.139</td>
<td>1.244</td>
<td>4.146</td>
</tr>
<tr>
<td>20</td>
<td>4.756</td>
<td>1.017</td>
<td>4.759</td>
<td>1.026</td>
<td>4.756</td>
</tr>
<tr>
<td>25</td>
<td>4.891</td>
<td>0.974</td>
<td>4.888</td>
<td>0.989</td>
<td>4.739</td>
</tr>
</tbody>
</table>

Table 2: This table reports sample means and standard deviations of the slope-adjusted yield changes. We refer to equation (3.3.3) for an exact specification. The yields are taken from a data set of yields with weekly observations fitted by selected yield curve shapes. Maturities are in years; means and standard deviation of slope-adjusted yield changes are in percent.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>NS</th>
<th>ANS</th>
<th>REP3</th>
<th>NSS</th>
<th>SLSC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. dev.</td>
<td>Mean</td>
<td>Std. dev.</td>
<td>Mean</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.013</td>
<td>0.089</td>
<td>-0.021</td>
<td>0.085</td>
<td>-0.020</td>
</tr>
<tr>
<td>1</td>
<td>-0.014</td>
<td>0.095</td>
<td>-0.017</td>
<td>0.095</td>
<td>-0.017</td>
</tr>
<tr>
<td>2</td>
<td>-0.015</td>
<td>0.114</td>
<td>-0.015</td>
<td>0.116</td>
<td>-0.015</td>
</tr>
<tr>
<td>3</td>
<td>-0.015</td>
<td>0.126</td>
<td>-0.015</td>
<td>0.127</td>
<td>-0.015</td>
</tr>
<tr>
<td>5</td>
<td>-0.014</td>
<td>0.134</td>
<td>-0.014</td>
<td>0.133</td>
<td>-0.013</td>
</tr>
<tr>
<td>7</td>
<td>-0.012</td>
<td>0.134</td>
<td>-0.012</td>
<td>0.132</td>
<td>-0.012</td>
</tr>
<tr>
<td>10</td>
<td>-0.010</td>
<td>0.130</td>
<td>-0.010</td>
<td>0.129</td>
<td>-0.010</td>
</tr>
<tr>
<td>15</td>
<td>-0.008</td>
<td>0.127</td>
<td>-0.008</td>
<td>0.127</td>
<td>-0.008</td>
</tr>
<tr>
<td>20</td>
<td>-0.007</td>
<td>0.127</td>
<td>-0.007</td>
<td>0.127</td>
<td>-0.007</td>
</tr>
<tr>
<td>25</td>
<td>-0.006</td>
<td>0.129</td>
<td>-0.006</td>
<td>0.129</td>
<td>-0.005</td>
</tr>
</tbody>
</table>
Table 3: This table shows the amount of variance explained by each factor. Factor estimation is done by principal factor analysis on the reduced correlation matrix of NS, ANS, REP3, NSS, and SLSC data, respectively.

<table>
<thead>
<tr>
<th>Factors</th>
<th>NS data</th>
<th>ANS data</th>
<th>REP3 data</th>
<th>NSS data</th>
<th>SLSC data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7526</td>
<td>0.7526</td>
<td>0.7612</td>
<td>0.7612</td>
<td>0.7577</td>
</tr>
<tr>
<td>2</td>
<td>0.1846</td>
<td>0.9372</td>
<td>0.1763</td>
<td>0.9376</td>
<td>0.1693</td>
</tr>
<tr>
<td>3</td>
<td>0.0518</td>
<td>0.9890</td>
<td>0.0457</td>
<td>0.9833</td>
<td>0.0483</td>
</tr>
<tr>
<td>4</td>
<td>0.0104</td>
<td>0.9993</td>
<td>0.0127</td>
<td>0.9960</td>
<td>0.0144</td>
</tr>
<tr>
<td>5</td>
<td>0.0006</td>
<td>0.9999</td>
<td>0.0037</td>
<td>0.9997</td>
<td>0.0060</td>
</tr>
<tr>
<td>6</td>
<td>0.0001</td>
<td>1.0000</td>
<td>0.0003</td>
<td>1.0000</td>
<td>0.0022</td>
</tr>
<tr>
<td>7</td>
<td>0.0000</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>0.0014</td>
</tr>
</tbody>
</table>
Table 4: This table shows the volatility parameter estimates of the unrestricted HL, HW, HJM2, EHJM3, and EHJM4 models with standard errors in parenthesis. The parameter $a_0$ is the constant volatility factor parameter of the HL, HJM2, EHJM3, and EHJM4 models. The parameters $a_1$ and $a_2$ are the parameters of the exponentially decreasing decaying volatility factor of the HW, HJM2, EHJM3, and EHJM4 models. The parameters $a_3$ and $a_4$ are the parameters of the humped volatility factor of the EHJM3 and EHJM4 models. The parameters $a_5$ and $a_6$ are the parameters of the humped volatility factor of the EHJM4 model. Factor estimation is done by maximum likelihood analysis on NS, ANS, REP3, and NSS data.

<table>
<thead>
<tr>
<th>Model</th>
<th>HL</th>
<th>HW</th>
<th>HJM2</th>
<th>EHJM3</th>
<th>EHJM4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: NS data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0013</td>
<td>0.0016</td>
<td>0.0015</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.2e-05)</td>
<td>(4.1e-05)</td>
<td>(3.7e-05)</td>
<td>(1.1e-05)</td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0013</td>
<td>0.0022</td>
<td>0.0020</td>
<td>0.0019</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.2e-05)</td>
<td>(5.4e-05)</td>
<td>(4.8e-05)</td>
<td>(4.6e-05)</td>
<td></td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.0000</td>
<td>0.1763</td>
<td>0.3819</td>
<td>0.8059</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.1e-08)</td>
<td>(3.9e-03)</td>
<td>(8.5e-03)</td>
<td>(1.5e-07)</td>
<td></td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.0017</td>
<td>0.0015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.7e-05)</td>
<td>(4.5e-05)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_4$</td>
<td>0.3812</td>
<td>0.3103</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.5e-03)</td>
<td>(1.6e-03)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_5$</td>
<td>0.0020</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.0e-05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_6$</td>
<td>0.5497</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.6e-04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **Panel B: ANS data** | | | | | |
| $a_0$ | 0.0013 | 0.0015 | 0.0014 | 0.0014 |
| | (3.2e-05) | (3.5e-05) | (3.6e-05) | (3.5e-05) |
| $a_1$ | 0.0013 | 0.0021 | 0.0021 | 0.0019 |
| | (3.1e-05) | (5.5e-05) | (5.2e-05) | (4.7e-05) |
| $a_2$ | 0.0000 | 0.2441 | 0.4489 | 0.2825 |
| | (2.4e-08) | (2.2e-03) | (3.2e-03) | (3.5e-03) |
| $a_3$ | 0.0019 | 0.0021 |
| | (4.8e-05) | (6.6e-05) |
| $a_4$ | 0.4241 | 0.4083 |
| | (2.9e-03) | (2.2e-03) |
| $a_5$ | 0.0046 |
| | (1.1e-04) |
| $a_6$ | 0.7557 |
| | (1.2e-06) |

| **Panel C: REP3 data** | | | | | |
| $a_0$ | 0.0012 | 0.0016 | 0.0015 | 0.0014 |
| | (3.8e-05) | (3.6e-05) | (3.6e-05) | (3.2e-05) |
| $a_1$ | 0.0012 | 0.0020 | 0.0018 | 0.0017 |
| | (3.0e-05) | (5.0e-05) | (4.3e-05) | (4.2e-05) |
| $a_2$ | 0.0000 | 0.2621 | 0.4471 | 0.1294 |
| | (3.1e-06) | (2.1e-03) | (1.2e-02) | (9.1e-03) |
| $a_3$ | 0.0018 | 0.0007 |
| | (5.7e-05) | (2.9e-05) |
| $a_4$ | 0.4440 | 0.1632 |
| | (7.1e-03) | (4.4e-03) |
| $a_5$ | 0.0026 |
| | (6.4e-05) |
| $a_6$ | 0.5931 |
| | (6.2e-05) |

| **Panel D: NSS data** | | | | | |
| $a_0$ | 0.0012 | 0.0017 | 0.0015 | 0.0017 |
| | (3.6e-05) | (4.9e-05) | (3.4e-05) | (2.7e-05) |
| $a_1$ | 0.0012 | 0.0021 | 0.0018 | 0.0024 |
| | (3.0e-05) | (4.9e-05) | (4.3e-05) | (4.3e-05) |
| $a_2$ | 0.0000 | 0.5482 | 0.3622 | 0.1736 |
| | (2.9e-07) | (6.8e-03) | (1.1e-02) | (6.9e-03) |
| $a_3$ | 0.0016 | 0.0015 |
| | (6.0e-05) | (4.5e-05) |
| $a_4$ | 0.3712 | 0.3017 |
| | (6.0e-03) | (4.9e-03) |
| $a_5$ | 0.0048 |
| | (1.1e-04) |
| $a_6$ | 0.7213 |
| | (7.3e-03) |
Table 5: This table shows the volatility parameter estimates of the restricted HL, HW, HJM2, EHJM3, and EHJM4 models with standard errors in parenthesis. The parameter $\theta_0$ is the constant volatility factor parameter of the HL, HJM2, EHJM3, and EHJM4 models. The parameters $\theta_1$ and $\theta_2$ are the parameters of the exponentially decreasing decaying volatility factor of the HW, HJM2, EHJM3, and EHJM4 models. The parameters $\theta_3$ and $\theta_4$ are the parameters of the humped volatility factor of the EHJM3 and EHJM4 models. The parameters $\theta_5$ and $\theta_6$ are the parameters of the humped volatility factor of the EHJM4 model. Factor estimation is done by maximum likelihood analysis on NS, ANS, REP3, and NSS data.

<table>
<thead>
<tr>
<th>Model</th>
<th>HL</th>
<th>HW</th>
<th>HJM2</th>
<th>EHJM3</th>
<th>EHJM4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($3.2 \times 10^{-5}$)</td>
<td>($3.2 \times 10^{-5}$)</td>
<td>($3.2 \times 10^{-5}$)</td>
<td>($3.2 \times 10^{-5}$)</td>
<td>($3.2 \times 10^{-5}$)</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0013</td>
<td>0.0016</td>
<td>0.0015</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0013</td>
<td>0.0022</td>
<td>0.0020</td>
<td>0.0019</td>
<td></td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.0000</td>
<td>0.1764</td>
<td>0.3821</td>
<td>0.8063</td>
<td></td>
</tr>
<tr>
<td>$a_3$</td>
<td></td>
<td>0.0017</td>
<td>0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_4$</td>
<td></td>
<td>0.3814</td>
<td>0.3103</td>
<td>0.0028</td>
<td>0.5496</td>
</tr>
<tr>
<td>$a_5$</td>
<td></td>
<td></td>
<td>0.0040</td>
<td>0.0046</td>
<td>0.7565</td>
</tr>
<tr>
<td>$a_6$</td>
<td></td>
<td></td>
<td></td>
<td>0.0026</td>
<td>0.5933</td>
</tr>
</tbody>
</table>

Panel B: ANS data

<table>
<thead>
<tr>
<th>Model</th>
<th>HL</th>
<th>HW</th>
<th>HJM2</th>
<th>EHJM3</th>
<th>EHJM4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($3.2 \times 10^{-5}$)</td>
<td>($3.2 \times 10^{-5}$)</td>
<td>($3.2 \times 10^{-5}$)</td>
<td>($3.2 \times 10^{-5}$)</td>
<td>($3.2 \times 10^{-5}$)</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0015</td>
<td>0.0014</td>
<td>0.0014</td>
<td>0.0014</td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0013</td>
<td>0.0021</td>
<td>0.0021</td>
<td>0.0019</td>
<td></td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.0000</td>
<td>0.2439</td>
<td>0.4486</td>
<td>2.2849</td>
<td></td>
</tr>
<tr>
<td>$a_3$</td>
<td></td>
<td>0.0019</td>
<td>0.0021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_4$</td>
<td></td>
<td>0.4239</td>
<td>0.4084</td>
<td>0.0040</td>
<td>0.7565</td>
</tr>
<tr>
<td>$a_5$</td>
<td></td>
<td></td>
<td>0.0040</td>
<td>0.0046</td>
<td>0.7565</td>
</tr>
<tr>
<td>$a_6$</td>
<td></td>
<td></td>
<td></td>
<td>0.0026</td>
<td>0.5933</td>
</tr>
</tbody>
</table>

Panel C: REP3 data

<table>
<thead>
<tr>
<th>Model</th>
<th>HL</th>
<th>HW</th>
<th>HJM2</th>
<th>EHJM3</th>
<th>EHJM4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($3.6 \times 10^{-5}$)</td>
<td>($3.6 \times 10^{-5}$)</td>
<td>($3.6 \times 10^{-5}$)</td>
<td>($3.6 \times 10^{-5}$)</td>
<td>($3.6 \times 10^{-5}$)</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0016</td>
<td>0.0015</td>
<td>0.0015</td>
<td>0.0014</td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0012</td>
<td>0.0020</td>
<td>0.0018</td>
<td>0.0017</td>
<td></td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.0000</td>
<td>0.2614</td>
<td>0.4466</td>
<td>0.1295</td>
<td></td>
</tr>
<tr>
<td>$a_3$</td>
<td></td>
<td>0.0018</td>
<td>0.0007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_4$</td>
<td></td>
<td>0.4437</td>
<td>0.1633</td>
<td>0.0026</td>
<td>0.5933</td>
</tr>
<tr>
<td>$a_5$</td>
<td></td>
<td></td>
<td>0.0026</td>
<td>0.0026</td>
<td>0.5933</td>
</tr>
<tr>
<td>$a_6$</td>
<td></td>
<td></td>
<td></td>
<td>0.0026</td>
<td>0.5933</td>
</tr>
</tbody>
</table>

Panel D: NSS data

<table>
<thead>
<tr>
<th>Model</th>
<th>HL</th>
<th>HW</th>
<th>HJM2</th>
<th>EHJM3</th>
<th>EHJM4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($3.6 \times 10^{-5}$)</td>
<td>($3.6 \times 10^{-5}$)</td>
<td>($3.6 \times 10^{-5}$)</td>
<td>($3.6 \times 10^{-5}$)</td>
<td>($3.6 \times 10^{-5}$)</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0015</td>
<td>0.0017</td>
<td>0.0015</td>
<td>0.0017</td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0012</td>
<td>0.0021</td>
<td>0.0018</td>
<td>0.0020</td>
<td></td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.0000</td>
<td>0.5470</td>
<td>0.3618</td>
<td>0.1737</td>
<td></td>
</tr>
<tr>
<td>$a_3$</td>
<td></td>
<td>0.0016</td>
<td>0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_4$</td>
<td></td>
<td>0.3709</td>
<td>0.3018</td>
<td>0.0048</td>
<td>0.7215</td>
</tr>
<tr>
<td>$a_5$</td>
<td></td>
<td></td>
<td>0.0048</td>
<td>0.0048</td>
<td>0.7215</td>
</tr>
<tr>
<td>$a_6$</td>
<td></td>
<td></td>
<td></td>
<td>0.7215</td>
<td>0.7215</td>
</tr>
</tbody>
</table>
Table 6: This table shows the log-likelihood function values of the unrestricted and restricted models, together with the test size and probability of the reduction from the general factor model. In addition, the table shows the Wald and LR test statistics for no-arbitrage within the HL, HW, HJM2, EHJM3, and EHJM4 models, respectively. Factor estimation is done by maximum likelihood analysis on NS, ANS, REP3, and NSS data.

<table>
<thead>
<tr>
<th>Model</th>
<th>HL/HW</th>
<th>HJM2</th>
<th>EHJM3</th>
<th>EHJM4</th>
<th>HL/HW</th>
<th>HJM2</th>
<th>EHJM3</th>
<th>EHJM4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factors</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>$ll_{GF}^U$</td>
<td>49829</td>
<td>55262</td>
<td>59319</td>
<td>62719</td>
<td>50299</td>
<td>55701</td>
<td>58868</td>
<td>62007</td>
</tr>
<tr>
<td>$ll_{Model}^U$</td>
<td>49089</td>
<td>52796</td>
<td>57321</td>
<td>61130</td>
<td>49426</td>
<td>53050</td>
<td>56638</td>
<td>59070</td>
</tr>
<tr>
<td>LR</td>
<td>1480.89</td>
<td>4932.93</td>
<td>3995.32</td>
<td>3177.54</td>
<td>1746.20</td>
<td>5302.10</td>
<td>4460.05</td>
<td>5873.39</td>
</tr>
<tr>
<td>Test Prob.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Wald test</td>
<td>25.41</td>
<td>4.32</td>
<td>5.00</td>
<td>0.25</td>
<td>35.92</td>
<td>2.85</td>
<td>11.13</td>
<td>23.34</td>
</tr>
<tr>
<td>Test Prob.</td>
<td>0.003</td>
<td>0.827</td>
<td>0.661</td>
<td>1.000</td>
<td>0.000</td>
<td>0.943</td>
<td>0.133</td>
<td>0.001</td>
</tr>
<tr>
<td>$ll_{Model}^R$</td>
<td>49076</td>
<td>52794</td>
<td>57319</td>
<td>61130</td>
<td>49409</td>
<td>53049</td>
<td>56633</td>
<td>59060</td>
</tr>
<tr>
<td>LR</td>
<td>25.23</td>
<td>4.18</td>
<td>4.88</td>
<td>0.25</td>
<td>35.30</td>
<td>2.84</td>
<td>8.62</td>
<td>21.66</td>
</tr>
<tr>
<td>Test Prob.</td>
<td>0.003</td>
<td>0.841</td>
<td>0.675</td>
<td>1.000</td>
<td>0.000</td>
<td>0.944</td>
<td>0.281</td>
<td>0.001</td>
</tr>
<tr>
<td>Factors</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>$ll_{GF}^U$</td>
<td>49973</td>
<td>52871</td>
<td>54671</td>
<td>55494</td>
<td>49298</td>
<td>51486</td>
<td>53130</td>
<td>55121</td>
</tr>
<tr>
<td>$ll_{Model}^U$</td>
<td>49011</td>
<td>51471</td>
<td>53529</td>
<td>53794</td>
<td>48414</td>
<td>49625</td>
<td>51816</td>
<td>52590</td>
</tr>
<tr>
<td>LR</td>
<td>1923.15</td>
<td>2798.52</td>
<td>2283.47</td>
<td>3398.23</td>
<td>1767.50</td>
<td>3722.59</td>
<td>2628.34</td>
<td>5061.35</td>
</tr>
<tr>
<td>Test Prob.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Wald test</td>
<td>34.04</td>
<td>3.86</td>
<td>8.37</td>
<td>3.37</td>
<td>25.00</td>
<td>7.64</td>
<td>5.47</td>
<td>5.17</td>
</tr>
<tr>
<td>Test Prob.</td>
<td>0.000</td>
<td>0.869</td>
<td>0.301</td>
<td>0.761</td>
<td>0.003</td>
<td>0.469</td>
<td>0.602</td>
<td>0.523</td>
</tr>
<tr>
<td>$ll_{Model}^R$</td>
<td>48994</td>
<td>51469</td>
<td>53525</td>
<td>53793</td>
<td>48402</td>
<td>49621</td>
<td>51813</td>
<td>52588</td>
</tr>
<tr>
<td>LR</td>
<td>33.68</td>
<td>3.78</td>
<td>7.88</td>
<td>3.23</td>
<td>24.88</td>
<td>6.95</td>
<td>5.06</td>
<td>4.22</td>
</tr>
<tr>
<td>Test Prob.</td>
<td>0.000</td>
<td>0.876</td>
<td>0.343</td>
<td>0.779</td>
<td>0.003</td>
<td>0.542</td>
<td>0.653</td>
<td>0.647</td>
</tr>
</tbody>
</table>
Table 7: This table shows the estimated market prices of risk of the unrestricted and restricted general factor models with standard errors in parenthesis. Factor estimation is done by maximum likelihood analysis on NS, ANS, REP3, and NSS data.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Panel A: NS data</th>
<th>Panel B: ANS data</th>
<th>Panel C: REP3 data</th>
<th>Panel D: NSS data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unrestricted models</td>
<td>Restricted models</td>
<td>Unrestricted models</td>
<td>Restricted models</td>
</tr>
<tr>
<td>λ₁</td>
<td>-0.070 (0.0255)</td>
<td>-0.066 (0.0253)</td>
<td>-0.100 (0.0372)</td>
<td>-0.087 (0.0345)</td>
</tr>
<tr>
<td>λ₂</td>
<td>-0.069 (0.0330)</td>
<td>-0.073 (0.0321)</td>
<td>0.027 (0.0052)</td>
<td>-0.073 (0.0306)</td>
</tr>
<tr>
<td>λ₃</td>
<td>-0.104 (0.0622)</td>
<td>-0.078 (0.0308)</td>
<td>-0.075 (0.0016)</td>
<td>-0.079 (0.0287)</td>
</tr>
<tr>
<td>λ₄</td>
<td>-0.119 (0.0007)</td>
<td>0.00016</td>
<td>-0.125 (0.0319)</td>
<td>0.00016</td>
</tr>
</tbody>
</table>

Panel B: ANS data

| λ₁      | -0.100 (0.0238) | -0.155 (0.0538) | -0.035 (0.0349) | -0.155 (0.0338) |
| λ₂      | -0.055 (0.0282) | 0.336 (0.0613)  | 0.078 (0.0365)  | 0.078 (0.0365)  |
| λ₃      | 0.085 (0.0012)  | -0.076 (0.0196) | 0.078 (0.0365)  | 0.078 (0.0365)  |
| λ₄      | -0.078 (0.0064) |                 |                  |                 |

Panel C: REP3 data

| λ₁      | -0.103 (0.0219) | -0.083 (0.0350) | -0.082 (0.0348) | -0.080 (0.0347) |
| λ₂      | -0.065 (0.0389) | 0.059 (0.0651)  | -0.074 (0.0051) | -0.065 (0.0601) |
| λ₃      | -0.068 (0.0325) | 0.221 (0.0509)  | 0.078 (0.0516)  | 0.205 (0.0441)  |
| λ₄      | 0.052 (0.0437)  |                 | 0.051 (0.0620)  |                 |

Panel D: NSS data

| λ₁      | -0.104 (0.0216) | 0.044 (0.0901)  | -0.076 (0.0706) | -0.082 (0.0445) |
| λ₂      | -0.077 (0.0027) | -0.105 (0.0318) | -0.010 (0.0872) | -0.104 (0.0347) |
| λ₃      | -0.085 (0.0566) | 0.058 (0.0959)  | -0.101 (0.0340) | 0.058 (0.0447)  |
| λ₄      | -0.075 (0.0419) |                 | -0.075 (0.0285) |                 |
Table 8: This table shows the estimated market prices of risk of the unrestricted and restricted parameterized factor models with standard errors in parenthesis. Factor estimation is done by maximum likelihood analysis on NS, ANS, REP3, and NSS data.

<table>
<thead>
<tr>
<th>Model</th>
<th>HL/HW</th>
<th>HJM2</th>
<th>EHJM3</th>
<th>EHJM4</th>
<th>HL/HW</th>
<th>HJM2</th>
<th>EHJM3</th>
<th>EHJM4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: NS data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Unrestricted models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.079</td>
<td>-0.025</td>
<td>-0.018</td>
<td>-0.026</td>
<td>-0.019</td>
<td>-0.018</td>
<td>-0.026</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.163)</td>
<td>(0.0112)</td>
<td>(0.0038)</td>
<td>(0.0060)</td>
<td>(0.0037)</td>
<td>(0.0062)</td>
<td>(0.0037)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.051</td>
<td>-0.050</td>
<td>-0.040</td>
<td>-0.054</td>
<td>-0.056</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0117)</td>
<td>(0.0035)</td>
<td>(0.0006)</td>
<td>(0.00037)</td>
<td>(0.0036)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>-0.050</td>
<td>-0.002</td>
<td>-0.043</td>
<td>-0.019</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0046)</td>
<td>(0.0011)</td>
<td>(0.0033)</td>
<td>(0.0037)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>-0.055</td>
<td>-0.055</td>
<td>-0.056</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0009)</td>
<td>(0.0009)</td>
<td>(0.0026)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Restricted models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.019</td>
<td>-0.018</td>
<td>-0.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0037)</td>
<td>(0.0062)</td>
<td>(0.0037)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.054</td>
<td>-0.060</td>
<td>-0.040</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00037)</td>
<td>(0.0036)</td>
<td>(0.0026)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>-0.043</td>
<td>-0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0033)</td>
<td>(0.0026)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>-0.056</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0026)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel B: ANS data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Unrestricted models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.102</td>
<td>-0.025</td>
<td>-0.035</td>
<td>-0.017</td>
<td>-0.017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0346)</td>
<td>(0.0060)</td>
<td>(0.0098)</td>
<td>(0.0057)</td>
<td>(0.0057)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.075</td>
<td>-0.075</td>
<td>-0.120</td>
<td>-0.071</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0069)</td>
<td>(0.0081)</td>
<td>(0.0093)</td>
<td>(0.0036)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>-0.010</td>
<td>-0.046</td>
<td>-0.051</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0134)</td>
<td>(0.0122)</td>
<td>(0.0036)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>0.056</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0131)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Restricted models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.026</td>
<td>-0.023</td>
<td>-0.025</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0346)</td>
<td>(0.0047)</td>
<td>(0.0034)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.071</td>
<td>-0.043</td>
<td>-0.109</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0036)</td>
<td>(0.0046)</td>
<td>(0.0034)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>-0.051</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0036)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>0.028</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0147)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel C: REP3 data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Unrestricted models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.103</td>
<td>-0.023</td>
<td>-0.029</td>
<td>0.015</td>
<td>0.015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0416)</td>
<td>(0.0053)</td>
<td>(0.0090)</td>
<td>(0.0095)</td>
<td>(0.0095)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.082</td>
<td>-0.089</td>
<td>-0.136</td>
<td>-0.074</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0069)</td>
<td>(0.0093)</td>
<td>(0.0103)</td>
<td>(0.0035)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>-0.017</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0137)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>0.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0072)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Restricted models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.029</td>
<td>-0.030</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0346)</td>
<td>(0.0048)</td>
<td>(0.0034)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.074</td>
<td>-0.070</td>
<td>-0.137</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0035)</td>
<td>(0.0048)</td>
<td>(0.0034)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>-0.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0032)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>0.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0034)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel D: NSS data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Unrestricted models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.105</td>
<td>-0.037</td>
<td>-0.030</td>
<td>0.007</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0361)</td>
<td>(0.0067)</td>
<td>(0.0086)</td>
<td>(0.0069)</td>
<td>(0.0069)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.073</td>
<td>-0.086</td>
<td>-0.088</td>
<td>-0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0099)</td>
<td>(0.0086)</td>
<td>(0.0083)</td>
<td>(0.0013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>-0.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0118)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>0.028</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0099)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Restricted models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.055</td>
<td>-0.023</td>
<td>-0.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0147)</td>
<td>(0.0026)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.050</td>
<td>-0.076</td>
<td>-0.066</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0071)</td>
<td>(0.0039)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>-0.027</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0147)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0290)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 9: This table shows the log-likelihood function value for the unrestricted and restricted general factor model. In addition, the table shows the Wald and LR test statistics for no-arbitrage. Factor estimation is done by maximum likelihood analysis on NS, ANS, REP3, and NSS data.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Panel A: NS data</th>
<th>Panel B: ANS data</th>
<th></th>
<th>Panel C: REP3 data</th>
<th>Panel D: NSS data</th>
</tr>
</thead>
<tbody>
<tr>
<td>ℓ_U</td>
<td>49829</td>
<td>55262</td>
<td>59319</td>
<td>62719</td>
<td>50299</td>
</tr>
<tr>
<td>ℓ_R</td>
<td>49808</td>
<td>55233</td>
<td>59304</td>
<td>62713</td>
<td>50251</td>
</tr>
<tr>
<td>LR</td>
<td>43.22</td>
<td>58.72</td>
<td>29.00</td>
<td>11.53</td>
<td>97.68</td>
</tr>
<tr>
<td>Test Prob.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.073</td>
<td>0.000</td>
</tr>
<tr>
<td>Wald-test</td>
<td>44.59</td>
<td>58.68</td>
<td>29.65</td>
<td>11.68</td>
<td>97.95</td>
</tr>
<tr>
<td>Test Prob.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.069</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 10: This table reports estimation of the DCHW model, both with and without the HJM drift restriction imposed, and results of tests for no-arbitrage. Panel A of the table provides parameter estimates, where figures for \( \mu \) are in basis points. Panel B shows the test of restrictions imposed by the HJM drift condition using the LR test.

<table>
<thead>
<tr>
<th></th>
<th>NS</th>
<th>ANS</th>
<th>SLSC</th>
<th>NSS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HJM drift</td>
<td>Unr. drift</td>
<td>HJM drift</td>
<td>Unr. drift</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>-0.0168</td>
<td>-0.0170</td>
<td>-0.0127</td>
<td>-0.0130</td>
</tr>
<tr>
<td></td>
<td>(0.0009)</td>
<td>(0.0009)</td>
<td>(0.0004)</td>
<td>(0.0006)</td>
</tr>
<tr>
<td>( a_1 )</td>
<td>0.0081</td>
<td>0.0080</td>
<td>0.0084</td>
<td>0.0083</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>-0.5123</td>
<td>-0.5029</td>
<td>-0.6498</td>
<td>-0.7216</td>
</tr>
<tr>
<td></td>
<td>(0.2436)</td>
<td>(0.2301)</td>
<td>(0.2626)</td>
<td>(0.2848)</td>
</tr>
</tbody>
</table>

Panel B: No-arbitrage test

<table>
<thead>
<tr>
<th></th>
<th>NS</th>
<th>ANS</th>
<th>SLSC</th>
<th>NSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ℓ</td>
<td>49093</td>
<td>49106</td>
<td>49683</td>
<td>49701</td>
</tr>
<tr>
<td>LR</td>
<td>27.58</td>
<td>36.10</td>
<td>26.29</td>
<td>19.90</td>
</tr>
<tr>
<td>Test Prob.</td>
<td>0.001</td>
<td>0.000</td>
<td>0.002</td>
<td>0.019</td>
</tr>
</tbody>
</table>
Table 11: This table reports estimation of the DCSLSC model, both with and without the HJM drift restriction imposed, and results of tests for no-arbitrage. Panel A of the table provides parameter estimates, where figures for \( \mu \) are in basis points. Panel B shows the test of restrictions imposed by the HJM drift condition using the LR test.

<table>
<thead>
<tr>
<th></th>
<th>NS</th>
<th>ANS</th>
<th>SLSC</th>
<th>NSS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HJM drift</td>
<td>Unr. drift</td>
<td>HJM drift</td>
<td>Unr. drift</td>
</tr>
<tr>
<td>( \sigma_2 )</td>
<td>0.3034</td>
<td>0.3030</td>
<td>0.3436</td>
<td>0.3438</td>
</tr>
<tr>
<td></td>
<td>(0.0022)</td>
<td>(0.0021)</td>
<td>(0.0026)</td>
<td>(0.0026)</td>
</tr>
<tr>
<td>( \sigma_0 )</td>
<td>0.0149</td>
<td>0.0150</td>
<td>0.0141</td>
<td>0.0143</td>
</tr>
<tr>
<td></td>
<td>(0.0094)</td>
<td>(0.0004)</td>
<td>(0.0003)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>( \sigma_1 )</td>
<td>0.0181</td>
<td>0.0181</td>
<td>0.0169</td>
<td>0.0169</td>
</tr>
<tr>
<td></td>
<td>(0.0094)</td>
<td>(0.0004)</td>
<td>(0.0004)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>( \sigma_3 )</td>
<td>0.0328</td>
<td>0.0327</td>
<td>0.0296</td>
<td>0.0296</td>
</tr>
<tr>
<td></td>
<td>(0.0008)</td>
<td>(0.0008)</td>
<td>(0.0007)</td>
<td>(0.0007)</td>
</tr>
<tr>
<td>( \lambda_1 )</td>
<td>-0.3199</td>
<td>-0.3849</td>
<td>-0.3182</td>
<td>-0.3182</td>
</tr>
<tr>
<td></td>
<td>(0.1443)</td>
<td>(0.1421)</td>
<td>(0.1419)</td>
<td>(0.1419)</td>
</tr>
<tr>
<td>( \lambda_2 )</td>
<td>-0.2702</td>
<td>-0.2708</td>
<td>-0.2834</td>
<td>-0.2834</td>
</tr>
<tr>
<td></td>
<td>(0.1487)</td>
<td>(0.1423)</td>
<td>(0.1424)</td>
<td>(0.1424)</td>
</tr>
<tr>
<td>( \lambda_3 )</td>
<td>-0.0298</td>
<td>-0.1164</td>
<td>-0.1180</td>
<td>-0.1180</td>
</tr>
<tr>
<td></td>
<td>(0.1473)</td>
<td>(0.1429)</td>
<td>(0.1441)</td>
<td>(0.1441)</td>
</tr>
</tbody>
</table>

Panel B: No-arbitrage test

| \( \ell \) | 57568   | 57578   | 56936   | 56948   | 53754   | 53756   | 51669   | 51672   |
| LR          | 18.70   | 23.63   | 4.98    | 4.98    | 5.13    | 5.13    |
| Test. Prob. | 0.009   | 0.001   | 0.662   | 0.644   |
A.3 Figures

Figure 1: Number of observations

This figure shows the number of observations included in the empirical application in each calendar week. The observations consist of weekly US Treasury bill, note, and bond prices retrieved from CRSP.
**Figure 2:** Plot of yields and slope-adjusted yield changes fitted by NS yield curve

This figure shows a three-dimensional plot of yields (upper panel) and slope-adjusted yield changes (lower panel) with weekly observations fitted by the NS yield curve. We refer to (3.3.3) for an exact specification for the slope-adjusted yield changes. The sample period is from January 1999 to December 2014 at maturities 0.25, 0.5, 1, 2, 3, 5, 7, 10, 15, 20, and 25 years.
**Figure 3:** Volatility structure in slope-adjusted yield changes extracted by PFA

This figure shows the estimated three-factor volatility structure extracted by PFA on NS, ANS, REP3, and NSS slope-adjusted yield changes.
**Figure 4:** Volatility structure in parameterized and general DTSMs extracted by maximum likelihood factor analysis

This figure shows the estimated volatility structure of the unrestricted general models and the unrestricted parameterized models. Estimation is done by maximum likelihood analysis on NS and ANS slope-adjusted yield changes.
Figure 5: Volatility structure in parameterized and general DTSMs extracted by maximum likelihood factor analysis

This figure shows the estimated volatility structure of the unrestricted general models and the unrestricted parameterized models. Estimation is done by maximum likelihood analysis on REP3 and NSS slope-adjusted yield changes.
Figure 6: Unique variances in parameterized and general DTSMs extracted by maximum likelihood factor analysis

This figure shows the estimated unique variances for the unrestricted HL/HW, HJM2, EHJM3 and EHJM4 models and the unrestricted general one-, two-, three- and four-factor models. Estimation is done by maximum likelihood analysis on slope-adjusted yield changes.
Figure 7: Sample mean and estimated HJM drift in NS slope-adjusted yield changes

Illustration of the difference between the sample mean and the $\Delta \hat{\gamma}$-estimates for the unrestricted factor models and the unrestricted specific models. The $\Delta \hat{\gamma}$-estimates have been calculated from the maximum likelihood estimates $\hat{\sigma}$ and the linear regression estimates $\hat{\lambda}$ on NS slope-adjusted yield changes.
**Figure 8:** Sample mean and estimated HJM drift in ANS slope-adjusted yield changes data

Illustration of the difference between the sample mean and the $\Delta\hat{\gamma}$-estimates for the unrestricted factor models and the unrestricted specific models. The $\Delta\hat{\gamma}$-estimates have been calculated from the maximum likelihood estimates $\hat{\sigma}$ and the linear regression estimates $\hat{\lambda}$ on ANS slope-adjusted yield changes.
**Figure 9:** Sample mean and estimated HJM drift in REP3 slope-adjusted yield changes

Illustration of the difference between the sample mean and the $\Delta \hat{\gamma}$-estimates for the unrestricted factor models and the unrestricted specific models. The $\Delta \hat{\gamma}$-estimates have been calculated from the maximum likelihood estimates $\hat{\sigma}$ and the linear regression estimates $\hat{\lambda}$ on REP3 slope-adjusted yield changes.
Figure 10: Sample mean and estimated HJM drift in NSS slope-adjusted yield changes

Illustration of the difference between the sample mean and the $\Delta \hat{\gamma}$-estimates for the unrestricted factor models and the unrestricted specific models. The $\Delta \hat{\gamma}$-estimates have been calculated from the maximum likelihood estimates $\hat{\sigma}$ and the linear regression estimates $\hat{\lambda}$ on NSS slope-adjusted yield changes.
This figure shows the estimated volatility structure of the restricted DCHW models. Estimation is done by maximum likelihood analysis on NS, ANS, SLSC, and NSS slope-adjusted yield changes.
**Figure 12:** Volatility structure in restricted DCSLSC model extracted by maximum likelihood factor analysis

This figure shows the estimated volatility structure of the restricted DCSLSC models. Estimation is done by maximum likelihood analysis on NS, ANS, SLSC, and NSS slope-adjusted yield changes.

**Figure 13:** Unique variances in restricted DCHW and DCSLSC models extracted by maximum likelihood factor analysis

This figure shows the unique variances for the restricted DCHW and DCSLSC models estimated on NS, ANS, SLSC, and NSS slope-adjusted yield changes.
Figure 14: Sample mean and estimated HJM drift in DCHW and DCSLSC models

Illustration of the difference between the sample mean and the HJM drift condition for the restricted DCHW and DCSLSC models. The HJM drifts are calculated from parameters $a_0, a_1, \tilde{a}_3, \lambda_1, \lambda_2, \lambda_3$ on NS, ANS, SLSC, and NSS slope-adjusted yield changes.
# Declaration of co-authorship

Full name of the PhD student: Jorge Wolfgang Hansen

This declaration concerns the following article/manuscript:

<table>
<thead>
<tr>
<th>Title</th>
<th>Immunization with term structure dynamics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors:</td>
<td>Daniel Borup, Bent Jesper Christensen, Jorge Wolfgang Hansen, Torben B. Rasmussen</td>
</tr>
</tbody>
</table>

The article/manuscript is: Published ☐ Accepted ☐ Submitted ☐ In preparation ☑

If published, state full reference:

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☐ Yes ☑ If yes, give details: A first version has been used in Torben B. Rasmussen’s PhD dissertation.

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

<table>
<thead>
<tr>
<th>Element</th>
<th>Extent (A-E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Formulation/identification of the scientific problem</td>
<td>C</td>
</tr>
<tr>
<td>2. Planning of the experiments/methodology design and development</td>
<td>C</td>
</tr>
<tr>
<td>3. Involvement in the experimental work/clinical studies/data collection</td>
<td>B</td>
</tr>
<tr>
<td>4. Interpretation of the results</td>
<td>C</td>
</tr>
<tr>
<td>5. Writing of the first draft of the manuscript</td>
<td>C</td>
</tr>
<tr>
<td>6. Finalization of the manuscript and submission</td>
<td>C</td>
</tr>
</tbody>
</table>

## Signatures of the co-authors

<table>
<thead>
<tr>
<th>Date</th>
<th>Name</th>
<th>Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/12/2019</td>
<td>Bent Jesper Christensen</td>
<td>[Signature]</td>
</tr>
<tr>
<td>14/12/2019</td>
<td>Daniel Borup</td>
<td>[Signature]</td>
</tr>
<tr>
<td>12/1/2019</td>
<td>Torben B. Rasmussen</td>
<td>[Signature]</td>
</tr>
</tbody>
</table>

Date: 28.02.20

In case of further co-authors please attach appendix

Date: 28.02.20

Signature of the PhD student
## Declaration of co-authorship

Full name of the PhD student: Jorge Wolfgang Hansen

This declaration concerns the following article/manuscript:

<table>
<thead>
<tr>
<th>Title:</th>
<th>The impact of curve-fitting procedure on estimation and testing of term structure models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors:</td>
<td>Bent Jesper Christensen, Jorge Wolfgang Hansen</td>
</tr>
</tbody>
</table>

The article/manuscript is: Published ☐ Accepted ☐ Submitted ☐ In preparation ☒

If published, state full reference:

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

<table>
<thead>
<tr>
<th>Element</th>
<th>Extent (A-E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Formulation/identification of the scientific problem</td>
<td>C</td>
</tr>
<tr>
<td>2. Planning of the experiments/methodology design and development</td>
<td>C</td>
</tr>
<tr>
<td>3. Involvement in the experimental work/clinical studies/data collection</td>
<td>B</td>
</tr>
<tr>
<td>4. Interpretation of the results</td>
<td>B</td>
</tr>
<tr>
<td>5. Writing of the first draft of the manuscript</td>
<td>B</td>
</tr>
<tr>
<td>6. Finalization of the manuscript and submission</td>
<td>B</td>
</tr>
</tbody>
</table>

### Signatures of the co-authors

<table>
<thead>
<tr>
<th>Date</th>
<th>Name</th>
<th>Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>19/2/2019</td>
<td>Bent Jesper Christensen</td>
<td></td>
</tr>
</tbody>
</table>

In case of further co-authors please attach appendix

Date: 28.02.20

Signature of the PhD student