Available online at www.sciencedirect.com
SCIENCE DIRECT

JOURNAL OF Econometrics

Journal of Econometrics 130 (2006) 337–364

www.elsevier.com/locate/jeconom

Forecasting the term structure of government bond yields

Francis X. Diebold[a,b], Canlin Li[c,*]

[a]*Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297, USA*

[b]*NBER, 1050 Massachusetts Ave., Cambridge, MA 02138, USA*

[c]*A. Gary Anderson Graduate School of Management, University of California, Riverside, CA 92521, USA*

Accepted 21 March 2005
Available online 23 May 2005

Abstract

Despite powerful advances in yield curve modeling in the last 20 years, comparatively little attention has been paid to the key practical problem of forecasting the yield curve. In this paper we do so. We use neither the no-arbitrage approach nor the equilibrium approach. Instead, we use variations on the Nelson–Siegel exponential components framework to model the entire yield curve, period-by-period, as a three-dimensional parameter evolving dynamically. We show that the three time-varying parameters may be interpreted as factors corresponding to level, slope and curvature, and that they may be estimated with high efficiency. We propose and estimate autoregressive models for the factors, and we show that our models are consistent with a variety of stylized facts regarding the yield curve. We use our models to produce term-structure forecasts at both short and long horizons, with encouraging results. In particular, our forecasts appear much more accurate at long horizons than various standard benchmark forecasts.

JEL classification: G1; E4; C5

Keywords: Term structure; Yield curve; Factor model; Nelson–Siegel curve

*Corresponding author.
E-mail addresses: fdiebold@sas.upenn.edu (F.X. Diebold), canlin.li@ucr.edu (C. Li).

0304-4076/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.jeconom.2005.03.005

1. Introduction

The last 25 years have produced major advances in theoretical models of the term structure as well as their econometric estimation. Two popular approaches to term structure modeling are no-arbitrage models and equilibrium models. The no-arbitrage tradition focuses on perfectly fitting the term structure at a point in time to ensure that no arbitrage possibilities exist, which is important for pricing derivatives. The equilibrium tradition focuses on modeling the dynamics of the instantaneous rate, typically using affine models, after which yields at other maturities can be derived under various assumptions about the risk premium.[1] Prominent contributions in the no-arbitrage vein include Hull and White (1990) and Heath et al. (1992), and prominent contributions in the affine equilibrium tradition include Vasicek (1977), Cox et al. (1985), and Duffie and Kan (1996).

Interest rate point forecasting is crucial for bond portfolio management, and interest rate density forecasting is important for both derivatives pricing and risk management.[2] Hence one wonders what the modern models have to say about interest rate forecasting. It turns out that, despite the impressive theoretical advances in the financial economics of the yield curve, surprisingly little attention has been paid to the key practical problem of yield curve forecasting. The arbitrage-free term structure literature has little to say about dynamics or forecasting, as it is concerned primarily with fitting the term structure at a point in time. The affine equilibrium term structure literature is concerned with dynamics driven by the short rate, and so is potentially linked to forecasting, but most papers in that tradition, such as de Jong (2000) and Dai and Singleton (2000), focus only on in-sample fit as opposed to out-of-sample forecasting. Moreover, those that *do* focus on out-of-sample forecasting, notably Duffee (2002), conclude that the models forecast poorly.

In this paper we take an explicitly out-of-sample forecasting perspective, and we use neither the no-arbitrage approach nor the equilibrium approach. Instead, we use the Nelson and Siegel (1987) exponential components framework to distill the entire yield curve, period-by-period, into a three-dimensional parameter that evolves dynamically. We show that the three time-varying parameters may be interpreted as factors. Unlike factor analysis, however, in which one estimates both the unobserved factors and the factor loadings, the Nelson–Siegel framework imposes structure on the factor loadings.[3] Doing so not only facilitates highly precise estimation of the factors, but, as we show, it also lets us interpret the factors as level, slope and curvature. We propose and estimate autoregressive models for the factors, and then we forecast the yield curve by forecasting the factors. Our results are encouraging; in

[1]The empirical literature that models yields as a cointegrated system, typically with one underlying stochastic trend (the short rate) and stationary spreads relative to the short rate, is similar in spirit. See Diebold and Sharpe (1990), Hall et al. (1992), Shea (1992), Swanson and White (1995), and Pagan et al. (1996).

[2]For comparative discussion of point and density forecasting, see Diebold et al. (1998) and Diebold et al. (1999).

[3]Classic unrestricted factor analyses include Litterman and Scheinkman (1991) and Knez et al. (1994).

particular, our models produce one-year-ahead forecasts that are noticeably more accurate than standard benchmarks.

Related work includes the factor models of Litzenberger et al. (1995), Bliss (1997a,b), Dai and Singleton (2000), de Jong and Santa-Clara (1999), de Jong (2000), Brandt and Yaron (2001) and Duffee (2002). Particularly relevant are the three-factor models of Balduzzi et al. (1996), Chen (1996), and especially the Andersen and Lund (1997) model with stochastic mean and volatility, whose three factors are interpreted in terms of level, slope and curvature. We will subsequently discuss related work in greater detail; for now, suffice it to say that little of it considers forecasting directly, and that our approach, although related, is indeed very different.

We proceed as follows. In Section 2 we provide a detailed description of our modeling framework, which interprets and extends earlier work in ways linked to recent developments in multifactor term structure modeling, and we also show how it can replicate a variety of stylized facts about the yield curve. In Section 3 we proceed to an empirical analysis, describing the data, estimating the models, and examining out-of-sample forecasting performance. In Section 4 we offer interpretive concluding remarks.

2. Modeling and forecasting the term structure I: methods

Here we introduce the framework that we use for fitting and forecasting the yield curve. We argue that the well-known Nelson and Siegel (1987) curve is well-suited to our ultimate forecasting purposes, and we introduce a novel twist of interpretation, showing that the three coefficients in the Nelson–Siegel curve may be interpreted as latent level, slope and curvature factors. We also argue that the nature of the factors and factor loadings implicit in the Nelson–Siegel model facilitate consistency with various empirical properties of the yield curve that have been cataloged over the years. Finally, motivated by our interpretation of the Nelson–Siegel model as a three-factor model of level, slope and curvature, we contrast it to various multi-factor models that have appeared in the literature.

2.1. Constructing "Raw" yields

Let us first fix ideas and establish notation by introducing three key theoretical constructs and the relationships among them: the discount curve, the forward curve, and the yield curve. Let $P_t(\tau)$ denote the price of a τ-period discount bond, i.e., the present value at time t of \$1 receivable τ periods ahead, and let $y_t(\tau)$ denote its continuously compounded zero-coupon nominal yield to maturity. From the yield curve we obtain the discount curve,

$$P_t(\tau) = \mathrm{e}^{-\tau y_t(\tau)},$$

and from the discount curve we obtain the instantaneous (nominal) forward rate curve,

$$f_t(\tau) = -P_t'(\tau)/P_t(\tau).$$

The relationship between the yield to maturity and the forward rate is therefore

$$y_t(\tau) = \frac{1}{\tau}\int_0^{\tau} f_t(u)\,\mathrm{d}u,$$

which implies that the zero-coupon yield is an equally-weighed average of forward rates. Given the yield curve or forward curve, we can price any coupon bond as the sum of the present values of future coupon and principal payments.

In practice, yield curves, discount curves and forward curves are not observed. Instead, they must be estimated from observed bond prices. Two popular approaches to constructing yields proceed by estimating a smooth discount curve and then converting to yields at the relevant maturities via the above formulae. The first discount-curve approach to yield construction is due to McCulloch (1975) and McCulloch and Kwon (1993), who model the discount curve with a cubic spline. The fitted discount curve, however, diverges at long maturities instead of converging to zero. Hence such curves provide a poor fit to yield curves that are flat or have a flat long end, which requires an exponentially decreasing discount function.

A second discount-curve approach to yield construction is due to Vasicek and Fong (1982), who fit exponential splines to the discount curve, using a negative transformation of maturity instead of maturity itself, which ensures that the forward rates and zero-coupon yields converge to a fixed limit as maturity increases. Hence the Vasicek–Fong model is more successful at fitting yield curves with flat long ends. It has problems of its own, however, because its estimation requires iterative nonlinear optimization, and it can be hard to restrict the implied forward rates to be positive.

A third and very popular approach to yield construction is due to Fama and Bliss (1987), who construct yields not via an estimated discount curve, but rather via estimated forward rates at the observed maturities. Their method sequentially constructs the forward rates necessary to price successively longer-maturity bonds, often called an "unsmoothed Fama–Bliss" forward rates, and then constructs "unsmoothed Fama–Bliss yields" by averaging the appropriate unsmoothed Fama–Bliss forward rates. The unsmoothed Fama–Bliss yields exactly price the included bonds. Throughout this paper, we model and forecast the unsmoothed Fama–Bliss yields.

2.2. Modeling yields: the Nelson–Siegel yield curve and its interpretation

At any given time, we have a large set of (Fama–Bliss unsmoothed) yields, to which we fit a parametric curve for purposes of modeling and forecasting. Throughout this paper, we use the Nelson and Siegel (1987) functional form, which is a convenient and parsimonious three-component exponential approximation. In particular, Nelson and Siegel (1987), as extended by Siegel and Nelson (1988), work with the forward rate curve,

$$f_t(\tau) = \beta_{1t} + \beta_{2t}\mathrm{e}^{-\lambda_t\tau} + \beta_{3t}\lambda_t\mathrm{e}^{-\lambda_t\tau}.$$

The Nelson–Siegel forward rate curve can be viewed as a constant plus a Laguerre function, which is a polynomial times an exponential decay term and is a popular mathematical approximating function.[4] The corresponding yield curve is

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau}\right) + \beta_{3t}\left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau}\right).$$

The Nelson–Siegel yield curve also corresponds to a discount curve that begins at one at zero maturity and approaches zero at infinite maturity, as appropriate.

Let us now interpret the parameters in the Nelson–Siegel model. The parameter λ_t governs the exponential decay rate; small values of λ_t produce slow decay and can better fit the curve at long maturities, while large values of λ_t produce fast decay and can better fit the curve at short maturities. λ_t also governs where the loading on β_{3t} achieves its maximum.[5]

We interpret β_{1t}, β_{2t} and β_{3t} as three latent dynamic factors. The loading on β_{1t} is 1, a constant that does not decay to zero in the limit; hence it may be viewed as a long-term factor. The loading on β_{2t} is $(1 - e^{-\lambda_t \tau})/\lambda_t \tau$, a function that starts at 1 but decays monotonically and quickly to 0; hence it may be viewed as a short-term factor. The loading on β_{3t} is $((1 - e^{-\lambda_t \tau})/\lambda_t \tau) - e^{-\lambda_t \tau}$, which starts at 0 (and is thus not short-term), increases, and then decays to zero (and thus is not long-term); hence it may be viewed as a medium-term factor. We plot the three factor loadings in Fig. 1. They are similar to those obtained by Bliss (1997a), who estimated loadings via a statistical factor analysis.[6]

An important insight is that the three factors, which following the literature we have thus far called long-term, short-term and medium-term, may also be interpreted in terms of level, slope and curvature. The long-term factor β_{1t}, for example, governs the yield curve level. In particular, one can easily verify that $y_t(\infty) = \beta_{1t}$. Alternatively, note that an increase in β_{1t} increases all yields equally, as the loading is identical at all maturities, thereby changing the level of the yield curve.

The short-term factor β_{2t} is closely related to the yield curve slope, which we define as the ten-year yield minus the three-month yield. In particular, $y_t(120) - y_t(3) = -0.78\beta_{2t} + 0.06\beta_{3t}$. Some authors such as Frankel and Lown (1994), moreover, define the yield curve slope as $y_t(\infty) - y_t(0)$, which is *exactly* equal to $-\beta_{2t}$. Alternatively, note that an increase in β_{2t} increases short yields more than long yields, because the short rates load on β_{2t} more heavily, thereby changing the slope of the yield curve.

We have seen that β_{1t} governs the level of the yield curve and β_{2t} governs its slope. It is interesting to note, moreover, that the instantaneous yield depends on *both* the level and slope factors, because $y_t(0) = \beta_{1t} + \beta_{2t}$. Several other models have the same implication. In particular, Dai and Singleton (2000) show that the

[4]See, for example, Courant and Hilbert (1953).

[5]Throughout this paper, and for reasons that will be discussed subsequently in detail, we set $\lambda_t = 0.0609$ for all t.

[6]Factors are typically not uniquely identified in factor analysis. Bliss (1997a) rotates the first factor so that its loading is a vector of ones. In our approach, the unit loading on the first factor is imposed from the beginning, which potentially enables us to estimate the other factors more efficiently.

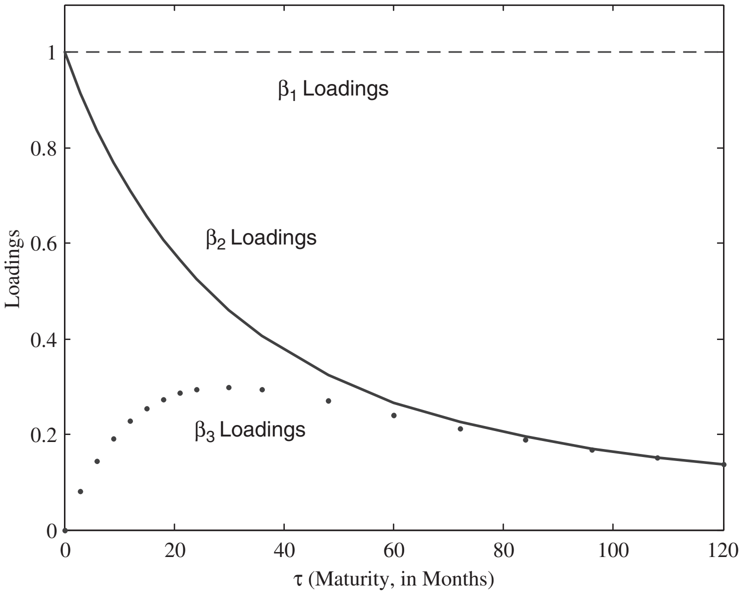

Fig. 1. Factor loadings. We plot the factor loadings in the three-factor model,

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right),$$

where the three factors are β_{1t}, β_{2t}, and β_{3t}, the associated loadings are 1, $(1-e^{-\lambda_t\tau})/\lambda_t\tau$, and $(1-e^{-\lambda_t\tau})/\lambda_t\tau - e^{-\lambda_t\tau}$, and τ denotes maturity. We fix $\lambda_t = 0.0609$.

three-factor models of Balduzzi et al. (1996) and Chen (1996) impose the restrictions that the instantaneous yield is an affine function of only two of the three state variables, a property shared by the Andersen and Lund (1997) three-factor nonaffine model.

Finally, the medium-term factor β_{3t} is closely related to the yield curve curvature, which we define as twice the two-year yield minus the sum of the ten-year and three-month yields. In particular, $2y_t(24) - y_t(3) - y_t(120) = 0.00053\beta_{2t} + 0.37\beta_{3t}$. Alternatively, note that an increase in β_{3t} will have little effect on very short or very long yields, which load minimally on it, but will increase medium-term yields, which load more heavily on it, thereby increasing yield curve curvature.

Now that we have interpreted Nelson–Siegel as a three-factor of level, slope and curvature, it is appropriate to contrast it to Litzenberger et al. (1995), which is highly related yet distinct. First, although Litzenberger et al. model the discount curve $P_t(\tau)$ using exponential components and we model the yield curve $y_t(\tau)$ using exponential components, the yield curve is a log transformation of the discount curve because $y_t(\tau) = -\log P_t(\tau)/\tau$, so the two approaches are equivalent in the one-factor case. In the multi-factor case, however, a sum of factors in the yield curve will not be a sum in the discount curve, so there is generally no simple mapping between the approaches.

Second, both we and Litzenberger et al. provide novel interpretations of the parameters of fitted curves. Litzenberger et al., however, do not interpret parameters directly as factors.

In closing this sub-section, it is worth noting that what we have called the "Nelson–Siegel curve" is actually a different factorization than the one originally advocated by Nelson and Siegel (1987), who used

$$y_t(\tau) = b_{1t} + b_{2t}\,\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - b_{3t} e^{-\lambda_t \tau}.$$

Obviously the Nelson–Siegel factorization matches ours with $b_{1t} = \beta_{1t}$, $b_{2t} = \beta_{2t} + \beta_{3t}$, and $b_{3t} = \beta_{3t}$. Ours is preferable, however, for reasons that we are now in a position to appreciate. First, $(1 - e^{-\lambda_t \tau})/\lambda_t \tau$ and $e^{-\lambda_t \tau}$ have similar monotonically decreasing shape, so if we were to interpret b_2 and b_3 as factors, then their loadings would be forced to be very similar, which creates at least two problems. First, conceptually, it would be hard to provide intuitive interpretations of the factors in the original Nelson–Siegel framework. Second, operationally, it would be difficult to estimate the factors precisely, because the high coherence in the factors produces multicolinearity.

2.3. *Stylized facts of the yield curve and the model's potential ability to replicate them*

A good model of yield curve dynamics should be able to reproduce the historical stylized facts concerning the average shape of the yield curve, the variety of shapes assumed at different times, the strong persistence of yields and weak persistence of spreads, and so on. It is not easy for a parsimonious model to accord with all such facts.

Let us consider some of the most important stylized facts and the ability of our model to replicate them, in principle:

(1) The average yield curve is increasing and concave. In our framework, the average yield curve is the yield curve corresponding to the average values of β_{1t}, β_{2t} and β_{3t}. It is certainly possible in principle that it may be increasing and concave.
(2) The yield curve assumes a variety of shapes through time, including upward sloping, downward sloping, humped, and inverted humped. The yield curve in our framework can assume all of those shapes. Whether and how often it does depends upon the variation in β_{1t}, β_{2t} and β_{3t}.
(3) Yield dynamics are persistent, and spread dynamics are much less persistent. Persistent yield dynamics would correspond to strong persistence of β_{1t}, and less persistent spread dynamics would correspond to weaker persistence of β_{2t}.
(4) The short end of the yield curve is more volatile than the long end. In our framework, this is reflected in factor loadings: the short end depends positively on both β_{1t} and β_{2t}, whereas the long end depends only on β_{1t}.
(5) Long rates are more persistent than short rates. In our framework, long rates depend only on β_{1t}. If β_{1t} is the most persistent factor, then long rates will be more persistent than short rates.

Overall, it seems clear that our framework is consistent, at least in principle, with many of the key stylized facts of yield curve behavior. Whether principle accords with practice is an empirical matter, to which we now turn.

3. Modeling and forecasting the term structure II: empirics

In this section, we estimate and assess the fit of the three-factor model in a time series of cross sections, after which we model and forecast the extracted level, slope and curvature components. We begin by introducing the data.

3.1. The data

We use end-of-month price quotes (bid-ask average) for U.S. Treasuries, from January 1985 through December 2000, taken from the CRSP government bonds files. CRSP filters the data, eliminating bonds with option features (callable and flower bonds), and bonds with special liquidity problems (notes and bonds with less than one year to maturity, and bills with less than one month to maturity), and then converts the filtered bond prices to unsmoothed Fama and Bliss (1987) forward rates. Then, using programs and CRSP data kindly supplied by Rob Bliss, we convert the unsmoothed Fama–Bliss forward rates into unsmoothed Fama–Bliss zero yields.

Although most of our analysis does not require the use of fixed maturities, doing so greatly simplifies our subsequent forecasting exercises. Hence we pool the data into fixed maturities. Because not every month has the same maturities available, we linearly interpolate nearby maturities to pool into fixed maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months, where a month is defined as 30.4375 days. Although there is no bond with exactly 30.4375 days to maturity, each month there are many bonds with either 30, 31, 32, 33, or 34 days to maturity. Similarly we obtain data for maturities of 3 months, 6 months, etc.[7]

The various yields, as well as the yield curve level, slope and curvature defined above, will play a prominent role in the sequel. Hence we focus on them now in some detail. In Fig. 2 we provide a three-dimensional plot of our yield curve data. The large amount of temporal variation in the level is visually apparent. The variation in slope and curvature is less strong, but nevertheless apparent. In Table 1, we present descriptive statistics for the yields. It is clear that the typical yield curve is upward sloping, that the long rates are less volatile and more persistent than short rates, that the level (120-month yield) is highly persistent but varies only moderately relative to its mean, that the slope is less persistent than any individual yield but quite highly variable relative to its mean, and the curvature is the least persistent of all factors and the most highly variable relative to its mean. It is also worth

[7]We checked the derived dataset and verified that the difference between it and the original dataset is only one or two basis points.

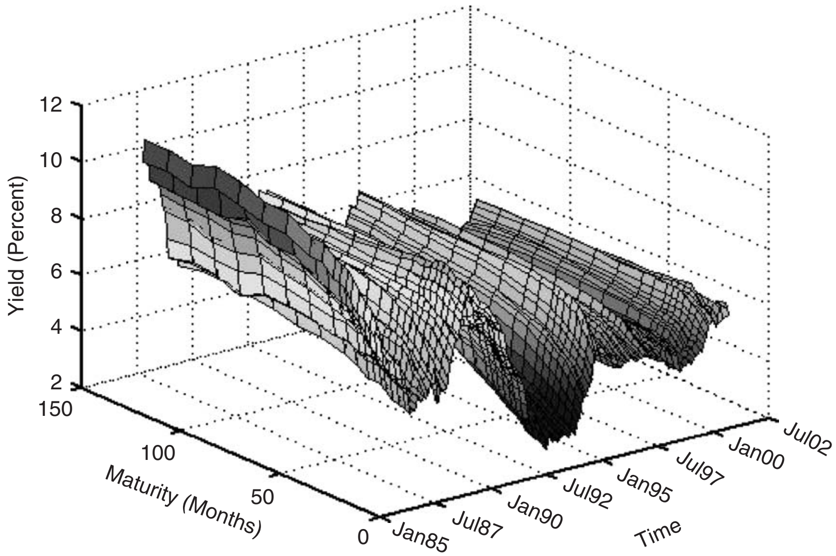

Fig. 2. Yield curves, 1985.01–2000.12. The sample consists of monthly yield data from January 1985 to December 2000 at maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months.

Table 1
Descriptive statistics, yield curves

Maturity (Months)	Mean	Std. dev.	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
3	5.630	1.488	2.732	9.131	0.978	0.569	−0.079
6	5.785	1.482	2.891	9.324	0.976	0.555	−0.042
9	5.907	1.492	2.984	9.343	0.973	0.545	−0.005
12	6.067	1.501	3.107	9.683	0.969	0.539	0.021
15	6.225	1.504	3.288	9.988	0.968	0.527	0.060
18	6.308	1.496	3.482	10.188	0.965	0.513	0.089
21	6.375	1.484	3.638	10.274	0.963	0.502	0.115
24	6.401	1.464	3.777	10.413	0.960	0.481	0.133
30	6.550	1.462	4.043	10.748	0.957	0.479	0.190
36	6.644	1.439	4.204	10.787	0.956	0.471	0.226
48	6.838	1.439	4.308	11.269	0.951	0.457	0.294
60	6.928	1.430	4.347	11.313	0.951	0.464	0.336
72	7.082	1.457	4.384	11.653	0.953	0.454	0.372
84	7.142	1.425	4.352	11.841	0.948	0.448	0.391
96	7.226	1.413	4.433	11.512	0.954	0.468	0.417
108	7.270	1.428	4.429	11.664	0.953	0.475	0.426
120 (level)	7.254	1.432	4.443	11.663	0.953	0.467	0.428
Slope	1.624	1.213	−0.752	4.060	0.961	0.405	−0.049
Curvature	−0.081	0.648	−1.837	1.602	0.896	0.337	−0.015

Note: We present descriptive statistics for monthly yields at different maturities, and for the yield curve level, slope and curvature, where we define the level as the 10-year yield, the slope as the difference between the 10-year and 3-month yields, and the curvature as the twice the 2-year yield minus the sum of the 3-month and 10-year yields. The last three columns contain sample autocorrelations at displacements of 1, 12, and 30 months. The sample period is 1985:01–2000:12.

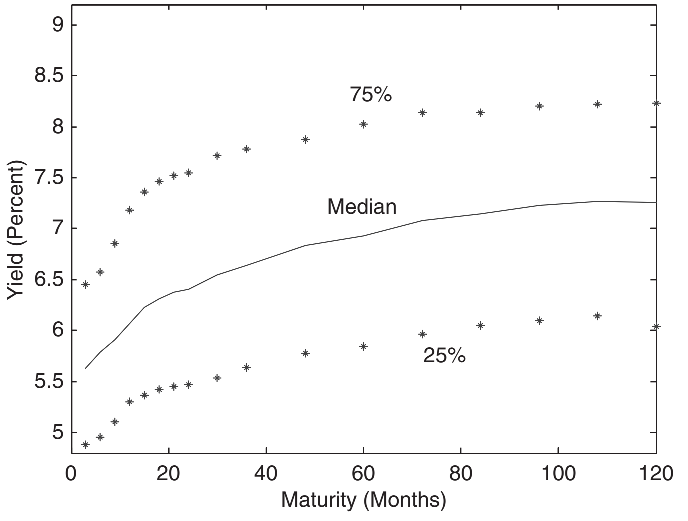

Fig. 3. Median data-based yield curve with pointwise interquartile range. For each maturity, we plot the median yield along with the 25th and 75th percentiles.

noting, because it will be relevant for our future modeling choices, that level, slope and curvature are not highly correlated with each other; all pairwise correlations are less than 0.40. In Fig. 3 we display the median yield curve together with pointwise interquartile ranges. The earlier-mentioned upward sloping pattern, with long rates less volatile than short rates, is apparent. One can also see that the distributions of yields around their medians tend to be asymmetric, with a long right tail.

3.2. Fitting yield curves

As discussed above, we fit the yield curve using the three-factor model,

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-\mathrm{e}^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3t}\left(\frac{1-\mathrm{e}^{-\lambda_t\tau}}{\lambda_t\tau} - \mathrm{e}^{-\lambda_t\tau}\right).$$

We could estimate the parameters $\theta_t = \{\beta_{1t}, \beta_{2t}, \beta_{3t}, \lambda_t\}$ by nonlinear least squares, for each month t. Following standard practice tracing to Nelson and Siegel (1987), however, we instead fix λ_t at a prespecified value, which lets us compute the values of the two regressors (factor loadings) and use ordinary least squares to estimate the betas (factors), for each month t. Doing so enhances not only simplicity and convenience, but also numerical trustworthiness, by enabling us to replace hundreds of potentially challenging numerical optimizations with trivial least-squares regressions. The question arises, of course, as to an appropriate value for λ_t. Recall that λ_t determines the maturity at which the loading on the medium-term, or curvature, factor achieves it maximum. Two- or three-year maturities are commonly

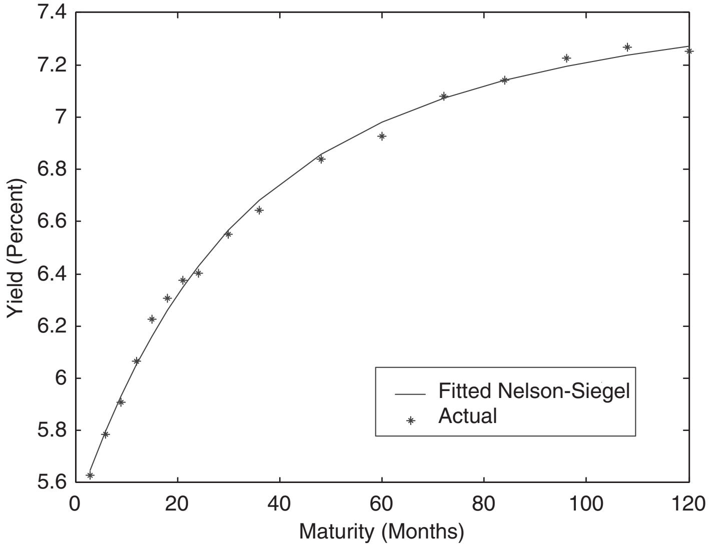

Fig. 4. Actual (data-based) and fitted (model-based) average yield curve. We show the actual average yield curve and the fitted average yield curve obtained by evaluating the Nelson–Siegel function at the mean values of $\hat{\beta}_{1t}$, $\hat{\beta}_{2t}$ and $\hat{\beta}_{3t}$ from Table 3.

used in that regard, so we simply picked the average, 30 months. The λ_t value that maximizes the loading on the medium-term factor at exactly 30 months is $\lambda_t = 0.0609$.

Applying ordinary least squares to the yield data for each month gives us a time series of estimates of $\{\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}\}$ and a corresponding panel of residuals, or pricing errors. Note that, because the maturities are not equally spaced, we implicitly weight the most "active" region of the yield curve most heavily when fitting the model.[8] There are many aspects to a full assessment of the "fit" of our model. In Fig. 4 we plot the implied average fitted yield curve against the average actual yield curve. The two agree quite closely. In Fig. 5 we dig deeper by plotting the raw yield curve and the three-factor fitted yield curve for some selected dates. Clearly the three-factor model is capable of replicating a variety of yield curve shapes: upward sloping, downward sloping, humped, and inverted humped. It does, however, have difficulties at some dates, especially when yields are dispersed, with multiple interior minima and maxima. Overall, however, the residual plot in Fig. 6 indicates a good fit.

In Table 2 we present statistics that describe the in-sample fit. The residual sample autocorrelations indicate that pricing errors are persistent. As noted in Bliss (1997b), regardless of the term structure estimation method used, there is a persistent

[8]Other weightings and loss functions have been explored by Bliss (1997b), Soderlind and Svensson (1997), and Bates (1999).

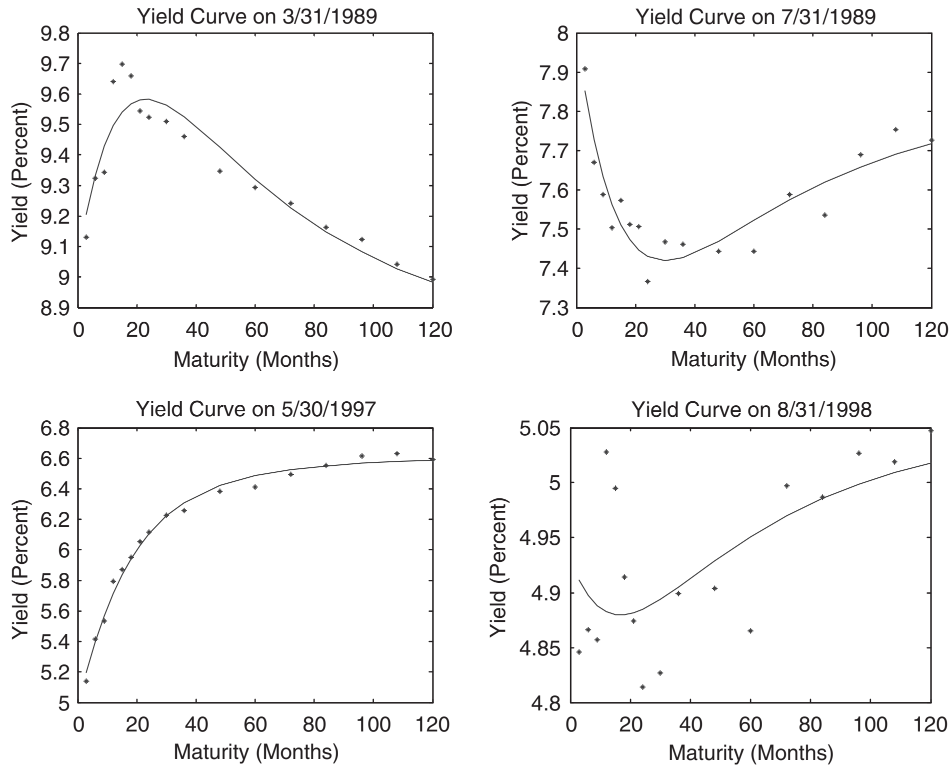

Fig. 5. Selected fitted (model-based) yield curves. We plot fitted yield curves for selected dates, together with actual yields. See text for details.

discrepancy between actual bond prices and prices estimated from term structure models. Presumably these discrepancies arise from persistent tax and/or liquidity effects.[9] However, because they persist, they should vanish from fitted yield changes.

In Fig. 7 we plot $\{\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}\}$ along with the empirical level, slope and curvature defined earlier. The figure confirms our assertion that the three factors in our model correspond to level, slope and curvature. The correlations between the estimated factors and the empirical level, slope, and curvature are $\rho(\hat{\beta}_{1t}, I_t) = 0.97$, $\rho(\hat{\beta}_{2t}, s_t) = -0.99$, and $\rho(\hat{\beta}_{3t}, c_t) = 0.99$, where (l_t, s_t, c_t) are the empirical level, slope and curvature of the yield curve. In Table 3 and Fig. 8 (left column) we present descriptive statistics for the estimated factors. From the autocorrelations of the three factors, we can see that the first factor is the most persistent, and that the second

[9] Although, as discussed earlier, we attempted to remove illiquid bonds, complete elimination is not possible.

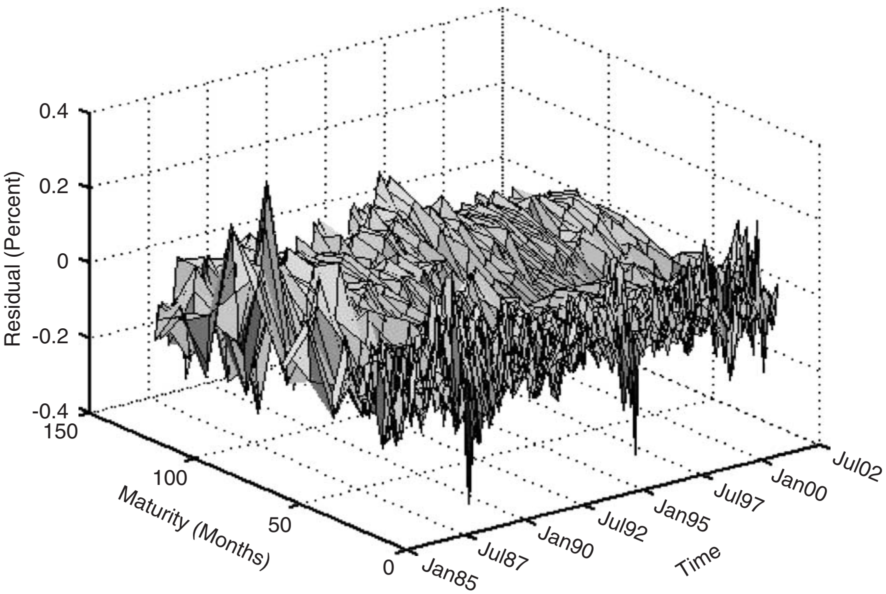

Fig. 6. Yield curve residuals, 1985.01–2000.12. We plot residuals from Nelson–Siegel yield curves fitted month-by-month. See text for details.

Table 2
Descriptive statistics, yield curve residuals

Maturity (Months)	Mean	Std. Dev.	Min.	Max.	MAE	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
3	−0.018	0.080	−0.332	0.156	0.061	0.082	0.777	0.157	−0.360
6	−0.013	0.042	−0.141	0.218	0.032	0.044	0.291	0.257	−0.046
9	−0.026	0.062	−0.200	0.218	0.052	0.067	0.704	0.216	−0.247
12	0.013	0.080	−0.160	0.267	0.064	0.081	0.563	0.322	−0.266
15	0.063	0.050	−0.063	0.243	0.067	0.080	0.650	0.139	−0.070
18	0.048	0.035	−0.048	0.165	0.052	0.059	0.496	0.183	−0.139
21	0.026	0.030	−0.091	0.101	0.033	0.040	0.370	−0.044	−0.011
24	−0.027	0.045	−0.190	0.082	0.037	0.052	0.667	0.212	0.056
30	−0.020	0.036	−0.200	0.098	0.029	0.041	0.398	0.072	−0.058
36	−0.037	0.046	−0.203	0.128	0.047	0.059	0.597	0.053	−0.017
48	−0.018	0.065	−0.204	0.230	0.052	0.067	0.754	0.239	−0.321
60	−0.053	0.058	−0.199	0.186	0.066	0.079	0.758	−0.021	−0.175
72	0.010	0.080	−0.133	0.399	0.056	0.081	0.904	0.278	−0.163
84	0.001	0.062	−0.259	0.263	0.044	0.062	0.589	0.019	0.000
96	0.032	0.045	−0.202	0.111	0.045	0.055	0.697	0.120	−0.144
108	0.033	0.046	−0.161	0.132	0.047	0.057	0.669	0.081	−0.176
120	−0.016	0.071	−0.256	0.164	0.057	0.073	0.623	0.252	−0.070

Note: We fit the three-factor model,

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-\mathrm{e}^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3t}\left(\frac{1-\mathrm{e}^{-\lambda_t\tau}}{\lambda_t\tau} - \mathrm{e}^{-\lambda_t\tau}\right),$$

using monthly yield data 1985:01–2000:12, with λ_t fixed at 0.0609, and we present descriptive statistics for the corresponding residuals at various maturities. The last three columns contain residual sample autocorrelations at displacements of 1, 12, and 30 months.

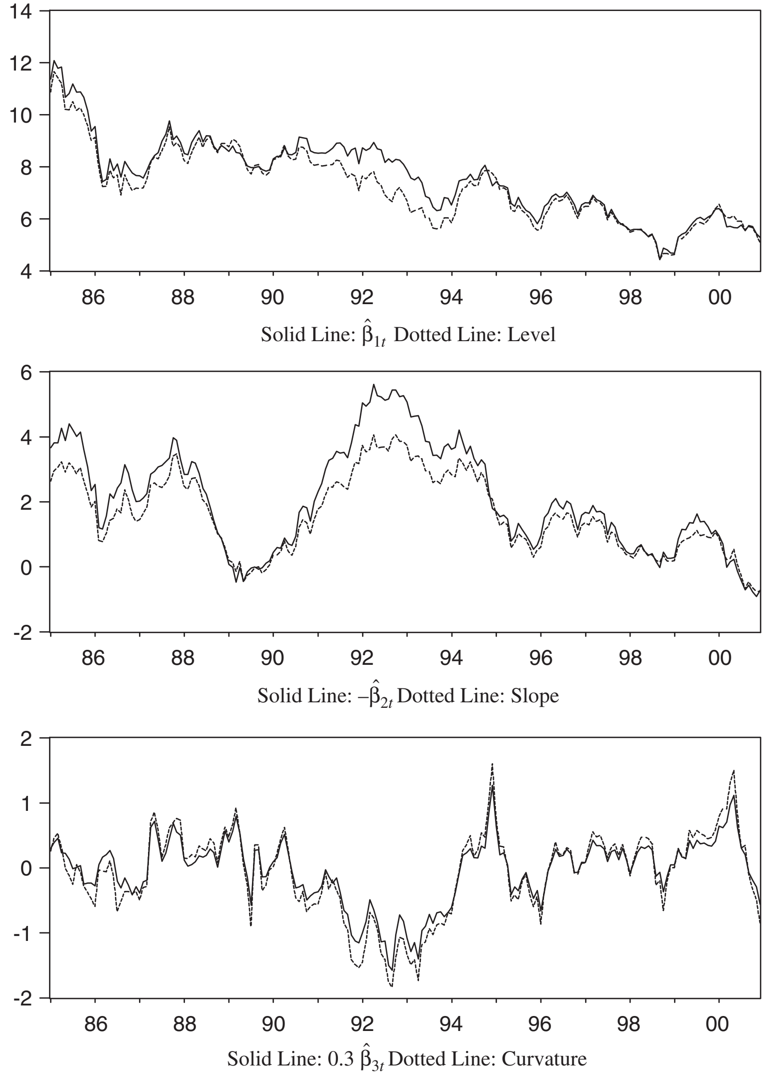

Fig. 7. Model-based level, slope and curvature (i.e., estimated factors) vs. data-based level, slope and curvature. We define the level as the 10-year yield, the slope as the difference between the 10-year and 3-month yields, and the curvature as the twice the 2-year yield minus the sum of the 3-month and 10-year yields.

factor is more persistent than the third. Augmented Dickey–Fuller tests suggest that $\hat{\beta}_1$ and $\hat{\beta}_2$ may have a unit roots, and that $\hat{\beta}_3$ does not.[10]Finally, the pairwise correlations between the estimated factors are not large.

[10]We use SIC to choose the lags in the augmented Dickey–Fuller unit-root test. The MacKinnon critical values for rejection of hypothesis of a unit root are −3.4518 at the one percent level, −2.8704 at the five percent level, and −2.5714 at the ten percent level.

Table 3
Descriptive statistics, estimated factors

Factor	Mean	Std. Dev.	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$	ADF
$\hat{\beta}_{1t}$	7.579	1.524	4.427	12.088	0.957	0.511	0.454	−2.410
$\hat{\beta}_{2t}$	−2.098	1.608	−5.616	0.919	0.969	0.452	−0.082	−1.205
$\hat{\beta}_{3t}$	−0.162	1.687	−5.249	4.234	0.901	0.353	−0.006	−3.516

Note: We fit the three-factor Nelson–Siegel model using monthly yield data 1985:01–2000:12, with λ_t fixed at 0.0609, and we present descriptive statistics for the three estimated factors $\hat{\beta}_{1t}$, $\hat{\beta}_{2t}$, and $\hat{\beta}_{3t}$. The last column contains augmented Dickey–Fuller (ADF) unit root test statistics, and the three columns to its left contain sample autocorrelations at displacements of 1, 12, and 30 months.

3.3. Modeling and forecasting yield curve level, slope and curvature

We model and forecast the Nelson–Siegel factors as univariate AR(1) processes. The AR(1) models can be viewed as natural benchmarks determined a priori: the simplest great workhorse autoregressive models. The yield forecasts based on underlying univariate AR(1) factor specifications are:

$$\hat{y}_{t+h/t}(\tau) = \hat{\beta}_{1,t+h/t} + \hat{\beta}_{2,t+h/t}\left(\frac{1-\mathrm{e}^{-\lambda\tau}}{\lambda\tau}\right) + \hat{\beta}_{3,t+h/t}\left(\frac{1-\mathrm{e}^{-\lambda\tau}}{\lambda\tau} - \mathrm{e}^{-\lambda\tau}\right),$$

where

$$\hat{\beta}_{1,t+h/t} = \hat{c}_i + \hat{\gamma}_i\hat{\beta}_{it}, \quad i = 1, 2, 3,$$

and $\hat{c}_i$ and $\hat{\gamma}_i$ are obtained by regressing $\hat{\beta}_{it}$ on an intercept and $\hat{\beta}_{i,t-h}$.[11]

For comparison, we also produce yield forecasts based on an underlying multivariate VAR(1) specification, as

$$\hat{y}_{t+h/t}(\tau) = \hat{\beta}_{1,t+h/t} + \hat{\beta}_{2,t+h/t}\left(\frac{1-\mathrm{e}^{-\lambda\tau}}{\lambda\tau}\right) + \hat{\beta}_{3,t+h/t}\left(\frac{1-\mathrm{e}^{-\lambda\tau}}{\lambda\tau} - \mathrm{e}^{-\lambda\tau}\right),$$

where

$$\hat{\beta}_{t+h/t} = \hat{c} + \hat{\Gamma}\hat{\beta}_t.$$

We include the VAR forecasts for completeness, although one might expect them to be inferior to the AR forecasts for at least two reasons. First, as is well-known from the macroeconomics literature, unrestricted VARs tend to produce poor forecasts of economic variables even when there is important cross-variable interaction, due to the large number of included parameters and the resulting potential for in-sample

[11]Note that we directly regress factors at $t+h$ on factors at t, which is a standard method of coaxing least squares into optimizing the relevant loss function, h-month-ahead RMSE, as opposed to the usual 1-month-ahead RMSE. We estimate all competitor models in the same way, as described below.

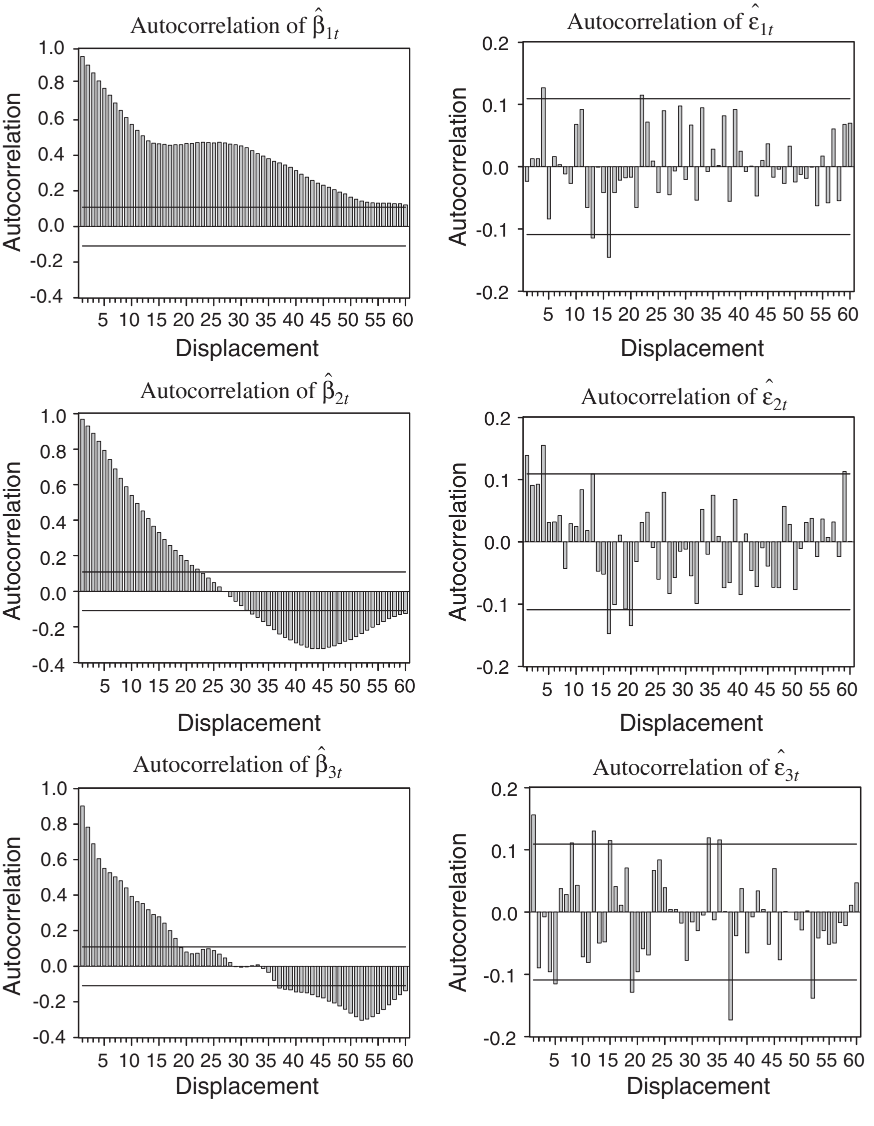

Fig. 8. Autocorrelations and residual autocorrelations of level, slope and curvature factors. We plot the sample autocorrelations of the three estimated factors, $\hat{\beta}_{1t}$, $\hat{\beta}_{2t}$, and $\hat{\beta}_{3t}$, as well as the sample autocorrelations of AR(1) models fit to the three estimated factors, along with Barlett's approximate 95% confidence bands.

overfitting.[12] Second, our factors indeed display little cross-factor interaction and are not highly correlated, so that an appropriate multivariate model is likely close to a stacked set of univariate models.

[12]That, of course, is the reason for the ubiquitous use of Bayesian analysis, featuring strong priors on the VAR coefficients, for VAR forecasting, as pioneered by Doan et al. (1984).

In Fig. 8 (right column) we provide some evidence on the goodness of fit of the AR(1) models fit to the estimated level, slope and curvature factors, showing residual autocorrelation functions. The autocorrelations are very small, indicating that the models accurately describe the conditional means of level, slope and curvature.

3.4. Out-of-sample forecasting performance of the three-factor model

A good approximation to yield-curve dynamics should not only fit well in-sample, but also forecast well out-of-sample. Because the yield curve depends only on $\{\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}\}$, forecasting the yield curve is equivalent to forecasting $\{\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}\}$. In this section we undertake just such a forecasting exercise. We estimate and forecast recursively, using data from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12.

In Tables 4–6 we compare h-month-ahead out-of sample forecasting results from Nelson–Siegel models to those of several natural competitors, for maturities of 3, 12,

Table 4
Out-of-sample 1-month-ahead forecasting results

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$
Nelson–Siegel with AR(1) factor dynamics					
3 months	−0.045	0.170	0.176	0.247	0.017
1 year	0.023	0.235	0.236	0.425	−0.213
3 years	−0.056	0.273	0.279	0.332	−0.117
5 years	−0.091	0.277	0.292	0.333	−0.116
10 years	−0.062	0.252	0.260	0.259	−0.115
Random walk					
3 months	0.033	0.176	0.179	0.220	0.053
1 year	0.021	0.240	0.241	0.340	−0.153
3 years	0.007	0.279	0.279	0.341	−0.133
5 years	−0.003	0.276	0.276	0.275	−0.131
10 years	−0.011	0.254	0.254	0.215	−0.145
Slope regression					
3 months	NA	NA	NA	NA	NA
1 year	0.048	0.242	0.247	0.328	−0.145
3 years	0.032	0.286	0.288	0.373	−0.146
5 years	0.019	0.284	0.285	0.318	−0.150
10 years	0.013	0.260	0.260	0.245	−0.159
Fama–Bliss forward rate regression					
3 months	0.066	0.159	0.172	0.178	0.036
1 year	0.066	0.233	0.242	0.313	−0.148
3 years	0.024	0.286	0.287	0.380	−0.157
5 years	0.038	0.277	0.280	0.273	−0.125
10 years	0.041	0.251	0.254	0.200	−0.159

Table 4 (*continued*)

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$
Cochrane–Piazzesi forward curve regression					
3 months	NA	NA	NA	NA	NA
1 year	−0.038	0.238	0.241	0.282	−0.088
3 years	−0.034	0.287	0.289	0.377	−0.108
5 years	−0.068	0.292	0.300	0.364	−0.084
10 years	−0.113	0.257	0.281	0.271	−0.097
Univariate AR(1)s on yield levels					
3 months	0.042	0.177	0.182	0.229	0.060
1 year	0.025	0.238	0.239	0.341	−0.147
3 years	−0.005	0.276	0.276	0.345	−0.125
5 years	−0.030	0.274	0.276	0.280	−0.127
10 years	−0.054	0.252	0.258	0.224	−0.144
VAR(1) on yield levels					
3 months	−0.013	0.176	0.176	0.229	0.128
1 year	−0.026	0.262	0.263	0.447	−0.162
3 years	−0.041	0.302	0.305	0.437	−0.154
5 years	−0.064	0.303	0.310	0.429	−0.133
10 years	−0.090	0.274	0.288	0.310	−0.123
VAR(1) on yield changes					
3 months	0.043	0.176	0.181	−0.019	0.156
1 year	0.029	0.230	0.232	0.157	−0.149
3 years	0.026	0.276	0.277	0.077	−0.049
5 years	0.021	0.276	0.277	0.010	−0.002
10 years	0.020	0.263	0.264	−0.017	−0.030

Note: We present the results of out-of-sample 1-month-ahead forecasting using eight models, as described in detail in the text. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors at $t+1$ as $y_{t+1}(\tau) - \hat{y}_{t+1/t}(\tau)$, and we report the mean, standard deviation and root mean squared errors of the forecast errors, as well as their first and 12th sample autocorrelation coefficients.

Table 5
Out-of-sample 6-month-ahead forecasting results

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}$ (6)	$\hat{\rho}$ (18)
Nelson–Siegel with AR(1) factor dynamics					
3 months	0.083	0.510	0.517	0.301	−0.190
1 year	0.131	0.656	0.669	0.168	−0.174
3 years	−0.052	0.748	0.750	0.049	−0.189
5 years	−0.173	0.758	0.777	0.069	−0.273
10 years	−0.251	0.676	0.721	0.058	−0.288
Random walk					
3 months	0.220	0.564	0.605	0.381	−0.214
1 year	0.181	0.758	0.779	0.139	−0.150

Table 5 (*continued*)

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}$ (6)	$\hat{\rho}$ (18)
3 years	0.099	0.873	0.879	0.018	−0.211
5 years	0.048	0.860	0.861	0.008	−0.249
10 years	−0.020	0.758	0.758	0.019	−0.271
Slope regression					
3 months	NA	NA	NA	NA	NA
1 year	0.422	0.811	0.914	0.109	−0.113
3 years	0.281	0.944	0.985	0.116	−0.198
5 years	0.209	0.939	0.962	0.103	−0.235
10 years	0.145	0.832	0.845	0.096	−0.256
Fama–Bliss forward rate regression					
3 months	0.494	0.549	0.739	0.208	−0.072
1 year	0.373	0.821	0.902	0.194	−0.150
3 years	0.255	0.964	0.997	0.092	−0.211
5 years	0.220	0.932	0.958	0.050	−0.248
10 years	0.223	0.794	0.825	0.038	−0.268
Cochrane–Piazzesi forward curve regression					
3 months	NA	NA	NA	NA	NA
1 year	−0.155	0.845	0.859	0.220	−0.110
3 years	−0.210	0.910	0.934	0.179	−0.218
5 years	−0.224	0.910	0.937	0.193	−0.270
10 years	−0.345	0.837	0.905	0.192	−0.287
Univariate AR(1)s on yield levels					
3 months	0.224	0.539	0.584	0.405	−0.210
1 year	0.160	0.707	0.725	0.193	−0.155
3 years	−0.030	0.800	0.801	0.075	−0.211
5 years	−0.144	0.789	0.802	0.061	−0.253
10 years	−0.286	0.699	0.755	0.073	−0.278
VAR(1) on yield levels					
3 months	−0.138	0.659	0.673	0.289	−0.160
1 year	−0.195	0.880	0.901	0.133	−0.169
3 years	−0.218	0.926	0.951	0.122	−0.240
5 years	−0.258	0.919	0.955	0.140	−0.273
10 years	−0.406	0.811	0.907	0.137	−0.293
VAR(1) on yield changes					
3 months	0.312	0.661	0.731	0.319	−0.256
1 year	0.310	0.845	0.900	0.172	−0.181
3 years	0.276	0.941	0.981	0.059	−0.210
5 years	0.246	0.917	0.949	0.048	−0.242
10 years	0.192	0.809	0.831	0.043	−0.259

Note: We present the results of out-of-sample 6-month-ahead forecasting using eight models, as described in detail in the text. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors at $t+6$ as $y_{t+6}(\tau) - \hat{y}_{t+6/t}(\tau)$, and we report the mean, standard deviation and root mean squared errors of the forecast errors, as well as their sixth and eighteenth sample autocorrelation coefficients.

Table 6
Out-of-sample 12-month-ahead forecasting results

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}$ (12)	$\hat{\rho}$ (24)
Nelson–Siegel with AR(1) factor dynamics					
3 months	0.150	0.724	0.739	−0.288	0.001
1 year	0.173	0.823	0.841	−0.332	−0.004
3 years	−0.123	0.910	0.918	−0.408	0.015
5 years	−0.337	0.918	0.978	−0.412	0.003
10 years	−0.531	0.825	0.981	−0.433	−0.003
Nelson–Siegel with VAR(1) factor dynamics					
3 months	−0.463	1.000	1.102	−0.163	−0.111
1 year	−0.416	1.224	1.293	−0.265	−0.065
3 years	−0.576	1.268	1.393	−0.317	−0.036
5 years	−0.673	1.210	1.385	−0.315	−0.039
10 years	−0.721	1.056	1.279	−0.299	−0.037
Random walk					
3 months	0.416	0.930	1.019	−0.118	−0.109
1 year	0.388	1.132	1.197	−0.268	−0.019
3 years	0.236	1.214	1.237	−0.419	0.060
5 years	0.130	1.184	1.191	−0.481	0.072
10 years	−0.033	1.051	1.052	−0.508	0.069
Slope regression					
3 months	NA	NA	NA	NA	NA
1 year	0.896	1.235	1.526	−0.187	−0.024
3 years	0.641	1.316	1.464	−0.212	0.024
5 years	0.515	1.305	1.403	−0.255	0.035
10 years	0.362	1.208	1.261	−0.268	0.042
Fama–Bliss forward rate regression					
3 months	0.942	1.010	1.381	−0.046	−0.096
1 year	0.875	1.276	1.547	−0.142	−0.039
3 years	0.746	1.378	1.567	−0.291	0.035
5 years	0.587	1.363	1.484	−0.352	0.040
10 years	0.547	1.198	1.317	−0.403	0.062
Cochrane–Piazzesi forward curve regression					
3 months	NA	NA	NA	NA	NA
1 year	−0.162	1.275	1.285	−0.179	−0.079
3 years	−0.377	1.275	1.330	−0.274	−0.028
5 years	−0.529	1.225	1.334	−0.301	−0.021
10 years	−0.760	1.088	1.327	−0.307	−0.020
Univariate AR(1)s on yield levels					
3 months	0.246	0.808	0.845	−0.213	−0.073
1 year	0.182	0.953	0.970	−0.271	−0.004

Table 6 (*continued*)

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}$ (12)	$\hat{\rho}$ (24)
3 years	−0.113	0.996	1.002	−0.380	0.061
5 years	−0.301	0.961	1.007	−0.433	0.058
10 years	−0.603	0.835	1.030	−0.431	0.020
VAR(1) on yield levels					
3 months	−0.276	1.006	1.043	−0.219	−0.099
1 year	−0.390	1.204	1.266	−0.322	−0.058
3 years	−0.467	1.240	1.325	−0.345	−0.015
5 years	−0.540	1.201	1.317	−0.348	−0.012
10 years	−0.744	1.060	1.295	−0.328	−0.010
VAR(1) on yield changes					
3 months	0.717	1.072	1.290	−0.068	−0.127
1 year	0.704	1.240	1.426	−0.223	−0.041
3 years	0.627	1.341	1.480	−0.399	0.051
5 years	0.559	1.281	1.398	−0.459	0.070
10 years	0.408	1.136	1.207	−0.491	0.072
ECM(1) with one common trend					
3 months	0.738	0.982	1.228	−0.163	−0.123
1 year	0.767	1.143	1.376	−0.239	−0.072
3 years	0.546	1.203	1.321	−0.278	−0.013
5 years	0.379	1.191	1.250	−0.278	−0.003
10 years	0.169	1.095	1.108	−0.224	0.009
ECM(1) with two common trends					
3 months	0.778	1.037	1.296	−0.175	−0.129
1 year	0.868	1.247	1.519	−0.286	−0.033
3 years	0.586	1.186	1.323	−0.288	−0.034
5 years	0.425	1.155	1.231	−0.304	−0.014
10 years	0.220	1.035	1.058	−0.274	0.015
Direct regression on three AR(1) principal components					
3 months	0.162	0.785	0.802	−0.298	−0.020
1 year	0.416	0.979	1.064	−0.305	0.042
3 years	−0.127	1.014	1.022	−0.372	0.054
5 years	−0.393	1.013	1.087	−0.335	0.038
10 years	−0.394	0.929	1.009	−0.284	0.066

Note: We present the results of out-of-sample 12-month-ahead forecasting using twelve models, as described in detail in the text. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors at $t+12$ as $y_{t+12}(\tau) - \hat{y}_{t+12/t}(\tau)$, and we report the mean, standard deviation and root mean squared errors of the forecast errors, as well as their 12th and 24th sample autocorrelation coefficients.

36, 60 and 120 months, and forecast horizons of $h = 1, 6$ and 12 months. Let us now describe the competitors in terms of how their forecasts are generated.

(1) Random walk:

$$\hat{y}_{t+h/t}(\tau) = y_t(\tau).$$

The forecast is always "no change."

(2) Slope regression:

$$\hat{y}_{t+h/t}(\tau) - y_t(\tau) = \hat{c}(\tau) + \hat{\gamma}(\tau)(y_t(\tau) - y_t(3)).$$

The forecasted yield change is obtained from a regression of historical yield changes on yield curve slopes.

(3) Fama–Bliss forward rate regression:

$$\hat{y}_{t+h/t}(\tau) - y_t(\tau) = \hat{c}(\tau) + \hat{\gamma}(\tau)(f_t^h(\tau) - y_t(\tau)),$$

where $f_t^h(\tau)$ is the forward rate contracted at time t for loans from time $t+h$ to time $t+h+\tau$. Hence the forecasted yield change is obtained from a regression of historical yield changes on forward spreads. Note that, because the forward rate is proportional to the derivative of the discount function, the information used to forecast future yields in forward rate regressions is very similar to that in slope regressions.

(4) Cochrane and Piazzesi (2002) forward curve regression:

$$\hat{y}_{t+h/t}(\tau) - y_t(\tau) = \hat{c}(\tau) + \hat{\gamma}_0(\tau)y_t(12) + \sum_{k=1}^{9} \hat{\gamma}_k(\tau)f_t^{12k}(12).$$

Note that the Fama–Bliss forward regression is a special case of the Cochrane–Piazzesi forward regression.[13]

(5) AR(1) on yield levels:

$$\hat{y}_{t+h/t}(\tau) = \hat{c}(\tau) + \hat{\gamma}y_t(\tau).$$

(6) VAR(l) on yield levels:

$$\hat{y}_{t+h/t} = \hat{c} + \hat{\Gamma}y_t,$$

where $y_t = [y_t(3), y_t(12), y_t(36), y_t(60), y_t(120)]'$.

(7) VAR(l) on yield changes:

$$\hat{z}_{t+h/t} = \hat{c} + \hat{\Gamma}z_t,$$

where $z_t \equiv [y_t(3) - y_{t-1}(3), y_t(12) - y_{t-1}(12), y_t(36) - y_{t-1}(36), y_t(60) - y_{t-1}(60), y_t(120) - y_{t-1}(120)]'$.

[13] Note that this is an unrestricted version of the model estimated by Cochrane and Piazzesi. Imposition of the Cochrane–Piazzesi restrictions produced qualitatively identical results.

(8) ECM(1) with one common trend:

$$\hat{z}_{t+h/t} = \hat{c} + \hat{\Gamma} z_t,$$

where $z_t \equiv [y_t(3) - y_{t-1}(3), y_t(12) - y_t(3), y_t(36) - y_t(3), y_t(60) - y_t(3), y_t(120) - y_t(3)]'$.

(9) ECM(1) with two common trends:

$$\hat{z}_{t+h/t} = \hat{c} + \hat{\Gamma} z_t,$$

where $z_t \equiv [y_t(3) - y_{t-1}(3), y_t(12) - y_{t-1}(12), y_t(36) - y_t(3), y_t(60) - y_t(3), y_t(120) - y_t(3)]'$.

(10) Direct regression on three AR(1) principal components
We first perform a principal components analysis on the full set of seventeen yields y_t, effectively decomposing the yield covariance matrix as $Q\Lambda Q^{\mathrm{T}}$, where the diagonal elements of Λ are the eigenvalues and the columns of Q are the associated eigenvectors. Denote the largest three eigenvalues by λ_1, λ_2, and λ_3, and denote the associated eigenvectors by q_1, q_2, and q_3. The first three principal components $x_t = [x_{1t}, x_{2t}, x_{3t}]$ are then defined by $x_{it} = q_i' y_t$, $i = 1, 2, 3$. We then use a univariate AR(1) model to produce h-step-ahead forecasts of the principal components:

$$\hat{x}_{i,t+h/t} = \hat{c}_i + \hat{\gamma}_i x_{it}, \quad i = 1, 2, 3,$$

and we produce forecasts for yields $y_t \equiv [y_t(3), y_t(12), y_t(36), y_t(60), y_t(120)]'$ as

$$\hat{y}_{t+h/t}(\tau) = q_1(\tau)\hat{x}_{1,t+h/t} + q_2(\tau)\hat{x}_{2,t+h/t} + q_3(\tau)\hat{x}_{3,t+h/t},$$

where $q_i(\tau)$ is the element in the eigenvector q_i that corresponds to maturity τ.

We define forecast errors at $t + h$ as $y_{t+h}(\tau) - \hat{y}_{t+h/t}(\tau)$. Note well that, in each case, the object being forecast ($y_{t+h}(\tau)$) is a future yield, not a future Nelson–Siegel fitted yield. We will examine a number of descriptive statistics for the forecast errors, including mean, standard deviation, root mean squared error (RMSE), and autocorrelations at various displacements.

Our model's 1-month-ahead forecasting results, reported in Table 4, are in certain respects humbling. In absolute terms, the forecasts appear suboptimal: the forecast errors appear serially correlated. In relative terms, RMSE comparison at various maturities reveals that our forecasts, although slightly better than the random walk and slope regression forecasts, are indeed only very slightly better. Finally, the Diebold and Mariano (1995) statistics reported in Table 7 indicate universal insignificance of the RMSE differences between our 1-month-ahead forecasts and those from random walks or Fama–Bliss regressions.

The 1-month-ahead forecast defects likely come from a variety of sources, some of which could be eliminated. First, for example, pricing errors due to illiquidity may be highly persistent and could be reduced by including variables that may explain mispricing. It is worth noting, moreover, that related papers such as Bliss (1997b) and de Jong (2000) also find serially correlated forecast errors, often with persistence much stronger than ours.

Table 7
Out-of-sample forecast accuracy comparisons

Maturity (τ)	1-Month horizon		12-Month horizon	
	Against RW	Against FB	Against RW	Against FB
3 months	−0.27	0.18	−1.65*	−2.43*
1 year	−0.64	−0.56	−2.04*	−2.31*
3 years	−0.02	−0.58	−2.11*	−2.18*
5 years	0.97	0.57	−1.61	−1.90*
10 years	0.49	0.34	−0.63	−1.35

Note: We present Diebold–Mariano forecast accuracy comparison tests of our three-factor model forecasts (using univariate AR(1) factor dynamics) against those of the random walk model (RW) and the Fama–Bliss forward rate regression model (FB). The null hypothesis is that the two forecasts have the same mean squared error. Negative values indicate superiority of our three-factor model forecasts, and asterisks denote significance relative to the asymptotic null distribution at the 10 percent level.

Matters improve radically, however, as the forecast horizon lengthens. Our model's 6-month-ahead forecasting results, reported in Table 5, are noticeably improved, and our model's 12-month-ahead forecasting results, reported in Table 6, are much improved. In particular, our model's 12-month ahead forecasts outperform those of all competitors at all maturities, often by a wide margin in both relative and absolute terms. Seven of the 10 Diebold–Mariano statistics in Table 7 indicate significant 12-month-ahead RMSE superiority of our forecasts at the five percent level. The strong yield curve forecastability at the 12-month-ahead horizon is, for example, very attractive from the vantage point of active bond trading and the vantage point of credit portfolio risk management.[14] Moreover, our 12-month-ahead forecasts, like their 1- and 6-month-ahead counterparts, could be improved upon, because the forecast errors remain serially correlated.[15]

It is worth noting that Duffee (2002) finds that even the simplest random walk forecasts dominate those from the Dai and Singleton (2000) affine model, which therefore appears largely useless for forecasting. Hence Duffee proposes a less-restrictive "essentially affine" model and shows that it forecasts better than the random walk in most cases, which is appropriately viewed as a victory. A comparison of our results and Duffee's, however, reveals that our three-factor model

[14]Note that Nelson–Siegel loadings imply a very smooth yield curve, which in turn suggests that our model, although not arbitrage-free, would not likely generate extreme portfolio positions. Competitors such as regression on principal components, in contrast, have no smooth cross-sectional restrictions and may well generate extreme portfolio positions in practice. This is one important way in which our approach is superior to directs regression on principal components, despite the fact that our estimated factors are close to the first three principal components. (Four more are given below.)

[15]We report 12-month-ahead forecast error serial correlation coefficients at displacements of 12 and 24 months, in contrast to those at displacements of 1 and 12 months reported for the 1-month-ahead forecast errors, because the 12-month-ahead errors would naturally have moving-average structure even if the forecasts were fully optimal, due to the overlap.

produces larger percentage reductions in out-of-sample RMSE relative to the random walk than does Duffee's best essentially affine model. Our forecasting success is particularly notable in light of the fact that Duffee forecasts only the smoothed yield curve, whereas we forecast the actual yield curve.[16]

Finally, we note that although our approach is closely related to direct principal components regression, neither our approach nor our results are identical. Interestingly, there is reason to prefer our approach on both empirical and theoretical grounds. Empirically, our results indicate that our approach has superior forecasting performance on our sample of yields. Theoretically, other methods, including regression on principal components and regression on ad hoc empirical level, slope and curvature, often have unappealing features, including:

(1) they cannot be used to produce yields at maturities other than those observed in the data,
(2) they do not guarantee a smooth yield curve and forward curve,
(3) they do not guarantee positive forward rates at all horizons, and
(4) they do not guarantee that the discount function starts at 1 and approaches 0 as maturity approaches infinity.

4. Concluding remarks

We have re-interpreted the Nelson–Siegel yield curve as a dynamic model that achieves dimensionality reduction via factor structure (level, slope and curvature), and we have explored the model's performance in out-of-sample yield curve forecasting. Although the 1-month-ahead forecasting results are no better than those of random walk and other leading competitors, the 1-year-ahead results are much superior.

A number of authors have proposed extensions to Nelson–Siegel to enhance flexibility, including Bliss (1997b), Soderlind and Svensson (1997), Björk and Christensen (1999), Filipovic (1999, 2000), Björk (2000), Björk and Landén (2000) and Björk and Svensson (2001). From the perspective of interest rate forecasting accuracy, however, the desirability of the above generalizations of Nelson–Siegel is not obvious, which is why we did not pursue them here. For example, although the Bliss and Soderlind–Svensson extensions can have in-sample fit no worse than that of Nelson–Siegel, because they include Nelson–Siegel as a special case, there is no guarantee of better out-of-sample forecasting performance. Indeed, accumulated experience suggest that parsimonious models are often more successful for out-of-sample forecasting.[17]

Some of the extensions alluded to above are designed to make Nelson–Siegel consistent with no-arbitrage pricing. It is not obvious to us, however, that use of

[16]We note, however, that our enthusiasm must be tempered by the fact that our in-sample and out-of-sample periods are not identical to Duffee's, so definitive comparisons cannot be made.

[17]See Diebold (2004).

arbitrage-free models is necessary or desirable for producing good forecasts.[18] Indeed we have shown that our model, which is not arbitrage-free, can produces good forecasts.

In closing, we would like to elaborate on the likely reason for the forecasting success of our approach, which relies heavily on a broad interpretation of the shrinkage principle. The essence of our approach is intentionally to impose substantial a priori structure, motivated by simplicity, parsimony, and theory, in an explicit attempt to avoid data mining and hence enhance out-of-sample forecasting ability. This includes our use of a tightly parametric model that places strict structure on factor loadings in accordance with simple theoretical desiderata for the discount function, our decision to fix λ, our emphasis on simple univariate modeling of the factors based upon our theoretically derived interpretation of the model as one of approximately orthogonal level, slope and curvature factors, and our emphasis on the simplest possible AR(1) factor dynamics. All of this is in keeping with a broad interpretation of the "shrinkage principle," which has a firm foundation in Bayes–Stein theory, in empirical intuition, and in an accumulated track record of good performance (e.g., Garcia-Ferrer et al., 1987; Zellner and Hong, 1989; Zellner and Min, 1993). Here we interpret the shrinkage principle as the insight that imposition of restrictions, which will of course degrade in-sample fit, may nevertheless be helpful for out-of-sample forecasting, even if the restrictions are false. The fact that the shrinkage principle works in the yield-curve context, as it does in so many other contexts, is precisely what theory and empirical experience would lead one to expect. This is not to say, of course, that our specification is in any sense uniquely best, and we make no claims to that effect. Rather, the broad lesson of the paper is to show in the yield-curve context that the shrinkage perspective, which tends to produce seemingly naive but truly sophisticatedly simple models (of which ours is one example), may be very appealing when the goal is forecasting. Put differently, the paper emphasizes in the yield curve context Zellner's (1992) "KISS principle" of forecasting —"Keep it sophisticatedly simple."

Acknowledgements

The National Science Foundation, the Wharton Financial Institutions Center, and the Guggenheim Foundation provided research support. For helpful comments we are grateful to the Editor (Arnold Zellner), the Associate Editor, and three referees, as well as Dave Backus, Rob Bliss, Michael Brandt, Todd Clark, Qiang Dai, Ron Gallant, Mike Gibbons, David Marshall, Monika Piazzesi, Eric Renault, Glenn Rudebusch, Til Schuermann, and Stan Zin, and seminar participants at Geneva, Georgetown, Wharton, the European Central Bank, and the National Bureau of Economic Research. We, however, bear full responsibility for all remaining flaws.

[18]See Dai and Singleton (2002) for an interesting analysis that explores certain aspects of the tradeoff between freedom from arbitrage and forecasting performance.

References

Andersen, T.G., Lund, J., 1997. Stochastic volatility and mean drift in the short term interest rate diffusion: source of steepness, level and curvature in the yield curve. Working Paper 214, Department of Finance, Kellogg School, Northwestern University.

Balduzzi, P., Das, S.R., Foresi, S., Sundaram, R., 1996. A simple approach to three-factor affine term structure models. Journal of Fixed Income 6, 43–53.

Bates, D., 1999. Financial markets' assessment of EMU. Carnegie–Rochester Conference Series on Public Policy 51, 229–269.

Björk, T., 2000. A geometric view of interest rate theory. Handbook of Mathematical Finance. Cambridge University Press, Cambridge forthcoming.

Björk, T., Christensen, B., 1999. Interest rate dynamics and consistent forward rate curves. Mathematical Finance 9, 323–348.

Björk, T., Landén, C., 2000. On the construction of finite dimensional realizations for nonlinear forward rate models. Manuscript, Stockholm School of Economics.

Björk, T., Svensson, L., 2001. On the existence of finite dimensional realizations for nonlinear forward rate models. Mathematical Finance 11, 205–243.

Bliss, R., 1997a. Movements in the term structure of interest rates. Economic Review Federal Reserve Bank of Atlanta 82, 16–33.

Bliss, R., 1997b. Testing term structure estimation methods. Advances in Futures and Options Research 9, 97–231.

Brandt, M.W., Yaron, A., 2001. Time-consistent no-arbitrage models of the term structure. Manuscript, University of Pennsylvania.

Chen, L., 1996. Stochastic Mean and Stochastic Volatility—A Three Factor Model of the Term Structure of Interest Rates and its Application to the Pricing of Interest Rate Derivatives. Blackwell Publishers, London.

Cochrane, J.H., Piazzesi, M., 2002. Bond risk premia. Manuscript, University of Chicago and UCLA.

Courant, E., Hilbert, D., 1953. Methods of Mathematical Physics. John Wiley, New York.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407.

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

Dai, Q., Singleton, K.J., 2002. Expectation puzzles, time-varying risk premia, and affine models of the term structure. Journal of Financial Economics 63, 415–441.

de Jong, F., 2000. Time series and cross section information in affine term structure models. Journal of Business and Economic Statistics 18, 300–314.

de Jong, F., Santa-Clara, P., 1999. The dynamics of the forward interest rate curve: a formulation with state variables. Journal of Financial and Quantitative Analysis 31, 131–157.

Diebold, F.X., 2004. Elements of Forecasting, third ed. South-Western, Cincinnati.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Diebold, F.X., Sharpe, S., 1990. Post-deregulation bank deposit rate pricing: the multivariate dynamics. Journal of Business and Economic Statistics 8, 281–293.

Diebold, F.X., Gunther, T., Tay, A.S., 1998. Evaluating density forecasts, with applications to financial risk management. International Economic Review 39, 863–883.

Diebold, F.X., Hahn, J., Tay, A.S., 1999. Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange. Review of Economics and Statistics 81, 661–673.

Doan, T., Litterman, R., Sims, C., 1984. Forecasting and conditional projection using realistic prior distributions. Econometric Reviews 3, 1–100.

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

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

Fama, E., Bliss, R., 1987. The information in long-maturity forward rates. American Economic Review 77, 680–692.

Filipovic, D., 1999. A note on the Nelson–Siegel family. Mathematical Finance 9, 349–359.

Filipovic, D., 2000. Exponential-polynomial families and the term structure of interest rates. Bernoulli 6, 1–27.

Frankel, J.A., Lown, C.S., 1994. An indicator of future inflation extracted from the steepness of the interest rate yield curve along its entire length. Quarterly Journal of Economics 109, 517–530.

Garcia-Ferrer, A., Highfield, R.A., Palm, F., Zellner, A., 1987. Macroeconomic forecasting using pooled international data. Journal of Business and Economic Statistics 5, 53–67.

Hall, A.D., Anderson, H.M., Granger, C.W.J., 1992. A cointegration analysis of treasury bill yields. Review of Economics and Statistics 74, 116–126.

Heath, D., Jarrow, R., Morton, A., 1992. Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation. Econometrica 60, 77–105.

Hull, J., White, A., 1990. Pricing interest-rate-derivative securities. Review of Financial Studies 3, 573–592.

Knez, P., Litterman, R., Scheinkman, J., 1994. Exploration into factors explaining money market returns. Journal of Finance 49, 1861–1882.

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

Litzenberger, R., Squassi, G., Weir, N., 1995. Spline models of the term structure of interest rates and their applications. Working Paper, Goldman, Sachs and Company.

McCulloch, J.H., 1975. The tax adjusted yield curve. Journal of Finance 30, 811–830.

McCulloch, J.H., Kwon, H., 1993. U.S. term structure data, 1947–1991. Working paper 93-6, Ohio State University.

Nelson, C.R., Siegel, A.F., 1987. Parsimonious modeling of yield curve. Journal of Business 60, 473–489.

Pagan, A.R., Hall, A.D., Martin, V., 1996. Modeling the term structure. In: Rao, C.R., Maddala, G.S. (Eds.), Handbook of Statistics. North-Holland, Amsterdam, pp. 91–118.

Shea, G.S., 1992. Benchmarking the expectations hypothesis of the interest-rate term structure: an analysis of cointegration vectors. Journal of Business and Economic Statistics 10, 347–366.

Siegel, A.F., Nelson, C.R., 1988. Long-term behavior of yield curves. Journal of Financial and Quantitative Analysis 23, 105–110.

Soderlind, P., Svensson, L.E.O., 1997. New techniques to extract market expectations from financial instruments. Journal of Monetary Economics 40, 383–430.

Swanson, N.R., White, H., 1995. A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. Journal of Business and Economic Statistics 13, 265–275.

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

Vasicek, O.A., Fong, H.G., 1982. Term structure modeling using exponential splines. Journal of Finance 37, 339–348.

Zellner, A., 1992. Statistics Science and Public Policy (ASA Presidential Address). Journal of the American Statistical Association 87, 1–6.

Zellner, A., Hong, C., 1989. Forecasting international growth rates using Bayesian shrinkage and other procedures. Journal of Econometrics 40, 183–202.

Zellner, A., Min, C., 1993. Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. Journal of Econometrics 56, 89–118.