Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 02-026

"Forecasting the Term Structure of Government Bond Yields"

by

Francis X. Diebold and Canlin Li

Forecasting the Term Structure of Government Bond Yields

Francis X. Diebold
University of Pennsylvania and NBER

Canlin Li
University of California, Riverside

First Draft, December 2000
This Draft/Print: August 6, 2002

Abstract: Despite powerful advances in yield curve modeling in the last twenty 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, which focuses on accurately fitting the cross section of interest rates at any given time but neglects time-series dynamics, nor the equilibrium approach, which focuses on time-series dynamics (primarily those of the instantaneous rate) but pays comparatively little attention to fitting the entire cross section at any given time and has been shown to forecast poorly. 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. Finally, we discuss a number of extensions, including generalized duration measures, applications to active bond portfolio management, and arbitrage-free specifications.

Acknowledgments: The National Science Foundation and the Wharton Financial Institutions Center provided research support. For helpful comments we are grateful to 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, as well as 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.

1. Introduction

The last twenty-five 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, Jarrow and Morton (1992), and prominent contributions in the affine equilibrium tradition include Vasicek (1977), Cox, Ingersoll and Ross (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), show that the models forecast miserably.

In this paper we take an explicit out-of-sample forecasting perspective, and we use a flexible modeling approach in hope of superior forecasting performance. We use neither the no-arbitrage approach nor the equilibrium approach. Instead, we use the Nelson-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

---

[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, Anderson, and Granger (1992), Shea (1992), Swanson and White (1995), and Pagan, Hall and Martin (1996).

[2] For comparative discussion of point and density forecasting, see Diebold, Gunther and Tay (1998) and Diebold, Hahn and Tay (1999).

factor loadings, we impose a particular functional form on the factor loadings. Doing so not only facilitates highly precise estimation of the factors, but also lets us interpret the estimated factors as level, slope and curvature. We propose and estimate autoregressive models for the factors, and we then forecast the yield curve by forecasting the factors. Our results are encouraging; in particular, our models produce one-year-ahead forecasts that are strikingly more accurate than standard benchmarks.

Closely related work includes the factor models of Litzenberger, Squassi and Weir (1995), Bliss (1997a, 1997b), 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, Das, Foresi and Sundaram (1996), Chen (1996), and especially the Andersen-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 multi-factor 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 conclude and discuss a number of variations and extensions that represent promising directions for future research, including state-space modeling and optimal filtering, generalized duration measures, applications to active bond portfolio management, and arbitrage-free specifications.

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-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 make it potentially consistent with various stylized empirical facts about 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.

Fitting 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,

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

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

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

or

$$f_t(\tau) = y_t(\tau) + \tau y_t^{/}(\tau),$$

which implies that the zero-coupon yield is an equally-weighed average of forward rates. Given the yield curve or forward rate 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 rate curves are not observed. Instead they are estimated from observed prices of bonds by interpolating for missing maturities and/or smoothing to reduce the impact of noise. One popular approach to yield curve fitting is due to McCulloch (1975) and McCulloch and Kwon (1993), who model the discount curve with a cubic spline, which can be conveniently estimated by least squares. 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 approach 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 forward rates to be positive.

A third approach to yield curve fitting is due to Fama and Bliss (1987), who develop an iterative method for piecewise-linear fitting of forward rate curves, sometimes called "unsmoothed Fama-Bliss." A natural extension, "smoothed Fama-Bliss," begins with the unsmoothed Fama-Bliss piecewise linear curve, and then smooths using the Nelson-Siegel (1987) model, which we discuss in detail below. Unsmoothed Fama-Bliss appears accurate and unrestrictive, but its lack of restrictions may be a vice rather than a virtue for forecasting, because it's not clear how to extrapolate a nonparametrically-fit curve, and even if it could be done it might lead to poor forecasts due to overfitting. Instead, we want to distill the entire term structure into just a few parameters. Smoothed Fama-Bliss effectively does so, but one may as well then go ahead and fit Nelson-Siegel to the raw term structure data, rather than to the Fama-Bliss term structure – which is precisely what we do.

This brings us to a fourth approach to yield curve fitting, which proves very useful for our purposes, due to Nelson and Siegel (1987). Nelson-Siegel is a three-component exponential approximation to the yield curve. It is parsimonious, easy to estimate by least squares, has a discount function that begins at one at zero horizon and approaches zero at infinite horizon, as appropriate, and it is from the class of functions that are solutions to differential or difference equations. Bliss (1997b) compares the different yield curve fitting methods and finds that the Nelson-Siegel approach performs admirably. We now proceed to examine the Nelson-Siegel approach in greater detail.

<u>The Nelson-Siegel Yield Curve and its Interpretation</u>

Nelson and Siegel (1987), as extended by Siegel and Nelson (1988), work with the instantaneous forward rate curve,

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

which implies the yield curve,

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

The Nelson-Siegel 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.[3] 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.

We work with the Nelson-Siegel model because of its ease of interpretation and its parsimony, which promote simplicity of modeling and accuracy of forecasting, as we shall demonstrate. We interpret β_1, β_2 and β_3 in the Nelson-Siegel model as three latent 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 $\frac{1-e^{-\lambda_t \tau}}{\lambda_t \tau}$, a function that starts at 1 but decays monotonically and quickly to 0; hence it may be viewed as a short-term factor.[4] The loading on β_{3t} is $\frac{1-e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau}$, 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

---

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

[4] The factor loading in the Vasicek (1977) model has exactly the same form, where λ_t is a mean reversion coefficient.

factor. We plot the three factor loadings in Figure 1, with λ_t=0.0609.[5] The loading on β_{3t} is maximized at a maturity of approximately three years. The factor loading plots also look very much like those obtained by Bliss (1997a), who estimated loadings via a statistical factor analysis.[6]

An important insight is that the three factors, which we have thus far called long-term, short-term and medium-term, may also be interpreted in terms of the aspects of the yield curve that they govern: 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, $y_t(120)-y_t(3) = -.78\beta_{2t}+.06\beta_{3t}$ when λ_t= 0.0609. 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 three-factor models of Balduzzi, Das, Foresi and Sundaram (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-Lund (1997) three-factor non-affine 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, $2y_t(24)-y_t(3)-y_t(120) = .00053\beta_{2t}+.37\,\beta_{3t}$ when λ_t=0.0609. 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, Squassi and Weir (1995), which is highly related yet distinct.

---

[5] In our subsequent empirical work, we find that we can fix λ_t = 0.0609 without significantly degrading the goodness of fit of the time series of fitted term structures. We do so throughout this paper, and we will provide details in due course.

[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.

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.

Finally, 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, $\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}$ 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 is hard to provide intuitive interpretations of the factors in the original Nelson-Siegel framework, and second, operationally, it is difficult to estimate the factors precisely, because the high coherence in the factors produces multicolinearity.

Stylized Facts of the Yield Curve and the Three-Factor 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. Duffee (2002), for example, shows that multi-factor affine models are inconsistent with many of the facts, perhaps because term premia may not be adequately captured by affine models.

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.

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. Following Fama and Bliss (1987), we filter the data before further analysis, 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). We then use the Fama-Bliss (1987) bootstrapping method to compute raw yields recursively from the filtered data.[7] At each step, we compute the forward rate necessary to price successively longer maturity bonds, given the yields fitted to previously included issues. We then calculate the yields by averaging the forward rates. The resulting yields, which we call "unsmoothed Fama-Bliss," exactly price the included bonds.

Although most of our analysis does not require the use of fixed maturities, doing so greatly

---

[7] We thank Rob Bliss for providing us with the computer programs and data.

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.[8] 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.[9]

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 Figure 2 we provide a three-dimensional plot of our term structure 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 monthly yields. It is clear that the average yield curve is upward sloping, that the long rates 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.[10] It is worth noting, because it will be relevant for our future modeling choices, that level, slope and curvature are not highly correlated with each other. In particular, *corr(level, slope)=0.38*, *corr(level, curvature)=0.13*, and *corr(slope, curvature)=-0.40*.

In Figures 3 and 4 we highlight and expand upon certain of the facts revealed in Table 1. In Figure 3 we display time-series plots of yield level, spread and curvature, and graphs of their sample autocorrelations to a displacement of sixty months. The very high persistence of the level, moderate persistence of the slope and comparatively weak persistence of the curvature are apparent, as is the presence of a stochastic cycle in the slope as evidenced by its oscillating autocorrelation function. In Figure 4 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.

---

[8] Due to potential problems of idiosyncratic behavior with the 1-month bill, we don't use it in the analysis. See also Duffee (1996) for additional discussion.

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

[10] That is why affine models don't fit the data well; they can't generate such high variability and quick mean reversion in curvature.

Fitting Yield Curves

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

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

We begin by estimating the parameters $\theta_t = \{\beta_{1t}, \beta_{2t}, \beta_{3t}, \lambda_t\}$ by nonlinear least squares, for each month t; that is,

$$\hat{\theta}_t = argmin_{\theta_t} \sum_{i=1}^{N_t} \epsilon_{it}^2,$$

where ϵ_{it} is the difference at time t between the observed and fitted yields at maturity τ_i. 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.[11]

We will subsequently examine the fitted series $\{\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}\}$ in detail, but first let us discuss the fitted values of λ_t. Although there is variation over time in the estimated value of λ_t, the variation is small relative to the standard error. Related, as in Nelson and Siegel (1987), we found that the sum-of-squares function is not very sensitive to λ_t. Both findings suggest that little would be lost by fixing λ_t. We verify this claim in Figure 5. In the top panel, we plot the three-factor RMSE over time (averaged over maturity), and in the bottom panel we plot it by maturity (averaged over time), with and without λ_t fixed. The differences are for the most part minor. The case for fixing λ_t becomes very strong when one adds to this the facts that estimation of lambda is fraught with difficulty in terms of getting to the global optimum, that allowing λ_t to vary over time can make the estimated $\{\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}\}$ change dramatically at various times, that a slightly better in-sample fit does not necessarily produce better out-of-sample forecasting, and that recent theoretical developments suggest that λ_t should in fact be constant (as we will discuss subsequently).

In light of the above considerations, and after some experimentation, we decided to fix λ_t=0.0609. We do so for the remainder of this paper. This lets us compute the values of the two regressors (factor loadings) and use ordinary least squares to estimate the betas (factors). Applying

---

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

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.

Assessing the Fit

There are many aspects to a full assessment of the "fit" of our model. In Figure 6 we plot the implied average fitted yield curve against the average actual yield curve. The two agree quite closely. In Figure 7 we dig deeper by plotting the raw yield curve and the three-factor fitted yield curve for some selected dates. It can be seen that the three-factor model is capable of replicating various yield curve shapes: upward sloping, downward sloping, humped, and inverted humped. It does, however, have difficulties at some dates, especially when yields are dispersed.[12] The model also has trouble fitting times when the yield curve has multiple interior optima.[13] Overall, however, the residual plot in Figure 8 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 discrepancy between actual bond prices and prices estimated from term structure models. Presumably these discrepancies arise from persistent tax and/or liquidity effects.[14] However, because they persist, they should vanish from fitted yield changes. Moreover, if interest centers on forecasting the *estimated* term structure rather than the actual term structure, then the pricing errors are irrelevant.[15]

In Figure 9 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 $\rho(\hat{\beta}_{1t}, l_t) = 0.97$, $\rho(\hat{\beta}_{2t}, s_t) = -0.99$, 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 Figure 10 we present descriptive statistics for

---

[12] Allowing λ to vary over time usually improves the fit by only a few basis points. However, it can improve the fit significantly when the short end of the yield curve is steep, which happens occasionally.

[13] Subsequently we will discuss extensions of the three-factor model that would better fit yield curves with multiple interior optima, although it's not obvious that they would produce better out-of-sample forecasts.

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

[15] Forecasting the estimated term structure is done quite commonly. See, for example, Duffee (2002).

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 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.[16] Finally, the pairwise correlations between the estimated factors are: $corr(\hat{\beta}_1, \hat{\beta}_2) = -0.55$, $corr(\hat{\beta}_1, \hat{\beta}_3) = -0.07$, and $corr(\hat{\beta}_2, \hat{\beta}_3) = 0.51$.

Modeling and Forecasting Yield Curve Level, Slope and Curvature

We model and forecast the Nelson-Siegel factors in two ways, first as univariate AR(1) processes and second as a VAR(1) process. The AR(1) or VAR(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:

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

where

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

and $\hat{c}_i$ and $\hat{\gamma}_i$ are obtained by regressing $\hat{\beta}_{it}$ on an intercept and $\hat{\beta}_{i,t-h}$. Similarly, the yield forecasts based on an underlying multivariate VAR(1) specification are:

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

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

The factors are not highly correlated, so that an appropriate multivariate model is close to a stacked set of univariate models. Hence we expect little, if any, forecasting improvement from moving to a multivariate model.

In Figure 11 we provide some evidence on goodness of fit of the models of level, slope and curvature, showing residual time series plots and autocorrelation functions. The autocorrelations are very small, indicating that the models accurately describe the conditional means of level, slope and curvature .

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

---

[16] 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.

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, 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 premia. 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-Piazzesi (2001) forward curve regression:

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

Note that the Fama-Bliss forward regression is a special case of the Cochrane-Piazzesi forward regression.[17]

[17] 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.

(5) AR(1) on yield levels:

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

(6) VAR(1) on yield levels:

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

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

(7) VAR(1) on yield changes:

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

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

(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) ECM(1) with three common trends:

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

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

We define forecast errors at $t+h$ as $y_{t+h}(\tau) - \hat{y}_{t+h/t}(\tau)$, and we report 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-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 forecasting the fitted yield curve rather than the "raw" curve, or by including variables that may explain mispricing. Second, as discussed above, we made no attempt to approximate the factor dynamics optimally; instead, we simply asserted and fit AR(1) conditional mean models. 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.

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 radically 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 ten 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. 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.[18]

Finally, we compare certain aspects of our forecasting results to those in two important and related papers. First, we note that we refute the Cochrane-Piazzesi (2001) claim that even an adequate model fit to monthly data would miss the 12-month yield forecastability. We find strong forecastability by optimizing the relevant loss function. In fact, our three-factor model fit to monthly data produces 12-month-ahead forecasts that appear much better that those from the Cochrane-Piazzesi model.

Second, we note that Duffee (2002) finds that even the simplest random walk forecasts dominate those from the Dai-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

[18] 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.

and Duffee's, however, reveals that our three-factor model 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.

4. Concluding Remarks and Directions for Future Research

We have provided a new interpretation of the Nelson-Siegel yield curve as a modern three-factor model of level, slope and curvature, and we have explored the model's performance in out-of-sample yield curve forecasting. The forecasting results at a 1-month horizon are no better than those of random walk and slope models, whereas the results at a 12-month horizon are strikingly superior. Here we discuss several variations and extensions of the basic modeling and forecasting framework developed thus far. In our view, all of them represent important directions for future research.

State-Space Representation and One-Step Estimation

Let us begin with a technical, but potentially useful and important, extension. To maximize clarity and intuitive appeal, in this paper we followed a two-step procedure, first estimating the level, slope and curvature factors, β_1, β_2, and β_3, and then modeling and forecasting them. Although we believe that we lost little by following the two-step approach, it is suboptimal relative to simultaneous estimation, which is facilitated by noticing that the model forms a state-space system. In an obvious vector notation, we have:

$$y_t = Z\beta_t + \epsilon_t$$

$$\beta_t = c + A\beta_{t-1} + R\eta_t$$

$$\begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} \sim WN(0,\ diag(Q,\ H))$$

$$E(\beta_0\ \varepsilon_t') = 0$$

$$E(\beta_0\ \eta_t') = 0,$$

$t = 1, \ldots, T.$

In particular, the measurement equation is

$$\begin{pmatrix} y_t(3) \\ y_t(6) \\ \dots \\ y_t(120) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-3\lambda}}{3\lambda} & \frac{1-e^{-3\lambda}}{3\lambda} - e^{-3\lambda} \\ 1 & \frac{1-e^{-6\lambda}}{6\lambda} & \frac{1-e^{-6\lambda}}{6\lambda} - e^{-6\lambda} \\ \dots & \dots & \dots \\ 1 & \frac{1-e^{-120\lambda}}{120\lambda} & \frac{1-e^{-120\lambda}}{120\lambda} - e^{-120\lambda} \end{pmatrix} \begin{pmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{pmatrix} + \begin{pmatrix} \epsilon_t(3) \\ \epsilon_t(6) \\ \dots \\ \epsilon_t(120) \end{pmatrix},$$

and the transition equation presently features diagonal R and A matrices. Maximum likelihood estimates are readily obtained via the Kalman filter in conjunction with the prediction-error decomposition of the likelihood, as are optimal extractions and forecasts of the latent level, slope and curvature factors. Extensions are readily accommodated, including allowing for richer dynamics (e.g., non-diagonal A and R matrices), and exogenous (e.g., macroeconomic) variables. Conditional heteroskedasticity in the latent level, slope and curvature factors, moreover, may be explicitly accommodated by allowing for a time-varying R matrix, allowing us to produce confidence "tunnels" for the entire yield curve.

Beyond Duration

The most important single source of risk associated with holding a government bond is variation in interest rates; that is, the shifting yield curve. For a discount bond, this risk is directly linked to maturity; longer maturity bonds suffer greater price fluctuations than shorter maturity bonds for a given change in the level of the interest rates. For a coupon bond paying x_i units at time t_i, i=1, 2, ..., n, where $t < t_1 < ... < t_n$, with price given by $P_{ct} = \sum_{i=1}^{n} P_t(\tau_i) x_i$ with $\tau_i = t_i - t$ as required to eliminate arbitrage opportunities, the corresponding risk measure is duration, which is a weighted average of the maturities of the underlying discount bonds,

$$D = \frac{\sum_{i=1}^{n} \tau_i P_t(\tau_i) x_i}{\sum_{i=1}^{n} P_t(\tau_i) x_i}.$$

It is well-known that duration is a valid measure of price risk only for parallel yield curve shifts. But in our model, and certainly in the real world, yield curves typically shift in non-parallel ways involving not only level, but also slope and curvature. However, we can easily generalize the notion of duration to our multi-factor framework. For a given shift in the yield curve, the risk of a discount bond with price $P_t(\tau)$ with respect to the three factors is

$$-\frac{dP_t(\tau)}{P_t(\tau)} = \tau\, dy_t(\tau) = \tau\, d\beta_{1t} + \left(\frac{1-e^{-\lambda\tau}}{\lambda}\right) d\beta_{2t} + \left(\frac{1-e^{-\lambda\tau}}{\lambda} - \tau e^{-\lambda\tau}\right) d\beta_{3t}\,.$$

Hence there are now three components of price risk, associated with the three loadings on $d\beta_{1t}$, $d\beta_{2t}$ and $d\beta_{3t}$ in the above equation. The traditional duration measure corresponds to maturity, τ, which is the loading on the level shock $d\beta_{1t}$; that is why it is an adequate risk measure only for parallel yield curve shifts. Generalized duration of course still tracks "level risk," but two additional terms, corresponding to slope and curvature risk, now feature prominently as well.

The notion of generalized duration is readily extended to coupon bonds. For a coupon bond paying x_i units at time t_i, i=1, 2, ..., n, where $t<t_1<...<t_n$, we define the corresponding three-factor generalized duration as

$$-\frac{dP_{ct}}{P_{ct}} = \sum_{i=1}^{n} w_i x_i \tau_i dy_t(\tau_i) = \sum_{i=1}^{n}\left(w_i x_i \tau_i\right) d\beta_{1t} + \sum_{i=1}^{n}\left(w_i x_i \frac{1-e^{-\lambda\tau_i}}{\lambda}\right) d\beta_{2t} + \sum_{i=1}^{n}\left(w_i x_i \frac{1-e^{-\lambda\tau_i}}{\lambda} - w_i x_i \tau_i e^{-\lambda\tau_i}\right) d\beta_{3t},$$

where $w_i = \frac{P_t(\tau_i)}{P_{ct}}$. Given that our three-factor model provides an accurate summary of yield-curve dynamics, our generalized duration should provide an accurate summary of the risk exposure of a bond portfolio, with immediate application to fixed income risk management. In future work, we plan to pursue this idea, comparing generalized duration to the traditional Macaulay duration and to the stochastic duration of Cox, Ingersoll and Ross (1979).

Active Bond Portfolio Management

Hedging and speculation are opposite sides of the same coin. Hence, as with any risk management tool, generalized duration may also be used in a speculative mode. In particular, generalized duration, by splitting bond price risk into components associated with yield curve level, slope and curvature shifts, suggests using separate forecasts of those components to guide active trading strategies.

Active bond portfolio managers attempt to profit from their views on changes in the level and shape of the yield curve. The so-called interest rate anticipation strategy, involving increasing portfolio duration when rates are expected to decline, and conversely, is an obvious way to take a position reflecting views regarding expected shifts in the level of the yield curve.

The yield curve can of course shift in ways more subtle than simple level shifts, but it nevertheless often remains straightforward to take appropriate positions reflecting views on likely movements, for example with bullet or barbell portfolios. Consider, for example, the commonly observed cases of (1) a downward shift combined with a steepening, and (2) an upward shift combined with a flattening. A bullet portfolio has maturities centered at a single point on the yield curve, and a barbell portfolio has maturities concentrated at two extreme points on the yield curve, with one maturity shorter and the other longer than that of the bullet portfolio. Hence, in general, the bullet will outperform if the yield curve steepens with long rates rising relative to short rates, because of the capital loss on the longer term bonds in the barbell portfolio. Conversely, if the yield curve flattens with long rates falling relative to short rates, the barbell will almost surely outperform because of the positive effect of capital gains on long term bonds.

Links of the Yield Curve to Macroeconomic and Other Financial Indicators

It will be of interest to explore the links between various yields and macroeconomic fundamentals, in both directions, within the framework developed here. This effectively amounts to characterizing the relationships among $\{\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}\}$ and various macroeconomic fundamentals. Consider first the effects of macroeconomic fundamentals on the term structure. Standard central bank reaction functions suggest that short rates should be related to output and inflation. For example, the Taylor (1993) rule is typically written as

$$\bar{i}_t = c + \phi_\pi(\pi_t - \pi^*) + \phi_z(z_t - z_t^*)$$

$$i_t = \theta i_{t-1} + (1-\theta)\bar{i}_t,$$

where $\bar{i}_t$ is the target short interest rate, i_t is the actual short rate set by the policy maker, π_t is the inflation rate, π^* is the target inflation rate, z_t is output, and z_t^* is target output. At issue, however, is how macroeconomic fundamentals affect yields as one moves out the yield curve. Along these lines, Ang and Piazessi (2000) find in an affine environment that macroeconomic fundamentals move yields at the short but not the long end of the curve. This result is intriguing, considering that the long rate is approximately the average of expected future short rates plus a term premium. Perhaps the reason is the restrictions placed on term premia in affine models.

Now consider the effects of the term structure on macroeconomic fundamentals. The predictive content of the yield curve has long been recognized; for recent developments see Estrella and Mishkin (1998) and Stock and Watson (2000). Typically only the term structure slope is used as a leading indicator. Our analysis suggests, in contrast, that not only slope, but also level and curvature, are necessary for a full summarization of the information in the term structure. It will be interesting, therefore, to see whether variables such as curvature have marginal predictive content.

Finally, it will be of interest to explore the interaction between the term structure and various financial, as opposed to macroeconomic, variables, as well as the interaction between different countries' term structures. For example, Clarida et al. (2001) explore the predictive content of forward premia for subsequent spot exchange rate movements. But by covered interest parity, the forward premium is just the interest rate differential, and of course there are many interest rate differentials as one moves out the term structure. Hence the ultimate issue is whether the term structure of credit spreads can be distilled efficiently via the three-factor model advocated in this paper, and then used to enhance spot exchange rate forecasts.

Generalizations to Enhance Flexibility and to Maintain Consistency with Standard Interest Rate Processes

A number of authors have proposed extensions to Nelson-Siegel that enhance flexibility. For example, Bliss (1997b) extends the model to include two decay parameters. Beginning with the instantaneous forward rate curve,

$$f_t(\tau) = \beta_{1t} + \beta_{2t} e^{-\lambda_{1t}\tau} + \beta_{3t}\lambda_{2t}\tau e^{-\lambda_{2t}\tau},$$

we obtain the yield curve

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

Obviously the Bliss curve collapses to the original Nelson-Siegel curve when $\lambda_{1t}=\lambda_{2t}$.

Soderlind and Svensson (1997) also extend Nelson-Siegel to allow for two decay parameters, albeit in a different way. They begin with the instantaneous forward rate curve,

$$f_t(\tau) = \beta_{1t} + \beta_{2t} e^{-\lambda_{1t}\tau} + \beta_{3t}\lambda_{1t}\tau e^{-\lambda_{1t}\tau} + \beta_{4t}\lambda_{2t}\tau e^{-\lambda_{2t}\tau},$$

which implies the yield curve

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_{1t}\tau}}{\lambda_{1t}\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_{1t}\tau}}{\lambda_{1t}\tau} - e^{-\lambda_{1t}\tau}\right) + \beta_{4t}\left(\frac{1-e^{-\lambda_{2t}\tau}}{\lambda_{2t}\tau} - e^{-\lambda_{2t}\tau}\right).$$

The Soderlind-Svensson model allows for up to two humps in the yield curve, whereas the original Nelson-Siegel model allows only one.

A number of authors have also considered generalizations of Nelson-Siegel to maintain consistency with arbitrage-free pricing for certain short-rate processes. Björk and Christensen (1999) show that in the Heath-Jarrow-Morton (1992) framework with deterministic volatility, Nelson-Siegel forward-rate dynamics are inconsistent with standard interest rate processes, such as those of Ho and Lee (1986) and Hull and White (1990). By "inconsistent" we mean that if we start with a forward rate curve that satisfies Nelson-Siegel, and if interest rates subsequently evolve according to the Ho-Lee or Hull-

White models, then the corresponding forward curves will not satisfy Nelson-Siegel. Filipovic (1999, 2000) extends this negative result to stochastic volatility environments.

Björk and Christensen (1999), however, show that a five-factor variant of the Nelson-Siegel forward rate curve,

$$f_t(\tau) = \beta_{1t} + \beta_{2t}\tau + \beta_{3t}e^{-\lambda\tau} + \beta_{4t}\tau e^{-\lambda\tau} + \beta_{5t}e^{-2\lambda\tau},$$

is consistent not only with Ho-Lee and Hull-White, but also with the two-factor models studied in Heath, Jarrow and Morton (1992) under deterministic volatility.[19] Björk (2000), Björk and Landén (2000) and Björk and Svensson (2001) and provide additional insight into the sorts of term structure dynamics that are consistent with various forward curves.

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, both the parsimony principle and accumulated experience suggest that parsimonious models are often more successful for out-of-sample forecasting.[20] Similarly, although consistency with some historically-popular interest rate processes is perhaps attractive ceterus paribus, it is not clear that insisting upon it would improve forecasts. Nevertheless, the serially correlated and forecastable errors produced by our three-factor model reveal the potential for improvement, perhaps by adding additional factors, and if doing so would not only improve forecasting performance, but also promote consistency with interest rate dynamics associated with modern arbitrage-free models, so much the better. We look forward to exploring this possibility in future work.

[19] Note that λ is constant in the above expression, providing further justification for our assumption of constancy in our earlier empirical work.

[20] See Diebold (2001).

References

Andersen, T.G. and 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.

Ang, A. and Piazzesi, M. (2000), "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," Working Paper, Columbia University.

Balduzzi, P., Das, S.R., Foresi, S., and 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*, forthcoming. Cambridge: Cambridge University Press.

Björk, T. and Christensen, B. (1999), "Interest Rate Dynamics and Consistent Forward Rate Curves," *Mathematical Finance*, 9, 323-348.

Björk, T. and Landén, C. (2000), "On the Construction of Finite Dimensional Realizations for Nonlinear Forward Rate Models," Manuscript, Stockholm School of Economics.

Björk, T. and 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. and 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*. London: Blackwell Publishers.

Clarida, R.H., Sarno, L., Taylor, M.P. and Valente, G. (2001), "The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond," Manuscript, Columbia University and University of Warwick.

Cochrane, J.H. and Piazzesi, M. (2001), "Bond Risk Premia," Manuscript, University of Chicago and UCLA.

Courant, E. and Hilbert, D. (1953), *Methods of Mathematical Physics*. New York: John Wiley.

Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1979), "Duration and Measurement of Basis Risk," *Journal of Business*, 52, 51-61.

Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-407.

Dai, Q. and Singleton, K. (2000), "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943-1978.

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. and 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. (2001), *Elements of Forecasting* (Second Edition). Cincinnati: South-Western.

Diebold, F.X., Gunther, T. and Tay, A.S. (1998), "Evaluating Density Forecasts, with Applications to Financial Risk Management," *International Economic Review*, 39, 863-883.

Diebold, F.X., Hahn, J. and 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.

Diebold, F.X. and Mariano, R.S. (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.

Diebold, F.X. and Sharpe, S. (1990), "Post-Deregulation Bank Deposit Rate Pricing: The Multivariate Dynamics," *Journal of Business and Economic Statistics*, 8, 281-293.

Duffie, D. and Kan, R. (1996), "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6, 379-406.

Duffee, G. (1996), "Idiosyncratic Variation of Treasury Bill Yields," *Journal of Finance*, 51, 527-551.

Duffee, G. (2002), "Term premia and interest rate forecasts in affine models," *Journal of Finance*, 57, 405-443.

Estrella, A. and Mishkin, F.S. (1998), "Predicting U.S. Recessions: Financial Variables as Leading Indicators," *Review of Economics and Statistics*, 80, 45-61.

Fama, E. and 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. and 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

Hall, A.D., Anderson, H.M. and Granger, C.W.J. (1992), "A Cointegration Analysis of Treasury Bill Yields," *Review of Economics and Statistics*, 74, 116-126.

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

Ho, T.S. and Lee, S.-B. (1986), "Term Structure Movements and the Pricing of Interest Rate Contingent Claims," *Journal of Finance*, 41, 1011-1029.

Hull, J. and White, A. (1990), "Pricing Interest-Rate-Derivative Securities," *Review of Financial Studies*, 3, 573-592.

Litzenberger, R., Squassi, G. and 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. and Kwon, H. (1993), "U.S. Term Structure Data, 1947-1991," Working Paper 93-6, Ohio State University.

Nelson, C.R. and Siegel, A.F. (1987), "Parsimonious Modeling of Yield Curves," *Journal of Business*, 60, 473-489.

Pagan, A.R., Hall, A.D. and Martin, V. (1996), "Modeling the Term Structure," in C.R Rao and G.S. Maddala (eds.), *Handbook of Statistics*, 91-118. Amsterdam: North-Holland.

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. and Nelson, C.R. (1988), "Long-term Behavior of Yield Curves," *Journal of Financial and Quantitative Analysis*, 23, 105-110.

Soderlind, P. and Svensson, L.E.O. (1997), "New Techniques to Extract Market Expectations from Financial Instruments," *Journal of Monetary Economics*, 40, 383-430.

Stock, J.H. and Watson, M.W. (2000), "Forecasting Output and Inflation: The Role of Asset Prices," Manuscript, Harvard University and Princeton University.

Swanson, N.R. and 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.

Taylor, J.B (1993), "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195-214

Vasicek, O. (1977), "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177-188.

Vasicek, O.A. and Fong, H.G. (1982), "Term Structure Modeling Using Exponential Splines," *Journal of Finance*, 37, 339-348.

Table 1
Descriptive Statistics, Yield Curves

Maturity (Months)	Mean	Standard Deviation	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

Table 2
Descriptive Statistics, Yield Curve Residuals

Maturity (Months)	Mean	Standard Deviation	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

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

Table 4
Out-of-Sample 1-Month-Ahead Forecasting Results

Nelson-Siegel with AR(1) Factor Dynamics

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$
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)
Out-of-Sample 1-Month-Ahead Forecasting Results

Cochrane-Piazzesi Forward Curve Regression

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$
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

Table 5
Out-of-Sample 6-month-Ahead Forecasting Results

Nelson-Siegel with AR(1) Factor Dynamics

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$
3 months	0.220	0.564	0.605	0.381	-0.214
1 year	0.181	0.758	0.779	0.139	-0.150
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$
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

Table 5 (Continued)
Out-of-Sample 6-month-Ahead Forecasting Results

Cochrane-Piazzesi Forward Curve Regression

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$
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

Table 6
Out-of-Sample 12-month-Ahead Forecasting Results

Nelson-Siegel with AR(1) Factor Dynamics

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
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

Table 6 (Continued)
Out-of-Sample 12-month-Ahead Forecasting Results

Fama-Bliss Forward Rate Regression

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
3 months	0.246	0.808	0.845	-0.213	-0.073
1 year	0.182	0.953	0.970	-0.271	-0.004
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
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

Table 6 (Continued)
Out-of-Sample 12-month-Ahead Forecasting Results

VAR(1) on Yield Changes

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
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

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
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

ECM(3) with Three Common Trends

Maturity (τ)	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
3 months	0.810	0.951	1.249	-0.245	-0.082
1 year	0.786	1.261	1.486	-0.248	-0.064
3 years	0.613	1.453	1.577	-0.289	0.028
5 years	0.306	1.236	1.273	-0.246	-0.069
10 years	0.063	1.141	1.143	-0.191	-0.086

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

Notes to Tables

Notes to Table 1:
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.

Notes to Table 2:
We fit the three-factor model,

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

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.

Notes to Table 3:
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.

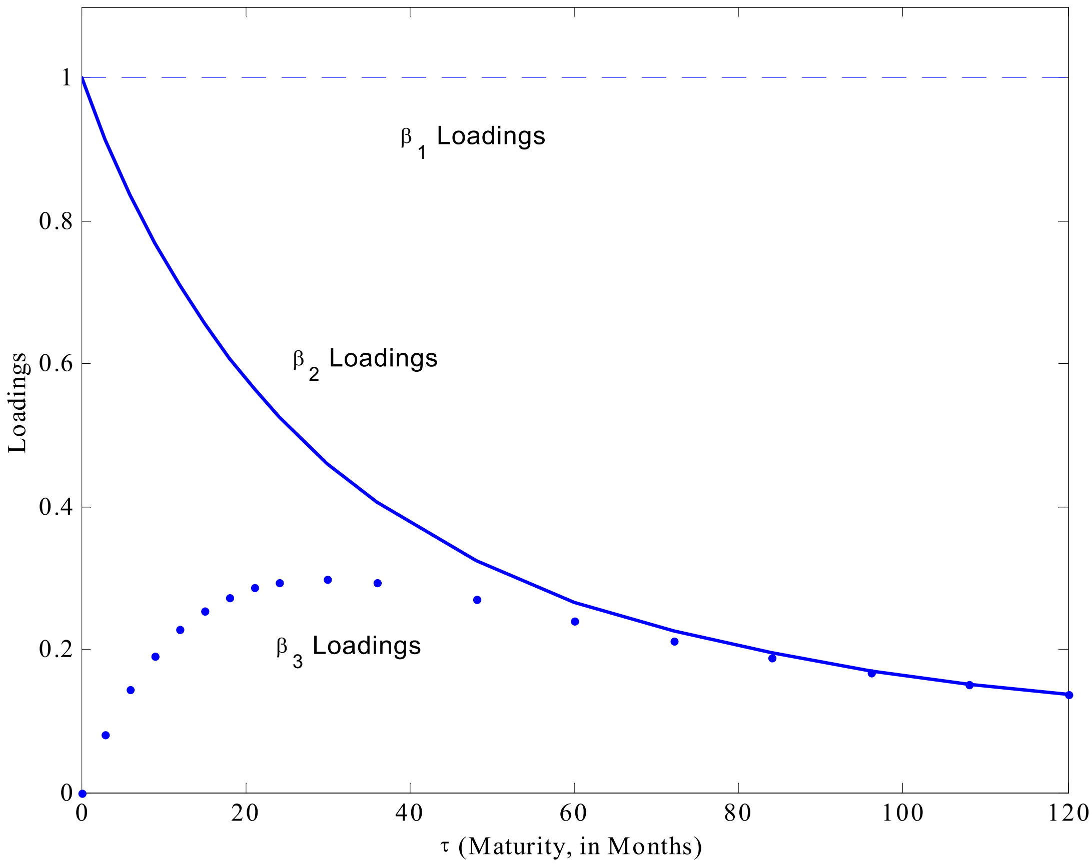
Notes to Table 4:
We present the results of out-of-sample 1-month-ahead forecasting using eight models, as described in detail in the text. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors 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 twelfth sample autocorrelation coefficients.

Notes to Table 5:
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.

Notes to Table 6:
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 twelfth and twenty-fourth sample autocorrelation coefficients.

Notes to Table 7:
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 ten percent level.

Figure 1
Factor Loadings

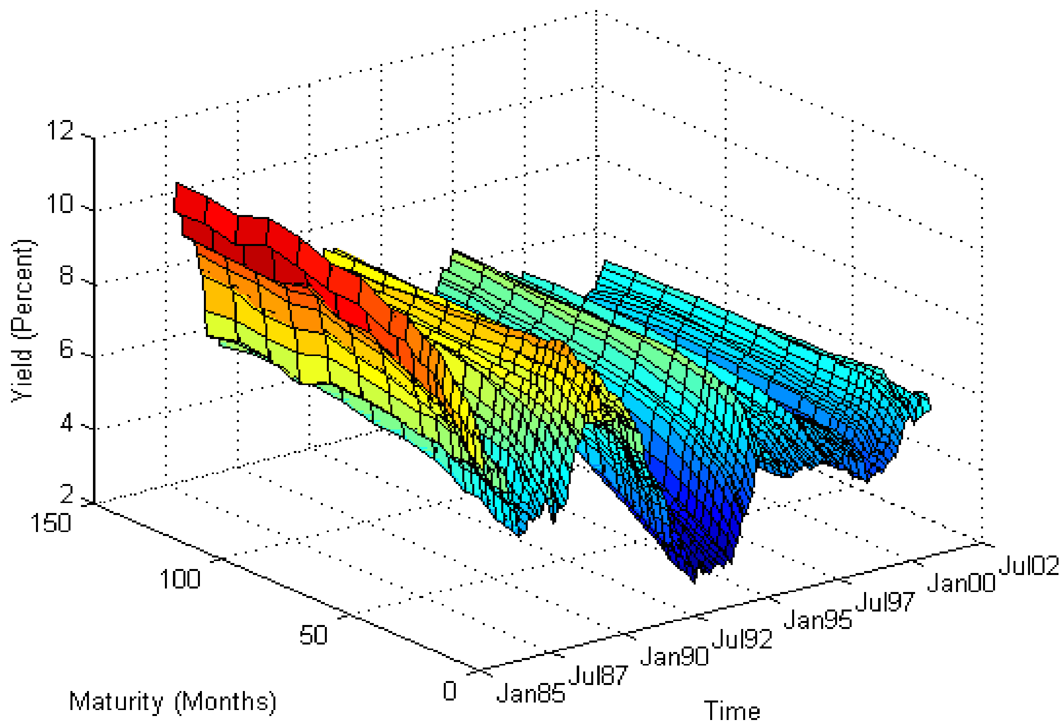


Notes to Figure 1: We plot the factor loadings in the three-factor model,

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

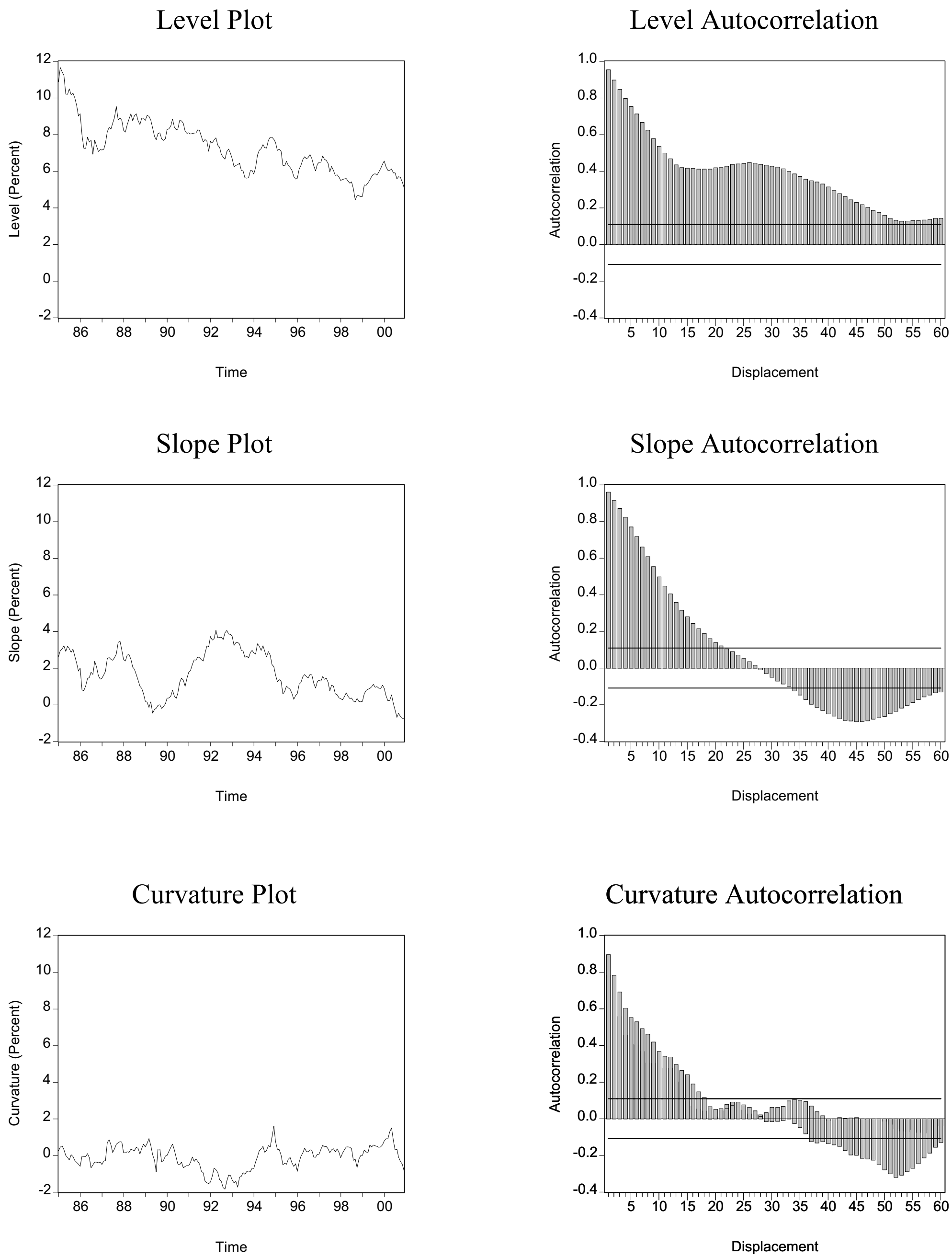
where the three factors are β_{1t}, β_{2t}, and β_{3t}, the associated loadings are 1, $\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}$, and $\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}$, and τ denotes maturity. We fix $\lambda_t = 0.0609$.

Figure 2
Yield Curves, 1985.01-2000.12



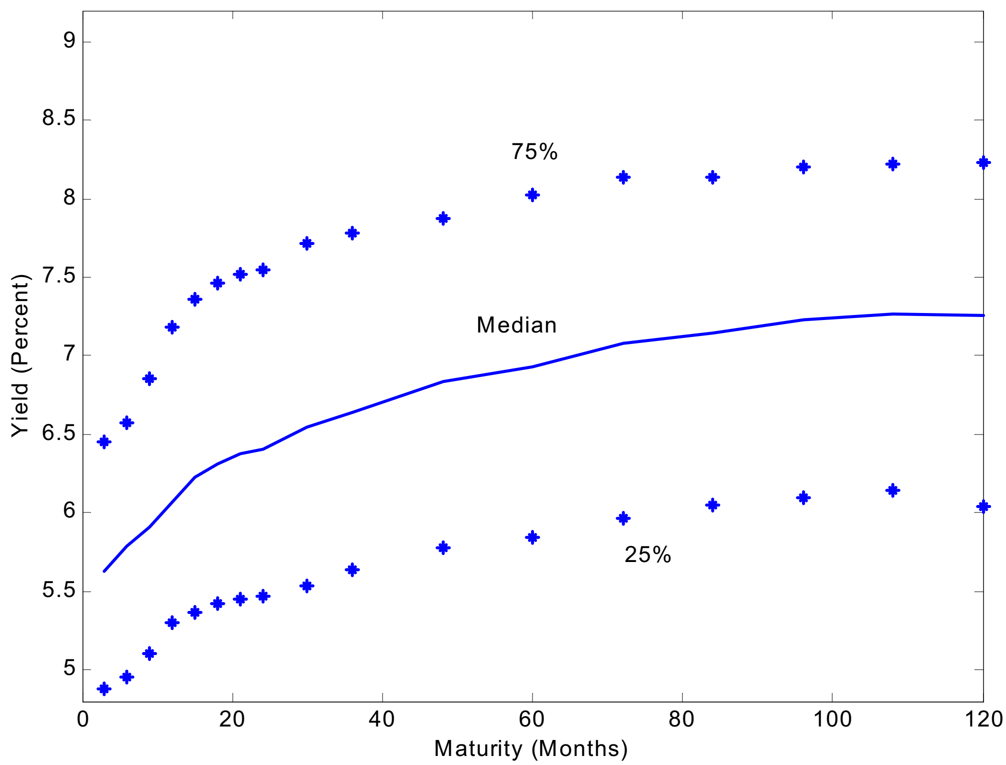
Notes to Figure 2: The sample consists of monthly yield data from January 1985 to December 2000 at maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months.

Figure 3
Yield Curve Level, Slope and Curvature (Data-Based)



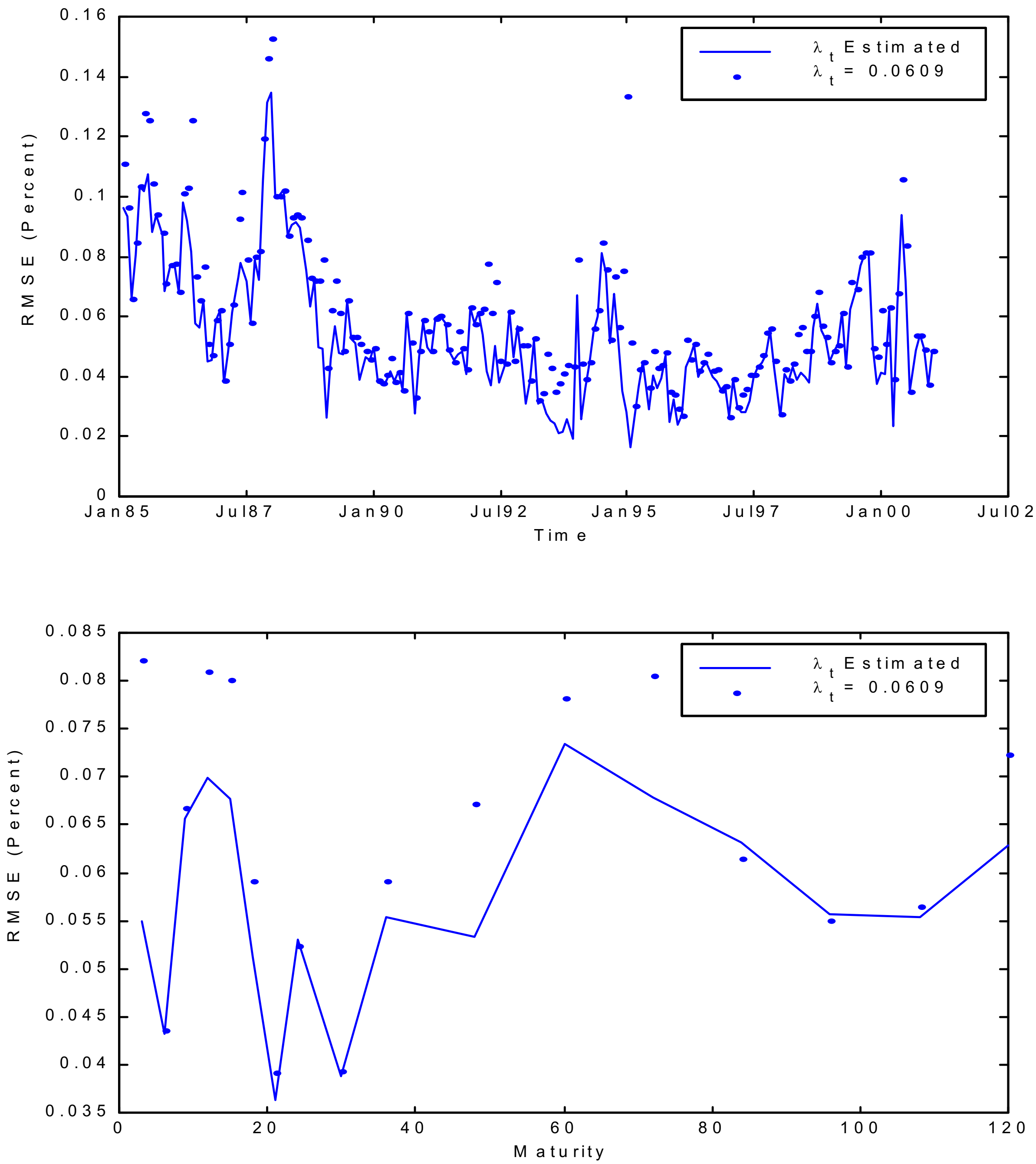
Notes to Figure 3: In the left panel we present time-series plots of yield curve level, slope and curvature, and in the right panel we plot their sample autocorrelations, to a displacement of 60 months, along with Bartlett's approximate 95% confidence bands. We define the level as the 10-year yield, 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.

Figure 4
Median Data-Based Yield Curve with Pointwise Interquartile Range



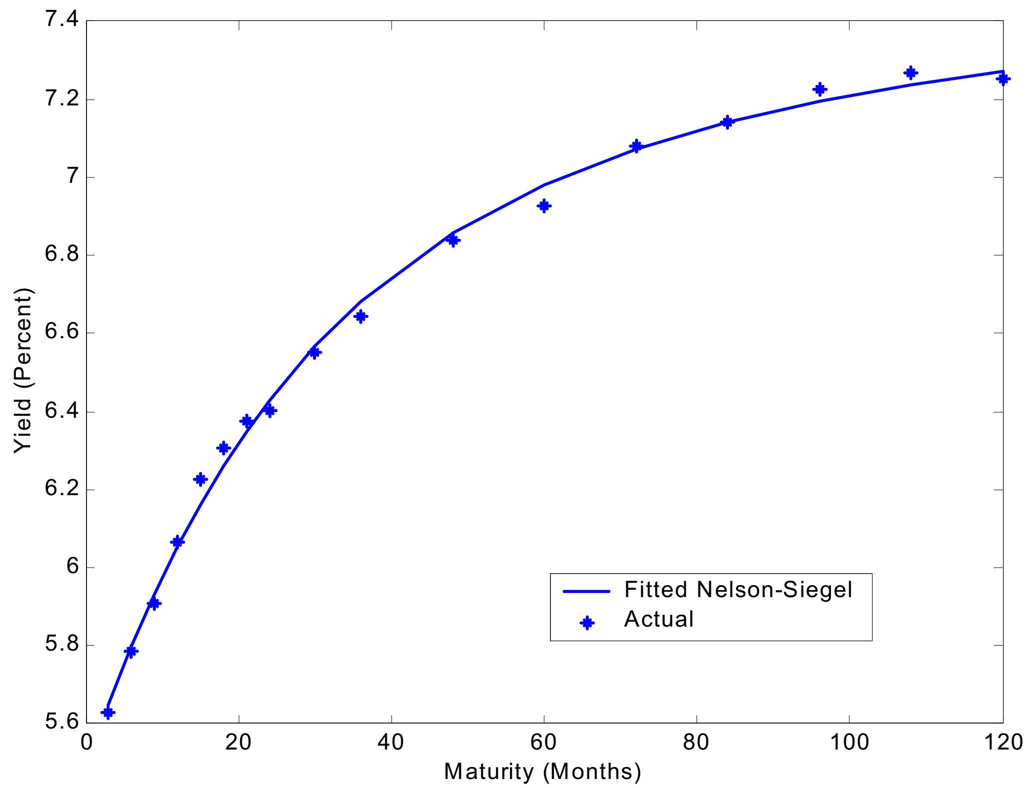
Notes to Figure 4: For each maturity, we plot the median yield along with the twenty-fifth and seventy-fifth percentiles.

Figure 5
RMSE of Yield Curve Residuals with Estimated vs. Fixed λ_t



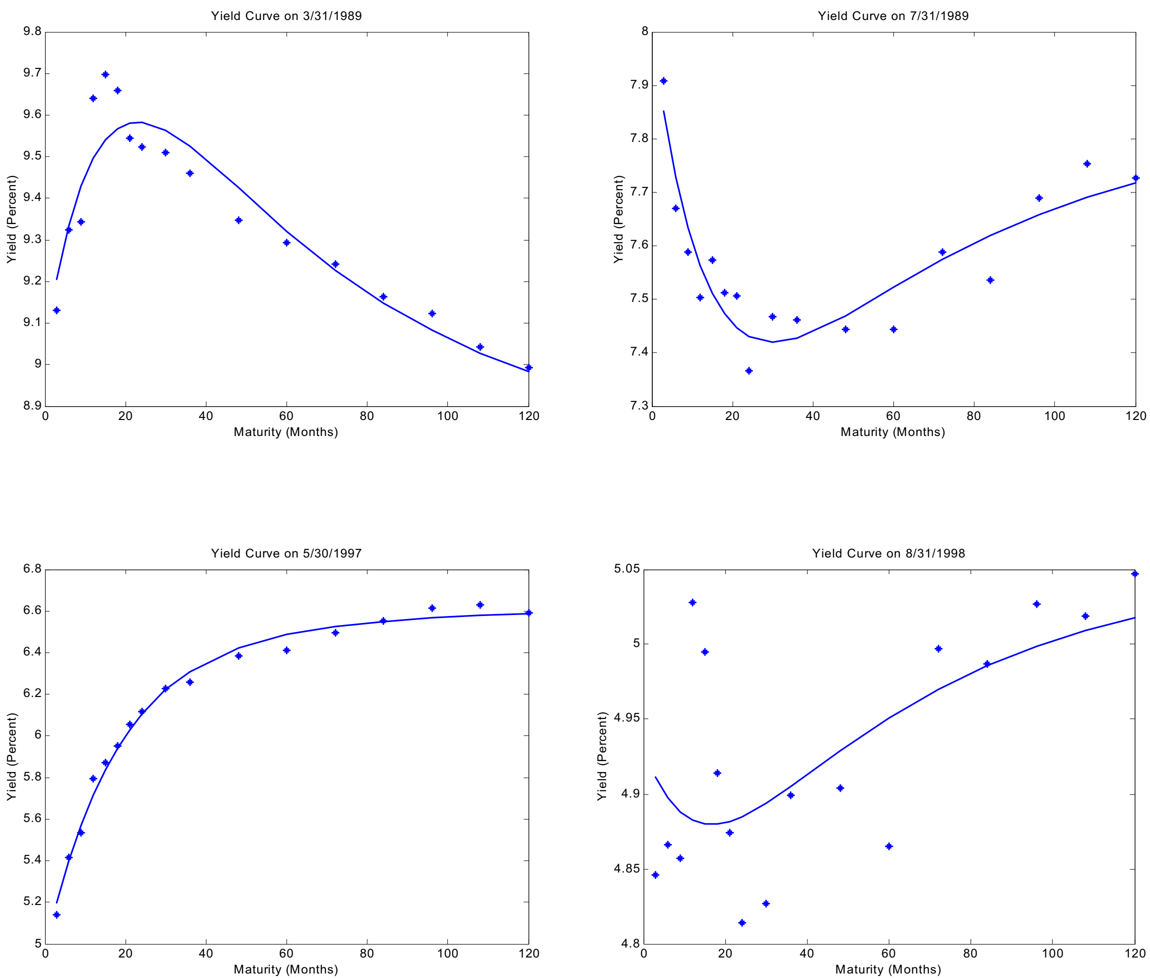
Notes to Figure 5: In the top panel, we plot RMSE over time (averaged over maturity), and in the bottom panel we plot it by maturity (averaged over time). In both panels, the solid line corresponds to λ_t estimated and the dotted line correspond to λ_t fixed at 0.0609.

Figure 6
Actual (Data-Based) and Fitted (Model-Based) Average Yield Curve



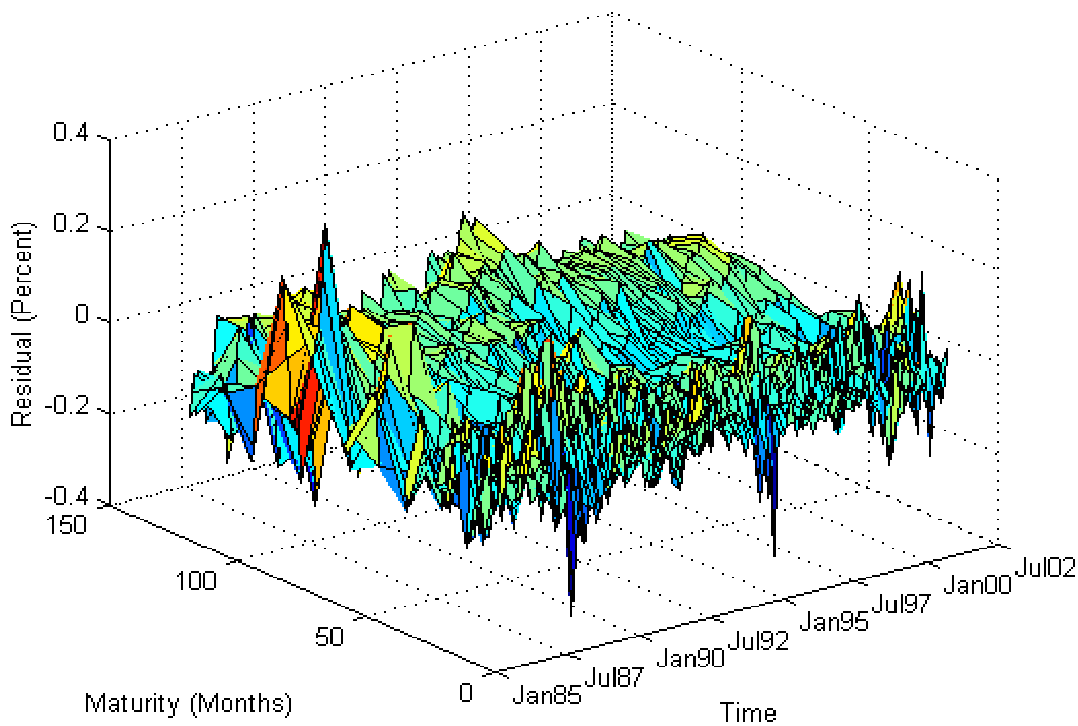
Notes to Figure 6: We show the actual average yield curve and the fitted average yield curve obtained by evaluating the Nelson-Siegel function at the mean values of $\hat{\beta}_{1t}$, $\hat{\beta}_{2t}$ and $\hat{\beta}_{3t}$ from Table 3.

Figure 7
Selected Fitted (Model-Based) Yield Curves



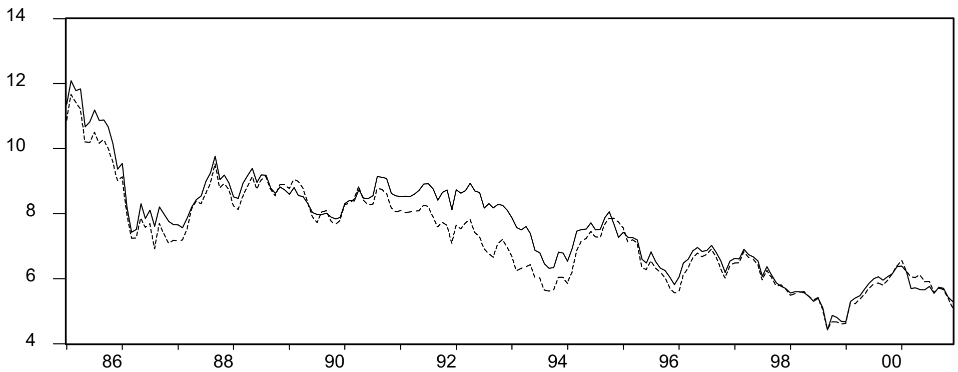
Notes to Figure 7: We plot fitted yield curves for selected dates, together with actual yields. See text for details.

Figure 8
Yield Curve Residuals, 1985.01 - 2000.12

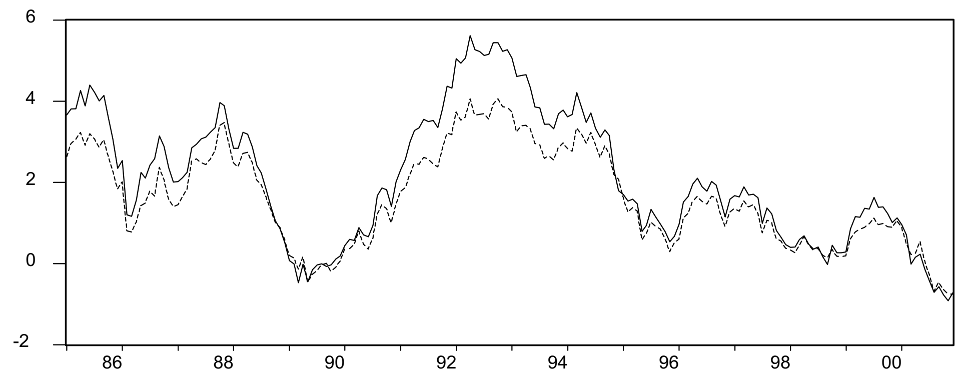


Notes to Figure 8: We plot residuals from Nelson-Siegel yield curves fitted month-by-month. See text for details.

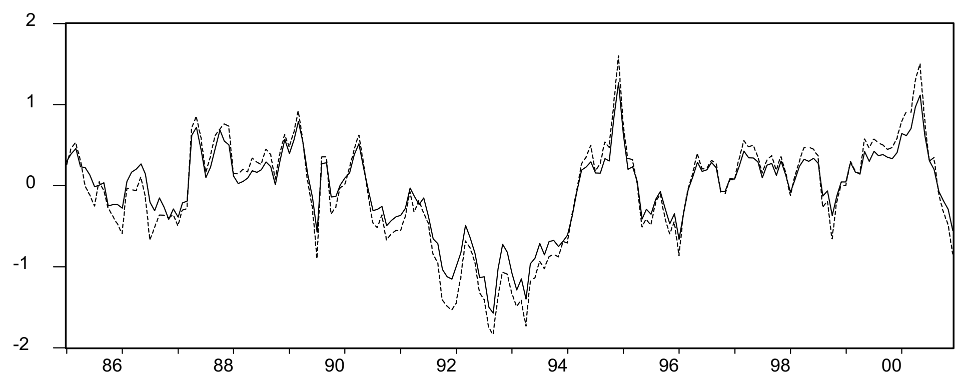
Figure 9
Model-Based Level, Slope and Curvature (i.e., Estimated Factors) vs. Data-Based Level, Slope and Curvature



Solid Line: $\hat{\beta}_{1t}$ Dotted Line: Level



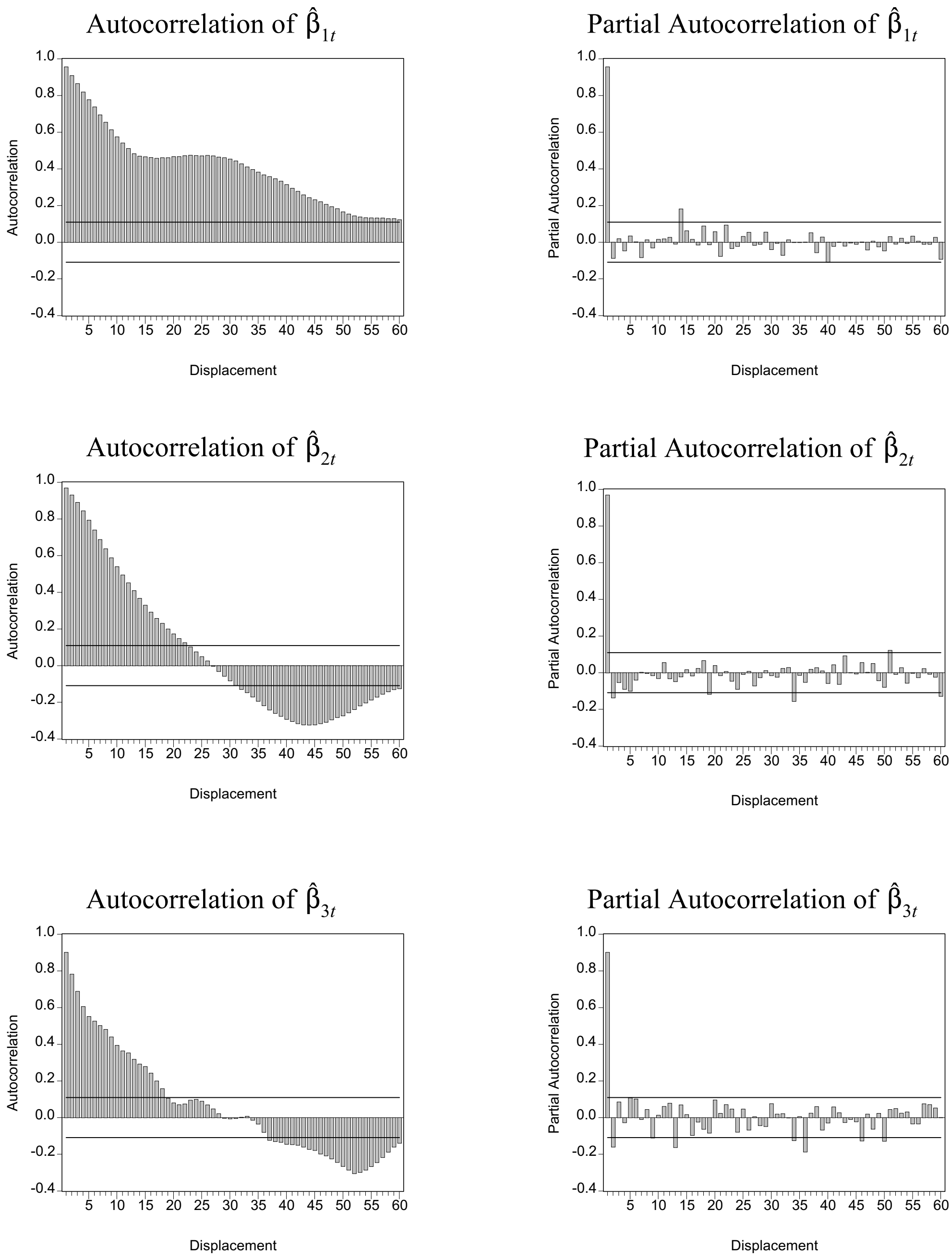
Solid Line: $-\hat{\beta}_{2t}$ Dotted Line: Slope



Solid Line: $0.3\hat{\beta}_{3t}$ Dotted Line: Curvature

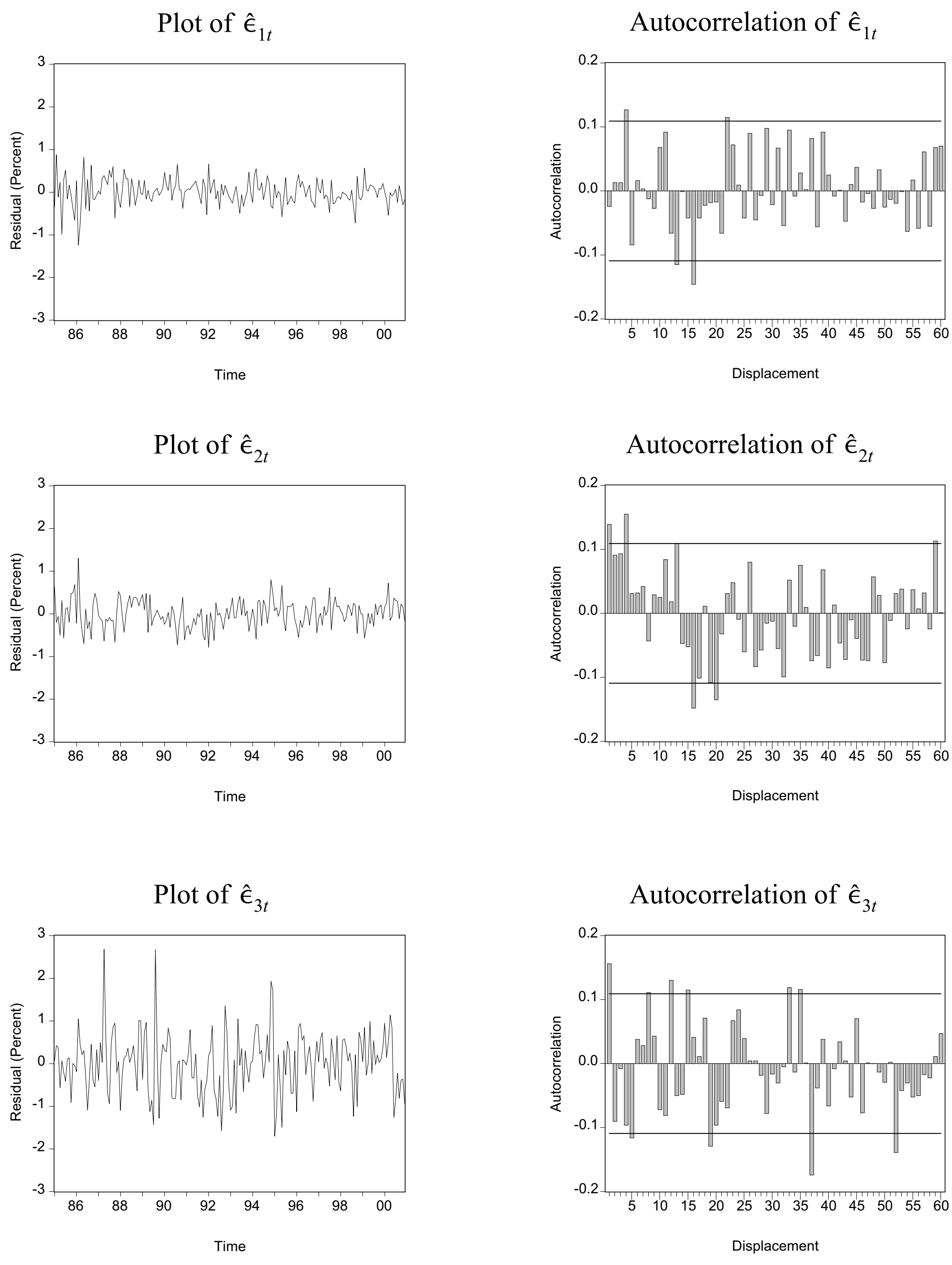
Notes to Figure 9: We define the level as the 10-year yield, the slope as the difference between the 10-year and 3-month yields, and the curvature as the twice the 2-year yield minus the sum of the 3-month and 10-year yields.

Figure 10
Sample Autocorrelations and Partial Autocorrelations of Estimated Factors



Notes to Figure 10: We plot the sample autocorrelations and partial autocorrelations of the three estimated factors $\hat{\beta}_{1t}$, $\hat{\beta}_{2t}$, and $\hat{\beta}_{3t}$, along with Barlett's approximate 95% confidence bands.

Figure 11
AR(1) Factor Model Residuals and Residual Autocorrelations



Notes to Figure 11: In the left panel, we show time series plots of residuals $\{\hat{\epsilon}_{1t}, \hat{\epsilon}_{2t}, \hat{\epsilon}_{3t}\}$ from univariate AR(1) models fit to the estimated factors $\{\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}\}$, and in the right panel we plot their sample autocorrelations, to a displacement of 60 months, along with Bartlett's approximate 95% confidence bands.