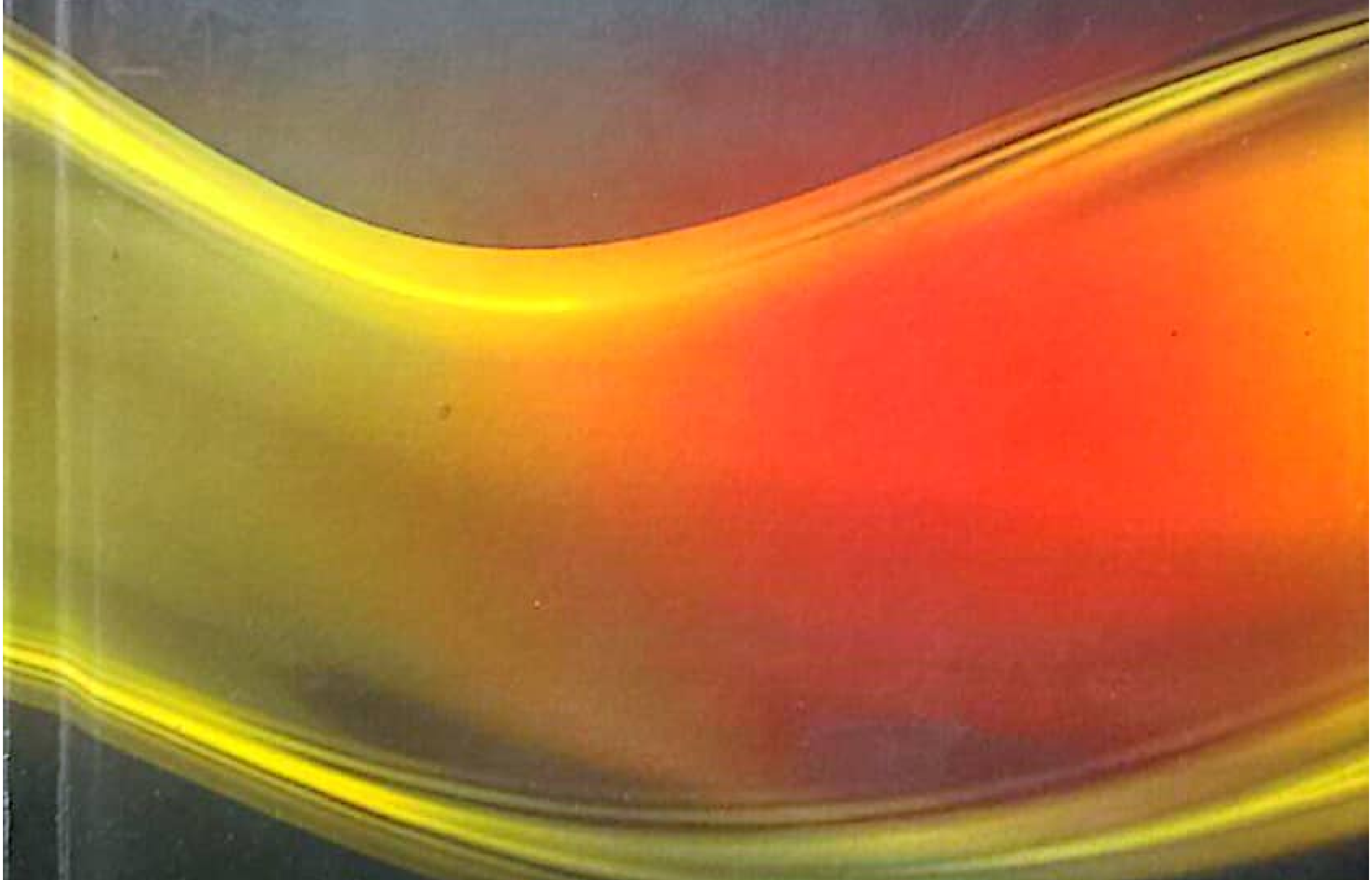
YIELD CURVE MODELING AND FORECASTING

THE DYNAMIC NELSON-SIEGEL APPROACH



FRANCIS X. DIEBOLD
GLENN D. RUDEBUSCH

simultaneously in close touch with modern statistical and financial economic thinking, and effective in a variety of situations. But we are getting ahead of ourselves. First we must lay the groundwork.

1.1 Three Interest Rate Curves

Here we fix ideas, establish notation, and elaborate on key concepts by recalling three key theoretical bond market constructs and the relationships among them: the discount curve, the forward rate curve, and the yield curve. Let $P(\tau)$ denote the price of a τ-period discount bond, that is, the present value of \$1 receivable τ periods ahead. If $y(\tau)$ is its continuously compounded yield to maturity, then by definition

$$P(\tau) = e^{-\tau y(\tau)}. \tag{1.1}$$

Hence the discount curve and yield curve are immediately and fundamentally related. Knowledge of the discount function lets one calculate the yield curve.

The discount curve and the forward rate curve are similarly fundamentally related. In particular, the forward rate curve is defined as

$$f(\tau) = \frac{-P'(\tau)}{P(\tau)}. \tag{1.2}$$

Thus, just as knowledge of the discount function lets one calculate the yield curve, so too does knowledge of the discount function let one calculate the forward rate curve.

Equations (1.1) and (1.2) then imply a relationship between the yield curve and forward rate curve,

$$y(\tau) = \frac{1}{\tau}\int_0^\tau f(u)du. \tag{1.3}$$

In particular, the zero-coupon yield is an equally weighted average of forward rates.

The upshot for our purposes is that, because knowledge of any one of $P(\tau)$, $y(\tau)$, and $f(\tau)$ implies knowledge of the other two, the three are effectively interchangeable. Hence with no loss of generality one can choose to work with $P(\tau)$, $y(\tau)$, or $f(\tau)$. In this book, following much of both academic and industry practice, we work with the yield curve, $y(\tau)$. But again, the choice is inconsequential in theory.

Complications arise in practice, however, because although we observe prices of traded bonds with various amounts of time to maturity, we do not directly observe yields, let alone the zero-coupon yields at fixed standardized maturities (e.g., six-month, ten-year, ...), with which we work throughout. Hence we now provide some background on yield construction.

1.2 Zero-Coupon Yields

In practice, yield curves are not observed. Instead, they must be estimated from observed bond prices. Two historically popular approaches to constructing yields proceed by fitting a smooth discount curve and then converting to yields at the relevant maturities using formulas (1.2) and (1.3).

The first discount curve approach to yield curve con-

2
Dynamic Nelson-Siegel

Here we begin our journey. We start with static Nelson-Siegel curve fitting in the cross section, but we proceed quickly to dynamic Nelson-Siegel modeling, with all its nuances and opportunities. Among other things, we emphasize the model's state-space structure, we generalize it to the multicountry context, and we highlight aspects of its use in risk management and forecasting.

2.1 Curve Fitting

As we will see, Nelson-Siegel fits a smooth yield curve to unsmoothed yields. One can arrive at a smooth yield curve in a different way, fitting a smooth discount curve to unsmoothed bond prices and then inferring the implied yield curve. That's how things developed historically, but there are problems, as discussed in Chapter 1.

So let us proceed directly to the static Nelson-Siegel representation. At any time, one sees a large set of yields and may want to fit a smooth curve. Nelson and Siegel (1987) begin with a forward rate curve and fit the function

$$f(\tau) = \beta_1 + \beta_2 e^{-\lambda\tau} + \beta_3 \lambda\tau e^{-\lambda\tau}.$$

The corresponding static Nelson-Siegel yield curve is

$$y(\tau) = \beta_1 + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right). \tag{2.1}$$

Note well that these are simply functional form suggestions for fitting the *cross section* of yields.

At first pass, the Nelson-Siegel functional form seems rather arbitrary—a less-than-obvious choice for approximating an arbitrary yield curve. Indeed many other functional forms have been used with some success, perhaps most notably the smoothing splines of Fisher et al. (1995).

But Nelson-Siegel turns out to have some very appealing features. First, it desirably enforces some basic constraints from financial economic theory. For example, the corresponding discount curve satisfies $P(0) = 1$ and $\lim_{\tau\to\infty} P(\tau) = 0$, as appropriate. In addition, the zero-coupon Nelson-Siegel curve satisfies

$$\lim_{\tau\to 0} y(\tau) = f(0) = r,$$

the instantaneous short rate, and $\lim_{\tau\to\infty} y(\tau) = \beta_1$, a constant.

Second, the Nelson-Siegel form provides a parsimonious approximation. Parsimony is desirable because it promotes smoothness (yields tend to be very smooth functions of maturity), it guards against in-sample overfitting (which is important for producing good forecasts), and it promotes empirically tractable and trustworthy estimation (which is always desirable).

Third, despite its parsimony, the Nelson-Siegel form also provides a flexible approximation. Flexibility is desirable because the yield curve assumes a variety of shapes at different times. Inspection reveals that, depending on the values of the four parameters (β_1, β_2, β_3, λ), the Nelson-Siegel curve can be flat, increasing, or decreasing linearly, increasing or decreasing at an increasing or decreasing rate, U-shaped, or upside-down U-shaped. It can't have more than one internal optimum, but that constraint is largely nonbinding, as the yield curve tends not to "wiggle" with maturity.

Fourth, from a mathematical approximation-theoretic viewpoint, the Nelson-Siegel form is far from arbitrary. As Nelson and Siegel insightfully note, the forward rate curve corresponding to the yield curve (2.1) is a constant plus a Laguerre function. Laguerre functions are polynomials multiplied by exponential decay terms and are well-known mathematical approximating functions on the domain $[0, \infty)$, which matches the domain for the term structure. Moreover, as has recently been discovered and as we shall discuss later, the desirable approximation-theoretic properties of Nelson-Siegel go well beyond their Laguerre structure.

For all of these reasons, Nelson-Siegel has become very popular for static curve fitting in practice, particularly among financial market practitioners and central banks, as discussed, for example, in Svensson (1995), BIS (2005), Gürkaynak et al. (2007), and Nyholm (2008). Indeed the Board of Governors of the U.S. Federal Reserve System fits and publishes on the Web daily yield Nelson-Siegel curves in real time, as does the European Central Bank.[1] We now proceed to dynamize the Nelson-Siegel curve.

[1] The FRB and ECB curves are actually based on an extension of Nelson-Siegel introduced in Svensson (1995), which we discuss and extend even further in Chapter 4.

2.2 Introducing Dynamics

Following Diebold and Li (2006), let us now recognize that the Nelson-Siegel parameters must be time-varying if the yield curve is to be time-varying (as it obviously is). This leads to a reversal of the perspective associated with static Nelson-Siegel (2.1), which then produces some key insights.

2.2.1 Mechanics

Consider a cross-sectional environment for fixed t. The Nelson-Siegel model is

$$y(\tau) = \beta_1 + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right).$$

This is a cross-sectional linear projection of $y(\tau)$ on variables $(1, ((1 - e^{-\lambda\tau})/\lambda\tau), ((1 - e^{-\lambda\tau})/\lambda\tau - e^{-\lambda\tau}))$ with parameters β_1, β_2, β_3.[2]

Alternatively, consider a time-series environment for fixed τ. The model becomes

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

This is a time-series linear projection of y_t on *variables* β_{1t}, β_{2t}, β_{3t} with *parameters* $(1, ((1 - e^{-\lambda\tau})/\lambda\tau), ((1 - e^{-\lambda\tau})/\lambda\tau - e^{-\lambda\tau}))$.

Hence from a cross-sectional perspective the βs are parameters, but from a time-series perspective the βs are variables. Combining the spatial and temporal perspectives produces the *dynamic Nelson-Siegel (DNS) model*:

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right). \tag{2.2}$$

[2] For now, assume that λ is known. We will elaborate on λ later.

set of parameter values. In practice the parameters are unknown and must be estimated, to which we now turn.

2.4 Estimation

Several procedures are available for estimating the DNS model, ranging from a simple two-step procedure, to exact maximum likelihood estimation using the state-space representation in conjunction with the Kalman filter, to Bayesian analysis using Markov-chain Monte Carlo methods. We now introduce them and provide some comparative assessment.

2.4.1 Static Nelson-Siegel in Cross Section

First consider estimation of the static Nelson-Siegel model in the cross section. The four-parameter Nelson-Siegel curve (2.1) is intrinsically nonlinear but may be estimated by iterative numerical minimization of the sum of squares function (nonlinear least squares). Importantly, moreover, note that if λ is known or can be calibrated, estimation involves just trivial *linear* least squares regression of $y(\tau)$ on 1, $((1 - e^{-\lambda\tau})/\lambda\tau)$, and $((1 - e^{-\lambda\tau})/\lambda\tau - e^{-\lambda\tau})$.

In practice λ *can* often be credibly calibrated and treated as known, as follows. Note that λ determines where the loading on the curvature factor c_t achieves its maximum. For c_t to drive curvature, its loading should be maximal at a medium maturity τ_m. One can simply choose a reasonable τ_m and reverse engineer the corresponding λ_m. Values of m in the range of two or three

[7] For details see standard texts such as Harvey (1990).

years are commonly used; for example, Diebold and Li (2006) use $m = 30$ months. Model fit is typically robust to the precise choice of λ.

2.4.2 Two-Step DNS

The first estimation approach, so-called two-step DNS, was introduced by Diebold and Li (2006). Consider first the case of calibrated λ. In step 1, we fit the static Nelson-Siegel model (2.1) for each period $t = 1, \ldots, T$ by OLS. This yields a three-dimensional time series of estimated factors, $\{\hat{l}_t, \hat{s}_t, \hat{c}_t\}_{t=1}^{T}$, and a corresponding N-dimensional series of residual pricing errors ("measurement disturbances"), $\{\hat{\varepsilon}_t(\tau_1),\ \hat{\varepsilon}_t(\tau_2),\ \hat{\varepsilon}_t(\tau_N)\}_{t=1}^{T}$.[8] The key is that DNS distills an N-dimensional time series of yields into a three-dimensional time series of yield factors, $\{\hat{l}_t, \hat{s}_t, \hat{c}_t\}_{t=1}^{T}$.

Next, in step 2, we fit a dynamic model to $\{\hat{l}_t, \hat{s}_t, \hat{c}_t\}_{t=1}^{T}$. An obvious choice is a vector autoregression, but there are many possible variations, some of which we will discuss subsequently. Step 2 yields estimates of dynamic parameters governing the evolution of the yield factors ("transition equation parameters"), as well as estimates of the factor innovations ("transition disturbances").

The benefits of two-step estimation with calibrated λ (relative to one-step estimation, which we will discuss shortly) are its simplicity, convenience, and numerical stability: Nothing is required but trivial linear

[8] 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, which seems desirable. It would be interesting to explore loss functions that go even further in reflecting such economic considerations, based, for example, on bond portfolio pricing or success of trading rules, such as that done in different but related contexts by Bates (1999) and Fabozzi et al. (2005). Thus far, the DNS literature has not pursued that route aggressively.

regressions. Moreover, one can of course estimate λ as well, if desired, with only a slight increase in complication. The first-step OLS regressions then become four-parameter nonlinear least squares regressions, and the second-step three-dimensional dynamic model for $\{\hat{l}_t, \hat{s}_t, \hat{c}_t\}_{t=1}^{T}$ becomes a four-dimensional dynamic model for $\{\hat{l}_t, \hat{s}_t, \hat{c}_t, \hat{\lambda}_t\}_{t=1}^{T}$.[9] The cost of two-step estimation is its possible statistical suboptimality, insofar as the first-step parameter estimation error is ignored in the second step, which may distort second-step inference.

2.4.3 One-Step DNS

The second estimation approach, which was introduced by Diebold et al. (2006b), is so-called one-step DNS. The basic insight is that exploitation to the state-space structure of DNS allows one to do all estimation simultaneously.

One-step estimation can be approa[illegible] achieved in several ways. On the classical sid[illegible] aximum likelihood estimation may be done u[illegible] alman ter, which delivers the innovations [illegible] evalua[illegible]

(2000) results are equally ... ever, because it still requires many runs of the Kalman filter.

Finally, one can often blend the methods productively. Two-step DNS, for example, may provide quick and accurate startup values for EM iteration. The EM algorithm, moreover, typically gets close to an optimum very quickly but is ultimately slow to reach full convergence.[12] Hence one may then switch to a gradient-based method, which can quickly move to an optimum when given highly accurate startup values from EM.

2.4.3.4 Bayesian estimation

Optimization in high-dimensional spaces is always a challenging problem. Some of the methods or combinations of methods discussed thus far may confront that problem better than others, but all must nevertheless grapple with it.

Moving to a Bayesian approach may therefore be helpful, because it replaces optimization with *averaging* in the estimation of moments (e.g., posterior means). Averaging is mathematically easier than optimization.[13]

Quite apart from the pragmatic motivation above, Bayesian analysis of DNS may also be intrinsically appealing for the usual reasons (see, for example, Koop (2003)) as long as one is willing and able to specify a credible prior and likelihood. The multimove Gibbs

---

[12] In the original EM paper, Dempster et al. (1977) show that EM converges at a linear rate, in contrast to the faster quadratic rate achieved by many gradient-based algorithms.

[13] Optimization and averaging are of course related, however, as emphasized by Chernozhukov and Hong (2003).

sampler of Carter and Kohn (1994) facilitates simple Bayesian analysis of state-space models such as DNS.[14]

In addition, Bayesian analysis may be especially appealing in the DNS context because there is a potentially natural shrinkage direction (i.e., a natural prior mean), corresponding to the restrictions associated with the absence of arbitrage possibilities. We shall subsequently have much to say about absence of arbitrage, what it implies in the DNS context, and whether its strict imposition is desirable. Bayesian shrinkage estimation is potentially appealing because it blends prior and data information, coaxing but not forcing the MLE toward the prior mean, with the exact amount depending on prior precision versus likelihood curvature. In any event, if one is going to shrink the maximum-likelihood estimates in one direction or another, a natural shrinkage direction—and one clearly motivated by financial economic theory—would appear to be toward no-arbitrage.

2.4.3.5 Discussion

In our view, there is no doubt that the state-space framework is a powerful and productive way to conceptualize the structure and estimation of DNS. There is also little doubt that the one-step estimation afforded by the state-space framework is superior to two-step estimation in principle.

We conjecture, however, that little is lost in practice by using two-step estimation, because there is typically enough cross-sectional variation such that $\hat{l}_t$, $\hat{s}_t$, and $\hat{c}_t$ are estimated very precisely at each time t. Moreover, in applications we have found that gradient-based

[14] See also the exposition and insightful applications in Kim and Nelson (1999).

one-step estimation is frequently intractable or incompletely trustworthy. In particular, if forced to choose between two-step and traditional gradient-based one-step estimation procedures, we would lean toward the two-step method. When traditional one-step converges, two-step nevertheless tends to match closely, but one-step doesn't always converge in a trustworthy fashion, whereas two-step is always simple and trustworthy.[15]

The EM and Bayesian one-step methods that we sketched above, however, are in certain respects much more sophisticated and more numerically stable than traditional gradient-based one-step methods. Bayesian Markov-chain Monte Carlo methods are, after all, now frequently and successfully employed in modeling environments with many hundreds of parameters, and there is no obvious reason why DNS should be an exception. Hence it will be interesting to see how the literature develops as experience accumulates, and which estimation approach is ultimately preferred.

2.5 Multicountry Modeling

The state-space structure of DNS makes it easy to generalize to incorporate a layered, hierarchical stru