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

Gregory R. Duffee[*]

Johns Hopkins University

First draft: October 2007
This Draft: January 2009

ABSTRACT

Does imposing no-arbitrage restrictions help when using the term structure to forecast future bond yields or macroeconomic activity? Standard intuition says that if we know the restrictions are correct, imposing them increases estimation efficiency and forecast accuracy. This paper argues that for affine models, if the restrictions hold exactly, they must be irrelevant. The restrictions can affect forecasts only if we weaken them by allowing for measurement error in yields. Even in this case, empirical evidence indicates the restrictions have no practical effect on forecast accuracy. In fact, the restrictions are more likely to be useful if they are false. Then deviations from the restrictions can help detect model misspecification.

[*]Voice 410-516-8828, email duffee@jhu.edu. Address correspondence to 440 Mergenthaler Hall, 3400 N. Charles St., Baltimore, MD 21218. Thanks to seminar participants at NYU, Penn State, Ohio State, Johns Hopkins, and a Wharton brown bag seminar.

1 Introduction

Dynamic no-arbitrage term structure models have long been recognized as powerful tools for cross-sectional asset pricing. Among these models, the affine class is particularly useful because of the tractable pricing formulas developed by Vasicek (1977), Cox, Ingersoll, and Ross (1985), and Duffie and Kan (1996). In a typical affine model, a few factors drive the entire term structure. A variety of fixed-income instruments can be valued using zero-coupon bond prices because exposures to the small number of risks must be priced consistently across the instruments.

Recently, researchers have turned their attention to a time-series application of affine models: forecasting. The term structure contains information about both future interest rates and future macroeconomic conditions. Information in bond yields has typically been exploited using predictive regressions, vector autoregressions, dynamic factor analysis, and structural macroeconomic models.[1] Dynamic term structure models appear to be a useful addition to an econometrician's toolkit. Duffee (2002) and Christensen, Diebold, and Rudebusch (2007) 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 models with no-arbitrage restrictions produce more accurate forecasts, at least in the context of Gaussian dynamics.

The no-arbitrage requirement is viewed as key to this forecasting success because it imposes cross-equation restrictions on yield dynamics. Duffee (2002) argues "...imposing these 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 the cross-sectional intuition than by either the logic of affine models or empirical analysis.

I argue here that the no-arbitrage restrictions of Gaussian models, and presumably affine models in general, are effectively useless in forecasting. The intuition is simple. The bite of no-arbitrage in a term structure model follows 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. In particular, if the term structure is described by an

[1] The relevant literature is vast, thus I mention only a few notable contributions here. 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.

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 derive cross-bond 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^2s 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^2s are around 0.999. When models are estimated with maximum likelihood, Monte Carlo analysis of out-of-sample forecasting performance reveals that 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 here 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 happens to be consistent with the Duffie-Kan restrictions is an empirical question. I use a three-factor Gaussian model to address this issue, using quarterly data from 1985 through 2006. The Duffie-Kan restrictions are statistically overwhelmingly rejected, although the economic significance of the rejection is small. Thus imposing the restrictions is not only unhelpful; it can be counterproductive to the extent that the restrictions are inconsistent with the correct affine mapping.

Is there any role for no-arbitrage restrictions in forecasting? The results here suggest a few possibilities. The restrictions help us better understand the sources of variation in expected excess bond returns. For example, Dai and Singleton (2002) use a Gaussian term structure model to interpret the failure of the expectations hypothesis of interest rates. That kind of analysis can be performed without imposing no-arbitrage, but the cross-sectional implications of no-arbitrage restrictions broaden the range of questions that can be asked. In addition, the restrictions can be used as an informal specification test of the broader class of models—the models that do not impose the no-arbitrage restriction.

The next section describes the modeling framework. The third section describes the general econometric testing procedure. Estimation results are in Section 4 and Monte Carlo simulation results are in Section 5. Section 6 considers circumstances in which no-arbitrage restrictions can be helpful. The final section contains concluding remarks.

2 The modeling framework

The ingredients of no-arbitrage term structure models are the physical dynamics of a state vector, a short-term interest rate that is a function of the state, and equivalent-martingale dynamics of the state vector. In the affine class of Duffie and Kan, the short-rate function and the state dynamics are chosen carefully so that zero-coupon bond yields are specific affine functions of the state.

We can study the usefulness of the no-arbitrage restriction by comparing a no-arbitrage model with a model that does not impose the restriction, but is otherwise identical. This section constructs such an "unrestricted" model, which nests a no-arbitrage model. The next section describes how to estimate both models and test statistically the hypothesis that the no-arbitrage restrictions are satisfied.

The specific case analyzed here uses a discrete-time Gaussian setting, although the generalization to other affine models is analytically identical. I focus on discrete-time Gaussian models because of the computational demands of the Monte Carlo simulations in Section 5. The concluding section briefly mentions some issues that arise in the context of non-Gaussian affine models.

2.1 The unrestricted model

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

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

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

$$y_t^{(m)} = A_m + B_m' x_t + \eta_{m,t}, \quad \eta_{m,t} \sim N(0, \sigma_\eta^2). \tag{2}$$

The idiosyncratic component $\eta_{m,t}$ is independent across time and bonds.

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

$$\tilde{y}_t^{(m)} = A_m + B_m' x_t, \tag{3}$$

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

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

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 of is derived from (4) and the specification of equivalent-martingale dynamics of the state vector. In this unrestricted model, (2) is simply an assumption.

2.2 The no-arbitrage restriction

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 RP 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, any covariation across bonds is driven only by the state vector.

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 coefficients A_m and B_m in (2). Using the discrete-time version of the essentially affine Gaussian framework of Duffee (2002), the equivalent martingale measure dynamics of x_t are

$$x_{t+1} = \mu^q + K^q x_t + \Sigma \epsilon_{t+1}^q, \quad \epsilon_{t+1}^q \sim MVN(0, I). \tag{5}$$

[2]The 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).

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

$$\begin{aligned} B'_m &= \mathbb{B}(m; \delta_1, K^q)' \\ &= \frac{1}{m}\delta'_1 (I - K^q)^{-1} (I - (K^q)^m) . \end{aligned} \tag{6}$$

The constant term for $m > 1$ is

$$\begin{aligned} A_m &= \mathbb{A}(m; \delta_0, \delta_1, \mu^q, K^q, \Sigma) \\ &= \delta_0 + \frac{1}{m}\delta'_1 \left[mI - (I - K^q)^{-1} (I - (K^q)^m)\right] (I - K^q)^{-1} \mu^q \\ &\quad -\frac{1}{2m}\sum_{i=1}^{m-1} i^2 B'_i \Sigma_x \Sigma'_x B_i. \end{aligned} \tag{7}$$

I refer to equations (6) and (7) 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.

The main point of this paper is that the restriction of no-arbitrage has no appreciable effect on forecasting performance. This does not mean that stronger assumptions about investors' attitudes towards risk have no effect on forecasts. Such assumptions correspond to restrictions on the equivalent-martingale dynamics. For example, constant risk premia over time, as in Vasicek (1977), corresponds to the assumption that K equals K^q. More recently, Christensen et al. (2007) find that a model with a parsimonious specification of K^q does a good job forecasting future interest rates. I return to this issue in Section 6.

2.3 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

$$z_t = A_z + B_z x_t + \eta_{z,t}. \tag{8}$$

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 (8). There are no Duffie-Kan restrictions associated with A_z and B_z. Thus if the no-arbitrage restrictions (6) and (7) 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.

2.4 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.

3 The econometric procedure

Parameter estimation and statistical tests of the Duffie-Kan restrictions are easily implemented with maximum likelihood using the Kalman filter.

3.1 A state space setting

Estimation uses observed yields on d zero-coupon bonds with maturities $M = (m_1, \ldots, m_d)'$, where $d > (n+1)$. This inequality is necessary to generate overidentifying restrictions. Stack the period-t yields in the d-vector y_t. The dynamics of y_t are conveniently written in state-space form as a combination of the transition equation (1) and the measurement equation

$$y_t = A + Bx_t + \eta_t, \quad \eta_t \sim MVN(0, \sigma_\eta^2 I). \tag{9}$$

In (9), 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. The vector can be scaled, rotated, and translated as desired. One identification approach is to tighten the description of the vector's dynamics; this restricts (1). Another is to tighten the link between the observed yields and the vector (e.g., identify each element of the state vector with a particular yield); this restricts (9).

The normalization I use in estimation is chosen for its tractability, while the normalization I use in explaining the results is chosen for its intuitive appeal. 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

$$x_{t+1} = Dx_t + \Sigma\epsilon_{t+1} \tag{10}$$

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.

On an intuitive level, it is easier to identify the factors as yields, and I use this kind of identification when interpreting the models. However, it is not useful for estimation because identification is done through nonlinear constraints on the parameters of the equivalent-martingale dynamics (5).

3.2 The role of measurement error

If the null hypothesis is true, but there is no measurement error in yields, ML estimation of the restricted model is identical to ML estimation of the unrestricted model. To make this point clearly, transform the latent state vector so that it equals a vector of n yields, y_t^n, all of which are in the observed data. Place these n yields at the beginning of the observed yields y_t, denoting the other $d-n$ yields as y_t^o. Then the measurement and transition equations are

$$\begin{pmatrix} y_t^n \\ y_t^o \end{pmatrix} = \begin{pmatrix} 0 \\ A^* \end{pmatrix} + \begin{pmatrix} I \\ B^* \end{pmatrix} y_t^n, \tag{11}$$

$$y_{t+1}^n = \mu^* + K^* y_t^n + \Sigma^* \epsilon_{t+1}. \tag{12}$$

The special case here is that there is no measurement error in (11). The vector A^* and matrix B^* are free parameters in the unrestricted model and functions of equivalent-martingale parameters in the restricted model.

Because there is no measurement error, A^* and B^* are coefficients from regressing y_t^o on y_t^n. There is no estimation error in the regressions. For the unrestricted model, the regression coefficients are parameters. For the restricted model, the equivalent-martingale parameters are chosen to reproduce these coefficients. (Of course, if there is no measurement error and the null hypothesis is *not* true, the regression coefficients cannot be reconciled any equivalent-martingale parameters, which would allow us to immediately reject the null.)

With these choices of A^* and B^*, the probability of observing y_t conditional on y_t^n is one. Therefore the conditional likelihood function for y_t can be written

$$l(y_{t+1}|y_t; A^*, B^*, \mu^*, K^*, \Sigma^*) = l(y_{t+1}^n|y_t^n; \mu^*, K^*, \Sigma^*). \tag{13}$$

Thus ML estimation of either the restricted or unrestricted models is identical to ML estimation of a first-order vector autoregression for y_t^n.

In the presence of measurement error, the two models need not coincide. The difference between the two sets of estimates can be used to test statistically the null hypothesis.

3.3 Testing the Duffie-Kan restrictions

The restrictions of the no-arbitrage model apply to the measurement equation (9). Under the null hypothesis that the Duffie-Kan restrictions hold, the matrix B and vector A satisfy the restrictions of (6) and (7) respectively. Formally,

$$\begin{aligned} H0: \quad A &= \mathbb{A}(M;\delta_0,\delta_1,\mu^q,K^q,\Sigma) = \begin{pmatrix} \mathbb{A}(m_1;\cdot) \\ \dots \\ \mathbb{A}(m_d;\cdot) \end{pmatrix}; \\ B &= \mathbb{B}(M;\delta_1,K^q) = \begin{pmatrix} \mathbb{B}(m_1;\cdot)' \\ \dots \\ \mathbb{B}(m_d;\cdot)' \end{pmatrix}. \end{aligned} \tag{14}$$

For estimation purposes, the parameters of the restricted model are those of (4), (5), and (10). They are stacked in the vector

$$\rho_0 = \{\delta_0, \delta_1, D, \Sigma, \mu^q, K^q, \sigma_\eta^2\}. \tag{15}$$

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

$$H1: \quad A,\ B \text{ unrestricted.} \tag{16}$$

A likelihood ratio test statistically evaluates $H0$ versus $H1$. The parameters of the unrestricted model are those of (9) and (10), stacked in

$$\rho_1 = \{D, \Sigma, A, B, \sigma_\eta^2\}. \tag{17}$$

There are $2+3n+n^2+n(n-1)/2$ parameters in ρ_0 and $1+n+(n+1)d+n(n-1)/2$ parameters in ρ_1. there are $(1+n)(d-n-1)$ overidentifying restrictions to test the 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. The narrow interpretation is the one suggested in Section 2.2. 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 latter interpretation is explored in Section 6.

3.4 Reparameterizing the alternative hypothesis

Although the unrestricted model nests the restricted model, the parameter vector ρ_1 do not nest the vector 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 (9) 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

$$\begin{pmatrix} y_t^x \\ y_t^v \end{pmatrix} = \begin{pmatrix} A^x \\ A^v \end{pmatrix} + \begin{pmatrix} B^x \\ B^v \end{pmatrix} x_t + \eta_t, \tag{18}$$

$$A^x = \mathbb{A}(M^x; \delta_0^\dagger, \delta_1^\dagger, \mu^{q^\dagger}, K^{q^\dagger}, \Sigma), \tag{19}$$

$$B^x = \mathbb{B}(M^x; \delta_1^\dagger, K^{q^\dagger}), \tag{20}$$

$$A^v = \mathbb{A}(M^v; \delta_0^\dagger, \delta_1^\dagger, \mu^{q^\dagger}, K^{q^\dagger}, \Sigma) + c_0, \tag{21}$$

$$B^v = \mathbb{B}(M^v; \delta_1^\dagger, K^{q^\dagger}) + C_1. \tag{22}$$

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 $\delta_1^\dagger$ and $K^{q^\dagger}$:

$$\{\delta_1^\dagger, K^{q^\dagger}\} = \mathbb{B}^{-1}(B^x; M^x). \tag{23}$$

The inverse mapping in (23) is done numerically. There are values of B^x which cannot be inverted using (23). If inversion is impossible for one set of bonds that comprise the "exact" group, a different set of bonds can be used.[3] The remaining equivalent-martingale parameters are determined numerically by the inversion

$$\{\delta_0^\dagger, \mu^{q^\dagger}\} = \mathbb{A}^{-1}(A^x; M^x, \delta_1^\dagger, K^{q^\dagger}, \Sigma). \tag{24}$$

[3]In 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.

After calculating these equivalent-martingale parameters, we can write the parameters A^v and B^v in (20) and (21) 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 (17) to

$$\rho_1^\dagger = \{D, \Sigma, \delta_0^\dagger, \delta_1^\dagger, \mu^{q^\dagger}, K^{q^\dagger}, c_0, C_1, \sigma_\eta^2\}. \tag{25}$$

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

Writing the null 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.

When the null hypothesis is correct, imposing it in estimation is likely to improve efficiency, in the sense that standard errors of the parameter estimates are reduced. One way to informally measure the efficiency gain is to estimate the unrestricted model and examine the standard errors of c_0 and C_1 in (20) and (21). If the standard errors on these parameters are large, fixing the parameters to zero (when this restriction is true) may produce a substantial increase in efficiency. If the standard errors are tiny, the efficiency gains are modest.

4 Empirical estimation

This paper studies two empirical questions. First, how closely do Treasury yields adhere to the Duffie-Kan restrictions of the discrete-time Gaussian model? Second, assuming the Duffie-Kan restrictions are true, what is the loss in forecast accuracy if they are not imposed? The answer to the first question necessarily depends on the data sample. The "true" model used to answer the second is based on observed Treasury yield behavior, and therefore also depends on a particular sample. For computational ease, the same sample of Treasury yields is used to test the Duffie-Kan restrictions and estimate the "true" model used in evaluating forecasting performance.

This section estimates both the unrestricted and restricted models and tests the restrictions. The next section uses the restricted model in Monte Carlo simulations of forecasting.

4.1 Data

The data used are quarter-end observations of Treasury yields from 1985 through 2006. The sample size is short and the observation frequency is low relative to other empirical analyses in the literature. My data choice is motivated by the demands of Monte Carlo simulation. As discussed below in Section 4.3, I compute standard errors of parameter estimates with Monte Carlo simulation. The simulations are time-consuming because each requires ML estimation of the restricted and unrestricted term structure models. To economize on computing time, I use the same set of Monte Carlo simulations to construct standard errors here and to evaluate forecast accuracy in the next section. Therefore the sample size here must be the same as the sample size used in studying forecasts.

A small sample size helps to strengthen the case, made in the next section, that the Duffie-Kan restrictions are unhelpful in forecasting. Parameter restrictions are more likely to play an important role in estimation when using a small sample than a large sample. Since this paper's main emphasis is on the usefulness of the restrictions rather than on the exact dynamics of the term structure, I give up examination of the Duffie-Kan restrictions over a long sample in order to produce more compelling evidence in the next section.

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.

4.2 Level, slope, and curvature factors

This paper follows much of the no-arbitrage empirical literature by using three state variables. I estimate unrestricted and restricted three-factor versions of the model. Estimation uses the measurement equation (9) and the normalized transition equation (10). The unrestricted model has 31 free parameters and the restricted model has 23 free parameters.

After estimation, the parameters of the unrestricted model are transformed into the equivalent representation (18), where the three-month, one-year, three-year, and five-year bonds are used to exactly identify an equivalent-martingale measure. Deviations from Duffie-Kan restrictions are allowed in the two-year and four-year bond yields. As explained in Section 3.4, this transformation makes it easier to express violations of the restrictions in economically meaningful terms.

A transformation of the latent state vector helps to interpret the factors. Since Litterman and Scheinkman (1991), financial economists have usually viewed the dynamics of Treasury yields in terms of "level," "slope," and "curvature" factors. I thus rotate the vector to roughly correspond to level, slope and curvature. The state vector used in describing the results is

$$f_t = \begin{pmatrix} \tilde{y}_t^{(20)} - E\tilde{y}_t^{(20)} \\ \left(\tilde{y}_t^{(20)} - E\tilde{y}_t^{(20)}\right) - \left(\tilde{y}_t^{(1)} - E\tilde{y}_t^{(1)}\right) \\ \left(\tilde{y}_t^{(8)} - E\tilde{y}_t^{(8)}\right) - \frac{1}{2}\left(\left(\tilde{y}_t^{(1)} - E\tilde{y}_t^{(1)}\right) + \left(\tilde{y}_t^{(20)} - E\tilde{y}_t^{(20)}\right)\right) \end{pmatrix}. \tag{26}$$

The first factor is the demeaned five-year yield, the second is the five-year yield less the three-month yield (both demeaned), and the third is the two-year yield less the average of the three-month and five-year yields (again, demeaned). This factor transformation is applied to both the restricted model and transformation (18) of the unrestricted model.

Appendix A contains the mechanical transformation between the state vector x_t used in estimation and the state vector f_t used in explaining the results. The physical dynamics of f_t are written in terms of μ_f, K_f, and Σ_f. The same subscript denotes parameters for equivalent-martingale dynamics and deviations from no-arbitrage The appendix describes the transformation from the parameters of the physical and equivalent-martingale dynamics of x_t to the corresponding parameters for f_t.

4.3 Optimization

The likelihood functions are maximized using Intel Fortran calling IMSL numerical optimization routines. Details are in Appendix B. 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. This issue is addressed in more detail in Section 5. Here it is sufficient to point out that Matlab has a much harder time maximizing the likelihood functions than does Fortran/IMSL.

4.4 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 (26) with their observable counterparts. For each maturity m, the approximate function is

$$y_t^{(m)} = a_m + b_m' \left(\begin{array}{c} y_t^{(20)} - \overline{y_t^{(20)}} \\ \left(y_t^{(20)} - \overline{y_t^{(20)}}\right) - \left(y_t^{(1)} - \overline{y_t^{(1)}}\right) \\ \left(y_t^{(8)} - \overline{y_t^{(8)}}\right) - \frac{1}{2}\left(\left(y_t^{(1)} - \overline{y_t^{(1)}}\right) + \left(y_t^{(20)} - \overline{y_t^{(20)}}\right)\right) \end{array} \right) + e_t^{(m)} \tag{27}$$

where the bars indicate sample means. We can think of (27) as a regression equation. Estimates of the coefficients a_m and b_m will be biased because of an errors-in-variables problem.

Panel A of Table 1 reports summary statistics for the observable version of the factors. Panel B reports OLS estimation results of applying (27) 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^2s 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. 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^2s may not be informative about the population properties of yields. However, these R^2s 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^2s 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.

4.5 Estimation results

Table 2 reports parameter estimates based on the level, slope, and curvature representation of the factors in (26). Although there are 23 and 31 free parameters in the restricted and unrestricted models, the table reports 29 and 37 respective 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.[4]

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 precision of the point estimates.

4.5.1 The economic importance of the restrictions

The vector $c_{f,0}$ and the matrix $C_{f,1}$ of the unrestricted model capture deviations from Duffie-Kan restrictions. (Recall the f subscript denotes parameters based on the factor rotation (26).) The estimate of $c_{f,0}$ 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.[5] 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 $c_{f,0}$ and $C_{f,1}$ 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 cal-

[4]Parameter estimates for each simulation of the unrestricted model are transformed into the set of parameters corresponding to (18). 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 7 and 8 indicates that this discrepancy does not have a material effect on the standard errors.

[5]The units of $c_{f,0}$ 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.

culate, for each quarter in the sample, the fitted deviation

$$\text{fitted deviation}_t = c_{f,0} + C_{f,1}\hat{f}_t. \tag{28}$$

In (28), $\hat{f}_t$ 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 $c_{f,0}$ and $C_{f,1}$ 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. (Section 4.6 explains the dotted-dashed lines.) 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.

4.5.2 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.[6] The source of the rejection is largely the mean yields. The standard errors on the two elements of $c_{f,0}$ are a basis point or less of annualized yield. The standard errors of $C_{f,1}$ 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 $c_{f,0}$ and $C_{f,1}$ are displayed in red in Fig. 1.

The tight standard errors on $c_{f,0}$ and $C_{f,1}$ 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 $c_{f,0}$ and $C_{f,1}$ are roughly coefficients of a cross-sectional regression of yields on yields. The standard errors on $c_{f,0}$ (rescaled to percent at an annual horizon) and $C_{f,1}$ 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.

[6] The asymptotic 95 percent critical value is 15.51. The finite-sample critical value is similar, as discussed in Section 5.1 and displayed in Table 4.

4.5.3 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-Kan restrictions. Standard errors for parameters identified by the physical measure (the mean short rate, K_f, Σ_f, 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_f^Q and K_f^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 errrors of parameters related to the curvature factor. These parameters are the third element of μ_f^Q and the third column of K_f^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 $\delta_{f,0}$ 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.[7] 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.[8]

4.6 An alternative model

Diebold and Li (2006) build a dynamic Nelson-Siegel model and examine its forecasting performance out of sample. They find that the parsimonious model performs well relative to more flexible models, even though the model violates no-arbitrage restrictions. They broadly attribute this performance to the shrinkage principle, writing "…the impositions of

[7] 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 4. The corresponding standard error using all 1000 simulations is also 1.54.

[8] The standard errors for the unrestricted model are based on the 960 simulations discussed in Footnote 4. The corresponding standard errors for all 1000 simulations are 0.0063, 0.0091 and 0.0454.

restrictions, which will of course degrade in-sample fit, may nevertheless be helpful for out-of-sample forecasting, even if the restrictions are false.” (p. 362) The modeling framework here allows us to interpret more precisely which restrictions in the Diebold and Li model are helpful in forecasting and which are not.

The starting point of this analysis is the model of Diebold, Rudebusch, and Aruoba (2006), which generalizes Diebold and Li. In the DRA model, a three-factor latent vector has the Gaussian dynamics of (1). The relation between factors and yields is

$$A_m = 0, \tag{29}$$

$$B'_m = \begin{pmatrix} 1 & \frac{1-e^{-m\lambda}}{m\lambda} & \frac{1-e^{-m\lambda}}{m\lambda} - e^{-m\lambda} \end{pmatrix}. \tag{30}$$

The Diebold-Li model further restricts the VAR(1) dynamics in (1) to a set of three AR(1) equations with correlated shocks. Both the Diebold-Lie and DRA model are nested in the unrestricted model used here. However, they are not nested in the restricted model because they do not satisfy the Duffie-Kan restrictions.[9]

I estimate the DRA model using the same data and same numerical optimization procedure used to estimate the other two models. There are 20 free parameters. The model is estimated using (1), (29), and (30). The three factors correspond to level, slope, and curvature. However, the precise definitions of these terms in the Nelson-Siegel framework differs slightly from the definition in (26). To simplify comparison with the results reported for the restricted and unrestricted models in Table 2, I therefore rotate the Nelson-Siegel factors to the factors of (26). Table 3 reports estimates of the dynamics of the rotated parameter vector along with estimates of λ and σ_η.

Two results are worth highlighting. First, the additional restrictions imposed by Diebold and Li on the VAR(1) dynamics are roughly consistent with the data. Both the level and curvature factors are indistinguishable from univariate AR(1) processes with correlated shocks. (The relevant estimates are in the first and third rows of the K matrix in the table.) The slope factor has more complicated estimated dynamics. The t-statistics on lagged level and curvature exceed conventional significance levels. The same matrix appears in the restricted and unrestricted models estimated in Section 4.5. Note that the parameter estimates and standard errors in Table 2 are very close to those for the estimated DRA model.

Second, the DRA model is statistically rejected in favor of the unrestricted model. The LR test statistic exceeds 100. With 17 degrees of freedom, the asymptotic one percent critical value is about 33. Since the DRA and unrestricted models have the same physical

[9]Christensen et al. (2007) construct a similar model that satisfies Duffie and Kan. It is a continuous-time model and is thus not nested here either.

factor dynamics, the rejection is entirely driven by the cross-sectional restrictions of the DRA model.

Yet economically, differences between the DRA model and the other two models do not appear large. Visual evidence is in Fig. 2, which was described in Section 4.5.1. The figure displays mean yields and factor loadings for the DRA model using dotted-dashed lines. The biggest disagreement between the DRA model and the other two models is the shape of the unconditional mean yield curve. The DRA yield curve is higher, steeper at the short end, and flatter at the long end. Loadings on the level and slope factors for the DRA model are almost identical to those of the other models. Loadings on the curvature factor do not match up quite as well, especially for the one-year maturity.

As Diebold and Li note, a parsimonious model that performs relatively poorly in-sample may nonetheless perform relatively well out-of-sample, even if the model is false. The next section evaluates the forecasting performance of the DRA model in such a setting. The restricted model estimated in this section is treated as truth, thus by construction the DRA model is false. However, the Fig. 2 shows that the "true" factor loadings are very close to factor loadings generated by a Nelson-Siegel model. Thus it is an empirical question whether imposing the false restrictions improve out-of-sample forecasts. In this exercise the DRA model is used for forecasting instead of the Diebold-Li model to isolate the role of the Nelson-Siegel cross-sectional restrictions.

5 Out-of-sample forecasting

As mentioned at the beginning of Section 4, the goal of this section is to examine the loss in forecast accuracy when Duffie-Kan restrictions are true but are not imposed in estimation. It also examines whether the DRA model produce more accurate forecasts owing to its parsimonious (albeit misspecified) cross-sectional restrictions. The answers necessarily depend on both the true model and the sample size. Both are taken from the previous section. 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 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, unrestricted, and DRA models are all 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.

5.1 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 Appendix B.

Because of computational constraints, optimization within Monte Carlo simulations cannot use an elaborate hands-on procedure. For example, the method in the appendix 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 maximimum 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 (10). 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 highight 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 4 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.[10] This problem is more severe with the restricted model. For this 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 4 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

[10]The Matlab routine is 'fminunc.' The maximum number of function evaluations is set to 100,000, the 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.

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 artifically 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.

5.2 Results

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. I also construct the difference between the restricted model's forecast and the DRA model's forecast. Table 5 reports the square roots of the mean squared differences. In the language of the table, Models 1, 2, and 3 refer to the restricted model, the unrestricted model, and the DRA model respectively.

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. The DRA model produces noticeably different forecasts. At the twelve-quarter-ahead horizon, the root mean squared differences are 65, 21, and 9 basis points for level, slope, and curvature.

Table 6 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, while the DRA model produces less accurate forecasts for all horizons and variables. 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. Forecasts produced by the DRA model have higher RMSEs. For horizons up to four quarters ahead the differences in RMSE are economically quite small (below three basis points), and even at longer horizons they are not dramatic. For example, at the twelve month horizon, the DRA model forecast of the level has a RMSE seven basis points higher than the other two models. Nonethless, there is a clear qualitative difference in forecast accuracy between the two models that nest the "true" model and the more parsimonious model that is inconsistent with it.

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 4.4, the R^2s from three-factor cross-sectional regressions for 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 When are no-arbitrage restrictions useful?

The Monte Carlo evidence in the previous section focused on a specific kind of forecast. We observe a panel of constant-maturity bond yields and construct forecasts of future yields on bonds with the same maturities. Although the no-arbitrage requirement is not helpful in constructing such forecasts, it can play an important role in other kinds of forecasts. If, say, we wish to forecast the bond yield for a maturity not included in the panel, some kind of interpolation is necessary. Duffie-Kan restrictions can be used to perform the interpolation. Three additional uses of these restrictions are discussed in this section.

6.1 Imposing tighter restrictions on preferences

No-arbitrage restrictions can be valuable if we want to impose additional structure on investors' preferences towards risk. Imposing additional structure can increase the precision of forecasts. To understand why, recall that in an n-factor model, the Duffie-Kan restrictions tell us how to infer the mean and factor loadings for one bond's yield in terms of the means and factor loadings of yields on $n + 1$ other bonds. In the model of Section 2, Duffie-Kan restrictions have no implications for the means or factor loadings of the $n+1$ bonds. But the Duffie-Kan restrictions also can be used to impose additional restrictions on risk preferences.

For example, imagine that we know investors are indifferent to interest rate exposure. In the context of Section 2.2, yields on all bonds are then given by the Duffie-Kan formulas evaluated using the physical-measure values of μ, K, and Σ. By imposing this structure in estimation, contemporaneous covariances among bond yields help to pin down the mean and persistence of short-term interest rates. Put differently, cross-sectional moments contain information about time-series parameters. Exploiting these moments increases the precision of parameter estimates.

Although unrealistic, this example illustrates the source of increased efficiency in estimation. Components of equivalent-martingale factor dynamics are equated to corresponding components of physical factor dynamics. Without the Duffie-Kan framework, there is no way to exploit restrictions on risk preferences.

6.2 Interpreting expected excess bond returns

No-arbitrage restrictions are also helpful if we want to better understand the economics linking the term structure to yield forecasts. These forecasts are effectively equivalent to forecasting future bond returns—both riskfree returns and expected excess returns to long-maturity bonds. For example, Dai and Singleton (2002) use an affine model to interpret the empirical failure of the expectations hypothesis. They show that after adjusting for model-implied risk premia, the hypothesis holds. Duffee (2002) shows how expected excess bond returns are related to the level, slope, and curvature of the term structure.

Here I show that although no-arbitrage restrictions are not critical to the analysis of expected excess bond returns, they can be useful. The log return from t to $t+k$ on an m-maturity bond in excess of the log return to rolling over a position in a one-period bond is

$$xr^{(m)}_{t,t+k} = my^{(m)}_t - (m-k)y^{(m-k)}_{t+k} - \sum_{i=1}^{k} r_{t+i-1}. \tag{31}$$

Recall r_t is the log return from t to $t+1$ to a one-period bond. We are interested in the period-t expectation of this excess return. Using the factor transformation into level, slope, and curvature factors of (26), a little algebra reveals

$$\begin{aligned} E_t\left(xr^{(m)}_{t,t+k}\right) &= \left(mA_{f,m} - (m-k)A_{f,m-k} - kA_{f,1}\right) \\ &+ \left(mB'_{f,m} - (m-k)B'_{f,m-k}\left(K_f\right)^k - B'_{f,1}\left(I - \left(K_f\right)^k\right)\left(I-K_f\right)^{-1}\right) f_t. \end{aligned} \tag{32}$$

The term on the right of (32) containing A's is the mean expected log excess return. The vector premultiplying the state vector represents the sensitivity of the expected excess return to the level, slope, and curvature of the term structure.

When no-arbitrage is not imposed, the properties of (32) can be studied only for combinations of m and $m-k$ for which factor loadings have been estimated. For example, the parameter estimates of the unrestricted model in Section 4 allow us to calculate the components of (32) for, say, a five-year bond held for one year (because loadings of five-year and four-year bonds are estimated), but not for a five-year bond held for one quarter. By constrast, imposing no-arbitrage allows interpolation of $A_{f,m}$ and $B_{f,m}$ for all maturities.

Even if we restrict our attention to expected excess returns that can be calculated from an unrestricted model, no-arbitrage restrictions are likely to be modestly helpful in sharpening statistical inferences about expected returns to long-maturity bonds. When m is large relative to k, sampling errors in unrestricted estimates of $B_{f,m}$ and $B_{f,m-k}$ can introduce errors in expected excess returns calculated with (32).

I document this point using the same Monte Carlo simulations studied in the previous section. I ask how accurately the restricted and unrestricted models estimate the unconditional mean and standard deviation of the expected excess annual log return to a five-year bond. I also ask how accurately the models estimate the loadings of the expected return on the level, slope, and curvature of the term structure.

For each simulation and both types of models (restricted and unrestricted), estimates of the constant term and vector on the right of (32) are calculated for $m = 20$ and $k = 4$ (a five-year bond held for one year). The models' estimated parameters are also used to calculate the implied unconditional covariance matrix of the state vector. This covariance matrix, in combination with the factor loading in (32), implies an unconditional standard deviation of the expected excess return. Table 7 reports means and standard errors of these statistics across 1000 simulations. (The standard errors are the standard deviations of the statistics across the simulations.)

The table documents that the no-arbitrage restrictions are unimportant in estimating the unconditional mean and standard deviation of the expected excess return. The standard error of the unconditional mean is 35 basis points for both models. The standard error of the unconditional standard deviation is 62 basis points for both models. However, the restricted model is slightly more accurate than the unrestricted model in its estimates of factor loadings. The standard errors of the loadings on the level, slope, and curvature are roughly 10 percent larger for the unrestricted model.

6.3 No-arbitrage restrictions as a specification test

The Duffie-Kan restrictions can be valuable when used as a specification test. Recall that the unrestricted model does not impose no-arbitrage restrictions because unobserved components of investors' returns create deviations of prices and yields from the restrictions. These deviations can be measured using the transformation of Section 3.4. The three-factor unrestricted model estimated in Section 4.5 has deviations from Duffie-Kan restrictions of a few basis points of annualized yields. Deviations of this magnitude are plausibly interpreted as evidence of unobserved components of returns.

If, however, estimated deviations are large, such an interpretation is less plausible. An alternative interpretation of large deviations from Duffie-Kan restrictions is that both models are misspecified, either in the dynamics of the state (e.g., too few factors or non-Gaussian dynamics) or in the mapping from yields to the state. Thus these deviations play the role of an informal specification test of model dynamics. I illustrate this specification test by using the data described in Section 4.1 to estimate a two-factor Gaussian model.

I follow the procedure of Section 4.3 to calculate parameter estimates for the unrestricted model. I then transform these estimates by solving for the equivalent-martingale parameters that are consistent with mean yields and factor loadings of bonds with maturities of three months, three years, and five years. Mean yields and factor loadings of bonds with maturities of one, two, and four years are allowed to deviate from the Duffie-Kan restrictions.

To conserve space, individual parameter estimates are not reported. Fig. 3 displays mean yields and factor loadings given by rotating factors into "level" (five-year bond yield) and "slope" (five-year less three-month). As in Fig. 1, the lines are mean yields and factor loadings consistent with no-arbitrage. The parameters c_0 and C_1 equal the differences between the lines and the dots. The red lines are plus/minus two standard error bounds on these parameters.[11]

Uncertainty in c_0 and C_1 is larger here than in the three-factor model because the cross-sectional fit is worse with only two factors. The fitted deviations from no-arbitrage are also larger. Using (28), fitted deviations of the one-year yield in the sample range from -9 basis points to 27 basis points. Remember that these deviations are not the same as random measurement error in yields. They are deviations from no-arbitrage of covariances between the one-year yield and other yields. Put differently, Fig. 3 shows that the one-year yield overreacts to the five-year yield—at least, it reacts more than the Duffie-Kan restrictions imply it should react in a two-factor Gaussian model. One interpretation is that unmeasured components of returns to a one-year bond are large and vary systematically with the level of the term structure. A more plausible interpretation is that the two-factor Gaussian model is misspecified.

[11]The standard errors are constructed with Monte Carlo simulations. A restricted two-factor model is estimated. It is treated as truth and 1000 Monte Carlo simulations are used to calculate standard errors for the unrestricted model.

7 Conclusion

Imposing the no-arbitrage restrictions of Duffie and Kan (1996) does not improve the forecasting performance of a three-factor discrete-time Gaussian model, either in practice or in Monte Carlo simulations. This conclusion depends on the use of at least three factors. A natural question is whether it also depends on either the discrete-time framework or the Gaussian structure.

The logic motivating the irrelevance of the Duffie-Kan restrictions does not rely on normally-distributed discrete shocks to interest rates. Instead, it comes directly from the idea that in any n-factor affine model, yields are linear functions of a constant and n other yields. Deviations from this linear equation are so small that its parameters can be estimated with minimal uncertainty even without imposing cross-equation restrictions. Thus it appears that the results should apply more generally to the class of affine term structure models. In principle this conjecture can be tested using Monte Carlo simulations. Computational difficulties in the estimation of non-Gaussian models prevent me from studying this issue here.

This paper does *not* argue that no-arbitrage restrictions are unimportant. They are used for pricing, hedging, and studying the dynamic properties of expected excess returns. Even in the realm of forecasting, no-arbitrage restrictions are the tool used to impose assumptions on risk preferences that go beyond the law of one price. The argument here is that if we are interested in extracting information from the term structure for the purpose of forecasting, the no-arbitrage restriction is irrelevant (except as a specification test) unless we have some reason to impose such additional assumptions about the dynamic behavior of risk compensation.

Appendix A. Factor transformation

This appendix describes how the model is written in terms of level, slope, and curvature factors.

Starting with the measurement equation (9) and the normalized transition equation (10), 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:

$$T_2 = \begin{pmatrix} B_1' \\ B_8' \\ B_{20}' \end{pmatrix}, \qquad T_1 = \begin{pmatrix} 1 & -1 & 0 \\ 1 & -1/2 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad Z = T_1^{-1} T_2. \tag{33}$$

The new state vector is

$$f_t = Z x_t. \tag{34}$$

The corresponding measurement and transition equations are

$$y_t = A_f + B_f f_t + \eta_t, \quad A_f = A, \quad B_f = B Z^{-1}, \tag{35}$$

$$f_{t+1} = K_f f_t + \Sigma_f \epsilon_{t+1}, \quad K_f = Z D Z^{-1},\ \Sigma_f = \sqrt{Z \Sigma \Sigma' Z'}, \tag{36}$$

where the square root in (36) indicates a Cholesky decomposition.

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

$$f_{t+1} = \mu_f^q + K_f^q f_t + \Sigma_f \epsilon_{t+1}^q, \quad \mu_f^q = Z \mu^q,\ K_f^q = Z K^q Z^{-1}. \tag{37}$$

Finally, the short rate is defined using

$$\delta_{f,0} = \delta_0, \quad \delta_{f,1} = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}'. \tag{38}$$

This factor rotation is applied to both the restricted model and transformation (18) of the unrestricted model. For the latter model, deviations from the no-arbitrage restrictions are measured by

$$c_{f,0} = c_0, \quad C_{f,1} = C_1 Z^{-1}. \tag{39}$$

Appendix B. Estimation details

Estimation uses a fairly elaborate hands-on procedure.

1. Choose a initial parameter vector of starting values based on OLS estimation. The

procedure is described below.

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 vaues. 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.

Here is the procedure for determining starting parameter values. A VAR(1) is estimated 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.

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.

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.

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, 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.

Diebold, Francis X., Glenn D. Rudebusch, and S. Boragan Aruoba, 2006, The macroeconomy and the yield curve: a dynamic latent factor approach, *Journal of Econometrics* 131, 309-338.

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.

Duffie, Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379-406.

Estrella, Arturo, and Gikas A. Hardouvelis, 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.

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	1.000	−0.809	1.016	0.998	0.098
	(0.010)	(0.007)	(0.011)	(0.050)		
Three year bond	5.748	1.001	−0.291	0.612	0.999	0.052
	(0.006)	(0.004)	(0.006)	(0.027)		
Four year bond	5.953	1.012	−0.112	0.249	0.999	0.056
	(0.006)	(0.004)	(0.006)	(0.029)		

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 f_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

$$f_{t+1} = K_f x_t + \Sigma_f \epsilon_{t+1}, \qquad f_{t+1} = \mu_f^Q + K_f^Q x_t + \Sigma_f \epsilon_{t+1}^q.$$

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 $c_{f,0} + C_{f,1} f_t$. 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	1.478			1.476		
(%/quarter)	(0.350)			(0.348)		
K_f	0.969	−0.107	0.079	0.968	−0.106	0.084
	(0.048)	(0.067)	(0.373)	(0.046)	(0.069)	(0.375)
	0.091	0.856	−1.294	0.086	0.872	−1.242
	(0.042)	(0.059)	(0.298)	(0.042)	(0.060)	(0.300)
	0.006	−0.001	0.785	0.007	0.000	0.766
	(0.010)	(0.012)	(0.094)	(0.010)	(0.012)	(0.096)
$\Sigma_f \times 10^3$	1.557	0	0	1.542		
	(0.118)			(0.118)		
	0.781	0.790	0	0.765	0.791	
	(0.109)	(0.066)		(0.109)	(0.065)	
	0.344	0.045	0.190	0.357	0.057	0.201
	(0.037)	(0.029)	(0.027)	(0.038)	(0.029)	(0.027)
$\mu_f^Q \times 10^4$	2.318			2.301		
	(0.445)			(0.447)		
	−8.112			−8.442		
	(3.331)			(3.343)		
	−2.399			−2.865		
	(1.087)			(1.142)		
K_f^Q	0.997	0.074	−0.056	0.997	0.076	−0.056
	(0.001)	(0.001)	(0.003)	(0.001)	(0.002)	(0.005)
	0.020	0.965	−1.398	0.009	0.986	−1.285
	(0.020)	(0.013)	(0.108)	(0.011)	(0.016)	(0.146)
	0.017	0.042	0.372	0.011	0.051	0.386
	(0.006)	(0.009)	(0.059)	(0.007)	(0.010)	(0.103)
$c_{f,0} \times 10^4$	-			−0.610		
				(0.251)		
				0.699		
				(0.229)		
$C_{f,1}$	-			−0.004	0.020	0.032
				(0.007)	(0.011)	(0.059)
				0.008	0.012	−0.042
				(0.006)	(0.009)	(0.044)
$\sigma_\eta \times 10^4$	1.394			1.306		
	(0.061)			(0.061)		

Table 3. Estimates of the Diebold, Rudebusch, and Aruoba term structure model

The three elements of a state vector x_t are 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). Their physical dynamics are

$$x_{t+1} = \mu + Kx_t + \Sigma\epsilon_{t+1}.$$

The mapping from the state vector to yields is

$$y_{m,t} = B(m,\lambda)'x_t + \eta_{m,t},\ \ \eta_{m,t} \sim N(0, \sigma_\eta^2),$$

where the function $B(m,\lambda)$ is described in the text. The model is 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. Yields are continuously compounded at quarterly rates. Standard errors, computed from Monte Carlo simulations, are in parentheses.

Log likelihood	3891.26		
$\mu' \times 10^3$	0.704	−0.554	−0.008
	(0.108)	(0.073)	(0.026)
K	0.976	−0.101	0.033
	(0.063)	(0.090)	(0.391)
	0.089	0.855	−1.252
	(0.043)	(0.066)	(0.290)
	0.006	0.002	0.799
	(0.015)	(0.024)	(0.101)
$\Sigma \times 10^3$	1.531	0	0
	(0.118)		
	0.741	0.780	0
	(0.108)	(0.066)	
	0.367	0.050	0.173
	(0.039)	(0.031)	(0.030)
λ	0.261		
	(0.030)		
$\sigma_\eta \times 10^4$	1.604		
	(0.090)		

Table 4. 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.

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

Table 5. Monte Carlo simulations of differences in 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 three term structure models. Model 1 imposes no-arbitrage restrictions and has the same structure as the true model. Model 2 does not impose these restrictions, and therefore nests the true model. Model 3 imposes the restrictions of Diebold, Rudebusch, and Aruoba, which are inconsistent with the true model. Each model is 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 Model 1 and forecasts produced with either Model 2 or Model 3. All yields are expressed in annualized percentage points.

Horizon	Level forecasts		Slope forecasts		Curvature forecasts	
(quarters)	Model 2	Model 3	Model 2	Model 3	Model 2	Model 3
1	0.015	0.077	0.009	0.048	0.009	0.037
2	0.025	0.138	0.011	0.054	0.011	0.051
3	0.034	0.194	0.012	0.061	0.012	0.061
4	0.043	0.249	0.016	0.077	0.013	0.068
5	0.052	0.302	0.021	0.100	0.014	0.074
6	0.061	0.354	0.026	0.126	0.014	0.078
7	0.070	0.407	0.030	0.149	0.014	0.081
8	0.079	0.458	0.034	0.169	0.015	0.083
9	0.088	0.508	0.037	0.185	0.015	0.084
10	0.096	0.556	0.039	0.197	0.015	0.085
11	0.104	0.602	0.041	0.206	0.015	0.086
12	0.112	0.645	0.042	0.211	0.015	0.086

Table 6. Monte Carlo simulations of RMSEs for 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. Model 1 imposes no-arbitrage restrictions and has the same structure as the true model. Model 2 does not impose these restrictions, and therefore nests the true model. Model 3 imposes the restrictions of Diebold, Rudebusch, and Aruoba, which are inconsistent with the true model. Each model is 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	Level forecasts			Slope forecasts			Curvature forecasts		
(quarters)	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
1	0.639	0.639	0.644	0.458	0.458	0.458	0.179	0.179	0.183
2	0.907	0.907	0.921	0.572	0.572	0.575	0.232	0.232	0.238
3	1.091	1.091	1.111	0.639	0.640	0.640	0.254	0.254	0.261
4	1.273	1.272	1.298	0.712	0.712	0.713	0.277	0.276	0.286
5	1.408	1.406	1.443	0.763	0.763	0.763	0.274	0.274	0.282
6	1.553	1.551	1.593	0.820	0.820	0.820	0.290	0.290	0.301
7	1.666	1.663	1.713	0.865	0.865	0.866	0.287	0.287	0.297
8	1.751	1.748	1.802	0.875	0.875	0.878	0.292	0.292	0.302
9	1.850	1.847	1.911	0.893	0.893	0.892	0.289	0.290	0.301
10	1.951	1.949	2.014	0.911	0.910	0.908	0.291	0.292	0.301
11	2.023	2.020	2.096	0.910	0.909	0.905	0.290	0.291	0.303
12	2.144	2.142	2.217	0.917	0.916	0.917	0.297	0.298	0.307

Table 7. Monte Carlo simulations of expected excess returns

This table summarizes the results of 1000 Monte Carlo simulations. Each simulation consists of maximum likelihood estimation of two three-factor term structure models. The restricted model imposes the Duffie-Kan restrictions and the unrestricted model does not. The simulated data are 88 quarters of bond yields generated by a model that imposes no-arbitrage. A model's parameter estimates are used to compute properties of the expected excess annual log return to a five-year zero-coupon bond. To express these properties, the term structure factors are rotated into 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).

		Restricted model		Unrestricted model	
Property	Truth	Mean	Std error	Mean	Std error
Unconditional mean (%)	1.601	1.600	0.349	1.598	0.353
Unconditional std dev (%)	2.961	2.879	0.619	2.873	0.618
Loading on:					
Level	3.571	4.373	2.000	4.326	2.040
Slope	9.920	10.070	2.817	10.118	2.859
Curvature	−27.211	−23.297	15.468	−22.875	15.471

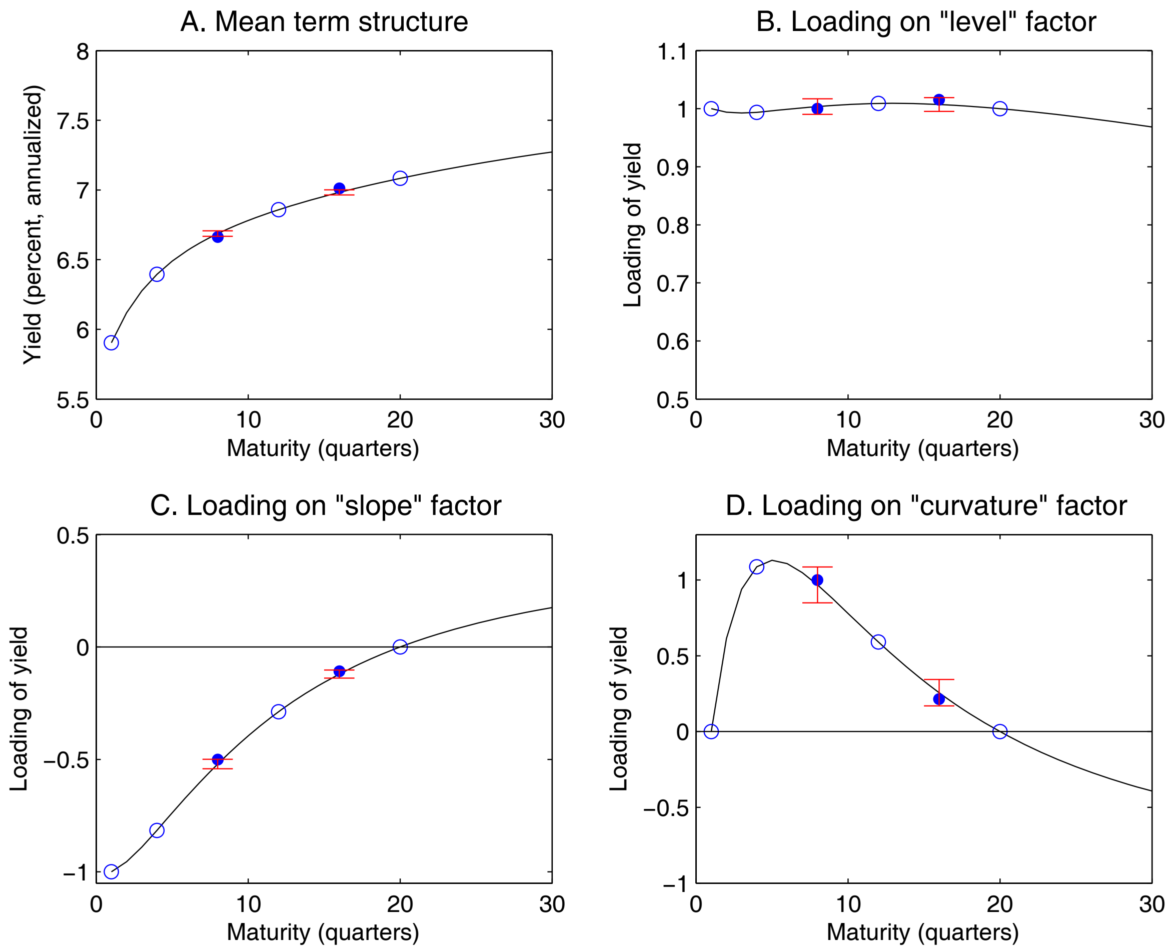

Fig. 1. Parameter estimates of a term structure model. A three-factor Gaussian model is estimated using quarterly Treasury bond yields from 1985Q1 through 2006Q4. The factors are rotated into level, slope, curvature. The lines in each panel are the no-arbitrage means and factor loadings consistent with 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.

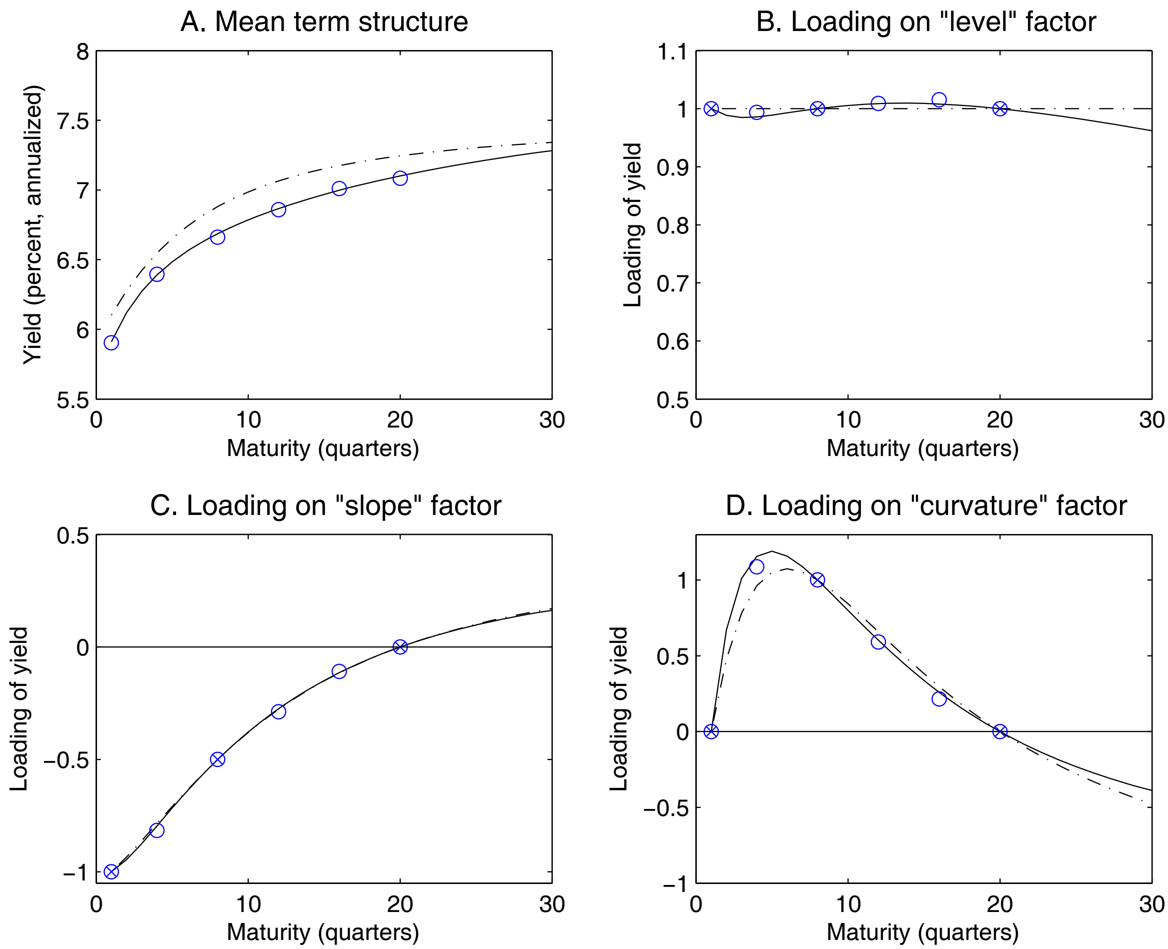

Fig. 2. Parameter estimates of three term structure models. Different three-factor Gaussian models are estimated using quarterly Treasury bond yields from 1985Q1 through 2006Q4. The factors are rotated into level, slope, curvature. The circles are estimates of mean yields and factor loadings from an unrestricted model. The solid lines are means and loadings implied by a model that imposes no-arbitrage. The dotted-dashed lines are implied by the Diebold, Rudebusch, and Aruoba model. 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.

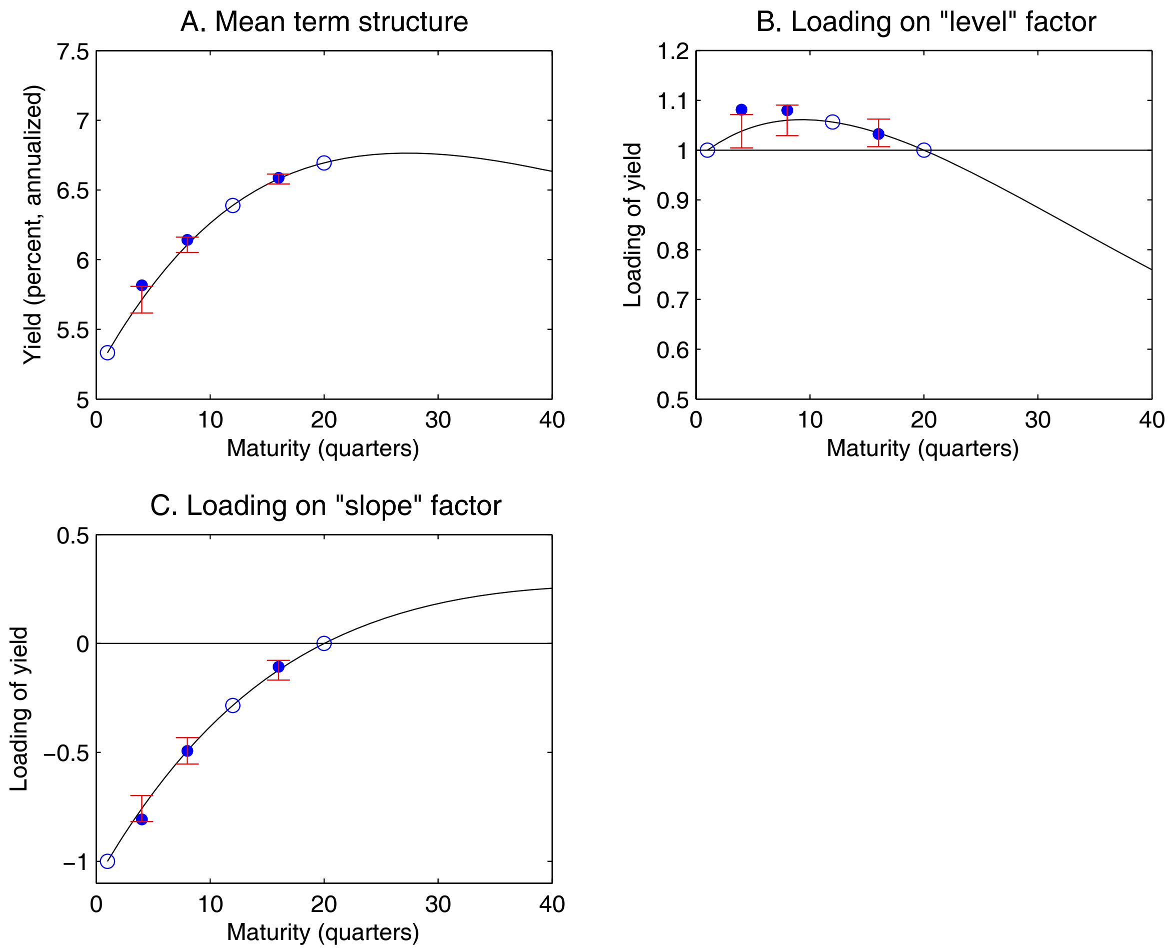

Fig. 3. Parameter estimates of a term structure model. A two-factor Gaussian model is estimated using quarterly Treasury bond yields from 1985Q1 through 2006Q4. The factors are rotated into level and slope. The lines in each panel are the no-arbitrage means and factor loadings consistent with yields on three-month, three-year, and five-year bonds. Means and factor loadings for one-year, 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.