THE JOURNAL OF FINANCE • VOL. LXIII, NO. 2 • APRIL 2008

Identification of Maximal Affine Term Structure Models

PIERRE COLLIN-DUFRESNE, ROBERT S. GOLDSTEIN, and CHRISTOPHER S. JONES*

ABSTRACT

Building on Duffie and Kan (1996), we propose a new representation of affine models in which the state vector comprises infinitesimal maturity yields and their quadratic covariations. Because these variables possess unambiguous economic interpretations, they generate a representation that is *globally identifiable*. Further, this representation has more identifiable parameters than the "maximal" model of Dai and Singleton (2000). We implement this new representation for select three-factor models and find that model-independent estimates for the state vector can be estimated directly from yield curve data, which present advantages for the estimation and interpretation of multifactor models.

THE AFFINE CLASS OF TERM STRUCTURE MODELS as characterized by Duffie and Kan (DK, 1996) owes much of its popularity to its analytic tractability.[1] In particular, the affine class possesses closed-form solutions for bond and bond option pricing (Duffie, Pan, and Singleton (2000)), efficient approximation methods for swaption pricing (Collin-Dufresne and Goldstein (2002b), Singleton and Umantsev (2002)), and closed-form moment conditions for empirical analysis (Singleton (2001), Pan (2002)). As such, it has generated much attention both theoretically and empirically.[2]

Typically, affine term structure models are written in terms of a Markov system of *latent* state variables $X = \{X_1, \ldots, X_n\}$ that describe the entire state of

*Collin-Dufresne is from the Haas School of Business, University of California at Berkeley, and NBER. Goldstein is from the Carlson School of Management, University of Minnesota, and NBER. Jones is from the Marshall School of Business, University of Southern California. We thank seminar participants at UCLA, Cornell University, McGill, the University of Minnesota, the University of Arizona, UNC, Syracuse University, the University of Pennsylvania, the USC Applied Math seminar, the University of Texas at Austin, the Federal Reserve Bank of San Francisco, the CIREQ-CIRANO-MITACS conference on Univariate and Multivariate Models for Asset Pricing, the Econometric Society Meetings in Washington DC, and the Math-finance workshop in Frankfurt for their comments and suggestions. We would like to thank Luca Benzoni, Michael Brandt, Mike Chernov, Qiang Dai, Jefferson Duarte, Greg Duffee, Garland Durham, Bing Han, Philipp Illeditsch, Mike Johannes, and Ken Singleton for many helpful comments. We are especially grateful to the editor, Suresh Sundaresan, and an anonymous referee for their extensive guidance.

[1] The affine class essentially includes all multifactor extensions of the models of Vasicek (1977) and Cox, Ingersoll, and Ross (1985).

[2] See the recent survey by Dai and Singleton (2003) and the references therein.

the term structure (see, for example, Piazzesi (2006) for a survey). One problem with these latent factor models is that the parameter vector used to define the dynamics of the state vector might not be identifiable even if a panel data set of all possible fixed income securities, observed continuously, was available to the researcher. Accordingly, much effort has gone into identifying the most flexible model (i.e., the model with the greatest number of free parameters) that is identifiable. To date, the literature offers two approaches to deal with identification.

One approach, introduced by Dai and Singleton (DS (2000)), consists of performing a set of "invariant transformations" that leave security prices unchanged but that reduce the number of free parameters to a set that is identifiable.[3] Unfortunately, since these representations are expressed in terms of a latent state vector, they possess the undesirable feature that neither the state variables nor the model parameters have any particular economic meaning. Hence, a rotation to a more meaningful state vector is eventually necessary in order to interpret the results of the model (beyond just goodness-of-fit). Moreover, these representations suffer from the problem that latent state variables often lead to models that are *locally* but not *globally* identifiable. That is, there exist multiple combinations of state vectors and parameter vectors that are observationally equivalent.[4] This means that two researchers with the same data could obtain different estimates for the state and parameter vectors even though both had successfully maximized the same likelihood function.[5] In addition, as DS point out,[6] these representations provide only sufficient conditions for identification. Thus, there may be more general models, not nested by their representation, that are identifiable.

The second approach, introduced by DK, is to obtain an identifiable model by rotating from a set of *latent* state variables to a set of *observable* zero coupon yields (with distinct finite maturities). As we discuss below, while the use of observable state variables circumvents all of the problems associated with latent variables, this approach is often difficult to implement and therefore has not been widely used. Further, the DK framework cannot incorporate those models that exhibit unspanned stochastic volatility (USV, Collin-Dufresne and Goldstein (2002a)).

Below, we combine insights from both DS and DK to identify an invariant transformation of latent variable affine models in which the resulting representation is both tractable and specified in terms of economically meaningful state variables. Specifically, we rotate the state vector so that it consists of

[3] DS identify three such types of invariant transformations: (i) rotation of the state vector $\mathcal{T}_{\mathcal{A}}$, (ii) diffusion rescaling $\mathcal{T}_{\mathcal{D}}$, and (iii) Brownian motion rotation $\mathcal{T}_{\mathcal{O}}$.

[4] A model is locally identifiable if the likelihood function possesses only a countable number of maxima, whereas a model is globally identifiable if the likelihood function has a unique global maximum.

[5] In pre-publication drafts it is apparent that DS realized the need for such inequality constraints. However, they did not identify these constraints for the general $A_m(N)$ model. Joslin (2007) takes this approach.

[6] See footnote 6 of DS.

two types of variables: (i) the first few terms in the Taylor series expansion of the yield curve around a maturity of zero (terms that have intuitive economic interpretations such as level, slope, and curvature) and (ii) their quadratic covariations. The resulting representation has several advantages over latent variable representations.

- First, because the state vector has a unique economic interpretation, both the state vector and the parameters are globally identifiable.
- Second, our representation naturally leads to specifications that are more flexible than the canonical model identified by DS. That is, we show that in some cases the maximal $A_m(N)$ model has more identifiable parameters than that reported by DS.
- Third, while latent variables can only be extracted from observed prices *conditional* on both a particular model *and* a particular choice of parameter vector, our state vector is "observable" in that model-independent estimates for it are readily obtainable. As we discuss below, this presents several advantages for estimation of large-scale models.
- Fourth, in our representation the state vector and the parameter vector values can be meaningfully compared across different countries, different sample periods, or even different models because the state variables have unique economic interpretations that are model- and parameter-independent. In contrast, the parameters and state variables obtained from a latent factor representation cannot be compared until a rotation to an economically meaningful representation is performed.[7]

Our representation also has several advantages over the approach of DK. First, it is easy to implement. As we discuss below, DK's yield factor representation requires imposing constraints on systems of nonlinear equations that are often not solvable in closed form. Second, our representation works for USV models, for which there does not exist a one-to-one mapping between state variables and yields. Without such a mapping, the DK approach is not implementable. We acknowledge that DK's state vector, which consists of finite maturity yields, also possesses a clear economic interpretation. Furthermore, observing their state vector only requires a relatively straightforward interpolation from whatever yields are available in the data. In contrast, observing our state variables (without first specifying and estimating a model) requires extrapolation of the yield curve down to very short maturities, which may be less accurate. However, we demonstrate using both simulated and actual data that it is possible to obtain accurate model-independent estimates of our state variables even in the presence of substantial measurement error.

[7] It is often the case that state variables are highly correlated with one or more principal components, and thus researchers interpret the state variables as such. However, such interpretations are approximate at best. Furthermore, as shown by Duffee (1996) and Tang and Xia (2005), the weights of such principal components change over time and across countries. Hence, attempting to compare models and/or parameters through their implied principal component dynamics is at best suggestive and likely somewhat misleading.

Our empirical results show that our observable representation also has some practical advantages. First, since we can estimate a time series for the state vector before attempting to identify parameter estimates, we can use simple econometric methods (e.g., OLS) to come up with a first guess for the parameter vector, simplifying the search over what is often a large dimensional parameter space. Second, we find that when the model-independent estimates for the state vector differ significantly from those obtained by a full-fledged econometric analysis, the model may be badly misspecified. For example, we find remarkable similarity between model-free and model-implied state variables for a Gaussian three-factor model. In contrast, the relation between the model-free and model-implied state variables depends on the way in which a model with stochastic volatility is estimated. Specifically, estimating the model using yield data only results in a close match to the model-independent state variables. In contrast, forcing the model to fit a proxy for the short-rate volatility process causes model-implied and model-independent state variables to differ sharply. These results point to model misspecification, which we interpret as suggesting that three-factor affine models cannot simultaneously fit the time series properties of the quadratic variation of the short rate and the dynamics of the third (i.e., curvature) factor. A companion paper, Collin-Dufresne, Goldstein, and Jones (2007), provides further evidence on this issue.

The rest of the paper is organized as follows. In Section I we begin by defining a few important terms and showing that latent state variables lead to models that are only locally identifiable. In Section II we propose a canonical representation for the $A_m(N)$ class in terms of m latent square root processes and $(N - m)$ Gaussian processes, identifying a larger parameter vector than that identified by the canonical representation of DS. We then show that the Gaussian variables possess simple, unambiguous economic interpretations such as the level, slope, and curvature of the yield curve, and that the covariances among these Gaussian variables are observable. As such, we show that we can rotate from the original m latent square root processes to processes that are economically meaningful. In Section III we provide some examples. We describe the data in Section IV, while in Section V we discuss the construction and properties of model-free estimates of the state vector. Section VI presents the estimation methods, and in Section VII we report the empirical results for several specifications written in terms of observable state variables. We conclude in Section VIII.

I. Background

Throughout this paper, we use terms that have different meanings in the applied and theoretical econometrics literatures. For clarity, we define our use of these terms here.

Identified, identifiable, and maximal models

A given model is said to be *identified* if the state vector and parameter vector can be inferred from a *particular* data set. In contrast, we say that a model is *identifiable* if the state vector and parameter vector can be inferred from observing *all fixed income security prices* (i.e., all conceivable securities) as frequently as necessary; that is, the term *identifiable* is defined as a theoretical construct. For concreteness, we assume that all fixed income securities are claims on cash flows occurring at finite maturities that depend only on the spot rate process $r(\cdot)$. Hence, their prices are solutions to

$$P(t) = \mathrm{E}_t^Q \left[e^{-\beta \int_t^{T_1} r(s)\,ds} \, CF \left\{ r(u), u \right\} \Big|_{u \in (t, T_2)} \right]. \tag{1}$$

The case $\beta = 1$ corresponds to standard risk-neutral discounting of the cash flows CF, which may depend on the entire path of the spot rate $r(u)|_{u \in (t, T_2)}$. The case $\beta = 0$ applies to futures prices.[8]

A special case of an identifiable model is a *maximal* model, which, as defined by DS, is the most general admissible model that is identifiable given sufficiently informative data. That is, a maximal model is an identifiable model that has the largest number of free parameters (within a particular class). Note that maximality is also a theoretical concept. Indeed, DS determine maximality by considering a series of invariant rotations of the fundamental PDE (satisfied by path-independent European contingent claims) that leave all security prices unchanged. As such, it is defined without ever making reference to what particular securities are actually available to the econometrician. Moreover, the concept of maximality is independent of whether the data are assumed to be measured with or without error. Below, we follow DS and interpret both maximality and identifiability as theoretical constructs, recognizing the possibility that a particular data set might be insufficient for all parameters of the model to be inferred.

In their definition of maximality, DS focus on identifying parameters used to specify state vector dynamics under both the historical (P-) measure and risk-neutral (Q-) measure. However, below we provide examples in which a model is not identifiable even with the most restrictive risk premium structure, namely, when the risk premia are set to zero, so that the P- and Q-dynamics are equivalent. Clearly, assuming more general risk premia structures in these cases cannot solve this problem. We therefore need to first understand what models are identifiable under the assumption that risk premia are equal to zero. In the spirit of DS, we refer to a model as being *Q-maximal* if it is the most general model (within a particular class) that is identifiable given all conceivable security data expressed in equation (1) when all risk premia are assumed to be zero. Given a Q-maximal model, it turns out to be a trivial matter to determine whether or not the risk premia are identifiable. This is because Q-maximality implies that the state variables can be observed or estimated on each date.[9] As such, parameters capturing the risk premia can be identified from the time

[8] The fact that the cash flow can depend upon the entire path of interest rates implies that Asian-type options are also permitted. Note that the prices of such securities are solutions to a PDE that is more general than that investigated by DS.

[9] Here, we do not consider the possibility that there are state variables that drive P-measure dynamics but not Q-measure dynamics. We thank Greg Duffee for pointing out this possibility.

series of these variables. Furthermore, because the concept of Q-maximality applies only to the risk-neutral dynamics, our approach remains valid even when risk premia do not preserve affine dynamics under the historical measure (e.g., Duarte 2004).

As we mention above, identification can be either local or global. The latter implies the existence of a single parameter vector that provides the best fit of the data; the former implies that there are a finite number of such parameter vectors that are observationally equivalent. While DS implicitly focus on local identification, here we seek representations that are globally identifiable, motivated by the fact that only these specifications lead to parameters and state variables to which meaningful interpretation can be attributed. We discuss this subject in detail below.

Observed, observable, and latent variables

A state variable is said to be *observed* from a particular data set if its value can be readily determined without reference to any particular model. In contrast, we define a state variable to be *observable* if, given the availability of *all fixed income securities prices* (as defined in equation (1) above) observed as frequently as desired, its value can be measured without reference to any particular model. Note that the term *observable* is also a theoretical construct.

Two important examples of variables that are observable according to this definition are the spot rate and its volatility. The former is the very short end of the continuously compounded term structure and the latter is its quadratic variation, which as Merton (1980) points out can be estimated perfectly with any finite span of data in continuous time. Note that observable variables are *economically meaningful* variables in that they have unambiguous definitions independent of any model, and in turn are independent of any model's parameter values. Indeed, throughout the paper, we use these two terms interchangeably.[10]

On the other hand, we refer to variables that are not observable as *latent*. The important distinction is that latent variables can only be measured conditional on choosing a model and estimating its parameters. Therefore, the values that latent variables take are inherently tied to a particular theory and a specific set of parameter values. In contrast, an observable state variable has an unambiguous economic interpretation that is independent of the particular model being considered. This implies that observable state variables can be estimated without knowledge of the correct model or its parameter vector. Moreover, this implies that observable variables can be compared across models, countries, data sets, etc., whereas latent variables cannot.

[10] However, not all economically meaningful state variables are observable. For example, if there were a state variable that drove expected changes in the spot rate but did not show up in risk-neutral dynamics, it would be economically meaningful but not observable.

A. Properties of Observable State Variables

We note that observable state variables have the following properties:

(P1) If $X(t)$ is an observable state variable that follows an Itô process, then its risk-neutral drift $\mu^Q(t) \equiv \lim_{\Delta\to 0} \frac{1}{\Delta}\mathrm{E}_t^Q[X(t+\Delta) - X(t)]$ is also observable.[11]

(P2) If $X(t)$ is an observable state variable, then its quadratic variation $V(t) = \langle X, X\rangle(t)$ is observable.

(P3) If $X(t)$ and $Y(t)$ are two observable state variables, then their quadratic covariation process $V_{XY}(t) = \langle X, Y\rangle(t)$ is observable.

Properties P2 and P3 follow directly from the definition of quadratic variation (e.g., Shreve 2004) and the assumption that we observe data continuously (recall that "observable" as defined above is a theoretical concept). We note that the observability of the instantaneous variance of a price series is not an original argument, being explicit in the theoretical work of Black (1976) and Merton (1980).

Property P1 is perhaps the most surprising given the well-known difficulty of measuring drifts empirically. We emphasize, however, that it is not the actual drift but rather the *risk-neutral* drift that is observable. This follows from the fact that since $X(t)$ is observable, one can write a futures contract on its value at some future date $(t + \Delta)$. By absence of arbitrage (Duffie 2006, ch. 8-D), the corresponding futures price is

$$F(t, \Delta) = \mathrm{E}_t^Q[X(t+\Delta)].$$

Therefore, given the entire term structure (as a function of the maturity) of such futures prices, we can measure the instantaneous slope

$$\lim_{\Delta\to 0} \frac{1}{\Delta}(F(t,\Delta) - X(t)) = \lim_{\Delta\to 0} \frac{1}{\Delta}\mathrm{E}_t^Q[X(t+\Delta) - X(t)] \equiv \mu^Q(t).$$

It follows that $\mu^Q(t)$ denotes the drift of the Itô process followed by $X(t)$ under the risk-neutral measure, that is

$$dX(t) = \mu^Q(t)\,dt + dM(t),$$

where $M(t)$ is some continuous Q-martingale.

We claim these properties imply that if one specifies affine term structure dynamics using observable state variables, then all parameters that show up in the risk-neutral dynamics are identifiable. Indeed, assume that some N-dimensional state vector $\{X_i\}$ is observable. Since we are assuming that the model is affine under the Q-measure, the risk-neutral drift of each variable

[11] We assume that a risk-neutral measure exists; see Harrison and Pliska (1981).

must be of the form

$$\mu_i^Q(t) \equiv \delta_{i,0} + \sum_{j=1}^{N} \delta_{i,j} X_j(t). \tag{2}$$

Note that P1 implies that $\mu_i^Q(t)$ is observable. By observing its value, along with the values of the state vector $X(t)$, equation (2) provides us with one equation for the $(N+1)$ unknowns $\{\delta_i\}$. Thus, by observing data on $(N+1)$ different dates, we obtain $(N+1)$ equations that are linear in the $(N+1)$ unknowns $\{\delta_i\}$, implying that we will be able to identify their values. Analogously, since covariances are also observable and affine in the state vector, a similar argument can be made to prove the identification of the parameters that show up in the covariance matrix.[12]

Below, we show that in addition to being observable, in many cases these risk-neutral drifts have clear economic interpretations. For example, the risk-neutral drift of the spot rate is intimately related to the slope of the yield curve at short maturities. Furthermore, in practice it is not necessary to have prices of exotic securities to identify the model. Indeed, for those models that do not exhibit USV, bond prices alone are sufficient for identifying all risk-neutral parameters since they are all easily extracted from appropriate regressions. We discuss this point further below.

B. Latent Variables and Model Identification

It is well known from many branches of econometrics and statistics that latent variable models often suffer from problems of identification. Affine term structure models with latent factors are no exception, and it is straightforward to write down a model in which some model parameters are not identifiable regardless of how many securities are available and how often they are observed. To address this issue, DS propose a set of invariant rotations in an attempt to eliminate the unidentified parameters. The resulting model is identified if all possible rotations have been performed. However, it is not clear that this approach delivers the most general identifiable model.[13]

Further, neither the model parameters nor the latent state variables of their representations have any particular economic meaning. Indeed, there are several examples in the literature where researchers have attempted to attribute an economic interpretation to latent variables when, in fact, they have none. A very elegant example illustrating this concern comes from Babbs and Nowman

[12] Of course, models that are written with obviously redundant parameters cannot be identified. For example, one cannot separately identify δ_0 and δ'_0 in $\mu^Q(t) \equiv (\delta_0 + \delta_0') + \sum_{i=1}^{N} \delta_i X_i(t)$. Fortunately, specifications like this are easily avoided and are ruled out by our canonical form.

[13] It might be difficult to prove that all possible rotations have been performed. Further, as DS point out in their footnote 6, they cannot rule out that their representation might be nested in a more general model. We confirm this below.

(BN, 1999). Consider the two-factor Gaussian (maximal $A_0(2)$) model:

$$dr(t) = \kappa_r(\theta(t) - r(t))\,dt + \sigma_r\,dz_r(t) \tag{3}$$

$$d\theta(t) = \kappa_\theta(\overline{\theta} - \theta(t))\,dt + \sigma_\theta\,dz_\theta(t), \tag{4}$$

with $dz_r\,dz_\theta = \rho\,dt$. BN show that one can find an invariant transformation of the model by defining another latent variable $\theta^*(t)$ by

$$\theta^*(t) = \left(1 - \frac{\kappa_r}{\kappa_\theta}\right) r(t) + \frac{\kappa_r}{\kappa_\theta}\,\theta(t) \tag{5}$$

so that the dynamics of the system become

$$dr(t) = \kappa_\theta(\theta^*(t) - r(t))\,dt + \sigma_r\,dz_r(t) \tag{6}$$

$$d\theta^*(t) = \kappa_r(\overline{\theta} - \theta^*(t))\,dt + \sigma_{\theta^*}\,dz_{\theta^*}(t). \tag{7}$$

Note that the system of equations (3) and (4) is identical to the system of equations (6) and (7). Hence, even though the model is maximal in the sense of DS, two researchers could obtain different estimates for the state and parameter vectors even though both had successfully maximized the same likelihood function. In particular, the prices of all fixed income securities are identical whether one uses the values $[\{\theta(t)\}, \kappa_r, \kappa_\theta]$ or $[\{\theta^*(t)\}, \kappa_\theta, \kappa_r]$. This duplicity is especially problematic when one wants to give economic meaning to θ. For example, this variable has been previously interpreted as a long-run target rate set by the central bank (e.g., Jegadeesh and Pennacchi (1996), Balduzzi, Das, and Foresi (1996)). The implication is that in their model there are two sets of solutions leading to two different time series for the state vector θ, both of which generate identical prices for all securities. Hence, the time series of θ by itself has no economic meaning!

In the parlance of system identification (e.g., Ljung (1999, ch. 4)), maximal latent variable models are only locally and not globally identifiable. We emphasize that the insights of BN are relevant not just for Gaussian models. For example, the same transformation can be applied to the maximal $A_1(3)$ model of DS (2000) in its Ar representation (equation (23), p. 1951) to show that the "central tendency" defined by DS is not uniquely determined, and the same issue arises for the canonical AY representation of DS (p. 1948).[14]

The example above is particularly salient because it emphasizes the difference between latent and observable state variables. In particular, the state

[14] The AY canonical $A_0(2)$ model of DS is given by:

$$r(t) = \overline{r} + \sigma_1 X_1(t) + \sigma_2 X_2(t) dX_1(t) = -\kappa_{11} X_1(t)\,dt + dz_1(t) dX_2(t)$$
$$= -(\kappa_{21} X_1(t) + \kappa_{22} X_2)dt + dz_2(t).$$

It is straightforward to show that the BN model given in equations (1) and (2) is an invariant transformation of the canonical AY model above, where, in particular, we have the relation $\kappa_{11} = \kappa_r$ and $\kappa_{22} = \kappa_\theta$. Yet, following the argument leading to the equivalent representation in (6) and (7), there is an equivalent AY representation with $\kappa_{22} = \kappa_r$ and $\kappa_{11} = \kappa_\theta$. This shows that the AY canonical representation is not globally identifiable.

variable r is by definition the short end of the term structure, and is therefore observable (or, equivalently, economically meaningful) in that its value cannot be changed without necessarily changing the prices of some fixed income securities (in particular, those with very short, but finite, maturities). In contrast, because θ is latent, its value can be replaced by θ^* and, provided the parameters are adjusted appropriately (i.e., $\kappa_\theta \longleftrightarrow \kappa_r$), the prices and price dynamics of *all* fixed income securities remain unchanged.[15]

C. Advantages of Observable State Variables

Here we illustrate the advantages of rotating from latent to observable state variables in a simple two-factor Gaussian case. We use the original approach of DK (1996), who propose rotating a latent state vector to an observable state vector defined in terms of yields of finite maturities.

Consider the following risk-neutral dynamics of a two-factor Gaussian model written in terms of the short rate r and a latent variable s

$$dr(t) = (\alpha_r + \beta_{rr}\, r(t) + \beta_{rs}\, s(t))\, dt + \sigma_r\, dZ_r^Q(t) \tag{8}$$

$$ds(t) = (\alpha_s + \beta_{sr}\, r(t) + \beta_{ss}\, s(t))\, dt + \sigma_s\, dZ_s^Q(t), \tag{9}$$

where $dZ_r^Q(t)\, dZ_s^Q(t) = \rho_{rs} dt$. This model has a total of nine risk-neutral parameters.

DK show that yields of all maturities τ are affine in r and s

$$Y(t,\tau) = -\frac{A(\tau)}{\tau} + \frac{B_r(\tau)}{\tau} r(t) + \frac{B_s(\tau)}{\tau} s(t).$$

As such, we can rotate from the latent state vector $(r(t), s(t))$ to the observable state vector $(r(t), Y(t,\hat{\tau}))$ for some specific choice of $\hat{\tau} > 0$. As DK demonstrate, the dynamics of this state vector are jointly Markov and affine,

$$dr(t) = (\hat{\alpha}_r + \hat{\beta}_{rr}\, r(t) + \hat{\beta}_{ry}\, Y(t,\hat{\tau}))\, dt + \sigma_r\, dZ_{r,t}^Q \tag{10}$$

$$dY(t,\hat{\tau}) = (\hat{\alpha}_y + \hat{\beta}_{yr}\, r(t) + \hat{\beta}_{yy}\, Y(t,\hat{\tau}))\, dt + \sigma_y\, dZ_{y,t}^Q, \tag{11}$$

and the yields are still affine in the state variables,

$$\forall\tau \;\; Y(t,\tau) = -\frac{\hat{A}(\tau)}{\tau} + \frac{\hat{B}_r(\tau)}{\tau} r(t) + \frac{\hat{B}_y(\tau)}{\tau} Y(t,\hat{\tau}). \tag{12}$$

[15] One could solve the identification problem for this model by imposing an additional constraint on the parameters, for example $\kappa_\theta > \kappa_r$. However, similar restrictions have not been identified by DS for the general $A_m(N)$ model. Further, this approach still leaves unaddressed the problem that neither the state variables nor the parameters have any intrinsic economic meaning. Finally, imposing such arbitrary restrictions only makes it more likely that investigators impute economic meaning to such variables (e.g., "central tendency"), when in fact they have none.

In particular, this equation must hold for the special case $\tau = \hat{\tau}$, which introduces three additional constraints, namely,

$$\hat{A}(\hat{\tau}) = 0, \ \hat{B}_r(\hat{\tau}) = 0, \ \text{and} \ \hat{B}_y(\hat{\tau}) = \hat{\tau}. \tag{13}$$

Although these constraints are nonlinear, one would (correctly) suspect that they will lead to three restrictions on the parameters in equations (10) and (11). Hence, while the latent state vector representation (equations (8) and (9)) seems to suggest that there are nine free risk-neutral parameters, in fact, there are only six—a fact that becomes obvious when we rotate to an observable state vector.

In summary, this discussion illustrates several key points.

- When writing down a model with latent variables as in equations (8) and (9), there is a risk of including more risk-neutral parameters than it is possible to identify. This arises independently of how the risk premia structure is specified, and even if we assume all conceivable fixed income data are available to the researcher.
- Rotating to observables as in equations (10) and (11) is a straightforward way to eliminate extra parameters if the latent factor model is not identifiable. Doing so also solves the local versus global identification issue, since observable state variables have a model-independent economic interpretation.
- For general affine models, in practice it may be difficult to rotate from a latent state vector to yields of *finite* maturities, since the constraints (equation (13)) are often written in terms of functions that do not have analytic solutions.

In the next section, we propose a representation that is similar in spirit to the original idea of DK (1996) of rotating to observable state variables but that avoids some of the shortcomings of that approach. First, for the subset of models exhibiting unspanned stochastic volatility, the rotation proposed by DK fails since not all state variables can be written as a linear combination of yields. Second, even for non-USV models, for which the rotation is in principle possible, our approach avoids the difficulties inherent in rotating to a vector of yields of finite maturities. Finally, we identify a Q-maximal model that is more flexible than that identified by DS.

II. Q-Maximal Affine Models with Observable State Variables

In the previous section we discussed some problems associated with latent variables. In this section we propose a canonical representation of affine models that is Q-maximal and that nests the canonical representation of DS (2000).[16] We show that our canonical representation leads to a fully observable representation in terms of the state variables $\{\mu_j\}$ and their quadratic covariates $\{V_{jj'}\}$

[16] A further technical advantage of our representation is that its admissibility (i.e., mathematical soundness) is easily verified.

that can be estimated independent of a model given a sufficiently rich panel of term structure data.

We proceed in several steps. First, we propose a canonical representation written in terms of Gaussian variables[17] $\{\mu_j\}$ and latent square root variables $\{x_i\}$. Second, we show that the $\{\mu_j\}$ variables are observable. From P3, it follows that the instantaneous covariances among the $\{\mu_j\}$ variables, which we refer to as $\{V_{jj'}\}$ variables, are also observable. We can therefore rotate from the $\{x_i\}$ variables to the $\{V_{jj'}\}$ variables to obtain a framework written completely in terms of observable variables. Finally, we show that the size of the parameter vector does not change due to this rotation, and that the parameter vector is in fact identifiable.

A. A Canonical Representation for Affine Term Structure Models

Following the nomenclature of DS, an $A_m(N)$ affine model has m square root state variables that show up in the covariance matrix and $(N - m)$ Gaussian variables that do not. Here, we propose a canonical representation that, as we show below, has the maximal number of risk-neutral parameters for a given $A_m(N)$ class of models.

Following DS, we first specify the dynamics of m latent square root processes $\{x_i\}|_{i\in(1,m)}$ as jointly Markov

$$dx_i(t) = \left(\kappa_{i0} + \sum_{i'=1}^{m} \kappa_{ii'}\, x_{i'}(t)\right) dt + \sqrt{x_i}\, dz_i^Q(t), \qquad dz_i^Q(t)\, dz_{i'}^Q(t) = \mathbf{1}_{\{i=i'\}}\, dt. \tag{14}$$

In order to guarantee that the $\{x_i\}$ remain positive, that is, in order to guarantee admissibility, we restrict the $(m + 1)$ risk-neutral drift coefficients $\kappa_{ii'}$ to be nonnegative for all $i' \neq i$. Note that equation (14) specifies that there is a total of $m(m + 1)$ risk-neutral parameters in the specification of all m square root processes.[18]

With the square root processes specified, we now turn to the $(N - m)$ Gaussian state variables in an $A_m(N)$ model. Note that for the case $N = m$, there are no Gaussian state variables, implying that the spot rate is an affine function of the x processes,

$$r(t) = \delta_0 + \sum_{i=1}^{m} \delta_i x_i(t), \tag{15}$$

and that the model is fully specified.

[17] We use the term Gaussian to indicate that, conditional upon the values of the square-root variables, these variables have Gaussian dynamics. As such, they can take on all real values. In contrast, square root variables are associated with a lower bound.

[18] We note that our specification rules out the special case of the Wishart Quadratic-affine term structure models identified by Gourieroux and Sufana (2003).

In contrast, when $N > m$, there are $(N - m)$ Gaussian variables to be specified. Here, we show that these can be chosen to be the spot rate r, its risk-neutral drift $\mu_1 \equiv \frac{1}{dt}\mathrm{E}^Q[dr]$, its risk-neutral drift $\mu_2 \equiv \frac{1}{dt}\mathrm{E}^Q[\mu_1]$, and so on, up to $\mu_{N-m-1} \equiv \frac{1}{dt}\mathrm{E}^Q[\mu_{N-m-2}]$. The proof follows from induction. Indeed, for a given set of m square root processes $\{x_i\}$, either the spot rate is an affine function of these $\{x_i\}$ or it is not. If it is, then the model falls into the $A_m(m)$ category, contrary to the assumption that we have an $A_m(N)$ model with $N > m$. Thus, r must be linearly independent of the square root processes. Therefore, we can choose it as the first of the Gaussian state variables.

Analogously, we now show that μ_1 can also be chosen as a Gaussian state variable if $N > (m + 1)$. Recall that by assumption, only the x processes show up in the covariance matrix. Hence, the spot rate variance, and all of its covariances with the x variables, are affine functions of the x variables. Thus, the only available channel for increasing the state space of the risk-neutral dynamics of r (and hence, of the entire system, since the $\{x\}$ are jointly Markov) is through its drift μ_1. Now, either μ_1 is an affine function of $\{r, \{x\}\}$ or it is not. If it is, then the model falls into the $A_m(m + 1)$ category, contrary to the assumption that we have an $A_m(N)$ model and that $N > (m + 1)$. Thus, μ_1 must be linearly independent of $\{r, \{x\}\}$.

This argument is repeated until we have $(N - m)$ state variables, each of which is the risk-neutral drift of the previously introduced state variable. Thus, we have specified the drifts of all the state variables except for the drift of r_{N-m-1}, which we specify here as generally as possible,

$$\frac{1}{dt}\mathrm{E}^Q\left[d\mu_{N-m-1}(t)\right] = \gamma + \sum_{j=0}^{N-m-1} \kappa_j\, \mu_j(t) + \sum_{i=1}^{m} \kappa_{N-m+i}\, x_i(t). \tag{16}$$

For tractability purposes, we define $r(t) \equiv \mu_0(t)$ in this equation and in many equations below so that $\{\mu\}$ denotes the entire set of Gaussian variables. Note that equation (16) specifies $(N + 1)$ risk-neutral drift parameters.

Equations (14) to (16) (along with the definitions of $\{\mu_j\}$) identify the drifts of all state variables as well as the covariance matrix among the x variables. This leaves only the covariance matrix among the μ variables and the covariance matrix between μ and x variables for the model to be completely specified. Following DS, we specify the covariance between variables μ_j and x_i as

$$\frac{1}{dt}\, d\mu_j(t)\, dx_i(t) = \rho_{ij}\, x_i(t). \tag{17}$$

Further, we specify the covariance between μ_j and $\mu_{j'}$ as an affine function of the m square root processes:

$$\frac{1}{dt}\, d\mu_j(t)\, d\mu_{j'}(t) \equiv V_{jj'}(\mathbf{x}(\mathbf{t})) = \omega^0_{jj'} + \sum_{i=1}^{m} \omega^i_{jj'}\, x_i(t).$$

We emphasize that these choices are not arbitrary. Rather, they are the most general that are simultaneously identifiable and consistent with the admissibility of the process. Indeed, any further generalization would either introduce unidentifiable parameters into the model or imply a negative definite covariance matrix for some values of x_i.

Below, we show that the $\{\mu_j\}$ variables are observable because they can be inferred from the shape of the yield curve at short maturities (e.g., slope and curvature). Thus, from P3, the quadratic covariates $\{V_{jj'}(\mathbf{x})\}$ are also observable. As such, we will eventually find it convenient to rotate from the latent square root processes $\{x\}$ to some subset of the $\{V_{jj'}(\mathbf{x})\}$. For now, however, we specify our state vector in terms of the m variables of $\{x\}$ and the $(N-m)$ variables of $\{\mu\}$.

B. Maximal Parameter Vector

Note that equations (14) to (18) uniquely specify the risk-neutral dynamics of the N-dimensional Markov system. We refer to this system of equations as our canonical representation. For $m < N$, the number of risk-neutral parameters is

(1) $m(m+1)$ drift parameters for the square root processes (equation (14))
(2) $(N+1)$ drift parameters for the last Gaussian variable (equation (16))
(3) $m(N-m)$ covariance parameters ρ_{ij} between the square root and Gaussian processes (equation (17))
(4) $\frac{1}{2}(N-m)(N-m+1)(m+1)$ covariance parameters $\omega^i_{jj'}$ between the Gaussian processes (equation (18)).

When $m = N$, the number of risk-neutral parameters is

(1) $m(m+1)$ drift parameters for the square root processes (equation (14))
(2) $(m+1)$ parameters of δ_0 and $\{\delta_i\}$ (equation (15)).

In both cases, the total number of risk-neutral parameters ($\#_{CGJ}$) is

$$\#_{CGJ} = m(m+1) + (N+1) + m(N-m) + \frac{1}{2}(N-m)(N-m+1)(m+1). \tag{18}$$

This contrasts with DS, who find

$$\#_{DS} = \begin{cases} N^2 + N + m + 1 & m \neq 0 \\ \frac{1}{2}(N+1)(N+2) & m = 0. \end{cases}$$

Note that the two formulas are in agreement in several cases ($m = 0$, $m = 1$, $m = (N-1)$, $m = N$) but in general differ for the non-Gaussian cases when $N > 3$.[19] For example, we find that the $A_2(4)$ model has 24 risk-neutral parameters, while DS find there are only 23. For the $A_2(5)$ model, we find 36

[19] The fact that models agree when $N \leq 3$ is also related to a mathematical result derived in independent work by Cheridito, Filipovic, and Kimmel (2005), who note that the form of diffusion matrix chosen by DS for their canonical representation only spans the entire space when $N \leq 3$. However, these authors do not identify the maximal model for $N > 3$.

risk-neutral parameters, while DS find only 33. As N and m get larger, so does the discrepancy.

The source of this discrepancy can be traced back to the number of parameters that appear in the covariances between Gaussian state variables. In particular, DS assume that any N-factor affine process can be written as

$$dX(t) = (a^Q + b^Q X(t))\,dt + \Sigma\sqrt{S(t)}\,dZ^Q(t), \tag{19}$$

where $S(t)$ is a diagonal matrix with components $S_{ii}(t) = \alpha_i + \beta_i^\top X(t)$, and $dZ^Q(t)$ is N-dimensional. It turns out, however, that the most general identifiable model cannot always be written in this form, which we demonstrate below for the $A_2(4)$ model. Instead, we argue that in order to identify the maximal model, one must specify the model in one of three ways: (i) with more Brownian motions than state variables, (ii) using a more general form than $\Sigma\sqrt{S(t)}$ for the diffusion matrix, (iii) in terms of a covariance matrix, as we have done in equations (17) to (18), rather than a system of SDEs. We note that specifying the stochastic components of the model in terms of a covariance matrix rather than as Itô diffusions has the advantage of introducing parameters that have clear economic interpretations.[20] It is also these parameters that show up in the fundamental partial differential equation that security prices satisfy.

C. *Proof that the $\{\mu_j\}$ Variables Are Observable*

In this subsection we show that the $\{\mu_j\}$ variables can be measured directly from the short end of the yield curve. Hence, they are observable in the sense defined in Section I. In the empirical sections below, we demonstrate that model-independent estimates for these variables are readily obtainable.

As we note in Section I.C, the risk-neutral drift of any observable state variable is itself observable, as one can design a futures contract with an associated arbitrage-free futures price equal to the risk-neutral drift. Therefore, since μ_{j+1} is by definition the risk-neutral drift of μ_j, all we have to show is that r ($\equiv \mu_0$) is observable. But note that r is defined as the shortest maturity bond yield. Thus, it is directly observable from the short end of the yield curve, with an economic meaning that is independent of any parameter vector and independent of any model. By induction, the observability of r implies that for all $j > 0$, all μ_j are also (theoretically) observable. In practice, we show below that the first few $\{\mu_j\}$ can be estimated accurately using empirical data.

Admittedly, the futures contracts used in our argument in Section I.C do not exist in practice. However, we can show that the μ_j have simple economic interpretations as they are directly tied to the shape of the yield curve. To do so, it is convenient to express the yield curve in terms of its Taylor series expansion

[20] While all three strategies are mathematically equivalent, we view the third approach as more convenient in practice.

with respect to time-to-maturity τ

$$Y(\tau,t) = Y^0(t) + \tau Y^1(t) + \frac{1}{2}\tau^2 Y^2(t) + \ldots, \tag{20}$$

where $Y^n \equiv \frac{\partial^n Y(\tau)}{\partial \tau^n}|_{\tau=0}$. We emphasize that since the entire yield curve is observable, it follows that the Taylor series components $Y^0(t), Y^1(t)$, and $Y^2(t)$ are also observable and have the interpretation of the level, slope, and curvature of the yield curve at very short maturities ($\tau \approx 0$). In the Appendix we show that the $\{\mu_j\}$ can be recursively obtained from the derivatives (e.g., slope and curvature) of the yield curve $\{Y^j\}$ and their quadratic covariations, with the first few terms given by [21]

$$Y^0(t) = r(t) \tag{21}$$

$$Y^1(t) = \frac{1}{2}\mu_1(t) \tag{22}$$

$$Y^2(t) = \frac{1}{3}\left[\mu_2(t) - V_{00}(t)\right] \tag{23}$$

$$Y^3(t) = \frac{1}{4}\left[\mu_3(t) - \frac{1}{dt}\mathrm{E}_t^Q\left[dV_{00}(t)\right] - 3V_{01}(t)\right]. \tag{24}$$

Here, $V_{00}(t)$ is the spot rate variance and $V_{01}(t) = \frac{1}{dt}dr(t)\,d\mu_1(t)$. Thus, r is the level of the yield curve at short maturities, μ_1 is twice the slope of the yield curve at short maturities, and μ_2 is equal to three times the curvature at short maturities minus the short rate variance.

D. Proof that the Canonical Representation Is Maximal

In Section II.A we wrote the canonical representation in terms of latent variables x. Thus, as we noted in Section I.C, there is a concern that only a smaller parameter vector will survive when the risk-neutral dynamics are specified in terms of an observable state vector. Here, we show that this is not the case, and that the size of the parameter vector is as given in equation (18).

To demonstrate this, first note (from P3) that the covariance terms from equation (18),

$$V_{jj'}(\mathbf{x}(\mathrm{t})) \equiv \omega_{jj'}^0 + \sum_{i=1}^{m} \omega_{jj'}^i\, x_i(t), \tag{25}$$

are observable. As such, we can obtain a (continuous) time series of this variable, which for convenience we define as $V^{(0)}(\mathbf{x}(\mathrm{t}))$. With this time series, we can (from

[21] We give the general relation in the Appendix.

P2) observe its variance

$$V^{(1)}(\mathbf{x}(\mathrm{t})) \equiv \frac{1}{dt}\left(dV^{(0)}(\mathbf{x}(\mathrm{t}))\right)^2 = \sum_{i=1}^{m} \frac{1}{dt}(\omega^i_{jj'} dx_i(t))^2 = \sum_{i=1}^{m} \left(\omega^i_{jj'}\right)^2 x_i(t). \tag{26}$$

Here, we use the conditional independence of the x processes and the fact that $dx_i^2 = x_i\,dt$ from equation (14). Equation (26) provides one equation for the $m + 1$ unknown parameters $\{\omega^i_{jj'}\}$ and the m unknown state variables $\{x_i(t)\}$. Note, however, that since $V^{(1)}(\mathbf{x}(\mathrm{t}))$ is observable, we can obtain a (continuous) time series of it. Therefore, from P3 we can also estimate the covariance

$$V^{(2)}(\mathbf{x}(\mathrm{t})) \equiv \frac{1}{dt} dV^{(0)}(\mathbf{x}(\mathrm{t}))\, dV^{(1)}(\mathbf{x}(\mathrm{t})) = \sum_{i=1}^{m} \left(\omega^i_{jj'}\right)^3 x_i(t). \tag{27}$$

Note that no new unknowns appear in going from equation (26) to equation (27). As such, by continuing this argument recursively, we can obtain as many equations as we like with which to infer the $2m + 1$ unknowns $\{\omega^i_{jj'}, x_i\}$. The implication is that both the parameters $\{\omega^i_{jj'}\}$ and the state variables $\{x_i\}$, are identifiable. Once the $\{x_i\}$ have been identified, P1 to P3 guarantee that all of the other risk-neutral parameters specified in the canonical representation are identifiable.

It is worth noting that if one were to apply this argument to the more general square root process

$$dx_i(t) = \left(\kappa_{i0} + \sum_{i'=1}^{m} \kappa_{ii'}\, x_{i'}(t)\right) dt + \sqrt{a_i + b_i x_i}\, dz_i^Q(t), \quad dz_i^Q(t)\, dz_{i'}^Q(t) = \mathbf{1}_{\{i=i'\}}\, dt,$$

rather than to equation (14), the state vector would not be identifiable. This follows from the fact that for all $n > 0$, the $V^{(n)}$ can be written as

$$V^{(n)} = \sum_{i=1}^{m} \left(\omega^i_{jj'} b_i\right)^{n+1} \left(a_i + b_i x_i(t)\right) b_i^{-2},$$

which for all $n > 0$ is a function of the quantities $\omega^i_{jj'} b_i$ and $(a_i + b_i x_i(t))\, b_i^{-2}$. Thus, while these quantities can be identified, the values of $\omega^i_{jj'}$, a_i, b_i, and x_i can never be identified separately. Therefore, without loss of generality, we assume (as do DS) that $a_i = 0$ and $b_i = 1$ in equation (14).

E. Rotation from Latent $\{x\}$ Variables to Observables

While our canonical representation in terms of $\{\mu_j, x_i\}$ is maximal, it is not fully observable since the x variables are latent. Fortunately, a variety of alternatives exist for rotating $\{x_i\}$ to observable variables. We note that the extant literature identifies some of the simplest alternatives. For example, a one-factor $A_1(1)$ model can be re-expressed as a translated Cox et al. (1985) process for the short rate. As another example, the two-factor $A_2(2)$ model can be rotated to the short rate and its variance following the approach of Longstaff and Schwartz (1992).

For models with a mixture of Gaussian and square root processes, it is often more straightforward to rotate from $\{\mu_j, x_i\}$ to $\{\mu_j, V_{jj'}\}$, where the new state variables,

$$V_{jj'}(x) = \frac{1}{dt}\mathrm{E}^Q[d\mu_j d\mu_{j'}], \tag{28}$$

represent quadratic covariation processes and are therefore observable by properties P2 and P3 in Section I.A. In several examples below, we find these rotations to be particularly tractable. A second alternative is to choose the new state variable to be the risk-neutral drift of μ_{N-m-1} from equation (16). Finally, one could also choose the drifts of $V_{jj'}(x)$ as some of the state variables, where the observability of these variables follows property P1.

Thus, there are many possible rotations from our canonical representation to replace the latent $\{x_i\}$ vector with observable state variables. Which choice is preferable depends on the particular model and estimation strategy to be employed. The important point is that all these distinct representations in terms of observables are equivalent in that they are all both maximal and globally identifiable. This follows from the fact that they are all invariant transformations of our canonical representation, which we have proved to be maximal and identifiable. The rotation to observable state variables simply guarantees global identification.[22,23]

III. Examples

In this section, we consider some examples to provide some intuition as to why specifying a model in terms of economically meaningful variables guarantees that both the state vector and the parameter vector are globally identifiable.

A. The $A_0(3)$ Model

We investigate the $A_0(3)$ model because it is a widely used benchmark that yields closed-form solutions for bond prices. Further, it allows us to demonstrate

[22] We note that our canonical representation is trivially only locally identified since the x_i variables are perfectly symmetric and therefore interchangeable.

[23] For the knife-edge case that some parameter values are exactly zero, not all rotations are possible.

in a transparent manner that the risk-neutral parameters of our canonical representation are identifiable from bond prices alone. That is, the prices of exotic securities are unnecessary for identification purposes.

Since the $A_0(3)$ model is, by definition, a three-factor model with no square root processes, the entire state vector comes from the $\{r, \mu\}$ variables. There are 10 risk-neutral parameters in the drift and covariance matrix:

$$\begin{bmatrix} dr \\ d\mu_1 \\ d\mu_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \gamma + \kappa_0 r + \kappa_1\mu_1 + \kappa_2\mu_2 \end{bmatrix} dt, \begin{bmatrix} V_{00} & V_{01} & V_{02} \\ V_{01} & V_{11} & V_{12} \\ V_{02} & V_{12} & V_{22} \end{bmatrix} dt \right). \quad (29)$$

Any additional parameter in the mean or the covariance matrix would either make the model unidentifiable or inconsistent with the definitions of μ_1 and μ_2.

To see how the model is identifiable from a panel of bond prices, note that the observability of the yield curve means that all of its Taylor series components are also observable. Thus, by observing $Y^0\ (t)$ and $Y^1\ (t)$, equations (21) and (22) imply that we also observe $r(t)$ and $\mu_1(t)$. By observing a time series of $r(t)$, we observe its variance V_{00}. Given V_{00}, and by observing $Y^2\ (t)$, equation (23) implies that we observe $\mu_2(t)$ as well. Thus, all of the state variables in the model are observable from bond data only.

To show that the parameters of the risk-neutral drift are identifiable from bond data alone, first note that equation (24), together with the form of the risk-neutral drift of μ_2, implies that

$$Y^3(t) = \frac{1}{4}\left(\gamma - 3V_{01} + \kappa_0\, r(t) + \kappa_1\, \mu_1(t) + \kappa_2\, \mu_2(t)\right). \quad (30)$$

Since $Y^3(t)$ is observable from the term structure and V_{01} from the quadratic variation of $r(t)$ and $\mu_1(t)$, all time-series variables in this equation are observed. The implication is that if we observe yield curves on four different dates, we will have four equations for the four unknown parameters $\{\gamma, \kappa_0, \kappa_1, \kappa_2\}$, implying that the parameters that make up the risk-neutral drift of $\mu_2(t)$ are identifiable from bond prices alone. Since the state variables are observable, all covariance matrix parameters are identifiable using time-series information. Finally, given the time series of the state vector, all parameters that show up in the risk premia (with the qualification given in footnote 12) are also identifiable, even if the implied historical dynamics of the state vector fall outside of the affine framework.

B. The $A_1(3)$ Model

The $A_1(3)$ model is a popular model for describing three-factor dynamics in a way that allows for the presence of stochastic volatility in interest rates. In our canonical form, the model is written in terms of the state vector $S = [x\ \ r\ \ \mu_1]^\top$, where we drop the subscript on x_1. Equations (14) and (16) imply that

$$\begin{bmatrix} dx \\ dr \\ d\mu_1 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} a_1^Q + b_{11}^Q x \\ \mu_1 \\ a_3^Q + b_{31}^Q x + b_{32}^Q r + b_{33}^Q \mu_1 \end{bmatrix} dt,\ [\Omega^0 + \Omega^1 x]\, dt\right), \tag{31}$$

where

$$\Omega^0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \omega_{22}^0 & \omega_{23}^0 \\ 0 & \omega_{23}^0 & \omega_{33}^0 \end{bmatrix} \quad \text{and} \quad \Omega^1 = \begin{bmatrix} 1 & \omega_{12}^1 & \omega_{13}^1 \\ \omega_{12}^1 & \omega_{22}^1 & \omega_{23}^1 \\ \omega_{13}^1 & \omega_{23}^1 & \omega_{33}^1 \end{bmatrix}. \tag{32}$$

The restrictions on the Ω_0 and Ω_1 matrices follow directly from our canonical representation (equations (14), (17), and (18)) above and the requirement that $\Omega_t = \Omega_0 + \Omega_1 x_t$ be a valid (i.e., positive definite) covariance matrix.

The $A_1(3)$ model is more easily interpreted by rotating to a system in which the three state variables are the short rate, the risk-neutral drift of the short rate, and variance of the short rate (denoted V instead of V_{00} for brevity). The observable state vector is therefore defined as $X = [V\ r\ \mu_1]^\top$, where by definition [24]

$$V(t) = \omega_{22}^0 + \omega_{22}^1 x(t).$$

Deriving the dynamics of the new rotation is straightforward.[25] We find

$$\begin{bmatrix} dV \\ dr \\ d\mu_1 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} a_V^Q + b_{VV}^Q V \\ \mu_1 \\ a_\mu^Q + b_{\mu V}^Q V + b_{\mu r}^Q r + b_{\mu\mu}^Q \mu_1 \end{bmatrix} dt,\ \left[\Omega_0 + \Omega_V(V - \underline{V})\right] dt\right), \tag{33}$$

where

$$\Omega_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \underline{V} & c_{r\mu}^0 \\ 0 & c_{r\mu}^0 & \sigma_\mu^0 \end{bmatrix} \text{ and } \Omega_V = \begin{bmatrix} \sigma_V^V & c_{Vr}^V & c_{V\mu}^V \\ c_{Vr}^V & 1 & c_{r\mu}^V \\ c_{V\mu}^V & c_{r\mu}^V & \sigma_\mu^V \end{bmatrix}. \tag{34}$$

[24] Clearly, in the degenerate case where $\omega_{22}^1 = 0$, one cannot rotate to this state variable. Instead, a different alternative, such as the quadratic variation of μ_1, must be chosen.

[25] At first glance, it appears that we lose the two parameters ω_{22}^0 and ω_{22}^1 under this rotation, as they are subsumed into the definition of V. We emphasize that this cannot actually be the case, since we have already shown in Section II.D that the canonical rotation is identifiable. Indeed, simple calculations show that the quadratic variation of V is $\frac{1}{dt}\mathrm{Var}(dV) = \omega_{22}^1(V(t) - \omega_{22}^0)$. Since this replaces the quadratic variation of x, which has no unknown parameters, the rotation from the canonical form to the observable state vectors does not result in any gain or loss in the total number of parameters. We confirm this result below.

The form of the matrices Ω_0 and Ω_V is easily understood. First, all covariances are affine in V because it is the only state variable that affects covariances in the $A_1(3)$ model. Second, the variance V must have a lower bound $\underline{V}$ that is positive. When V approaches its lower bound, its variance and covariances with other variables must vanish, though the volatilities of the other two state variables need not. Finally, since V is the variance of the short rate, the variance of dr must have a unit loading on V and no intercept. The most general covariance matrix with all of these properties is Ω_t.

We note that the model has a total of 14 risk-neutral parameters (six in the drift and eight in the covariance matrix), which agrees with the prediction of equation (18).

Using equations (21)–(24), we can show that in most cases this model can be fully identified from bond prices alone. Given the specification of the $A_1(3)$ risk-neutral drift, equation (23) implies that the second-order Taylor series term is equal to

$$Y^2(t) = \frac{1}{3}\left[\mu_2(t) - V(t)\right] = \frac{1}{3}\left[a^Q_\mu + b^Q_{\mu V}V(t) + b^Q_{\mu r}r(t) + b^Q_{\mu\mu}\mu_1(t) - V(t)\right].$$

Because Y^2, V, r, and μ_1 can be observed from short-maturity yields and their quadratic variations, all four parameters in the risk-neutral drift of $d\mu_1$ can be identified given at least four independent observations.

To identify a^Q_V and b^Q_{VV} from the risk-neutral drift of V, note that the form of

$$\mu_3 \equiv \frac{1}{dt}\mathrm{E}^Q\left[d\mu_2\right] = \frac{1}{dt}\mathrm{E}^Q\left[b^Q_{\mu V}dV(t) + b^Q_{\mu r}dr(t) + b^Q_{\mu\mu}d\mu_1(t)\right]$$

combined with equation (24) implies that

$$Y^3(t) - \frac{1}{4}\left(b^Q_{\mu r}\mu_1(t) + b^Q_{\mu\mu}\mu_2(t) - 3V_{01}(t)\right) = \frac{1}{4}\left(b^Q_{\mu V} - 1\right)\left(a^Q_V + b^Q_{VV}V(t)\right).$$

Since all time-series variables are again observable and all parameters except a^Q_V and b^Q_{VV} were identified already, we need just two observations to identify the two unknown parameters. Parameters associated with the historical measure can then be identified using time-series information.

Unfortunately, this scheme fails when $b^Q{}_{\mu V} = 1$, which turns out to be one of the conditions that is required for the $A_1(3)$ model to display unspanned stochastic volatility. If all USV conditions hold, then it is not possible to infer a^Q_V and b^Q_{VV} from bond prices alone. This illustrates the impossibility of proving generally that models that are identifiable given all possible fixed income derivatives data are identifiable given bond data alone.

C. The $A_2(4)$ Model

In addition to providing a tangible example of a model with complex mean and covariance dynamics, we investigate the $A_2(4)$ model because it is an example of a model whose Q-maximal representation is more general than that obtained by DS. In particular, we find 24 risk-neutral parameters rather than the 23 predicted by DS. To demonstrate, we start with latent square root processes and then show that rotation to an observable state vector does not affect the number of identifiable risk-neutral parameters.

Since the $A_2(4)$ model by definition has two Gaussian variables and two square root variables, our canonical state vector is $S = [x_1\ x_2\ r\ \mu]^\top$, where we now drop the subscript on μ_1. Following equations (14) and (16), the risk-neutral dynamics are

$$\begin{bmatrix} dx_1 \\ dx_2 \\ dr \\ d\mu \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} a_1^Q + b_{11}^Q x_1 + b_{12}^Q x_2 \\ a_2^Q + b_{21}^Q x_1 + b_{22}^Q x_2 \\ \mu \\ a_4^Q + b_{41}^Q x_1 + b_{42}^Q x_2 + b_{43}^Q r + b_{44}^Q \mu \end{bmatrix} dt,\ [\Omega^0 + \Omega^1 x_1 + \Omega^2 x_2]\, dt \right), \tag{35}$$

where Ω^0, Ω^1, and Ω^2 are given by

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \omega_{33}^0 & \omega_{34}^0 \\ 0 & 0 & \omega_{34}^0 & \omega_{44}^0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & \omega_{13}^1 & \omega_{14}^1 \\ 0 & 0 & 0 & 0 \\ \omega_{13}^1 & 0 & \omega_{33}^1 & \omega_{34}^1 \\ \omega_{14}^1 & 0 & \omega_{34}^1 & \omega_{44}^1 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & \omega_{23}^2 & \omega_{24}^2 \\ 0 & \omega_{23}^2 & \omega_{33}^2 & \omega_{34}^2 \\ 0 & \omega_{24}^2 & \omega_{34}^2 & \omega_{44}^2 \end{bmatrix},$$

respectively.[26]

Consistent with our general result above, the total number of parameters is 24, which is one more than obtained by DS. The source of this discrepancy is related to the way the affine class of models is described, specifically, whether it is in the mean-covariance form used above or the traditional stochastic differential equation (SDE) form (19).

We note that the dynamics of the model expressed in equations (35) can be written as a system of SDEs. However, they cannot be written in the traditional affine form (19), that is, with diffusion coefficients linear in square root terms, using only four Brownian motions. Because DS restrict themselves to writing $A_m(N)$ using only N Brownian motions to represent their N state variables' diffusion matrix in traditional affine form (19), they obtain a less general characterization than we do for certain N-factor models.

For the case of the maximal $A_2(4)$ model, somewhat tedious calculations show that the model can be written in a form consistent with (19) using five Brownian

[26] As before, ones on the diagonals of Ω^1 and Ω^2 follow from our standardization of the square root variables. Zeros on the diagonals of all three matrices ensure that the variances of x_1 and x_2 approach zero as the processes themselves approach zero, while off-diagonal zeros are required for positive definiteness of Ω_t.

motions (i.e, one more than the number of state variables) as

$$
\begin{aligned}
dx_1 &= (\kappa_{10} + \kappa_{11}x_1 + \kappa_{12}x_2)\,dt + \sqrt{x_1}dz_1^Q \\
dx_2 &= (\kappa_{20} + \kappa_{21}x_1 + \kappa_{22}x_2)\,dt + \sqrt{x_2}dz_2^Q \\
dr &= \mu dt + \nu_{r1}\sqrt{x_1}\,dz_1^Q + \nu_{r2}\sqrt{x_2}\,dz_2^Q \\
&\quad + \sqrt{a_1 + b_1x_1 + b_2x_2}dz_3^Q + \sqrt{c_1 + d_1x_1 + d_2x_2}dz_4^Q \\
d\mu &= \left(\kappa_{\mu 0} + \kappa_{\mu r}r + \kappa_{\mu\mu}\mu + \kappa_{\mu 1}x_1 + \kappa_{\mu 2}x_2\right)dt + \nu_{\mu 1}\sqrt{x_1}\,dz_1^Q + \nu_{\mu 2}\sqrt{x_2}\,dz_2^Q \\
&\quad + \sqrt{a_1 + b_1x_1 + b_2x_2}dz_3^Q + \sqrt{e_1 + f_1x_1 + f_2x_2}dz_5^Q.
\end{aligned}
$$

There are multiple ways to rewrite (35) in terms of four Brownian motions. For example, a simple formulation in terms of four Brownian motions is

$$
\begin{aligned}
dx_1 &= (\kappa_{10} + \kappa_{11}x_1 + \kappa_{12}x_2)\,dt + \sqrt{x_1}dw_1^Q \\
dx_2 &= (\kappa_{20} + \kappa_{21}x_1 + \kappa_{22}x_2)\,dt + \sqrt{x_2}dw_2^Q \\
dr &= \mu\,dt + \nu_{r1}\sqrt{x_1}\,dw_1^Q + \nu_{r2}\sqrt{x_2}\,dw_2^Q \\
&\quad + \sqrt{a_1 + c_1 + (b_1 + d_1)x_1 + (b_2 + d_2)x_2}dw_3^Q \\
d\mu &= \left(\kappa_{\mu 0} + \kappa_{\mu r}r + \kappa_{\mu\mu}\mu + \kappa_{\mu 1}x_1 + \kappa_{\mu 2}x_2\right)dt + \nu_{\mu 1}\sqrt{x_1}\,dw_1^Q + \nu_{\mu 2}\sqrt{x_2}\,dw_2^Q \\
&\quad + \frac{a_1 + b_1x_1 + b_2x_2}{\sqrt{a_1 + c_1 + (b_1 + d_1)x_1 + (b_2 + d_2)x_2}}dw_3^Q \\
&\quad + \sqrt{a_1 + e_1 + (b_1 + f_1)x_1 + (b_2 + f_2)x_2 - \frac{(a_1 + b_1x_1 + b_2x_2)^2}{a_1 + c_1 + (b_1 + d_1)x_1 + (b_2 + d_2)x_2}}dw_4^Q.
\end{aligned}
$$

Clearly, this is not of the form (19) that was assumed by DS, but it nevertheless implies instantaneous means and covariances that are affine in S. The variance of $d\mu$, for instance, is easily shown to equal

$$
a_1 + e_1 + \left(\nu_{\mu 1}^2 + b_1 + f_1\right)x_1 + \left(\nu_{\mu 2}^2 + b_2 + f_2\right)x_2.
$$

While all of these representations have 24 parameters, they are expressed in terms of the latent state variables x_1 and x_2. As we note in Section II.E, the square root processes can be rotated to a number of different observables. One possibility is to rotate x_1 and x_2 to the variance of dr and the risk-neutral drift of that variance. However, given that we have two Gaussian processes and two square root processes, arguably a more natural choice is to rotate to V, the instantaneous variance of dr (previously defined as V_{00}), and U, the instantaneous variance of $d\mu$ (previously defined as V_{11}). Let $X = [V\ U\ r\ \mu]^\top$ denote the state vector under this new rotation. Note that from equation (35) we have

$$
V = \omega_{33}^0 + \omega_{33}^1 x_1 + \omega_{33}^2 x_2 \tag{36}
$$

$$U = \omega_{44}^0 + \omega_{44}^1 x_1 + \omega_{44}^2 x_2. \tag{37}$$

Since all of these coefficients need to be positive to guarantee admissibility, and since both x_1 and x_2 have minimum values of zero, $\omega_{33}^0 \equiv \underline{V}$ is the lowest value that V can take. Similarly, $\omega_{44}^0 \equiv \underline{U}$ is the lowest value that U can take. As such, it is convenient to write

$$\begin{bmatrix} V - \underline{V} \\ U - \underline{U} \end{bmatrix} = \begin{bmatrix} \omega_{33}^1 & \omega_{33}^2 \\ \omega_{44}^1 & \omega_{44}^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{38}$$

This matrix form can be inverted to write the $\{x_i\}$ in terms of $\{V, U\}$, allowing for a simple rotation to observables that preserves the same form for the drift (though with different parameter values) and that allows the covariance matrix to be written as

$$\frac{1}{dt}\mathrm{Cov}(dX, dX^\top) = \Omega^0 + \Omega^V (V - \underline{V}) + \Omega^U (U - \underline{U}),$$

where Ω^0 is unchanged from above. Unfortunately, while they are easily derived, the restrictions on Ω^V and Ω^U are somewhat messy, and we do not report them here. Nevertheless, a rotation to this form may be desirable given that the observability of r, μ_1, V, and U makes all the parameters in this representation globally identifiable.

IV. Data

So far, our discussion has been mostly theoretical. In the next few sections, we demonstrate that our representations have practical advantages as well.

Our data set is derived from weekly observations of all maturities of LIBOR and swap contracts that were available over the entire 1988 to 2005 sample period. These consisted of LIBOR rates with maturities of 1, 3, 6, 9, and 12 months and swap rates with maturities of 2, 3, 4, 5, 7, and 10 years. Swap rates, which are recorded in London at 5:30 pm, are Wednesday values. LIBOR rates, meanwhile, are fixed at 11:00 am in London, and it turns out that Thursday morning LIBOR rates are most highly correlated with Wednesday evening swap rates. We therefore build our term structures using Wednesday swap rates and Thursday LIBOR rates.[27]

Swap rates are converted to zero coupon rates assuming that they can be valued as par bond yields.[28] In order to minimize concerns about pre-smoothing the

[27] In other work, we also use a simple model to impute approximate LIBOR values at the close. Doing so does not have any material effect on our results.

[28] If swaps were free of default risk, this would directly follow from absence of arbitrage. In the presence of credit risk, this assumption is warranted if credit quality is homogeneous across swap and LIBOR markets. In that case, the zero coupon curve corresponds to a risk-adjusted corporate curve for issuers with refreshed AA credit quality (see Duffie and Singleton (1997), Collin-Dufresne and Solnik (2001), and Johannes and Sundaresan (2007)).

data, we employ the Unsmoothed Fama-Bliss method (Fama and Bliss 1987), which assumes a piecewise constant forward rate curve. From the bootstrapped yield curves, we extract the zero coupon yields with the same maturities as the LIBOR and swap rate inputs. This results in zero yields up to 1 year that are unchanged from observed LIBOR values and estimated for maturities of 2, 3, 4, 5, 7, and 10 years, a total of 11 yields.

We note that both the bootstrapped zero coupon yields we use and the results of our model estimations are essentially unchanged when we instead use the extended Nelson–Siegel method of Bliss (1997). This is likely due to the fact that differences between bootstrapping methods are most pronounced for maturities that do not correspond to any observed yield. Since both swap and LIBOR rates are quoted on a constant maturity basis, we are able to construct a panel of constant maturity zeros with maturities that are always the same as some underlying instrument. We believe that this aspect of LIBOR and swap markets is an additional advantage over using Treasuries.

In some of our analysis, we also employ a measure of implied volatility from the Eurodollar futures options market. Since the Eurodollar futures price is equal to $100 - F$, where F is the annualized futures rate, an option is considered at the money (ATM) when the strike price is equal to $100 - F$. From closing prices on each Wednesday in our sample, we find the shortest maturity call options (excluding options within a week to expiration) with strike prices that are within \$1 of ATM and that can be matched with a put of the same strike and maturity. To reduce measurement problems arising from asynchronous recording of futures and options prices, we back out the implied futures price from each pair of options using put-call parity. We use the Black (1976) formula to impute a single measure of volatility for each pair. Finally, to ascribe a single ATM implied volatility to each day, we fit a quadratic polynomial to these implied volatilities as a function of the strike prices (again, all of which are within \$1 of ATM). This procedure results in an estimate of the *proportional* volatility of $100 - F$. We take this value and multiply by $1 - F/100$ to obtain an estimate of the *level* of yield volatility.

It is known (see Ledoit, Santa-Clara, and Yan 2002) that even when the Black (1976) model is misspecified, ATM-implied volatilities from that model often converge to true volatilities as both the option and underlying maturities approach zero. However, since the options we use have on average 7 weeks until expiration, and because the underlying futures are themselves 3-month contracts, we do not consider this theoretical result as being obtained in our sample. Rather, we view our implied volatility series as simply a reasonable approximation of current yield volatility. We will see below that this series appears to have good empirical properties, at least for the period after 1990; prior to 1990, most likely due to the newness of the market, the series contained a much higher level of volatility that appeared to be mostly measurement error. We therefore restrict our use of these data to the 1991 to 2005 subsample.

V. Model-Free Estimates of Gaussian State Variables

When a model is specified in terms of latent state variables, estimates of the state vector depend on the assumed values of the parameters, which are not initially available. In contrast, as demonstrated above, the Gaussian state variables (r, μ_1, μ_2) in our representation are proportional to the level, slope, and (to a close approximation) curvature of the term structure at zero. In theory, this suggests that it should be possible to obtain model-independent estimates of these state variables simply by observing the yield curve. Such estimates can be quite valuable. For example, they can be used to obtain reasonable estimates of the parameters, which in turn can be used as first guesses for a full-fledged estimation. This should be especially useful for multifactor models, for which estimation can be computationally burdensome.

In practice, however, we do not observe the entire (continuous) term structure of zero coupon yields. Rather, we observe only discrete points along the curve, and these may be contaminated by noise resulting from bid-ask spreads or inexact bootstrapping procedures, for example. In fact, prior work (e.g., Duffee (1996) and Dai and Singleton (2002)) indicates that the shortest maturity yields, which are in theory the ones most directly related to our state variables, appear especially contaminated by idiosyncratic noise. Thus, any attempt to directly measure the level, slope, and curvature of the term structure at zero, say by fitting a spline to the yield curve and extrapolating down to near-zero maturities, is likely to result in large errors.

We propose an alternative approach to estimating the first three $\{\mu\}$ variables that dramatically reduces estimation error by imposing some parametric yet still model-free structure. Our approach is based on Litterman and Scheinkman's (1991) finding that three principal components can explain the vast majority of the variation in bond yields. For our analysis, we use all but 3 of the 11 zero coupon yields in the sample to extract principal components. The other three yields, with 9-month, 4-year, and 10-year maturities, are reserved for use elsewhere in an effort to avoid utilizing the same data more than once. Over our sample period, for the eight yields that are analyzed, the first three principal components explain 98.8% of the variation in yield levels. Thus, if $Y(t, \tau)$ denotes the time-t value of the τ-year zero rate and $\mathcal{P}_k(t)$ the contemporaneous realization of the k^{th} principal component, then the approximation

$$Y(t,\tau) \approx \sum_{k=1}^{3} f_k(\tau)\mathcal{P}_k(t) \tag{39}$$

holds with great accuracy.

The result above is useful because all yield curve derivatives (e.g., slope and curvature) can be written as sums of the derivatives of the $f_k(\tau)$ "loading" functions

$$\frac{\partial^n Y(t,\tau)}{\partial \tau^n} \approx \sum_{k=1}^{3} \frac{\partial^n f_k(\tau)}{\partial \tau^n}\mathcal{P}_k(t), \tag{40}$$

given estimates of those derivatives. This means that the level, slope, and curvature of the yield curve at zero maturity in *every* period can now be obtained by extrapolating just three functions down to zero. This need be done only once for the entire sample; that is, we do not need to extrapolate one yield curve for each day. Furthermore, since these three functions are estimated using the entire sample, they are likely to be estimated with relatively little error.

While principal component (PC) analysis results in estimates of the $f_k(\tau)$ for a discrete set of $\tau > 0$, we are interested in the local behavior of the $f_k(\tau)$ around $\tau = 0$. We extrapolate each of the $f_k(\tau)$ down to zero using low-order polynomials. Since a global approximation is unnecessary, we fit these curves only out to 1 year using the yields with maturities of 1, 3, 6, and 12 months. Polynomial orders were initially chosen subjectively, but a simple cross-validation exercise confirmed these choices under both mean squared and mean absolute error criteria.[29] This results in a linear polynomial being used to approximate $f_1(\tau)$, a quadratic for $f_2(\tau)$, and a cubic for $f_3(\tau)$.

Given these fitted polynomials, we can calculate the values in (39) and derivatives in (40) for $\tau = 0$ for each day in the sample. From our results above, the three state variables r, μ_1, and μ_2 are equal to, respectively, the level of the yield curve at zero, twice the first derivative at zero, and three times the second derivative at zero plus the variance of the short rate. Since both the level and variation of the short-rate variance are extremely small relative to the variance in the second derivative, a precise accounting of short-rate variance is unimportant. Thus, we simply assume that it is constant and equal to the annualized sample variance of changes in the 1-month yield.[30]

Figure 1 illustrates our approach. The top panel contains results for $f_1(\tau)$, the loadings on the first principal component. Each circle represents a value of $f_1(\tau)$ corresponding to one of the maturities used in the PC analysis. The line denotes the best-fit linear polynomial to the maturities out to 1 year. With all loadings having similar values, the standard interpretation of the first principal component as a level factor is clear. The next two columns contain results for the second and third principal components, which also have their usual interpretations as slope and curvature factors.

[29] To select the polynomial order used to approximate $f_k(\tau)$, we use a separate "leave-one-out" cross-validation for each k. This entails a series of regressions of $f_k(\tau_i)$ on a constant and the first P powers of τ_i, where $\tau_i = \{1, 3, 6, 9, 12\}$ months. In each regression, a different observation is held out and then used to compute an error. Squared or absolute values of these errors are then summed across maturities, and the polynomial order P with the smallest sum is then returned as the preferred value. In our data, both criteria agree for each k. Note that to choose the polynomial order we include the 9-month yield in the PC analysis. This is because we need at least five different maturities to estimate the cubic polynomial given that one of the five will be left out of each regression. Once the polynomial order is chosen, we repeat the PC analysis without the 9-month rate.

[30] The unconditional standard deviation of three times the second derivative is approximately 0.045. The variance of the short rate is something close to 0.0001, possibly taking values three to four times as large in the most volatile periods in our sample. This degree of variation is completely dominated by the movements in curvature, and it is without consequence to ignore it here.

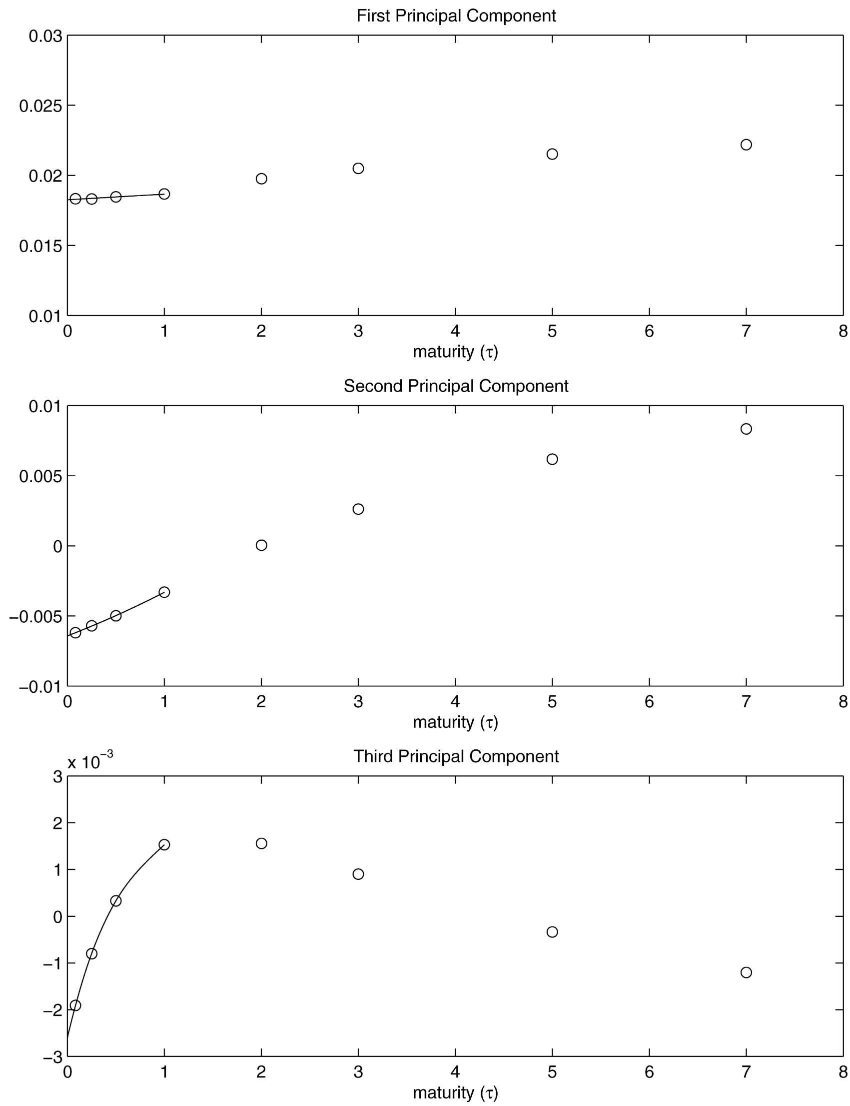

Figure 1. Principal component loadings. This figure displays a principal components analysis of yield levels performed over the 1988 to 2005 sample period. Circles denote the loadings ($f_k(\tau)$) of each included maturity on the first three principal components. Smooth lines are the result of a polynomial approximation procedure designed to extrapolate each loading (and its derivatives) down to a zero maturity.

All three curves are generally smooth and well behaved. As such, one would anticipate that extrapolation down to zero generates reasonably accurate results. To verify this conjecture, we consider a brief Monte Carlo exercise designed to replicate our procedure as closely as possible. We begin by simulating zero coupon and par bond yields that are identical in maturity and length to

Table I

Accuracy of Model-Free State Variable Estimation

This table reports Monte Carlo results based on simulated interest rates that are identical in maturity and sample length to the full sample of data used in this paper. In addition to a zero measurement error case, two levels of measurement errors (*SD*s of 0.5 b.p. and 2 b.p.) are considered. The data construction procedure used in the empirical analysis is then repeated on each simulated data set, leading to a set of model-free state variables. The table reports means and standard deviations (in parentheses) of the coefficients and R-squares from regressions of the form

$$\text{Actual State Variable}(t) = \alpha + \beta \times \text{Estimated State Variable}(t) + \epsilon(t).$$

Yields are simulated using Duffee's (2002) preferred version of the $A_0(3)$ model, a three-factor model with constant covariances, with parameter values from table III of his paper. The state variables r and μ_1 denote the short rate and its risk-neutral drift, respectively, while μ_2 denotes the risk-neutral drift of μ_1.

	State Variable $= r$			State Variable $= \mu_1$			State Variable $= \mu_2$		
Error *SD*	$1000 \times \hat{\alpha}$	$\hat{\beta}$	R^2	$1000 \times \hat{\alpha}$	$\hat{\beta}$	R^2	$1000 \times \hat{\alpha}$	$\hat{\beta}$	R^2
0	−0.034	1.000	1.000	0.012	1.008	1.000	0.500	1.096	1.000
	(0.017)	(0.000)	(0.000)	(0.049)	(0.001)	(0.000)	(0.225)	(0.003)	(0.000)
0.5	−0.034	1.000	1.000	0.011	1.008	1.000	0.499	1.097	0.999
	(0.023)	(0.000)	(0.000)	(0.118)	(0.005)	(0.000)	(0.854)	(0.023)	(0.001)
2	−0.048	1.000	1.000	0.096	1.014	0.998	−0.334	1.131	0.987
	(0.065)	(0.001)	(0.000)	(0.493)	(0.021)	(0.001)	(3.888)	(0.128)	(0.015)

the data in our sample. For this purpose we use Duffee's (2002) preferred version of the essentially affine $A_0(3)$ model, with parameter values taken from table III of that paper.[31] We then add independent and identically distributed (i.i.d.) Gaussian measurement errors to the simulated zero coupon and par bond yields. We consider a low-error case in which the errors have a standard deviation of one-half of a basis point and a high-error case in which the standard deviation is two basis points. The latter represents an extremely large value, as i.i.d. errors of that magnitude generate negative serial correlation in daily yield changes that are between −0.10 and −0.25, which are clearly at odds with the near-zero autocorrelations observed in the data.

Every aspect of our bootstrapping, principal components, cross-validation, and polynomial extrapolation procedure is repeated on each set of simulated data. In the end, we are left with both actual and estimated values of the three state variables r, μ_1, and μ_2. To assess the closeness between the actual and estimated time series, we regress the former on the latter. Table I reports the means and standard deviations of the estimated intercepts and slopes, as well as the regression R-squares, from 1,000 Monte Carlo simulations. If the model-independent estimates are unbiased and accurate, then we expect to find intercepts close to zero, slopes close to one, and high R-squares.

[31] We also use parameter values from column III of our own Table II, which generate essentially the same results.

The results reported in Table I are very encouraging. They show that the estimate of r is unbiased and accurate even given a high level of noise, with R-squares that on average are greater than 0.999 even for the highest level of error. The results for μ_1 are also extremely promising, with average R-squares still no lower than 0.998 and essentially no bias. Only for μ_2 do we see minor evidence of bias, with the average slope being too high by 10% to 13%. Even in this case, however, the average R-squares for the highest level of measurement error are still at least 0.987. We demonstrate below that similar accuracy obtains using actual data, since we find our model-independent estimates to be extremely highly correlated with estimates from standard estimation procedures, at least for models that are not clearly misspecified.

As illustrated most clearly in Section III. A, a major theoretical advantage of our rotation is the ability to easily prove identifiability of the parameters of the Q drift of the state vector. To see how this idea may be applied in practice, consider a three-factor affine model written in terms of r, μ_1, and μ_2, whose drift by definition must be of the form given in equation (29), even if the model is outside the $A_0(3)$ class and the covariance matrix differs.

Using the procedure described above, we obtain model-free proxies of $r(t), \mu_1(t), \mu_2(t)$, and $\frac{\partial^3 Y(t,\tau)}{\partial\tau^3}|_{\tau=0} \equiv Y^3(t)$. Assuming the variance terms are small enough to ignore (which is true in practice), we can then use equation (30) to obtain rough estimates of the four unknown parameters of the risk-neutral drift of μ_2 simply by running an OLS regression.

In unreported simulation results, we find that this approach is reasonably accurate when the degree of measurement error in yields is not too large. In general, however, $Y^3(t)$ and higher order terms appear to be estimated with successively lower accuracy, and this makes model-free estimates of Gaussian state variables beyond $r(t), \mu_1(t)$, and $\mu_2(t)$ appear less reasonable. Nevertheless, below we use model-free estimates of $Y^3(t)$ to obtain starting values for our estimations, and the result is a set of values that are surprisingly close to our maximum-likelihood estimates.

VI. Models and Estimation Methods

We investigate the usefulness of our rotation and the observability of our state vectors in two examples. We first consider the three-factor Gaussian $A_0(3)$ model, which Duffee (2002) finds to be most accurate in forecasting future bond yields. This model is estimated over the 1988 to 2005 sample period. Second, we consider the $A_1(3)$ model, which also has three factors, one of which is a square root process that can drive conditional volatilities. Since it is not our goal to compare these specifications, we estimate the latter model over the 1991 to 2005 period, the subsample over which our implied volatility data appear to be most reliable.

For both models, we compare alternative estimation approaches. The most familiar is the now-standard inversion method in which a vector of three yields, denoted $Y_o(t)$, is assumed to be observed without measurement error. The state vector, denoted $X(t)$, is then obtained by inverting the relation

$$Y_o(t) = H_0 + H_1 X(t), \tag{41}$$

where H_0 and H_1 are implied by

$$Y(t, \tau) = -\frac{A(\tau)}{\tau} + \frac{B(\tau)^\top}{\tau} X(t). \tag{42}$$

In our analysis these yields are taken to be the 3-month, 2-year, and 10-year maturities.

The vectors of remaining zero coupon yields, $Y_e(t)$, are assumed to be equal to their model-implied values plus Gaussian measurement errors that are uncorrelated both cross-sectionally and over time. This results in a log-likelihood function of the form

$$l\left(\phi, D\right) = \sum_{t=1}^{T} \left[-\ln|\det(H_1)| + \ln p\left(X(t)|X(t-1), \phi\right) + \ln p\left(Y_e(t)|X(t), D, \phi\right)\right], \tag{43}$$

where ϕ denotes the affine model parameters and D the diagonal covariance matrix of measurement errors. The transition density of $X(t)$ is a Gaussian approximation (exact in the case of the $A_0(3)$ model), where the true discrete-time means and covariances are derived according to Fisher and Gilles (1996). The mean and covariances of the stationary distribution of $X(t)$ are used to compute the likelihood of the initial observation. In sum, this approach is identical to that of Duffee (2002) and others.

We compare this standard approach with several alternatives in which the state vector is treated as observable and the Gaussian variables equal to the model-free values computed in Section V. Recall that the Gaussian state variables are computed using a principal components analysis on all but 3 of the 11 zero coupon yields in our sample. To avoid using the same data twice, only those three yields not used to compute principal components, namely the 9-month, 4-year, and 10-year yields, are directly included in the estimation. The details of the estimation methods based on observed state variables are somewhat model specific, so we describe them separately for each specification.

It is worth noting that merely rotating the state vector does not change the likelihood of the data as long as that rotation is invariant. For example, parameterizing a model in terms of $\{r, \mu_1, V_{00}\}$ rather than $\{r, \mu_1, \mu_2\}$ does not, by itself, alter the ability of the model to explain the data. Different estimation methods only result in different parameter estimates because they either compute the likelihood differently (exactly or using an approximation) or because they make different assumptions about what variables are observed with and without error. Thus, while assuming that V_{00} is exactly observed instead of μ_2 can have a major effect on our results, choosing to write the model in terms of V_{00} rather than μ_2 is done solely for reasons of convenience and taste.

A. $A_0(3)$ Specification and Estimation

As Section III.A demonstrates, the $A_0(3)$ model is represented by a state vector that consists of the three Gaussian state variables $X = [r\ \mu_1\ \mu_2]^\top$, whose dynamics are given in equation (29). For ease of interpretation, we write the covariance matrix in terms of standard deviations and correlations as

$$\frac{1}{dt}\text{Cov}\left(dX(t), dX(t)^\top\right) = \begin{bmatrix} \sigma_0^2 & \sigma_0\sigma_1\rho_{01} & \sigma_0\sigma_2\rho_{02} \\ \sigma_0\sigma_1\rho_{01} & \sigma_1^2 & \sigma_1\sigma_2\rho_{12} \\ \sigma_0\sigma_2\rho_{02} & \sigma_1\sigma_2\rho_{12} & \sigma_2^2 \end{bmatrix}. \tag{44}$$

We assume Duffee's (2002) essentially affine risk premia, which allows the drift under the P-measure, written as $a + b\,X(t)$, to be completely unrestricted, so that

$$a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \text{ and } b = \begin{bmatrix} b_{00} & b_{01} & b_{02} \\ b_{10} & b_{11} & b_{12} \\ b_{20} & b_{21} & b_{22} \end{bmatrix}. \tag{45}$$

Note that a and b, which determine the drift of the process under the *historical* measure, are not restricted by the definitions of μ_1 and μ_2, which are the drifts of r and μ_1, respectively, under the *risk-neutral* measure.

While standard estimation methods can be applied here without difficulty, when the state variables $X(t)$ are treated as observable it is natural to propose an even simpler estimation approach that might also perform reasonably well.

Method I

The simplest approach consists of four steps.

1. Estimate $X(t)$ using the model-free approach of Section V.
2. Estimate a and b, the parameters of the P drift of $X(t)$, by running OLS regressions of $\Delta\,X(t)$ on $X(t-1)$ and a constant, then annualizing.
3. Estimate Cov $(dX(\text{t}), dX(\text{t})^\top)/dt$ by annualizing the sample covariance matrix of the residuals from these regressions.
4. Estimate the measurement error variances and the four parameters of the risk-neutral drift of $X(t)$ by maximizing

$$\sum_{t=1}^{T} \ln p(Y_e(t)\,|\,X(t)), \tag{46}$$

   the likelihood of the yields observed with error, treating as given the observed $X(t)$ and the estimated values of the other parameters.

Several possible shortcomings of this approach are notable. First, the covariance parameters affect bond prices, so estimating them using only time-series information is inefficient. Second, the likelihood of the first observation, $X(1)$, is ignored in this approach. Third, discretization bias is ignored when estimating the P drift and covariance parameters. Finally, the values computed for the state variables may be imperfect. While the inversion approach (i.e., Method III below) addresses all four problems, it is possible to consider an intermediate case (i.e., Method II below) in which the first three problems are dealt with but the state variables are still fixed at their estimated values.

Method II

To do so, we consider an estimation approach based on the exact likelihood function of the observed $X(t)$. Following Fisher and Gilles (1996), we compute the discrete-time means and covariances of $X(t)$ given $X(t-1)$. The density of the stationary distribution is used to compute the likelihood of $X(1)$. With three yields, $Y_e(t)$, observed with error, our log likelihood function is of the form

$$l\left(\phi, D\right) = \sum_{t=1}^{T} \left[\ln p\left(X(t) \mid X(t-1), \phi\right) + \ln p\left(Y_e(t) \mid X(t), D, \phi\right)\right], \qquad (47)$$

where ϕ again denotes the parameters of the affine process and D the diagonal covariance matrix of measurement errors.

Method III

This method is the inversion-based approach described above.

B. $A_1(3)$ Specification and Estimation

Section III.B derives the rotation of the $A_1(3)$ model to the observable state vector $X = [V\ r\ \mu_1]^\top$, where V denotes the instantaneous variance of dr.

The model is completed by the specification of the drift under the historical measure. Risk premia are assumed to be of the "extended affine" form,[32] so that the P drift of $X(t)$ is $a + b\,X(t)$, where

$$a = \begin{bmatrix} a_V \\ a_r \\ a_\mu \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} b_{VV} & 0 & 0 \\ b_{rV} & b_{rr} & b_{r\mu} \\ b_{\mu V} & b_{\mu r} & b_{\mu\mu} \end{bmatrix}. \qquad (48)$$

The zero restrictions here also keep $V(t)$ from crossing its lower bound.

We impose several constraints on this system, all of which are designed to impose admissibility and nondegeneracy. First, we require that Ω_0 be positive semidefinite and that Ω_V be positive definite, where both matrices are described in (32). This generates positive definite covariance matrices for all $V(t) > \underline{V}$. In addition, we require that $a_V > -b_{VV}\underline{V}$, which is necessary to ensure $V(t) > \underline{V}$. The same inequality is required for the risk-neutral drift of V as well.

[32] We use the risk premium specification introduced by Duffee (2002) and generalized by Cheridito, Filipovic, and Kimmel (2006) based on theorem 7.19 in Liptser and Shiryaev (1974, p. 294), which shows that if zero is not accessible by $V_t - \underline{V}$ under both measures, then the two measures defined by the P- and Q-measure dynamics of the Markov process above are equivalent. Liptser and Shiryaev's result applies to any process of the diffusion type (see their definition 7, p. 118).

Given the state dependence of the covariance matrix, even if $X(t)$ were observable, this model is somewhat more complicated to estimate. In addition to the standard inversion approach (Method III), we consider three estimation methods that use observable state variables.

Method I
The simplest approach we consider is one in which we have observable proxies for $V(t), r(t)$, and $\mu_1(t)$ and we estimate the parameters of a, b, Ω_0, and Ω_V by assuming that

$$\Delta X(t) \sim \mathrm{N}(\Delta t(a + bX(t-1)), \Delta t(\Omega_0 + \Omega_V(V(t-1) - \underline{V}))), \tag{49}$$

where $\Delta t = 1/52$ is the length of time between observations. As in the $A_0(3)$ model, the parameters of the risk-neutral drift of $X(t)$ are then estimated by maximizing the likelihood of the yields measured with error ($Y_e(t)$) given the observed $X(t)$ and the other model parameters.

This approach ignores nonnormality, discretization bias, the initial distribution of $X(1)$, and the information in bond prices about the parameters of the covariance matrix. To address all but the first of these problems, we propose a second approach.

Method IIa
We next use a Quasi-maximum likelihood (QML) approach with the exact discrete time means and covariances, where the first observation $X(1)$ is assumed to come from a Gaussian distribution with moments equal to the true moments of the stationary distribution of $X(t)$. With three yields again observed with error, we maximize the joint likelihood of all parameters simultaneously. This is identical to the approach taken for the $A_0(3)$ model in maximizing the likelihood in (47), except that now the Gaussian density is an approximation.

Method IIb
Related to the previous approach, we also consider a method that is appropriate when we have observable proxies for r, μ_1, and μ_2, but not V. Given the definition of μ_2 as the risk-neutral drift of μ_1, we note that there is a simple transformation between the two rotations. Namely, if

$$X(t) = \begin{bmatrix} V(t) \\ r(t) \\ \mu_1(t) \end{bmatrix} \text{ and } X^*(t) = \begin{bmatrix} r(t) \\ \mu_1(t) \\ \mu_2(t) \end{bmatrix}, \tag{50}$$

then (33) and the fact that $\mu_2(t) \equiv \mathrm{E}\ [d\mu_1(t)]/dt$ together imply that the transformation from $X(t)$ to $X^*(t)$ is linear:

$$X^*(t) = G_0 + G_1 X(t), \tag{51}$$

where

$$G_0 = \begin{bmatrix} 0 \\ 0 \\ a_\mu^Q \end{bmatrix} \text{ and } G_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ b_{\mu V}^Q & b_{\mu r}^Q & b_{\mu\mu}^Q \end{bmatrix}. \tag{52}$$

Rather than writing a new parameterization of the $A_1(3)$ model in terms of $X^*(t)$, we instead choose to transform model-free observations of $X^*(t)$ into $X(t)$ by inverting (51). Using the change of variables formula, the log likelihood can now be written as

$$l(\phi, D) = \sum_{t=1}^{T} [-\ln \mid \det(G_1) \mid + \ln p(X(t) \mid X(t-1), \phi) + \ln p(Y_e(t) \mid X(t), D, \phi)], \tag{53}$$

where ϕ again denotes the affine model parameters and D the covariance matrix of yield errors. Maintaining the previous parameterization makes it straightforward to compare the resulting estimates with those from other approaches.

VII. Empirical Results

In trying to assess the practical usefulness of the methods developed earlier in the paper, our empirical analysis primarily addresses three questions. First, do simple estimation procedures based on model-free estimates of the state variables provide reasonably good parameter estimates? Second, do state variables extracted using the standard inversion method resemble their model-independent estimates? Finally, to what extent are model deficiencies more easily diagnosed in a system in which each state variable has a clear economic interpretation?

A. A_0(3) Results

As described above, we estimate the $A_0(3)$ model using three different techniques. A problem with interpreting the resulting drift parameter estimates is that r, μ_1, and μ_2 are highly collinear. In particular, the vast majority of the variation in μ_2 is explained by r and μ_1. We therefore report parameter estimates based on an orthogonalization of μ_2. This is obtained by running the regression

$$\mu_2(t) = \alpha + \beta_0 r(t) + \beta_1 \mu_1(t) + e(t) \tag{54}$$

for a given set of state variables. We then define $\mu^*{}_2(t)$ as the residual of that regression or the component of μ_2 that is orthogonal to the other state variables.[33]

Using this orthogonalization, the drift of μ_2 (not μ_2^*) can be re-expressed as

$$\gamma^* + \kappa_0^* r(t) + \kappa_1^* \mu_1(t) + \kappa_2 \mu_2^*(t), \tag{55}$$

where $\gamma^* = \gamma + \kappa_2 \hat{\alpha}$, $\kappa_0^* = \kappa_0 + \kappa_2 \hat{\beta}_0$, and $\kappa_1^* = \kappa_1 + \kappa_2 \hat{\beta}_1$. The P drifts of r, μ_1,

[33] An alternative would be to express the model in terms of r, μ_1, and $2\ \mu_1 + \mu_2$. Clearly, this state vector is also observable, and the degree of collinearity is much lower.

and μ_2 can be rewritten similarly, so that

$$\mathrm{E}\left[dX(t)\right]/dt = \begin{bmatrix} a_0^* \\ a_1^* \\ a_2^* \end{bmatrix} + \begin{bmatrix} b_{00}^* & b_{01}^* & b_{02} \\ b_{10}^* & b_{11}^* & b_{12} \\ b_{20}^* & b_{21}^* & b_{22} \end{bmatrix} \begin{bmatrix} r(t) \\ \mu_1(t) \\ \mu_2^*(t) \end{bmatrix}. \tag{56}$$

When the model is reasonably well specified, one of the main advantages of our rotation is the ease with which we can compute reasonable starting values for a full-blown likelihood maximization. Specifically, parameter values estimated using model-free r, μ_1, and μ_2 may be used as starting values for the inversion method. For the estimations using model-free r, μ_1, and μ_2, few starting values are required, as most parameters can be estimated via regression. The exceptions are the parameters of the risk-neutral drift of $X(t)$. As we demonstrate in Section V, however, rough estimates of these parameters may be obtained by regressing a model-free measure of $\frac{\partial^3 Y(t)(\tau)}{\partial \tau^3}|_{\tau=0}$ on the three state variables. Doing so results in the values $\gamma* = -0.0598$, $\kappa_0^* = 0.8085$, $\kappa_1^* = 4.9132$, and $\kappa_2 = -1.7587$. All of these are reasonably close to our final estimates.

Table II reports parameter estimates using this orthogonalization for the three different estimation methods. The table also contains estimates obtained by applying Method III to two halves of the sample separately, which we discuss below. Standard errors for all columns except column I are computed using the outer product of the gradients estimator. Standard errors for column I are somewhat ad hoc, though they are generally similar to those in column II, which are also based on model-free r, μ_1, and μ_2.[34] In some cases, standard errors are much smaller for the estimates based on model-free state variables. This naturally reflects the (perceived) gain in information that results from knowing the exact values of all state variables.

The overall similarity of the three sets of estimates in columns I to III is immediately apparent. In particular, the parameters of the risk-neutral dynamics are extremely close. While some P drift parameters are noticeably different, all estimation methods result in the same sign for each parameter and with similar magnitudes.

The fact that estimates based on model-free and inverted state variables are generally close must indicate that the state variables are themselves similar. Figure 2 confirms this conjecture for each state variable, including both μ_2 and μ_2^*. In each panel, we plot the model-free state variable along with the corresponding inverted variable. The two are in all cases extremely similar, with correlations of 0.95 and above. In particular, the two time series of instantaneous short rates agree nearly perfectly, with a correlation of 0.999.

[34] Parameters of the P drift of $X(t)$ are estimated via OLS regression, so OLS standard errors are used for these parameters. The standard errors for each standard deviation parameter is approximated by applying the chi-squared distribution to the inverse of the sum of squared errors. We use Fisher's z transformation to approximate standard errors for the three correlation coefficients. Finally, standard errors of the risk-neutral drift parameters are computed using an outer product of gradients estimator applied to (46).

Table II

$A_0(3)$ Parameter Estimates

This table contains parameter estimates and asymptotic standard errors (in parentheses) from the 1988 to 2005 full sample and the two subsamples ((1988 to 1996) and (1997 to 2005)). Parameters are defined in equations (29), (44), and (45), where an asterisk denotes a parameter that has been orthogonalized as described in (55). The state variables r, μ_1, and μ_2 are defined in Table I. Five sets of estimates are reported:

I Parameters estimated over the full sample using a regression-based approach, treating the model-free r, μ_1, and μ_2 as actual.
II Parameters estimated over the full sample using MLE, treating the model-free r, μ_1, and μ_2 as actual.
III Parameters estimated over the full sample using MLE by inverting three bootstrapped zero coupon yields (3-month, 2-year, and 10-year).
III-HI Identical to III but estimated using only the first half of the sample (1988 to 1996).
III-H2 Identical to III but estimated using only the second half of the sample (1997 to 2006).

	I	II	III	III-H1	III-H2		I	II	III	III-H1	III-H2
γ^*	−0.0784	−0.0784	−0.1027	−0.1662	−0.0974	a_0^*	−0.0061	−0.0067	−0.0056	−0.0058	−0.0050
	(0.0001)	(0.0001)	(0.0008)	(0.0021)	(0.0012)		(0.0053)	(0.0091)	(0.0074)	(0.0165)	(0.0079)
κ_0^*	1.1529	1.1529	1.1181	1.6891	1.5376	a_1^*	0.0375	0.0395	0.0361	0.0704	0.0319
	(0.0009)	(0.0011)	(0.0089)	(0.0213)	(0.0191)		(0.0216)	(0.0264)	(0.0199)	(0.0435)	(0.0224)
κ_1^*	4.8096	4.8096	4.2511	5.3723	3.7070	a_2^*	−0.0865	−0.0922	−0.0811	−0.1758	−0.0682
	(0.0040)	(0.0054)	(0.0337)	(0.0676)	(0.0458)		(0.0492)	(0.0592)	(0.0425)	(0.1016)	(0.0447)
κ_2	−2.3611	−2.3611	−2.3836	−2.6545	−2.2686	b_{00}^*	0.0408	0.0520	0.0404	0.0641	−0.0039
	(0.0021)	(0.0028)	(0.0191)	(0.0339)	(0.0290)		(0.0946)	(0.1332)	(0.1094)	(0.2244)	(0.1488)
σ_0	0.0096	0.0094	0.0085	0.0104	0.0063	b_{01}^*	1.0410	1.0728	0.8503	0.7767	0.9675
	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0001)		(0.1183)	(0.1088)	(0.1143)	(0.1737)	(0.2286)
σ_1	0.0379	0.0379	0.0357	0.0418	0.0312	b_{02}	−0.6177	−0.6377	−0.2069	−0.2148	−0.6728
	(0.0009)	(0.0008)	(0.0009)	(0.0015)	(0.0011)		(0.2690)	(0.3302)	(0.2036)	(0.4530)	(0.5943)
σ_2	0.0862	0.0859	0.0772	0.1011	0.0631	b_{10}^*	−0.6407	−0.5594	−0.4789	−1.0089	−0.4536
	(0.0020)	(0.0016)	(0.0022)	(0.0044)	(0.0027)		(0.3848)	(0.4234)	(0.3462)	(0.6679)	(0.5493)
ρ_{01}	−0.5528	−0.5733	−0.4688	−0.4507	−0.5516	b_{11}^*	−2.2199	−2.2388	−2.1568	−2.3696	−2.1892
	(0.0227)	(0.0150)	(0.0201)	(0.0326)	(0.0276)		(0.4813)	(0.4501)	(0.4619)	(0.6693)	(0.6820)
ρ_{02}	0.5134	0.5393	0.4127	0.3939	0.5048	b_{12}	0.9587	1.1886	0.4616	0.6137	0.1060
	(0.0241)	(0.0142)	(0.0207)	(0.0354)	(0.0258)		(1.0941)	(1.1983)	(0.7930)	(1.3293)	(1.6110)
ρ_{12}	−0.9945	−0.9944	−0.9892	−0.9908	−0.9831	b_{20}^*	1.4900	1.3455	1.1161	2.5412	1.0431
	(0.0004)	(0.0003)	(0.0007)	(0.0009)	(0.0015)		(0.8761)	(0.9477)	(0.7459)	(1.5523)	(1.1146)
						b_{21}^*	4.4351	4.4732	4.1021	5.2214	3.7419
							(1.0956)	(1.0124)	(0.9909)	(1.6094)	(1.3246)
						b_{22}	−2.4097	−2.8909	−1.1901	−1.8564	−0.6329
							(2.4909)	(2.7111)	(1.6959)	(3.1928)	(3.1857)

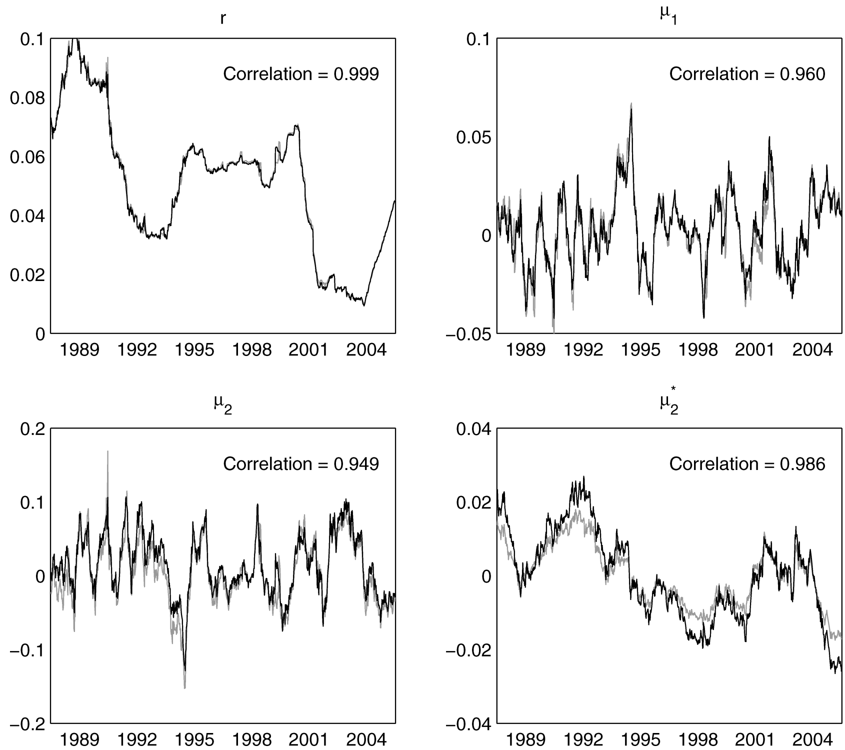

Figure 2. Model-free versus inverted state variables for the $A_0(3)$ model. Each panel compares the time series of one model-free state variable with an alternative that has been inverted from the 3-month, 2-year, and 10-year yields using the parameter values in column III of Table II. Model-free state variables are gray lines, while inverted variables are black lines.

Given the arguably impressive performance of even the simplest estimation method, a reasonable question to ask is whether there is any potential gain in using the slightly more involved inversion approach. Table III, which reports root mean squared errors (RMSE) for all yields, confirms that there is. For columns I to III, which correspond to the same estimates in Table II, model-implied zero coupon yields are calculated via (42). Consistent with the way they are estimated, columns I and II use model-free values of the state vector $X(t)$, while column III inverts $X(t)$ from the 3-month, 2-year, and 10-year yields. Because we believe that inverted $X(t)$ are, by construction, likely to be more accurate in explaining longer-term yields, we also use parameters I and II, which are estimated using model-free $X(t)$, to invert $X(t)$ for the purpose of computing yield fits. These results are in the last two columns of the table.[35] Finally, we report the results for all maturities, not only those used in the estimation.

[35] We also compute yield fits by combining the inversion-based parameters with model-free state variables. These fits are the poorest of all.

Table III

$A_0(3)$ Yield Errors

This table contains root mean squared errors computed from yields over the 1988 to 2005 sample for the $A_0(3)$ model. For each of the three sets of parameters described in columns I to III of Table II, model-implied yields are computed as

$$Y(t,\tau) = -\frac{A(\tau)}{\tau} + \frac{B(\tau)}{\tau} X(t)$$

and errors are defined as actual minus model-implied yields. Consistent with the three estimation methods, columns I and II use model-free estimates of the state vector $X(t)$ while column III inverts $X(t)$ from the 3-month, 2-year, and 10-year yields. In addition, we report yield fits that combine parameter estimates I and II with inverted $X(t)$. These are reported in columns I* and II*. Entries with a dash denote maturities with errors that are zero by construction.

	I	II	III	I*	II*
1-month	6.10	6.10	14.16	17.65	17.64
3-month	5.43	5.43	–	–	–
6-month	5.26	5.26	6.72	10.39	10.37
9-month	5.40	5.40	9.78	17.56	17.53
1-year	5.26	5.26	10.94	19.22	19.19
2-year	23.91	23.85	–	–	–
3-year	33.69	33.57	3.89	12.68	12.61
4-year	37.14	36.96	5.41	20.24	20.11
5-year	36.38	36.15	6.22	23.55	23.37
7-year	26.80	26.55	4.78	20.28	20.05
10-year	9.24	9.24	–	–	–

The results are easily summarized. In comparing the RMSEs, it is clear that inverted state variables are superior at explaining longer maturity yields, while model-free state variables are best at explaining short maturity yields. This result is not surprising given that the model-free variables are explicitly constructed to describe the shape of the term structure at near-zero maturities. In contrast, the inverted state variables are computed to provide the best fit over a much wider range of maturities, and this allows a much closer fit at the long end of the yield curve, where the model-free state variables perform poorly.

Among fits based on inverted $X(t)$ (in columns III, I*, and II*), it is also clear (and unsurprising) that column III, which uses parameters estimated via the inversion method, contains the best fit, with RMSEs between 4 and 14 basis points. In contrast, columns I* and II*, which are based on parameters estimated with model-free $X(t)$, produce RMSEs between 10 and 24 basis points.

While it is clear that each set of state variables has difficulty in explaining some range of maturities, it is important to realize that the errors analyzed in Table III are in some respects quite small. For instance, regressions of actual yields on model-implied yields result in slope coefficients that are *in every case* between 0.98 and 1.03, and the R-squares of all these regressions are *at least* 0.995. Thus, while the errors implied by each set of estimates may be large enough to motivate further improvements, all of the estimates are successful in explaining the vast majority of the variation in yields.

An additional benefit of using observable state variables is the stability of the parameter estimates and inverted state variables. We illustrate this by re-estimating the model using different sample periods. Inversion-based estimates that use data from the first half of our sample (1988 to 1996) and estimates that use data from the second half (1997 to 2005) appear in columns III-H1 and III-H2 of Table II. Each of these sets of estimates is then used to invert time series of r, μ_1, and μ_2 for the entire sample (1988 to 2005). Figure 3, which plots both sets of time series in the three left-hand panels, shows that the state variables are close to identical, a striking result given that they are based on parameters estimated over completely nonoverlapping sample periods. The result is consistent with the idea that our state variables, both in theory and in practice, have interpretations that are essentially independent of the parameters of the model.

When we look at Table II, it is clear that the full-sample and two half-sample estimates are generally similar. While there are differences, these differences reflect changing characteristics of the data rather than changing interpretations of the state variables. The state variable volatility parameters σ_0, σ_1, and σ_2, for example, are all higher for the earlier sample than they are for the later sample. This implies that yields were more volatile in the earlier period, which in fact is true. Such a straightforward inference, we emphasize, is impossible in a latent factor framework.

To understand why, we note that there are six sets of parameters under the canonical rotation of Dai and Singleton (2000) that are each exactly equivalent to the parameters in column III-H1 of Table II.[36] Another six parameter vectors are equivalent to column III-H2. Thus, if the researcher were to re-estimate the model using different sample periods, it is entirely possible that the optimization might result in a jump from one maximum of the likelihood function to another. Such a jump would completely redefine the interpretation of each state variable, making comparisons of the state variable volatility parameters meaningless. We note that this problem is not merely confined to comparisons across sample periods, but that it also affects samples that differ by country, sampling frequency, bond maturities, or usage of derivatives.

As an example of this phenomenon, we invert state variables under the DS rotation using one of the six sets of parameters that are equivalent to the estimates in column III-H1 of Table II and another that is equivalent to the estimates in column III-H2. A representative set of these results is plotted in the right three panels of Figure III. It is clear from the figure that, in some cases, the two sets of state variables bear little relation to one another (even though they both generate identical likelihood and bond prices). Comparing parameter values estimated from these two samples therefore makes little sense.

[36] The DS rotation of the $A_0(3)$ model is written as $dX = -\mathcal{K}X(t) + dB(t)$, where $\mathcal{K}$ is lower triangular with positive diagonal elements and $B(t)$ is a vector of standard Brownian motions. The short rate is then specified as $r(t) = \delta_0 + \delta_X^\top X(t)$, where δ_X has all positive elements. Since μ_1 and μ_2 are also simple linear combinations of $X(t)$, it is easy to deduce the implied dynamics of $\{r, \mu_1, \mu_2\}$ for a given combination of $\mathcal{K}$, δ_0, and δ_X.

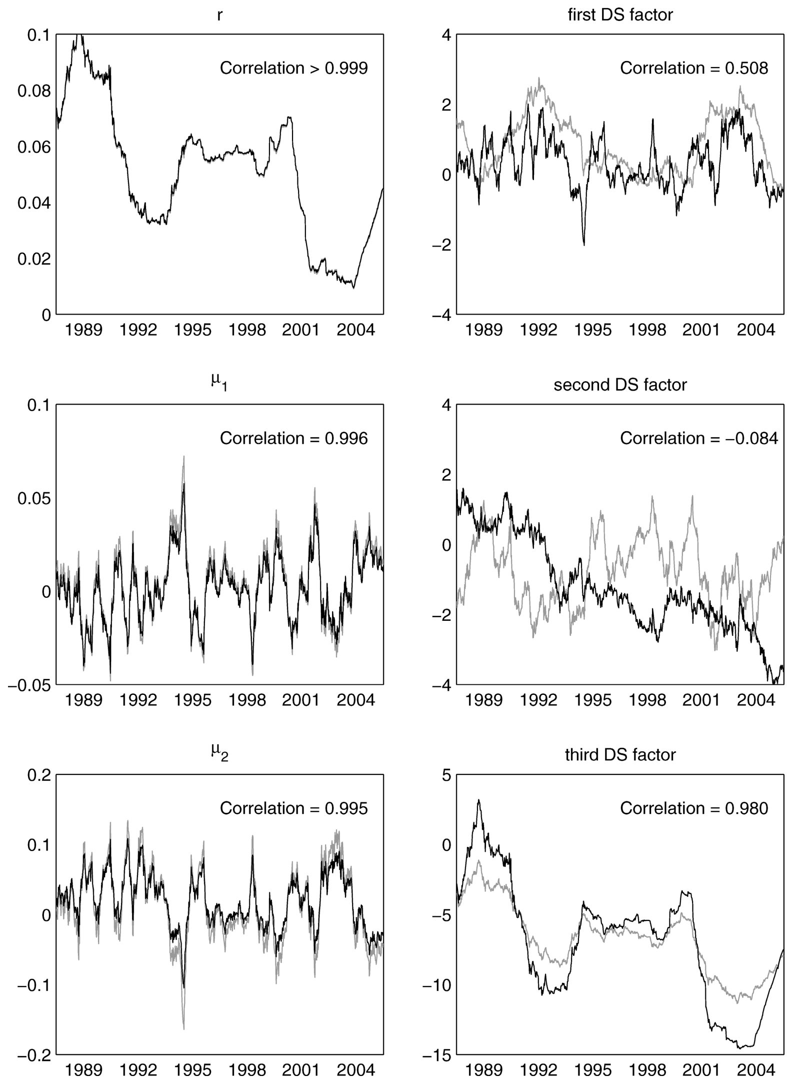

Figure 3. $A_0(3)$ state variables inverted using parameters from nonoverlapping sample periods. Gray lines denote state variables that are inverted using parameter values that are estimated using the 1988 to 1996 sample period. Black lines are constructed identically but use parameters estimated over 1997 to 2005. The three panels on the left correspond to our canonical rotation of the $A_0(3)$ model. Panels on the right correspond to state variables from the DS rotation. Note that each set of DS state variables corresponds to a nonunique set of parameter estimates and thus the results for DS are only representative.

Overall, we find a number of practical advantages in using state variables that have a clear economic interpretation and that can be reliably inferred from term structure data without the use of a particular model. Our model-free state variables appear to be accurately estimated, and they produce excellent starting values for full-blown maximum likelihood estimation. While yield fits using model-free state variables are more accurate for short maturities, they are less so for longer maturities. Possible explanations of this result include measurement errors in the model-free estimates and the existence of an idiosyncratic factor that only affects short-maturity LIBOR rates. Another possibility is that the $A_0(3)$ model is somewhat misspecified, so that factors extracted from the long and short ends of the term structure do not coincide. In this case, the inversion method naturally fits the long end better since it matches, exactly by construction, the 10-year yield.

B. $A_1(3)$ Results

Given that we have written the $A_1(3)$ model in terms of the short-rate variance V (in addition to the short rate r and its risk-neutral drift μ_1), the simplest estimation approaches involve using an observed proxy for V. Many possible proxies might be entertained, including GARCH estimates from daily data, so-called realized variances computed from high frequency intraday data, and implied variances from options. None of these proxies is likely to be as accurate as our model-free $\{\mu\}$ variables. GARCH variances, for instance, are filtered estimates that effectively smooth out squared innovations over many lags. Thus, these estimates are unlikely to represent the instantaneous variance denoted by $V(t)$. Since realized variances are computed over much shorter intervals, they avoid this issue to a great degree, but they can be sensitive to measurement error and may tend to capture a transient component that is less important for the pricing of most fixed income securities. For these reasons, we choose to proxy for V using an implied variance from Eurodollar futures options. As discussed previously in Section IV, the implied variance measure we are able to construct is imperfect at best, mostly because there is often no option with a sufficiently short time until expiration. As such, we view the results that are based on this proxy as somewhat exploratory.[37]

Two of the four estimation methods we propose for this rotation in Section VI.B assume observable r, μ_1, and V. The first is an approximate two-step method in which the P parameters are estimated using a discrete-time heteroskedastic VAR and the risk-neutral drift parameters are estimated conditional on the first step. The second is a standard QML approach given that the state variables are assumed to be observed. A third estimation method treats V as unobserved and computes it as a linear combination of model-free r, μ_1, and μ_2. The last one is the standard inversion method.

Table IV reports the estimation results. In contrast to the $A_0(3)$ case, the four different methods in some cases result in very different estimated parameters. Columns I and IIa, which are both based on the implied variance proxy for V, are in substantial agreement. Columns IIb and III, which do not use the implied variance series, are also generally similar. Very large differences, however, are

[37] Bikbov and Chernov (2005) provide an extensive study of the relation between volatility and yields implied from Eurodollar futures option prices.

Table IV

$A_1(3)$ Parameter Estimates

This table contains parameter estimates and asymptotic standard errors (in parenthesis) from the 1991 to 2005 sample. Parameter values are defined in (33), (34), and (48). The variable V denotes the instantaneous variance of the short rate, while other state variables are defined in Table I. Three sets of estimates are reported:

I Parameters estimated using naive discretization, treating implied V and model-free r and μ_1 as actual.
IIa Parameters estimated using QML, treating implied V and model-free r and μ_1 as actual.
IIb Parameters estimated using QML, treating model-free r,μ_1, and μ_2 as actual.
III Parameters estimated using QML by inverting three bootstrapped zero coupon yields (3-month, 2-year, and 10-year).

	I	IIa	IIb	III		I	IIa	IIb	III
$a_V^Q \times 10^5$	2.145	2.243	1.833	1.832	$a_V \times 10^4$	2.397	2.089	0.113	0.120
	(0.180)	(0.195)	(0.063)	(0.074)		(0.361)	(0.410)	(0.107)	(0.109)
b_{VV}^Q	0.000	0.000	0.360	0.523	b_{VV}	0.007	0.007	0.127	0.036
	(N/A)	*(N/A)*	(0.002)	(0.006)		(0.005)	(0.005)	(0.038)	(0.017)
a_μ^Q	0.039	0.039	−0.096	−0.032	a_r	0.023	0.020	−0.122	−0.028
	(0.001)	(0.001)	(0.020)	(0.005)		(0.019)	(0.019)	(0.135)	(0.054)
$b_{\mu V}^Q$	180.224	181.346	1974.271	1094.626	b_{RV}	−5.880	−4.292	−0.223	−0.285
	(5.355)	(6.045)	(383.049)	(116.260)		(1.235)	(1.194)	(0.207)	(0.232)
$b_{\mu r}^Q$	−0.733	−0.733	0.042	−0.064	b_{rr}	−194.779	−209.415	−1872.613	−537.017
	(0.015)	(0.016)	(0.002)	(0.001)		(65.546)	(68.310)	(613.400)	(221.025)
$b_{\mu\mu}^Q$	−2.544	−2.544	−1.993	−1.732	$b_{r\mu}$	−0.091	−0.081	−0.734	−0.329
	(0.023)	(0.024)	(0.003)	(0.022)		(0.089)	(0.088)	(0.195)	(0.165)
$V \times 10^5$	0.128	0.128	2.230	0.440	a_μ	0.865	0.841	0.664	0.674
	(0.056)	(0.108)	(0.790)	(0.740)		(0.086)	(0.085)	(0.113)	(0.098)
$\sigma_V^V \times 10^4$	3.699	4.075	0.012	0.037	$b_{\mu V}$	255.709	374.602	1966.431	601.661
	(0.115)	(0.248)	(0.003)	(0.006)		(348.841)	(358.206)	(2057.559)	(797.411)
$\sigma_\mu^0 \times 10^4$	4.440	4.379	1.472	1.949	$b_{\mu r}$	−0.614	−0.554	0.654	0.244
	(0.637)	(0.718)	(1.358)	(1.085)		(0.409)	(0.386)	(0.716)	(0.468)
σ_V^μ	16.249	16.563	32.833	19.469	$b_{\mu\mu}$	−1.545	−1.240	−1.835	−1.655
	(2.414)	(2.448)	(7.354)	(3.005)		(0.468)	(0.450)	(0.429)	(0.386)
$c_V^V r \times 10^3$	0.626	1.255	0.249	0.304					
	(0.736)	(0.693)	(0.044)	(0.073)					
$c_{V\mu}^V \times 10^3$	2.149	2.032	−4.754	−6.453					
	(3.558)	(3.355)	(0.725)	(0.981)					
$c_{r\mu}^0 \times 10^5$	−2.385	−2.369	−5.728	−2.929					
	(4.731)	(5.648)	(3.260)	(2.251)					
$a_{r\mu}^V$	−2.040	−2.150	−2.296	−1.790					
	(0.224)	(0.227)	(0.906)	(0.425)					

Table V

$A_1(3)$ Yield Errors

This table contains root mean squared errors computed from yields over the 1991 to 2005 sample for the $A_1(3)$ model. For parameter vectors IIa, IIb, and III from Table IV, model-implied yields are computed as

$$Y(t,\tau) = -\frac{A(\tau)}{\tau} + \frac{B(\tau)}{\tau}X(t)$$

and errors are defined as actual minus model-implied yields. Consistent with the three estimation methods, column IIa uses an implied variance proxy for V and model-free estimates of r and μ_1. Column IIb uses model-free estimates of r, μ_1, and μ_2, while column III inverts the state vector X from the 3-month, 2-year, and 10-year yields. In addition, we report yield fits that combine parameter estimates IIa and IIb with inverted X. These are reported in columns IIa* and IIb*. Entries with a dash denote maturities with errors that are zero by construction.

	IIa	IIb	III	IIa*	*IIb**
1-month	6.04	6.02	13.74	17.51	16.96
3-month	5.37	5.40	–	–	–
6-month	5.25	5.35	6.67	12.33	10.10
9-month	6.21	5.19	9.13	20.51	16.32
1-year	7.76	5.07	9.53	23.64	17.22
2-year	29.07	22.96	–	–	–
3-year	42.30	32.27	3.73	27.87	12.13
4-year	50.46	35.35	5.09	31.89	19.20
5-year	56.39	34.77	5.64	35.38	22.54
7-year	63.72	24.73	4.73	41.98	18.65
10-year	72.17	8.85	–	–	–

observed between these two pairs of estimates, especially for the model-implied variance dynamics. Surprisingly, under the risk-neutral measure, V has zero mean reversion ($b^Q_{VV} \approx 0$) for estimates I and IIa but substantial mean reversion for IIb and III.[38] Through their higher values for $b^Q{}_{\mu V}$, estimates IIb and III also imply a term structure that is much more sensitive to changes in V, yet these parameters also suggest (through small values of σ^V_V) that V is itself not very volatile. Many other differences are notable, but most of them reinforce the idea that volatility dynamics look very different depending on whether or not implied variance is used in the estimation.

To investigate this issue further, Figure 4 plots the time series of the square root of V corresponding to each estimation method. Alongside each time series is another model-free estimate of short-rate volatility, a simple standard deviation calculated by applying a 30-day centered rolling window to daily changes in 3-month LIBOR rates.

The top panel plots the implied volatility from Eurodollar options used in estimates I and IIa against the rolling LIBOR volatility. The correlation between the two time series is 0.65, and the two are also similar in terms of level and

[38] Our estimates are constrained to be stationary under both measures, but the constraint binds for estimates I and IIa. To compute standard errors for these estimates, we simply treat b^Q_{VV} as being fixed at zero.

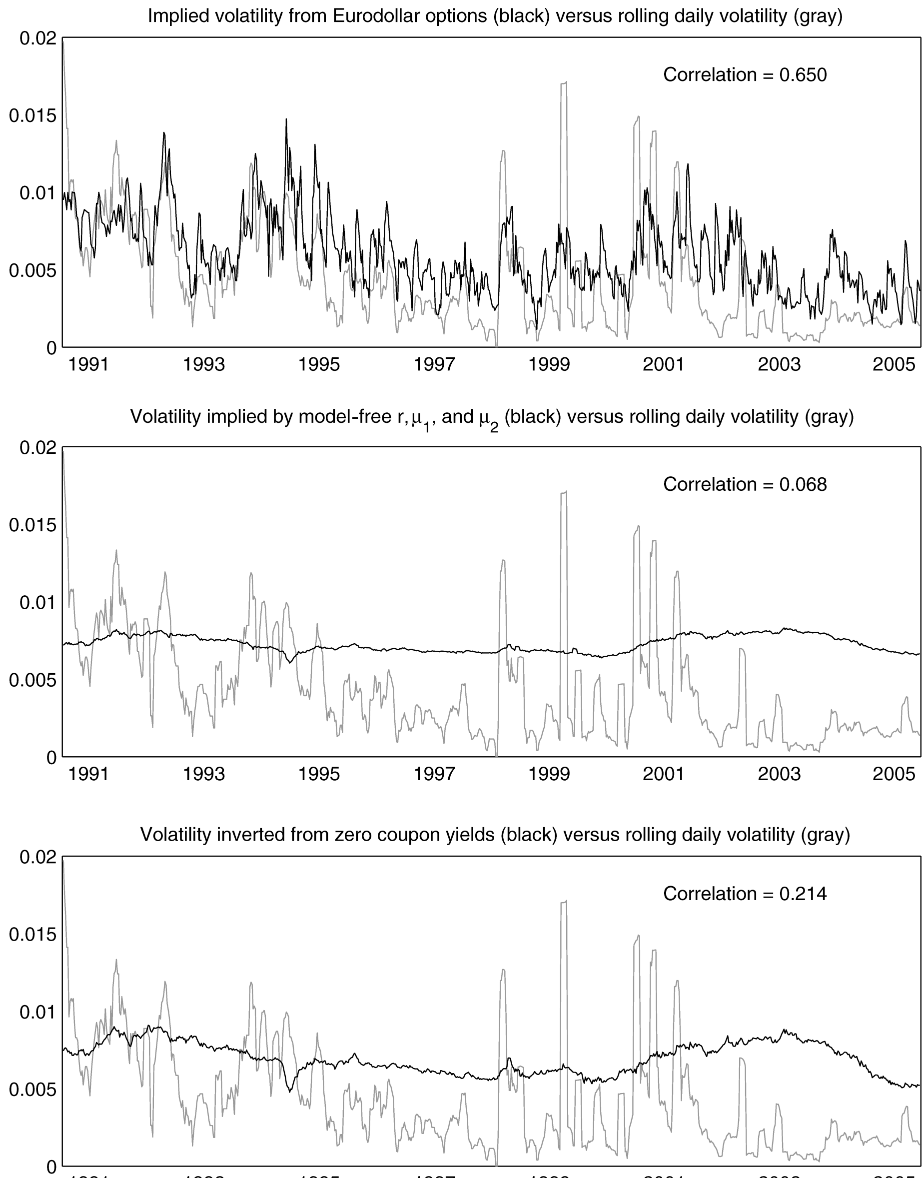

Figure 4. Estimated and implied short rate volatilities for the $A_1(3)$ model. Each panel contains the measure of volatility assumed or implied by each estimation method (the black lines) along with a rolling volatility estimate computed from daily 3-month LIBOR rates using a centered 30-day moving window (the gray lines). In the top panel, the black line denotes an implied volatility from Eurodollar options. The black line in the middle panel denotes a volatility that is inferred from model-free r, μ_1, and μ_2, while in the bottom panel it is a volatility measure inverted from the 3-month, 2-year, and 10-year yields.

degree of variation. In the middle panel, we have the short-rate volatility that is implied by model-free estimates of r, μ_1, and μ_2. What is most striking here is the near constancy of the volatility process. Matters improve slightly for the inversion approach, displayed in the bottom panel, but the volatility process is still just barely fluctuating and only marginally related to our ex post measure.

It is not clear, however, that using the implied volatility measure leads to an overall improvement in fit. Table V contains root mean squared errors to gauge the ability of each set of parameter estimates to fit the cross-section of yields.[39] As in Table III, we first compute model-implied yields in a manner consistent with the estimation of each set of parameters, so that parameters estimated using implied variance and model-free r and μ_1, for example, use those proxies in computing yields via (42). We then compute yield fits by using parameters estimated from model-free proxies for $X(t)$ to invert those state variables instead and use the inverted values to compute model-implied yields.

In short, the results suggest that using an implied variance proxy for $V(t)$ does not lead to an adequate fit of the cross-section of yields. In column IIa we see that implied volatility-based estimates generate the largest RMSEs. In column IIa*, we use the same parameters with inverted state variables and find results that are only marginally improved. Thus, our results suggest that as long as our implied variance measure at least roughly matches actual short-rate variance, it is unlikely that a variance factor can help span the three factors identified in most term structure analysis.

In comparing yield fits based on model-free r, μ_1, and μ_2 to those based on inverted $X(t)$, we find results very similar to those for the $A_0(3)$ model. Specifically, model-free state variables are best at explaining short-term yields, while inverted state variables provide a better fit over the yield curve as a whole. Here, however, we believe that there is much stronger evidence that the main advantage of the inversion approach is its ability to compensate for clear model misspecifications. In particular, the inversion method allows the state variable V to adopt a dynamic that is inconsistent with its definition as the variance of the short rate. While this inconsistency would be implicit in any representation of the $A_1(3)$ model, in our rotation it is very hard to miss, which is a benefit of our approach.

Finally, in an attempt to diagnose why our use of implied variance is so ineffective in explaining yield curve variation, we return to analyzing the relations between model-implied and model-free state variables. As we found for the $A_0(3)$ model, inverted time series of r and μ_1 match the model-free estimates precisely, with correlation coefficients of 0.96 or above, so we do not examine them further. Instead, Figure 5 focuses on μ_2 and μ_2^*, the latter of which is orthogonalized to r and μ_1 (but not V) using the regression approach of Section VII.A.

The top panels of Figure 5 plot measures of μ_2 and μ_2^* that are implied by model-free r and μ_1 and the implied variance proxy for V. Since μ_2 is, by

[39] For brevity, we drop estimate I because it is very similar to IIa.

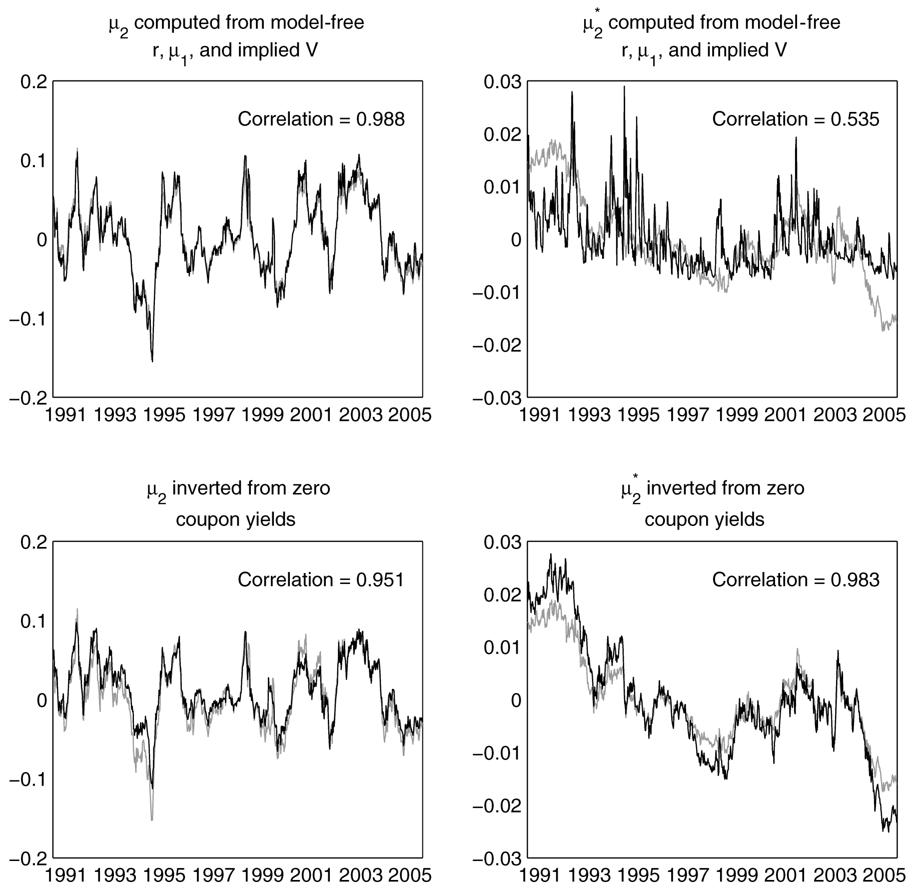

Figure 5. Estimated versus model-free μ_2 and μ_2^* for the $A_1(3)$ model. Each panel displays a time series of μ_2 or μ_2^* (the black lines) that is estimated using the $A_1(3)$ model against model-free estimates (the gray lines) of the same series. The top two panels compare model-free estimates of μ_2 and μ_2^* with series that are computed as linear combinations of model-free r and μ_1 and the implied variance proxy for V. The bottom panels compare model-free μ_2 and μ_2^* with series that are inverted from, 3-month, 2-year, and 10-year yields using the parameter values in column III of Table IV.

definition, the risk-neutral drift of μ_1, it can be computed from these series following (33) as

$$b_{\mu r}^{Q} r(t) + b_{\mu\mu}^{Q} \mu_1(t) + b_{\mu V}^{Q} V(t).$$

The parameter values used are from column IIa of Table IV because these estimates are obtained by treating the same series as observable. The bottom panels of Figure 5 compute μ_2 similarly, but there we use the inversion-based parameters from column III of Table IV along with inverted values of r, μ_1, and V.

Surprisingly, the top left panel suggests that observable proxies for r, μ_1, and V do an adequate job in fitting the time series of μ_2. Unfortunately, the goodness of this fit is illusory, as it merely reflects the high degree of correlation between model-free μ_1 and μ_2. In the top right panel, which displays the orthogonalized series μ_2^*, we find a much worse fit. Because models that fit μ_2^* also fit yields reasonably well, the poor fit reveals the source of the large RMSEs in Table V: Implied volatilities cannot explain yields because they do not adequately explain independent variation in μ_2.

In contrast, the bottom panels show that inversion-based time series of μ_2 and μ_2^* match quite closely with model-free estimates. Given the poor performance of inversion-based estimates in explaining volatility, their greater success in matching μ_2 is perhaps not surprising.

Although our results are based on a somewhat simplistic empirical approach, the fact that they are so sensitive to the choice of μ_1 versus V indicates that the model is likely misspecified. A companion paper, Collin-Dufresne et al. (2007), provides further evidence on this issue using a more robust econometric approach.

VIII. Conclusion

Typically, affine models of the term structure are written in terms of a latent state variable whose components have no economic meaning independent of the model and the model's parameters. Often, this leads to representations that are not globally identifiable. Outside of the goodness-of-fit estimates, values for the state vector and parameter vector, taken individually, are often meaningless.

To circumvent these concerns, we propose a representation in which the state vector is written in terms of theoretically observable state variables that have unambiguous economic interpretations. As such, global identification of the model is guaranteed. Further, the representation naturally leads to a canonical representation that is more flexible (in that there are more free parameters) than that identified by Dai and Singleton (2000). In addition, we suspect that by rotating to economically meaningful variables, the likelihood function becomes "steeper," since changes in parameter values can no longer be offset by changes in the state variables, whose economic interpretations are now fixed. Thus, we suspect that, starting at the same initial first guess, the search for the maximal likelihood parameter vector will generally be faster for our proposed specifications than for latent specifications.

We demonstrate using simulations and actual data that our state variables can be estimated independent of the model extremely well. This provides a simple method for obtaining good first-guess values for the parameter vector of a complex model, which is useful for the computationally burdensome estimation methods often employed in this literature. Furthermore, being able to compute model-free state variables allows for a comparison with the state variables obtained from inverting yields, and our results suggest that deviations between the two can be informative about the form of any model misspecification. We study one Gaussian model and one stochastic volatility model. For the Gaussian model, we find that model-implied state variables and their model-free proxies are close matches.[40] For the stochastic volatility case, we find significant discrepancies between model-implied and observable state variables, which suggests to us a tension in fitting level, slope, curvature, and volatility within affine three-factor models. We investigate this question further in a companion paper (Collin-Dufresne et al. (2007)). Finally, we provide evidence that observable state variables, even when estimated using the model- and parameter-dependent inversion method, are often not very sensitive to the model or the parameter vector. This facilitates the comparison of parameter values obtained from different countries or sample periods, as they can be made directly without having to correct for differing economic interpretations of the underlying state variables.

Appendix: Proof of Generality of Equations (21)–(24)

Consider a Markov state vector $\{X(t)\}$ of length N with general (i.e., non-affine) risk-neutral dynamics

$$dX_i = m_i^Q\left(\{X\}\right)\,dt + \sum_{k=1}^{N}\sigma_{ik}\left(\{X\}\right)\,dz_k^Q. \tag{A1}$$

Further, assume the spot rate is some arbitrary function of the state vector $r = r\left(\{X\}\right)$. Using the shorthand notations $m_i^Q = m_i^Q(\{X\})$ and $\sigma_{ik} = \sigma_{ik}(\{X\})$, we obtain from Ito's lemma the dynamics for $r(t)$:

$$dr = \sum_{i=1}^{N}\frac{\partial r}{\partial X_i}\left[m_i^Q\,dt + \sum_{k=1}^{N}\sigma_{ik}\,dz_k^Q\right] + \frac{1}{2}\sum_{i,j,k=1}^{N}\frac{\partial^2 r}{\partial X_i \partial X_j}\sigma_{ik}\,\sigma_{jk}\,dt. \tag{A2}$$

Note that this allows us to define the drift and the variance of the spot rate as

$$\begin{aligned}\mu_1(t) &\equiv \frac{1}{dt}\mathrm{E}_t^Q\left[dr\right]\\ &\equiv \sum_{i=1}^{N}\frac{\partial r}{\partial X_i}\,m_i^Q + \frac{1}{2}\sum_{i,j,k=1}^{N}\frac{\partial^2 r}{\partial X_i \partial X_j}\sigma_{ik}\,\sigma_{jk}\end{aligned} \tag{A3}$$

$$\begin{aligned}V(t) &\equiv \frac{1}{dt}\mathrm{Var}_t^Q\left[dr\right]\\ &\equiv \sum_{i,j,k=1}^{N}\frac{\partial r}{\partial X_i}\,\frac{\partial r}{\partial X_j}\,\sigma_{ik}\,\sigma_{jk}.\end{aligned} \tag{A4}$$

[40] In unreported results (available upon request), we find large discrepancies between model-implied and observable state variables for the two-factor Gaussian model, suggesting that the model is not well specified.

Further, from Ito's lemma we have

$$\begin{aligned}\mu_2(t) &= \frac{1}{dt}\,\mathrm{E}_t^Q\,[d\mu_1(t)] \\ &= \sum_{i=1}^{N} \frac{\partial \mu_1}{\partial X_i}\, m_i^Q + \frac{1}{2} \sum_{i,\,j,\,k=1}^{N} \frac{\partial^2 \mu_1}{\partial X_i \partial X_j}\, \sigma_{ik}\, \sigma_{jk}.\end{aligned} \tag{A5}$$

If we define $\tau \equiv (T - t)$, then the date-t price $P^T\,(t, \{X_t\})$ of a zero coupon bond with maturity T can be written in terms of yield to maturity as

$$P^T(t, \{X_t\}) \equiv e^{-\tau\, Y(\{X_t\},\tau)}. \tag{A6}$$

Using the notation $P_\tau \equiv \frac{\partial P}{\partial \tau}$, $P_i \equiv \frac{\partial P}{\partial X_i}$, etc., we have

$$P_\tau = [-Y - \tau\, Y_\tau]\, P \tag{A7}$$

$$P_i = -\tau\, Y_i\, P \tag{A8}$$

$$P_{ij} = [\tau^2\, Y_i\, Y_j - \tau\, Y_{ij}]\, P. \tag{A9}$$

Bond prices satisfy the PDE

$$rP = -P_\tau + \sum_{i=1}^{N} P_i\, m_i^Q + \frac{1}{2} \sum_{ijk=1}^{N} P_{ij} \sigma_{ik} \sigma_{jk}. \tag{A10}$$

Plugging in equations (A7) to (A9), we find

$$r(t) = \left[Y + \tau\, Y_\tau\right] - \tau \sum_{i=1}^{N} Y_i\, m_i^Q + \frac{1}{2} \sum_{ijk=1}^{N} \left[\tau^2\, Y_i\, Y_j - \tau\, Y_{ij}\right] \sigma_{ik}\, \sigma_{jk}. \tag{A11}$$

Now we use a Taylor series expansion to write yields as

$$Y(\{X_t\}, \tau) \equiv Y^0(\{X_t\}) + \tau\, Y^1(\{X_t\}) + \frac{1}{2}\, \tau^2\, Y^2(\{X_t\}) + \ldots. \tag{A12}$$

Plugging this expansion into equation (A11) and collecting terms of different orders of τ, we find that the general relation between $\{r_j\}$ and $\{Y^h\}$ is given by

$$Y^0 = r \tag{A13}$$

and for $h > 0$

$$Y^h = \frac{h!}{1+h}\left(\sum_{i=1}^{N}\frac{Y_i^{h-1}}{(h-1)!}m_i + \frac{1}{2}\sum_{ijk=1}^{N}\left\{\frac{Y_{ij}^{h-1}}{(h-1)!} - \sum_{\substack{(m,n)>0\\ s.t.\\ m+n+2=h}}\frac{Y_i^m}{m!}\frac{Y_j^n}{n!}\right\}\sigma_{ik}\sigma_{jk}\right)$$

$$r_h(t) = \sum_{i=1}^{N}\frac{\partial r_{h-1}}{\partial X_i}m_i^Q + \frac{1}{2}\sum_{i,j,k=1}^{N}\frac{\partial^2 r_{h-1}}{\partial X_i \partial X_j}\sigma_{ik}\,\sigma_{jk}. \tag{A14}$$

For example, we obtain for the first few terms:

$$\tau^0 : Y^0(\{X_t\}) = r(\{X_t\}) \tag{A15}$$

$$\tau^1 : Y^1(\{X_t\}) = \frac{1}{2}\mu_1(\{X_t\}) \tag{A16}$$

$$\tau^2 : Y^2(\{X_t\}) = \frac{1}{3}\left[\mu_2(t) - V(t)\right] \tag{A17}$$

$$\tau^3 : Y^3(t) = \frac{1}{4}\left\{r_3(t) - \mathrm{E}_t[dV_{00}(t)]/dt - 3V_{01}(t)\right\}. \tag{A18}$$

That is, the level, slope, and curvature of the yield curve at $\tau = 0$ are intimately related to r, μ_1, μ_2, and the variance of the spot rate $(dr)^2$.

REFERENCES

Babbs, Simon H., and K. Ben Nowman, 1999, Kalman filtering of generalized Vasicek term structure models, *Journal of Financial and Quantitative Analysis* 34, 115–130.

Balduzzi, Pierluigi, Sanjiv Das, and Silverio Foresi, 1996, A simple approach to three factor affine term structure models, *Journal of Fixed Income* 6, 43–53.

Bikbov, Ruslan, and Mikhail Chernov 2005, Term structure and volatility: Lessons from the Eurodollar market, Working paper, Columbia University.

Black, Fisher, 1976, The pricing of commodity contracts, *Journal of Financial Economics* 3, 167–179.

Bliss, Robert, 1997, Testing term structure estimation methods, *Advances in Futures and Options Research* 9, 197–231.

Cheridito, Patrick, Damir Filipovic, and Robert L. Kimmel, 2005, A note on the canonical representation of affine diffusion processes, Working paper, Princeton University.

Cheridito, Patrick, Damir Filipovic, and Robert L. Kimmel, 2006, Market price of risk specifications for affine models: Theory and evidence, *Journal of Financial Economics* 83, 123–170.

Collin-Dufresne, Pierre, and Robert S. Goldstein, 2002a, Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility, *Journal of Finance* 57, 1685–1730.

Collin-Dufresne, Pierre, and Robert S. Goldstein, 2002b, Pricing swaptions within an affine framework, *Journal of Derivatives* 10, 9–26.

Collin-Dufresne, Pierre, Robert S. Goldstein, and Christopher S. Jones, 2007, Can interest rate volatility be extracted from the cross section of bond yields? An investigation of unspanned stochastic volatility, Working paper, USC.

Collin-Dufresne, Pierre, and Bruno Solnik, 2001, On the term structure of default premia in the swap and LIBOR markets, *Journal of Finance* 56, 1095–1115.

Cox, John C., Jonathan E. Ingersoll Jr., and Steve A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–407.

Dai, Qiang, and Kenneth J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943–1978.

Dai, Qiang, and Kenneth Singleton, 2002, Expectations puzzle, time-varying risk premia, and affine models of the term structure, *Journal of Financial Economics* 63, 415–441.

Dai, Qiang, and Kenneth Singleton, 2003, Term structure dynamics in theory and reality, *Review of Financial Studies* 16, 631–678.

Duarte, Jefferson, 2004, Evaluating an alternative risk preference in affine term structure models, *Review of Financial Studies* 17, 370–404.

Duffee, Gregory R. 1996, Idiosyncratic variation of treasury bill yields, *Journal of Finance* 51, 527–551.

Duffee, Gregory R., 2002, Term premia and interest rate forecasts in affine models, *Journal of Finance* 57, 405–443.

Duffie, Darrell, 2006, *Dynamic Asset Pricing Theory*, (Princeton University Press, Princeton).

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

Duffie, Darrell, Jun Pan, and Kenneth Singleton, 2000, Transform analysis and option pricing for affine jump-diffusions, *Econometrica* 68, 1343–1376.

Duffie, Darrell, and Kenneth Singleton, 1997, An econometric model of the term structure of interest rate swap yields, *Journal of Finance* 52, 1287–1381.

Fama, Eugene F., and Robert R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680–692.

Fisher, Mark, and Christian Gilles, 1996, Estimating exponential affine models of the term structure, Working paper, Federal Reserve Board, Washington, DC.

Gourieroux, Christian, and Razvan Sufana, 2003, Wishart quadratic term structure models, Working paper, Les Cahiers du Cref.

Harrison, Michael, and Stanley Pliska, 1981, Martingales and stochastic integrals in the theory of continuous trading, *Stochastic processes and their applications* 11, 313–316.

Jegadeesh, Narasimhan, and George G. Pennacchi, 1996, The behavior of interest rates implied by the term structure of eurodollar futures, *Journal of Money, Credit and Banking* 28, 426–446.

Johannes, Michael, and Suresh Sundaresan, 2007, The impact of collateralization on swap rates, *Journal of Finance* 62, 383–410.

Joslin, Scott, 2007, Can unspanned stochastic volatility explain the cross section of bond volatilities, Working paper, Stanford University.

Ledoit, Olivier, Pedro Santa-Clara, and Shu Yan, 2002, Relative pricing of options with stochastic volatility, Working paper, UCLA.

Liptser, R. S., and A. N. Shiryaev, 1974, *Statistics of Random Processes I and II* (Springer-Verlag, Berlin).

Litterman, Robert, and Jose Scheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income* 1, 54–61.

Ljung, Lennart, 1999, *System Identification: Theory for the User* (Prentice Hall, New Jersey).

Longstaff, Francis, and Eduardo S. Schwartz, 1992, Interest rate volatility and the term structure: A two-factor general equilibrium model, *Journal of Finance* 47, 1259–1282.

Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Pan, Jun, 2002, The jump-risk premia implicit in options: Evidence from an integrated time-series study, *Journal of Financial Economics* 63, 3–50.

Piazzesi, Monika, 2006, Affine term structure models, inYacine Aït-Sahalia and Lars Hansen, eds.: *Handbook of Financial Econometrics* (Elsevier, Amsterdam).

Shreve, Steven, 2004, *Stochastic Calculus Models for Finance II: Continuous-Time Models* (Springer-Verlag, Berlin).

Singleton, Kenneth, 2001, Estimation of affine asset pricing models using the empirical characteristic function, *Journal of Econometrics* 102, 111–141.

Singleton, Kenneth J., and Len Umantsev, 2002, Pricing coupon-bond options and swaptions in affine term structure models, *Mathematical Finance* 12, 427–446.

Tang, Huarong, and Yihong Xia, 2005, An international examination of affine term structure models and the expectations hypothesis, Working paper, University of Pennsylvania.

Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177–188.