Estimation of Dynamic Term Structure Models

Gregory R. Duffee and Richard H. Stanton*

Haas School of Business
U.C. Berkeley
545 Student Services Building #1900
Berkeley, CA 94720-1900
tel. (510) 642-1435 (Duffee)
tel. (510) 642-7382 (Stanton)
fax (510) 643-1420
e-mail: duffee@haas.berkeley.edu (Duffee)
e-mail: stanton@haas.berkeley.edu (Stanton)

This Draft: April 24, 2001

ABSTRACT

In this paper we estimate a multifactor term structure model that exhibits three key features: correlated factors, nonlinear dynamics, and a flexible price of interest rate risk. We use this estimation as a test bed to investigate the small sample properties of the most commonly used technique for estimating such flexible term structure models, the Efficient Method of Moments (EMM) of Gallant and Tauchen (1996). We find that the standard combination of EMM with a SNP auxiliary model [Gallant and Tauchen (1992)], despite its good asymptotic properties, behaves extremely poorly in samples of the size and type usual in term structure estimation. Substantially better results are obtained using EMM with an alternative auxiliary model based on an approximate Kalman filter. However, the best results are produced by the approximate Kalman filter alone, despite its theoretical inconsistency.

Our results also reveal substantially biased parameter estimates, regardless of the estimation method. The bias is introduced by the combination of the near-unit-root behavior of interest rates and a more general specification of the price of risk than in Dai and Singleton (2000). It makes the flexible risk specification appear to fit the data better than the Dai-Singleton specification, even when the true model is in the more restrictive class. This bias is likely to become increasingly pervasive as researchers estimate ever more general term structure models.

Finally, our model, while qualitatively possessing the potential to resolve some of the empirical shortcomings of more restrictive term structure models, is unable fully to overcome their deficiencies in matching the observed relationship between bond returns and the slope of the yield curve, nor in forecasting future interest changes. It allows us, however, to shed useful light on the kind of flexibility that must be built into the price of risk in order to fit the empirical features of Treasury bond yield and return dynamics.

JEL classification: C14, C52, G12, G13.

*Both authors U.C. Berkeley. We gratefully acknowledge financial assistance from the Berkeley Program in Finance, from the U.C. Berkeley Committee on Research, and from the Fisher Center for Real Estate and Urban Economics.
1 Introduction

There is now an enormous literature devoted to the development, estimation and testing of dynamic term-structure models. By specifying the behavior of interest rates under both the physical and risk-neutral measures, these models describe the evolution of the entire term structure of interest rates, allowing for a rich set of testable restrictions.

Much of this research has focused on the affine term structure models characterized by Duffie and Kan (1996). These allow for multiple factors driving interest rates, and have the computationally convenient feature that bond yields are linear functions of the underlying set of state variables. Initial research on these models made two specializing assumptions. First, the market price of risk was assumed to be a multiple of interest rate volatility, and second, the state variables were assumed to move independently. These assumptions allow the calculation of closed form likelihood functions, facilitating estimation of the models.

Despite their advantages, certain implications of these models appear to be at odds with the data. Dai and Singleton (2000) find strong evidence of nonzero correlations among the state variables. Ghysels and Ng (1998) and Balduzzi and Eom (1997) find evidence against (particular) affine models using semiparametric tests. Duffee (2001) finds that assuming the price of risk to be proportional to volatility prevents the model from generating realistic behavior of expected excess returns to bonds. Moreover, there is evidence of nonlinearity in expected interest rate movements (see, for example, Pfann et al. (1996), Aït-Sahalia (1996), Conley, Hansen, Luttmer, and Scheinkman (1997), and Stanton (1997)) that is inconsistent with the affine models that have been estimated to date.

Partially in response to these perceived limitations, recent research has considered more flexible models. For example, Dai and Singleton (2000) estimate affine models in which the state variables are allowed to be correlated, but retain the assumption that the price of risk is proportional to volatility. There are now also a number of models that allow either nonlinear dynamics or a more flexible form for the price of risk. However, as the models become more sophisticated, estimation has become more challenging. In particular, in all but a few special

---

2 The early models of Vasicek (1977) and Cox, Ingersoll, and Ross (1985) are single factor examples of this class of model.
3 This assumption for the form of the price of risk is not required for the affine property to hold. This requires restrictions only on the risk-neutral drift. The assumption implies that interest rate dynamics are affine under the physical measure as well as the risk-neutral measure, simplifying estimation.
4 Parametric nonlinear term structure models include Longstaff (1989), Beaglehole and Tenney (1992), Constantinides (1992), Ahn and Gao (1999), Duarte (1999), Leippold and Wu (2000), and Ahn, Dittmar, and Gallant (2001). All but the first two have prices of risk that are more flexible than those in Dai and Singleton (2000). Stanton (1997) and Boudoukh, Richardson, Stanton, and Whitelaw (1998) are nonparametric (and therefore nonlinear) models of both physical drifts and prices of risk. Duffee (2001) constructs an affine model with linear dynamics where the prices of risk are more general than Dai and Singleton.
cases we cannot write down a closed form likelihood function. These estimation problems are exacerbated as researchers estimate their models using ever broader cross-sections of bond yields, often requiring them to assume some form of measurement error in the data.

These issues are by no means unique to the term structure literature. For example, many multifactor option pricing models also possess intractable likelihood functions, and estimation of these models is also performed using a broad cross-section of asset prices (in this case, options with different strike prices). The method that has emerged as the standard estimation technique in such situations is the “Efficient Method of Moments” (EMM), a simulation-based GMM estimator developed by Gallant and Tauchen (1996). It is well known that the best possible GMM estimator is the maximum likelihood estimator, in which the moments used are the score vector corresponding to the likelihood function. However, this score vector is usually impossible to calculate, or even simulate efficiently, for the sort of models we are interested in estimating. The insight of Gallant and Tauchen was to suggest using the score vector from some other, more tractable, auxiliary model for the data. If the distribution implied by the auxiliary model comes close to that given by the actual likelihood function, then parameter estimates from using its score vector will be almost as efficient as MLE estimators. One other convenient feature of EMM is that, since it is a path simulation technique, it can very easily handle measurement errors in the data.

The usual implementation of EMM uses a semi-nonparametric (SNP) auxiliary model, which summarizes substantial information about the conditional distribution of the data without imposing any structure from the model of interest. Using SNP as the auxiliary model has the useful property that, asymptotically, EMM then attains the same efficiency as maximum likelihood. However, while its small sample properties have been investigated in a few cases involving a single data series, little is known about its small sample properties when used with multidimensional, highly correlated data, as when estimating multifactor term structure or option pricing models.

In this paper we estimate a multifactor term structure model that simultaneously exhibits three of the key features needed to address some of the shortcomings of more restrictive models: correlated factors, nonlinear dynamics, and a flexible price of interest rate risk. The model’s likelihood function is intractable, ruling out the use of maximum likelihood, and

For example, it is the estimation method used in Dai and Singleton (2000), Ahn, Dittmar, and Gallant (2001), Chernov and Ghysels (2000), Andersen and Lund (1997), and Gallant and Tauchen (1997)

I.e. parameters are estimated by minimizing the distance between certain population moments implied by the model and sample estimates of those moments calculated from the data.

This is made more formal by Gallant and Long (1997).

See Section 2.3.


See, for example, Zhou (1999).
making it a natural candidate for the use of EMM. The first contribution of the paper is to use this estimation as a test bed to investigate the small-sample properties of EMM in the context of term structure estimation. We find that, despite its attractive asymptotic properties, SNP with EMM performs very poorly in practice when applied to the estimation of dynamic term structure models using multiple points on the yield curve.

One problem with the SNP/EMM approach is that it requires a large number of moment conditions. Estimation of term structure models requires information about both the time-series and cross-sectional properties of bond yields. Yields of different maturity bonds are not perfectly correlated, thus the amount of information available to the econometrician increases with the breadth of the cross-section. But to summarize the information in a broad cross-section without imposing much structure requires the use of many moments.\footnote{The number of moment conditions implied by a given SNP auxiliary model increases with the square of the number of points on the term structure considered.} It is well known that, as the number of moment conditions increases, the small-sample properties of GMM estimators can differ, sometimes dramatically, from their asymptotic properties.\footnote{See Tauchen (1986), Kocherlakota (1990), Ferson and Foerster (1994), and Hansen, Heaton, and Yaron (1996).} Our results show that the samples of interest rates available to econometricians are definitely “small”, especially when we take into account the high persistence of interest rates.

We also investigate the use of a linearized Kalman filter as an alternative auxiliary model. This reduces the above problem significantly, since the number of moment conditions is now equal to the number of parameters being estimated. We use Monte Carlo simulations to compare the performance of SNP as an auxiliary model with the performance of the Kalman filter, and find that for the models examined in this paper, and for reasonable sample sizes, the results strongly support the choice of the Kalman filter.

However, we also conclude that (again, for the models we examine, and for reasonable sample sizes) this simulation-based estimation procedure, even when linearized Kalman filter moments are used, tends to produce poor estimates of the variance-covariance matrix of the estimated parameters. Thus statistical inference is unreliable. By contrast, the direct use of analytic moments from a linearized Kalman filter, even though they are misspecified, produces more accurate estimates of this variance-covariance matrix. Because the magnitude of the misspecification is small in practice, we advocate the use of a linearized Kalman filter when estimating dynamic term structure models, rather than using EMM at all.

The second contribution of our paper concerns the interpretation of results from estimating dynamic term structure models. We document a small-sample bias against the restrictive Dai-Singleton type of risk specification in favor of more general specifications, even when the true model fits into the Dai-Singleton class. This bias is related to the well-known downward
bias in small-sample estimates of the speed of mean reversion of near unit-root processes such as bond yields. Ball and Torous (1996) argued that using a combination of the cross-sectional and time-series information in yields could substantially reduce the small-sample bias in estimates of drifts of bond yields. However, we find that their conclusion relies heavily on the assumption of a restrictive functional form for the price of risk, which closely links the cross-sectional behavior of yields to their time-series behavior.

The use of more general risk specifications weakens these links, with the result that in small samples, both the physical drift and the risk adjustment to this drift are biased. We use Monte Carlo simulations to measure the magnitude of these biases. For typical sample sizes and standard test sizes, we are frequently able to reject restrictive models of the price of risk in favor of more general models, even through the data are known to be generated by the more restrictive models.

The final contribution of our paper is to estimate our model using US Treasury rates from January 1974 to December 1998. The model is in the affine class of Duffie and Kan, but allows for a more general specification of the price of risk than Dai and Singleton (2000). This specification produces nonlinear interest-rate dynamics under the physical measure, and also allows for more flexibility in the relationship between bond returns and the slope of the term structure. Unfortunately, while qualitatively right, it is unable to match the magnitude of the relation that we observe in the data. The estimation of the model does, however, allow us to offer some insights on what features must be built into a term structure model in order for it to generate real-world interest-rate dynamics. In particular, we explain why the form for the price of risk examined in this paper, although more general than the Dai and Singleton (2000) framework, is nonetheless too restrictive to generate realistic behavior for expected excess returns to bonds. We describe the kind of flexibility that must be built into the price of risk in order to fit the empirical features of Treasury bond yield and return dynamics.

The remainder of the paper is organized as follows. Section 2 discusses various econometric methods used to estimate diffusion models, and presents a motivating example in which SNP/EMM is used to estimate a simple one factor model using simulated data. The poor results of this estimation clearly motivate our search for a better combination. Section 3 presents the general interest rate model. In Section 4, special cases of the model are used to investigate small-sample properties of the econometric methods. Results from fitting the model to U.S. Treasury bond data are presented in Section 5. Some concluding comments are offered in Section 6.

---

13Duarte (1999) has independently investigated the same kind of model, although he focuses on issues that are largely different from those considered here.
2 Methods of Estimating Term Structure Models

2.1 Setup

We consider estimating stationary, continuous-time dynamic term structure models using observations of bond yields. At equally spaced intervals \( t = 1, \ldots, T \) we observe a length-\( m \) vector of yields \( \tilde{y}_t \). (The tilde denotes a realization of the random variable \( y_t \).) Denote the history of observed yields through \( t \) as \( \tilde{Y}_t = (\tilde{y}'_1, \ldots, \tilde{y}'_t)' \). Yields are a function of a length-\( n \) latent state vector \( x_t \) and (perhaps) a latent noise vector \( w_t \):

\[
y_t = y(x_t, w_t, \rho_1)
\]

(1)

The noise may represent market microstructure effects or measurement error. The vector \( \rho_1 \) contains the parameters of this function. The state vector follows a diffusion process

\[
dx_t = \mu(x_t, \rho_2) \, dt + \sigma(x_t, \rho_2) \, dB_t
\]

(2)

where \( \rho_2 \) is the parameter vector and \( B_t \) is a vector of independent Brownian motions. The density function associated with the noise is

\[
g_w(w_1, \ldots, w_T).
\]

(3)

Typically, we will have candidate functional forms for (1), (2), and (3). We are interested in the resulting probability distribution of yields. Stack the parameter vectors \( \rho_1 \) and \( \rho_2 \) into \( \rho \). Then we can always write the log density function of the observed data as

\[
\ln g_{Y_T}(\tilde{Y}_T) = \sum_{t=1}^{T} \ln g_t(\tilde{y}_t \mid \tilde{Y}_{t-1}, \rho)
\]

(4)

where \( g_t(y_t \mid Y_{t-1}, \rho) \) is the unconditional distribution of \( y_t \). The primary difficulty in estimating models with this structure is that the functional form for \( g_t(\cdot) \) is often unknown.

2.2 Some standard estimation methods

Here we consider standard techniques for estimating these kinds of models. (We exclude methods that use path simulations, which are discussed in the next subsection.) The techniques use information about the conditional densities \( g_t(y_t \mid Y_{t-1}, \rho) \) in (4). Therefore the first task is to calculate these densities. We start with the dynamics of the latent state vector.
Given (2), the transition density from value \( x_t \) at time \( t \) to value \( x_s \) at \( s \), \( p(s, x_s \mid t, x_t) \), must satisfy the Kolmogorov forward equation (see Øksendal (1985)),

\[
\frac{\partial p(s, x_s \mid t, x_t)}{\partial s} = -\frac{\partial}{\partial x_s} (\mu(x_s, \rho_2)p(s, x_s \mid t, x_t)) + \frac{1}{2} \frac{\partial^2}{\partial x_s^2} (\sigma^2(x_s, \rho_2)p(s, x_s \mid t, x_t)),
\]

which can, in principle, be solved for the conditional density \( p \) as a function of \( \rho_2 \). Now, if the density function of noise (3) allows us to infer \( x_t \) from \( y_t \) (say, if we make the extreme assumption that there is no noise in yields), the conditional density of yields can be expressed in terms of \( p(s, x_s \mid t, x_t) \), (1) and (3). We can then use maximum likelihood to estimate the model’s parameters (see, for example, Lo (1988)). This approach was followed by Chen and Scott (1993) and Pearson and Sun (1994) in estimating the parameters of term structure models in which the state variables followed independent CIR processes.

Unfortunately, we can only solve equation (5) numerically (except in rare cases such as CIR), making this technique very burdensome.\(^\text{14}\) When the state density \( p \) is too complicated or time-consuming to calculate, an alternative approach is the Generalized Method of Moments (GMM) of Hansen (1982). For fairly general diffusion processes, Hansen and Scheinkman (1995) show how to derive analytic moment restrictions from equation (2), but these do not take advantage of all of the information contained in the discretely observed data. Chan, Karolyi, Longstaff, and Sanders (1992) use approximate moment conditions, obtained by assuming that the size of the time interval between observations is “small”, so the expected change and variance over the next period are given by

\[
E[\Delta x_t] \approx \mu(x_t, \rho_2) \Delta t,
\]
\[
\text{var}[\Delta x_t] \approx \sigma^2(x_t, \rho_2) \Delta t.
\]

This approach is extended and used in a nonparametric setting by Stanton (1997) and Boudoukh et al. (1998). These approximations are very simple to implement, and very close to the true moments for small time intervals, but estimators based on these approximations are, strictly speaking, inconsistent.

The second problem is that all of these methods (ML or GMM using moments of \( x_t \)) rely on the ability to infer \( x_t \) from \( y_t \). In many dynamic term structure models, there is

\(\text{\textsuperscript{14}}\)Some progress has recently been made on this front. Pedersen (1995) and Santa-Clara (1995) \[\text{see also Brandt and Santa-Clara (1999) and Duffie, Pedersen, and Singleton (2000)}\] develop a simulation-based approach that allows the approximation of the likelihood function by splitting each observation interval into small pieces, and using the fact that the distribution of the variables approaches conditional normality as the length of the intervals shrinks towards zero. A¨ıt-Sahalia (1999) develops a series of approximations to the likelihood function that are tractable to estimate, but his method only applies to a single variable (or to multiple variables, but only if they are all independent).
a one-to-one mapping between a length-$n$ $x_t$ and $n$ bond yields. Thus we can pick any $n$ points on the yield curve, assume these yields have no noise, and invert the appropriate pricing equations to infer $x_t$. This commonly-adopted approach was first used by Pearson and Sun (1994) and Chen and Scott (1993). However, when we observe $m > n$ bond yields, there is in general no set of values for the $n$ factors that exactly matches the $m$ bond yields every period. One solution is to assume that $n$ of the yields are measured without error, and allow for noise in the remaining yields. This has the disadvantage that the choice of the yields estimated without error is, necessarily, somewhat ad hoc.

If the term structure model does not imply a one-to-one mapping between $x_t$ and bond yields even in the absence of noise, as in Ahn, Dittmar, and Gallant (2001), or if we allow all the yields to be measured with noise, the underlying state variables cannot be observed. In this case, a natural approach is to use some sort of filtering. Given an analytic conditional density $p(s, x_s | t, x_t)$ for the state vector, exact nonlinear filtering is possible but numerically demanding, especially for non-scalar $x_t$. We are unaware of any empirical term-structure implementations of exact filtering when the dynamics of $x_t$ are nonlinear. Approximate linear filtering is easier to implement. The Kalman filter has been applied to term structure models in which $x_t$ has affine dynamics and thus analytic expressions of the first two moments of the conditional density $p(s, x_s | t, x_t)$ are available. The Kalman filter is exact only when the underlying model is entirely linear and homoskedastic. In a term-structure setting, these conditions hold only in the case of a multifactor version of the Vasicek (1977) model. For all other models, parameter estimates obtained directly from Kalman filter estimation are inconsistent. There is Monte Carlo evidence that when the underlying model is linear but heteroskedastic, the inconsistency may be of limited importance in practice.

The main point to take away from this discussion is that none of these techniques is generically applicable to dynamic term structure models. In other words, for each technique, there is some class of term structure models for which it is ill-suited. By contrast, path simulation techniques can be used to estimate any dynamic term structure model. We turn to these techniques next.

15This holds for the true model, but may not always be possible in practice. For example, observed bond yields might imply negative values for state variables that ought never to be negative.

16The exact filter of Kitagawa (1987) is implemented by Lu (1999) for a Constantinides (1992) model, which has Gaussian dynamics for the state vector. For a discussion of the high computational cost of Kitagawa’s filter, see the comments by Kohn and Ansley (1987) and Martin and Raftery (1987).


18See, e.g., Babbs and Nowman (1999).

19See, e.g., de Jong (1998) and Duan and Simonato (1997). In addition, in certain cases, as in Lund (1997), the approximation error can be reduced using iterative techniques or numerical integration.
2.3 Path simulation and EMM

Path simulation means that simulations produced with the dynamic model are used to draw inferences about the density function $g_{Y_T}(Y_T)$. These simulations can be used to calculate arbitrary population moments as functions of the parameters of the process being estimated, which can be compared with sample moments estimated from the data [see, for example, Duffie and Singleton (1993)]. The density function $g_{Y_T}(Y_T)$ can be intractable or unknown. There remains, however, the question of which moments to simulate.

The optimal moments, if available, would be the score vector corresponding to likelihood function. Unfortunately, this is impossible to calculate for many term-structure models, particularly when we allow for measurement error in the observations. Gallant and Tauchen (1996) suggest replacing the score vector from the true, but intractable, model with the score vector from some other, more tractable, auxiliary model. First, the parameters of the auxiliary model are estimated from the observed data using quasi maximum likelihood (QML). Next, the parameters of the true model are estimated by finding the values that set the score vector of the auxiliary model, evaluated at their estimated values, (close) to zero in a large simulated dataset. This leaves only the decision of what auxiliary model to use.

The insight of Gallant and Tauchen (1996) was to note that if the distribution implied by the auxiliary model was close to that implied by the true underlying model, then the estimates obtained should be close to those obtained using maximum likelihood, leading to the name Efficient Method of Moments (EMM) usually associated with this technique. Indeed, it is straightforward to show that if the auxiliary model is the score vector from the true likelihood function, then the EMM estimator is identical to the ML estimator.

As defined by Gallant and Tauchen (1996), EMM allows considerable freedom in choosing the auxiliary model. In particular, "... the auxiliary model need not nest the structural model and in fact may be misspecified". By far the most common implementation of EMM is to use a semi-nonparametric (SNP) auxiliary model. This is in large part motivated by its very desirable asymptotic properties. Gallant and Long (1997) show that with this choice of auxiliary model, EMM asymptotically attains the same efficiency as maximum likelihood [see also Gallant and Tauchen (1999)]. However, although it is desirable to use estimation

---

20 This is closely related to the “indirect inference” approach of Gouriéroux, Monfort, and Renault (1993), which involves matching parameter estimates in the auxiliary model (rather than the score vector). It has the advantage that this may help to correct any finite sample bias in the estimated parameters of the true model [See Gouriéroux, Monfort, and Touzi (1994) and Broze, Scaillet, and Zakoian (1998)]. However, it has the disadvantage that it requires repeated estimation of the auxiliary model, whereas the EMM approach only requires the model to be estimated once.

21 Ignoring simulation error, i.e. this result holds if the simulation size is large enough that sample means equal their population counterparts.

techniques that have good asymptotic properties, the finite-sample properties are more important in practice. While we know something about the finite-sample properties of EMM estimation when the underlying data is a univariate time series,\(^{23}\) little is known about these properties when the underlying data is a multivariate time series, such as when estimating dynamic term structure or option pricing models using panels of bond yields or option prices. There is good reason to suspect that there may be problems in this setting. Efficient Method of Moments is a GMM estimator, and it is well-known that the finite-sample properties of GMM can deteriorate seriously as the number of overidentifying restrictions increases [See, for example, Tauchen (1986), KocHerlakota (1990), Ferson and Foerster (1994), and Hansen, Heaton, and Yaron (1996)]. Because SNP puts little structure on data, the number of SNP parameters that are used to summarize a multivariate time series can be large. As we will see below, an SNP specification uses a minimum of \(m(m + 1)(3/2)\) parameters to fit an \(m\)-dimensional time series, and usually many more than this.

This curse of dimensionality leads us to consider a more structural auxiliary model as an alternative to SNP. In different contexts, both Gouriéroux and Monfort (1996) and Duan and Simonato (1997) suggest the use of the Kalman filter. Neither the standard nor extended (i.e., linearized discrete-time dynamics) Kalman filter can be applied directly to general diffusion models because the first two moments of the conditional density \(p(s, x_s \mid t, x_t)\) are typically unknown. Therefore we use linearized instantaneous dynamics in the Kalman filter, producing a misspecified model, but one that is structurally similar to the true dynamic model.\(^{24}\) We discuss the simulation technique and auxiliary models in more detail in the next subsection.

\[\text{2.3.1 EMM and score vector simulation}\]

This discussion is largely drawn from Gallant and Tauchen (1996) and Gallant and Tauchen (1998a), which contain additional details. Our goal is to estimate the vector \(\rho\) in (4), but we cannot use this density directly, perhaps because it is unknown or intractable. We introduce an auxiliary function that (perhaps approximately) expresses the density of \(y_t\) as a function of \(Y_t\) and a parameter vector \(\theta\):

\[f_t(y_t \mid Y_{t-1}, \theta).\]

\(^{23}\)See, for example, Zhou (1999), Chumacero (1997), Andersen, Chung, and Sørensen (1999), and Andersen and Sørensen (1996).

\(^{24}\)Using this auxiliary model agrees with Gallant and Tauchen’s usage of the term “EMM”. However, other authors define EMM more strictly. For example, Gouriéroux and Monfort (1996) use it only when the auxiliary model is a high dimensional descriptive model capable of approximating the true distribution arbitrarily closely as the sample size and number of parameters grow to infinity. By their definition, the use of SNP as the auxiliary model would qualify as EMM, but using an approximate Kalman filter would not. In this paper we use Gallant and Tauchen’s broader definition.
The observed data are summarized by estimating $\theta$ with QML. (The form of $f_t(\cdot)$ is chosen to make this estimation tractable.)

$$\tilde{\theta}_T = \operatorname*{argmax}_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \ln f_t(\tilde{y}_t | \tilde{Y}_{t-1}, \theta). \quad (6)$$

The next step is to simulate a long time series $\hat{Y}_N(\rho) = (\hat{y}_1(\rho), \ldots, \hat{y}_N(\rho))'$ using (1), (2) and (3). If the discrete density $p(s, x_s | t, x_t)$ is known, $x_t$ can be generated from this density. Otherwise the continuous process (2) is discretized.\(^{25}\) The simulated time series is used to calculate the expectation of the score vector associated with the auxiliary model:

$$m_T(\rho, \tilde{\theta}_T) = \frac{1}{N} \sum_{\tau=1}^{N} \frac{\partial}{\partial \theta} \ln f_\tau[\hat{y}_\tau(\rho) | \hat{Y}_{\tau-1}(\rho), \tilde{\theta}_T].$$

As the length of the simulation, $N$, approaches infinity, this sample mean approaches the unconditional expectation of the score vector. The estimator of $\rho$ is

$$\hat{\rho}_T = \operatorname*{argmin}_{\rho \in \mathbb{R}} m_T'(\rho, \tilde{\theta}_T) \tilde{I}_{OP,T}^{-1} m_T(\rho, \tilde{\theta}_T), \quad (7)$$

where $\tilde{I}_{OP,T}$ is the outer-product estimate of auxiliary model’s information matrix

$$\tilde{I}_{OP,T} = \frac{1}{T} \sum_{t=1}^{T} \left[ \frac{\partial}{\partial \theta} \ln f_t(\tilde{y}_t | \tilde{Y}_{t-i}, \tilde{\theta}_T) \right] \left[ \frac{\partial}{\partial \theta} \ln f_t(\tilde{y}_t' | \tilde{Y}_{t-i}, \tilde{\theta}_T) \right]' \quad (8)$$

An estimate of the asymptotic variance-covariance matrix of $\hat{\rho}_T$ is

$$\hat{\Sigma}_T = \frac{1}{T} [(\hat{M}_T)' \tilde{I}_{OP,T}^{-1} (\hat{M}_T)]^{-1} \quad (9)$$

where

$$\hat{M}_T = \left( \frac{\partial}{\partial \rho} \right) m_T(\rho, \tilde{\theta}_T). \quad (10)$$

If there are more moment conditions (length of $\theta$) than parameters (length of $\rho$), then under the null hypothesis,

$$C = Tm_T'(\hat{\rho}_T, \tilde{\theta}_T) \tilde{I}_{OP,T}^{-1} m_T(\hat{\rho}_T, \tilde{\theta}_T)$$

is asymptotically distributed as a $\chi^2(q)$ random variable, where $q$ is the number of overidentifying moment conditions.

\(^{25}\)Discretization techniques are discussed in Kloeden and Platen (1992).
There are two features of this estimation technique that are worth highlighting (to foreshadow some results we will see later). First, if the parameter estimate of the auxiliary model (6) is biased in a finite sample, this bias will also be reflected in the parameter estimate of the true model (7). Second, note that the only information from the observed data that is used in constructing $\hat{\rho}_T$ and its associated variance-covariance matrix is the information in the auxiliary parameter estimate $\tilde{\theta}_t$ and in the estimate of the information matrix $\tilde{I}_{OPT}$.

### 2.3.2 SNP as an auxiliary model

The standard auxiliary model for use with EMM is the SNP (for SemiNonParametric) model of Gallant and Tauchen (1992) [see also the user’s guide, Gallant and Tauchen (1998b)]. The use of SNP requires a slight modification to the above setup, so we need some additional notation. The $L$ lags of observed data at time $t$ are denoted $\tilde{Y}^L_{t-1} = (\tilde{y}_{t-1}, \ldots, \tilde{y}_{t-L})'$. The SNP approach is to express the conditional density of the dataset under analysis in the form of a Hermite polynomial multiplied by a normal density, i.e.,

$$f(y_t \mid Y^L_{t-1}, \theta) = c(Y^L_{t-1}) [h(z_t \mid Y^L_{t-1})]^2 \phi(z_t), \quad (11)$$

where

- $\phi(.)$ represents the standard normal p.d.f.,
- $h(z_t \mid Y^L_{t-1})$ is a Hermite polynomial in $z_t$,
- $c(Y^L_{t-1})$ is a normalization constant (equal to $1/\int [h(s \mid Y^L_{t-1})]^2 \phi(s) \, ds$), and
- $z_t$ is a normalized version of $y_t$, defined by

$$z_t = R^{-1}_{t-1} (y_t - \mu_{t-1}),$$

where $\mu_{t-1}$ is the conditional mean and $R^{-1}_{t-1}$ is the Cholesky decomposition of the conditional variance of $y_t$. In both cases, the conditioning information is $Y^L_{t-1}$. The conditional mean is specified as a vector autoregression (VAR). Because (11) is not defined for $t \leq L$, the sum in (6) begins with $t = 1 + L$.

The SNP specification allows great flexibility in fitting the conditional distribution. In particular, we are free to choose:

- the dimensionality of the Hermite polynomial in $z$, $K_z$. This allows for non-Gaussian behavior.
- the degree of the polynomial in $Y^L_{t-1}$ that makes up each of the coefficients in the Hermite polynomial, $K_y$ (another way to allow for conditional heterogeneity).
- the number of lags in the Hermite polynomial, $K_p$.  

11
• the number of lags in a VAR specification for \( \mu_{t-1}, L_\mu \).
• the degree of an ARCH (or GARCH, setting \( L_g > 0 \)) specification for the scale transformation \( R_{t-1}, L_r \).

Choice of an appropriate specification is performed by using a model selection criterion, such as the Schwarz Bayes information criterion [see Schwarz (1978)], which rewards good fit, while penalizing over-parametrization. Gallant and Tauchen (1998b) discuss a search strategy for finding an appropriate parametrization for a given problem.

2.3.3 An example of SNP’s finite sample properties

Because SNP has nice asymptotic properties, it is important to justify why we want to consider alternatives to SNP. To illustrate that SNP has finite-sample properties that are not so desirable in a dynamic term structure context, we use the following example. One thousand weekly observations (approximately 20 years)\(^{26}\) on the instantaneous riskless interest rate were simulated from a standard one factor Cox, Ingersoll, and Ross (1985) model,

\[
dr_t = \kappa(\mu - r_t) \, dt + \sigma \sqrt{r_t} \, dB_t, \tag{12}
\]

where the price of interest rate risk is given by the parameter \( \lambda \). The parameter values used for the simulation were

\[
\kappa = 0.29368, \quad \mu = 0.07935, \quad \sigma = 0.11425, \quad \lambda = -0.12165,
\]

corresponding to the values estimated in Pearson and Sun (1989). The simulation was performed using the fact [See Cox, Ingersoll, and Ross (1985)] that \( r_t \) is conditionally distributed as a multiple of a non-central \( \chi^2 \) distribution. For each simulated value of \( r_t \), the one year yield was calculated, and to prevent perfect correlation between the two series, some i.i.d. noise was added to the one year interest rate, in the form of a normally distributed random variable with standard deviation \( \sigma_\epsilon = 0.0001 \) (one basis point), making a total of five parameters determining the joint distribution of the two series, \( \{r_t, y_t^1\} \).

The first step in estimating \( \kappa, \mu, \sigma, \lambda \) and \( \sigma_\epsilon \) was to determine the appropriate SNP specification using the search methodology described in Gallant and Tauchen (1998b). The

\(^{26}\)Very similar results were obtained using 100 years of simulated data.
optimal specification was 30314411, corresponding to

\[
\begin{align*}
L_u &= 3 \quad \text{(order of VAR)}, \\
L_g &= 0 \quad \text{(order of GARCH)}, \\
L_r &= 3 \quad \text{(order of ARCH)}, \\
L_p &= 1 \quad \text{(lags in Hermite polynomial)}, \\
K_z &= 4 \quad \text{(degree of Hermite polynomial)}, \\
I_z &= 4 \quad \text{(\# cross terms to suppress)}, \\
K_y &= 1 \quad \text{(order of polynomial in Hermite polynomial coefficients)}, \\
I_y &= 1 \quad \text{(\# cross terms to suppress)}.
\end{align*}
\]

An important issue in estimating all but the simplest dynamic term structure model is the challenge of finding a global optimum. We first estimated this model using a starting parameter vector of \((0.3, 0.11, 0.05, 0.01, 0.01)\). These are not equal to the true parameter values, but would not be an unreasonable set of values to start from if we did not already know the true parameter values. The minimization in (7) converged to a local, but not global, minimum. Table 1 shows that the estimated parameter values are substantially different from the true values, the reported standard errors are essentially zero, and the true parameter values are outside any conceivable confidence interval. (For example, the estimated value for \(\kappa\) is approximately 2 million standard errors from the true parameter value.) In addition, the \(\chi^2(44)\) test statistic has the value \(7.2 \times 10^{10}\).

We then reestimated the model starting from the true parameter values. The final estimated parameter values are different, but again a long way from the true parameters, and again the reported standard errors seem very small, though much larger than with the previous estimation (for example, the estimated value for \(\mu\) is now “only” 478 standard errors from the truth). The \(\chi^2(44)\) test statistic is much smaller, at 1557.8, but again this statistic would imply a huge rejection of the model, even though it is true by construction.

At least in this case, the use of SNP as a score vector leads to biased parameter estimates and standard errors, and to overstated \(\chi^2\) statistics. Moreover, the parameter estimates are highly dependent on where the estimation starts from. This in turn suggests that the finite sample properties of EMM as an estimator of dynamic term structure models may not be nearly as attractive as their asymptotic properties.\(^{27}\)

\(^{27}\)When used only to estimate the parameters \(\kappa, \mu\) and \(\sigma\) from a single series of observations on the short rate, \(r_t\), the procedure produced much more reasonable results. The problems seem to arise only when using multidimensional data. This, of course, is a standard feature of term structure estimation.
2.3.4 The Kalman filter as an auxiliary model

We first briefly review the standard Kalman filter. For details see, for example, Harvey (1989) or Hamilton (1994). The “observation equation” expresses \( y_t \) as a linear function of \( x_t \) and measurement error \( \epsilon_t \). The “transition equation” expresses the discrete-time evolution of \( x_t \) as linear in \( x_t \). These equations are determined by some underlying parameter vector \( \theta \).

\[
\begin{align*}
y_t &= H_0(\theta) + H_1(\theta)'x_t + \epsilon_t; \quad (13) \\
x_{t+1} &= F_0(\theta) + F_1(\theta)x_t + v_{t+1}; \quad (14) \\
E(\epsilon_t) &= 0; \quad (15) \\
E(v_{t+1}) &= 0; \\
E(\epsilon_t\epsilon_t') &= R(\theta); \\
E(v_{t+1}v_{t+1}') &= Q(\theta). \quad (18)
\end{align*}
\]

The vectors and matrices in (13) through (18) are functions only of \( \theta \), not of the observation or state vectors. With this setup, the Kalman filter recursion can be used to produce one-step-ahead forecasts of the state vector and observable vector, which we denote \( x_{t+1|t} \) and \( y_{t+1|t} \), and the variance-covariance matrices of these forecasts, which we denote \( P_{t+1|t} \) and \( V_{t+1|t} \) respectively. The recursion also produces the contemporaneous prediction of the state vector and its associated variance-covariance matrix, which we denote \( x_{t|t} \) and \( V_{t|t} \), respectively.

Quasi-maximum likelihood estimation using the Kalman filter is straightforward. The estimated parameter vector \( \tilde{\theta}_T \) solves (6) with

\[
\ln f_t(\tilde{y}_t \mid \tilde{Y}_{t-1}, \theta) = -\frac{1}{2}[m \ln(2\pi) + \ln |V_{t|t-1}| + (\tilde{y}_t - y_{t|t-1})V_{t|t-1}^{-1}(\tilde{y}_t - y_{t|t-1})].
\]

An estimate of the asymptotic variance-covariance matrix of the estimated parameters \( \tilde{\theta}_T \) is

\[
\tilde{\Sigma}_T = \frac{1}{T}[\tilde{I}_{2D,T}I_{OP,T}^{-1}\tilde{I}_{2D,T}]^{-1}
\]

where \( I_{2D,T} \) is the second-derivative estimate of the information matrix

\[
\tilde{I}_{2D,T} = -\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 \ln f_t(\tilde{y}_t \mid \tilde{Y}_{t-1}, \theta)}{\partial \theta \partial \theta'} \bigg|_{\theta = \tilde{\theta}_T}.
\]

As noted earlier, the only term structure models that satisfy the assumptions underlying
the Kalman filter are in the Vasicek (1977) class. The “extended” Kalman filter handles a nonlinear model through a first-order Taylor series approximation of the model. The appropriateness of this linearization depends, of course, on the amount of nonlinearity in the true model. Two classes of term structure models are partially linear and therefore better suited to Kalman filter estimation. Quadratic term structure models have linear transition equations with a constant variance covariance matrix. Duffie-Kan type models have linear measurement equations for zero-coupon bond yields (if the yields are observed with serially uncorrelated measurement error $w_t$ with a constant variance-covariance matrix $R$).

From the perspective of estimating many dynamic term structure models, a major problem with this extended Kalman filter approach is that it requires an expression for the discrete-time dynamics of $x_t$. As discussed in Section 2.2, analytic expressions are available for only a few models, such as quadratic term structure models. Accordingly, to use the Kalman filter as an auxiliary model we modify it by replacing (14) with a linearization of the instantaneous dynamics of $x_t$ in (2). The linearization is taken in the neighborhood of $x_{t|t}^p$. The time between discrete observations is denoted $\Delta t$. The linearization is (suppressing the dependence on parameters)

\[
x_{t+1} = F_{0t} + F_{1t} x_t + v_{t+1};
\]

\[
F_{0t} = \left( \mu(x_{t|t}^p) - \frac{\partial \mu(x_t)}{\partial x_t'} \bigg|_{x_t=x_{t|t}^p} \right) x_{t|t}^p \Delta t; \tag{22}
\]

\[
F_{1t} = I + \frac{\partial \mu(x_t)}{\partial x_t'} \bigg|_{x_t=x_{t|t}^p} \Delta t; \tag{23}
\]

\[
Q_t = \sigma(x_{t|t}^p) \sigma(x_{t|t}^p)' \Delta t. \tag{24}
\]

The Kalman filter recursion begins with an initial $x_{0|0}^p$ and $P_{0|0}$. The Appendix describes how they are computed. In (21) through (24), three approximation errors are introduced. The first is that we use the instantaneous dynamics of $x_t$ as a proxy for the discrete-time dynamics of $x_t$. The second is the linearization of these dynamics. The third is the evaluation of these dynamics at filtered value of $x_t$ instead of an exactly-identified value of $x_t$. Therefore the parameter vector that maximizes the Kalman filter quasi-log-likelihood function will not, in general, be a consistent estimator of the true parameter vector. Nonetheless, because the auxiliary model is closely related to the (assumed) true data generation model, the auxiliary model should be an efficient method of compressing the information in the observed data.
3 A nonlinear term structure model

In this section, we describe a multifactor term structure model that, at least a priori, has the potential to overcome some of the problems that have been noted with prior models. The model is independently developed in Duarte (1999). Like the models considered by Dai and Singleton (2000), it is a multifactor affine model under the risk neutral measure, with correlated factors. However, we allow for a more general form for the price of risk. This allows the model to exhibit more flexibility in its relationship between bond returns and the slope of the term structure. In addition, it allows for the possibility of nonlinear behavior under the physical measure. As noted in the introduction, earlier research has documented that interest rates appear to be characterized by nonlinear dynamics.

The price for this flexibility is that the model has an intractable likelihood function. Its estimation thus provides a perfect test bed for analyzing the small sample properties of EMM and the Kalman filter with linearized instantaneous dynamics.

3.1 Interest rates under the equivalent martingale measure

This model uses the framework of Duffie and Kan (1996). There are \( n \) state variables, denoted \( x_t \equiv (x_{t,1}, \ldots, x_{t,n})' \). Uncertainty is generated by \( n \) independent Brownian motions. Under the equivalent martingale measure these are denoted \( \tilde{B}_t \equiv (\tilde{B}_{t,1}, \ldots, \tilde{B}_{t,n})' \); corresponding Brownian motions under the physical measure are represented without the tildes. The instantaneous nominal interest rate, denoted \( r_t \), is affine in the state:

\[
r_t = \delta_0 + \delta x_t.
\]

(25)

Here, \( \delta_0 \) is a scalar and \( \delta \) is an \( n \)-vector. The evolution of the state variables under the equivalent martingale measure is given by equation (26):

\[
dx_t = (K\theta - K x_t)dt + \Sigma S_t d\tilde{B}_t.
\]

(26)

In (26), \( K \) and \( \Sigma \) are \( n \times n \) matrices and \( \theta \) is an \( n \)-vector. Dai and Singleton (2000) show that \( \Sigma \) can be normalized to a diagonal matrix; we adopt their normalization here. The matrix \( S_t \) is also diagonal, with elements described in (27):

\[
S_{t(ii)} \equiv \sqrt{\alpha_i + \beta_i'x_t},
\]

(27)

where \( \beta_i \) and \( \alpha \) are \( n \)-vectors. This discussion assumes that the dynamics of (26) and (27) are well-defined, which requires that \( \alpha_i + \beta_i'X_t \) is nonnegative for all \( i \) and all possible \( x_t \).
See Dai and Singleton (2000) for the required restrictions.

Using the results of Duffie and Kan (1996), we can write the price and yield of a zero-coupon bond that matures at time \( t + \tau \) as

\[
P(x_t, \tau) = \exp[A(\tau) - B(\tau)'x_t],
\]

\[
Y(x_t, \tau) = \frac{1}{\tau}[\frac{1}{\tau}[-A(\tau) + B(\tau)'x_t]].
\]

In (28) and (29), \( A(\tau) \) is a scalar function and \( B(\tau) \) is an \( n \)-valued function. Both can be represented as solutions to a set of ordinary differential equations (ODEs).

### 3.2 The price of risk

The dynamics of \( x_t \) under the physical measure are determined by specifying the market price of risk. Defining \( \pi_s/\pi_t \) as the state price deflator for time-\( t \) pricing of time-\( s \) payoffs, we can write

\[
\frac{d\pi_t}{\pi_t} = -r_t dt - \Lambda_t dB_t.
\]

The element \( i \) of the \( n \)-vector \( \Lambda \) represents the price of risk associated with the Brownian motion \( Z_{t,i} \). We parameterize the model as follows. Let \( \lambda_1 \) and \( \lambda_2 \) be \( n \)-vectors of constants. Define \( \Lambda \) as

\[
\Lambda_t = \Sigma^{-1} \lambda_1 + S_t \Sigma^{-1} \lambda_2
\]

This form, which is equivalent to the form adopted by Duarte (1999), nests the class of “completely affine” models estimated by Dai and Singleton (2000). For this more restrictive class, \( \lambda_1 \) is zero. When this vector is nonzero it introduces nonlinear dynamics into the physical measure as long as \( S_t \) is a nontrivial function of \( x_t \) (i.e., \( S_t \) is not simply a constant matrix).

One of the main advantages of this more general form of \( \Lambda \) is that individual elements of \( \Lambda \) can change sign, depending on the shape of the term structure (i.e., the depending on the elements of \( X_t \)). Thus investors’ willingness to face certain types of interest-rate risk can switch sign. As discussed at length in Duffee (2001), the structure of \( \Lambda \) in completely affine models is at odds with the stylized fact that excess bond returns tend to be positive when the yield curve is more steeply sloped than usual and negative when the yield curve is less steeply sloped than usual. It is an open question as to whether the more general form is sufficiently flexible to capture this stylized behavior of bond returns. Duarte concludes that it seems sufficient, while our conclusion is more negative.

\[\text{This class of models is affine under both the true and the risk-neutral probability distributions. It includes the models of Cox, Ingersoll, and Ross (1985) and Vasicek (1977).}\]
3.3 Interest rate dynamics under the physical measure

The general representation of the state price deflator’s dynamics in (30) allow us to write the dynamics of \( x_t \) under the physical measure as follows:

\[
dx_t = K(\theta - x_t)dt + \Sigma S_t \Lambda_t dt + \Sigma S_t dB_t
\]  

(32)

Combining (32) and (31), we can express the physical dynamics of the state vector as

\[
dx_t = [(K\theta + \psi) + S_t \lambda_1 - K^p x_t] dt + \Sigma S_t dB_t
\]  

(33)

where element \( i \) of the vector \( \psi \) is \( \alpha_i \lambda_{2i} \), the matrix \( K^p \) is defined as \( K - \Phi \), and row \( i \) of matrix \( \Phi \) is \( \beta_i \lambda_{2i} \).

Stationarity of \( x_t \) is determined largely by the eigenvalues of \( K^p \), because the affine function of \( x_t \), \( K^p x_t \), dominates the square-root function of \( x_t \), \( \Sigma S_t \lambda_1 \), for large \( x_t \). If the eigenvalues are all positive, \( x_t \) is stationary. If any of the eigenvalues of \( K^p \) are negative, \( x_t \) is nonstationary. If any of the eigenvalues are zero, stationarity will depend on the signs on the square-root terms in \( \Sigma S_t \lambda_1 \).

The combination of (25), (28), (33), and the structure of the ODEs in Duffie and Kan (1996) imply that the instantaneous bond-price dynamics are

\[
dP(x_t, \tau) = [r_t - B(\tau)'(\psi + S_t \lambda_1 + \Phi x_t)] dt - B(\tau)'\Sigma S_t dB_t
\]  

(34)

Instantaneous expected excess returns to a \( \tau \)-maturity bond, denoted \( e_{t,\tau} \), can be inferred from (34):

\[
e_{t,\tau} = -B(\tau)'(\psi + S_t \lambda_1 + \Phi x_t).
\]  

(35)

3.4 An example

To illustrate this framework, here we take a closer look at a one-factor version. The model is a simple extension of Cox et al. (1985), but its features (and limitations) provide a useful framework for interpreting the more complex models we estimate later in the paper. The instantaneous interest rate has CIR dynamics under the equivalent martingale measure:

\[
 dr_t = (\kappa \theta - \kappa r_t)dt + \sigma \sqrt{r_t} dB_t.
\]  

(36)
This model fits into the general framework by setting $\delta_0 = \alpha = 0$ and $\delta = \beta = 1$. Under the physical measure, the dynamics of $r_t$ can be written as

$$dr_t = (\kappa \theta + \lambda_1 \sqrt{r_t} - \kappa^p r_t) + \sigma \sqrt{r_t} dB_t, \quad \kappa^p = \kappa - \lambda_2. \quad (37)$$

The instantaneous interest rate is stationary if $\kappa^p > 0$ or if $\kappa^p = 0, \lambda_1 < 0$. Expected excess returns to bonds are given by

$$e_{t, \tau} = -B(\tau)(\lambda_1 \sqrt{r_t} + \lambda_2 r_t) \quad (38)$$

where $B(\tau)$ is given by the usual CIR pricing formula. The signs of $\lambda_1$ and $\lambda_2$ can be pinned down by two features of the empirical behavior of expected excess returns to bonds. First, mean excess bond returns are positive. Therefore, since $B(\tau) > 0$, we require

$$\lambda_1 E(\sqrt{r_t}) + \lambda_2 E(r_t) < 0. \quad (39)$$

Second, expected excess returns to bonds move in the same direction as the slope of the term structure. The parametric restrictions necessary to reproduce this feature in our one-factor example depend on the sign of $\kappa$. If $\kappa > 0$, investors price bonds as if shocks to interest rates die out over time. Therefore an increase in the short rate corresponds to a smaller increase in long-term yields, and thus a decrease in the slope of the term structure. This pins down the sign of the derivative of expected excess returns with respect to $r_t$:

$$\kappa > 0, \quad (1/2)\lambda_1 E(1/\sqrt{r_t}) + \lambda_2 > 0. \quad (40)$$

Multiply the second inequality in (40) by $E(r_t)$, and ignore the difference between $E(\sqrt{r_t})$ and $E(r_t)E(1/\sqrt{r_t})$:

$$(1/2)\lambda_1 E(\sqrt{r_t}) + \lambda_2 E(r_t) > 0. \quad (41)$$

A comparison of (39) and (41) reveals that in order to produce excess bond returns that are, on average, positive, and are positively correlated with the slope of the term structure, we require $\lambda_1 < 0$ and $\lambda_2 > 0$ (as long as $\kappa > 0$). The additional flexibility provided by $\lambda_1$ is vital here. As noted by Backus, Foresi, Mozumdar, and Wu (1998), a standard CIR-type model cannot simultaneously fit both of these empirical regularities. If $\lambda_1 = 0$ and $\lambda_2 > 0$, mean excess bond returns are negative, and if $\lambda_1 = 0$ and $\lambda_2 < 0$, expected excess returns move inversely with the slope of the term structure.

Stationarity implies $\lambda_2 \leq \kappa$. Thus imposing stationarity limits the magnitude of the relationship between $r_t$ (and hence the slope of the term structure) and expected excess
returns to bonds. Recall from (38) that expected excess returns are the product of exposure to interest rate risk \((-B(\tau))\) and the price of interest rate risk. If \(\kappa\) is close to zero, \(\lambda_2\) must also be close to zero, thus the price per unit of interest-rate risk will not fluctuate much with \(r_t\). If \(\kappa\) is large, \(\lambda_2\) can be large, and the price per unit of interest-rate risk can fluctuate substantially. However, with large \(\kappa\), \(-B(\tau)\) is small. Recall that with CIR pricing, \(-B(\tau) \approx \int_0^\tau \exp(-ks)ds\). (The formula is exact aside from a Jensen’s inequality term.) When \(\kappa\) is large, bonds are priced as if shocks to interest rates die off quickly, and bonds have little risk exposure. (Note that the standard CIR model has exactly the same limitation.)

This logic assumed \(\kappa > 0\). If \(\kappa < 0\), it is much easier to fit the above two facts about bond returns. The reason is that bonds are priced as if shocks to \(r_t\) are explosive. This produces ‘the tail wags the dog’ behavior in the term structure, in which longer-maturity yields fluctuate much more than do shorter-maturity yields. Then an increase in the short-term interest rate corresponds to an increase in the slope of the yield curve. A negative value of \(\lambda_2\) produces the correct correlation between the slope of the term structure and expected excess bond returns, and there is no lower bound on \(\lambda_2\). However, this implication is at odds with two facts about the term structure. First, shorter-maturity bond yields (say, two-year maturities) are more volatile than are longer-maturity bond yields, and second, short-term interest rates and the slope of the term structure tend to move in opposite directions.

As we shall see, these same limitations appear in the estimation of our multifactor term-structure model. Whether these limitations of the model are binding is, of course, an empirical question. In the empirical work that follows, we find that they are binding; the limitations are at the heart of the failure of this class of models to fit the empirical behavior of bond returns.

4 Finite-sample properties of the estimation methods

In this section we use Monte Carlo simulations to consider two questions. First, what kinds of biases show up in small-sample estimation of flexible term-structure models, such as the model examined here? Second, for the kinds of data samples we typically use in term-structure estimation, how do the small-sample properties of the linearized Kalman filter compare to those of the Kalman filter/EMM and SNP/EMM?
4.1 A one-factor example

Here we document small-sample biases associated with estimating models such as ours, in which the price of risk vector $\Lambda_t$ is more flexible than it is in the completely affine case. The most straightforward way to illustrate the biases is when the true term-structure model is a one-factor affine model, with a restrictive functional form for the price of risk. When the data are fit to a more flexible term structure model, estimates of the parameters that determine the drift of $r_t$ are biased.

We generate instantaneous interest rates using a CIR process. Equation (36) describes the equivalent-martingale dynamics and (37), with $\lambda_1 = 0$, describes the physical dynamics. The parameters of the process are based on the results of fitting a one-factor CIR model to monthly U.S. interest rate data from 1974 through 1998. We assume that the econometrician observes, with noise, 240 monthly observations (twenty years) of $r_t$ and the continuously-compounded one-year bond yield. The measurement errors are normally-distributed, independent across maturities and time, and have standard deviations $D_r$ and $D_1$.

The data are fit to the nonlinear term-structure model of (36) and (37), where $\lambda_1$ is allowed to be nonzero. The model is estimated with EMM using the linearized Kalman filter as an auxiliary model. Each EMM simulation has length 30,000 months (2,500 years of data). A small-scale Monte Carlo simulation (250 simulations) is used to generate distributions of parameter estimates and associated standard errors. The parameter estimates from both the auxiliary model and the EMM stage are displayed in Panels A and B of Table 2.

There are three points to take from these tables. First, the mean parameter estimates from the Kalman filter and EMM procedure are almost identical. This is not too surprising, given that the underlying data are generated by a model that is affine under both the equivalent and physical measures. Second, the mean parameter estimates imply that the interest rate dynamics are nonlinear. The estimates of $\lambda_1$ and $\kappa^p$ are biased upward. The table also documents that estimate of $\kappa$ is not biased, thus the bias in $k^p = \kappa - \lambda_2$ is equivalent to a downward bias in the estimate of $\lambda_2$.

These biases are created by the well-known small-sample bias in the estimation of the speed of mean reversion of a persistent process. Because this bias is an inherent feature of the generalization of affine models we consider, it is worth discussing it in detail here. Ball and Torous (1996) point out that if a standard CIR model is estimated using only time-series information, the speed of mean reversion under the physical measure, $\kappa^p$, is biased upwards; this is the standard near-unit-root problem. They also note that if both time-series and cross-sectional information are used in the estimation procedure, the bias is substantially reduced (and the precision of the estimate increases). The intuition is straightforward. The speed of mean reversion under the equivalent martingale measure, $\kappa$, determines the
contemporaneous covariances among changes in yields of different maturities. Therefore the sample covariances determine the estimate of $\kappa$. The mean slope of the term structure is a function of the price of risk $\lambda_2$, therefore the sample mean slope determines the estimate of $\lambda_2$. Thus $\kappa_2$ can be estimated using relatively precise cross-sectional information; the less informative speed of mean reversion in the sample is not needed to pin down the model’s parameters.

The point we make here is the conclusion of Ball and Torous depends critically on the restrictive form of the price of risk in the CIR model. When both time-series and cross-sectional information are used to estimate our generalization of a CIR model, the small-sample bias remains. The reason is that the price of interest rate risk is a function of both $\lambda_1$ and $\lambda_2$. The sample mean slope of the yield curve cannot determine them both. Therefore $\kappa_2$ cannot be estimated without using the speed of mean reversion in the sample, leading to a small-sample upward bias in this speed. Since $\kappa$ is determined with precision in the cross-section, the upward bias in $\kappa_2$ corresponds to a downward bias in $\lambda_2$. Then $\lambda_1$ is biased upwards to match the mean slope of the term structure.

A graphical view of this bias is informative. Figure 1 plots $r_t$ versus its drift, $\kappa \theta + \lambda_1 \sqrt{r_t} - \kappa_2 r_t$. The solid line is constructed with the parameters of the CIR model used to generate the true data and the dashed/dotted lines (they are virtually indistinguishable) are constructed using the mean estimated parameters from the two estimation procedures. The nonlinear shape of the dashed line allows it to produce both the correct value of $\kappa \theta$ (the leftmost point in the figure) and the correct mean instantaneous interest rate—note that the dashed line crosses the x-axis at essentially the same value of $r_t$ as does the solid line. In addition, it implies faster mean reversion than does the solid line. The additional parameter $\lambda_1$ allows the model to fit both the cross-sectional behavior of yields and the mean slope of the term structure, while simultaneously fitting the upward-biased speed of mean reversion in the sample.

Although the bias documented here is in the context of a nonlinear model of interest rates, it is not created simply by allowing for nonlinear dynamics in $r_t$, as in Chapman and Pearson (2000). Instead, it is created by loosening the tight restrictions that the CIR model places on the form of $\Lambda_t$. Presumably, any loosening of these restrictions, whether nonlinear or linear (as in Duffee (2001)) will give rise to the same kind of bias in the parameters that determine the drift of $r_t$.

The third point to take from this table is that the standard errors produced by the Kalman filter tend to be smaller and somewhat more accurate than those produced by EMM with a Kalman filter auxiliary model. A comparison of the mean Kalman filter standard errors with the corresponding standard deviations of the parameter estimates (across the 250
simulations) reveals that the standard errors typically underestimate the standard deviations, but the magnitudes of the underestimates are small. The worst case is the estimate of $\lambda_1$, for which the standard deviation is 1.16 times the mean standard error.

The standard deviations of the Kalman filter/EMM parameter estimates are similar to those of the Kalman filter parameter estimates (except for $\kappa \theta$), and the mean Kalman filter/EMM standard errors typically overestimate the corresponding Kalman filter/EMM standard deviations (the notable exception is $\kappa \theta$). The extreme overestimate is for $\kappa \theta$, where the mean standard error is 1.45 times the standard deviation.\footnote{The Kalman filter/EMM standard errors are also more variable across the Monte Carlo simulations. For example, the Kalman filter standard errors for $\lambda_1$ have a standard deviation of 0.14, compared with a standard deviation of 0.56 for the Kalman filter/EMM standard errors. This greater variability corresponds to a higher frequency of rejection of the nested linear model in favor of the nonlinear model. Of the 250 Kalman filter $t$-statistics for $\lambda_1$, thirteen imply rejection of the hypothesis that $\lambda_1 = 0$ at the one percent level. The corresponding number for the Kalman filter/EMM $t$-statistics is 31.}

Because the main differences between the Kalman filter and Kalman filter/EMM results are in the standard errors, we now take a closer look at the source of these differences. The variance-covariance matrices (9) and (19) differ from each other because (9) uses (10), the derivative of the expectation of the auxiliary model’s score vector with respect to the true model’s parameters, while (19) uses (20), the derivative of the actual score vector with respect to the auxiliary model’s parameters. There are three reasons why these score-vector derivatives will differ from each other. First, the true model differs from the auxiliary model. In this Monte Carlo simulation, however, the true model and the auxiliary model are very similar—they have almost the same structure and nearly identical parameter estimates. Second, because the length of the simulated series $N$ is finite, $m_t(\rho, \tilde{\theta}_T)$ will not exactly equal the expectation of the auxiliary model’s score vector. Third, and the reason we emphasize here, is that the EMM derivative uses less information from the data sample. The only information from the data used in (10) is that contained in the auxiliary model’s parameter estimate $\tilde{\theta}_T$. The Kalman filter derivative (20) is a function of the entire sample data $\tilde{Y}_T$.

Thus even in the case in which the auxiliary model is identical to the (assumed) true model, Kalman filter standard errors will differ from EMM standard errors in small samples. We have no small-sample theory to tell us which set of standard errors is more appropriate. Asymptotically, the two sets of standard errors will converge. As the sample size goes to infinity, the estimated auxiliary model parameters will converge to the true parameters, and the true parameters completely describe the probability distribution of the data. Therefore there will be no information in an asymptotically-large sample that is not contained in the auxiliary model’s parameter estimates.

The Monte Carlo results for this simple model suggest that estimating the model with
the linearized Kalman filter is superior to estimating the model with EMM using the Kalman filter as an auxiliary model. The parameter estimates are nearly identical, the Kalman filter standard errors are slightly more accurate, and the EMM procedure is substantially more computationally intensive. However, this comparison relies on an underlying data-generating process that is linear. Presumably if the underlying data-generating process is sufficiently nonlinear, the magnitude of misspecification in the linearized Kalman filter will outweigh any other advantages this procedure has relative to EMM. The important question is whether, for realistic data-generating processes, this misspecification is important. We consider this issue next, along with a comparison of SNP as an auxiliary model versus the Kalman filter as an auxiliary model.

4.2 A two-factor example

In this subsection we compare the performance of SNP/EMM to both the linearized Kalman filter and the Kalman filter/EMM, using a model that is closer to the kinds of nonlinear models that econometricians are likely to estimate in practice. The instantaneous interest rate is the sum of a constant and two independent factors.

\[ r_t = \delta_0 + x_{1,t} + x_{2,t} \]  

(42)

Under the equivalent martingale measure, the dynamics of the independent factors can be written as

\[ dx_{i,t} = (\kappa \theta_i - \kappa_i x_{i,t})dt + \sigma_i \sqrt{x_{i,t}} d\tilde{z}_{i,t}, \quad i = 1, 2. \]  

(43)

Thus under the equivalent martingale measure, interest rates are generated by a two-factor CIR process. But under the physical measure, the factors follow nonlinear processes.

\[ dx_{i,t} = (\kappa \theta_i - \kappa_i^p x_{i,t} + \lambda_{1i} \sqrt{x_{i,t}})dt + \sigma_i \sqrt{x_{i,t}} dz_{i,t}, \quad i = 1, 2. \]  

(44)

To identify the parameters in the model we use yields on bonds of three maturities \((m = 3)\). We assume that the available data are 240 monthly observations (20 years) of continuously-compounded yields on zero-coupon bonds with maturities of six months, two years, and ten years. All yields are observed with normally-distributed, iid measurement error. The standard deviations of the measurement errors are \(D_{1/2}, D_2,\) and \(D_{10}\).

The assumption that we observe three maturities is important because it, in part, determines the number of SNP parameters. The other determinant is the desired complexity of the SNP model. The base SNP model is a simple VAR with one lag and no ARCH, GARCH, or higher-order terms in the variance. The VAR part of the SNP model has \(m\) equations de-
scribing the bond yields, each with $m + 1$ parameters (the constant term and the coefficients on the $m$ bond yields). The unconditional variance-covariance matrix has $m(m + 1)/2$ free elements, thus the base SNP model has $3m(m + 1)/2$ parameters. Experimentation with the two-factor, three-bond model described here revealed that the preferred SNP model is this base model, thus there are 18 SNP parameters to be estimated. If we had chosen to use only two bonds in the estimation, we could find an SNP specification that would require fewer than 18 parameters. But this would require throwing away the information in the other bond’s yield. Duffee (2001) found that, in estimating a flexible three-factor model, using more bonds than factors was vital to produce reasonable-looking estimation results. Thus if we were attempting to fit this model to real-world data, the use of (at least) three bond yields would be prudent.

The true parameters and a summary of the parameter estimates from 150 Monte Carlo simulations are displayed in Table 3. The true parameters are taken from the results of fitting the model to monthly U.S. interest rate data from 1974 through 1998. The parameter $\delta_0$ is not estimated. Instead, it is set to $-0.04$ because the Kalman filter likelihood surface is nearly flat in the direction of changes in $\delta_0$. The EMM simulations (both for the Kalman filter and SNP as auxiliary models) have length 30,000 months (2,500 years of data).

The results from Kalman filter estimation are, on balance, reasonable. The mean estimates of the parameters that are identified under the equivalent martingale measure are very close to their true values. Estimates of the parameters identified under the physical measure exhibit the small-sample bias discussed in Section 4.1. The bias is graphically displayed in Figure 2. The solid lines are the true drifts and the dotted lines are the drifts implied by the Kalman filter estimates. For both factors, the estimated drifts imply faster mean reversion than the true drifts. This bias makes the dynamics of the first factor appear more linear than in the true model, and makes the dynamics of the second factor appear less linear. The mean standard errors are typically close to the standard deviations of the parameter estimates.

The mean parameter estimates from the Kalman filter/EMM are not better than the Kalman filter estimates at reproducing the true drifts. The implied drifts are the dashed lines in Figure 2. The true drift of the first state variable is closer to the Kalman filter drift than to the Kalman filter/EMM drift. This pattern is reversed for the second state variable, but the difference between the two estimated drifts is quite small. The mean Kalman filter/EMM standard errors are very large. The typical mean standard error is about six to seven times the size of the corresponding standard deviation. Moreover, the distribution of standard errors is asymmetric; median standard errors are significantly less than mean standard errors.
By comparison, SNP/EMM performs terribly. The mean parameter estimates are often far away from both truth and the corresponding Kalman filter estimates. The standard deviations of the parameter estimates are very large, indicating that the estimation technique has a hard time pinning down the parameters. The mean standard errors are nowhere near these standard deviations. The distribution of standard errors is strongly asymmetric, with median standard errors well below mean standard errors. The table does not show the overall goodness-of-fit tests associated with SNP/EMM. Of the 150 test statistics produced with the simulations, 26, or 17 percent, exceeded the $\chi^2(5)$ one percent critical value, while 36, or 24 percent, exceeded the five percent critical value.

The clearest message of these results is that the SNP/EMM technique is inappropriate for estimating this kind of dynamic term structure model on a reasonably-sized data sample (both in its time series and cross section). There are too many SNP moment conditions, and the moments are not closely tied to the parameters of interest. The results also indicate that, notwithstanding the misspecification of the linearized Kalman filter, its overall performance is superior to the Kalman filter/EMM. The linearized Kalman filter misspecifies the drift function, but the estimates of the drift function are biased in small samples even for a correctly-specified model. Given the amount of nonlinearity in the “true” model, the EMM correction for misspecification is swamped by the bias. In addition, the Kalman filter/EMM standard errors are so poor that the estimation technique cannot be used to draw statistical inferences. By contrast, the Kalman filter’s standard errors are reasonably accurate. We take these messages to heart in the next section, where we turn our attention to estimating a nonlinear term structure model on real-world data.

5 Estimating a nonlinear term-structure model

5.1 The data

We use month-end yields on five zero-coupon Treasury bonds (interpolated from coupon bonds) from Bliss (1997), who uses the interpolation method of McCulloch and Kwon (1993). The maturities are six months, one year, two years, five years, and ten years. The observed yields, which are annualized and expressed in decimal form, are stacked in the vector $Y_t$.

$$Y_t = \begin{pmatrix} Y_{1/2,t} & Y_{1,t} & Y_{2y,t} & Y_{5,t} & Y_{10,t} \end{pmatrix}'$$

The time period is January 1974 through December 1998, for a total of 300 months. We chose this period to include both recent data and the highly volatile interest rate regime of the late 1970s and early 1980s. We include this volatile period because nonlinear behavior,
if it exists, is more likely to be discovered in a sample of data with a large range.

5.2 A three-factor model

In our three-factor model, the instantaneous interest rate is the sum of a constant and the factors:

\[ r_t = \delta_0 + X_{t,1} + X_{t,2} + X_{t,3} \]  (45)

We restrict our attention to the class of models denoted as \( A_3(3) \) by Dai and Singleton (2000), which means that matrix \( \Sigma S_t \) has the following form:

\[
\Sigma S_t = \begin{pmatrix}
\sigma_i \sqrt{X_{t,1}} & 0 & 0 \\
0 & \sigma_i \sqrt{X_{t,2}} & 0 \\
0 & 0 & \sigma_i \sqrt{X_{t,3}}
\end{pmatrix}
\]

Our primary goal is to estimate a nonlinear model in which the factors are correlated. Dai and Singleton (2000) note that in the linear version of this model, nonzero correlations among factors is necessary to reproduce the hump in the term structure of yield volatilities. The importance of allowing for correlation in this more general model is an open empirical question. Duarte (1999) estimates a version of this model with uncorrelated factors, for which simulation techniques are not necessary.\(^{30}\) Our correlated-factor model requires a ‘feedback’ matrix \( K \) with some nonzero off-diagonal elements. However, to avoid overfitting, we do not want to give the model too much freedom in choosing these off-diagonal elements. We assume that a single off-diagonal element is nonzero.

\[
K = \begin{pmatrix}
\kappa_{11} & 0 & 0 \\
0 & \kappa_{22} & 0 \\
0 & 0 & \kappa_{33}
\end{pmatrix}
\]  (46)

We experimented with allowing \( K \) to be lower triangular. This extension improved the fit of the model, especially in reducing cross-sectional errors at the short end of the yield curve. However, it did not materially affect the estimates of the nonlinear parameters in the vector \( \lambda_1 \).

\(^{30}\)To use his estimation approach (an application of the approximate maximum likelihood technique developed by Ait-Sahalia (1999)) Duarte has to assume the three factors in his model are uncorrelated, and also that certain bond yields are observed without error, while others are observed with some measurement error.
The physical dynamics of $X_t$ are:

$$dX_t = \left[ K\theta + \begin{pmatrix} X_{1,t}^{1/2} \lambda_{11} \\ X_{2,t}^{1/2} \lambda_{12} \\ X_{3,t}^{1/2} \lambda_{13} \end{pmatrix} - K^p X_t \right] dt + \Sigma S_t dZ_t,$$

(47)

where $K^p = K - \text{diag}(\lambda_2)$ refers to the diagonal matrix with $\lambda_2$ along the diagonal.

Certain restrictions on the parameters are necessary for these dynamics to not admit arbitrage opportunities. They are $(K\theta)_i \geq 0 \ \forall \ i$, and nonpositivity of $\kappa_{21}$. Stationarity requires that the diagonal elements of $K^p$ are nonnegative. If any diagonal element of $K^p$ is zero, the corresponding element of $\lambda_1$ must be negative. Finally, the parameter vector $X_t$ must also be nonnegative for all $t$, otherwise elements of $S_t$ are not real.

We close the model by specifying the behavior of measurement error in yields. The variance-covariance matrix of the measurement error, which is $R$ in (17), is assumed to be diagonal. The diagonal elements are $D^2_{1/2}, D^2_1, D^2_2, D^2_3$, and $D^2_{10}$. Thus there are a total of 22 parameters. The specification of $r_t$ and its equivalent-martingale dynamics require eleven parameters, while an additional six price-of-risk parameters determine the physical dynamics. The five measurement-error standard deviations complete the description of observed yields.

### 5.3 Estimation results

We first estimate the model using the misspecified, linearized Kalman filter. The appendix gives details of our estimation procedure, which is complicated owing to the shape of the likelihood surface. The parameter estimates and associated standard errors, in parentheses, are displayed in Table 4. We next estimate the model with EMM using the Kalman filter as an auxiliary model. The resulting parameter estimates are in brackets in the same table.

Two features of the EMM estimation deserve special mention. First, note that two of the parameter estimates from the Kalman filter are on the boundary of their parameter space: $K^p_{33}$ and $D_1$. Therefore the derivatives of the Kalman filter likelihood function with respect to these parameters are not zero. This implies that these derivatives cannot be used as moment conditions in the EMM stage. Thus we have fewer EMM moment conditions than model parameters. Our response is to set the same parameters in the EMM step to their Kalman-filter fixed values. However, this is not a completely satisfactory solution because the auxiliary model is known to be misspecified. It is possible that the estimated parameters
would not lie on their boundaries if a correctly-specified model were used.

Second, determining the solution to the EMM optimization problem is simplified considerably because we do not need to perform a comprehensive search of the parameter space to find the parameters that minimize the EMM objective function. That search was already performed in the Kalman filter stage. As long as the linearized model is sufficiently close to the true model, we can be confident that the parameters that minimize the EMM objective function are in the same local neighborhood as are the parameters that maximize the Kalman filter likelihood function.

In the next subsection we take a detailed look at whether the nonlinear parameters allow this model to fit some important features of bond returns. Here we simply note that the parameter estimates from the Kalman filter are little different from those of EMM using the Kalman filter as an auxiliary model. This indicates that the misspecification in the Kalman filter is not an important concern here. We do not report the standard errors from the Kalman filter/EMM, owing to the evidence in the previous section that they are inferior to those of the Kalman filter.\footnote{It suffices to mention that the Kalman filter/EMM standard errors are substantially different from those of the Kalman filter.}

The parameter estimates and corresponding standard errors lead us to reject independence among the factors ($\kappa_{21}$ is nonzero) and reject linearity in the dynamics of the factors ($\lambda_{13}$ is nonzero). What we cannot evaluate is the magnitude of the bias in the estimates of $\lambda_1$ or $\kappa_0$.

5.4 The price of risk

As noted in Section 3, completely affine models—those for which the vector $\lambda_1$ is identically zero—cannot reproduce the stylized fact that excess bond returns (over short-term interest rates) are positively correlated with the slope of the term structure. Our model, by using the more general form of the price of risk in (31), is, at least in principle, capable of producing this behavior (See Section 3.4.) In this section, we look at whether the model can match the magnitude of the relation that we observe in the data.

Table 5 summarizes the relation, implied by the model, between the slope of the term structure and excess bond returns. Instantaneous expected excess returns to bonds are constructed for each month in the data sample using (35), with the contemporaneous predictions $X_{it}$ used as observations of the state vector. The table reports statistics for expected excess returns to bonds with maturities of six months, two years, and ten years. The model-produced expected excess returns are positively correlated with the slope of the term structure, as measured by the difference between yields on five-year and six-month bonds. However, they are not very volatile. Expected excess returns do not differ much from steep-
slope regimes to low-slope regimes. The evidence discussed in Duffee (2001) documents that actual expected excess returns to Treasury bonds are much more sensitive to the slope of the term structure.

Because the model does not generate much volatility in expected excess returns, it does a poor job forecasting future yields with the slope of the term structure. Campbell and Shiller (1991) note that when the slope of the term structure is steeper than usual, short-term yields subsequently tend to rise, while long-term yields subsequently tend to fall. Table 6 reproduces this result for the 1974–1998 period examined in this paper. The change in a bond yield from month $t$ to month $t + k$ is regressed on the month-$t$ slope of the term structure, for $k = 1, 6$. In the columns labeled “Actual coefficients,” we observe that six-month yields tend to rise when the slope is more steeply sloped, while longer-maturity bond yields tend to fall. As the horizon lengthens from one month to six months, these patterns become more pronounced.

Table 6 also reports the corresponding regression coefficients for yield forecasts from our model. The month $t + k$ bond yield forecasted as of month $t$ is approximated by extrapolating from $X_{t|t}$ using the instantaneous dynamics of (47). Then the change from the month-$t$ yield to the forecasted month $t + k$ yield is regressed on the month-$t$ slope. The model implies that all yields are expected to rise over the next six months when the slope of the term structure is steep. The only prediction of declining long-term yields is at the one-month horizon for the ten-year bond.

We can interpret Tables 5 and 6 in terms of the expectations hypothesis of interest rates. If expected excess returns to bonds are to remain constant over time, an increase in the slope of the term structure must be accompanied by an expected future increase in yields on long-maturity bonds. In the data, this hypothesis is rejected. Long-bond yields tend to fall; their expected excess returns rise substantially. The model we estimate allows for a small increase in their expected excess returns, but not enough to forecast declining yields over the next few months.

The reason behind the failure of the model to produce sufficiently-volatile expected excess returns is essentially that given in the one-factor example in Section 3. Although including the vector $\lambda_1$ in (31) frees up the sign of the vector $\lambda_2$, it does not free up the magnitude of $\lambda_2$.

To see this clearly, focus on the behavior of $X_{t,3}$. This factor drives the slope of the term structure, and changes in expected excess returns are largely driven by this shock. Under the equivalent martingale measure, this shock dies out fairly quickly; $\kappa_{33}$ is approximately one, which implies a half-life of seven to eight months. Under the physical measure, $\kappa_{33}^p$ is estimated to be on the boundary of its parameter space. In other words, given $\kappa_{33}$,
the estimate of \( \lambda_{23} \) is positive and as large as possible. Thus a negative shock to factor three corresponds to higher expected excess returns to long bonds—they are priced as if the negative shock to the instantaneous interest rate dies off quickly, but in fact it is expected to persist. This drives the correlation, in Table 5, between the term-structure slope and expected excess returns.

The additional flexibility afforded by \( \lambda_1 \) in (31) is that \( \lambda_{23} \) can be positive, and hence generate the correct sign of the relation between the slope and expected excess returns. Because \( \lambda_{13} \) is negative, investors, on average, receive compensation for facing the bond-price risk associated with slope shocks. Without \( \lambda_{13} \), a positive value of \( \lambda_{23} \) would make investors pay to face slope risk on average. However, \( \lambda_{23} \) is bounded above by the requirement that \( \kappa_{33}^p \) be nonnegative. This bound makes it impossible for the model to generate a sufficiently large covariance between expected excess returns and the slope to the term structure.

Duarte (1999) comes to a different conclusion using a related model and a different data sample. (His model does not allow for correlations among the factors.) He finds that his estimated model captures the forecast power of the slope for expected excess returns and future changes in yields. The difference between his results and ours are probably the result of differing sample periods. In the 1974–1998 period we examine, the failure of the expectations hypothesis is pronounced in the data. In the 1983-1998 period studied by Duarte, actual term structure dynamics are much closer to those implied by the expectations hypothesis. Thus over his sample, the model’s parameters do not need to be pushed to their limits to capture the empirical dynamics of the term structure slope and expected excess returns. However, the results in our paper document that the limited flexibility offered by the model is insufficient to capture the dynamics of the term structure over a longer time period.

### 6 Concluding comments

In many areas of asset pricing, such as term structure modeling and option pricing, researchers are developing and estimating ever more complex models in an attempt to fit the behavior of multiple asset prices simultaneously. Unfortunately, as the models become more flexible, and as researchers confront them with broader cross-sections of asset prices, their estimation becomes harder. The method that has emerged as the standard estimation technique in such situations is the “Efficient Method of Moments” (EMM) of Gallant and Tauchen (1996), usually used in conjunction with a SNP auxiliary model [See Gallant and Tauchen (1992)]. However, while this is known to have desirable asymptotic properties, little is known about its small-sample properties in these settings.

We investigate these small sample properties in the context of estimating a multifactor
term structure model that (a priori) has the potential to overcome some of the problems that have been noted with prior models. We find that, despite its good asymptotic properties, the usual combination of EMM with a SNP auxiliary model actually behaves extremely poorly in samples of the size and type usual in term structure estimation. We find substantially better results when we use an alternative auxiliary model based on an approximate Kalman filter. However, using the approximate Kalman filter alone, and not as an auxiliary model for EMM, gives better results in our experiments than EMM with either auxiliary model. Thus we advocate the direct use of the linearized Kalman filter as the estimation method of choice; the benefit of more accurate standard errors outweighs the cost of misspecification.

The second contribution of the paper is to document pervasive biases with all the estimation methods considered. This bias leads to rejecting models of the Dai-Singleton class in favor of more flexible alternatives even when using data known to be generated from the more restrictive class. This bias is related to the well-known downward bias in small-sample estimates of the speed of mean reversion of near unit-root processes such as bond yields. Unfortunately, while Ball and Torous (1996) suggested that using a combination of the cross-sectional and time-series information in yields could substantially reduce this bias, we find that this is not the case when estimating models with more general specifications for the price of interest rate risk than they considered.

Finally, we also estimate our model using Treasury data from 1974–1998. While the additional flexibility of the model allows it to fit the behavior of bond yields qualitatively better than the models considered by Dai and Singleton (2000), it is unable to match the magnitude of the relations that we observe in the data. Estimating the model does, however, shed useful light on the kind of flexibility that must be built into the price of risk in order to fit the empirical features of Treasury bond yield and return dynamics.
A  Estimation details

A.1 Initializing the linearized Kalman filter

In the usual Kalman filter setup, an initial $y_{1|0}$ and $V_{1|0}$ are calculated from an initial $x_{0|0}^p$ and $P_{0|0}$, which are in turn calculated using the first two unconditional moments of $x_t$. This method is unavailable to us because we do not have analytic expressions for these moments. We therefore use simulations to approximate them. We generate a long time series of the state vector by discretizing (2), using a time step of $\Delta t$ (the same time step for our discrete observations). Smaller time steps could be used to reduce the approximation error in these moments, but in practice we found that they were unnecessary.

Our method for beginning the Kalman filter recursion requires some justification. There are two obvious alternatives. First, we could set aside an initial set of observations and estimate the unconditional population moments using the unconditional moments of the state implied by this initial set. However, the highly persistent nature of interest rates requires that a very long initial set of observations be used to estimate these moments accurately, and the available time series are short.

Second, we could condition the recursion on the first observation, bypassing the use of unconditional moments. A consequence of this alternative is that the resulting parameter estimates need not imply a stationary process for interest rates. Because of the persistence of interest rates, this is not an unlikely result. Although there is nothing economically or econometrically wrong with implied nonstationarity from such a conditional recursion, it is highly inconvenient. A very useful feature of the Kalman filter as a auxiliary model is that the structure of the auxiliary model is the same as the structure of the model estimated in the simulation stage. Therefore the estimated parameters from the auxiliary model can be used as the initial parameters to search for the parameters of the true model. However, if the auxiliary parameters imply nonstationarity, they will not produce sensible results in the simulation stage, because the simulation methodology imposes stationarity on the underlying process.

A.2 Estimation details for the three-factor model

A number of practical problems arise in estimating the linearized Kalman filter. First, the Kalman filter recursion can produce estimates of the state vector, $\hat{X}_{t|t}$ or $\hat{X}_{t|t-1}$, that violate the state vector’s lower bound of zero. If the contemporaneous prediction, $\hat{X}_{t|t}$, violates the bound, the Kalman filter recursion will fail because $F_{0t}, F_{1t}$, and $Q_t$ will be undefined. We address this problem by evaluating $\mu(X_t)$ and $\sigma(X_t)$ at an ‘adjusted’ $\hat{X}_{t|t}$. The adjustment replaces each negative component in $X_{t|t}$ with a small positive number. We make no other adjustments to the estimated states. In particular, we allow $X_{t|t-1}$ to violate the lower bound.

The second problem is that the Kalman filter log-likelihood function has a large number of local maxima, making estimation problematic. We therefore adopted the following maximization technique, adapted from the procedure in Duffee (2001).

1. Randomly generate a parameter vector from a multivariate normal distribution with a diagonal variance-covariance matrix. The means and variances were arbitrarily set to reasonable values.

2. Using this parameter vector as the starting value, maximize the Kalman filter log-likelihood function using two successive applications of the Simplex algorithm.
To keep the maximization algorithm out of unpromising regions of the parameter space, we set the value of the log-likelihood to a very large negative number if any values of $\hat{X}_{t|t}$ violated the lower bound of zero.

3. If the parameter estimates from the above step produce admissible values of $\hat{X}_{t|t}$ for all $t$, use NPSOL to make any final improvements in the estimates. Otherwise, this step is skipped. In this stage, a few violations of admissibility of $\hat{X}_{t|t}$ are allowed.\(^{32}\)

4. Repeat these steps until NPSOL is used to produce 500 sets of log-likelihood values and parameter estimates. The global maximum from this set is used as the estimate of the linearized Kalman filter.

In the EMM stage, the length of the EMM simulation is 50,000 months. The state vector $X_t$ is simulated by discretizing (47). The discretization interval is 1/60th of a month and the simulations are produced using the weak order-2 method of Kloeden and Platen (1992). The simulation is started from $X_0 = 1$ and the first 500 months are discarded. (We cannot start the simulation with $E(X_t)$ because we do not have an analytic expression for this mean.) This simulation method can produce values of $X_t$ which violate the lower bound of zero. Our algorithm replaced any such values with zero. Given a time series of simulated $X_t$, bond yields are produced using (29).

\(^{32}\)We cannot allow an unlimited number of violations of $\hat{X}_{t|t}$, because then the “optimum” can be one in which all state vectors are entirely negative numbers with zero drift and zero variance. Since negative states are replaced with zeros, the likelihood value is infinite (all forecasts of future states are zero with no variance). In practice, we found that the alternative of allowing no violations resulted in implausible standard errors. We arbitrarily chose to allow five violations. At the optimum, five violations occurred.
References


Gouriéroux, C., A. Monfort, and N. Touzi, 1994, Calibration by simulation for small sample bias correction, Discussion Paper 9554, CREST.


<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Estimates (start near truth)</th>
<th>Estimates (start at truth)</th>
</tr>
</thead>
<tbody>
<tr>
<td>κ</td>
<td>0.29368</td>
<td>0.37807</td>
<td>0.35942</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.00000003)</td>
<td>(0.00013770)</td>
</tr>
<tr>
<td>μ</td>
<td>0.07935</td>
<td>0.11127</td>
<td>0.06404</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.00000000)</td>
<td>(0.00003200)</td>
</tr>
<tr>
<td>σ</td>
<td>0.11425</td>
<td>0.10611</td>
<td>0.10892</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.00000003)</td>
<td>(0.00005381)</td>
</tr>
<tr>
<td>λ</td>
<td>-0.12165</td>
<td>0.00883</td>
<td>-0.18976</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.00000001)</td>
<td>(0.00017812)</td>
</tr>
<tr>
<td>D₁</td>
<td>0.00100</td>
<td>0.00203</td>
<td>0.00010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.00000000)</td>
<td>(0.00000024)</td>
</tr>
<tr>
<td>χ²(44)</td>
<td>-</td>
<td>7.2 × 10¹⁰</td>
<td>1557.8</td>
</tr>
</tbody>
</table>

Table 1: Estimating a one-factor CIR model using SNP/EMM

Twenty years of monthly observations of instantaneous interest rates and one-year bond yields are generated by a one-factor CIR process. The one year bond yield is observed with iid measurement error (with standard deviation $D₁$), and the data are fit to a one-factor CIR interest rate model using EMM with an SNP auxiliary model. This table summarizes the results of estimation starting from two different sets of starting values. The first set are close to, but not equal to, the true parameter values. The second set are the true parameter values.
Panel A. Kalman filter estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Mean Estimate</th>
<th>Std. Dev. of Estimates</th>
<th>Mean Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa \theta$</td>
<td>0.012</td>
<td>0.012</td>
<td>0.0006</td>
<td>0.0007</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.098</td>
<td>0.098</td>
<td>0.0110</td>
<td>0.0106</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.073</td>
<td>0.072</td>
<td>0.0038</td>
<td>0.0035</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.000</td>
<td>0.111</td>
<td>0.1173</td>
<td>0.1007</td>
</tr>
<tr>
<td>$\kappa^p$</td>
<td>0.192</td>
<td>0.656</td>
<td>0.4614</td>
<td>0.4079</td>
</tr>
<tr>
<td>$D_r \times 10^3$</td>
<td>1.000</td>
<td>0.987</td>
<td>0.1930</td>
<td>0.1840</td>
</tr>
<tr>
<td>$D_1 \times 10^3$</td>
<td>1.000</td>
<td>0.977</td>
<td>0.1864</td>
<td>0.1710</td>
</tr>
</tbody>
</table>

Panel B. Kalman filter/EMM estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Mean Estimate</th>
<th>Std. Dev. of Estimates</th>
<th>Mean Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa \theta$</td>
<td>0.012</td>
<td>0.012</td>
<td>0.0015</td>
<td>0.0007</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.098</td>
<td>0.098</td>
<td>0.0110</td>
<td>0.0115</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.073</td>
<td>0.074</td>
<td>0.0039</td>
<td>0.0040</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.000</td>
<td>0.114</td>
<td>0.1196</td>
<td>0.1559</td>
</tr>
<tr>
<td>$\kappa^p$</td>
<td>0.192</td>
<td>0.664</td>
<td>0.4683</td>
<td>0.6704</td>
</tr>
<tr>
<td>$D_r \times 10^3$</td>
<td>1.000</td>
<td>0.984</td>
<td>0.1886</td>
<td>0.1980</td>
</tr>
<tr>
<td>$D_1 \times 10^3$</td>
<td>1.000</td>
<td>0.984</td>
<td>0.1868</td>
<td>0.1848</td>
</tr>
</tbody>
</table>

Table 2: Fitting CIR data to a nonlinear model: Monte Carlo simulation results

Twenty years of monthly observations of instantaneous interest rates and one-year bond yields are generated by a one-factor CIR process. The data, which are observed with iid measurement error (with standard deviations $D_r$ and $D_1$), are fit to a nonlinear term-structure model described by equations (36) and (37) in the paper, using EMM with a linearized Kalman filter as the auxiliary model. This table summarizes the results of 250 Monte Carlo simulations.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Kalman filter</th>
<th>Kalman filter/EMM</th>
<th>SNP/EMM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Mean</td>
<td>Std. Error Estimate</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>Mean</td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>Mean</td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>Mean</td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td>$\kappa_1$</td>
<td>0.040</td>
<td>0.041</td>
<td>0.0064</td>
<td>0.0064</td>
</tr>
<tr>
<td></td>
<td>0.0055</td>
<td>0.041</td>
<td>0.0065</td>
<td>0.0499</td>
</tr>
<tr>
<td></td>
<td>0.0064</td>
<td>0.035</td>
<td>0.0221</td>
<td>0.2047</td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>0.750</td>
<td>0.748</td>
<td>0.0297</td>
<td>0.0291</td>
</tr>
<tr>
<td></td>
<td>0.0283</td>
<td>0.745</td>
<td>0.2136</td>
<td>0.0340</td>
</tr>
<tr>
<td></td>
<td>0.0177</td>
<td>0.758</td>
<td>0.3067</td>
<td>1.1418</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.100</td>
<td>0.098</td>
<td>0.0108</td>
<td>0.0130</td>
</tr>
<tr>
<td></td>
<td>0.0120</td>
<td>0.100</td>
<td>0.0112</td>
<td>0.1515</td>
</tr>
<tr>
<td></td>
<td>0.0177</td>
<td>0.175</td>
<td>0.1373</td>
<td>0.9002</td>
</tr>
<tr>
<td>$\lambda_{11}$</td>
<td>-0.200</td>
<td>-0.116</td>
<td>0.1195</td>
<td>0.0921</td>
</tr>
<tr>
<td></td>
<td>0.0975</td>
<td>-0.113</td>
<td>0.1212</td>
<td>0.2751</td>
</tr>
<tr>
<td></td>
<td>0.0316</td>
<td>-0.100</td>
<td>0.5596</td>
<td>3.5119</td>
</tr>
<tr>
<td>$\kappa_p$</td>
<td>0.050</td>
<td>0.502</td>
<td>0.6099</td>
<td>0.5258</td>
</tr>
<tr>
<td></td>
<td>0.5303</td>
<td>0.657</td>
<td>0.6473</td>
<td>0.7849</td>
</tr>
<tr>
<td></td>
<td>0.1242</td>
<td>0.578</td>
<td>2.8484</td>
<td>15.5495</td>
</tr>
<tr>
<td>$\kappa_2 \times 10^2$</td>
<td>0.220</td>
<td>0.214</td>
<td>0.0402</td>
<td>0.0411</td>
</tr>
<tr>
<td></td>
<td>0.0381</td>
<td>0.212</td>
<td>0.0400</td>
<td>0.1896</td>
</tr>
<tr>
<td></td>
<td>0.0395</td>
<td>0.258</td>
<td>0.1913</td>
<td>0.7954</td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>0.016</td>
<td>0.015</td>
<td>0.0056</td>
<td>0.0055</td>
</tr>
<tr>
<td></td>
<td>0.0051</td>
<td>0.015</td>
<td>0.0057</td>
<td>0.0258</td>
</tr>
<tr>
<td></td>
<td>0.0043</td>
<td>0.013</td>
<td>0.0252</td>
<td>0.0635</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.050</td>
<td>0.050</td>
<td>0.0052</td>
<td>0.0053</td>
</tr>
<tr>
<td></td>
<td>0.0042</td>
<td>0.051</td>
<td>0.0055</td>
<td>0.0418</td>
</tr>
<tr>
<td></td>
<td>0.0060</td>
<td>0.057</td>
<td>0.0318</td>
<td>0.1014</td>
</tr>
<tr>
<td>$\lambda_{12}$</td>
<td>0.060</td>
<td>0.147</td>
<td>0.1014</td>
<td>0.0878</td>
</tr>
<tr>
<td></td>
<td>0.0831</td>
<td>0.136</td>
<td>0.0903</td>
<td>0.3002</td>
</tr>
<tr>
<td></td>
<td>0.0324</td>
<td>0.709</td>
<td>2.3167</td>
<td>2.5582</td>
</tr>
<tr>
<td>$\kappa_p$</td>
<td>0.276</td>
<td>0.655</td>
<td>0.4183</td>
<td>0.3546</td>
</tr>
<tr>
<td></td>
<td>0.3492</td>
<td>0.616</td>
<td>0.3857</td>
<td>0.7214</td>
</tr>
<tr>
<td></td>
<td>0.1384</td>
<td>3.225</td>
<td>12.7729</td>
<td>8.1993</td>
</tr>
<tr>
<td>$D_{1/2} \times 10^3$</td>
<td>1.000</td>
<td>0.957</td>
<td>0.2260</td>
<td>0.2518</td>
</tr>
<tr>
<td></td>
<td>0.2244</td>
<td>0.990</td>
<td>0.2216</td>
<td>1.8320</td>
</tr>
<tr>
<td></td>
<td>0.3166</td>
<td>1.017</td>
<td>0.3797</td>
<td>0.8808</td>
</tr>
<tr>
<td>$D_2 \times 10^3$</td>
<td>1.000</td>
<td>0.992</td>
<td>0.0934</td>
<td>0.1037</td>
</tr>
<tr>
<td></td>
<td>0.1024</td>
<td>0.993</td>
<td>0.0946</td>
<td>0.9332</td>
</tr>
<tr>
<td></td>
<td>0.1354</td>
<td>0.927</td>
<td>0.1996</td>
<td>0.5352</td>
</tr>
<tr>
<td>$D_{10} \times 10^3$</td>
<td>1.000</td>
<td>0.999</td>
<td>0.1118</td>
<td>0.1258</td>
</tr>
<tr>
<td></td>
<td>0.1259</td>
<td>1.000</td>
<td>0.1111</td>
<td>1.5538</td>
</tr>
<tr>
<td></td>
<td>0.1918</td>
<td>1.026</td>
<td>0.1983</td>
<td>0.2427</td>
</tr>
</tbody>
</table>

Table 3: A comparison of estimation methods: Monte Carlo simulation results

Twenty years of monthly observations of six-month, two-year, and ten-year zero-coupon bond yields are generated by a nonlinear, two-factor term-structure model. The data, which are observed with iid measurement error (with standard deviations $D_{1/2}, D_2,$ and $D_{10}$), are fit to the model using two methods. The first is a misspecified Kalman filter, which uses instantaneous dynamics in place of one-month-ahead dynamics. It also begins the filter with simulated unconditional moments of the states instead of the analytic unconditional moments. The second is SNP/EMM. This table summarizes the results of 150 Monte Carlo simulations.
Table 4: Estimates of a three-factor nonlinear term-structure model

Monthly yields on zero-coupon Treasury bonds (interpolated from coupon bonds) are fit to the three-factor model summarized by equations (45) through (47). The sample period is January 1974 through December 1998. A linearized Kalman filter produced the parameter estimates and associated standard errors in parentheses. Using the Kalman filter as an auxiliary model, EMM produced the parameter estimates in brackets. If there is no standard error, the parameter is set to zero by assumption. If the standard error is a dashed line, the parameter was estimated to be on the boundary of its parameter space.
Table 5: Summary of monthly fitted values of excess returns to bonds, 1974–1998

The term-structure model of Section 5 is used to construct implied instantaneous expected excess (over the instantaneous interest rate) returns to zero-coupon Treasury bonds from January 1974 through December 1998. The table reports summary statistics for these returns, which are annualized and expressed in percent. The slope of the term structure is measured by the difference between the five-year and the six-month zero-coupon yields. To produce the columns labeled “slope-sorted quartile means,” expected excess returns are sorted into quartiles based on the contemporaneous slope of the term structure. Means of the first and fourth quartiles are reported above.
Changes in zero-coupon bond yields from month $t$ to month $t+n$ are regressed on the slope of the yield curve in month $t$. The regression coefficients for monthly data from January 1974 through December 1998 are reported in the “Actual coefficients” columns. In addition, the term-structure model of Section 5 is used to construct forecasts, in month $t$, of month $t+k$ bond yields. The same regression is run, with the forecasted yields used in place of the actual month $t+k$ yields. The slope of the term structure is measured by the difference between the five-year and the six-month zero-coupon yields.

<table>
<thead>
<tr>
<th></th>
<th>One month ahead</th>
<th></th>
<th>Six months ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actual coef</td>
<td>Model-implied coef</td>
<td>Actual coef</td>
</tr>
<tr>
<td>6 months</td>
<td>0.029</td>
<td>0.116</td>
<td>0.144</td>
</tr>
<tr>
<td>2 yrs</td>
<td>-0.014</td>
<td>0.106</td>
<td>-0.081</td>
</tr>
<tr>
<td>5 yrs</td>
<td>-0.036</td>
<td>0.002</td>
<td>-0.188</td>
</tr>
<tr>
<td>10 yrs</td>
<td>-0.044</td>
<td>-0.065</td>
<td>-0.225</td>
</tr>
</tbody>
</table>

Table 6: Regressions of changes in bond yields on the slope of the term structure, 1974–1998
Figure 1: Actual and estimated drifts for a one-factor term structure model.

The data are generated by a CIR model for which the drift in the instantaneous interest rate is given by the solid line. The dotted line displays the drift implied by Kalman filter parameter estimates of a nonlinear term structure model. The dashed line displays the drift implied by Kalman filter/EMM parameter estimates.
Figure 2: Actual and estimated drifts for a two-factor, nonlinear interest rate model.

The true drifts are given by the solid lines. The dotted lines display the drifts implied by Kalman filter parameter estimates. The dashed lines display the drifts implied by Kalman filter/EMM parameter estimates.