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

Gregory R. Duffee∗

Haas School of Business
University of California – Berkeley

First draft: October 17, 2007
This Draft: October 29, 2007

ABSTRACT

Does imposing no-arbitrage help when using the term structure to forecast? Standard intuition says the restrictions should be imposed to increase estimation efficiency when we are confident the restrictions are correct, but ignoring them is often preferable if they are likely to be false. This paper argues that something close to the reverse is true, at least in the context of Gaussian models. Imposing the restrictions does not noticeably increase forecast accuracy when they are true, but examination of no-arbitrage restrictions can help detect misspecification of a broader model that nests the restrictions.

∗Voice 510-642-1435, email duffee@haas.berkeley.edu. Address correspondence to 545 Student Services Building #1900, Berkeley, CA 94720. Thanks to participants at a Wharton brown bag seminar.
1 Introduction

The beliefs of forward-looking investors determine Treasury yields. Thus it is not surprising that the term structure contains information about both future interest rates and macroeconomic conditions. Researchers have long attempted to exploit this information econometrically, using techniques such as straightforward predictive regressions, vector autoregressions, dynamic factor analysis, and structural macroeconomic models.

No-arbitrage dynamic term structure models appear to be a powerful addition to the econometrian’s toolkit. The starting point of these models is the assumption that the dynamics of the entire term structure are driven by a few common factors. The requirement of no-arbitrage imposes restrictions on the relation between the factors and bond yields, which should reduce the problem of overfitting. Although in principle these restrictions are complicated to impose, they are particularly easy to handle within the class of affine term structure models. Duffie and Kan (1996) derive the relevant restrictions.

Recent empirical work indicates that the Duffie-Kan restrictions are valuable in forecasting. 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 techniques that do not impose no-arbitrage. 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.

However, it is not clear that the greater forecast accuracy is actually a consequence of imposing no-arbitrage. These comparisons involve models that differ along multiple dimensions. For example, Duffee (2002) compares interest-rate forecasts from a three-factor Gaussian term structure model to those produced by a univariate forecasting regression. Is the greater accuracy of the former model a result of imposing no-arbitrage or simply a consequence of using a richer set of dynamics? To answer this question properly, we need to compare models that differ only in the imposition of no-arbitrage restrictions. Put differently, we want to nest no-arbitrage affine models in a slightly broader class of models that relaxes only the Duffie-Kan restrictions.

I construct these less restrictive models. The critical assumption underlying this class is that investors’ payoffs to holding bonds are not necessarily entirely captured by bond prices. Part of the return to bondholders may be unobserved, such as the ability to borrow at below-market rates in the RP market. Because we cannot observe the entire return to investors, we cannot impose Duffie-Kan restrictions on the part of the return that we do observe. This class of models allows us to construct statistical tests of the Duffie-Kan restrictions. It also allows us to evaluate the effects of these restrictions on forecasting performance.
The main conclusion is that the no-arbitrage restrictions of Gaussian models (and presumably affine models in general) are, in practice, effectively irrelevant to forecasting performance. The main evidence consists of Monte Carlo simulations, but the intuition is transparent. In affine models, bond yields are affine functions of a length-$n$ state vector. If we take this feature of affine models literally, then the yield on an $m$-maturity bond can be written as an exact affine function of yields on $n$ other bonds. Duffie and Kan derive cross-bond restrictions on the loadings of this “yield-factor” model.

But given data on all of these yields, the factor loadings can be precisely calculated without using any information about no-arbitrage. Simply regress the maturity-$m$ yield on a constant and the yields of the $n$ bonds. There is no estimation error in the regression because the $R^2$ is one. In practice, these $R^2$s are not quite one, which is why empirical applications of term structure models include measurement error in yields. Yet with a reasonable choice of $n$ (say, three), the variances of measurement errors are tiny relative to the variances of yields. For example, regressing the four-year yield on the three-month, two-year, and five-year yields using quarterly data from 1985 through 2006 results in an $R^2$ in excess of 0.999. The standard errors on the factor loadings are correspondingly small—too small to give a meaningful advantage to models that impose no-arbitrage.

I use Monte Carlo simulations to formally study the issue of out-of-sample accuracy of interest rate forecasts. I first estimate restricted and unrestricted three-factor Gaussian models using quarterly data from 1985 through 2006. From an economic perspective, the estimated models are almost indistinguishable. Differences between the models’ factor loadings amount to only a few basis points in implied bond yields. Yet because the standard errors on these differences are so small, statistical tests overwhelmingly reject the Duffie-Kan restrictions. Monte Carlo simulations treat the estimates of the restricted model as truth. Using 22 years of simulated quarterly data, I estimate both the restricted and unrestricted restrictive forms of the model. Simulations are used to calculate root mean squared forecast errors of the term structure’s level, slope, and curvature for one to twelve quarters ahead. Differences between RMSEs of the restricted and unrestricted models are never greater than a basis point.

Is there any role for no-arbitrage restrictions in forecasting? The results here suggest that they can be used as an informal specification test of the broader class of models. Under the maintained hypothesis of this broader class, any differences between the restricted and unrestricted models must be explained by unobserved components of returns. This puts an informal bound on the economic distance between the two models. If fitted yields from the models differ by only a few basis points, this maintained hypothesis is plausible. If, however, these differences are economically large, a more reasonable interpretation is that some other form of misspecification is present.
The next section describes the unrestricted and restricted models. 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 term structure models are a state vector and its physical measure dynamics, a short-term interest rate that is a function of the state, and equivalent-martingale dynamics of the state vector. In the affine class, the short-rate function and the state dynamics are chosen so that zero-coupon bond yields are affine functions of the state.

The class of affine models is large. In this paper I focus on the special case of discrete-time Gaussian models. This choice is necessitated by 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

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

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 bond yields are affine functions of the state vector plus an idiosyncratic component. Denoting the continuously-compounded yield on an \( m \)-maturity bond by \( y_t^{(m)} \), yields are

\[
    y_t^{(m)} = A_m + B'_m x_t + \eta_{m,t}, \quad \eta_{m,t} \sim N(0, \sigma^2_{\eta}). \tag{2}
\]

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
\[ r_t = \delta_0 + \delta'_1 x_t + \eta_{r,t}, \quad \eta_{r,t} \sim N(0, \sigma^2_\eta). \]  \hspace{1cm} (4)

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

### 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.\(^1\) 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 true yields 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 essentially affine Gaussian framework of Duffee (2002), the equivalent martingale measure dynamics of \( x_t \) are

\[ x_{t+1} = \mu^q + K^q x_t + \Sigma^q_{t+1}, \quad \epsilon^q_{t+1} \sim MVN(0, I). \]  \hspace{1cm} (5)

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

by

\[ B'_m = \mathbb{B}(m; \delta_1, K^q)' = \frac{1}{m} \delta'_1 (I - K^q)^{-1} (I - (K^q)^m) . \tag{6} \]

The constant term for \( m > 1 \) is

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

I refer to equations (6) and (7) as the Duffie-Kan restrictions.

The essence of the no-arbitrage restrictions is that in an \( n \)-shock model, any one bond can be priced in terms of \( n + 1 \) other bond prices. (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 \( n + 1 \) 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 \( f_t \) of variables such as inflation, output growth, and the
unemployment rate. The relation between the macro variables and the state vector is

\[ f_t = A_f + B_fx_t + \eta_{f,t}. \]  

(8)

Adding this affine relation allows us to use the model to forecast future realizations of \( f_t \).

Given the objectives of this paper, there is no reason to include (8). There are no Duffie-Kan restrictions associated with \( A_f \) and \( B_f \). 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 \( f_t \). Conversely, if imposing the restrictions affects estimated factor loadings of bond yields, the estimated dynamics of \( x_t \) are also likely to be affected. In this case, the restrictions will indirectly affect macroeconomic forecasts.

### 3 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 \( d \) 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

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

(9)

In (9), \( A \) is a \( d \)-vector and \( B \) is a \( d \times n \) matrix. This state space formulation is underidentified because the state vector is unobserved. For estimation purposes, it is convenient to normalize the transition equation to

\[ x_{t+1} = Dx_t + \Sigma \epsilon_{t+1}, \]  

(10)

where \( D \) is diagonal and \( \Sigma \) is lower triangular with ones along the diagonal. There is nothing economically interesting about this normalization; it is simply the easiest to use in estimation. (An additional normalization orders the diagonal of \( D \), but I do not apply this in estimation.) In Section 4, a different normalization is used to explain the empirical results.
3.2 The hypotheses

The null hypothesis is that the Duffie-Kan restrictions hold. Formally, this hypothesis is

\[ H_0 : \ A = A(M; \delta_0, \delta_1, \mu^q, K^q, \Sigma) = \begin{pmatrix} A(m_1; \cdot) \\ \vdots \\ A(m_d; \cdot) \end{pmatrix}; \]

\[ B = B(M; \delta_1, K^q) = \begin{pmatrix} B(m_1; \cdot)' \\ \vdots \\ B(m_d; \cdot)' \end{pmatrix}. \] (11)

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

\[ H_1 : \ A, B \text{ unrestricted}. \] (12)

For estimation purposes, the parameters of the model that imposes no-arbitrage are

\[ \rho_0 = \{\delta_0, \delta_1, D, \Sigma, \mu^q, K^q, \sigma_\eta^2\}. \] (13)

I refer to this model as the “restricted” model. The parameters of the “unrestricted” model are

\[ \rho_1 = \{D, \Sigma, A, B, \sigma_\eta^2\}. \] (14)

There are \(2 + 3n + n^2 + n(n - 1)/2\) parameters in \(\rho_0\) and \(1 + n + (n + 1)d + n(n - 1)/2\) parameters in \(\rho_1\). Thus there are \((1 + n)(d - n - 1)\) overidentifying restrictions. (Recall that the number of observed bond yields \(d\) exceeds \(n + 1\).)

3.3 A useful transformation of the alternative hypothesis

A likelihood ratio test statistically evaluates \(H_0\) versus \(H_1\). However, there is a more intuitive way to compare these two hypotheses. We can almost always write the unrestricted parameters \(A\) and \(B\) as sums of two pieces. One piece represents parameters consistent with no-arbitrage, while the other piece represents deviations from the no-arbitrage restrictions.

The procedure begins by splitting observed yields into two vectors. The first, denoted \(y^x_t\), is an \((n + 1)\)-vector of yields assumed to satisfy exactly the usual no-arbitrage restrictions. (The superscript \(x\) denotes eXact.) The second, denoted \(y^v_t\) (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

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

(15)

$$
A^x = A(M^x; \delta_0^1, \delta_1^1, \mu^{q^1}, K^{q^1}, \Sigma),
$$

(16)

$$
B^x = B(M^x; \delta_1^1, K^{q^1}),
$$

(17)

$$
A^v = A(M^v; \delta_0^1, \delta_1^1, \mu^{q^1}, K^{q^1}, \Sigma) + c_0,
$$

(18)

$$
B^v = B(M^v; \delta_1^1, K^{q^1}) + C_1.
$$

(19)

The parameters $\delta_0^1, \delta_1^1, \mu^{q^1}$, and $K^{q^1}$ reconcile the “exact-identification” bond yields with the absence of arbitrage. The parameters $c_0$ and $C_1$ are the deviations of the other bond yields from no-arbitrage.

To implement this representation, invert the functional form of the $(n + 1) \times n$ matrix $B^x$ to determine implied equivalent-martingale parameters $\delta_1^1$ and $K^{q^1}$:

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

(20)

The inverse mapping in (20) is done numerically. There are values of $B^x$ which cannot be inverted using (20). If inversion is impossible for one set of bonds that comprise the “exact” group, a different set of bonds can be used. The remaining equivalent-martingale parameters are determined numerically by the inversion

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

(21)

After calculating these equivalent-martingale parameters, we can write the parameters $A^v$ and $B^v$ in (17) and (18) as the sum of parameters implied by no-arbitrage 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 (14) to

$$
\rho_1^\dagger = \{D, \Sigma, \delta_0^1, \delta_1^1, \mu^{q^1}, K^{q^1}, c_0, C_1, \sigma_\eta^2\}.
$$

(22)

2In 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.
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 no-arbitrage restrictions can be imposed, any $d - n - 1$ yields must be set consistently with the other $n + 1$ yields.

### 3.4 Discussion

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 alternative model and examine the standard errors of $c$ and $C$ in (17) and (18). These are free parameters under the alternative hypothesis but not under the null. If the standard errors on these parameters are large, fixing them to zero (when this restriction is true) represents a substantial increase in efficiency. If the standard errors are tiny, the efficiency gains are modest.

The terminology “unrestricted model” is a bit of a misnomer. Although not as restrictive as the null hypothesis, the alternative hypothesis (12) imposes strong limitations 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. If these common factors were the only factors allowed to affect observed yields, the $d \times d$ covariance matrix of observed yields would have rank $n$. The role of the idiosyncratic shock is to weaken this requirement, and thus allow an observed set of data to have a nonzero likelihood.

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.

### 4 Empirical estimation

There are two empirical questions addressed in this paper. First, is the behavior of the Treasury term structure consistent with the Duffie-Kan restrictions of the discrete-time Gaussian model? Second, are out-of-sample forecasts of Treasury yields noticeably improved by imposing the Duffie-Kan restrictions when they are true? This section answers the first question and the next section answers the second.
4.1 Data

The empirical analysis uses quarterly data from 1985 through 2006. The choice of this relatively short sample period is motivated by two considerations. First, parameter restrictions are more likely to play an important role in estimation when using a small sample than a large sample. I want to give the Duffie-Kan restrictions a reasonable opportunity to bite. Second, there is considerable evidence of a regime switch during the late 1970s and early 1980s. The post-disinflation period is a more homogeneous sample.

I use quarter-end observations of yields on zero-coupon Treasury bonds with maturities of three months and one through five years. 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 A three-factor transformation

Since Litterman and Scheinkman (1991), financial economists have usually viewed the dynamics of Treasury yields in terms of “level,” “slope,” and “curvature” factors. This paper follows much of the no-arbitrage literature by using three state variables. The vector $x_t$ is latent, which allows us to rotate it into any convenient interpretation. To help interpret the parameter estimates, I rotate the vector to roughly correspond to level, slope and curvature.

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 $Z$:

$$
T_2 = \begin{pmatrix}
B_{1}' & B_{8}' & B_{20}'
\end{pmatrix}, \quad T_1 = \begin{pmatrix}
1 & -1 & 0 \\
1 & -1/2 & 1 \\
1 & 0 & 0
\end{pmatrix}, \quad Z = T_1^{-1}T_2. \tag{23}
$$

The new state vector is

$$
x_t^* = \begin{pmatrix}
\tilde{y}_t^{(20)} - E\tilde{y}_t^{(20)} \\
(\tilde{y}_t^{(20)} - E\tilde{y}_t^{(20)}) - (\tilde{y}_t^{(1)} - E\tilde{y}_t^{(1)}) \\
(\tilde{y}_t^{(8)} - E\tilde{y}_t^{(8)}) - \frac{1}{2} ( (\tilde{y}_t^{(1)} - E\tilde{y}_t^{(1)}) + (\tilde{y}_t^{(20)} - E\tilde{y}_t^{(20)}) )
\end{pmatrix} = ZX_t. \tag{24}
$$

The factors are versions of the level, slope, and curvature. 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). The corresponding measurement and transition equations are

\[ y_t = A + B^* x_t^* + \eta_t, \quad B^* = B Z^{-1}, \]  
\[ x_{t+1}^* = K^* x_t^* + \Sigma^* \epsilon_{t+1}, \quad K^* = Z D Z^{-1}, \Sigma^* = \sqrt{Z \Sigma \Sigma' Z'}, \]  

where the square root in (25) indicates a Cholesky decomposition.

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

\[ x_{t+1}^* = \mu^q + K^q x_t^* + \Sigma^* \epsilon_{t+1}, \quad \mu^q = Z \mu^*, \quad K^q = Z K^q Z^{-1}. \]  

Since the short rate is the three-month yield, the constant term \( \lambda_0 \) is the mean short rate and the loading of the short rate on the state vector is

\[ \delta_1 = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}. \]  

4.3 A preliminary look at bond yields

The measurement equation (25) says that all yields are affine functions of the level, slope, and curvature, plus noise. These functions can be approximated by replacing the latent vector \( x_t^* \) with its observable counterpart. For each maturity \( m \), the approximate function is

\[ y_t^{(m)} = a_m + b_m \begin{pmatrix} y_t^{(20)} - y_t^{(1)} \\ y_t^{(8)} - y_t^{(8)} \\ y_t^{(1)} - y_t^{(1)} \end{pmatrix} + e_t \]  

where the bars indicate sample means. We can think of (29) 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 (29) to the one-year, three-year, and four-year bond yields. The three factors explain almost all of the variation in the dependent yields. The adjusted \( R^2 \)s range from 0.998 to 0.999. The standard errors of the point estimates are correspondingly small. The estimated factor loadings range from around one to minus one (a consequence of the definition of the factors). The standard errors for level and slope range from 0.004 to 0.011. The standard errors for curvature are somewhat higher because, as seen in Panel A, curvature contributes relatively little to the variation in yields.

These regression results foreshadow what we will see in Section 5. Imposing cross-
equation restrictions on factor loadings is of little practical importance under the assumption that the restrictions are correct. 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.

### 4.4 Model estimation details

I estimate both unrestricted and restricted three-factor versions of the model. When estimating the models I use the factor rotation (9) and (10). There are six bond yields observed at each of 88 quarterly observations. The unrestricted model has 31 free parameters and the restricted model has 23 free parameters.

The likelihood functions are maximized with Matlab. The method is

1. Choose an initial vector of starting values. Details are in the appendix.
2. Given a vector of starting values, use five successive rounds of Simplex optimization. Each round uses 4000 iterations. A derivative-based optimizer with analytic first derivatives refines the parameter estimates. The function tolerance for the derivative method is $10^{-12}$.
3. Repeat the previous step 99 times, using starting values that are drawn from a multivariate normal distribution with a mean given by the vector from Step 1.
4. Choose the parameter vector with the highest likelihood among these 100 optimizations. Using this as a starting value, repeat Step 2.
5. Repeat Step 4 at the new parameter vector. In practice, the results are unaffected by this step. Therefore the parameter estimates from Step 4 are treated as the maximum likelihood estimates.

It is worth noting that estimation of the unrestricted model takes about half the time necessary to estimate the restricted model. Estimation of the unrestricted model is also better behaved, in the sense that most of the 100 separate maximization problems in Step 3 result in the same value of the likelihood.

After estimation, the estimated measurement equation is transformed into the equivalent representation (15), where the three-month, one-year, three-year, and five-year bonds are used to exactly identify an equivalent-martingale measure. Deviations from no-arbitrage are
allowed in the two-year and four-year bond yields. Finally, the factors are rotated into level, slope and curvature, resulting in measurement and transition equations (25) and (26). This factor rotation is also applied to the restricted model.

4.5 Estimation results

Table 2 reports parameter estimates for the rotation (25) and (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 loadings are six nonlinear restrictions on the reported parameter estimates. Thus the covariance matrix of the reported estimates is singular. Standard errors are in parentheses. They are computed from Monte Carlo simulations and are discussed in detail in Section 5.

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 raises the precision of almost all the point estimates, but this effect is economically important only for estimates of mean yields.

4.5.1 The economic importance of the restrictions

The vector $c_0$ and the matrix $C_1$ of the unrestricted model capture deviations from Duffie-Kan restrictions. The estimate of $c_0$ implies that mean yields on the two-year and four-year bonds deviate from no-arbitrage by two to three basis points of annualized yields. (Recall the reported parameters are in units of quarterly yields.) Deviations in factor loadings are economically even smaller. Visual evidence is in Fig. 1. The circles are the means and loadings of the three-month, one-year, three-year, and five-year bonds yields. The lines are drawn by calculating the equivalent-martingale parameters consistent with the circles. The dots are the means and loadings of the two-year and four-year bond yields. The parameters $c_0$ and $C_1$ equal the differences between the lines and the dots. They are almost undetectable in the figure.

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

$$\text{fitted deviation}_t = c_0 + C_1 \hat{x}_t. \quad (30)$$

In (30), $\hat{x}_t$ represents the filtered values of the state vector. Across the 88 quarters in the
sample, absolute fitted deviations never exceed 7.5 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.

4.5.2 The statistical importance of the restrictions

The likelihood ratio test statistic of the Duffie-Kan restrictions is 35.56, which rejects the
null hypothesis at any conceivable asymptotic significance level. The source of the rejection
is largely the mean yields. The standard errors on the two elements of $c_0$ are both less
than a basis point of annualized yield. The standard errors of $C_1$ are also quite small, but
the individual $t$-statistics are typically less than two in absolute value. Two-standard-error
bounds on the estimates of $c_0$ and $C_1$ are displayed in red in Fig. 1.

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

A quick comparison of the two sets of standard errors in Table 2 reveals that for all but
one parameter (the standard deviation of measurement error) the standard errors for the
restricted model are smaller than those for the unrestricted model. This is not surprising.
However, a more detailed examination reveals that only parameters associated with uncondition-
ditional means have economically large differences in uncertainty. For example, the standard
error of the mean short rate is 80 basis points for the restricted model (multiplying the stan-
dard error in Table 2 by four to put it in terms of annualized yield) and 117 basis points for
the unrestricted model. The corresponding comparison for the mean five-year yield, which
incorporates uncertainty in mean risk compensation, is 96 basis points and 126 basis points
respectively.4

Differences in uncertainty in dynamics are more muted. In particular, consider the level
factor, which accounts for a large majority of the variation in yields. According to the
restricted model, the estimated standard deviation of shocks to this factor has a standard
error of 4.76 basis points (annualized). The corresponding standard error for the unrestricted
model is 4.80 basis points. The standard error of the one-quarter persistence of this shock is
small for both models; 0.029 with the restricted model and 0.032 with the unrestricted model.

---

3The finite-sample five percent critical value, based on the Monte Carlo simulations in Section 5, is 19.80.
4These standard errors cannot be read directly off of Table 2. They are discussed in more detail in
Section 5.
(These are the standard errors of the upper left element of $K$.) Finally, consider the loadings of yields on this factor. Because of the chosen factor rotation, the loadings on this factor of the three-month, two-year, and five-year yields are all normalized to one. To illustrate uncertainty in factor loadings for other bonds, consider the three-year bond. For both restricted and unrestricted models, the point estimate of the loading is 1.009 and is estimated with high precision. The restricted model’s standard error is 0.003. The unrestricted model’s standard error of 0.007 is more than twice as large, but is economically tiny, as we saw in Panel B of Fig. 1.

4.5.3 Comparing forecasts of future interest rates

Given the small economic importance of deviations from no-arbitrage, we anticipate that the parameter estimates of the restricted model will be very close to those of the unrestricted model. A comparison of the two sets of parameter estimates in Table 2 generally confirms this. However, the estimated models do not quite agree on mean yields. 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. (Ignore the dotted-dashed lines.) The chosen factor rotation implies that the loadings of the two models coincide at maturities of three months, two years, and five years. These points are marked with an $x$. Even for the factor loadings not marked with an $x$, the unrestricted loadings are indistinguishable from the loading functions of the restricted model. But the mean yield function from the unrestricted model lies a few basis points below the unrestricted mean yields.

The proper interpretation of this result is that mean yields are estimated with little precision. As mentioned in Section 4.5.2, standard errors for these mean yields are around one annualized percentage point. When estimated, the two models give slightly different weights to the information in each observation. These different weights lead to noticeably different mean yields because the values of the likelihood functions are insensitive to mean yields.

The near-equivalence of the two models carries over to their respective out-of-sample forecasts of yields. Fig. 3 displays these forecasts. The parameter estimates and filtered value of the 2006Q4 state vector are used to predict yields on three-month, two-year, and five-year bonds over the twenty-quarter span 2007Q1 through 2011Q4. The solid lines (the restricted model) and dashed lines (the unrestricted model) do not differ from each other by more than three basis points for any yield and forecast horizon. (Ignore the dotted-dashed lines.)

The evidence here tells us that Treasury yield behavior during 1985 through 2006 is
economically consistent (although statistically inconsistent) with the Duffie-Kan restrictions of a three-factor Gaussian model. Imposing the restrictions has only a minimal effect on forecasts of future interest rates. These conclusions lead to two questions. First, for this sample of data, would any reasonable three-factor Gaussian model generate nearly identical forecasts? Second, is it generally the case that the Duffie-Kan restrictions are irrelevant to forecasting? The first question is addressed in the next subsection and the second is addressed in Section 5.

4.6 An alternative model

Diebold, Rudebusch, and Aruoba (2006) slightly generalize the model of Diebold and Li (2006). In the DRA model, a three-factor latent vector has the Gaussian dynamics of Equation (1). The relation between factors and yields is

\[ A_m = 0, \]
\[ B'_m = \begin{pmatrix} 1 & \frac{1-e^{-m\lambda}}{m\lambda} & \frac{1-e^{-m\lambda}}{m\lambda} - e^{-m\lambda} \end{pmatrix}. \]

The DRA model is nested in the unrestricted model used here. However, it is not nested in the restricted model because it does not satisfy the Duffie-Kan restrictions.\(^5\)

I estimate this model using the same data and same numerical optimization procedure. There are 20 free parameters to estimate. The model is estimated using (1), (31), and (32). However, to simplify comparison with the other two models, I then transform the estimates by rotating the state vector into level, slope, and curvature, as defined in (24). Defining \( T_1, T_2, \) and \( Z \) as in (23), where the factor loadings in \( T_1 \) are taken from (32), the rotation is

\[ x^*_t = ZX_t. \]

This state vector is the non-demeaned version of the state vector used in the restricted and unrestricted models. Table 3 reports estimates of the dynamics of the rotated parameter vector along with estimates of \( \lambda \) and \( \sigma_\eta. \)

Statistically, the DRA model is 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. One sign of the poorer fit of the DRA model is in the standard deviation of measurement error. It is 1.3 basis points of quarterly yields in the unrestricted model (1.4 basis points in the restricted model), and 1.6 basis points in the DRA model. Other

\(^5\)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.
differences between the DRA model and the unrestricted model are revealed in Fig. 2. The figure displays mean yields and factor loadings for the DRA model using dotted-dashed lines. The mean yield curve is higher than it is for the unrestricted model, but as noted in the discussion of the restricted model, we should not read much into this because of low precision. Loadings on the level factor for the DRA model are economically close to those of the unrestricted model, but statistically the deviations are large. Loadings on the slope factor are indistinguishable for the DRA and unrestricted models. Loadings on the curvature factor do not match up as well.

Differences between the DRA model and the other two models are noticeable in out-of-sample forecasts displayed in Fig. 3. The forecasts of the DRA model, shown as dotted-dashed lines, differ from those of the other models by around five to fifteen basis points, depending on the maturity and the horizon. Qualitatively, the models all agree that long rates are expected to rise and short rates are expected to slightly fall, then rise. But the magnitude of the expected increase in rates is moderately lower for the DRA model.

In some sense, the point of the Diebold-Li model and its DRA extension is to produce different out-of-sample forecasts than the other two models studied here. As discussed by Diebold et al., the appeal of the Diebold-Li and DRA models is based on their parsimonious, yet empirically reasonable, specifications. Even though the DRA model is strongly rejected statistically, it may nonetheless produce more accurate forecasts because the danger of overfitting is lower. The next section uses Monte Carlo simulations to study out-of-sample forecasting performance.

5 Simulation evidence on forecasting performance

If term structure dynamics satisfy the Duffie-Kan restrictions, what is the loss in forecast accuracy of not imposing them? Does the DRA model, which is misspecified in such a setting, nonetheless produce more accurate forecasts owing to its parsimonious specification? Here I address these questions. The answers necessarily depend on both the true model and the sample size. Both are taken from the previous section. The estimated restricted model from Section 4.5 is the assumed true model. The sample size is 88 quarterly observations of yields on maturities with three months and one through five years.

The Monte Carlo procedure is straightforward. A single 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.

Results from 500 Monte Carlo simulations are used to calculate finite-sample standard errors for the parameter estimates in Tables 2 and 3. (Note that the standard errors for the DRA model are calculated assuming the restricted model, not the DRA model, is true.) Similarly, standard errors are calculated for population means of bond yields. Finally, root mean squared forecast errors are constructed across the Monte Carlo simulations for each forecast horizon and forecasted variable.

5.1 Some numerical optimization details

Unfortunately, it is important to discuss some details of the estimation procedure. If computer processing time were not a constraint, each model would be estimated using an extensive search such as that described in Section 4.4. But current computer speeds preclude such a method. The approach here is to use a single starting value for numerical optimization. Since we know the true parameters, it is sensible to use these as starting values in numerical estimation. Put differently, we are reasonably confident that the global maximum is in the neighborhood of truth.

Given a starting value, each likelihood function is maximized using a derivative-based optimizer and analytic first derivatives. The function tolerance is $10^{-12}$. This method is sufficient to locate the maximum for the DRA model. However, it is not sufficient for the restricted and unrestricted models. For these models, I also use truth as a starting value for five successive rounds of Simplex optimization, followed by the derivative-based optimizer. In the great majority of the simulations (but not all), this algorithm produces a higher likelihood than the single round of the derivative-based procedure. I use the set of parameters that produces the higher likelihood as the starting value for another five rounds of Simplex, followed by the derivative-based procedure.

The importance of refining the search is documented in Table 4. The table reports the mean and standard deviation, across 500 simulations, of the log-likelihood values of the restricted and unrestricted models. The single application of a derivative-based approach is denoted as Method A. The more intensive optimization procedure is denoted Method B.

Not surprisingly, Method B results higher log-likelihoods. The average increases for the restricted and unrestricted models are 1.8 and 1.4 respectively. The larger effect on the
restricted model is not an accident. Numerical optimization algorithms have greater difficulty with this model than with the unrestricted model. In particular, numerical optimization techniques tend to not move sufficiently far away from the starting point along the dimension of mean yields.

Consider the population mean of the five-year yield. Table 4 says that when using Method A, the restricted model estimate of this value is close to unbiased. The mean value, across simulations, is 7.06 percent, which is three basis points below the true population mean. But Method B raises the mean value by 12 basis points. It also increases the standard error from 0.78 percent to 0.96 percent. The effect of shifting from Method A to Method B is smaller for the unrestricted model. The mean value increases only two basis points and the standard error increases from 1.12 percent to 1.26 percent.

Given this evidence, it is not surprising that the finite-sample distribution of a likelihood ratio test is sensitive to the optimization method. Table 4 shows that the five percent and ten percent critical values are lower using Method B than Method A. Method B allows the restricted model to move further away from truth and closer to the true maximum.

The relevance of this comparison for forecasting performance is clear. Out-of-sample forecasts generated by models estimated using Method A will be more accurate than those based on Method B, because the estimated parameters will be closer to the true parameters. Moreover, the relative effect of estimation method on forecasting performance is greater for the restricted model than the unrestricted model. A similar conclusion can be drawn for simulation-based standard errors of parameter estimates. The standard errors of estimated parameters are lower for the restricted model than for the unrestricted model. However, differences between these standard deviations are smaller when parameter estimates are computed with Method B.

An even more extensive parameter search will allow us to determine whether it is sufficient to stop with Method B. It is underway, but not yet completed. The search across hundreds of simulations takes many weeks of processing time.

5.2 Results

Recall that in Section 4.5.3, the restricted and unrestricted models produce nearly identical out-of-sample forecasts. The first set of results helps us see whether that is generally true (assuming the restricted model is correct) or whether it is sample-specific. 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 500 simulations, the different models produce noticeably different 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 45 basis points (annualized). Replacing the unrestricted model with the DRA model raises this root mean squared difference to almost a full percentage point. The divergence between the DRA and restricted model forecasts is largely driven by misspecification in the DRA model, as discussed in the next set of results. The divergence between the restricted and unrestricted model forecasts has a different source.

The main reason for the differences in forecasts between the restricted and unrestricted models is that the two sets of estimates tend to disagree about unconditional mean yields. For these two models, the root mean squared differences in estimated population means are around 90 basis points for maturities ranging from three months to five years. (These results are not reported in any table.) The models forecasts all incorporate mean reversion, thus their forecasts diverge slightly at short horizons and diverge substantially at long horizons. However, differences in estimated population means are largely noise in both sets of parameter estimates. Thus these differences do not show up in differences in forecast accuracy.

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 noticeably lower-quality 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 basis point of annualized yield. Moreover, the restricted model does not always produce the more accurate forecast, although this result may be a consequence of the finite number of Monte Carlo simulations.

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 five 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 17 basis points higher than the other two models. Nonetheless, 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.

As noted at the beginning of this section, these results necessarily depend on the true model and the sample size. Econometric intuition suggests that the no-arbitrage restriction is likely to have more relevance when the cross-sectional fit of the model is worse. We can informally measure this fit using cross-sectional yield regressions as in Table 1. The Duffie-
Kan restrictions have a negligible effect on forecasts here because the “true” model is based on the 1985–2006 sample where cross-sectional adjusted $R^2$s are 0.998 and higher. If the baseline sample had lower $R^2$s, the Duffie-Kan restrictions would take on more relevance.

Although presumably such a data sample could be located, the $R^2$s in Table 1 are more the norm than the exception. For example, if the sample period for the regressions is extended back to 1952Q2 (the first quarter for which the CRSP longer-horizon yields are available), the corresponding adjusted $R^2$s range from 0.997 to 0.999. Almost identical results are obtained when using the Federal Reserve Board’s zero-coupon bond yields for maturities up to ten years. (These results are not reported in any table.) Thus the simulation evidence presented here appear to be robust to plausible variations in term structure behavior.

6 Are no-arbitrage restrictions useful?

Explicit no-arbitrage restrictions such as those of Duffie and Kan play important roles in finance. One is to determine the price of a particular instrument in terms of other instruments’ prices. If, say, we want to price a zero-coupon bond or a bond option, we need the Duffie-Kan restrictions. The unrestricted model of Section 2.1 is not helpful, at least not directly. That model says how to infer factor loadings given zero-coupon bond prices, but it does not tell you how to interpolate prices of other instruments.

The question addressed in this section is whether the Duffie-Kan restrictions can be helpful in forecasting. In one trivial sense, the answer is yes. The unrestricted model can only forecast yields on bonds included in the Kalman filter. If you want to forecast the yield on a bond for which you have no historical data, interpolation is necessary.

The restrictions are also necessary 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 consistently 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 $\mu$, $K$, and $\Sigma$. 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.

Finally, 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 3.3. 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 4.4 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. 4 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.\(^6\)

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 (30), fitted deviations of the one-year yield in the sample range from $-10$ to $15$ basis points.

\(^6\)The standard errors are constructed with Monte Carlo simulations. A restricted two-factor model is estimated. It is treated as truth and 500 Monte Carlo simulations are used to calculate standard errors for the unrestricted model.
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. 4 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.

7 Conclusion

Imposing the no-arbitrage restrictions of Duffie and Kan (1996) does not noticeably 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

Starting values for numerical optimization are chosen as follows. 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 \( \Sigma \) 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 \( \delta_0 \) is the sample mean of the short rate. The starting value for \( \delta_1 \) is the coefficient vector from a regression of the three-month yield on \( \hat{x}_t \). The starting value for \( \mu^q \) 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


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

<table>
<thead>
<tr>
<th>Level</th>
<th>Slope</th>
<th>Curvature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>6.076</td>
<td>1.318</td>
</tr>
<tr>
<td>Std dev</td>
<td>1.872</td>
<td>0.948</td>
</tr>
</tbody>
</table>

Panel B. Regression results

<table>
<thead>
<tr>
<th></th>
<th>Constant</th>
<th>Level</th>
<th>Slope</th>
<th>Curvature</th>
<th>Adj $R^2$</th>
<th>Resid SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>One year bond</td>
<td>5.183</td>
<td>1.000</td>
<td>−0.809</td>
<td>1.016</td>
<td>0.998</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.007)</td>
<td>(0.011)</td>
<td>(0.050)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three year bond</td>
<td>5.748</td>
<td>1.001</td>
<td>−0.291</td>
<td>0.612</td>
<td>0.999</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.006)</td>
<td>(0.027)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four year bond</td>
<td>5.953</td>
<td>1.012</td>
<td>−0.112</td>
<td>0.249</td>
<td>0.999</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.006)</td>
<td>(0.029)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Yields are affine functions of three factors with joint Gaussian dynamics. The factors $x_t$, are normalized to the five-year yield (level), the difference between the five-year yield and the three-month yield (slope), and the difference between the two-year yield and the average of the five-year and three-month yields (curvature). The factors are also demeaned. Their dynamics under the physical and equivalent-martingale measures are

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

Two models are estimated with the Kalman filter using quarter-end yields from 1985Q1 to 2006Q4. The maturities are three months and one, two, three, four, and five years. One model allows for affine deviations in yields from the no-arbitrage restrictions: the two-year and four-year yields deviate from no-arbitrage yields by $c_0 + C_1 x_t$. All yields are contaminated by iid measurement error with standard deviation $\sigma_\eta$. Yields are continuously compounded at quarterly rates. Standard errors, computed from Monte Carlo simulations, are in parentheses.
<table>
<thead>
<tr>
<th></th>
<th>No-arbitrage imposed</th>
<th>No-arbitrage not imposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log likelihood</td>
<td>3925.44</td>
<td>3943.22</td>
</tr>
<tr>
<td>Mean short rate (%/quarter)</td>
<td>1.475 (0.198)</td>
<td>1.498 (0.293)</td>
</tr>
<tr>
<td>$K$</td>
<td>0.969 (0.029) -0.107 (0.028) 0.091 (0.033) 0.006 (0.004)</td>
<td>0.974 (0.032) -0.110 (0.029) 0.088 (0.037) 0.009 (0.005)</td>
</tr>
<tr>
<td></td>
<td>(0.215) 0.856 -1.294 (0.246) 0.785 (0.046)</td>
<td>(0.245) 0.869 -1.257 (0.257) 0.754 (0.066)</td>
</tr>
<tr>
<td>$\Sigma \times 10^3$</td>
<td>1.557 (0.119) 0 0</td>
<td>1.541 (0.120)</td>
</tr>
<tr>
<td></td>
<td>(0.120)</td>
<td>(0.104) (0.063)</td>
</tr>
<tr>
<td>$\mu^Q \times 10^4$</td>
<td>2.321 (0.429) -8.086 (1.642) -2.390 (0.309)</td>
<td>2.281 (0.469) -8.614 (2.809) -2.940 (1.042)</td>
</tr>
<tr>
<td>$K^Q$</td>
<td>0.997 (0.001) 0.074 (0.001) 0.020 (0.008) 0.017 (0.006)</td>
<td>0.997 (0.002) 0.076 (0.002) 0.009 (0.013) 0.010 (0.008)</td>
</tr>
<tr>
<td></td>
<td>(0.056) (0.002) -1.398 (0.104) 0.372 (0.057)</td>
<td>(0.056) (0.018) (0.173)</td>
</tr>
<tr>
<td>$c_0 \times 10^4$</td>
<td>- (0.268)</td>
<td>-0.615 (0.268)</td>
</tr>
<tr>
<td></td>
<td>0.709 (0.226)</td>
<td></td>
</tr>
<tr>
<td>$C_1$</td>
<td>- (0.007)</td>
<td>-0.004 (0.007)</td>
</tr>
<tr>
<td></td>
<td>0.020 (0.010)</td>
<td>0.020 (0.010)</td>
</tr>
<tr>
<td>$\sigma\eta \times 10^4$</td>
<td>1.394 (0.062)</td>
<td>1.305 (0.062)</td>
</tr>
</tbody>
</table>

29
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

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

The mapping from the state vector to yields is

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

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.

<table>
<thead>
<tr>
<th></th>
<th>$\mu' \times 10^3$</th>
<th>0.704</th>
<th>-0.554</th>
<th>-0.008</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(0.106)</td>
<td>(0.073)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>$K$</td>
<td>0.976</td>
<td>-0.101</td>
<td>0.033</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.090)</td>
<td>(0.381)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.089</td>
<td>0.855</td>
<td>-1.252</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.064)</td>
<td>(0.287)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.006</td>
<td>0.002</td>
<td>0.799</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.024)</td>
<td>(0.100)</td>
<td></td>
</tr>
<tr>
<td>$\Sigma \times 10^3$</td>
<td>1.531</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.121)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.741</td>
<td>0.780</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td>(0.065)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.367</td>
<td>0.050</td>
<td>0.173</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.034)</td>
<td>(0.030)</td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.261</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\eta} \times 10^4$</td>
<td>1.604</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Effects of numerical optimization on simulated properties of term structure models

This table summarizes the results of 500 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 multiple steps. Starting at truth, five rounds of Simplex optimization are followed by a derivative-based optimizer. The resulting likelihood is compared to that produced with Method A. The point with the higher likelihood is then used as a starting value for another five rounds of Simplex optimization, followed by a derivative-based optimizer. 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 tenth and fifth percentile critical values for an LR test of the Duffie-Kan restrictions.

<table>
<thead>
<tr>
<th>Optimization Method</th>
<th>Log-likelihood restricted</th>
<th>Log-likelihood unrestricted</th>
<th>Population mean of 5-year bond yield (true mean is 7.09%) restricted</th>
<th>Population mean of 5-year bond yield (true mean is 7.09%) unrestricted</th>
<th>Critical values of likelihood ratio test 10th prctl</th>
<th>Critical values of likelihood ratio test 5th prctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3932.7</td>
<td>3938.5</td>
<td>7.062</td>
<td>7.119</td>
<td>19.21</td>
<td>21.43</td>
</tr>
<tr>
<td></td>
<td>(15.7)</td>
<td>(15.9)</td>
<td>(0.776)</td>
<td>(1.092)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>3934.5</td>
<td>3939.9</td>
<td>7.186</td>
<td>7.139</td>
<td>17.70</td>
<td>19.80</td>
</tr>
<tr>
<td></td>
<td>(15.7)</td>
<td>(15.9)</td>
<td>(0.959)</td>
<td>(1.261)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Monte Carlo simulations of differences in out-of-sample forecasts

Each of 500 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.

<table>
<thead>
<tr>
<th>Horizon (quarters)</th>
<th>Level forecasts</th>
<th>Slope forecasts</th>
<th>Curvature forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 2</td>
<td>Model 3</td>
<td>Model 2</td>
</tr>
<tr>
<td>1</td>
<td>0.055</td>
<td>0.118</td>
<td>0.048</td>
</tr>
<tr>
<td>2</td>
<td>0.101</td>
<td>0.209</td>
<td>0.071</td>
</tr>
<tr>
<td>3</td>
<td>0.142</td>
<td>0.295</td>
<td>0.083</td>
</tr>
<tr>
<td>4</td>
<td>0.181</td>
<td>0.379</td>
<td>0.091</td>
</tr>
<tr>
<td>5</td>
<td>0.219</td>
<td>0.462</td>
<td>0.100</td>
</tr>
<tr>
<td>6</td>
<td>0.255</td>
<td>0.544</td>
<td>0.110</td>
</tr>
<tr>
<td>7</td>
<td>0.290</td>
<td>0.623</td>
<td>0.122</td>
</tr>
<tr>
<td>8</td>
<td>0.323</td>
<td>0.699</td>
<td>0.134</td>
</tr>
<tr>
<td>9</td>
<td>0.356</td>
<td>0.770</td>
<td>0.145</td>
</tr>
<tr>
<td>10</td>
<td>0.387</td>
<td>0.837</td>
<td>0.154</td>
</tr>
<tr>
<td>11</td>
<td>0.417</td>
<td>0.899</td>
<td>0.162</td>
</tr>
<tr>
<td>12</td>
<td>0.446</td>
<td>0.956</td>
<td>0.169</td>
</tr>
</tbody>
</table>
Table 6. Monte Carlo simulations of RMSEs for out-of-sample forecasts

Each of 500 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.

<table>
<thead>
<tr>
<th>Horizon (quarters)</th>
<th>Level forecasts</th>
<th>Slope forecasts</th>
<th>Curvature forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
</tr>
<tr>
<td>1</td>
<td>0.639</td>
<td>0.640</td>
<td>0.646</td>
</tr>
<tr>
<td>2</td>
<td>0.848</td>
<td>0.849</td>
<td>0.867</td>
</tr>
<tr>
<td>3</td>
<td>1.054</td>
<td>1.055</td>
<td>1.081</td>
</tr>
<tr>
<td>4</td>
<td>1.156</td>
<td>1.154</td>
<td>1.197</td>
</tr>
<tr>
<td>5</td>
<td>1.289</td>
<td>1.283</td>
<td>1.340</td>
</tr>
<tr>
<td>6</td>
<td>1.391</td>
<td>1.383</td>
<td>1.449</td>
</tr>
<tr>
<td>7</td>
<td>1.512</td>
<td>1.509</td>
<td>1.596</td>
</tr>
<tr>
<td>8</td>
<td>1.628</td>
<td>1.631</td>
<td>1.732</td>
</tr>
<tr>
<td>9</td>
<td>1.716</td>
<td>1.722</td>
<td>1.848</td>
</tr>
<tr>
<td>10</td>
<td>1.827</td>
<td>1.828</td>
<td>1.984</td>
</tr>
<tr>
<td>11</td>
<td>1.926</td>
<td>1.927</td>
<td>2.099</td>
</tr>
<tr>
<td>12</td>
<td>2.044</td>
<td>2.040</td>
<td>2.218</td>
</tr>
</tbody>
</table>
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. Out of sample forecasts of Treasury yields as of year-end 2006. Three different three-factor Gaussian term structure models are estimated using quarterly Treasury bond yields from 1985Q1 through 2006Q4. Each estimated model is used to forecast the three-month yield (blue), the two-year yield (red), and the five-year yield (black), for the next twenty quarters. The solid lines are forecasts from a no-arbitrage model, the dashed lines are from an unrestricted model, and the dotted-dashed lines are from the model of Diebold, Rudebusch, and Aruoba.
Fig. 4. 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.