Journal of Empirical Finance 16 (2009) 852–861

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Evaluating stochastic discount factors from term structure models ☆

Heber K. Farnsworth *

Olin Business School Washington University in St. Loius, USA

ARTICLE INFO

Article history:
Accepted 16 June 2009
Available online 27 June 2009

JEL classification:
G12
C1
C5

Keywords:
Stochastic discount factors
Term structure models
Monte Carlo method
GMM estimator

ABSTRACT

This paper examines the feasibility of applying the stochastic discount factor methodology to fixed-income data using modern term structure models. Using this approach the researcher can examine returns on bond portfolios whose exact composition is unknown, as is often the case. This paper proposes an observable proxy for the SDF from continuous-time models and documents via Monte Carlo methods the properties of the GMM estimator based on using this proxy.

1. Introduction

Most empirical term structure work has focused on models for which the prices of some set of claims, usually zero-coupon bonds, are known functions of the state vector so the paths of the state variables can be inferred from bond prices. Typically likelihood-based time-series analysis is then applied to the paths of this state vector in order to estimate parameters (see Dai and Singleton (2003) for a survey).

By contrast, recent work in equity pricing models has been focused on constructing the stochastic discount factor (SDF) implied by a given model as a known function of observable variables and applying it to a set of returns. I will call this approach the SDF approach. The SDF approach has gained popularity partially because it lends itself well to straightforward testing, based on the Generalized Method of Moments (GMM) as in Hansen and Singleton (1982). An advantage of this approach is that very few assumptions need to be made on the distribution of the data.

Heretofore most applications of the SDF approach have been in discrete time. However an SDF is guaranteed to exist in any model, either in discrete or continuous time, in which arbitrage is not allowed. The purpose of this paper is to examine to what extent the SDF approach can be useful in estimating and testing continuous-time term structure models.

Typically term structure models fully specify the dynamics of the state vector under the true probability measure and so likelihood-based inference is (in principle) feasible. However in order to apply likelihood-based techniques the researcher must know the exact functional relationship between asset prices and the state vector. In many cases this is not known, as when dealing

☆ This paper has benefitted from useful comments by my chairman, Wayne Ferson, as well as Richard Bass. Comments by seminar participants at the University of California at Berkeley, the University of Illinois at Urbana-Champagne, Ohio State University, Columbia University, Arizona State University, the University of Arizona, the University of Utah, Brigham Young University and the University of Southern California are also acknowledged. All remaining errors are my own. I also gratefully acknowledge financial support provided by the Thomas Foster as well as the Samuel Stroum fellowships at the University of Washington.

* 15217 Isleview Dr. Chesterfield, MO 63017, USA.
E-mail address: farnsworth@wustl.edu.

0927-5398/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2009.06.004

with returns on portfolios of bonds. The SDF approach allows including such returns in the analysis, as was done by Ferson et al. (2006) in their study of bond mutual funds. Such an analysis is not possible using the likelihood-based approaches that are generally used in the term structure literature.

There are other advantages to the SDF approach over likelihood-based approaches. The first is the difficulty of evaluating the likelihood of the state vector. Except for a few special cases the likelihood of the state vector from a continuous-time model is typically not known in closed form. Numerical techniques to evaluate this likelihood are computationally expensive and present their own difficulties.

In addition, since no model fits the data exactly some joint distribution of pricing errors must be assumed in a likelihood-based approach. But this increases the number of parameters that need to be estimated. The SDF approach is robust to pricing errors, at least those which are orthogonal to the SDF, so there is no need to specify a statistical model for these errors or estimate any additional parameters.

Finally likelihood-based approaches do not provide an overall test of model fit. In the SDF approach there are over-identifying restrictions which can be tested so a given model can be rejected by the data.

Whether these considerations override the advantages of the likelihood-based methods is an open question and the answer may depend on the particular model and data set under consideration. In this paper I endeavor to show that the SDF approach is a feasible and somewhat attractive alternative. To be concrete I will restrict attention to models which fall into the affine class of Duffie and Kan (1996). This class of model is well-known in the literature and can be estimated using either approach.

I find that parameters are estimated with more accuracy using the SDF approach compared with a commonly used alternative (QML). This may be due to the fact that the SDF approach allows the use of more data as mentioned above. However I find that there are biases in the SDF approach which seem to be responsible for throwing off the distribution of the test for over-identifying restrictions. So the approach must be used with caution.

2. Methods

To understand the issues involved in applying the SDF methodology in continuous time one must understand what the SDF from a continuous-time model looks like.

Consider a continuous-time economy in which the uncertainty is driven by a d-dimensional Brownian Motion $W(t)$. Let $p_i(t)$ be the time t price of the ith non-dividend paying security and let $\boldsymbol{\sigma}_i(t)$ be the vector of sensitivities to $W(t)$. The absence of arbitrage implies that there exists some vector process $\Lambda(t)$ such that $p(t)$ solves

$$\frac{dp_i(t)}{p_i(t)} = r(t)dt + \sigma_i(t)'\Lambda(t)dt + \sigma_i(t)'dW(t)$$

where $r(t)$ is the instantaneous risk-free rate (see, for example, chapter 6 of Duffie (1996)).

A stochastic discount factor process is a scalar process $m(t)$ having the property that $m(t)p_i(t)$ is a martingale for all i. Therefore we must have

$$m(t)p_i(t) = \mathbb{E}_t[m(T)p_i(T)]$$

or

$$\mathbb{E}_t\left[\frac{m(T)}{m(t)}R_i(t,T)\right] = 1 \tag{1}$$

where $R_i(t,T)$ is the gross return on asset i between time t and T. By applying Ito's formula we can verify that $m(t)$ must satisfy

$$\frac{dm(t)}{m(t)} = -r(t)dt - \Lambda(t)'dW(t). \tag{2}$$

Solving the SDE (2) yields the following expression for the stochastic discount factor.

$$\frac{m(T)}{m(t)} = exp\left(-\int_t^T r(s)ds - \frac{1}{2}\int_t^T \Lambda(s)'\Lambda(s)ds - \int_t^T \Lambda(s)'dW(s)\right) \tag{3}$$

In order to test the stochastic discount factor for a particular model we must be able to construct $m(T)/m(t)$ from observed data. An examination of Eq. (3) suggests that in general it will not be possible to recover the SDF directly from discretely sampled data. Therefore we must construct an SDF proxy which is based on the observed data.

An affine model of the term structure is one in which bond prices satisfy

$$P(t,s) = exp(A(s-t) - B(s-t)'\boldsymbol{X}(t))$$

where $\boldsymbol{X}(t)$ is the state vector and A and B are deterministic functions of time. In an affine model the state vector, $\boldsymbol{X}(t)$, is assumed to solve[1]

$$X(s) = X(t) + \int_t^s K(\Theta - X(u))du + \int_t^s \Sigma \sqrt{\mathbf{S}(u)} dW(u) \tag{4}$$

where $\mathbf{S}(t)$ is a diagonal matrix whose diagonal is given by $\alpha + \beta' X(t)$. The short rate $r(t)$ is also assumed to satisfy $r(t) = \delta_0 + \delta_1' \boldsymbol{X}(t)$. The market price of risk process, $\Lambda(t)$ is assumed to be given by $\sqrt{\mathbf{S}(t)}\lambda$ for some vector $\boldsymbol{\lambda}$ so we have

$$\frac{m(T)}{m(t)} = exp\left(-\int_t^T (\delta_0 + \delta_1' X(u))du - \frac{1}{2}\int_t^T \lambda' \mathbf{S}(u)\lambda du - \int_t^T \lambda' \sqrt{\mathbf{S}(u)} dW(u)\right) \tag{5}$$

The difficulty is that this continuous-time process is not observed continuously so some discretization is necessary. There is a large literature on approximating diffusions via discrete processes using normally distributed random numbers as proxies for Brownian increments. Our aim is somewhat the reverse. Assuming that we have identified state variables we can only observe their paths at discrete points. We must approximate the terms in the SDF expression using these discrete observations.

If we did observe this process continuously then we could easily compute the integrals with respect to time in the SDF formula above. In that case we could obtain the stochastic integral using Eq. (4)

$$\lambda' \Sigma^{-1}[X(T) - X(t) - \int_t^T K(\Theta - X(u))du] = \int_t^T \lambda' \sqrt{\mathbf{S}(u)} dW(u). \tag{6}$$

Now suppose we only observe the state vector at discrete intervals of length Δ. Then we could approximate the integrals with respect to time by a (left) Riemann sum.

$$\int_t^{t+\Delta} r(u)du \approx (\delta_0 + \delta_1' X(t))\Delta \tag{7}$$

$$\int_t^{t+\Delta} \lambda' \mathbf{S}(u)\lambda du \approx \lambda' \mathbf{S}(t)\lambda\Delta \tag{8}$$

$$\int_t^{t+\Delta} \lambda' \sqrt{\mathbf{S}(u)} dW(u) \approx \lambda' \Sigma^{-1}[X(t+\Delta) - X(t) - K\Theta\Delta + KX(t)\Delta] \tag{9}$$

In what follows I will use $M(t,T)$ to denote this proxy to the true SDF, $m(T)/m(t)$.

Now clearly the quality of these approximations will be better for small Δ. In our simulation results we shall take Δ to be equal to 1 day since daily is a common frequency of observation that researchers have access too.

One of the peculiarities of fixed-income markets is that there are a few bonds which are very liquid and the others are much less so. In talking to bond traders one discovers that the reported daily prices on the less liquid bonds are viewed with considerable skepticism. However month-end numbers are considered more reliable. The extent to which this is true is beyond the scope of this paper. However it is the case that researchers typically have at their disposal daily series on bonds of a few maturities and monthly data on a larger set of bonds and portfolios of bonds. The SDF approach allows the researcher to utilize both types of data in one estimation. For instance, Ferson et al. (2006) use these approximations to study monthly returns on bond mutual funds with daily data on the state variable proxies.

In the next section we will use the approximations above at a daily frequency, assuming that the bonds we observe daily are observed without error, and then take products of daily SDFs to obtain monthly SDFs which we will then apply to monthly returns on portfolios.

3. Finite sample properties

In order to evaluate the performance of the SDF approach we must examine estimates using this approach when the true model is known. To do this we will simulate from a realistic model of the term structure that has multiple factors but is one which is difficult to estimate using the time-series approach because it is non-Gaussian and has no closed-form likelihood. We will simulate 15 years worth of data and then estimate parameters. I will repeat this two hundred times to get an idea of the sampling distribution of the estimators.

There is fairly broad agreement that three factors are adequate to describe the dynamics of U.S. yield curves. We will simulate daily yields on three bonds that we will use to solve for the three factors. In addition we will simulate monthly returns on coupon bonds of varying maturities. These monthly returns will be polluted with pricing errors so the covariance matrix of these returns will be full rank even though there are more than three bond returns.

[1] We are only considering affine models with state variable that have continuous sample paths. Affine jump diffusion models exist but examining these is beyond the scope of this paper.

The model we will use is from the $\mathbb{A}_1(3)$ branch of affine models. This notation means that the model is a three factor affine model with one process which drives the conditional variances of yields. The parameter values will be those estimated in Dai and Singleton (2000) using the parameterization which the authors call their "preferred" model. The state vector is the short rate, $r(t)$, its "central tendency" $\theta(t)$, and its volatility $v(t)$.

$$\begin{aligned} dr(t) &= \kappa(\theta(t)-r(t))dt + \sqrt{v(t)}dW_r(t)\ldots + \sigma_{r\theta}\zeta dW_\theta(t) + \sigma_{rv}\eta\sqrt{v(t)}dW_v(t) \\ d\theta(t) &= \nu(\overline{\theta}-\theta(t))dt + \sigma_{\theta r}\sqrt{v(t)}dW_r(t) + \zeta dW_\theta(t) \\ dv(t) &= \mu(\overline{v}-v(t))dt + \eta\sqrt{v(t)}dW_v(t) \end{aligned} \tag{10}$$

In terms of the general framework of affine models we have

$$X(t) = \begin{bmatrix} r(t) \\ \theta(t) \\ v(t) \end{bmatrix}, \quad K = \begin{bmatrix} \kappa & -\kappa & 0 \\ 0 & \nu & 0 \\ 0 & 0 & \mu \end{bmatrix}, \quad \Theta = \begin{bmatrix} \theta \\ \theta \\ v \end{bmatrix}$$

$$\sum = \begin{bmatrix} 1 & \sigma_{r\theta} & \sigma_{rv} \\ \sigma_{\theta r} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \alpha = \begin{bmatrix} 0 \\ \zeta^2 \\ 0 \end{bmatrix}, \quad \beta = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & \eta^2 \end{bmatrix}$$

where we note that α and β, defined as above, give us

$$S(t) = \begin{bmatrix} v(t) & 0 & 0 \\ 0 & \zeta^2 & 0 \\ 0 & 0 & \eta^2 v(t) \end{bmatrix}.$$

3.1. Simulation strategy

Since this model has no closed-form density it is also difficult to sample from. However the process which determines the conditional variance, $v(t)$ is Markovian and in fact of CIR form. We can sample paths of this state variable which have exactly the right conditional distributions using well-known methods (see Glasserman, 2004 for details). I sampled 4 points per day and then simulated the other factors as Gaussian conditional on the path of this conditional volatility process.

Ideally we would have liked to initialize each path by drawing from the unconditional distribution of the state variables. However this unconditional distribution is not known analytically and so sampling from it directly is not feasible. So instead I started each path at the unconditional mean of the process and simulated 25 years of data. Then I threw away the first 10 years so that the remaining portion of each path is started at a random point drawn from a distribution which ought to be close to the unconditional distribution.

For each path of the state variables I calculated two kinds of price data: (1) daily yields on three zero-coupon bonds with maturities of 6 months, 3 years, and 20 years and (2) monthly returns on 6 coupon bonds. The daily data is simulated exactly according to the model. The 6 coupon bonds were simulated to be at par at the beginning of each month with maturities of 2, 5, 7, 10, 15, and 20 years. At the end of each month the bonds are re-priced to compute a return. Then pricing errors were added as IID normal noise with mean zero and standard deviation of 3 basis points per month.

3.2. Estimation

In the estimation stage of the Monte Carlo the econometrician is assumed to know that the daily data has no pricing error and that the monthly data has error. But the econometrician does not know either how much error is in the monthly returns or even what kind of bond or portfolio the monthly returns correspond to.

Since the yields of the three bonds observed daily are assumed to be observed without error we have that the yield of the discount bond with time-to-maturity τ, y^τ satisfies.

$$y(t)^\tau = -a(\tau) + b(\tau)' X(t)$$

where $\alpha(\tau) = A(\tau)/\tau$ and $b(\tau) = B(\tau)/\tau$. By stacking these equations for the 3 τs we obtain the following system of equations.

$$Y(t) = -\mathbf{a} + \mathbf{b}X(t).$$

The estimation routine uses the current guess of the parameters to form **a** and **b** and solves for the implied levels of the state variables for each date.

$$X(t) = \mathbf{b}^{-1}(Y(t) + \mathbf{a}) \tag{11}$$

The SDF proxy is formed by the econometrician from these daily observations of the state variable.

Table 1
Monte Carlo results using SDF method.

	5%	25%	Median	True	Mean	75%	95%
κ	5.494	15.788	16.535	17.400	15.311	17.122	18.122
ν	0.206	0.220	0.227	0.226	0.225	0.232	0.261
μ	0.176	0.285	0.361	0.365	0.350	0.405	0.504
$\overline{\theta}$	0.049	0.068	0.078	0.083	0.077	0.084	0.110
$\overline{\nu}$	0.007	0.012	0.015	0.015	0.015	0.018	0.025
$\sigma_{r\theta}$	−4.437	−3.573	−2.777	−3.420	−1.434	−1.755	7.033
σ_{rv}	3.213	4.682	6.103	4.270	15.806	8.130	15.035
$\sigma_{\theta r}$	−0.134	−0.112	−0.098	−0.094	−0.097	−0.081	−0.057
ξ^2	4.8e−5	1.5e−5	2.1e−4	2.0e−4	3.0e−4	3.1e−4	8.1e−4
η^2	0.000	0.001	0.005	0.008	0.024	0.023	0.122
λ_r	8.510	10.355	12.337	9.320	14.552	16.630	28.032
λ_θ	4.113	25.055	36.591	31.700	43.829	50.844	142.893
λ_v	−16.961	−5.191	−0.731	−0.344	−22.073	2.290	17.552

Means and percentiles of estimates compared with true values.

Since coupon bond is a portfolio of zero-coupon bonds the gross return on a coupon bond is the gross return on a portfolio of zeros. Using these portfolio weights and taking weighted averages of equations like Eq. (1) we see that the same equation holds for coupon bonds. By taking unconditional expectations and plugging in our proxy we obtain the following

$$\mathbb{E}[M(t,T)R_c(t,T)-1] = 0$$

where $R_c(t,T)$ is the gross return on the coupon bond from t to T (one month). In order to minimize the impact of any bias that the discretization introduces into the mean of our SDF proxy we instead use the following moment condition

$$\mathbb{E}[M(t,T)(R_c(t,T)-R_f(t))] = 0 \tag{12}$$

where R_f is the return on a 1-period risk-free asset observed at time t. Using the 6 coupon bond returns gives us 6 moment conditions. Additional moment conditions were obtained by imposing orthogonality with the beginning of month values of $\boldsymbol{X}(t)$, i.e.

$$\mathbb{E}[M(t,T)(R_c(t,T)-R_f(t))\otimes X(t)] = 0 \tag{13}$$

where ⊗ is the Kronecker product. This gives us an additional 18 moments.

Besides the pricing moments above there are moment conditions we can impose based on the assumed stochastic process for the state variables. Some interesting moments are not known in closed form and so cannot be used.[2] But for all affine models the conditional mean is known in closed form.

$$\mathbb{E}[X(t+\Delta)\,|\,X(t)] = \Theta - exp(-K\Delta)(X(t)-\Theta)$$

We could derive moment conditions directly from this equation but notice that these moment conditions are assumed to hold at a daily frequency rather than a monthly frequency like the pricing moments. The only complication here is in calculating the covariance of these daily moments with the monthly moments to form the weighting matrix. So instead I defined the following monthly moment conditions

$$\mathbb{E}\left[\sum_{i=0}^{d} X(t+(i+1)\Delta)-\Theta-exp(-K\Delta)(X(t+i\Delta)-\Theta)\right] = 0 \tag{14}$$

$$\mathbb{E}\left[\left(\sum_{i=0}^{d} X(t+(i+1)\Delta)-\Theta-exp(-K\Delta)(X(t+i\Delta)-\Theta)\right)\otimes X(t)\right] = 0 \tag{15}$$

where exp is the matrix exponential and d is the number of trading days in a month (assumed to be 21 in the simulation). This gives us 12 addition moment conditions. Combining Eqs. (12), (13), and (14) gives us 36 moments to use in estimating 13 parameters.

The estimation employs the so-called *two-step* GMM technique. In the first step an identity matrix is used as the weighting matrix. In the second step a new weighting matrix is calculated using the first-stage estimates which are then used in a second-

[2] For instance the scores of the likelihood of the transition density of the state variables are arguably the best moments to match. But for this affine model the scores are not known in closed form.

Table 2
Monte Carlo results using quasi-maximum likelihood.

	5%	25%	Median	True	Mean	75%	95%
κ	2.286	5.258	9.372	17.400	12.639	17.757	29.593
ν	0.001	0.079	0.161	0.226	0.197	0.246	0.619
$\underline{\mu}$	0.000	0.015	0.118	0.365	0.459	0.518	1.946
$\overline{\theta}$	0.007	0.025	0.051	0.083	0.328	0.152	0.956
$\overline{v}$	0.000	0.004	0.068	0.015	3.550	1.467	18.839
$\sigma_{r\theta}$	−6.419	0.038	5.786	−3.420	12.809	19.228	54.005
σ_{rv}	−18.087	−1.031	2.799	4.270	7.128	13.364	51.508
$\sigma_{\theta r}$	−0.026	−0.015	−0.009	−0.094	−0.011	−0.004	−0.001
ξ^2	5.6e−4	0.004	0.032	2.0e−4	61.804	0.130	282.234
η^2	0.001	0.007	0.033	0.008	1.643	0.273	4.861
λ_r	0.399	1.880	8.144	9.320	9.760	14.014	25.751
λ_θ	−9.099	−1.217	4.770	31.700	47.775	42.348	200.587
λ_v	−4.025	−1.349	−0.503	−0.344	−0.809	−0.000	1.964

Means and Percentiles of Estimates Compared with True Values.

step estimation. To deal with time-series dependence in the moment conditions this second-step weighting matrix was calculated using the Newey and West (1987) estimator with 5 moving average terms.

The mean and median parameter values across the 200 simulations along with the true values are shown in Table 1. Also included in the table are 5th, 25th, 75th, and 95th percentiles of the estimated parameter values to give an idea of the distribution of the estimates.

In order to gauge the accuracy of the SDF approach we need to compare to some other technique. Exact maximum likelihood estimation is not possible for this model but another common technique, quasi-maximum likelihood (or QML), is. The QML approach to affine term structure models is due to Fisher and Gilles (1996). The idea is to compute the conditional mean and variance of the state vector at each date and use the Gaussian likelihood with this conditional mean and variance in place of the true likelihood.[3]

To find the true moments of an affine diffusion note that we can rewrite the process as

$$\boldsymbol{X}(s) = \Theta + \Phi(s-t)(\boldsymbol{X}(t)-\Theta) + \int_t^s \Phi(u-t)\sum\sqrt{\mathbf{S}(u)}dW(u)$$

where $\Phi(u-t) = \exp(-K(u-t))$. Since the last term on the right is a martingale the conditional mean is clearly

$$\mathbb{E}_t(\boldsymbol{X}_s) = \Theta + \Phi(s-t)(\boldsymbol{X}_t-\Theta)$$

as noted above. The conditional variance is a bit more complex. But recall that for integrals with respect to Brownian motion we have that

$$Var_t\left[\int_t^s g(u)dW(u)\right] = \mathbb{E}_t\left[\left(\int_t^s g(u)dW(u)\right)^2\right] = \mathbb{E}_t\left[\int_t^s g(u)g'(u)du\right].$$

Using this fact we can write the conditional variance as

$$Var_t(\boldsymbol{X}(s)) = \mathbb{E}_t\left[\int_t^s \Phi(u-t)\sum\mathbf{S}(u)\sum{}'\Phi(u-t)'du\right] = \int_t^s \Phi(u-t)\sum\mathbb{E}_t[\mathbf{S}(u)]\sum{}'\Phi(u-t)'du$$

which is easy to evaluate since **S** is linear in $\boldsymbol{X}$. Computing the conditional variance is simply a matter of a numerical integration. However this numerical integration must be done for each day for each guess of the parameters for each simulated path. This means that the QML estimation takes much longer than the SDF estimation.

Since the econometrician is assumed not to know the nature of the gross returns $R_c(t,T)$ there is no way to incorporate this data into the QML estimation. So this comparison may seem somewhat artificial since the two methods are not using the same data. But this is exactly the situation econometricians find themselves in. In order to implement methods like QML one must throw away or ignore some data because the exact cash-flow structure is not known. So the only data used in the QML estimation is the 15 years of daily observations of the 3 bonds assumed to have no pricing errors.

One could argue that in most cases there are more than 3 bonds available at daily frequency whose cash-flow structures are known. In fact these other bonds could be used to help back out $\boldsymbol{X}(t)$ if we think that all bonds are priced with some error. For instance in an affine model with Gaussian state variables the Kalman filter could be used. Even in a non-Gaussian model like the one we are examining an approximate Kalman filter could be used to extract the paths of the factors from the bond data. However this would not help our comparison because whatever method is used to back out state variables can be used in both the QML and SDF framework.

[3] Keeping in mind that we solve for the factors using Eq. (11) we must also include the Jacobian of the transformation $1/|\mathbf{b}|$ in the likelihood.

Table 3
Bias (expressed in basis points per month) in conditional pricing moment $\mathbb{E}_t[m(R-R_f)]$ evaluated at various percentiles of the distribution of estimates.

	5%	25%	Median	75%	95%
κ	2.570	1.969	1.933	1.904	1.857
ν	1.883	1.889	1.891	1.893	1.905
μ	1.888	1.890	1.891	1.891	1.893
$\overline{\theta}$	19.620	9.654	4.483	1.023	−12.553
$\overline{v}$	8.680	4.492	1.879	−0.987	−6.757
$\sigma_{r\theta}$	2.128	1.919	1.792	1.693	2.001
σ_{rv}	2.001	1.847	1.695	1.470	0.643
$\sigma_{\theta r}$	−9.723	−3.433	0.762	6.037	13.289
ξ^2	−9.453	−2.164	2.547	10.025	47.747
η^2	2.418	2.342	2.083	0.889	−4.636
λ_r	4.186	−1.039	−6.641	−18.701	−49.955
λ_θ	−10.833	−1.181	4.154	10.760	53.721
λ_v	−3.024	0.462	1.777	2.666	7.131

The state vector was set to its sample mean and the asset return was the monthly return on a zero-coupon bond with 5 years to maturity at the beginning of the month.

The distribution of the QML estimates is given in Table 2. Comparing this with Table 1 we see that while some bias is apparent in both methodologies the bias seems more severe in the QML table. Also comparing the difference between the 25% and 75% columns for each method it seems that the distribution of the estimates is tighter for the SDF method. This seems to favor the SDF method over the QML method.

However the test of over-identifying restrictions in the SDF approach seems to have a problem. The mean across simulation of the *J*-statistic is about what one would expect (21.01 for a χ^2 with $36-13=23$ degrees of freedom) but seems to have too small a variance (Fig. 1). Using the asymptotic distribution the 5% critical value of this test is 35.17 but there were no observed J-stats this high. The Monte Carlo 5% critical value is about 25. It may be that the data series is simply not long enough for the *J*-statistic to get close to its asymptotic distribution. Alternatively the bias that the SDF proxy introduces into the moment conditions may be the source of the problem. I examine this bias in the next section.

4. Evaluating discretization bias

With an affine model it is possible to obtain analytical expressions for the bias introduced by the discrete approximation to the SDF. Note that the proxy examined in this paper has an exponential-affine form, i.e.

$$exp(a+b'\boldsymbol{X}(t_i)-c'\boldsymbol{X}(t_{i+1}))$$

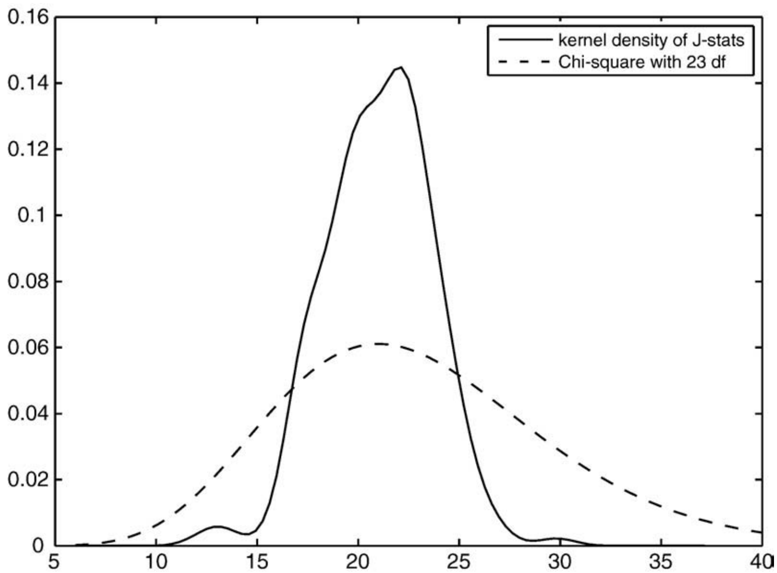

Fig. 1. Kernel density of *J*-statistic and asymptotic distribution.

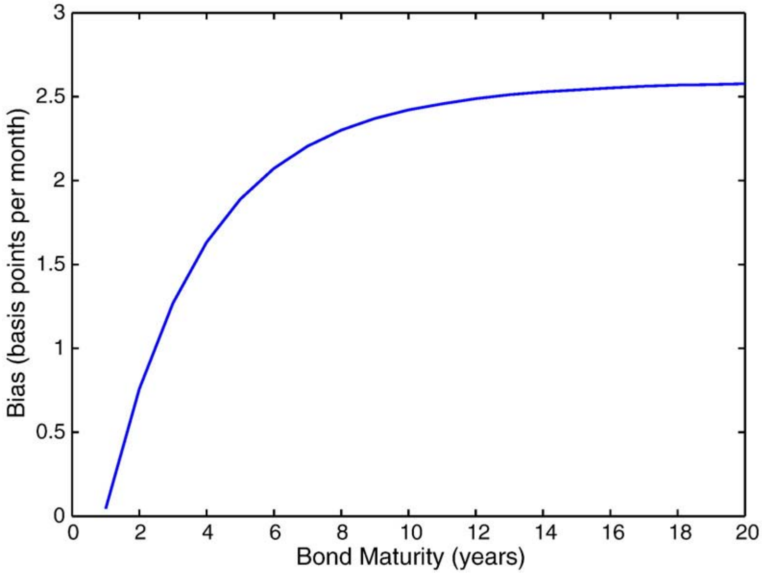

Fig. 2. Term structure of bias.

where

$$a = (-\delta + \lambda^{'}\textstyle\sum^{-1}K\theta - \frac{1}{2}\alpha^{'}\lambda 2)\Delta$$

$$b = -(\delta + \textstyle\sum^{-1}\lambda - K^{'}\textstyle\sum^{-1}\lambda - \frac{1}{2}\beta\lambda 2$$

$$c = \textstyle\sum^{-1}\lambda.$$

Now consider applying the SDF proxy to the gross return between time $t = t_0$ and time t_d of a pure discount bond which matures at time $s > t_d$. This gross return is given by

$$R = exp(A(s-t_d) - B(s-t_d)^{'}X(t_d) - A(s-t) + B(s-t)^{'}\boldsymbol{X}(t)).$$

The SDF approach is based on the following conditional moment

$$\mathbb{E}_t\left[exp\left(\sum_{i=0}^{d-1}(a + b^{'}\boldsymbol{X}(t_i) - c^{'}\boldsymbol{X}(t_{i+1}) + A(s-t_d) - B(s-t_d)^{'}\boldsymbol{X}(t_d) - A(s-t_0) + B(s-t_0)^{'}\boldsymbol{X}(t_0)\right)\right]. \quad (16)$$

Conditional moments of the form

$$\mathbb{E}_t[exp(-v^{'}\boldsymbol{X}(u))]$$

can be computed in closed form for affine models as

$$exp(C(u-t) - D(u-t)^{'}\boldsymbol{X}(t))$$

where the function C and D solve the following ODEs

$$\frac{dC}{d\tau} = -\theta^{'}K^{'}D(\tau) + \frac{1}{2}\sum_{k=1}^{N}[\textstyle\sum^{'}D(\tau)]_k^2\alpha_k$$

$$\frac{dD}{d\tau} = -K^{'}D - \frac{1}{2}\sum_{k=1}^{N}[\textstyle\sum^{'}D]^2\beta k$$

where $\tau = u - t$, $\boldsymbol{\beta}_k$ is the kth column of the matrix $\boldsymbol{\beta}$ and where the subscript τ indicates partial derivative. The initial conditions for these ODEs are $C(0) = 0$ and $D(0) = v$. We can use this fact recursively to evaluate the above conditional moment as follows.

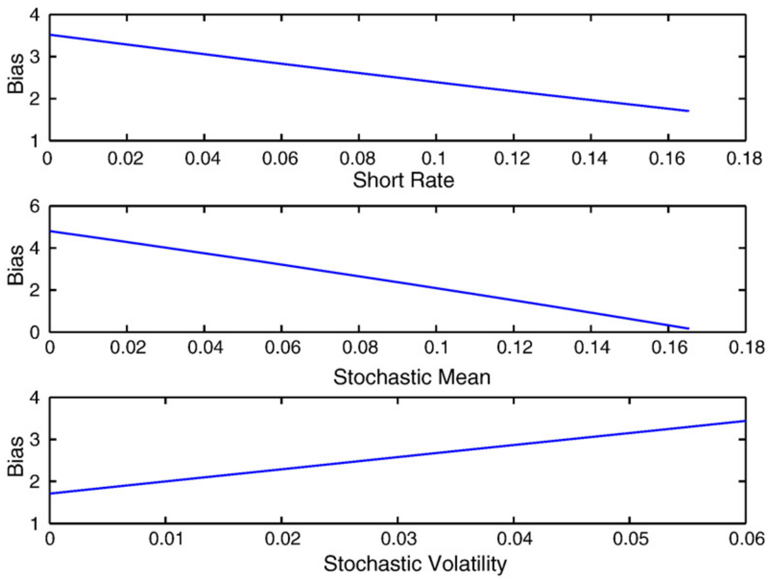

Fig. 3. Bias as a function of the state variables. The parameters are set to their true values and the asset is a discount bond with 5 years to maturity.

First we take t to be t_{d-1} and set $v = B(s - t_d) + c$ because the vectors that multiply $\boldsymbol{X}(t_d)$ in Eq. (16) are $B(s - t_d)$ and c. Using this v we compute C and D. Then we step back a day and take $t = t_{d-2}$, set $v = c - b + D$ and compute a new C and a new D. At t_{d-3} we have $v = c - b + D$ (using the D we computed in the previous step) and so forth until date t_0. Once we have the conditional mean of the SDF proxy times the gross return of the risky asset we calculate the conditional mean of the SDF proxy times the risk-free return (using a similar recursion) and subtract one from the other to compute the bias.

This recursion does not give us the bias in the unconditional moments used in the GMM estimation. It gives the bias in the conditional moments used to derive the unconditional moments.

Applying the above recursion to bonds of various maturities we obtain a term structure of bias. This is plotted in Fig. 2. Note that if the term structure of bias were flat, even at a non-zero level then the bias would be a less serious problem. We could simply incorporate it into the moment conditions used in estimation by replacing Eqs. (12) and (13) with

$$\mathbb{E}[M(t,T)(R_c(t,T) - R_f(t) - bias(t,T))] = 0$$

and

$$\mathbb{E}[M(t,T)(R_c(t,T) - R_f(t,T) - bias(t,T)) \otimes \boldsymbol{X}(t)] = 0$$

respectively. However, since the term structure of bias is not flat we would have to know the exact cash-flow structure of each portfolio $R_c(t,T)$ to calculate the right bias to subtract. This effectively eliminates one of the advantages of the SDF approach. And it complicates analyses like that of Ferson et al. (2006) since typically the researcher does not know the exact composition of a bond mutual fund.[4] An approximate solution would be to use returns net of a benchmark index which has very similar yield curve exposures (similar maturity and sensitivities to the state variables). However it might be hard to determine, even approximately, the sensitivities of a fund about whose composition we know very little.

The magnitude of the bias is also a function of the state variables and parameters of the model. In Fig. 3 I plot the bias as a function of each state variable while holding the other state variables at their unconditional means. In Table 3 I have calculated the bias for various values of the parameters. Each row of the table corresponds to choosing values for the named variable from the distribution the SDF estimates in Table 1 and setting all the other parameters equal to their true values.

Ferson et al. (2006) find that the average bias associated with bond index returns using this proxy and a similar 3-factor model was between 5 and 6 basis points per month. Fig. 2 suggests that less than half of that could be due to the discretization. However Fig. 3 shows that this bias could vary substantially with the path of the state vector so it is possible that all of the bias they experienced could be due to discretization bias in the proxy. Ferson et al. (2006) also find that bond mutual funds have risk-adjusted performance of about −6 to −8 basis points per month. The existence of this bias may indicate that the true risk-adjusted performance could be much worse.

[4] Even if the holdings were known this would not solve the problem. The holding would have to be constant during the month in order to calculate the bias. This is seldom the case for a mutual fund.

5. Conclusion

This paper has examined a new technique for estimating and testing term structure models. The method is based on approximating the SDF from a model using daily observations on a few securities and then applying this proxy SDF to monthly returns on other securities. The monthly returns can be portfolios or other assets whose exact cash-flow pattern is not known. So data can be analyzed using this method that cannot be analyzed using likelihood-based techniques.

Within the realm of affine models it appears that the method performs well, especially when compared with QML, a commonly used alternative. The bias in the estimates is less and the distribution of the estimated parameters appears to be tighter. There is, however a bias introduced by the discretization inherent in the proxy. Also the distribution of the test for over-identifying restrictions is far from its theoretical asymptotic distribution. A more detailed analysis may determine whether this is related to the bias issue. I leave this to future research.

Another possible fruitful direction for future research is finding better SDF proxies. For instance if we replace the left Riemann sum approximation to the integrals in Eq. (5) with a more accurate approximation we might get better results. Preliminary results suggest that using a trapezoidal sum reduces the bias substantially.

The technique could also be extended to models that are not *strictly affine*. In a strictly affine model the price of risk is $\sqrt{\mathbf{S}(t)}\lambda$, a linear function of the volatility of the state vector, $(\sum\sqrt{\mathbf{S}(t)})$. Duffee (2002) and Duarte (2004) have argued for more general classes of models (called respective *Essentially Affine* and *Semi Affine Square-Root*) in which the volatility process is the same but the price of risk can depend on the state vector in a more general way. In that case getting from

$$\int_t^{t+\Delta}\sum\sqrt{\mathbf{S}(u)}dW(u)$$

to

$$\int_t^{t+\Delta}\Lambda(u)dW(u)$$

is a bit more complicated. A likely approximation might be

$$\int_t^{t+\Delta}\Lambda(u)dW(u) = \Lambda_t\sqrt{\mathbf{S}(u)}^{-1}\sum{}^{-1}\int_t^{t+\Delta}\sqrt{\mathbf{S}(u)}dW(u).$$

How much of a bias this approximation might introduce is an open question which shall be left to future research.

References

Dai, Q., Singleton, K.J., 2000. Specification analysis of affine term structure models. Journal of Finance 55 (5), 1943–1978.
Dai, Q., Singleton, K.J., 2003. Term structure dynamics in theory and reality. Review of Financial Studies 16 (3), 631–678.
Duarte, J., 2004. Evaluating an alternative risk preference in affine term structure models. Review of Financial Studies 17, 379–404.
Duffee, G.R., 2002. Term premia and interest rate forecasts in affine models. Journal of Finance 57, 405–443.
Duffie, D., 1996. Dynamic Asset Pricing Theory, 2nd edn. Princeton University Press, Princeton, New Jersey.
Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical Finance 6, 379–406.
Ferson, W., Henry, T., Kisgen, D., 2006. Evaluating government bond fund performance with stochastic discount factors. Review of Financial Studies 19 (2), 423–455.
Fisher, M., Gilles, C., 1996. Estimating exponential-affine models of the term structure. Working Paper Federal Reserve Board, Washington, D.C.
Glasserman, P., 2004. Monte Carlo Methods in Financial Engineering. Springer-Verlag, New York.
Hansen, L.P., Singleton, K.J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50 (5), 1269–1286.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.