EMM Estimation of Affine and Nonaffine 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
email: duffee@haas.berkeley.edu (Duffee)
email: stanton@haas.berkeley.edu (Stanton)

This Draft: November 18, 2000

ABSTRACT

We use the Efficient Method of Moments (EMM) of Gallant and Tauchen (1996) to estimate a three factor term structure model which is affine under the risk-neutral probability distribution, but non-affine under the true probability distribution. Unlike most previous research, in which the model is affine under both distributions, this allows us to retain the analytical convenience for pricing of this class of model, while allowing greater flexibility in matching the observed time-series properties of interest rates. We find that the fully affine specification is statistically rejected in favor of the more flexible alternative. We also shed new light on the implementation of EMM for estimating models using high dimensional, very persistent series such as term structure models. We find, in particular, that the auxiliary model most commonly used in conjunction with EMM, the SNP model of Gallant and Tauchen (1992), has serious problems in this environment, and that substantially better results are obtained using moment conditions derived from an approximate maximum likelihood estimator based on the Kalman filter.

JEL classification: G12, G13, G23.

*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.

Estimation of dynamic term structure models typically involves minimizing a distance between certain population moments implied by the model and sample estimates of those moments calculated from the data.[1] This estimation is complicated by the fact that analytic moment conditions are often hard (or impossible) to calculate. While it is possible in certain cases to calculate analytic approximations to true moment conditions [see, for example, Duffie, Pedersen, and Singleton (2000) and Aït-Sahalia (1999)], in general numerical procedures such as Monte Carlo simulation must be used to calculate approximations to the true moments.

Not surprisingly, the choice of moment conditions is important. The first contribution of our paper concerns this choice of moments. The most common current approach to simulation-based estimation is the Efficient Method of Moments (EMM), which uses moments that summarize substantial information about the conditional distribution of the data without imposing any structure from the model of interest. This approach is embodied in Gallant and Tauchen (1996) and implemented in Dai and Singleton (1998), where moment conditions from a semi-nonparametric (SNP) description of the data are used. However, despite the attractive asymptotic behavior of this approach [see, for example, Gallant and Long (1997)], we find that it may perform poorly in practice when used for modeling interest rate behavior. Instead, we advocate the use of moments from a linearized Kalman filter for both simulation-based methods and as analytic approximations to true moments.

The problem with the "agnostic" SNP approach is that it requires the use of a very large number of moment conditions, since 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 large cross-section without imposing much structure, a very large number of moment conditions will be required. If, say, we observe n bond yields at each point in time, denoted $y_{i,t}$, then using only unconditional first and second moments requires n^2 combinations of $E(y_{i,t+1}|y_{j,t})$ and another $n^2(n+1)/2$ combinations of $E(y_{i,t+1}y_{j,t+1}|y_{k,t})$.

[1] This description includes maximum likelihood, where the moments used are the score vector corresponding to the likelihood function.

In small samples, the use of a large number of moment conditions is problematic. The small-sample properties of GMM and EMM estimators can differ, sometimes dramatically, from the asymptotic properties of these estimators.[2] Unfortunately, the time series of interest rates available to econometricians are short, especially when we take into account the well-documented persistence of interest rates.

Accordingly, we advocate that the information in a panel of interest rates should be compressed by first imposing the no-arbitrage restrictions implied by the term-structure model of interest. These restrictions are built into the linearized Kalman filter through parametric restrictions on the filter's observation and transition equations.[3] We use Monte Carlo simulations to compare the combination of the Kalman filter/EMM to the combination of SNP/EMM. For the models examined in this paper, and for reasonable sample sizes, the results strongly support the choice of the Kalman filter.

Due primarily to their flexibility, combined with the computational ease with which they can handle multiple factors, much of the academic literature has focused largely on affine models characterized by Duffie and Kan (1996). In this class, risk-neutral dynamics are specified such that bond yields are linear (affine) functions of some underlying set of state variables. Until very recently, the term-structure literature followed Vasicek (1977) and Cox, Ingersoll, and Ross (1985) by augmenting these risk-neutral dynamics with restrictive functional forms for market prices of risk. Essentially, the price of risk is assumed to be a multiple of interest rate volatility, which implies that interest-rate dynamics are affine under the physical measure as well as the risk-neutral measure. This class of models was fully characterized by Dai and Singleton (1998). Recent research has highlighted the limitations of this class of affine models. For example, Ghysels and Ng (1998) and Balduzzi and Eom (1997) find evidence against particular affine specifications using semiparametric tests. Duffee (1999) finds that these restrictive forms are too inflexible to generate 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 (1996b), Conley, Hansen, Luttmer, and Scheinkman (1997), and Stanton (1997)] that is inconsistent with the predictions of the affine models that have been estimated to date.

Partly in response to these empirical problems, several non-affine term structure models have recently been proposed [see, for example, Longstaff (1989) Aït-Sahalia (1996a), Aït-Sahalia (1996b), Stanton (1997), Boudoukh, Richardson, Stanton, and Whitelaw (1998), Andersen and Lund (1996b), Constantinides (1992), Conley, Hansen, Luttmer, and Scheinkman

[2] See, for example, Zhou (1999), Hansen, Heaton, and Yaron (1996), Chumacero (1997), Andersen, Chung, and Sørensen (1999), and Andersen and Sørensen (1996).

[3] If we take an agnostic approach to moment selection, such as using SNP, these restrictions are not imposed when choosing moments; they are only imposed when data are simulated.

(1997)]. While these are able to overcome some of the empirical drawbacks of existing affine models, most unfortunately lack the analytical tractability of these models. This is not the case for the recent quadratic term structure models proposed by, for example, Ahn, Dittmar, and Gallant (2000) and Leippold and Wu (2000), but these require the estimation of many more parameters than the affine models.

In this paper, we take an alternative, more parsimonious route, first suggested by Dai and Singleton (1998). We retain the affine form of Duffie and Kan, and generalize the functional form for the price of risk. Unlike Duffee (1999), which also generalizes this form, we here describe the price of risk in a way that produces nonlinear interest-rate dynamics under the physical measure.[4] We use this model to address two questions. First, how does EMM perform with different auxiliary models? Second, what features must be built into a term structure model in order for it to generate real-world interest-rate dynamics?

For all but pure Gaussian term-structure models, the linearized Kalman filter is a misspecified description of the dynamic behavior of yields. This misspecification motivates the use of this filter as an auxiliary model for use with Gallant and Tauchen's EMM methodology. However, we find that for the models that we analyze, the magnitude of the misspecification is small for reasonable parameterizations. Monte Carlo evidence presented here suggests that the combination of the Kalman filter and EMM does not result in more accurate parameter estimates than does the Kalman filter alone. Put differently, the approximation error introduced by using the analytic Kalman filter moment conditions is not larger than is the approximation error introduced by simulating moment conditions with EMM.

Another contribution of our paper relates 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 the combination of the cross-sectional and time-series information in yields could be used to substantially reduce the small-sample bias in estimates of drifts of bond yields. We argue here that their conclusion relies on the assumption of restrictive functional forms for the price of risk. These restrictions closely link the cross-sectional behavior of yields to their time-series behavior. When we use more general forms, such as those estimated in this paper, these links are weakened, and the small-sample bias contaminates estimates of drifts even when information in both the time-series and cross-section is used.

This small-sample bias also affects statistical tests of the more-restrictive functional forms for the price of risk. We use Monte Carlo simulations to measure the magnitude of this contamination. For typical sample sizes, we find that standard econometric techniques fre-

[4]Duarte (1999) has independently pursued a similar idea, although he focuses on issues that are largely different from those considered here.

quently reject the restrictive models of the price of risk in favor of more general models, even through the data are generated by the more restrictive models. In particular, linear models are often rejected in favor of nonlinear models because in finite samples, the nonlinear parameters are biased towards more mean reversion, and thus away from zero.

The final contribution of this paper is to explain why the form for the price of risk examined in this paper, although more general than standard models, 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 EMM/SNP 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. Section3 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.

2 Estimation

2.1 Estimating Continuous-Time Models

Given a general diffusion model

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dZ_t, \tag{1}$$

the transition density from value x at time t to value y at some later time s, $p(s, y \mid t, x)$, must satisfy the Kolmogorov forward equation,

$$\frac{\partial p(s, y \mid t, x)}{\partial s} = -\frac{\partial}{\partial y}\left(\mu(y)p(s, y \mid t, x)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y)p(s, y \mid t, x)\right), \tag{2}$$

and the Kolmogorov backward equation (see Øksendal (1985)),

$$-\frac{\partial p(s, y \mid t, x)}{\partial t} = \mu(x)\frac{\partial}{\partial x}\left(p(s, y \mid t, x)\right) + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}\left(p(s, y \mid t, x)\right). \tag{3}$$

In principle, for a given parametrization of μ and σ, we can solve equation (2) for the conditional density p as a function of the parameters, then use maximum likelihood to estimate

the model's parameters (see, for example, Lo (1988)). This approach was followed by Pearson and Sun (1994) in estimating the parameters of a 2 factor CIR interest rate model, using the fact that, under this process, interest rates are conditionally distributed as a multiple of a non-central χ^2 random variable [see Feller (1951)]. However, except in rare cases, we can only solve equation (2) numerically, making this technique very burdensome. Recently several authors have extended the range of problems to which we can apply maximum likelihood. Pedersen (1995) and Santa-Clara (1995) [see also Brandt and Santa-Clara (1999)] 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. This method can only be applied, however, when the variables being estimated are observed without error. Aït-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).

As an alternative approach, the Generalized Method of Moments (GMM) of Hansen (1982) often allows us to estimate models when either the full likelihood function is too complicated or time-consuming to calculate, or where we wish to specify only certain properties of the distribution, rather than the full likelihood function [see, for example, Gibbons and Ramaswamy (1993)]. Sometimes we can calculate the moment conditions in closed form. For example, for fairly general diffusion processes, Hansen and Scheinkman (1995) show how to derive certain analytic moment restrictions from equation (1), but these do not take advantage of all of the information contained in the discretely observed data. Chan et al. (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

$$\begin{aligned} E[\Delta X_t] &\approx \mu(X_t)\,\Delta t, \\ \mathrm{var}[\Delta X_t] &\approx \sigma^2(X_t)\,\Delta t. \end{aligned}$$

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.

One issue in estimating term structure models is the exact observability of the data. In general, given an n factor model, and a set of parameter values, we can pick any n points on the yield curve, and invert the appropriate pricing equations to calculate exactly

the implied values of the underlying state variables.[5] This was used to allow maximum likelihood estimation by Pearson and Sun (1994), Chen and Scott (1993), and Duffie and Singleton (1997). However, when we have $m > n$ bond yields, there is in general no set of values for the n factors that exactly matches the m bond yields. One solution is to assume that n of the yields are measured without error, and allow for some sort of "measurement error" in the remaining yields [as in Chen and Scott (1993), and early versions (though not the published version) of Pearson and Sun (1994)]. This has the disadvantage that the specification of the yields estimated without error is, necessarily, somewhat ad hoc. An alternative approach is to allow all the yields to be measured with error, but it now becomes impossible to observe the underlying state variables exactly. Instead, some sort of filtering must be performed to calculate an expected value for these variables, given current and past values of the bond yields. A natural approach to this problem is the use of the linear Kalman Filter [see, for example, Harvey (1989) or Hamilton (1994)]. This has been used in a term structure context by, among others, Pennacchi (1991), Chen and Scott (1995), Duan and Simonato (1997), de Jong (1998), Geyer and Pichler (1999), and Jegadeesh and Pennacchi (1996). The standard Kalman filter assumes normally distributed dynamics for both the true state variables and measurement error(s). In addition, it assumes that the variance matrix is constant. These conditions hold only in the case of a multifactor version of the Vasicek (1977) model [see, for example, Babbs and Nowman (1999)], but not for other affine models, where volatility is stochastic, and conditional distributions are non-normal. This makes estimates obtained directly from the Kalman filter approach inconsistent, though there is Monte Carlo evidence that this may be of limited importance in practice [see, for example, de Jong (1998) and Duan and Simonato (1997)].

An alternative approach is to use simulation 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)]. There remains, however, the question of which moments to simulate. We would like to use the vector of derivatives of the likelihood function (score vector) of the data, as we would then be calculating maximum likelihood estimators. In general we cannot do this, but we can come close using the Efficient Method of Moments (EMM) approach of Gallant and Tauchen (1996). They use moment conditions derived from an auxiliary semi-nonparametric (SNP) description of the data, and Gallant and Long (1997) show that this approach asymptotically attains the same efficiency as maximum likelihood [see also Gallant and Tauchen (1999)].[6] This has led many authors to adopt this procedure. For example, in an interest

[5] This is true of 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.

[6] This is closely related to the "indirect inference" approach of Gourieroux, Monfort, and Renault (1993).

rate context, see Dai and Singleton (1998), Andersen and Lund (1997), and Andersen and Lund (1996a), and for estimation of stochastic volatility models see, for example, Andersen, Chung, and Sørensen (1999). In our case, this approach has several advantages. First, we observe zero-coupon bond yields at discrete intervals, and want to use an estimation technique that exploits the information in the conditional density of these discretely-observed yields. The nonlinear physical dynamics of the state vector prevent us from obtaining analytic expressions related to this density. This is not a problem with a simulation based approach such as EMM. Second, we would like to use multiple points on the term structure without necessarilyassuming that a particular subset is measured without error. Again, EMM allows us to include measurement error in all bond yields without additional complication. We now discuss the estimation methodology in more detail. We shall implement it using two different choices of auxiliary model.

2.2 EMM

The "Efficient Method of Moments" of Gallant and Tauchen (1996) is a simulation based GMM estimator which uses the expectation of the score vector from some auxiliary model as the vector of moment conditions for GMM. To the extent that the auxiliary model nests the structural model being estimated, the estimator achieves the same asymptotic efficiency as maximum likelihood [see Gallant and Long (1997)], hence the name of the approach. For details see Gallant and Tauchen (1996) and the user manual, Gallant and Tauchen (1998a), on which this discussion is based.

Let $\{\tilde{y}_t, \tilde{x}_{t-1}\}_{t=1}^n$, be the observed dataset, where

$$\tilde{x}_{t-1} = \left(\tilde{y}'_{t-L}, \ldots \tilde{y}'_{t-1}\right)',$$

for some $L \geq 1$. The first step in EMM estimation is quasi-maximum likelihood of the auxiliary model (also referred to as the "score generator"), using the observed dataset,

$$\tilde{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}} \frac{1}{n} \sum_{t=1}^{n} \ln f(\tilde{y}_t \mid \tilde{x}_{t-1}, \theta).$$

To estimate the parameters of the structural model, ρ, GMM is now performed using (simulated) expectations of this score vector to generate moment conditions, i.e., the moment

This also uses an auxiliary model, but matches estimated parameter values, rather than score vectors, making it in general much more computationally burdensome.

vector is

$$m_n(\rho, \tilde{\theta}_n) = \frac{1}{N} \sum_{\tau=1}^{N} \frac{\partial}{\partial \theta} \ln f[\hat{y}_\tau(\rho) \mid \hat{x}_{\tau-1}(\rho), \tilde{\theta}_n],$$

where N is the size of the simulated dataset. The EMM estimator of ρ is then

$$\hat{\rho}_n = \underset{\rho \in \mathcal{R}}{\operatorname{argmin}}\, m_n'(\rho, \tilde{\theta}_n) \tilde{I}_n^{-1} m_n(\rho, \tilde{\theta}_n),$$

where $\tilde{I}_n^{-1}$ is the weighting matrix.

In implementing EMM as described above, there remains the question of what auxiliary model to use. The only real requirement is that the auxiliary model should provide a good statistical description of the distribution of the data. We shall consider two different auxiliary models. One is the now-standard SNP model of Gallant and Tauchen (1992); the other is an approximate maximum likelihood estimation of a linearized version of the true model using a Kalman filter.[7]

2.3 Auxiliary Model 1 - SNP

The standard auxiliary model for use with EMM has become the SNP (for **S**emi**N**on**P**arametric) model of Gallant and Tauchen (1992) [see also the user's guide, Gallant and Tauchen (1998b)]. It consists of writing the conditional density of the dataset under analysis in the form of a Hermite polynomial multiplied by a normal density, i.e.,

$$f(\tilde{y}_t \mid \tilde{x}_{t-1}, \rho) = c(\tilde{x}_{t-1}) \left[h(\tilde{z}_t \mid \tilde{x}_{t-1})\right]^2 \phi(\tilde{z}_t),$$

where

- $\phi(.)$ represents the standard normal p.d.f.,
- $h(\tilde{z}_t \mid \tilde{x}_{t-1})$ is a Hermite polynomial in $\tilde{z}_t$,
- $c(\tilde{x}_{t-1})$ is a normalization constant (equal to $1/\int \left[h(s \mid \tilde{x}_{t-1})\right]^2 \phi(s)\, ds$), and
- $\tilde{z}_t$ is a normalized version of $\tilde{y}_t$, defined by

$$\tilde{z}_t = R_{x,t-1}^{-1} \left(\tilde{y}_t - \mu_{x,t-1}\right),$$

where $\mu_{x,t-1}$ is the conditional mean, and $R_{x,t-1}^{-1}$ the Cholesky decomposition of the conditional variance of $\tilde{y}_t$. This specification allows great flexibility in fitting the conditional distribution. In particular, we are free to choose:

[7] This auxiliary model was also suggested by Duan and Simonato (1997).

- the dimensionality of the Hermite polynomial in z, K_z (allows for non-Gaussian behavior).
- the degree of the polynomial in x that makes up each of the coefficients in the Hermite polynomial, K_x (another way to allow for conditional heterogeneity).
- the number of lags of x in the Hermite polynomial, K_p.
- the number of lags in a VAR specification for $\mu_{x,t-1}$, L_μ.
- the degree of an ARCH (or GARCH, setting $L_g > 0$) specification for the scale transformation $R_{x,t-1}$, 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.

2.3.1 Example - One Factor CIR model

As an example of this procedure, 1,000 weekly observations (approximately 20 years) on the instantaneous riskless interest rate were simulated from a standard one factor Cox, Ingersoll, and Ross (1985) model,

$$dr_t = \kappa(\mu - r_t)\,dt + \sigma\sqrt{r_t}\,dZ, \tag{4}$$

where the price of interest rate risk is given by the parameter λ. The parameter values used for the simulation were

$$\begin{aligned} \kappa &= 0.29368, \\ \mu &= 0.07935, \\ \sigma &= 0.11425, \\ \lambda &= -0.12165, \end{aligned}$$

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 χ^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 $\sigma_\epsilon = 0.0001$, making a total of five parameters determining the joint distribution of the two series, $\{r_t, y_t^1\}$.

The first step in estimating κ, μ, σ, λ and σ_ϵ was to determine the appropriate SNP specification using the search methodology described in Gallant and Tauchen (1998b). The

optimal specification was 30314411, corresponding to

$$\begin{aligned}
L_u &= 3 \quad \text{(order of VAR)},\\
L_g &= 0 \quad \text{(order of GARCH)},\\
L_r &= 3 \quad \text{(order of ARCH)},\\
L_p &= 1 \quad \text{(lags in Hermite polynomial)},\\
K_z &= 4 \quad \text{(degree of Hermite polynomial)},\\
I_z &= 4 \quad \text{(\# cross terms to suppress)},\\
K_x &= 1 \quad \text{(order of polynomial in Hermite polynomial coefficients)},\\
I_x &= 1 \quad \text{(\# cross terms to suppress)}.
\end{aligned}$$

The model was estimated using EMM/SNP, with two different sets of starting values. The first estimation was performed starting from parameter values (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. Table 1 shows that at the end of the estimation, the estimated parameter values are substantially different from the true values, the reported standard errors are essentially zero, and the true parameter values are way outside any reasonable confidence intervals one might construct (for example, the estimated value for κ is approximately 2 million standard errors from the true parameter value!). In addition, the $\chi^2(44)$ test statistic has the value 7.2×10^{10}. The model was reestimated, this time 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 μ 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.

This one example suggests that, at least in this case, we seem to be obtaining biased parameter estimates and standard errors, overstated χ^2 statistics, and that the final parameter estimates are highly dependent on where the estimation starts from.[8] This in turn suggests that the small sample properties of the EMM/SNP combination, when used in conjunction with datasets of the size and type we actually see in practice, may not be nearly as attractive as their asymptotic properties, motivating a more thorough investigation of these small sample properties, and the search for a possibly better behaved auxiliary model than SNP. The alternative auxiliary model we shall investigate is the linearized Kalman filter.

[8] Another test with 100 *years* of weekly data produced very similar results.

2.4 Auxiliary Model 2 - Linearized Kalman filter

We first briefly review the standard Kalman filter. For details see, for example, Harvey (1989) or Hamilton (1994). At discrete time intervals $t = 1, \ldots, T$, a vector m-vector Y_t is observed. We refer to the entire panel of observations as Y. The observables depend on an underlying state n-vector X_t. The 'observation equation' expresses Y_t as linear in X_t and the 'transition equation' expresses the discrete-time evolution of X_t as linear in X_t. These equations determined by some underlying parameter vector ρ.

$$
\begin{aligned}
Y_t &= H_0(\rho) + H_1(\rho)' X_t + w_t; && (5)\\
X_{t+1} &= F_0(\rho) + F_1(\rho) X_t + v_{t+1}; && (6)\\
E(w_t) &= 0; && (7)\\
E(v_{t+1}) &= 0; && (8)\\
E(w_t w_t') &= R(\rho); && (9)\\
E(v_{t+1} v_{t+1}') &= Q(\rho). && (10)
\end{aligned}
$$

The vectors and matrices in (5) through (10) are functions only of ρ, 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 $\hat{X}_{t+1|t}$ and $\hat{Y}_{t+1|t}$, 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 $\hat{X}_{t|t}$ and $P_{t|t}$, respectively.

Quasi-maximum likelihood estimation using the standard Kalman filter is straightforward. An initial $\hat{Y}_{1|0}$ and $V_{1|0}$ are calculated from an initial $\hat{X}_{0|0}$ and $P_{0|0}$, which are in turn calculated using the first two unconditional moments of X_t. Then the quasi log-likelihood value associated with $Y_t, t = 1, \ldots, T$ is

$$L(\rho, Y) = \sum_{t=1}^{T} l_t,$$

$$l_t = -\frac{1}{2}[m \log(2\pi) + \log |V_{t|t-1}| + (Y_t - \hat{Y}_{t|t-1})' V_{t|t-1}^{-1} (Y_t - \hat{Y}_{t|t-1})].$$

The estimated parameter vector $\hat{\rho}$ is chosen to maximize L.

This Kalman filter approach is best suited for models that express zero-coupon bond yields as affine functions of an underlying state variable. Such models, such as those in

the Duffie and Kan (1996) class, fit naturally into the linear structure of (5), if we use zero-coupon bonds as the observable vector Y_t. The Kalman filter structure then implies that yields are observed with serially uncorrelated measurement error w_t with a constant variance-covariance matrix R. If the dynamics of the state vector are nonlinear, there will be no exact analog to (6). For these nonlinear models, a natural alternative is to replace (6) with a linearization of (1) around $\hat{X}_{t|t}$. The time between discrete observations is denoted Δt. The linearization is

$$X_{t+1} = F_{0t} + F_{1t}X_t + v_{t+1}; \tag{11}$$

$$F_{0t} = \left(\mu(\hat{X}_{t|t}) - \left. \frac{\partial \mu(X_t)}{\partial X_t'} \right|_{X_t = \hat{X}_{t|t}} \hat{X}_{t|t} \right) \Delta t; \tag{12}$$

$$F_{1t} = I + \left. \frac{\partial \mu(X_t)}{\partial X_t'} \right|_{X_t = \hat{X}_{t|t}} \Delta t; \tag{13}$$

$$Q_t = \sigma(\hat{X}_{t|t})\sigma(\hat{X}_{t|t})' \Delta t. \tag{14}$$

In (11) through (14), 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.

The usual method of starting the Kalman filter recursion at the analytic unconditional first and second moments of X_t is unavailable to us because we do not have analytic expressions for these moments. We use simulations to approximate them. We generate a long time series of the state vector by discretizing (1), using a time step of Δt (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 accurately these moments, 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 EMM stage. Therfore the estimated parameters from the auxiliary model can be used as the initial parameters to search for the EMM optimum. However, if the auxiliary parameters imply nonstationarity, they will not produce sensible results in the EMM stage, because EMM imposes stationarity on the underlying process.

3 A nonlinear term structure model

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 $\tilde{Z}_t \equiv (\tilde{Z}_{t,1}, \ldots, \tilde{Z}_{t,n})'$; corresponding Brownian motions under the physical measure are represented without the tildas. The instantaneous nominal interest rate, denoted r_t, is affine in the state:

$$r_t = \delta_0 + \delta X_t. \tag{15}$$

Here, δ_0 is a scalar and δ is an n-vector. The evolution of the state variables under the equivalent martingale measure is given by equation (16):

$$dX_t = (K\theta - KX_t)dt + \Sigma S_t d\tilde{Z}_t. \tag{16}$$

In (16), K and Σ are n x n matrices and θ is an n-vector. Dai and Singleton (1998) show that Σ can be normalized to a diagonal matrix; we adopt their normalization here. The matrix S_t is also diagonal, with elements described in (17):

$$S_{t(ii)} \equiv \sqrt{\alpha_i + \beta_i' X_t}, \tag{17}$$

where β_i and α are n-vectors. This discussion assumes that the dynamics of (16) and (17) 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 (1998) 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

$$P(X_t, \tau) = \exp[A(\tau) - B(\tau)' X_t], \tag{18}$$

$$Y(X_t, \tau) = (1/\tau)[-A(\tau) + B(\tau)' X_t]. \tag{19}$$

In (18) and (19), $A(\tau)$ 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 π_s/π_t as the state price deflator for time-t pricing of time-s payoffs, we can write

$$\frac{d\pi_t}{\pi_t} = -r_t dt - \Lambda_t' dZ_t. \tag{20}$$

The element i of the n-vector Λ_t represents the price of risk associated with the Brownian motion $Z_{t,i}$. We parameterize the model as follows. Let λ_1 and λ_2 be n-vectors of constants. Define Λ_t as

$$\Lambda_t = \Sigma^{-1}\lambda_1 + S_t \Sigma^{-1}\lambda_2 \tag{21}$$

This form, which is equivalent to the form adopted by Duarte (1999), nests the completely affine class of Dai and Singleton (1998) (which includes the models of Cox, Ingersoll, and Ross (1985) and Vasicek (1977)). For this more restrictive class, λ_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 Λ_t is that individual elements of Λ_t 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 (1999), the structure of Λ_t 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.

3.3 Interest rate dynamics under the physical measure

The general representation of the state price deflator's dynamics in (20) allow us to write the dynamics of X_t under the physical measure as follows:

$$dX_t = K(\theta - X_t)dt + \Sigma S_t \Lambda_t dt + \Sigma S_t dZ_t \tag{22}$$

Combining (22) and (21), we can express the physical dynamics of the state vector as

$$dX_t = \left[(K\theta + \psi) + S_t\lambda_1 - K^p X_t\right] dt + \Sigma S_t dZ_t \tag{23}$$

where element i of the vector ψ is $\alpha_i\lambda_{2i}$, the matrix K^p is defined as $K - \Phi$, and row i of matrix Φ 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^pX_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 (15), (18), (23), and the structure of the ODEs in Duffie and Kan (1996) imply that the instantaneous bond-price dynamics are

$$\frac{dP(X_t,\tau)}{P(X_t,\tau)} = [r_t - B(\tau)'(\psi + S_t\lambda_1 + \Phi X_t)]dt - B(\tau)'\Sigma S_t dZ_t \tag{24}$$

Instantaneous expected excess returns to a τ-maturity bond, denoted $e_{t,\tau}$, can be inferred from (24):

$$e_{t,\tau} = -B(\tau)'(\psi + S_t\lambda_1 + \Phi X_t). \tag{25}$$

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:

$$dr_t = (k\theta - kr_t)dt + \sigma\sqrt{r_t}d\tilde{z}_t. \tag{26}$$

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

$$dr_t = (k\theta + \lambda_1\sqrt{r_t} - k^p r_t) + \sigma\sqrt{r_t}dz_t,\ k^p = k - \lambda_2. \tag{27}$$

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

$$e_{t,\tau} = -B(\tau)(\lambda_1\sqrt{r_t} + \lambda_2 r_t) \tag{28}$$

where $B(\tau)$ is given by the usual CIR pricing formula. The signs of λ_1 and λ_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

$$\lambda_1 E(\sqrt{r_t}) + \lambda_2 E(r_t) < 0. \tag{29}$$

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 k. If $k > 0$, 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:

$$k > 0, \quad (1/2)\lambda_1 E(1/\sqrt{r_t}) + \lambda_2 > 0. \tag{30}$$

Multiply the second inequality in (30) by $E(r_t)$, and ignore the difference between $E(\sqrt{r_t})$ and $E(r_t)E(1/\sqrt{r_t})$:

$$(1/2)\lambda_1 E(\sqrt{r_t}) + \lambda_2 E(r_t) > 0. \tag{31}$$

A comparison of (29) and (31) 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 $k > 0$). The additional flexibility provided by λ_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 $\lambda_2 \leq k$. 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 (28) that expected excess returns are the product of exposure to interest rate risk ($-B(\tau)$) and the price of interest rate risk. If k is close to zero, λ_2

must also be close to zero, thus the price per unit of interest-rate risk will not fluctuate much with r_t. If k is large, λ_2 can be large, and the price per unit of interest-rate risk can fluctuate substantially. However, with large k, $-B(\tau)$ 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 k 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 $k > 0$. If $k < 0$, 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 λ_2 produces the correct correlation between the slope of the term structure and expected excess bond returns, and there is no lower bound on λ_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 Properties of the estimation methods: Monte Carlo evidence

In this section we use Monte Carlo simulations to consider three 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 large is the misspecification in the linearized Kalman filter? Third, how do the small-sample properties of the Kalman filter compare to those of SNP/EMM?

4.1 Small-sample estimation biases in extended affine models

Here we document small-sample biases associated with estimating models such as ours, in which the price of risk vector Λ_t is more flexible than it is in the completely affine case. The

biases show up in the parameters that determine the drift of r_t. We begin by using a small Monte Carlo simulation to illustrate the biases.

We generate instantaneous interest rates using a CIR process. Equation (26) describes the equivalent-martingale dynamics and (27), with $\lambda_1 = 0$, describes the physical dynamics. The parameters used in the data-generating 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 (20 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 (26) and (27), where λ_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 12,000 months (1,000 years of data). A small-scale Monte Carlo simulation (100 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.

The main point to take from this table is that the mean parameter estimates imply that the interest rate dynamics are nonlinear. The estimates of λ_1 are biased upward (from its true value of zero), and the estimates of k^p are also biased upward. The table also documents that estimate of k is not biased, thus the bias in k^p is equivalent to a downward bias in the estimate of λ_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.

First consider the bias in the speed of mean reversion when estimating a standard CIR model. Ball and Torous (1996) point out that if a CIR model is estimated using only time-series information, the speed of mean reversion under the physical measure, k^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, k, is determined by the contemporaneous covariances among changes in yields of different maturities. The mean slope of the term structure determines the price of risk λ_2, and thus also determines the speed of mean reversion under the physical measure. Thus k^p is pinned down without using any information about the speed of mean reversion in the sample.

However, when both time-series and cross-sectional information are used to estimate our

nonlinear 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 λ_1 and λ_2. The mean slope of the yield curve cannot determine them both, and thus cannot determine k^p. Therefore the speed of mean reversion in the sample is important in determining these parameter estimates, and the resulting estimates will reflect the small-sample downward-bias in this speed.

This bias can be easily seen by examining the relation between r_t and the drift in r_t. Figure 1 plots r_t versus its drift, $\kappa\theta + \lambda_1\sqrt{r_t} - k^p r_t$. The solid line is constructed with the parameters of the CIR model used to generate the true data and the dashed line is constructed with the mean EMM parameter estimates.

The nonlinear shape of the dashed line allows it to produce both the correct value of $k\theta$ (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. In other words, the additional parameter λ_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 Λ_t. Presumably, any loosening of these restrictions, whether nonlinear or linear (as in Duffee (1999)) will give rise to the same kind of bias in the parameters that determine the drift of r_t.

4.2 The magnitude of misspecification in the Kalman filter

A comparison of Panels A and B of Table 2 reveals that the mean parameter estimates from the linearized Kalman filter auxiliary model are very similar to those from the second-stage, EMM procedure. Moreover, the standard deviations of the estimated parameters are lower with linearized Kalman filter estimation than with the Kalman filter/EMM estimation. This suggests we may be better off simply estimating the (slightly misspecified) Kalman filter.

Of course, the accuracy of our linearized Kalman filter approach depends on the nonlinearity of the true data-generation process. The results in Table 2 are based on a linear data-generation process, so it is not particularly surprising to find that the linearized Kalman filter works so well. The only source of misspecification is in the use of instantaneous first and second moments of yields as proxies for the one-month-ahead first and second moments of yields.

A more appropriate comparison between the linearized Kalman filter and the combined Kalman filter/EMM procedure requires the simulation of a truly nonlinear model. Therefore we generate data using the one-factor nonlinear model of (26) and (27). The parameters of the process were chosen to capture the kinds of nonlinearities that we are likely to observe in term-structure data. To do so, we first fit a three-independent-factor version of the model to monthly U.S. interest rate data from 1974 through 1998. The parameters of the factor that were the most nonlinear (the largest absolute second derivative of the drift function, evaluated at the unconditional mean and plus/minus one unconditional standard deviation from the mean) were rounded off and used in this simulation.

As with the earlier Monte Carlo simulation, 240 months of the instantaneous interest rate and the continuously-compounded one-year bond yield are observed with iid measurement error. The observed data are fit to the nonlinear term-structure model of (26) and (27). The model is estimated with EMM using the linearized Kalman filter as an auxiliary model. Each EMM simulation has length 12,000 months (1,000 years of data). A small-scale Monte Carlo simulation (100 simulations) is used to generate distributions of parameter estimates and associated standard errors. The parameter estimates from the auxiliary model and the EMM stage are displayed in Panels A and B of Table 3.

The mean estimates for those parameters identified by the equivalent martingale measure are very similar across the two panels. However, the mean estimates for both λ_1 and k^p clearly differ. The estimated drift functions are displayed in Figure 2. As seen in the table and the figure, the Kalman filter estimates are closer to the true parameters than are the EMM estimates.

The reason for this ordering is straightforward. The estimated drift functions (the dotted and dashed lines in Figure 2) exhibit more mean reversion than the true drift function, for the reason discussed in the previous subsection.[9] As in Table 2, the bias shows up in Table 3 as an upward bias to both λ_1 and k^p. However, there is another, offsetting bias introduced by the misspecification of the linearized Kalman filter. The linearized model implies that shocks die out linearly, while in the data-generating model they die out exponentially. Therefore the linearized model produces more mean reversion for a given k^p than does the true model.[10] Thus for a given speed of mean reversion in the data, the linearized Kalman filter will fit it with a smaller value of k^p.

Table 3 documents that the parameter estimates of λ_1 and k^p are imprecise. Both the

[9]The magnitude of the difference between the true and estimated drifts is not as large in Figure 2 as it is in Figure 1 because the true speed of mean reversion is substantially higher in Figure 2, resulting in a smaller bias.

[10]For example, in a CIR model with $k^p = 1.5$, the linearized model implies that $1 - 1.5/12 = 0.875$ of a shock remains after one month, while the correct calculation is $e^{1.5/12} = 0.882$.

standard errors of the estimates and the means of their associated standard errors are similar in magnitude to the parameters themselves. Note that the Kalman filter estimates of these parameters are more precise than are the corresponding EMM estimates. This fact, combined with the fact that the Kalman filter estimates are closer to the true parameters than are the EMM estimates, indicates that we may be better off simply estimating the linearized Kalman filter rather than taking the additional step of using the Kalman filter as an auxiliary model for EMM.

4.3 Comparing the linearized Kalman filter with SNP/EMM

We argued earlier that the SNP/EMM procedure is likely to perform poorly in econometric settings where the number of SNP moment conditions is large relative to the number of parameters to be estimated. In order to test this claim, the one-factor examples considered in previous subsections are inappropriate. In those examples, the data consisted of only two bond yields. The simplest SNP description of two time series has nine parameters, which is only slightly larger than the seven parameters (including measurement error standard deviations) in the one-factor nonlinear model of (26) and (27).

Therefore here we compare the performance of SNP/EMM to the linearized Kalman filter in the context of a model that is closer to the kinds of models that econometricians will estimate in practice. (We would like to include the performance of the Kalman filter/EMM, but Monte Carlo simulation of this two-step procedure using realistic models is currently prohibitively time-consuming.) The process driving bond yields is a two-factor CIR process, with the instantaneous interest rate is the sum of the independent factors.

$$r_t = x_{1,t} + x_{2,t}. \tag{32}$$

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

$$dx_{i,t} = (k\theta_i - k_i x_{i,t})dt + \sigma_i \sqrt{x_{i,t}} d\tilde{z}_{i,t},\ i = 1, 2. \tag{33}$$

Under the physical measure, the dynamics are also affine (there is no λ_1 vector):

$$dx_{i,t} = (k\theta_i - (k_i - \lambda_{2i})x_{i,t})dt + \sigma_i \sqrt{x_{i,t}} dz_{i,t},\ i = 1, 2. \tag{34}$$

To identify the parameters in the model we use yields on bonds of three maturities. 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 parameters of (32), (33), and (34) are estimated in two ways. First, the data are fit to the misspecified Kalman filter. Because the true dynamics of r_t are affine, the misspecification consists in using the instantaneous dynamics to construct moments instead of using the true one-month-ahead moments. The unconditional moments used to begin the Kalman filter are estimated with simulations.

Second, the SNP/EMM methodology is used to estimate the parameters. Experimentation revealed that the preferred SNP model is a simple VAR with one lag and no ARCH, GARCH, or higher-order terms in the variance. Because the observed data are three-dimensional, there are eighteen SNP parameters (4x3 VAR terms plus the six free elements of the unconditional variance-covariance matrix). The resulting eighteen moment conditions are used in the EMM stage. Each EMM simulation has length 12,000 months (1,000 years of data).[11]

The true parameters and a summary of the parameter estimates from 100 Monte Carlo simulations are displayed in Table 4. First consider the Kalman filter estimates. There are three main conclusions to draw from the table. First, the mean Kalman filter estimates are very close to the true parameters values, except for a slight overestimate of the speeds of mean reversion under the physical measure. The true speeds of mean reversion for factors one and two are $k_1 - \lambda_{11} = 0.6$ and $k_2 - \lambda_{12} = 0.2$ respectively, while corresponding mean point estimates are 0.649 and 0.267. Second, the distributions of the Kalman filter parameter estimates appear to be reasonably symmetric; the mean and median estimates are very similar. Third, the mean standard errors of the estimates are reasonably close to the standard deviations of the estimated parameters. The standard errors are typically somewhat larger than the corresponding standard deviations, but the differences are not large.

The table also documents that the SNP/EMM estimates are much less precise than are the Kalman filter estimates. For every parameter, the standard deviation of the SNP/EMM estimates is larger, sometimes by a factor of 40, than the corresponding standard deviation of the Kalman filter estimates. (Because only 100 Monte Carlo simulations are generated, these large standard deviations make it impossible to evaluate the potential bias in SNP/EMM parameters.) However, the standard errors from the SNP/EMM estimation process do not reflect the imprecision in the estimates. They understate, sometimes drastically, the true uncertainty in the parameters. For example, the standard deviations of the parameter estimates

[11]This length is much shorter than those typically used in higher-frequency SNP/EMM estimation (e.g., Zhou (1999)). Here, however, the only features of the data that the simulation needs to reproduce are the first-order VAR coefficients and the unconditional variance-covariance matrix of shocks to this VAR. For these features, 1,000 years should be sufficient.

for $k\theta_1$ and k_1 are fifteen times greater than their mean standard errors.

The combination of imprecise parameter estimates and overly precise standard errors leads to overrjection of the true model. One measure of overrejection is fraction of Type 1 errors associated with the GMM general χ^2 goodness-of-fit test. When a 99th percent confidence level is used, the null hypothesis that the model is correct is rejected in 42 of the 100 Monte Carlo simulations. Half of the time, the null hypothesis is rejected at the 95% confidence level.

These Monte Carlo results are strong evidence for the conclusion that the SNP/EMM technique is inappropriate for the estimation of typical no-arbitrage term-structure models. There are too many SNP moment conditions, and the moments are inefficient in the sense that they do not exploit fully the implications of the models. By contrast, the linearized Kalman filter procedure produces both more accurate parameter estimates and more accurate estimates of the uncertainty in the parameters.

5 Fitting a nonlinear term-structure model to Treasury yields

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 data range is January 1974 through December 1998, for a total of 300 months. We chose this range to include both recent data and the highly volatile interest rate period 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:

$$r_t = \delta_0 + X_{t,1} + X_{t,2} + X_{t,3} \tag{35}$$

We restrict our attention to the class of models denoted as $A_3(3)$ by Dai and Singleton (1998), which means that matrix ΣS_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 (1998) 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.[12] Therefore the 'feedback' matrix K will have 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.

$$K = \begin{pmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ 0 & 0 & \kappa_{33} \end{pmatrix} \tag{36}$$

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 λ_1.

The physical dynamics of X_t are:

$$\begin{aligned} dX_t &= \left[K\theta + \begin{pmatrix} X_{1,t}^{1/2}\lambda_{11} \\ X_{2,t}^{1/2}\lambda_{12} \\ X_{3,t}^{1/2}\lambda_{13} \end{pmatrix} - K^p X_t \right] dt + \Sigma S_t dZ_t, \qquad (37) \\ K^p &= K - diag(\lambda_2) = \begin{pmatrix} \kappa_{11}^p & 0 & 0 \\ \kappa_{21} & \kappa_{22}^p & 0 \\ 0 & 0 & \kappa_{33}^p \end{pmatrix}, \end{aligned}$$

[12]To use his estimation approach (an application of the approximate maximum likelihood technique developed by Aït-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.

where $diag(\lambda_2)$ refers to the diagonal matrix with λ_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 k_{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 λ_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 (9), is assumed to be diagonal. The diagonal elements are $D^2_{1/2}, D^2_1, D^2_2, D^2_5$, 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 of the auxiliary model

In order to estimate the parameters of the above model, we employ the Kalman filter using a linearized version of the same model. The score vector used in the EMM step is the vector of derivatives of the Kalman filter log-likelihood function with respect to the estimated parameters.

A number of practical problems arise in estimating the auxiliary model. 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 $\hat{X}_{t|t-1}$ 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 (1999).

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.[13]

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 auxiliary model.

The parameter estimates and associated standard errors from the Kalman filter are displayed in Table 5. The parameter estimates are in square brackets and their standard errors, which are computed assuming that the Kalman filter is correctly specified, are in curly brackets. We defer a detailed discussion of these results until a little later. However, one feature of these parameter estimates deserves special mention here. Two of the estimates are on the boundary of their parameter space: K_{33}^{p} 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.

5.4 EMM estimation

The main motivation for the choice of the Kalman filter as an auxiliary model for EMM is because of its small-sample properties. However, there is another implication of this choice that has large practical implications. 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

[13]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, for reasons discussed below. We arbitrarily chose to allow five violations. At the optimum, five violations occurred.

the EMM objective function are in the same local neighborhood as are the parameters that maximize the Kalman filter likelihood function.

This extremely useful. Because the objective functions have so many local maxima and minima, we must conduct a comprehensive search over the parameter space. Estimation of the auxiliary model takes much less CPU time (by a factor of ten or more) than does EMM estimation. Thus the ability to conduct this search in the auxiliary model stage instead of the EMM stage is an important advantage of our choice of auxiliary model.

To recap, there are 20 free parameters to estimate with EMM. They are the original 22 parameters excluding k^p_{33} and D_1, which are set to zero. We estimate the remaining parameters by simulating a long time series of data and minimizing the EMM objective function.

The length of the simulation is 50,000 months. The state vector X_t is simulated by discretizing (37). 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 (19).

The parameter estimates from EMM estimation are displayed in Table 5. (They are the values without parentheses or brackets.) Standard errors are in parentheses. The parameter estimates from the EMM stage are very close to those from the Kalman filter stage.

Although their parameter estimates are similar, the standard errors produced by the Kalman filter and EMM procedures are typically substantially different. Our interpretation of these differences is that they are indirect evidence of model misspecification. The problem is related to the contemporaneous Kalman filter forecasts $\hat{X}_{t|t}$. When the model is simulated in the EMM stage with the optimal EMM parameters, none of the 50,000 months of forecasted states are particularly close to their lower bound of zero. Moreover, perturbing the parameters around their optimal values did not alter this result. However, the state forecasts generated with actual bond yields are frequently very close to their lower bounds, and on five occasions the lower bound is violated.

When the state forecasts are very close to the lower bound, a tiny change in a parameter can cause a large change in the likelihood function, by pushing the state forecast either further away from the bound or through the bound. The standard errors on the Kalman filter parameters reflect this sensitivity.[14] But if the model were correctly specified, the

[14]If the likelihood function is set to a large negative value when the lower bound on the state vector is

EMM simulations indicate that small perturbations of the parameters should not have such an effect.

Regardless of which set of standard errors are used, the results in Table 5 do not give strong support for nonlinearities in interest-rate dynamics—at least the kind of nonlinearities examined in this paper. None of the three elements of λ_1 is statistically different from zero under both the Kalman filter and EMM estimates. We now take a closer look at the role λ_1 plays in fitting the data.

5.5 The price of risk

As noted in Section 3, completely affine models–those for which the vector λ_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. The question we address here is whether the more general form of the price of risk in (21) provides the needed flexibility. The short answer is no. More precisely, the model gets the sign of the relation correct, but the magnitude is too small.

Table 6 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 (25), with the contemporaneous predictions $X_{t|t}$ 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 (1999) 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 7 reproduces this result for the 1974–1998 period examined in this paper. The change in a bond yield from month t to month $t+n$ is regressed on the month-t slope of the term structure, for $n = 1, 6$. 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

violated, this problem is exacerbated. This is why we allow for a few violations of the bound at the optimum set of Kalman filter parameters.

to fall. As the horizon lengthens from one month to six months, these patterns become more pronounced.

Table 7 also reports the corresponding regression coefficients for yield forecasts from our model. The month $t+n$ bond yield forecasted as of month t is approximated by extrapolating from $X_{t|t}$ using the instantaneous dynamics of (37). Then the change from the month-t yield to the forecasted month $t+n$ 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 6 and 7 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 λ_1 in (21) frees up the sign of the vector λ_2, it does not free up the magnitude of λ_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; k_{33} is approximately one, which implies a half-life of seven to eight months. Under the physical measure, k^p_{33} is estimated to be on the boundary of its parameter space. In other words, given k_{33}, the estimate of λ_{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 6, between the term-structure slope and expected excess returns.

The additional flexibility afforded by λ_1 in (21) is that λ_{23} can be positive, and hence generate the correct sign of the relation between the slope and expected excess returns. Because λ_{13} is negative, investors, on average, receive compensation for facing the bond-price risk associated with slope shocks. Without λ_{13}, a positive value of λ_{23} would make investors pay to face slope risk on average. However, λ_{23} is bounded above by the requirement that k^p_{33} 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.) His parameter estimates generate substantial covariation between expected excess returns and the term-structure slope. But the source of this relation in his model is a factor that is explosive under the equivalent martingale measure.[15] As discussed in the example in Section 3, an explosive factor makes it fairly simple to generate large variations between expected excess returns to bonds and the slope of the term structure. However, as also noted in the example, a shock to such a nonstationary factor creates a positive relation between short-term interest rates and the slope of the term structure, as interest-rate movements are magnified at the long end. Over the 1974–1998 data sample examined in this paper, we do not observe such behavior.

6 Concluding comments

The search continues for a model of the term structure that is simultaneously parsimonious, does not admit arbitrage, and fits the behavior of Treasury yields. We make several contributions to this search. First, we show that (currently) the most commonly used procedure for estimating these models, EMM combined with an SNP auxiliary model, may have poor small sample properties when used to estimate term stucture models using both cross-sectional and time series data. We instead advocate using an alternative auxiliary model, a linearized Kalman filter. This has the key advantage over SNP that the restrictions imposed by no-arbitrage are imposed when constructing moments. This offers advantages both because of its small-sample properties, and its practical value in estimating models for which the likelihood functions have a large number of local maxima.

The model we use to investigate the small sample properties of these estimators is of interest in its own right. It is a a three-factor term-structure model that parsimoniously generalizes the usual affine formulation by being affine under the equivalent martingale measure, but nonlinear under the physical measure. Using this model we find that the biases noted by Ball and Torous (1996) are a pervasive feature of term structure estimation, and do not (as they hypothesize) in general go away when we use a combination of the cross-sectional and time-series information in yields.

Finally, estimating the model using real data, we find that the price of risk examined in this paper, although more general than standard models, is nonetheless too restrictive to generate realistic behavior for expected excess returns to bonds. We are able to describe

[15] Duarte calls the factor a "curvature" factor.

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.

References

Ahn, D.-H., R. F. Dittmar, and A. R. Gallant, 2000, Quadratic term structure models: Theory and evidence, Working paper, University of North Carolina.

Aït-Sahalia, Y., 1996a, Nonparametric pricing of interest rate derivative securities, *Econometrica* 64, 527–560.

Aït-Sahalia, Y., 1996b, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385–426.

Aït-Sahalia, Y., 1999, Transition densities for interest rate and other diffusions, *Journal of Finance* 54, 1361–1395.

Andersen, T. G., H.-J. Chung, and B. E. Sørensen, 1999, Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study, *Journal of Econometrics* 91, 61–87.

Andersen, T. G., and J. Lund, 1996a, The short rate diffusion revisited: An investigation guided by the efficient method of moments, Working paper, Northwestern University.

Andersen, T. G., and J. Lund, 1996b, Stochastic volatility and mean drift in the short term interest rate diffusion: Sources of steepness, level and curvature in the yield curve, Working paper, Northwestern University.

Andersen, T. G., and J. Lund, 1997, Estimating continuous time stochastic volatility models of the short term interest rate, *Journal of Econometrics* 77, 343–377.

Andersen, T. G., and B. E. Sørensen, 1996, GMM estimation of a stochastic volatility model: A Monte Carlo study, *Journal of Business and Economic Statistics* 14, 328–352.

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

Backus, D. K., S. Foresi, A. Mozumdar, and L. Wu, 1998, Predictable changes in yields and forward rates, Working Paper 6379, NBER.

Balduzzi, P., and Y. H. Eom, 1997, Nonlinearities in U.S. Treasury yields: A semi-nonparametric approach, Working paper, NYU.

Ball, C. A., and W. N. Torous, 1996, Unit roots and the estimation of interest rate dynamics, *Journal of Empirical Finance* 3, 215–238.

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

Boudoukh, J., M. Richardson, R. Stanton, and R. F. Whitelaw, 1998, The stochastic behavior of interest rates: Implications from a multifactor, nonlinear, continuous-time model, Working paper, NYU.

Brandt, M. W., and P. Santa-Clara, 1999, Simulated likelihood estimation of multivariate diffusions with an application to interest rates and exchange rates with stochastic volatility, Working paper, Wharton.

Campbell, J. Y., and R. J. Shiller, 1991, Yield spreads and interest rate movements: A bird's eye view, *Review of Economic Studies* 58, 495–514.

Chan, K. C., A. G. Karolyi, F. A. Longstaff, and A. B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* 47, 1209–1227.

Chapman, D. A., and N. D. Pearson, 2000, Is the short rate drift *actually* nonlinear?, *Journal of Finance* 55, 355–388.

Chen, R.-R., and L. Scott, 1993, Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates, *Journal of Fixed Income* 3, 14–31.

Chen, R.-R., and L. Scott, 1995, Multifactor Cox-Ingersoll-Ross models of the term structure: Estimates and tests from a Kalman filter model, Working paper, University of Georgia.

Chumacero, R. A., 1997, Finite sample properties of the efficient method of moments, *Studies in Nonlinear Dynamics and Econometrics* 2, 35–51.

Conley, T. G., L. P. Hansen, E. G. J. Luttmer, and J. A. Scheinkman, 1997, Short-term interest rates as subordinated diffusions, *Review of Financial Studies* 10, 525–577.

Constantinides, G. M., 1992, A theory of the nominal term structure of interest rates, *Review of Financial Studies* 5, 531–552.

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

Dai, Q., and K. J. Singleton, 1998, Specification analysis of affine term structure models, Working paper, Stanford University, Forthcoming, *Journal of Finance*.

Dai, Q., and K. J. Singleton, 2000, Expectation puzzles, time-varying risk premia, and dynamic models of the term structure, Working paper, Stanford University.

de Jong, F., 1998, Time-series and cross-section information in affine term structure models, Working paper, University of Amsterdam.

Duan, J.-C., and J.-G. Simonato, 1997, Estimating and testing exponential-affine term structure models by Kalman filter, Working paper, Hong Kong University of Science and Technology.

Duarte, J., 1999, The relevance of the price of risk in affine term structure models, Working paper, Chicago GSB.

Duffee, G. R., 1999, Forecasting future interest rates: Are affine models failures?, Working paper, U.C. Berkeley.

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

Duffie, D., L. H. Pedersen, and K. J. Singleton, 2000, Modeling sovereign yield spreads: A case study of Russian debt, Working paper, Stanford University.

Duffie, D., and K. J. Singleton, 1993, Simulated moments estimation of Markov models of asset prices, *Econometrica* 61, 929–952.

Duffie, D., and K. J. Singleton, 1997, An econometric model of the term structure of interest-rate swap yields, *Journal of Finance* 52, 1287–1321.

Feller, W., 1951, Two singular diffusion problems, *Annals of Mathematics* 54, 173–182.

Gallant, A. R., and J. R. Long, 1997, Estimating stochastic differential equations efficiently by minimum chi-square, *Biometrika* 84, 125–141.

Gallant, A. R., and G. E. Tauchen, 1992, A nonparametric approach to nonlinear time series analysis: Estimation and simulation, in D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, and M. S. Taqqu, eds., *New Directions in Time Series Analysis, Part II*, 71–92 (Springer-Verlag, New York, NY).

Gallant, A. R., and G. E. Tauchen, 1996, Which moments to match?, *Econometric Theory* 12, 657–681.

Gallant, A. R., and G. E. Tauchen, 1998a, EMM: A program for efficient method of moments estimation, version 1.4, User's guide, University of North Carolina, Chapel Hill.

Gallant, A. R., and G. E. Tauchen, 1998b, SNP: A program for nonparametric time series analysis, version 8.7, User's guide, University of North Carolina, Chapel Hill.

Gallant, A. R., and G. E. Tauchen, 1999, The relative efficiency of method of moments estimators, *Journal of Econometrics* 92, 149–172.

Geyer, A. L. J., and S. Pichler, 1999, A state-space approach to estimate and test multifactor Cox-Ingersoll-Ross models of the term structure, *Journal of Financial Research* 22, 107–130.

Ghysels, E., and S. Ng, 1998, A semiparametric factor model of interest rates and tests of the affine term structure, *Review of Economics and Statistics* 80, 535–548.

Gibbons, M. R., and K. Ramaswamy, 1993, A test of the Cox, Ingersoll, and Ross model of the term structure, *Review of Financial Studies* 6, 619–658.

Gill, P. E., W. Murray, M. A. Saunders, and M. H. White, 1983, User's guide for SOL/NPSOL: A FORTRAN package for nonlinear programming, Technical Report SOL 83–12, Systems Optimization Laboratory, Department of Operations Research, Stanford University.

Gourieroux, C., A. Monfort, and E. Renault, 1993, Indirect inference, *Journal of Applied Econometrics* 8, S85–118.

Hamilton, J. D., 1994, *Time Series Analysis* (Princeton University Press, Princeton).

Hansen, L. P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

Hansen, L. P., J. C. Heaton, and A. Yaron, 1996, Finite-sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics* 14, 262–280.

Hansen, L. P., and J. A. Scheinkman, 1995, Back to the future: Generating moment implications for continuous-time Markov processes, *Econometrica* 63, 767–804.

Harvey, A. C., 1989, *Forecasting, Structural Models and the Kalman Filter* (Cambridge University Press, New York, NY).

Heath, D., R. A. Jarrow, and A. Morton, 1992, Bond pricing and the term structure of interest rates: A new methodology, *Econometrica* 60, 77–105.

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

Jiang, G. J., 1999, Estimation of jump-diffusion processes based on indirect inference - with applications to currency exchange rate models, in S. Holly, and S. Greenblatt, eds., *Issues in Computational Economics and Finance* (Elsevier, Amsterdam).

Jones, C. S., 1998, Nonlinear mean reversion in the short-term interest rate, Working paper, Wharton.

Kloeden, P. E., and E. Platen, 1992, *Numerical Solution of Stochastic Differential Equations* (Springer-Verlag, New York, NY).

Leippold, M., and L. Wu, 2000, Quadratic term structure models, Working paper, University of St. Gallen.

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

Lo, A. W., 1988, Maximum likelihood estimation of generalized Ito processes with discretely sampled data, *Econometric Theory* 4, 231–247.

Longstaff, F. A., 1989, A nonlinear general equilibrium model of the term structure, *Journal of Financial Economics* 23, 195–224.

McCulloch, J. H., and H.-C. Kwon, 1993, U.S. term structure data, 1947–1991, Working Paper 93–6, Ohio State University.

Nelder, J. A., and R. Mead, 1965, A simplex method for function minimization, *Computer Journal* 7, 308–313.

Øksendal, B., 1985, *Stochastic Differential Equations: An Introduction with Applications*, third edition (Springer-Verlag, New York, NY).

Pearson, N. D., and T.-S. Sun, 1989, A test of the Cox, Ingersoll, Ross model of the term structure of interest rates using the method of maximum likelihood, Working paper, MIT.

Pearson, N. D., and T.-S. Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model, *Journal of Finance* 49, 1279–1304.

Pedersen, A. R., 1995, A new approach to maximum-likelihood estimation of stochastic differential equations based on discrete observations, *Scandinavian Journal of Statistics* 22, 55–71.

Pennacchi, G. G., 1991, Identifying the dynamics of real interest rates and inflation: Evidence using survey data, *Review of Financial Studies* 4, 53–86.

Pfann, G. A., P. C. Schotman, and R. Tschernig, 1996, Nonlinear interest rate dynamics and implications for the term structure, *Journal of Econometrics* 74, 149–176.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1992, *Numerical Recipes in C*, second edition (Cambridge University Press, New York).

Santa-Clara, P., 1995, Simulated likelihood estimation of diffusions with an application to the short term interest rate, Working paper, INSEAD.

Schwarz, G., 1978, Estimating the dimension of a model, *Annals of Statistics* 6, 461–464.

Stanton, R. H., 1997, A nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance* 52, 1973–2002.

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

Zhou, H., 1999, Finite sample properties of efficient method of moments and maximum likelihood estimations for a square-root diffusion process, Working paper, Duke University.

Variable	True Value	Estimate 1 Set 1	Estimate 2 Truth
κ	0.29368	0.37807	0.35941669
		(0.00000003)	0.00013770
μ	0.07935	0.11127186	0.06404175
		(0.00000000)	0.00003200
σ	0.11425	0.10611242	0.10892350
		(0.00000003)	0.00005381
λ	-0.12165	0.00883312	-0.18975737
		(0.00000001)	0.00017812
D_1	0.001	0.00203092	0.00009934
		(0.00000000)	0.00000024
$\chi^2(44)$	-	7.2×10^{10}	1557.8

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

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 D_1), and the data are fit to a standard 1 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, one close to, but not equal to, the true parameter values, and another where the estimation was started at the population parameter values.

Panel A. Kalman filter estimates

Parameter	True value	Mean Estimate	Std. Dev. of Estimates	Mean Std. Error
$k\theta$	0.021	0.0210	0.00058	0.00058
k	0.200	0.2005	0.00890	0.00802
σ	0.100	0.0985	0.00480	0.00465
λ_1	0.0	0.1062	0.13058	0.10968
k^p	0.3	0.7200	0.50388	0.41255
D_r	0.001	9.822×10^{-4}	3.233×10^{-4}	3.074×10^{-4}
D_1	0.001	9.223×10^{-4}	2.830×10^{-4}	2.301×10^{-4}

Panel B. EMM estimates

Parameter	True value	Mean Estimate	Std. Dev. of Estimates	Mean Std. Error
$k\theta$	0.021	0.0207	0.00146	0.00058
k	0.200	0.1985	0.00882	0.00791
σ	0.100	0.1010	0.00507	0.00484
λ_1	0.0	0.1202	0.14776	0.13182
k^p	0.3	0.7660	0.57373	0.51702
D_r	0.001	1.027×10^{-3}	3.071×10^{-4}	2.348×10^{-4}
D_1	0.001	8.870×10^{-4}	2.818×10^{-4}	2.350×10^{-4}

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 (26) and (27) in the paper, using EMM with a linearized Kalman filter as the auxiliary model. This table summarizes the results of 100 Monte Carlo simulations.

Panel A. Kalman filter estimates

Parameter	True value	Mean Estimate	Std. Dev. of Estimates	Mean Std. Error
$k\theta$	0.065	0.0648	0.00155	0.00169
k	1.500	1.4953	0.04280	0.04595
σ	0.100	0.0933	0.00479	0.00515
λ_1	-0.150	-0.1185	0.14202	0.18580
k^p	1.200	1.3758	0.78385	1.02168
D_r	0.001	1.001×10^{-3}	3.469×10^{-4}	4.468×10^{-4}
D_1	0.001	9.676×10^{-4}	1.017×10^{-4}	1.012×10^{-4}

Panel B. EMM estimates

Parameter	True value	Mean Estimate	Std. Dev. of Estimates	Mean Std. Error
$k\theta$	0.065	0.0642	0.00153	0.00155
k	1.500	1.4809	0.04205	0.04210
σ	0.100	0.1009	0.00523	0.00584
λ_1	-0.150	-0.0358	0.17058	0.19622
k^p	1.200	1.8088	0.95025	1.10605
D_r	0.001	1.070×10^{-3}	3.321×10^{-4}	2.597×10^{-4}
D_1	0.001	9.522×10^{-4}	1.039×10^{-4}	9.041×10^{-5}

Table 3: Comparing linearized Kalman filter estimation with Kalman filter/EMM: Monte Carlo simulation results

Twenty years of monthly observations of instantaneous interest rates and one-year bond yields are generated by a nonlinear extension of a one-factor CIR process. The model is described by equations (26) and (27) in the paper. The data, which are observed with iid measurement error (with standard deviations D_r and D_1), are fit to the model using EMM with a linearized Kalman filter as the auxiliary model. This table summarizes the results of 100 Monte Carlo simulations.

Panel A. Parameter estimates

Parameter	True Value	Kalman filter estimates Mean	Median	Std. Dev.	SNP/EMM estimates Mean	Median	Std. Dev.
$k\theta_1$	0.018	0.018	0.018	0.0013	0.023	0.019	0.0476
k_1	0.300	0.299	0.296	0.0235	0.380	0.294	0.8584
σ_1	0.100	0.099	0.098	0.0077	0.111	0.093	0.0584
λ_{21}	-0.300	-0.350	-0.331	0.1404	-0.584	-0.317	1.0381
$k\theta_2$	0.006	0.006	0.006	0.0006	0.006	0.006	0.0024
k_2	-0.010	-0.009	-0.010	0.0099	-0.013	-0.006	0.0388
σ_2	0.100	0.100	0.100	0.0075	0.101	0.100	0.0248
λ_{22}	-0.210	-0.276	-0.241	0.1335	-0.285	-0.219	0.1855
$D_{1/2}$	0.001	0.001	0.001	0.0001	0.001	0.001	0.0003
D_2	0.001	0.001	0.001	0.0001	0.001	0.001	0.0002
D_{10}	0.001	0.001	0.001	0.0001	0.001	0.001	0.0002

Table 4: A comparison of estimation methods: Monte Carlo simulation results

Table 4 continues on the next page.

Panel B. Estimated standard errors

	Kalman filter estimates		SNP/EMM estimates	
Parameter	Mean	Median	Mean	Median
$k\theta_1$	0.0039	0.0011	0.0062	0.0031
k_1	0.0873	0.0233	0.0539	0.0258
σ_1	0.0120	0.0072	0.0297	0.0191
λ_{21}	0.1305	0.1189	0.3152	0.1101
$k\theta_2$	0.0008	0.0006	0.0015	0.0011
k_2	0.0238	0.0080	0.0277	0.0200
σ_2	0.0095	0.0060	0.0185	0.0142
λ_{22}	0.1242	0.0886	0.1164	0.0750
$D_{1/2}$	0.0003	0.0001	0.0003	0.0002
D_2	0.0002	0.0001	0.0001	0.0001
D_{10}	0.0002	0.0001	0.0002	0.0001

Table 4: 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 two-independent-factor CIR process. 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 100 Monte Carlo simulations.

	Index number (i)		
Parameter	1	2	3
$(K\theta)_i$	0.0026 (0.0069) [0.0026] {0.0026}	0.0040 (0.0245) [0.0040] {0.0012}	0.0440 (0.2917) [0.0424] {0.0136}
k_{1i}	0.0427 (0.0488) [0.0424] {0.0416}	0	0
k_{2i}	-0.2667 (0.8500) [-0.2612] {0.1051}	0.2869 (0.1289) [0.2843] {0.1844}	0
k_{3i}	0	0	1.1278 (0.8152) [1.121] {0.563}
σ_i	0.0432 (0.0260) [0.0428] {0.0082}	0.0799 (0.0318) [0.0791] {0.0552}	0.1198 (0.6101) [0.1164] {0.0127}
λ_{1i}	0.0013 (0.0340) [0.0013] {0.0018}	0.1440 (0.0585) [0.1411] {0.6416}	-0.2818 (0.5428) [-0.2791] {0.0935}
k_{ii}^p	0.1176 (0.5073) [0.1160] {0.0639}	0.9043 (0.4581) [0.8992] {2.4674}	0 [0] {–}

Table 5: Results of Kalman filter/EMM estimation of a nonlinear term-structure model

Table 5 continues on the next page.

Constant term and measurement error standard deviations

δ_0	$D_{1/2}$	D_1	D_2	D_5	D_{10}
-0.0409	0.00205	0	0.00072	0.00085	0.00118
(0.2085)	(0.00011)		(0.00016)	(0.00015)	(0.00024)
$[-0.0404]$	[0.00204]	[0]	[0.00071]	[0.00086]	[0.00117]
{0.0538}	{0.00014}	{–}	{0.00018}	{0.00016}	{0.00041}

Table 5: Results of Kalman filter/EMM estimation of a 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 (35) through (37). The sample period is January 1974 through December 1998. A linearized Kalman filter produced the parameter estimates in square brackets and associated standard errors in curly brackets. Using the Kalman filter as an auxiliary model, EMM produced the parameter estimates at the top of each set of figures, with associated standard errors in parentheses. 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.

Maturity	Mean	Std. dev.	Corr. with term slope	Slope-sorted quartile means 1st	4th
2 yrs	0.38	0.77	0.57	-0.26	0.72
5 yrs	0.19	1.14	0.38	-0.42	0.60
10 yrs	0.26	1.38	0.27	-0.25	0.64

Table 6: 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.

	One month ahead		Six months ahead	
	Actual coef	Model-implied coef	Actual coef	Model-implied coef
6 months	0.029	0.116	0.144	0.509
2 yrs	-0.014	0.106	-0.081	0.317
5 yrs	-0.036	0.002	-0.188	0.117
10 yrs	-0.044	-0.065	-0.225	0.014

Table 7: Regressions of changes in bond yields on the slope of the term structure, 1974–1998

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 $t+n$ bond yields. The same regression is run, with the forecasted yields used in place of the actual month $t+n$ yields. The slope of the term structure is measured by the difference between the five-year and the six-month zero-coupon yields.

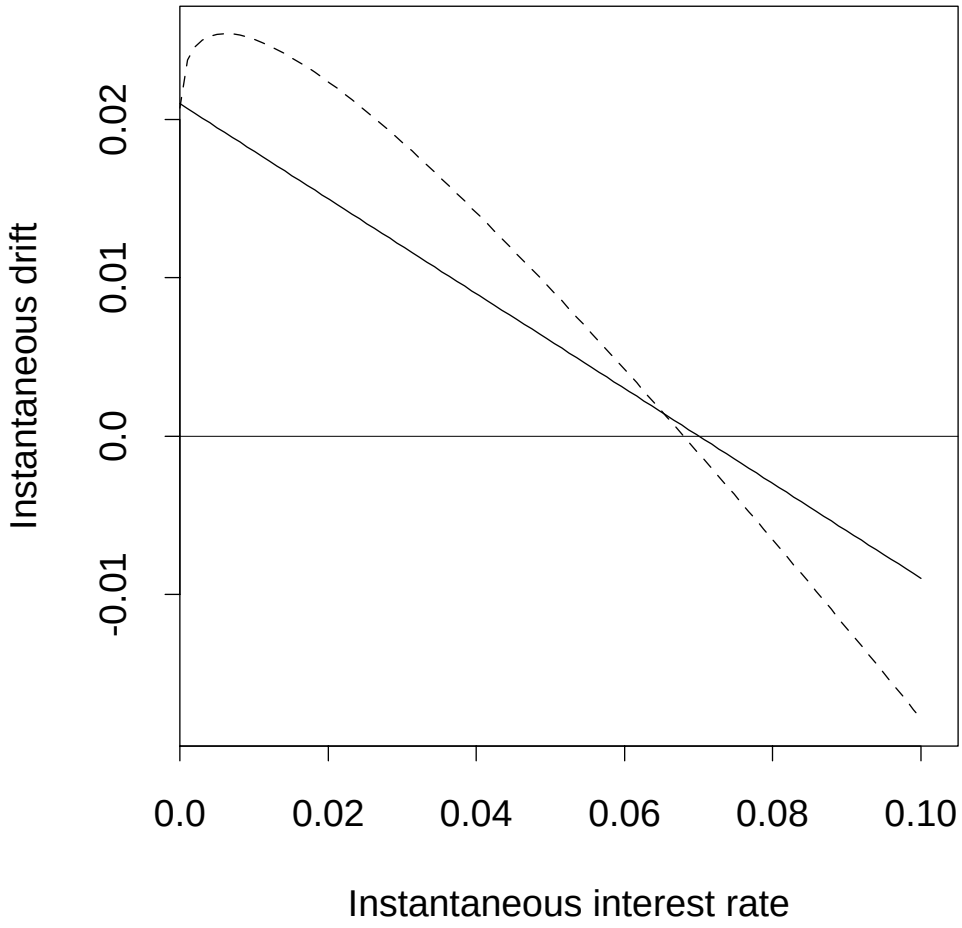

Figure 1: Actual and estimated drifts for a one-factor term-structure model.

The data are generated by a CIR model for which the drift in the instantaneous interest rate is given by the solid line. The dashed line displays the drift implied by EMM parameter estimates of a nonlinear extension to the CIR model.

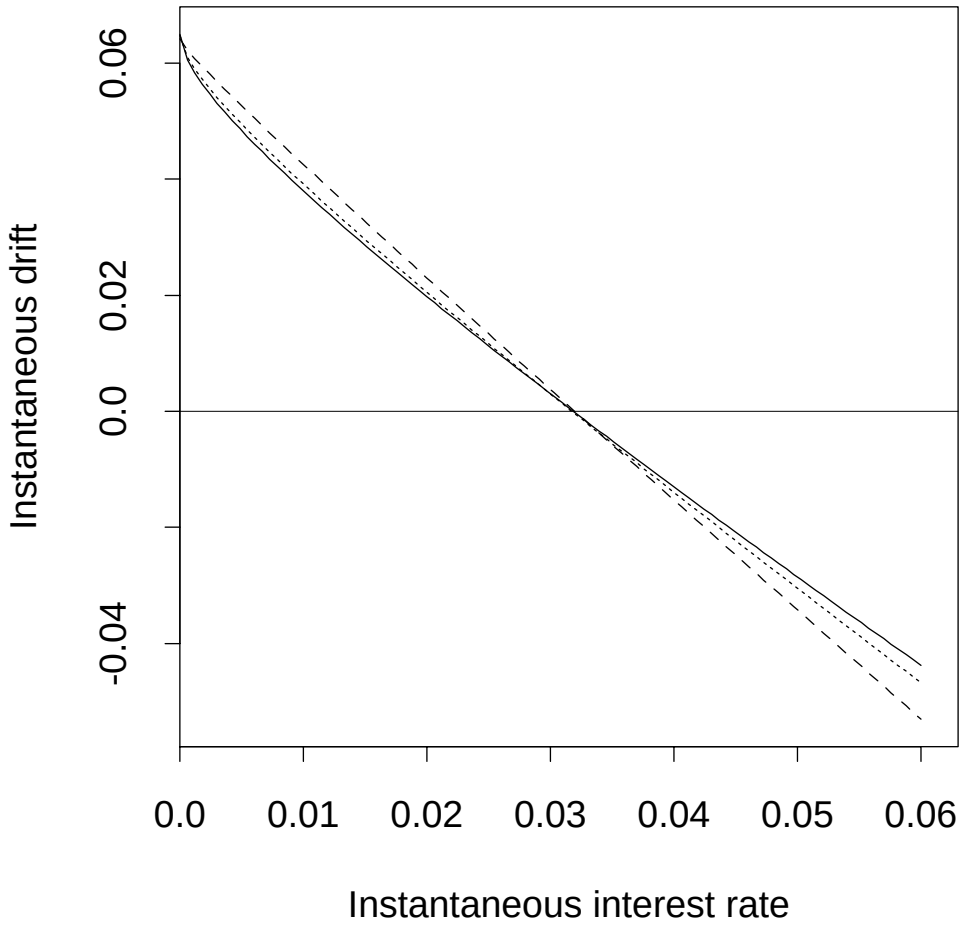

Figure 2: Actual and estimated drifts for a nonlinear one-factor term-structure model.

The data are generated by a nonlinear extension to 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. The dashed line displays the drift implied by EMM parameter estimates.