

Forecasting with the term structure: The role of no-arbitrage restrictions

*Gregory R. Duffee**

Johns Hopkins University

First draft: October 2007

This Draft: July 2009

ABSTRACT

No-arbitrage term structure models impose cross-sectional restrictions among yields and can be used to impose dynamic restrictions on risk compensation. This paper evaluates the importance of these restrictions when using the term structure to forecast future bond yields or macroeconomic activity. It concludes that no cross-sectional restrictions are helpful, because cross-sectional properties of yields are easy to infer with high precision. Dynamic restrictions are useful, but can be imposed without reliance on the no-arbitrage structure. In practice, the assumption that long-term bond yields follow a random walk appears to be the only dynamic restriction that improves the forecasting performance of term structure models.

*Voice 410-516-8828, email duffee@jhu.edu. Address correspondence to 440 Mergenthaler Hall, 3400 N. Charles St., Baltimore, MD 21218. Thanks to Philippe Mueller and seminar participants at NYU, Penn State, Ohio State, Johns Hopkins, the Spring 2009 Adam Smith Asset Pricing Conference, and a Wharton brown bag seminar.

1 Introduction

No-arbitrage term structure models are rapidly becoming important forecasting tools. For example, Gaussian versions of affine models are employed by Duffee (2002), Dai and Singleton (2002), and Christensen, Diebold, and Rudebusch (2009) to predict Treasury yields, by Cochrane and Piazzesi (2008) to predict excess bond returns, and by Ang and Piazzesi (2003) to predict macroeconomic activity. This literature argues that models satisfying no-arbitrage should produce more accurate forecasts than models that do not impose such restrictions. No-arbitrage implies the existence of an equivalent-martingale measure, which imposes restrictions on the cross-section of yields. No-arbitrage also provides a mechanism to specify a functional form for risk compensation, which imposes restrictions on yield dynamics.

Here I take a close look at the role that no-arbitrage plays in forecasting. We can nest an n -factor affine model that satisfies no-arbitrage restrictions in an n -factor affine model that does not impose no-arbitrage. In theory and practice, how do forecasts produced with the former model differ from those produced by the latter?

I make two main points. First, the theory of no-arbitrage affine models implies that the cross-sectional restrictions are useless in forecasting. If the restrictions are true, they are effectively irrelevant. If they are false, they produce misspecification. Second, empirically valuable restrictions on yield dynamics can be imposed without relying on a researcher's ability to intuit the correct functional form of risk compensation. Thus in practice, we need not look to no-arbitrage when building a parsimonious dynamic model of yields that produces more accurate forecasts than, say, regression-based methods.

There is simple intuition behind the irrelevance of cross-sectional restrictions. Recall that the restrictions are derived from standard contingent-claims logic. When n shocks drive uncertainty in prices of bonds of all maturities, prices of any n of these bonds determine prices of all other bonds. When the term structure is described by an n -factor affine model, this restriction takes the form of an affine relation between the yield on an m -maturity bond and yields on n other bonds. Duffie and Kan (1996) derive restrictions on the loadings of this "yield-factor" model.

Yet if we take the n -factor setting literally, the affine relation among yields can be determined without bothering to apply no-arbitrage. Simply regress the maturity- m yield on a constant and the yields of the n other bonds. There is no estimation error in the loadings because the R^2 is one. Since a model is never literally true, R^2 s of such regressions are not quite one, which is why empirical applications of term structure models include measurement error in yields. But with a reasonable choice of n right-hand-side yields (three is sufficient), the variances of measurement errors are tiny relative to the variances of yields. Typical R^2 s

are around 0.999. Monte Carlo analysis of out-of-sample forecasting performance reveals that when models are estimated with maximum likelihood, such small deviations from an exact affine relation are too small to give any advantage to models that impose no-arbitrage.

Some readers have incorrectly interpreted this argument as meaning “no-arbitrage holds so strongly in the data that it need not be imposed.” Instead, the point is that if there is an affine mapping from the yield on one bond to yields on n other bonds, that mapping can be determined without imposing additional restrictions. Whether the mapping is consistent with the formulas of Duffie and Kan is irrelevant. Thus this point also applies to models that impose cross-sectional restrictions that are not derived from no-arbitrage, as in the dynamic Nelson-Siegel model of Diebold and Li (2006). If cross-sectional restrictions affect yield forecasts, it must be because the restrictions are inconsistent with the true cross-sectional properties of yields.

In contrast to cross-sectional properties of yields, time-series dynamics cannot be estimated precisely without imposing some restrictions on these dynamics. For the purpose of forecasting, I advocate imposing the assumption that long-maturity yields, but not short-maturity yields, follow a random walk. When applied to fifty years of Treasury yield data, this assumption produces more reliable out-of-sample forecasts of long-maturity yields than any of a wide variety of forecasting regressions and term structure models. In addition, the assumption improves substantially the estimation precision of other aspects of yield dynamics, such as the interplay between the slope and curvature of the term structure.

An implication of these two points is that a reasonable affine model for predicting yields is one that imposes a random walk on long-maturity yields, but imposes no other restrictions, including those of no-arbitrage. I use a three-factor Gaussian version in horse races of out-of-sample yield forecasting. The competing methods are forecasting regressions using lagged yields and forward rates, the parsimonious Nelson-Siegel models of Diebold and Li (2006) and Christensen et al. (2009), an unrestricted three-factor Gaussian model, and the simple assumption that the entire term structure follows a random walk. The robust result, again based on fifty years of Treasury yield data, is that the model proposed here generates more accurate estimates of future yields.

The next section frames the issues in the context of the existing literature. The third section describes the modeling framework. The fourth and fifth sections describe the methodology and empirical results, respectively, used to evaluate cross-sectional restrictions. The sixth section examines restrictions on yield dynamics and the final section contains concluding remarks.

2 Earlier research

Duffie and Kan (1996), building on the work of Vasicek (1977) and Cox, Ingersoll, and Ross (1985), develop tractable pricing formulas for the affine class of term structure models. Pricing is determined by the equivalent-martingale dynamics of yields. Dai and Singleton (2000) also build on Vasicek and Cox et al. to construct models in which the physical dynamics of yields are affine. Duffee (2002) extends the completely affine setting of Dai and Singleton to the essentially affine framework. Researchers typically use Gaussian essentially affine models when forecasting with the term structure.

A long literature has established that the term structure contains information about both future interest rates and future macroeconomic conditions. Prior to the use of no-arbitrage models, information in bond yields was typically exploited using predictive regressions, vector autoregressions, dynamic factor analysis, and structural macroeconomic models.¹ Duffee (2002) and Christensen et al. (2009) compare the accuracy of interest rate forecasts produced with no-arbitrage affine models to those produced by more standard techniques. Ang, Piazzesi, and Wei (2006) make a similar comparison in forecasting output growth. All note that the Gaussian essentially affine models produce more accurate forecasts than are produced using older techniques.

Because this research compares forecasts produced by non-nested models, the source of the greater accuracy is unclear. No-arbitrage models impose cross-sectional restrictions on the term structure. Duffee (2002) claims "...imposing these [cross-sectional] restrictions should allow us to exploit more of the information in the current term structure, and thus improve forecasts." Similarly, Ang et al. (2006) state that the superior out-of-sample performance of their model is driven in part by these restrictions. But (at least in the former article) this conclusion is motivated more by casual intuition than by either the logic of affine models or empirical analysis.

Restrictions on risk premia may also play a role in forecast accuracy. Gaussian essentially affine models allow the drift of the term structure under the physical measure to be specified separately from the drift under the equivalent-martingale measure. Put differently, risk premia dynamics have an affine structure but are otherwise unconstrained. However, researchers can impose restrictions on risk premia that link drifts under the two measures. Ball and Torous (1996) note that such restrictions allow us to infer physical dynamics from

¹The relevant literature is vast, thus I mention only a few notable contributions. Early examples of predictive regressions using the slope of the term structure are Campbell and Shiller (1991), who forecast future bond yields, and Estrella and Hardouvelis (1991), who forecast business cycles. Cochrane and Piazzesi (2005) estimate predictive regressions using many points on the term structure. Evans and Marshall (1998, 2002) extract information from multiple points on the term structure using both atheoretic and structural VARs. Singleton (1980) is the first application of dynamic factor analysis to the term structure.

covariances among yields, improving estimation precision.

Researchers building no-arbitrage forecasting models typically do not use utility theory to motivate restrictions on risk premia dynamics. Instead, restrictions are based either on estimation tractability, as in Ang and Piazzesi (2003), or sample properties of yields. For example, Duffee (2002) sets to zero any parameters with small t -statistics. Christensen et al. (2009) use the empirical results of Diebold and Li (2006) to motivate restrictions on vector autoregression dynamics of level, slope, and curvature. Cochrane and Piazzesi (2008) assume that variations in risk premia are driven by a single factor, a choice based on the regression evidence of Cochrane and Piazzesi (2005). None of this research attempts to disentangle the effects of the cross-sectional and dynamic restrictions on forecasts.

3 The modeling framework

This section describes a general Gaussian affine term structure model. The model nests Gaussian essentially affine models, thus the cross-sectional restrictions can be imposed and tested statistically. Dynamic restrictions can also be imposed, either with or without imposing cross-sectional restrictions. I use the term “unrestricted model” to refer to this general model.

3.1 The unrestricted model

The term structure is driven by n -dimensional state vector x_t . Its physical measure dynamics are

$$x_{t+1} = \mu + Kx_t + \Sigma\epsilon_{t+1}, \quad \epsilon_{t+1} \sim MVN(0, I). \quad (1)$$

Instead of immediately proceeding to the equivalent-martingale measure, I follow the spirit of the dynamic factor analysis approach in Singleton (1980) by assuming that observed zero-coupon bond yields are affine functions of the state vector plus an idiosyncratic component. Denoting the continuously-compounded yield on an m -maturity zero-coupon bond by $y_t^{(m)}$, yields are

$$y_t^{(m)} = A^{(m)} + B^{(m)'}x_t + \eta_t^{(m)}, \quad \eta_t^{(m)} \sim N(0, \sigma_\eta^2). \quad (2)$$

The idiosyncratic component $\eta_t^{(m)}$ is independent across time and bonds.

I use separate notation for the non-idiosyncratic component of yields. Define

$$\tilde{y}_t^{(m)} = A^{(m)} + B^{(m)'}x_t, \quad (3)$$

where for the moment the yields with tildes are simply one piece of observed yields.

Special notation is used for the one-period bond. Its yield is the short rate r_t and its relation to the state vector is written as

$$r_t = \delta_0 + \delta_1' x_t + \eta_{r,t}, \quad \eta_{r,t} \sim N(0, \sigma_\eta^2). \quad (4)$$

Similarly, \tilde{r}_t is defined as r_t excluding its idiosyncratic component.

It is worth emphasizing that in affine term structure models, the affine relation in (2) between bond yields and the state vector is derived from (4) and the specification of equivalent-martingale dynamics of the state vector. In this unrestricted model, (2) is simply an assumption.

3.2 No-arbitrage cross-sectional restrictions

There are no arbitrage opportunities. But the absence of arbitrage does not restrict yields in (2) unless we assume that equations (1) and (2) capture all of the information relevant to investors about costs and payoffs of Treasury securities. The real world is not so simplistic. These functional forms abstract from both transaction costs and institutional features of the market. For example, owners of on-the-run Treasury bonds usually have the ability to borrow at below-market interest rates in the repurchase market. Certain Treasury securities trade at a premium because they are the cheapest to deliver in fulfillment of futures contract obligations. Treasury debt is more liquid than non-Treasury debt, which is one reason why Treasury bonds are perceived to offer a “convenience yield” to investors in addition to the yield calculated from price. In a nutshell, returns calculated from bond yields do not necessarily correspond to returns realized by investors. Evidence suggests that these market imperfections can have significant effects on observed yields.² The mapping from factors to yields in (2) implicitly assumes that if these effects vary over time, they do so in lockstep with the state vector x_t .

Imposing testable no-arbitrage restrictions requires assuming away (or measuring) these market imperfections. If market imperfections are ruled out, the idiosyncratic term $\eta_{m,t}$ is treated as measurement error. Then $\tilde{y}_t^{(m)}$ denotes a true yield and n factors drive realized returns on all bonds. The absence of arbitrage across the term structure restricts the

²The first academic evidence appears to be Park and Reinganum (1986). Early research focused on prices of securities with remaining maturities of only a few weeks or months. Duffee (1996) contains evidence and references to earlier work. Evidence of market imperfections at longer maturities is in Krishnamurthy (2002), Greenwood and Vayanos (2007), and Krishnamurthy and Vissing-Jorgensen (2007).

coefficients $A^{(m)}$ and $B^{(m)}$ in (2). The stochastic discount factor is

$$M_{t+1} = \exp \left\{ -\tilde{r}_t - \Lambda_t' \epsilon_{t+1} - \frac{1}{2} \Lambda_t' \Lambda_t \right\}. \quad (5)$$

The vector Λ_t is the compensation investors require to face epsilon shocks. Using the discrete-time version of the essentially affine Gaussian framework, the compensation required to face shocks to the state vector has the functional form

$$\Sigma \Lambda_t = \lambda_0 + \lambda_1 x_t. \quad (6)$$

Then under the equivalent martingale measure, the dynamics of x_t are

$$x_{t+1} = \mu^q + K^q x_t + \Sigma \epsilon_{t+1}^q, \quad \epsilon_{t+1}^q \sim MVN(0, I), \quad (7)$$

where

$$\mu^q = \mu - \lambda_0, \quad K^q = K - \lambda_1. \quad (8)$$

Solving recursively using the law of one price, the loadings of a yield on the factors are given by

$$\begin{aligned} B^{(m)'} &= \mathbb{B}(m; \delta_1, K^q)' \\ &= \frac{1}{m} \delta_1' (I - K^q)^{-1} (I - (K^q)^m). \end{aligned} \quad (9)$$

The constant term for $m > 1$ is

$$\begin{aligned} A^{(m)} &= \mathbb{A}(m; \delta_0, \delta_1, \mu^q, K^q, \Sigma) \\ &= \delta_0 + \frac{1}{m} \delta_1' [mI - (I - K^q)^{-1} (I - (K^q)^m)] (I - K^q)^{-1} \mu^q \\ &\quad - \frac{1}{2m} \sum_{i=1}^{m-1} i^2 B^{(i)'} \Sigma_x \Sigma_x' B^{(i)}. \end{aligned} \quad (10)$$

I refer to equations (9) and (10) as the Duffie-Kan restrictions.

The essence of the no-arbitrage restrictions is that in an n -factor model, the mapping from one bond's yield to the n factors can be written in terms of similar mappings for $n + 1$ other “base” bonds. (We need $n + 1$ bonds instead of n because the restrictions are tied to expected excess returns, not expected returns.) By themselves, the Duffie-Kan restrictions do not pin down yields on the base bonds, for the same reason that the Black-Scholes formula takes a stock price as given. The law of one price says that compensation for risk must be the

same across assets—it does not say what that compensation should be. In the math of the n -factor Gaussian model, this corresponds to treating as free parameters each of $\delta_0, \delta_1, \mu^q$, and K^q .

3.3 Dynamic restrictions

Researchers often impose parameter restrictions on the determinants of the price of risk λ_0 and λ_1 in (8). A recent example is Cochrane and Piazzesi (2008). These restrictions allow information about μ^q and K^q from the cross section to help estimate μ and K in (1), as shown by Ball and Torous (1996). An alternative approach in the no-arbitrage framework is to impose restrictions directly on μ or K . Then estimates of μ^q and K^q implicitly determine the dynamics of risk compensation.

Dynamic restrictions do not require the Duffie-Kan restrictions. Restrictions on μ and K can also be imposed on the more general model that does not impose no-arbitrage. Section 6 contains further discussion of these restrictions.

3.4 A macro-finance extension

Following Ang and Piazzesi (2003), a branch of the no-arbitrage term structure literature incorporates macro variables into this type of model. The model described above can be extended by defining a vector z_t of variables such as inflation, output growth, and the unemployment rate. The relation between the macro variables and the state vector is

$$z_t = A_z + B_z x_t + \eta_{z,t}. \tag{11}$$

Adding this affine relation allows us to use the model to forecast future realizations of z_t .

Given the objectives of this paper, there is no reason to include (11). There are no Duffie-Kan restrictions associated with A_z and B_z . Thus if the no-arbitrage restrictions (9) and (10) turn out to be irrelevant for the purposes of forecasting future bond yields, they will also be irrelevant for forecasting future realizations of z_t . Conversely, if imposing the restrictions affects estimated factor loadings of bond yields, the estimated dynamics of x_t are also likely to be affected. In this case, the restrictions will indirectly affect macroeconomic forecasts.

3.5 Discussion

Some of the language used in this section is a little ambiguous. I clarify two terms here. First, “unrestricted model” is a bit of a misnomer. Although not as restrictive as a model

that satisfies the Duffie-Kan restrictions, the unrestricted model imposes strong requirements on the behavior of yields. There are n common factors with Gaussian dynamics, and yields are affine functions of these factors. These common factors pick up all joint variation in yields, including any joint time-variation in convenience yields. The role of the idiosyncratic shock (or measurement error if the Duffie-Kan restrictions are true) is to allow the covariance matrix of observed bond yields to have rank greater than n .

Second, the phrase “term structure model” is a little loose. The restricted model is a model of fixed income. It not only describes the dynamics of zero-coupon bond yields; it can also be used to price all claims contingent on these yields, such as coupon bonds and bond options. Any of these data could be used to estimate the model, and the model can be used to forecast prices of any fixed-income instrument. The unrestricted model is only a model of zero-coupon bond yields. Removing the no-arbitrage restriction generalizes the description of zero-coupon bonds; the cost is an ability to say anything about other fixed-income instruments.

In practice, researchers who use no-arbitrage models for forecasting typically do not apply the models to fixed-income instruments other than zero-coupon bonds. Thus the estimation procedure described in the next section assumes that only zero-coupon bonds are used to estimate both models.

4 Cross-sectional restrictions: Methodology

This section provides the intuition behind the irrelevance of cross-sectional restrictions. It also sets up a formal econometric framework to test, both economically and statistically, the validity of the restrictions. Section 5 applies the econometric framework to Treasury yields.

4.1 A state space setting

To set up the empirical problem, assume that we have a panel of yields on zero-coupon bonds. At dates $t = 1, \dots, T$, we observe a d -vector of yields y_t , where $d > (n + 1)$. This inequality is necessary to produce overidentifying restrictions. The yields on the bonds are fixed over time and are given by m_1, \dots, m_d . Using state-space language, the transition equation of the underlying state is (1) and the measurement equation is

$$y_t = A + Bx_t + \eta_t, \quad \eta_t \sim MVN(0, \sigma_\eta^2 I). \quad (12)$$

In (12), A is a d -vector and B is a $d \times n$ matrix. The transition and measurement equations are underidentified because the state vector is latent. For identification, the vector can be

scaled, rotated, and translated.

Estimation and hypothesis testing are performed with maximum likelihood. The focus here is on testing the null hypothesis that the Duffie-Kan cross-sectional restrictions are correct, against the alternative hypothesis of the general Gaussian affine model. Before getting into the details of estimation, it is helpful to study the intuition of a special case.

4.2 The intuition

For the moment, assume that measurement error η_t in (12) is identically zero. There are two immediate implications of the assumption that all yields are perfectly observed. First, if the Duffie-Kan restrictions are true, imposing them has no empirical implications. Second, if the restrictions are false, we will be able to reject the restrictions empirically with probability one.

An easy way to demonstrate these implications is to apply an identification transformation to the state so that it equals a vector of n yields, y_t^n , all of which are in the observed data.³ Place these n yields at the beginning of the observed yields y_t , denoting the vector of the other $d - n$ yields as y_t^o . The corresponding measurement and transition equations are

$$\begin{pmatrix} y_t^n \\ y_t^o \end{pmatrix} = \begin{pmatrix} 0 \\ A^o \end{pmatrix} + \begin{pmatrix} I \\ B^o \end{pmatrix} y_t^n, \quad (13)$$

$$y_{t+1}^n = \mu + K y_t^n + \Sigma \epsilon_{t+1}. \quad (14)$$

Cross-sectional implications of no-arbitrage affect only the vector A^o and the matrix B^o . In the unrestricted model, these are free parameters. The no-arbitrage pricing formulas (9) and (10) require that they are functions of equivalent-martingale parameters. Dynamic restrictions, in the form of constraints on the parameters of (14), may be present, but are irrelevant to the argument here.

The unrestricted parameter estimates of A^o and B^o are simply coefficients from regressing y_t^o on y_t^n . There is no estimation error because there is no error term. There is nothing to be gained by imposing Duffie-Kan restrictions; we cannot improve on perfect estimation. If the Duffie-Kan restrictions are false, there will be no set of equivalent-martingale parameters consistent with regression coefficients. In other words, (9) and (10) cannot be inverted for all d bonds.

Introducing measurement error weakens both of these implications. But the analysis to

³There is a small caveat to this transformation. It rules out the existence of hidden factors as defined by Duffee (2008). More generally, rotate $n - k$ of the factors into $n - k$ yields, where k is the number of hidden factors.

follow shows that in Treasury data, the magnitude of measurement error is very small. Thus there are no practical implications to imposing Duffie-Kan restrictions when they are true, and it is easy to statistically reject the restrictions when yields deviate from them by only a few basis points.

4.3 The general econometric framework

Under the null hypothesis that the Duffie-Kan restrictions hold, the matrix B and vector A in (12) satisfy the restrictions of (9) and (10) respectively. Formally,

$$\begin{aligned}
 H0 : \quad A &= \mathbb{A}(M; \delta_0, \delta_1, \mu^q, K^q, \Sigma) = \begin{pmatrix} \mathbb{A}(m_1; \cdot) \\ \dots \\ \mathbb{A}(m_d; \cdot) \end{pmatrix}; \\
 B &= \mathbb{B}(M; \delta_1, K^q) = \begin{pmatrix} \mathbb{B}(m_1; \cdot)' \\ \dots \\ \mathbb{B}(m_d; \cdot)' \end{pmatrix}.
 \end{aligned} \tag{15}$$

For estimation purposes, the parameters of the restricted model are those of the physical dynamics (1), the definition of the short rate (4), and the equivalent-martingale dynamics (7). They are stacked in the vector

$$\rho_0 = \{\mu' \quad \text{vec}(K)' \quad \text{vech}(\Sigma)' \quad \delta_0 \quad \delta_1' \quad \mu^q' \quad \text{vec}(K^q)' \quad \sigma_\eta^2\}'. \tag{16}$$

There are $2 + 3n + 2n^2 + n(n - 1)/2$ parameters in (16). Of these, $n + n^2$ are determined by the desired identification transformation.

The alternative hypothesis does not impose cross-sectional restrictions and thus nests the null. The statement of this hypothesis is

$$H1 : \quad A, B \text{ unrestricted.} \tag{17}$$

A likelihood ratio test statistically evaluates $H0$ versus $H1$. The parameters of the unrestricted model are those of (1) and (12), stacked in

$$\rho_1 = \{\mu' \quad \text{vec}(K)' \quad \text{vech}(\Sigma) \quad A' \quad \text{vec}(B)' \quad \sigma_\eta^2\}'. \tag{18}$$

There $1 + 2n + n^2 + (n + 1)d + n(n - 1)/2$ parameters in ρ_1 with $n + n^2$ determined by identification. Thus there are $(1 + n)(d - n - 1)$ overidentifying restrictions to test the cross-sectional null hypothesis. (Recall that the number of observed bond yields d exceeds

$n + 1$.)

Statistical rejection of the null in favor of the alternative can be interpreted in two ways. One interpretation is suggested in Section 3.2. The unrestricted model (1) and (2) holds, but returns computed from Treasury bond prices do not represent the only payoff relevant to investors. Another interpretation is that both models are misspecified. The true model may have more factors, non-Gaussian shocks, or nonaffine dynamics.

4.4 Reparameterizing the alternative hypothesis

Although the unrestricted model nests the cross-sectional restrictions, the parameter vector ρ_1 does not nest ρ_0 . To understand the economics underlying the test of the null hypothesis, we want nested parameters, where a subset are zero under the null and unrestricted under the alternative. Here I transform ρ_1 to a vector that nests the parameters of the null.

We can almost always write the unrestricted parameters A and B in (12) as sums of two pieces. One piece represents parameters consistent with Duffie-Kan, while the other piece represents deviations from the restrictions.

The procedure begins by splitting observed yields into two vectors. The first, denoted y_t^x , is an $(n + 1)$ -vector of yields assumed to satisfy exactly the Duffie-Kan restrictions. (The superscript x denotes eXact.) The second, denoted y_t^v (the v denotes oVer), is a $(d - n - 1)$ vector of yields that provide overidentifying restrictions. The choice of bonds included in the first vector is arbitrary; in particular, they need not be split according to maturity. Stack the corresponding bond maturities in the vectors M^x and M^v . Then rewrite the unrestricted model as

$$\begin{pmatrix} y_t^x \\ y_t^v \end{pmatrix} = \begin{pmatrix} A^x \\ A^v \end{pmatrix} + \begin{pmatrix} B^x \\ B^v \end{pmatrix} x_t + \eta_t, \quad (19)$$

$$A^x = \mathbb{A}(M^x; \delta_0^\dagger, \delta_1^\dagger, \mu^{q^\dagger}, K^{q^\dagger}, \Sigma), \quad (20)$$

$$B^x = \mathbb{B}(M^x; \delta_1^\dagger, K^{q^\dagger}), \quad (21)$$

$$A^v = \mathbb{A}(M^v; \delta_0^\dagger, \delta_1^\dagger, \mu^{q^\dagger}, K^{q^\dagger}, \Sigma) + c_0, \quad (22)$$

$$B^v = \mathbb{B}(M^v; \delta_1^\dagger, K^{q^\dagger}) + C_1. \quad (23)$$

The parameters $\delta_0^\dagger, \delta_1^\dagger, \mu^{q^\dagger}$, and K^{q^\dagger} reconcile the “exact-identification” bond yields with the Duffie-Kan restrictions. The parameters c_0 and C_1 are the deviations of the other bond yields from these restrictions.

To implement this representation, invert the functional form of the $(n + 1) \times n$ matrix

B^x to determine implied equivalent-martingale parameters δ_1^\dagger and K^{q^\dagger} :

$$\{\delta_1^\dagger, K^{q^\dagger}\} = \mathbb{B}^{-1}(B^x; M^x). \quad (24)$$

The inverse mapping in (24) is done numerically. There are values of B^x which cannot be inverted using (24). If inversion is impossible for one set of bonds that comprise the “exact” group, a different set of bonds can be used.⁴ The remaining equivalent-martingale parameters are determined numerically by the inversion

$$\{\delta_0^\dagger, \mu^{q^\dagger}\} = \mathbb{A}^{-1}(A^x; M^x, \delta_1^\dagger, K^{q^\dagger}, \Sigma). \quad (25)$$

After calculating these equivalent-martingale parameters, we can write the parameters A^v and B^v in (21) and (22) as the sum of parameters implied by Duffie-Kan and the error terms c_0 and C_1 . The vector c_0 is the average yield error for the overidentified bonds and the matrix C_1 is the error in the factor loadings. Thus we can transform the parameters of the unrestricted model from (18) to

$$\rho_1^\dagger = \{\mu, K, \Sigma, \delta_0^\dagger, \delta_1^\dagger, \mu^{q^\dagger}, K^{q^\dagger}, c_0, C_1, \sigma_\eta^2\}. \quad (26)$$

The null hypothesis is that both c_0 and C_1 are zero.

Writing the alternative hypothesis in this way does not require that only the overidentified yields are potentially contaminated by convenience yield effects. All yields may be contaminated. This version of the model simply says that *if* the Duffie-Kan restrictions can be imposed, any $d - n - 1$ yields must be set consistently with the other $n + 1$ yields.

Using this reparameterization, we can restate the conclusions for the case of zero measurement error. In the absence of measurement error, the standard errors on c_0 and C_1 are zero. Under the null hypothesis that the Duffie-Kan restrictions hold, the values are also zero. If, however, yields are contaminated by convenience yield effects, c_0 and C_1 are not identically zero, and a t -statistic on at least one of these parameters is infinite.

5 Cross-sectional restrictions: Evidence

When yields are measured with negligible error, cross-sectional relations among yields are inferred precisely without relying on no-arbitrage restrictions. But if measurement error is

⁴In rare circumstances, there is no set of bonds for which this inversion is possible. For example, consider a one-factor model estimated using data on three bonds. The unrestricted model has scalar B 's for each of the three bonds. If the estimated B 's are positive, zero, and negative respectively, then inversion is impossible regardless of which two bonds are placed in the “exact” group.

large and sample sizes are small, cross-sectional restrictions (if true) are likely to noticeably improve estimation efficiency. The main question asked in this section is whether, in practice, there is enough measurement error to matter.

The main tool used to answer this question is Monte Carlo simulation. Take a parameterized model that satisfies the Duffie-Kan restrictions. Randomly generate a panel of yields from the model, with d yields at each of $T + k$ time series observations. Use the first T observations to estimate two versions of the model: one that imposes the cross-sectional restrictions and one that does not. Then use the two estimated models to forecast the d yields from one to k periods ahead. Save the forecast errors, then repeat the process. A comparison of root mean squared forecast errors reveals the marginal contribution of the cross-sectional restrictions.

The main conclusion of this section is that the contribution of the restrictions is miniscule. To help convince skeptics, I attempt to tilt the playing field in the direction of a large contribution. Parameter restrictions are more likely to play an important role in estimation when using a small sample. I therefore use small values of d and T ; smaller than usually employed in empirical work. There are six yields observed at each of 88 observations.

The parameterized no-arbitrage model used to generate the simulations is a three-factor model estimated on quarterly Treasury yields from 1985 through 2006. To focus exclusively on the role of cross-sectional restrictions, no dynamic restrictions are imposed on the estimated model. The choice of data sample is somewhat arbitrary. Its length is the same used in the simulations. This choice economizes on Monte Carlo computing time. I compute standard errors of parameter estimates with Monte Carlo simulation. Since the data samples are the same, I can use the same set of simulations to construct standard errors and to evaluate forecast accuracy. Section 5.3 briefly discusses the robustness of the results to the use of a much longer sample of Treasury yields.

I also estimate the unrestricted model, without cross-sectional or dynamic restrictions, using the same sample of Treasury yields. By comparing the two estimated models, we can evaluate empirically the Duffie-Kan restrictions. The precise question is whether the Treasury term structure during the period 1985 through 2006 is consistent with the Duffie-Kan restrictions. The conclusion is that Treasury yields deviate from the restrictions by only a few basis points, but the standard errors for cross-sectional restrictions are so tight that the restrictions are overwhelmingly statistically rejected.

5.1 Data description

The data are yields on zero-coupon Treasury bonds with maturities of three months and one through five years. There are six bond yields observed at each of 88 quarterly observations from 1985Q1 through 2006Q4. All data are from the Center for Research in Security Prices (CRSP). Because the model specifies the length of a period as one unit of time, model estimation uses continuously compounded rates per quarter. When discussing estimation results, I typically refer to the model's implications for annualized yields.

5.2 Level, slope, and curvature factors

To help provide some intuition for the results, I normalize the factors to versions of level, slope, and curvature. Level is measured by the five-year yield, slope by the five-year yield less the three-month yield, and curvature by the two-year yield less the average of the three-month and five-year yields. All of these factors are demeaned. Formally, the state vector is

$$x_t = \begin{pmatrix} \tilde{y}_t^{(20)} - E\tilde{y}_t^{(20)} \\ \left(\tilde{y}_t^{(20)} - E\tilde{y}_t^{(20)} \right) - \left(\tilde{y}_t^{(1)} - E\tilde{y}_t^{(1)} \right) \\ \left(\tilde{y}_t^{(8)} - E\tilde{y}_t^{(8)} \right) - \frac{1}{2} \left(\left(\tilde{y}_t^{(1)} - E\tilde{y}_t^{(1)} \right) + \left(\tilde{y}_t^{(20)} - E\tilde{y}_t^{(20)} \right) \right) \end{pmatrix}. \quad (27)$$

In (27), the yields are 'true' yields instead of measured yields, and demeaning uses model-implied unconditional expectations instead of sample means. This normalization sets the vector μ in the physical dynamics (1) to zero.

Section 4.4 describes how to express the unrestricted model in the form of deviations from no-arbitrage restrictions. This three-factor model requires four cross-sectional points on the yield curve to pin down the equivalent-martingale parameters. I use the three-month, one-year, three-year, and five-year bonds to identify the equivalent-martingale measure. Deviations from Duffie-Kan restrictions are allowed in the two-year and four-year bond yields.

5.3 A preliminary look at bond yields

The assumption of three latent factors says that all yields are affine functions of the level, slope, and curvature, plus noise. These functions can be approximated by replacing the latent factors in (27) with their observable counterparts. For each maturity m , the approximate

function is

$$y_t^{(m)} = a_m + b'_m \left(\begin{array}{c} y_t^{(20)} - \overline{y_t^{(20)}} \\ \left(y_t^{(20)} - \overline{y_t^{(20)}} \right) - \left(y_t^{(1)} - \overline{y_t^{(1)}} \right) \\ \left(y_t^{(8)} - \overline{y_t^{(8)}} \right) - \frac{1}{2} \left(\left(y_t^{(1)} - \overline{y_t^{(1)}} \right) + \left(y_t^{(20)} - \overline{y_t^{(20)}} \right) \right) \end{array} \right) + e_t^{(m)} \quad (28)$$

where the bars indicate sample means. We can think of (28) as a regression equation. Estimates of the coefficients a_m and b_m will be biased because of an errors-in-variables problem owing to measurement error.

Panel A of Table 1 reports summary statistics for the observable version of the factors. Panel B reports OLS estimation results of applying (28) to the one-year, three-year, and four-year bond yields. The three factors explain almost all of the variation in the dependent yields. The adjusted R^2 s range from 0.998 to 0.999. The standard errors of the point estimates are correspondingly small. The estimated factor loadings range from around one to minus one (a consequence of the definition of the factors). The standard errors for level and slope range from 0.004 to 0.011. The standard errors for curvature are somewhat higher because, as seen in Panel A, curvature contributes relatively little to the variation in yields.

These regression results foreshadow what we will see in Section 5.5. Imposing cross-equation restrictions on factor loadings is of no practical importance under the assumption that the restrictions are correct, because the standard errors are so small. One potential criticism of these results is that the CRSP zero-coupon bond yields are constructed from coupon bond yields by filtering outliers from the data. The filtering procedure probably reduces slightly the standard error of the residual. Thus the forecasting exercise studied here should be thought of as forecasting with zero-coupon bond yields that are inferred and smoothed from coupon bond yields.

Another potential criticism is that the sample period studied here may be unusual; the high R^2 s may not be informative about the population properties of yields. However, these R^2 s appear to be more the norm than the exception. For example, if the sample period for the regressions is extended back to 1952Q2 (the first quarter for which the CRSP longer-horizon yields are available), the corresponding adjusted R^2 s range from 0.997 to 0.999. Almost identical results are obtained when using the Federal Reserve Board's zero-coupon bond yields for maturities up to ten years. (These results are not reported in any table.) Thus the sample period here seems representative from the perspective of the cross-sectional explanatory power of a three-factor model.

5.4 Estimation results

Table 2 reports parameter estimates based on the level, slope, and curvature representation of the factors in (27).⁵ Although there are 23 and 31 free parameters in the restricted and unrestricted models, the table reports 29 and 37 parameter estimates respectively. The rotation into level, slope, and curvature pins down the factor loadings for the three-month, two-year, and five-year bond yields. These fixed loadings are six nonlinear restrictions on the reported parameter estimates. Thus the covariance matrix of the reported estimates is singular. Standard errors, in parentheses, are based on 1000 Monte Carlo simulations.⁶

The results are discussed in detail below, but can be summarized in three main points. First, deviations from Duffie-Kan are economically tiny in the unrestricted model. Second, notwithstanding the first point, the Duffie-Kan restrictions are overwhelmingly rejected statistically. Third, imposing the restrictions has an economically trivial effect on the estimation precision of the parameters.

5.4.1 The economic importance of the restrictions

The vector c_0 and the matrix C_1 of the unrestricted model capture deviations from Duffie-Kan restrictions. The estimate of c_0 in Table 2 implies that mean yields on the two-year and four-year bonds deviate from the restricted means by two to three basis points of annualized yields.⁷ Deviations in factor loadings are economically even smaller. Visual evidence is in Fig. 1. The circles are the means and loadings of the three-month, one-year, three-year, and five-year bonds yields. The lines are drawn by calculating the equivalent-martingale parameters consistent with the circles. The dots are the means and loadings of the two-year and four-year bond yields. The parameters c_0 and C_1 equal the differences between the lines and the dots. They are almost undetectable in the figure.

Another way to judge the economic importance of the Duffie-Kan restrictions is to calculate, for each quarter in the sample, the fitted deviation

$$\text{fitted deviation}_t = c_0 + C_1 \hat{x}_t. \tag{29}$$

⁵Details of the estimation procedure are contained in the Appendix.

⁶Parameter estimates for each simulation of the unrestricted model are transformed into the set of parameters corresponding to (19). This transformation could not be performed for 40 of the 1000 simulations. In other words, for 40 of the simulations, no set of equivalent-martingale parameters could reconcile the behavior of the three-month, one-year, three-year, and five-year bonds with the Duffie-Kan restrictions. The standard errors for the unrestricted model in Table 2 are based only on the 960 observations for which the inversion was successful. The evidence in Footnotes 9 and 10 indicates that this discrepancy does not have a material effect on the standard errors.

⁷The units of c_0 are decimal points per quarter. The estimates reported in Table 2 are multiplied by 10^4 to put them in basis points per quarter. For example, the deviation of the mean four-year yield from the Duffie-Kan restricted mean is 0.699 basis points per quarter, or 2.8 basis points per year.

In (29), \hat{x}_t represents the filtered values of the state vector. Across the 88 quarters in the sample, absolute fitted deviations never exceed seven basis points of annualized yields for either the two-year or four-year bonds. These deviations are within the range of microstructure-induced effects on yields.

This analysis based on c_0 and C_1 is based entirely on the estimates of the unrestricted model. The same message is conveyed by comparing estimates of the two models. The two sets of parameters in Table 2 are almost identical. Visual evidence is in Fig. 2. The circles are the unrestricted mean yields and factor loadings. The solid lines are mean yield and slope functions from the estimated restricted model. The estimated mean yields for the unrestricted model lie on the estimated mean term structure for the restricted model. Similarly, the estimated factor loadings coincide. The chosen factor rotation implies that by definition, the loadings of the two models match at maturities of three months, two years, and five years. These points are marked with an x . Yet even for the factor loadings not marked with an x , the unrestricted loadings are indistinguishable from the loading functions of the restricted model.

5.4.2 The statistical importance of the restrictions

The likelihood ratio test statistic of the Duffie-Kan restrictions is 35.64, which strongly rejects the null hypothesis.⁸ The source of the rejection is largely the mean yields. The standard errors on the two elements of c_0 are a basis point or less of annualized yield. The standard errors of C_1 are also quite small, but the individual t -statistics are typically less than two in absolute value. Two-standard-error bounds on the estimates of c_0 and C_1 are displayed in red in Fig. 1.

The tight standard errors on c_0 and C_1 may be surprising, especially since a simple comparison of 88 observations to 31 free parameters in the unrestricted model suggests the standard errors will be large. But c_0 and C_1 are roughly coefficients of a cross-sectional regression of yields on yields. The standard errors on c_0 (rescaled to percent at an annual horizon) and C_1 are similar to those of the OLS regression coefficients reported in Panel B of Table 1. Cross-sectional deviations from a three-factor model are tiny, thus standard errors for cross-sectional regressions are tiny.

5.4.3 The effect of the restrictions on estimation precision

A quick comparison of the two sets of standard errors in Table 2 reveals that for most of the parameters, estimation precision is largely unaffected by the imposition of the Duffie-

⁸The asymptotic 95 percent critical value is 15.51. The finite-sample critical value is similar, as discussed in the appendix.

Kan restrictions. Standard errors for parameters identified by the physical measure (the mean short rate, K , Σ , and σ_η) are almost identical. For example, the standard error of the unconditional mean of the annualized short rate is 1.40 percent for the restricted model and 1.39 percent for the unrestricted model. (These are calculated by multiplying the reported standard errors in Table 2 by four to express them as annualized yields.) Standard errors of most of the parameters identified only by the equivalent-martingale measure (μ^Q and K^Q) are smaller when no-arbitrage is imposed than when it is not imposed. However, differences across these standard errors are tiny except for standard errors of parameters related to the curvature factor. These parameters are the third element of μ^Q and the third column of K^Q . Recall that the curvature factor plays a very small role in overall term structure dynamics.

From an economic perspective, it is more meaningful to consider estimated properties of yields rather than individual parameter estimates. Here I focus on unconditional means and factor loadings. Unconditional mean yields are determined by the mean short rate δ_0 and the equivalent-martingale dynamics of the state vector. Standard errors of the estimated unconditional means are almost identical across the two models. For example, the standard error of the unconditional mean of the four-year annualized bond yield is 1.55 percent for the restricted model and 1.54 percent for the unrestricted model.⁹ Standard errors of yield loadings on factors are close to zero for both models. Consider, for example, the four-year bond yield. The restricted model's standard errors of the yield's loadings on level, slope, and curvature are 0.0024, 0.0033, and 0.0081. The standard errors for the unrestricted model are 0.0062, 0.0091, and 0.0454 respectively.¹⁰

5.5 Out-of-sample forecasting

We are now in a position to examine the loss in forecast accuracy when Duffie-Kan restrictions are true but are not imposed in estimation. The magnitude of the loss necessarily depends on both the true model and the sample size. The true model used here is the restricted model estimated in Section 5.4. The sample size is 88 quarterly observations of yields on maturities with three months and one through five years. If forecast accuracy is not improved in such a small sample, we can confidently rule out the possibility that the restrictions are useful in sample sizes more commonly used in empirical work.

A single Monte Carlo simulation proceeds in three steps. First, 100 quarters of yields are

⁹ These standard errors cannot be read off of Table 2. They are the standard deviations, across the Monte Carlo simulations, of the population mean of the four-year bond yield implied by each simulation's parameter estimates. The value of 1.54 is based on the 960 simulations discussed in Footnote 6. The corresponding standard error using all 1000 simulations is also 1.54.

¹⁰ The standard errors for the unrestricted model are based on the 960 simulations discussed in Footnote 6. The corresponding standard errors for all 1000 simulations are 0.0063, 0.0091 and 0.0454.

generated with the true model. The first observation is drawn from the unconditional distribution of yields. All other observations are drawn from the conditional distribution given by the transition and measurement equations. Second, the restricted and unrestricted models are estimated with maximum likelihood using the first 88 quarters of data. Third, the estimated models are used to calculate out-of-sample forecasts of the three-month, two-year, and five-year bond yields at horizons of one through twelve quarters. These are transformed into forecasts of “level” (five-year yield), “slope” (five-year less three-month), and “curvature” (two-year less average of five-year and three-month). Forecast errors are then calculated using the final 12 observations of the sample. After generating 1,000 simulations, root mean squared forecast errors are constructed across the Monte Carlo simulations for each forecast horizon and forecasted variable.

I first ask whether the models produce similar forecasts. For each forecast horizon and variable, I construct the difference between the forecast of the restricted model and the forecast of the unrestricted model. Table 3 reports the square roots of the mean squared differences. Across 1000 simulations, the restricted and unrestricted models produce similar forecasts. For example, the table reports that at the twelve-quarter-ahead horizon, the root mean squared difference between the restricted model’s estimate of the level and the unrestricted model’s estimate is 11 basis points (annualized). Root mean squared differences for forecasts of the slope and curvature are four and two basis points respectively.

Table 4 reports the root mean squared forecast errors. The results are easily summarized. The choice of whether to impose Duffie-Kan restrictions is irrelevant to forecast accuracy. Regardless of the forecast horizon and forecasted variable, the RMSEs of the restricted and unrestricted models differ by no more than a third of a basis point of annualized yield.

These results necessarily depend on the sample size and the true model. In particular, they depend on how well the cross-section of yields lines up with a three-factor representation. As discussed in Section 5.3, the R^2 s from three-factor cross-sectional regressions for the data sample underlying the “true” model here are similar to those for other samples. Thus the simulation evidence presented here appears to be a robust feature of the term structure.

6 The role of dynamic restrictions

The previous two sections argue that for the purpose of forecasting, there is no reason to impose restrictions on the *cross-section* of yields. But restrictions on yield *dynamics* can be valuable. This section argues that forecast accuracy is improved by requiring that the level of yields follows a random walk, while slope and curvature follow unrestricted stationary processes. The resulting model dominates commonly used forecasting regressions

and dynamic term structure models. A brief discussion of earlier empirical analysis helps to put this argument in context.

6.1 Existing evidence on yield dynamics

There is not enough information in the time series of Treasury yields to estimate term structure dynamics with high precision. The main difficulty is that the dynamics are close, both economically and statistically, to nonstationary. The survey of Martin et al. (1996) concludes that the level of yields appears to have a unit root, while spreads between yields of different maturities are stationary.

By restricting the functional form of risk compensation, no-arbitrage term structure models can use information in the cross section to help estimate dynamics. But it is not clear how to choose a reasonable model of risk compensation. Duffee (2002; henceforth DU2002) finds that the functional forms of risk compensation in the entire class of completely affine term structure models are incompatible with the empirical behavior of yields. They counterfactually imply that when the slope of the yield curve is steep, long-maturity yields are expected to rise. DU2002 sets a forecasting benchmark by showing that the assumption that yields follow a random walk produces more accurate forecasts of future yields.

Essentially affine models perform better. In fact, DU2002 concludes that a three-factor essentially affine Gaussian model requires no restrictions on risk compensation in order to forecast more accurately than a random walk. The model's forecasts are also more accurate than forecasts produced by regressing future changes in yields on the slope of the term structure. In other words, the no-arbitrage model studied in Section 5 is a valuable forecasting tool, even without imposing restrictions on its dynamics. This evidence is based on out-of-sample forecasts for 1995 through 1998.

An alternative approach to increasing estimation precision is to imposing restrictions directly on term structure dynamics, bypassing restrictions on risk compensation. Diebold and Li (2006; henceforth DL2006) build a dynamic version of the term structure introduced by Nelson and Siegel (1987). They restrict the level, slope, and curvature of the yield curve to follow univariate AR(1) processes and find that the resulting model is more accurate than both a random walk and the three-factor Gaussian model of DU2002. Their out-of-sample forecasts are produced for January 1994 through December 2000. Similar restrictions are imposed on the no-arbitrage modified Nelson-Siegel model developed in Christensen et al. (2009).

6.2 A broader look at forecast accuracy

This section makes three empirical points. First, the assumption that long-maturity yields follow a random walk produces more accurate out-of-sample forecasts of these yields than both of the models advocated in DU2002 and DL2006. The contradictory conclusions in this earlier work are not robust to alternative sample periods. Standard forecasting regressions are also inferior to the random walk model.

Second, both the slope and curvature of the term structure contain predictable components. Term structure models and forecasting regressions outperform the assumption of a random walk in out-of-sample forecasts of slope and curvature. Third, a dynamic model's forecasts of slope and curvature tend to be improved by imposing the restriction that the level of the term structure follows a random walk.

6.2.1 The horse races

A total of seven forecasting tools are studied here. They are all applied to monthly observations of Treasury yields.

1. Random walk

This model assumes that yields at all maturities follow a random walk.

$$E_t y_{t+j}^{(m)} = y_t^{(m)}$$

2. AR(1) regressions

The j -period ahead yield is predicted using its lag.

$$y_{t+j}^{(m)} = b_{0,m,j} + b_{1,m,j} y_t^{(m)} + \epsilon_{t+j}^{(m)}$$

3. Slope regressions

The j -period ahead yield is predicted using the slope of the term structure, as measured by the difference between the five-year and three-month yields.

$$y_{t+j}^{(m)} = b_{0,m,j} + b_{1,m,j} \text{slope}_t + \epsilon_{t+j}^{(m)}$$

4. Five forward rates

The j -period ahead yield is predicted using the five annual forward rates popularized by Cochrane and Piazzesi (2005). Denote the period- t forward rate from year y_1 to

year y_2 by $F_{y_1, y_2, t}$.

$$y_{t+j}^{(m)} = b_{0,m,j} + \sum_{i=1}^5 b_{i,m,j} F_{i-1,i,t} + \epsilon_{t+j}^{(m)}$$

5. Dynamic Nelson-Siegel model

DL2006 estimate their model both in unrestricted form, which allows VAR(1) dynamics for level, slope, and curvature, and in restricted form, imposing AR(1) dynamics on each factor. They advocate the latter for out-of-sample forecasting, and I use this restricted form here. I also estimate the no-arbitrage Nelson-Siegel model of Christensen et al. (2009) with the same set of restrictions. Results for this model are very close to those of the DL2006 model, and thus are not discussed further.

6. Unrestricted three-factor affine Gaussian model

I estimate the unrestricted model from Section 5. Recall that for this model, “unrestricted” refers to cross-sectional restrictions. Here this term also implies unrestricted dynamics, to differentiate it from the next forecasting tool. Aside from the lack of cross-sectional restrictions, this model is identical to the forecasting model in DU2002.

7. Constrained three-factor affine Gaussian model

The model from Section 5 is estimated with dynamic restrictions. Using the level, slope, and curvature factor normalization of (27), this model has a feedback matrix K in (1) with the form

$$K = \begin{pmatrix} 0.9999 & 0 & 0 \\ 0 & k_{22} & k_{23} \\ 0 & k_{32} & k_{33} \end{pmatrix}. \quad (30)$$

The first row implies that the level of yields is approximately a random walk. The use of 0.9999 instead of 1.0 simplifies the software code. With this value at a monthly frequency, the half-life of a shock to the level is about 580 years. The zeros in the first column prevent the (near) random walk in the level from inducing (near) nonstationarity in slope and curvature.

The data are month-end observations of yields on zero-coupon bonds with maturities of three months and one through five years. The sample is June 1952 through December 2008. Each regression and model is estimated on rolling samples of T months. The regressions are estimated with ordinary least-squares. The dynamic models are estimated with the Kalman filter.

The fitted regressions and models are used to predict yields on these same bonds, at forecast horizons of three and twelve months. For example, using ten years of data in estimation ($T=120$), the first rolling sample is June 1952 through May 1972. Information as of the end of May 1972 is then used to predict yields at month-end August 1972 and May 1973. The final rolling sample is January 1998 through December 2007. The results of that sample are used to predict yields as of March 2008 and December 2008.

There are a total of $667 - T$ rolling samples and associated forecast errors for each bond yield. I do not calculate root mean squared forecast errors for each bond, because these errors are highly correlated across bonds. Instead, I compute RMSEs for level (the five-year yield), slope (the five-year yield less the three-month yield), and curvature (the two-year yield less the average of the three-month and five-year yields).¹¹ Results are displayed in Table 5 for $T = 120$ and $T = 480$. Results for intermediate choices of T are similar and not shown.

Perhaps the most obvious message to take from the table is that forecasts using forty years of data are more accurate than forecasts using ten years of data. Although part of the difference may be attributable to greater estimation precision, most is driven by the periods over which the forecasts are produced. With ten years of data, out-of-sample forecasts begin in the late 1970s and include the gyrations of the Fed’s monetarist experiment. Models estimated with forty years of data produce forecasts beginning in 1992.

6.2.2 Forecasting the level of the term structure

First consider the columns labeled “Level.” The random walk assumption dominates the other forecasting tools. At the twelve-month horizon, there are clear qualitative differences in forecast accuracy. The regressions using five forward rates are overparameterized, producing the worst forecasts. When ten years of monthly data are used to estimate this regression, the RMSE is 50 basis points above the random walk RMSE. The DL2006 model, with its parsimonious dynamics, produces more accurate forecasts than the unrestricted three-factor model; the RMSEs differ by a few basis points. The constrained three-factor model, by design, produces forecasts almost identical to the random walk model. Any differences between the random walk and constrained forecasts are due to filtering. The three-factor model forms forecasts using a filtered value of the level instead of the observed level.

The superiority of the random walk model documented here runs counter to the evidence of DU2002 and DL2006. However, the out-of-sample periods studied in this earlier work

¹¹This may introduce some ambiguity into the AR(1) forecasting tool. To be precise, I forecast the individual yields with AR(1) regressions instead of forecasting slope and curvature with AR(1) regressions. The forecasts of slope and curvature are the appropriate linear combinations of the forecasts of the individual yields.

are quite short. I briefly summarize an analysis similar to that of DU2002 to illuminate the sensitivity of those results to the sample period. To best match my procedure with that of DU2002, I estimate the three-factor unrestricted model over the period June 1952 through t , where t ranges from December 1994 to December 2007. These expanding samples are used to generate out-of-sample forecasts of the five-year yield twelve months ahead. For the first three years of out-of-sample forecasts (1995 through 1997), the RMSE of the level is 92 basis points with the random walk model and only 78 basis points for the three-factor model. This period corresponds to the period studied in DU2002, leading to the conclusion that the dynamic model outperforms the random walk.

We now have ten years of data unavailable at the writing of DU2002. Including these data reverses the RMSE comparison. For the full 13-year period, the RMSEs for the random walk and unconstrained models are 103 and 105 basis points respectively. Similarly, the DL2006 result breaks down with a longer sample. Again to match my results to DL2006, I estimate the DL2006 model over the period January 1985 through t , where t ranges from December 1993 through December 2007. I then construct twelve-month-ahead forecasts and forecast errors for the five-year yield through December 2000. Over this period studied by DL, the RMSE of the level is 115 basis points with the random walk model and only 108 basis points for the DL model. But when an additional eight years of data are included, the ordering of the RMSEs is reversed.

6.2.3 Forecasting slope and curvature

Since both the slope and curvature of the term structure follow stationary processes, the assumption that they follow random walks produces relatively high RMSEs. When ten years of data are used to estimate the regressions and models, the constrained three-factor model is unambiguously more accurate than the other six tools. From an economic perspective, its advantage over some of the other tools is modest. In particular, the unrestricted three-factor model has RMSEs all within three basis points of the constrained model.

The superiority of the constrained model also holds when estimating with forty years of data. But at this horizon, regressions using five forward rates, as in Cochrane and Piazzesi (2005), is the closest competitor. The RMSEs of these two tools are almost identical.

The constrained model's form of K in (30) has two effects on forecasts of slope and curvature. The direct effect is that the level of the term structure cannot feed back into expectations of future slope and curvature. The indirect effect is that the sampling uncertainty in the free elements of K is reduced. The evidence is in Panel A of Table 6, which displays means and standard deviations of parameter estimates for the unrestricted and constrained three-factor models. The sample is the 547 sets of parameters estimated using rolling sam-

ples of 120 monthly observations. Note that standard deviations are in parentheses, not standard errors of the means.

For the free parameters that they have in common, the estimates from the constrained model are less sensitive to the particular data sample. The standard deviations of the unrestricted parameters are around 1.5 times the standard deviations of the constrained parameters.

The economic implications of these estimates boil down to the effect of level, slope, and curvature on expected future yields. At the j -month horizon, this effect is summarized by K^j . Panel B reports means and standard deviations for the twelve-month horizon. The two sets of mean estimates differ substantially in the forecasting implications of the level of the term structure. The mean estimate for the unrestricted model implies that close to half of a shock to the level dies out within twelve months. By contrast, they largely agree about the forecasting implications of the slope and curvature. As in Panel A, the estimates from the constrained model typically are much more precise than those from the unconstrained model.

6.3 Summarizing dynamic properties

Among the forecasting tools studied here, the constrained three-factor model, which does not impose no-arbitrage restrictions, is the single best choice from the perspective of out-of-sample forecasting over the past fifty years. Two limitations are built into this statement. First, there is certainly some other dynamic model that produces more accurate forecasts; we just do not know what it is. For example, the evidence in Section 6.2.2 does not imply that the assumption of a random walk is the best possible model for forecasting long-term bond yields. A model in which long-term yields are nonstationary yet predictable may be developed that performs better than the random walk model. However, until this model appears, the random walk benchmark remains the one to beat.

Second, there is some evidence that combinations of different forecasting tools can perform must as well as the constrained three-factor model. For example, when using forty years of data to estimate the models, the constrained model is no better than the combination of a random walk description of long-maturity yields and regression forecasts of slope and curvature using five forward rates.

7 Conclusion

Dynamic no-arbitrage term structure models have long been recognized as powerful tools for cross-sectional asset pricing. For example, they allow us to price exotic term structure instruments given the properties of standard instruments. But they have nothing special to offer from the perspective of forecasting. The reason is that Treasury yield dynamics are consistent with an affine factor model. Cross-sectional relations among yields in an affine model are easy to infer from yield data without imposing *a priori* restrictions, whether the restrictions are those of no-arbitrage or some alternative model. Thus for the purpose of forecasting, cross-sectional restrictions are unneeded.

By contrast, restrictions on dynamics can improve forecast accuracy. But the most important restriction from an empirical perspective is not one derived from no-arbitrage. Instead, it is the assumption that long-maturity yields follow a random walk. A Gaussian three-factor model that satisfies this restriction and is otherwise unconstrained produces out-of-sample forecasts that are more accurate than forecasts produced by other standard tools.

8 Estimation appendix

Note: This appendix is likely to be put on-line instead of in the paper.

8.1 Normalizations

Factor rotations are arbitrary. I use one factor rotation to estimate the model, then rotate the factors (and transform the parameters accordingly) when discussing the results. For estimation, the elements of the state vector are rotated so each follows a univariate autoregressive process with an unconditional mean of zero. The innovations of the processes are correlated. The identified transition equation is

$$x_{t+1} = Dx_t + \Sigma\epsilon_{t+1} \quad (31)$$

where D is diagonal and Σ is lower triangular with ones along the diagonal. An additional normalization orders the diagonal of D , but I do not apply this in estimation.

A three-factor unrestricted model applied to d bond yields has $7 + 4d$ free parameters in

$$\rho_1 = \{\text{diag}(D)' \ \Sigma_{21} \ \Sigma_{31} \ \Sigma_{32} \ A' \ \text{vec}(B)' \ \sigma_\eta^2\}'.$$

The restricted model has 23 free parameters in

$$\rho_0 = \{\delta_0 \ \delta_1' \ \text{diag}(D)' \ \Sigma_{21} \ \Sigma_{31} \ \Sigma_{32} \ \mu^{q'} \ \text{vec}(K^q)' \ \sigma_\eta^2\}'.$$

8.2 Optimization overview

The likelihood functions are maximized using Intel Fortran calling IMSL numerical optimization routines. The shapes of the likelihood functions determine the choice of technique used to compute standard errors. A close look at the functions (not detailed here) reveals that they are locally quadratic in only tiny neighborhoods around the optimal parameter estimates. Thus for many of the parameters, asymptotic standard errors are inappropriate. I estimate standard errors with Monte Carlo simulations. Asymptotic and Monte Carlo standard errors differ substantially. The former are typically much larger. For example, for the restricted model, asymptotic standard errors of individual parameter estimates are up to 300 times the corresponding Monte Carlo standard errors.

Early versions of this paper were written based on Matlab optimization routines. Unfortunately, Matlab performs relatively poorly in this setting, as discussed in Section 8.6.

8.3 Optimization algorithm

Estimation of the models on actual Treasury data requires a fairly elaborate hands-on procedure. A simpler procedure is used when the models are applied to simulated data. Here I describe the former procedure.

1. Choose a initial parameter vector of starting values based on OLS estimation. The procedure is described in Section 8.4.
2. Given a starting parameter vector, estimate the parameters with Simplex using 5000 iterations. A derivative-based optimizer with analytic first derivatives refines the parameter estimates.

Researchers often use derivative-based optimizers in combination with numerical approximations to first derivatives. However, extensive experiments (not detailed here) reveal that such an approach does not work well in this setting. Along particular dimensions of the likelihood surface, numerical imprecision creates large errors in these approximations in the neighborhood of local optima.

3. Repeat Step 2 many times, using starting values that are drawn from a multivariate normal distribution with a mean given by the vector from Step 1.

Step 3 creates a sequence of independently-drawn local optima. I terminate Step 3 when I am confident that the highest value in this sequence corresponds to the global optimum. In practice, this is when the highest value has been reached many times from different starting values. For the restricted model, this subjective termination point is reached after 100 repetitions. For the unrestricted model, it is reached after 50 repetitions.

4. Start from the parameter vector with the highest likelihood among these repetitions. Using this as a starting value, repeat Step 2.

It is worth noting that a single application of Step 2 for the unrestricted model takes about half the time necessary for the restricted model. Estimation of the unrestricted model is also better behaved, which is why fewer repetitions are necessary in Step 3.

8.4 Starting values

To determine starting values using Treasury data, I estimate a VAR(1) using yields on three of the bonds. Denoting the estimated VAR as

$$s_t = b_0 + b_1 s_{t-1} + e_t, \quad e_t \sim N(0, \Omega),$$

diagonalize the matrix b_1 into

$$b_1 = \hat{P}\hat{D}\hat{P}'$$

where \hat{D} is a diagonal matrix of eigenvalues of b_1 and the columns of \hat{P} are the eigenvectors of b_1 . Define $\hat{\Sigma}$ as a Cholesky decomposition,

$$\hat{\Sigma}\hat{\Sigma}' = \hat{P}'\Omega\hat{P}.$$

The matrices \hat{D} and $\hat{\Sigma}$ are the starting values for D and Σ in the transition equation. Fitted values of the latent factors are

$$P's_t = \hat{x}_t.$$

For the unrestricted model, the starting value for the constant vector in the measurement equation is the sample mean of the bond yield vector. Starting values for each bond's factor loadings in the unrestricted model are coefficients of regressions of the bond's yield on \hat{x}_t .

For the restricted model, the starting value for δ_0 is the sample mean of the short rate. The starting value for δ_1 is the coefficient vector from a regression of the three-month yield on \hat{x}_t . The starting value for μ^q is zero and the starting value for K^q is a diagonal matrix with 0.8, 0.6, and 0.4 along the diagonal. The starting value for the standard deviation of measurement error is 10 basis points.

8.5 Factor rotation into level, slope, and curvature

Given parameters of the model using the normalization (31), the following procedure rotates the model into one where the factors are demeaned level, slope, and curvature. This description assumes quarterly observations are used.

Starting with the measurement equation

$$y_t = A + Bx_t + \eta_t, \quad \eta_t \sim MVN(0, \sigma_\eta^2 I)$$

and the normalized transition equation (31), pick out the factor loadings for bonds with maturities of one, eight, and twenty quarters. Put them in the matrix T_2 , and define two other matrices T_1 and Z :

$$T_2 = \begin{pmatrix} B'_1 \\ B'_8 \\ B'_{20} \end{pmatrix}, \quad T_1 = \begin{pmatrix} 1 & -1 & 0 \\ 1 & -1/2 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad Z = T_1^{-1}T_2.$$

The new state vector is

$$f_t = Zx_t.$$

The corresponding measurement and transition equations are

$$\begin{aligned} y_t &= A_f + B_f f_t + \eta_t, & A_f &= A, & B_f &= BZ^{-1}, \\ f_{t+1} &= K_f f_t + \Sigma_f \epsilon_{t+1}, & K_f &= ZDZ^{-1}, & \Sigma_f &= \sqrt{Z\Sigma\Sigma'Z'}, \end{aligned} \quad (32)$$

where the square root in (32) indicates a Cholesky decomposition.

When no-arbitrage is imposed, the equivalent-martingale dynamics of the new state vector are

$$f_{t+1} = \mu_f^q + K_f^q f_t + \Sigma_f^q \epsilon_{t+1}^q, \quad \mu_f^q = Z\mu^q, \quad K_f^q = ZK^qZ^{-1}.$$

Finally, the short rate is defined using

$$\delta_{f,0} = \delta_0, \quad \delta_{f,1} = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}'.$$

This rotation of the no-arbitrage parameters is to both the restricted model and the form of the unrestricted model that specifies deviations from no-arbitrage as the parameters c_0 and C_1 . For the latter model, deviations from the no-arbitrage restrictions are measured by

$$c_{f,0} = c_0, \quad C_{f,1} = C_1Z^{-1}.$$

8.6 A comment on optimization software

Researchers who estimate no-arbitrage term structure models are well aware of the practical difficulties of numerical optimization. There are many parameters and the likelihood surface has many local maxima. This optimization requires extensive searching using many different starting points, such as the procedure described in Section 8.3.

Because of computational constraints, optimization within Monte Carlo simulations cannot use an elaborate hands-on procedure. For example, the method in Section 8.3 requires about two days per optimization at current processing speeds. However, optimizing using simulated data can take advantage of the fact that we know the data-generating process. As long as the global maximum is in a local neighborhood of truth, we can use the true parameters as the single starting value for numerical optimization.

Yet even in a local neighborhood, numerical optimization of these models is difficult. The likelihood surface is almost flat along certain dimensions of the parameter space. In addition, numerical imprecision creates artificial bumpiness along the dimensions of the elements of

Σ in (31). Unfortunately, some optimization software performs poorly in this setting. This poor performance creates systematic biases in the results of Monte Carlo simulations. In particular, the out-of-sample tests are biased in favor of more complicated models, such as the restricted model studied here.

In my conversations with other researchers in this area, it became apparent that there is little appreciation for the role of the software, and there is no discussion in the term structure literature about this problem. I therefore digress from the main point of this paper to highlight the poor performance of the commonly-used optimization packages in Matlab.

Optimization within each simulation uses the true parameter vector as a starting point. Given this starting point, both the restricted and unrestricted likelihood functions are maximized using a derivative-based optimizer and analytic first derivatives. Denote this procedure as “Method A.” The parameter estimates are then refined by using five rounds of Simplex optimization and a final round of derivative-based optimization. Denote this entire procedure as “Method B.”

Simulated yields are produced using code written in Fortran/IMSL. Optimization is then performed separately with Fortran/IMSL and Matlab. (Thus optimization routines of Fortran/IMSL and Matlab are applied to the same panel of simulated data.) Table 1A reports the means and standard deviations, across 1000 simulations, of the log-likelihood values of the fitted models. It also reports means and standard deviations of the model-implied population mean of the five-year bond yield, calculated from the parameter estimates. Finally, it reports the finite-sample 95 percent critical value for the LR test of the restricted model relative to the unrestricted model. Results are reported separately for Methods A and B, as well as Fortran/IMSL and Matlab.

One conclusion to draw from the table is that the single round of derivative-based estimation (Method A) is acceptable when optimization is performed with Fortran/IMSL. For the unrestricted model, the additional refinement of Method B is irrelevant. Across the 1000 simulations, the largest improvement in log-likelihood produced by Method B is 0.05. (This number is not reported in the table.) The IMSL algorithms have greater difficulty with the restricted model, thus the refinement is slightly more important. For this model, the mean improvement in the log-likelihood is 0.04 and the maximum improvement across the 1000 simulations is 3.02.

Another conclusion is that Matlab optimization routines tend to terminate prior to reaching the optimum values located with Fortran/IMSL. This occurs when Matlab cannot locate a parameter vector with a higher likelihood value, although the score vector indicates that such a vector exists.¹² This problem is more severe with the restricted model. For this

¹²The Matlab routine is ‘fminunc.’ The maximum number of function evaluations is set to 100,000, the

model, the mean log-likelihood reached using Method A with Matlab is 3.8 less than the corresponding mean using Fortran/IMSL. Refining the estimates with Method B cuts that gap to 2.3.

The early termination of Matlab optimization routines means that the estimated parameter vector has not moved sufficiently far away from the starting point. The example highlighted in Table 1A is the estimated population mean of the five-year bond yield. The likelihood surfaces of both models are close to flat along this dimension. According to estimates from Fortran/IMSL, the standard deviation of this mean across 1000 simulations is 1.52 annualized percentage points for both models. With Matlab using Method A, the standard deviation is only 0.70 percentage points for the restricted model and 1.18 percentage points for the unrestricted model. Application of Method B raises both of these standard deviations, but neither is close to its Fortran/IMSL counterpart.

Early termination is a critical problem when evaluating out-of-sample forecast accuracy with Monte Carlo simulations. The forecasts are artificially too accurate because the starting point for optimization is truth. Moreover, because early termination afflicts the restricted model more than the unrestricted model, comparisons of out-of-sample forecasts will artificially favor the restricted model. Early termination also affects the interpretation of statistical tests of no-arbitrage restrictions. With three factors and six bonds, an LR test of the null hypothesis that the Duffie-Kan restrictions are valid has eight degrees of freedom. The asymptotic 95 percent critical value of a $\chi^2(8)$ distribution is 15.51. The finite-sample critical value, based on estimation with Fortran/IMSL, is 15.59 for Method A and 15.55 for Method B. With Matlab, there is an artificially larger wedge between the estimated restricted and unrestricted models. Thus the finite-sample critical value is higher: 21.96 using Method A and 20.18 using Method B.

The narrow message of these results is that Matlab should not be used in Monte Carlo simulations of ML estimation. The broader message is that term structure researchers need to check their Monte Carlo simulation results to ensure that the optimization procedure actually locates global optima.

maximum number of iterations is set to 10,000, and numerical tolerances are 10^{-12} . Early termination is triggered when the line search of `fminunc` cannot find an acceptable point along the current search direction. The exit flag code is `-2`.

References

- Ang, Andrew, and Monika Piazzesi, 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* 50, 745-787.
- Ang, Andrew, Monika Piazzesi, and Min Wei, 2006, What does the yield curve tell us about GDP growth?, *Journal of Econometrics* 131, 359-403.
- Ball, Clifford A., and Walter Torous, 1996, On unit roots and the estimation of interest rate dynamics, *Journal of Empirical Finance* 3, 215-238.
- Campbell, John Y., and Robert J. Shiller, 1991, Yield spreads and interest rate movements: A bird's eye view, *Review of Economic Studies* 58, 495-514.
- Christensen, Jens H.E., Francis X. Diebold, and Glenn D. Rudebusch, 2007, The affine arbitrage-free class of Nelson-Siegel term structure models, Working paper, Federal Reserve Bank of San Francisco.
- Cochrane, John H., and Monika Piazzesi, 2005, Bond risk premia, *American Economic Review* 95, 138-160.
- Cochrane, John H., and Monika Piazzesi, 2008, Decomposing the yield curve, Working paper, Chicago Booth.
- Cox, John C., Jonathan E. Ingersoll, Jr., and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385-407.
- Dai, Qiang, and Kenneth J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943-1978.
- Dai, Qiang, and Kenneth J. Singleton, 2002, Expectation puzzles, time-varying risk premia, and affine models of the term structure, *Journal of Financial Economics* 63, 415-441.
- Diebold, Francis X., and Canlin Li, 2006, Forecasting the term structure of government bond yields, *Journal of Econometrics* 130, 337-364.
- Duffee, Gregory R., 1996, Idiosyncratic variation of Treasury bill yields, *Journal of Finance* 51, 527-552.
- Duffee, Gregory R., 2002, Term premia and interest rate forecasts in affine models, *Journal of Finance* 57, 405-443.

- Duffee, Gregory R., 2008, Information in (and not in) the term structure, Working paper, Johns Hopkins University.
- Duffie, Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379-406.
- Estrella, Arturo, and Gikas A. Hardouvelis, 1991, The term structure as a predictor of real economic activity, *Journal of Finance* 46, 555-576.
- Evans, Charles L., and David A. Marshall, 1998, Monetary policy and the term structure of nominal interest rates: evidence and theory, *Carnegie-Rochester Conference on Public Policy* 49, 53-111.
- Evans, Charles L., and David A. Marshall, 2002, Economic determinants of the nominal Treasury yield curve, Working paper, Federal Reserve Bank of Chicago.
- Greenwood, Robin, and Dimitri Vayanos, 2007, Bond supply and excess bond returns, Working paper, LSE.
- Krishnamurthy, Arvind, 2002, The bond/old bond spread, *Journal of Financial Economics* 66, 463-506.
- Krishnamurthy, Arvind, and Annette Vissing-Jorgensen, 2007, The demand for Treasury debt, Working paper, Northwestern University.
- Litterman, Robert, and Jose Scheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income* 1, 54-61.
- Martin, Vance, Anthony D. Hall, and Adrian R. Pagan, 1996, Modelling the term structure, *Handbook of Statistics* 14, G.S. Maddala and C.R. Rao, Eds., 91-118.
- Nelson, Charles R., and Andrew F. Siegel, 1987, Parsimonious modeling of yield curves, *Journal of Business* 60, 473-489.
- Park, Sang Yong, and Marc R. Reinganum, 1986, the puzzling price behavior of Treasury bills that mature at the turn of calendar months, *Journal of Financial Economics* 16, 267-283.
- Singleton, Kenneth J., 1980, A latent time series model of the cyclical behavior of interest rates, *International Economic Review* 21, 559-575.
- Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177-188.

Table 1. Estimates of noise in a three-factor term structure model, 1985 through 2006

The level, slope, and curvature of the Treasury term structure are measured by the five-year yield, the five-year yield less the three-month yield, and the two-year yield less the average of the three-month and five-year yields. Yields are continuously compounded at annual rates. The sample is 88 observations of quarter-end data from 1985 through 2006. Summary statistics of these measures are in Panel A. In Panel B, yields on one-year, three-year, and four-year bonds are regressed on demeaned versions of these measures. Yields are continuously compounded at annual rates, and are expressed in percent. Ordinary least-squares standard errors are in parentheses.

Panel A. Summary statistics

	Level	Slope	Curvature
Mean	6.076	1.318	0.079
Std dev	1.872	0.948	0.255

Panel B. Regression results

	Constant	Level	Slope	Curvature	Adj R^2	Resid SD
One year bond	5.183 (0.010)	1.000 (0.007)	-0.809 (0.011)	1.016 (0.050)	0.998	0.098
Three year bond	5.748 (0.006)	1.001 (0.004)	-0.291 (0.006)	0.612 (0.027)	0.999	0.052
Four year bond	5.953 (0.006)	1.012 (0.004)	-0.112 (0.006)	0.249 (0.029)	0.999	0.056

Table 2. Estimates of dynamic term structure models for 1985 through 2006

Yields are affine functions of three factors with joint Gaussian dynamics. The factors x_t , are normalized to the five-year yield (level), the difference between the five-year yield and the three-month yield (slope), and the difference between the two-year yield and the average of the five-year and three-month yields (curvature). The factors are also demeaned. Their dynamics under the physical and equivalent-martingale measures are

$$x_{t+1} = Kx_t + \Sigma\epsilon_{t+1}, \quad x_{t+1} = \mu^Q + K^Qx_t + \Sigma\epsilon_{t+1}^q.$$

Two models are estimated with the Kalman filter using quarter-end yields from 1985Q1 to 2006Q4. The maturities are three months and one, two, three, four, and five years. One model allows for affine deviations in yields from the no-arbitrage restrictions: the two-year and four-year yields deviate from no-arbitrage yields by $c_0 + C_1x_t$. All yields are contaminated by iid measurement error with standard deviation σ_η . Yields are continuously compounded at quarterly rates. Standard errors, computed from Monte Carlo simulations, are in parentheses.

	No-arbitrage imposed			No-arbitrage not imposed		
Log likelihood	3925.44			3943.26		
Mean short rate (%/quarter)	1.478 (0.350)			1.476 (0.348)		
K	0.969 (0.048)	-0.107 (0.067)	0.079 (0.373)	0.968 (0.046)	-0.106 (0.069)	0.084 (0.375)
	0.091 (0.042)	0.856 (0.059)	-1.294 (0.298)	0.086 (0.042)	0.872 (0.060)	-1.242 (0.300)
	0.006 (0.010)	-0.001 (0.012)	0.785 (0.094)	0.007 (0.010)	0.000 (0.012)	0.766 (0.096)
$\Sigma \times 10^3$	1.557 (0.118)			1.542 (0.118)		
	0.781 (0.109)	0.790 (0.066)	0	0.765 (0.109)	0.791 (0.065)	
	0.344 (0.037)	0.045 (0.029)	0.190 (0.027)	0.357 (0.038)	0.057 (0.029)	0.201 (0.027)
$\mu^Q \times 10^4$	2.318 (0.445)			2.301 (0.447)		
	-8.112 (3.331)			-8.442 (3.343)		
	-2.399 (1.087)			-2.865 (1.142)		
K^Q	0.997 (0.001)	0.074 (0.001)	-0.056 (0.003)	0.997 (0.001)	0.076 (0.002)	-0.056 (0.005)
	0.020 (0.020)	0.965 (0.013)	-1.398 (0.108)	0.009 (0.011)	0.986 (0.016)	-1.285 (0.146)
	0.017 (0.006)	0.042 (0.009)	0.372 (0.059)	0.011 (0.007)	0.051 (0.010)	0.386 (0.103)
$c_0 \times 10^4$	-			-0.610 (0.251)		
				0.699 (0.229)		
C_1	-			-0.004 (0.007)		
				0.020 (0.011)		
				0.032 (0.059)		
				0.008 (0.006)		
				0.012 (0.009)		
				-0.042 (0.044)		
$\sigma_\eta \times 10^4$	1.394 (0.061)			1.306 (0.061)		

Table 3. Monte Carlo simulations of differences in quarterly out-of-sample forecasts

Each of 1000 Monte Carlo simulations begins with 100 quarters of simulated bond yields. These data are generated by a “true” model that imposes no-arbitrage. The first 88 quarters of data are used to estimate two term structure models. The first imposes no-arbitrage restrictions and has the same structure as the true model. The second does not impose these restrictions, and therefore nests the true model. Both models are used to construct forecasts of the term structure’s future level (the five-year yield), slope (five-year less three month), and curvature (two-year less the average of the five-year and three-month). The forecast horizon ranges from one to twelve quarters. This table reports, for each combination of horizon and variable, the square root of the mean squared difference between forecasts produced with the two models. All yields are expressed in annualized percentage points.

Horizon (quarters)	Level forecasts	Slope forecasts	Curvature forecasts
1	0.015	0.009	0.009
2	0.025	0.011	0.011
3	0.034	0.012	0.012
4	0.043	0.016	0.013
5	0.052	0.021	0.014
6	0.061	0.026	0.014
7	0.070	0.030	0.014
8	0.079	0.034	0.015
9	0.088	0.037	0.015
10	0.096	0.039	0.015
11	0.104	0.041	0.015
12	0.112	0.042	0.015

Table 4. Monte Carlo simulations of RMSEs for quarterly out-of-sample forecasts

Each of 1000 Monte Carlo simulation begins with 100 quarters of simulated bond yields. These data are generated by a “true” model that imposes no-arbitrage. The first 88 quarters of data are used to estimate three term structure models. One model imposes no-arbitrage restrictions and has the same structure as the true model. The other does not impose these restrictions, and therefore nests the true model. Both models are then used to construct forecasts of the term structure’s future level (the five-year yield), slope (five-year less three month), and curvature (two-year less the average of the five-year and three-month). The forecast horizon ranges from one to twelve quarters. This table reports, for each combination of horizon, variable, and model, the square root of the mean squared simulated forecast error. All yields are expressed in annualized percentage points.

Horizon (quarters)	Level forecasts		Slope forecasts		Curvature forecasts	
	Restricted	Unrestricted	Restricted	Unrestricted	Restricted	Unrestricted
1	0.639	0.639	0.458	0.458	0.179	0.179
2	0.907	0.907	0.572	0.572	0.232	0.232
3	1.091	1.091	0.639	0.640	0.254	0.254
4	1.273	1.272	0.712	0.712	0.277	0.276
5	1.408	1.406	0.763	0.763	0.274	0.274
6	1.553	1.551	0.820	0.820	0.290	0.290
7	1.666	1.663	0.865	0.865	0.287	0.287
8	1.751	1.748	0.875	0.875	0.292	0.292
9	1.850	1.847	0.893	0.893	0.289	0.290
10	1.951	1.949	0.911	0.910	0.291	0.292
11	2.023	2.020	0.910	0.909	0.290	0.291
12	2.144	2.142	0.917	0.916	0.297	0.298

Table 5. Root mean squared errors of monthly out-of-sample forecasts

Forecasting regressions and term structure models are estimated on rolling panels of Treasury yields with T monthly observations. The regressions and models are defined in the text. The results are used to forecast the level, slope, and curvature of the term structure three and twelve months ahead. The baseline forecast is the assumption that yields at all maturities follow a random walk. This table reports root mean squared forecast errors in basis points of annualized yields. The entire sample is 1952:6 through 2008:12.

Model	Obs (T)	Date of first forecast	RMSE (b.p.)			RMSE (b.p.)		
			Three months ahead Level	Slope	Curve	Twelve months ahead Level	Slope	Curve
Random walk	120	1972:5	68	72	25	128	124	36
AR(1) regressions			72	72	25	165	112	37
Slope regressions			70	70	25	137	115	39
5 forward rates			77	76	25	177	121	37
Diebold-Li model			71	69	24	149	112	28
Unrestr. 3-factor model			72	70	24	156	104	30
Constrained 3-factor			68	69	23	130	101	27
Random walk	480	1992:5	55	51	19	111	113	32
AR(1) regressions			56	51	19	118	106	31
Slope regressions			56	50	20	119	91	35
5 forward rates			56	45	21	122	80	27
Diebold-Li model			56	50	20	115	91	28
Unrestr. 3-factor model			58	43	20	118	91	26
Constrained 3-factor			56	44	20	113	80	27

Table 6. Estimated monthly term structure dynamics

A three-factor term structure model is estimated on 547 rolling panels of Treasury yields, each with 120 monthly observations. One estimated version has unrestricted dynamics and the other imposes the restriction that the level of the term structure is approximately a random walk. The dynamics of filtered and demeaned level, slope, and curvature satisfy

$$E_t \begin{pmatrix} \text{level} \\ \text{slope} \\ \text{curvature} \end{pmatrix}_{t+j} = K^j \begin{pmatrix} \text{level} \\ \text{slope} \\ \text{curvature} \end{pmatrix}_t.$$

Panel A reports means and standard deviations of the elements of K across the 547 sets of estimates. Panel B reports the same statistics for the elements of K^{12} .

Panel A. Estimates of K

Unrestricted			Constrained		
0.970 (0.048)	0.013 (0.071)	-0.164 (0.343)	1.000 -	0.000 -	0.000 -
0.014 (0.061)	0.966 (0.072)	-0.451 (0.313)	0.000 -	0.969 (0.040)	-0.453 (0.225)
0.009 (0.020)	0.024 (0.027)	0.748 (0.150)	0.000 -	0.012 (0.021)	0.819 (0.114)

Panel B. Estimates of K^{12}

Unrestricted			Constrained		
0.585 (0.228)	-0.076 (0.356)	-0.027 (1.411)	1.000 -	0.000 -	0.000 -
0.096 (0.243)	0.465 (0.228)	-1.413 (1.055)	0.000 -	0.509 (0.122)	-1.542 (0.668)
0.016 (0.055)	0.038 (0.043)	0.032 (0.247)	0.000 -	0.025 (0.057)	0.111 (0.231)

Table 1A. Effects of numerical optimization on simulated properties of term structure models

This table summarizes the results of 1000 Monte Carlo simulations. Each simulation consists of maximum likelihood estimation of two term structure models. The restricted model imposes the Duffie-Kan restrictions. This model is also used to generate simulated data. Numerical maximization of the likelihood functions is performed with both Method A and Method B. Method A uses a derivative-based optimizer in combination with analytic first derivatives. The starting point is truth. Method B has refines the solution of Method A with five rounds of Simplex optimization and a final round of derivative-based estimation. Optimization is performed either with Fortran/IMSL or Matlab. The table reports means and standard deviations (in parentheses), across the simulations, of the log-likelihood values and point estimates of the unconditional mean of the five-year bond yield. It also reports a critical value for an LR test of the Duffie-Kan restrictions.

Method	Software	Log-likelihood		Population mean of 5-year bond yield		95 percent crit val of LR test
		restricted	unrestricted	restricted	unrestricted	
A	Fortran/IMSL	3936.4 (16.8)	3940.7 (17.0)	7.064 (1.519)	7.063 (1.516)	15.59
B	Fortran/IMSL	3936.5 (16.8)	3940.7 (17.0)	7.067 (1.524)	7.063 (1.516)	15.55
A	Matlab	3932.6 (16.7)	3938.6 (16.9)	6.968 (0.697)	7.093 (1.175)	21.96
B	Matlab	3934.2 (16.8)	3939.8 (16.9)	7.139 (0.974)	7.079 (1.287)	20.18

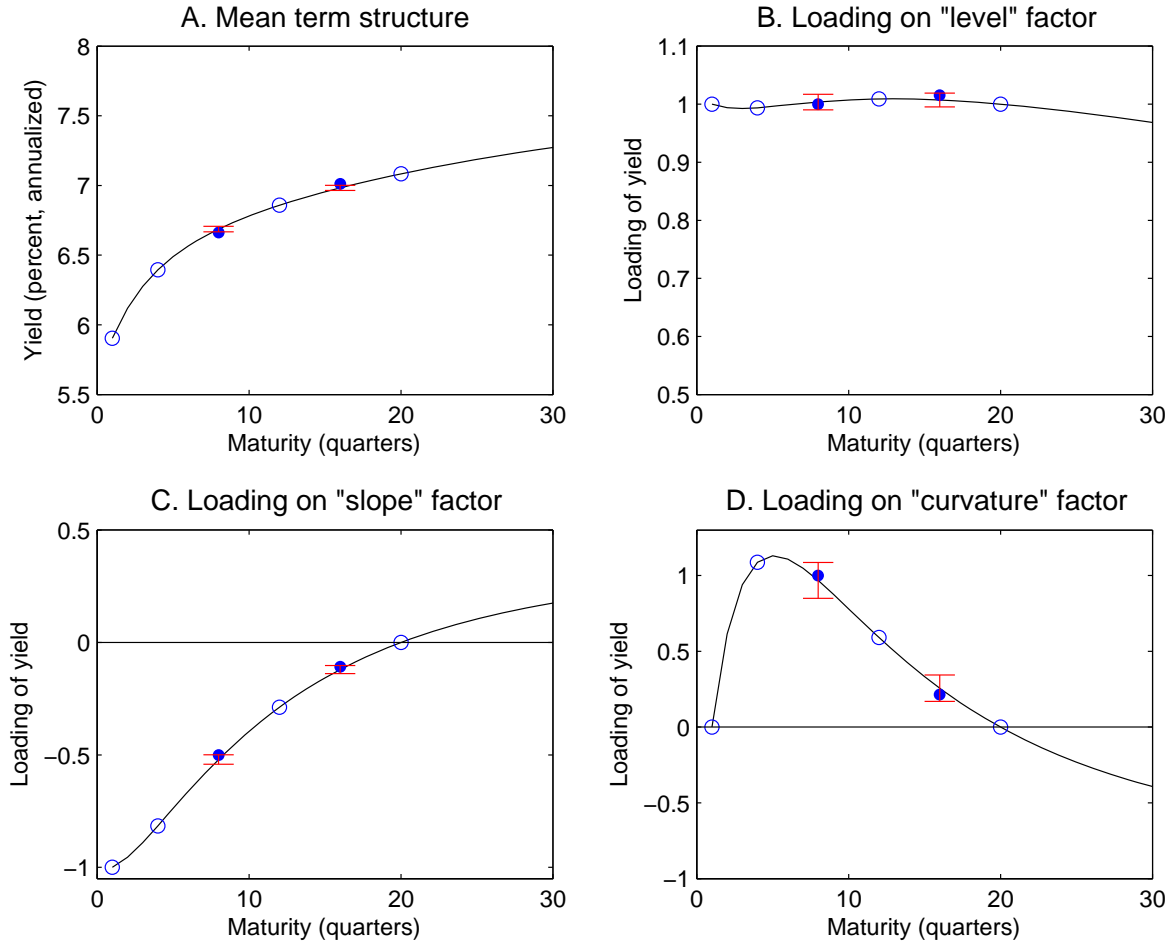


Fig. 1. Parameter estimates of a term structure model. A three-factor Gaussian model is estimated using quarterly Treasury bond yields from 1985Q1 through 2006Q4. The factors are rotated into level, slope, curvature. The lines in each panel are the no-arbitrage means and factor loadings consistent with yields on three-month, one-year, three-year, and five-year bonds. Means and factor loadings for two-year and four-year bond yields are allowed to deviate from the restrictions of no-arbitrage. Their point estimates are represented with dots and plus/minus two standard error bounds are shown in red.

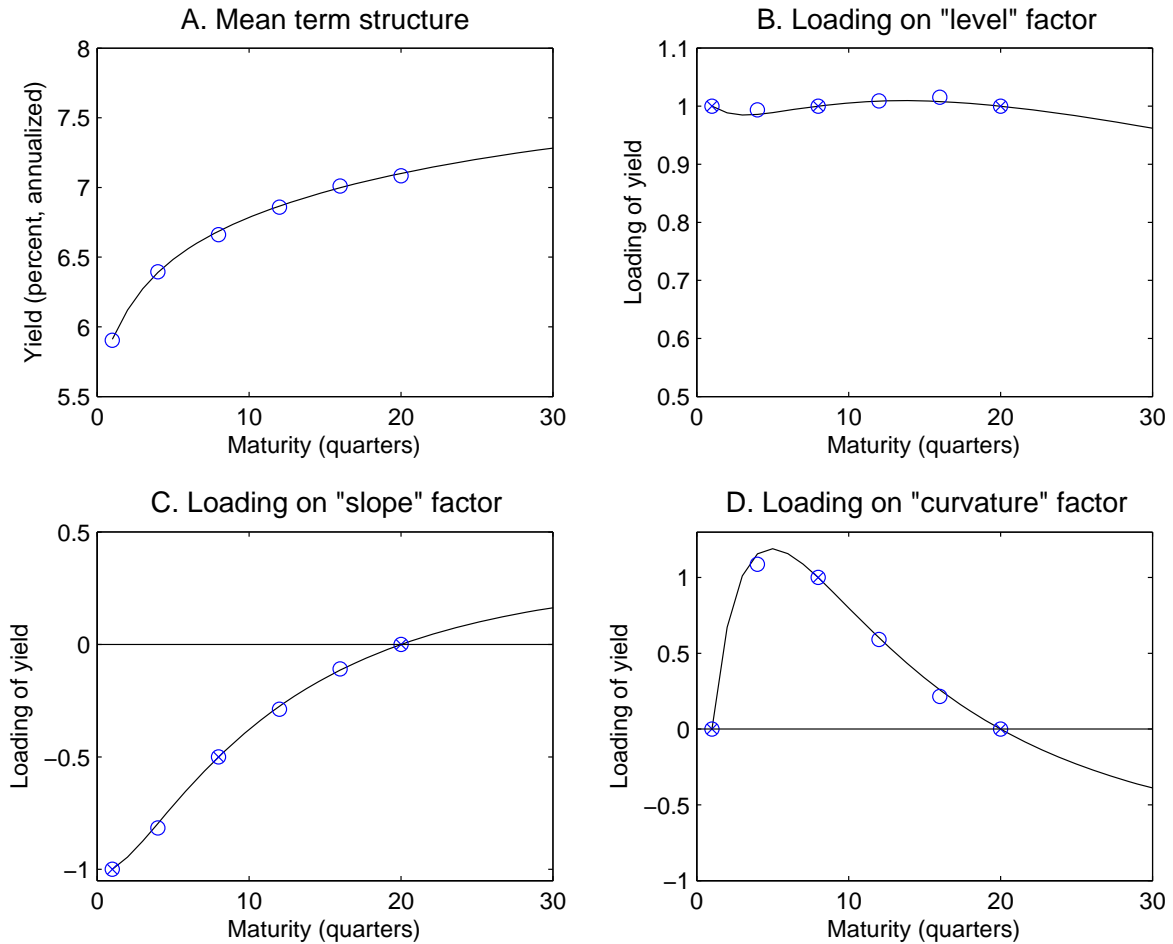


Fig. 2. Parameter estimates of two term structure models. Three-factor Gaussian models, with and without no-arbitrage restrictions, are estimated using quarterly Treasury bond yields from 1985Q1 through 2006Q4. The factors are rotated into level, slope, curvature. The circles are estimates of mean yields and factor loadings from the unrestricted model. The solid lines are means and loadings implied by the model that imposes no-arbitrage. The factor rotation implies that all loadings coincide at maturities of three months, two years, and five years. These points are marked with an x .