Evaluating Government Bond Fund Performance with Stochastic Discount Factors

Wayne Ferson
Boston College and NBER

Tyler R. Henry
University of Georgia

Darren J. Kisgen
Boston College

This article shows how to evaluate the performance of managed portfolios using stochastic discount factors (SDFs) from continuous-time term structure models. These models imply empirical factors that include time averages of the underlying state variables. The approach addresses a performance measurement bias, described by Goetzmann, Ingersoll, and Ivkovic (2000) and Ferson and Khang (2002), arising because fund managers may trade within the return measurement interval or hold positions in replicable options. The empirical factors contribute explanatory power in factor model regressions and reduce model pricing errors. We illustrate the approach on US government bond funds during 1986–2000.

Recent years have witnessed an explosion of research on the performance of mutual funds, with most of the attention focussed on equity-style funds. The amount of work on fixed-income funds is small in relation to their importance in the economy. As of June 2002, there were 2057 bond funds in the United

We are grateful to Warren Bailey, Jeffrey Busse, Yong Chen, Campbell Harvey (the Editor), Edie Hotchkiss, Clifton Green, Eric Jacquier, Raymond Kan, Stanley Kon, Anthony Lynch, Shingo Goto, Steven Mann, Fabio Moneta, Meijun Qian, Richard Roll, Rudi Schadt, Russ Wermers, and an anonymous referee for help or suggestions. This paper was presented at Boston College, Harvard, HEC Paris, the University of Maryland, MIT, Princeton, the University of South Carolina, the 2003 Conference on Financial Economics and Accounting, the 2003 Northern Finance Association meetings, and the 2004 Conference on Emerging Markets: Innovations in Portfolio Management at the University of Virginia, and the 2004 CIREQ-CIRANO Conference on Macroeconomics and Finance: The Term Structure of Interest Rates. An earlier pilot study benefited from workshops at Babson College, Boston University, Brandeis, the Universities of California at Los Angeles and Riverside, Cornell, the Federal Reserve Bank of Atlanta, the London Business School, London School of Economics, New York University, Northwestern, the Said Business School at Oxford University, the University of Texas at Dallas, Utah, the Warwick Business School, and Wharton. Portions of this work were also presented at the 2003 Berkeley Program in Finance, the 2003 Boston College Finance Advisory Conference, the 2003 Chicago Quantitative Alliance conference, the 2003 European Finance Meetings, and the 2003 McGill Conference on Global Asset Management. The authors appreciate financial support from the Gutmann Center for Portfolio Management at the University of Vienna and the Institute for Quantitative Research in Finance. Address correspondence to Wayne Ferson, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, or email: wayne.ferson@bc.edu.

doi:10.1093/rfs/hhj015 Advance Access publication January 19, 2006

States, representing 25% of all mutual funds. Assets under management by these funds totalled just over $1 trillion, or 15% of the $6.6 trillion in mutual fund assets. The number of fixed-income funds and assets under management increased by 97 and 245%, respectively, between 1990 and 2002.[1]

One may expect that the variation in fixed-income fund performance should be small, relative to that of equity funds. However, given the relatively low volatility of fixed-income returns, the standard errors associated with measures of performance are also small. Fixed-income managers have a variety of tools to use in their quest to outperform their peer funds and benchmarks. They can tune the interest rate sensitivity (duration, convexity) of the portfolio to time changes in the level or the shape of the yield curve in response to economic developments. They can vary the allocation to asset classes (e.g., Treasury versus agency securities). They can attempt to exploit liquidity differences, such as those associated with on-the-run or off-the-run issues. They can strategically manage their securities lending operations. They can trade a host of interest rate derivative products in implementing these strategies. Active fixed-income fund management would seem to present a rich, undeveloped field for research.[2]

This article measures the performance of government bond mutual funds, using continuous-time models, in a *stochastic discount factor* (SDF) framework. Continuous-time term structure models identify the SDFs, and we temporally aggregate them for pricing monthly returns. The SDFs are simple exponential functions of term structure factors. But they also contain new empirical factors—time averages of the models' state variables. We find that these new factors are statistically and economically significant for explaining returns. The main goal of the article is to motivate and illustrate this approach.

Using continuous-time models addresses an "interim trading bias" in performance measurement that arises when monthly returns are measured, while managers trade during the month [Goetzmann, Ingersoll, and Ivkovic (2000), Ferson and Khang (2002)]. A related problem is option-like positions. Interim trading biases may be especially relevant for fixed-income portfolios that use derivatives. Ferson and Khang propose a weight-based measure of performance to avoid these biases. With our approach, portfolio weights are not needed. This is convenient since data on the holdings of funds typically are available on a less-frequent basis than the funds' returns.

Given an SDF derived from a particular model, a measure of abnormal return or *SDF alpha* is easily constructed [Chen and Knez (1996)]. This

[1] These figures exclude balanced funds, which hold a mix of bonds and stocks. Sources: the Investment Company Institute, *Trends in Mutual Fund Investing*, June 2002, and 2002 *Mutual Fund Handbook*.

[2] Well-known studies that focus on US fixed-income funds include Blake, Elton, and Gruber (1993) and Elton, Gruber, and Blake (1995). Cornell and Green (1991) examine low-grade bond funds. Dahlquist, Engstrom, and Soderlind (2000) include bond and money market funds in their Swedish fund sample. The practitioner-oriented literature contains additional references.

approach lends itself naturally to *Conditional Performance Evaluation*, where funds' alphas are conditioned on ex-ante economic states. Term structure models in particular prescribe the state variables on which to condition. This removes some of the ambiguity in instrument selection that is typical of the conditional asset pricing literature.

Ideally, one would like an SDF model or a set of factors to price both stocks and bonds. Empirically, however, this is challenging. Roll (1970) found that the capital asset pricing model [CAPM, Sharpe (1964)] does not work well for bonds. Mehra and Prescott (1985) observe that simple consumption models can not price both Treasury bills and stocks. Multiple-factor models with both bond and stock-related factors appear to fare better [e.g., Ferson and Harvey (1991), Campbell (1996)]. However, it is more common to find bond factors used for pricing bonds and stock factors for pricing stocks. We stick with this tradition, using term structure models to price government bond funds.

We evaluate the SDF models using passive benchmarks and dynamic bond portfolio strategies. The SDF models explain large fractions of the conditional expected returns on these portfolios. For example, returns of the Lehman Government bond index vary between 61 and 132 basis points (bp) per month over the 1986–2000 period, depending on the state of the term structure. After risk adjustment with a three-factor SDF model the conditional alphas are less than 3 bp.

We evaluate the performance of a sample of US government bond funds. During 1986–2000, the funds returned less on average than benchmark portfolios that do not pay expenses. Conditioning on the level, slope, and convexity of the term structure, we find more variation in funds' conditional expected returns across the term structure states than we find across fund groups formed according to characteristics like expense ratios, size, age, and past returns. When we adjust for risk, we find that the conditional performance is precisely estimated, slightly negative but economically small. We can not reject the hypothesis that performance gross of fund costs is neutral.

The rest of the article is organized as follows. Section 1 presents the models. Section 2 discusses the interim trading bias. Section 3 describes our estimation methods. Section 4 describes the data. Section 5 presents a comparison of our approach with linear factor models and results on the estimation of the SDF models for benchmark strategies. Section 6 evaluates the government bond fund performance. Section 7 offers concluding remarks.

1. The Models

Most asset pricing models specify an SDF, ${}_tm(\phi)_{t+1}$, a scalar random variable that depends on data observed up to time $t+1$ and parameters ϕ, such that the following equation holds

$$E_t\left[{}_tm(\phi)_{t+1}R_{t+1}\right]=\underline{1} \tag{1}$$

The notation ${}_tm(\phi\,)_{t+1}$ is chosen to emphasize that the SDF refers to a discrete time interval, in our case one month, that begins at time t and ends at time $t+1$. R_{t+1} is an N-vector of gross (i.e., one plus) "primitive" asset returns over the same period, 1 is an N-vector of ones and $E_t(.)$ denotes the conditional expectation, given the information in the model at time t. We say that the SDF "prices" the primitive assets if Equation (1) is satisfied. We allow that a mutual fund, with return $R_{p,t+1}$, may not be priced exactly by the SDF. Its SDF alpha is defined following Chen and Knez (1996) and Farnsworth, et al. (2002) as

$$\alpha_{pt} \equiv E_t\left[{}_tm(\phi)_{t+1}R_{p,t+1} - 1\right] \tag{2}$$

The SDF alpha is proportional to the traditional alpha in a beta pricing representation, when the SDF is linear in the factors. For example, in the case of the CAPM, the SDF is linear in the market return and α_p is proportional to Jensen's (1968) alpha.[3]

1.1 SDFs from term structure models

Popular term structure models specify a continuous-time stochastic process for the underlying state variable(s). For example, let X be a state variable following a diffusion process:

$$dX = \mu(X_t)dt + \sigma(X_t)dw, \tag{3}$$

where dw is the local change in a standard Wiener process. The state variable may be the level of an interest rate, the slope of the term structure, and so on. A term structure model specifies the functions μ (.) and σ (.) and the form of a market price of risk, $q(X)$, associated with the state variable. The market price of risk is the expected return in excess of the instantaneous interest rate, per unit of state variable risk.[4] Multiple state variable models are described below.

Term structure models based on Equation (3) can be shown [using Girsanov's Theorem, see Cox, Ingersoll, and Ross (1985b) or Farnsworth (1997)] to imply SDFs of the following form:

[3] See Ferson (1995, 2003) and Cochrane (2001) for discussions and interpretations of SDF alphas.

[4] The literature has directed a lot of firepower at accurately modelling continuous-time processes like Equation (3). For example, when X is a short-term interest rate and $\sigma(X) = X^{\gamma}$, studies following Chan et al. (1992) debate whether γ is 0.5, 1.0, 1.5, or some other number. Other studies ask whether the drift of the process is linear or nonlinear in X [see, e.g, Ait-Sahalia (1996), Stanton (1997), Chapman and Pearson (2000), Balduzzi and Eom (2002), Duarte (2004), and Jones, (2003)]. As we show below, our approach does not require all of the structure implied by a process like Equation (3) and should therefore be robust to some misspecifications of the stochastic process.

$$
\begin{aligned}
{}_t m(\phi)_{t+1} &= \exp(-\mathrm{A}_{t+1} - B_{t+1} - C_{t+1}), \quad \text{where} \\
A_{t+1} &= \int_t^{t+1} r_s ds, \\
B_{t+1} &= \int_t^{t+1} q(X_S) dw_S \\
C_{t+1} &= \left(\frac{1}{2}\right) \int_t^{t+1} q(X_S)^2 ds,
\end{aligned}
\tag{4}
$$

where r_s is the instantaneous interest rate at time s. We apply a first-order Euler approximation to Equation (3), to use the term structure models with monthly mutual fund data:

$$
X(t+\Delta) - X(t) \approx \mu(X_t)\Delta + \sigma(X_t)[w(t+\Delta) - w(t)]. \tag{5}
$$

The period between t and $t+1$ is divided into $1/\Delta$ increments of length Δ. The period is one month, and we divide it into increments of one day. The terms A_{t+1}, B_{t+1}, and C_{t+1} in Equation (4) are approximated using daily data by[5]

$$
\begin{aligned}
A_{t+1} &\approx \sum_{i=1,\ldots\frac{1}{\Delta}} r\,[t+(i-1)\Delta]\Delta \\
B_{t+1} &\approx \sum_{i=1,\cdots\frac{1}{\Delta}} q\{X[t+(i-1)\Delta]\}\{w(t+i\Delta) - w[t+(i-1)\Delta]\} \\
C_{t+1} &\approx \left(\frac{1}{2}\right) \sum_{i=1,\cdots\frac{1}{\Delta}} q\{X[t+(i-1)\Delta]\}^2\Delta.
\end{aligned}
\tag{6}
$$

We illustrate our approach on a collection of classical models. Three models are special cases of the three-factor affine model:

---

[5] Stanton (1997) evaluates the accuracy of similar first-order approximation schemes. He concludes that with daily data approximating a monthly integral, the approximations are almost indistinguishable from the true functions over a wide range of values. He also evaluates higher order approximation schemes and finds that they offer negligible improvements. Sun (1997) evaluates the approximations in (6), to estimate the parameters of continuous-time models of the short rate and finds that they can be improved upon with a test function method. Thus, the approximations may not be adequate for all applications. However, Gourieroux, Montfort, and Polimenis (2002) show that if we start with the assumption that an affine model holds for the discrete period of length Δ, then the summations in Equation (6) apply to the longer holding period returns without any approximation.

$$
\begin{aligned}
dr &= K_1(\theta_1 - r_t)dt + \left(\frac{q_1}{\lambda_1}\right)dw_1, \\
dl &= K_2(\theta_2 - l_t)dt + \left(\frac{q_2}{\lambda_2}\right)dw_2, \\
dc &= K_3(\theta_3 - c_t)dt + \left(\frac{q_3}{\lambda_3}\right)dw_3, \\
q_1 &= \lambda_1(\alpha_1 + \beta_1 r_t + \gamma_1 l_t + \delta_1 c_t)^{\frac{1}{2}}, \\
q_2 &= \lambda_2(\alpha_2 + \beta_2 r_t + \gamma_2 l_t + \delta_2 c_t)^{\frac{1}{2}}, \\
q_3 &= \lambda_3(\alpha_3 + \beta_3 r_t + \gamma_3 l_t + \delta_3 c_t)^{\frac{1}{2}}, \\
E\left(dw_i dw_j\right) &= \rho_{ij} dt,
\end{aligned}
\tag{7}
$$

where $\{K_i, \theta_i, \lambda_i, \alpha_i, \beta_i, \delta_i, \gamma_i; i = 1, ..., 3\}$ are constant parameters, and the Wiener terms, dw_i, may be correlated. In this model, r_t is the level of the instantaneous interest rate, l_t is a "long-term" interest rate and c_t is the convexity of the term structure. The drifts, the squared diffusion terms, and the squared prices of risk are affine functions of the state variables r_t, l_t and c_t.[6] The motivation for the third factor is provided by studies such as Litterman and Scheinkman (1991).

The simpler models are obtained by setting various terms in Equation (7) to zero. In the first model, the short-term interest rate r_t is the only state variable. There is only the single Wiener term dw_1, and δ_1, γ_1, and all the coefficients with subscripts 2 or 3 are set to zero. This case includes the models of Vasicek (1977), where $\beta_1 = 0$, and of Cox, Ingersoll, and Ross (1985a), where $\alpha_1 = 0$.

We include two versions of two-factor term structure models. The first is an affine model, obtained by setting δ_1, δ_2, and all the terms with subscript 3 equal to zero in Equation (7). The second is the Brennan and Schwartz (1979) two-factor model:

$$
\begin{aligned}
dr &= r_t\left[\alpha \ln\left(\frac{l_t}{\kappa r_t}\right)\right]dt + r_t\sigma_1 dw_1, \\
dl &= \left[l_t^2 - r_t l_t + l_t\sigma_2^2 + q_2 l_t \sigma_2\right]dt + l_t\sigma_2 dw_2, \\
& q_1,\ q_2 \text{ are constants}, \\
& E(dw_1 dw_2) = \rho dt,
\end{aligned}
\tag{8}
$$

and the fixed parameters are $\{\alpha, \kappa, \sigma_1, \sigma_2, q_1, q_2, \rho\}$.

[6] Dybvig, Ingersoll, and Ross (1996) show that infinite-maturity forward rates and zero coupon yields can never fall to avoid arbitrage in frictionless term structure models. The long rates in the models examined here can move up or down stochastically, so they should be interpreted as rates for a finite-maturity index. Empirically, we use a ten-year fixed-maturity yield as the long rate.

1.2 Empirical SDF models

The empirical SDF models are derived by substituting the market price of risk specifications of (7)–(8) into Equation (6). The change in the Wiener processes are found using Equation (5) and substituted into the B_{t+1} terms of (6). There is a term like B_{t+1} and C_{t+1} for each state variable, $j = 1,\dots,K$ ($K = 1$, 2, or 3). The empirical SDF is given as $\exp\left(-A_{t+1} - \Sigma_j B_{j,t+1} - \Sigma_j C_{j,t+1}\right)$. The results may be written in reduced form, as follows:

Affine

$$
\begin{aligned}
{}_t m(\phi)_{t+1} = \exp\big[&\mathrm{a} + \mathrm{b}\, A^r_{t+1} + c(r_{t+1} - r_t) + dA^l_{t+1} + e(l_{t+1} - l_t) \\
&+ fA^c_{t+1} + g(c_{t+1} - c_t)\big],
\end{aligned}
\tag{9a}
$$

Brennan and Schwartz:

$$
{}_t m(\phi)_{t+1} = \exp\left(a + bA^r_{t+1} + cA^l_{t+1} + dD^r_{t+1} + eD^l_{t+1} + f\, D^{rl}_{t+1}\right), \tag{9b}
$$

where

$$
A^r_{t+1} = \sum_{i=1,\cdots\frac{1}{\Delta}} r[t+(i-1)\Delta]\Delta,
$$

$$
A^l_{t+1} = \sum_{i=1,\cdots\frac{1}{\Delta}} l[t+(i-1)\Delta]\Delta,
$$

$$
A^c_{t+1} = \sum_{i=1,\cdots\frac{1}{\Delta}} c[t+(i-1)\Delta]\Delta
$$

$$
D^r_{t+1} = \sum_{i=1,\cdots\frac{1}{\Delta}} \left\{\frac{r(t+i\Delta)}{r[t+(i-1)\Delta]} - 1\right\}
$$

$$
D^l_{t+1} = \sum_{i=1,\cdots\frac{1}{\Delta}} \left\{\frac{l(t+i\Delta)}{l[t+(i-1)\Delta]} - 1\right\}
$$

$$
D^{rl}_{t+1} = \sum_{i=1,\cdots\frac{1}{\Delta}} \ln\left\{\frac{r[t+(i-1)\Delta]}{l[t+(i-1)\Delta]}\right\}\Delta
$$

The coefficients {a,b,c,...} are constant functions of the underlying fixed parameters of the models.[7] The two-factor affine model is nested in the three-factor model by setting $f = g = 0$. The single-factor affine model is nested in the two-factor affine model by further setting $d = e = 0$.

Note that the single-factor model actually depends on two "factors." Because of the effects of time aggregation, there is both a discrete change in the short rate, $(r_{t+1} - r_t)$, and an average of the daily short-rate levels over the month, A^r_{t+1}. Similarly, the two-factor affine model depends on changes in the long and short rates and on their averages. The three-factor model includes a discrete change in convexity and an average convexity factor. The Brennan and Schwartz model uses no discrete changes, only the daily averages including D^r_{t+1}, D^ℓ_{t+1}, and $D^{r\ell}_{t+1}$; a total of five empirical "factors." (We still refer to the models according to the number of instantaneous factors.)

The simple affine structure is analytically convenient. Because the risk premiums and instantaneous volatility are proportional to the square root of affine functions of the state, the two cancel in the B_{t+1} terms, and the C_{t+1} terms are linear. This allows the time-aggregated terms to be computed directly from the data. There is a tradeoff between the simplicity and tractability of the affine structure versus the likely misspecification of the simple models. While our general approach can be used in more complex models, the time-aggregated terms may be functions of the model parameters in such cases.[8] The estimation and time aggregation steps must then be combined. However, the simplification exploited here

[7] In the three-factor affine model the coefficients are

$$a = (K_1\theta_1\lambda_1 + K_2\theta_2\lambda_2 + K_3\theta_3\lambda_3) - \frac{1}{2}\left(\alpha_1\lambda_1^2 + \alpha_2\lambda_2^2 + \alpha_3\lambda_3^2\right),$$
$$b = -\left\{1 + K_1\lambda_1 + \left(\frac{1}{2}\right)\left[\beta_1\lambda_1^2 + \beta_2\lambda_2^2 + \beta_3\lambda_3^2\right]\right\}, \quad c = -\lambda_1,$$
$$d = -K_2\lambda_2 - \frac{1}{2}\left(\gamma_1\lambda_1^2 + \gamma_2\lambda_2^2 + \gamma_3\lambda_3^2\right), \quad e = -\lambda_2,$$
$$f = -K_3\lambda_3 - \frac{1}{2}\left(\lambda_1\delta_1 + \lambda_2\delta_2 + \lambda_3^2\delta_3\right), \quad \text{and} \quad g = -\lambda_3.$$

The two-factor affine model sets all the parameters with a subscript "3" as well as δ_1 and δ_2 to zero. The one-factor affine model sets all these parameters, γ_1 and the parameters with subscripts "2" or "3" to zero. In the two-factor Brennan and Schwartz model the coefficients are

$$a = -\alpha\left(\frac{q_1}{\sigma_1}\right)\ln(\kappa) + q_2\sigma_2 + \frac{1}{2}\left(q_2^2 - q_1^2\right),$$
$$b = -\left(1 + \frac{q_2}{\sigma_2}\right),$$
$$c = \frac{q_2}{\sigma_2},$$
$$d = -\left(\frac{q_1}{\sigma_1}\right),$$
$$e = -\frac{q_2}{\sigma_2}, \quad \text{and} \quad f = -\alpha\left(\frac{q_1}{\sigma_1}\right)$$

[8] This would occur in the essentially affine models studied by Duffee (2002) and Dai and Singleton (2002).

can be obtained in some models that are not affine as the Brennan and Schwartz (1979) example illustrates.

The number of parameters identified in the reduced form models is smaller than the number of underlying parameters in the term structure models. It would be possible to incorporate additional restrictions derived from the interest rate processes to identify additional parameters.[9] However, if we use more of its structure and the interest rate process is misspecified, the misspecification will spill over the performance measures. By not using all of the structure of the interest rate processes, the reduced-form models should be robust to some misspecification in those processes.

2. Addressing Interim Trading Bias

Our approach addresses the interim trading problem discussed by Goetzmann, Ingersoll, and Ivkovic (2000) and Ferson and Khang (2002), and the related problem of derivative securities. To see why, consider a model with state vector X and assume that an SDF, ${}_t m_{t+\Delta}$, prices asset returns for the period t to $t+\Delta$. Equation (1) implies this SDF must also price all portfolios of the primitive assets with weights that depend on the information generated by the state variables at time t. Similarly, at time $t+\Delta$, there is an SDF ${}_{t+\Delta}m_{t+2\Delta}$ that prices all portfolios formed at time $t+\Delta$ and held until $t+2\Delta$, with weights that may be any function of $X_{t+\Delta}$. From the law of iterated expectations, it follows that the SDF ${}_t m_{t+2\Delta} = ({}_t m_{t+\Delta})({}_{t+\Delta}m_{t+2\Delta})$ prices all returns measured over the t to $t+2\Delta$ period, including strategies that trade at t and $t+\Delta$ based on the information at these two dates.[10] In the term structure models, ${}_t m_{t+2\Delta}$ is the product of exponentials, which implies summing the exponents, thus generating the time-averaged terms.[11] As Δ approaches zero in the

[9] For example, in the Cox–Ingersoll–Ross model, the first and second moments of the discrete changes, $r_{t+1} - r_t$, conditional on the current value of the state variable r_t, may be expressed as a function of r_t and the parameters of the square root interest rate process. See Farnsworth (1997) for an illustration. Sun (1997) shows how a test function approach can be used to estimate all of the parameters of CIR-type models by equating additional empirical moments of interest rates to those implied by the models.

[10] Let $w(X_t)$ be a portfolio weight vector that sums to 1.0 and consider the two-period strategy with gross return ${}_t R_{pt+2\Delta}=[w(X_t)'\,{}_t\mathrm{R}_{t+\Delta}][w(X_{t+\Delta})'\,{}_{t+\Delta}R_{t+2\Delta}]$. At time $t+\Delta$, the Euler equation implies $1 = E_{t+\Delta}\{{}_{t+\Delta}m_{t+2\Delta}[w(X_{t+\Delta})'\,{}_{t+\Delta}R_{t+2\Delta}]\}$. At time t we have

$$\begin{aligned} 1 &= E_t\left\{{}_t m_{t+\Delta}\left[w(X_t)'_t R_{t+\Delta}\right]1\right\} \\ &= E_t\left({}_t m_{t+\Delta}\left[w(X_t)'_t R_{t+\Delta}\right]E_{t+\Delta}\left\{{}_{t+\Delta}m_{t+2\Delta}\left[w(X_{t+\Delta})'_{t+\Delta}R_{t+2\Delta}\right]\right\}\right) \\ &= E_t\left\{({}_t m_{t+\Delta})({}_{t+\Delta}m_{t+2\Delta})\left({}_t R_{pt+2\Delta}\right)\right\}. \end{aligned}$$

[11] In theory, any model in which ${}_t m_{t+\Delta}$ is a ratio of observables at the two dates can resolve interim trading bias. Consider a consumption-based asset pricing model in which ${}_t m_{t+\Delta} = \beta^{\Delta}(C_{t+\Delta}/C_t)^{-\alpha}$, and C_t is the aggregate consumption at time t. In this model, the SDF ${}_t m_{t+2\Delta} = ({}_t m_{t+\Delta})({}_{t+\Delta}m_{t+2\Delta}) = \beta^{2\Delta}(C_{t+2\Delta}/C_t)^{-\alpha}$, and the intermediate consumption at $t+\Delta$ cancels out.

continuous-time limit, the SDF should correctly price all dynamic strategies that are non-anticipating functions of the state variables. In practice, we are limited by the data, and with daily data we implicitly assume that managers trade only at the end of each day. Funds actually engage in intradaily trading, of course, but our empirical work cannot control for trading within the day.[12]

The time-averaged terms in the SDF can also be motivated by arbitrage considerations when interim trading is allowed. To see this, take the special case of no risk and consider two trading strategies over the t to $t+1$ period. The first strategy buys and holds the one-period bond with price $E_t({}_tm_{t+1}) = {}_tm_{t+1}$ and gross return equal to $1/{}_tm_{t+1}$. The second strategy rolls over instantaneous-term bonds, earning a gross return of $\exp\left(\int_t^{t+1} r_s ds\right)$. To avoid arbitrage, the returns of the two strategies must be equal, implying ${}_tm_{t+1} = \exp\left(-\int_t^{t+1} r_s ds\right)$. The no-arbitrage SDF depends on the time-averaged short-term interest rate. In the special case of no risk, $q(X) = 0$, and we see that Equation (4) gives the same solution.

A common instinct in empirical finance is to focus on end-of-period data only. The preceding illustrates that such an approach relies strongly on the assumption that the trading interval in the model is of the same length as the period over which the data are measured. In contrast, many macroeconomic data series are reported as time averages. These are often considered to be biased or "noisy" series, to be adjusted or otherwise viewed with suspicion [e.g., Working (1960), Breeden, Gibbons, and Litzenberger (1989)]. When dynamic trading strategies are involved, time-averaged data can be appropriate and useful.

3. Empirical Methods

We estimate the conditional performance of a fund and the parameters of the SDF model simultaneously using the following system of moment conditions and the generalized method of moments [GMM, see Hansen (1982)].

$$E\left\{\left[{}_tm(\phi)_{t+1}R_{t+1} - \underline{1}\right] \otimes D_t\right\} = 0 \tag{10a}$$

$$E\left\{\left[{}_tm(\phi)_{t+1}R_{pt+1} - 1 - \alpha_p' D_t\right] \otimes D_t\right\} = 0. \tag{10b}$$

Equation (10a) says that the SDF prices the primitive asset returns, R_{t+1}, while Equation (10b) identifies the fund's time-varying SDF alpha, $\alpha_{pt} = \alpha_p' D_t$. D_t is the *Conditioning Dummy Variable*, a vector of predetermined (0,1) instruments for the states of the term structure, described in more

[12] It remains an open question for future research as to the effects of intradaily trading on performance measurement, a topic that awaits the availability of intradaily fund trading data.

detail below. Christopherson, Ferson, and Glassman (1998) also allow time-varying alphas, assuming the conditional alphas are linear functions of lagged instruments. Here, the performance measure is "nonparametric," avoiding the linearity assumption. Using conditioning dummy variables and a small number of states, we obtain simplicity and interpretability. The cost is a coarse representation of the conditioning information. Of course, one can define more dummies to refine the information, relative to the examples we use here.

Farnsworth, et al. (2002) show that estimating a system like (10a, 10b) for one fund at a time produces the same point estimates and standard errors for alpha as a system that includes an arbitrary number of funds. This is convenient, as the number of available funds exceeds the number of monthly time series, and joint estimation with all of the funds is therefore not feasible.[13]

Farnsworth, et al. (2002) find small biases in SDF alphas for equity funds. To the extent that there are biases and they are similar for the fund and a benchmark return $R_{B,t+1}$, the biases may be controlled by measuring the abnormal performance of a fund relative to the benchmark. This is accomplished by replacing Equation (10b) with the expression, $E\{[{}_tm(\phi)_{t+1}(R_{pt+1} - R_{Bt+1}) - \alpha_p' D_t] \otimes D_t\} = 0$. Measuring performance relative to a benchmark may also increase the precision of the alpha, because the variance of the excess return is smaller than the raw return. Of course, if the model correctly prices the benchmark return, the point estimate of the fund's alpha is not changed by the introduction of the benchmark. We present both versions of the conditional alphas.

4. The Data

First, we describe our sample of US government bond mutual funds. We then describe the conditioning dummy variables for the states of the term structure. The Appendix describes the daily interest rate data that we use to construct the empirical factors implied by the models, the monthly bond return data, the dynamic interim-trading strategies that we use for model diagnostics, and the mutual fund performance benchmarks.

4.1 Government bond mutual fund data

The government bond fund data are selected by their objective codes, from the Center for Research in Security Prices (CRSP) mutual fund data base.[14] The number of funds with some monthly return data in a given year is less than 40 before 1986. Our fund sample starts in January of 1986

[13] Farnsworth et al. (2002) provide the invariance result for the special case where there is only a constant in D_t, so the alpha is a constant. An Appendix to this article, available by request, refines and extends the result for a general time-varying alpha.

[14] The Government bond funds include the ICDI_OBJ code GS, OBJ codes GOV or LTG, POLICY code of GS or SI_OBJ codes of GGN, GIM, GSM, GMB, or GS.

where the number of funds, based on the objective codes for year end 1985, is 67. The characteristics we use for sorting funds become available in 1988. The number of funds rises to 878 in June of 2001. Starting in the year 2000, many funds report quarterly data on fund characteristics. For these cases, we use only the data for the last quarter of the year (which includes all of the monthly returns). There are a total of 6552 fund-year records.

We subject the sample of funds to many screens. To address back-fill bias, we delete years before and including the year of fund organization. Data may be reported before the year of fund organization, for example, if a fund is incubated before it is made publicly available [see Elton, Gruber, and Blake (2001) and Evans (2004)]. This removes 539 records. We also delete all cases where the total net assets (TNA) of the fund is reported to be less than five million dollars. This removes another 809 records.[15] We delete all cases where the reported equity holdings at the end of the previous year exceeds 10% (555 records) and all cases where the reported holdings of bonds plus cash is less than 85% (1273 records). After these screens, we are left with 3376 fund years. The screens no doubt reduce the cross-sectional variation of performance in our sample but provide more consistency with our focus on default-free term structure factors.

We examine an equally weighted portfolio of all the funds, from 1986 through 2000. We also group the funds into portfolios according to fund characteristics measured at the end of the previous year, starting in 1988 (returns for 1989). The characteristics include fund age, TNA, percentage cash holdings, income yield, annual turnover, total load charges, annual expense ratios, and the annual return over the previous year. Each year we sort the funds with non-missing characteristic data from high to low and break them into thirds. We form equally weighted portfolio returns from the funds in the high group and the low group for each month of the next year. The returns are based on the end-of-month net asset values of the funds, plus any distributions.

We examine the time series of the cutoff values for the fund characteristics that define the upper and lower thirds of the distributions (graphs are available by request), and we find that government bond funds have experienced some trends different from equity funds. The age distribution of the funds increased, the TNA per fund shows a downward trend, while the turnover and expense ratio distributions show no trend. The upper-third cutoff for expense ratios is about one percent per year and the lower-third cutoff is about 0.8%. In contrast, a sample of equity funds over a similar period has grown younger—reflecting the large number of new funds starting in the mid-1980s—while the turnover and expense ratios have increased [e.g., Ferson and Qian, (2004)].

[15] Without these last two screens, the youngest third of the funds have higher average returns, lower volatility, and significant positive alphas after risk adjustment.

Summary statistics for the government bond fund returns, grouped according to characteristics, are reported in Table 1. Not surprisingly, the returns look very different from equity mutual fund returns. On average, the funds earned 0.60% per month, about 8 bp less than the Lehman Government bond index. The means are all between 0.5 and 0.7% per month and the first-order sample autocorrelations are all about 0.2. The standard deviations are between 0.98 and 1.3% per month. The mean returns differ across characteristics by 17 bp or less. The standard errors of the mean are $\simeq$ 8 bp, so none of the differences is strongly significant.

Table 1
Summary statistics for the fixed-income funds, lagged instruments and factors

	Mean	Minimum	Maximum	Standard deviation	ρ_1	Skew	Kurtosis	Normality
Panel A: Equally weighted fund portfolios grouped by characteristics, January, 1989–December, 2000 (144 observations). All funds January, 1986–December, 2000 (180 observations). Percent per month								
Fund group								
All funds	0.60	–3.42	4.32	1.20	0.19	–0.11	0.46	0.31
Old	0.60	–2.33	3.55	1.09	0.24	–0.26	–0.13	0.35
Young	0.57	–2.35	3.22	1.07	0.23	–0.25	0.06	0.43
Big	0.60	–2.08	3.45	1.09	0.22	–0.18	–0.20	0.28
Small	0.58	–2.00	3.14	1.06	0.23	–0.18	–0.19	0.34
High cash	0.57	–2.07	3.11	1.05	0.22	–0.19	0.05	0.48
Low cash	0.62	–2.38	3.73	1.12	0.23	–0.18	–0.07	0.27
High income	0.62	–2.69	3.80	1.14	0.25	–0.25	–0.03	0.35
Low income	0.54	–2.07	3.30	1.00	0.21	–0.08	0.24	0.42
High turnover	0.54	–2.07	3.64	0.99	0.26	0.10	0.12	0.44
Low turnover	0.56	–2.15	3.58	1.04	0.29	0.05	0.13	0.49
High load	0.58	–2.32	3.28	1.12	0.22	–0.24	–0.21	0.38
Low load	0.59	–2.14	3.21	1.04	0.25	–0.23	–0.13	0.26
High expense	0.54	–2.28	3.29	1.09	0.23	–0.22	–0.19	0.33
Low expense	0.62	–2.10	3.33	1.05	0.24	–0.18	–0.15	0.29
High lagret	0.51	–3.19	3.24	1.23	0.24	–0.63	0.61	1.20***
Low lagret	0.68	–1.85	4.84	1.09	0.20	0.61	1.15	0.62
Panel B: Bond returns, January, 1986–December, 2000 (180 observations)								
Bond								
r4mo	0.47	0.15	0.95	0.14	0.76	0.54	0.75	0.99**
r1y	0.55	–0.19	1.57	0.31	0.33	0.36	0.41	0.83**
r3y	0.63	–1.93	2.94	0.96	0.21	–0.01	–0.29	0.20
r4y	0.68	–2.94	3.77	1.30	0.17	–0.13	–0.22	0.19
r7y	0.72	–3.30	4.72	1.61	0.16	–0.01	–0.18	0.12
r10y	0.72	–4.40	6.48	2.03	0.12	0.15	0.15	0.29
r20y	0.86	–7.05	11.6	2.73	0.11	0.35	1.30	0.64
Le24	0.57	–0.65	2.01	0.52	0.24	0.07	–0.16	0.20
Le60	0.65	–2.44	3.84	1.26	0.17	–0.07	–0.32	0.16
Le120	0.71	–3.07	5.10	1.56	0.15	0.05	–0.04	0.15
gt120	0.83	–5.31	10.44	2.48	0.10	0.31	0.99	0.49
Lbgov	0.68	–2.46	4.03	1.33	0.15	0.04	–0.07	0.19
Sharpe	0.68	–2.40	3.83	1.26	0.16	0.05	–0.10	0.17
Implied monthly from daily data								
Roll3	0.45	0.15	0.88	0.14	0.78	0.38	0.05	1.01**
Randr	0.60	–2.80	4.38	1.29	0.13	0.11	0.51	0.82**
BuyhiY	0.67	–3.61	7.54	1.95	0.07	0.48	1.55	2.20***
BuyhiR	0.66	–3.06	6.42	1.52	0.14	0.82	1.40	3.37***
MV	0.46	0.09	1.19	0.18	0.55	0.75	1.25	0.67*

Table 1
(continued)

Factor	Mean	Minimum	Maximum	Standard deviation	ρ_1
Panel C: Term structure factors, January, 1986–December, 2000 (180 observations)					
$100^*(r_{t+1}-r_t)$	–0.06	–11.61	4.58	2.09	0.17
$100^*(\ell_{t+1}-\ell_t)$	–0.19	–7.86	5.79	2.51	0.12
$100^*(c_{t+1}-c_t)$	0.02	–2.38	1.61	0.65	0.02
A^r	0.46	0.24	0.76	0.12	0.99
A^ℓ	0.59	0.38	0.79	0.10	0.96
A^c	–0.02	–0.06	0.02	0.01	0.90
100^*D^r	–0.52	–10.09	64.71	23.52	0.15
$100^*D^{\ell 1}$	–1.23	–63.10	60.45	21.07	0.07
$D^{r\ell}$	–26.32	–81.11	12.37	21.58	0.98

Variable	Mean	Minimum	Maximum	Standard deviation	ρ_1	ρ_1Dhi	ρ_1Dlo
Panel D: Lagged instruments to predict returns, 1986–2000 (180 observations)							
R_t	0.46	0.23	0.77	0.12	0.98	0.77	0.91
ℓ_t-r_t	0.13	–0.06	0.32	0.09	0.96	0.79	0.78
C_t	–0.02	–0.06	0.02	0.01	0.88	0.72	0.49

Panel E: Correlation matrix of the conditioning dummy variables, 1986–2000 (180 observations)						
High r_t	1.00					
High ℓ_t-r_t	–0.11	1.00				
High c_t	0.12	–0.25	1.00			
Low r_t	–0.21	0.51	–0.29	1.00		
Low ℓ_t-r_t	0.23	–0.27	0.71	–0.50	1.00	
Low c_t	–0.01	0.56	–0.20	0.28	–0.22	1.00

Δr	$\Delta\ell$	Δc	A^r	A^ℓ	A^c	D^r	D^ℓ	$D^{r\ell}$
Panel F: Correlation matrix of the factors, 1986–2000 (180 observations)								
1.00								
0.49	1.00							
0.01	–0.60	1.00						
–0.02	–0.07	0.05	1.00					
0.03	0.01	0.02	0.68	1.00				
–0.13	–0.15	0.05	0.39	–0.326	1.00			
0.93	0.52	–0.09	–0.01	–0.002	–0.08	1.00		
0.45	0.96	–0.62	–0.06	0.008	–0.12	0.52	1.00	
–0.04	–0.09	0.05	0.72	–0.008	0.79	–0.02	–0.08	1.00

The symbol ρ_1 is the first-order sample autocorrelation. Dhi is the high state conditioning dummy variable and Dlo is the low-state variable. Skew refers to the maximum likelihood estimate of $E\{[r-E(r)]^3\}/\sigma(r)^3$. Kurtosis refers to the maximum likelihood estimate of $E\{[r-E(r)]^4\}/\sigma(r)^4 - 3$. Normality is the modified Anderson-Darling test, evaluated at the finite sample critical values in Stephens (1974, Table 1.3). Significance at the 10, 5, and 1% levels are denoted by ***, **, and *, respectively. The bond returns are discrete monthly percent. r*n*y denotes the *n*-year discount bond return, le*n* denotes less than or equal *n* months to maturity and gt*n* denotes at least *n* months to maturity. Lbgov denotes the Lehmann government securities index. Five strategies are based on daily trading. Randr chooses one maturity at random each day, buyhiY holds the maturity with the highest yield, buyhiR holds the maturity with the highest expected return, based on a regression of return on the level, slope, and convexity over the previous 60 days. Roll3 rolls over each day a notional 91-day Treasury bill. Mv is a mean variance efficient strategy. The instruments include r_t, the yield to maturity on a three-month Treasury bill, ℓ_t, a ten-year constant-maturity Treasury yield, and c_t, the convexity. The factors are expressed as continuously compounded monthly percent. The time averaged factors, denoted as A^x, are the monthly averages of daily factors. $D^{r\ell}$, D^r, and D^ℓ are the monthly averages of the Brennan and Schwartz factors.

The minimum return month since January of 1989 is March of 1994, where the high lagged-return funds lost 3.2%. October of 1987 was a relatively high return month, where a portfolio of all the funds earned 2.9%.

The three right-hand columns of Panel A present statistics to detect departures from normality. While 14 of the 18 cases show left skewness, there is little significant evidence of non-normality. Of course, portfolios are likely to be closer to normal than individual fund returns. The fund group with high lagged returns over the past year displays left skewness and peakedness relative to the normal, and the Anderson-Darling test for normality is significant at the 1% level, based on the finite sample critical values in Stephens (1974, Table 1.3). The low-lagged-return fund group has the opposite pattern: right skewness and fat tails. The means of this group are also higher, suggesting mean reversion in the fund returns.

4.2 Term structure factors

In the Cox–Ingersoll–Ross and Vasicek models, the short-term interest rate is the relevant state variable. In the two-factor models, we use the short rate and long-term interest rate, measured as the ten-year Treasury yield. In the three-factor model, we add convexity as the third state variable. These data are described in the Appendix, and summary statistics are presented in Panel D of Table 1.

The average monthly change in both short and long rates is slightly negative over this period. The fact that we observe the funds over a falling-rate period implies a caveat for our performance analysis. The results may not be valid in a rising interest rate environment. We would argue that using factors motivated by theoretical models is likely to be especially important in such a setting as theoretical factors should have better external validity than empirical factors.[16] We use the longer sample period, August 1974 through December 2000, for SDF model diagnostics. With these data, the diagnostics include both rising and falling interest rate environments.

4.3 Measuring term structure states

The conditioning dummy variables are a simple non-parametric approach to conditioning on the states of the term structure. Consider the monthly short rate series, r_t. We first convert the short rate into a deviation from its average level over the last 60 months: $x_t = r_t - (1/60)\Sigma_{j=1,\dots 60} r_{t-j}$. We then use the last 60 months to estimate a rolling standard deviation, $\sigma\ (r_t)$. The dummy variable $D_{t,\text{hi}}$ for a "higher than normal" level of the spot rate is defined as the indicator function: $I\{[x_t/\sigma\ (r_t)] > 1\}$. The dummy variable $D_{t,lo}$ for a "lower than normal" level of the spot rate is $I\{[x_t/\sigma\ (r_t)] < -1\}$. The lagged instrument in Equation (10) is

[16] Critiques by Lo and MacKinlay (1990), MacKinlay (1995), and Ferson, Sarkissian, and Simin (1999) illustrate the pitfalls of asset pricing factors motivated by empirical regularities.

defined as: $D_t = (1, D_{t,\text{lo}}, D_{t,\text{hi}})$. Dummy variables for the other state variables are similarly constructed.

Figure 1 presents plots of the three state variables and their associated conditioning dummies over time. The figure shows that the short rate is either normal or low over most of the sample period after 1985, with brief exceptions in 1989 and 2000. For the other state variables, the high and low states are not concentrated in discrete subperiods, but occur in episodes throughout the sample. Panel E of Table 1 shows the correlations among the conditioning dummy variables. The average absolute correlation is less than 0.30. The highest correlations are between the convexity and slope dummies (0.56 and 0.71).

A significant feature of the conditioning variables is the high persistence of the short rate and slope as indicated by the first-order sample autocorrelations of 98 and 96%. The averaged state variables are also highly persistent. This raises concerns about small-sample biases [e.g., Stambaugh (1999), Bekaert, Hodrick, and Marshall (1997)] and spurious regression problems [e.g., Ferson, Sarkissian, and Simin (2003)]. One advantage of the conditioning dummy variables is that their autocorrelations are smaller than those of the underlying instruments. The maximum first-order autocorrelation of a dummy variable, shown in Table 1, is 91%, and the rest are below 80%.[17]

We examine the mean returns and standard errors for the bond returns described in Table 1, conditional on the dummy variables. We find strong evidence that variation in both the conditional means and variances can be tracked by the dummy variables (tables are available by request). These results motivate our use of the dummy variables in a conditional analysis.

5. Evaluating the SDF Models

Blake, Elton, and Gruber (1993) and Elton, Gruber, and Blake (1995) use linear beta pricing models in their studies of fixed-income funds. Linear factor models are commonly used in practice as well. Such models provide interesting contrasts with the approach of this article. There are three essential differences. First, with linear models the factors must be excess returns to factor-mimicking portfolios to get the right alpha. One advantage of the SDF approach is that mimicking portfolios are not required. The second difference is that linear beta models correspond to

[17] Using the conditioning dummy variables lowers the autocorrelation of the lagged instruments in the GMM system, but it does not address the persistence of the variables in the empirical SDF. This raises the issue of finite sample performance including the accuracy of the asymptotic properties of the GMM estimators. Ferson and Foerster (1994) study finite sample properties of GMM estimators in models with persistent implied SDFs. They find that Hansen's J-statistic rejects too frequently, and the coefficient estimators can be unreliable in a two-stage GMM approach while an iterated GMM approach is reasonably accurate. We use iterated GMM in this article.

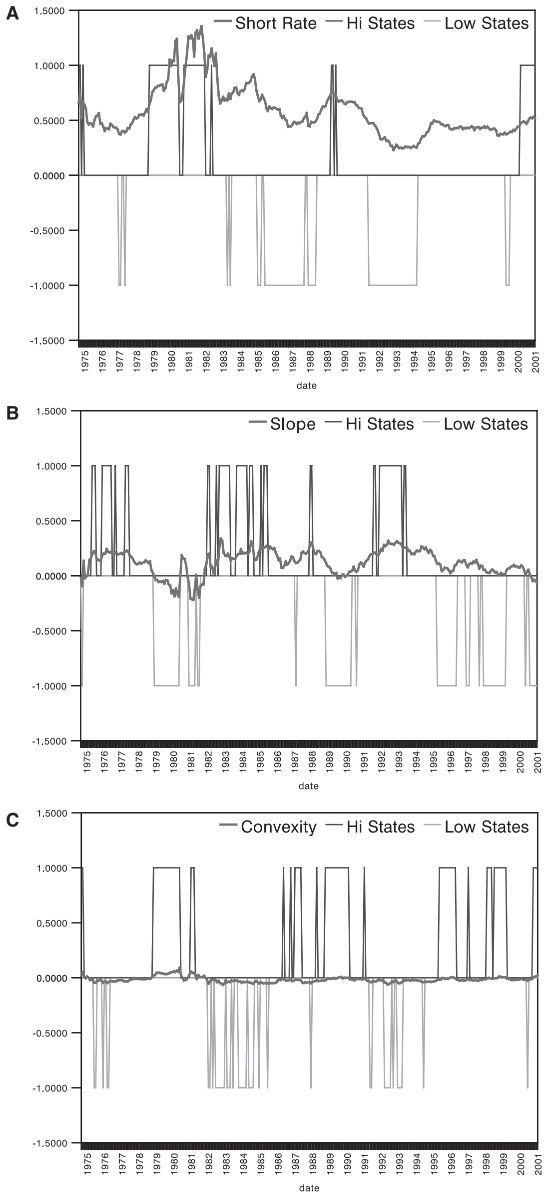

Figure 1
Dummy variables for the short term interest rate, slope, and convexity of the term structure are plotted over time in figures A, B and C respectively. The dummy variable is shown as +1 when the state variable is more than one trailing 60-month standard deviation above its trailing 60-month mean. The value is shown as –1 when the variable is more than one trailing 60-month standard deviation below the mean. The state variables are defined in Table 1. (**A**) Short rate and associated high-, low-state dummies, (**B**) slope and associated high-, low-state dummies, and (**C**) convexity and associated high-, low-state dummies.

SDFs that are linear in the factors, whereas the term structure models SDFs are exponentials of linear (more properly, affine) functions. A linear SDF can take negative values, and a negative SDF can assign negative prices to positive payoffs. The exponential function has the advantage that the SDF is restricted to positive values. The third feature of the term structure SDFs is the additional variables that appear because of time aggregation.

5.1 Linear factor model regressions

We now examine the empirical explanatory power of the new time-aggregated factors introduced by our approach. Suppose that the SDF were linear in the factors so that a beta pricing model applied. Under this approximation, the slope coefficients in a linear regression of bond returns on the factors are the unconditional betas that describe the cross-section of average returns. Factor model regressions are used in practice for fund risk measurement, hedging, and performance attribution. Here, the slope coefficients are elasticities or empirical duration measures [see Kon and Polek (1998) for examples].

Table 2 summarizes the factor model regressions over the 1974–2000 period; the Appendix discusses additional regression experiments.[18] The time-averaged term A^r is significant in the regression for the short-term bond return, le24, with a t-ratio larger than three. For this return, the A^ℓ term also has a t-ratio larger than two in the two-factor regressions (not shown in the table), and the three time-averaged terms are jointly significant in the three-factor regressions. Using the long-term bond return, gt120, the A^r term is not significant, and the one-factor regression R-squares are less than 30%. Bringing in the long-rate factors raises the adjusted R-squares to over 90%, and the time-averaged terms are jointly significant. In the three-factor models, the R-squares approach 95%, and the time-averaged terms are statistically significant.

When the dynamic strategies denoted buyhiY or buyhiR (described in the Appendix) are the dependent variables, the time-averaged term A^r is marginally significant in the one-factor regressions—the p-values for the step-down F-tests for the exclusion of A^r are 5.9 and 3.6%. The three time-averaged terms, however, are jointly significant in the three-factor

[18] We apply a version of Stambaugh's (1999) adjustment to the regression coefficients to address the high autocorrelation in some of the variables. Stambaugh (1999) considers a regression system:

$$r_{t+1} = a + b\, Z_t + u_{t+1}$$
$$Z_{t+1} = \delta + \rho Z_t + v_{t+1},$$

with $E(u_{t+1}v_{t+1}) = \sigma_{uv}$ and $E(v_t^2) = \sigma_v^2$. He shows that the OLS slope estimator $\hat{b}$ has bias $E(\hat{b} - \mathrm{b}) = E(\hat{\rho} - \rho)\sigma_{uv}/\sigma_v^2$, where $E(\hat{\rho} - \rho) \approx -(1+3\rho)/T$. Our adjusted estimator is $\mathrm{b}^* = \hat{b} + (1+3\rho^*)\sigma_{uv}/(\mathrm{T}\sigma_v^2)$, where $\rho^* = (T\hat{\rho}+1)/(T-3)$. We compute the regression R-squares using the adjusted slopes as the ratio of the variance of the fitted values to the variance of the dependent variable. The effect of the adjustment is typically on the order of 1% of the coefficient, never exceeding 10%, and it slightly increases the explanatory power of the factors as measured by their t-ratios. [Campbell, Chan, and Viciera (2003) find a similar effect.]

Table 2
Linear factor model regressions

	Δr	Δl	Δc	A^r	A^l	A^c	Rsq	p-value (%)
Le24	−13.3						61.4	
	1.23							
	−13.3			0.67			64.5	0.00
	1.15			0.18				
	−7.55	−16.1	2.53				84.5	
	1.17	1.39	4.66					
	−7.44	−16.5	3.29	0.95	0.01	−2.20	89.9	0.00
	1.12	1.29	4.27	0.36	0.38	1.82		
gt120	−29.6						29.5	
	3.50							
	−29.5		−0.28				29.4	59.6
	3.51		0.68					
	5.65	−89.7	−12.9				94.5	
	1.75	2.29	5.90					
	5.77	−89.9	−12.4	0.33	0.45	−1.12	94.7	0.30
	1.76	2.33	6.11	0.74	0.83	3.55		
Buyhiy	−16.4						18.5	
	2.82							
	−16.4			0.81			19.2	5.90
	2.86			0.39				
	−0.245	−51.3	26.8				71.3	
	3.08	4.18	12.9					
	0.0143	−52.3	27.4	−0.74	2.39	8.45	74.1	0.00
	2.78	3.93	12.8	1.43	1.55	8.12		
BuyhiR	−15.2						21.3	
	3.19							
	−15.2			0.76			22.1	3.60
	3.03			0.48				
	0.027	−38.2	−7.43				52.8	
	2.85	3.87	13.2					
	0.376	−39.3	−7.32	−2.85	4.46	15.3	56.3	0.00
	2.70	3.61	12.5	1.51	1.61	7.34		
Roll3	−3.09						31.5	
	0.440							
	−3.12			1.01			99.9	0.00
	0.0437			0.01				
	−3.22	0.613	−0.845				31.6	
	0.664	0.836	2.72					
	−3.12	−0.002	−0.062	0.98	0.03	0.153	99.9	0.00
	0.075	0.096	0.238	0.04	0.04	0.193		
All funds	−27.4						22.8	
	3.09							
	−27.3			0.82			23.0	20.6
	3.09			0.65				
	−3.14	−41.2	18.8				91.8	
	1.50	2.00	6.84					
	−3.18	−41.1	18.8	−0.43	1.47	3.44	92.7	0.02
	1.49	1.96	6.67	0.59	0.67	4.20		

Le24 is the bond portfolio with less than or equal to 24 months to maturity, and gt120 denotes at least 120 months to maturity. Three strategies are based on daily trading of the returns constructed from constant-maturity yields with maturities of three months and one, three, seven, and ten years. BuyhiY buys the bond with the highest yield to maturity. BuyhiR buys the bond with the highest expected return, based on a linear regression of returns on the level, slope, and convexity of the term structure over the last sixty days. Roll3 rolls over a notional 91-day Treasury bill. All Funds is an equally weighted portfolio of government bond funds. The coefficients are adjusted for finite sample bias, and heteroskedasticity-consistent standard errors are shown on the second line. The regressors are Δr, the monthly change in the three-month spot rate, $\Delta \ell$, the change in the ten-year Treasury yield, Δc, the discrete change in the monthly convexity, A^r, the daily average spot rate over the month, A^{ℓ}, the daily average ten-year yield and, A^c, the daily average convexity. Rsq is the adjusted R-squared, corrected for finite sample bias, in percent. P-val is the right-tail probability value of the step-down F statistic for the marginal explanatory power of the time-averaged terms, in percent. The sample period is August, 1974 through December, 2000 (317 observations), except for the All Funds regression, which is over the January, 1986 through December, 2000 period (180 observations).

regressions. Using the Roll3 strategy, the A^r factor adds more than 65% to the adjusted R-squared in the one-factor regression. The dependence of the SDF on the time-averaged short rate is especially important when short-term bonds and interim trading strategies using the short rate are involved.

Because the Roll3 strategy rolls over a notional three-month bill, the level of the three-month yield during the month explains much of its return. Thus, the Roll3 strategy results can only be considered illustrative. The last panel of Table 2 reports regression results for the return of the equally weighted portfolio of all the mutual funds. In these regressions, the A^r term is not marginally significant, but the A^ℓ term generates t-ratios between two and three in all of its regressions. Based on the F-statistics, the time-averaged terms are jointly significant in the two and three-factor models. Overall, the factor model regressions show that the time-averaged versions of the factors have significant explanatory power for the returns.

5.2 The economic significance of time-averaged factors

While the tests in Table 2 present evidence for the statistical significance of the time-averaged factors, it is also useful to evaluate the explanatory power in economic terms. Table 3 summarizes estimates of the average risk premiums, λ, associated with the betas on the factors in a linear model. These are estimated with the following system, which refines the approach used by Ferson and Harvey (1994):

$$\begin{aligned} f &= \mu_F + \varepsilon, \quad E(\varepsilon) = 0, \\ \sigma^2 &= (f - \mu_F)^2 + v, \quad E(v) = 0, \\ r &= \beta(f - \mu_F + \lambda.^*\sigma) + u, \quad E(u) = 0, \quad E(uf') = 0, \end{aligned} \tag{11}$$

where the f's are the factors and .* denotes element-by-element multiplication. The asset returns, r, are the 18-bond portfolios and dynamic strategies summarized in Panel B of Table 1, measured in excess of a one-month bill return. The risk premiums are percent per month per unit volatility, and heteroskedasticity-consistent standard errors are shown below.

Table 3 summarizes that the risk premiums attached to the betas on the time-averaged factors are economically large; the magnitudes exceed those of the premiums on the discrete factor changes in each example. However, the standard errors are also large, and the t-ratios are distributed around 1.0 for both the discrete changes and the time-averaged factors. The right-tail p-values of Hansen's (1982) goodness of fit statistics are the largest for the three-factor affine model, but none of the models can be rejected. Overall, the results suggest that the economic significance

Table 3
Economic significance of factors in a linear factor model

Model	Δr	Δl	Δc	A^r	A^l	A^c	D^r	D^l	D^{rl}
1FA	−23.9			45.9					
	97.6			129.0					
2FA	−14.4	−9.52		39.0	66.8				
	8.84	5.86		36.1	50.8				
3FA	−15.1	−9.58	−8.83	28.3	54.0	−27.0			
	9.57	6.12	17.4		55.6	48.6			
2BS				16.4	55.1		−14.7	−5.28	−51.6
				45.5	52.3		10.9	6.72	70.0

Estimates of the risk premiums per unit of volatility, λ, are shown based on the following system estimated using the Generalized Method of Moments:

$$f = \mu F + \varepsilon,\ E(\varepsilon) = 0,$$

$$\sigma^2 - (f - \mu_F)^2 = v, \quad E(v) = 0;$$

$$r = \beta(f - \mu F + \lambda .* \sigma) + u,\ E(u) = 0,\ E(uf') = 0$$

where ".*" denotes element-by-element multiplication. The risk premiums are percent per month and heteroscedasticity-consistent standard errors are shown on the second lines. The 18 asset returns are the bond returns and dynamic strategies summarized in Table 1, measured in excess of a one-month bill return. The factors, f, are Δr, the monthly change in the three-month spot rate, $\Delta \ell$, the change in the ten-year Treasury yield, Δc, the discrete change in the monthly convexity, A^r, the daily average spot rate over the month, A^ℓ, the daily average ten-year yield, A^c, the daily average convexity, and the daily averaged factors from the Brennan and Schwartz model: D^r, D^ℓ, and $D^{r\ell}$. The sample period is August, 1974 through December, 2000 (317 observations).

of the time-averaged factors is comparable to that of the discrete factor changes, but the precision of the linear factor model is low.[19]

5.3 Term structure models meet benchmark strategies

In this section, we explore the performance of the SDF models for pricing the benchmark bond-strategy returns. To this end, the benchmark returns replace the mutual fund returns in system (10) and become the test assets. We impose zero alphas for the primitive assets while allowing non-zero, time-varying alphas for the test assets, mirroring the way we treat the mutual funds below. Thus, we fit the models on the assumption that they correctly price the primitive assets, while allowing the test assets to have abnormal returns.

The primitive assets include the four-month Treasury bill and the 20-year bond return, which span the maturity spectrum, in the one and two-factor models. In the three-factor models, we include the three-year bond return.[20] Since we are interested in the ability of the SDF models to

[19] We experimented with exactly-identified versions of the model using a smaller number of asset returns. The impressions were similar.

[20] We experimented with the le24 and gt120 bonds as alternative primitive assets and found similar results. We estimate systems with the Lehman Government index and another benchmark return simultaneously, to facilitate the computation of standard errors for the alphas in excess of the Lehman benchmark and to mirror the way we treat the mutual funds below. The results are invariant to which return is paired with the Lehman index in the system.

handle interim trading and dynamic strategies, we include the dynamic strategies as test assets. We also include a range of bond maturities to evaluate the models over the maturity spectrum and the Lehman Government Bond index.[21] We summarize the results for seven test assets including the Lehman index, the gt120 bond portfolio, and the dynamic strategies whose construction is described in the Appendix: randr, buyhiY, buyhiR, roll3, and mv.

None of the models with time-averaged factors are rejected by Hansen's J-statistic, the smallest p-value being 27%. The estimated SDF models always have at least one factor coefficient with a t-ratio larger than two. The time-averaged short rate A^r gets a t-ratio of –13 in the one-factor affine model. The discrete change in the long rate is also an important factor, with t-ratios of 3.4 to 5.8. We can not reject the hypothesis that the convexity factors may be excluded from the three-factor affine model, reducing it to a two-factor model. However, deleting the time-averaged factors from the model, the discrete one-factor model is strongly rejected, with a p-value of 0.002%.[22]

Table 4 reports summary statistics of the diagnostics taken across cases, where each case is a particular test asset given a high- or low-state variable. Panel A covers the August 1974 through December 2000 sample period. The column labelled Avg($|\alpha|/|r|$) shows that the models explain large fractions of the conditional mean returns. The average absolute conditional mean return before risk adjustment is about 80 bp per month while the average absolute alpha is 8 bp or less. Thus, the SDF models can attribute about 90% of the expected returns to the risk factors. The average standard error for an alpha varies from 7.4 bp to 9 bp across the models. In the three-factor affine model, the average absolute alpha is 5.2 bp, and only five of the 42 cases present absolute alphas larger than 10 bp. The two-factor Brennan and Schwartz models perform similarly to the two-factor affine models over this period, and the one-factor model has the largest pricing errors.

The individual test asset results are not shown, but the Lehman index is illustrative. The raw returns of the Lehman index are 94 bp in the high spot rate state and 76 bp in the low spot rate state. The conditional alphas from the one-factor affine model are 13 and 8 bp, respectively. The two-factor affine model assigns alphas of 3.7 bp and 1.9 bp, respectively while the three-factor model assigns 2.3 and 0.8 bp, respectively, in the two states.

[21] We tried an equally weighted portfolio of the bond returns as an alternative, with similar results.

[22] We examine both the raw return alphas and alphas measured in excess of the Lehman Government index. While the excess alpha standard errors tend to be smaller, there is no systematic reduction in the magnitudes when moving to excess returns.

Table 4
Estimating stochastic discount factors (SDF) models on benchmark strategies

Model	Number of cases	Avg|r|	Avg|α|	Avg[$\sigma(\alpha)$]	%|α|>10 bp	Signed max|α|	%|$t(\alpha)$|>2	Avg(|α|/|r|)	HJD
Panel A: August, 1974 through December, 2000 (317 months)									
1FA	14	0.801	0.079	0.090	28.6	0.280	7.1	0.089	0.0330
1FD	14	0.801	0.072	0.089	35.7	0.144	7.1	0.090	0.0336
2FA	28	0.804	0.062	0.085	14.3	0.287	10.7	0.073	0.0270
2FD	28	0.804	0.064	0.082	14.3	0.293	10.7	0.076	0.0275
2BS	28	0.804	0.060	0.085	17.9	0.284	10.7	0.072	0.0241
3FA	42	0.793	0.052	0.074	11.9	0.225	14.3	0.068	0.0328
3FD	42	0.793	0.061	0.069	11.9	0.242	19.0	0.077	0.0324
Panel B: January, 1986 through December, 2000 (180 months)									
1FA	14	0.843	0.116	0.089	42.9	–0.552	21.4	0.132	0.0701
1FD	14	0.843	0.117	0.086	42.9	–0.576	28.6	0.134	0.0710
2FA	28	0.766	0.084	0.081	21.4	–0.576	3.6	0.115	0.0530
2FD	28	0.766	0.087	0.079	32.1	–0.689	25.0	0.117	0.0777
2BS	28	0.766	0.133	0.108	46.4	0.413	21.4	0.179	0.0490
3FA	42	0.714	0.073	0.068	21.4	–0.567	16.7	0.120	0.0460
3FD	42	0.714	0.078	0.069	21.4	–0.535	21.4	0.124	0.0863

Generalized method of moments (GMM) estimation of SDF models using the conditioning dummy for the relevant state variable(s) as the instruments. The primitive asset returns are a four-month Treasury bill and a 20-year Treasury bond return in the one- and two-factor models. In the three-factor models, a three-year bond return is the third primitive asset. There are seven benchmark test assets, whose returns are denoted as r and reported as percent per month. The seven test or benchmark assets include the Lehman Government Bond index return, a bond portfolio with at least ten years to maturity and five strategies based on daily trading of the returns constructed from constant-maturity yields with maturities of three months and one, three, seven and ten years to maturity. The dynamic strategies are randr, which chooses one maturity at random each day, buyhiY, which holds the maturity with the highest yield, buyhiR, which holds the maturity with the highest expected return, based on a regression of returns on the level, slope, and convexity over the previous 60 days. Roll3 rolls over each day a notional 91-day Treasury bill. Mv is a mean variance efficient strategy based on a linear regression function on past yields, with a target expected return set equal to the tangency point on the mean variance frontier drawn from the global minimum variance point minus 0.1% per month. The conditional alphas are denoted as α, also in percent per month. The averages are taken across the number of cases, which is the number of tests assets times the number of high or low-state dummies for which the conditional returns and alphas are estimated. The symbol σ (α) denotes the efficient GMM estimate of the standard error of the alpha. %|α|>10 bp denotes the percentage of the number of cases in which the absolute value of a conditional alpha exceeds 10 bp per month. The other columns have similar interpretations. Signed max |α| is the minimum or maximum value, depending on which is farther from zero. HJD is the Hansen-Jagannathan distance measure. NFA is the N-factor affine model, nFD is the n-factor discrete model, from which the time-averaged terms implied by the affine models are dropped. 2BS is the two-factor Brennan and Schwartz model.

Table 4 summarizes that the time-averaged factors help the models control expected returns by almost any measure, although the magnitudes of these improvements are modest, as the alphas are already small. In the three-factor discrete model (with no time-averaged factors), the average absolute alpha is 6.1 bp per month while it is 5.2 bp under the affine model with the time-averaged factors. In the discrete model, there are eight absolute t-ratios for alphas larger than two versus six under the affine model.

Panel B of Table 4 presents diagnostics for the subperiod 1986–2000 over which we evaluate the mutual funds. Most of the model comparisons are similar, although the alphas are generally larger in this period. The three-factor affine model produces average absolute alphas of 7.3 bp per month compared with the average absolute returns of over 71 bp. Thus, the best SDF models can still attribute almost 90% of the conditional expected returns to risk, and the three-factor affine model delivers the best overall performance. The time-averaged terms seem to improve the performance of the affine models to a greater degree in this sample. The Brennan and Schwartz model generates the largest average absolute alphas, and its performance relative to the two-factor affine model is not as good over this period.

Most of the statistics in Table 4 summarize the conditional alphas. The right-hand column presents a summary measure of the unconditional alphas using the distance measure of Hansen and Jagannathan (1997). The measure is the sample version of $HJD = E(mR - \underline{1})'[E(RR')^{-1}]E(mR - \underline{1})$, where R is the vector of the seven test assets' gross returns. HJD is a weighted sum of squared alphas, weighted by the second moments of the returns. As Hansen and Jagannathan explain, the measure is useful for comparisons across models, because the weights attached to the alphas do not vary across the models to be compared.[23]

The Hansen–Jagannathan distances confirm many of the previous comparisons. The models without the time-averaged terms produce larger pricing errors in all but one experiment. The differences are greater in the shorter subsample. The one-factor models perform the worst in both samples, although according to the HJD the two-factor and three-factor affine models swap positions in the performance rankings, depending on the sample period. The two-factor Brennan and Schwartz model performs relatively well in the full sample, but the three-factor affine model beats it in the shorter subsample. The HJDs are uniformly larger in the subsample period, reflecting partly the larger estimates of alpha but also the "larger" sample value of $E(RR')^{-1}$ in that period.[24]

We conclude from this section that the term structure models can explain large fractions of the conditional mean returns on benchmark test assets. The time-averaged factors help the models' performance, both in statistical and economic terms.

6. Government Bond Fund Performance

We estimate the SDF models for the government bond funds, concentrating on the three-factor affine model. Hansen's J-statistic produces p-values larger than 30% for the equal-weighted portfolio of all the funds,

[23] In contrast, Hansen's (1982) J statistic and the classical quadratic form tests for vectors of alphas [e.g., Jobson and Korkie (1982)] use the inverse of the covariance matrix of alpha as the weighting matrix. Thus, a "noisy" alpha can produce a small distance measure even when the economic magnitudes of the pricing errors are large. See Hansen and Jagannathan (1997) for additional interpretations of the measure.

[24] The maximum eigenvalue of the sample estimate of $[E(RR')^{-1}]$ is 68% larger in the subsample than in the full sample period.

the results for which are summarized in Table 5 for the 1986–2000 period. We also examine, but do not report in the table, the results for each high– and low–characteristic-based group of funds starting in 1989. The p-values are all between 7 and 8.5%. All of the models continue to display at least one factor coefficient with a t-ratio larger than two, and the discrete change in the long rate remains an important factor. Thus, estimating the models for mutual funds in many respects produces results similar to estimating the models on the test assets.

The first two columns of Table 5 report the conditional mean returns before risk adjustment. On average, and in most of the term structure states, the funds return less than the Lehman index. The differences in the conditional mean returns of the funds across the states are economically significant. The mean return for the portfolio of all funds in high spot rate states exceeds that in low spot rate states by 43 bp per month. The average difference for the characteristics-grouped portfolios is 51 bp. High term slopes predict relatively high future returns. The differences between the high and low slope states are 30 bp for all funds and they average 24 bp across the characteristics groups. The convexity states are the least important predictors.

Table 5
Conditional returns and alphas of government bond funds

State variable	r_{hi}	r_{lo}	α_{hi}	α_{lo}	Excess α_{hi}	Excess α_{lo}	α_u
Panel A: Three-factor affine model							
Short rate	1.11	0.68	–0.06	–0.06	–0.08	–0.07	
	0.22	0.17	0.03	0.05	0.02	0.06	
Slope	0.89	0.59	–0.06	–0.07	–0.07	–0.06	
	0.24	0.14	0.03	0.02	0.04	0.02	
Convexity	0.53	0.77	–0.09	–0.03	–0.11	–0.03	
	0.16	0.20	0.05	0.05	0.05	0.05	
Unconditional							–0.08
							0.02
Panel B: Discrete three-factor model							
Short rate	1.11	0.68	–0.06	–0.02	–0.10	–0.02	
	0.22	0.17	0.03	0.05	0.02	0.05	
Slope	0.89	0.59	–0.04	–0.07	–0.04	–0.07	
	0.24	0.14	0.03	0.02	0.04	0.02	
Convexity	0.53	0.77	–0.07	–0.02	–0.08	–0.00	
	0.16	0.20	0.03	0.05	0.04	0.05	
Unconditional							–0.07
							0.02

Generalized method of moments (GMM) estimation of SDF models. The instruments are the conditioning dummy variables for the level, slope, and convexity of the term structure. The benchmark returns are a four-month Treasury bill, a three-year zero coupon bond, and a 20-year Treasury bond. Funds are summarized through an equally weighted portfolio of all the government bond funds. Conditional mean fund returns are in the columns labelled r_{hi} and r_{lo}, abnormal returns conditional on the high and low states are denoted by α_{hi} and α_{lo}, and α_u is the unconditional alpha. Standard errors of the conditional means and alphas are shown below the point estimates. The excess alpha is the alpha for the return difference between the fund portfolio and the Lehman Government index returns. The sample period is January of 1986 through December, 2000 (180 observations). The units are percent per month.

The conditional mean returns vary much less across the characteristics groups. In no case, do we find a t-ratio larger than two, for the hypothesis that a high-characteristic differs from a low-characteristic conditional mean return. In only three cases, of 108 that we examine, do the conditional means of the characteristics groups differ by more than 20 bp. To save space, we do not report results for all the characteristic groups in the table.

Table 5 summarizes that the unconditional alpha for the all-fund portfolio is –8 bp per month, with a t-ratio of four.[25] The conditional alphas given the term structure states range from –3 to –9 bp. Four of the six have t-ratios larger than two. The alphas are economically small but estimated with high precision, compared with equity style funds where alphas are notoriously imprecise. This reflects the relatively low volatility of the funds' returns.[26] For the characteristics-based fund groups almost all of the conditional and unconditional alphas are negative.[27] The average absolute conditional alpha is 6.6 bp per month. Alphas measured in excess of the Lehman Government index leave similar impressions as the raw return alphas, both for the portfolio of all funds shown in the table and for the characteristics-grouped funds.[28] Performance is negative but seems economically small.

Panel B of Table 5 presents alphas using a discrete three-factor model, where the time-averaged factors are excluded. While the results for some of the fund groups differ, the overall impression is similar to that obtained from the affine model. The conditional alphas vary from –2 to –7 bp across the states, and half of them sport t-ratios larger than two.

Blake, Elton, and Gruber (1993) and Elton, Gruber and Blake (1995) argue that expenses are the main determinant of fixed-income fund after-cost performance. They do not conduct a conditional analysis. We therefore explore the role of fund costs in conditional performance (at the fund-group level). Following Bogle (1994), the total cost measure combines expense ratios and trading costs. We take the average monthly expense ratio for each group and add 1/12 of the annual turnover multiplied by a round-trip trading cost of 4/32 of a bp. The cost figures

[25] The conditional SDF alpha given Z is $(Z) = E(mR\text{-}1|Z)$ and the unconditional alpha is $\alpha_u = E(mR–1)$, so $E[(Z)] = \alpha_u$. In this respect, SDF alphas differ from beta pricing model alphas. The conditional alpha of a beta pricing model is the SDF alpha divided by the risk-free rate. When the risk-free rate is time varying, Jensen's inequality implies that the expected conditional alpha is not the unconditional alpha. Ferson and Schadt (1996) show that average conditional alphas and unconditional alphas can differ empirically for equity style funds.

[26] Future research should examine the finite sample performance of SDF term structure models as the reported standard errors may be inaccurate. See Farnsworth et al. (2002) for finite sample evidence in the context of equity funds and equity-style SDF models.

[27] The only exception is the low–lagged-return fund group where the alphas are positive, 2-5 bp depending on the state.

[28] We check the robustness of the fund performance results using an alternative style-matched benchmark, formed with an approach similar to Sharpe (1992) and described in the Appendix. The results are very close to those using the Lehman index return as the benchmark.

come from the end of 1988, the first year when all of the fund-group characteristics become available, and they range from 8.7 bp to 11.3 bp per month across the fund-characteristic groups.[29] The average conditional alpha is –7 bp and about two standard errors from zero. Thus, we could not reject the hypothesis that fund costs explain the negative conditional performance on average. The conditional performance before costs would be slightly positive but not significant.

We examine the cross-sectional relation between the fund-group alphas and the cost measures, running regressions of the alphas on the total costs. Using unconditional alphas as the dependent variable the slope coefficient is –0.81, but the R-squared of the regression is only 4.9%. Averaging across twelve regressions, one for each version of the conditional alphas, the slope coefficient is –0.24 and the R-squared is 14.5%. Thus, there is a weak negative cross-sectional relation between alphas and total costs. These results are similar to those reported by Blake, Elton, and Gruber (1993) and Elton, Gruber, and Blake (1995), who used unconditional alphas in linear factor models. They conclude that on average, investors are better off selecting low cost funds.

7. Concluding Remarks

We show how to evaluate the performance of mutual funds using SDFs derived from continuous-time term structure models. The SDFs are simple exponential functions of term structure factors. The solutions imply additional empirical factors, formed as time averages of the underlying state variables in the model. We find that these new factors are statistically and economically significant in factor model regressions, and they help reduce the pricing errors of the SDF models. Our approach addresses a bias in performance measurement described by Goetzmann, Ingersoll, and Ivkovic (2000) and Ferson and Khang (2002), that arises when fund managers may trade dynamically or hold positions in derivative securities.

We illustrate our approach on a sample of US government bond funds during 1986–2000, providing the first conditional performance evaluation for US fixed-income mutual funds. We find that high spot rates, high term structure slopes, and low term convexity predict higher conditional expected returns. Fund returns vary more across the term structure states than across fund groups formed on the basis of size, age, expenses, and other common characteristics. After risk adjustment, the conditional

[29] Our total cost measure does not allow trading costs to be state dependent. Accommodating state-dependent costs would require more accurate and higher frequency data, and we leave that for future research.

alphas average –7 bp per month and are precisely estimated, with a typical *t*-ratio of two. An estimate of total costs is between 8 and 12 bp per month, so the results are consistent with the view that before-cost performance is neutral.

Our article suggests many opportunities for future research. We illustrate the approach using simple term structure models. The term structure literature has identified several shortcomings of these simple models and has suggested models that may address these shortcomings. It should be interesting to apply our approach with more refined term structure models.

We limited our fund sample to government bond funds. However, a broader sample of fixed-income funds is likely to offer more cross-sectional variation in performance. Fixed-income funds may hold corporate bonds subject to default risk, mortgage-related securities and other positions exposed to risk factors that go beyond pure term structure models. The approach of this article can be extended beyond pure term structure models, to study additional factors.[30] International fixed-income funds and hedge funds present other important settings where our approach should be applied and refined.

Statistical issues, such as the sampling properties and power of alternative models, should also be addressed in future research. This article has just started plowing what we think should be the fertile field of conditional performance measurement for fixed-income funds.

Appendix A

A.1 Interest rate data

We use daily interest rate data to construct the monthly empirical factors that involve time averages, and we use their values on the last day of the month for the discrete state variable measures. Our daily short rate series is the three-month Treasury bill rate from the Federal Reserve Database (FRED), H.15 release. Chapman, Long, and Pearson (1999) find that the errors in using this rate to approximate an instantaneous interest rate are economically insignificant in affine term structure models. The remaining daily interest rate data, from the H.15 release, are the fixed maturity yields for one-, three-, seven- and ten-year bonds. We use the ten-year yield as the long rate, and we construct our convexity measure from the remaining yields as $(y_7 + 2y_1)/3 - y_3$, where y_j is the *j*-year fixed-maturity yield.

A.2 Bond return data

We take the returns for bonds with four months, and 1, 3, 4, 7, 10, and 20 years to maturity from the CRSP term structure files.[31] We also include four portfolios from the CRSP Fama-

[30] Ferson, Henry, and Kisgen (2004) and Chen, Ferson, and Peters (2005) are two working papers along these lines.

[31] The four-month return is from the CRSP Fama file of original issue six-month bill returns, based on the ask yields. The one-, ten-, and twenty-year returns are from the CRSP mcti files. The three- and four-year returns are based on the monthly CRSP Fama-Bliss price files. Here, we compute the one-month return by "buying" at the month-end-price of a *j*-year bond and selling one month later at a price which we interpolate exponentially from the next month's prices of *j*- and *j*–1 year bonds.

Bliss files: Bonds with 24 months or less to maturity (denoted le24), 60 months or less (le60), 120 months or less (le120), and at least ten years to maturity (gt120).

A.3 Benchmark dynamic strategies

We construct dynamic strategy portfolios for diagnostic purposes. We first approximate daily returns from the daily fixed-maturity yield series. If we were willing to assume that a particular term structure model holds exactly, we could invert the cross-section of yields each day for the state variables, solve for the term structure of discount bond prices as functions of the state variables, and figure our dynamic strategy returns based on the daily prices. However, we do not wish to impose the assumption that any particular term structure model holds exactly, so we adopt a simple approximate scheme.[32]

Assume that each fixed-maturity yield represents a newly issued coupon bond trading at par on a given day; thus, the yield is the coupon rate. After one day, we use the new yield to compute a selling price as the present value of the semi-annual coupons and maturity payment, discounted by the new yield, and we add one day's accrued interest.[33] The result is a series of approximate daily returns for bonds of three months, 1, 3, 7, and 10 years to maturity.

We model four dynamic strategies using the constructed daily returns. The first, denoted "randr," picks one of the five maturities at random each day and holds that bond for one day. In the remaining three strategies, interim trading is "optimal" according to some simple criterion.[34] The strategy denoted "buyhiY" purchases each day the bond with the highest yield. The strategy buyhiR picks each day the bond with the highest expected return, naively estimated by regressing the last 60 days' returns on the lagged short rate, slope, and convexity. The strategy denoted "mv" is based on daily mean-variance optimization. Using the same regression for the conditional means, the sample covariance matrix of the residuals is the conditional covariance matrix for that day. A notional risk-free rate is set equal to the estimated global minimum variance return on that day, less 0.1% per month. The notional risk-free rate is used only to determine a point on the mean variance boundary as we normalize the strategy to be fully invested in the five bonds. We constrain the weights to be non-negative, thus ruling out short sales.[35]

A.4 Fund benchmarks

We include the Lehman Government index as a benchmark return. We also construct an alternative style-matched benchmark using an approach similar to Sharpe (1992). The problem is to combine asset class returns, denoted by R_i, using a set of portfolio weights, denoted by w_i, so as to minimize the "tracking error" between the return of the fund or fund group, R_p, and the style-matched benchmark portfolio, $\Sigma_i w_i R_i$. The portfolio weights are

[32] Our thanks to Richard Roll for suggesting this approach.

[33] For the three-month Treasury return, we assume a bill is purchased at the end of each day at the price $100 - y_t * 91/360$, where y_t is the quoted discount yield. After one day, the bond is sold at the price $100 - y_{t+1} * (91-d)/360$, where d is the number of calendar days the bond is held. Over a weekend or on a day in which a fixed maturity yield is missing, the bond is assumed to be held until the next non-missing yield is observed.

[34] For each strategy that involves any "optimization", we impose a one-day lag between the calculation of the strategy and its implementation, so that the strategies are not allowed to trade at the same recorded price that enters the calculations.

[35] We experimented with two alternative versions of the daily mean variance strategy. In one, we set the notional risk-free rate equal to zero, which results in a tangency point close to the global minimum variance portfolio and a strategy more concentrated in the shorter maturities. In the second, we restricted the conditional covariance matrix to be diagonal as suggested by Elton and Gruber (1991). Both alternatives produced similar conclusions for the model diagnostics.

required to sum to 1.0 and must be non-negative, which rules out short positions. The problem is

$$\begin{aligned} &\mathrm{Min}_{wi}\mathrm{Var}(R_p - \sum_i w_i R_i), \\ &\text{subject to}: \sum_i w_i = 1, \quad w_i \geq 0 \text{ for all } i, \end{aligned} \tag{A1}$$

where Var[.] denotes the variance.

We solve the problem numerically, where the asset class returns are the bonds {r4mo, r3y, r7y, gt120 and rLB} and the fund is the all-funds portfolio, over 1986–2000. A regression of the all-fund portfolio return on these five bond returns has an adjusted R-square of 94.7%. The solution delivers the weight vector (0, 0.27, 0, 0.04, 0.69). Summary statistics for the benchmark portfolio formed using these weights are shown in Table 1, under the heading "Sharpe."

A.5 Additional experiments

One question that arises in the interpretation of the Table 2 regressions is the importance of time averaging versus the levels of the state variables. The levels of the state variables would capture the expected component of returns in a model where time aggregation is not an issue. The continuous-time theory specifies that time-averaged terms are the right way to incorporate the paths of the state variables. However, the end-of-month levels and the time averages over the month are highly correlated, so a horse race between the two would be hard to interpret. The one-month bill return is a convenient control for a common levels effect of the state variables as the log of the bill price is linear in the state variables in an affine model. We ran the Table 2 regressions, substituting the returns in excess of the one-month Treasury bill for the raw returns. The overall R-squares are similar. The time-averaged term A^r becomes statistically significant in the gt120 regressions where it was not before. The same term becomes insignificant in the buyhiY and buyhiR strategy regressions, but the time-averaged terms remain jointly significant in these regressions. The A^r term remains strongly significant in the regressions for the Roll3 strategy. These results show that the explanatory ability of the time-averaged factors is not simply a common levels effect.

We ran the regressions in Table 2 for returns and factors measured over non-overlapping quarterly and annual frequencies as well to see whether the time aggregation terms are more or less important at the longer holding periods. We find that the overall patterns in the adjusted R-squares are similar. The time-averaged terms have greater explanatory power for the buyhiY and buyhiR strategies at the longer horizons.

When we examine the SDF models' performance in pricing benchmark returns we find that the Brennan and Schwartz model is over parameterized when all of its time-aggregation terms are used.[36] We use the factor model regressions to explore the effects of dropping one of the D-terms from the list of Brennan and Schwartz variables. It is possible to drop one of the variables without degrading the explanatory power, but the regressions give mixed signals on which one to drop. Regressions for the shorter term bonds suggest that D^{ℓ} or $D^{r\ell}$ can be dropped while longer term bond regressions suggest dropping either D^{r} or $D^{r\ell}$. We drop the $D^{r\ell}$ term in the empirical SDF models.

References

Ait-Sahalia, Y., 1996, "Testing Continuous Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385–426.

Balduzzi, P., and Y. H. Eom, 2002, "Nonlinearities in U.S. Treasury Rates: A Semi-Nonparametric Approach," Working paper, Boston College, Chestnut Hill, MA.

[36] The gradient matrix becomes rank deficient.

Bekaert, G., R. J. Hodrick, and D. Marshall, 1997, "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Financial Economics*, 44, 309–348.

Blake, C. R., E. J. Elton, and M. J. Gruber, 1993, "The Performance of Bond Mutual Funds," *Journal of Business*, 66, 371–403.

Bogle, J. C., 1994, *Bogle on Mutual Funds*, Irwin Publishers, New York.

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger, 1989, "Empirical Tests of the Consumption Oriented CAPM," *Journal of Finance*, 44, 231–262.

Brennan, M. J., and E. Schwartz, 1979, "A Continuous Time Approach to the Pricing of Bonds," *Journal of Banking and Finance*, 3, 133–155.

Campbell, J. Y., 1996, "Understanding Risk and Return," *Journal of Political Economy*, 104(2), 298–435.

Campbell, J. Y., Y. L. Chan, and L. Viciera, 2003, "A Multivariate Model of Strategic Asset Allocation," *Journal of Financial Economics*, 67, 41–80.

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., J. B. Long, and N. D. Pearson, 1999, "Using Proxies for the Short Rate: When are Three Months Like an Instant?" *Review of Financial Studies*, 12, 763–806.

Chapman, D. A., and N. D. Pearson, 2000, "Is the Short Rate Drift Actually Nonlinear?," *Journal of Finance*, 55, 355–388.

Chen, Z., and P. J. Knez, 1996, "Portfolio Performance Measurement: Theory and Applications," *Review of Financial Studies*, 9, 511–556.

Chen, Y., W. Ferson, and H. Peters, 2005, "The Timing Ability of Fixed-Income Mutual Funds," Working paper, Boston College, Chestnut Hill, MA.

Christopherson, J. A., W. Ferson, and D. A. Glassman, 1998, "Conditioning Manager Alpha on Economic Information: Another Look at the Persistence of Performance," *Review of Financial Studies*, 11, 111–142.

Cochrane, J., 2001, *Asset Pricing*, Princeton University Press, Princeton, NJ.

Cornell, B., and K. Green, 1991, "The Investment Performance of Low-Grade Bond Funds," *Journal of Finance*, 46, 29–48.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross, 1985a, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–346.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross, 1985b, "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53, 363–384.

Dahlquist, M., S. Engstrom, and P. Soderlind, 2000, "Performance Characteristics of Swedish Mutual Funds," *Journal of Financial and Quantitative Analysis*, 35, 409–423.

Dai, Q., and K. Singleton, 2002, "Expectations Puzzles, Time-Varying Risk Premia, and Dynamic Models of the Term Structure," *Journal of Financial Economics*, 63, 415–441.

Duarte, J., 2004, "Evaluating Alternative Risk Preferences in Affine Term Structure Models," *Review of Financial Studies*, 17, 379–404.

Duffee, G., 2002, "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 67, 405–443.

Dybvig, P. H., J. E. Ingersoll, and S. A. Ross, 1996, "Long Forward and Zero-Coupon Rates can Never Fall," *Journal of Business*, 69, 1–25.

Elton, E. J., and M. J. Gruber, 1991, *Modern Portfolio Theory and Investment Analysis*, John Wiley and Sons, New York.

Elton, E. J., M. J. Gruber, and C. R. Blake, 1995, "Fundamental Economic Variables, Expected Returns and Bond Fund Performance," *Journal of Finance*, 50, 1229–1256.

Elton, E. J., M. J. Gruber, and C. R. Blake, 2001, "A First Look at the Accuracy of the CRSP Mutual Fund Data Base and a Comparison of the CRSP and Morningstar Databases," *Journal of Finance*, 56(6), 2415–2430.

Evans, R. B., 2004, "Does Alpha Really Matter? Evidence from Fund Incubation, Termination and Manager Change," Working paper, Boston College, Chestnut Hill, MA.

Farnsworth, H. K., 1997, "Evaluating Stochastic Discount Factors from Term Structure Models," Unpublished Ph.D. Dissertation, University of Washington.

Farnsworth, H. K., W. Ferson, D. Jackson, and S. Todd, 2002, "Performance Evaluation with Stochastic Discount Factors," *Journal of Business*, 75, 473–504.

Ferson, W. E., 1995, "Theory and Empirical Testing of Asset Pricing Models," in R. A. Jarrow, V. Maksimovic, and W. T. Ziemba (eds.), *Handbooks in Operations Research and Management Science*, Elsevier, Amsterdam, 145–200.

Ferson, W. E., 2003, "Tests of Multifactor Pricing Models, Volatility Bounds and Portfolio Performance," in G. M. Constantinides, M. Harris, and R. M. Stulz (eds.), *Handbook of the Economics of Finance*, Elsevier Science Publishers, Amsterdam, North Holland.

Ferson, W., and C. R. Harvey, 1991, "The Variation of Economic Risk Premiums," *Journal of Political Economy*, 99, 385–415.

Ferson, W., and C. R. Harvey, 1994, "Sources of Risk and Expected Returns in Global Equity Markets," *Journal of Banking and Finance*, 18, 775–803.

Ferson, W. E., S. Sarkissian, and T. Simin, 1999, "The Alpha Factor Asset Pricing Model: A Parable," *Journal of Financial Markets*, 2, 49–68.

Ferson, W. E., S. Sarkissian, and T. Simin, 2003, "Spurious Regressions in Financial Economics?," *Journal of Finance*, 58, 1393–1414.

Ferson, W., and K. Khang, 2002, "Conditional Performance Measurement Using Portfolio Weights: Evidence for Pension Funds," *Journal of Financial Economics*, 65, 249–282.

Ferson, W., and M. Qian, 2004, "Conditional Performance Evaluation Revisited," Monograph, The Research Foundation of the CFA Institute, Charlottesville, VA.

Ferson, W., and R. Schadt, 1996, "Measuring Fund Strategy and Performance in Changing Economic Conditions," *Journal of Finance*, 51, 425–462.

Ferson, W., and S. R. Foerster, 1994, "Finite Sample Properties of the Generalized Methods of Moments Tests of Conditional Asset Pricing Models," *Journal of Financial Economics*, 36, 29–56.

Ferson, W., T. Henry, and D. Kisgen, 2004, "Fixed Income Fund Performance Across Economic States," Working paper, Boston College and the University of Georgia, Chestnut Hill, MA and Athens, GA.

Goetzmann, W., J. Ingersoll, and Z. Ivkovic, 2000, "Monthly Measurement of Daily Timers," *Journal of Financial and Quantitative Analysis*, 35, 257–290.

Gourieroux, C., A. Montfort, and V. Polimenis, 2002, "Affine Term Structure Models," Working paper, University of California, Riverside.

Hansen, L. P., 1982, "Large Sample Properties of the Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hansen, L. P., and R. Jagannathan, 1997, "Assessing the Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52, 557–590.

Jensen, M. C., 1968, "The Performance of Mutual Funds in the period 1945–1964," *Journal of Finance*, 23, 389–416.

Jobson, J.D., and B. Korkie, 1982, "Potential Performance and Tests of Portfolio Efficiency," *Journal of Financial Economics*, 10, 433–466.

Jones, C. S., 2003, "Nonlinear Mean Reversion in the Short Interest Rate," *Review of Financial Studies*, 16, 793–843.

Kon, S. J., and C. Y. Polek, 1998, "Time-Varying Empirical Duration and Slope Effects for Mortgage-Backed Securities," *Journal of Fixed Income*, 8, 7–28.

Litterman, R., and J. Scheinkman, 1991, "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, 1, 54–61.

Lo, A. W., and A. C. MacKinlay, 1990, "Data Snooping in Tests of Financial Asset Pricing Models," *Review of Financial Studies*, 3, 431–467.

MacKinlay, A. C., 1995, "Multifactor Models do not Explain Deviations from the CAPM," *Journal of Financial Economics*, 38, 3–28.

Mehra, R., and E. Prescott, 1985, "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–162.

Roll, R., 1970, The Behavior of Interest Rates: An Application of the Efficient Market Model to U.S. Treasury Bills, Basic Books, Inc., New York.

Sharpe, W. F., 1964, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19, 425–442.

Sharpe, W. F., 1992, "Asset Allocation: Management Style and Performance Measurement," *Journal of Portfolio Management*, 18(2), 7–19.

Stambaugh, R. S., 1999, "Predictive Regressions," *Journal of Financial Economics*, 54, 315–421.

Stanton, R., 1997, "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance*, 52, 1973–2002.

Stephens, M. A., 1974, "EDF Statistics for Goodness of Fit and Some Comparisons," *Journal of the American Statistical Association*, 69, 730–737.

Sun, G., 1997, "The Test Function Method for Estimating Continuous-Time Models of short Term Interest Rate," Working paper, University of Minnesota.

Vasicek, O. A., 1977, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.

Working, H., 1960, "A Note on the Correlation of First Differences of Averages of a Random Chain," *Econometrica*, 28, 916–918.